ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

Özge KURAN

DOĞRUSAL KARMA VE HİYERARŞİK MODELLERDE REZİDÜ VE ETKİ ANALİZİ

İSTATİSTİK ANABİLİM DALI

ADANA, 2013

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DOĞRUSAL KARMA VE HİYERARŞİK MODELLERDE REZİDÜ VE ETKİ ANALİZİ

Özge KURAN

YÜKSEK LİSANS TEZİ

İSTATİSTİK ANABİLİM DALI

Bu Tez 12/07/2013 Tarihinde Aşağıdaki Jüri Üyeleri Tarafından Oybirliği/Oyçokluğu ile Kabul Edilmiştir.

.....	.....	.....
Doç. Dr. Mahmude Revan ÖZKALE	Prof. Dr. Sadullah SAKALLIOĞLU	Yrd. Doç. Dr. Gülsen KIRAL
DANIŞMAN	ÜYE	ÜYE

Bu Tez Enstitümüz İstatistik Anabilim Dalında hazırlanmıştır.
Kod No:

Prof. Dr. Mustafa GÖK
Enstitü Müdürü

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

ÖZ

YÜKSEK LİSANS TEZİ

DOĞRUSAL KARMA VE HİYERARŞİK MODELLERDE REZİDÜ VE ETKİ ANALİZİ

Özge KURAN

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI

Danışman : Doç. Dr. Mahmude Revan ÖZKALE
Yıl: 2013, Sayfa: 157
Jüri : Doç. Dr. Mahmude Revan ÖZKALE
: Prof. Dr. Sadullah SAKALLIOĞLU
: Yrd. Doç. Dr. Gülsen KIRAL

İstatistiksel modeller, verinin tanımlanması, veri yapısını en iyi temsil eden matematiksel formun belirlenmesi ve değişkenler arasındaki ilişkinin yaklaşık olarak tanımlanması amacıyla kullanılan modellerdir. En yaygın kullanılan istatistiksel model sıradan doğrusal (lineer) regresyon modeli olmasına rağmen bazen sabit etkilere ilave olarak rastgele etkileri de içeren modellere uyan veri yapıları ile de karşılaşmak mümkündür. Bu durumda, lineer karma modeller ile verinin iç içe geçmiş bir yapıya sahip olduğu lineer karma model tiplerinden biri olan hiyerarşik lineer modeller ortaya çıkmaktadır.

Bu çalışmada ilk olarak, lineer karma ve hiyerarşik modellerde rezidü ve etki analizinin daha iyi bir şekilde anlaşılabilmesi amacıyla lineer regresyon modellerindeki rezidü ve etki analizi üzerinde durulmuştur. Daha sonra, sırasıyla, lineer karma model tipleri, lineer karma modellerin parametre tahminleri ve güncelleme formülleri, tahmin edilen modelin geçerliliğinin araştırılması amacıyla rezidülerin incelenmesi, lineer karma modellerdeki etkili birimlerin ve gözlemlerin belirlenmesinde kullanışlı olabileceği düşünülen genelleştirilmiş leverage matrisi, lineer karma modellerdeki etkili noktaların ortaya çıkarılması için geliştirilen ölçütler ve hiyerarşik lineer modellerdeki rezidü analizi üzerinde durulmuştur. Çalışma, R ve Matlab programlarının kullanılmasıyla yapılan örneklerle desteklenmiştir.

Anahtar Kelimeler: Sabit Etkiler, Rastgele Etkiler, Sapan Değerler, Genelleştirilmiş Leverage, Etkili Gözlemler

ABSTRACT

MSc THESIS

RESIDUAL AND INFLUENCE ANALYSIS IN LINEAR MIXED AND HIERARCHICAL MODELS

Özge KURAN

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATUREL AND APPLIED SCIENCES
DEPARTMENT OF STATISTICS

Supervisor : Assoc. Prof. Dr. Mahmude Revan ÖZKALE
Year: 2013, Pages: 157
Jury : Assoc. Prof. Dr. Mahmude Revan ÖZKALE
: Prof. Dr. Sadullah SAKALLIOĞLU
: Asst. Prof. Dr. Gülsen KIRAL

Statistical models are used for the identification of the data, the determination of the mathematical form that best represents the data structure and the definition approximately the relationship between the variables. Although the most widely-used statistical model is the ordinary linear regression model, it is sometimes possible to encounter with the data structures that accord with the models that include both fixed and random effects. In this case, linear mixed models and hierarchical linear models, which is one of the linear mixed model types that has nested data structure, are arised.

In this study firstly, residual and influence analysis in linear regression models is emphasized in order to understand the residual and influence analysis in linear mixed and hierarchical models. Later, linear mixed model types, parameter estimates and update formulas of linear mixed models, examination of the residuals for searching estimated model validity, generalized leverage matrix that is considered to be useful for the identification of influential units and observations in linear mixed models, measures developed for the identification of influential observations in linear mixed models and last, residual analysis in hierarchical linear models are emphasized, respectively. The study is supported with examples done using R and Matlab programs.

Key Words: Fixed Effects, Random Effects, Outliers, Generalized Leverage, Influence Observations

TEŞEKKÜR

Tez çalışmam boyunca değerli destek ve yardımlarını esirgemeyerek her zaman yanımda olan danışman hocam Doç. Dr. M. Revan ÖZKALE' ye teşekkür ederim.

Ayrıca, eğitim hayatım boyunca aldığım her kararın arkasında olarak beni destekleyen ve yanımda olan sevgili aileme de teşekkür ederim.

İÇİNDEKİLER **SAYFA**

ÖZ I
ABSTRACT II
TEŞEKKÜR III
İÇİNDEKİLER ..... …IV
ÇİZELGELER DİZİNİ VIII
ŞEKİLLER DİZİNİ X
SİMGELER VE KISALTMALAR XII
1. GİRİŞ 1
2. LİNEER REGRESYON MODELLERİNDE REZİDÜ VE ETKİ ANALİZİ 5
2.1. Giriş 5
2.2. X-yönündeki Sapan Değerleri Belirleyen Ölçüler 10
2.3. Yanıt Değişkeni y-yönündeki Sapan Değerleri Belirleyen Ölçüler 11
2.3.1. Standartlaştırılmış Rezidüler 12
2.3.2. Studentized Rezidüler 12
2.3.3. PRESS Rezidüler 14
2.3.4. R-Student 15
2.4. X ve y-yönündeki Sapan Değerleri Belirleyen Ölçüler 15
2.5. Etkili Gözlemleri Belirleyici Ölçüler 17
2.5.1. Etki Eğrisine Dayalı Ölçüler 17
2.5.1.1. Cook Uzaklığı 19
2.5.1.2. Welsch-Kuh Uzaklığı (DFFITS_i) ve $\mathrm{DFBETAS}_{j,i}$ 20
2.5.1.3. Welsch Uzaklığı 21
2.5.2. Güven Elipsoidinin Hacmine Dayalı Ölçüler 22
2.5.2.1. Andrew-Pregibon İstatistiği 22
2.5.2.2. Varyans Oranı (COVRATIO) 23
2.5.2.3. Cook Weisberg İstatistiği 24
2.5.3. Likelihood Fonksiyonuna Dayalı Ölçüler 26
2.6. X ve y-yönündeki Sapan Değerlerin Belirlenmesi için Grafiksel Metotlar 26

3. LİNEER KARMA MODELLERDE REZİDÜ VE ETKİ ANALİZİ 29
3.1. Giriş.... 29
3.1.1. Lineer Karma Model Tipleri 30
3.1.1.1. Gaussian Lineer Karma Modeller 30
3.1.2. Lineer Karma Modellerde Parametre Tahminleri 34
3.1.2.1. H Biliniyorken β ve σ_e^2 nin ML Tahminleri ve u un ML Ön Tahmini 35
3.1.2.2. H Bilinmiyorken β ve σ_e^2 nin ML Tahminleri ve u nun ML Ön Tahmini 37
3.1.3. Lineer Karma Modellerde Güncelleme Formülleri 39
3.2. Lineer Karma Modellerde Rezidü Analizi 47
3.2.1. Lineer Karma Modellerde Rezidü Tipleri 47
3.2.1.1. Marjinal Rezidüler 48
3.2.1.2. EBLUP 50
3.2.1.3. Koşullu Rezidüler 51
3.2.2. Lineer Karma Modellerde Sapan Değerlerin Belirlenmesi 62
3.2.2.1. Sapan Değerlerin Belirlenmesi İçin Bir Test 62
3.2.2.2. Sapan Değerlerin Belirlenmesi İçin Grafiksel Metotlar 64
3.3. Lineer Karma Modellerde Leverage Analizi.... 65
3.3.1. H_2 ile L_2 nin Karşılaştırılması 69
3.4. Lineer Karma Modellerde Etki Analizi.... 69
3.4.1. Varyans Bileşenleri Oranları Üzerinde Etki Analizi 70
3.4.1.1. Cook Uzaklığının Benzeri 70
3.4.1.2. Bilgi Oranının Uzaklığının Benzeri 70
3.4.2. Sabit Etkili Parametre Tahminleri Üzerinde Etki Analizi 71
3.4.2.1. Cook Uzaklığının Benzeri 71
3.4.2.2. Varyans Oranının Benzeri 72
3.4.2.3. Cook-Weisberg İstatistiğinin Benzeri 73
3.4.2.4. Andrews-Pregibon İstatistiğinin Benzeri 75
3.4.3. Rastgele Etkili Parametrelerin Ön Tahmini Üzerinde Etki Analizi. 77
3.4.4. Likelihood Fonksiyonu Üzerindeki Etki Analizi 77

4. HİYERARŞİK LİNEER MODELLERDE REZİDÜ ANALİZİ..... 81
4.1. Giriş.... 81
4.2. Hiyerarşik Lineer Modellerde Parametre Tahminleri.... 86
4.3. Hiyerarşik Lineer Modellerde Rezidü Analizi..... 89
4.3.1. Gözlem-Düzeyli Rezidüler..... 89
4.3.1.1. EB Rezidüler 90
4.3.1.2. LS Rezidüler..... 91
4.3.1.3. Karma Rezidüler..... 92
4.3.2. Birim-Düzeyli Rezidüler (Rastgele Etkiler)..... 94
4.4. Çift Rezidü Grafiği ve Uyumun Değerlendirilmesi.... 95
5. UYGULAMA..... 97
5.1. Lineer Karma Modellerde Rezidü Analizi Örneği..... 97
5.2. Lineer Karma Modellerde Leverage Analizi Örneği.....106
5.3. Lineer Karma Modellerde Etki Analizi Örneği.... 110
6. SONUÇLAR VE ÖNERİLER 117
KAYNAKLAR.....119
ÖZGEÇMİŞ…..125
EKLER…..126

ÇİZELGELER DİZİNİ **SAYFA**

Çizelge 3.1. Tanılama önerileri için rezidülerin kullanımı (Nobre, 2007) 62
Çizelge 5.1. Bakteri plak indeksleri (Nobre ve Singer, 2007) 98
Çizelge 5.2. REML yardımıyla uydurulan lineer karma model sonuçları 99
Çizelge 5.3. Rastgele etkiler için uydurulan lineer karma model sonuçları 99
Çizelge 5.4. Sabit etkiler için uydurulan lineer karma model sonuçları 99
Çizelge 5.5. Sabit etkilerin korelasyonu 99
Çizelge 5.6. Uydurulan lineer karma model için ANOVA tablosu 99
Çizelge 5.7. 12 ve 29 numaralı denekler modelde iken ve modelde değilken tahminler (± tahmin edilen standart hatalar) ve model parametrelerine ilişkin değişimler 103
Çizelge 5.8. 12 ve 29 numaralı denekler modelde iken ve modelde değilken AIC ve BIC değerleri 104
Çizelge 5.9. Bakteri plak indeksleri (Nobre ve Singer, 2011) 107
Çizelge 5.10. REML yardımıyla uydurulan lineer karma model sonuçları 107
Çizelge 5.11. Rastgele etkiler için uydurulan lineer karma model sonuçları 107
Çizelge 5.12. Sabit etkiler için uydurulan lineer karma model sonuçları 107
Çizelge 5.13. Uydurulan lineer karma model için ANOVA tablosu 108
Çizelge 5.14. İmalatçıların yararlandığı üç filtre tipinin kullanılması ile iki aerosolden elde edilen yüzdelik penetrasyonlar (Beckman ve ark., 1987) 112
Çizelge 10.1. X matrisinin elde edilmesi 141
Çizelge 10.2. Z matrisinin elde edilmesi 143
Çizelge 10.3. y vektörünün elde edilmesi 146
Çizelge 11.1. Gözlemler için R analiz sonuçları 149
Çizelge 11.2. Denekler için R analiz sonuçları 151
Çizelge 13.1. Gözlemler için R analiz sonuçları 155
Çizelge 13.2. Eşik noktaları 156
Çizelge 14.1. Gözlemler için Matlab analiz sonuçları 157

ŞEKİLLER DİZİNİ **SAYFA**

Şekil 1.1. İstatistiksel modelleme süreci (Box ve Jenkins, 1970) 1

Şekil 2.1. Sapan değer tipleri arasındaki farklılıkları gösteren bir örnek (Chatterjee ve Hadi, 1988)..... 8

Şekil 2.2. Sapan değer tipleri arasındaki farklılıkları gösteren diğer bir örnek (Chatterjee ve Hadi, 1988)..... 9

Şekil 2.3. Potansiyel-Rezidü Grafiği (Hadi, 1992)..... 27

Şekil 5.1. (5.1) modeli için marjinal rezidü grafiği 100

Şekil 5.2. (5.1) modeli için denekler-içi kovaryans matris yapısına ait rezidülerin grafiği..... 100

Şekil 5.3. (5.1) modeli için standartlaştırılmış koşullu rezidülerin grafiği 101

Şekil 5.4. (5.1) modeli için standartlaştırılmış en küçük bozulmuş rezidülerin 95% lik güven aralığının grafiği 102

Şekil 5.5. (5.1) modeli için EBLUP grafiği..... 103

Şekil 5.6. (5.2) modeli altında marjinal uydurulmuş değerlerin genelleştirilmiş leverage matrisi olan L_1 in grafiği..... 108

Şekil 5.7. (5.2) modeli altında rastgele etkilerin varlığıyla açıklanan birimler içi değişkenliğin parçası olan L_2 nin grafiği..... 109

Şekil 5.8. (5.2) modeli altında koşullu uydurulmuş değerlerin rastgele etki bileşeni için genelleştirilmiş leverage matris olan H_2 nin grafiği..... 109

Şekil 5.9. (5.3) modeli altında $CD(\beta)$ nın indeks grafiği..... 113

Şekil 5.10. (5.3) modeli altında $|VR - 1|$ nın indeks grafiği 113

Şekil 5.11. (5.3) modeli altında $|CW|$ nun indeks grafiği..... 114

Şekil 5.12. (5.3) modeli altında AP nin indeks grafiği 115

Şekil 5.13. (5.3) modeli altında $CD(u)$ nun indeks grafiği..... 115

Şekil 5.14. (5.3) modeli altında LD nin indeks grafiği..... 116

SİMGELER VE KISALTMALAR

EKK : En Küçük Kareler (Ordinary Least Squares-OLS)

BLUE : En İyi Lineer Yansız Tahmin Edici (Best Linear Unbiased Estimator)

BLUP : En İyi Lineer Yansız Ön Tahmin Edici (Best Linear Unbiased Predictor)

MS_{Res} : Rezidü Kareler Ortalaması (Residual Mean Square)

SSE : Hata Kareler Toplamı (Sum of Squared Error)

MLE : Maksimum Likelihood Tahmin (Maksimum Likelihood Estimation)

REML : Sınırlandırılmış Maksimum Likelihood (Restricted MLE)

PRESS : Ön Tahmin Hata Kareler Toplamı (Prediction Error Sum of Squares)

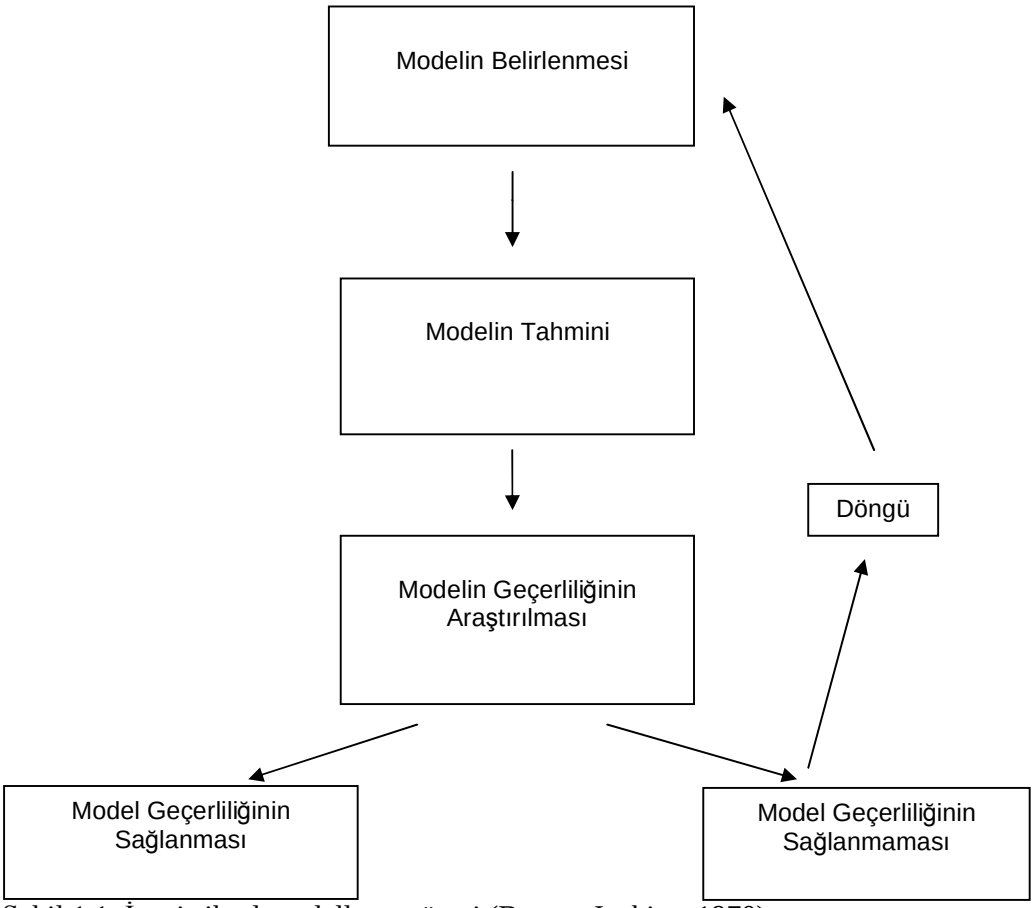
AIC : Akaike Bilgi Kriteri (Akaike Information Criterion)

BIC : Bayesian Bilgi Kriteri (Bayesian Information Criterion)

1. GİRİŞ

İstatistiksel modeller, verinin tanımlanması ve veri yapısının modellenmesi amacıyla oluşturulan modellerdir. Verinin modellenmesi, modern istatistiksel analizin ayrılmaz bir parçasını oluşturmaktadır. Bilgisayar paket programlarında meydana gelen gelişmeler sayesinde durumların ve olayların modellenme süreci hızlanmaktadır.

Box ve Jenkins (1970)'e göre, istatistiksel modelleme süreci üç adımdan oluşmaktadır. Birinci adım modelin belirlenmesi, ikinci adım modelin tahmin edilmesi ve üçüncü adım ise, modelin geçerliliğinin araştırılmasıdır. Modelleme sürecinin adımları Şekil 1.1 ile gösterilmektedir.



Şekil 1.1. İstatistiksel modelleme süreci (Box ve Jenkins, 1970)

Şekil 1.1 ile verilen istatistiksel modelleme sürecinin ilk adımı olan modelin belirlenmesi aşamasında, ilgilenilen veri kümesine uygun olabilecek modeller seçilir. Modelin tahmini aşamasında, seçilen modeldeki bilinmeyen parametrelerin veri kümesinden yararlanılarak elde edilmesi ve bulunan parametre değerlerinin modelde yerine konulması üzerinde durulur. Modelin geçerliliğinin araştırılması aşamasında ise, belirlenen ve tahmin edilen modelin veriye uygun olup olmadığı değerlendirilir. Tahmin edilen modelin veriye uygun olup olmadığı değerlendirilirken de rezidü ve etki analizinden yararlanılır. Bu analizler yardımı ile analiz sonuçlarını etkileyen noktaların (sapan değerler, yüksek leverage ve etkili noktalar) olup olmadığı araştırılır. Analiz sonuçlarını etkileyen bu noktaların ortaya çıkarılması amacıyla da çeşitli ölçütlerden yararlanılır. Modelin geçerliliğinin araştırılması aşaması, verilen modelin tahmin sürecini nasıl etkilediğini de içerir. Modelleme sürecinin tamamlanması açısından bu aşama oldukça önemlidir. Eğer model veriye uygunsa, tahmin edilen model kullanılabilir ve modelleme süreci tamamlanmış olur. Eğer model veriye uygun değilse, uygun model elde edilene kadar modelleme süreci tekrarlanır.

Bu tez çalışmasında, istatistiksel analizlerde en sık yararlanılan modellerden biri olan lineer (doğrusal) karma modellerde rezidü ve etki analizi ele alınacaktır. Lineer karma modeller, hem sabit hem de rastgele etkilerin bir arada bulunduğu, lineer regresyon modellerinin genişletilmiş şekli olan modellerdir. Bu nedenle model tahmini, sabit etkili parametrelerin, rastgele etkili parametrelerin ve varyans bileşenlerinin tahminini içermektedir. Lineer karma modeller, uzun süreli verilerin (longitudinal data) veya tekrarlı ölçümlerin analizinde, istatistiksel genetik çalışmalarında çok sık kullanılan modellerdendir. Çünkü, lineer regresyon modellerinin metodolojisi, uzun süreli veriler veya tekrarlı ölçümler gibi karmaşık verilerin analizi için oldukça kısıtlayıcıdır.

Lineer karma modellerin teorisi ise, Ronald Fisher tarafından tanımlanan rastgele etkili modellerin teorisine dayanmaktadır. Daha sonra 1950 lerde, Charles Roy Henderson tarafından sabit etkilerin en iyi lineer yansız tahmin edicisi (BLUE) ile rastgele etkilerin en iyi lineer yansız ön tahmin edicisi (BLUP) geliştirilmiştir.

Bu tez, lineer karma modellerde analiz sonuçlarını etkileyen noktaların ortaya çıkarılması için lineer regresyon modellerinde tanımlanan ölçütlerin, lineer karma modellere genişletilmesi amacıyla oluşturulan ve altı bölümden oluşan bir çalışmadır.

Çalışmanın ikinci bölümünde, lineer karma modellerdeki rezidü ve etki analizinin daha iyi bir şekilde anlaşılabilmesi amacıyla lineer regresyon modellerindeki rezidü ve etki analizi ele alınacaktır.

Çalışmanın üçüncü bölümünde, sırasıyla, lineer karma model tipleri, lineer karma modellerin parametre tahminleri ve güncelleme formülleri, tahmin edilen modelin geçerliliğinin araştırılması amacıyla rezidülerin incelenmesi, lineer karma modellerdeki etkili birimlerin ve gözlemlerin belirlenmesinde kullanışlı olabileceği düşünülen genelleştirilmiş leverage matrisi ve lineer karma modellerdeki etkili noktaların ortaya çıkarılması için geliştirilen ölçütler üzerinde durulacaktır.

Çalışmanın dördüncü bölümünde, lineer karma model tiplerinden biri olan hiyerarşik lineer modellerde rezidü analizi incelenecektir.

Çalışmanın beşinci bölümünde, üçüncü bölümde incelenen lineer karma modellerdeki rezidü, leverage ve etki analizi ile ilgili üç farklı örnek R ve Matlab programları yardımıyla analiz edilerek, konunun daha iyi bir şekilde algılanması sağlanacaktır.

2. LİNEER REGRESYON MODELLERİNDE REZİDÜ VE ETKİ ANALİZİ

2.1. Giriş

Regresyon analizinin amacı, elde edilen veri yardımıyla yanıt ve açıklayıcı değişkenler arasındaki ilişkinin açıklanmasını sağlayacak uygun bir modelin oluşturulmasıdır. Yanıt değişken y, açıklayıcı değişkenler $X_1, X_2, \cdots, X_p$ ile gösterilsin. y ile $X_1, X_2, \cdots, X_p$ arasındaki gerçek ilişkiye,

$$y = f(X_1, X_2, \cdots, X_p) + \varepsilon$$

regresyon modeli ile yaklaşılır. Burada $f(X_1, X_2, \cdots, X_p)$, y ile $X_1, X_2, \cdots, X_p$ arasındaki ilişkiyi tanımlayan fonksiyon ve ε, modelin veriye uyumundan kaynaklanan hata terimidir. Açıklayıcı değişkenler ve yanıt değişken y arasındaki ilişki lineerdir veya yaklaşık olarak lineerdir.

Bir regresyon modeli, bir tek açıklayıcı değişken içeriyorsa ***basit lineer regresyon modeli (simple linear regression model),*** birden fazla açıklayıcı değişken içeriyorsa ***çoklu lineer regresyon modeli (multiple linear regression model)*** olarak adlandırılır.

Genel olarak çoklu lineer regresyon modeli,

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

şeklinde ifade edilir. Burada y, $n \times 1$ tipinde yanıt değişkenin değerlerinin vektörü, X, $n \times p$ tipinde ilk sütunu 1 ler den oluşan tam sütun ranklı bilinen açıklayıcı değişkenlerin matrisi, β, $p \times 1$ tipinde bilinmeyen regresyon katsayılarının vektörü, ve ε, $n \times 1$ tipinde sıfır ortalamalı, bilinmeyen σ^2 varyanslı normal dağılıma sahip bağımsız rastgele hataların bir vektörüdür.

Uydurulan (2.1) modeline en küçük kareler (EKK) yöntemi uygulandığında,

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

$$Var(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$$

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$$

$$P = X(X^T X)^{-1} X^T$$

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

$$Var(\hat{y}) = \sigma^2 P$$

$$\hat{y} \sim N(X\beta, \sigma^2 P)$$

$$e = y - \hat{y} = (I - P)y$$

$$Var(e) = \sigma^2 (I - P)$$

$$e \sim N\big(0, \sigma^2 (I - P)\big)$$

$$\hat{\sigma}^2 = \frac{e^T e}{n - p} = MS_{Res}$$

elde edilir. Burada e, ***rezidüler vektörü*** ve P, ***şapka matrisi*** olarak adlandırılır. Bu şapka matrisinin i-inci köşegen elemanı ise, X matrisinin i-inci sütunu x_i^T olmak üzere, $p_{ii} = x_i^T (X^T X)^{-1} x_i$ ile gösterilir. EKK yöntemiyle elde edilen tahmin ediciler, en iyi lineer yansız tahmin edicilerdir (BLUE).

Genel olarak regresyon analizi çalışmalarında, tüm gözlemlerin eşit güvenilirlikte olduğu ve regresyon sonuçlarının belirlenmesi üzerinde yaklaşık olarak aynı etkiye sahip olduğu varsayılır. Bu varsayımın sağlanmaması durumunda bu

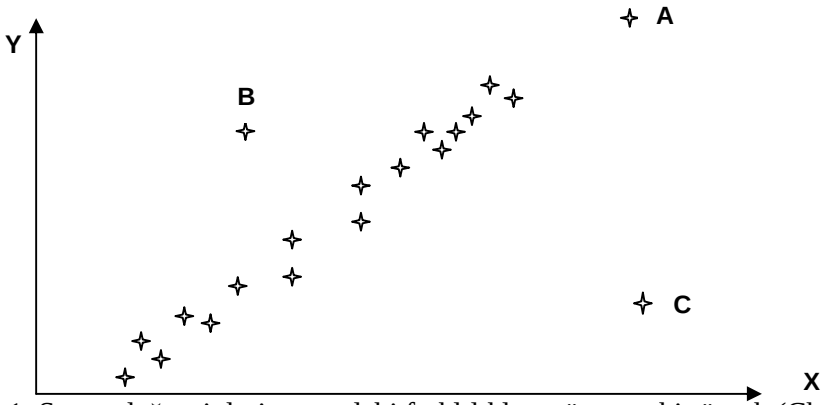
problemlere neden olan gözlemlerin belirlenmesi, modelin uygunluğu, güvenilirliği ve kararlılığı için gereklidir. Bu gözlemler y-yönünde sapan değerler (outliers), X-yönünde sapan değerler (yüksek leverage noktalar (high-leverage points)) ve etkili gözlemler (influential observations) olmak üzere üç grupta incelenebilir.

Lineer regresyonda, uydurulmuş regresyon doğrusunun uzağında olan, başka bir deyişle rezidüsü büyük olan gözlemlere ***y-yönünde sapan değerler*** denir. Eldeki X veri grubunun yoğunlaştığı bölgenin uzağında bulunan gözlemlere ***X-yönünde sapan değerler (yüksek leverage noktalar)*** denir. Ayrı ayrı ya da hep beraber diğer gözlemlerle karşılaştırıldığında regresyon denkleminin uydurulmasında etkili olan gözlemlere ***etkili gözlemler*** denir (Chatterjee ve Hadi, 1988).

y-yönündeki, X-yönündeki sapan değerler ve etkili gözlemler arasındaki ilişkiler aşağıdaki gibi verilebilir (Chatterjee ve Hadi, 1988):

- y-yönündeki sapan değerler etkili gözlem olmayabilir.
- Etkili gözlemler y-yönünde sapan değer olmayabilir.
- Büyük rezidüye sahip olan gözlemler istenmediği gibi, küçük rezidüye sahip gözlem de olağan bir gözlem olmayabilir. Genel olarak, yüksek leverage noktaları daha düşük rezidüye sahiptir.
- Yüksek leverage gözlemleri, etkili gözlem olmadığı gibi, etkili gözlemlerde yüksek leverage değerine sahip olmayabilir.

Bu üç grup arasındaki ilişkilerin daha iyi bir şekilde anlaşılabilmesi amacıyla bir sonraki sayfada verilen örnekler incelenebilir.

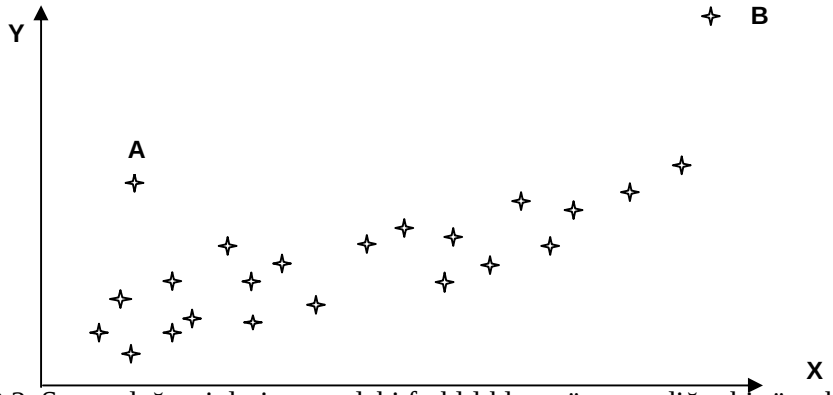


Şekil 2.1. Sapan değer tipleri arasındaki farklılıkları gösteren bir örnek (Chatterjee ve Hadi, 1988)

Şekil 2.1 incelendiğinde, uydurulmuş regresyon doğrusunda bulunan diğer noktalarla aynı hizada yer alan A noktasının, küçük bir rezidü değerine sahip olduğu ve bu nedenle, y-yönünde sapan bir değer olmadığı sonucuna varılır. Aynı zamanda A noktası, uydurulmuş regresyon doğrusu ve tahmin edilen regresyon katsayıları üzerinde büyük bir etkiye sahip olmaması nedeniyle etkili bir nokta da değildir. A noktası, X-yönünde sapan bir değer olup, yüksek leverage bir noktadır.

B noktası, X uzayının merkezine yakın olması nedeniyle yüksek leverage bir nokta değildir. Fakat B noktası, uydurulmuş regresyon doğrusuna uzak olması nedeniyle büyük bir rezidü değerine sahiptir ve dolayısıyla, hem y-yönünde sapan bir değer hem de etkili bir gözlemdir. Bunun bir sonucu olarak, uydurulmuş regresyon doğrusunun eğimi değişmeyebilir fakat, uydurulmuş doğrunun y eksenini kestiği nokta değişecektir. Aynı zamanda tahmin edilen hata varyansı ve dolayısıyla, tahmin edilen katsayıların varyansı da değişecektir.

C noktası, uydurulmuş regresyon doğrusunun uzağında olması nedeniyle büyük bir rezidü değerine sahiptir ve y-yönünde sapan bir değerdir. Aynı zamanda, X uzayının merkezindeki noktaların uzağında olması nedeniyle yüksek leverage bir noktadır. Son olarak, uydurulmuş regresyon eşitliğinin özelliklerini büyük ölçüde değiştirebileceğinden dolayı etkili bir gözlemdir.



Şekil 2.2. Sapan değer tipleri arasındaki farklılıkları gösteren diğer bir örnek (Chatterjee ve Hadi, 1988)

Şekil 2.2 incelendiğinde, y-yönündeki sapan değerlerin her zaman için etkili gözlem, etkili gözlemlerin ise y-yönünde sapan değer olmadığı görülebilir. A noktası, uydurulmuş regresyon doğrusuna uzak olması nedeniyle y-yönünde sapan bir değer olmasına rağmen, bu gözlemin çıkarılması durumunda uydurulmuş doğruda önemli bir değişme olmaz, yani A noktası, etkili bir gözlem değildir. Diğer taraftan B noktası, uydurulmuş regresyon doğrusuna yakın olması nedeniyle y-yönünde sapan bir değer olmamasına rağmen, bu gözlemin çıkarılması durumunda uydurulmuş doğrunun eğimi önemli derecede değişebilir ve bu nedenle, B noktasının etkili bir gözlem olduğu söylenebilir.

y-yönünde sapan değer, model ve/veya dağılım varsayımlarının geçersizliği hakkında bilgi verir. Bu tür gözlemlerin belirlenmesi regresyon sonuçları açısından oldukça önemlidir ve bu gözlemlerin saptanabilmesi için rezidü analizi yapılır. X-yönünde sapan değerler (yüksek leverage noktalar) ve etkili gözlemler ise veri noktalarının seyrek bulunduğu alan hakkında bilgi vermektedir. Bu tür gözlemlerin belirlenmesi regresyon sonuçları açısından oldukça önemlidir ve bu gözlemlerin saptanabilmesi için tanılama ölçüleri adı verilen çok sayıda istatistiksel ölçü vardır.

Sırasıyla, X-yönündeki, y-yönündeki, X ve y-yönündeki sapan değerleri ve etkili gözlemleri belirlemeye yönelik ölçüler incelenecektir.

2.2. X-yönündeki Sapan Değerleri Belirleyen Ölçüler

Bir gözlem X-uzayındaki gözlemlerin ortalamasından ne kadar uzaksa, p_{ii} değeri o kadar büyük olur ve bu nedenle p_{ii}, X uzayının uzağında bulunan sapan değerlerin belirlenmesinde bir ölçü olarak kullanılabilir. $p_{ii} \geq 2p/n$ olan gözlemler yüksek leverage noktaları olarak düşünülebilir (Montgomery ve ark., 2001).

P şapka matrisi ve leverage değerleri olan p_{ii} ler bir takım özelliklere sahiptir. Bu özelliklerden bazıları aşağıda verilmiştir.

X, $n \times p$ tipinde açıklayıcı değişkenlerin matrisi ve $P = X(X^TX)^{-1}X^T$ olmak üzere;

- P, simetrik ($P^T = P$) ve idempotent ($P^2 = P$) bir matristir.
- $trace(P) = rank(P) = p$.
- $X = (X_1, X_2)$ ve X_1, $n \times r$ ve X_2, $n \times (p-r)$ tipinde tam kolon ranklı matrisler olsun. Bu durumda, X için şapka matrisi $P_1 = X_1(X_1^TX_1)^{-1}X_1^T$ olmak üzere;

  $$P = P_1 + (I - P_1)X_2[X_2^T(I - P_1)X_2]^{-1}X_2^T(I - P_1)$$

  şeklinde hesaplanabilir.
- Sabit terim içermeyen modeller için $0 \leq p_{ii} \leq 1$, $i = 1, \cdots, n$ şeklindedir (Cook ve Weisberg, 1982).

X uzayının uzağında bulunan sapan değerleri belirlemek için kullanılan başka ölçüler de vardır. Bu ölçülerin birisi olan Mahalonobis uzaklığında amaç, x_i nin (i-inci satır) diğer gözlemlerden ne kadar uzakta olduğunu ölçmektir. Mahalonobis uzaklığı $i = 1, \cdots, n$ için,

$$M_i = \frac{n(n-2)}{n-1}\frac{(p_{ii} - 1/n)}{(1 - p_{ii})}$$

olarak tanımlanır (Bkz. Ek 1). Bir diğer ölçü, ağırlıklı kareler standartlaştırılmış uzaklığı olup (WSSD) $i = 1, \cdots, n$ için,

$$WSSD_i = \frac{\sum_{j=1}^{p}\left[\hat{\beta}_j\left(x_{ij} - \overline{X}_j\right)\right]^2}{S_Y^2}, \qquad S_Y^2 = \frac{\sum_{j=1}^{n}\left(y_j - \overline{y}\right)^2}{n-1}$$

ile verilir (Topçubaşı, 2001).

2.3. Yanıt Değişkeni y-Yönündeki Sapan Değerleri Belirleyen Ölçüler

Daha önce de tanımlandığı gibi, yüksek rezidüye sahip gözlemler uydurulmuş doğrudan uzak olduklarından ***y-yönünde sapan değer*** olarak adlandırılırlar. y-yönündeki sapan değerlerin belirlenebilmesi için temel olarak kullanılan ölçü, rezidüler üzerindeki çalışmalara dayalıdır.

y_i gözlenen değer, $\hat{y}_i$ karşılık gelen uydurulmuş değer olmak rezidü,

$$e_i = y_i - \hat{y}_i$$

olarak tanımlanır. Rezidü, veri ile uyum arasındaki sapmayı ölçtüğü gibi, regresyon modeli tarafından açıklanmayan yanıt değişkendeki değişimin de bir ölçüsüdür. Rezidüler, model hatalarının gözlenen veya uygun değerleri olarak da düşünülür. Bu nedenle, hatalar üzerindeki varsayımlardan herhangi bir sapma rezidülerle görülür. Rezidü analizleri, çeşitli model yetersizliklerini belirlemede etkilidir. Rezidüler, sıfır ortalamaya sahiptirler. Varyansları ise yaklaşık olarak,

$$\hat{\sigma}^2 = \frac{e^T e}{n-p} = MS_{Res}$$

ile tahmin edilir.

Maskeleme problemi, veride sapan değer olmasına karşın bu gözlemlerin belirlenememesidir. Bu genellikle bir sapan değerin, verideki diğer sapan değerler

tarafından gizlenmesiyle ortaya çıkar. Bazı gözlemler sorunsuz olmalarına karşın, sapan değer olarak belirlenebilir. Bu durum ise ***swamping problemi*** olarak bilinir. Genellikle verideki gerçek sapan değerlerin uydurulmuş regresyon doğrusunu kendilerine doğru çekmeleri ve böylece gerçek gözlemlerin uydurulmuş doğrudan uzaklaşmaları nedeniyle ortaya çıkar. Eğer veride sadece bir sapan değer varsa bu problemle karşılaşılmaz ancak birden fazla sapan değer olması durumunda oldukça sık karşılaşılabilecek bir problemdir.

Ölçeklendirilmiş rezidüler, y-yönündeki sapan değerlerin veya uç değerlerin yani veriden ayrı duran gözlemlerin belirlenmesinde yararlıdır. Dört farklı ölçeklendirme incelenecektir.

2.3.1. Standartlaştırılmış Rezidüler

Bir rezidünün varyansı yaklaşık olarak MS_{Res} ile tahmin edildiğinden standartlaştırılmış rezidü,

$$e_i^* = \frac{e_i}{\sqrt{MS_{Res}}}$$

olarak tanımlanır. Standartlaştırılmış rezidüler, sıfır ortalama ve yaklaşık olarak birim varyansa sahiptirler (Montgomery ve ark., 2001).

2.3.2. Studentized Rezidüler

MS_{Res}, i-inci rezidü e_i için sadece bir yaklaşımdır. e_i, i-inci rezidü için tam standart sapma ile bölünerek de ölçeklendirilebilir. $P = X(X^TX)^{-1}X^T$ şapka matrisi olmak üzere, rezidü vektörü,

$$e = y - \hat{y} = y - X\hat{\beta} = y - X(X^TX)^{-1}X^Ty = (I - P)y$$

dir. $y = X\beta + \varepsilon$, rezidü vektöründe yerine yazılırsa,

$$e = (I - P)y = (I - P)(X\beta + \varepsilon) = X\beta - P\,X\beta + (I - P)\varepsilon$$
$$= X\beta - X\overbrace{(X^TX)^{-1}X^TX}^{I}\beta + (I - P)\varepsilon = (I - P)\varepsilon$$

elde edilir. Yani rezidüler, y gözlemlerinin ve ε hatalarının aynı lineer dönüşümüdür. Şapka matrisinin simetrik ve idempotent olmasından dolayı,

$$var(e) = var\big((I - P)\varepsilon\big) = (I - P)\underbrace{var(\varepsilon)}_{\sigma^2 I}(I - P)^T = \sigma^2(I - P)$$

elde edilir. $(I - P)$, genellikle köşegen değildir. Bu nedenle, rezidüler farklı varyanslara sahiptir ve ilişkilidir. p_{ii}, P şapka matrisinin i-inci köşegen elemanı, p_{ij} ise şapka matrisinin j-inci elemanı olmak üzere,

$$var(e_i) = \sigma^2(1 - p_{ii})$$

$$cov\big(e_i, e_j\big) = -\sigma^2 p_{ij}$$

dir. $0 \le p_{ii} \le 1$ olduğundan, rezidünün varyans tahmini için MS_{Res} nin kullanılması aşırı tahmine neden olacaktır. Ayrıca p_{ii}, i-inci gözlemin x uzayındaki konumunu gösterir. Bu nedenle e_i nin varyansı, x_i gözleminin konumuna bağlıdır. x uzayının merkezine yakın noktalar uzaktakilerden daha büyük varyansa sahip olur. Model varsayımlarındaki sapmalar, uzak duran gözlemlerde daha belirgin olur ve e_i veya e_i^* lar incelenerek bu varsayımları belirlemek zordur. Bu nedenle, bu ikisi yerine studentized rezidüler,

$$t_i = \frac{e_i}{\sqrt{MS_{Res}(1 - p_{ii})}}, i = 1, \cdots, n$$

önerilir. Bu rezidüler, model doğru iken x_i nin konumu ne olursa olsun sabit varyansa sahiptir. Büyük veri kümelerinde rezidülerin varyansı durağan olur. Bu gibi

durumlarda, e_i^* ve t_i arasında çok az bir farklılık olacaktır. Dolayısıyla, e_i^* ve t_i aynı bilgiyi içerir. Fakat büyük p_{ii} ve büyük rezidüye sahip herhangi bir nokta EKK uyumda etkili gözlem ise, studentized rezidünün incelenmesi önerilir (Montgomery ve ark., 2001).

2.3.3. PRESS Rezidüler

Standartlaştırılmış ve studentized rezidüler dışında sapan değerleri belirlemede bir diğer ölçeklendirme PRESS yani ön tahmin hata kareler toplamıdır. i-inci PRESS rezidüyü hesaplayabilmek için i-inci gözlem hariç diğer tüm $n-1$ gözlemden, $\hat{y}_{(i)}$ hesaplanır ve buna karşılık gelen ön tahmin hatası $e_{(i)} = y_i - \hat{y}_{(i)}$ hesaplanır. Bu ön tahmin hatası, tüm $i = 1, \cdots, n$ gözlemler için tekrarlanır. Ve bu ön tahmin hatalarına ***PRESS rezidüler*** denir. PRESS rezidü hesaplamak n farklı regresyon uydurmayı gerektirir. Fakat PRESS rezidü, n gözlemi EKK uydurularak da elde edilebilir:

$$e_{(i)} = \frac{e_i}{1-p_{ii}}, i = 1, \cdots, n$$

Büyük p_{ii} değerleri, büyük PRESS rezidülere neden olur. Böyle noktalar, yüksek etkili noktalardır.

$$var(e_{(i)}) = var\left(\frac{e_i}{1-p_{ii}}\right) = \frac{1}{(1-p_{ii})^2} var(e_i) = \frac{1}{(1-p_{ii})^2}\sigma^2(1-p_{ii})$$
$$= \frac{\sigma^2}{(1-p_{ii})}$$

olup, standartlaştırılmış PRESS rezidü,

$$\frac{e_{(i)}}{\sqrt{var(e_{(i)})}} = \frac{e_i/(1-p_{ii})}{\sqrt{\sigma^2/(1-p_{ii})}} = \frac{e_i}{\sqrt{\sigma^2(1-p_{ii})}}$$

şeklindedir. σ^2 yi tahmin etmek için MS_{Res} kullanılırsa, studentized rezidü elde edilir (Montgomery ve ark., 2001).

2.3.4. *R*-Student

Studentized rezidü t_i ler hesaplanırken σ^2 nin bir tahmini olan MS_{Res} kullanılmıştı. Bu ölçeklendirmeye ***iç veya dahili ölçeklendirme*** denir. σ^2 nin tahmini için bir diğer yaklaşım da i-inci gözlemin veriden atılmasına dayalıdır. Bu şekilde elde edilen σ^2 nin tahmini,

$$S_{(i)}^2 = \frac{(n-p)MS_{Res} - {e_i}^2/(1-p_{ii})}{n-p-1}$$

şeklinde olur. Buna göre,

$$t_i^* = \frac{e_i}{\sqrt{S_{(i)}^2(1-p_{ii})}}, i = 1, \cdots, n$$

olarak tanımlanan R-Student t dağılımına sahiptir ve ***dıştan studentized rezidü*** olarak adlandırılır. t_i^*, t_i den çok farklı değildir. Bununla birlikte, i-inci gözlem etkili ise MS_{Res} den önemli derecede farklı olur. Ve R-Student, bu noktada daha hassastır (Montgomery ve ark., 2001).

2.4. *X* ve *y*-yönündeki Sapan Değerleri Belirleyen Ölçüler

İyi bir veri analizi için rezidüler mutlaka incelenmelidir. Ancak sadece rezidülerin incelenmesi etkili gözlemlerin saptanmasında özellikle de yüksek leverage noktaları için yeterli değildir. Çünkü rezidü (e_i) ile leverage değerleri (p_{ii}) arasında,

$$p_{ii} + \frac{{e_i}^2}{e^T e} \leq 1$$

şeklinde bir bağıntı vardır. Bu bağıntı, leverage değeri yüksek gözlemlerin, küçük rezidüye sahip olduklarını göstermektedir. Sonuç olarak, bu iki ölçünün bazı noktalardaki problemleri belirlemede tamamlayıcı rol oynadıkları bilindiğinden, rezidü ve p_{ii} değerlerinin birlikte değerlendirilmeleri gerekmektedir (Chatterjee ve ark., 2000). Bazı ölçüler hem p_{ii} hem de rezidüler hakkında bilgi verebilir.

Bu ölçülerden biri $Z = (X:y)$ şeklindeki eklemeli matris olmak üzere, $P_z = Z(Z^T Z)^{-1} Z^T$ şapka matrisinin köşegen elemanlarıdır (Topçubaşı, 2001). Bölüm 2.2'de tanımlanan özelliklerden, $X_1 = X$ ve $X_2 = y$ alınırsa, $P_z = P_X + \frac{ee^T}{e^T e}$ olarak elde edilir. P_z matrisinin i-inci köşegen elemanı $i = 1, \cdots, n$ için,

$$p_{zii} = p_{ii} + \frac{e_i^2}{e^T e}$$

şeklinde tanımlanır. p_{zii}, p_{ii} ya da e_i büyük olduğunda yine büyük olacaktır. Dolayısıyla, p_{zii} değeri büyük olan bir gözlemin sapan değer olduğu söylenebilir; ancak X-yönünde ya da y-yönünde sapan değer olup olmadığı konusunda kesin bir yargıya varılamaz.

Diğer bir ölçü, Hadi (1992) tarafından verilen H_i^2 ölçüsüdür ve

$$H_i^2 = \frac{p_{ii}}{1 - p_{ii}} + \frac{p}{1 - p_{ii}} \frac{{e_i^*}^2}{1 - {e_i^*}^2}, i = 1, \cdots, n \qquad (2.2)$$

ile tanımlanır. Bu eşitliğin ilk terimi potansiyel fonksiyon olup, X-yönündeki sapan değerlerin; ikinci terim ise standartlaştırılmış rezidünün fonksiyonu olup, y-yönündeki sapan değerlerin belirlenmesini sağlar. Dolayısıyla, ${e_i^*}^2$ veya p_{ii} ya da her ikisi de büyükse H_i^2 büyük olacaktır. Büyük H_i^2 değerine sahip bir gözlem, genel olarak tüm regresyon sonuçları üzerinde daha etkilidir.

2.5. Etkili Gözlemleri Belirleyici Ölçüler

Tüm gözlemler regresyon sonuçları üzerinde eşit etkiye sahip olmayabilir. Bunun yanında bir gözlem veya gözlemler bazı regresyon sonuçları için etkili iken, tüm regresyon sonuçları için etkili olmayabilir ($\hat{\beta}$ ya da $\hat{\beta}$ nın varyansı üzerinde, uydurulmuş değerler üzerinde etkili olabilir). Dolayısıyla, "ne üzerinde etkili?" sorusu önemlidir. O halde, araştırmacının öncelikle bu soruyu yanıtlayarak, çalışmaya başlaması gerekmektedir. Bununla birlikte bir gözlem, regresyon sonuçları üzerinde etkili ise y-yönünde sapan değer, X-yönünde sapan değer ya da ikisinin kombinasyonu olabilir. Dolayısıyla etki ölçüleri, leverage ve rezidü değerlerinin bir fonksiyonu olarak ifade edilir. Bu kısımda açıklanacak etki ölçüleri üç ana başlık altında incelenecektir.

2.5.1. Etki Eğrisine Dayalı Ölçüler

i-inci gözlemin etkisini ölçmek için Hampel (1974) tarafından tanımlanan etki fonksiyonu,

$$IF_i(x_i; y_i; F; T) = \lim_{\varepsilon \to 0} \frac{T\left[(1-\varepsilon)F + \varepsilon\delta_{x_i,y_i}\right] - T[F]}{\varepsilon} \tag{2.3}$$

şeklindedir. Burada $T(\cdot)$, kümülatif F den elde edilen rastgele örnekleme dayalı vektör değerli bir istatistik ve (x_i, y_i) de $\delta_{x_i,y_i} = 1$, diğer yerlerde $\delta_{x_i,y_i} = 0$ dır. IF_i, çok büyük bir örnekleme (x_i, y_i) tek gözlemi eklendiğinde T üzerindeki etkiyi ölçer. Sınırlı bir örneklem için, (2.3) eşitliğine alternatif yaklaşımlar da mevcuttur. Bu yaklaşımlardan üç tanesi deneysel etki eğrisi, örneklem etki eğrisi ve duyarlılık eğrisidir (Chatterjee ve Hadi, 1986).

$\hat{F}$, tam örnekleme dayalı deneysel dağılım fonksiyonu ve $\hat{F}_{(i)}$, i-inci gözlem çıkarıldığında elde edilen dağılım fonksiyonu olsun.

a) Deneysel etki eğrisi (EIC), (2.3) eşitliğinde F yerine $\hat{F}_{(i)}$ ve $T(\hat{F}_{(i)})$ yerine $\hat{\beta}_{(i)}$ alınmasıyla,

$$EIC_i = (n-1)\left(X_{(i)}^T X_{(i)}\right)^{-1} x_{(i)}^T \left(y_i - x_i \hat{\beta}_{(i)}\right)$$
$$= (n-1)(X^T X)^{-1} x_{(i)}^T \frac{e_i}{(1-p_{ii})^2}$$
$$= (n-1)\frac{1}{1-p_{ii}}\left(\hat{\beta} - \hat{\beta}_{(i)}\right)$$

olarak elde edilir. Burada, $\hat{\beta}_{(i)} = \left(X_{(i)}^T X_{(i)}\right)^{-1} X_{(i)}^T y_{(i)}$ olup, i-inci gözlem çıkarıldığında β nın tahminini göstermektedir.

b) Örneklem etki eğrisi (SIC), (2.3) eşitliğinin limitinin alınması ve $F = \hat{F}$, $T(\hat{F}) = \hat{\beta}, \varepsilon = -1/(n-1)$ alınmasıyla,

$$SIC_i = (n-1)(X^T X)^{-1} x_{(i)}^T \left(y_i - x_i \hat{\beta}_{(i)}\right) = (n-1)(X^T X)^{-1} x_{(i)}^T \frac{e_i}{(1-p_{ii})}$$
$$= (n-1)\left(\hat{\beta} - \hat{\beta}_{(i)}\right)$$

olarak elde edilir.

c) Duyarlılık eğrisi (SC), (2.3) eşitliğinde $F = \hat{F}_{(i)}$, $T(\hat{F}_{(i)}) = \hat{\beta}_{(i)}$, $\varepsilon = -1/n$ alınmasıyla,

$$SC_i = n(X^T X)^{-1} x_{(i)}^T \frac{e_i}{(1-p_{ii})} = n\left(\hat{\beta} - \hat{\beta}_{(i)}\right)$$

olarak elde edilir.

Açıkça görülüyor ki SIC_i ile SC_i eşdeğerdir. Miller (1974),

$$\hat{\beta} - \hat{\beta}_{(i)} = (X^T X)^{-1} x_{(i)}^T \frac{e_i}{(1 - p_{ii})} \tag{2.4}$$

olduğunu göstermiştir. (2.4) eşitliği yardımıyla SIC_i ile SC_i nin $\hat{\beta}$ ile $\hat{\beta}_{(i)}$ arasındaki uzaklık ile orantılı olduğu söylenebilir. Bununla birlikte, EIC_i, p_{ii} ye karşı SIC_i ile SC_i ye göre daha hassastır.

IF_i bir vektör olduğu için, gözlemlerin anlamlı bir şekilde sıralanabilmeleri için IF_i normalleştirilmelidir. Bu normalleştirme M ve c nin uygun seçimleri ile yapıldığında,

$$D_i(M, c) = \frac{(IF_i)^T M (IF_i)}{c}$$

elde edilir. $D_i(M, c)$ nin büyük bir değeri, i-inci gözlemin M ve c ile ilgili tahmin edilen katsayılar üzerinde güçlü bir etkiye sahip olduğunu gösterir (Chatterjee ve Hadi, 1986).

Burada M ve c nin uygun seçimleri ile Cook Uzaklığı, Welsch-Kuh Uzaklığı (DFFITS_i) ile $\text{DFBETAS}_{j,i}$ ve Welsch uzaklığı olarak bilinen ve yaygın olarak kullanılan üç etki ölçüsü incelenecektir.

2.5.1.1. Cook Uzaklığı

Cook (1977), tüm n gözleme dayalı $\hat{\beta}$, EKK tahmini ve i-inci gözlemi silerek elde edilen $\hat{\beta}_{(i)}$ tahmini arasındaki uzaklığın karesini etkili gözlemi belirlemek için önermiştir. Bu uzaklık ölçüsü genel olarak,

$$D_i(M, c) = \frac{\left(\hat{\beta}_{(i)} - \hat{\beta}\right)^T M \left(\hat{\beta}_{(i)} - \hat{\beta}\right)}{c}, \qquad i = 1, \cdots, n \tag{2.5}$$

formunda ifade edilir. $M = X^T X$ ve $c = pMS_{Res}$ uygun seçimleri yapıldığında (2.5) eşitliği,

$$D_i(X^TX, pMS_{Res}) \equiv D_i = \frac{\left(\hat{\beta}_{(i)} - \hat{\beta}\right)^T X^T X\left(\hat{\beta}_{(i)} - \hat{\beta}\right)}{pMS_{Res}}, i = 1, \cdots, n$$

şeklinde elde edilir.

D_i istatistiği,

$$D_i = \frac{{t_i}^2}{p}\frac{p_{ii}}{1 - p_{ii}} \quad i = 1, \cdots, n$$

olarak da hesaplanabilir ve $D_i > 1$ olan gözlemler etkilidir (Montgomery ve ark., 2001).

2.5.1.2. Welsch-Kuh Uzaklığı ($\mathbf{DFFITS}_i$) ve $\mathbf{DFBETAS}_{j,i}$

Belsley ve ark. (1980), i-inci gözlemin silinmesinin ya da atılmasının uydurulmuş gözlem üzerindeki etkisini ölçmek için DFFITS_i yi aşağıdaki gibi tanımlamışlardır.

$$\text{DFFITS}_i = WK_i = \frac{\hat{y}_i - \hat{y}_{(i)}}{\sqrt{S^2_{(i)} p_{ii}}} \quad i = 1, \cdots, n$$

Burada $\hat{y}_{(i)}$, i-inci gözlem kullanılmaksızın elde edilen y_i nin uydurulmuş değeridir.

$$\text{DFFITS}_i = WK_i = \left(\frac{p_{ii}}{1 - p_{ii}}\right)^{1/2} \quad t_i^*$$

biçiminde de ifade edilebilir. Bu ifade, rezidülerle leverage değerler arasındaki ilişkiyi gösterir. Veri noktası bir sapan değer ise t_i^* değeri büyük olacaktır. Veri noktası yüksek leverage noktası ise p_{ii}, 1 e yakın olur. Her iki durumda da DFFITS_i

büyük olabilir. p_{ii} yaklaşık 0 ise t_i^* kabul edilebilir ölçüde etkili olacaktır. Benzer şekilde, yüksek leverage noktada, t_i^* 0 a yakınsa DFFITS$_i$ küçük olur.

Bu nedenle, DFFITS$_i$ hem leverage hem de ön tahmin hatasından etkilenir. $|$ DFFITS$_i|$ > $2\sqrt{p/n}$ ise i-inci gözlem $\hat{y}$ üzerinde etkili gözlemdir.

Belsley ve ark. (1980), i-inci gözlemin silinmesinin ya da atılmasının $\hat{\beta}_j$ regresyon katsayısını ne kadar değiştirdiğini göstermek için DFBETAS$_{j,i}$ yi,

$$\text{DFBETAS}_{j,i} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{S_{(i)}^2 C_{jj}}}$$

olarak tanımlamışlardır. Burada C_{jj}, $(X^TX)^{-1}$ in j-inci esas köşegen elemanı ve $\hat{\beta}_{j(i)}$, i-inci gözlem kullanılmaksızın hesaplanan j-inci regresyon katsayısıdır. DFBETAS$_{j,i}$ nin büyük değerleri, i-inci gözlemin j-inci regresyon katsayısı üzerinde etkili olduğunu gösterir.

t_i^*, R-Student rezidü ve ŗ_j^T, $p \times n$ tipindeki R = $(X^TX)^{-1}X^T$ matrisinin j-inci satırını göstermek (Montgomery ve ark., 2001) üzere,

$$\text{DFBETAS}_{j,i} = \frac{\text{ŗ}_{j,i}}{\sqrt{\text{ŗ}_j \text{ŗ}_j}} \frac{t_i^*}{\sqrt{1-p_{ii}}}$$

olarak da yazılabilir.

Bu nedenle, DFBETAS$_{j,i}$ hem leverage hem de ön tahmin hatasından etkilenir. $\left|\text{DFBETAS}_{j,i}\right|$ > $2\sqrt{1/n}$ ise i-inci gözlem $\hat{\beta}_j$ nın belirlenmesinde etkilidir.

2.5.1.3. Welsch Uzaklığı

Eğer $(n-1)$ gözleme dayalı deneysel eğri kullanılırsa, $M = X_{(i)}^T X_{(i)} = X^TX - x_i^T x_i$ ve $c = (n-1)S_{(i)}^2$ alınırsa, $W_i{}^2 = D_i\left(X_{(i)}^T X_{(i)}, (n-1)S_{(i)}^2\right)$ ve buradan,

$$W_i = WK_i\sqrt{\frac{n-1}{1-p_{ii}}}$$

şeklinde tanımlanan Welsch uzaklığı elde edilir (Welsch, 1982).

2.5.2. Güven Elipsoidinin Hacmine Dayalı Ölçüler

Tahmin edilen regresyon katsayıları üzerinde i-inci gözlemin etkisinin bir ölçümü, i-inci gözlem modeldeyken ve modelde değilken güven elipsoidlerinin hacminde meydana gelen değişmeye dayalı olabilir. Andrews-Pregibon istatistiği, varyans oranı (COVRATIO) ve Cook-Weisberg ölçüleri bu sınıfta incelenebilir.

2.5.2.1. Andrew-Pregibon İstatistiği

β için güven elipsoidinin hacmi, X^TX matrisinin determinantı ile ters orantılıdır. Bu durumda, i-inci gözlemin güven elipsoidinin hacmi üzerindeki etkisini araştırmak için $det(X^TX)$ ile $det\left(X_{(i)}^TX_{(i)}\right)$ karşılaştırılabilir. Bu amaçla, Andrews ve Pregibon (1978) tarafından,

$$\frac{SSE_{(i)}det\left(X_{(i)}^TX_{(i)}\right)}{SSEdet(X^TX)} \tag{2.6}$$

oranı önerilmiştir. $Z = (X:y)$ matrisi ele alınırsa Z^TZ,

$$Z^TZ = \begin{pmatrix} X^TX & X^Ty \\ y^TX & y^Ty \end{pmatrix}$$

olup, Ek 2 deki teoremin (a) şıkkının kullanılmasıyla $det(Z^TZ) = det(X^TX)SSE$ elde edilir. Benzer olarak, $det\left(Z_{(i)}^TZ_{(i)}\right) = det\left(X_{(i)}^TX_{(i)}\right)SSE_{(i)}$ olduğu için (2.6) aynı zamanda,

$$\frac{det\left(Z_{(i)}^{T}Z_{(i)}\right)}{det(Z^{T}Z)}$$

şeklinde de ifade edilebilir.

Verinin merkezinden uzak olan bir gözlemin çıkarılması durumunda determinant azalacak ve dolayısıyla hacim artacaktır. Bu durumda $i = 1, \cdots, n$ için

$$AP_i = 1 - \frac{det\left(Z_{(i)}^{T}Z_{(i)}\right)}{det(Z^{T}Z)}$$

şeklinde tanımlanan AP_i nin küçük değerleri etkili gözlemlere karşılık gelir.

2.5.2.2. Varyans Oranı (COVRATIO)

Belsley ve ark. (1980), i-inci gözlemin, tahminin doğruluğu üzerindeki etkisini ifade etmek için,

$$VR_i = COVRATIO = \frac{det\left(\left(X_{(i)}^{T}X_{(i)}\right)^{-1}S_{(i)}^{2}\right)}{det((X^{T}X)^{-1}MS_{Res})}, i = 1, \cdots, n$$

tanımlamışlardır. Bu oran 1 den büyükse, i-inci gözlem tahminin doğruluğunu daha iyi hale getirir, 1 den küçükse i-inci gözlemin dahil edilmesi tahminin doğruluğunu düşürür.

$$VR_i = COVRATIO_i = \frac{\left(S_{(i)}^{2}\right)^{p}}{MS_{Res}{}^{p}}\left(\frac{1}{1-p_{ii}}\right), i = 1, \cdots, n$$

olarak da yazılabileceğinden yüksek leverage nokta $COVRATIO$ değerini büyütür. Yüksek etkili gözlem y-yönünde sapan bir gözlem değilse tahmin daha iyi hale gelecektir.

$$COVRATIO_i > 1 + \frac{3p}{n}$$

veya

$$COVRATIO_i < 1 - \frac{3p}{n}$$

ise i-inci gözlem etkili gözlem olarak belirlenir. Alt sınır sadece $n > 3p$ iken geçerlidir. Bu eşik değer sadece büyük örneklemler için geçerlidir.

2.5.2.3. Cook Weisberg İstatistiği

Normallik varsayımı altında β için $\%(1-\alpha)$'lık güven elipsoidi,

$$E = \left\{\beta: \frac{(\beta-\hat{\beta})^T X^T X(\beta-\hat{\beta})}{(pMS_{Res})} \leq F(\alpha, p, n-p)\right\}$$

olup, i-inci gözlem silindiğinde,

$$E_{(i)} = \left\{\beta: \frac{(\beta-\hat{\beta}_{(i)})^T X_{(i)}^T X_{(i)}(\beta-\hat{\beta}_{(i)})}{pMS_{Res}} \leq F(\alpha, p, n-p-1)\right\}$$

olarak elde edilir.

Cook ve Weisberg (1980), çoklu lineer regresyon modeli için güven elipsoidinin hacmi üzerindeki etkinin bir ölçümü olarak $E_{(i)}$ nin E ye oranının logaritmasını önermişlerdir.

$$CW_i = log\left(\frac{volume(E_{(i)})}{volume(E)}\right) \quad i = 1, \cdots, n \tag{2.7}$$

Dikkat edilirse, elipsoidin hacmi, karesel matrisin determinantının karekökünün tersi ile orantılıdır. Yani Cook ve Weisberg (1982) tarafından tanımlanan,

$$volume(E) \propto (p\hat{\sigma}_e^2 F(\alpha, p, n-p))^{p/2} det(X^T X)^{1/2}$$

$$volume(E_{(i)}) \propto (p\hat{\sigma}_{e(i)}^2 F(\alpha, p, n-p-1))^{p/2} det(X_{(i)}^T X_{(i)})^{1/2}$$

ifadelerin kullanılması sonucunda (2.7) eşitliğinin,

$$CW_i = log\left\{\frac{det(X_{(i)}^T X_{(i)})^{1/2} \hat{\sigma}_e^p (F(\alpha, p, n-p))^{p/2}}{det(X^T X)^{1/2} \hat{\sigma}_{e(i)}^p (F(\alpha, p, n-p-1))^{p/2}}\right\}$$

şeklinde ifade edilmesi ile,

$$CW_i = \frac{1}{2} log(1-p_{ii}) + \frac{p}{2} log\left(\frac{n-p-1}{n-p-t_i^2}\right) + \frac{p}{2} log\left(\frac{F(\alpha, p, n-p)}{F(\alpha, p, n-p-1)}\right)$$

olarak elde edilir. Eğer bu değer büyük ve pozitifse, i-inci gözlemin çıkarılması hacimde önemli derecede azalmaya; büyük ve negatifse hacimde önemli derecede artışa neden olacaktır. Fakat p_{ii} ve t_i^2 nin her ikisi de büyükse CW_i değerinin sıfıra yakınsadığı görülür. Bu durumlarda CW_i nin çok sağlıklı sonuç vermediği söylenebilir, yani bu ölçü maskeleme problemine karşı dayanıklı değildir. Varyans oranı (VR_i) ve Cook Weisberg ölçüsü arasında,

$$CW_i = -\frac{1}{2} log(VR_i) + \frac{p}{2} log\left(\frac{F(\alpha, p, n-p)}{F(\alpha, p, n-p-1)}\right)$$

ile verilen bir ilişki vardır. Dolayısıyla, bu iki ölçü birbirine denk ölçülerdir.

2.5.3. Likelihood Fonksiyonuna Dayalı Ölçüler

$L(\hat{\beta})$ ve $L(\hat{\beta}_{(i)})$ sırasıyla $\hat{\beta}$ ve $\hat{\beta}_{(i)}$ daki log likelihood değerleri olmak üzere, $\hat{\beta}$ üzerindeki i-inci gözlemin etkisinin bir ölçümü $L(\hat{\beta})$ ve $L(\hat{\beta}_{(i)})$ arasındaki uzaklığa dayalı olabilir. Cook ve Weisberg (1982), likelihood uzaklığını,

$$LD_i = 2\left[L(\hat{\beta}) - L(\hat{\beta}_{(i)})\right]$$
$$= nlog\left[\left(\frac{n}{n-1}\right)\frac{n-p-1}{t_i^{*2}+n-p-1}\right] + \frac{t_i^{*2}(n-1)}{(1-p_{ii})(n-p-1)} - 1$$

olarak tanımlamışlardır. Burada $t_i^* = t_i\sqrt{\{(n-p-1)/(n-p-t_i^2)\}}$ dir (Hadi, 1986).

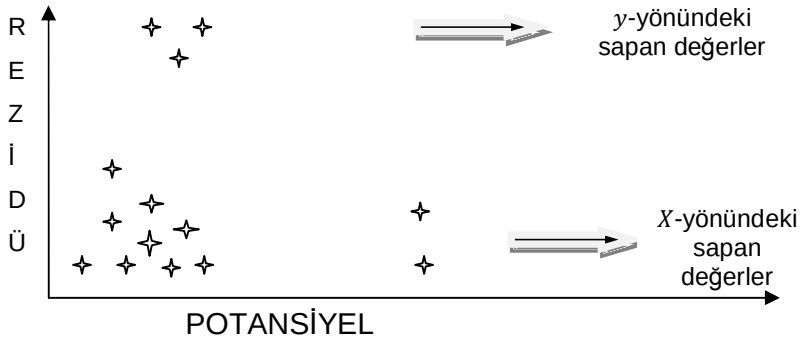
Likelihood uzaklığı, $\{\beta: 2\left[L(\hat{\beta}) - L(\beta)\right] \leq \mathcal{X}^2(\alpha, p+1)\}$ olan asimptotik güven bölgesiyle ilgilidir. $\mathcal{X}^2(\alpha, p+1)$, $(p+1)$ serbestlik dereceli ki-kare dağılımlı üst sınırdır. Sonuç olarak LD_i, $\mathcal{X}^2(\alpha, p+1)$ ile karşılaştırılır. Etkinin diğer ölçümlerinde kullanılan eşik değerler sayısal olarak ifade edilirken, etkinin bu ölçümünde kullanılan eşik değer olasılık modeline dayalıdır. Likelihood uzaklığının bir avantajı, normal doğrusal modelin dışındaki modellere de uygulanabilir olmasıdır.

2.6. *X* ve *y*-yönündeki Sapan Değerlerin Belirlenmesi için Grafiksel Metotlar

Verilen ölçüler için çeşitli eşik değerlerin var olduğu, ancak bunların her zaman sağlıklı sonuçlar vermediği bilinmektedir. Herhangi bir belirleyici ölçü için belli bir eşik değere bağlı kalmak araştırmacıyı yanlış sonuçlara götürebilir. Çünkü hesaplanan ölçü değerleri, gözlemlerden hiçbiri için eşik değeri geçmediği halde, birbirleriyle karşılaştırıldığında farklı davranan gözlemler ortaya çıkabilir. Bu nedenle önerilen ölçüleri birbirleriyle karşılaştırmak, yani indis grafikleri, kutu çizimleri ya da dal-yaprak gösterimlerini kullanmak, belli bir eşik değer seçmekten

daha sağlıklı sonuç verecektir. Böylece grafikte diğerlerinden ayrı davranan gözlemler sapan değer olarak düşünülebilir.

X ve y-yönündeki sapan değerlerin aynı anda görülmesini sağlayan grafiksel yöntemler de vardır. Bu grafiksel gösterimlerden birisi leverage-rezidü grafiğidir (Gray, 1983, 1986). Bu grafik p_{ii} değerlerine karşı e_i^{*2} lerin grafiğidir. Bu grafiğin bir benzeri ise potansiyel-rezidü grafiği (Hadi, 1992) olup, $P-R$ ile gösterilir. Bu grafik, (2.2) ile verilen ${H_i}^2$ ölçüsündeki birinci terim $\left(\frac{p_{ii}}{1-p_{ii}}\right)$ potansiyelin, ikinci terim $\left(\frac{p}{1-p_{ii}}\frac{e_i^{*2}}{1-e_i^{*2}}\right)$ rezidüye karşı grafiğidir. Bu grafikte sağ alt köşede yer alan gözlemler X-yönünde, sol üst köşedekiler ise y-yönündeki sapan değerlerdir (Şekil 2.3).



Şekil 2.3. Potansiyel-Rezidü Grafiği (Hadi, 1992)

3. LİNEER KARMA MODELLERDE REZİDÜ VE ETKİ ANALİZİ

3.1. Giriş

Lineer karma modeller, lineer regresyon modellerinin genişletilmiş hali olarak düşünülebilir. Lineer regresyon modellerde sadece sabit etkiler yer alırken, lineer karma modellerde hem sabit etkiler hem de rastgele etkiler bir arada bulunur.

Genel olarak lineer karma bir model,

$$y = X\beta + Zu + \varepsilon \tag{3.1}$$

şeklinde ifade edilir. Burada, y, $n \times 1$ tipinde gözlemlenebilir yanıt vektörü, β, $p \times 1$ tipinde sabit etkilere ilişkin parametreler vektörü, X ve Z sırasıyla $n \times p$ ve $n \times q$ tipinde sabit ve rastgele etkilere ilişkin tasarım matrisleri, u, $q \times 1$ tipinde rastgele etkilere ilişkin parametreler vektörü ve ε, $n \times 1$ tipinde rastgele hataların bir vektörüdür. u, $E(u) = 0$ ve $Var(u) = \sigma^2 D$ ile normal dağılıma sahiptir. Benzer şekilde ε, $E(\varepsilon) = 0$ ve $Var(\varepsilon) = \sigma^2 \Sigma$ ile normal dağılıma sahip olup, u ile ε karşılıklı olarak bağımsızdırlar. D ve Σ, bilinen pozitif tanımlı matrislerdir.

Lineer karma modellerde sabit ve rastgele etkilerin bir arada bulunması, parametre tahminlerinin elde edilmesinin daha karmaşık olmasına neden olur. Karma modellerin getirdiği avantajların yanında, model varsayımlarının çok olması nedeniyle uygulanma güçlüğü gibi dezavantajları da bulunmaktadır. Bununla birlikte lineer karma modeller, uzun süreli verilerin (longitudinal data) veya tekrarlı ölçümlerin analizinde, istatistiksel genetik çalışmalarında çok sık kullanılan modellerdendir.

3.1.1. Lineer Karma Model Tipleri

Lineer karma modellerin farklı tipleri ve onları sınıflandırmanın farklı yolları vardır. Sınıflandırma yollarından birisi, normallik varsayımının yapılıp yapılmamasına bağlıdır.

Normallik varsayımının yapıldığı modeller, bu varsayımın yapılmadığı modellere göre daha esnek davranabiliyorlarken, bu varsayımın yapılmadığı modellerde, dağılımsal varsayımlardan sapmalar daha fazladır. Bu açıdan değerlendirme yapıldığında karma modeller, Gaussian lineer karma modeller ve Gaussian olmayan lineer karma modeller olarak iki sınıfa ayrılır (Jiang, 2007). Bu çalışmada Gaussian lineer karma modeller üzerinde durulacaktır.

3.1.1.1. Gaussian Lineer Karma Modeller

Gaussian lineer karma model altında, (3.1)'de tanımlanan modelde olduğu gibi, rastgele etkilerin ve hataların normal dağıldığı varsayımı yapılır. Bu varsayım altında bazı özel lineer karma model tipleri incelenecektir.

1. Varyans Bileşenleri Modeli

Varyans bileşenleri modeli için (3.1) ile verilen lineer karma modelde,

$$u = \begin{bmatrix} u_1 \\ \vdots \\ u_r \end{bmatrix}, Z = [Z_1 \cdots Z_r]$$

olmak üzere $y = X\beta + Z_1u_1 + \cdots + Z_ru_r + \varepsilon$ şeklindedir. Öyle ki bu modelde, $i = 1, \cdots, r$ için $q = q_1 + q_2 + \cdots + q_r$ olmak üzere, u_i ve Z_i sırasıyla $q_i \times 1$ ve $n \times q_i$ tipindedir. $u_i \sim N(0, \sigma_e^2 D_i)$ dağılımından alınan bağımsız rastgele değişkenlerin vektörü ve $\varepsilon \sim N(0, \sigma_e^2 I_n)$ şeklinde olup, u_i ile ε karşılıklı olarak bağımsızdırlar. Aynı zamanda, $u \sim N(0, \sigma_e^2 D)$ olarak yazılabilir. Burada D, i-inci

bloğu $\gamma_i I_{qi}$ olan blok-köşegen bir matristir. $\gamma_i = \frac{\sigma_i^2}{\sigma_e^2}$ şeklinde olup, ***varyans bileşen oranları*** adını alır. Yani,

$$\sigma_e^2 D = var(u) = \begin{bmatrix} var(u_1) & \cdots & cov(u_1, u_r^T) \\ \vdots & \ddots & \vdots \\ cov(u_r, u_1^T) & \cdots & var(u_r) \end{bmatrix}$$

$$= \begin{bmatrix} var(u_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & var(u_r) \end{bmatrix} = \begin{bmatrix} \sigma_e^2 D_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_e^2 D_r \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_e^2 \frac{\sigma_1^2}{\sigma_e^2} I_{q1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_e^2 \frac{\sigma_r^2}{\sigma_e^2} I_{qr} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 I_{q1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r^2 I_{qr} \end{bmatrix}$$

şeklindedir. Böylece y,

$$E(y) = E(X\beta + Zu + \varepsilon) = X\beta + ZE(u) + E(\varepsilon) = X\beta$$

$$Var(y) = Var(X\beta + Zu + \varepsilon) = ZVar(u)Z^T + Var(\varepsilon) = \sigma_e^2 ZDZ^T + \sigma_e^2 I_n$$

$$= \sigma_e^2 \underbrace{(ZDZ^T + I_n)}_{H} = \sigma_e^2 \underbrace{\left(\sum_{i=1}^{r} \gamma_i Z_i Z_i^T + I_n \right)}_{H} = \sigma_e^2 H$$

olmak üzere, $y \sim N(X\beta, \sigma_e^2 H)$ şeklindedir (Zewotir ve Galphin, 2005).

2. Uzun Süreli Model

Bu modeller sıklıkla isimlerini, uzun süreli verilerin analizinin yapılmasıyla kazanırlar (Diggle ve ark., 1996). Bu modellerin bir özelliği gözlemlerin, her bir gruba karşılık gelen bir rastgele etki veya rastgele etkilerin bir vektörü ile bağımsız gruplara bölünebilmesidir. Uygulamada bu gruplar, uzun süreli çalışmadaki farklı bireylere karşılık gelebilir. Dahası, rastgele etkilere ek olarak, her bir grubun içinde

ciddi korelasyonlar olabilir. Uzun süreli modellerin diğer bir özelliği, sıklıkla zamana bağlı açıklayıcı değişkenlerden oluşmalarıdır.

Nobre ve Singer (2007), genel uzun süreli modeli,

$$y_i = X_i\beta + Z_i u_i + \varepsilon_i \qquad i = 1, \cdots, m \tag{3.2}$$

şeklinde ifade etmişlerdir. Burada, y_i, $n_i \times 1$ tipinde gözlemlenebilir yanıt vektörü, β, $p \times 1$ tipinde sabit etkilere ilişkin parametreler vektörü, X_i ve Z_i sırasıyla $n_i \times p$ ve $n_i \times q$ tipinde sabit ve rastgele etkilere ilişkin tasarım matrisleri, u_i, $q \times 1$ tipinde rastgele etkilere ilişkin parametreler vektörü ve ε_i, $n_i \times 1$ tipinde rastgele hataların bir vektörüdür. Genellikle,

$$u_i \sim N_q(0, \sigma^2 G) \; ve \; \varepsilon_i \sim N_{ni}(0, \sigma^2 T_i), \quad i = 1, \cdots, m$$

ve u_i ile ε_i nin bağımsız olduğu varsayımı yapılır. G ve T_i ise sırasıyla $q \times q$ ve $n_i \times n_i$ boyutlu pozitif tanımlı matrislerdir. Bu durum,

$$\begin{bmatrix} u \\ \varepsilon \end{bmatrix} \sim N_{qm+n}\left(\begin{bmatrix} 0_{qm} \\ 0_n \end{bmatrix}, \begin{bmatrix} \sigma^2 D & 0_{qm \times n} \\ 0_{n \times qm} & \sigma^2 \Sigma \end{bmatrix} \right)$$

olarak da ifade edilebilir, öyle ki $n = \sum_{i=1}^{m} n_i$, $D = I_m \otimes G$ ve $\Sigma = \begin{bmatrix} T_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & T_m \end{bmatrix}$ biçimindedir. $\otimes$, Kronecker çarpımı göstermektedir (Searle, 1992).

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}, X = \begin{bmatrix} X_1 \\ \vdots \\ X_m \end{bmatrix}, Z = \begin{bmatrix} Z_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Z_m \end{bmatrix}, u = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

alındığında (3.1) ile verilen model elde edilir.

Belirtilen varsayımlar altında y,

$$E(y) = E(X\beta + Zu + \varepsilon) = X\beta + ZE(u) + E(\varepsilon) = X\beta$$

$$V = Var(y) = Var(X\beta + Zu + \varepsilon) = ZVar(u)Z^T + Var(\varepsilon)$$
$$= \sigma^2 ZDZ^T + \sigma^2 \Sigma = \sigma^2 (ZDZ^T + \Sigma)$$

olmak üzere, $y \sim N\left(X\beta, \sigma^2 (ZDZ^T + \Sigma)\right)$ şeklindedir.

3. Marjinal Model

(3.1) ile verilen lineer karma modelde $\zeta = Zu + \varepsilon$ olarak alındığında,

$$y = X\beta + \zeta$$

marjinal modeli elde edilir (Hilden-Minton, 1995). Burada, $E(\zeta) = 0$ ve $Var(\zeta) = \sigma^2 (ZDZ^T + \Sigma) = V$ ile normal olarak modellenir. Böylece D ve Σ biliniyorsa, marjinal model V^{-1} ağırlıklı matrisi ile sadece lineer bir modeldir. Marjinal modelin bir dezavantajı, rastgele etkilerin açıkça tanımlanmamış olmasıdır.

4. Hiyerarşik Lineer Model

Bayesian düşüncesinden bakıldığında, lineer karma bir model üç aşamalıdır. İlk aşama, verilen rastgele etkilerin gözlemlerinin dağılımı olarak tanımlanır. İkinci aşama, model parametreleri verilmişken rastgele etkilerin dağılımı olarak tanımlanır. Son aşama, parametreler için önsel dağılım varsayımının yapılmasıdır. Normallik varsayımı altında, bu aşamalar belirlenebilir. Son aşamadan da anlaşılabileceği gibi, klasik lineer karma bir model ile Bayesian hiyerarşik model arasındaki tek farklılığın önsel dağılım olduğu söylenebilir (Jiang, 2007). Hiyerarşik lineer model ile ilgili detaylı inceleme 4. bölümde yapılacaktır.

3.1.2. Lineer Karma Modellerde Parametre Tahminleri

Lineer regresyon modellerinde parametre tahmini, EKK ve maksimum likelihood (ML) yöntemleri ile kolaylıkla elde edilebiliyordu. Lineer karma modellerde ise parametre tahminleri daha karmaşıktır, çünkü lineer regresyon modellerinde sadece sabit etkili parametrelerin tahmini incelenirken, lineer karma modellerde buna ek olarak hem rastgele etkili parametrelerin hem de varyans bileşenlerinin tahmini eklenir.

Lineer karma modellerde varyans bileşenlerinin tahmini için kullanılan standart yöntemler, ML ve sınırlandırılmış maksimum likelihood (REML) dir. Her iki yöntem de normallik varsayımı altında büyük örneklem özelliklerine sahiptir. Ancak ML yöntemi istatistiksel olarak daha çok tercih edilir. Çünkü bu yöntem kullanılarak elde edilen tahminler asimptotik olarak etkindir ve asimptotik olarak normal dağılırlar. Aynı zamanda, ML tahmin edicilerinin birçoğu asimptotik olarak yansızdır. ML yöntemiyle tahmin edilecek parametreler β, σ_e^2 ve γ_i dir.

Rastgele etkiler, lineer karma modeldeki rastgele değişkenlere karşılık gelen etkiler olduğu için bu etkilerin tahminlerine *ön tahmin* ve tahmin edicilerine de *ön tahmin edici* denilmektedir (McCulloch ve Searle, 2001). u nun ön tahmininin elde edilmesinde en iyi ön tahmin edici (BP), en iyi doğrusal ön tahmin edici (BLP), en iyi doğrusal yansız ön tahmin edici (BLUP) ve Henderson (1953) tarafından geliştirilen karma model eşitliklerinden elde edilen tahmin ediciler gibi çeşitli ön tahmin ediciler kullanılır. Henderson nın karma model model eşitliklerinin elde edilmesi için, normallik varsayımı altında u ve y nin ortak olasılık yoğunluk fonksiyonu aynı zamanda u için verilen ML fonksiyonudur.

Bu bölümde, (3.1) modeli için, sırasıyla, γ_i ler (ve böylece H) biliniyorken ve bilinmiyorken β, σ_e^2 nin ML tahminleri ile u un ML ön tahmini incelenecektir. Bununla birlikte, ML yönteminin uygulanmasıyla elde edilen sonuçlar istenirse kolaylıkla REML yöntemine de uygulanabilir.

3.1.2.1. H Biliniyorken β ve σ_e^2 nin ML Tahminleri ve u nun ML Ön Tahmini

Hartley ve Rao (1967) tarafından önerilen ML yönteminin kullanılabilmesi için tasarım matrislerinin bazı varsayımları sağlaması gerekmektedir. X ve $i = 1, \cdots, r$ için Z_i tasarım matrislerinin tam ranklı $(rank(X) = p, rank(Z_i) = q_i)$ olduğu varsayılır. Genellikle varyans analizinde her bir y, i-inci rastgele etkenin bir düzeyi ile ilişkilidir. Bu nedenle, Z_i tasarım matrisinin her satırının sıfırdan farklı ilk elemanı 1, diğer $q_i - 1$ elemanı sıfır değerini alacaktır. Hartley ve Rao (1967), Z_i nin bu özelliği sağladığını varsayarlar. Z_i nin bu özelliği sağlaması, ${Z_i}^T Z_i$ nin $q_i \times q_i$ boyutlu köşegen bir matris olduğunu gösterir.

y yanıt vektörünün normal dağıldığı varsayımı altında, olasılık yoğunluk fonksiyonu,

$$f(y) = (2\pi)^{-n/2}(\sigma_e^2)^{-n/2}|H|^{-1/2} exp\left[\frac{-(y - X\beta)^T H^{-1}(y - X\beta)}{2\sigma_e^2}\right] \qquad (3.3)$$

şeklindedir. Burada $|H|$, H matrisinin determinantıdır. Bu fonksiyonun logaritması daha sonra da β ve σ_e^2 ye göre türevi alınıp, sıfıra eşitlenir.

$$\begin{aligned} L(\beta, \sigma_e^2; y) &= \log f(y) \\ &= -\frac{1}{2} nlog(2\pi) - \frac{1}{2} nlog\sigma_e^2 - \frac{1}{2} log|H| \\ &\quad -\frac{1}{2\sigma_e^2}(y - X\beta)^T H^{-1}(y - X\beta) \end{aligned} \qquad (3.4)$$

(3.4) eşitliğinden,

$$L_\beta = \frac{\partial L(\beta, \sigma_e^2; y)}{\partial \beta} = \frac{1}{\sigma_e^2}(X^T H^{-1} y - X^T H^{-1} X\beta) = 0 \qquad (3.5)$$

$$L_{\sigma_e^2} = \frac{\partial L(\beta, \sigma_e^2; y)}{\partial \sigma_e^2} = -\frac{n}{2\sigma_e^2} + \frac{1}{2\sigma_e^4}(y - X\beta)^T H^{-1}(y - X\beta) = 0 \qquad (3.6)$$

elde edilir. Devam edilirse,

$$X^T H^{-1} X\hat{\beta} = X^T H^{-1} y$$

$$\hat{\beta} = (X^T H^{-1} X)^{-1} X^T H^{-1} y = W(X,X)^{-1} W(X,y)$$

$$\hat{\sigma}_e^2 = \left(y - X\hat{\beta}\right)^T H^{-1}\left(y - X\hat{\beta}\right)/n = W(d,d)/n$$

bulunur. Burada, sırasıyla, $d = y - X\hat{\beta}$ ve $W(A,B) = A^T H^{-1} B$ şeklindedir (Hemmerle ve Hartley, 1973) ve $\hat{\beta}$ ya, β nın en iyi yansız tahmin edicisi (BLUE) denir.

u ve y nin ortak olasılık yoğunluk fonksiyonu,

$$f(u;y) = f(y|u)f(u)$$
$$= \frac{exp\left\{-\frac{1}{2\sigma_e^2}[(y - X\beta - Zu)^T I^{-1}(y - X\beta - Zu) + u^T D^{-1} u]\right\}}{(2\pi)^{\frac{n+q}{2}}(\sigma_e^2)^{\frac{1}{2}}|D|^{\frac{1}{2}}}$$

şeklindedir. Bu ifadeyi maksimum yapan parametre tahmininin bulunması için, önce ifadenin logaritmasının alınması gerekir.

$$L(u;y) = log\{f(y|u)f(u)\}$$
$$= \left\{-\frac{1}{2\sigma_e^2}[(y - X\beta - Zu)^T(y - X\beta - Zu) + u^T D^{-1} u]\right\}$$
$$- log\left((2\pi)^{\frac{n+q}{2}}(\sigma_e^2)^{\frac{1}{2}}|D|^{\frac{1}{2}}\right)$$

Daha sonra, u ya göre türevin alınıp, bu türevin sıfıra eşitlenmesi ile,

$$\frac{\partial L(u;y)}{\partial u} = -Z^T y + Z^T X\beta + Z^T Zu + D^{-1} u = 0$$

$$\hat{u} = (D^{-1} + Z^T Z)^{-1} Z^T \left(y - X\hat{\beta}\right)$$

elde edilir. $(D^{-1} + Z^T Z)^{-1}$ ifadesi $(D^{-1} + Z^T IZ)^{-1}$ şeklinde yazılırsa matris tersi, Schur Tümleyenleri yöntemi (Bkz. Ek 3) kullanılarak,

$$(D^{-1} + Z^T IZ)^{-1} = D - DZ^T (I + ZDZ^T)^{-1} ZD = D - DZ^T H^{-1} ZD$$

ifadesi bulunur. $\hat{u}$ yeniden düzenlenirse,

$$\hat{u} = (D - DZ^T H^{-1} ZD) Z^T \left(y - X\hat{\beta}\right) = \left[DZ^T - DZ^T H^{-1} \underbrace{ZDZ^T}_{(H-I)}\right] \left(y - X\hat{\beta}\right)$$
$$= DZ^T H^{-1} \left(y - X\hat{\beta}\right)$$

olarak elde edilir. $\hat{u}$, u nun en iyi yansız ön tahmin edicisidir (BLUP).

3.1.2.2. H Bilinmiyorken β ve σ_e^2 nin ML Tahminleri ve u nun ML Ön Tahmini

γ_i ler (ve böylece H) bilinmiyorsa, γ_i lerin ML tahminleri, D ve H matrisinde yerine konularak, β ve σ_e^2 nin ML tahminleri elde edilir. Bu nedenle, γ_i lerin ML tahminlerinin bulunması gerekir. Eşitlik (3.4) ün γ_i ye göre türevinin alınması için önce H, $log|H|$ ve H^{-1} in γ_i ye göre türevi alınmalıdır. Bu türevler,

$$\frac{\partial H}{\partial \gamma_i} = Z_i {Z_i}^T, i = 1, \cdots, r$$

$$\frac{\partial log|H|}{\partial \gamma_i} = trace\left(H^{-1} \frac{\partial H}{\partial \gamma_i}\right)$$

ve

$$\frac{\partial H^{-1}}{\partial \gamma_i} = -H^{-1}\frac{\partial H}{\partial \gamma_i}H^{-1}$$

olarak elde edilir. Buna göre, Eşitlik (3.4) deki ifadenin γ_i lere göre türevi,

$$\frac{\partial L(\beta, \sigma_e^2; y)}{\partial \gamma_i} = -\frac{1}{2}trace\left(Z_i{}^T H^{-1} Z_i\right) + \frac{1}{2\sigma_e^2}(y - X\beta)^T H^{-1} Z_i Z_i{}^T H^{-1}(y - X\beta) \quad (3.7)$$

olarak bulunur. Eşitlik (3.5) in, Eşitlik (3.6) nın ve Eşitlik (3.7) nin sıfıra eşitlenmesiyle,

$$X^T \tilde{H}^{-1} X\tilde{\beta} = X^T \tilde{H}^{-1} y$$

$$\tilde{\sigma}_e^2 = \left(y - X\tilde{\beta}\right)^T \tilde{H}^{-1}\left(y - X\tilde{\beta}\right)/n$$

ve $i = 0,1,\cdots,r$ için

$$trace\left(Z_i{}^T \tilde{H}^{-1} Z_i\right) = \frac{1}{\tilde{\sigma}_e^2}\left(y - X\tilde{\beta}\right)^T \tilde{H}^{-1} Z_i Z_i{}^T \tilde{H}^{-1}\left(y - X\tilde{\beta}\right)$$

elde edilir. Bu eşitlikler yardımıyla, β ve σ_e^2 nin ML tahminleri bulunur. Ancak ML eşitlikleri, parametrelere göre doğrusal olmadığından, varyans bileşenleri tahminlerini elde etmek için Newton-Raphson algoritması, Skorlama yöntemi algoritması ve W dönüşümü kullanılır (Türkan, 2008).

γ_i ler (ve böylece H) bilinmiyorsa, γ_i lerin ML tahminleri, D ve H matrisinde yerine konularak, u nun ML ön tahmini aşağıdaki gibi elde edilir.

$$\tilde{u} = \tilde{D} Z^T \tilde{H}^{-1}\left(y - X\tilde{\beta}\right)$$

3.1.3. Lineer Karma Modellerde Güncelleme Formülleri

Bu bölümde, (3.1) modeli altında lineer karma modellerin güncelleme formülleri incelenecektir. Lineer karma modellerin analizinde, parametre tahminlerini elde etmek için kullanılan ML yöntemi sapan, leverage ve etkili noktalara karşı oldukça duyarlıdır. Bu nedenle, lineer karma modellerde analiz sonuçlarını önemli ölçüde etkileyen bu noktaların ortaya çıkarılması oldukça önemlidir. Regresyon analizinde olduğu gibi lineer karma modellerde de bu tür noktaların ortaya çıkarılması için bazı ölçütler önerilmiştir. Önerilen bu ölçütler, regresyon modelinde kullanılan ölçütlerin genişletilmiş halidir. Lineer regresyon modelinde önerilen ölçütler, lineer karma modeller için genişletilirken güncelleme formüllerinden yararlanılır. Güncelleme formülleri kullanılarak, lineer karma modellerde sabit etki parametrelerinin, rastgele etki parametrelerinin ML tahminlerinde ve varyans bileşenlerinin tahminlerini elde etmek için kullanılan iteratif yöntemlerde işlem sayısı azaltılır.

Christensen ve ark. (1992), veri kümesinden çıkarılan i-inci gözlemin etkisini araştırmak için kullanılan gözlem silme tanılamalarının türetilmesinde güncelleme formüllerini kullanmışlardır. Bu formüller türetilirken, i-inci gözlemin çıkarıldığı parçalanmış matrislerin yazılması uygundur. Parçalanmış matrisler $i = 1$ olduğu varsayılarak yazılmıştır ancak elde edilen sonuçlar bütün i değerleri için uygulanabilir (Christensen ve ark., 1992).

Zewotir ve Galphin (2005) de, lineer karma modellerde güncelleme formüllerini türetmek için çıkartılan i-inci gözlemi, matrisin ilk satırıymış gibi düşünerek $(i = 1)$ matrisleri parçalamışlardır. Bu durum,

$$X = \begin{bmatrix} x_i^T \\ X_{(i)} \end{bmatrix}, Z_j = \begin{bmatrix} z_{ji}^T \\ Z_{j(i)} \end{bmatrix}, Z = \begin{bmatrix} z_i^T \\ Z_{(i)} \end{bmatrix}, y = \begin{bmatrix} y_i \\ \mathrm{y}_{(i)} \end{bmatrix}, H = \begin{bmatrix} h_{ii} & h_i^T \\ h_i & H_{[i]} \end{bmatrix} \qquad (3.8)$$

olarak ifade edilebilir. (3.8) eşitliğinde yer alan $X_{(i)}, Z_{(i)}, \mathrm{y}_{(i)}$ sırasıyla, i-inci satırı çıkartıldıktan sonra elde edilen X, Z matrislerini ve y vektörünü, x_i, z_i, h_i, y_i sırasıyla, X, Z, H matrislerinin ve y vektörünün i-inci satırını, $H_{[i]}$ ise i-inci satır ve i-inci sütun

çıkartıldıktan sonra elde edilen H matrisini göstermektedir. Zewotir ve Galphin (2005), geliştirdikleri ölçütler için Christensen ve ark. (1992) nın önerdiği gözlem silme istatistiklerinin temel yapı bloklarına benzer tanımlamalar yapmışlardır. Bunlar,

$$A = \begin{bmatrix} {a_i}^T \\ A_{(i)} \end{bmatrix}, H^{-1} = \begin{bmatrix} 1/m_i & -{h_i}^T H_{[i]}^{-1}/m_i \\ -H_{[i]}^{-1}h_i/m_i & H_{[i]}^{-1} + H_{[i]}^{-1}h_i{h_i}^T H_{[i]}^{-1}/m_i \end{bmatrix}, B = \begin{bmatrix} {b_i}^T \\ B_{(i)} \end{bmatrix}$$

olmak üzere,

$$\check{a}_i = \mathrm{a}_i - A_{(i)}^T H_{[i]}^{-1} h_i$$

$$\check{b}_i^T = {b_i}^T - {h_i}^T H_{[i]}^{-1} B_{(i)}$$

$$p_{ii} = \check{x}_i^T W(X,X)^{-1} \check{x}_i$$

$$m_i = \mathrm{h}_{ii} - h_i^T H_{[i]}^{-1} h_i$$

$$\check{z}_{ji} = z_{ji} - Z_{j(i)}^T H_{[i]}^{-1} h_i$$

$$W_{[i]}(A,B) = A_{(i)}^T H_{[i]}^{-1} B_{(i)}$$

şeklinde ifade edilir. Buradaki H^{-1}, H parçalanmış matrisinin tersi yardımıyla elde edilir.

i-inci gözlem silindiğinde $\beta, \sigma_e^2, u, \gamma$ nın tahminleri olan $\hat{\beta}_{(i)}, \hat{\sigma}_{e(i)}^2, \hat{u}_{(i)}, \hat{\gamma}_{(i)}$ gösterilmek istensin. Bunun için, öncelikle, bazı teoremlerin incelenmesi gerekmektedir.

Teorem: (Zewotir ve Galphin, 2005) Sabit (veya bilinen) varyans bileşen oranları için,

$$W_{[i]}(A,B) = W(A,B) - \frac{\check{a}_i \check{b}_i^T}{m_i}$$

dir.

İspat. $W(A,B) = A^T H^{-1} B$ şeklinde ifade edilmişti (Hemmerle ve Hartley, 1973).

$$W(A,B) = \begin{bmatrix} a_i & A_{(i)}^T \end{bmatrix} \begin{bmatrix} 1/m_i & -h_i{}^T H_{[i]}^{-1}/m_i \\ -H_{[i]}^{-1} h_i/m_i & H_{[i]}^{-1} + H_{[i]}^{-1} h_i h_i{}^T H_{[i]}^{-1}/m_i \end{bmatrix} \begin{bmatrix} b_i{}^T \\ B_{(i)} \end{bmatrix}$$

$$= \frac{a_i b_i{}^T}{m_i} - \frac{A_{(i)}^T H_{[i]}^{-1} h_i b_i{}^T}{m_i} - \frac{a_i h_i{}^T H_{[i]}^{-1} B_{(i)}}{m_i} + A_{(i)}^T H_{[i]}^{-1} B_{(i)} + \frac{A_{(i)}^T H_{[i]}^{-1} h_i h_i{}^T H_{[i]}^{-1} B_{(i)}}{m_i}$$

$$= A_{(i)}^T H_{[i]}^{-1} B_{(i)} + a_i \left(b_i{}^T - h_i{}^T H_{[i]}^{-1} B_{(i)}\right)/m_i - A_{(i)}^T H_{[i]}^{-1} h_i \left(b_i{}^T - h_i{}^T H_{[i]}^{-1} B_{(i)}\right)/m_i$$

$$= A_{(i)}^T H_{[i]}^{-1} B_{(i)} + \left(a_i - A_{(i)}^T H_{[i]}^{-1} h_i\right)\left(b_i{}^T - h_i{}^T H_{[i]}^{-1} B_{(i)}\right)/m_i$$

$$= A_{(i)}^T H_{[i]}^{-1} B_{(i)} + \frac{\check{a}_i \check{b}_i^T}{m_i}$$

$$= W_{[i]}(A,B) + \frac{\check{a}_i \check{b}_i^T}{m_i}$$

son eşitliğin dikkate alınmasıyla ispat tamamlanır.

$W_{[i]}(A,B)$ nin tersi Sherman-Morrison-Woodbury Teoremi (Bkz. Ek 3) yardımıyla,

$$W_{[i]}(A,B)^{-1} = \left[W(A,B) - \frac{\check{a}_i \check{b}_i^T}{m_i}\right]^{-1}$$

$$W_{[i]}(A,B)^{-1} = W(A,B)^{-1} + \frac{W(A,B)^{-1}\check{a}_i\check{b}_i^T W(A,B)^{-1}}{m_i - \check{b}_i^T W(A,B)^{-1}\check{a}_i} \tag{3.9}$$

olarak bulunur. $\hat{\beta} = (X^T H^{-1} X)^{-1} X^T H^{-1} y$ eşitliğinden yararlanılarak,

$$\hat{\beta}_{(i)} = \left(X_{(i)}{}^T H_{[i]}^{-1} X_{(i)}\right)^{-1} X_{(i)}{}^T H_{[i]}^{-1} y_{(i)} = W_{[i]}(X,X)^{-1} W_{[i]}(X,y) \tag{3.10}$$

yazılır. Bu eşitlikte, (3.9) ile verilen Sherman-Morrison-Woodbury Teoremi (Bkz. Ek 3) kullanıldığında,

$$\left(X_{(i)}{}^T H_{[i]}^{-1} X_{(i)}\right)^{-1} = W(X,X)^{-1} + \frac{W(X,X)^{-1}\check{x}_i\check{x}_i^T W(X,X)^{-1}}{(m_i - \check{x}_i^T W(X,X)^{-1}\check{x}_i)} \tag{3.11}$$

elde edilir. Zewotir ve Galphin (2005) tarafından geliştirilen ölçütlerden yararlanıldığında $\check{x}_i = x_i - X_{(i)}^T H_{[i]}^{-1} h_i$ ve $\check{x}_i{}^T = x_i{}^T - h_i{}^T H_{[i]}^{-1} X_{(i)}$ olduğu açıkça görülür. (3.11) eşitliği ve,

$$W_{[i]}(X,y) = X_{(i)}^T H_{[i]}^{-1} y_{(i)} = W(X,y) - \frac{\check{x}_i\check{y}_i}{m_i} = (X^T H^{-1} y) - \frac{\check{x}_i\check{y}_i}{m_i}$$

(3.10) eşitliğinde yerine yazılıp, gerekli işlemler yapıldığında,

$$\hat{\beta}_{(i)} = \hat{\beta} - \frac{(X^T H^{-1} X)^{-1}\check{x}_i\left(\check{y}_i - \check{x}_i^T\hat{\beta}\right)}{m_i - p_{ii}} \tag{3.12}$$

olarak bulunur. Benzer şekilde, Zewotir ve Galphin (2005) tarafından geliştirilen ölçütlerden yararlanıldığında $\check{y}_i = y_i - y_{(i)}^T H_{[i]}^{-1} h_i$ ve $\check{z}_i = z_i - Z_{(i)}^T H_{[i]}^{-1} h_i$ şeklinde elde edilir.

$\hat{u} = DZ^T H^{-1}\left(y - X\hat{\beta}\right)$ eşitliğinden yararlanarak,

$$\hat{u}_{(i)} = DZ_{(i)}^T H_{[i]}^{-1} y_{(i)} - DZ_{(i)}^T H_{[i]}^{-1} X_{(i)} \hat{\beta}_{(i)} \tag{3.13}$$

yazılır. (3.13) eşitliğinde,

$$W_{[i]}(Z,y) = Z_{(i)}^T H_{[i]}^{-1} y_{(i)} = W(Z,y) - \frac{\check{z}_i \check{y}_i}{m_i} = (Z^T H^{-1} y) - \frac{\check{z}_i \check{y}_i}{m_i}$$

$$W_{[i]}(Z,X) = Z_{(i)}^T H_{[i]}^{-1} X_{(i)} = W(Z,X) - \frac{\check{z}_i \check{x}_i^T}{m_i} = (Z^T H^{-1} X) - \frac{\check{z}_i \check{x}_i^T}{m_i}$$

ve

$$\hat{\beta}_{(i)} = \hat{\beta} - \frac{(X^T H^{-1} X)^{-1} \check{x}_i \left(\check{y}_i - \check{x}_i^T \hat{\beta}\right)}{m_i - p_{ii}}$$

yerine yazılıp, gerekli işlemler yapıldığında,

$$\hat{u}_{(i)} = \hat{u} - D[\check{z}_i - (Z^T H^{-1} X)(X^T H^{-1} X)^{-1} \check{x}_i] \frac{\left(\check{y}_i - \check{x}_i^T \hat{\beta}\right)}{(m_i - p_{ii})} \tag{3.14}$$

bulunur.

$\hat{\sigma}_e^2 = \left(y - X\hat{\beta}\right)^T H^{-1} \left(y - X\hat{\beta}\right)/n$ ifadesi üzerinde biraz oynama yapıldığında,

$$n\hat{\sigma}_e^2 = \left[y^T H^{-1} y - y^T H^{-1} X\hat{\beta} - \hat{\beta}^T X^T H^{-1} y + \hat{\beta}^T X^T H^{-1} X\hat{\beta}\right]$$

$$n\hat{\sigma}_e^2 = y^T H^{-1} y - y^T H^{-1} X\hat{\beta} - \hat{\beta}^T X^T H^{-1} y + \hat{\beta}^T X^T H^{-1} y$$

$$n\hat{\sigma}_e^2 = y^T H^{-1} y - y^T H^{-1} X\hat{\beta}$$

olarak elde edilir. i-inci gözlem çıkarıldıktan sonra,

$$(n-1)\hat{\sigma}^2_{e(i)} = y^T_{(i)}H^{-1}_{[i]}y_{(i)} - y^T_{(i)}H^{-1}_{[i]}X_{(i)}\hat{\beta}_{(i)} \tag{3.15}$$

olur. (3.15) eşitliğinde,

$$W_{[i]}(y,y) = y^T_{(i)}H^{-1}_{[i]}y_{(i)} = W(y,y) - \frac{\check{y}_i\check{y}_i}{m_i} = (y^TH^{-1}y) - \frac{\check{y}_i\check{y}_i}{m_i}$$

$$W_{[i]}(y,X) = y^T_{(i)}H^{-1}_{[i]}X_{(i)} = W(y,X) - \frac{\check{y}_i\check{x}^T_i}{m_i} = (y^TH^{-1}X) - \frac{\check{y}_i\check{x}^T_i}{m_i}$$

ve

$$\hat{\beta}_{(i)} = \hat{\beta} - \frac{(X^TH^{-1}X)^{-1}\check{x}_i\left(\check{y}_i - \check{x}^T_i\hat{\beta}\right)}{m_i - p_{ii}}$$

yerine yazılıp, gerekli işlemler yapıldığında,

$$\hat{\sigma}^2_{e(i)} = \frac{1}{(n-1)}\left(n\hat{\sigma}^2_e - \frac{1}{m_i - p_{ii}}\left(\check{y}_i - \check{x}^T_i\hat{\beta}\right)^2\right) \tag{3.16}$$

elde edilir.

Lineer karma modelde, varyans bileşenleri oranlarının tahmininin iteratif olarak elde edilmesinden dolayı, i-inci gözlem çıkarıldığında varyans bileşenleri oranlarındaki değişim için basit analitik gösterimler elde edilemez. Ancak eğer $\hat{\gamma}$ başlangıç değeri olarak kullanılırsa ve i-inci gözlem çıkarılarak bir adım iterasyon uygulanırsa, $\hat{\gamma}$ daki değişimler için analitik gösterimler türetilebilir. Parametreleri doğrusal olmayan modeller için genelde bu bir adım ölçütleri i-inci gözlem çıkarıldığında parametrelerdeki iteratif değişimleri kestirir (Pregibon, 1981).

Cox ve Hinkley (1974), Newton-Raphson algoritması ve Fisher skorlama algoritması yönteminde eğer tutarlı bir tahminle başlanırsa, bir adımda asimptotik olarak etkili tahminlerin elde edilebileceğini söylemişlerdir. γ nın ML tahmini $\hat{\gamma}$

tutarlı bir tahmindir ve bir adımda, i-inci gözlem çıkarıldıktan sonra γ nın etkili tahminlerinin elde edilmesini sağlar (Hartley ve Rao, 1967).

Newton-Raphson ve Fisher skorlama algoritmalarının kullanılmasıyla, varyans bileşenleri oranları için bir adım güncelleme formülü,

$$\hat{\gamma}_{(i)} = \hat{\gamma} - (Q - \mathcal{G})^{-1} g_{(i)}$$

olarak elde edilir. Burada, $g_{(i)}$, ML fonksiyonun i-inci gözlem çıkarıldıktan sonra γ ya göre türevi Q, ML fonksiyonunun $\beta, \sigma_e^2, \gamma$ ya göre ikinci türevlerinden oluşan Hessian matrisinin (F) negatif beklenen değeri yani bilgi matrisi ve $j, k = 1, \cdots, r$ için,

$$Q(j,k) = E\big(F(i,j)\big)$$

$$\mathcal{G} = \mathcal{G}(j,k) = \frac{1}{m_i}\left\{\frac{1}{2m_i} ssq\left(\check{z}_{ji}\check{z}_{ki}^T\right) - \check{z}_{ji}^T W\left(Z_j, Z_k\right)\check{z}_{ji}\right\}$$

şeklindedir. Tahminlerin varyansını içeren matrise ***asimptotik yayılım matrisi (asymtotic dispersion matrix)*** denir ve bu matris, bilgi matrisinin tersine eşittir. Dolayısıyla, $Var(\hat{\gamma}) = -Q^{-1}$ şeklindedir. $ssq\left(\check{z}_{ji}\check{z}_{ki}^T\right)$ ise, $\check{z}_{ji}\check{z}_{ki}^T$ nın elemanlarının kareler toplamı anlamına gelmektedir (Zewotir ve Galphin, 2005).

$\check{x}_i, \check{z}_i, \check{z}_{ji}\check{y}_i, p_{ii}, m_i$ lerin hepsi H ve $H_{[i]}^{-1}$ in satırının (veya sütununun) bir fonksiyonudur. Fakat tam model için iteratif tahmin prosedüründe $n \times n$ boyutlu H matrisinin açıkça hesaplanması gerekli değildir. Dahası, $(n-1) \times (n-1)$ boyutlu olan $H_{[i]}$ nin ve tersinin her $i = 1, \cdots, n$ için hesaplanması kolay değildir. Bu nedenle, Christensen ve ark. (1992), bu problemin üstesinden gelmek amacıyla,

$$H_{[i]}^{-1} = C_{[i]} - \frac{1}{c_{ii}} c_i c_i^T \quad , \quad C = H^{-1} = \begin{bmatrix} c_{ii} & c_i^T \\ c_i & C_{[i]} \end{bmatrix}$$

tanımlamışlardır. Bu tanımlamada, sırasıyla silinen her n gözlem için, $H_{[i]}^{-1}$ nin hesaplanmasında hala $\check{x}_i, \check{z}_i, \check{z}_{ji}\check{y}_i, p_{ii}, m_i$ gözlem silme istatistikleri mevcuttur. Bu gözlem silme istatistiklerinin tekrardan hesaplanması aşırı bir zaman kaybına neden olmaktadır. Bu nedenle, problemin çözümüne en kısa zamanda ulaşılması amacıyla başka temel yapı blokları tanımlanmıştır.

C_i, H^{-1} in i-inci sütunu ve c_{ii}, H^{-1} in i-inci köşegen elemanı olmak üzere, $C_i^T = [c_{ii} \quad c_i^T]$ olarak alınırsa,

$$C = H^{-1} = \begin{bmatrix} c_{ii} & c_i^T \\ c_i & C_{[i]} \end{bmatrix} = \begin{bmatrix} 1/m_i & -h_i{}^T H_{[i]}^{-1}/m_i \\ -H_{[i]}^{-1}h_i/m_i & H_{[i]}^{-1} + H_{[i]}^{-1}h_i h_i{}^T H_{[i]}^{-1}/m_i \end{bmatrix}$$

eşitliğinden,

$$m_i = \frac{1}{c_{ii}}$$

şeklinde direkt olarak ifade edilebilir. Dahası, $c_i = -H_{[i]}^{-1}h_i/m_i$, $c_i/c_{ii} = -H_{[i]}^{-1}h_i$ dır. Dolayısıyla,

$$\check{x}_i = x_i - X_{(i)}^T H_{[i]}^{-1}h_i = x_i + \frac{1}{c_{ii}}X_{(i)}^T c_i = \frac{1}{c_{ii}}\left(c_{ii}x_i + X_{(i)}^T c_i\right)$$
$$= \frac{1}{c_{ii}}\begin{bmatrix} x_i & X_{(i)}^T \end{bmatrix}\begin{bmatrix} c_{ii} \\ c_i \end{bmatrix} = \frac{1}{c_{ii}}X^T C_i$$

şeklindedir. Benzer şekilde,

$$\check{z}_{ji} = z_{ji} - Z_{j(i)}^T H_{[i]}^{-1}h_i = \frac{1}{c_{ii}}Z_j^T C_i \quad , \quad \check{y}_i = y_i - y_{(i)}^T H_{[i]}^{-1}h_i = \frac{1}{c_{ii}}y^T C_i$$

elde edilir (Zewotir ve Galphin, 2005).

3.2. Lineer Karma Modellerde Rezidü Analizi

Klasik lineer regresyonda rezidü analizi, etkilerin lineerliği, normallik, sabit varyanslılık ve hataların bağımsızlığı gibi model varsayımlarının kontrol edilmesinde ve ek olarak, sapan değerlerin belirlenmesinde kullanılır. Klasik lineer regresyonda tanımlanan bu rezidü analizi, lineer karma modellere de genişletilebilir. Lineer karma modellerde birbirine denk olmayan çok sayıda rezidü tanımlanabilir ve bu rezidülerin çoğu, modeldeki hatanın kaynağı ile ilgili olmayan bozulmalara maruz kalabilir. Bu bölümde, lineer karma modellerdeki bozulmuş rezidü tipleri ve sapan değerlerin belirlenmesinde kullanılacak yöntemler incelenecektir.

3.2.1. Lineer Karma Modellerde Rezidü Tipleri

(3.2) ile verilen uzun süreli lineer karma model altında hatanın üç tipi,

- Marjinal hata, $\zeta = y - E[y] = y - X\beta$,
- Rastgele etkili hata, $Zu = E[y|u] - E[y]$ ve
- Tek gözlemli veya koşullu hata, $\varepsilon = y - E[y|u] = y - X\beta - Zu$

olarak tanımlanır. Bu tanımlamalarda sadece y, X ve Z bilinir veya gözlemlenir fakat hatalar, modelin içinde tanımlanan yapılardır ve modelden ayrı olduklarında direkt olarak gözlemlenemezler. Böylece, incelenen modelin onaylanması için gözlemlenmiş veriden alındığı varsayılan bu rastgele nicelikler tahmin edilmeye çalışılır. İdeal olarak, görülemeyen özel bir hata için bir rezidü tanımlanmak istenir. Bu rezidü, sadece görülemeyen hataya ve bilinen sabit niceliklere bağlı olan bir fonksiyondur. Böyle bir fonksiyon ***saf bir rezidü*** olarak tanımlanır. Diğer yandan, bir fonksiyon başka varsayımlı hatalara bağlı ise ***bozulmuş bir rezidü*** olarak tanımlanır. Klasik lineer regresyonda sadece bir hata modellenir. Sonuç olarak, klasik lineer model için tanımlanan tüm rezidülerin saf olduğu söylenebilir. Bu çalışmada hata, tüm gözlemler için rastgele bir vektör olarak düşünülür. Bununla birlikte, her gözlem kendi ilişkisiz hatasına sahip olursa, her skaler rezidü diğer gözlemlerin hataları

tarafından bozulmuş yapılacaktır. Bu durum, Cook ve Weisberg (1982) tarafından ***süpernormallik*** olarak ifade edilir. Yine de, lineer karma modeller için vektörler olarak tanımlanan rezidülerin çoğu bozulmuş bir yapıya sahip olacaktır.

(3.2) ile verilen uzun süreli lineer karma model altında tanımlanan üç hata tipine karşılık gelen bozulmuş rezidü tipleri,sırasıyla,

- Marjinal rezidü $\hat{\zeta} = y - X\hat{\beta}$,
- Rastgele etkili rezidü veya EBLUP, $Z\hat{u}$ ve
- Tek gözlemli veya koşullu rezidü, $e = y - X\hat{\beta} - Z\hat{u}$

şeklindedir. Burada, $\hat{\zeta} = Z\hat{u} + e$ olduğuna dikkat edilmelidir.

3.2.1.1. Marjinal Rezidüler

Bölüm 3.1.1.1 altında incelenen uzun süreli model için $S = \sigma^2 V^{-1} = (ZDZ^T + \Sigma)^{-1}$, $\mathrm{Q} = S - SX(X^T SX)^{-1}X^T S$ olmak üzere, $\hat{\beta} = (X^T SX)^{-1}X^T Sy$ ve $\hat{u} = DZ^T \mathrm{Q} y$ eşitliklerinin kullanılmasıyla marjinal rezidü,

$$\hat{\zeta} = y - X\hat{\beta} = S^{-1}\mathrm{Q}y$$

olarak bulunur. Şimdi de, D ve Σ nin bilindiği varsayımı altında, marjinal rezidünün varyansı elde edilecektir. Q, simetrik bir matris ve $\mathrm{Q}V\mathrm{Q} = \sigma^2 \mathrm{Q}\underbrace{S^{-1}\mathrm{Q}}_{\mathrm{Q}} = \sigma^2\mathrm{Q}$ olduğundan marjinal rezidünün varyansı,

$$Var(\hat{\zeta}) = Var(S^{-1}\mathrm{Q}y) = S^{-1}\mathrm{Q}Var(y)\mathrm{Q}S^{-1} = S^{-1}\underbrace{\mathrm{Q}V\mathrm{Q}}_{\sigma^2\mathrm{Q}}S^{-1} = \sigma^2 S^{-1}\mathrm{Q}S^{-1}$$
$$= \sigma^2[ZDZ^T + \Sigma - X(X^T SX)^{-1}X^T]$$

şeklinde elde edilir.

Açıkça görülüyor ki $\hat{\zeta}$, ζ marjinal hatası için saf bir rezidü olarak tanımlanır. Çünkü,

$$\begin{aligned}\hat{\zeta} - \zeta &= y - X\hat{\beta} - \zeta \\ &= X\beta + Zu + \varepsilon - X(X^TSX)^{-1}X^TS(X\beta + Zu + \varepsilon) - (Zu + \varepsilon) \\ &= -X(X^TSX)^{-1}X^TSZu - X(X^TSX)^{-1}X^TS\varepsilon \\ &= -X(X^TSX)^{-1}X^TS\underbrace{(Zu + \varepsilon)}_{\zeta} \\ &= -X(X^TSX)^{-1}X^TS\zeta\end{aligned}$$

eşitliğinin sağ kısmı sadece görülemeyen hataya ve bilinen sabit niceliklere bağlıdır.

$y = X\beta + \zeta$ marjinalliği verildiğinde, X deki açıklayıcı değişkenlere karşılık $\hat{\zeta} = \left(\hat{\zeta}_1^T, \dots, \hat{\zeta}_m^T\right)^T$ marjinal rezidülerin vektörünün elemanlarının grafiği, marjinal rezidüler standart lineer modeldeki rezidülerle aynı anlama sahip olduğunda, y nin lineerliğini kontrol etmek için kullanılabilir. Lineer ilişki sürdüğünde sıfır etrafında rastgele ilişki beklenir.

ζ, $V = \sigma^2(ZDZ^T + \Sigma)$ kovaryans yapısına sahip olduğu için, V kovaryans yapısının kontrol edilmesinde $\hat{\zeta}$ kullanılabilir. Yani, marjinal rezidüler aynı zamanda $i = 1, \cdots, m$ için $V_i = \sigma^2(Z_iDZ_i^T + T_i)$ gibi denekler içindeki kovaryans yapısının geçerliliğini kontrol etmek için de kullanılabilirler. Bu yönden Lesaffre ve Verbeke (1998) $\left\|I_{ni} - T_iT_i^T\right\|^2$ ifadesini, denekler içindeki kovaryans matrisi için uygun rezidü olarak düşünmüşlerdir. Burada, $\hat{V}_i$ denekler içindeki kovaryans yapısının tahminini göstermek üzere $T_i = \hat{V}_i^{-1/2}\hat{\zeta}_i^T$ şeklindedir ve $\|\cdot\|$, Frobenius uzunluğunu göstermektedir (Bkz. Ek 4). i-inci denek için ortalamaların vektörü $X_i\beta$ tarafından doğru modellendiğinde $var(y_i)$, $\hat{\zeta}_i\hat{\zeta}_i^T$ ile tahmin edilebilir. Bunun bir sonucu olarak, $\left\|I_{ni} - T_iT_i^T\right\|^2$ ifadesinin sıfıra yakın olması beklenir ve deneklere karşı $\left\|I_{ni} - T_iT_i^T\right\|^2$ değerlerinin bir grafiği, modele iyi uyum sağlamayan kovaryans yapısı varsayımı altındaki gözlemlerin belirlenmesine yardımcı olabilir.

3.2.1.2. EBLUP

Bölüm 3.1.1.1 altında incelenen uzun süreli model için $S = \sigma^2 V^{-1} = (ZDZ^T + \Sigma)^{-1}$, $\mathrm{q} = S - SX(X^T SX)^{-1}X^T S$ olmak üzere, $\hat{\beta} = (X^T SX)^{-1}X^T Sy$ ve $\hat{u} = DZ^T \mathrm{q} y$ eşitliklerinin kullanılmasıyla EBLUP,

$$Z\hat{u} = ZDZ^T \mathrm{q} y$$

olarak bulunur. Şimdi de, D ve Σ nin bilindiği varsayımı altında, EBLUP varyansı elde edilecektir. q, simetrik bir matris ve $\mathrm{q}V\mathrm{q} = \sigma^2 \mathrm{q}\underbrace{S^{-1}\mathrm{q}}_{\mathrm{q}} = \sigma^2 \mathrm{q}$ olduğundan EBLUP varyansı,

$$Var(Z\hat{u}) = Var(ZDZ^T \mathrm{q} y) = ZDZ^T \mathrm{q} Var(y) \mathrm{q} ZDZ^T = ZDZ^T \underbrace{\mathrm{q}V\mathrm{q}}_{\sigma^2 \mathrm{q}} ZDZ^T$$
$$= \sigma^2 ZDZ^T \mathrm{q} ZDZ^T$$

şeklinde elde edilir. Bununla birlikte,

$$\begin{aligned} Z\hat{u} - Zu &= ZDZ^T \mathrm{q} y - Zu = ZDZ^T \mathrm{q}(X\beta + Zu + \varepsilon) - Zu \\ &= ZDZ^T \underbrace{\mathrm{q} X}_{0} \beta + ZDZ^T \mathrm{q}\, Zu + ZDZ^T \mathrm{q}\, \varepsilon - Zu \\ &= [ZDZ^T \mathrm{q} - I]Zu + ZDZ^T \mathrm{q}\, \varepsilon \end{aligned} \qquad (3.17)$$

eşitliğinde $Z\hat{u}$, u için ε un olmasından dolayı bozulmuş bir rezidü olarak tanımlanır ve $Z^T \mathrm{q} = 0$ olmadığı sürece, Z nin sütunları X in kolon uzayında olur.

$Zu = E[y|u] - E[y]$ olarak ifade edildiği için $Z_i \hat{u}_i$, merkezden uzakta duran deneklerin belirlenmesinde kullanılabilir (Pinheiro ve Bates, 2000; Waternaux ve ark., 1989; Verbeke ve Lesaffre, 1996a; Longford, 2001). Örneğin, Pinheiro ve Bates (2000), denek indekslerine karşı $\hat{u}_i$ nın elemanlarının grafiğini önermişlerdir. Böyle bir grafik, BLUP ın j-inci elemanına karşılık gelen sapan deneklerin belirlenmesinde kullanılabilir. $\hat{u}_i,\ i = 1, \cdots, m$ değerleri sadece Z_i nin kovaryansları tüm deneklerde

aynı olduğu zaman karşılaştırılabilir (Verbeke ve Lesaffre, 1996a). Waternaux ve ark. (1989) tarafından merkezden uzakta duran deneklerin belirlenmesi için önerilen denek indekslerine karşı $\hat{\xi}_i = \hat{u}_i{}^T var[\hat{u}_i - u_i]\hat{u}_i$ olarak ifade edilen Mahalanobis uzaklığı (Bkz. Ek 1) veya $Z_i\hat{u}_i$ nın elemanlarının grafiği kullanılabilir.

Deneklerin rastgele etkilerin kovaryans matrislerinin homojenliğine karşı duyarlı olup olmadığının incelenmesi için Pinheiro ve Bates (2000), rastgele etkilerin ön tahminlerinin matris saçılım grafiğini önermişlerdir.

EBLUP aynı zamanda rastgele etkilerin normallik varsayımının kontrol edilmesinde de kullanılabilir. Lange ve Ryan (1989), böyle öneriler için rastgele etkilerin standartlaştırılmış lineer kombinasyonlarının ağırlıklandırılmış normal Q_Q grafiklerini önermişlerdir.

3.2.1.3. Koşullu Rezidüler

Bölüm 3.1.1.1 altında incelenen uzun süreli model için $S = \sigma^2 V^{-1} = (ZDZ^T + \Sigma)^{-1}$, $\mathrm{Q} = S - SX(X^TSX)^{-1}X^TS$ olmak üzere, $\hat{\beta} = (X^TSX)^{-1}X^TSy$ ve $\hat{u} = DZ^T\mathrm{Q}y$ eşitliklerinin kullanılmasıyla koşullu rezidü,

$$e = y - X\hat{\beta} - Z\hat{u} = y - X\hat{\beta} - ZDZ^T\mathrm{Q}y = S^{-1}\mathrm{Q}y - ZDZ^T\mathrm{Q}y$$
$$= (ZDZ^T + \Sigma - ZDZ^T)\mathrm{Q}y = \Sigma\mathrm{Q}y$$

olarak bulunur. Şimdi de, D ve Σ nin bilindiği varsayımı altında, koşullu rezidünün varyansı elde edilecektir. Q, simetrik bir matris ve $\mathrm{Q}V\mathrm{Q} = \sigma^2\,\mathrm{Q}\underbrace{S^{-1}\mathrm{Q}}_{\mathrm{Q}} = \sigma^2\mathrm{Q}$ olduğundan koşullu rezidünün varyansı,

$$Var(e) = Var(\Sigma\mathrm{Q}y) = \Sigma\mathrm{Q}Var(y)\mathrm{Q}\Sigma = \Sigma\,\underbrace{\mathrm{Q}V\mathrm{Q}}_{\sigma^2\mathrm{Q}}\Sigma = \sigma^2\,\Sigma\mathrm{Q}\Sigma$$

şeklinde elde edilir. Bununla birlikte,

$$e - \varepsilon = \Sigma^{-1}y - \varepsilon = \Sigma^{-1}(X\beta + Zu + \varepsilon) - \varepsilon = \Sigma^{-1}\underbrace{X}_{0}\beta + \Sigma^{-1}Zu + \Sigma^{-1}\varepsilon - \varepsilon$$

$$= [\Sigma^{-1} - I]\,\varepsilon + \Sigma^{-1}Zu \qquad (3.18)$$

eşitliğinde de e, ε için u un olmasından dolayı bozulmuş bir rezidü olarak tanımlanır.

Bozulmuş bir rezidüye dikkat edilmesinin nedeni nedir? $Z\hat{u}$ ve e rezidüleri, Zu ve ε hataları için yansız ön tahmin ediciler olmalarına rağmen bu durum, koşullu olarak doğru değildir. (3.17) ve (3.18) eşitlikleri açıkça gösteriyor ki $Z\hat{u}|\varepsilon$, $ZDZ^T\Sigma^{-1}\varepsilon$ ile yanlı iken $e|u$, $\Sigma^{-1}Zu$ ile yanlıdır. e bozulmuş rezidüsü incelenirken ε koşullu hatasının normalliğinin kontrol edilmek istendiği varsayılsın. Zu büyük ölçüde normallikten uzaksa, ε gerçekte normal olsa bile e normal olmayabilir. Böyle bir durumda, koşullu hataların dağılımı yeniden modellenebilir. Diğer yandan, Zu büyük ölçüde normal ise, ε un normal olmama durumunu gizleyebilir. Bu durum, bozulma nedeniyle kaybedilen güç miktarını gösterir. Bununla birlikte e, normallik varsayımını değiştirme durumu üzerinde ciddi kuşkulara neden olan bozulmuş bir rezidüdür.

Açıkça e koşullu rezidüsünün değerlendirilmesinde ε koşullu hatasının normal olmama durumunun belirlenme gücü, $\Sigma^{-1}\varepsilon$ un varyansı ile ilişkili olan $\Sigma^{-1}Zu$ nun varyansının artmasından dolayı azalır. Bu ifadenin anlaşılabilmesi için (3.18) eşitliği ile elde edilen $e = \Sigma^{-1}Zu + \Sigma^{-1}\varepsilon$ ye bakılması yeterlidir. Böylece, $Var(\Sigma^{-1}Zu) = \sigma^2\Sigma^{-1}ZDZ^T\Sigma^{-1}$ ile $Var(\Sigma^{-1}\varepsilon) = \sigma^2\Sigma^{-1}\Sigma^{-1}\Sigma$ nın karşılaştırılması önerilir (Hilden-Minton, 1995). Bu iki varyansın tam ranklı olmamaları nedeniyle determinantlarının birbirine oranlanması mümkün değildir. Dolayısıyla, bu iki varyansın köşegen elemanlarına veya bazı fonksiyonlarına karşılık gelen oranlarına bakılabilir. Örneğin, i-inci koşullu rezidü olan e_i için ***bozulma oranı***,

$$CF_{e_i} = \frac{Var[e_i|u]}{Var[e_i]} = \frac{k_i^T\Sigma^{-1}ZDZ^T\Sigma^{-1}k_i}{k_i^T\Sigma^{-1}\Sigma\,k_i} = 1 - \frac{k_i^T\Sigma^{-1}\Sigma\Sigma^{-1}k_i}{k_i^T\Sigma^{-1}\Sigma\,k_i} \qquad (3.19)$$

olarak tanımlanır (Hilden-Minton, 1995). Burada k_i, $n \times n$ boyutlu birim matrisin i-inci sütunu olup, $e_i = k_i^T e$ şeklindedir. Bu tanım herhangi bir bozulmuş rezidüye genişletilebilir.

(3.19) eşitliği ile tanımlanan bozulma oranı 0 ile 1 arasında bir değer alır ve 0 değeri, e_i nin bozulmamış bir rezidü olduğu anlamına gelir. Bu oran aynı zamanda saf bir rezidünün formülleştirilmiş bir şeklidir. Bozulma oranlarının uygulanabilir değeri, rezidü oranında meydana gelen değişmenin hataya katkıda bulunmadığını gösterir. Basit olarak bu oranlar, $Var(\Sigma Zu)$ nun $Var(\Sigma\varepsilon)$ u ne kadar engellediğinin bir göstergesidir.

Rezidülerde ortaya çıkan bozulma problemleri her ne olursa olsun yine de, koşullu rezidü vektörü olan e incelenmek istenebilir. Bozulma etkisinin minimum yapılması için, CF_{e_i} bozulma oranının küçük olmasını sağlayacak e_i değerine bakılabilir. Ancak bu durumda, üç problem ortaya çıkar. Birincisi, CF_{e_i}, nasıl küçük yapılabilir? İkincisi, e_i göz ardı edilirse ne kaybedilir? Üçüncüsü ise, bu yaklaşım bazı yanlı seçimlere neden olabilir mi? şeklindedir. Bütün bu problemlerden kaçınılması için e nin daha az bozulmaya sahip Le lineer dönüşümü yapılabilir. Örneğin, L nin satırları olan l_i için, $l_i^T e$ nin en küçük bozulmaya sahip olması $l_i^T e$ nin,

$$\lambda_i = \frac{l_i^T \Sigma \Sigma l_i}{l_i^T \Sigma l_i} \qquad (3.20)$$

eşitliğinde maksimum olması anlamına gelir. Burada, $Var[l_i^T e] \propto l_i^T \Sigma \Sigma l_i > 0$ dır.

Sonuç olarak (3.20) eşitliğinin tekil değer ayrışımı (Bkz. Ek 7) yardımıyla maksimum yapılmasını sağlayan $l_i^T e$ değerlerine ***en küçük bozulmuş rezidüler***, $\frac{l_i^T e}{\sigma}$ değerlerine ise ***standartlaştırılmış en küçük bozulmuş rezidüler*** adı verilir (Hilden-Minton, 1995). Nobre ve Singer (2007), koşullu hataların normallik varsayımının kontrol edilmesinde ($\Sigma = I_n$ alınarak), 95% lik güven aralığı ile standartlaştırılmış en küçük bozulmuş rezidülerin Q_Q grafiğini önermişlerdir.

m_{ii}, $Var(e) = \Sigma Q \Sigma$ nin i-inci esas köşegen elemanı olmak üzere, e nin elemanları σ^2 ve m_{ii} ye bağlı olan farklı varyanslara sahip olabilir. Pek çok tanılama prosedürü için, ölçeğe bağlı olmayan rezidülerin studentized versiyonlarını tanımlamak ve kendi aralarında karşılaştırmak kullanışlı olabilir. Örneğin, Nobre ve Singer (2007) tarafından i-inci standartlaştırılmış koşullu rezidü,

$$e_i^* = \frac{e_i}{\hat{\sigma}\sqrt{\hat{m}_{ii}}}$$

olarak tanımlanır. Burada e_i, i-inci koşullu rezidü ve $\hat{m}_{ii}$, m_{ii} nin tahminini göstermektedir.

Nobre ve Singer (2007), koşullu hataların sabit varyanslılığının kontrol edilmesinde ($\Sigma = I_n$ alınarak) uydurulmuş değerlere karşı i-inci standartlaştırılmış koşullu rezidü olan e_i^* nin grafiğini önermişlerdir. Aynı zamanda gözlem indekslerine karşı e_i^* nin grafiğinin kullanılmasıyla merkezden uzakta duran gözlemler belirlenebilir. Benzer öneriler Pinheiro ve Bates (2000), Weiss ve Lazaro (1992), Oman (1995) tarafından da verilmiştir.

Varyans bileşenleri modeli altında koşullu rezidüler, PRESS rezidüler ve studentized rezidüler olmak üzere iki grupta incelenebilir. Bu rezidü tiplerinin daha iyi anlaşılabilmesi için öncelikle iki teorem üzerinde durulacaktır.

Teorem: (Zewotir ve Galphin, 2007) Lineer karma modelde koşullu rezidü vektörü olan $e = y - X\hat{\beta} - Z\hat{u}$, ε nun en iyi lineer yansız ön tahmin edicisidir.

İspat: Sabit ve rastgele etkilerin lineer bir kombinasyonu olan $b^T\beta + a^T u$ için $b^T\beta$ tahmin edilebilir ise, en iyi lineer yansız tahmin edici $b^T\hat{\beta} + a^T\hat{u}$ (Harville, 1977, 1990; McLean ve ark., 1991; Robinson, 1991; Searle ve ark., 1992) olur. Böylece, y verildiğinde,

$$BLUP(\varepsilon) = BLUP(y - X\beta - Zu) = y - X\hat{\beta} - Z\hat{u} = Ry$$

ve

$$
\begin{aligned}
e = y - \hat{y} &= y - \left(X\hat{\beta} + Z\hat{u}\right) = y - X\beta - ZDZ^T H^{-1}\left(y - X\hat{\beta}\right) \\
&= (I_n - ZDZ^T H^{-1})\left(y - X\hat{\beta}\right) = (I_n - (H - I)H^{-1})\left(y - X\hat{\beta}\right) \\
&= H^{-1}\left(y - X\hat{\beta}\right) = H^{-1}y - H^{-1}X(X^T H^{-1}X)^{-1}X^T H^{-1}y \\
&= \underbrace{(H^{-1} - H^{-1}X(X^T H^{-1}X)^{-1}X^T H^{-1})}_{R} = Ry \qquad (3.21)
\end{aligned}
$$

şeklinde elde edilir. (3.21) eşitliğindeki R, simetrik bir matristir ve gözlem değerlerini rezidülere dönüştürür. Rezidüler, rastgele hata terimlerinin gerçekleşen değerleridir. Buna göre, i-inci rezidü e_i ile i-inci rastgele hata terimi ε_i arasındaki kovaryans ve korelasyon,

$$
\begin{aligned}
Cov(e, \varepsilon) = \; & Cov(Ry, \varepsilon) = Cov(R(X\beta + Zu + \varepsilon), \varepsilon) \\
&= E(R(X\beta + Zu + \varepsilon)\varepsilon) - E\left(R(X\beta + Zu + \varepsilon)\right)E(\varepsilon) = E(R\varepsilon^2) \\
&= R\sigma_e^2
\end{aligned}
$$

$$
Corr\left(e_i\ , \varepsilon_j\right) = \frac{Cov\left(e_i\ , \varepsilon_j\right)}{\sqrt{Var(e_i)Var\left(\varepsilon_j\right)}} = \frac{r_{ij}}{\sqrt{r_{ii}}}
$$

$$
Corr(e_i\ , \varepsilon_i) = \pm\sqrt{r_{ii}}
$$

olarak elde edilir. Burada, $0 \leq r_{ii} \leq 1$ dir. Bu, e ile ε arasındaki ilişkinin R matrisine bağlı olduğunu gösterir. Dahası, $e = Ry$, $e \sim N(0, R\sigma_e^2)$ olur. Böylece,

a. $e_i \sim N(0, \sigma_e^2 r_{ii})$

b. $Cov\left(e_i, e_j\right) = \sigma_e^2 r_{ij}$

c. $Corr\left(e_i\ , e_j\right) = \frac{r_{ij}}{\sqrt{r_{ii}}\sqrt{r_{jj}}}$

şeklinde elde edilir. Eğer R matrisinin köşegen elemanları 1 e yakınsa, e koşullu rezidü vektörü yerine ε hata vektörü kullanılabilir (Zewotir ve Galphin, 2007). i-inci rezidü olan $e_i = r_{ii}y_i + \sum_{j\neq i}^{n} r_{ij}\, y_i$ şeklinde yazılabilir.

Teorem: (Zewotir ve Galphin, 2007) $r_{ii} \to 0$ veya 1 ise $i \neq j$ için $r_{ij} \to 0$ dır.

İspat: $R_i = \begin{bmatrix} r_{i1} \\ \vdots \\ r_{in} \end{bmatrix}$ olmak üzere, $R_i^T(Z^TDZ)R_i = R_i^T(H - I)R_i \geq 0$ dır. Çünkü D, i-inci elemanı $\gamma_i I_{qi}$ olan blok köşegen bir matristir ve varyans bileşen oranları $\gamma_i = \frac{\sigma_i^2}{\sigma_e^2}$, $\sigma_i^2 \geq 0$, $\sigma_e^2 > 0$ dır.

$$R^THR = (H^{-1} - H^{-1}X(X^TH^{-1}X)^{-1}X^TH^{-1})H(H^{-1} - H^{-1}X(X^TH^{-1}X)^{-1}X^TH^{-1})$$
$$= R$$

$\left(r_{ii} - ssq(R_i)\right) \geq 0$ $ssq(R_i)$, R_i nin elemanlarının kareler toplamını gösterir. O halde,

$$r_{ii} - \sum_{j=1}^{n} r_{ij}^2 \geq 0 \quad , \quad r_{ii} - r_{ii}^2 - \sum_{j\neq i}^{n} r_{ij}^2 \geq 0$$

$$r_{ii}(1 - r_{ii}) \geq \sum_{j\neq i}^{n} r_{ij}^2$$

olur. Bu durumda, $r_{ii} \to 0$ veya 1 iken $\sum_{j\neq i}^{n} r_{ij}^2 \leq 0$ gibi bir durum, herhangi bir ifadenin karesi sıfırdan küçük olamayacağından dolayı, mümkün değildir. Dolayısıyla, $i \neq j$ için $r_{ij} \to 0$ olur.

1. PRESS Rezidüler

i-inci PRESS rezidü $e_{(i)}$, tüm veriden i-inci gözlemin silinmesiyle,

$$
\begin{aligned}
e_{(i)} &= y_i - x_i^T\hat{\beta}_{(i)} - z_i^T\hat{u}_{(i)} \\
&= y_i - x_i^T\hat{\beta} + x_i^T W(X,X)^{-1}X^T C_i \frac{e_i}{r_{ii}} - z_i^T\hat{u} + z_i^T DZ^T R_i \frac{e_i}{r_{ii}} \\
&= \underbrace{y_i - x_i^T\hat{\beta} - z_i^T\hat{u}}_{e_i} + \left(x_i^T W(X,X)^{-1}X^T C_i + \underbrace{z_i^T DZ^T R_i}_{1-x_i^T W(X,X)^{-1}X^T C_i - r_{ii}} \right)\frac{e_i}{r_{ii}} \\
&= \frac{e_i}{r_{ii}}
\end{aligned}
$$

olarak elde edilir. Burada, $1 - x_i^T \underbrace{W(X,X)^{-1}X^T C_i}_{[\hat{\beta}-\hat{\beta}_{(i)}]\frac{r_{ii}}{e_i}} - r_{ii} = z_i^T DZ^T R_i$ şeklindedir. Bu eşitliğin gösterilmesi,

$$
\begin{aligned}
1 - x_i^T\left[\hat{\beta} - \hat{\beta}_{(i)}\right]\frac{r_{ii}}{e_i} - r_{ii} &= 1 - \left[x_i^T\hat{\beta} - x_i^T\hat{\beta}_{(i)}\right]\frac{r_{ii}}{e_i} - r_{ii} \\
&= 1 - \left[\left(y_i - x_i^T\hat{\beta}_{(i)}\right) - \left(y_i - x_i^T\hat{\beta}\right)\right]\frac{r_{ii}}{e_i} - r_{ii} \\
&= 1 - \left[\left(e_{(i)} + z_i^T\hat{u}_{(i)}\right) - \left(e_i + z_i^T\hat{u}\right)\right]\frac{r_{ii}}{e_i} - r_{ii} \\
&= 1 - \left[\left(e_{(i)} - e_i\right) - z_i^T\left(\hat{u} - \hat{u}_{(i)}\right)\right]\frac{r_{ii}}{e_i} - r_{ii} \\
&= 1 - \left[(1 - r_{ii})\frac{e_i}{r_{ii}} - z_i^T DZ^T R_i \frac{e_i}{r_{ii}}\right]\frac{r_{ii}}{e_i} - r_{ii} = z_i^T DZ^T R_i
\end{aligned}
$$

şeklindedir ($W(X,X)^{-1}X^T C_i = \left[\hat{\beta} - \hat{\beta}_{(i)}\right]\frac{r_{ii}}{e_i}$ için Ek 5 ve Ek 6 incelenebilir). PRESS rezidünün varyansı,

$$
Var\left(e_{(i)}\right) = Var\left(\frac{e_i}{r_{ii}}\right) = \frac{1}{r_{ii}^2}Var(e_i) = \frac{1}{r_{ii}^2}[\sigma_e^2 r_{ii}] = \frac{\sigma_e^2}{r_{ii}}
$$

olup, $e_{(i)} \sim N\left(0, \frac{\sigma_e^2}{r_{ii}}\right)$ şeklindedir ve e_i ile aynı korelasyon yapısına sahiptir. $e_{(i)}$ ön tahmin hatalarına ***PRESS Rezidüler*** denir (Zewotir ve Galphin, 2007). PRESS rezidülerin kullanımı, küçük r_{ii} değerlerinin elde edilmesini sağlar.

2. Studentized Rezidüler

Studentized rezidüler kendi aralarında ikiye ayrılır. Bunlardan ilki, i-inci iç studentized rezidü, ikincisi ise i-inci dış studentized rezidü olup, sırasıyla,

$$t_i = \frac{e_i}{\sqrt{var(e_i)}} = \frac{e_i}{\hat{\sigma}_e \sqrt{r_{ii}}} \tag{3.22}$$

ve

$$t_i^* = \frac{e_i}{\hat{\sigma}_{e(i)}\sqrt{r_{ii}}} \tag{3.23}$$

şeklinde tanımlanırlar (Zewotir ve Galphin, 2007).

Model varsayımlarından sapmalar, uzak duran gözlemlerde daha belirgindir. Bu nedenle, klasik rezidülerle bu varsayımları belirlemek zordur. Studentized rezidü işte bu amaçla kullanır. Küçük r_{ii} ve küçük rezidüye sahip herhangi bir nokta EKK uyumda etkin gözlem ise, studentized rezidünün incelenmesi önerilir.

Teorem: (Zewotir ve Galphin, 2007) Bilinen γ için $t_i^{*2} \sim \frac{n-1}{n-p-1} F(1, n-p-1)$ dir.

İspat: (3.23) ile verilen i-inci dış studentized rezidünün kullanılmasıyla,

$$t_i^* = \frac{e_i}{\hat{\sigma}_e \sqrt{\frac{n-t_i^2}{n-1}}\sqrt{r_{ii}}}$$

$$t_i^{*2} = \frac{{e_i}^2}{\hat{\sigma}_e^2 \left(\frac{n-t_i^2}{n-1}\right) r_{ii}}$$

$$\frac{t_i^{*2}}{n-1} = \frac{{e_i}^2/\hat{\sigma}_e^2 r_{ii}}{n - {e_i}^2/\hat{\sigma}_e^2 r_{ii}} = \frac{{e_i}^2}{n\hat{\sigma}_e^2 r_{ii} - {e_i}^2} = \frac{{e_i}^2/r_{ii}}{n\hat{\sigma}_e^2 - {e_i}^2/r_{ii}}$$

olur. $e = Ry$ olduğundan $e_i = R_i^T y$ dir. Ve,

$$\hat{\sigma}_e^2 = e^T H^{-1} e/n$$

$$n\hat{\sigma}_e^2 = e^T H^{-1} e$$

$$n\hat{\sigma}_e^2 = y^T \underbrace{R^T H^{-1} R}_{R} y = y^T R y$$

dir. Dolayısıyla,

$$\frac{t_i^{*2}}{n-1} = \frac{y^T \frac{R_i R_i^T}{r_{ii}} y}{y^T \left(R - \frac{R_i R_i^T}{r_{ii}}\right) y}$$

yazılabilir. $y \sim N(X\beta, \sigma_e^2 H)$ varsayımı altında ve

$$\frac{1}{2}\beta^T X^T \frac{R_i R_i^T}{r_{ii}\sigma_e^2} X\beta = 0 \;\; X^T R = X^T (H^{-1} - H^{-1}X(X^T H^{-1} X)^{-1} X^T H^{-1}) = 0$$

$$\left(\frac{R_i R_i^T}{r_{ii}\sigma_e^2}\sigma_e^2 H\right)\left(\frac{R_i R_i^T}{r_{ii}\sigma_e^2}\sigma_e^2 H\right) = \frac{R_i R_i^T}{r_{ii}} H, \qquad R_i^T H R_i = r_{ii}$$

olduğundan (Bkz. Ek 8),

$$\frac{y^T \frac{R_i R_i^T}{r_{ii}} y}{\sigma_e^2} \sim \mathcal{X}^2(1)$$

olarak elde edilir. Benzer şekilde,

$$\frac{y^T\left(R-\frac{R_iR_i^T}{r_{ii}}\right)y}{\sigma_e^2}\sim\mathcal{X}^2(n-p-1)$$

olup,

$$\left(R-\frac{R_iR_i^T}{r_{ii}}\right)\frac{1}{\sigma_e^2}\sigma_e^2H\frac{R_iR_i^T}{r_{ii}\sigma_e^2}=\frac{\overbrace{RH}^{I}R_iR_i^T}{r_{ii}\sigma_e^2}-\frac{R_i\overbrace{R_i^THR_i}^{r_{ii}}R_i^T}{r_{ii}^2\sigma_e^2}=0$$

eşitliği yardımıyla,

$$\left(R-\frac{R_iR_i^T}{r_{ii}}\right)\frac{1}{\sigma_e^2}H\frac{R_iR_i^T}{r_{ii}}=0$$

elde edilir ve

$$t_i^{*2}\sim\frac{n-1}{n-p-1}F(1,n-p-1)$$

dir (Gerekli yerlerde Ek 8 incelenebilir).

Hurtado (1993), t_i^{*2} nin bilinen γ ve γ nın ML tahmini ile aynı limit dağılımına sahip olduğunu göstermiştir. Böylece,

$$t_i^{*2}=\frac{n-1}{n-p-1}F(1,n-p-1)=\frac{n-1}{n-p-1}\left(\frac{\mathcal{X}^2(1)/1}{\frac{\mathcal{X}^2(n-p-1)}{n-p-1}}\right)$$

$$=\frac{\mathcal{X}^2(1)}{\frac{\mathcal{X}^2(n-p-1)}{n-1}}$$

ve

$$\frac{\mathcal{X}^2(n-p-1)}{n-1} \xrightarrow{p} 1 \, , t_i^{*2} \xrightarrow{d} \mathcal{X}^2(1)$$

olur.

$$t_i^* = t_i \sqrt{\frac{n-1}{n-t_i^2}}$$

$$t_i^{*2} = \frac{t_i^2(n-1)}{n-t_i^2}$$

$$nt_i^{*2} - t_i^{*2}t_i^2 = t_i^2 n - t_i^2$$

$$nt_i^{*2} = t_i^2\left(n-1+t_i^{*2}\right)$$

$$\frac{nt_i^{*2}}{\left(n-1+t_i^{*2}\right)} = t_i^2$$

$$\frac{n}{1+\frac{n-1}{t_i^{*2}}} = t_i^2$$

olup, t_i^{*2} ile t_i^2 ölçütleri arasındaki ilişki gösterilmiş olur.

PRESS rezidünün studentized versiyonu,

$$\frac{e_{(i)}}{\sqrt{Var\left(e_{(i)}\right)}} = \frac{e_i / r_{ii}}{\sqrt{\hat{\sigma}_e^2 / r_{ii}}} = \frac{e_i}{\hat{\sigma}_e \sqrt{r_{ii}}}$$

olup, (3.22) ile verilen i-inci iç studentized rezidü versiyonunun aynısıdır.

Her bir rezidü tipinin kullanımı Çizelge 3.1. ile özetlenmektedir.

Çizelge 3.1. Tanılama önerileri için rezidülerin kullanımı (Nobre, 2007)

Tanılama İçin	Rezidü Tipi	Grafik
Etkilerin Lineerliği(Xβ)	Marjinal	Açıklayıcı değişkenlere karşı $\hat{\zeta}_k$
Denekler İçindeki Kovaryans Matrisi ($\boldsymbol{V_i}$)	Marjinal	Denek indekslerine karşı $\left\|I_{ni} - T_i T_i^T\right\|^2$
Merkezden Uzakta Duran Deneklerin Varlığı	EBLUP	Denek indekslerine karşı $\hat{\xi}_i$
Rastgele Etkilerin Normalliği ($\boldsymbol{u_i}$)	EBLUP	$\hat{u}_i$ için ağırlıklandırılmış Q_Q
Merkezden Uzakta Duran Gözlemlerin Varlığı	Koşullu	Gözlem indekslerine karşı e_i^*
Koşullu Hataların Sabit Varyanslılığı ($\boldsymbol{\varepsilon_i}$)	Koşullu	Uydurulmuş değerlere karşı e_i^*
Koşullu Hataların Normalliği ($\boldsymbol{\varepsilon_i}$)	Koşullu	Standartlaştırılmış en küçük bozulmuş rezidü için Q_Q

3.2.2. Lineer Karma Modellerde Sapan Değerlerin Belirlenmesi

(3.1) modeli altında sapan değerlerin belirlenebilmesi için iki temel yoldan yararlanılacaktır. Bunlar, sırasıyla, bir sapan değer için test ve grafiksel metotlardır.

3.2.2.1. Sapan Değerlerin Belirlenmesi İçin Bir Test

Burada tek bir sapan değerin varlığının belirlenebilmesi için bir test incelenecektir. Sapan değere aday olan gözlem bilinmediğinde, formal test prosedürü genellikle t_i nin maksimum değerine dayalıdır. t_i nin maksimum değerine dayalı test istatistiğinin yaklaşık olarak bulunması zor olduğu için $|t_i|$ nin maksimum değeri kullanılır. Bir α düzeyinde, $max|t_i|$ için kritik c değeri

$$P(max|t_i| > c) = \alpha$$

nın çözülmesiyle bulunur. Bonferroni eşitsizliğinden (Galphin ve Hawkins, 1981; Barnett ve Lewis, 1995),

$$\alpha = P(max|t_i| > c) \leq \sum_{i=1}^{n} P\left(|t_i| > c\right) = \sum_{i=1}^{n} P\left({t_i}^2 > c^2\right)$$

$$= \sum_{i=1}^{n} P\left(\frac{n}{1 + \frac{n-1}{t_i^{*2}}} > c^2\right) = \sum_{i=1}^{n} P\left(t_i^{*2} > \frac{(n-1)c^2}{n - c^2}\right)$$

$$= n\left[P\left\{F(1, n-p-1) > \frac{(n-p-1)c^2}{n - c^2}\right\}\right]$$

elde edilir. O halde,

$$P\left\{F(1, n-p-1) > \frac{(n-p-1)c^2}{n - c^2}\right\} \geq \frac{\alpha}{n}$$

$$\frac{(n-p-1)c^2}{n - c^2} > F\left(1 - \frac{\alpha}{n}; 1, n-p-1\right)$$

$$c^2 > \frac{nF\left(1 - \frac{\alpha}{n}; 1, n-p-1\right)}{F\left(1 - \frac{\alpha}{n}; 1, n-p-1\right) + (n-p-1)}$$

$$c > \sqrt{\frac{n\left\{t\left(1 - \frac{\alpha}{n}; n-p-1\right)\right\}^2}{\left\{t\left(1 - \frac{\alpha}{n}; n-p-1\right)\right\}^2 + (n-p-1)}}$$

şeklindedir. Burada, $t\left(1 - \frac{\alpha}{n}; n-p-1\right)$, $n-p-1$ serbestlik derecesi ile student t dağılımının alt $\left(1 - \frac{\alpha}{n}\right)$ düzeyini gösterir.

$$max|t_i| > \sqrt{\frac{n\left\{t\left(1-\frac{\alpha}{n};n-p-1\right)\right\}^2}{\left\{t\left(1-\frac{\alpha}{n};n-p-1\right)\right\}^2+(n-p-1)}} \tag{3.24}$$

ise i-inci gözlem olası bir sapan değer olarak tanımlanacaktır. Büyük n değerleri için, γ_i ler bilinmiyorsa ve ML tahminleri kullanılıyorsa, (3.24) eşitliğinin limit dağılımı kullanılarak,

$$max|t_i| > \sqrt{\mathcal{X}^2\left(1-\frac{\alpha}{n};1\right)} \tag{3.25}$$

elde edilir ve (3.25) eşitliği sağlanıyorsa i-inci gözlem, olası bir sapan değer olarak tanımlanacaktır.

Yukarıda da bahsedildiği gibi, buradaki test çoğunlukla verilmiş veri setinde bir tane sapan değer olduğunu ve onun da bilinmediğini varsayar. Eğer birden fazla sapan değer varsa sabit kritik bir değer yerine birbiriyle ilişkili rezidülerin karşılaştırılması önerilir. Bu da en iyi rezidülerin grafiksel gösterimleri kullanılarak başarılabilir.

3.2.2.2. Sapan Değerlerin Belirlenmesi İçin Grafiksel Metotlar

Rezidülerin grafiksel gösterimleri sıklıkla formal test prosedürlerinden daha kullanışlıdır. Bu grafikler sadece sapan değerlerin varlığıyla ilgili bilgi sağlamayıp aynı zamanda modelin ve/veya onunla ilgili varsayımların geçerliliği ile ilgili de bilgi sağlarlar. Rezidülerin şekilleri, bozukluğun tipi hakkında bilgilendiricidir. Histogramlar, dal-yaprak gösterimleri, nokta grafikleri ve kutu grafikleri modelin uygunluğunun değerlendirilmesi için rezidülere bakmanın kolay ve hızlı yollarıdır. Sapan gözlem sıklıkla grafikten ayrı bir yerde durarak kendini açıkça belli eder.

Daha önce de bahsedildiği gibi, küçük r_{ii} değerine sahip gözlemler küçük rezidülere sahip olur ve rezidü grafiklerinde tespit edilemezler. Bu durumda, R nin

tüm köşegen elemanları olan r_{ii} lerin kutu grafikleri, dal-yaprak gösterimleri gibi grafiksel gösterimlerle en iyi sonuç elde edilir.

Tanılama ölçüsü olarak sadece r_{ii} yi kullanmanın dezavantajı, y nin içerdiği bilgiyi göz ardı etmesidir. Çünkü R, sadece X ve Z (H aracılığıyla) nin bir fonksiyonudur. Küçük r_{ii} değeri, X ve Z tasarım uzaylarındaki i. noktanın uyumda, özellikle de uydurulmuş değer üzerinde, çok önemli bir etkiye sahip olduğunu gösterir. Diğer bir deyişle, küçük r_{ii} değerine sahip olan noktaların sabit ve rastgele etkili tasarım uzaylarındaki yüksek leverage noktalar olduğu söylenebilir.

Rezidü ve leverage değerlerinin birlikte gösterilmesinde etkili olan serpilme grafiği, $e_i^2/e^T e$ ye karşı r_{ii} nin grafiğidir. Burada, $0 \leq r_{ii} \leq 1$ ve aynı zamanda $0 \leq e_i^2/e^T e \leq 1$ dir.

Grafiklerde studentized rezidüler, klasik rezidülere tercih edilir. Çünkü, studentized rezidüleri kullanarak oluşturulan grafiklerin şekilleri, sabit terim içermeyen rezidü varyansı tarafından karıştırılmayacak ve klasik rezidülerden daha aydınlatıcı olacaktır.

3.3. Lineer Karma Modellerde Leverage Analizi

Bu bölümde, (3.2) ile verilen uzun süreli model altında lineer karma modeller için leverage analizi incelenecektir. Uydurulmuş değerler üzerinde büyük bir etkiye sahip olan birimlerin veya gözlemlerin yani yüksek leverage noktaların belirlenmesi üzerinde durulacaktır. Genelleştirilmiş ortak leverage matrisin, lineer karma modellerdeki etkili birimlerin ve gözlemlerin belirlenmesinde önemli bir rol oynadığı düşünülür. Genelleştirilmiş ortak leverage matris, kendi arasında genelleştirilmiş marjinal leverage matrisi $(\hat{y} = X\hat{\beta})$ ve koşullu uydurulmuş değerlerin rastgele etki bileşeni için genelleştirilmiş leverage matrisi $(\hat{y}^* = X\hat{\beta} + Z\hat{u})$ olmak üzere iki parçaya ayrılır. Genelleştirilmiş ortak leverage matrisin bu iki farklı ayrışımı Nobre (2004) ve Demidenko ve Stukel (2005) tarafından birbirinden bağımsız ve yüksek leverage noktaların belirlenmesinde bir kriter olarak düşünülmüştür.

y_i gözlenmişse ve $\hat{y}_i$, $y_i = \mu(\beta) + e_i$, $i = 1, \cdots, n$ modeli altında i-inci birim için uydurulmuş değerlere karşılık geliyorsa, y_i nin leverage değeri $\partial \hat{y}_i / \partial y_i$ ***kısmi türevi*** olarak tanımlanır ve i-inci gözlenmiş değere karşılık i-inci uydurulmuş değerin değişiminin oranını yansıtır.

(3.2) ile verilen uzun süreli lineer karma model altında, V nin bilindiği varsayımı ile Wei ve ark. (1998) incelenirse, ***marjinal uydurulmuş değerler için genelleştirilmiş leverage matris*****,**

$$L_1 = \frac{\partial \hat{y}}{\partial y^T} = \frac{\partial X(X^T S X)^{-1} X^T S y}{\partial y^T} = X(X^T S X)^{-1} X^T S \qquad (3.26)$$

olarak tanımlanır. (3.26) eşitliği, $S = \sigma^2 V^{-1}$ in neden olduğu uzaklığın kullanılmasıyla X in sütun uzayındaki şapka matrisine karşılık gelir ve marjinal uydurulmuş değerlerle ilgili olarak gözlemlerin leverage değerleri hakkında bilgi verir (Demidenko, 2004). Bunun bir sonucu olarak, L_1, ***genelleştirilmiş marjinal leverage matris*** olarak adlandırılır. Marjinal uydurulmuş değerlerle ilgili olarak gözlemlerin leverage değerlerinin incelenmesinde diğer bir yaklaşım, Christensen ve ark. (1992) tarafından gösterilmiştir. Christensen ve ark. (1992) tarafından gösterilen yaklaşım altında, her bir gözlemin leverage değeri silinmiş gözlemin tahmin edilen kovaryans matrisinin bir fonksiyonu olarak elde edilir. Ancak, özellikle büyük veri setlerinde kolay bir şekilde uygulanamayan bir yaklaşımdır.

Demidenko (2004)'nun önerdiği gibi i-inci birimin genelleştirilmiş marjinal leverage matrisi, L_1 in i-inci köşegen bloğu olan $L_{1i} = X_i(X^T S X)^{-1} X_i^T S_i^{-1}$ dir.

(i) Birimler içi yüksek leverage gözlemlerin belirlenmek istenmesi ve

(ii) Birimler içi yüksek leverage gözlemlerin olmasının, bunlara karşılık gelen birimlerin yüksek leverage olduğu anlamına gelmemesi nedenleriyle, L_{1i} nin esas köşegeninin j-inci elemanı olarak, i-inci birimin genelleştirilmiş marjinal leverage değerinin j-inci gözlemi $L_{1i(jj)}$ şeklinde tanımlanır.

$L_{1i(jj)}$ nin yüksek değerleri, marjinal uydurulmuş değerlerle ilgili i-inci birimin j-inci gözleminin yüksek leverage olduğunu gösterir. Uygulamada, yüksek leverage gözlemler veya birimler öznel temele dayandırılarak tanımlanmalarına

rağmen, literatürde bazı nesnel kriterler de bulunmaktadır. Örneğin, X_i nin $n_i \times p$ boyutlu olduğu hatırlatması yapıldığında,

$$trace(L_1) = \sum_{i=1}^{m}\sum_{j=1}^{n_i} L_{1i(jj)} = p$$

şeklindedir. Burada, $L_{1i(jj)} \geq 2p/n$ ise i-inci birimden alınan j-inci gözlem marjinal uydurulmuş değerler için yüksek leverage nokta olarak düşünülebilir. Demidenko (2004)'de önerildiği gibi,

$$\frac{trace(L_{1i})}{n_i} \geq \frac{2p}{n}$$

ise marjinal uydurulmuş değerler için yüksek leverage nokta i-inci birimdir, denilebilir. Literatürde, $3p/n$ gibi alternatif eşik noktaları da kullanılmıştır.

(3.2) ile verilen uzun süreli lineer karma model altında, bir gözlem hem marjinal hem de koşullu uydurulmuş değerleri etkiliyorsa, her ikisi üzerinde her gözlemin ortak etkisini değerlendirmek uygun olacaktır. Böylece, $\hat{y}^* = X\hat{\beta} + Z\hat{u}$ modeli kullanır ve ***genelleştirilmiş ortak leverage matrisi***,

$$\begin{aligned} L = \frac{\partial \hat{y}^*}{\partial y^T} &= \frac{\partial \hat{y}}{\partial y^T} + \frac{\partial Z\hat{u}}{\partial y^T} \\ &= L_1 + \underbrace{ZDZ^T}_{L_2} \text{Q} = L_1 + L_2 \text{Q} \\ &= L_1 + L_2(S - SX(X^TSX)^{-1}X^TS) \\ &= L_1 + L_2S - L_2SL_1 \\ &= L_2S + [I_n - L_2S]L_1 \end{aligned}$$

olarak gösterilir. Burada, L_1 genelleştirilmiş marjinal leverage matris ve $L_2 = ZDZ^T$ şeklindedir. L_2 matrisi, rastgele etkilerin varlığıyla açıklanan birimler içi değişkenliğin parçasını göstermektedir.

$L_1 + ZDZ^T$ ayrışımına dayalı olmak üzere, Nobre (2004) ve Demidenko ve Stukel (2005), bağımsız olarak,

$$H_2 = ZDZ^T \quad (3.27)$$

önermişlerdir. (3.27) eşitliği, ***koşullu uydurulmuş değerlerin rastgele etki bileşeni için genelleştirilmiş leverage matris*** olarak kullanılır.

$$H_2 = ZDZ^T(S - SL_1) = ZDZ^TS - ZDZ^TS\,L_1 = L_2S\,[I_n - L_1]$$

eşitliğinde H_2 nin genelleştirilmiş marjinal leverage matris olan L_1 e bağlı olduğu görülür. Koşullu uydurulmuş değerlerin hem sabit hem de rastgele etki bileşenlerinin etkisi altında olması modelin doğru bir şekilde algılanmasını güçleştirebilir. Bu nedenle, koşullu uydurulmuş değerlerin rastgele etki bileşeni için oluşturulan genelleştirilmiş leverage matrisin sadece rastgele model matrisi olan Z ve onun kovaryans matrisi olan D ye bağlı olması gerekir. Böylece, ***koşullu uydurulmuş değerlerin rastgele etki bileşeni için genelleştirilmiş leverage matris*** olarak H_2 yerine L_2 nin kullanılması önerilir. Genelleştirilmiş ortak leverage matrisi L nin hem L_1 i hem de L_2 yi içerdiğine dikkat edilmelidir. Z deki açıklayıcı değişkenler tarafından gerilen vektör uzayındaki merkezden uzakta duran gözlemlerin, koşullu uydurulmuş değerlerin rastgele etki bileşenini $(Z\hat{u})$ etkilemesi beklenir ve sonuç olarak, rastgele etkilerin varlığıyla açıklanan birimler içi değişkenliğin (L_2) tahmini etkilenir (Tan ve ark., 2001). Böylece, koşullu uydurulmuş değerlerin rastgele etki bileşenine karşılık gelen yüksek leverage gözlemlerin, modelin varyans bileşenlerinin tahmini üzerinde orantısız bir ağırlığa neden olduğu söylenebilir.

i-inci birimin genelleştirilmiş rastgele bileşen leverage matrisi $L_{2i} = \sigma^2 Z_i G Z_i^T$ şeklinde ifade edilir. L_{2i} nin esas köşegeninin j-inci elemanı ise $L_{2i(jj)}$ ile gösterilir. $(n_i)^{-1}trace(L_{2i}) \geq 2trace(L_2)/n$ ise koşullu uydurulmuş değerlerin rastgele bileşenlerine karşılık gelen i-inci birimin yüksek leverage olduğu söylenebilir. Aynı yorum $L_{2i(jj)}$ için de geçerlidir.

3.3.1. H_2 ile L_2 nin Karşılaştırılması

Koşullu uydurulmuş değerlerin hem sabit hem de rastgele etki bileşenlerinin etkisi altında olması modelin doğru bir şekilde algılanmasını güçleştirebilir. Aynı zamanda H_2 deki gözlemler, marjinal uydurulmuş değerler için yüksek leverage olan gözlemlerin varlığı tarafından maskelenebilir ve dolayısıyla, H_2 grafikleri yüksek leverage noktaları tam olarak gösteremeyebilirler. Bu nedenle, koşullu uydurulmuş değerlerin rastgele etki bileşeni için oluşturulan genelleştirilmiş leverage matrisin sadece rastgele model matrisi olan Z ve onun kovaryans matrisi olan D ye bağlı olması gerekir. Sonuç olarak, yüksek etkili gözlemlerin belirlenmesinde H_2 yerine L_2 yi kullanmanın daha doğru sonuçlara götüreceği söylenebilir.

3.4. Lineer Karma Modellerde Etki Analizi

Bu bölümde, (3.1) modeli altında lineer karma modeller için etki analizi incelenecektir. Bir analizden elde edilen sonuçlar, bir veya daha fazla gözlemden büyük ölçüde etkilenebilir, yani, gözlemlerin tümü istatistiksel modellerde eşit etkiye sahip değillerdir. Gözlem silme tanılamaları, bireysel gözlemlerin istatistiksel model üzerindeki etkisini ölçmek için kullanılırlar. Christensen ve ark. (1992), gözlem silme tanılamaları üzerinde durmuşlardır. Kullanışlı silme tanılamaları yapmanın anahtarı, etkili hesaplama formüllerinin geliştirilmesi, temel yapı bloklarının kullanılmasıyla gözlem silme tanılamalarının yapılmasına olanak sağlanması ve hesaplamalar için verinin tamamının kullanılmasıdır. Bu kısımda, i-inci gözlemin silinmesinin sabit etkilere ilişkin parametre tahminlerini, rastgele etkilere ilişkin parametre tahminlerini ve varyans oranlarının tahminlerini değiştirip değiştirmediğini ortaya çıkarmak amacıyla klasik regresyon etki tanılamalarının lineer karma modele genişletilmiş olan biçimleri incelenmiştir.

3.4.1. Varyans Bileşenleri Oranları Üzerinde Etki Analizi

Varyans bileşenleri oranları için genel tanılama araçları Cook uzaklığı ve bilgi oranının benzerleridir.

3.4.1.1. Cook Uzaklığının Benzeri

Varyans bileşenleri olan γ lar için Cook uzaklık ölçümünün benzeri, Zewotir ve Galphin (2005) tarafından,

$$\begin{aligned} CD_i(\gamma) &= \left(\hat{\gamma}_{(i)} - \hat{\gamma}\right)^T [Var(\hat{\gamma})]^{-1} \left(\hat{\gamma}_{(i)} - \hat{\gamma}\right) \\ &= -g(i)^T (Q - \mathcal{G})^{-1} [-Q^{-1}]^{-1} (-(Q - \mathcal{G})^{-1}) g(i) \\ &= -g(i)^T (Q - \mathcal{G})^{-1} Q (Q - \mathcal{G})^{-1} g(i) \\ &= -g(i)^T [Q(I_r - Q^{-1}\mathcal{G})]^{-1} Q [Q(I_r - Q^{-1}\mathcal{G})]^{-1}\, g(i) \\ &= -g(i)^T (I_r - Q^{-1}\mathcal{G})^{-1} \underbrace{Q^{-1}Q}_{I_r} (I_r - Q^{-1}\mathcal{G})^{-1} Q^{-1} g(i) \\ &= -g(i)^T (I_r - Q^{-1}\mathcal{G})^{-2} Q^{-1} g(i) \\ &= g(i)^T (I_r + Var(\hat{\gamma})\mathcal{G})^{-2} Var(\hat{\gamma}) g(i) \end{aligned}$$

şeklinde ifade edilmiştir. $CD_i(\gamma)$ nın büyük değerleri varyans oranlarının tahminleri üzerinde etkili olan noktaları göstermektedir.

3.4.1.2. Bilgi Oranının Benzeri

Bilgi oranının benzeri, ML tahmininin bilgi matrisinin determinantında meydana gelen değişmeyi ölçer ve Zewotir ve Galphin (2005) tarafından,

$$\begin{aligned} IR(\gamma) &= \frac{det\left(Inf\left(\hat{\gamma}_{(i)}\right)\right)}{det\left(Inf(\hat{\gamma})\right)} = \frac{det(Q - \mathcal{G})}{det(Q)} = \frac{det(Q)det(I_r - Q^{-1}\mathcal{G})}{det(Q)} \\ &= det(I_r + Var(\hat{\gamma})\mathcal{G}) \end{aligned}$$

olarak verilmiştir. $det(A)$, karesel A matrisinin determinantını gösterir. İdeal olarak tüm gözlemler bilgi matrisi üzerinde aynı etkiye sahipse $IR(\gamma)$, 1 e yakın olur. 1 den sapma, i-inci gözlemin etkili gözlem olduğunu gösterir. Dengeli veri setinde $IR(\gamma)$ nın değerleri benzer olacaktır. $Var(\hat{\gamma})$ ve I_r tüm $i = 1, \cdots, n$ için sabit olduğundan $IR(\gamma)$, $\mathcal{G}$ nin bir fonksiyonu olur. $\mathcal{G}$ de C_i ve $c_{ii} = 1 - z_i^T(D^{-1} + Z^TZ)^{-1}z_i$ nin bir fonksiyonudur.

3.4.2. Sabit Etkili Parametre Tahminleri Üzerinde Etki Analizi

3.4.2.1. Cook Uzaklığının Benzeri

Lineer karma modellerdeki sabit etkili parametreler üzerindeki etkinin ölçülmesi için, lineer regresyonda tanımlanan Cook uzaklığına (Cook, 1977) benzer bir ölçü, Zewotir ve Galphin (2005) tarafından,

$$
\begin{aligned}
CD_i(\beta) &= \frac{\left(\hat{\beta}_{(i)} - \hat{\beta}\right)^T W(X,X)\left(\hat{\beta}_{(i)} - \hat{\beta}\right)}{p\hat{\sigma}_e^2} \\
&= \frac{\frac{e_i}{r_{ii}} C_i^T X \overbrace{W(X,X)^{-1}W(X,X)}^{I} W(X,X)^{-1}X^T C_i \frac{e_i}{r_{ii}}}{p\hat{\sigma}_e^2} \\
&= \frac{\overbrace{C_i^T X W(X,X)^{-1}X^T C_i}^{(c_{ii}-r_{ii})} e_i^2}{pr_{ii}^2\hat{\sigma}_e^2} = \frac{(c_{ii} - r_{ii})t_i^2}{pr_{ii}}
\end{aligned}
\tag{3.28}
$$

olarak ifade edilmiştir. $CD_i(\beta)$ nın büyük değerleri, $\hat{\beta}$ üzerinde etkili olan noktaları göstermektedir. (3.28) eşitliğinin daha iyi bir şekilde anlaşılabilmesi için Ek 5 incelenebilir.

3.4.2.2. Varyans Oranının Benzeri

Varyans oranı, i-inci gözlem silindiğinde $\hat{\beta}$ nın varyansının determinantının değişimini ölçer. Bu ölçü, lineer regresyon modellerindeki Covratio oranının lineer karma modellerdeki karşılığıdır ve Zewotir ve Galphin (2005) tarafından,

$$VR_i(\beta) = \frac{det\left(Var\left(\hat{\beta}_{(i)}\right)\right)}{det\left(Var\left(\hat{\beta}\right)\right)} = \frac{det\left(\hat{\sigma}^2_{e(i)} W_{[i]}(X,X)^{-1}\right)}{det\left(\hat{\sigma}^2_e W(X,X)^{-1}\right)}$$

$$= det\left(\hat{\sigma}^2_e \left(\frac{n-t_i^2}{n-1}\right) W_{[i]}(X,X)^{-1} (\hat{\sigma}^2_e)^{-1} W(X,X)\right)$$

$$= det\left(\left(\frac{n-t_i^2}{n-1}\right)\left(I + \frac{W(X,X)^{-1}X^T C_i C_i^T X}{c_{ii} - C_i^T X W(X,X)^{-1} X^T C_i}\right)\right)$$

$$= \left(\frac{n-t_i^2}{n-1}\right)^p det\left(I + \frac{W(X,X)^{-1}X^T C_i C_i^T X}{c_{ii} - C_i^T X W(X,X)^{-1} X^T C_i}\right)$$

$$= \left(\frac{n-t_i^2}{n-1}\right)^p \frac{c_{ii}}{r_{ii}}$$

olarak elde edilmiştir. Burada,

$$\hat{\sigma}^2_{e(i)} = \frac{n}{n-1}\hat{\sigma}^2_e - \frac{e_i^2}{(n-1)r_{ii}} = \hat{\sigma}^2_e \left(\frac{n}{n-1} - \frac{t_i^2}{n-1}\right)$$

$$W_{[i]}(X,X) = W(X,X) - \frac{\check{x}_i \check{x}_i^T}{m_i} \quad \text{ve}$$

$$W_{[i]}(X,X)^{-1} W(X,X) = \left(W(X,X)^{-1} + \frac{W(X,X)^{-1}X^T C_i C_i^T X W(X,X)^{-1}}{(m_i - p_{ii})}\right) W(X,X)$$

$$= I + \frac{W(X,X)^{-1}X^T C_i C_i^T X}{c_{ii} - C_i^T X W(X,X)^{-1} X^T C_i}$$

şeklindedir (Bkz. Ek 3 ve Ek 5, Ek 6).

Eğer nokta etkili değilse $VR_i(\beta)$, 1 e yakınsar. Bu durum, $\hat{\beta}$ nın varyansı üzerinde i-inci gözlemin etkisinin değerlendirilmesi için kriter olarak $|VR_i(\beta)-1|$ ölçüsünün kullanılabileceğini gösterir. $|VR_i(\beta)-1|$ istatistiğinin büyük değerleri, i-inci gözlemin etkili gözlem olduğunu gösterir.

Determinant yerine trace işlemcisi kullanılırsa,

$$
\begin{aligned}
VR_i(\beta) &= trace\left\{\left[Var(\hat{\beta})\right]^{-1}\left(Var(\hat{\beta}_{(i)})\right)\right\} \\
&= \frac{\hat{\sigma}^2_{e(i)}}{\hat{\sigma}^2_e} trace\left\{W(X,X)W_{[i]}(X,X)^{-1}\right\} \\
&= \frac{n-t_i^2}{n-1} trace\left\{I + \frac{X^T C_i C_i^T X W(X,X)^{-1}}{r_{ii}}\right\} \\
&= \frac{n-t_i^2}{n-1} trace\left\{I + \frac{C_i^T X W(X,X)^{-1} X^T C_i}{r_{ii}}\right\} \\
&= \frac{n-t_i^2}{n-1} trace\left\{I + \frac{(c_{ii}-r_{ii})}{r_{ii}}\right\} = \frac{n-t_i^2}{n-1} trace\left\{I_p + \frac{c_{ii}}{r_{ii}} - 1\right\} \\
&= \frac{n-t_i^2}{n-1}\left(\frac{c_{ii}}{r_{ii}} + p - 1\right)
\end{aligned}
$$

elde edilir.

Silinen i-inci nokta trace değerini değiştirmiyorsa $VR_i(\beta)$, p ye yakın olacaktır. $|VR_i(\beta)-p|$, $\hat{\beta}$ nın varyansı üzerinde i-inci gözlemin etkisinin değerlendirilmesi için kriter olarak kullanılabilir. $|VR_i(\beta)-p|$ istatistiğinin büyük değerleri, i-inci gözlemin etkili gözlem olduğunu gösterir.

3.4.2.3. Cook-Weisberg İstatistiğinin Benzeri

β nın güven elipsoit hacminde meydana gelen değişimi ölçmek için kullanılan diğer bir istatistik Cook-Weisberg istatistiğidir. $y \sim N(X\beta, \sigma_e^2 H)$ varsayımı altında $\hat{\beta} \sim N(\beta, \sigma_e^2 W(X,X)^{-1})$ dir. β için $100(1-\alpha)\%$ lık güven elipsoidi,

$$E = \left\{\beta: \frac{(\beta - \hat{\beta})^T W(X,X)(\beta - \hat{\beta})}{p\hat{\sigma}_e^2} \leq F(\alpha, p, n-p)\right\}$$

şeklindedir.

i-inci gözlem silindiğinde,

$$E_{(i)} = \left\{\beta: \frac{(\beta - \hat{\beta}_{(i)})^T W_{[i]}(X,X)(\beta - \hat{\beta}_{(i)})}{p\hat{\sigma}_{e(i)}^2} \leq F(\alpha, p, n-p-1)\right\}$$

şeklinde olur.

Cook ve Weisberg (1980), çoklu lineer regresyon modeli için güven elipsoidinin hacmi üzerindeki etkinin bir ölçümü olarak $E_{(i)}$ nin E ye oranının logaritmasını önermişlerdir. Çoklu lineer regresyon modeli için önerilen bu yaklaşımın lineer karma modellere taşınması sonucunda,

$$CW_i = log\left(\frac{volume(E_{(i)})}{volume(E)}\right) \quad i = 1, \cdots, n$$

olarak elde edilir. Cook ve Weisberg (1982) tarafından tanımlanan,

$$volume(E) \propto (p\hat{\sigma}_e^2 F(\alpha, p, n-p))^{p/2} det(W(X,X))^{1/2}$$

$$volume(E_{(i)}) \propto (p\hat{\sigma}_{e(i)}^2 F(\alpha, p, n-p-1))^{p/2} det\left(W_{[i]}(X,X)\right)^{1/2}$$

ifadelerinin kullanılması sonucunda sabit etkiler için benzer ölçü, Zewotir ve Galphin (2005) tarafından,

$$CW_i = log\left\{\frac{det(W(X,X))^{1/2} \hat{\sigma}_{e(i)}^p (F(\alpha, p, n-p))^{p/2}}{det\left(W_{[i]}(X,X)\right)^{1/2} \hat{\sigma}_e^p (F(\alpha, p, n-p-1))^{p/2}}\right\}$$

$$CW_i = \frac{1}{2} log\left(\frac{c_{ii}}{r_{ii}}\right) + \frac{p}{2} log\left(\frac{n - t_i^2}{n-1}\right) + \frac{p}{2} log\left(\frac{F(\alpha, p, n-p)}{F(\alpha, p, n-p-1)}\right)$$

şeklinde elde edilmiştir.

CW_i, tüm gözlemler modelde iken negatif ise güven elipsoidinin hacmi azalır ve i-inci gözlem silindikten sonra güven elipsoidinin hacmi artar (CW_i, tüm gözlemler modelde iken pozitif ise bu durumun tersi söz konusu olur). CW_i nin işareti göz ardı edilirse, büyük $|CW_i|$ $\hat{\beta}$ ve/veya $Var(\hat{\beta})$ üzerinde güçlü bir etkiyi gösterir. $\frac{p}{2} log\left(\frac{F(\alpha,p,n-p)}{F(\alpha,p,n-p-1)}\right)$ sabiti, etkili gözlemlerin belirlenmesinde kullanışlı olmamasına rağmen CW_i nin işaretinin belirlenmesinde önemli bir role sahiptir. Açıkça CW_i istatistiği,

$$CW_i = \frac{1}{2} log\big(VR_i(\beta)\big) + \frac{p}{2} log\left(\frac{F(\alpha, p, n-p)}{F(\alpha, p, n-p-1)}\right)$$

şeklinde basitleştirilebilir. F değerleri göz ardı edildiğinde, CW_i, $VR_i(\beta)$ ya benzerdir. Gerçekte büyük n değerleri için F değerlerinin oranı 1 e yakınsar ve böylece, $CW_i \approx \frac{1}{2} log\big(VR_i(\beta)\big)$ olur.

3.4.2.4. Andrews-Pregibon İstatistiğinin Benzeri

Güven elipsoidinin hacmine dayalı diğer bir ölçü, Zewotir ve Galphin (2005) tarafından,

$$AP_i = -\frac{1}{2} log\left\{\frac{e_{(i)}^{V\,T} e_{(i)}^{V} det\left(W_{[i]}(X,X)\right)}{e^T e\ det(W(X,X))}\right\} \qquad i = 1, \dots, n$$

olarak tanımlanan Andrews-Pregibon istatistiğidir.

(3.8) eşitliği dikkate alındığında,

$$e_{(i)}^{\vee} = y_{(i)} - X_{(i)}\hat{\beta}_{(i)} - Z_{(i)}\hat{u}_{(i)}$$
$$= y_{(i)} - X_{(i)}\left(\hat{\beta} - W(X,X)^{-1}X^T C_i \frac{e_i}{r_{ii}}\right) - Z_{(i)}\left(\hat{u} - DZ^T R_i \frac{e_i}{r_{ii}}\right)$$
$$= y_{(i)} - X_{(i)}\hat{\beta} - Z_{(i)}\hat{u} + \left(X_{(i)}W(X,X)^{-1}X^T C_i + Z_{(i)}DZ^T R_i\right)\frac{e_i}{r_{ii}}$$

olarak ifade edilir. Burada tanımlanan $e_{(i)}^{\vee}$ nin Bölüm 3.2.1.3 de tanımlanan $e_{(i)}$ den farkı, $e_{(i)}$ nin formülünde y_i , x_i^T ve z_i^T yerine $y_{(i)}$, $X_{(i)}$ ve $Z_{(i)}$ nin gelmesidir.

$R_{i(i)}$ ve $C_{i(i)}$, i-inci eleman olmaksızın R ve H^{-1} in i-inci sütununu göstermek üzere,

$$Z_{(i)}DZ^T R_i = \left(H_{(i)} - I_{(i)}\right)H^{-1}(I_i - XW(X,X)^{-1}X^T C_i)$$
$$= -R_{i(i)} - X_{(i)}W(X,X)^{-1}X^T C_i$$

$$e_{(i)}^{\vee} = y_{(i)} - X_{(i)}\hat{\beta}_{(i)} - Z_{(i)}\hat{u}_{(i)} - R_{i(i)}\frac{e_i}{r_{ii}}$$

ve

$$e_{(i)}^{\vee\,T} e_{(i)}^{\vee} = ssq\left(y_{(i)} - X_{(i)}\hat{\beta}_{(i)} - Z_{(i)}\hat{u}_{(i)} - R_{i(i)}\frac{e_i}{r_{ii}}\right)$$
$$= e^T e + \left(\frac{e_i}{r_{ii}}\right)^2 ssq(R_i) - 2\frac{e_i}{r_{ii}}e^T R_i = ssq\left(e - \frac{e_i}{r_{ii}}R_i\right)$$

olup,

$$AP_i = \frac{1}{2}log\left(\frac{c_{ii}}{r_{ii}}\right) + \frac{1}{2}log(e^T e) - \frac{1}{2}log\left(ssq\left(e - \frac{e_i}{r_{ii}}R_i\right)\right)$$

şeklindedir. AP_i istatistiğinin büyük değerleri i-inci gözlemin etkili gözlem olduğunu gösterir.

3.4.3. Rastgele Etkili Parametrelerin Ön Tahmini Üzerindeki Etki Analizi

Rastgele etkilerin tüm veriden elde edilen ön tahmini ile i-inci gözlemin silinmesiyle elde edilen ön tahmini arasındaki karesel uzaklığın değerlendirilmesi amacıyla önerilen tanılama ölçüsüdür. Rastgele etkilerin varyansı, $var(u) = \sigma_e^2 D$ dir. Lineer regresyonda tanımlanan Cook uzaklığına (Cook, 1977) benzer bir ölçü, Zewotir ve Galphin (2005) tarafından,

$$CD_i(u) = \frac{\left(\hat{u} - \hat{u}_{(i)}\right)^T [Var(\hat{u})]^{-1} \left(\hat{u} - \hat{u}_{(i)}\right)}{\hat{\sigma}_e^2} = \frac{\left(\hat{u} - \hat{u}_{(i)}\right)^T D^{-1} \left(\hat{u} - \hat{u}_{(i)}\right)}{\hat{\sigma}_e^2}$$
$$= \frac{\frac{e_i}{r_{ii}} R_i^T ZDD^{-1} DZ^T R_i \frac{e_i}{r_{ii}}}{\hat{\sigma}_e^2} = \frac{e_i^2 R_i^T ZDZ^T R_i}{r_{ii}^2 \hat{\sigma}_e^2} = \frac{t_i^2 R_i^T ZDZ^T R_i}{r_{ii}}$$
$$= \frac{t_i^2}{r_{ii}} R_i^T (H - I) R_i = t_i^2 \left(1 - \frac{ssq(R_i)}{r_{ii}}\right)$$

olarak ifade edilmiştir (Bkz. Ek 5). $CD_i(u)$ nun büyük değerleri, i-inci gözlemin rastgele etkili parametrelerin ön tahmininde etkili bir gözlem olduğunu gösterir.

3.4.4. Likelihood Fonksiyonu Üzerindeki Etki Analizi

Zewotir ve Galphin (2005), varyans bileşenleri oranlarının tahminlerinin elde edilmesi zor olduğu için $\theta^T = [\beta^T, \sigma_e^2]$ nin tahminini γ sabitken düşünmüşlerdir. $\hat{\theta}(\gamma)$ ve $\hat{\theta}_{(i)}(\gamma)$, n ve $n-1$ gözlemlerine dayalı θ nın ML tahminleridir. $\lambda\left(\hat{\theta}(\gamma)\right)$ ve $\lambda\left(\hat{\theta}_{(i)}(\gamma)\right)$, $\hat{\theta}(\gamma)$ ve $\hat{\theta}_{(i)}(\gamma)$ da hesaplanan log likelihood fonksiyonlarıdır. Likelihood fonksiyonları arasındaki uzaklık, likelihood fonksiyonu üzerinde i-inci gözlemin etkisini ölçer. Likelihood uzaklığı, LD_i ile gösterilirse (Zewotir ve Galphin, 2005),

$$LD_i = 2\left\{\lambda\left(\hat{\theta}(\gamma)\right) - \lambda\left(\hat{\theta}_{(i)}(\gamma)\right)\right\} \tag{3.29}$$

şeklinde olur.

y yanıt vektörünün normal dağıldığı varsayımı altında, (3.3) eşitliğinin logaritması, $\hat{\sigma}_e^2 = \left(y - X\hat{\beta}\right)^T H^{-1}\left(y - X\hat{\beta}\right)/n$ kullanılarak,

$$logf(y) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\hat{\sigma}_e^2) - \frac{1}{2}log(\det(H)) - \frac{n}{2} \qquad (3.30)$$

elde edilir. (3.29) eşitliğinde yer alan $\lambda\left(\hat{\theta}(\gamma)\right)$ ve $\lambda\left(\hat{\theta}_{(i)}(\gamma)\right)$ ifadeleri, (3.30) eşitliğinden yararlanılarak,

$$\lambda\left(\hat{\theta}(\gamma)\right) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\hat{\sigma}_e^2) - \frac{1}{2}log(\det(H)) - \frac{n}{2} \qquad (3.31)$$

$$\lambda\left(\hat{\theta}_{(i)}(\gamma)\right) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\hat{\sigma}_e^2) - \frac{1}{2}log(\det(H)) - \frac{\left(y - X\hat{\beta}_{(i)}\right)^T H^{-1}\left(y - X\hat{\beta}_{(i)}\right)}{2\hat{\sigma}_{e(i)}^2} \qquad (3.32)$$

olarak bulunur. (3.32) eşitliğindeki $\left(y - X\hat{\beta}_{(i)}\right)^T H^{-1}\left(y - X\hat{\beta}_{(i)}\right)$ ifadesi, Ek 5 de yer alan

$$\hat{\beta}_{(i)} = \hat{\beta} - W(X,X)^{-1}X^T C_i \frac{e_i}{r_{ii}}$$

eşitliğinden yararlanılarak,

$$
\begin{aligned}
&\left(y - X\hat{\beta}_{(i)}\right)^T H^{-1}\left(y - X\hat{\beta}_{(i)}\right) \\
&\quad = \left(y - X\hat{\beta}\right)^T H^{-1}\left(y - X\hat{\beta}\right) + \left(y - X\hat{\beta}\right)^T H^{-1} XW(X,X)^{-1}X^T C_i \frac{e_i}{r_{ii}} \\
&\quad + \frac{e_i}{r_{ii}} C_i^T XW(X,X)^{-1}X^T H^{-1}\left(y - X\hat{\beta}\right) \\
&\quad + \frac{e_i}{r_{ii}} C_i^T XW(X,X)^{-1}W(X,X)W(X,X)^{-1}X^T C_i \frac{e_i}{r_{ii}} \\
&\quad = n\hat{\sigma}_e^2 + \underbrace{y^T H^{-1} XW(X,X)^{-1}}_{\hat{\beta}^T} X^T C_i \frac{e_i}{r_{ii}} - \underbrace{y^T H^{-1} XW(X,X)^{-1}}_{\hat{\beta}^T} X^T C_i \frac{e_i}{r_{ii}} \\
&\quad + \frac{e_i}{r_{ii}} C_i^T X\hat{\beta} - \frac{e_i}{r_{ii}} C_i^T X\hat{\beta} + \frac{e_i}{r_{ii}} \underbrace{C_i^T XW(X,X)^{-1}X^T C_i}_{(c_{ii} - r_{ii})} \frac{e_i}{r_{ii}} \\
&\quad = n\hat{\sigma}_e^2 + \left(\frac{e_i}{r_{ii}}\right)^2 (c_{ii} - r_{ii}) \qquad (3.33)
\end{aligned}
$$

olarak elde edilir. Son olarak, (3.31), (3.32) ve (3.33) eşitliklerindeki ifadeler (3.29) da yazılırsa,

$$
LD_i = n\, log\left(\frac{n - t_i^2}{n-1}\right) + \frac{[nr_{ii} + {t_i}^2(c_{ii} - r_{ii})](n-1)}{r_{ii}(n - t_i^2)} - n
$$

elde edilir.

LD_i nin büyük değerleri, i-inci gözlemin likelihood fonksiyonu üzerinde etkili olduğunu gösterir.

4. HİYERARŞİK LİNEER MODELLERDE REZİDÜ ANALİZİ

4.1. Giriş

Birden fazla düzeyde gerçekleşen, gruplara veya kümelere ayrılmış gözlemlere sahip veriden oluşan modellere *hiyerarşik lineer modeller* adı verilir. Hiyerarşik lineer modeller, günlük hayatta bir çok farklı alanda kullanılabilen modellerdir. Örneğin tarım alanında, gübreleme denemesinde parsellere ayrılmış toprağın her bir parselinden pek çok bitki rastgele seçilip, her bir bitkinin ağırlığı ölçülmek istenebilir veya eğitim alanında, öğrencilerin sınıflarına veya okullarına göre gruplandırılmasıyla başarı düzeyleri karşılaştırılmak istenebilir.

Hiyerarşik lineer modeller, lineer karma modellerin özel bir durumu olmalarına rağmen, model geçerliğinin değerlendirilmesinde bu özel yapı göz ardı edilmemelidir. Çünkü, hiyerarşik lineer bir model düzeylere ayrılır ve her bir düzey, ayrı ayrı incelenir. Bu sayede, her bir düzeyin çok daha basit ve kolay bir biçimde anlaşılması sağlanır. Dolayısıyla, verinin iç içe geçmiş bir yapıya sahip olduğu durumlarda hiyerarşik lineer modellerin kullanılması daha doğru sonuçlara ulaşılmasını sağlar.

Bu bölümde, iki düzeyli bir hiyerarşik lineer model tanımlanacaktır. İlk olarak, birinci düzey olan ***birimler-içi model*** daha sonra ise, ikinci düzey olan ***birimler-arası model*** incelecektir. N deneğin J birime gruplandırıldığı yani j-inci birimde n_j deneğin ve $\sum_{j=1}^{J} n_j = N$ olduğu varsayımı yapılsın. Bu varsayım altında, *birimler-içi model*,

$$y_j = X_j \; _j + \varepsilon_j \tag{4.1}$$

şeklindedir. Burada y_j $n_j \times 1$ tipinde, her bir X_j $n_j \times q'$ tipinde, $_j$ $q' \times 1$ tipinde ve ε_j, $n_j \times 1$ tipinde olup $E(\varepsilon_j) = 0$, $Var(\varepsilon_j) = \sigma_j^2 \Sigma_j$ ile normal dağılıma sahiptir. Aynı zamanda, ε_j ile $\varepsilon_{j'}$ her $j \neq j'$ için bağımsızdır. Fakat bağımsızlık varsayımı göz ardı edilirse, birimler-içi model genel karma bir modele dönüşebilir. Σ_j, bilinen

pozitif tanımlı bir matris olmak üzere, sıklıkla $\Sigma_j = I_{n_j}$ olduğu varsayımı yapılır ve bu varsayım, gözlemlerin birimler-içinde sabit hataya sahip olduğu anlamına gelir. Eğer $\sigma_1^2 = \cdots = \sigma_J^2 = \sigma^2$ varsayımı yapılırsa, gözlemler birimler-arasında sabit hataya sahip olurlar. Hatanın hem birimler-içi hem de birimler-arası sabit varyanslı olması, gözlemlerin de sabit varyanslı olması anlamına gelir.

Birimler-içi model incelendikten sonra, bir birimden diğer bir birime geçildiğinde $_j$ de meydana gelen değişme yani birimler-içi modellerle ilişkili olan birimler-arası model incelenmek istenir. $_j$ nin k-ıncı elemanına karşılık w_{jk} gelir ve,

$$_{jk} = w_{jk}^T \beta_k + u_{jk}$$

olarak modellenebilir. □ nın q' elemanlarının her biri için böyle bir eşitlik yazılabilir. Dolayısıyla $k = 1, \cdots, q'$ için,

$$W_j = \begin{pmatrix} w_{j1}^T & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_{jq'}^T \end{pmatrix} = w_{j1}^T \oplus w_{j2}^T \oplus \cdots \oplus w_{jq'}^T$$

$$\beta = \left(\beta_1^T, \beta_2^T, \cdots, \beta_{q'}^T\right)^T$$

$$u_j = \left(u_{j1}, u_{j2}, \cdots, u_{jq'}\right)^T$$

olarak alınırsa, ***birimler-arası model***,

$$_j = W_j \beta + u_j \tag{4.2}$$

olarak elde edilir. Burada $_j$ $q' \times 1$ tipinde, her bir W_j matrisi $q' \times p$ tipinde, β p × 1 tipinde ve u_j, $q' \times 1$ tipinde olup $E(u_j) = 0$, $Cov(u_j) = \sigma_j^2 D$ ile normal dağılıma sahiptir. Aynı zamanda her j, j' ve j'' ile $j \neq j'$ için $Cov(u_j, u_{j'}) = 0$,

$Cov(\varepsilon_j, \varepsilon_{j'}) = 0$ ve $Cov(u_j, \varepsilon_{j''}) = 0$ dır. Gözlemler birimler-arası sabit varyanslılığa sahipse, $Cov(u_j) = \sigma^2 D$ olarak belirlenerek de hatanın sabit varyansa sahip olması sağlanabilir. D köşegen ise model, varyans bileşenli hiyerarşik lineer model olarak adlandırılır.

Blok köşegen bir yapı $\oplus$ sembolü ile gösterilir. Örneğin,

$$A \oplus B = \begin{pmatrix} A & 0 \\ O & B \end{pmatrix}$$

ifadesi, esas köşegende A ve B, esas köşegen dışında ise sıfır matrislerini içeren blok köşegen bir matris anlamına gelir.

(4.1) ile (4.2) eşitliklerinin birleştirilmesi sonucunda,

$$y_j = X_j W_j \beta + X_j u_j + \varepsilon_j$$

elde edilir. Elde edilen bu eşitlik hiyerarşik lineer modellerin, karma modellerin özel bir durumu olduğunu gösterir. Aslında,

$$y = \left[y_1^T, y_2^T, \cdots, y_J^T\right]^T$$

$$X = \left[W_1^T X_1^T, W_2^T X_2^T, \cdots, W_J^T X_J^T\right]^T$$

$$Z = X_1 \oplus X_2 \oplus \cdots \oplus X_J$$

$$u = \left[u_1^T, u_2^T, \cdots, u_J^T\right]^T$$

$$\varepsilon = \left[\varepsilon_1^T, \varepsilon_2^T, \cdots, \varepsilon_J^T\right]^T$$

olarak alındığında,

$$y = X\beta + Zu + \varepsilon$$

lineer karma modeli elde edilir. (3.2) eşitliği ile verilen lineer karma modelde,

$$S = \sigma^2 V^{-1} = \sigma^2\left(\sigma^2(ZDZ^T + \Sigma)\right)^{-1} = (ZDZ^T + \Sigma)^{-1}$$

olarak elde edildi. Bu durum, hiyerarşik lineer modele uygulanırsa,

$$V_j = \sigma_j^2\left[X_j D X_j^T + \Sigma_j\right]$$

ve

$$S_j = \Sigma_j^{-1} - \Sigma_j^{-1}X_j\left(D^{-1} + X_j^T\Sigma_j^{-1}X_j\right)^{-1}X_j^T\Sigma_j^{-1}$$

olarak elde edilir. İşlem kolaylığı sağlaması açısından, $C_j = D^{-1} + X_j^T\Sigma_j^{-1}X_j$ olarak alınabilir ve dolayısıyla $C_j^{-1} = \left[I + DX_j^T\Sigma_j^{-1}X_j\right]^{-1}D$ şeklinde yazılabilir.

Hiyerarşik lineer modellere örnek olarak, y-eksenini kesim noktası rastgele olan model ve üç düzeyli hiyerarşik model verilebilir.

y Eksenini Kesim Noktası Rastgele Olan Model: Her bir birimdeki eğimlerin aynı fakat y eksenini kesim noktalarının farklı olduğu m birime sahip lineer bir model uydurulduğu varsayılsın. j-inci birim,

$$y_j = \alpha_j + X_j \quad + \varepsilon_j$$

şeklinde modellenebilir. n_j gözlemlerin sayısı olmak üzere, $\varepsilon_j \sim N\left(0, \sigma^2 I_{n_j}\right)$ dir. Dahası, y eksenini kesim noktaları α_j ye bağlı olarak değişebilir. Burada,

$$\alpha_j = w_j^T\beta + u_j$$

şeklindedir ve tüm hatalar bağımsız olmak üzere, $u_j \sim N(0, \gamma\sigma^2)$ dir. y eksenini kesim noktası rastgele olan model için, o halde,

$$V_j = \sigma^2 \left(I_{n_j} + \gamma J_{n_j} \right)$$

dir. Sonuç olarak,

$$S_j = I_{n_j} - \frac{\gamma}{1 + n_j\gamma} J_{n_j}$$

dir.

Üç Düzeyli Hiyerarşik Model: Bir Bayesian analizi, bilinmeyen parametreler için önsel dağılımları içerir. Yapısal olarak böyle bir önsellik, hiyerarşik bir modeldeki ek düzeyi oluşturur. Smith (1973), Σ_j ve $\sigma_j^2 D$ nin bilindiği varsayımı altında bilgilendirici bir önsele (Γ) yer vermiştir ve bu sayede, üç düzeyli hiyerarşik modelin gösterimini,

$$y_j \big| \; _j \sim N\left(X_j \; _j, \Sigma_j\right)$$

$$_j \big| \beta \sim N\left(W_j\beta, \sigma_j^2 D\right) \; j = 1, \cdots, J$$

$$\beta \sim N(\beta_0, \Gamma)$$

olarak elde etmiştir. β üzerindeki önsellik, diğer tüm parametrelerin sonsal dağılımlarını kendine çektiğinden dolayı Smith (1973) ün önerdiği gösterimler kullanılabilirdir.

4.2. Hiyerarşik Lineer Modellerde Parametre Tahminleri

Hiyerarşik lineer modellerin, lineer karma modeller olarak formülleştirilebilmelerinden dolayı, hiyerarşik lineer modellerin parametre tahminlerinin elde edilmesinde, lineer karma modellerdeki parametre tahminlerinden yaralanılabilir. Gözlemlerin birimler-arasında sabit varyanslı hataya sahip olduğu varsayımı yapılır. Yani $\sigma_1^2 = \cdots = \sigma_J^2 = \sigma^2$ şeklindedir. Böylece, lineer karma modellerdeki $\hat{\beta} = (X^T S X)^{-1} X^T S y$, hiyerarşik lineer modellerde,

$$\hat{\beta} = \mathbb{G}^{-1} \sum W_j^T X_j^T S_j \, y_j$$

olarak elde edilir. Burada,

$$\mathbb{G} = \sum W_j^T X_j^T S_j \, X_j W_j$$

dir. Lineer karma modellerdeki,

$$\begin{aligned}
\hat{u} = DZ^T \mathbb{Q} y &= DZ^T S\left(y - X\hat{\beta}\right) \\
&= DZ^T \left(\Sigma^{-1} - \Sigma^{-1} Z \left(\underbrace{D^{-1} + Z^T \Sigma^{-1} Z}_{C} \right)^{-1} Z^T \Sigma^{-1} \right) \left(y - X\hat{\beta}\right) \\
&= D(C - Z^T \Sigma^{-1} Z) C^{-1} Z^T \Sigma^{-1} \left(y - X\hat{\beta}\right) \\
&= D(D^{-1} + Z^T \Sigma^{-1} Z - Z^T \Sigma^{-1} Z) C^{-1} Z^T \Sigma^{-1} \left(y - X\hat{\beta}\right) \\
&= C^{-1} Z^T \Sigma^{-1} \left(y - X\hat{\beta}\right)
\end{aligned}$$

şeklinde elde edilmişti. Bu tahminin hiyerarşik lineer modeldeki karşılığı ise,

$$\hat{u}_j = C_j^{-1} X_j^T \Sigma_j^{-1} \left(y_j - X_j W_j \hat{\beta}\right) \tag{4.3}$$

olarak elde edilir. Aynı şekilde, $\hat{u}_j$ başka bir şekilde de gösterilebilir.

$$\mathbb{Q}_{ij} = \begin{cases} S_j - \mathbb{P}_{jj} & i = j \\ -\mathbb{P}_{ij} & i \neq j \end{cases}$$

parçalanışı verilsin. Burada, $\mathbb{P}_{ij} = S_i X_i W_i \mathbb{G}^{-1} W_j^T X_j^T S_j$ şeklinde olup,

$$\hat{u}_j = D X_j^T \sum_{k=1}^{J} \mathbb{Q}_{jk} y_k \tag{4.4}$$

veya eş değer olarak,

$$\hat{u}_j = D X_j^T \left[S_j y_j - \sum_{k=1}^{J} \mathbb{P}_{jk} y_k \right] \tag{4.5}$$

şeklindedir. (4.4) ve (4.5) eşitlikleri Eşitlik (4.3) e göre y_k nın $\hat{u}_j$ yı nasıl etkilediğini daha açık bir şekilde gösterir.

$\hat{\beta} = \mathbb{G}^{-1} \sum W_j^T X_j^T S_j \, y_j$ ile (4.4) ve (4.5) eşitliklerinin dikkate alınmasıyla

$$Var(\hat{\beta}) = \sigma^2 \mathbb{G}^{-1}$$

$$Var(\hat{u}_j - u_j) = \sigma^2 \left[D - D X_j^T \mathbb{Q}_{jj} X_j D \right]$$

$$Cov(\hat{\beta}, \hat{u}_j - u_j) = -\mathbb{G}^{-1} W_j^T X_j^T S_j \, X_j D$$

ve $j \neq k$ iken,

$$Cov(\hat{u}_j - u_j, \hat{u}_k - u_k) = \sigma^2 \left[D X_j^T \mathbb{Q}_{jk} X_k D \right]$$

olarak elde edilir (Hilden-Minton, 1995).

Alternatif olarak, sabit-gözlem varsayımı altında, verinin J birime ve J sabite ayrıldığı, yani

$$y_j = 0\beta + X_j \; _j + \varepsilon_j$$

$$0 = W_j\beta - I \; _j + u_j$$

şeklinde yazıldığı ifade edilsin. Verilen bu ifadelerin matris gösterimiyle yazılıp, uygun ağırlıklandırma yapılması sonucunda elde edilen normal eşitliklerin çözülmesiyle,

$$\widehat{\;}_j = C_j^{-1}\left(X_j^T \Sigma_j^{-1} y_j + D^{-1} W_j \hat{\beta}\right) \tag{4.6}$$

olarak elde edilir (Hilden-Minton, 1995).

Böylece, $_j$ için BLUP elde edilmiş olur. Aynı zamanda $\widehat{\;}_j$, deneysel bir Bayes tahmin edicisidir. Dolayısıyla Eşitlik (4.6),

$$\widehat{\;}_j = C_j^{-1}\left(X_j^T \Sigma_j^{-1} X_j \widehat{\;}_j^{LS} + D^{-1} \widehat{\;}_j^{Prior}\right)$$

$$\widehat{\;}_j = \Lambda_j \widehat{\;}_j^{LS} + \left(1 - \Lambda_j\right) \widehat{\;}_j^{Prior} \tag{4.7}$$

olarak ifade edilir. Burada $\widehat{\;}_j^{LS} = \left[X_j^T \Sigma_j^{-1} X_j\right]^{-1} X_j^T \Sigma_j^{-1} y_j$ şeklinde olup, sadece j-inci birim üzerindeki bilgiye bağlı olan $_j$ nin EKK tahmin edicisidir. $\widehat{\;}_j^{Prior} = W_j \hat{\beta}$ şeklinde olup, birimler arası modelden gelen $_j$ için deneysel önsel yoğunluğun tahmin edicisidir. Ve Λ_j, j-inci birimin büzülme katsayısını gösterir. Büzülme katsayısı, önsel ortalamayı daraltan birimler-içi tahmin edicinin ne kadarlık bir ağırlığa sahip olması gerektiğini gösterir. Aynı zamanda,

$$\widehat{\;}_j = W_j \hat{\beta} + \hat{u}_j = \widehat{\;}_j^{Prior} + \hat{u}_j \tag{4.8}$$

şeklinde yazılabilir. Bu nedenle $\hat{u}_j$, $_j$ nin deneysel önsel ve sonsal dağılımlarının ortalamalarının farkı olarak düşünülebilir. Veya $\hat{u}_j$, birimler-arası modelin bir rezidüsü olarak düşünülebilir. Bu iki düşünceden biri, $\hat{u}_1, \cdots, \hat{u}_J$ nin değerlendirilmesinde tanılama ölçüsü olarak kullanılabilir.

4.3. Hiyerarşik Lineer Modellerde Rezidü Analizi

Hiyerarşik lineer modellerin lineer karma modellerin özel bir durumu olmalarından dolayı, lineer karma modellerde rezidülerde meydana gelen bozulma hiyerarşik lineer modellere de uygulanabilir. Hiyerarşik lineer modellerin, küçük modellerin bir ağı olarak düşünülmesi oldukça doğaldır. Yani, birimler-arası modelle ilişkilendirilecek J birim-içi modelleri vardır. Birim-içi modelleri ayrı bir şekilde değerlendirilir, böylece birimler-arası modelde tanımlanan rastgele hataların neden olduğu bozulmalar engellenir. Birim-içi modellerin ayrı bir şeklide uydurulmasından sonra, birimler-arası model incelenebilir. Bu düzeyde bozulma kaçınılamaz, fakat birimler-arası modelin hataları çok değişkenli normal dağılıma sahip olacak şekilde modellenebilir ve rastgele etkilerin en küçük bozulmuş rezidülere sahip olduğu belirli izdüşümlere bakılabilir. Sonuç olarak, bu bölümde gözlem-düzeyli rezidüler ve birim-düzeyli rezidüler (rastgele etkiler) olmak üzere iki tip rezidünün analizi incelenecektir.

4.3.1. Gözlem-Düzeyli Rezidüler

j-inci birim model için,

$$\varepsilon_j = y_j - X_j \ _j$$

şeklinde elde edilmişti. $_j$ koşulu üzerinde ε_j nin dağılımıyla ilgili üç varsayım yapılabilir. Bu varsayımlar,

1. $E[\varepsilon_j |\ _j] = 0$, lineer model yansızdır;
2. $Var[\varepsilon_j |\ _j] = \sigma_j^2 \Sigma_j$; ve
3. $\varepsilon_j |\ _j$ çok değişkenli normal dağılıma sahiptir,

şeklindedir.

Bu varsayımların doğruluğunun belirlenebilmesi amacıyla ε_j nin dikkatli bir şekilde değerlendirilmesi gerekir. $_j$ ile ilgili gerekli olan bilgiye sahip olunmadığı için en iyisi, model varsayımları verilmişken $_j$ nin tahmin edilmesidir. Fakat sadece bu varsayımlar, $_j$ nin tahmin edilebilmesi için yeterli değildir. $_j$ nin tahmin edilebilmesi için daha iyi bir yöntem, rezidülerden yararlanılmasıdır. Bu amaçla, $_j$ nin dört tahmin edicisine karşılık gelen üç farklı rezidü incelenecektir. Bu rezidüler, sırasıyla, deneysel Bayes (EB) rezidüleri, en küçük kareler (LS) rezidüleri ve önsel (prior) veya ön tahmin rezidüleridir.

4.3.1.1. EB Rezidüler

(4.8) eşitliği ile verilen deneysel Bayes tahmini yardımıyla elde edilen EB rezidü,

$$e_j = y_j - X_j \widehat{\ }_j = y_j - X_j W_j \hat{\beta} - X_j \hat{u}_j$$

şeklindedir. EB rezidülerin temel avantajı, varsayılan hiyerarşik lineer model altında BLUP olmalarıdır. Temel dezavantajları ise, EB rezidüler oluşturulurken hiyerarşik lineer model varsayımının yapılmasıdır. j-inci birim-içi model belirlenmek istendiği için, birimler-arası modeldeki herhangi bir problemin, analizde bozulmaya neden olması istenen bir durum değildir. Birimler-arası modelin ortalama yapısının yanlış belirlenmesi, $W_j \hat{\beta}$ dan dolayı e_j nin yanlı olmasına neden olacaktır. Ortalama yapısı doğru belirlense bile, diğer birimlerin gözlem-düzeyli hataları ve rastgele etkilerdeki bozulma durumu, EB rezidünün de bozulmasına neden olacaktır.

EB rezidülerin kullanımı ile daha fazla karışıklık ortaya çıkar. Çünkü, $\widehat{\ }_j$ büzülen bir tahmin edicidir ve e_j, $\Sigma_j{}^{-1}$ ile ilişkili X_j nin sütunlarına ortogonal değildir. $e_j^T X_j = 0$ doğru olsaydı, $\widehat{\ }_j = \widehat{\ }_j{}^{LS}$ şeklinde olurdu. Bu, X_j nin sütunlarının herhangi bir lineer kombinasyonuna karşı e_j nin grafiğinin sıfırdan farklı lineer bir trende sahip olacağı anlamına gelir. Böyle bir grafik incelendiğinde, uyumla ilgili herhangi bir eksikliğin olmadığı sonucuna varılır.

Lineer karma modellerde incelendiği gibi, en küçük bozulmuş olacak şekilde e_j nin izdüşümleri belirlenebilir. Benzer işlemler yapıldığında,

$$CF\left(l^T e_j\right) = 1 - \frac{l^T \Sigma_j \mathrm{Q}_{jj} \Sigma_j \mathrm{Q}_{jj} \Sigma_j l}{l^T \Sigma_j \mathrm{Q}_{jj} \Sigma_j l}$$

olarak elde edilir. Böylece, tekil değer ayrışımı yardımıyla ilişkisiz standartlaştırılmış rezidülerin bir kümesi elde edilir (Hilden-Minton, 1995).

4.3.1.2. LS Rezidüler

Hiyerarşik lineer modellerde, birim-içi modellerin ayrı olarak incelenmesi hiyerarşik yapının kullanılmasının temel avantajlarından biridir. Kullanılan her birim-iç model, klasik bir lineer modeldir. Bu nedenle, $_j$ sadece j-inci birimin içindeki gözlemlerden tahmin edilebilir. Dolayısıyla,

$$\widehat{\ }_j{}^{LS} = \left[X_j^T \Sigma^{-1} X_j\right]^{-1} X_j^T \Sigma^{-1} y_j$$

eşitliği kullanılabilir. Burada, $X_j^T \Sigma^{-1} X_j$ ifadesi singüler olmayan bir yapıya sahiptir. $\widehat{\ }_j{}^{LS}$, LS rezidünün,

$$e_j^{LS} = y_j - X_j \widehat{\ }_j{}^{LS}$$

olarak elde edilmesine yardımcı olur.

LS rezidüler, saf veya bozulmamış rezidülerin avantajlarına sahiplerdir. Aynı zamanda, birimler-arası modeldeki yetersizlikler nedeniyle yansız olup, birim-içi tasarım matrislerine ortogonaldirler. Bu avantajlar, EB rezidülerdeki dezavantajlara karşılık gelmektedir. EB rezidüler, hem birim-içi hem de birimler-arası modeller doğru olduğunda BLUP ları kullanırken LS rezidüler, sadece birim-içi modeller doğru olduğunda BLUP ları kullanır. Bu durum EB rezidü için bir avantaj olmasına rağmen, birim-içi modelin belirlenme amacı için LS tahminlerinin kullanılması gerekir. Çünkü LS tahminler, varsayılan birimler-arası modeldeki herhangi bir sapmadan etkilenmezler. Birimler-arası modelin doğru bir şekilde oluşturulduğu belirleninceye kadar, LS rezidüler EB rezidülere tercih edilir.

LS rezidüleri kullanmanın diğer bir avantajı, klasik lineer regresyon modellerinin tanılanmasında herhangi bir yeni tekniğe ihtiyaç duyulmamasıdır. Standart regresyon tanılama tekniklerinin tümü, birim-içi modeller için de geçerlidir.

4.3.1.3. Karma Rezidüler

Gözlemlenmiş veya modellenmiş J birimin olduğu varsayımı altında, J + 1-inci birimin tahmini veya ön tahmini belirlenmek istensin. (4.2) ile verilen birimler-arası modelden yararlanılarak,

$$\widehat{}_{J+1}^{Prior} = W_{J+1}\hat{\beta}$$

olarak elde edilir. Böylece, bu tahmin veya ön tahminin doğruluğu,

$$c_{J+1} = y_{J+1} - X_{J+1}\widehat{}_{J+1}^{Prior}$$

değerlendirilmesiyle yapılabilir. c_{J+1}, Geisser (1992) tarafından ön tahmin veya tahmin için model kritiğinin hesaplanması amacıyla elde edilmiştir. Esas olarak, c_{J+1} in Euclidean formu $V_{J+1}^{-1} = \sigma^{-2}S_{J+1}$ yardımıyla,

$$\left\|c_{J+1}\right\|^2 = c_{J+1}^T V_{J+1}^{-1} c_{J+1}$$

olarak elde edilir. c_{J+1} in n_{J+1} serbestlik dereceli $\mathcal{X}^2$ dağılımına sahip olduğu varsayımı altında,

$$p_{J+1} = P\left[\mathcal{X}^2_{n_{J+1}} \geq \left\|c_{J+1}\right\|^2 \middle| \beta = \hat{\beta}, \sigma^2 = \hat{\sigma}^2, D = \widehat{D}\right]$$

olarak hesaplanabilir. Model kritiği olan p_{J+1}, bir p değeri olarak yorumlanabilir. Küçük p_{J+1} değeri, hiyerarşik lineer modelin yanlış belirlendiğini, tahminlerin zayıf olduğunu ve $J + 1$-inci birimin modelle uyumsuz olduğunu gösterir.

Belirli bir zamanda, her birimin teker teker silinmesiyle yukarıda ifade edilen ön tahmin süreci gösterilebilir. Silinmiş model kritikleri $p_{(1)}, p_{(2)}, \cdots, p_{(J)}$ olmak üzere, silinen-birim tahminlerin üzerine koşul konulmasıyla,

$$p_{(j)} = P\left[\mathcal{X}^2_{n_j} \geq \left\|c_j\right\|^2 \middle| \beta = \hat{\beta}_{(j)}, \sigma^2 = \hat{\sigma}^2_{(j)}, D = \widehat{D}_{(j)}\right]$$

olarak elde edilir. Veya bu yorum tamamen göz ardı edilip, yine de

$$c_j = y_j - X_j \widehat{\ }_j^{Prior}$$

incelenebilir. Silinmeyen model kritiklerinin gerçek değerleri olan $p_1, p_2, \cdots, p_J$ karşılaştırılabilir. Burada,

$$p_j = P\left[\mathcal{X}^2_{n_j} \geq \left\|c_j\right\|^2 \middle| \beta = \hat{\beta}, \sigma^2 = \hat{\sigma}^2, D = \widehat{D}\right]$$

şeklindedir. Silinmeyen model kritiklerinin gerçek değerlerinin ranklarına karşı silinmeyen model kritiklerinin gerçek değerlerini grafiklendirmek, sıradan olmayan birimlerin belirlenmesini sağlayabilir. Böyle bir grafikteki noktalar, model iyi

olduğunda ortak doğru boyunca uzanacaktır. Ek olarak, küçük p_j veya $p_{(j)}$ değerleri, birim-içi modellerdeki uyumsuz noktaların belirlenmesinde kullanılabilir.

4.3.2. Birim-Düzeyli Rezidüler (Rastgele Etkiler)

Birimler-arası modelin belirlenmesi için,

$$u_j = \quad_j - W_j\beta$$

rezidülerinin değerlendirilmesi gerekir. Bununla birlikte, $\quad_j$ direkt olarak gözlemlenemez ve j-inci birimde gerekli olmamasına rağmen bazı gözlem-düzeyleri oluşturulmaksızın $\quad_j$ hakkında herhangi bir çıkarımda bulunulamaz. Bu nedenle u_j, kolaylıkla tahmin edilebilen bozulmuş rezidülere sahip olacaktır. Esas olarak,

1. EB rastgele etki, $\hat{u}_j = \widehat{\quad}_j - W_j\hat{\beta}$;
2. LS rastgele etki, $\hat{u}_j^{LS} = \widehat{\quad}_j^{LS} - W_j\hat{\beta}$ ve
3. u_j nin sonsal yoğunluğunun değerlendirilmesi

şeklinde üç seçime sahip olunur.

Waternaux ve ark. (1989), ilk iki seçimi karşılaştırmışlardır. EB rastgele etkinin sıfır civarında yanlı büzülen bir tahmin olduğunu, EB rastgele etkilerinin değişkenliğinin LS rastgele etkilerinkinden daha az olduğunu ve LS rastgele etkilerinin küçük Λ_j için ulaşılamaz olabileceğini ifade etmişlerdir. (4.7) ve (4.8) eşitlikleri birleştirildiğinde,

$$\hat{u}_j = \Lambda_j\left(\widehat{\quad}_j^{LS} - W_j\hat{\beta}\right) = \Lambda_j\hat{u}_j^{LS}$$

olarak elde edilir. Burada, $\Lambda_j = C_j^{-1}X_j^T\Sigma^{-1}X_j$ şeklindedir.

Birim-içi ve birimler-arası modelleri doğru olduğunda, EB rastgele etkileri BLUP ları gösterecektir. Birimler-arası modellerle birim-içi modellerin birbirinden

bağımsız oldukları belirlendikten sonra, birim-içi modelin doğru bir şekilde oluşturulduğu kabul edilirse, EB rastgele etkilerin birimler-arası modelin geçerliliğinin belirlenmesinde en iyi tek etki olduğu kararına varılabilir. Öncelikle birim-içi modelin belirlenmek istenmesinin nedeni, birim-içi modelin yanlış belirlenmesinin EB veya LS rastgele etkilerin değerlendirilmesinde yanıltıcı yorumlara neden olabilmesidir.

u_j vektör-değerli ise, $\hat{u}_j$ nın birkaç izdüşümü incelenmek istenebilir. $\hat{u}_j$ nın marjinal dağılımları biraz kapsamlı bir içeriğe sahip oldukları için, $\hat{u}_j$ nın marjinal dağılımları ve aynı zamanda iki veya üç boyutlu saçılım grafikleri, $\hat{u}_j$ nın birkaç izdüşümünün belirlenmesinde yardımcı olabilir.

Rastgele etkilerdeki normal olmama durumunun belirlenmesinde, daha az bozulmuş olan $\hat{u}_j$ için rastgele etkilere daha fazla ağırlık verilmesi gerekebilir. Çünkü aşırı derecede bozulmuş olan rezidüler, sadece bu bozulma nedeniyle normal olarak ortaya çıkabilirler. $Var(\hat{u}_j)$ ya karşı $Var(\hat{u}_j - u_j)$ nın oranının tersten ağırlıklandırılması bu konuda yardımcı olabilir.

4.4. Çift Rezidü Grafiği ve Uyumun Değerlendirilmesi

Birimler-arası model geçerli bir şekilde belirlenmiş ve tahmin edilmiş ise, $\widehat{}_j^{Prior} = W_j\hat{\beta}$ iyi bir ön tahmin edicidir. Aynı zamanda $\widehat{}_j$, $\widehat{}_j^{LS}$ ye yakın olacağından dolayı $\widehat{}_j^{Prior}$, $\widehat{}_j^{LS}$ için iyi bir ön tahmin edici olacaktır. Sonuç olarak, birimler-arası model geçerli ise, LS rezidüler EB rezidülerle oldukça ilişkili olacaktır.

EB rezidülere karşı LS rezidülerin grafiklendirilmesi ile, birimler-arası modeldeki uyumun iyiliği değerlendirilebilir. Jan de Leeuw (1986), buna ***çift rezidü grafiği*** adını vermiştir. EB rezidüler, model doğru olduğunda doğru rastgele hataların ön tahminlerini tam olarak verdikleri için yatay eksen boyunca uzanırlar. Bu grafikte yüksek korelasyon varsa, birimler-arası modelin birim-içi modelin parametreleri üzerinde fazla yanlılığa sahip olmadığı yani birimler-arası modelin daha iyi olduğu sonucuna ulaşılır.

Teknik olarak, her j birim için $\Sigma_j^{-\frac{1}{2}}\left[y_j - X_j\widehat{\ }_j\right]$ ya karşı $\Sigma_j^{-\frac{1}{2}}\left[y_j - X_j\widehat{\ }_j^{LS}\right]$ nin grafiği çizilmek istenir (Hilden-Minton, 1995). Bu,

$$\rho^2 = \frac{\sum S_j^2}{\sum\left[S_j^2 + \left(\widehat{\ }_j^{LS} - \widehat{\ }_j\right)^T X_j^T \Sigma_j^{-1} X_j\left(\widehat{\ }_j^{LS} - \widehat{\ }_j\right)\right]}$$

olarak ifade edilen korelasyon tahmininin karesinin elde edilmesini sağlar. Burada, $S_j^2 = \left[y_j - X_j\widehat{\ }_j^{LS}\right]^T \Sigma_j^{-1}\left[y_j - X_j\widehat{\ }_j^{LS}\right]$ dir. ρ^2, aynı zamanda çift rezidü grafiğindeki eğimin EKK tahminidir ve 1 üst sınırına sahiptir. Grafiğin dik bir eğime sahip olması daha iyi bir uyum ve $\widehat{\ }_j^{LS}$ ile $\widehat{\ }_j$ nın birbirine yakın olması anlamına gelir.

EB rezidülere dayalı bir rezidü analizi ile LS rezidülere dayalı bir rezidü analizi farklı sonuçlara sahip olabilir. Çift rezidü grafiğini kullanmanın bir avantajı, her iki rezidü analizi sonucunu aynı grafikte göstermesidir.

Hiyerarşik lineer modellerde etki analizi ile ilgili literatürde yeterli düzeyde bir çalışmaya ulaşılamadığı için bu bölümde, hiyerarşik lineer modellerde etki analizi kısmına değinilmeyecektir.

5. UYGULAMA

5.1. Lineer Karma Modellerde Rezidü Analizi Örneği

Brezilya'da bulunan Sao Paulo Üniversitesi'nin Diş Hekimliği Fakültesi'nde Nobre ve Singer (2007) tarafından yürütülen bir çalışma sonucunda elde edilen verinin analiz edilmesiyle, günlük kullanım sonucunda temizlenen bakteri plağı kapasitesine göre, iki tip diş fırçasının karşılaştırılması amaçlanmaktadır. Bu diş fırçalarından biri, Brezilya'daki hemen hemen her dükkanda bulunan ve daha ucuz fiyatlı olan "monoblock" diş fırçası (monoblock tootbrush) diğeri ise, piyasaya yeni çıkarılan ve daha pahalı olan "conventional" diş fırçası (conventional tootbrush) dır. Dişlerin fırçalanmasından önce (before) ve sonra (after) yapılan dört değerlendirme seansında (session), yaşları 4 ile 6 arasında değişen 32 çocuktan elde edilen bakteri plak indekslerine karşılık gelen veri Çizelge 5.1 ile gösterilmektedir.

Nobre ve Singer (2007) tarafından model,

$$In\, y_{ijd} = \alpha_j + \beta In\, x_{ijd} + u_i + \varepsilon_{ijd} \qquad (5.1)$$

olarak tanımlanmaktadır. Bu model, lineer karma model tiplerinden uzun süreli bir modeldir. (5.1) eşitliğinde, $y_{ijd}(x_{ijd})$ j-inci diş fırçası tipi (conventional diş fırçası için $j = 0$, monoblock diş fırçası için $j = 1$) ile d-inci seansta değerlendirilen i-inci denek (subject) için tedavi sonrası (tedavi öncesi) elde edilen bakteri plak indeksini, α_j j-inci diş fırçası tipi ile ilgili sabit bir etkiyi, β beklenen bakteri plak indeks azalma oranının sabit bir katsayısını ifade etmektedir. $u_i \sim N(0, \tau^2)$ ve $\varepsilon_{ijd} \sim N(0, \sigma^2)$ $i = 1, \cdots, 32$ ve $d = 1, \cdots, 4$ için birbirinden bağımsız rastgele değişkenler olup, sırasıyla, deneklerin rastgele etkilerine ve rastgele ölçüm hatalarına karşılık gelmektedirler.

Çizelge 5.1. Bakteri plak indeksleri (Nobre ve Singer, 2007)

Subject	Toothbrush	1st session		2nd session		3rd session		4th session	
		Before	After	Before	After	Before	After	Before	After
1	Conventional	1.05	1.00	1.13	0.84	1.15	0.86	1.13	0.94
2	Conventional	1.07	0.62	0.92	0.62	1.02	0.57	1.15	0.85
3	Conventional	0.82	0.62	1.52	1.07	1.39	0.97	1.78	1.39
4	Conventional	1.37	0.90	1.65	1.20	1.75	1.40	1.92	1.67
5	Conventional	1.97	1.52	1.30	1.07	1.50	1.15	1.65	1.37
6	Conventional	1.30	0.82	1.17	0.70	0.75	0.50	1.47	1.12
7	Conventional	1.61	1.19	1.52	1.13	1.22	1.00	1.63	1.22
8	Conventional	1.02	0.73	1.08	0.64	0.94	0.73	1.14	0.97
9	Conventional	1.62	1.25	1.45	1.10	1.10	0.75	1.70	1.32
10	Conventional	1.65	1.22	1.57	1.22	1.47	1.10	1.62	1.17
11	Conventional	1.02	0.78	0.60	0.47	0.88	0.75	1.36	1.08
12	Conventional	0.71	0.60	1.13	0.39	0.84	0.65	1.65	1.31
13	Conventional	1.70	1.55	1.85	1.37	1.87	1.55	1.60	1.30
14	Conventional	1.30	1.02	1.65	0.97	1.72	1.20	1.37	1.22
15	Conventional	1.40	0.80	1.83	1.03	1.76	1.38	1.96	1.15
16	Conventional	1.40	1.12	1.25	0.67	1.50	1.10	1.50	1.22
17	Monoblock	1.66	1.63	1.36	1.16	1.52	0.88	1.41	1.20
18	Monoblock	1.02	0.80	0.92	0.82	1.10	0.76	1.28	1.15
19	Monoblock	0.75	0.67	1.00	0.92	1.00	0.87	1.15	1.10
20	Monoblock	1.29	1.23	0.91	0.76	1.14	0.94	1.35	0.97
21	Monoblock	1.27	1.20	1.20	0.95	1.10	1.00	1.37	1.17
22	Monoblock	1.07	0.85	1.39	1.25	1.39	1.25	1.28	1.21
23	Monoblock	1.35	1.21	1.42	1.17	1.42	1.19	1.42	1.23
24	Monoblock	1.32	1.02	1.60	1.40	1.35	1.02	1.50	1.25
25	Monoblock	1.66	1.61	1.50	1.36	1.72	1.41	1.69	1.44
26	Monoblock	1.30	1.07	0.84	0.61	0.88	0.61	0.96	0.57
27	Monoblock	1.57	1.20	1.50	1.07	1.15	1.00	1.25	1.05
28	Monoblock	1.67	1.50	1.47	1.32	1.07	0.97	1.50	1.37
29	Monoblock	0.91	0.67	0.96	0.62	1.09	0.53	1.12	0.37
30	Monoblock	1.06	0.70	1.00	0.85	1.15	0.93	1.12	1.00
31	Monoblock	2.30	2.00	1.37	1.25	1.40	1.32	2.15	1.90
32	Monoblock	1.15	1.00	1.23	1.11	1.15	1.07	1.26	1.00

Çizelge 5.1 ile verilen bakteri plak indeks verisi ilk olarak Microsoft Excel programına kaydedilir (Bkz. Ek 9), daha sonra R programına yazılan kod yardımıyla çağrılır.

(5.1) eşitliği ile verilen uzun süreli model, matris gösterimi yardımıyla $y = X\beta + Zu + \varepsilon$ formunda ifade edilir. Burada, y, 128×1 tipinde tedavi sonrası bakteri plak indekslerinin logaritma değerlerinden oluşan yanıt vektörünü, X, 128×3 tipinde sabit etkilere ve Z, 128×32 tipinde rastgele etkilere ilişkin tasarım matrislerini oluşturmaktadır. β, 3×1 tipinde sabit etkilerden oluşan parametreler

vektörü olup, $\begin{bmatrix} \beta \\ \alpha_0 \\ \alpha_1 \end{bmatrix}$ şeklinde ifade edilir. Son olarak u, 32×1 tipinde rastgele etkilerden oluşan parametreler vektörü olup, $\begin{bmatrix} u_1 \\ \vdots \\ u_{32} \end{bmatrix}$ şeklindedir (Bkz. Ek 10).

Çizelge 5.2. REML yardımıyla uydurulan lineer karma model sonuçları

AIC	BIC	LogLik	Deviance	REMLdev
-83.12	-68.86	46.56	-107.4	-93.12

Çizelge 5.3. Rastgele etkiler için uydurulan lineer karma model sonuçları

Groups Name	Variance	Std.Dev.
Subject (Intercept)	0.0072958	0.085415
Residual	0.0208129	0.144267

Çizelge 5.4. Sabit etkiler için uydurulan lineer karma model sonuçları

	Estimate	Std.Error	t value
Log(x) (β)	1.05141	0.06605	15.918
Conventional Tootbrush (α_0)	-0.32999	0.03376	-9.774
Monoblock Tootbrush (α_1)	-0.21527	0.03182	-6.765

Çizelge 5.5. Sabit etkilerin korelasyonu

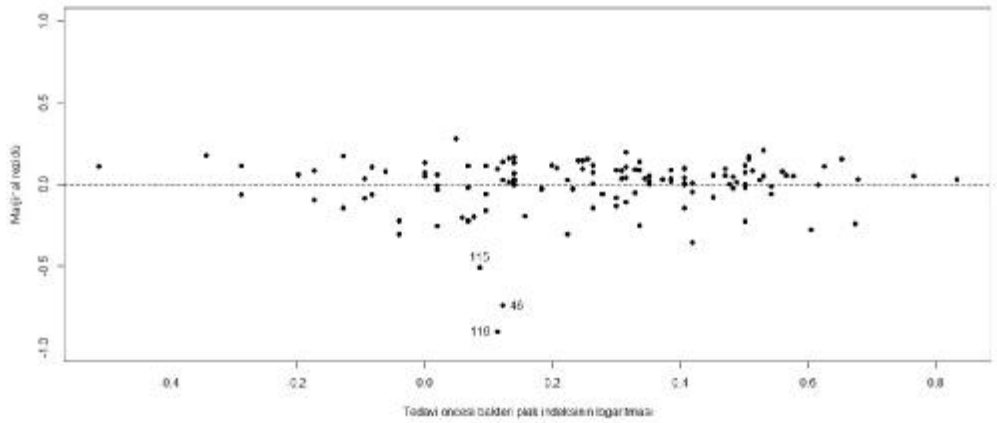
	Log(x)	Conventional Tootbrush
Conventional Tootbrush	-0.561	
Monoblock Tootbrush	-0.478	0.268

Çizelge 5.6. Uydurulan lineer karma model için ANOVA tablosu

	Df	Sum Sq	Mean Sq	F Value
Log(x)	1	2.9407	2.9407	141.292
Tootbrush	2	2.3734	1.1867	57.017

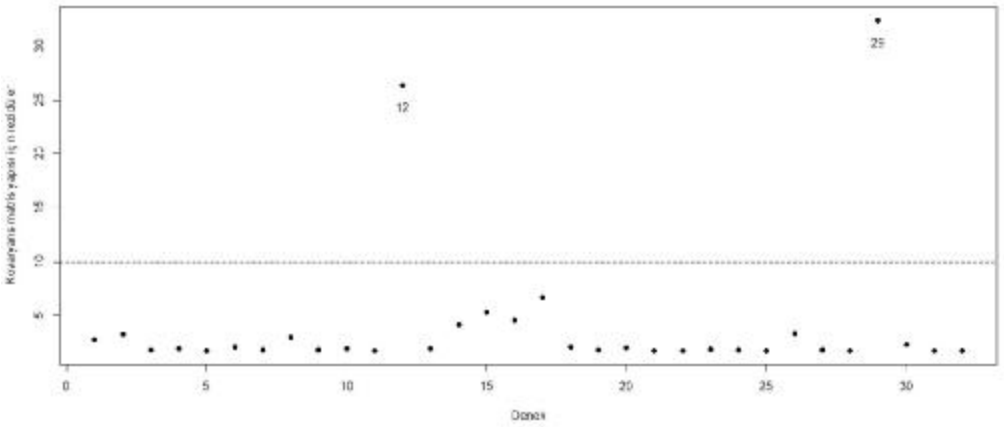
Uydurulan lineer karma modelin değerlendirilmesinde, sırasıyla, AIC (Akaike bilgi kriteri) ve BIC (Bayesian bilgi kriteri) den yararlanılabilir. Küçük AIC ve BIC değerleri, iyi bir şekilde uydurulmuş olan bir modelin göstergesidir.

Etkilerin lineer bir yapıya sahip olup olmadığının belirlenebilmesi amacıyla, Şekil 5.1 ile gösterilen tedavi öncesi bakteri plak indeksinin logaritmasına karşı marjinal rezidülerin grafiği incelenebilir. Modele iyi uyum sağlayamayan kovaryans yapısına sahip deneklerin belirlenmesi amacıyla da Şekil 5.2 incelenebilir.



Şekil 5.1. (5.1) modeli için marjinal rezidü grafiği

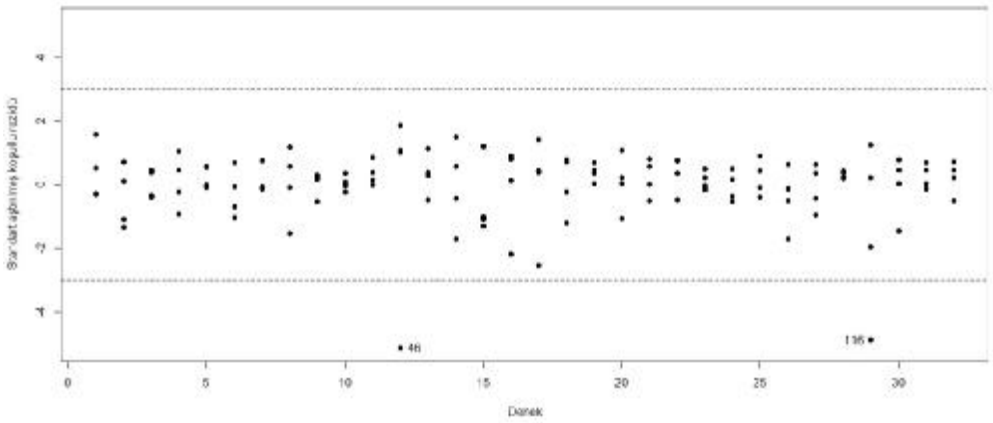
Şekil 5.1 incelendiğinde, 46 (12 numaralı deneğin ikinci seansına karşılık gelen 12.2 numaralı gözlem), 115 ve 116 (29 numaralı deneğin üçüncü ve dördüncü seanslarına karşılık gelen 29.3 ve 29.4 numaralı gözlemler) numaralı gözlemlerin sıfır etrafındaki dağılımlardan uzaklaşmaları sebebiyle lineer ilişkiden saptıkları sonucuna ulaşılır (Bkz. Ek 11).



Şekil 5.2. (5.1) modeli için denekler-içi kovaryans matris yapısına ait rezidülerin grafiği

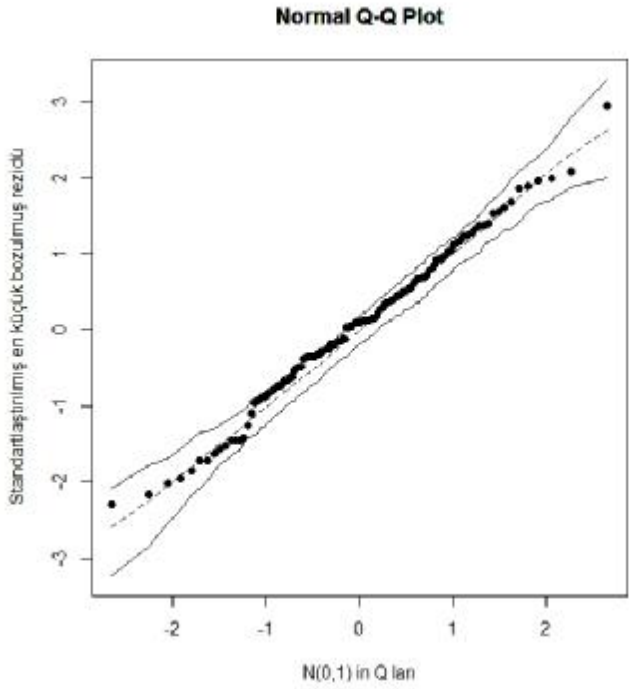
Şekil 5.2 incelendiğinde, 12 ve 29 numaralı deneklerin modele iyi uyum sağlayamayan kovaryans yapısına sahip oldukları sonucuna varılır (Bkz. Ek 11).

Merkezden uzakta duran gözlemlerin belirlenebilmesi ve koşullu hataların normallik varsayımının değerlendirilebilmesi amacıyla, sırasıyla, Şekil 5.3 ile Şekil 5.4 incelenebilir.



Şekil 5.3. (5.1) modeli için standartlaştırılmış koşullu rezidülerin grafiği

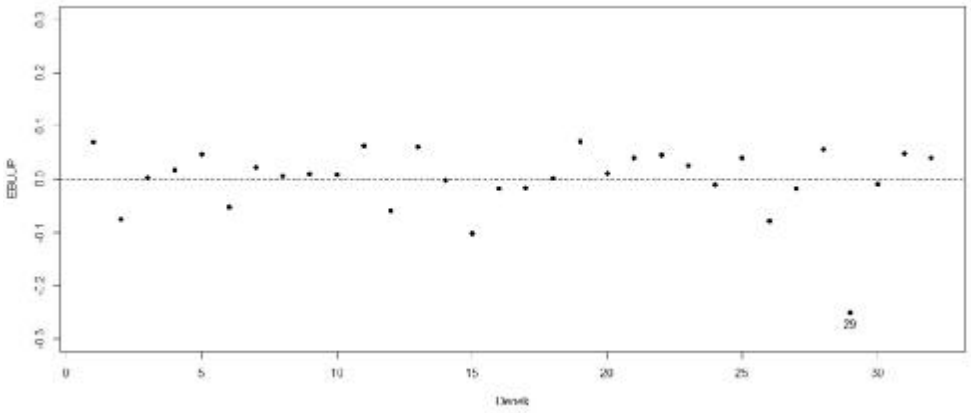
Şekil 5.3 incelendiğinde, 46 (12 numaralı deneğin ikinci seansına karşılık gelen 12.2 numaralı gözlem) ve 116 (29 numaralı deneğin dördüncü seansına karşılık gelen 29.4 numaralı gözlem) numaralı gözlemlerin grafikte bulunan diğer standartlaştırılmış koşullu rezidülerden uzakta durdukları ve dolayısıyla, sapan değer olabilecekleri sonucuna varılır (Bkz. Ek 11).



Şekil 5.4. (5.1) modeli için standartlaştırılmış en küçük bozulmuş rezidülerin 95% lik güven aralığının grafiği

Şekil 5.4 incelendiğinde, güven sınırlarının dışında herhangi bir gözlem bulunmaması nedeniyle, koşullu hataların normallik varsayımının doğru olduğu sonucuna ulaşılır.

Merkezden uzakta duran deneklerin belirlenebilmesi amacıyla, Şekil 5.5 ile verilen denek indekslerine karşı EBLUP değerlerinin grafiği incelenebilir.



Şekil 5.5. (5.1) modeli için EBLUP grafiği

Şekil 5.5 incelendiğinde, 29 numaralı deneğin sapan bir değer olduğu sonucuna varılır (Bkz. Ek 11).

EBLUP aynı zamanda rastgele etkilerin normallik varsayımının kontrol edilmesinde de kullanılabiliyordu. Ancak rastgele etkilerin normallik varsayımının kontrolü amacıyla kullanılan ağırlıklandırılmış normal quantil grafiğinin uygulanması kolay değildir. Bununla birlikte normallik varsayımı, rastgele etkilerin dağılımlarının hatalı olarak belirlenmesinin BLUE değerini etkilemediği göz önünde bulundurulduğunda, çok da önemli değildir.

Sonuç olarak, Şekil 5.1-5.5 ile verilen grafiklerin ortak özelliklerinin 12 ve 29 numaralı deneklerin diğer deneklerden uzakta olmalarıdır, denilebilir. Bu nedenle, 12 ve 29 numaralı denekler üzerinde ayrıntılı bir inceleme yapılacaktır.

Çizelge 5.7. 12 ve 29 numaralı denekler modelde iken ve modelde değilken tahminler (± tahmin edilen standart hatalar) ve model parametrelerine ilişkin değişimler

Parametreler	α_0	α_1	β	τ^2	σ^2
Tüm veri	-0.32±0.03	-0.21±0.03	1.06±0.06	0.00634±0.0028	0.021±0.02
Denek #12 kaldırıldığında	-0.32±0.03 (0.0%)	-0.22±0.03 (-4.8%)	1.06±0.06 (0.0%)	0.00691±0.0027 (-9.0%)	0.015±0.02 (**28.6**%)
Denek #29 kaldırıldığında	-0.33±0.03 (0.0%)	-0.19±0.03 (9.5%)	1.07±0.05 (0.9%)	0.00148±0.0013 (**76.7**%)	0.017±0.02 (**-**19.1%)
Denek #12 ve Denek #29 kaldırıldığında	-0.32±0.03 (0.0%)	-0.19±0.03 (9.5%)	1.07±0.05 (0.9%)	0.00299±0.0014 (**52.8**%)	0.012±0.01 (**42.9**%)

Çizelge 5.8. 12 ve 29 numaralı denekler modelde iken ve modelde değilken AIC ve BIC değerleri

Parametreler	AIC	BIC
Tüm veri	-83.12	-68.12
Denek #12 kaldırıldığında	-107	-92.9
Denek #29 kaldırıldığında	-116.3	-102.2
Denek #12 ve Denek #29 kaldırıldığında	-149	-135.1

Çizelge 5.7 de, verinin tamamının kullanılmasıyla elde edilen tahminlerle birlikte 12 ve 29 numaralı denekler analize dahil edilmediğinde elde edilen parametre tahminleri ve model parametrelerine ilişkin değişimler (parantez içinde verilen değerler) yer almaktadır. Parantez içinde verilen model parametrelerine ilişkin değişimler $\left[\left(\hat{\theta}_{(I)} - \hat{\theta}\right)/\hat{\theta}\right] \times 100\%$ eşitliği yardımıyla hesaplanır. Bu eşitlikte $\hat{\theta}_{(I)}$, I indeks kümesine ait gözlemler olmaksızın elde edilen parametre tahminlerini göstermektedir. Örneğin $\left[\left(\hat{\theta}_{(12)} - \hat{\theta}\right)/\hat{\theta}\right] \times 100\%$ eşitliği, 12 numaralı denek analize dahil edilmediğinde model parametrelerinde meydana gelen değişmeyi gösterir.

Çizelge 5.7 incelendiğinde, 12 numaralı deneğin σ^2 ve 29 numaralı deneğin de τ^2 üzerinde yani 12 ve 29 numaralı deneklerin kovaryans parametreleri üzerinde etkili oldukları sonucuna ulaşılır. Elde edilen bu sonuçlar, Şekil 5.1-5.5 ile verilen grafiklerden elde edilen sonuçlarla da bağdaşmaktadır.

Çizelge 5.8 de, verinin tamamının kullanılmasıyla elde edilen AIC ve BIC değerleriyle birlikte 12 ve 29 numaralı denekler analize dahil edilmediğinde elde edilen AIC ve BIC değerleri yer almaktadır.

Çizelge 5.8 incelendiğinde, 12 ve 29 numaralı deneklerin analize dahil edilmemesi durumunda daha küçük AIC ve BIC değerlerinin elde edildiği görülür. Küçük AIC ve BIC değerlerinin daha iyi bir modelin göstergesi oldukları daha önce ifade edilmişti. Dolayısıyla, 12 ve 29 numaralı deneklerin analize dahil edilmemesi durumunda daha iyi bir model elde edildiği sonucuna varılır.

Etkili deneklerin özelliklerinin değerlendirilmesi aşağıdaki gibi yapılabilir:

- Çizelge 5.1 e bakıldığında, 12 numaralı deneğin conventional diş fırçasını kullandığı görülmektedir. Denek, ikinci en düşük tedavi öncesi (0.71) ve tedavi sonrası (0.39) bakteri plak indeksine sahiptir. Denek, ikinci en düşük

tedavi sonrası bakteri plak indeksine sahip olmasına rağmen, conventional diş fırçasının ikinci seansında yüksek tedavi sonrası bakteri plak indekslerine de rastlanmaktadır. Monoblock diş fırçasını kullanan deneklerin tedavi sonrası daha yüksek bakteri plak indekslerine sahip olacakları düşünülmesine rağmen bu durum, uydurulan modelden elde edilen sonuçlarla çelişmekte ve conventional diş fırçasını kullanan deneklerin de tedavi sonrası yüksek bakteri plak indekslerine sahip olabilecekleri ifade edilebilmektedir. Sonuç olarak, conventional diş fırçasını kullanan 12 numaralı denek, tedavi sonrası bakteri plak indeksleri arasında en yüksek çeşitliliğe ve bakteri plak indeksinin ikinci en düşük azalışına $(y|x)$ sahiptir.

- Çizelge 5.1 e bakıldığında, 29 numaralı deneğin monoblock diş fırçasını kullandığı görülmektedir. Denek, en düşük tedavi sonrası bakteri plak indeksine (0.37) sahiptir. Monoblock diş fırçasının hemen hemen her yerde bulunabilmesi ve conventional diş fırçasına göre daha ucuz fiyatlı olması nedeniyle monoblock diş fırçasını kullananların, tedavi sonrası daha yüksek bakteri plak indeksine sahip olması beklenmektedir. Ancak bu durum, elde edilen sonuçlarla çelişmektedir. Sonuç olarak, monoblock diş fırçasını kullanan 29 numaralı denek, en düşük tedavi sonrası bakteri plak indeksine ve bakteri plak indeksinin en düşük azalışına $(y|x)$ sahiptir.

Bütün sonuçlar ele alındığında, en düşük tedavi sonrası bakteri plak indeksine ve bakteri plak indeksinin en düşük azalışına $(y|x)$ sahip olan, aynı zamanda, bütün marketlerde kolaylıkla bulunabilen ve daha ekonomik olan monoblock diş fırçasının, conventional diş fırçasına göre daha tercih edilebilir olduğu sonucuna varılır. Bununla birlikte, conventional diş fırçası ile monoblock diş fırçası arasındaki sayısal farklılık çok ciddi boyutlarda değildir. Conventional diş fırçasının monoblock diş fırçasına göre daha pahalı olmasının nedeni, piyasaya yeni çıkmış olması ve her yerde kolaylıkla bulunamaması olabilir.

5.2. Lineer Karma Modellerde Leverage Analizi Örneği

Brezilya'da bulunan Sao Paulo Üniversitesi'nin Diş Hekimliği Fakültesi'nde Nobre ve Singer (2011) tarafından yürütülen bir çalışma sonucunda elde edilen verinin analiz edilmesiyle, günlük kullanım sonucunda temizlenen bakteri plağı üzerinde "experimental" diş fırçasının (experimental toothbrush) etkisi değerlendirilmek istenmektedir. Dişlerin fırçalanmasından önce (before) ve sonra (after) yapılan iki değerlendirme seansında (session-occasion), yaşları 4 ile 6 arasında değişen 32 çocuktan elde edilen bakteri plak indekslerine karşılık gelen veri Çizelge 5.9 ile gösterilmektedir.

Nobre ve Singer (2011) tarafından rastgele eğimli model,

$$y_{ij} = \beta x_{ij} + u_i x_{ij} + \varepsilon_{ij} \tag{5.2}$$

olarak tanımlanmaktadır. Bu model, lineer karma model tiplerinden uzun süreli bir modeldir. (5.2) eşitliğinde, $y_{ij}(x_{ij})$ j-inci seansta değerlendirilen i-inci denek (subject-child) için tedavi sonrası (tedavi öncesi) elde edilen bakteri plak indeksini, β sabit eğimi ve u_i, eğimin rastgele bileşenini göstermektedir. $u_i \sim N(0, \tau^2)$ ve $\varepsilon_{ij} \sim N(0, \sigma^2)$ $i = 1, \cdots, 32$ ve $j = 1,2$ için birbirinden bağımsız rastgele değişkenler olup, sırasıyla, deneklerin rastgele etkilerine ve rastgele ölçüm hatalarına karşılık gelmektedirler.

(5.2) eşitliği ile verilen uzun süreli model, matris gösterimi yardımıyla $y = X\beta + Zu + \varepsilon$ formunda ifade edilir. Burada, y, 64×1 tipinde tedavi sonrası bakteri plak indeks değerlerinden oluşan yanıt vektörünü, X, 64×1 tipinde tedavi öncesi bakteri plak indeks değerlerinden oluşan sabit etkilere ilişkin tasarım matrisini ve Z, 64×32 tipinde esas köşegendeki elemanları tedavi öncesi bakteri plak indeks değerlerinden oluşan blok köşegen bir matris olup, rastgele etkilere ilişkin tasarım matrisini göstermektedir. Son olarak u, 32×1 tipinde rastgele etkilerden oluşan parametreler vektörü olup, $\begin{bmatrix} u_1 \\ \vdots \\ u_{32} \end{bmatrix}$ şeklindedir.

Çizelge 5.9. Bakteri plak indeksleri (Nobre ve Singer, 2011)

	First occasion		Second occasion	
Child	Before toothbrushing	After toothbrushing	Before toothbrushing	After toothbrushing
1	1.05	1.00	1.13	0.84
2	1.07	0.62	0.92	0.62
3	0.82	0.62	1.53	1.07
4	1.37	0.90	1.65	1.20
5	1.87	1.52	1.30	1.07
6	1.30	0.82	1.17	0.70
7	1.61	1.18	1.52	1.15
8	1.02	0.73	1.08	0.64
9	1.62	1.25	1.45	1.10
10	1.65	1.22	1.57	1.22
11	1.02	0.78	0.60	0.47
12	0.68	0.60	1.13	0.39
13	1.70	1.55	1.85	1.37
14	1.30	1.02	1.65	0.97
15	0.90	0.80	1.10	1.05
16	1.40	1.12	1.25	0.67
17	1.65	1.63	1.36	1.16
18	1.02	0.80	0.92	0.82
19	0.68	0.67	1.00	0.92
20	1.25	1.23	0.91	0.78
21	1.27	1.20	1.20	0.95
22	1.07	0.85	1.39	1.25
23	1.35	1.21	1.42	1.17
24	1.32	1.02	1.60	1.40
25	1.65	1.61	1.50	1.36
26	1.30	1.07	0.84	0.61
27	1.57	1.20	1.50	1.07
28	1.67	1.50	1.47	1.32
29	0.91	0.67	0.96	0.62
30	1.05	0.70	1.00	0.85
31	2.25	2.24	1.57	1.55
32	1.15	1.00	1.23	1.11

Çizelge 5.9 ile verilen bakteri plak indeks verisi ilk olarak Microsoft Excel programına kaydedilir (Bkz. Ek 12), daha sonra R programına yazılan kod yardımıyla çağrılır.

Çizelge 5.10. REML yardımıyla uydurulan lineer karma model sonuçları

AIC	BIC	LogLik	Deviance	REMLdev
-50.16	-43.69	28.08	-62.26	-56.16

Çizelge 5.11. Rastgele etkiler için uydurulan lineer karma model sonuçları

Groups Name	Variance	Std.Dev.
Subject (x)	0.0072249	0.084999
Residual	0.0134660	0.116043

Çizelge 5.12. Sabit etkiler için uydurulan lineer karma model sonuçları

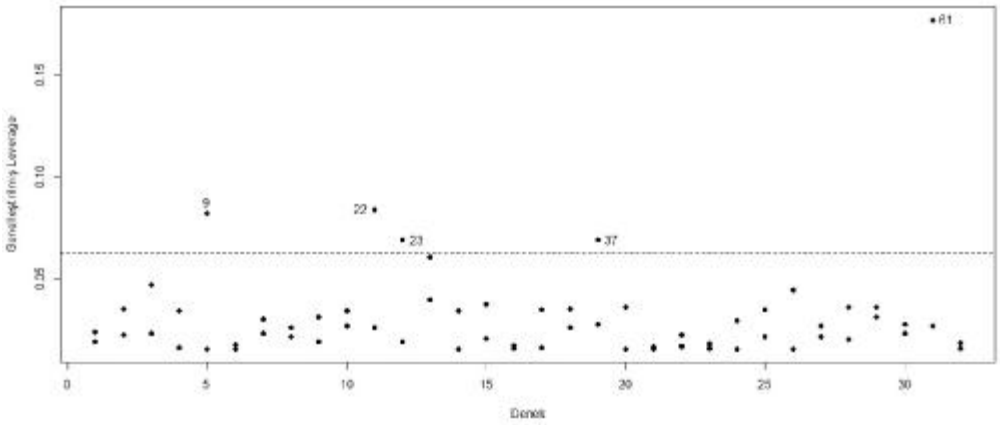
	Estimate	Std.Error	t value
x (β)	0.79160	0.01898	41.71

Çizelge 5.13. Uydurulan lineer karma model için ANOVA tablosu

	Df	Sum Sq	Mean Sq	F Value
X	1	23.429	23.429	1739.9

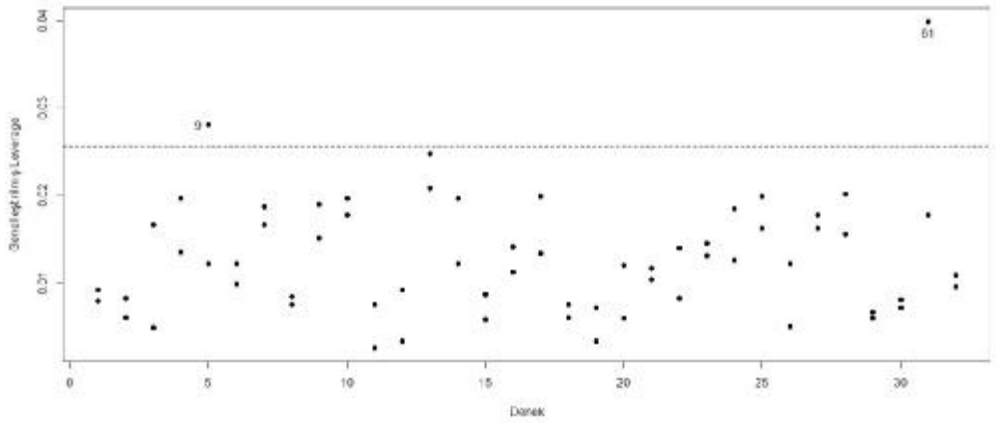
Uydurulan lineer karma modelin değerlendirilmesinde, sırasıyla, AIC (Akaike bilgi kriteri) ve BIC (Bayesian bilgi kriteri) den yararlanılabilir. Küçük AIC ve BIC değerleri, iyi bir şekilde uydurulmuş olan bir modelin göstergesidir.

(5.2) modeli altında, denekler ve gözlemler için genelleştirilmiş leverage grafikleri incelenecektir. İncelenen tüm grafiklerde, kesik çizgi yardımıyla gösterilen eşik noktaları, ortalama leverage değerinin iki katına karşılık gelmektedir.



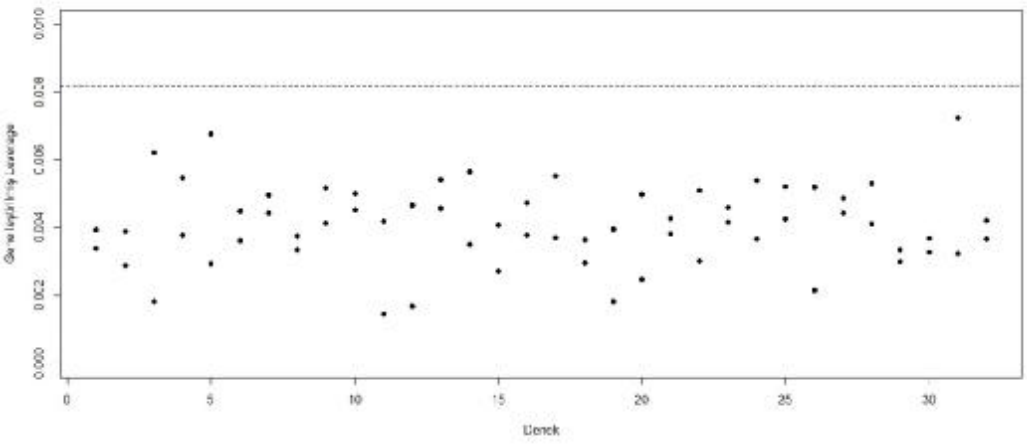
Şekil 5.6. (5.2) modeli altında marjinal uydurulmuş değerlerin genelleştirilmiş leverage matrisi olan L_1 in grafiği

Şekil 5.6 incelendiğinde, 9 (5 numaralı deneğin birinci seansına karşılık gelen 5.1 numaralı gözlem), 22, 23, 37 ve 61 (11 numaralı deneğin ikinci seansına karşılık gelen 11.2 numaralı gözlem, 12 numaralı deneğin birinci seansına karşılık gelen 12.1 numaralı gözlem, 19 numaralı deneğin birinci seansına karşılık gelen 19.1 numaralı gözlem ve 31 numaralı deneğin birinci seansına karşılık gelen 31.1 numaralı gözlem) numaralı gözlemlerin yüksek genelleştirilmiş marjinal leverage (L_1) değerine sahip olduğu sonucuna varılır (Bkz. Ek 13).



Şekil 5.7. (5.2) modeli altında rastgele etkilerin varlığıyla açıklanan birimler içi değişkenliğin parçası olan L_2 nin grafiği

Şekil 5.7 incelendiğinde, 9 (5 numaralı deneğin birinci seansına karşılık gelen 5.1 numaralı gözlem) ve 61 (31 numaralı deneğin birinci seansına karşılık gelen 31.1 numaralı gözlem) numaralı gözlemlerin yüksek genelleştirilmiş rastgele bileşen leverage (L_2) değerine sahip olduğu sonucuna ulaşılır (Bkz. Ek 13).



Şekil 5.8. (5.2) modeli altında koşullu uydurulmuş değerlerin rastgele etki bileşeni için genelleştirilmiş leverage matris olan H_2 nin grafiği

Şekil 5.8 incelendiğinde, eşik noktasının $2trace(H_2)/n = 0.008171392$ olarak bulunması nedeniyle tüm gözlemlerin ve deneklerin eşik noktasının altında

kaldığı, dolayısıyla H_2 grafiğinin yüksek leverage noktaları göstermediği sonucuna varılır. Bu durum, H_2 deki gözlemlerin marjinal uydurulmuş değerler için yüksek leverage olan gözlemlerin varlığı tarafından maskelenmesi sonucunda meydana gelmiş olabilir. Dolayısıyla, koşullu uydurulmuş değerlerin rastgele etki bileşeni için yüksek etkili gözlemlerin belirlenmesinde, H_2 yerine L_2 yi kullanmanın daha doğru sonuçlara götüreceği söylenebilir (Bkz. Ek 13).

5.3. Lineer Karma Modellerde Etki Analizi Örneği

Beckman ve ark. (1987), New Mexico'da kurulu ABD Enerji Bakanlığı'na bağlı ulusal bir laboratuar olan Los Alamos tarafından yüksek etkili partikül hava (HEPA-High Efficiency Particular Air) filtreleri üzerinde yapılan çalışma sonucunda elde edilen verinin bir kısmını kullanarak, lineer karma modellerde yerel etki tanılamalarını geliştirmişlerdir. Bu çalışmada, lineer karma modeldeki etki tanılamalarının incelenmesi amacıyla Beckman ve ark. (1987) tarafından kullanılan aerosol penetrasyon verisi Matlab yardımıyla analiz edilecektir.

HEPA filtreleri havayı hem virüslerden, solunabilir tozlardan, polen ve duman partiküllerinden hem de bakterilerden, çeşitli zehirli tozlardan ve aerosollerden korumak amacıyla oluşturulan filtrelerdir. Bu çalışmanın amacı, filtrelerin penetrasyon (duman, toz ve benzeri yapıları çekim gücü) çeşitliliğine en fazla katkıda bulunan faktörleri belirlemek ve kalite güvenlik testinde kullanılması amacıyla standart aerosol ile yer değiştirebilecek alternatif diğer bir aerosol tipi elde etmektir. Bu amaçla, iki filtre imalatçısı tarafından üretilen iki aerosol tipinden elde edilen yüzdelik penetrasyonlar üzerinde analiz yapılacaktır. Her bir filtre imalatçısı, iki aerosol tipinin yüzdelik penetrasyonlarını değerlendirirken üç filtre tipinden yararlanacaktır. Aerosollerin penetrasyon yüzdelerinin çok büyük olması, filtrelerin kalite güvenlik testinde başarısız olduğu anlamına gelmektedir. Veri, Çizelge 5.14 ile gösterilmektedir.

Beckman ve ark. (1987) tarafından aerosol verisi için lineer karma model,

$$y_{ijkl} = \mu + A_i + M_j + F_{jk} + \varepsilon_{ijkl} \tag{5.3}$$

olarak tanımlanmaktadır. Bu model, lineer karma model tiplerinden varyans bileşenleri modelinin bir örneğidir. (5.3) eşitliğinde, y_{ijkl} yüzdelik penetrasyonu, A_i aerosol için sabit etkiyi, M_j imalatçı için sabit etkiyi, F_{jk} j-inci imalatçının ürettiği k-inci filtre için rastgele etkiyi ve ε_{ijkl}, ijk-inci hücredeki l-inci ölçüme karşılık gelen hatayı göstermektedir. Ayrıca model, i = 1,2 ve j = 1,2 için $\sum_i A_i = \sum_j M_j = 0$ kısıtlamalarına sahiptir ve $F_{jk} \sim N(0, \sigma_F^2)$ ve $\varepsilon_{ijkl} \sim N(0, \sigma_e^2)$ şeklindedir. Aerosoller ile imalatçılar sabit etkileri ve imalatçılar içinde iç içe geçmiş filtreler de rastgele etkileri oluşturmaktadır.

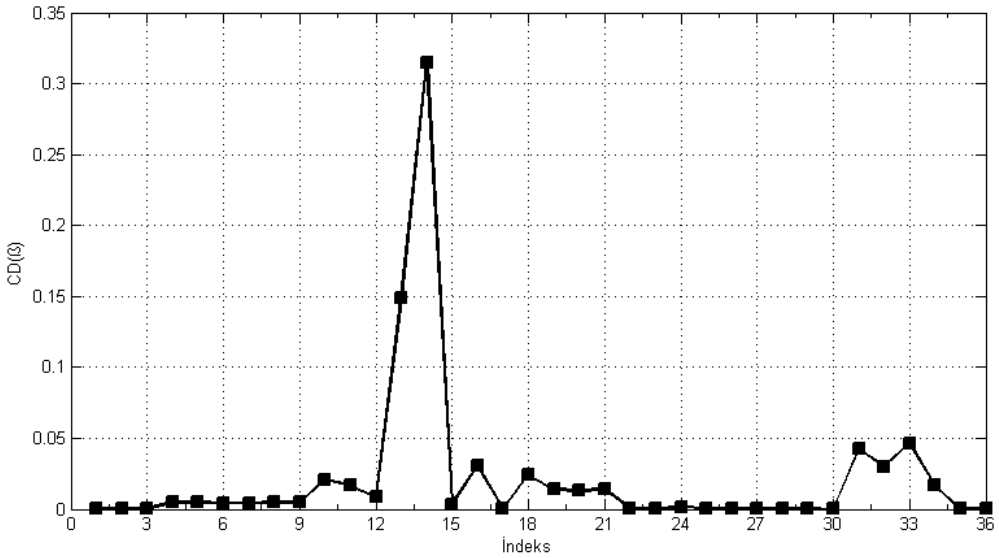
(5.3) eşitliği ile verilen varyans bileşenleri modeli, matris gösterimi yardımıyla $y = X\beta + Zu + \varepsilon$ formunda ifade edilir. Burada, y, 36×1 tipinde yüzdelik penetrasyon değerlerinden oluşan yanıt vektörünü, X, 36×3 tipinde sabit etkilere ilişkin tasarım matrisini ve Z, 36×6 tipinde rastgele etkilere ilişkin tasarım matrisini göstermektedir. X matrisinin ilk sütunu 1 değerlerinden oluşan μ lere, ikinci sütunu sırasıyla 1 ve -1 değerlerinden oluşan Aerosol 1 ve Aerosol 2 lere, üçüncü sütunu ise sırasıyla -1 ve 1 değerlerinden oluşan İmalatçı 1 ve İmalatçı 2 lere karşılık gelmektedir. $\sum_i A_i = \sum_j M_j = 0$ kısıtlamaları nedeniyle, X matrisinin ikinci ve üçüncü sütunu 1 ve -1 değerlerinden oluşmaktadır. Z matrisinin ilk sütunu Filtre 1 e, ikinci sütunu Filtre 2 ye, üçüncü sütunu Filtre 3 e, dördüncü sütunu Filtre 4 e, beşinci sütunu Filtre 5 e ve altıncı sütunu da Filtre 6 ya karşılık gelmektedir. β, 3×1 tipinde sabit etkilerden oluşan parametreler vektörü olup, $\begin{bmatrix} \mu \\ A_1 \\ M_1 \end{bmatrix}$ şeklinde ifade edilir. Son olarak u, 6×1 tipinde rastgele etkilerden oluşan parametreler vektörü olup, $\begin{bmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \end{bmatrix}$ şeklindedir.

Çizelge 5.14. İmalatçıların yararlandığı üç filtre tipinin kullanılması ile iki aerosolden elde edilen yüzdelik penetrasyonlar (Beckman ve ark., 1987)

Aerosol 1 (i=1)		Aerosol 2 (i=2)	
İmalatçı 1 (j=1)	İmalatçı 2 (j=2)	İmalatçı 1 (j=1)	İmalatçı 2 (j=2)
Filtre 1 (k=1)	Filtre 4 (k=1)	Filtre 1 (k=1)	Filtre 4 (k=1)
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 (k=2)	Filtre 5 (k=2)	Filtre 2 (k=2)	Filtre 5 (k=2)
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 (k=3)	Filtre 6 (k=3)	Filtre 3 (k=3)	Filtre 6 (k=3)
0.082	1.000	0.150	2.170
0.076	1.800	0.120	1.520
0.077	2.700	0.120	1.580

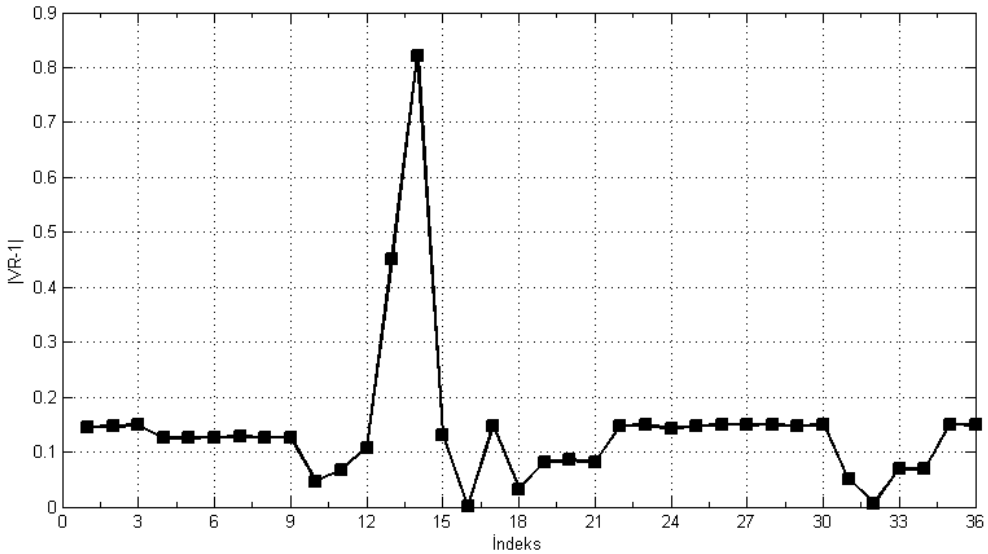
Çizelge 5.14 ile verilen veri Matlab yardımıyla analiz edilmiştir. Çalışmanın amacının lineer karma modellerde tahmin prosedürleri olmaması nedeniyle, D nin bilindiği varsayımı altında koşullu rezidülerin kullanılmasıyla yapılan analiz sonucunda $\hat{\beta} = \begin{bmatrix} \hat{\mu} \\ \hat{A}_1 \\ \widehat{M}_1 \end{bmatrix} = \begin{bmatrix} 0.9920 \\ 0.1969 \\ 0.5970 \end{bmatrix}$, $\hat{u} = \begin{bmatrix} 0.3123 \\ -0.1482 \\ -0.1640 \\ -0.4408 \\ 0.3250 \\ 0.1157 \end{bmatrix}$, $\hat{\sigma}_F^2 = 0.1176$ ve $\hat{\sigma}_e^2 = 0.5475$ olarak elde edilir. Aynı zamanda, $var(u) = \begin{bmatrix} 0.1176 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0.1176 \end{bmatrix}_{6\times 6}$ ve $var(\varepsilon) = 0.5475 \times I_{36}$ şeklinde olduğu da ifade edilebilir.

(5.3) modeli altında, sırasıyla, sabit ve rastgele etkili parametre tahminleri ve likelihood fonksiyonu üzerindeki etki ölçümlerinin indeks grafikleri incelenecektir.



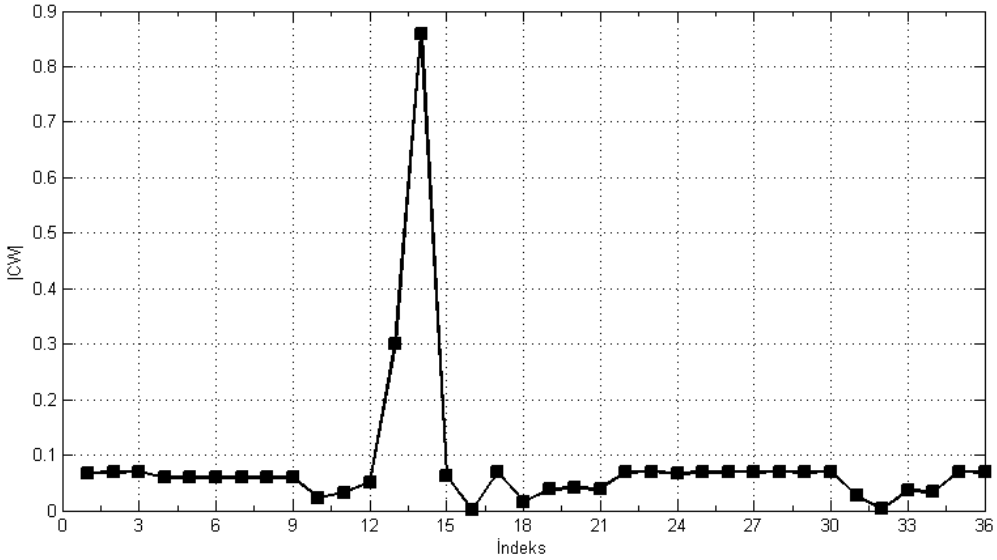
Şekil 5.9. (5.3) modeli altında $CD(\beta)$ nın indeks grafiği

Şekil 5.9 incelendiğinde, 13 ve 14 numaralı gözlemlerin büyük $CD(\beta)$ değerlerine sahip olmaları nedeniyle $\hat{\beta}$ üzerinde etkili olan noktalar oldukları sonucuna ulaşılır (Bkz. Ek 14).



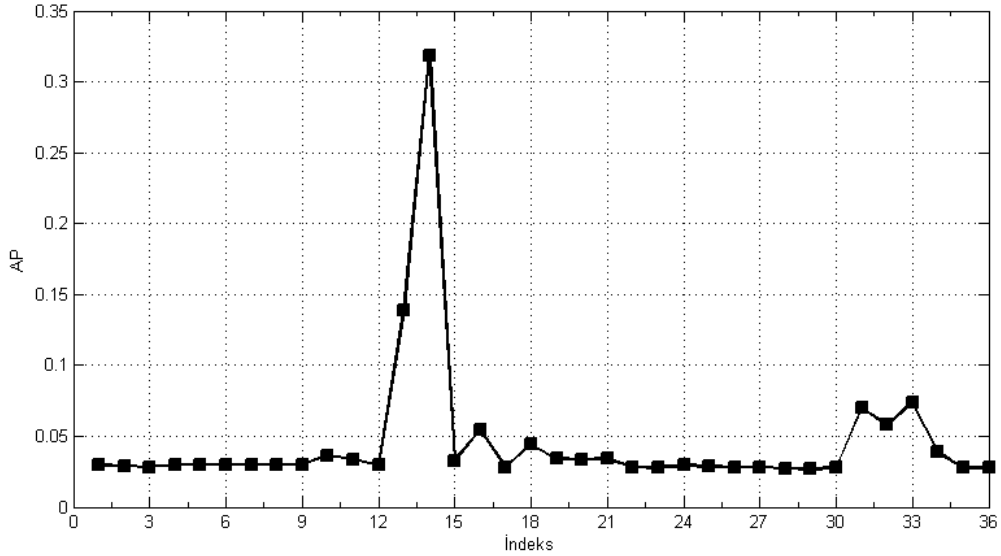
Şekil 5.10. (5.3) modeli altında $|VR-1|$ nın indeks grafiği

Şekil 5.10 incelendiğinde, 13 ve 14 numaralı gözlemlerin $\hat{\beta}$ nın varyansı üzerinde etkili olan noktalar oldukları söylenebilir. Çünkü $|VR-1|$ istatistiğinin büyük değerleri, $\hat{\beta}$ nın varyansı üzerinde etkili olan noktaları göstermekteydi (Bkz. Ek 14).



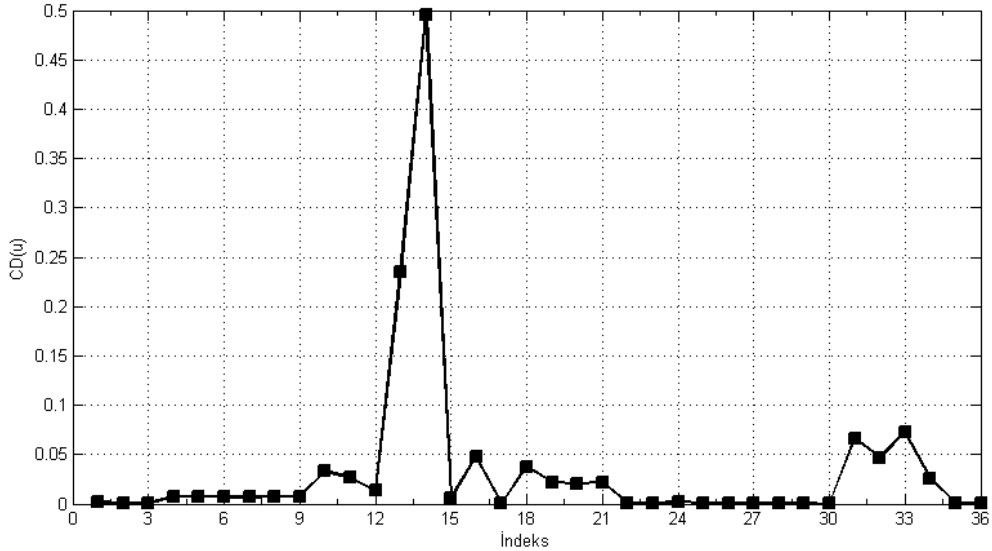
Şekil 5.11. (5.3) modeli altında $|CW|$ nun indeks grafiği

Şekil 5.11 incelendiğinde, 13 ve 14 numaralı gözlemlerin $\hat{\beta}$ veya $Var(\hat{\beta})$ üzerinde etkili olan noktalar oldukları sonucu elde edilir. Çünkü CW un işareti göz ardı edildiğinde, büyük $|CW|$, $\hat{\beta}$ veya $Var(\hat{\beta})$ üzerinde etkili olan noktaları göstermekteydi ve büyük n değerleri için $CW_i \approx \frac{1}{2} log(VR_i(\beta))$ olarak ifade edilmekteydi (Bkz. Ek 14).



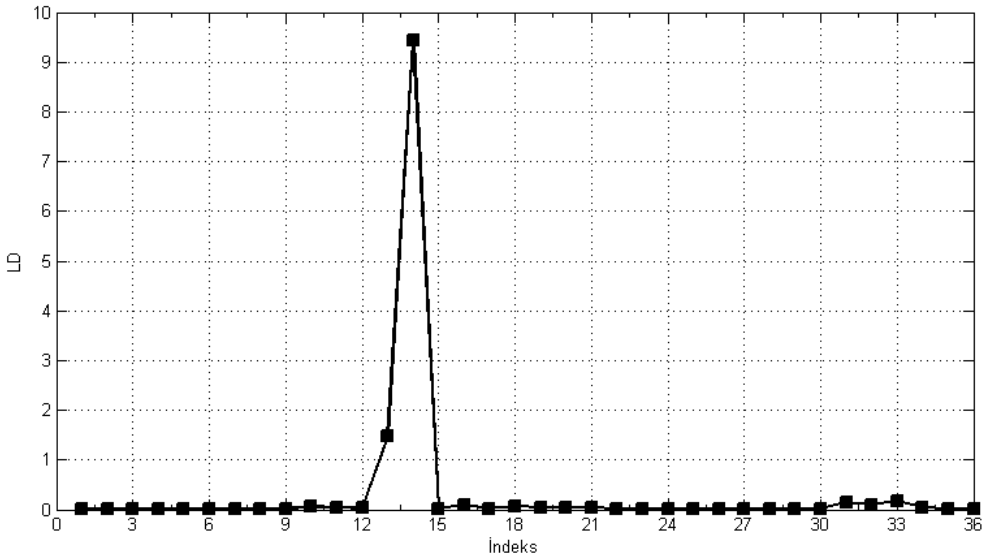
Şekil 5.12. (5.3) modeli altında AP nin indeks grafiği

Şekil 5.12 incelendiğinde, AP nin büyük değerlerinin etkili olan noktaları göstermesi sebebiyle 13 ve 14 numaralı gözlemlerin etkili olan noktalar oldukları söylenebilir (Bkz. Ek 14).



Şekil 5.13. (5.3) modeli altında $CD(u)$ nun indeks grafiği

Şekil 5.13 incelendiğinde, 13 ve 14 numaralı gözlemlerin büyük $CD(u)$ değerlerine sahip olmalarından dolayı rastgele etkili parametrelerin ön tahmini üzerinde etkili olan noktalar oldukları sonucuna ulaşılır (Bkz. Ek 14).



Şekil 5.14. (5.3) modeli altında LD nin indeks grafiği

Şekil 5.14 incelendiğinde, 13 ve 14 numaralı gözlemlerin likelihood fonksiyonu üzerinde etkili olan noktalar oldukları söylenebilir. Çünkü LD nin büyük değerleri etkili olan noktaları göstermekteydi (Bkz. Ek 14).

Sabit ve rastgele etkili parametre tahminleri ve likelihood fonksiyonu üzerindeki etki ölçümlerinin indeks grafiklerinin incelenmesi sonucunda 14 numaralı gözlemin en etkili, 13 numaralı gözlemin ise ikinci en etkili nokta olduğu açıkça görülmektedir.

6. SONUÇLAR VE ÖNERİLER

Bu çalışmanın konusunu, lineer karma modellerdeki rezidü ve etki analizi oluşturmaktadır.

Bu çalışmada ilk olarak, lineer regresyonda rezidü ve etki analizi çalışmalarına yer verilmiş, daha sonra lineer karma model ile hiyerarşik lineer model açıklanmış ve lineer regresyonda rezidü ve etki analiziyle ilgili tanımlanan ölçütler lineer karma model için verilmiştir.

Lineer karma modellerde yer alan etki tanılamaları için eşik noktaları geliştirilmemiştir. Aynı zamanda, hiyerarşik lineer modellerde etki analizi ile ilgili de literatürde yeterli düzeyde bir çalışmaya ulaşılamamaktadır. Dolayısıyla, lineer karma modellerdeki etki tanılamaları için eşik noktalarının belirlenmesi ve hiyerarşik lineer modellerde etki analizi gelecek çalışmaların konusunu oluşturabilir.

Ayrıca, lineer karma modellerde çoklu iç ilişki olması durumunda rezidü ve etki analizi ile ilgili de gelecekte çalışmalar yapılabilir.

KAYNAKLAR

ANDREWS, D. F., ve PREGIBON, D., 1978. Finding Outliers that Matter. Journal of Royal Statistical Society (B), 40: 85-93.

BARNETT, V., ve LEWIS, T., 1995. Outliers in Statistical Data. Wiley, New York.

BECKMAN, R. J., NACHTSHEIM, C. J., ve COOK, R. D., 1987. Diagnostics for Mixed Model Analysis of Variance. Technometrics, 29: 413-426.

BELSLEY, D. A., KUH, E., ve WELSCH, R. E., 1980. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. Wiley, New York.

BOX, G. E. P., ve JENKINS, G., 1970. Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco.

CHATTERJEE, S., HADI, A. S., ve PRICE, B., 2000. Regression Analysis by Example. New York, John Wiley and Sons.

CHATTERJEE, S., ve HADI, A. S., 1986. Influential Observations, High Leverage Points, and Outliers in Linear Regression. Statistical Science, 3:379-416.

CHATTERJEE, S., ve HADI, A. S, 1988. Sensitivity Analysis in Linear Regression. New York, John Wiley and Sons.

CHRISTENSEN, R., JOHNSON, W., ve PEARSON, L. M., 1992. Prediction Diagnostics for Spatial Linear Models. Biometrika, 79: 583-591.

CHRISTENSEN, R., JOHNSON, W., ve PEARSON, L. M, 1992. Case-deletion Diagnostics for Mixed Models. Technometrics, 34: 38-45.

COOK, R. D., 1977. Detection of Influential Observations in Linear Regression. Technometrics, 19: 15-18.

COOK, R. D., ve WEISBERG, S., 1980. Characterization of an Empirical Influence Function for Detecting Influential Cases in Regression. Technometrics, 22: 495-508.

COOK, R. D., ve WEISBERG, S, 1982. Residuals and Influence in Regression. Chapmann and Hall, London.

COX, D. R., ve HINKLEY, D. V., 1974. Theoretical Statistics. Chapman and Hall, London.

DATTA, G. S., ve LAHIRI, P., 2000. A Unified Measure of Uncertainty Estimated Best Linear Unbiased Predictors in Small Area Estimation Problems. Statist. Snica, 10: 613-627.

DE LEEUW, J.,ve KREFT, I. G. G., 1986. Random Coefficient Models for Multilevel Analysis. Journal of Educational Statistics, 11: 57-86.

DEMIDENKO, E., 2004. Mixed Models: Theory and Applications. John Wiley and Sons, New York.

DEMIDENKO, E., ve STUKEL, T. A., 2005. Influence Analysis for Linear Mixed-Effects Models. Stat. Med., 24: 893-909.

DIGGLE, P. J., LIANG, K. Y., ve ZEGER, S. L., 1996. Analysis of Longitudinal Data. Oxford Univ. Press.

GALPHIN, J., ve HAWKINS, D., 1981. Rejection of a Single Outlier in Two or Three Way Table. Technometrics, 23: 65-70.

GEISSER, S., 1992. Predictive Inference: An Introduction. Chapman ve Hall, New York.

GOLUB, G. H., ve VAN LOAN, C. F., 1983. Matrix Computations. The John Hopkins University Press.

GRAY, J. B., 1983. The L-R Plot, a Graphical Tool for Assessing Influence. Proceedings of the Statistical Computing Section, Amer. Statist. Asso., 159-164.

GRAY, J. B, 1986. A Simple Graphic for Assessing Influence in Regression. Journal of the Statistical Computation and Simulation, 24: 121-134.

HADI, A. S., 1992. A New Measure of Potential Influence in Linear Regression. Computational Statistics, Data Analysis, 14: 1-27.

HAMPER, F. R., 1974. The Influence Curve and its Role in Robust Estimation. Journal of the American Statistical Assocation, 62: 1179-1186.

HARTLEY, H. O., ve RAO, J. N. K., 1967. Maksimum Likelihood Estimation for the Mixed Analysis of Variance Model. Biometrika, 54: 93-108.

HARVILLE, D. A., 1977. Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. Journal of the American Statistical Association, 72: 320-340.

HARVILLE, D. A, 1990. BLUP and Beyond (D. Gianola, K. Hammond, Edts.). Statistical Methods for Genetic Improvement of Livestock. Springer, New York, 239-272.

HASLETT, J., ve HAYES, K., 1998. Residuals for the Linear Model with General Covariance Structure. Journal of Royal Statistical Society Series B, 60(1): 201-215.

HEMMERLE, W. J., ve HARTLEY, H. O., 1973. Computing MLE for the Mixed A. O. V. Model Using the W Transformation. Technometrics, 15(4): 819-831.

HENDERSON, C. R., 1953. Estimation of Variance and Covariance Components. Biometrics, 9: 226-252.

HENDERSON, C. R, 1975. Best Linear Unbiased Estimation and Prediction under a Selection Model. Biometrics, 31: 423-447.

HILTEN-MINTON, J. A., 1995. Multilevel Diagnostics for Mixed and Hierarchical Linear Models - PhD Thesis. University of California, Los Angeles.

HURTADO, G. I., 1993. Detection of Influential Observations in Linear Mixed Models - PhD Thesis. Department of Statistics, North Carolina State University.

JENNRICH, R. I., ve SAMPSON, P. F., 1976. Newton Raphson and Related Algorithims for Maximum Likelihood Variance Component Estimation. Technometrics, 18(1): 11-17.

JIANG, J., 2007. Linear and Generalized Linear Mixed Models and Their Applications. Springer Series in Statistics, New York.

LANGE, N., ve RYAN, L., 1989. Assessing Normality in Random Effects Models. The Annals of Statistics, 17: 624-642.

LAWRANCE, A. J., 1990. Local and Deletion Influence (W. A. Stahel, S. Weisberg, Edts.). In Robust Statistics and Diagnostics. Springer Verlag, Part I: 141-157.

LESAFFRE, E., ve VERBEKE, G., 1998. Local Influence in Linear Mixed Models. Biometrics, 54: 570-582.

LONGFORD, N. T., 2001. Simulation-Based Diagnostics in Random-Coefficient Models. Journal of the Royal Statistical Society A, 164: 259-273.

MARTIN, R. J., 1992. Leverage, Influence and Residuals in Regression Models When Observations are Correlated. Common Stat Theory Methods, 21: 1183-1212.

MCCULLOCH, C. E., ve SEARLE, S. R., 2001. Generalized, Linear and Mixed Models. John Wiley and Sons, New York.

MCLEAN, R. A., SANDERS, W. L., ve STROUP W. W., 1991. A Unified Approach to Mixed Linear Models. Am Stat, 45: 54-64.

MILLER, R. G., 1974. An Unbalanced Jackknife. Ann. Statist., 2: 880-891.

MONTGOMERY, D. C., PECK, E. A., ve VINING, G. G., 2001. Introduction to Linear Regression Analysis. New York, John Wiley and Sons.

NOBRE, J. S., 2004. Métodos de diagnóstico para modelos lineares mistos (in Portuguese), M. Sc. dissertation. Instituto de Mathemática e Estatistica, Universidade de São Paulo, São Paulo.

NOBRE, J. S., ve SINGER, J. M., 2007. Residual Analysis for Linear Mixed Models. Biometrical Journal, 49(6): 863-875.

NOBRE, J. S., ve SINGER, J. M, 2011. Leverage Analysis for Linear Mixed Models. Journal of Applied Statistics, 5: 1063-1072.

OMAN, S. D., 1995. Checking the Assumptions in Mixed-Model Analysis of Variance: A Residual Analysis Approach. Computational Statistics and Data Analysis, 20: 309-330.

PINHEIRO, J. C., ve BATES, D. M., 2000. Mixed-effects in S and S-PLUS. Springer, New York.

PREGİBON, D., 1981. Logistic Regression Diagnostics. Annal of Statistics, 9: 705-724.

ROBINSON, G. K., 1991. That BLUP is a Good Thing: The Estimation of Random Effects (with Discussion). Stat Sci, 6: 15-51.

SARIPINAR, E., 2004. Filbert Matrislerinin Normları, Yüksek Lisans Tezi. Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü, Isparta.

SATMAN, M. H., 2005. Doğrusal Regresyonda Aykırı Gözlemlerin Teşhis Yöntemleri, Yüksek Lisans Tezi. İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.

SEARLE, S. R., 2006. Matrix Algebra Useful for Statistics. Ithaca, John Wiley and Sons, New York.

SEARLE, S. R., CASELLA, G., ve MCCULLOCH, C. E., 1992. Variance Components. Ithaca, John Wiley and Sons, New York.

SMITH, A. F. M., 1973. A General Bayesian Linear Model. Journal of Royal Statistical Society, Series B, 35: 67-75.

ŞAHİNLER, S., ve BEK, Y., 2006. A Classification of Single Influential Observation Statistics in Regression Analysis. Kırgızistan-Türkiye Manas Üniversitesi, Fen Bilimleri Dergisi, 7: 1-18.

TAN, F. E. S., OUWENS, M. J. N., ve BERGER, M. P. F., 2001. Detection of Influential Observations in Longitudinal Mixed Effects Regression Models. Statistician, 50: 271-284.

TOPÇUBAŞI, A. S., 2001. Yanlı Regresyon Kestiriminde Sapan Değerlerin Belirlenmesi için Tanılama Yöntemleri, Yüksek Lisans Tezi. Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Adana.

TÜRKAN, S., 2008. Karışık Doğrusal Modellerde Artık Analizi ve Etki Analizi, Yüksek Lisans Tezi. Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

VERBEKE, G., ve LESAFFRE, E., 1996a. A Linear Mixed-Effects Model with Heterogeneity in the Random-Effects Population. Journal of the American Statistical Association, 91: 217-221.

WATERNAUX, C., LAIRD, N. M., ve WARE, J. H., 1989. Methods for Analysis of Longitudinal Data: Blood-Lead Concentrations and Cognitive Development. Journal of the American Statistical Association, 84: 33-41.

WEI, B. C., HU, Y. Q., ve FUNG, W. K., 1998. Generalized Leverage and its Applications. Scand. J. Stat., 25: 25-37.

WEISS, R. E., ve LAZARO, C. G., 1992. Residual Plots for Repeated Measures. Statistics in Medicine, 11: 115-124.

WELSLEY, R. E., 1982. Influence Functions and Regression Diagnostics. Modern Data Analysis, New York.

ZEWOTIR, T., ve GALPHIN, J. S., 2005. Influence Diagnostics for Linear Mixed Models. Journal of Data Science, 3: 153-177.

ZEWOTIR, T., ve GALPHIN, J. S, 2007. A Unified Approach on Residuals, Leverages and Outliers in the Linear Mixed Models. Springer, 16: 58-75.

ÖZGEÇMİŞ

23/08/1989 yılında İskenderun'da doğdu. İlk, orta ve lise öğrenimini İskenderun'da tamamladı. 2007 yılında Çukurova Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü'nde lisans öğrenimine başladı. 2008 yılında Çukurova Üniversitesi Fen Edebiyat Fakültesi Matematik Bölümü'nde çift anadal yapmaya hak kazandı. 2011 yılında Çukurova Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü'nden bölüm ikincisi olarak mezun oldu. Aynı yıl Çukurova Üniversitesi Fen Edebiyat Fakültesi Matematik Bölümü'nden de mezun olduktan sonra, Çukurova Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü'nde yüksek lisans öğrenimine başladı.

EKLER

EK 1. Mahalanobis Uzaklığı (Satman, 2005)

Mahalanobis uzaklığı, verinin varyans-kovaryans yapısını da göz önüne alan ve tek bir aykırı gözlemin testi için kullanılabilecek olan bir uzaklık formülü tanımlar. Bu uzaklık, bir regresyon denkleminden hesaplanılan kalıntılar üzerinden değil çok değişkenli veri üzerinden hesaplanır. X gözlem değerlerinden oluşan veri matrisi, $\bar{x}$ veriden hesaplanan ortalama vektörü ve $\mathcal{S}$ de aynı veriden hesaplanan örneklem varyans-kovaryans matrisi olmak üzere i-inci gözlem için Mahalanobis uzaklığı,

$$M_i = \sqrt{(x_i - \bar{x})^T \mathcal{S}^{-1} (x_i - \bar{x})}$$

olarak tanımlanır. Mahalanobis uzaklığının Frobenius yani Euclidean normundan farkı, veri kümesindeki korelasyonları da içermesidir.

EK 2. Determinantla İlgili Teorem (Searle, 2006)

Teorem: $X = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ şeklinde parçalanmış matris olsun. Bu durumda,

a) Eğer A tekil olmayan bir matris ise $det(X) = det(A)det(D - CA^{-1}B)$

b) Eğer D tekil olmayan bir matris ise $det(X) = det(D)det(A - BD^{-1}C)$

olacaktır.

EK 3. Sherman-Morrison-Woodbury Teoremi (Golub ve Van Loan, 1983)

A, tekil olmayan bir matris ve c, d sütun vektörleri olsun. Buna göre;

$$(A - cd^T)^{-1} = A^{-1} + \frac{A^{-1}cd^T A^{-1}}{1 - d^T A^{-1}c}$$

dir. i-inci gözlem çıkarıldıktan sonra elde edilen $W_{[i]}(X,X)$ matrisi,

$$W_{[i]}(X,X) = W(X,X) - \frac{\check{x}_i \check{x}_i^T}{m_i}$$

şeklinde, Bölüm 3.1.3'de ifade edilmişti. $W_{[i]}(X,X)^{-1}$ matrisi, Sherman-Morrison-Woodbury Teoremi yardımıyla ve

$$A = W(X,X)\,, c = \check{x}_i\,, d = \frac{\check{x}_i^T}{m_i}$$

alınmasıyla, $W_{[i]}(X,X)^{-1}W(X,X)$ ifadesi,

$$W_{[i]}(X,X)^{-1}W(X,X) = \left(W(X,X)^{-1} + \frac{W(X,X)^{-1}X^T C_i C_i^T X W(X,X)^{-1}}{(m_i - p_{ii})}\right)W(X,X)$$
$$= I + \frac{W(X,X)^{-1}X^T C_i C_i^T X}{c_{ii} - C_i^T X W(X,X)^{-1}X^T C_i}$$

olarak elde edilir.

EK 4. Frobenius (Euclidean) Normu (Sarıpınar, 2004)

Tanım: $M_n(G)$ $n \times n$ matrislerin bir kümesini göstermek üzere,

$$\|.\|: M_n(G) \to R^+$$

şeklinde tanımlanan dönüşüm,

i) $A \epsilon M_n(G)$ için $A \neq 0$ ise $\|A\| > 0$ ve $A = 0 \leftrightarrow \|A\| = 0$ dır.
ii) $A \epsilon M_n(G)$ ve $\alpha \epsilon G$ için $\|\alpha A\| = |\alpha| \|A\|$ dır.
iii) $A, B \epsilon M_n(G)$ için $\|A + B\| \leq \|A\| + \|B\|$ dır.
iv) $A, B \epsilon M_n(G)$ $\|AB\| \leq \|A\| \|B\|$ dır.

şartlarını sağlıyorsa, ***matris normu*** adını alır ve matris normu $A \epsilon M_n(G)$ için $M_n(A) = \|A\|$ şeklinde gösterilir.

$A = \left(a_{ij}\right)_{n \times n}$ bir matris olmak üzere,

$$\|A\|_F = \left(\sum_{i=1}^{n} \left|a_{ij}\right|^2 \right)^{\frac{1}{2}}$$

ifadesi ***Frobenius*** veya ***Euclidean normu*** adını alır.

EK 5. $\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)}$ ile $\widehat{\boldsymbol{u}} - \widehat{\boldsymbol{u}}_{(i)}$ nın Elde Edilmesi

(3.12) ile verilen

$$\hat{\beta}_{(i)} = \hat{\beta} - \frac{(X^T H^{-1} X)^{-1} \check{x}_i \left(\check{y}_i - \check{x}_i^T \hat{\beta}\right)}{m_i - p_{ii}}$$

eşitliğinde, Zewotir ve Galphin (2005) tarafından geliştirilen

$$\check{x}_i = \frac{1}{c_{ii}} X^T C_i \, , \check{y}_i = \frac{1}{c_{ii}} y^T C_i \, , p_{ii} = \check{x}_i^T W (X, X)^{-1} \check{x}_i \, , m_i = \frac{1}{c_{ii}}$$

ölçütleri ile

$$r_{ii} = c_{ii} - C_i^T X (X^T H^{-1} X)^{-1} X^T C_i$$

ifadesi yerine konulup, gerekli işlemler yapıldığında,

$$\hat{\beta} - \hat{\beta}_{(i)} = W (X, X)^{-1} X^T C_i \frac{e_i}{r_{ii}}$$

olarak elde edilir. Burada,

$$\begin{aligned} e = y - X\hat{\beta} - Z\hat{u} &= y - X\hat{\beta} - Z(DZ^T H^{-1}\left(y - X\hat{\beta}\right)) \\ &= y - X\hat{\beta} - \underbrace{ZDZ^T}_{H-I} H^{-1} y + ZDZ^T H^{-1} X\hat{\beta} \\ &= H^{-1}\left(y - X\hat{\beta}\right) = C\left(y - X\hat{\beta}\right) \end{aligned}$$

olup, e nin i-inci elemanı ise,

$$e_i = y^T C_i - C_i^T X \hat{\beta}$$

dir.

$\hat{u} - \hat{u}_{(i)}$ incelensin. (3.14) ile verilen

$$\hat{u}_{(i)} = \hat{u} - D[\check{z}_i - (Z^T H^{-1} X)(X^T H^{-1} X)^{-1} \check{x}_i] \frac{(\check{y}_i - \check{x}_i^T \hat{\beta})}{(m_i - p_{ii})}$$

eşitliğinde, Zewotir ve Galphin (2005) tarafından geliştirilen

$$\check{x}_i = \frac{1}{c_{ii}} X^T C_i \, , \check{y}_i = \frac{1}{c_{ii}} y^T C_i \, , p_{ii} = \check{x}_i^T W (X, X)^{-1} \check{x}_i \, , m_i = \frac{1}{c_{ii}} , \check{z}_i = \frac{1}{c_{ii}} Z^T C_i$$

ölçütleri ile

$$r_{ii} = c_{ii} - C_i^T X (X^T H^{-1} X)^{-1} X^T C_i \, , R_i = C_i - H^{-1} X (X^T H^{-1} X)^{-1} X^T C_i$$

ifadeleri yerine konulup, gerekli işlemler yapıldığında,

$$\hat{u} - \hat{u}_{(i)} = D Z^T R_i \frac{e_i}{r_{ii}}$$

olarak bulunur.

EK 6. $\hat{\sigma}^2_{e(i)}$ ile $\hat{\sigma}^2_e$ Arasındaki İlişki

(3.16) ile verilen

$$\hat{\sigma}^2_{e(i)} = \frac{1}{(n-1)}\left(n\hat{\sigma}^2_e - \frac{1}{m_i - p_{ii}}\left(\check{y}_i - \check{x}_i^T\hat{\beta}\right)^2\right)$$

eşitliğinde, Zewotir ve Galphin (2005) tarafından geliştirilen

$$\check{x}_i = \frac{1}{c_{ii}}X^T C_i \,, \check{y}_i = \frac{1}{c_{ii}}y^T C_i \,, p_{ii} = \check{x}_i^T W(X,X)^{-1}\check{x}_i \,, m_i = \frac{1}{c_{ii}}$$

ölçütleri ile

$$r_{ii} = c_{ii} - C_i^T X(X^T H^{-1} X)^{-1} X^T C_i$$

ifadesi yerine konulup, gerekli işlemler yapıldığında,

$$\hat{\sigma}^2_{e(i)} = \frac{n\hat{\sigma}^2_e}{n-1} - \frac{\left(y^T C_i - C_i^T X\hat{\beta}\right)^2 / {c_{ii}}^2}{(n-1)r_{ii}/{c_{ii}}^2} = \frac{n\hat{\sigma}^2_e}{n-1} - \frac{e_i^2}{(n-1)r_{ii}}$$
$$= \frac{n\hat{\sigma}^2_e}{n-1} - \frac{t_i^2\hat{\sigma}^2_e r_{ii}}{(n-1)r_{ii}} = \left(\frac{n-t_i^2}{n-1}\right)\hat{\sigma}^2_e$$

olarak elde edilir.

EK 7. En Küçük Bozulmuş Rezidülerin Elde Edilmesi (Nobre ve Singer, 2007)

(3.20) eşitliğinin tekil değer ayrışımı yardımıyla maksimum yapılmasını sağlayan $l_i^T e$ en küçük bozulmuş rezidüleri elde edilmek istensin. Özel olarak $\Sigma = I_n$ olarak alınması sebebiyle, Q un sıfırdan farklı uzayı ile ilgilenilmesi yeterlidir. $n \times n$ boyutlu ve $n - p$ ranklı Q matrisi verildiğinde $K^T K = I_{(n-p)}$ olmak üzere,

$$\Sigma^{\frac{1}{2}} Q \Sigma^{\frac{1}{2}} = K \Pi K^T$$

tekil değer ayrışımı yapılabilir. Burada K, $n \times (n-p)$ boyutlu bir matris, Π ise $(n-p) \times (n-p)$ boyutlu köşegen bir matris olup, $\Pi = diag(\pi_1, \cdots, \pi_{n-p})$ şeklindedir ve $\pi_{n-p} \leq \cdots \leq \pi_1 \leq 1$ dir. v_i nin herhangi bir $(n-p) \times 1$ vektörü için $l_i = \Sigma^{-\frac{1}{2}} K \Pi^{-\frac{1}{2}} v_i$ olarak alınıp, (3.20) eşitliğinde yerine yazılırsa,

$$\lambda_i = \frac{v_i^T \Pi v_i}{v_i^T v_i}$$

olarak bulunur. v_i, $I_{(n-p)}$ nin i-inci sütunu olarak alındığında,

$$l_i = \Sigma^{-\frac{1}{2}} K \Pi^{-\frac{1}{2}} v_i = \pi_i{}^{-\frac{1}{2}} \Sigma^{-\frac{1}{2}} K_i \quad , \quad (i = 1, \cdots, n-p)$$

şeklinde olup K_i, K matrisinin i-inci sütununu göstermektedir. Böylece, (3.20) eşitliğinin maksimum olmasını sağlayan $l_i^T e$ en küçük bozulmuş rezidüleri,

$$l_i^T e = \pi_i{}^{\frac{1}{2}} K_i^T \Sigma^{-\frac{1}{2}} y$$

olarak elde edilir.

EK 8. Karesel Formların Dağılımıyla İlgili Teoremler (McCulloch ve Searle, 2001)

Teorem: $y \sim N_p(\mu, \Sigma)$ olsun. J , $p \times p$ tipinde rankı r olan simetrik bir matris ve $\lambda = \frac{1}{2}\mu^T J \mu$ iken $y^T J y \sim \chi^2(r, \lambda)$ olması için gerek ve yeter koşul $J\,\Sigma$ nın idempotent olmasıdır.

Teorem: $y \sim N_p(\mu, \Sigma)$ olsun. $B\,\Sigma J = 0$ ise $y^T J y$ ile $y^T B y$ karesel formları bağımsızdır.

EK 9. Microsoft Excel'e Bakteri Plak İndeks Verisinin Girişi

Observations	Tootbrush	Session	Before (x)	After (y)	Subject
1	Conventional	1	1,05	1	1
2	Conventional	2	1,13	0,84	1
3	Conventional	3	1,15	0,86	1
4	Conventional	4	1,13	0,94	1
5	Conventional	1	1,07	0,62	2
6	Conventional	2	0,92	0,62	2
7	Conventional	3	1,02	0,57	2
8	Conventional	4	1,15	0,85	2
9	Conventional	1	0,82	0,62	3
10	Conventional	2	1,52	1,07	3
11	Conventional	3	1,39	0,97	3
12	Conventional	4	1,78	1,39	3
13	Conventional	1	1,37	0,9	4
14	Conventional	2	1,65	1,2	4
15	Conventional	3	1,75	1,4	4
16	Conventional	4	1,92	1,67	4
17	Conventional	1	1,97	1,52	5
18	Conventional	2	1,3	1,07	5
19	Conventional	3	1,5	1,15	5
20	Conventional	4	1,65	1,37	5
21	Conventional	1	1,3	0,82	6
22	Conventional	2	1,17	0,7	6
23	Conventional	3	0,75	0,5	6
24	Conventional	4	1,47	1,12	6
25	Conventional	1	1,61	1,19	7
26	Conventional	2	1,52	1,13	7
27	Conventional	3	1,22	1	7
28	Conventional	4	1,63	1,22	7
29	Conventional	1	1,02	0,73	8
30	Conventional	2	1,08	0,64	8
31	Conventional	3	0,94	0,73	8
32	Conventional	4	1,14	0,97	8
33	Conventional	1	1,62	1,25	9
34	Conventional	2	1,45	1,1	9
35	Conventional	3	1,1	0,75	9
36	Conventional	4	1,7	1,32	9
37	Conventional	1	1,65	1,22	10
38	Conventional	2	1,57	1,22	10
39	Conventional	3	1,47	1,1	10
40	Conventional	4	1,62	1,17	10
41	Conventional	1	1,02	0,78	11
42	Conventional	2	0,6	0,47	11
43	Conventional	3	0,88	0,75	11
44	Conventional	4	1,36	1,08	11
45	Conventional	1	0,71	0,6	12
46	Conventional	2	1,13	0,39	12
47	Conventional	3	0,84	0,65	12
48	Conventional	4	1,65	1,31	12
49	Conventional	1	1,7	1,55	13

50	Conventional	2	1,85	1,37	13
51	Conventional	3	1,87	1,55	13
52	Conventional	4	1,6	1,3	13
53	Conventional	1	1,3	1,02	14
54	Conventional	2	1,65	0,97	14
55	Conventional	3	1,72	1,2	14
56	Conventional	4	1,37	1,22	14
57	Conventional	1	1,4	0,8	15
58	Conventional	2	1,83	1,03	15
59	Conventional	3	1,76	1,38	15
60	Conventional	4	1,96	1,15	15
61	Conventional	1	1,4	1,12	16
62	Conventional	2	1,25	0,67	16
63	Conventional	3	1,5	1,1	16
64	Conventional	4	1,5	1,22	16
65	Monoblock	1	1,66	1,63	17
66	Monoblock	2	1,36	1,16	17
67	Monoblock	3	1,52	0,88	17
68	Monoblock	4	1,41	1,2	17
69	Monoblock	1	1,02	0,8	18
70	Monoblock	2	0,92	0,82	18
71	Monoblock	3	1,1	0,76	18
72	Monoblock	4	1,28	1,15	18
73	Monoblock	1	0,75	0,67	19
74	Monoblock	2	1	0,92	19
75	Monoblock	3	1	0,87	19
76	Monoblock	4	1,15	1,1	19
77	Monoblock	1	1,29	1,23	20
78	Monoblock	2	0,91	0,76	20
79	Monoblock	3	1,14	0,94	20
80	Monoblock	4	1,35	0,97	20
81	Monoblock	1	1,27	1,2	21
82	Monoblock	2	1,2	0,95	21
83	Monoblock	3	1,1	1	21
84	Monoblock	4	1,37	1,17	21
85	Monoblock	1	1,07	0,85	22
86	Monoblock	2	1,39	1,25	22
87	Monoblock	3	1,39	1,25	22
88	Monoblock	4	1,28	1,21	22
89	Monoblock	1	1,35	1,21	23
90	Monoblock	2	1,42	1,17	23
91	Monoblock	3	1,42	1,19	23
92	Monoblock	4	1,42	1,23	23
93	Monoblock	1	1,32	1,02	24
94	Monoblock	2	1,6	1,4	24
95	Monoblock	3	1,35	1,02	24
96	Monoblock	4	1,5	1,25	24
97	Monoblock	1	1,66	1,61	25
98	Monoblock	2	1,5	1,36	25
99	Monoblock	3	1,72	1,41	25
100	Monoblock	4	1,69	1,44	25
101	Monoblock	1	1,3	1,07	26
102	Monoblock	2	0,84	0,61	26

103	Monoblock	3	0,88	0,61	26
104	Monoblock	4	0,96	0,57	26
105	Monoblock	1	1,57	1,2	27
106	Monoblock	2	1,5	1,07	27
107	Monoblock	3	1,15	1	27
108	Monoblock	4	1,25	1,05	27
109	Monoblock	1	1,67	1,5	28
110	Monoblock	2	1,47	1,32	28
111	Monoblock	3	1,07	0,97	28
112	Monoblock	4	1,5	1,37	28
113	Monoblock	1	0,91	0,67	29
114	Monoblock	2	0,96	0,62	29
115	Monoblock	3	1,09	0,53	29
116	Monoblock	4	1,12	0,37	29
117	Monoblock	1	1,06	0,7	30
118	Monoblock	2	1	0,85	30
119	Monoblock	3	1,15	0,93	30
120	Monoblock	4	1,12	1	30
121	Monoblock	1	2,3	2	31
122	Monoblock	2	1,37	1,25	31
123	Monoblock	3	1,4	1,32	31
124	Monoblock	4	2,15	1,9	31
125	Monoblock	1	1,15	1	32
126	Monoblock	2	1,23	1,11	32
127	Monoblock	3	1,15	1,07	32
128	Monoblock	4	1,26	1	32

EK 10. X ve Z Matrisi ile y Vektörünün Elde Edilmesi

Çizelge 10.1. X matrisinin elde edilmesi

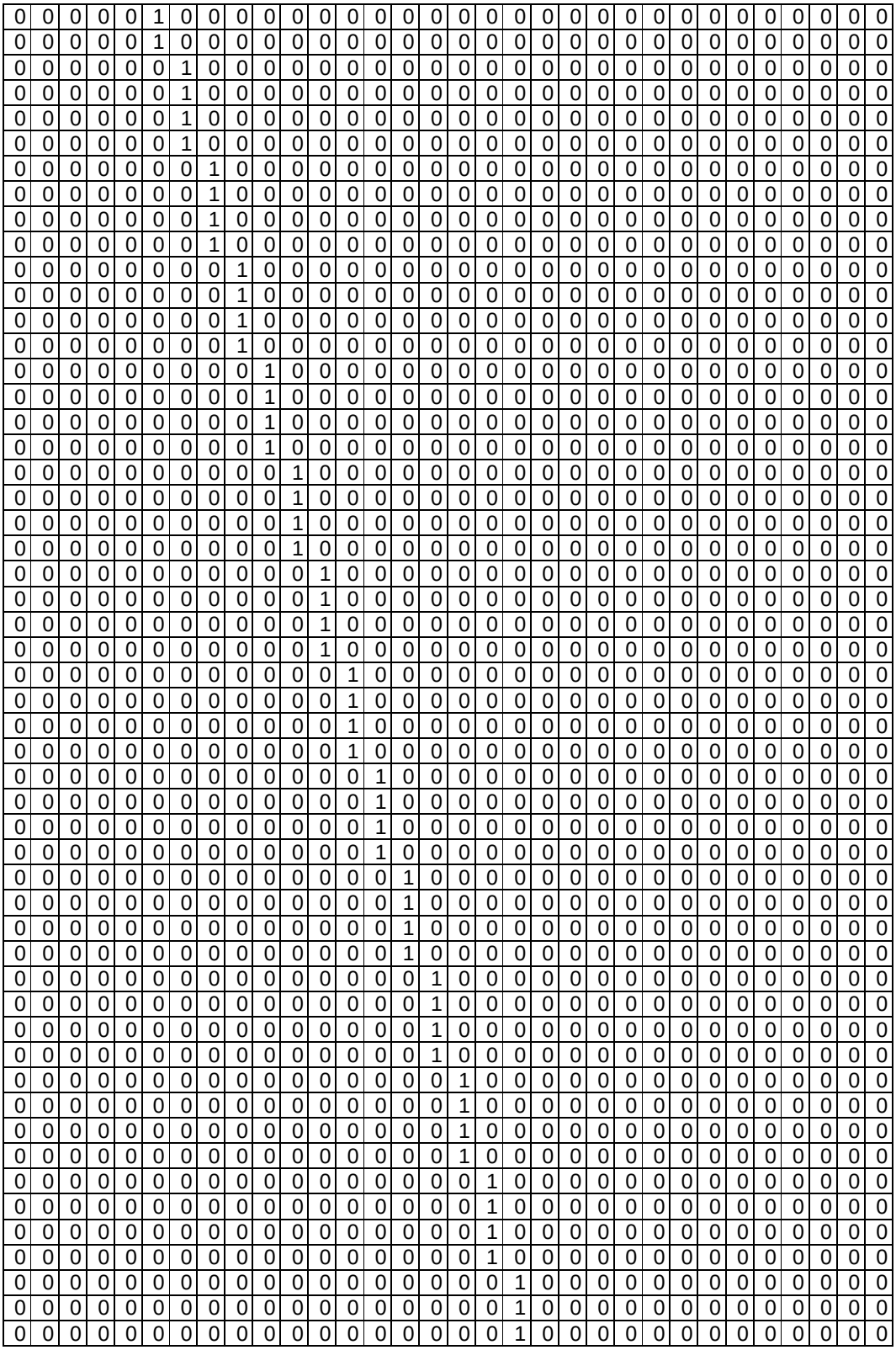
Observations	log(x)	α_0	α_1
1	0.04879016	1	0
2	0.12221763	1	0
3	0.13976194	1	0
4	0.12221763	1	0
5	0.06765865	1	0
6	-0.08338161	1	0
7	0.01980263	1	0
8	0.13976194	1	0
9	-0.19845094	1	0
10	0.41871033	1	0
11	0.32930375	1	0
12	0.57661336	1	0
13	0.31481074	1	0
14	0.50077529	1	0
15	0.55961579	1	0
16	0.65232519	1	0
17	0.67803354	1	0
18	0.26236426	1	0
19	0.40546511	1	0
20	0.50077529	1	0
21	0.26236426	1	0
22	0.15700375	1	0
23	-0.28768207	1	0
24	0.38526240	1	0
25	0.47623418	1	0
26	0.41871033	1	0
27	0.19885086	1	0
28	0.48858001	1	0
29	0.01980263	1	0
30	0.07696104	1	0
31	-0.06187540	1	0
32	0.13102826	1	0
33	0.48242615	1	0
34	0.37156356	1	0
35	0.09531018	1	0
36	0.53062825	1	0
37	0.50077529	1	0
38	0.45107562	1	0
39	0.38526240	1	0
40	0.48242615	1	0
41	0.01980263	1	0
42	-0.51082562	1	0
43	-0.12783337	1	0
44	0.30748470	1	0
45	-0.34249031	1	0
46	0.12221763	1	0
47	-0.17435339	1	0

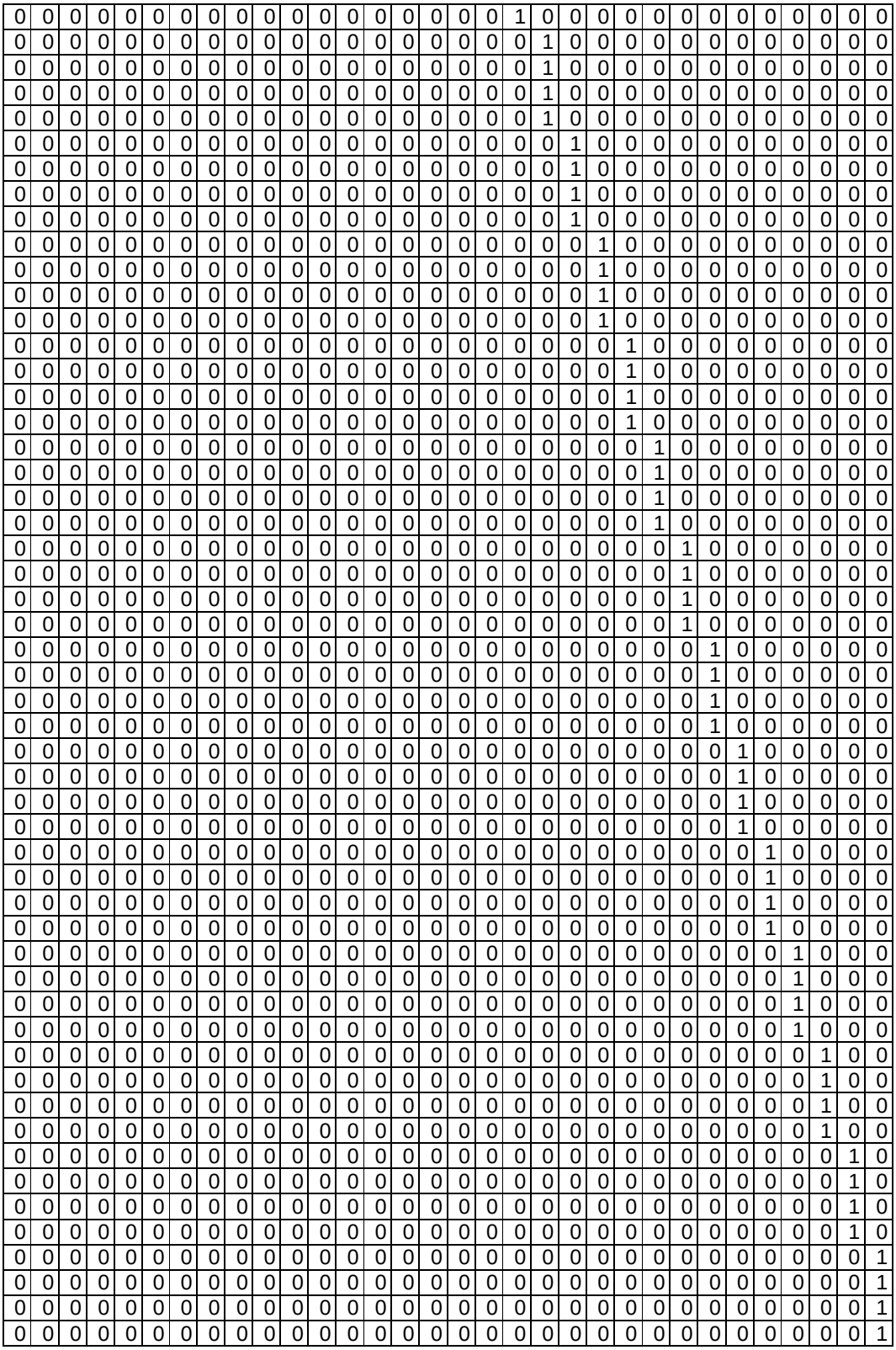
48	0.50077529	1	0
49	0.53062825	1	0
50	0.61518564	1	0
51	0.62593843	1	0
52	0.47000363	1	0
53	0.26236426	1	0
54	0.50077529	1	0
55	0.54232429	1	0
56	0.31481074	1	0
57	0.33647224	1	0
58	0.60431597	1	0
59	0.56531381	1	0
60	0.67294447	1	0
61	0.33647224	1	0
62	0.22314355	1	0
63	0.40546511	1	0
64	0.40546511	1	0
65	0.50681760	0	1
66	0.30748470	0	1
67	0.41871033	0	1
68	0.34358970	0	1
69	0.01980263	0	1
70	-0.08338161	0	1
71	0.09531018	0	1
72	0.24686008	0	1
73	-0.28768207	0	1
74	0.00000000	0	1
75	0.00000000	0	1
76	0.13976194	0	1
77	0.25464222	0	1
78	-0.09431068	0	1
79	0.13102826	0	1
80	0.30010459	0	1
81	0.23901690	0	1
82	0.18232156	0	1
83	0.09531018	0	1
84	0.31481074	0	1
85	0.06765865	0	1
86	0.32930375	0	1
87	0.32930375	0	1
88	0.24686008	0	1
89	0.30010459	0	1
90	0.35065687	0	1
91	0.35065687	0	1
92	0.35065687	0	1
93	0.27763174	0	1
94	0.47000363	0	1
95	0.30010459	0	1
96	0.40546511	0	1
97	0.50681760	0	1
98	0.40546511	0	1
99	0.54232429	0	1
100	0.52472853	0	1

101	0.26236426	0	1
102	-0.17435339	0	1
103	-0.12783337	0	1
104	-0.04082199	0	1
105	0.45107562	0	1
106	0.40546511	0	1
107	0.13976194	0	1
108	0.22314355	0	1
109	0.51282363	0	1
110	0.38526240	0	1
111	0.06765865	0	1
112	0.40546511	0	1
113	-0.09431068	0	1
114	-0.04082199	0	1
115	0.08617770	0	1
116	0.11332869	0	1
117	0.05826891	0	1
118	0.00000000	0	1
119	0.13976194	0	1
120	0.11332869	0	1
121	0.83290912	0	1
122	0.31481074	0	1
123	0.33647224	0	1
124	0.76546784	0	1
125	0.13976194	0	1
126	0.20701417	0	1
127	0.13976194	0	1
128	0.23111172	0	1

Çizelge 10.2. Z matrisinin elde edilmesi

1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0





Çizelge 10.3. y vektörünün elde edilmesi

Observations	log(y)
1	0.00000000
2	-0.17435339
3	-0.15082289
4	-0.06187540
5	-0.47803580
6	-0.47803580
7	-0.56211892
8	-0.16251893
9	-0.47803580
10	0.06765865
11	-0.03045921
12	0.32930375
13	-0.10536052
14	0.18232156
15	0.33647224
16	0.51282363
17	0.41871033
18	0.06765865
19	0.13976194
20	0.31481074
21	-0.19845094
22	-0.35667494
23	-0.69314718
24	0.11332869
25	0.17395331
26	0.12221763
27	0.00000000
28	0.19885086
29	-0.31471074
30	-0.44628710
31	-0.31471074
32	-0.03045921
33	0.22314355
34	0.09531018
35	-0.28768207
36	0.27763174
37	0.19885086
38	0.19885086
39	0.09531018
40	0.15700375
41	-0.24846136
42	-0.75502258
43	-0.28768207
44	0.07696104
45	-0.51082562
46	-0.94160854
47	-0.43078292
48	0.27002714
49	0.43825493
50	0.31481074
51	0.43825493

52	0.26236426
53	0.01980263
54	-0.03045921
55	0.18232156
56	0.19885086
57	-0.22314355
58	0.02955880
59	0.32208350
60	0.13976194
61	0.11332869
62	-0.40047757
63	0.09531018
64	0.19885086
65	0.48858001
66	0.14842001
67	-0.12783337
68	0.18232156
69	-0.22314355
70	-0.19845094
71	-0.27443685
72	0.13976194
73	-0.40047757
74	-0.08338161
75	-0.13926207
76	0.09531018
77	0.20701417
78	-0.27443685
79	-0.06187540
80	-0.03045921
81	0.18232156
82	-0.05129329
83	0.00000000
84	0.15700375
85	-0.16251893
86	0.22314355
87	0.22314355
88	0.19062036
89	0.19062036
90	0.15700375
91	0.17395331
92	0.20701417
93	0.01980263
94	0.33647224
95	0.01980263
96	0.22314355
97	0.47623418
98	0.30748470
99	0.34358970
100	0.36464311
101	0.06765865
102	-0.49429632
103	-0.49429632
104	-0.56211892

105	0.18232156
106	0.06765865
107	0.00000000
108	0.04879016
109	0.40546511
110	0.27763174
111	-0.03045921
112	0.31481074
113	-0.40047757
114	-0.47803580
115	-0.63487827
116	-0.99425227
117	-0.35667494
118	-0.16251893
119	-0.07257069
120	0.00000000
121	0.69314718
122	0.22314355
123	0.27763174
124	0.64185389
125	0.00000000
126	0.10436002
127	0.06765865
128	0.00000000

EK 11. R Analiz Sonuçları

Çizelge 11.1. Gözlemler için R analiz sonuçları

Observations	Tootbrush	resm	resc	rescp	Subject
1	Conventional	0.278690928	0.208987803	1.576993202	1
2	Conventional	0.027134832	-0.042568293	-0.320647431	1
3	Conventional	0.032218987	-0.037484138	-0.282287635	1
4	Conventional	0.139612816	0.069909691	0.526597643	1
5	Conventional	-0.219183472	-0.144470589	-1.088452279	2
6	Conventional	-0.060377548	0.014335335	0.108607257	2
7	Conventional	-0.252950072	-0.178237189	-1.344392610	2
8	Conventional	0.020522947	0.095235831	0.717045063	2
9	Conventional	0.060608019	0.057055063	0.442485089	3
10	Conventional	-0.042589872	-0.046142828	-0.348202323	3
11	Conventional	-0.046704341	-0.050257297	-0.378492235	3
12	Conventional	0.053033685	0.049480729	0.376565320	3
13	Conventional	-0.106367490	-0.123067956	-0.927754783	4
14	Conventional	-0.014211249	-0.030911714	-0.232949967	4
15	Conventional	0.078073673	0.061373208	0.463284372	4
16	Conventional	0.156949054	0.140248588	1.063372094	4
17	Conventional	0.035805622	-0.010791469	-0.082104351	5
18	Conventional	0.121794661	0.075197570	0.567296963	5
19	Conventional	0.043439646	-0.003157445	-0.023773696	5
20	Conventional	0.118277934	0.071680843	0.540534449	5
21	Conventional	-0.144314926	-0.091998219	-0.693060123	6
22	Conventional	-0.191761353	-0.139444646	-1.050085298	6
23	Conventional	-0.060684449	-0.008367742	-0.064897920	6
24	Conventional	0.038247809	0.090564517	0.684954308	6
25	Conventional	0.003223380	-0.018847749	-0.142186897	7
26	Conventional	0.011969112	-0.010102017	-0.076103732	7
27	Conventional	0.120914931	0.098843802	0.746328888	7
28	Conventional	0.015140341	-0.006930789	-0.052306873	7
29	Conventional	-0.005541899	-0.011283194	-0.085114248	8
30	Conventional	-0.197215445	-0.202956739	-1.528930990	8
31	Conventional	0.080335571	0.074594276	0.564586106	8
32	Conventional	0.161765388	0.156024093	1.174751905	8
33	Conventional	0.045903297	0.035346372	0.266843405	9
34	Conventional	0.034632468	0.024075543	0.181297890	9
35	Conventional	-0.057902966	-0.068459891	-0.519156304	9
36	Conventional	0.049711091	0.039154166	0.296198428	9
37	Conventional	0.002318053	-0.005982837	-0.045120714	10
38	Conventional	0.054573008	0.046272118	0.348578823	10
39	Conventional	0.020229304	0.011928414	0.089810844	10
40	Conventional	-0.020236505	-0.028537395	-0.215115735	10
41	Conventional	0.060707486	-0.002498477	-0.018818230	11
42	Conventional	0.112056523	0.048850559	0.384417288	11
43	Conventional	0.176713410	0.113507446	0.858949190	11
44	Conventional	0.083656771	0.020450808	0.155026768	11
45	Conventional	0.179263285	0.238340117	1.847305665	12
46	Conventional	**-0.740120321**	-0.681043488	**-5.127780855**	12
47	Conventional	0.082524387	0.141601220	1.078969787	12

48	Conventional	0.073494332	0.132571164	1.015100775	12
49	Conventional	0.210334286	0.149536945	1.126873599	13
50	Conventional	-0.002014773	-0.062812114	-0.474595395	13
51	Conventional	0.110123776	0.049326436	0.372874287	13
52	Conventional	0.098185228	0.037387888	0.281518350	13
53	Conventional	0.073938640	0.075851153	0.571712306	14
54	Conventional	-0.226992013	-0.225079499	-1.699080523	14
55	Conventional	-0.057896475	-0.055983961	-0.423337360	14
56	Conventional	0.197843884	0.199756398	1.504308693	14
57	Conventional	-0.246925739	-0.144359209	-1.088272305	15
58	Conventional	-0.275838179	-0.173271649	-1.309110826	15
59	Conventional	0.057693954	0.160260484	1.209077264	15
60	Conventional	-0.237792049	-0.135225519	-1.025196272	15
61	Conventional	0.089546498	0.106188212	0.799536175	16
62	Conventional	-0.305104326	-0.288462612	-2.174368989	16
63	Conventional	-0.001012116	0.015629598	0.117785696	16
64	Conventional	0.102528563	0.119170277	0.898075183	16
65	Monoblock	0.170978171	0.186312208	1.408482526	17
66	Monoblock	0.040399675	0.055733712	0.419645282	17
67	Monoblock	-0.352797952	-0.337463915	-2.543524972	17
68	Monoblock	0.036339899	0.051673937	0.389075884	17
69	Monoblock	-0.028690766	-0.030428743	-0.229489706	18
70	Monoblock	0.104491253	0.102753276	0.778314316	18
71	Monoblock	-0.159373800	-0.161111777	-1.213290100	18
72	Monoblock	0.095483221	0.093745244	0.706886984	18
73	Monoblock	0.117269104	0.045938752	0.351624654	19
74	Monoblock	0.131891947	0.060561594	0.456289047	19
75	Monoblock	0.076011488	0.004681135	0.035269066	19
76	Monoblock	0.163635996	0.092305643	0.695351835	19
77	Monoblock	0.154553193	0.142957446	1.077047665	20
78	Monoblock	0.039996330	0.028400583	0.215883264	20
79	Monoblock	0.015633131	0.004037384	0.030407922	20
80	Monoblock	-0.130719986	-0.142315733	-1.073363858	20
81	Monoblock	0.146289267	0.106171350	0.799427207	21
82	Monoblock	-0.027715276	-0.067833192	-0.510803332	21
83	Monoblock	0.115063046	0.074945129	0.565313189	21
84	Monoblock	0.041280713	0.001162797	0.008765204	21
85	Monoblock	-0.018382661	-0.063959114	-0.483293004	22
86	Monoblock	0.092182357	0.046605904	0.351266143	22
87	Monoblock	0.092182357	0.046605904	0.351266143	22
88	Monoblock	0.146341638	0.100765186	0.758683939	22
89	Monoblock	0.090359581	0.065830093	0.495647700	23
90	Monoblock	0.003591569	-0.020937920	-0.157709205	23
91	Monoblock	0.020541127	-0.003988361	-0.030041250	23
92	Monoblock	0.053601990	0.029072501	0.218980735	23
93	Monoblock	-0.056829863	-0.046964371	-0.353642072	24
94	Monoblock	0.057577140	0.067442633	0.509443152	24
95	Monoblock	-0.080458151	-0.070592658	-0.531506887	24
96	Monoblock	0.012105194	0.021970687	0.165615537	24
97	Monoblock	0.158632335	0.118948405	0.897236270	25
98	Monoblock	0.096446343	0.056762413	0.427395570	25
99	Monoblock	-0.011344388	-0.051028318	-0.385385595	25
100	Monoblock	0.028209461	-0.011474469	-0.086603212	25

101	Monoblock	0.007078600	0.085465889	0.646029124	26
102	Monoblock	-0.095705079	-0.017317790	-0.131479066	26
103	Monoblock	-0.144616900	-0.066229611	-0.501446392	26
104	Monoblock	-0.303924524	-0.225537235	-1.701391661	26
105	Monoblock	-0.076672355	-0.058828564	-0.444660091	27
106	Monoblock	-0.143379708	-0.125535917	-0.947210857	27
107	Monoblock	0.068325816	0.086169607	0.650231137	27
108	Monoblock	0.029447344	0.047291135	0.356175338	27
109	Monoblock	0.081548444	0.025099118	0.190083661	28
110	Monoblock	0.087834800	0.031385475	0.236539304	28
111	Monoblock	0.113677061	0.057227735	0.433681333	28
112	Monoblock	0.103772383	0.047323058	0.356831858	28
113	Monoblock	-0.086044391	0.164108473	1.241730088	29
114	Monoblock	-0.219841407	0.030311457	0.228824401	29
115	Monoblock	**-0.510213200**	-0.260060337	-1.958122117	29
116	Monoblock	**-0.898134146**	-0.647981283	**-4.878788600**	29
117	Monoblock	-0.202666166	-0.194196814	-1.463386110	30
118	Monoblock	0.052754626	0.061223978	0.462107374	30
119	Monoblock	-0.004244876	0.004224476	0.031806729	30
120	Monoblock	0.096118127	0.104587480	0.787531328	30
121	Monoblock	0.032687969	-0.015693448	-0.120678023	31
122	Monoblock	0.107420516	0.059039099	0.445173248	31
123	Monoblock	0.139133488	0.090752071	0.683936949	31
124	Monoblock	0.052303418	0.003922001	0.029964149	31
125	Monoblock	0.068325816	0.027675598	0.208469255	32
126	Monoblock	0.101975862	0.061325644	0.461732431	32
127	Monoblock	0.135984465	0.095334247	0.718114895	32
128	Monoblock	-0.027720670	-0.068370888	-0.514834042	32

Çizelge 11.2. Denekler için R analiz sonuçları

Subject	Tootbrush	eblup	resmcov
1	Conventional	0.069703125	2.763874
2	Conventional	-0.074712883	3.241516
3	Conventional	0.003552956	1.802934
4	Conventional	0.016700465	1.969681
5	Conventional	0.046597091	1.741148
6	Conventional	-0.052316707	2.059107
7	Conventional	0.022071129	1.786948
8	Conventional	0.005741295	2.981459
9	Conventional	0.010556925	1.832378
10	Conventional	0.008300890	1.924992
11	Conventional	0.063205964	1.758125
12	Conventional	-0.059076832	**26.357726**
13	Conventional	0.060797341	1.977721
14	Conventional	-0.001912514	4.156869
15	Conventional	-0.102566530	5.289914
16	Conventional	-0.016641714	4.609627
17	Monoblock	-0.015334037	6.677817
18	Monoblock	0.001737977	2.119505
19	Monoblock	0.071330353	1.774420

20	Monoblock	0.011595747	2.006548
21	Monoblock	0.040117917	1.750463
22	Monoblock	0.045576453	1.741157
23	Monoblock	0.024529489	1.848904
24	Monoblock	-0.009865493	1.777530
25	Monoblock	0.039683930	1.741576
26	Monoblock	-0.078387289	3.345593
27	Monoblock	-0.017843791	1.784948
28	Monoblock	0.056449325	1.746053
29	Monoblock	**-0.250152864**	**32.385059**
30	Monoblock	-0.008469352	2.309929
31	Monoblock	0.048381417	1.735135
32	Monoblock	0.040650218	1.735247

EK 12. Microsoft Excel'e Bakteri Plak İndeks Verisinin Girişi

Observations	Toothbrush	Session	Before (x)	After (y)	Subject
1	Experimental	1	1,05	1	1
2	Experimental	2	1,13	0,84	1
3	Experimental	1	1,07	0,62	2
4	Experimental	2	0,92	0,62	2
5	Experimental	1	0,82	0,62	3
6	Experimental	2	1,52	1,07	3
7	Experimental	1	1,37	0,9	4
8	Experimental	2	1,65	1,2	4
9	Experimental	1	1,97	1,52	5
10	Experimental	2	1,3	1,07	5
11	Experimental	1	1,3	0,82	6
12	Experimental	2	1,17	0,7	6
13	Experimental	1	1,61	1,19	7
14	Experimental	2	1,52	1,13	7
15	Experimental	1	1,02	0,73	8
16	Experimental	2	1,08	0,64	8
17	Experimental	1	1,62	1,25	9
18	Experimental	2	1,45	1,1	9
19	Experimental	1	1,65	1,22	10
20	Experimental	2	1,57	1,22	10
21	Experimental	1	1,02	0,78	11
22	Experimental	2	0,6	0,47	11
23	Experimental	1	0,68	0,6	12
24	Experimental	2	1,13	0,39	12
25	Experimental	1	1,7	1,55	13
26	Experimental	2	1,85	1,37	13
27	Experimental	1	1,3	1,02	14
28	Experimental	2	1,65	0,97	14
29	Experimental	1	0,90	0,8	15
30	Experimental	2	1,10	1,03	15
31	Experimental	1	1,4	1,12	16
32	Experimental	2	1,25	0,67	16
33	Experimental	1	1,66	1,63	17
34	Experimental	2	1,36	1,16	17
35	Experimental	1	1,02	0,8	18
36	Experimental	2	0,92	0,82	18
37	Experimental	1	0,68	0,67	19
38	Experimental	2	1	0,92	19
39	Experimental	1	1,29	1,23	20
40	Experimental	2	0,91	0,76	20
41	Experimental	1	1,27	1,2	21
42	Experimental	2	1,2	0,95	21
43	Experimental	1	1,07	0,85	22
44	Experimental	2	1,39	1,25	22
45	Experimental	1	1,35	1,21	23
46	Experimental	2	1,42	1,17	23
47	Experimental	1	1,32	1,02	24
48	Experimental	2	1,6	1,4	24
49	Experimental	1	1,66	1,61	25

50	Experimental	2	1,5	1,36	25
51	Experimental	1	1,3	1,07	26
52	Experimental	2	0,84	0,61	26
53	Experimental	1	1,57	1,2	27
54	Experimental	2	1,5	1,07	27
55	Experimental	1	1,67	1,5	28
56	Experimental	2	1,47	1,32	28
57	Experimental	1	0,91	0,67	29
58	Experimental	2	0,96	0,62	29
59	Experimental	1	1,06	0,7	30
60	Experimental	2	1	0,85	30
61	Experimental	1	2,35	2,14	31
62	Experimental	2	1,57	1,55	31
63	Experimental	1	1,15	1	32
64	Experimental	2	1,23	1,11	32

EK 13. R Analiz Sonuçları

Çizelge 13.1. Gözlemler için R analiz sonuçları

Observations	L_1	L_2	H_2
1	0.02390558	0.007965418	0.003401004
2	0.01930765	0.009225435	0.003938995
3	0.02258381	0.008271753	0.003896164
4	0.03529675	0.006115129	0.002880351
5	0.04736136	0.004858002	0.001810896
6	0.02320970	0.016692337	0.006222329
7	0.01654028	0.013560357	0.003771758
8	0.03421591	0.019669706	0.005471049
9	**0.08198264**	**0.028038994**	0.006766640
10	0.01563891	0.012210029	0.002946641
11	0.01563891	0.012210029	0.004479748
12	0.01769784	0.009890123	0.003628596
13	0.03031252	0.018727583	0.004972341
14	0.02320970	0.016692337	0.004431965
15	0.02610359	0.007516754	0.003348717
16	0.02196599	0.008427087	0.003754271
17	0.03124530	0.018960946	0.005170393
18	0.01929329	0.015190287	0.004142186
19	0.03421591	0.019669706	0.005008588
20	0.02686856	0.017808580	0.004534681
21	0.02610359	0.007516754	0.004200186
22	**0.08401106**	0.002600953	0.001453352
23	**0.06907588**	0.003340779	0.001686542
24	0.01930765	0.009225435	0.004657323
25	0.03974123	0.020879871	0.004576502
26	0.06062437	0.024727114	0.005419750
27	0.01563891	0.012210029	0.003498855
28	0.03421591	0.019669706	0.005636469
29	0.03747995	0.005852144	0.002735610
30	0.02081651	0.008742091	0.004086528
31	0.01735730	0.014160743	0.004740317
32	0.01585651	0.011288858	0.003778952
33	0.03526355	0.019908849	0.005532527
34	0.01632537	0.013363118	0.003713515
35	0.02610359	0.007516754	0.003641699
36	0.03529675	0.006115129	0.002962643
37	**0.06907588**	0.003340779	0.001830961
38	0.02771250	0.007224869	0.003959690
39	0.01562500	0.012022905	0.004997587
40	0.03637399	0.005982914	0.002486931
41	0.01568333	0.011652991	0.004280695
42	0.01679198	0.010403811	0.003821812
43	0.02258381	0.008271753	0.003023497
44	0.01705625	0.013959169	0.005102366
45	0.01613917	0.013167324	0.004159099
46	0.01804555	0.014568226	0.004601595
47	0.01575287	0.012588612	0.003672756

48	0.02940846	0.018495665	0.005396152
49	0.03526355	0.019908849	0.005205530
50	0.02194715	0.016255955	0.004250414
51	0.01563891	0.012210029	0.005193058
52	0.04471872	0.005097868	0.002168179
53	0.02686856	0.017808580	0.004865146
54	0.02194715	0.016255955	0.004440983
55	0.03633989	0.020149437	0.005312192
56	0.02026869	0.015612219	0.004116001
57	0.03637399	0.005982914	0.003011463
58	0.03127491	0.006658439	0.003351484
59	0.02323033	0.008117863	0.003693779
60	0.02771250	0.007224869	0.003287450
61	**0.17689606**	**0.039899339**	0.007243794
62	0.02686856	0.017808580	0.003233178
63	0.01844532	0.009554889	0.003675733
64	0.01614456	0.010930504	0.004204927

Çizelge 13.2. Eşik Noktaları

	L_1	L_2	H_2
Eşik Noktaları	0.0625	0.02562419	0.008171392

EK 14. MATLAB Analiz Sonuçları

Çizelge 14.1. Gözlemler için Matlab analiz sonuçları

Gözlemler	CD(β)	|VR-1|	|CW|	AP	CD(u)	LD
1	0.0009	0.1454	0.0679	0.0294	0.0014	0.0155
2	0.0007	0.1465	0.0684	0.0291	0.0011	0.0152
3	0.0002	0.1492	0.0695	0.0282	0.0002	0.0144
4	0.0049	0.1251	0.0589	0.0302	0.0078	0.0223
5	0.0049	0.1255	0.0591	0.0301	0.0077	0.0222
6	0.0046	0.1270	0.0598	0.0299	0.0072	0.0216
7	0.0045	0.1272	0.0599	0.0298	0.0071	0.0215
8	0.0047	0.1264	0.0595	0.0299	0.0074	0.0218
9	0.0047	0.1266	0.0596	0.0299	0.0073	0.0218
10	0.0210	0.0471	0.0230	0.0362	0.0330	0.0633
11	0.0167	0.0675	0.0326	0.0337	0.0263	0.0501
12	0.0085	0.1073	0.0510	0.0293	0.0134	0.0295
13	**0.1491**	**0.4519**	**0.3007**	**0.1389**	**0.2350**	**1.4854**
14	**0.3147**	**0.8207**	**0.8594**	**0.3182**	**0.4958**	**9.4488**
15	0.0037	0.1316	0.0618	0.0328	0.0058	0.0200
16	0.0307	0.0015	0.0007	0.0543	0.0483	0.0995
17	0.0004	0.1480	0.0690	0.0282	0.0006	0.0147
18	0.0240	0.0326	0.0161	0.0442	0.0378	0.0737
19	0.0140	0.0806	0.0387	0.0340	0.0220	0.0427
20	0.0131	0.0850	0.0408	0.0335	0.0206	0.0403
21	0.0140	0.0806	0.0387	0.0340	0.0220	0.0427
22	0.0005	0.1477	0.0689	0.0283	0.0007	0.0148
23	0.0001	0.1493	0.0696	0.0279	0.0002	0.0143
24	0.0017	0.1416	0.0662	0.0296	0.0026	0.0167
25	0.0005	0.1474	0.0688	0.0284	0.0008	0.0149
26	0.0003	0.1486	0.0693	0.0282	0.0004	0.0146
27	0.0003	0.1486	0.0693	0.0282	0.0004	0.0146
28	0.0001	0.1496	0.0697	0.0272	0.0001	0.0142
29	0.0006	0.1471	0.0686	0.0268	0.0009	0.0150
30	0.0000	0.1500	0.0699	0.0276	0.0000	0.0142
31	0.0422	0.0506	0.0260	0.0697	0.0665	0.1544
32	0.0297	0.0060	0.0030	0.0580	0.0468	0.0955
33	0.0463	0.0687	0.0356	0.0736	0.0729	0.1772
34	0.0165	0.0684	0.0331	0.0387	0.0260	0.0496
35	0.0000	0.1500	0.0699	0.0276	0.0000	0.0142
36	0.0002	0.1490	0.0694	0.0276	0.0003	0.0144