

155726

**BÜYÜME EĞRİSİ MODELLERİNDE RAO 'NUN BASİT
KOVARYANS YAPISI VE AYKIRI GÖZLEMLER**

Ramazan ARSLAN

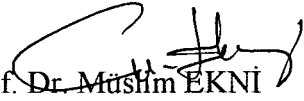
**YÜKSEK LİSANS TEZİ
(İSTATİSTİK)**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

Aralık 2004

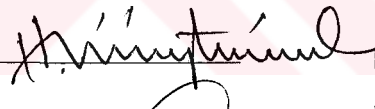
ANKARA

Ramazan ARSLAN tarafından hazırlanan...BÜYÜME EĞRİSİ MODELLERİNDE
RAO'NUN BASİT KOVARYANS YAPISI VE AYKIRI GÖZLEMLER...adlı bu
tezin...YÜKSEK LİSANS...tezi olarak uygun olduğunu onaylarım.

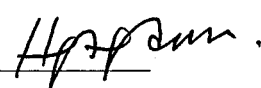

Prof. Dr. Müslim EKNİ

Tez Yöneticisi

Bu çalışma, jürimiz tarafından...İSTATİSTİK...Anabilim Dalında...YÜKSEK
LİSANS...tezi olarak kabul edilmiştir.

Başkan: : _Prof. Dr. Hüseyin TATLIDİL 

Üye : _Prof. Dr. Müslim EKNİ (Danışman) 

Üye : _Prof. Dr. Hamza GAMGAM 

Bu tez, Gazi Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygundur.



İÇİNDEKİLER

Sayfa

ÖZET	i
ABSTRACT	ii
TEŞEKKÜR	iii
ÇİZELGELERİN LİSTESİ	iv
ŞEKİLLERİN LİSTESİ.....	v
EKLERİN LİSTESİ.....	vi
SİMGELER VE KISALTMALAR	vii
1. GİRİŞ.....	1
2. BÜYÜME EĞRİSİ MODELLERİ.....	5
2.1. Model Yapıları.....	5
2.1.1. Gruplaşmamış Panel Veri	5
2.1.2. Gruplaşmış Panel Veri	7
2.1.3. Farklı Derecelere Sahip Olan Polinomlar.....	8
2.1.4. Eşlik Değişkenlerine Sahip Olan Büyüme Eğrileri	12
2.2. Kovaryans Yapısının Seçimi.....	13
3. GENELLEŞTİRİLMİŞ EN KÜÇÜK KARELER.....	17
3. 1. Genelleştirilmiş En Küçük Kareler Tahmini.....	17
4. EN ÇOK OLABİLİRLİK	23
4.1. Beklenen Değer ve Varyans-Kovaryans.....	24
4.2. Rao'nun Basit Kovaryans Yapısı.....	26
4.2.1. En çok olabilirlik tahmininin genelleştirilmiş en küçük kareler tahminine özdeşlik koşulu	26

4.2.2. Yayılış Bileşenlerinin Tahminleri.....	31
5. AYKIRI VE ETKİLİ GÖZLEMLER.....	35
5.1. Özet.....	35
5.2. Aykırı Gözlemler	41
5.3. Etkili Gözlemler	42
5.4. Aykırı ve Etkili Gözlemler Arasındaki İlişki	43
5.5. Çok Değişkenli Veri Kümesinde Çoklu Aykırı Gözlemler.....	44
5.6. Uyumsuz Aykırı Gözlem Üreten Model.....	48
6. UYGULAMA	51
6.1. Genelleştirilmiş En Küçük Kareler.....	55
6.1.1. Baş Çevresi	55
6.1.2. Boy Uzunluğu.....	56
6.1.3. Ağırlık.....	58
6.2. En Çok Olabilirlik Tahminleri.....	59
6.2.1. Baş Çevresi	59
6.2.2. Boy Uzunluğu.....	61
6.2.3. Ağırlık.....	63
6.3. Rao'nun Basit Kovaryans Yapısı.....	65
6.3.1. Baş Çevresi	65
6.3.2. Boy Uzunluğu.....	70
6.3.3. Ağırlık.....	73
6.4. Uyumsuz Aykırı Gözlem Bulma.....	75
6.4.1. Baş Çevresi	75
6.4.2. Boy Uzunluğu.....	79

6.4.3. Ağırlık.....	82
7. SONUÇ VE ÖNERİLER.....	86
KAYNAKLAR	89
EKLER	92
ÖZGEÇMİŞ	110



BÜYÜME EĞRİSİ MODELLERİNDE RAO'NUN BASİT KOVARYANS YAPISI VE AYKIRI GÖZLEMLER

(Yüksek Lisans Tezi)

Ramazan ARSLAN

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Aralık 2004

ÖZET

Bu çalışmanın amacı, Büyüme Eğrisi Modellerinde Rao 'nun Basit Kovaryans Yapısı 'nın incelenmesi ve aykırı değer bulma tekniğinin kullanılmasıdır. İlk olarak, Büyüme Eğrisi Modeli kullanılarak En Küçük Kareler ve En Çok Olabilirlik tahminlerinin hangi koşul altında daha iyi sonuçlar verdiği gösterilmiştir. İkinci olarak, Rao 'nun Basit Kovaryans Yapısı ve bu yapıdan türetilen bazı faydalı kovaryans yapılarına değinilmiştir. Son olarak, Basit Kovaryans Yapısına Sahip Olan Büyüme Eğrisi Modelleri 'nde Çoklu Uyumsuz Aykırı Değer Bulma probleminin çözümü bulunmaya çalışılmıştır. Teorik yapı verildikten sonra kapsamlı bir uygulama gerçekleştirilmiştir. Uygulama bölümünde, Gazi Üniversitesi Çocuk Sağlığı Hastalıkları Anabilim Dalı 'na ait üç farklı veri kümesi için hem kız hem de erkek bebeklerin büyüme süreçleri incelenmiştir. Bu veri kümelerinin Büyüme Eğrisi Modeli 'ne uygunluğuyla ilgilenilmiş ve böylece aykırı değerler belirlenmiştir.

Bilim Kodu : 0303
Anahtar Kelimeler : Büyüme Eğrisi Modelleri, Rao 'nun Basit Kovaryans Yapısı, Uyumsuz Aykırı Değerler
Sayfa Adedi : 110
Tez Yöneticisi : Prof. Dr. Müslim Ekni

**RAO'S SIMPLE COVARIANCE STRUCTURE IN GROWTH CURVE
MODELS AND OUTLIERS**

(M.Sc. Thesis)

Ramazan ARSLAN

**GAZI UNIVERSITY
INSTITUTE OF SCIENCE AND TECHNOLOGY**

December 2004

ABSTRACT

The aim of this study is to examine Rao 's simple covariance structure in Growth Curve Models and to use the technique of outlier identification. Firstly, by using Growth Curve Models it is showed that under which conditionals Least Square Estimation and Maximum Likelihood Estimation give good results. Secondly, Rao 's simple covariance structure and many useful covariance structures which are included in the simple covariance structure as special cases are discussed. Finally, in Growth Curve Models with simple covariance structure solution of the problem of multiple discordant outlier identification is discussed. Having given the theoretical structure, an extensive application is realized. In the application part, both girl and boy baby 's growth process is observed for each three data sets of Gazi University Medical Faculty Department Child Healty Diseases. These data sets are considered for corresponding to Growth Curve Models. Thus Outliers are identified.

Science Code : 0303

Key Words : Growth Curve Models, Rao' s Simple Covariance Structure,
Discordant Outliers

Page Number: 110

Adviser : Prof. Dr. Müslim Ekni

TEŞEKKÜR

Çalışmalarım boyunca değerli yardım ve katkılarıyla beni yönlendiren hocam Prof. Dr. Müslim Ekni 'ye yine uygulamamı gerçekleştirmek için gerekli olan verileri elde etme aşamasında yardımlarını esirgemeyen Gazi Üniversitesi Çocuk Sağlığı Hastalıkları Anabilim Dalı Başkanı Prof. Dr. Sedef Tunaoglu 'na ve Prof. Dr. Ufuk Beyazova 'ya teşekkürü bir borç bilirim.

Tüm yaşamım boyunca bana maddi ve manevi olanaklarını ellerinden geldiğince sunan ve başarılarımın altında her zaman imzaları olan sevgili anne ve babama şükranlarımı sunarım.

Hayatımın son beş yılında ve özellikle tezimin hazırlık aşamasında, gerek sabrıyla gerek yardımlarıyla beni hiçbir zaman yalnız bırakmayan ve bundan sonraki hayatımda da varlığıyla hep yanımda olacağına inandığım Sayın Hülya Akgönüllü 'ye teşekkür ederim.

Saygılarımla

2004

Ramazan ARSLAN

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 5.1. Anscombe'nin Veri Seti.....	38



ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. Gruplaşmamış Veri İçin Ortalama Büyüme Eğrisi	6
Şekil 2.2. Gruplaşmış Veri İçin Ortalama Büyüme Eğrisi	7
Şekil 2.3. Farklı Derecelere Sahip Büyüme Eğrisi	12
Şekil 5.1.(a,b) Anscombe'nin Veri Seti $D(x_1, y_1)$ ve $D(x_1, y_2)$ İçin Serpme Diyagramları ve Regresyon Doğruları	39
Şekil 5.1.(c) Anscombe'nin Veri Seti $D(x_1, y_3)$ İçin Serpme Diyagramı ve Regresyon Doğrusu.....	40
Şekil 5.1.(d) Anscombe'nin Veri Seti $D(x_2, y_4)$ İçin Serpme Diyagramı ve Regresyon Doğrusu.....	40
Şekil 5.2.(a) Etkili Gözlem İçin Serpme Diyagramı ve Regresyon Doğrusu.....	43
Şekil 5.2.(b) Aykırı Gözlem İçin Serpme Diyagramı ve Regresyon Doğrusu	43
Şekil 5.2.(c) Aykırı Gözlem İçin Serpme Diyagramı ve Regresyon Doğrusu	44
Şekil 6.1. Kız Bebeklerin Baş Çevrelerine İlişkin Doğrusallık Grafiği	52
Şekil 6.2. Erkek Bebeklerin Baş Çevrelerine İlişkin Doğrusallık Grafiği.....	52
Şekil 6.3. Kız Bebeklerin Boy Uzunluklarına İlişkin Doğrusallık Grafiği.....	53
Şekil 6.4. Erkek Bebeklerin Boy Uzunluklarına İlişkin Doğrusallık Grafiği.....	53
Şekil 6.5. Kız Bebeklerin Ağırlıklarına İlişkin Doğrusallık Grafiği	54
Şekil 6.6. Erkek Bebeklerin Ağırlıklarına İlişkin Doğrusallık Grafiği.....	54

EKLERİN LİSTESİ

Ekler	Sayfa
Ek-1. Matris Değişkenli Normal ve t Dağılımları.....	93
Ek-2. Kronecker Çarpımı ve $\text{vec}(\cdot)$ Operatörleri İle İlgili Bazı Özellikler.....	97
Ek-3. Q Ortogonal Matrisinin Bulunması.....	101
Ek-4. Uygulama İçin Kullanılan Veriler.....	104



SİMGELER VE KISALTMALAR

Bu çalışmada kullanılan bazı simgeler, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler	Açıklama
R^p	p boyutlu Eucliden uzayı
$\text{iz}(A)$	Karesel A matrisinin izi
$\det(A)$	Karesel A matrisinin determinanı
$\Gamma_m(\alpha)$	Çok değişkenli gamma fonksiyonu
	$\Gamma_m(\alpha) = \pi^{m(m-1)/4} \prod_{i=1}^m \Gamma(\alpha - (i - \frac{1}{2}))$
1_p	$1_p = (1,1,...,1) \in R^p$ olmak üzere p elemanlı birler vektörü
$\text{vec}(A)$	A matrisinin vektör operatörü
$\text{vec}'(A)$	$\text{vec}(A)$ matrisinin transpozu
A'	A matrisinin transpozu
A^{-1}	A matrisinin transpozu
I_p	(pxp) boyutlu özdeş (birim) matris
P_A	A'nın izdüşüm matrisi
	$P_A = A(A' A)^{-1} A'$
$Q_{(S)}$	$Q_{(S)} = S Q (Q' S Q)^{-1}$ yarı izdüşüm matrisi
$A \otimes B$	A ve B matrislerinin kronecker çarpımı

$\Sigma > 0$	Σ pozitif tanımlı matris
$\Sigma^{1/2}$	Σ matrisinin karekökü
$\text{Cov}(X)$	Rastgele X matrisinin varyans kovaryans matrisi
	$\text{Cov}(X) = \text{Cov}(\text{vec}(X))$
$N(\mu, \sigma^2)$	Beklenen değeri μ , varyansı σ^2 olan tek değişkenli normal dağılım
t_p	p serbestlik dereceli t dağılımı
χ_p^2	p serbestlik dereceli ki-kare dağılımı
$F_{n,m}$	n, m serbestlik dereceli F dağılımı
$\Lambda(p, m, n)$	Parametreleri p, m, n olan Wilks dağılımı
$N_p(\mu, \Sigma)$	Beklenen değer vektörü μ , yayılış matrisi ve $\Omega > 0$ olan matris değişkenli normal dağılım
$N_p(\mu, \Sigma, \Omega)$	Konum matrisi μ , yayılış matrisleri $\Sigma > 0$ ve $\Omega > 0$ olan matris değişkenli normal dağılım
$W_p(n, \Sigma)$	Parametreleri n ve $\Sigma > 0$ olan Wishard dağılımı

Kısaltmalar**GEKK****EÇOB****AKT****BLUE****Açıklama**

Genelleştirilmiş En Küçük Kareler

En Çok Olabilirlik

Artık Kareler Toplamı

(Best Lineer Unbiased Estimator)

En İyi Doğrusal Yansız Tahmin Edici



1. GİRİŞ

Büyüme eğrisi modelleri (GCM), çok değişkenli varyans analizi (MANOVA) modeli olarak genelleştirilmiştir. Bu model özellikle ekonomi, biyoloji, tıbbi araştırmalar ve epidemiolojik çalışmaların kısa zaman serilerinde araştırılan büyüme problemleri için bilhassa faydalıdır. Büyüme eğrisi modeli aşağıdaki gibi tanımlanmaktadır:

$$Y_{pxn} = X_{pxm} B_{mxr} Z_{rxn} + \varepsilon_{pxn}. \quad [1.1]$$

X ve Z sırasıyla (pxm) ve (rxn), (m<p) ve (r<n), boyutlu bilinen tasarım içi (within-design) ve tasarım arası (between-design) matrislerdir. Burada, regresyon katsayısı B bilinmiyor. Ayrıca normal dağılıma sahip olan ε 'ın sütunları; ortalama vektörü sıfır, ortak kovaryans matrisi $\Sigma > 0$ olan bağımsız p değişkenli vektördür. Bu gösterim;

$$Y \sim N_{p,n}(XBZ; \Sigma, I_n)$$

$$\varepsilon \sim N_{p,n}(0; \Sigma, I_n)$$

(Bkz. Ek-1)

biçimindedir. Büyüme eğrisi modeli [1.1] eşitliğinde, 2.1.1 - 2.1.4 tarafından anlatılacağı gibi, p, genellikle n durumun her birinde gözlenmiş zaman noktalarının sayısını, (m-1), zaman içindeki polinomun derecesini ve r, deneme gruplarının sayısını belirtmektedir. Açıkça, büyüme eğrisi modeli [1.1]'deki regresyon katsayı matrisi B 'nin sol tarafına eklenmiş olan tasarım içi matrisi X daha çok, çok değişkenli regresyon modelinin genişletilmiş bir durumudur. Örneğin, 2.1.2–2.1.3 'te görüleceği gibi, büyüme eğrisi modeli, panel çalışmaları gibi değişik alanlarda kümeleme çalışmaları ya da ilişkili veri (correlation data) sağlamada kullanılabilir. Bununla birlikte, regresyon modelindeki genişlemenin çok fazla olması modellerin istatistiksel sonuç çıkarmasının daha zor sağlanmasına neden olur. Çünkü B 'nin en çok olabilirlik tahmini, bağımlı değişken Y 'nin doğrusal bir fonksiyonu değildir. Bu nedenle, bu anlamdaki olağan çok değişkenli regresyon modeline göre büyüme

eğrisi modelinin istatistiksel sonuç çıkarması çok farklıdır ve bunun için de farklı şekillerde geniş çapta çalışmalara ihtiyaç duyulur.

Literatürde, büyüme eğrisi modelindeki B ve Σ parametreleri üzerinde tahmin metodlarının farklı olduğu belirtilir. Bunlar arasında başlıca kullanılan tahmin edicilerden ikisi, en çok olabilirlik (EÇOB) tahmini ve genelleştirilmiş en küçük kareler (GEKK) tahminidir. Regresyon katsayısı B 'nin genelleştirilmiş en küçük kareler tahmini, artık kare toplamı (AKT) üzerinden kayıp fonksiyon matrisinin izinin (trace) en küçüklenmesi ile , $\arg \min_B \text{iz}\{(Y-XBZ)(Y-XBZ)'\}$, bulunur ve bu yapı aşağıdaki gibi gösterilir:

$$\hat{B} = (X'X)^{-1} X'YZ'(ZZ')^{-1} . \quad [1.2]$$

Burada $\hat{B}$, bağımlı değişken Y 'nin doğrusal bir fonksiyonudur. En küçük kareler tahmini $\hat{B}$ 'ya dayanan kovaryans Σ 'nın tahmini de aşağıdaki gibi oluşturulabilir:

$$\hat{\Sigma} = \frac{1}{n} (Y - XBZ)(Y - XBZ)' . \quad [1.3]$$

$\hat{B}$, Gauss-Markov tahmini B 'nin en iyi doğrusal yansız tahmin edicisi (BLUE) olarak gösterilir. Alternatif olarak, iz kaybından ziyade artık kareler toplamı (AKT) üzerinde kayıp fonksiyon matrisinin determinantını en küçük yapan tahminle ilgileniyorsa, $\arg \min_B \det\{(Y - XBZ)(Y - XBZ)'\}$, bu durumda tahmin;

$$\hat{B}_{(S)} = (X'S^{-1}X)^{-1} X'S^{-1}YZ'(ZZ')^{-1} \quad [1.4]$$

biçiminde olur. $\hat{B}_{(S)}$ gösterimi, S matrisinin bir fonksiyonudur ve $S = Y(I_n - P_Z)Y'$ biçimindedir. Eğer $A'A$ tekil değilse, $P_A = A(A'A)^{-1}A'$ ifadesi A matrisinin iz düşümünü (projection) gösterir. Burada S matrisi, $n > p+r$ olduğu sürece pozitif

tanımlıdır. Aslında $\hat{B}_{(s)}$ tahmini, normal dağılım varsayımı altında B 'nin en çok olabilirlik (EÇOB) tahminine karşılık gelmektedir. Ayrıca, kovaryans Σ 'nın en çok olabilirlik tahmini;

$$\begin{aligned}\hat{\Sigma}_{(s)} &= \frac{1}{n}(Y - X\hat{B}_{(s)}Z)(Y - X\hat{B}_{(s)}Z)' \\ &= \frac{1}{n}(S + Q_S Y P_Z' Y' Q_S')\end{aligned}\quad [1.5]$$

biçimindedir. Burada,

$$Q_S = SQ(Q'SQ)^{-1}Q' \text{ ve } Q \in Q, \quad [1.6]$$

$$Q = \{Q \mid Q : p \times (p - m), \text{rank}(Q) = p - m \text{ ve } X'Q = 0\}$$

biçimindedir. Eğer yığın simetrik dağılımlar sınıfına genişletilirse bazı hipotez testleri için hem olabilirlik oran kriteri hem de bu sonuçlar denk tutulur. Normallik varsayımı altında, [1.4] ve [1.5] 'deki en çok olabilirlik tahminlerinin bazı asimptotik dağılım özellikleri ele alınmaktadır (1). Ayrıca $\hat{B}_{(s)}$ ve $\hat{\Sigma}_{(s)}$ 'nin yüksek momentleri için bazı formüller türetilmiştir (1,2). Çoğunlukla, genelleştirilmiş en küçük kareler tahminleri $\hat{B}$ ve $\hat{\Sigma}$ 'ya göre en çok olabilirlik tahminleri $\hat{B}_{(s)}$ ve $\hat{\Sigma}_{(s)}$ farklılık gösterir. Çünkü, en çok olabilirlik tahmini $\hat{B}_{(s)}$, bağımlı değişken Y 'nin doğrusal fonksiyonu değildir. Bu fark, büyüme eğrisi modeli için istatistiksel sonuç çıkarmayı olağan regresyon analizine göre daha zor hale getirir. Doğal olarak, genelleştirilmiş en küçük kareler tahminleri $\hat{B}$ ve $\hat{\Sigma}$ ile en çok olabilirlik tahminlerinin sırasıyla $\hat{B}_{(s)}$ ve $\hat{\Sigma}_{(s)}$ ile bütünüyle çakışması şartıyla ilgilenilir. Ancak ve ancak kovaryans Σ aşağıdaki şekilde ifade edilen Rao 'nun basit

kovaryans yapısında (simple covariance structure = SCS) ise $\hat{B}_{(S)} \equiv \hat{B}_{(I_p)} = \hat{B}$ ve $\hat{\Sigma}_{(S)} \equiv \hat{\Sigma}_{(I_p)} = \hat{\Sigma}$ 'dır.

$$\hat{\Sigma} = \Lambda \Gamma \Lambda' + Q \Theta Q' \quad [1.7]$$

Burada Γ ve Θ , sırasıyla boyutları $(m \times m)$ ve $(p-n) \times (p-n)$ olan bilinmeyen pozitif tanımlı yayılış matrislerini göstermektedir ve $Q \in \mathcal{Q}$ 'dur. Büyüme analizinde, Rao'nun basit kovaryans yapısı çok önemlidir çünkü, doğrusal olmayan bir istatistiksel sonucu, $\hat{B}_{(S)}$ tahminine dayanarak doğrusal olana indirgeyebilir.



2. BÜYÜME EĞRİSİ MODELLERİ (Growth Curve Models)

2.1. Model Yapıları

2.1.1. Gruplaşmamış panel (longitudinal) veri

Varsayalım ki n tane birey, hepsi aynı şartlara bağlı ve her biri $t_1, t_2, \dots, t_p$ olacak şekilde p zaman içinde gözlenmiş olsun. p gözlemlili bireylerin bağımsız olmadığı ancak bilinmeyen kovaryans matrisi $\Sigma > 0$ 'a sahip çok değişkenli normal dağıldığı varsayılır. Ayrıca, farklı bireylerin gözlem değerlerinin bağımsız olduğu varsayılmaktadır. Bu durumların amacı, her bir birey için büyüme çalışması yapmak ve bireyler için büyümenin ortalama oranının araştırılmasını sağlamaktır. Panel (longitudinal) verinin modellenmesi, genellikle zamanın polinom fonksiyonunu büyüme eğrisine uydurma gibi değerlendirilebilir. Örneğin, büyüme eğrisi $(m-1)$ zaman dereceli polinom fonksiyonu gibi seçilebilir. Böylece t zamandaki herhangi bir bireyin ölçümünün beklentisi;

$$b_0 + b_1 t + \dots + b_{m-1} t^{m-1} \quad [2.1]$$

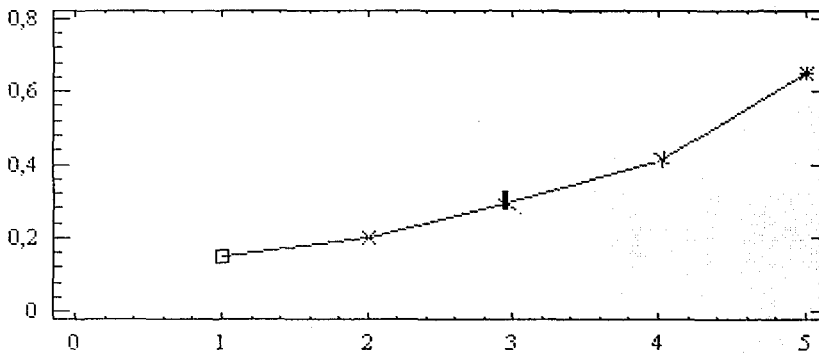
şeklinde olur. Burada m , bireylerin ortalama büyüme oranlarını yansıtan önceden seçilmiş bir tamsayı değeridir. Bireylerin bağımlı matrisi $Y = (y_1, y_2, \dots, y_n)$ şeklinde belirtilmekle beraber Y , $(p \times n)$ boyutludur. Burada $y_j = (y_{1j}, y_{2j}, \dots, y_{pj})'$, p zaman noktasında j .nci birey üzerinde ölçülen p değişkenli yanıt vektörünü göstermektedir.

$$X = \begin{bmatrix} 1 & t_1 & . & . & . & t_1^{m-1} \\ 1 & t_2 & . & . & . & t_2^{m-1} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & t_p & . & . & . & t_p^{m-1} \end{bmatrix} \quad [1]$$

X, boyutu (p x m) olan açıklayıcı matrisi (explanatory matrix) göstermektedir. Regresyon katsayı vektörü B, boyutu m olan $B = (b_0, b_1, \dots, b_{m-1})$ 'lerden oluşur. Her biri birlerden meydana gelen n elemanlı vektör Z ise, $Z = 1'_n = (1, 1, \dots, 1)$ biçiminde gösterilir. Bağımlı matris olan Y 'nin beklenen değeri, [2.1] ve [1] 'e dayandırılarak matris biçiminde açıkça aşağıdaki gibi yazılabilir:

$$E(Y) = XBZ. \quad [2.2]$$

Burada, Y 'nin sütunları $N_p(0, \Sigma)$ olacak şekilde bağımsız dağılır. Şekil 2.1. 'de bu model ayrıca resmedilerek verilir. Grafikte ölçümler beş zaman noktasında alınır ve ortalama büyüme eğrisi ikinci dereceden polinom tarafından uygunlaştırılır.



Şekil 2.1.

Şekil 2.1. Gruplaşmamış veri için ortalama büyüme eğrisi.

Eğri ikinci dereceden polinoma uymaktadır, (Not: y-ekseni büyümeyi, x-ekseni de zamanı göstermektedir).

2.1.2. Gruplaşmış panel veri

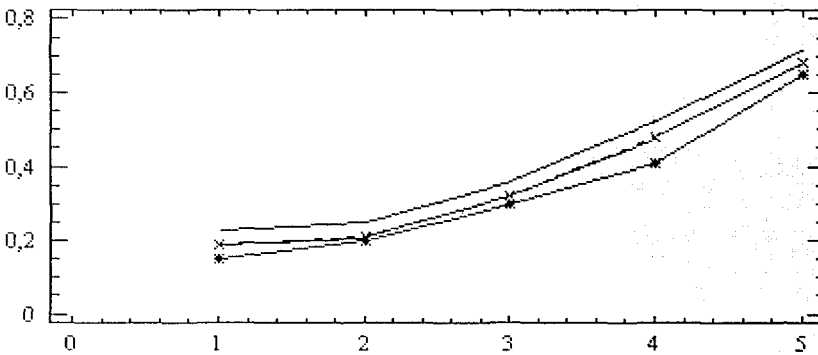
Varsayalım, her grup farklı deneme şartlarından dolayı yanıt ölçümleri $y_1, y_2, \dots, y_n$ olacak şekilde r grupta sınıflandırılmış olsun. Bu, panel veri çalışmalarında genel bir durumdur. j .nci grupta aynı şartlar altında denenen n_j bireyin olduğu varsayılır.

Burada, $j = 1, 2, \dots, r$ ve $\sum_{j=1}^r n_j = n$ 'dir. Gruplaşmamış panel veri kümesinde olduğu

gibi tüm gruptaki bireyler, $t_1, t_2, \dots, t_p$ olmak üzere zaman içinde aynı noktalarda ölçüldükleri ve temel kovaryans matrisinin $\Sigma > 0$ olduğu varsayılır. Genelde, gruplaşmış panel veri kümesinin ortalama büyüme eğrisi, gruplara göre farklılık gösterir ancak, her veri grubuna ayarlanmış olan polinomların derecesi aynı şekilde tespit edilebilir. Örneğin, ortalama büyüme eğrisi ile ilgili olan j .nci grubun beklentisinin:

$$b_{0,j} + b_{1,j}t + \dots + b_{m-1,j}t^{m-1} \quad [2.3]$$

şeklinde olduğunu varsayalım. Aşağıdaki gösterimler bu ifadeyi iyi bir biçimde temsil etmektedir:



Şekil 2.2.

Şekil 2.2. Gruplaşmış veri için ortalama büyüme eğrisi. Her grup, ortalama büyüme eğrisi için ikinci dereceden polinoma uymaktadır, (Not: y-ekseni büyümeyi, x-ekseni de zamanı göstermektedir).

[2.3] eşitliğini [2.2] 'deki gibi aynı matris biçiminde yazmak için X matrisi [1] 'deki gibi alınırsa, regresyon katsayı matrisi B ve tasarım matrisi Z aşağıdaki gibi yazılabilir:

$$B = \begin{bmatrix} b_{01} & b_{02} & . & . & . & b_{0r} \\ b_{11} & b_{12} & . & . & . & b_{1r} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ b_{m-1,1} & b_{m-1,2} & . & . & . & b_{m-1,r} \end{bmatrix} \quad [2]$$

$$Z = \begin{bmatrix} 1_{n_1} & 0 & . & . & . & 0 \\ 0 & 1_{n_2} & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & 1_{n_r} \end{bmatrix} \quad [3]$$

Başka bir deyişle, Z matrisi, r satır n sütunda oluşmaktadır. n_1 .nci sütunu $(1, 0, \dots, 0)'$, n_2 .nci sütunu $(0, 1, \dots, 0)'$, ..., ve n_r .nci sütunu $(0, 0, \dots, 1)'$ 'lerden meydana gelir. Bu şekilde ortalama büyüme eğrisi $E(Y)$, tekrar [2.2] 'deki birleştirilmiş biçim gibi ifade edilebilir. Şekil 2.2., bu gruplaşmış büyüme eğrisi modelini örneklemektedir. Bireyler üç farklı deneme grubunda sınıflandırılarak beş zaman noktasında gözlenmiştir. Grupların ortalama büyüme eğrisine uygunluğu için ikinci dereceden, $m=3$, polinomlar kullanılır.

2.1.3. Farklı derecelere sahip olan polinomlar

Farklı gruplardaki ortalama büyüme eğrisine uygunluk için kullanılan polinomlar aynı, yani $(m-1)$ derecelidir. Pratikte, farklı gruplardaki büyüme eğrilerine dereceleri farklı olan polinomların ayarlanması pek yaygın bir yöntem değildir. Örneğin, gruplaşmış

panel veride, verilen gruplaşmış cevaplar için ilk iki deneme grubunda ortalama büyüme eğrilerine, dereceleri (m-3) ve (m-2) olan polinomlar ayarlanmışsa bu durumda ortalama büyüme eğrilerinin beklentileri;

$$b_{0,1} + b_{1,1}t + \dots + b_{m-3,1}t^{m-3} \quad [2.4]$$

ve

$$b_{0,2} + b_{1,2}t + \dots + b_{m-3,2}t^{m-3} + b_{m-2,2}t^{m-2} \quad [2.5]$$

biçiminde gösterilir ve burada $m, r \geq 3$ olmalıdır. Geriye kalan ortalama büyüme eğrilerinin tümüne derecesi (m-1) olan polinom ayarlanmış olsun. Bu durumda ifademiz;

$$b_{0,j} + b_{1,j}t + \dots + b_{m-3,j}t^{m-3} + b_{m-2,j}t^{m-2} + b_{m-1,j}t^{m-1} \quad [2.6]$$

biçiminde olur. Burada, $j = 3, 4, \dots, r$ 'dir. Kullanılan matris notasyonları tarafından, [2.4], [2.5] ve [2.6] 'ün büyüme eğrilerinin beklenen değeri;

$$E(Y) = X_1 B_1 Z_1 + X_2 B_2 Z_2 + X_3 B_3 Z_3 \quad [2.7]$$

gibi yazılabilir.

Tasarım matrisi X_i , ($i = 1, 2, 3$) olmak üzere;

$$X_1 = \begin{bmatrix} 1 & t_1 & . & . & . & t_1^{m-3} \\ 1 & t_2 & . & . & . & t_2^{m-3} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & t_p & . & . & . & t_p^{m-3} \end{bmatrix} \quad X_2 = \begin{bmatrix} t_1^{m-2} \\ t_2^{m-2} \\ . \\ . \\ . \\ t_p^{m-2} \end{bmatrix} \quad X_3 = \begin{bmatrix} t_1^{m-1} \\ t_2^{m-1} \\ . \\ . \\ . \\ t_p^{m-1} \end{bmatrix} \quad [4]$$

şeklinde, regresyon katsayı matrisi B_i ($i = 1, 2, 3$) olmak üzere;

$$B_1 = \begin{bmatrix} b_{01} & b_{02} & \cdot & \cdot & \cdot & b_{0r} \\ b_{11} & b_{12} & \cdot & \cdot & \cdot & b_{1r} \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ b_{m-3,1} & b_{m-3,2} & \cdot & \cdot & \cdot & b_{m-3,r} \end{bmatrix} \quad B_2 = \begin{bmatrix} b_{m-2,1} \\ b_{m-2,2} \\ \cdot \\ \cdot \\ \cdot \\ b_{m-2,r} \end{bmatrix} \quad B_3 = \begin{bmatrix} b_{m-1,1} \\ b_{m-1,2} \\ \cdot \\ \cdot \\ \cdot \\ b_{m-1,r} \end{bmatrix} \quad [5]$$

şeklinde ve tasarım matrisi Z_i de ($i = 1, 2, 3$) olmak üzere;

$$Z_1 = \begin{bmatrix} 1_{n_1} & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1_{n_2} & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & 1_{n_r} \end{bmatrix}$$

$$Z_2 = \begin{bmatrix} 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1_{n_2} & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1_{n_3} & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 1_{n_r} \end{bmatrix} \quad Z_3 = \begin{bmatrix} 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1_{n_3} & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 1_{n_r} \end{bmatrix} \quad [6]$$

şeklinde verilir. Büyüme eğrisi modeli [1.7] için tasarım matrisi X , $X = (X_1: X_2: X_3)$ biçiminde gösterilirse B ve Z aşağıdaki biçimde yazılabilir:

$$B = \begin{bmatrix} B_1 & 0 & 0 \\ 0 & B_2 & 0 \\ 0 & 0 & B_3 \end{bmatrix} \quad \text{ve} \quad Z = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix}. \quad [7]$$

Bu durumda, birleştirilmiş matris biçimi [2.2] yani, $E(Y) = XBZ$ elde edilir. Alternatif olarak, gruplaşmış panel veride olduğu gibi X, B ve Z matrisleri aynı biçimde alınabilir. Fakat bu sefer de regresyon katsayı matrisi B aşağıdaki kısıtlara bağlı olur:

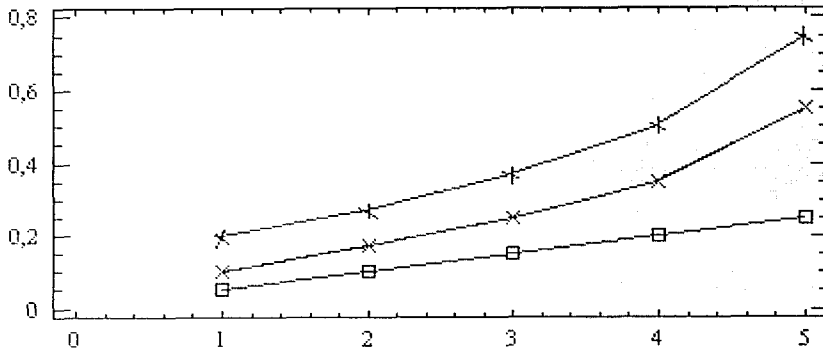
$$b_{m-3,r} = 0, b_{m-2,r} = 0 \text{ ve } b_{m-1,r} = 0.$$

$$K_1 B L_1 = 0 \quad \text{ve} \quad K_2 B L_2 = 0 \quad [2.8]$$

Burada, $K_1 = (0, 0, \dots, 1, 0)$ ve $K_2 = (0, 0, \dots, 0, 1)$ 'lerden oluşan $(1 \times m)$ boyutlu satır vektörlerini göstermekte ve L_i ;

$$L_1 = \begin{bmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \quad \text{ve} \quad L_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 0 & 0 \end{bmatrix} \quad [8]$$

biçiminde $(r \times 1)$ ve $(r \times 2)$ boyutlu, r değişkenli vektörleri gösterir. Bu durumda, büyüme eğrisini ayarlamak için yayılım matrisi (dispersion matrix) Σ 'nın yanı sıra [2.8] kısıtını sağlayan regresyon katsayısı matrisi B'nin de tahminine ihtiyaç duyulur. Şekil 2.3., gruplaşmış büyüme eğrisinin bu türünü örneklemektedir. Bireyler üç grupta ve beş zaman noktasında ölçülüp, ortalama büyüme eğrilerine ya birinci dereceden, $(m-1)$, ya da ikinci dereceden, $(m-2)$, polinomlar ayarlanır. Bu çeşit gruplaşmış büyüme eğrisi için Şekil 2.3. aşağıdaki gibi verilir:



Şekil 2.3.

Şekil 2.3. Farklı derecelere sahip büyüme eğrisi.

Beş eşit zaman noktasında ($t = 1, 2, 3, 4, 5$) bütün veriler gözlemlenerek, üç farklı deneme grubu içinde bireyler sınıflandırılmıştır, (Not: y-ekseni büyümeyi, x-ekseni de zamanı göstermektedir).

2.1.4. Eşlik değişkenlerine (concomitant variables) sahip olan büyüme eğrileri

Zamanın tümünde gözlemlenen bireylerin büyümelerinde etkin olan bazı arka plan değişkenlerini hesaba katmaya ihtiyaç duyulur. Örneğin, gruplaşmış panel veride, panel veri çalışmalarında her bir birey için başlangıç ağırlıklarının (starting weights) (mesela hayvanlar için doğum ağırlıkları) hesaba katılması zorunludur. Genellikle, bireylerin büyümeleri üzerindeki ağırlıkların etkisi gruptan gruba farklı olabilir. Bu durumda ortalama büyüme eğrileri, eşlik değişkenlerine sahip olan büyüme eğrisi gibi aşağıdaki biçimde modellenenabilir:

$$E(Y) = XBZ + XB_1Z_1 . \quad [2.9]$$

Burada, X , B ve Z matrisleri [1], [2] ve [3] 'deki gibi alınır. Büyüme üzerindeki başlangıç ağırlıklarının etkisini açıklayan yeni parametre B_1 , ($m \times r$) boyutlu matrisi, başlangıç ağırlıklarını veren Z_1 ise ($r \times n$) boyutlu matrisi göstermektedir.

$$Z_1 = \begin{bmatrix} w_{11} & . & . & . & w_{1n_1} & 0 & . & . & . & 0 & . & . & . & 0 & . & . & . & 0 \\ 0 & . & . & . & 0 & w_{21} & . & . & . & w_{2n_2} & . & . & . & 0 & . & . & . & 0 \\ . & . & . & . & . & . & . & . & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . & . & . & . & . & . & . & . & . \\ 0 & . & . & . & 0 & 0 & . & . & . & 0 & . & . & . & w_{r1} & . & . & . & w_{rn_r} \end{bmatrix} \quad [9]$$

$$(i = 1, 2, \dots, r) \quad (j = 1, 2, \dots, n_i)$$

Burada w_{ij} , i.nci gruptaki j.nci bireyin başlangıç ağırlığını göstermektedir. Başka bir deyişle, büyüme üzerinde başlangıç ağırlıklarının etkisi her grup için aynı olduğunda Z_1 aşağıdaki gibi varsayılabilir:

$$Z_1 = (w_{11} \dots w_{1n_1}; w_{21} \dots w_{2n_2}; \dots; w_{r1} \dots w_{rn_r}) . \quad [10]$$

Bu durum ile her bir bireyin büyüme üzerindeki başlangıç ağırlığının etkisi izah edilebilir. Regresyon katsayısı ve tasarım arası (between-design) matrisleri ($B:B_1$) ve $(Z' : Z_1')$ biçiminde gösterildiği sürece [2.9] eşitliğindeki büyüme eğrisi modeli [2.2] eşitliğinin birleştirilmiş biçimi gibi yazılabilir.

Yukarıda verilen 2.1.1 - 2.1.4, gruplaşmış ya da gruplaşmamış ölçümler için bütün büyüme eğrilerinin birleştirilmiş matris modeli şeklinde yani $E(Y)=XBZ$ biçiminde olduğunu göstermektedir.

2.2. Kovaryans Yapısının Seçimi

Uygulamalı büyüme veri kümesi için doğal problem, büyüme eğrisi modeli için en uygun kovaryans yapısının nasıl seçileceğidir. Yani, bir bakıma da kovaryans matrisi Σ 'la ilgili olarak modelin seçilmesidir. İncelenen noktanın kestiriminden, büyüme eğrisi ortalamaları için birkaç seçim ölçütü sağlanarak uygun modelin seçimi için doğru örnek metodlarının kullanılması ileri sürülmektedir (3). Büyüme

eğrisi modeli için, kuşkulu (diagnostics) gözlemler üzerindeki araştırmada kovaryans yapı seçimi çok önemli rol oynadığından son derece yaygın biçimde kullanılan iki metodun nasıl takip edileceği Lee tarafından müzakere edilmektedir (3). Büyüme eğrisi modeli için, model seçimi ile ilgili daha detaylı bilgiler bulunabilir (3,4).

Elimizde büyüme veri kümesi için üretilmiş q olası modelin $M_1, M_2, \dots, M_q$ biçiminde olduğunu varsayalım. Örneğin, oluşturulan büyüme eğrisi modelinin olası kovaryans yapıları;

$M_1 : \Sigma$, keyfi pozitif tanımlı matris

$M_2 : \Sigma = ATX' + Q\Theta Q'$ (SCS, $Q \in Q$)

$M_3 : \Sigma = \sigma^2 \{(1 - \rho)I_p + \rho 1_p 1_p'\}$

$M_4 : \Sigma = ATX' + \sigma^2 I_p$

$M_5 : \Sigma = \sigma^2 (\rho^{|i-j|}) \quad 1 \leq i, j \leq p$

$M_6 : \Sigma = \sigma^2 G$ (G , verilen küresel (spherical) kovaryans yapısı)

biçiminde türetmelere devam edilebilir. $q=2$ iken, uygun kovaryans yapısı seçimi için olabilirlik oran testi yöntemi genelde kabul edilebilir. Bununla birlikte, uygulamada $q > 2$ için ikiden çok alternatif model bulunmaktadır. Bundan dolayı, seçilen diğer kriterlere dikkat edilmelidir. Büyüme eğrisi modeli, yapısal olmayan (unstructure) kovaryans yapısında uygun olur. Bunun anlamı; Σ , M_1 'de tanımlandığı gibi pozitif tanımlı keyfi bir matris olarak ele alınır. Literatürde söz konusu M_1 , yapısal olmayan kovaryans (unstructure covariance = UC) yapısını belirtir. Bu çalışmada, M_2 ile gösterilen Rao 'nun basit kovaryans yapısı üzerinde

durulacaktır. Diğer taraftan, Rao 'nun basit kovaryans yapısından bazı faydalı kovaryans yapıları türetilmiştir. Örneğin $\sigma^2 > 0$ ve $\Gamma, \Gamma > 0$ olan $(p \times p)$ boyutlu matrisi göstermek üzere $\Sigma = X\Gamma X' + \sigma^2 I_p$ biçiminde ifade edilen rastgele regresyon katsayısı yapısı Rao 'nun basit kovaryans yapısının özel bir durumudur. Bu kovaryans yapısı, doğrusal karma modellerin özel bir biçimi olan rastgele katsayılı regresyon modellerinin bağlamından ortaya çıkar. Bir başka örnek ise, X sütun uzayında tanımlanan $1_p, 1_p = (1, 1, \dots, 1)$ biçiminde p değişkenli vektör olduğu sürece (3,5);

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & . & . & . & \rho \\ \rho & 1 & . & . & . & \rho \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \rho & \rho & . & . & . & 1 \end{bmatrix} \quad [11]$$

tekdüze kovaryans yapısı olarak adlandırılan ya da basitçe $\hat{\Sigma} = \sigma^2 \{(1 - \rho)I_p + \rho 1_p 1_p'\}$ eşitliği de diğer bir özel durumudur. Burada, $\sigma^2 > 0$ ve $\rho \in (-1 / (p-1), 1)$ bilinmeyen parametrelerdir. Bazı yığınların gözlem değerleri karma olduğunda, özellikle zaman serisi verisi için bu kovaryans yapısı faydalı olmadığı halde büyüme eğrisi verisi için bu kovaryans yapısının faydalı olabileceği vurgulanmaktadır (5). Ayrıca, Rao 'nun basit kovaryans yapısı için, yayılış parametreleri (dispersion parameters) Γ ve Θ 'nın en çok olabilirlik tahminleri de sırasıyla aşağıdaki gibidir (6):

$$\begin{aligned} \hat{\Gamma} &= \frac{1}{n} (X'X)^{-1} X' S X (X'X)^{-1} \\ \hat{\Theta} &= \frac{1}{n} (Q'Q)^{-1} Q' Y Y' Q (Q'Q)^{-1} \end{aligned} \quad [2.10]$$

Rao 'nun basit kovaryans yapısından üretilen diğer bir kovaryans yapısı da seri kovaryans yapısı olarak adlandırılır ve aşağıdaki gibi gösterilir:

$$\Sigma = \sigma^2(\rho^{|i-j|}) \quad 1 \leq i, j \leq p. \quad [2.11]$$

Burada, $\sigma^2 > 0$, $|\rho| < 1$ ve her ikisi de bilinmiyor. Her ne kadar bu yapı, doğrusal olmayan $\hat{B}_{(s)}$ ve $\hat{\Sigma}_{(s)}$ tahminlerini doğrusal olan tahminlere indirgeyemese de büyüme problemlerinin kısa zaman serilerinde önemli rol oynar. [2.11] yapısı altında B , σ^2 ve ρ parametrelerinin en çok olabilirlik tahminleri Lee tarafından elde edilmiştir (5). Gerçekte yapısal olmayan kovaryans (UC), pratikte en genel durumdur ve herhangi bir yapısal kovaryans matrisi ile mukayese için ana ilke gibi hizmet edebilir. Büyüme eğrisi modeli hakkında daha fazla detay için von Rosen 'nin inceleme makalesine bakılabilir (2).

3. GENELLEŞTİRİLMİŞ EN KÜÇÜK KARELER

3.1. Genelleştirilmiş En Küçük Kareler Tahmini (GEKK)

Bu bölümde, büyüme eğrisi modeli [1.1] 'deki regresyon katsayısı B 'nin genelleştirilmiş en küçük kareler tahminini inceleyeceğiz. Bu tahmin temel alınarak yayılış parametresi Σ 'nın yansız tahmini oluşturulur. Bu konuyu anlatmadan önce olağan çok değişkenli regresyon modelindeki (OMRM) genelleştirilmiş en küçük kareler tahminini kısaca gözden geçirelim:

$$Y_{pxn} = B_{pxr} Z_{rxn} + \varepsilon_{pxn}. \quad [3.1]$$

Burada, boyutu rxn ($r < n$) olan açıklayıcı tasarım matrisi Z ve regresyon katsayısı B bilinmemektedir. Ayrıca normal dağılıma sahip olan hata matrisi ε 'un sütunları; ortalama vektörü sıfır, ortak kovaryans matrisi $\Sigma > 0$ olan p değişkenli vektördür ve ortak kovaryans matrisi $\Sigma > 0$ bilinmemektedir. Büyüme eğrisi modeli [1.1] ile karşılaştırıldığında bu iki model arasındaki tek fark, olağan çok değişkenli regresyon modeli [3.1] 'deki regresyon katsayısının sol tarafındaki çoklu X tasarım matrisinin olmayışdır. Bu farklılık bu iki modelin istatistiksel sonucunu da farklı yapar. Olağan çok değişkenli regresyon modeli [3.1] için, çapraz çarpımlar ve çok değişkenli artık kareler toplamı;

$$Q(B) = (Y - BZ)(Y - BZ)' \quad [3.2]$$

biçiminde tanımlanmaktadır. Matrisin kesin nicelik ölçüleri olan determinantı ve izi, negatif olmayan matris tanımından modelin uygunluğunun ölçümü için kullanılabilir. Örneğin, regresyon katsayısı B 'nin en küçük kareler tahmini $\hat{B}$, elde edilen artık kare toplamlarının izini en küçük yapan B 'nin değerine karşılık gelir ve bu ifade aşağıdaki gibidir:

$$\hat{B} = \arg \min_B \{iz(Q(B))\}. \quad [3.3]$$

Burada $Q(B)$ matrisinin izi, $iz(Q(B))$ biçiminde gösterilir ve B , $(m \times r)$ boyutlu gerçekte matris uzayına karşılık gelir. Bilindiği gibi en küçük kareler tahmini $\hat{B}$;

$$\hat{B} = YZ'(ZZ')^{-1} \quad [3.4]$$

biçimindedir. Dikkat edileceği gibi [3.4]’de verilen $\hat{B}$ tahmini aşağıda ifade edilen dört niceliğin gerçekten de en küçük olmalarını sağlar (6):

1. Artık kareler toplamı (negatif olmayan tanımlı)
2. $Q(B)$ ’nin izi , $iz(Q(B))$
3. $Q(B)$ ’nin determinantı, $\det(Q(B))$
4. $Q(B)$ ’nin en büyük özdeğeri, $\lambda_{\max}(Q(B))$.

Birinci nicelikten, $Q_1 \geq Q_2$ olduğu durum yani $Q_1 - Q_2$ farkının negatif olmadığı durum için, negatif olmayan tanımlı matris anlamında $Q(B)$ ’yi $\hat{B}$ en küçüklemektedir. Bunun yanı sıra ölçümün dört farklı kriteri altında, olağan çok değişkenli regresyon modeli için regresyon katsayısı B ’nin en küçük hata kareleri tümüyle aynıdır. Buna rağmen, büyüme eğrisi modeli için çoğunlukla bu sonuçlar doğru değildir.

Olağan çok değişkenli regresyon modeli açısından düşünüldüğünde, büyüme eğrisi modeli [1.1] için çapraz çarpımlar ve çok değişkenli artık karelerin toplamı;

$$Q(B) = (Y - XBZ)(Y - XBZ)' \quad [3.5]$$

biçiminde tanımlanabilir.

Tanım 3.1.1.

Büyüme eğrisi modeli [1.1] 'de, eğer aşağıdaki eşitlik sağlanıyorsa, tahmin çifti $(\hat{B}, \hat{\Sigma})$ parametre çifti (B, Σ) 'nın en çok olabilirlik tahmini olarak bilinir ve bu tahmin çifti aşağıdaki gibi gösterilir:

$$(\hat{B}, \hat{\Sigma}) = \arg \min_{(B, \Sigma)} \{L(B, \Sigma)\} .$$

Burada $L(B, \Sigma)$, genelleştirilmiş en küçük kareler tahmininin olabilirlik fonksiyonudur ve bu fonksiyon aşağıdaki gibi tanımlanır (6):

$$L(B, \Sigma) = (2\pi)^{-pn/2} [\det(\Sigma)]^{-n/2} \exp\left\{-\frac{1}{2} \text{iz}(\Sigma^{-1}(Y - XBZ)(Y - XBZ)')\right\}.$$

Ayrıca B , $(m \times r)$ boyutlu matris uzayında ve Σ , $(p \times p)$ boyutlu pozitif tanımlı matris uzayında yer alır.

Regresyon katsayısı B 'nin genelleştirilmiş en küçük kareler tahmini öyle tahmin edilir ki, artık kareler toplamı $Q(B) = (Y - XBZ)(Y - XBZ)'$ 'nin iz fonksiyonunu en küçükler. Doğal olarak, en çok olabilirlik tahmini $\hat{B}$ 'nin artık kareler toplamı $Q(B)$ ile nasıl bir ilişki içerisinde olduğu sorulabilir? Gerçekte en çok olabilirlik tahmini $\hat{B}$ önemsizdir ancak, çapraz çarpımlar ve artık kareler toplamının determinant fonksiyonunu en küçükleyen tahmin aşağıdaki gibidir:

$$\hat{B} = \arg \min_{(B)} \{\det[(Y - XBZ)(Y - XBZ)']\}.$$

Tanım 3.1.2.

Büyüme eğrisi modeli [1.1] 'de, eğer $\hat{B}$ aşağıdaki şartı yeterince yerine getirebiliyorsa, regresyon katsayısı B 'nin genelleştirilmiş en küçük kare tahmini olarak bilinir ve bu tahmin edici aşağıdaki gibi gösterilir:

$$\hat{B} = \arg \min_B \{iz(Q(B))\}. \quad [3.6]$$

Burada, gözden geçirilen (run over) B, (mxr) boyutlu gerçek matris uzayıdır.

Regresyon katsayısı B 'nin genelleştirilmiş en küçük kareler tahmini nasıl öğrenilebilir?, [3.6] 'nin bulunan bu çözümü doğrusu nasıl gerçekleştirilebilir?. Bu soruların cevabı takip eden teoremden bulunabilir.

Teorem 3.1.1.

Büyüme eğrisi modeli [1.1] için, ancak ve ancak aşağıdaki tahmin edici eşitlik yeteri derecede sağlanıyorsa $\hat{B}$ tahmini regresyon katsayısı B 'nin genelleştirilmiş en küçük kareler tahmini olur:

$$(X'X)\hat{B}(ZZ') = X'YZ' . \quad [3.7]$$

Buradan açıklayıcı matrisler X ve Z, tasarım içi ve tasarım arası matrisleri, büyüme eğrisi modelinde tam ranklıysa $\hat{B}$;

$$\hat{B} = (X'X)^{-1} X'YZ'(ZZ')^{-1} \quad [3.8]$$

olarak elde edilir (6). X ve Z matrislerinin tam ranklı olmadığı varsayılırsa tahmin edici eşitlik [3.7] 'nin çözümü bu durumda aşağıdaki gibi ifade edilir:

$$\hat{B} = (X'X)^{-} X'YZ'(ZZ')^{-} + H - (X'X)^{-} (X'X)H(ZZ')(ZZ')^{-}.$$

Burada A^{-} gösterim biçimi ile, A matrisinin genelleştirilmiş tersi belirtilmekte ve $(m \times r)$ boyutlu H matrisi keyfi olarak seçilmektedir. Başka bir deyişle, bu biçimde olan her hangi bir tahmin bu durumdaki B 'nin genelleştirilmiş en küçük kareler tahmini olmalıdır. Ek olarak, ancak açıklayıcı tasarım matrisleri X ve Z tam ranklıysa tahmin edici eşitlik [3.7], [3.8] tarafından verilen tek çözüme sahiptir (6).

Genelleştirilmiş en küçük kareler tahmini [3.8] için $\hat{B}$, bağımlı matris Y 'nin doğrusal birleşimi olduğundan dolayı $\hat{B}$ normal dağılmaktadır. Bu durumda, $\hat{B}$ 'nin beklenen değeri aşağıdaki gibi bulunabilir (6):

$$\begin{aligned} E(\hat{B}) &= (X'X)^{-1} X' E(Y) Z'(ZZ')^{-1} \\ &= (X'X)^{-1} X' XBZZ'(ZZ')^{-1} \\ &= B \end{aligned} \quad [3.9]$$

Başka bir deyişle genelleştirilmiş en küçük kareler tahmini $\hat{B}$, regresyon katsayısı B 'nin yansız bir tahminidir. Ayrıca kovaryansı aşağıdaki gibi hesaplanabilir (6):

$$\begin{aligned} Cov(\hat{B}) &\equiv Cov(vec(\hat{B})) \\ &= cov([(ZZ')^{-1} Z \otimes (X'X)^{-1} X'] vec(Y)) \\ &= [(ZZ')^{-1} Z \otimes (X'X)^{-1} X'] [Cov(vec(Y))] [Z'(ZZ')^{-1} \otimes X(X'X)^{-1}] \\ &= [(ZZ')^{-1} Z \otimes (X'X)^{-1} X'] [I_n \otimes \Sigma] [Z'(ZZ')^{-1} \otimes X(X'X)^{-1}] \\ &= (ZZ')^{-1} \otimes (X'X)^{-1} X' \Sigma X (X'X)^{-1} \end{aligned} \quad [3.10]$$

(Bkz. Ek-1,2).

Teorem 3.1.2.

Büyüme eğrisi modeli [1.1] 'de regresyon katsayısı B için, [3.8] tarafından verilen genelleştirilmiş en küçük kareler tahmini $\hat{B}$ 'nın dağılımı;

$$\hat{B} \sim N_{m,r}(B; (X'X)^{-1} X' \Sigma X (X'X)^{-1}, (ZZ')^{-1}) \quad [3.11]$$

şeklinde yazılabilir. Özellikle bağımsız kovaryans yapısı $\Sigma = \sigma^2 I_n$ ise, burada $\sigma^2 > 0$ ve bilinmemekte, bu durumda regresyon katsayısı $\hat{B}$ 'nın dağılımı aşağıdaki gibi basitçe ifade edilebilir (6):

$$\hat{B} \sim N_{m,r}(B; \sigma^2 (X'X)^{-1}, (ZZ')^{-1}). \quad [3.12]$$

Teorem 3.1.3.

Bağımsız kovaryans yapısı $\Sigma = \sigma^2 I_n$ 'e sahip olan büyüme eğrisi modeli [1.1] için ($\sigma^2 > 0$ bilinmemekte), yayılım bileşeni σ^2 'nin yansız tahmin edicisi;

$$\hat{\sigma}^2 = \frac{1}{np - mr} \text{tr}[(Y - XBZ)(Y - XBZ)'] \quad [3.13]$$

biçiminde ifade edilir. Burada, m ve r sırasıyla X ve Z açıklayıcı tasarım matrislerinin ranklarını göstermektedir. $\hat{B}$, regresyon katsayısı B 'nin genelleştirilmiş en küçük kareler tahmin edicisidir (6).

4. EN ÇOK OLABİLİRLİK TAHMİNİ

En çok olabilirlik tahmini, istatistiksel sonuç çıkarımında yaygın kullanılan tekniklerdendir. En çok olabilirlik tahmini asimptotik olarak normal dağılımlıdır. Ayrıca, en çok olabilirlik tahmini tutarlılık gibi bazı istatistiksel özelliklere de sahiptir. Olağan tek değişkenli regresyon modeli ya da çok değişkenli regresyon modeli için, genelleştirilmiş en küçük kareler tahminine dayandırılarak çıkartılan sonuçlara karşı, en çok olabilirlik tahmin analizine dayandırılarak çıkartılan sonuçlar tamamen aynıdır. En küçük kareler tahmini $\hat{B}$, bağımlı değişken Y 'nin doğrusal bir fonksiyonuyken en çok olabilirlik tahmini $\hat{B}$ ise bağımlı değişken Y 'nin doğrusal bir fonksiyonu değildir. Büyüme eğrisi modelinde, parametrelerin en çok olabilirlik tahminleri olabilirlik temeline dayanan analizde önemli rol oynar.

$$Y \sim N_{p,n}(XBZ; \Sigma, I_n)$$

(Bkz.Ek-1)

Normal dağılıma sahip olan büyüme eğrisi modeli [1.1] ele alınırsa, B ve Σ parametrelerinin olabilirlik fonksiyonunu $L(B, \Sigma)$, Tanım 3.1.1.'de de verildiği gibi, tekrar gösterilebilir:

$$L(B, \Sigma) = (2\pi)^{-pn/2} [\det(\Sigma)]^{-n/2} \exp\left\{-\frac{1}{2} \text{tr}(\Sigma^{-1}(Y - XBZ)(Y - XBZ)')\right\}. \quad [4.1]$$

Teorem 4.1.

Normal dağılıma sahip olan büyüme eğrisi modeli [1.1] için, X ve Z açıklayıcı tasarım matrisleri tam ranklı ve $n > p + r$ ise;

$$S = Y(I_n - P_{Z'})Y', \text{ pozitif tanımlı matris olmak üzere,}$$

$$\hat{B} = (X'S^{-1}X)^{-1}X'S^{-1}YZ'(ZZ')^{-1} \quad [4.2]$$

ve

$$\hat{\Sigma} = \frac{1}{n}(Y - X\hat{B}Z)(Y - X\hat{B}Z)' \quad [4.3]$$

biçimindedir (6).

Teorem 4.2.

Normal dağılıma sahip olan büyüme eğrisi modeli [1.1] için, $n > p+r$ ise, [4.2] ve [4.3] eşitliklerinde verilen regresyon katsayısı $\hat{B}$ 'nin ve yayılım bileşeni $\hat{\Sigma}$ 'nin olabilirlik fonksiyonu;

$$L(\hat{B}, \hat{\Sigma}) = (2\pi)^{-pn/2} [\det(\hat{\Sigma})]^{-n/2} \exp(-pn/2) \quad [4.4]$$

biçimindedir (6).

4.1. Beklenen Değer ve Varyans-Kovaryans

Regresyon katsayısı B 'nin genelleştirilmiş en küçük kare tahmini için Teorem 3.1.2., B 'nin ortalamasını ve varyans-kovaryans yapısını göstererek normal dağıldığını sağlamıştır. Bu dağılım, genelleştirilmiş en küçük kareler tahmini $\hat{B}$ 'nın, regresyon katsayısı B 'nin yansız tahmin edicisi olduğunu ifade eder. Benzer şekilde, bu bölümde regresyon katsayısı B ve yayılım bileşeni Σ 'nin en çok olabilirlik tahminleri için dağılımlar ve momentler ile ilgilenilmektedir. En çok olabilirlik tahminlerinin dağılımları yaygın olarak kullanılmak için oldukça karışık olduklarından en çok olabilirlik tahminlerinin beklenen değerleri ve varyans-kovaryansları ile ilgilenilir.

İlk olarak, büyüme eğrisi modelindeki regresyon katsayısı B için en çok olabilirlik tahmini $\hat{B}$:

$$\hat{B} = (X'S^{-1}X)^{-1} X'S^{-1}YZ'(ZZ')^{-1}$$

biçimindedir. Burada, $S = Y(I_n - P_{Z'})Y'$ pozitif tanımlı matristir. $\hat{B}$, bağımlı değişken Y'nin doğrusal olmayan fonksiyonu olduğundan dolayı momentlerinin hesaplanması daha karışık hale gelir. En çok olabilirlik tahminlerinin beklenen değeri ve varyans-kovaryansı, rastgele matris S'de verilen koşullu değişkenlerin yardımıyla hesaplanır. Gerçekte, $\hat{B}$ 'nin beklenen değeri;

$$\begin{aligned} E(\hat{B}) &= E_S(E_Y(\hat{B}|S)) \\ &= E_S[(X'S^{-1}X)^{-1} X'S^{-1}E_Y(Y|S)Z'(ZZ')^{-1}] \\ &= E_S[(X'S^{-1}X)^{-1} (X'S^{-1}X)B(ZZ')(ZZ')^{-1}] \\ &= E(B) \\ &= B \end{aligned} \quad [4.5]$$

biçiminde bulunur. Buradan da görüldüğü üzere en çok olabilirlik tahmini $\hat{B}$, regresyon katsayısı B'nin yansız bir tahminidir. Aynı şekilde, en çok olabilirlik tahmini $\hat{B}$ 'nin varyans-kovaryansı ise aşağıdaki gibidir (6):

$$\begin{aligned} \text{Cov}(\hat{B}) &= E_S[\text{Cov}_Y(\hat{B}|S)] + \text{Cov}_S[E_Y(\hat{B}|S)], (\text{Cov}_S[E_Y(\hat{B}|S)] = 0) \\ &= E_S[\text{Cov}_Y(\hat{B}|S)] \\ &= \frac{n-r-1}{n-r-(p-m)-1} (ZZ')^{-1} \otimes (X'\Sigma^{-1}X)^{-1} \end{aligned} \quad [4.6]$$

Teorem 4.3.

Normal dağılıma sahip olan büyüme eğrisi modeli [1.1] için, $n > p+r$ ise, [4.2] tarafından verilen en çok olabilirlik tahmini $\hat{B}$, regresyon katsayısı B'nin yansız tahmin edicisidir. Ayrıca, en çok olabilirlik tahmini $\hat{B}$ 'nin varyans-kovaryansı;

$$Cov(\hat{B}) = \frac{n-r-1}{n-r-(p-m)-1} (ZZ')^{-1} \otimes (X'\Sigma^{-1}X)^{-1}$$

biçiminde ifade edilebilir. Eğer $\Sigma = \sigma^2 I_n$ ise, en çok olabilirlik tahmininin varyans kovaryans matrisi;

$$Cov(\hat{B}_{MLE}) = \frac{n-r-1}{n-r-(p-m)-1} \sigma^2 (ZZ')^{-1} \otimes (X'X)^{-1} \quad [4.7]$$

biçimindedir. Genelleştirilmiş en küçük kareler tahmini $\hat{B}$ ile en çok olabilirlik tahmini $\hat{B}$ 'yı karşılaştırılırsa;

$$\begin{aligned} Cov(\hat{B}_{GLSE}) &= \sigma^2 (ZZ')^{-1} \otimes (X'X)^{-1} \\ &\leq \frac{n-r-1}{n-r-(p-m)-1} \sigma^2 (ZZ')^{-1} \otimes (X'X)^{-1} \\ &= Cov(\hat{B}_{MLE}) \end{aligned}$$

olduğu görülür. Buradan, bağımsız kovaryans yapısına sahip olan büyüme eğrisi modeli için, genelleştirilmiş en küçük kareler tahmini $\hat{B}$ 'nın en çok olabilirlik tahmini $\hat{B}$ 'ya göre daha iyi bir tahmin edici olduğu söylenebilir (6).

4.2. Rao 'nun Basit Kovaryans Yapısı (BKY)

4.2.1. En çok olabilirlik tahmininin genelleştirilmiş en küçük kareler tahminine özdeşlik koşulu

Bu alt bölümde, daha önce büyüme eğrisi modelinde bahsedilen Rao 'nun basit kovaryans yapısını açıklanmaktadır. Regresyon katsayısı B ve yayılış bileşeni Σ fonksiyonlarının her ikisi için en çok olabilirlik tahminleri olan $\hat{B}_{(S)}$ ve $\hat{\Sigma}_{(S)}$, $S = Y(I_n - P_Z)Y'$ matrisinin fonksiyonlarıdır. $n > p+r$ olmak üzere S matrisi $(p \times p)$

boyutlu pozitif tanımlı matristir. Takiben, Teorem 4.1. 'de verilen en çok olabilirlik tahminleri tekrar anımsanırsa;

$$\hat{B}_{(S)} = (X'S^{-1}X)^{-1}X'S^{-1}YZ'(ZZ')^{-1}$$

$$\hat{\Sigma}_{(S)} = \frac{1}{n}(Y - X\hat{B}_{(S)}Z)(Y - X\hat{B}_{(S)}Z)' \quad [4.8]$$

biçiminde olduğu görülür. S 'ye bağımlı olan en çok olabilirlik tahminleri bağımsız değildir. Diğer taraftan, B' nin genelleştirilmiş en küçük kareler tahmini ve Σ 'nın tahmini ise;

$$\hat{B} = (X'X)^{-1}X'YZ'(ZZ')^{-1}$$

$$\hat{\Sigma} = \frac{1}{n}(Y - X\hat{B}Z)(Y - X\hat{B}Z)' \quad [4.9]$$

biçimindedir. Önceden de belirtildiği gibi, [4.9] 'daki genelleştirilmiş en küçük kareler tahminleri ile [4.8] 'deki en çok olabilirlik tahminleri arasındaki genel fark; genelleştirilmiş en küçük kareler tahmini $\hat{B}$, bağımlı fonksiyon Y 'nin doğrusal fonksiyonuyken en çok olabilirlik tahmini $\hat{B}$ 'nin bağımlı fonksiyon Y 'nin doğrusal fonksiyonu değildir. Aslında bu iki tür tahmin edici arasındaki tek fark, S matrisinin en çok olabilirlik tahminine dahil edilmesidir. Burada S matrisi, ağırlık matrisine benzer bir davranış sergilemektedir. Böylece (pxp) boyutlu Σ matrisi bilinmediğinde, parametre tahmini üzerinde açıklayıcı tasarım matrisi X'in etkinliği azalır. Buna karşın [4.9] eşitliğindeki genelleştirilmiş en küçük kareler tahmini, $S = I_p$ olduğu sürece, [4.8] 'in özel bir biçimi gibi görülebilir. Böylece genelleştirilmiş en küçük kareler tahmini [4.9], $\hat{B} \equiv \hat{B}_{(I_p)}$ ve $\hat{\Sigma} \equiv \hat{\Sigma}_{(I_p)}$ şeklinde yazılabilir ve burada $\hat{B}_{(I_p)}$ ve $\hat{\Sigma}_{(I_p)}$ en çok olabilirlik tahminlerini göstermek üzere özdeş matris I_p üzerinde değerlendirilebilir.

Eğer en çok olabilirlik tahminleri gerçekten genelleştirilmiş en küçük kareler tahminlerine benzerse, en çok olabilirlik tahminlerine dayandırılarak çıkartılan istatistiksel sonuç çıkarsaması daha kolay ifade edilebilir. Örneğin, Rao 'nun tekdüze kovaryans yapısı $\hat{\Sigma} = \sigma^2 \{(1 - \rho)I_p + \rho 1_p 1_p'\}$ 'da ilk olarak I_p ele alınır;

$$X(X'X)^{-1}X' + Q(Q'Q)^{-1}Q' = I_p \quad [4.10]$$

şeklinde düşünülebilir. [4.10] 'deki I_p terimini $\hat{\Sigma} = \sigma^2 \{(1 - \rho)I_p + \rho 1_p 1_p'\}$ 'da yerine yazılırsa;

$$\Sigma = \sigma^2 [(1 - \rho)X(X'X)^{-1}X' + (1 - \rho)Q(Q'Q)^{-1}Q' + \rho 1_p 1_p'] \quad [4.11]$$

eşitliği elde edilebilir. Eğer p değişkenli vektör 1_p , açıklayıcı matris X ya da onun ortogonal tamamlayıcı uzayı tarafından kapsanmışsa, [4.11] 'e dayandırılan bu aşıkâr tekdüze kovaryans yapısı, Rao 'nun basit kovaryans yapısı [1.7] 'de de içerilir. Örneğin, $1_p \in \mu(X)$ ise a gibi m değişkenli vektörler mevcuttur ve $1_p = Xa$ şeklinde yazılır. Bu durumda [4.11] eşitliği, Rao 'nun basit kovaryans yapısı [1.7] 'nin özel bir durumu olarak aşağıdaki gibi tekrar yazılabilir:

$$\Sigma = \sigma^2 [X\{(1 - \rho)(X'X)^{-1} + \rho aa'\}X' + (1 - \rho)Q(Q'Q)^{-1}Q']$$

Benzer anlamda, $1_p \in \mu(Q)$ ise o zaman [4.11], [1.7] 'nin özel bir durumu olur. Diğer taraftan;

$$1_p \in \mu(X) \text{ ya da } 1_p \in \mu(Q) \text{ ise } X'1_p 1_p'Q = 0$$

biçimindedir.

Teorem 4.4.

Büyüme eğrisi modeli [1.1] için, $\hat{\Sigma} = \sigma^2 \{(1 - \rho)I_p + \rho 1_p 1_p'\}$ 'da verilen tekdüze kovaryans yapısının [1.7] 'de verilen Rao 'nun basit kovaryans yapısında özel bir durum gibi içerilmesi için ancak ve ancak aşağıdaki [4.12] eşitliğinin sağlanması gerekmektedir:

$$X' 1_p 1_p' Q = 0 \quad . \quad [4.12]$$

$1_p \in \mu(X)$ ya da $1_p \in \mu(Q)$ ve $Q \in Q$ 'dur. 1_p ise, tüm bileşenleri bir olan p değişkenli vektörü göstermektedir (6).

Genellikle [4.12] şartı doğru olmayabilir. Eğer şartlar gerçekten sağlanıyorsa eksik olan bu model için tekdüze kovaryans yapısına bakmak yeterlidir. Ek olarak, modelin çözümlemeli sonucuna göre olası tekdüze kovaryans yapısı ile bağımlı fonksiyon Y 'nin kestirimi Lee tarafından gösterilmektedir (5). X açıklayıcı matris olduğu sürece, $X = (1_p : X_1)$ biçiminin [4.12] şartı daima sağlanır. Burada X_1 , $p \times (m-1)$ boyutlu matrisi göstermektedir. Bu şekilde, eğer büyüme eğrisi modelindeki artık terim dikkate alınmışsa, özel durumda olduğu gibi Rao 'nun basit kovaryans yapısı her zaman tekdüze kovaryans yapısını içerir. Bu durumda, Rao 'nun basit kovaryans yapısı kullanılarak türetilen istatistiksel sonuç, tekdüze kovaryans yapısında olan modele uygulanabilir.

Rao 'nun basit kovaryans yapısındaki özel durumlardan bir diğeri ise aşağıdaki gibi verilebilir:

$$\Sigma = X\Gamma X' + \sigma^2 I_p. \quad [4.13]$$

Bu yapı, 'Rastgele Regresyon Katsayısı Yapısı' olarak isimlendirilir. Burada, Γ $(m \times m)$ boyutlu bilinmeyen yayılım matrisidir ve $\sigma^2 > 0$ bilinmemektedir. Rastgele

regresyon katsayısı yapısı [4.13], rastgele etkili olan büyüme eğrisi modeline uygulanırsa model;

$$Y_{pxm} = X_{pxm} B_{mxr} Z_{rxn} + X_{pxm} A_{mxn} + E_{pxn} \quad [4.14]$$

biçimini alır. Burada A, $A \sim N_{m,n}(0; \Gamma, I_n)$ biçiminde normal dağılımlı rastgele etki matrisi ve E, $E \sim N_{p,n}(0; \sigma^2 I_p, I_n)$ dağılımına sahip olan hata matrisini göstermektedir. Ayrıca hata matrisi E, A matrisinden bağımsızdır. [4.13] eşitliği, rastgele etkili olan doğrusal modeller ya da doğrusal karışık modellerin çözümlemelerinde çok faydalıdır. Bu tür modeller, verinin farklı kaynaklardan meydana gelen değişkenliklerini açıklamada kullanılır. Küçük alan tahmini, ilişkili veri ya da panel verinin çözümlemesinde istatistiksel araçlar ortak olarak kullanılır. Rao'nun basit kovaryans yapısı [1.7] 'nin özel bir hali olmasından dolayı [4.13] eşitliğindeki rastgele regresyon katsayısı yapısının daha anlaşılır olması için [4.13] eşitliğinde özdeş (identity) matris olan I_p 'nin yerine basitçe [4.10] eşitliği kullanıldığında bu eşitlik;

$$\Sigma = X[\Gamma + \sigma^2(X'X)^{-1}]X' + \sigma^2Q(Q'Q)^{-1}Q' \quad [4.15]$$

biçimini alır. Yani, [4.15] eşitliği [1.7] 'deki Rao'nun basit kovaryans yapısının özel bir biçimidir (6). Bağımlı matris Y'nin beklentisinin XBZ biçimini açıkladığı durumda varyans-kovaryans matrisi, $Cov(Y) = I_n \otimes \Sigma$, aşağıdaki gibi hesaplanabilir:

$$\begin{aligned} Cov(Y) &= Cov[vec(Y)] = Cov[vec(XA)] + Cov[vec(E)] \\ &= (I_n \otimes X)Cov[vec(A)](I_n \otimes X)' + \sigma^2(I_n \otimes I_p) \\ &= (I_n \otimes X)(I_n \otimes \Gamma)(I_n \otimes X)' + (I_n \otimes \sigma^2 I_p) \\ &= (I_n \otimes X\Gamma X') + (I_n \otimes \sigma^2 I_p) \\ &= [I_n \otimes X\Gamma X' + \sigma^2 I_p] \end{aligned}$$

(Bkz. Ek-1,2)

Bu da gösteriyor ki, $\Sigma = XTX' + \sigma^2 I_p$ 'dir. Açıkça, çok değişkenli model için rastgele etkili büyüme eğrisi modeli [4.14], olağan karma modelin bir uzantısıdır (7). Genellikle bu rastgele etkiler modeli için, sabit etki matrisi B 'nin yanı sıra varyans bileşeni Γ 'nın tahminiyle de ilgilenilir. Bu durum, Rao 'nun basit kovaryans yapısına dayanan sonuç çıkarmayı da içerebilir.

4.2.2. Yayılış bileşenlerinin tahminleri

Rao 'nun basit kovaryans yapısı [1.7] 'ye sahip olan büyüme eğrisi modeli için, regresyon katsayısı B 'nin en çok olabilirlik tahmini genelleştirilmiş en küçük kareler tahminine denktir ve bu tahmin, $\hat{B} = (X'X)^{-1} X'YZ'(ZZ')^{-1}$ biçimindedir. Yayılış bileşenleri (dispersion components) Γ ve Θ 'nın en çok olabilirlik tahminleri;

$$\hat{\Gamma} = (X'X)^{-1} X' \hat{\Sigma} X (X'X)^{-1} \quad \text{ve} \quad \hat{\Theta} = (Q'Q)^{-1} Q' \hat{\Sigma} Q (Q'Q)^{-1}$$

biçiminde olmalıdır (6).

$$\hat{\Sigma} = \frac{1}{n} (S + P_Q Y P_Z' Y' P_Q) \quad [4.16]$$

Burada $\hat{\Sigma}$, yayılış bileşeni Σ 'nın en çok olabilirlik tahminidir ve $P_Q = Q(Q'Q)^{-1}Q'$ ile $P_Z = Z'(ZZ')^{-1}Z$ biçiminde gösterilir. [4.16] eşitliği Rao 'nun basit kovaryans yapısı varsayımı altında yazılabilir. Dolaylı olarak Γ ve Θ 'nın en çok olabilirlik tahminleri;

$$\hat{\Gamma} = \frac{1}{n} (X'X)^{-1} X' S X (X'X)^{-1} \quad \text{ve} \quad \hat{\Theta} = \frac{1}{n} (Q'Q)^{-1} Q' Y Y' Q (Q'Q)^{-1}$$

biçiminde olmaktadır. Bu sonuçlar, Rao 'nun basit kovaryans yapısıyla parametrelerin olabilirlik fonksiyonundan doğrudan doğruya türetilebilir (6).

Teorem 4.5.

Tekdüze kovaryans yapısına sahip olan büyüme eğrisi modeli için, eğer $n > p+r$, $-1/(p-1) < \rho < 1$ ve $X = (1_p : X_1)$ ise, regresyon katsayısı B 'nin en çok olabilirlik tahmini ve $\hat{\Sigma} = \sigma^2 \{(1-\rho)I_p + \rho 1_p 1_p'\}$ 'deki yayılış bileşenleri σ^2 ve ρ 'nun en çok olabilirlik tahminleri aşağıdaki gibidir:

$$\hat{B} = (X'X)^{-1} X'YZ'(ZZ')^{-1}$$

$$\hat{\sigma}^2 = \frac{1}{pn} iz(S^*) \quad [4.17]$$

$$\hat{\rho} = \frac{1}{p-1} \frac{1_p' S^* 1_p - iz(S^*)}{iz(S^*)}$$

Burada, $S^* = S + P_Q Y P_Z Y' P_Q$ ve Q , X 'in ortogonal uzay matrisi Q 'nun elemanıdır (6).

Teorem 4.6.

Rastgele regresyon katsayısı yapısına sahip olan büyüme eğrisi modeli için, eğer $n > p+r$ ise, regresyon katsayısı B 'nin en çok olabilirlik tahmini ve [4.13] 'daki yayılış bileşenleri σ^2 ve ρ 'nun en çok olabilirlik tahminleri aşağıdaki gibidir:

$$\hat{B} = (X'X)^{-1} X'YZ'(ZZ')^{-1}$$

$$\hat{\Gamma} = \frac{1}{n} (X'X)^{-1} X' S X (X'X)^{-1} - \hat{\sigma}^2 (X'X)^{-1} \quad [4.18]$$

$$\hat{\sigma}^2 = \frac{1}{(p-m)n} iz[Y'(I_p - P_X)Y]$$

Burada P_X , X tarafından biçimlendirilen izdüşüm matrisidir. Yani, $P_X = X(X'X)^{-1}X'$ biçimindedir (6).

Rastgele etki matrisi A , $A = (a_1, a_2, \dots, a_n)$ şeklinde gösterilir. Sıralanmış ve seçilmiş bireylerin söz konusu olduğu durumda genellikle A matrisiyle ilgilenilir. Rastgele etkiler vektörünün kestirimleri olan $\hat{a}_i$ 'ler gelecek gözlemlerin kestirimlerinde de önemlidir. Rastgele etkili büyüme modeli [4.14] için, rastgele etkilerin $A \sim N_{n,m}(0; \Gamma, I_n)$ ve hata terimi $\varepsilon \sim N_{p,n}(0; \sigma^2 I_p, I_n)$ biçiminde dağıldığı ve her ikisinin de karşılıklı bağımsız olduğu varsayılır. Burada, Γ ve σ^2 'nin bilindiği kabul edilir. Eksiksiz veri ya da eklenmiş (augmented) verinin ortak dağılım yoğunluk fonksiyonu bağımlı değişken Y ve rastgele etkiler matrisi A olmak üzere aşağıdaki gibidir:

$$\begin{aligned} f(Y, A) &= f(Y \setminus A) f(A) \\ &= (2\pi)^{-pn/2} (\sigma^2)^{-pn/2} \exp\left[-\frac{1}{2\sigma^2} \text{tr}\{(Y - XBZ - XA)(Y - XBZ - XA)'\}\right] \\ &\quad \cdot (2\pi)^{-mn/2} [\det(\Gamma)]^{-n/2} \exp\left[-\frac{1}{2} \text{tr}(\Gamma^{-1} AA')\right]. \end{aligned}$$

A ve B ile ilgili olan bu ortak yoğunluk fonksiyonun maksimize edilmesi ile rastgele etkiler matrisi A 'nın kestirimi, $\hat{A} = \Gamma X' \Sigma^{-1} (Y - X\hat{B}Z)$ olarak bulunur. Burada $\hat{B} = (X'X)^{-1} X'YZ'(ZZ')^{-1}$, B 'nin en çok olabilirlik tahminidir ve $\Sigma = \sigma^2 I_p + X\Gamma X'$ biçimindedir. Genellikle yayılış bileşenleri Γ ve σ^2 bilinmemektedir. Bu durumda, [4.18]'de verilen en çok olabilirlik tahminleri $\hat{\Gamma}$ ve $\hat{\sigma}^2$, Γ ve σ^2 'nin yerine geçici olarak kullanılabilir. Diğer bir deyişle $\hat{A}$ 'nın;

$$\hat{A} = \hat{\Gamma} X' \hat{\Sigma}^{-1} (Y - X\hat{B}Z) \quad [4.19]$$

biçiminde olduğu söylenebilir. Burada $\hat{\Sigma} = \hat{\sigma}^2 I_p + X\hat{\Gamma}X'$, Σ 'nın en çok olabilirlik tahminidir ve $\hat{A}$, rastgele etkiler matrisi A 'nın kestirimi olarak kullanılabilir.

Özet olarak Rao' nun basit kovaryans yapısı, büyüme eğrisi modeli bağlamında oldukça önemlidir. Çünkü, doğrusal olmayan analizi doğrusal olan analize indirger. Diğer yandan, tekdüze kovaryans yapısı ve rastgele regresyon katsayı yapısı gibi bazı faydalı kovaryans yapıları özel durum olarak Rao'nun basit kovaryans yapısında içerilir. İstatistiksel çıkarımları daha etkin yapan cimri (parsimon) karakterler yüzünden bu özel kovaryans yapıları genellikle büyüme eğrileri modeli analizinde kullanılmaktadır.



5. AYKIRI (OUTLIERS) VE ETKİLİ (INFLUENTIAL) GÖZLEMLER

5.1. Özet

İstatistik biliminde önemli ve gerekli tekniklerden biri istatistiksel tanımlamalardır. Bu tanımlamaların amacı, modelden sapan / aykırı (outlier) gözlemleri belirlemek, etkili (influential) gözlemleri tanımlamak ve seçilen istatistiksel modelin geçerliliğini kabul etmektir. İstatistiksel model; verilerin analizinde, istatistiksel anlamlılıkta ve gelecekle ilgili tahmin yapmakta önemli rol oynar.

İstatistiksel bir çalışmada, örnekte yer alan gözlemlerden bir ya da birkaçı araştırmacının beklentileri üzerinde şüphe doğurabilir. Araştırmacının beklentisi ya da başlangıçtaki varsayımı, gözlemlerin tamamının aynı yapıya sahip olmasıdır. Aynı yapıya ait olmadığı düşünülen, diğerleri ile tutarsız olan, sıra dışı gözlem ya da gözlemler verinin kaynağı hakkındaki bilgiyi azaltmakta ya da çarpıtmaktadır.

‘Aynı yapıya ait olmadığı’ ya da ‘diğerlerinden farklı olduğu’ düşünülen gözlem şeklindeki genel tanımlamaların dışında, çok daha özel bir tanımlamayı aykırı gözlem için yapabilmek her probleme ve her araştırmacıya göre değişmektedir. Tüm gözlemlerin aynı yapının elemanı olması tanımı, istatistikte gözlemlerin her birinin tesadüfi değişken için öngörülen model tarafından üretilmiş olmasına ya da modele uygunluk göstermesine karşılık gelir. Öngörülen model tesadüfi değişken için yanlış bir seçim değilse tesadüfi karakteristiğe sahip bir gözleme ‘aykırıdır’ diyebilmek için kullanılan yöntemler, gözlemin öngörülen model için çok uç bir gözlem olduğu ya da öngörülen model dışında bir modele uyduğu görüşüne dayanmaktadır.

Cevap aranması gereken ilk soru diğerlerinden farklı olduğu ya da onlarla aynı yapıya ait olmadığı düşünülen gözlemin karakteristiğinin ne olduğudur. Yani aykırı olduğu düşünülen gözlemin karakteristiği tesadüfi mi yoksa deterministik (her olayın başka olayların kaçınılmaz sonucu olarak ortaya çıktığı düşüncesi) midir? Aykırı gözlemin karakteristiğinin deterministik olup olmadığını incelemek çoğu zaman mümkün olmayabilir. Mümkün olduğu durumlarda gözlemi düzeltmek, çıkarmak ya

da deneyi tekrarlamak yoluyla problem giderilebilir. Mümkin olmadığı durumlarda ise, gerçekte deterministik bir karakteristiğin elemanı olan aykırı gözlemi, karakteristiği tesadüfiymiş gibi değerlendirilmek zorunda kalınabilir. Aykırı gözlemin karakteristiğinin tesadüfi olduğunun düşünülmesi problemin istatistiksel açıdan değerlendirilmesini gerektirmektedir.

İyi bir istatistiksel model nasıl olmalıdır? Seçilen modelden istatistiksel sonuç çıkarmak veri kümesi ile ilişkili midir? Orijinal veride doğal ya da dışarıdan kaynaklı bir hata varsa modelin değiştirilmesi gerekir mi? Veri kümesindeki bazı gözlemler veri kümesinden çıkartıldığında sonuç değişir mi? Eğer cevap evet ise, çıkartılacak bu gözlem ya da gözlemlere nasıl karar verilir? Yani veri kümesindeki her bir gözlemin modele etkisi nasıldır? Bu ve benzeri problemlerin giderilmesi için çeşitli istatistiksel yaklaşımlar geliştirilmiştir (8,9).

İstatistiksel yaklaşımlardan birisi, modelden sapan / aykırı gözlemlere ve modelden istatistiksel sonuç çıkarmada etkisi büyük olan etkili gözlemlere karar vermektir. Bu yaklaşımlar literatürde ‘kuşkulu istatistikler’ olarak tanımlanmaktadır.

İstatistiksel yaklaşımlardan diğeri de, tesadüfi örneğin herhangi bir gerçekleşişi üzerinden yığın hakkında çıkarsama yapmak ve bu çıkarsamayı yaparken aykırı gözlemin varlığından çok fazla etkilenmeyecek istatistiksel yöntemler geliştirmektir. Bu yaklaşımda literatürde ‘sağlam (robust) istatistik’ olarak tanımlanmaktadır. Bazı istatistikçiler sağlam istatistikte aykırı gözlemlerin göz ardı edildiğini düşünürler ama bu doğru bir yaklaşım değildir. Aslında sağlam regresyon ve regresyon analizinde aynı amaca ulaşılmaya çalışılır ama sıralamaları farklıdır. Regresyon analizinde önce aykırı gözlemler belirlenir ve veri kümesinden çıkartılır, daha sonrada en küçük kareler yöntemiyle uygun model belirlenir. Sağlam analizde ise veriye uygun regresyon modeli belirlenir, daha sonrada sağlam çözüm ile en büyük artığa sahip olan noktalar yani aykırı gözlemler belirlenir.

‘Tesadüfî değişken için öngörülen dağılım gerçeği yansıtmakta mıdır?’ sorusu açısından aykırı gözlem problemine bakıldığında, sağlam istatistiksel yöntemlerin,

uyuşmazlık testi yöntemlerine göre daha uygun tahminler vereceği düşünülmektedir. Tesadüfî deęişken için, öngörülen model üzerinden bir gözleme ‘aykırıdır’ denebilir. Ancak gerçekte tesadüfî deęişkenin sahip olduęu model bizim öngördüğümüz modelden farklı olabilir. Örneğin, normal dağılım öngörölmüşken tesadüfî deęişken gerçekte daha **yüksek** kuyrukları olan bir dağılıma sahip olabilir (student t dağılımı gibi). Bu durumda aykırı gözlemlerden korunmak için uç gözlemleri yok saymak, gözlemleri örnekten çıkartmak ya da gözlemlerin örnekte hiç yokmuş gibi olduğunu düşünerek dağılımın parametrelerini tahmin etmek çok fazla bilgi kaybına sebep olacak bir tercih olabilir. Oysaki uç gözlemlerin etkisini azaltacak sağlam yöntemlerin kullanılması sonucu elde edilen tahminler, gözlemlerin çıkartılması sonucu elde edilen tahminlerden daha çok gerçeęi yansıtacaktır.

Amacımız kestirim yapmaksa, regresyon modelini destekleyen bir gözlemi ‘aykırıdır’ diye yok saymak (sadece uç gözlem olup olmaması kriteri ile hareket edilirse) ya da birtakım sağlam yöntemlerle bu gözlemin sağladığı bilgiyi düşürmek istenmez. Böyle bir durumda regresyon doğrusunun parametre tahminleri üzerinde büyük deęişmeye sebep olacak uç ve etkili (influential) bir gözlemi aykırı bir gözlemdir diye niteleyecek bir yöntem önermek daha mantıklı olacaktır.

Aykırı gözlemlerin varlığı, model yapısının yanlış olmasından, gerekli bazı dönüşümlerin yapılmamış olmasından, ölçüm, tartım, kaydetme yanlışlarından ya da örneklem rastgeleliğinden kaynaklanabilir. Örneklemdeki bazı gözlemlerin, öteki gözlemlerden farklı etkenlerin etkisinde kalmış olmaları da aykırı gözlemleri ortaya çıkarabilir.

Aykırı ve etkili gözlemleri incelenmeden önce bir veri seti üzerinde istatistiksel tanımlamanın önemi aşağıda gösterilmektedir. Bu veri seti Anscombe (1973) tarafından istatistikte tanımlamaların önemini göstermek için hazırlanmıştır.

Örnek 5.1.

Çizelge 5.1: Anscombe'nin veri seti

No	x_1	y_1	y_2	y_3	x_2	y_4
1	10 ,00	8 ,04	9 ,14	7 ,46	8 ,00	6 ,58
2	8 ,00	6 ,95	8 ,14	6 ,77	8 ,00	5 ,76
3	13 ,00	7 ,58	8 ,74	12 ,70	8 ,00	7 ,71
4	9 ,00	8 ,81	8 ,77	7 ,11	8 ,00	8 ,84
5	11 ,00	8 ,33	9 ,26	7 ,81	8 ,00	8 ,47
6	14 ,00	9 ,96	8 ,10	8 ,84	8 ,00	7 ,04
7	6 ,00	7 ,24	6 ,13	6 ,08	8 ,00	5 ,25
8	4 ,00	4 ,26	3 ,10	5 ,39	19 ,00	12 ,05
9	12 ,00	10 ,08	9 ,13	8 ,15	8 ,00	5 ,56
10	7 ,00	4 ,82	7 ,26	6 ,42	8 ,00	7 ,91
11	5 ,00	5 ,68	4 ,74	5 ,73	8 ,00	6 ,89

Veri seti dört farklı grubu içermektedir ve bu gruplar aşağıdaki gibidir:

$$D(x_1, y_1) \quad D(x_1, y_2) \quad D(x_1, y_3) \quad \text{ve} \quad D(x_2, y_4) .$$

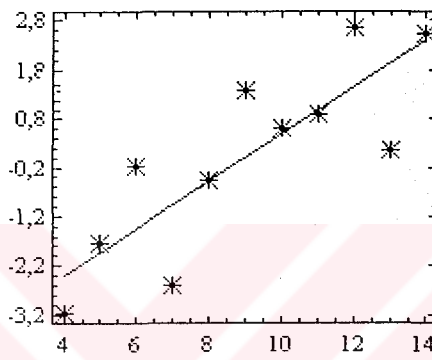
Bu dört veri setinin her birine uyan basit doğrusal regresyon modeli ise aşağıdaki gibi verilir:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i , \quad (i=1, 2, \dots, 11) .$$

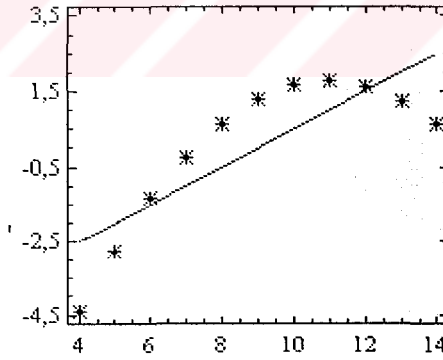
Dört veri seti için de β_0 ve β_1 'in en küçük kareler tahmin gösterimleri bütünüyle aynı olup en küçük kareler tahminleri aşağıdaki gibi elde edilir:

$$\hat{\beta}_0 = 3,0 \quad \hat{\beta}_1 = 0,5 \quad \hat{\sigma}^2 = 1,531 \quad \text{ve} \quad \hat{R} = 0,667 .$$

Burada $\hat{\sigma}^2$ ve $\hat{R}$, sırasıyla veri kümesinin varyansının ve korelasyon katsayısının tahminleridir. Verinin biçimine bakılmaksızın her bir veri setine uyan $\hat{y} = 3,0 + 0,5x$ regresyon doğrusu kullanılır. Veri setlerine ilişkin grafikler çizildiğinde, $D(x_1, y_1)$ veri seti için doğrusal regresyon modelinin uygun olduğu, $D(x_1, y_2)$ veri seti için karesel bir eğri elde edildiğinden, $\hat{y}_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$ regresyon modelinin bu veri seti için daha iyi bir model olduğu söylenebilir (Şekil 5.1.(b)).

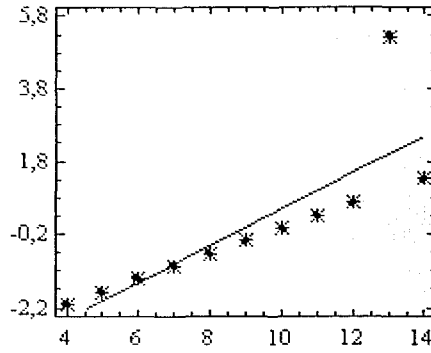


Şekil 5.1.(a)



Şekil 5.1.(b)

Şekil 5.1.(a,b) Anscombe'nin veri seti $D(x_1, y_1)$ ve $D(x_1, y_2)$ için serpmeye diyagramı ve regresyon doğruları.



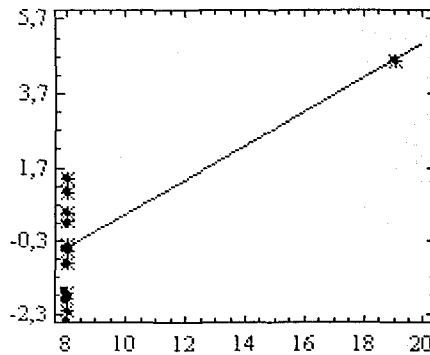
Şekil 5.1.(c)

Şekil 5.1.(c) Anscombe'nin veri seti $D(x_1, y_3)$ için serpmeye diyagramı ve regresyon doğrusu.

Üçüncü veri seti için, $(13,00 ; 12,70)$ gözlem değeri diğer gözlemlerden farklıdır ve regresyon doğrusunu kendisine doğru çekerek istatistiksel sonuçları büyük ölçüde etkilemektedir (Şekil 5.1.(c)). Eğer bu gözlem veri setinden çıkartılır ve tekrar hesaplama yapılırsa elde edilen tahmin değerleri aşağıdaki gibi bulunabilir:

$$\hat{\beta}_0 = 4,0 \quad \hat{\beta}_1 = 0,346 \quad \hat{\sigma}^2 = 9,497 * 10^{-6} \quad \text{ve} \quad \hat{R} = 1.$$

Dolayısıyla, $\hat{y} = 4,0 + 0,346x$ regresyon doğrusu bu veri seti için daha uygun bir model olacaktır.



Şekil 5.1.(d)

Şekil 5.1.(d) Anscombe'nin veri seti $D(x_2, y_4)$ için serpmeye diyagramı ve regresyon doğrusu.

Dördüncü veri setine bakıldığında, 11 gözlemden 10'unun $x = 8$ gözlemini içerdiği dolayısıyla Şekil 5.1.(d) 'nin bu veri seti için uygun bir regresyon doğrusu olmadığı görülmektedir. Çünkü örnekten elde edilen bilgi yeterli değildir.

5.2. Aykırı Gözlemler (Outliers)

Literatürde, aykırı gözlemler tanımlanırken genellikle iki kategoride sınıflandırılır. Birinci yaklaşımda, veri setindeki diğer gözlemlerle tutarsız olan gözlem değeri aykırı gözlem olarak kabul edilebilir (örneğin, yukarıda $D(x_1, y_3)$ veri setindeki (13,00 ; 12,70) gözlemi gibi). Bu tutarsız gözlemler analizde önemli rol oynar örneğin, tek değişkenli veri seti için tutarsız olan bir gözlem uç gözlemdir. Yani, öngörülen modelden hareketle aykırı gözlemi belirlemeye çalışmak, uç (extrem) gözlemleri dikkate almaktır. Gerçekte, her uç gözlemin aykırı olduğunu söylemek doğru olmayabilir. Tesadüfî değişken için bir kritik değer belirlenmeli ve bu kritik değere göre uç gözlemin aykırı olup olmadığı analiz edilmelidir. Yani tüm gözlemler öngörülen model tarafından üretilmiştir, ancak uç gözlemlerden bazılarının belirli bir ölçü üzerinden diğerlerine göre daha uç olduğu belirlenebilir. Belirli bir sınırdan daha uça yer alan bu gözlemler aykırı gözlem olarak adlandırılır. Gerçekten, tek değişkenli örneklerde genellikle uç (ya da göreceli uç) değerler aykırı gözlemlerdir (10).

Aykırı gözlemler için diğer bir yaklaşımda veri kümesindeki bir gözlemin öngörülen model dışında bir başka modele daha uygun olduğu durumdur (11). Bu duruma uyan gözlem aykırı gözlem olarak kabul edilir ve veri kümesindeki bu gözlem kirletici gözlem olarak tanımlanır. Yani tüm gözlemler için öngörülen bir model olmasına rağmen, eğer herhangi bir gözlem kendine daha uygun özelliklerdeki başka bir modele uygunluk gösteriyorsa o gözlem kirletici olarak adlandırılır. Kirletici gözlem uç bir değer olabileceği gibi merkezde yer alan bir gözlemde olabilir. Bu yüzden her kirletici gözlem aykırıdır denilemez. Dolayısıyla aykırı gözlem için alternatif bir modelin tanımladığı yöntemlerin bu ayrımı ortaya koyması gerekmektedir. Yani, bir gözleme aykırıdır diyebilmek için gözlemin hem uç, hem de kirletici olması gerekir. Aşağıdaki gözlem değerlerinin rastgele olarak alınan gözlemler olduğunu varsayalım:

1,74 1,46 -1,28 -0,02 -0,40 0,02 3,89 -0,10 1,71 .

Normal dağılıma sahip olan yığının $N(\Theta ; 1,5)$ dağılımından geldiğini ve örnek ortalamasının $\bar{x} = 0,837$ olarak verildiğini kabul edelim. Çok küçük bir olasılıkla 3,89 değeri dağılımı $N(0,837 ; 1,5)$ olan normal dağılımdan gelmiştir. Diğer bir deyişle, 3,89 değeri araştırmacıyı varsayılan dağılımdan şüpheye götürmektedir. Gerçekte bu veri kümesi, $f(x) = (1/\pi) * (1 + x^2)^{-1}$ olasılık yoğunluk fonksiyonuna sahip olan Cauchy dağılımından rastgele örnek gibi genelleştirilmiştir. Buradaki örnek ortalaması, makul etkinliği şöyle dursun Cauchy dağılımının beklenen değerinin bulunamamasından dolayı uygun değildir. Eğer örnek ortalaması, konum (location) tahmin edicisi olarak kullanılmış olsaydı tahmin prosedüründe verinin çok az kullanımı sağlanabilirdi.

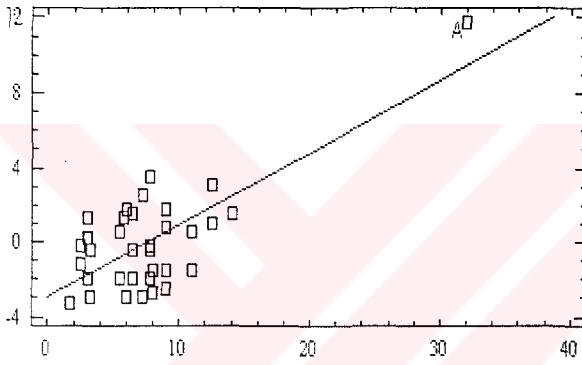
5.3. Etkili Gözlemler (Influent Observes)

Biliyoruz ki, istatistiksel modellerin tümü genellikle daha karmaşık süreçlerin yaklaşık olarak tanımlamalarıdır. Hassas olmayan istatistiksel sonuçlar güçlü ve sağlamdır, bu bakımdan yararlıdır ama çok küçük bir tedirginliğe bile hassas olan bu sonuçlar için özel tedbirler alınması gerekir. Gerçekten bir istatistiksel modelin etkinliği, veri kümesindeki beklenmeyen bir veya birkaç gözlemden oldukça etkilenebilir. Örneğin, $D(x_1, y_3)$ veri kümesindeki (13,00 ; 12,70) noktası beklenmeyen bir noktadır. Örnek 5.1'de bu nokta çok güzel görülmektedir (Şekil 5.1.(c)). Diğer bir deyişle, bu gözlem veri kümesinden çıkarıldığında veya silindiğinde istatistiksel çıkarımlar büyük ölçüde değişebilir. Bu gözlem literatürde etkili gözlem olarak bilinir (12,13). Etkili gözlem analizlerinde temel düşünce çok basittir. Küçük bir şüphe ortaya çıkarılır ve problem formülüne uygulanıp, elde edilen analizlerle bu kuşkulu gözlemlerin nasıl değişime uğradığı takip edilir. Etki analizleri için tasarlamadaki en önemli sorun, kuşkulu gözlemlerin seçim planı, takip edilecek analizin genel yönleri ve ölçme metodlarının tespit edilmesidir. Bu sorunların olası çözümleri bizi çeşitli tanılara yönlendirir. Örneğin, kuşkulu gözlemlerin planlanmasından elde edilen sonuçlar veriyeye

uygulandığında, hataların normal dağılması durumuna göre farklı değerler elde etmemiz sağlanır.

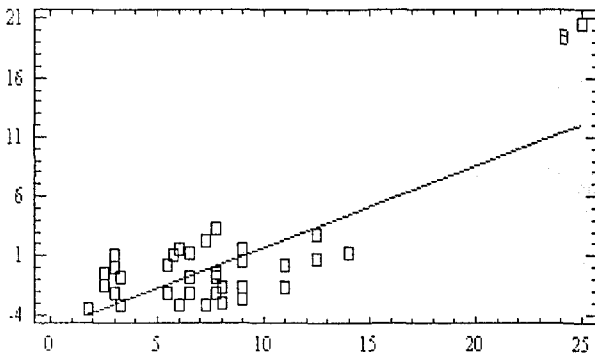
5.4. Aykırı ve Etkili Gözlemler Arasındaki İlişki

Genelde, aykırı ve etkili gözlemler arasında yakın bir ilişki yoktur. Aykırı gözlemler etkili gözlem olabilir de olmayabilir de tabi terside geçerlidir, yani her etkili gözlem aykırı gözlem değildir.



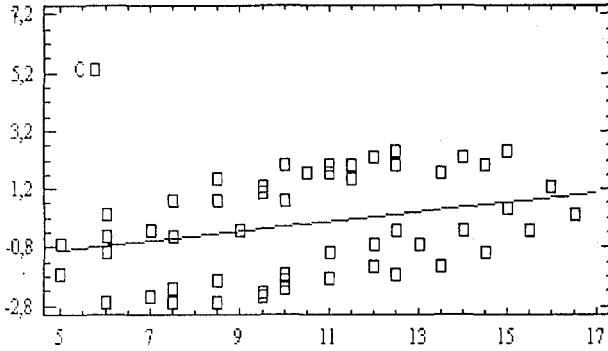
Şekil 5.2.(a)

Şekil 5.2.(a) Yalnızca A gözlemi etkili bir gözlemdir, fakat artışı küçük olduğundan aykırı bir gözlem değildir.



Şekil 5.2 (b)

Şekil 5.2 (b) B gözlemi sadece etkili bir gözlem değil, ayrıca aykırı bir değerdir.



Şekil 5.2.(c)

Şekil 5.2.(c) C noktası aykırı gözlemdir, fakat etkili gözlem değildir.

5.5. Çok Değişkenli Veri Kümesinde Çoklu Aykırı Gözlemler

Sosyal, biomedikal, eğitim ve fizik bilimlerinde araştırmacılar birkaç değişkenle çalışırlar. Çok değişkenli istatistiksel metodlar buna benzer verilerin analizinde kullanılır. Çok değişkenli veri analizleri ilgi çekicidir çünkü, veriler birbirleriyle ilişkili ve çok boyutludur.

Çok değişkenli analizde, istatistiksel özellikler ve modeller birbirleriyle ilişkilidir. Modelin etkinliğinde bir ya da birden fazla beklenmeyen gözlem, aykırı ve / veya etkili gözlemlerden etkilenir. Birçok istatistikçi, veri kümesindeki gözlemlerin %10 veya %15 i oranında beklenmeyen kuşkulu gözlemlerin olduğunu söyler. Ascombe 'nin açısından bakıldığında, aslında tüm veri kümesi aykırı gözlem içerebilir ve asıl sorun ise bunların aldatıcı olup olmamasıdır. Yani, aykırı gözlemler dağılımlardan üretilmezler. Örneğin normal dağılım $N_p \sim (\mu, \Sigma)$ iken, alternatif dağılım $N_p \sim (\mu + a, \Sigma)$ (a , sıfırdan farklı olmak üzere) ortalaması değiştirilmiş bir yığından elde edilmiş bir dağılım olabilir. Dolayısıyla, istatistikte böyle aldatıcı gözlemlerin tanımlanması gereklidir.

Çok değişkenli-çoklu (multivariate-multiple) analizde aykırı ve/veya etkili gözlemleri belirlemek zor bir durumdur. Bu zorluk bağımlı değişken sayısının birden fazla olmasından kaynaklıdır. Bu durumda, çoklu aykırı gözlemler sadece ve sadece dağılış ölçülerini değil aynı zamanda veri kümesindeki korelasyonda etkiler (14). Buna

ek olarak, çok değişkenli analizde aykırı gözlemleri karakterize etmek, tek değişkenli analizde aykırı gözlemleri karakterize etmekten daha zordur. Tek değişkenli analizde sıralanmış gözlemlerden en uçta yer alan gözlemler kuşku uyandırmakta ve en uçta yer alan bu gözlemlerin aykırılığının istatistiksel bakımdan mantıklı olup olmadığını, uygun bir uzaklık ölçüsüyle ya da alternatif modelin dağılımı ile değerlendirebiliriz. Tek değişkenli durumda gözlemlerin büyüklüklerine göre sıralanması mümkün iken, birden fazla değişken olduğunda gözlemleri büyüklüklerine göre sıralamak mümkün değildir. Örneğin, bir grup insanın boyları ölçülüp sayı doğrusu üzerinde sıralanabilir. Ancak hem boy hem de kiloları ölçüldüğünde bu iki tesadüfi değişkenin, düzlem üzerinde yer alan gözlemlerin büyüklüklerine göre sıralanıp sıralanmayacağından bahsetmek mümkün olmayacaktır. Düzlem üzerinde hangi gözlemin daha büyük ya da daha küçük olduğunu yani uç gözlem olduğunu bilmemizi sağlayacak şekilde düzlemde yer alan tüm noktalar için tanımlı bir sıralamanın mümkün olmaması, düzlemde yer alan gözlemlerin belirli bir ölçü bakımından aynı olan gözlemlerden oluşan alt gruplara ayrılmasını gerekli kılmaktadır. Böylece bu alt gruplar üzerinden bir sıralama yapılabilir. Belirli bir ölçü bakımından aynı özellikleri gösteren gözlemlerden oluşan alt grupların, kullanılan ölçü üzerinden sıralanması alt-sıralama (sub-ordering) olarak adlandırılır. Fakat bu şekilde sıralama yapmanın da bazı olumsuzlukları vardır. Bunlardan biri, kullanılacak tek ve değişmez ölçünün olmayışı ve aynı grupta yer alan gözlemlerin tek bir değerle ifade ediliyor olmasıdır. Ayrıca grupta yer alan gözlemlerin tek tek irdelenmesi de mümkün olmamaktadır.

Literatürde çok değişkenli aykırı gözlemler için çeşitli alt sıralama ilkeleri önerilmektedir. Bu ilkeler arasında en yaygın olarak bilineni Mahalanobis uzaklığıdır. $x_1, x_2, \dots, x_n$, p boyutlu yığından gelen n birimlik tesadüfi örnek olsun. Yani, $F_p \sim (\mu, \Sigma)$. Burada μ , p değişkenli konum parametresi ve Σ , $(p \times p)$ boyutlu $(\Sigma > 0)$ yayılış parametresidir. i .nci gözlem x_i ve konum tahmin edicisi $\tau(x)$ arasındaki uzaklık $C(x)$ 'e bağlıdır. Bu durumda, dağılım ölçümü aşağıda verildiği gibidir:

$$D_i(\tau(x), C(x)) = \sqrt{(x_i - \tau(x))' C^{-1}(x) (x_i - \tau(x))} . \quad [5.1]$$

$\tau(x)$ ve $C(x)$ 'in geleneksel seçiminde, aritmetik ortalama $\bar{x}$ olarak, örneğin kovaryans matrisi de S^2 olarak gösterilmektedir. Bu durumda eşitlik [5.1], $MD_i \equiv D_i(\bar{x}, S^2)$ 'yi işaret eder ki bu bilinen Mahalanobis uzaklığıdır. MD_i 'nin büyük değerinin aykırı gözleme karşılık gelen bir gözlem olduğu makul görülür. Peki MD_i 'nin seçimi nasıl sağlanır? $F_p \equiv N_p$, MD_i 'nin olabilirlik (likelihood) oran test istatistiğine denk olduğu durumdur, burada yokluk hipotezi $N_p \sim (\mu, \Sigma)$ 'dir. Alternatif hipotez ise ya ortalaması kayan normal model ya da varyansı şişirilmiş normal modeldir. MD_i üzerindeki bu sonuç, yığın simetrik dağılıma genişlese bile geçerlidir.

Ne yazık ki, MD_i ile aykırı gözlemleri belirleme yaklaşımı maskeleye (masking) ve batırma (swamping) etkilerinden zarar görmektedir. Çoklu aykırı gözlemler, gereken MD_i değerlerine sahip değildir ayrıca MD_i 'nin büyük değerlerine karşılık gelen tüm gözlemlerinde aykırı gözlem olması gerekli değildir.

Maskelme (masking) ve batırma (swamping) olaylarının nedeni, örnek ortalaması $\bar{x}$ ve kovaryans matrisi S^2 'nin tekli aykırı gözlemlere hassasiyet göstermesidir. Yani, bu bakımdan bu değerler güçlü ve sağlam değildirler. Bu belirleme kriterleri daha güvenilir, daha sağlam tahmin edicilerle yer değiştirmeli, yani M-tahmin edicileri gibi tahmin ediciler kullanılmalıdır. Literatürde birçok M-tahmin edicisi önerilmektedir. Ne yazık ki M-tahmin edicilerinin kırılma noktası, yani tolere (müsaade) edilebilen aykırı gözlemlerin oranı $1 / (p+1)$ 'dir. Bundan dolayı, veri kümesinin aşırı koordinatlı olması durumunda M-tahmin edicileri düşer. M-tahmin edicilerinin bu dezavantajı ile baş edebilmek için yeni bir tahmin edici tanımlamışlardır (15). Bu tahmin edici daha sağlam bir tahmin edicidir ve minimum hacimli elipsoit tahmin edicisi (MVE) olarak bilinir. Bu tahmin edici $(\tau(x), C(x))$ çiftiyle tanımlanır. $C(x)$ 'in determinanı aşağıdaki formüle bağlı olarak minimize edilir:

$$\# \{i : (x_i - \tau(x))' C^{-1}(x) (x_i - \tau(x)) \leq a^2\} \geq h. \quad [5.2]$$

$h = [(n+p+1) / 2]$ olduğu durumda, $([q])$, q 'nun tamsayı kısmıdır), a^2 sayısı sabit değişmezdir ve $\chi^2_{p,0.05}$ gibi seçilebilir. Bu durum, normal dağılımdan gelen verinin en

büyükü istenildiğinde seçilebilir (15). Bu durumda [5.2] eşitliği kolaylık olması için RD_i olarak belirlenmiştir. Literatürde, MVE kriteri çoğunlukla kabul edilmiş durumdadır. Bunun yanında MVE tahmin edicisinin birçok avantajı vardır. Örneğin, bütün tekil olmayan $(p \times p)$ boyutlu matrisler için, $b \in R^p$ durumunda $\tau(Ax + b) = A\tau(x) + b$ ve $C(Ax + b) = AC(x)A'$ durumları eş varyantlıdır (16). Daha da önemlisi, MVE tahmin edicisi, %50 'ye yakın bir kırılma noktasına sahiptir. Bu nedenle çok değişkenli aykırı gözlemlerde belirleme yapılırken RD_i daha etkili olmalıdır. Ayrıca MVE tahmin edicisinin asimptotik özellikleri dolayısıyla $n^{-1/3}$ oranında non - Gaussian dağılımına (17) yaklaştığı görülür. Bununla birlikte, pratikte MVE tahmin edicisini kullanmak birçok problemi de beraberinde getirir. Birincisi; MVE tahmin edicisi, veri noktalarının genel bir pozisyonda olması durumunda p noktalarının en fazla $(p-1)$ boyutlu hiperdüzlemde olmasını gerektirir. İkincisi ise; MVE, hesaplamada oldukça pahalıya mal olur ve tam bir MVE tahmin edicisi bulmak mümkün olmayabilir. Örneğin, h noktalarını örten elipsoit sayısı $n! / h! (n-h)!$ olduğu için, MVE tahmin edicisi bütün elipsoitlerin hacmi hesaplanana kadar elde edilemez. İki boyutlu veri seti için, $n = 20$ olduğunda hacmi hesaplanması gereken 167.960 tane elipsoit ortaya çıkar. $n = 30$ için elipsoit sayısı 145 milyona kadar çıkabilir. Bu hesapsal birikimi indirmek amacıyla MVE tahmin edicisine yakın bir çok basit algoritma önerilmektedir. Bu algoritmalar, izdüşüm (projection) ve tekrarlı örnek (resampling) algoritmalarıdır (15). Bununla beraber, algoritmalarda birtakım sınırlamalar olduğu görülür. Örneğin, tekrarlı örnek (resampling) algoritmasında ilerlenmiş tekrar örnek (resample) sayıları verilmelidir. Tekrar örnek (resample) sayılarının etkili bir biçimde nasıl seçileceği henüz açık değildir. İzdüşüm algoritması için, bütün olası yaklaşımlar uygulanıp denenebileceği gibi, sadece izdüşüm yaklaşımlarını almayı önermektedir (15). Buna göre $i = 1, 2, \dots, n$ durumunda $(x_i - m)$, izdüşüm yaklaşımlarının bir biçimidir ve m , ortanca içeren bir vektördür. Bu pratik açıdan yeterliliğe son derece uzaktır. İzdüşüm araştırması (PP) açısından, çok değişkenli veri kümesinde çoklu aykırı gözlemleri tanımlamak için bazı yöntemler ileri sürülmektedir (18). Pan, Fung ve Fang 'ın geliştirdiği bu test istatistiği ki-kare dağılımı gösterir. Aykırı gözlemleri tanımlayan yapı, yüksek boyutlu kümeye indekslenerek merkezlenmiş Gaussian sürecini içermektedir. Merkezlenmiş Gaussian süreci de, kümede farklı noktalar

üretmeye yönlendirmektedir. Bu dağılım asimptotik dağılım gibidir. İzdüşüm araştırması tipinde, aykırı gözlemleri belirlemek için en önemli soru alanda kaç tane yön düzgün olarak üretilmelidir? Yeni koordinat sisteminin araştırması, aykırı gözlemlerle ilgili mümkün olan en fazla bilgiyi ortaya çıkaran kesin izdüşüm belirleyicilerini içermektedir. Örneğin, örnek basıklığı ve örnek çarpıklığı birer izdüşüm belirleyicisidir (19,20).

5.6. Uyumsuz Aykırı Gözlem Üreten Model

İstatistik literatüründe, 'outlier' (aykırı gözlem) kelimesinin açık olarak bir tanımının verilmemesine rağmen, özellikle tek değişkenli örnekte aykırı gözlemlerin bulunması kapsamlı olarak araştırılmıştır (10,13). Kabul edilen en geniş tanımlardan birisi Roussessuw ve Leroy tarafından verilir (11). Bu tanıma göre, verinin çoğunluğunun dağılımsal örüntüsüne (distributional pattern) uygunluk göstermeyen gözlemler aykırı gözlem gibi tanımlanır ve uzak gözlem değerleri gerçek olmayan gözlemler olarak verilir. Barnett ve Lewis tarafından ileri sürülen 'uyumsuz aykırı gözlem' kavramı, dağılımın örüntüsünün sahteliğini / tutarsızlığını (spuriousness / inconsistency) yansıtmaktadır (10). Aykırı gözlem uyumsuzdur çünkü, farklı dağılımdan, gerçeğe aykırı (sahte) bir şekilde oluşmaktadır. Diğer bir yandan, dağılım örüntüsü, örneğin yığın, genellikle dağılım parametreleri ve ortalama gibi tipik parametrelerle nitelendirilir. Dolayısıyla, karakteristik parametrelerdeki değişim ile dağılım örüntüsünden sapma sık sık yansıtılabilir. Gerçekte, istatistik literatüründe uyumsuz (discordant) aykırı gözlem üreten ortak iki model vardır. Bunlar; ortalaması kayan model (mean-shift model) ve varyans şişirme modeli (variance inflation model)'dir.

Varsayalım ki , p değişkenli örnek $y_1, y_2, \dots, y_n$, $F_p(\mu, \Sigma)$ dağılımından gelsin. Burada her iki konum parametresi μ ve Σ bilinmiyor. Ortalaması kayan modeli için, bazı gözlemler $y_{i1}, y_{i2}, \dots, y_{ik}$ ($1 \leq k \leq n/2$) olarak verilir ve ortalama vektörleri μ 'den $\mu + m_i$ 'ye ($1 \leq i \leq k$) kayar. Burada, p değişkenli vektörler, $m_{i1}, m_{i2}, \dots, m_{ik}$ gibi ortalaması kayan parametreler olarak bilinir. Bu gösterim aşağıdaki gibidir:

$$y_i \sim \begin{cases} F_p(\mu, \Sigma) & , i \notin I = \{i_1, i_2, \dots, i_k\} \\ F_p(\mu + m, \Sigma) & , d.h. \end{cases} \quad [5.3]$$

Başka bir deyişle, $y_{i1}, y_{i2}, \dots, y_{ik}$ gözlem değerleri, $F_p(\mu, \Sigma)$ dağılımın örüntüsüne uygunluk göstermeyerek ortalama vektörlerinde kayma oluşur. Bu yüzden, $y_{i1}, y_{i2}, \dots, y_{ik}$ 'ların uyumsuz aykırı gözlemler olup olmadığının bulunması aşağıdaki hipotezin test edilmesi ile eş durumdadır:

$$H : m_{i1} = m_{i2} = \dots = m_{ik} = 0. \quad [5.4]$$

Yokluk hipotezi H , α büyüklüğünde reddedildiğinde $y_{i1}, y_{i2}, \dots, y_{ik}$ gözlem değerleri α düzeyinde uyumsuz aykırı gözlemler şeklinde ifade edilebilir. Şişirilmiş varyans modeli için, uyumsuz aykırı gözlem üreten model aşağıdaki gibidir:

$$y_i \sim \begin{cases} F_p(\mu, \Sigma) & , i \notin I = \{i_1, i_2, \dots, i_k\} \\ F_p(\mu, w_i^2 \Sigma) & , d.h. \end{cases} \quad [5.5]$$

Burada, $w_{i1}^2, w_{i2}^2, \dots, w_{ik}^2$ şişirilmiş varyans parametreleridir. Bu durumda uyumsuz aykırı gözlemler varyansların şiştiği durumda ortaya çıkar. Benzer olarak, uyumsuz aykırı gözlemleri bulmak aşağıdaki hipotezin test edilmesi ile eş durumdadır:

$$H : w_{i1} = w_{i2} = \dots = w_{ik} = 1. \quad [5.6]$$

Tekrarlarsak yokluk hipotezi H , α büyüklüğünde reddedilirse $y_{i1}, y_{i2}, \dots, y_{ik}$ gözlem değerleri α düzeyinde uyumsuz aykırı gözlemler olarak kabul edilebilir. Uyumsuz aykırı gözlemler üreten model için daha fazla bilgi için Barnett ve Lewis'e bakılabilir (10).

Açıklama 5.1.

$k=1$ ve $I=\{i\}$, $1 \leq i \leq n$, olduğunda yalnız i .nci bireyin (durumun) tek uyumsuz aykırı gözlem olup olmadığını bulmaya ihtiyaç duyarız. Wilks'in dağılımsal özelliklerinden birine göre H yokluk hipotezi altında yokluk dağılımı $\Lambda_i \sim \Lambda(m, n-k-r, k)$ aşağıdaki ifadedeki gibidir:

$$\frac{n-r-m}{m} \cdot \frac{1-\Lambda_i}{\Lambda_i} \sim F_{m, n-r-m}.$$

Bundan dolayı, eğer aşağıdaki [5.7] eşitsizliği sağlanırsa i .nci birey tek uyumsuz aykırı gözlem olarak açıklanır.

$$T_i \equiv \frac{e_i' X (X' S X)^{-1} X' e_i}{1 - p_{ii}} \geq \frac{m C_{\alpha}^*}{(n-r-m) + m C_{\alpha}^*} \quad [5.7]$$

Burada, p_{ii} , P_Z izdüşüm matrisinin i .nci köşegen (diagonal) elemanıdır ve $p_{ii} = z_i' (Z Z')^{-1} z_i$ biçimindedir. e_i , E artığının i .nci sütunu ve C_{α}^* , $F_{m, n-r-m}$ dağılımının α anlamlılık düzeyinde üst kritik noktasıdır (6).

6. UYGULAMA

Ek-4 'de verilen ölçümler 2003 yılının Kasım ayında Gazi Üniversitesi Tıp Fakültesinde doğan, doğum anı, 2.ay, 4.ay ve 6.ay itibarı ile Gazi Üniversitesi Çocuk Sağlığı Hastalıkları Anabilim Dalına kontrole gelen kız ve erkek bebeklerin muayeneleri sonucunda elde edilmiştir.

$$Y_{pxn} = X_{pxm} B_{mxr} Z_{rxn} + \varepsilon_{pxn}$$

$$Y \sim N_{p,n}(XBZ; \Sigma, I_n)$$

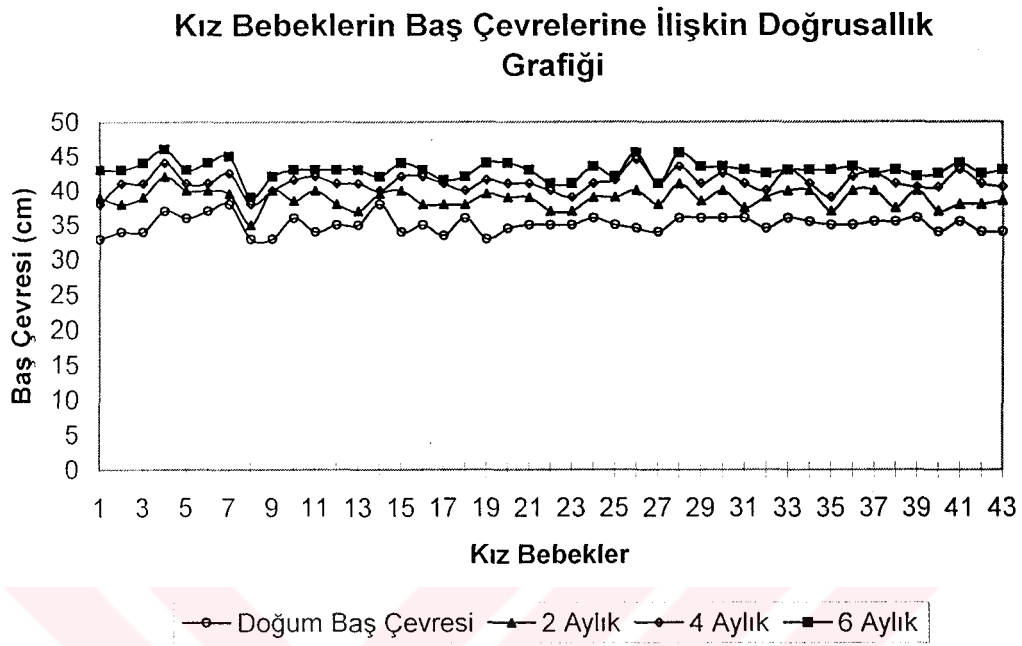
$$\varepsilon \sim N_{p,n}(0; \Sigma, I_n)$$

Ölçümler eş zamanlı yapıldığından, gözlemler X ve Z tasarım matrisine göre iki grupta toplanarak aşağıdaki gibi sınıflandırılmıştır:

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 4 \\ 1 & 6 \end{bmatrix} \quad Z = \begin{bmatrix} 1'_{43} & 0 \\ 0 & 1'_{37} \end{bmatrix}.$$

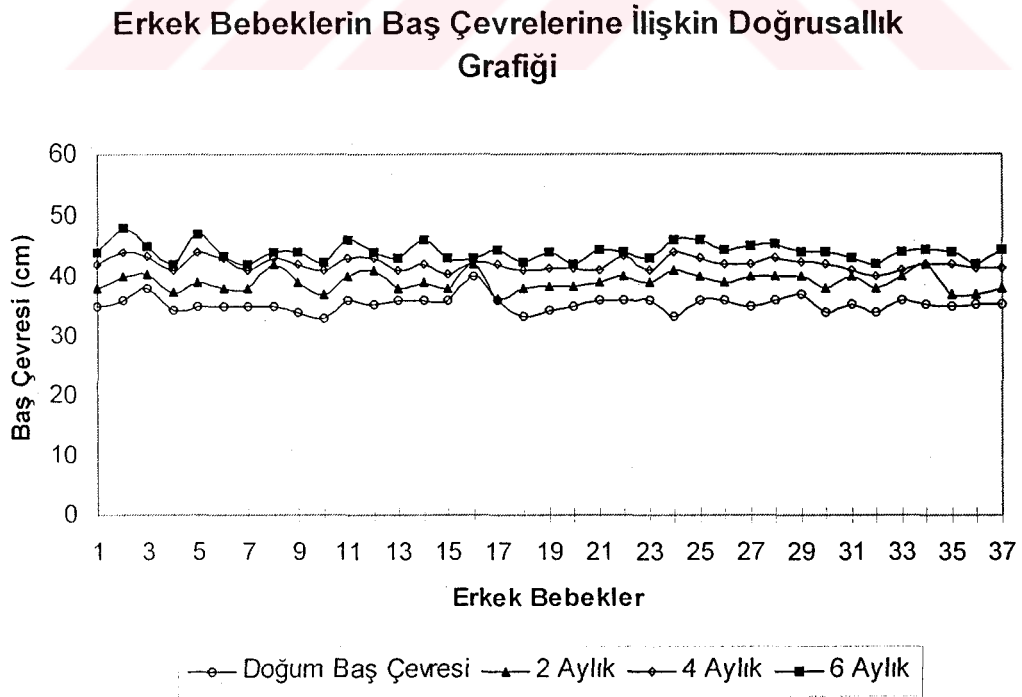
Bu veri çalışmasının amacı, ayrı ayrı kız ve erkek bebeklerin gelişimleri arasındaki baş çevresi, boy uzunluğu ve ağırlık değişkenlerine ilişkin büyüme özelliklerini araştırmak ve bu iki grup arasında belirgin bir fark olup olmadığını göstermektir.

Kız ve erkek bebeklerin ilgili değişkenlere göre doğrusallık grafikleri aşağıdaki gibi verilmektedir.



Şekil 6.1.

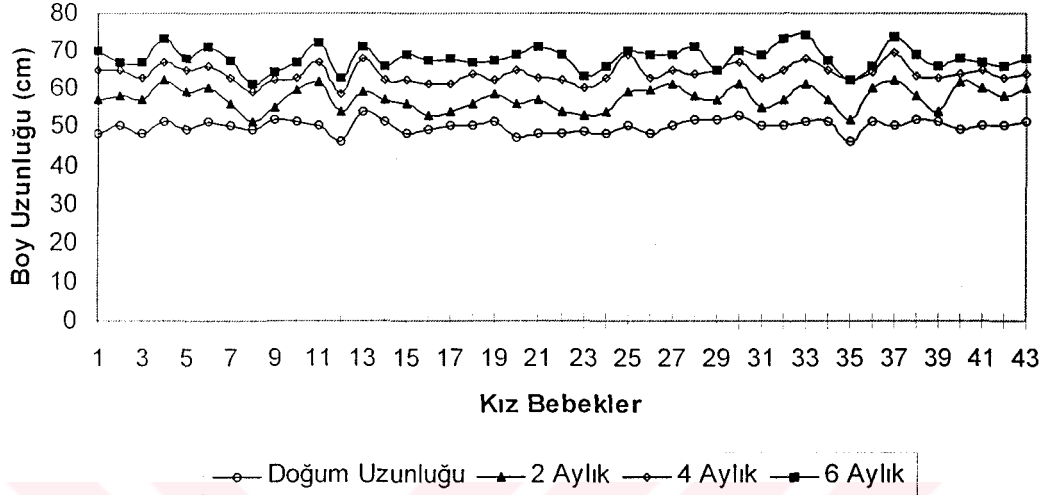
Şekil 6.1. Kız bebeklerin baş çevrelerine ilişkin doğrusallık grafiği



Şekil 6.2.

Şekil 6.2. Erkek bebeklerin baş çevrelerine ilişkin doğrusallık grafiği

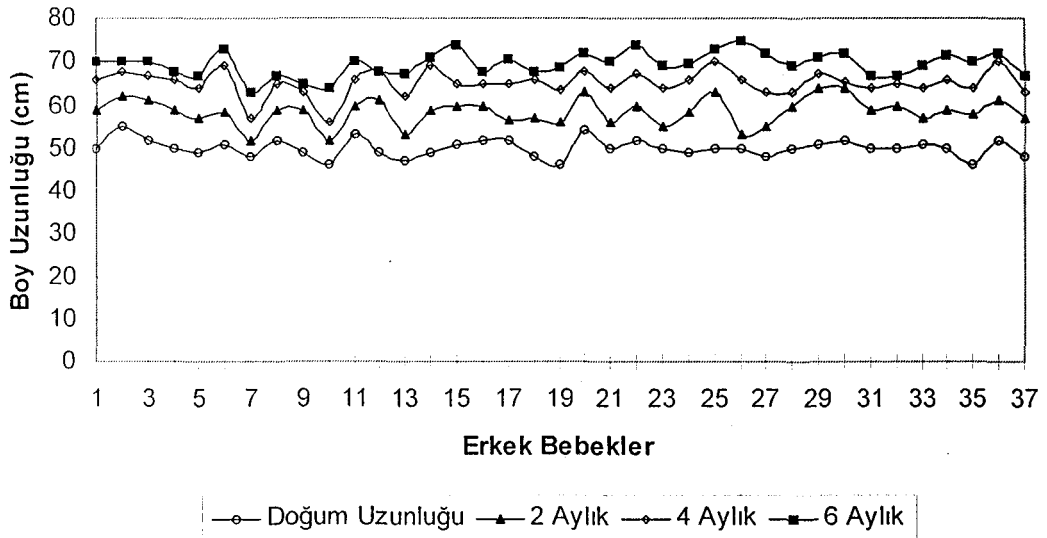
Kız Bebeklerin Boy Uzunluklarına İlişkin Doğrusallık Grafiği



Şekil 6.3.

Şekil 6.3. Kız bebeklerin boy uzunluklarına ilişkin doğrusallık grafiği

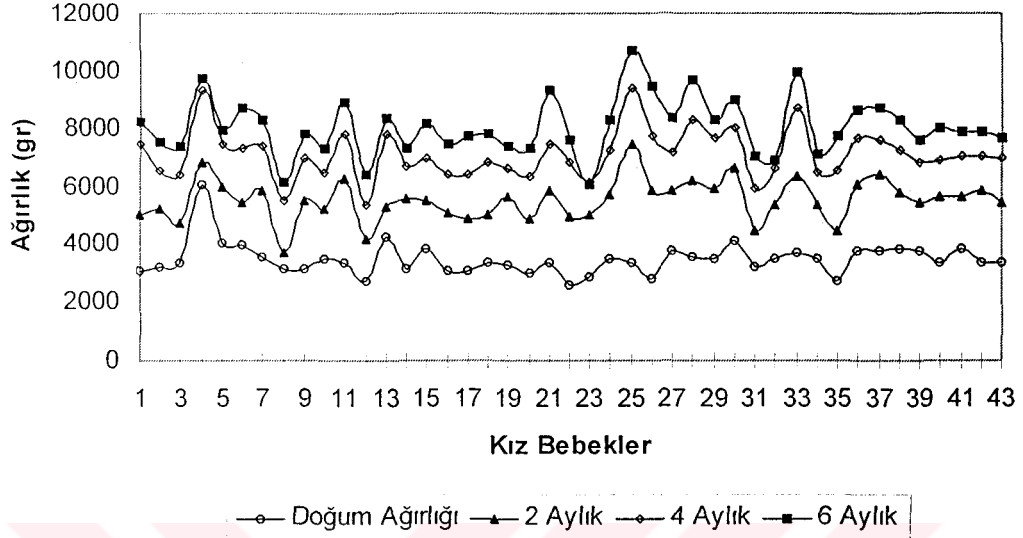
Erkek Bebeklerin Boy Uzunluğuna İlişkin Doğrusallık Grafiği



Şekil 6.4.

Şekil 6.4. Erkek bebeklerin boy uzunluklarına ilişkin doğrusallık grafiği

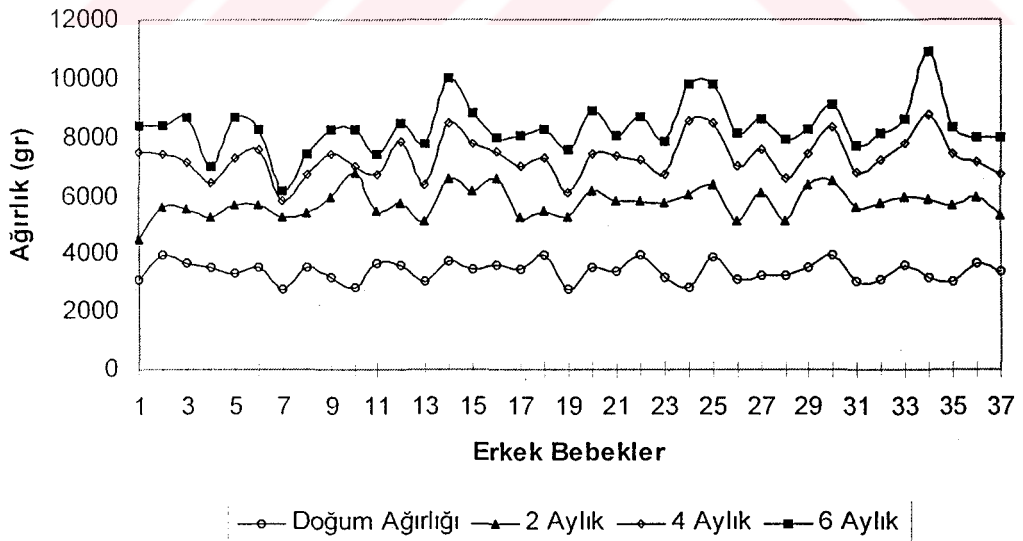
Kız Bebeklerin Ağırlıklarına İlişkin Doğrusallık Grafiği



Şekil 6.5.

Şekil 6.5. Kız bebeklerin ağırlıklarına ilişkin doğrusallık grafiği

Erkek Bebeklerin Ağırlıklarına İlişkin Doğrusallık Grafiği



Şekil 6.6.

Şekil 6.6. Erkek bebeklerin ağırlıklarına ilişkin doğrusallık grafiği

6.1. Genelleştirilmiş En Küçük Kareler Tahminleri (GEKK)

Uygulamamıza baş çevresi, boy uzunluğu ve ağırlığa ilişkin GEKK tahminlerini aşağıdaki formüller yardımıyla hesaplayarak başlayalım:

$$\hat{B} = (X'X)^{-1} X'YZ'(ZZ')^{-1}$$

$$\hat{\Sigma} = \frac{1}{n} (Y - X\hat{B}Z)(Y - X\hat{B}Z)'$$

$$\text{Cov}(\hat{\beta}) = (ZZ')^{-1} \otimes (X'X)^{-1} X' \hat{\Sigma} X (X'X)^{-1}$$

6.1.1. Baş çevresi

Baş çevresi için genelleştirilmiş en küçük kareler tahminleri olan $\hat{B}$, $\hat{\Sigma}$ ve $\text{Cov}(\hat{\beta})$ değerleri aşağıda bulunmuştur.

$$\hat{B} = (X'X)^{-1} X'YZ'(ZZ')^{-1}$$

$$\hat{B} = \begin{bmatrix} 35,5802 & 35,8405 \\ 1,3047 & 1,4405 \end{bmatrix}$$

$$\hat{\Sigma} = \frac{1}{n} (Y - X\hat{B}Z)(Y - X\hat{B}Z)'$$

$$\hat{\Sigma} = \begin{bmatrix} 1,7206 & 0,4074 & 0,3093 & 0,5985 \\ 0,4074 & 2,1418 & 1,0674 & 0,7496 \\ 0,3093 & 1,0674 & 1,4523 & 0,9287 \\ 0,5985 & 0,7496 & 0,9287 & 1,9749 \end{bmatrix}$$

$$Cov(\hat{\beta}) = (ZZ')^{-1} \otimes (X'X)^{-1} X' \hat{\Sigma} X (X'X)^{-1}$$

$$Cov(\hat{\beta}) = \begin{bmatrix} 0,0305 & -0,0036 & 0 & 0 \\ -0,0036 & 0,0015 & 0 & 0 \\ 0 & 0 & 0,0354 & -0,0042 \\ 0 & 0 & -0,0042 & 0,0017 \end{bmatrix}$$

Burada t zaman değişkenlerini göstermek üzere sırasıyla $\hat{B}$ 'yı içeren kız grubu için büyüme ortalamasına uyan $\hat{y} = 35,5802 + 1,3047t$ ve erkek grubu için büyüme ortalamasına uyan $\hat{y} = 35,8405 + 1,4405t$ doğruları kullanılmıştır.

$Cov(\hat{\beta})$, her iki grubun da birbirlerinden bağımsız olduklarını göstermektedir. Eğim ve başlangıç noktaları arasındaki ilişki ise her iki grup için de negatiftir. Yani, başlangıç değeri büyük olanın eğimi küçüktür. Bu da, erkek grubunu ele aldığımızda baş çevresi büyük olan erkek bebeklerin baş çevrelerinde daha yavaş bir büyüme, baş çevresi küçük olan erkek bebeklerin baş çevrelerinde daha hızlı bir büyüme olduğunu gösterir. Bu durum kız grubu için de geçerlidir.

6.1.2. Boy uzunluğu

Şimdi baş çevresi için yaptığımız hesaplamaları boy uzunluğu için yapalım:

$$\hat{B} = (X'X)^{-1} X'YZ'(ZZ')^{-1}$$

$$\hat{B} = \begin{bmatrix} 50,6465 & 51,0676 \\ 3,0616 & 3,2331 \end{bmatrix}$$

$$\hat{\Sigma} = \frac{1}{n}(Y - X\hat{B}Z)(Y - X\hat{B}Z)'$$

$$\hat{\Sigma} = \begin{bmatrix} 4,3654 & 2,3191 & 1,6187 & 2,8300 \\ 2,3191 & 9,1897 & 6,6016 & 4,0078 \\ 1,6187 & 6,6016 & 8,2024 & 4,1834 \\ 2,8300 & 4,0078 & 4,1834 & 8,6030 \end{bmatrix}$$

$$Cov(\hat{\beta}) = (ZZ')^{-1} \otimes (X'X)^{-1} X' \hat{\Sigma} X (X'X)^{-1}$$

$$Cov(\hat{\beta}) = \begin{bmatrix} 8,3498 & -0,4924 & 0 & 0 \\ -0,4924 & 0,3499 & 0 & 0 \\ 0 & 0 & 9,7039 & -0,5722 \\ 0 & 0 & -0,5722 & 0,4066 \end{bmatrix}$$

Burada t zaman değişkenlerini göstermek üzere sırasıyla $\hat{B}$ 'yı içeren kız grubu için büyüme ortalamasına uyan $\hat{y} = 50,6465 + 3,0616t$ ve erkek grubu için büyüme ortalamasına uyan $\hat{y} = 51,0676 + 3,2331t$ doğruları kullanılmıştır.

$Cov(\hat{\beta})$, her iki grubun da birbirlerinden bağımsız olduklarını göstermektedir. Eğim ve başlangıç noktaları arasındaki ilişki ise her iki grup için de negatiftir. Yani, başlangıç değeri büyük olanın eğimi küçüktür. Bu da, erkek grubunu ele aldığımızda boy uzunluğu büyük olan erkek bebeklerin boy uzunluklarında daha yavaş bir büyüme, boy uzunluğu küçük olan erkek bebeklerin boy uzunluklarında daha hızlı bir büyüme olduğunu gösterir. Bu durum kız grubu için de geçerlidir.

6.1.3 Ağırlık

Son olarak ağırlık için de aynı hesapları yaparak aşağıdaki sonuçları elde ederiz:

$$\hat{B} = (X'X)^{-1} X'YZ'(ZZ')^{-1}$$

$$\hat{B} = \begin{bmatrix} 3,7189 & 3,7443 \\ 0,7633 & 0,8175 \end{bmatrix}$$

$$\hat{\Sigma} = \frac{1}{n} (Y - X\hat{B}Z)(Y - X\hat{B}Z)'$$

$$\hat{\Sigma} = \begin{bmatrix} 0,3133 & 0,0327 & 0,0945 & 0,2502 \\ 0,0327 & 0,4744 & 0,4541 & 0,2954 \\ 0,0945 & 0,4541 & 0,6589 & 0,5520 \\ 0,2502 & 0,2954 & 0,5520 & 0,9203 \end{bmatrix}$$

$$Cov(\hat{\beta}) = (ZZ')^{-1} \otimes (X'X)^{-1} X' \hat{\Sigma} X (X'X)^{-1}$$

$$Cov(\hat{\beta}) = \begin{bmatrix} 0,0047 & -0,0001 & 0 & 0 \\ -0,0001 & 0,0005 & 0 & 0 \\ 0 & 0 & 0,0054 & -0,0001 \\ 0 & 0 & -0,0001 & 0,0005 \end{bmatrix}$$

Burada t zaman değişkenlerini göstermek üzere sırasıyla $\hat{B}$ 'yı içeren kız grubu için büyüme ortalamasına uyan $\hat{y} = 3,7189 + 0,7633t$ ve erkek grubu için büyüme ortalamasına uyan $\hat{y} = 3,7443 + 0,8175t$ doğruları kullanılmıştır.

$Cov(\hat{\beta})$, her iki grubun da birbirlerinden bağımsız olduklarını göstermektedir. Eğim ve başlangıç noktaları arasındaki ilişki ise her iki grup için de negatiftir. Yani, başlangıç değeri büyük olanın eğimi küçüktür. Bu da, erkek grubunu ele aldığımızda

ağırlığı büyük olan erkek bebeklerin ağırlıklarında daha yavaş bir büyüme, ağırlığı küçük olan erkek bebeklerin ağırlıklarında daha hızlı bir büyüme olduğunu gösterir. Bu durum kız grubu için de geçerlidir.

Her üç örnek için de görüldüğü üzere, erkek grubunun başlangıç noktası (intercept) kız grubuna göre çok az farkla da olsa büyük olmasına rağmen erkek grubunun ortalama büyüme hızı kız grubuna göre daha hızlı olmaktadır. Bebekler üzerinde sürekli izlenen ölçümlerdeki ortak değişimleri, genelleştirilmiş en küçük kareler tahmini olan $\hat{\Sigma}$ 'da görebiliriz. $\hat{\Sigma}$ 'da gözlemlenen büyük değerler, sürekli izlenen verilerde aşırı yayılım (overdispersion) içerebilir.

6.2. En Çok Olabilirlik Tahminleri (EÇOB)

Yukarıda hesaplamış olduğumuz tahminleri şimdi de baş çevresi, boy uzunluğu ve ağırlığa ilişkin olarak en çok olabilirlik tahmin yöntemini kullanarak aşağıdaki formüller yardımıyla hesaplayalım:

$$\hat{B} = (X'S^{-1}X)^{-1}X'S^{-1}YZ'(ZZ')^{-1}$$

$$S = Y(I_n - P_{Z'})Y'$$

$$P_{Z'} = Z'(ZZ')^{-1}Z$$

$$\hat{\Sigma} = \frac{1}{n}(Y - X\hat{B}Z)(Y - X\hat{B}Z)'$$

$$Cov(\hat{\beta}) = (ZZ')^{-1} \otimes (X'S^{-1}X)^{-1}X'S^{-1}\hat{\Sigma}S^{-1}X(X'S^{-1}X)^{-1}$$

6.2.1. Baş çevresi

İlk olarak baş çevresine ilişkin en çok olabilirlik tahminleri olan $\hat{B}$, $\hat{\Sigma}$ ve $Cov(\hat{\beta})$ değerlerini hesaplayarak gerekli yorumları yapalım:

$$S = Y(I_n - P_Z)Y'$$

$$P_Z = Z'(ZZ')^{-1}Z$$

$$S = \begin{bmatrix} 120,6826 & 51,7473 & 37,3369 & 33,1003 \\ 51,7473 & 149,4305 & 71,7593 & 76,3565 \\ 37,3369 & 71,7593 & 105,6543 & 85,8580 \\ 33,1003 & 76,3565 & 85,8580 & 144,8253 \end{bmatrix}$$

$$\hat{B} = (X'S^{-1}X)^{-1}X'S^{-1}YZ'(ZZ')^{-1}$$

$$S^{-1} = \begin{bmatrix} 0,0099 & -0,0027 & -0,0019 & 0,0003 \\ -0,0027 & 0,0111 & -0,0045 & -0,0026 \\ -0,0019 & -0,0045 & 0,0211 & -0,0097 \\ 0,0003 & -0,0026 & -0,0097 & 0,0139 \end{bmatrix}$$

$$\hat{B} = \begin{bmatrix} 35,5857 & 35,9107 \\ 1,2615 & 1,4016 \end{bmatrix}$$

$$\hat{\Sigma} = \frac{1}{n}(Y - X\hat{B}Z)(Y - X\hat{B}Z)'$$

$$\hat{\Sigma} = \begin{bmatrix} 1,7515 & 0,3694 & 0,2295 & 0,5077 \\ 0,3694 & 2,2023 & 1,1596 & 0,8557 \\ 0,2295 & 1,1596 & 1,5561 & 0,9775 \\ 0,5077 & 0,8557 & 0,9775 & 1,8508 \end{bmatrix}$$

$$\text{Cov}(\hat{\beta}) = (ZZ')^{-1} \otimes (X'S^{-1}X)^{-1}X'S^{-1}\hat{\Sigma}S^{-1}X(X'S^{-1}X)^{-1}$$

$$Cor(\hat{\beta}) = \begin{bmatrix} 0,0297 & -0,0036 & 0 & 0 \\ -0,0036 & 0,0015 & 0 & 0 \\ 0 & 0 & 0,0346 & -0,0042 \\ 0 & 0 & -0,0042 & 0,0017 \end{bmatrix}$$

Burada t zaman değişkenlerini göstermek üzere sırasıyla $\hat{B}$ 'yı içeren kız grubu için büyüme ortalamasına uyan $\hat{y} = 35,5857 + 1,2615t$ ve erkek grubu için büyüme ortalamasına uyan $\hat{y} = 35,9107 + 1,4016t$ doğruları kullanılmıştır.

$Cor(\hat{\beta})$, her iki grubun da birbirlerinden bağımsız olduklarını göstermektedir. Eğim ve başlangıç noktaları arasındaki ilişki ise her iki grup için de negatiftir. Yani, başlangıç değeri büyük olanın eğimi küçüktür. Bu da, erkek grubunu ele aldığımızda baş çevresi büyük olan erkek bebeklerin baş çevrelerinde daha yavaş bir büyüme, baş çevresi küçük olan erkek bebeklerin baş çevrelerinde daha hızlı bir büyüme olduğunu gösterir. Bu durum kız grubu için de geçerlidir.

6.2.2. Boy uzunluğu

Yukarıdaki baş çevresine ilişkin en çok olabilirlik tahminlerini bulmak için yaptığımız işlemleri boy uzunluğu için tekrarlayalım:

$$S = Y(I_n - P_{Z'})Y'$$

$$P_{Z'} = Z'(ZZ')^{-1}Z$$

$$S = \begin{bmatrix} 285,4384 & 242,3564 & 207,2280 & 155,6328 \\ 242,3564 & 684,4381 & 459,1273 & 383,5354 \\ 207,2280 & 459,1273 & 560,9921 & 421,1367 \\ 155,6328 & 383,5354 & 421,1367 & 609,6213 \end{bmatrix}$$

$$\hat{B} = (X'S^{-1}X)^{-1}X'S^{-1}YZ'(ZZ')^{-1}$$

$$S^{-1} = \begin{bmatrix} 0,0052 & -0,0013 & -0,0010 & 0,0002 \\ -0,0013 & 0,0036 & -0,0021 & -0,0005 \\ -0,0010 & -0,0021 & 0,0056 & -0,0023 \\ 0,0002 & -0,0005 & -0,0023 & 0,0035 \end{bmatrix}$$

$$\hat{B} = \begin{bmatrix} 49,8111 & 49,9856 \\ 3,0781 & 3,2508 \end{bmatrix}$$

$$\hat{\Sigma} = \frac{1}{n} (Y - X\hat{B}Z)(Y - X\hat{B}Z)'$$

$$\hat{\Sigma} = \begin{bmatrix} 3,5721 & 3,1399 & 2,7180 & 1,9365 \\ 3,1399 & 11,5113 & 9,1354 & 4,5637 \\ 2,7180 & 9,1354 & 10,9289 & 4,9938 \\ 1,9365 & 4,5637 & 4,9938 & 7,6404 \end{bmatrix}$$

$$Cov(\hat{\beta}) = (ZZ')^{-1} \otimes (X'S^{-1}X)^{-1} X'S^{-1}\hat{\Sigma}S^{-1}X(X'S^{-1}X)^{-1}$$

$$Cov(\hat{\beta}) = \begin{bmatrix} 0,0829 & -0,0061 & 0 & 0 \\ -0,0061 & 0,0044 & 0 & 0 \\ 0 & 0 & 0,0963 & -0,0071 \\ 0 & 0 & -0,0071 & 0,0051 \end{bmatrix}$$

Burada t zaman değişkenlerini göstermek üzere sırasıyla $\hat{B}$ 'yı içeren kız grubu için büyüme ortalamasına uyan $\hat{y} = 49,8111 + 3,0781t$ ve erkek grubu için büyüme ortalamasına uyan $\hat{y} = 49,9856 + 3,2508t$ doğruları kullanılmıştır.

$Cov(\hat{\beta})$, her iki grubun da birbirlerinden bağımsız olduklarını göstermektedir. Eğim ve başlangıç noktaları arasındaki ilişki ise her iki grup için de negatiftir. Yani, başlangıç değeri büyük olanın eğimi küçüktür. Bu da, erkek grubunu ele aldığımızda boy uzunluğu büyük olan erkek bebeklerin boy uzunluklarında daha yavaş bir

büyüme, boy uzunluğu küçük olan erkek bebeklerin boy uzunluklarında daha hızlı bir büyüme olduğunu gösterir. Bu durum kız grubu için de geçerlidir.

6.2.3. Ağırlık

Son olarak ağırlığa ilişkin en çok olabilirlik tahminleri olan $\hat{B}$, $\hat{\Sigma}$ ve $Cov(\hat{\beta})$ değerlerini aşağıdaki şekilde hesaplarız:

$$S = Y(I_n - P_{Z'})Y'$$

$$P_{Z'} = Z'(ZZ')^{-1}Z$$

$$S = \begin{bmatrix} 17,4372 & 10,4996 & 14,6644 & 12,6464 \\ 10,4996 & 29,7573 & 29,0616 & 31,2050 \\ 14,6644 & 29,0616 & 45,9246 & 51,1034 \\ 12,6464 & 31,2050 & 51,1034 & 66,4667 \end{bmatrix}$$

$$\hat{B} = (X'S^{-1}X)^{-1}X'S^{-1}YZ'(ZZ')^{-1}$$

$$S^{-1} = \begin{bmatrix} 0,0885 & -0,0062 & -0,0612 & 0,0331 \\ -0,0062 & 0,0895 & -0,0638 & 0,0082 \\ -0,0612 & -0,0638 & 0,2451 & -0,1469 \\ 0,0331 & 0,0082 & -0,1469 & 0,1178 \end{bmatrix}$$

$$\hat{B} = \begin{bmatrix} 3,4513 & 3,4712 \\ 0,7131 & 0,7644 \end{bmatrix}$$

$$\hat{\Sigma} = \frac{1}{n}(Y - X\hat{B}Z)(Y - X\hat{B}Z)'$$

$$\hat{\Sigma} = \begin{bmatrix} 0,2205 & 0,1040 & 0,1555 & 0,1479 \\ 0,1040 & 0,8490 & 0,8907 & 0,5828 \\ 0,1555 & 0,8907 & 1,1629 & 0,8541 \\ 0,1479 & 0,5828 & 0,8541 & 0,9096 \end{bmatrix}$$

$$Cov(\hat{\beta}) = (ZZ')^{-1} \otimes (X'S^{-1}X)^{-1} X'S^{-1}\hat{\Sigma}S^{-1}X(X'S^{-1}X)^{-1}$$

$$Cov(\hat{\beta}) = \begin{bmatrix} 0,0044 & -0,0001 & 0 & 0 \\ -0,0001 & 0,0005 & 0 & 0 \\ 0 & 0 & 0,0051 & -0,0002 \\ 0 & 0 & -0,0002 & 0,0005 \end{bmatrix}$$

Burada t zaman değişkenlerini göstermek üzere sırasıyla $\hat{B}$ 'yı içeren kız grubu için büyüme ortalamasına uyan $\hat{y} = 3,4513 + 0,7131t$ ve erkek grubu için büyüme ortalamasına uyan $\hat{y} = 3,4712 + 0,7644t$ doğruları kullanılmıştır.

$Cov(\hat{\beta})$, her iki grubun da birbirlerinden bağımsız olduklarını göstermektedir. Eğim ve başlangıç noktaları arasındaki ilişki ise her iki grup için de negatiftir. Yani, başlangıç değeri büyük olanın eğimi küçüktür. Bu da, erkek grubunu ele aldığımızda ağırlığı büyük olan erkek bebeklerin ağırlıklarında daha yavaş bir büyüme, ağırlığı küçük olan erkek bebeklerin ağırlıklarında daha hızlı bir büyüme olduğunu gösterir. Bu durum kız grubu için de geçerlidir.

6.3. Rao 'nun Basit Kovaryans Yapısı

Önceki uygulamalarda en çok olabilirlik ve genelleştirilmiş en küçük kareler tahminleri, tahmin teknikleri açısından üç farklı veri kümesi için çözümlendi. Bu veri kümelerinde, en çok olabilirlik tahmini $\hat{B}$ ile genelleştirilmiş en küçük kareler tahmini $\hat{B}$ arasındaki farkın önemsenmeyecek kadar az olduğu görüldü. Rao 'nun basit kovaryans yapısı;

$$\hat{\Sigma} = X'X^{-1} + Q\Theta Q'$$

idi. Uygulamanın bu kısmında, Rao 'nun basit kovaryans yapısı ile Rao 'nun basit kovaryans yapısından türetilen rastgele regresyon katsayısı yapısı ve tekdüze kovaryans yapısından yararlanılarak kovaryans tahminleri gösterilecektir. Çalışmamızda regresyon katsayısı B 'nin, en çok olabilirlik tahmini $\hat{B}$ ile genelleştirilmiş en küçük kareler tahmini $\hat{B}$ arasındaki farkın önemsenmeyecek kadar küçük olmasından dolayı takip eden örneklerde regresyon katsayısı B 'nin en küçük kareler tahmini olan $\hat{B}$ değeri kullanılacaktır.

6.3.1. Baş çevresi

Öncelikle baş çevresi için üç farklı kovaryans yapısını ele alarak kovaryans tahminlerini bulalım:

a) Basit kovaryans yapısı için;

$$Q = \begin{bmatrix} 2 & 3 \\ -3 & -5 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

(Bkz.Ek-3)

Q , $X'Q = 0$ eşitliğini sağlayan X 'in ortagonal uzayıdır.

$$\hat{B} = (X'X)^{-1} X'YZ'(ZZ')^{-1}$$

$$\hat{B} = \begin{bmatrix} 35,5802 & 35,8405 \\ 1,3047 & 1,4405 \end{bmatrix}$$

$\hat{\Sigma} = X\Gamma X' + Q\Theta Q'$ Rao 'nun basit kovaryans yapısını göstermek üzere eşitliğin sağ taraftaki kısmının hesaplanmasında aşağıdaki eşitlikten yararlanılır:

$$\Gamma = \frac{1}{n} (X'X)^{-1} X' SX (X'X)^{-1} .$$

Buradan, yayılış bileşeni Γ ;

$$\Gamma = \begin{bmatrix} 1,3115 & -0,1541 \\ -0,1541 & 0,0640 \end{bmatrix}$$

olarak bulunur. Yayılış bileşeni $\hat{\Sigma}$ 'da parametrenin diğer kısmı Θ 'dır. (2x2) boyutlu Θ matrisi tek olmayabilir zira Q matrisi tek değildir. Yayılış bileşeni $\hat{\Sigma}$, $Q\Theta Q'$ matrisinde baştan sona kadar yalnızca Θ matrisi üzerinde bağımlıdır.

$$\Theta = \frac{1}{n} (Q'Q)^{-1} Q'YY'Q(Q'Q)^{-1} , \text{ den,}$$

$$\Theta = \begin{bmatrix} 1,4128 & -0,7706 \\ -0,7706 & 0,4523 \end{bmatrix}$$

ve

$$Q\Theta Q' = \begin{bmatrix} 0,4748 & -0,6202 & -0,1842 & 0,3295 \\ -0,6202 & 0,9052 & 0,0502 & -0,3352 \\ -0,1842 & 0,0502 & 0,4523 & -0,3183 \\ 0,3295 & -0,3352 & -0,3183 & 0,3239 \end{bmatrix}$$

olarak bulunur. Buradan yayılış bileşeni $\hat{\Sigma}$ aşağıdaki gibi bulunur.

$$\hat{\Sigma} = X\Gamma X' + Q\Theta Q'$$

$$\hat{\Sigma} = \begin{bmatrix} 1,7863 & 0,3832 & 0,5109 & 0,7165 \\ 0,3832 & 1,8564 & 0,9493 & 0,5120 \\ 0,5109 & 0,9493 & 1,5556 & 0,9890 \\ 0,7165 & 0,5120 & 0,9890 & 2,0913 \end{bmatrix}$$

b) Rastgele regresyon katsayısı için;

Rastgele regresyon katsayısı için kovaryans yapısı $\hat{\Sigma} = X\Gamma X' + \hat{\sigma}^2 I_p$ biçimindeydi. Eşitliğin sağ tarafındaki yayılış parametresi σ 'nın tahmini aşağıdaki formülden yararlanılarak kolaylıkla bulunabilir:

$$\hat{\sigma}^2 = \frac{1}{(p-m)n} \text{iz}(Y'(I_p - P_X)Y)$$

ve

$$P_X = X(X'X)^{-1}X' \text{ olacak şekilde,}$$

$$\hat{\sigma}^2 = 1,0781$$

olarak bulunur. $\hat{\Sigma} = X\Gamma X' + \hat{\sigma}^2 I_p$ rastgele regresyon yapısını göstermek üzere eşitliğin sağ taraftaki kısmının hesaplanmasında aşağıdaki eşitlikten yararlanılır:

$$\hat{\Gamma} = \frac{1}{n}(X'X)^{-1}X'SX(X'X)^{-1} - \hat{\sigma}^2(X'X)^{-1}$$

Buradan, yayılış bileşeni Γ ;

$$\hat{\Gamma} = \begin{bmatrix} 0,7185 & -0,0463 \\ -0,7471 & 0,1718 \end{bmatrix}$$

olarak bulunur. Rastgele regresyon katsayısı için kovaryans yapısı $\hat{\Sigma} = X\Gamma X' + \hat{\sigma}^2 I_p$ den;

$$\hat{\Sigma} = \begin{bmatrix} 1,7967 & 0,6260 & 0,5334 & 0,4409 \\ 0,6260 & 0,8973 & 0,4140 & 1,0088 \\ 0,5334 & 0,4140 & 1,3727 & 1,5768 \\ 0,4409 & 1,0088 & 1,5768 & 3,2228 \end{bmatrix}$$

biçiminde bulunur.

c) Tekdüze kovaryans yapısı için;

Tekdüze kovaryans yapısı $\hat{\Sigma} = \sigma^2 \{(1-\rho)I_p + \rho 1_p 1_p'\}$ biçimindeydi. Eşitliğin sağ tarafındaki yayılış parametresi σ ve ρ 'nun tahminlerini aşağıdaki formüllerden yararlanarak kolaylıkla bulabiliriz:

$$\hat{\sigma}^2 = \frac{1}{pn} iz(S^*)$$

$$\rho = \frac{1}{p-1} \frac{1_p' S^* 1_p - iz(S^*)}{iz(S^*)} \quad \text{ve} \quad 1_p' = [1 \quad 1 \quad 1 \quad 1]$$

$$S^* = S + P_Q Y P_Z' Y' P_Q.$$

$$P_Q = Q(Q'Q)^{-1}Q'$$

$$P_Q = \begin{bmatrix} 0,3000 & -0,4000 & -0,1000 & 0,2000 \\ -0,4000 & 0,7000 & -0,2000 & -0,1000 \\ -0,1000 & -0,2000 & 0,7000 & -0,4000 \\ 0,2000 & -0,1000 & -0,4000 & 0,3000 \end{bmatrix}$$

$$S = \begin{bmatrix} 120,6826 & 51,7473 & 37,3369 & 33,1003 \\ 51,7473 & 149,4305 & 71,7593 & 76,3565 \\ 37,3369 & 71,7593 & 105,6543 & 85,8580 \\ 33,1003 & 76,3565 & 85,8580 & 144,8253 \end{bmatrix}$$

$$S^* = S + P_Q Y P_Z Y' P_Q$$

$$S^* = \begin{bmatrix} 137,6484 & 32,5959 & 24,7423 & 47,8804 \\ 32,5959 & 171,3432 & 85,3881 & 59,9664 \\ 24,7423 & 85,3881 & 116,1804 & 74,2976 \\ 47,8804 & 59,9664 & 74,2976 & 157,9955 \end{bmatrix}$$

$$\hat{\sigma}^2 = \frac{1}{pn} iz(S^*)$$

$$\hat{\sigma}^2 = 1,8224$$

$$\rho = \frac{1}{p-1} \frac{1_p' S^* 1_p - iz(S^*)}{iz(S^*)}$$

$$\rho = 0,3714$$

olarak bulunur. Tekdüze kovaryans yapısı $\hat{\Sigma} = \sigma^2 \{(1-\rho)I_p + \rho 1_p 1_p'\}$ 'yı basitçe aşağıdaki gibi bulabiliriz:

$$\hat{\Sigma} = \hat{\sigma}^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

$$\hat{\Sigma} = \begin{bmatrix} 1,8224 & 0,6768 & 0,6768 & 0,6768 \\ 0,6768 & 1,8224 & 0,6768 & 0,6768 \\ 0,6768 & 0,6768 & 1,8224 & 0,6768 \\ 0,6768 & 0,6768 & 0,6768 & 1,8224 \end{bmatrix}.$$

6.3.2. Boy uzunluğu

İkinci değişken olan boy uzunluğu için, üç farklı kovaryans yapısına ilişkin tahminleri baş çevresinde olduğu gibi aynı şekilde hesaplarsak, sonuçlar aşağıdaki gibi kısaca gösterilebilir.

a) Basit kovaryans yapısı için;

Bulunan tahmin değerlerini kısaca gösterirsek;

$$\hat{B} = \begin{bmatrix} 50,6465 & 51,0676 \\ 3,0616 & 3,2331 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} 4,4880 & -0,2647 \\ -0,2647 & 0,1881 \end{bmatrix}$$

$$\Theta = \begin{bmatrix} 7,5509 & -4,1379 \\ -4,1379 & 2,3496 \end{bmatrix} \text{ ise } Q\Theta Q' = \begin{bmatrix} 1,6948 & -1,9287 & -1,2271 & 1,4610 \\ -1,9287 & 2,5600 & 0,6659 & -1,2973 \\ -1,2271 & 0,6659 & 2,3496 & -1,7883 \\ 1,4610 & -1,2973 & -1,7883 & 1,6247 \end{bmatrix}$$

$\hat{\Sigma} = X\Gamma X' + Q\Theta Q'$ eşitliğinden;

$$\hat{\Sigma} = \begin{bmatrix} 6,1829 & 2,0301 & 2,2023 & 4,3611 \\ 2,0301 & 6,7417 & 5,0704 & 3,3302 \\ 2,2023 & 5,0704 & 7,7292 & 4,5664 \\ 4,3611 & 3,3302 & 4,5664 & 9,7068 \end{bmatrix}$$

olarak bulunabilir.

b) Rastgele regresyon katsayısı için;

$$\hat{\sigma}^2 = 4,1145$$

$$\Gamma = \begin{bmatrix} 1,6079 & 0,3525 \\ 0,3525 & -0,0177 \end{bmatrix}$$

Buradan da, rastgele regresyon katsayısı için kovaryans yapısı;

$$\hat{\Sigma} = X\Gamma X' + \hat{\sigma}^2 I_p \text{ eşitliğinden,}$$

$$\hat{\Sigma} = \begin{bmatrix} 8,6026 & 3,9587 & 3,4294 & 2,9001 \\ 3,9587 & 8,2962 & 4,4045 & 4,6274 \\ 3,4294 & 4,4045 & 9,4942 & 6,3548 \\ 2,9001 & 4,6274 & 6,3548 & 12,1966 \end{bmatrix}$$

olarak bulunabilir.

c) Tekdüze kovaryans yapısı için;

$$P_Q = \begin{bmatrix} 0,3000 & -0,4000 & -0,1000 & 0,2000 \\ -0,4000 & 0,7000 & -0,2000 & -0,1000 \\ -0,1000 & -0,2000 & 0,7000 & -0,4000 \\ 0,2000 & -0,1000 & -0,4000 & 0,3000 \end{bmatrix}$$

$$S = \begin{bmatrix} 285,4384 & 242,3564 & 207,2280 & 155,6328 \\ 242,3564 & 684,4381 & 459,1273 & 383,5354 \\ 207,2280 & 459,1273 & 560,9921 & 421,1367 \\ 155,6328 & 383,5354 & 421,1367 & 609,6213 \end{bmatrix}$$

$$S^* = \begin{bmatrix} 349,2333 & 185,5305 & 129,4951 & 226,3967 \\ 185,5305 & 735,1775 & 528,1261 & 320,6230 \\ 129,4951 & 528,1261 & 656,1933 & 334,6697 \\ 226,3967 & 320,6230 & 334,6697 & 688,2368 \end{bmatrix}$$

$\hat{\sigma}^2 = 7,5901$ ve $\rho = 0,4734$ olarak bulunur.

Buradan, tekdüze kovaryans yapısı $\hat{\Sigma} = \sigma^2 \{(1 - \rho)I_p + \rho 1_p 1_p'\}$ 'yı basitçe;

$$\hat{\Sigma} = \hat{\sigma}^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix} \text{ eşitliğinden,}$$

$$\hat{\Sigma} = \begin{bmatrix} 7,5901 & 3,5934 & 3,5934 & 3,5934 \\ 3,5934 & 7,5901 & 3,5934 & 3,5934 \\ 3,5934 & 3,5934 & 7,5901 & 3,5934 \\ 3,5934 & 3,5934 & 3,5934 & 7,5901 \end{bmatrix}$$

olarak bulunabilir.

6.3.3. Ağırlık

Üçüncü değişken olan ağırlık için, üç farklı kovaryans yapısına ilişkin tahminleri baş çevresinde olduğu gibi aynı şekilde hesaplırsak, sonuçlar aşağıdaki gibi kısaca gösterilebilir.

a) Basit Kovaryans Yapısı;

Bulunan tahmin değerlerini kısaca gösterirsek;

$$\hat{B} = \begin{bmatrix} 3,7189 & 3,7443 \\ 0,7633 & 0,8175 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} 0,2013 & -0,0039 \\ -0,0039 & 0,0200 \end{bmatrix}$$

$$\Theta = \begin{bmatrix} 0,4172 & -0,2116 \\ -0,2116 & 0,1113 \end{bmatrix} \text{ ise } Q\Theta Q' = \begin{bmatrix} 0,1310 & -0,1518 & -0,0894 & 0,1102 \\ -0,1518 & 0,1886 & 0,0784 & -0,1151 \\ -0,0894 & 0,0784 & 0,1113 & -0,1003 \\ 0,1102 & -0,1151 & -0,1003 & 0,1053 \end{bmatrix}$$

$\hat{\Sigma} = X\Gamma X' + Q\Theta Q'$ eşitliğinden;

$$\hat{\Sigma} = \begin{bmatrix} 0,3323 & 0,0417 & 0,0964 & 0,2881 \\ 0,0417 & 0,4542 & 0,4162 & 0,2948 \\ 0,0964 & 0,4162 & 0,6012 & 0,5416 \\ 0,2881 & 0,2948 & 0,5416 & 0,9792 \end{bmatrix}$$

olarak bulunabilir.

b) Rastgele regresyon katsayısı için;

$$\hat{\sigma}^2 = 0.2681$$

$$\Gamma = \begin{bmatrix} 0,0136 & 0,0363 \\ 0,0363 & 0,0066 \end{bmatrix}$$

Buradan da, rastgele regresyon katsayısı için kovaryans yapısı;

$$\hat{\Sigma} = X\Gamma X' + \hat{\sigma}^2 I_p \text{ eşitliğinden,}$$

$$\hat{\Sigma} = \begin{bmatrix} 0,2817 & 0,0863 & 0,1589 & 0,2315 \\ 0,0863 & 0,4533 & 0,2842 & 0,3831 \\ 0,1589 & 0,2842 & 0,6775 & 0,5347 \\ 0,2315 & 0,3831 & 0,5347 & 0,9544 \end{bmatrix}$$

olarak bulunabilir.

c) Tekdüze kovaryans yapısı için;

$$P_Q = \begin{bmatrix} 0,3000 & -0,4000 & -0,1000 & 0,2000 \\ -0,4000 & 0,7000 & -0,2000 & -0,1000 \\ -0,1000 & -0,2000 & 0,7000 & -0,4000 \\ 0,2000 & -0,1000 & -0,4000 & 0,3000 \end{bmatrix}$$

$$S = \begin{bmatrix} 17,4372 & 10,4996 & 14,6644 & 12,6464 \\ 10,4996 & 29,7573 & 29,0616 & 31,2050 \\ 14,6644 & 29,0616 & 45,9246 & 51,1034 \\ 12,6464 & 31,2050 & 51,1034 & 66,4667 \end{bmatrix}$$

$$S^* = \begin{bmatrix} 25,0634 & 2,6135 & 7,5579 & 20,0128 \\ 2,6135 & 37,9530 & 36,3287 & 23,6284 \\ 7,5579 & 36,3287 & 52,7099 & 44,1575 \\ 20,0128 & 23,6284 & 44,1575 & 73,6228 \end{bmatrix}$$

$\hat{\sigma}^2 = 0,5917$ ve $\rho = 0,4728$ olarak bulunur.

Buradan, tekdüze kovaryans yapısı $\hat{\Sigma} = \sigma^2 \{(1 - \rho)I_p + \rho 1_p 1_p'\}$ 'yı basitçe;

$$\hat{\Sigma} = \hat{\sigma}^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix} \text{ eşitliğinden,}$$

$$\hat{\Sigma} = \begin{bmatrix} 0,5917 & 0,2810 & 0,2810 & 0,2810 \\ 0,2810 & 0,5917 & 0,2810 & 0,2810 \\ 0,2810 & 0,2810 & 0,5917 & 0,2810 \\ 0,2810 & 0,2810 & 0,2810 & 0,5917 \end{bmatrix}$$

olarak bulunabilir.

6.4. Uyumsuz Aykırı Gözlem Bulma

Uygulamanın bu kısmında baş çevresi, boy uzunluğu ve ağırlık değişkenleri için varsa aykırı gözlem araştırması yapılmaktadır.

6.4.1. Baş çevresi

Baş çevresi değişkenini ele alarak tüm gözlemlerimiz için $\hat{e}_{11}, \hat{e}_{12}, \hat{e}_{13}, \hat{e}_{14}$ değerleri hesaplanırsa:

$\hat{e}_{11}$	$\hat{e}_{12}$	$\hat{e}_{13}$	$\hat{e}_{14}$	$\hat{e}_{11}$	$\hat{e}_{12}$	$\hat{e}_{13}$	$\hat{e}_{14}$
-2,5802	0,8104	-2,7990	-0,4084	0,4198	-0,6896	0,2010	-0,4084
-1,5802	-0,1896	0,2010	-0,4084	-1,0802	0,8104	-0,7990	-0,9084
-1,5802	0,8104	0,2010	0,5916	0,4198	1,8104	2,2010	-0,4084
1,4198	3,8104	3,2010	2,5916	-0,0802	1,8104	0,2010	-0,4084
0,4198	1,8104	0,2010	-0,4084	-0,5802	-1,1896	-1,7990	-0,4084
1,4198	1,8104	0,2010	0,5916	-0,5802	1,8104	1,2010	0,0916
2,4198	1,3104	1,7010	1,5916	-0,0802	1,8104	1,7010	-0,9084
-2,5802	-3,1896	-2,7990	-4,4084	-0,0802	-0,6896	0,2010	-0,4084
-2,5802	1,8104	-0,7990	-1,4084	0,4198	1,8104	-0,2990	-1,4084
0,4198	0,3104	0,7010	-0,4084	-1,5802	-1,1896	-0,2990	-0,9084
-1,5802	1,8104	1,2010	-0,4084	-0,0802	-0,1896	2,2010	0,5916
-0,5802	-0,1896	0,2010	-0,4084	-1,5802	-0,1896	0,2010	0,0916
-0,5802	-1,1896	0,2010	-0,4084	-1,5802	0,3104	-0,2990	-0,4084
2,4198	1,3104	-0,7990	-1,4084	-0,8405	-0,7215	0,3975	-0,4835
-1,5802	1,8104	1,2010	0,5916	0,1595	1,2785	2,3975	3,5165
-0,5802	-0,1896	1,2010	-0,4084	2,1595	1,7785	1,8975	0,5165
-2,5802	-0,1896	0,2010	-1,9084	-1,3405	-1,2215	-0,6025	-2,4835
0,4198	-0,1896	-0,7990	-1,4084	-0,8405	0,2785	2,3975	2,5165
-2,5802	1,3104	0,7010	0,5916	-0,8405	-0,7215	1,3975	-0,9835
-1,0802	0,8104	0,2010	0,5916	-0,8405	-0,7215	-0,6025	-2,4835
-0,5802	0,8104	0,2010	-0,4084	-0,8405	3,2785	1,3975	-0,4835
-0,5802	-1,1896	-0,7990	-2,4084	-1,8405	0,2785	0,3975	-0,4835
-0,5802	-1,1896	-1,7990	-2,4084	-2,8405	-1,7215	-0,6025	-1,9835
0,4198	0,8104	0,2010	0,0916	0,1595	1,2785	1,3975	1,5165
-0,5802	0,8104	0,7010	-1,4084	-0,3405	2,2785	1,3975	-0,4835
-1,0802	1,8104	0,7010	2,0916	0,1595	-0,7215	-0,6025	-1,4835
-1,5802	-0,1896	0,2010	-2,4084	0,1595	0,2785	0,3975	1,5165
0,4198	2,8104	0,2010	2,0916	0,1595	-0,7215	-1,1025	-1,4835
0,4198	0,3104	0,2010	0,0916	4,1595	3,2785	0,8975	-1,4835
0,4198	1,8104	1,7010	0,0916	0,1595	-2,2215	0,3975	0,0165

$\hat{e}_{11}$	$\hat{e}_{12}$	$\hat{e}_{13}$	$\hat{e}_{14}$
-2,3405	-0,7215	-0,6025	-1,9835
-1,3405	-0,2215	-0,1025	-0,4835
-0,8405	-0,2215	-0,1025	-2,4835
0,1595	0,2785	-0,6025	0,0165
0,1595	1,2785	1,8975	-0,4835
0,1595	0,2785	-0,6025	-1,4835
-2,3405	2,2785	2,3975	1,5165
0,1595	1,2785	1,3975	1,5165
0,1595	0,2785	0,3975	0,0165
-0,8405	1,2785	0,3975	0,5165
0,1595	1,2785	1,3975	1,0165
1,1595	1,2785	0,8975	-0,4835
-1,8405	-0,7215	0,3975	-0,4835
-0,3405	1,2785	-0,6025	-1,4835
-1,8405	-0,7215	-1,6025	-2,4835
0,1595	1,2785	-0,6025	-0,4835
-0,3405	3,2785	0,3975	0,0165
-0,3405	-1,7215	0,3975	-0,4835
-0,3405	-1,7215	-0,1025	-2,4835
-0,3405	-0,7215	-0,1025	0,0165

değerleri elde edilir.

Tüm gözlem değerlerine karşılık gelen T değerleri aşağıdaki formülden yararlanılarak hesaplanırsa:

$$T_i \equiv \frac{e_i' X (X' S X)^{-1} X' e_i}{1 - p_{ii}}$$

T1= 0,0287	T11= 0,0047	T21= 1,7815e-005	T31= 0,0014	T41= 0,0134
T2= 0,0128	T12= 0,0014	T22= 0,0294	T32= 0,0036	T42= 0,0186
T3= 0,0180	T13= 0,0060	T23= 0,0423	T33= 0,0177	T43= 0,0088
T4= 0,1012	T14= 0,1017	T24= 0,0038	T34= 0,0063	T44= 0,0056
T5= 0,0139	T15= 0,0191	T25= 0,0034	T35= 0,0134	T45= 0,0942
T6= 0,0258	T16= 0,0019	T26= 0,0412	T36= 0,0057	T46= 0,0524
T7= 0,0471	T17= 0,0146	T27= 0,0141	T37= 0,0102	T47= 0,0280
T8= 0,1433	T18= 0,0201	T28= 0,0593	T38= 8,2431e-004	T48= 0,0824
T9= 0,0087	T19= 0,0410	T29= 0,0017	T39= 0,0310	T49= 0,0031
T10= 0,0036	T20= 0,0098	T30= 0,0153	T40= 0,0201	T50= 0,0258

T51= 0,0106	T61= 0,0292	T71= 0,0143
T52= 0,0128	T62= 0,0088	T72= 0,0245
T53= 0,0558	T63= 0,0198	T73= 0,0227
T54= 0,0215	T64= 8,9237e-004	T74= 0,0146
T55= 0,0086	T65= 0,0086	T75= 0,0379
T56= 0,0154	T66= 0,0179	T76= 0,0080
T57= 0,0118	T67= 0,0742	T77= 0,0127
T58= 0,0197	T68= 0,0215	T78= 0,0131
T59= 0,2592	T69= 7,1997e-004	T79= 0,0259
T60= 0,0057	T70= 0,0060	T80= 0,0030

değerleri elde edilir. Buradan, aşağıdaki kısıt altında aykırı gözlem / gözlemler bulunabilir.

$$\frac{mC_{\alpha}^*}{(n-r-2) + mC_{\alpha}^*} = \frac{2(7,02)}{(80-2-2) + 2(7,02)} \cong 0,156$$

$$T_i \geq \frac{mC_{\alpha}^*}{(n-r-2) + mC_{\alpha}^*} \Rightarrow \text{gözlem değeri aykırıdır.}$$

$T_{59} = 0,2592 \geq 0,156$ olduğundan 59.ncu gözlem değeri aykırıdır. Yani 59.ncu bebeğin baş çevresine ilişkin gözlem değerleri diğer bebeklerin gözlem değerlerinden oldukça farklıdır.

6.4.2. Boy uzunluğu

Boy uzunluğu değişkeni ele alınarak tüm gözlem değerleri için $\hat{e}_{11}, \hat{e}_{12}, \hat{e}_{13}, \hat{e}_{14}$ değerleri hesaplanırsa;

$\hat{e}_{11}$	$\hat{e}_{12}$	$\hat{e}_{13}$	$\hat{e}_{14}$	$\hat{e}_{11}$	$\hat{e}_{12}$	$\hat{e}_{13}$	$\hat{e}_{14}$
-2,6465	0,2303	2,1071	0,9839	-2,6465	-0,7697	-0,8929	-0,0161
-0,6465	1,2303	2,1071	-2,0161	-1,6465	-3,7697	-1,8929	-1,5161
-2,6465	0,2303	0,1071	-2,0161	-0,6465	-2,7697	-1,8929	-1,0161
0,3535	5,2303	4,1071	3,9839	-0,6465	-0,7697	1,1071	-2,0161
-1,6465	2,2303	2,1071	-1,0161	0,3535	1,7303	-0,8929	-1,5161
0,3535	3,2303	3,1071	1,9839	-3,6465	-0,7697	2,1071	-0,0161
-0,6465	-0,7697	0,1071	-1,5161	-2,6465	0,2303	0,1071	1,9839
-1,6465	-5,2697	-3,8929	-8,0161	-2,6465	-2,7697	-0,8929	-0,0161
1,3535	-1,7697	-0,8929	-4,5161	-2,1465	-3,7697	-2,8929	-5,5161
0,3535	2,7303	0,1071	-2,0161	-2,6465	-2,7697	0,1071	-3,0161
-0,6465	4,7303	4,1071	2,9839	-0,6465	2,2303	6,1071	0,9839
-4,6465	-2,7697	-4,3929	-6,0161	-2,6465	2,7303	0,1071	-0,0161
3,3535	2,2303	5,1071	1,9839	-0,6465	4,2303	2,1071	-0,0161
0,3535	0,2303	-0,8929	-3,0161	1,3535	1,2303	1,1071	1,9839

$\hat{e}_{11}$	$\hat{e}_{12}$	$\hat{e}_{13}$	$\hat{e}_{14}$	$\hat{e}_{11}$	$\hat{e}_{12}$	$\hat{e}_{13}$	$\hat{e}_{14}$
1,3535	0,2303	2,1071	-4,0161	-2,0676	3,4662	4,0000	-2,4662
2,3535	4,2303	4,1071	0,9839	-4,0676	-4,5338	-2,0000	-2,9662
-0,6465	-1,7697	-0,3929	-0,0161	-2,0676	1,4662	5,0000	0,5338
-0,6465	0,2303	2,1071	3,9839	-0,0676	2,4662	1,0000	3,5338
0,3535	4,2303	5,1071	5,4839	0,9324	2,4662	1,0000	-2,4662
0,3535	0,2303	2,1071	-1,5161	0,9324	-1,0338	1,0000	0,0338
-4,6465	-4,7697	-0,8929	-7,0161	-3,0676	-1,5338	20000	-2,4662
0,3535	3,2303	1,6071	-3,0161	-5,0676	-1,5338	-0,5000	-1,9662
-0,6465	5,2303	6,6071	4,4839	2,9324	5,4662	4,0000	1,5338
1,3535	1,2303	0,6071	-0,0161	-1,0676	-1,5338	0	-0,4662
0,3535	-27697	0,1071	-3,0161	0,9324	2,4662	35000	3,5338
-1,6465	4,7303	11071	-1,0161	-1,0676	-2,5338	0	-1,4662
-0,6465	3,2303	2,1071	-2,0161	-2,0676	0,9662	2,0000	-0,9662
-0,6465	1,2303	0,1071	-3,0161	-1,0676	5,4662	6,0000	2,5338
0,3535	3,2303	1,1071	-1,0161	-1,0676	-4,5338	2,0000	-0,9662
-1,0676	1,4662	2,0000	-0,4662	-3,0676	-2,5338	-1,0000	1,5338
3,9324	4,4662	4,0000	-0,4662	-1,0676	2,4662	-1,0000	-1,4662
0,9324	3,4662	3,0000	-0,4662	-0,0676	6,4662	3,5000	0,5338
-1,0676	1,4662	20000	-2,4662	0,9324	6,4662	1,5000	1,5338
-2,0676	-0,5338	0	-3,4662	-1,0676	1,4662	0	-3,4662
-0,0676	0,9662	50000	2,5338	-1,0676	2,4662	1,0000	-3,4662
-3,0676	-5,5338	-70000	-7,4662	-0,0676	-0,5338	0	-1,4662
0,9324	1,4662	1,0000	-3,4662	-1,0676	1,4662	2,0000	1,4338
-2,0676	1,4662	-1,0000	-5,4662	-5,0676	0,4662	0	-0,4662
-5,0676	-5,5338	-8,0000	-6,4662	0,9324	3,4662	6,0000	1,5338
1,9324	2,4662	2,0000	-0,4662	-3,0676	-0,5338	-1,0000	-3,4662

değerleri elde edilir.

Tüm gözlem değerlerine karşılık gelen T değerleri aşağıdaki formülden yararlanılarak hesaplanırsa;

$$T_i \equiv \frac{e_i' X (X' S X)^{-1} X' e_i}{1 - p_{ii}}$$

T1= 0,0300	T11= 0,0301	T21= 0,0366	T31= 0,0047	T41= 0,0083
T2= 0,0023	T12= 0,0553	T22= 0,0335	T32= 0,0424	T42= 0,0115
T3= 0,0052	T13= 0,0334	T23= 0,0397	T33= 0,0644	T43= 0,0128
T4= 0,0379	T14= 0,0215	T24= 0,0159	T34= 0,0035	T44= 0,0012
T5= 7,7197e-004	T15= 0,0193	T25= 0,0198	T35= 0,0534	T45= 0,0836
T6= 0,0138	T16= 0,0186	T26= 0,0052	T36= 0,0302	T46= 0,0175
T7= 0,0015	T17= 0,0076	T27= 0,0064	T37= 0,0686	T47= 0,0025
T8= 0,0872	T18= 0,0014	T28= 0,0056	T38= 0,0081	T48= 0,0070
T9= 0,0477	T19= 0,0126	T29= 0,0379	T39= 0,0107	T49= 0,0282
T10= 0,0197	T20= 0,0410	T30= 0,0360	T40= 0,0032	T50= 0,1027
T51= 0,0355	T61= 0,0090	T71= 0,0037		
T52= 0,0292	T62= 0,0481	T72= 0,0230		
T53= 0,1103	T63= 0,0538	T73= 0,0272		
T54= 0,0237	T64= 0,0052	T74= 0,0128		
T55= 0,0020	T65= 0,0247	T75= 0,0134		
T56= 0,0523	T66= 0,0062	T76= 0,0024		
T57= 0,0223	T67= 0,0036	T77= 0,0083		
T58= 0,0179	T68= 0,0389	T78= 0,0473		
T59= 0,0294	T69= 0,0169	T79= 0,0249		
T60= 3,2986e-004	T70= 0,0595	T80= 0,0116		

değerleri elde edilir. Buradan, aşağıdaki kısıt altında aykırı gözlem / gözlemler bulunabilir.

$$\frac{mC_{\alpha}^*}{(n-r-2) + mC_{\alpha}^*} = \frac{2(7,02)}{(80-2-2) + 2(7,02)} \cong 0,156$$

$$T_i \geq \frac{mC_{\alpha}^*}{(n-r-2) + mC_{\alpha}^*} \Rightarrow \text{gözlem değeri aykırıdır.}$$

T_i değerlerinden hiçbirisi kritik değer olan 0,156 değerinden büyük olmadığı için bir aykırı gözlem gözlenmemiştir. Dolayısıyla uyumsuz aykırı gözlem çiftlerinden de bahsedemeyiz.

6.4.3. Ağırlık

Ağırlık değişkeni ele alınarak tüm gözlem değerleri için $\hat{e}_{11}, \hat{e}_{12}, \hat{e}_{13}, \hat{e}_{14}$ değerleri hesaplanırsa

$\hat{e}_{11}$	$\hat{e}_{12}$	$\hat{e}_{13}$	$\hat{e}_{14}$	$\hat{e}_{11}$	$\hat{e}_{12}$	$\hat{e}_{13}$	$\hat{e}_{14}$
-0,6389	-0,2455	0,6279	-0,0987	0,0911	0,2545	0,1779	-0,1987
-0,5189	-0,0155	-0,2721	-0,7987	-0,6989	-0,1955	-0,3621	-0,9087
-0,3689	-0,4955	-0,3721	-0,9287	-0,6389	-0,3955	-0,4121	-0,5687
2,3411	1,5545	2,5279	1,4013	-0,4189	-0,2455	0,0279	-0,4987
0,2811	0,7545	0,6279	-0,3987	-0,4689	0,3545	-0,1721	-0,9187
0,2411	0,1545	0,5179	0,3813	-0,7689	-0,4155	-0,4721	-1,0187
-0,1889	0,6045	0,5779	-0,0487	-0,4189	0,5545	0,6579	1,0013
-0,6189	-1,5855	-1,2721	-2,1787	-1,1389	-0,2955	0,0279	-0,7487
-0,6189	0,2545	0,1779	-0,5487	-0,8689	-0,2455	-0,6721	-2,2787
-0,2789	-0,0755	-0,3421	-1,0487	-0,2189	0,4545	0,4279	-0,0287
-0,4189	1,0045	1,0279	0,5913	-0,3689	2,1545	2,5779	2,4013
-0,9889	-1,0955	-1,4221	-1,8987	-0,9589	0,5545	0,9079	1,1013
0,5011	0,0045	0,9779	0,0013	0,0311	0,5545	0,3979	0,0513
-0,6189	0,3345	-0,0921	-1,0387	-0,1689	0,9245	1,4779	1,3313

$\hat{e}_{11}$	$\hat{e}_{12}$	$\hat{e}_{13}$	$\hat{e}_{14}$	$\hat{e}_{11}$	$\hat{e}_{12}$	$\hat{e}_{13}$	$\hat{e}_{14}$
-0,2789	0,6845	0,8779	-0,0487	-0,1543	0,3707	0,8357	-0,1993
0,3611	1,3245	1,2279	0,6513	-0,6643	-0,2793	-0,6143	-0,8693
-0,5189	-0,7955	-0,8721	-1,2987	-0,0143	1,2207	1,4357	1,3507
-0,2189	0,1245	-0,2121	-1,3987	-0,2743	0,8207	0,7657	0,1707
-0,0189	1,0945	1,8979	1,5913	-0,1343	1,2107	0,4857	-0,6493
-0,2389	0,1145	-0,3421	-1,2487	-0,2443	-0,1293	-0,0143	-0,5993
-0,9989	-0,8055	-0,2321	-0,5687	0,2357	0,0707	0,2457	-0,3993
0,0311	0,8045	0,8479	0,3313	-0,9843	-0,0793	-0,9143	-1,1193
0,0411	1,1545	0,7679	0,3513	-0,2143	0,8207	0,4357	0,2007
0,0811	0,5445	0,4279	-0,0487	-0,3543	0,4207	0,3357	-0,5993
0,0511	0,1545	0,0079	-0,7687	0,1857	0,4207	0,1657	0,0007
-0,4189	0,3545	0,1279	-0,3487	-0,5443	0,3507	-0,2543	-0,8293
0,1211	0,4045	0,2679	-0,4487	-0,8943	0,6707	1,5357	1,1607
-0,4189	0,6045	0,2079	-0,4687	0,1057	0,9707	1,4657	1,1207
-0,3689	0,1645	0,1379	-0,6987	-0,6443	-0,2193	-0,0343	-0,4993
-0,6443	-0,8993	0,4857	-0,2493	-0,5043	0,7207	0,5757	-0,0193
0,2057	0,2707	0,4357	-0,2493	-0,4943	-0,2793	-0,3943	-0,7493
-0,0943	0,1707	0,1357	0,0507	-0,2343	1,0207	0,3857	-0,4193
-0,2143	-0,1393	-0,5643	-1,6493	0,1757	1,1207	1,3357	0,4507
-0,3843	0,3207	0,2857	0,0507	-0,6843	0,2707	-0,2443	-0,9393
-0,2043	0,3207	0,5357	-0,3993	-0,6543	0,4007	0,1857	-0,5493
-0,9943	-0,0793	-1,2143	-2,4993	-0,1443	0,6207	0,7357	-0,0493
-0,1843	0,0207	-0,2643	-1,2493	-0,5843	0,5507	1,7557	2,2507
-0,5743	0,6207	0,3857	-0,3993	-0,6643	0,3307	0,3757	-0,3193
-0,8743	1,4207	-0,0143	-0,3993	-0,0443	0,5607	0,1357	-0,6493
-0,0543	0,0907	-0,3143	-1,1993	-0,3743	-0,0293	-0,2643	-0,6493

değerleri elde edilir.

Tüm gözlem değerlerine karşılık gelen T değerleri aşağıdaki formülden yararlanılarak hesaplanırsa:

$$T_i \equiv \frac{e_i' X (X' S X)^{-1} X' e_i}{1 - p_{ii}}$$

T1= 0,0222	T11= 0,0159	T21= 0,0322	T31= 0,0274	T41= 0,0131
T2= 0,0058	T12= 0,0693	T22= 0,0393	T32= 0,0249	T42= 6,9003e-004
T3= 0,0105	T13= 0,0131	T23= 0,0431	T33= 0,0586	T43= 0,0018
T4= 0,3187	T14= 0,0054	T24= 9,0644e-004	T34= 0,0197	T44= 0,0409
T5= 0,0320	T15= 0,2394	T25= 0,1358	T35= 0,0587	T45= 0,0095
T6= 0,0039	T16= 0,0128	T26= 0,0852	T36= 0,0102	T46= 2,6272e-004
T7= 0,0023	T17= 0,0181	T27= 0,0047	T37= 0,0166	T47= 0,0359
T8= 0,0715	T18= 0,0053	T28= -3,4375e-004	T38= 0,0070	T48= 0,0034
T9= 0,0026	T19= 0,0057	T29= 0,0034	T39= 0,0141	T49= 0,0011
T10= 0,0110	T20= 0,0208	T30= 0,0401	T40= 3,0482e-004	T50= 0,0617
T51= 0,0197	T61= 0,0099	T71= 0,0094		
T52= 2,0263e-004	T62= 0,0256	T72= 0,0104		
T53= 7,6885e-005	T63= 0,0035	T73= 0,0267		
T54= 0,0255	T64= 0,0014	T74= 0,0057		
T55= 0,0020	T65= 0,0071	T75= 0,0018		
T56= 0,0158	T66= 0,0043	T76= 0,0038		
T57= 0,0397	T67= 0,0880	T77= 0,1601		
T58= 0,0050	T68= 0,0314	T78= 0,0050		
T59= 0,0274	T69= 0,0128	T79= 0,0144		
T60= 0,0023	T70= 0,0028	T80= 0,0039		

değerleri elde edilir. Buradan, aşağıdaki kısıt altında aykırı gözlem / gözlemler bulunabilir.

$$\frac{mC_{\alpha}^*}{(n-r-2) + mC_{\alpha}^*} = \frac{2(7,02)}{(80-2-2) + 2(7,02)} \cong 0,156$$

$$T_i \geq \frac{mC_{\alpha}^*}{(n-r-2) + mC_{\alpha}^*} \Rightarrow \text{gözlem değeri aykırıdır.}$$

$T_4 = 0,3187 \geq 0,156$, $T_{15} = 0,2394 \geq 0,156$ ve $T_{77} = 0,1601 \geq 0,156$ olduğundan 4.ncü, 15.nci ve 77.nci gözlem değerleri aykırıdır.Yani, 4.ncü, 15.nci ve 77.nci bebeklerin ağırlıklarına ilişkin gözlem değerleri diğer bebeklerin gözlem değerlerinden oldukça farklıdır.



7. SONUÇ ve ÖNERİLER

Bu çalışmada, Büyüme Eğrisi Modelleri 'nde Rao 'nun Basit Kovaryans Yapısı ve bu yapıdan türetilen bazı faydalı kovaryans yapıları incelenerek Büyüme Eğrisi Modelleri üzerinde durulmuştur. Öncelikle bir ön çalışma olarak Anscombe 'nin (1973) veri kümesi üzerinde 'aykırı ve etkili gözlemler nasıl belirlenir?' sorusunun yanıtları bulunmaya çalışılmış, 'Aykırı ve Etkili Gözlemler' başlığı altında incelenmiştir. Çok değişkenli analizde aykırı gözlemler için açıklamalar yapıldıktan sonra sürekli izlenen veri üzerinde Büyüme Eğrisi Modelleri tanıtılmaya çalışılmıştır.

Daha sonra, regresyon katsayısı B ve yayılış bileşeni Σ için En Küçük Kareler ve En Çok Olabilirlik tahminlerinin nasıl bulunması gerektiği üzerinde durularak bu tahmin edicilerin karşılaştırmaları yapılmıştır. Bu karşılaştırmaları sayısal değerler yardımıyla daha açık bir şekilde gösterebilmek için Gazi Üniversitesi Çocuk Hastalıkları Anabilim Dalı 'nda kayıtlı bebeklerin doğum anı, 2. ay, 4. ay ve 6.ay itibariyle ölçülen baş çevresi, boy uzunluğu ve ağırlık değişkenlerinden faydalanılmıştır. Bu değişkenlerden sadece baş çevresi için elde ettiğimiz değerlerden faydalanarak sonuçlarımızı yorumlamaya çalışalım. Baş çevresi için;

$$\hat{B}_{EKK} = \begin{bmatrix} 35,5802 & 35,8405 \\ 1,3047 & 1,4405 \end{bmatrix}$$

ve

$$\hat{B}_{EKO} = \begin{bmatrix} 35,5857 & 35,9107 \\ 1,2615 & 1,4016 \end{bmatrix}$$

$$\hat{\Sigma}_{EKK} = \begin{bmatrix} 1,7206 & 0,4074 & 0,3093 & 0,5985 \\ 0,4074 & 2,1418 & 1,0674 & 0,7496 \\ 0,3093 & 1,0674 & 1,4523 & 0,9287 \\ 0,5985 & 0,7496 & 0,9287 & 1,9749 \end{bmatrix}$$

ve

$$\hat{\Sigma}_{E\zeta OB} = \begin{bmatrix} 1,7515 & 0,3694 & 0,2295 & 0,5077 \\ 0,3694 & 2,2023 & 1,1596 & 0,8557 \\ 0,2295 & 1,1596 & 1,5561 & 0,9775 \\ 0,5077 & 0,8557 & 0,9775 & 1,8508 \end{bmatrix}$$

değerleri elde edilmiştir. Bu iki tahmin edici için regresyon katsayısı $\hat{\beta}$ 'lar arasındaki fark önemsenmeyecek kadar küçük iken yayılış bileşeni $\hat{\Sigma}$ 'lar arasında ise belirgin bir fark olduğu görülmektedir. Bunun nedeni olarak, yayılış bileşeni $\hat{\Sigma}_{E\zeta OB}$ 'nin, Σ 'nın yanlış bir tahmin edicisi olduğu söylenebilir. Ayrıca, regresyon katsayısı $\hat{\beta}_{EKK}$ bağımlı değişken Y 'nin doğrusal olduğu durumlarda, $\hat{\beta}_{E\zeta OB}$ ise bağımlı değişken Y 'nin doğrusal olmadığı durumlarda kullanılır.

Ayrıca, Rao 'nun basit kovaryans yapısı ve bu yapıdan türetilen rastgele regresyon katsayısı ve tekdüze kovaryans yapıları karşılaştırılarak en iyi kovaryans yapısının Rao 'nun basit kovaryans yapısı olduğu aşağıdaki gibi gösterilebilir.

Rao 'nun basit kovaryans yapısı için yayılış bileşeni Σ 'nın tahmininin;

$$\hat{\Sigma} = \begin{bmatrix} 1,7863 & 0,3832 & 0,5109 & 0,7165 \\ 0,3832 & 1,8564 & 0,9493 & 0,5120 \\ 0,5109 & 0,9493 & 1,5556 & 0,9890 \\ 0,7165 & 0,5120 & 0,9890 & 2,0913 \end{bmatrix}$$

biçiminde olduğunu, rastgele regresyon katsayısı yapısı için yayılış bileşeni Σ 'nın tahmininin;

$$\hat{\Sigma} = \begin{bmatrix} 1,7967 & 0,6260 & 0,5334 & 0,4409 \\ 0,6260 & 0,8973 & 0,4140 & 1,0088 \\ 0,5334 & 0,4140 & 1,3727 & 1,5768 \\ 0,4409 & 1,0088 & 1,5768 & 3,2228 \end{bmatrix}$$

biçiminde olduğunu ve son olarak tekdüze kovaryans yapısı için yayılım bileşeni Σ 'nın tahmininin;

$$\hat{\Sigma} = \begin{bmatrix} 1,8224 & 0,6768 & 0,6768 & 0,6768 \\ 0,6768 & 1,8224 & 0,6768 & 0,6768 \\ 0,6768 & 0,6768 & 1,8224 & 0,6768 \\ 0,6768 & 0,6768 & 0,6768 & 1,8224 \end{bmatrix}$$

biçiminde olduğu gösterilmektedir.

Son olarak, aykırı gözlem bulma problemiyle ilgilenildiğinde baş çevresi değişkeni için;

$$T_i \equiv \frac{e_i' X (X' S X)^{-1} X' e_i}{1 - p_{ii}}$$

değerleri hesaplanarak;

$$T_i \geq \frac{mC_{\alpha}^*}{(n - r - 2) + mC_{\alpha}^*}$$

eşitliğini sağlayan 59. gözlem değerinin aykırı gözlem olduğu söylenebilir. Diğer değişkenler için de aynı işlemler uygulandığında boy uzunluğu değişkeni için herhangi bir aykırı gözlem gözlenmemişken ağırlık değişkeni için 4.,15. ve 77. gözlem değerlerinin aykırı olduğunu söyleyebilir.

KAYNAKLAR

1. von Rosen, D., “ Moments for a Multivariate Linear Normal Models with Application to Growth Curve Model ”, *Journal of Multivariate Analysis*, 31 : 187-200 (1990).
2. von Rosen, D., “ The Growth Curve Model: a Review ”, *Communication in Statistics: Theory and Methods*, 20(9) : 2791-2822 (1991).
3. Lee, J. C., “ Prediction and Estimations of Growth Curve with Special Covariance Structure ”, *Journal of American Statistical Association*, 83 : 432-440 (1988a).
4. Keremidas, E. M. and Lee, J. C., “ Forecasting Technological Substitutions with Short Time Series ”, *Journal of American Statistical Association*, 85 : 625- 632 (1990).
5. Lee, J. C., “ Tests and Model Selection for the General Growth Curve Model ”, *Biometrics*, 47 : 147-159 (1991).
6. Pan J. X., Fang K. T., “ Growth Curve Models and Statistical Diagnostics ”, *Springer Beijing*, New York, 55-128 (2000).
7. Laird, N. M. And Ware, J., “ Random Effects Models for Longitudinal Data ”, *Biometrics*, 38 : 963-974 (1982).
8. Huber, P. J., “ Generalities ”, *Robust Statistics*, *John Wiley and Sons*, New York, 1-20 (1981).

9. Cook, R. D., “ Detection of Influential Observation in Linear Regression ”, *Techometrics*, 19 : 15-18 (1977).
10. Barnett, V. And Lewis, T., “ Outliers In Relation to Probability Models ”, Outliers in Statistical Data, *John Wiley and Sons*, New York, 7-10 (1984).
11. Rousseuw, P. J. And Leroy, A., “ Outliers in Regression Analysis ” and “ Outliers Diagnostics ”, Robust Regression and Outlier Detection, *John Wiley and Sons*, New York, 216-242 (1987).
12. Cook, R. D. And Weisberd, S., “ Characterization of an Emprical Influence Function for Detecting Influential Cases in Regression ”. *Technometrics*, 23 : 59-64 (1980).
13. Cook, R. D. And Weisberd, S., “ Assessment of Influence ”. Residual and Influence in Regression, *Chapman and Hall*. New York. 101-156 (1982).
14. Gnanadesikan, R., “ Multivariate Observation ”. Methods for Statistical Data Analysis of Multivariate Observation, *John Wiley and Sons*, New York, 208-231 (1977).
15. Rousseeuw, P.J. and van Zomeren, B.C., “ Unmasking Multivariate Outliers and Leverage Points ”. *Journal of the American Statistical Association*, 85 : 633-651 (1990).

16. Woodruff, D.L. and Rocke, D.M., “ Computable Robust Estimation of Multivariate Location and Shape in High Dimension Using Compound Estimators ”, *Journal of the American Statistical Association*, 89 : 888-896 (1994).
17. Davies, L., “ The Asymptotics of Rousseeuw’s Minimum Value Ellipsoid estimator ”, *Annals of Statistics*, 14 : 1828-1843 (1992).
18. Pan, J.X., Fung, W.K. and Fang, K.T., “ Multiple outlier detection in multivariate data using projection pursuit techniques ”, *Journal of Statistical Inference and Planning*, 83 : 153-167 (2000).
19. Sinha, B.K., “ Detection of multivariate outliers in elliptically symmetric distributions ”, *Annals of Statistics*, 12 : 1558-1565 (1984).
20. Das, R. And Sinha, B.K., “ Detection of The Multivariate Outliers with Dispersion Slippage in Elliptically Symmetric Distribution ”, *Annals of Statistics*, 14 : 1619-1624 (1986).



EKLER

EK-1

1. Matris Değişkenli Normal ve t Dağılımları

Matris değişkenli normal ve t dağılımları matrissel simetrik dağılımlar sınıfında oldukça sık kullanılan dağılımlardan iki tanesidir. Bu dağılımlar büyüme eğrisi modelleri araştırmasında oldukça önemli rol oynar.

Tanım 1.1.

$(p \times n)$ boyutlu rastgele X matrisi matris değişkenli normal dağılıma uygunluk göstermesi için olabilirlik yoğunluk fonksiyonu aşağıdaki biçimde olmalıdır:

$$f(X) = (2\pi)^{-pn/2} [\det(\Sigma)]^{-n/2} [\det(\Omega)^{-p/2} \exp(-\frac{1}{2} \text{tr}\{\Omega^{-1}(X-M)'\Sigma^{-1}(X-M)\})].$$

Burada M , $\Sigma > 0$, $\Omega > 0$ ve boyutları sırasıyla $(p \times n)$, $(p \times p)$ ve $(n \times n)$ olan matrisleri göstermektedir. Bu durumda X , $X \sim N_{p,n}(M, \Sigma, \Omega)$ biçiminde gösterilir.

$x \sim N_p(\mu, \Sigma)$ şeklinde ifade edilen vektör değişkenli normal dağılım temel alındığında, aşağıdaki önerme: matris değişkenli ve vektör değişkenli normal dağılımlar arasındaki ilişkiyi verir.

Önerme 1.1.

X , $(p \times n)$ boyutlu rastgele matris ve $x = \text{vec}(X)$ olmak üzere ancak ve ancak $x \sim N_{pn}(\text{vec}(M), \Omega \otimes \Sigma)$ ise $X \sim N_{p,n}(M, \Sigma, \Omega)$ 'dır.

Önerme 1.1. , $N_{p,n}(M, \Sigma, \Omega)$ normal dağılımın beklenen değeri ve kovaryansının sırasıyla $E(X) = M$ ve $\text{cov}(X) = \text{cov}(\text{vec}(X)) = \Omega \otimes \Sigma$ olduğunu ima etmektedir. Bunun yanı sıra, vektör değişkenli normal dağılım durumuna benzer olarak matris

değişkenli normal değişkenin dönüşümü de matris değişkenli normal dağılıma sahip olur.

Önerme 1.2.

Varsayalım ki, $X \sim N_{p,n}(M, \Sigma, \Omega)$ ve $C, A > 0$ ve $B > 0$ boyutları sırasıyla $(p \times n)$, $(p \times p)$ ve $(n \times n)$ olarak verilen matrisler olsun. $AXB + C \sim N_{p,n}(AMB + C, A\Sigma A', B\Omega B')$ 'dir.

Önerme 1.2. aşağıdaki gibi ifade edilebilir:

$$\Sigma^{-1/2}(X - M) \sim N_{p,n}(0, I_p, \Omega), \quad [1.1]$$

$$(X - M)\Omega^{-1/2} \sim N_{p,n}(0, \Sigma, I_n), \quad [1.2]$$

$$\Sigma^{-1/2}(X - M)\Omega^{-1/2} \sim N_{p,n}(0, I_p, I_n). \quad [1.3]$$

Burada, $\Sigma^{-1/2}$ ve $\Omega^{-1/2}$ ifadeleri Σ ve Ω 'nın karekök matrisleridir. $Y = \Sigma^{-1/2}(X - M)$ matrisi için: [1.1], Y 'nin satırlarının ortak dağılımı $N_p(0, \Omega)$ olan bağımsız aynı dağılımlı rastgele vektörler olduklarını ifade etmektedir. Ayrıca [1.2], $Z = (X - M)\Omega^{-1/2}$ 'in sütunlarının ortak dağılımı $N_p(0, \Sigma)$ olan bağımsız aynı dağılımlı rastgele vektörler olduklarını ifade etmektedir. Son olarak [1.3], $\Sigma^{-1/2}(X - M)\Omega^{-1/2}$ matrisinin öğelerinin $N(0,1)$ dağılımına sahip bağımsız aynı dağılımlı rastgele değişkenler olduğunu ifade etmektedir. Bundan dolayı, $\text{vec}\{\Sigma^{-1/2}(X - M)\Omega^{-1/2}\} \sim N_{pn}(0, I_{pn})$ 'dir. Diğer bir ifadeyle:

$$(\Omega^{-1/2} \otimes \Sigma^{-1/2})\text{vec}(X - M) \sim N_{pn}(0, I_{pn})$$

biçimindedir. Böylece ;

$$\text{vec}'(X - M)(\Omega^{-1} \otimes \Sigma^{-1})\text{vec}(X - M) \sim \chi_{pn}^2$$

yani, pn serbestlik dereceli ki-kare dağılımına sahip olduğunu buluruz. Bu sonuç $iz\{(X - M)\Omega^{-1}(X - M)'\Sigma^{-1}\} \sim \chi_{pn}^2$ ifadesine denktir. Böylece, aşağıdaki önermeyi elde etmiş oluruz.

Önerme 1.3.

$X \sim N_{p,n}(M, \Sigma, \Omega)$ olmak üzere;

$$iz\{(X - M)\Omega^{-1}(X - M)'\Sigma^{-1}\} = vec'(X - M)(\Omega^{-1} \otimes \Sigma^{-1})vec(X - M) \sim \chi_{pn}^2 \quad [1.4]$$

biçimindedir.

Tanım 1.2.

Eğer (pxn) boyutlu X matrisi $N_{p,n}(0, \Sigma, I_n)$ biçiminde dağılmak üzere $A=XX'$ ise A 'nın kovaryans matrisi $\Sigma > 0$ olan n serbestlik dereceli Wishart dağılımına sahip olduğu söylenir. Bu durumda A 'yı, $A \sim W_p(n, \Sigma)$ şeklinde gösteririz.

Bu tanıma göre, [1.1] ve [1.2]'nin yardımıyla aşağıdaki önermeyi hemen bulabiliriz.

Önerme 1.4.

$X \sim N_{p,n}(M, \Sigma, \Omega)$ olmak üzere;

$$\begin{aligned} (X - M)\Omega^{-1}(X - M)' &\sim W_p(n, \Sigma) \\ (X - M)'\Sigma^{-1}(X - M) &\sim W_n(p, \Omega) \end{aligned}$$

biçimindedir.

Wishart dağılımı $W_p(n, \Sigma)$ için eğer $n \geq p$ ise olasılık yoğunluk fonksiyonu mevcuttur. Aşağıdaki önerme bu ifadeyi açık bir şekilde göstermektedir.

Önerme 1.5.

$A \sim W_p(n, \Sigma)$ ve $n \geq p$ olmak üzere A 'nın olasılık yoğunluk fonksiyonu aşağıdaki gibidir:

$$f(A) = 2^{-mn/2} [\Gamma_m(\frac{n}{2})]^{-1} [\det(\Sigma)]^{-n/2} [\det(A)]^{(n-m-1)/2} \exp\{-\frac{1}{2} \text{tr}(\Sigma^{-1} A)\}. \quad [1.5]$$

Burada $\Gamma_m(\cdot)$, çok değişkenli gamma fonksiyonunu gösterir ve $\Gamma_m(\cdot)$ aşağıdaki gibidir:

$$\Gamma_m(\alpha) = \pi^{m(m-1)/4} \prod_{i=1}^m \Gamma(\alpha - \frac{1}{2}(i-1)).$$

Bu durumda A 'yı $A \sim W_m(n, \Sigma)$ biçiminde gösterebiliriz.

Eğer $A \sim W_m(n, \Sigma)$ biçiminde ise A 'nın beklenen değeri $E(A) = n\Sigma$ 'dır. Ayrıca literatürde Wishart dağılımı üzerinde değişik özellikler vardır.

Önerme 1.6.

(mxm) boyutlu rastgele $A_1, A_2, \dots, A_r$ matrisleri karşılıklı bağımsız ve

$A_i \sim W_m(n_i, \Sigma)$, $i=1, 2, \dots, r$ olmak üzere $\sum_{i=1}^r A_i \sim W_m(n, \Sigma)$ 'dır. Burada $n = \sum_{i=1}^r n_i$ 'dir.

EK-2*Kronecker Çarpımı ve vec(.) Operatörleri İle İlgili Bazı Özellikler*

1) $\text{vec}(\cdot)$ İşlemi:

$$\text{vec}(A) = \begin{bmatrix} a_{11} \\ \vdots \\ a_{1M} \\ a_{21} \\ \vdots \\ a_{2M} \\ \vdots \\ a_{M1} \\ \vdots \\ a_{MM} \end{bmatrix}$$

Eğer A matrisi $M \times M$ boyutlu bir matris ise, $\text{vec}(\cdot)$ işlemi boyut indirgemesi yaparak A matrisini $M^2 \times 1$ boyutlu vektöre çevirir. Bu durum bize matris işlemlerinde kolaylık sağlar.

2) Kronecker Çarpımı:

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdot & \cdot & \cdot & a_{1M}B \\ a_{21}B & a_{22}B & \cdot & \cdot & \cdot & a_{2M}B \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{M1}B & a_{M2}B & \cdot & \cdot & \cdot & a_{MM}B \end{bmatrix}$$

Eğer A matrisi $M \times M$ boyutlu ve B matrisi de $N \times N$ boyutlu matrisler ise, $A \otimes B$ kronecker çarpımı $NM \times NM$ boyutlu matris halini alır. Örneğin;

$$H_3 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}_{3 \times 3} \quad \text{ve} \quad B = \begin{bmatrix} 1 & -\alpha \\ -\alpha & 1 \end{bmatrix}_{2 \times 2} \quad \text{ise};$$

$$H_3 \otimes B = \begin{bmatrix} 0 & 0 & 1 & -\alpha & 0 & 0 \\ 0 & 0 & -\alpha & 1 & 0 & 0 \\ 1 & -\alpha & 0 & 0 & 1 & -\alpha \\ -\alpha & 1 & 0 & 0 & -\alpha & 1 \\ 0 & 0 & 1 & -\alpha & 0 & 0 \\ 0 & 0 & -\alpha & 1 & 0 & 0 \end{bmatrix}_{6 \times 6}$$

biçimindedir. Yukarıda da görüldüğü üzere 3×3 boyutlu H_3 matrisi ile 2×2 boyutlu B matrisinin kronecker çarpımı 6×6 ($M=3, N=2$) boyutlu matris biçimini almıştır.

$$3) \quad A \otimes (B \otimes C) = (A \otimes B) \otimes C$$

$$4) \quad \begin{aligned} A \otimes (B + C) &= (A \otimes B) + (A \otimes C) \\ (A + B) \otimes C &= (A \otimes C) + (B \otimes C) \end{aligned}$$

5) a, sabit bir sayı olmak üzere;

$$a \otimes A = A \otimes a = aA$$

dır.

6) a ve b sabit sayılar olmak üzere;

$$aA \otimes bB = abA \otimes B$$

dır.

7) Matrisler için;

$$(A \otimes B)(C \otimes D) = AC \otimes BD$$

dır.

$$8) (A \otimes B)^T = A^T \otimes B^T, (A \otimes B)^H = A^H \otimes B^H$$

9) a ve b vektör olmak üzere;

$$a^T \otimes b = ba^T = b \otimes a^T$$

dır.

10) Parçalı matrisler için;

$$[A_1, A_2] \otimes B = [A_1 \otimes B, A_2 \otimes B]$$

eşitliği geçerlidir. Fakat;

$$A \otimes [B_1, B_2] \neq [A \otimes B_1, A \otimes B_2]$$

dır.

11) Tekil olmayan karesel matrisler A ve B için;

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$$

dır.

12) $m \times m$ boyutlu matris A ve $n \times n$ boyutlu matris B için;

$$|A \otimes B| = |A|^n |B|^m$$

dır.

$$13) \text{iz}(A \otimes B) = \text{iz}(A) \text{iz}(B)$$

$$14) \text{rank}(A \otimes B) = \text{rank}(A) \text{rank}(B)$$

$$15) \text{iz}(ABC) = \text{vec}(A^T)^T (I \otimes B) \text{vec}(C)$$

dır. Ayrıca;

$$\text{iz}(AB) = \text{vec}(A^T)^T \text{vec}(B)$$

ve

$$\text{iz}(A^T BCD^T) = \text{vec}(A)^T (D \otimes B) \text{vec}(C)$$

dır.

EK-3*Q Ortogonal Matrisinin Bulunması*

X matrisinin ortogonalı olan Q matrisinin bulunuşu aşağıdaki gibidir.

$X' = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 4 & 6 \end{bmatrix}$ olmak üzere $X'Q = 0$ eşitliğinden (4x2) boyutlu Q matrisini;

$$Q = \begin{bmatrix} a & b \\ c & d \\ e & f \\ g & h \end{bmatrix}$$

biçiminde gösterelim. $X'Q = 0$ matris çarpımını açarsak aşağıdaki dört eşitliği yazabiliriz:

$$a+c+e+g=0$$

$$2c+4e+6g=0$$

$$b+d+f+h=0$$

$$2d+4f+6h=0$$

Buradan, matrislerin eşelon form özelliklerinden yararlanılarak çözüme kolaylıkla ulaşılır.

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & 0 & 4 & 0 & 6 & 0 \\ 0 & 0 & 0 & 2 & 0 & 4 & 0 & 6 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Yukarıdaki (4x8) boyutlu matrisin herbir sütunu sırasıyla a, b, c, d, e, f, g, h bilinmeyenlerine karşılık gelmektedir.

8 bilinmeyenli 4 denklemden 4 faktör bulunur. Bu faktörler e, f, g, h bilinmeyenlerine karşılık gelmektedir. Amaç, ilk 4 bilinmeyene karşılık gelen sütunların oluşturduğu kısmi birim matris yapmaktır. Bunun için gerekli satır işlemleri aşağıda verilmiştir.

Birinci satır için; üçüncü satırın (-1/2) katı ile birinci satır toplanır.

İkinci satır için; dördüncü satırın (-1/2) katı ile ikinci satır toplanır.

Üçüncü satır için; (1/2) ile çarpılır.

Dördüncü satır için; (1/2) ile çarpılır.

Bu işlemlerin sonucunda aşağıdaki matris elde edilir.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & -1 & 0 & -2 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & -2 \\ 0 & 0 & 1 & 0 & 2 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 & 0 & 2 & 0 & 3 \end{bmatrix}$$

Bulunan matris $\begin{bmatrix} a & b & c & d & e & f & g & h \end{bmatrix}$ vektörü ile çarpılırsa;

$$\begin{bmatrix} 1 & 0 & 0 & 0 & -1 & 0 & -2 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & -2 \\ 0 & 0 & 1 & 0 & 2 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 & 0 & 2 & 0 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$a - e - 2g = 0$$

$$c + 2e + 3g = 0$$

$$b - f - 2h = 0$$

$$d + 2f + 3h = 0$$

eşitlikleri elde edilir. Eşitlikleri kullanarak çözüme ulaşmak için denklemler aşağıdaki uygun hale dönüştürülür.

$$a = e + 2g$$

$$c = -2e - 3g$$

$$b = f + 2h$$

$$d = -2f - 3h$$

Bu dört eşitlikten (e, f, g, h) bilinmeyenleri yerine keyfi olarak (0, 1, 1, 1) alınarak X matrisinin ortogonali olan Q matrisi aşağıdaki gibi bulunur. Burada, Q matrisi tek değildir.

$$Q = \begin{bmatrix} 2 & 3 \\ -3 & -5 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

EK-4

Gazi Üniversitesi Çocuk Hastalıkları Anabilim Dalından alınan veriler aşağıda verilmektedir. 0: kız bebekleri, 1: erkek bebekleri göstermektedir.

Baş çevresi için ölçüm değerleri santimetre cinsindendir.

Cins	Doğum Baş Çevresi	2 Aylık	4 Aylık	6 Aylık
0	33	39	38	43
0	34	38	41	43
0	34	39	41	44
0	37	42	44	46
0	36	40	41	43
0	37	40	41	44
0	38	39,5	42,5	45
0	33	35	38	39
0	33	40	40	42
0	36	38,5	41,5	43
0	34	40	42	43
0	35	38	41	43
0	35	37	41	43
0	38	39,5	40	42
0	34	40	42	44
0	35	38	42	43
0	33,5	38	41	41,5
0	36	38	40	42
0	33	39,5	41,5	44
0	34,5	39	41	44
0	35	39	41	43
0	35	37	40	41
0	35	37	39	41
0	36	39	41	43,5
0	35	39	41,5	42
0	34,5	40	44,5	45,5
0	34	38	41	41
0	36	41	43,5	45,5
0	36	38,5	41	43,5
0	36	40	42,5	43,5
0	36	37,5	41	43
0	34,5	39	40	42,5
0	36	40	43	43
0	35,5	40	41	43
0	35	37	39	43
0	35	40	42	43,5
0	35,5	40	42,5	42,5

0	35,5	37,5	41	43
0	36	40	40,5	42
0	34	37	40,5	42,5
0	35,5	38	43	44
0	34	38	41	42,5
0	34	38,5	40,5	43
1	35	38	42	44
1	36	40	44	48
1	38	40,5	43,5	45
1	34,5	37,5	41	42
1	35	39	44	47
1	35	38	43	43,5
1	35	38	41	42
1	35	42	43	44
1	34	39	42	44
1	33	37	41	42,5
1	36	40	43	46
1	35,5	41	43	44
1	36	38	41	43
1	36	39	42	46
1	36	38	40,5	43
1	40	42	42,5	43
1	36	36,5	42	44,5
1	33,5	38	41	42,5
1	34,5	38,5	41,5	44
1	35	38,5	41,5	42
1	36	39	41	44,5
1	36	40	43,5	44
1	36	39	41	43
1	33,5	41	44	46
1	36	40	43	46
1	36	39	42	44,5
1	35	40	42	45
1	36	40	43	45,5
1	37	40	42,5	44
1	34	38	42	44
1	35,5	40	41	43
1	34	38	40	42
1	36	40	41	44
1	35,5	42	42	44,5
1	35	37	42	44
1	35,5	37	41,5	42
1	35,5	38	41,5	44,5

Doğum uzunluğu için ölçüm değerleri santimetre cinsindendir,

Cins	Doğum Uzunluğu	2 Aylık	4 Aylık	6 Aylık
0	48	57	65	70
0	50	58	65	67
0	48	57	63	67
0	51	62	67	73
0	49	59	65	68
0	51	60	66	71
0	50	56	63	67,5
0	49	51,5	59	61
0	52	55	62	64,5
0	51	59,5	63	67
0	50	61,5	67	72
0	46	54	58,5	63
0	54	59	68	71
0	51	57	62	66
0	48	56	62	69
0	49	53	61	67,5
0	50	54	61	68
0	50	56	64	67
0	51	58,5	62	67,5
0	47	56	65	69
0	48	57	63	71
0	48	54	62	69
0	48,5	53	60	63,5
0	48	54	63	66
0	50	59	69	70
0	48	59,5	63	69
0	50	61	65	69
0	52	58	64	71
0	52	57	65	65
0	53	61	67	70
0	50	55	62,5	69
0	50	57	65	73
0	51	61	68	74,5
0	51	57	65	67,5
0	46	52	62	62
0	51	60	64,5	66
0	50	62	69,5	73,5
0	52	58	63,5	69
0	51	54	63	66
0	49	61,5	64	68
0	50	60	65	67
0	50	58	63	66
0	51	60	64	68
1	50	59	66	70
1	55	62	68	70

1	52	61	67	70
1	50	59	66	68
1	49	57	64	67
1	51	58,5	69	73
1	48	52	57	63
1	52	59	65	67
1	49	59	63	65
1	46	52	56	64
1	53	60	66	70
1	49	61	68	68
1	47	53	62	67,5
1	49	59	69	71
1	51	60	65	74
1	52	60	65	68
1	52	56,5	65	70,5
1	48	57	66	68
1	46	56	63,5	68,5
1	54	63	68	72
1	50	56	64	70
1	52	60	67,5	74
1	50	55	64	69
1	49	58,5	66	69,5
1	50	63	70	73
1	50	53	66	75
1	48	55	63	72
1	50	60	63	69
1	51	64	67,5	71
1	52	64	65,5	72
1	50	59	64	67
1	50	60	65	67
1	51	57	64	69
1	50	59	66	71,5
1	46	58	64	70
1	52	61	70	72
1	48	57	63	67

Doğum ağırlığı için ölçüm değerleri gram cinsindendir,

Cins	Doğum Ağırlığı	2 Aylık	4 Aylık	6 Aylık
0	3080	5000	7400	8200
0	3200	5230	6500	7500
0	3350	4750	6400	7370
0	6060	6800	9300	9700
0	4000	6000	7400	7900
0	3960	5400	7290	8680
0	3530	5850	7350	8250
0	3100	3660	5500	6120
0	3100	5500	6950	7750
0	3440	5170	6430	7250
0	3300	6250	7800	8890
0	2730	4150	5350	6400
0	4220	5250	7750	8300
0	3100	5580	6680	7260
0	3810	5500	6950	8100
0	3020	5050	6410	7390
0	3080	4850	6360	7730
0	3300	5000	6800	7800
0	3250	5600	6600	7380
0	2950	4830	6300	7280
0	3300	5800	7430	9300
0	2580	4950	6800	7550
0	2850	5000	6100	6020
0	3500	5700	7200	8270
0	3350	7400	9350	10700
0	2760	5800	7680	9400
0	3750	5800	7170	8350
0	3550	6170	8250	9630
0	3440	5930	7650	8250
0	4080	6570	8000	8950
0	3200	4450	5900	7000
0	3500	5370	6560	6900
0	3700	6340	8670	9890
0	3480	5360	6430	7050
0	2720	4440	6540	7730
0	3750	6050	7620	8630
0	3760	6400	7540	8650
0	3800	5790	7200	8250
0	3770	5400	6780	7530
0	3300	5600	6900	7950
0	3840	5650	7040	7850
0	3300	5850	6980	7830
0	3350	5410	6910	7600
1	3100	4480	7500	8400
1	3950	5650	7450	8400

1	3650	5550	7150	8700
1	3530	5240	6450	7000
1	3360	5700	7300	8700
1	3540	5700	7550	8250
1	2750	5300	5800	6150
1	3560	5400	6750	7400
1	3170	6000	7400	8250
1	2870	6800	7000	8250
1	3690	5470	6700	7450
1	3590	5750	7850	8450
1	3080	5100	6400	7780
1	3730	6600	8450	10000
1	3470	6200	7780	8820
1	3610	6590	7500	8000
1	3500	5250	7000	8050
1	3980	5450	7260	8250
1	2760	5300	6100	7530
1	3530	6200	7450	8850
1	3390	5800	7350	8050
1	3930	5800	7180	8650
1	3200	5730	6760	7820
1	2850	6050	8550	9810
1	3850	6350	8480	9770
1	3100	5160	6980	8150
1	3240	6100	7590	8630
1	3250	5100	6620	7900
1	3510	6400	7400	8230
1	3920	6500	8350	9100
1	3060	5650	6770	7710
1	3090	5780	7200	8100
1	3600	6000	7750	8600
1	3160	5930	8770	10900
1	3080	5710	7390	8330
1	3700	5940	7150	8000
1	3370	5350	6750	8000

ÖZGEÇMİŞ

Ramazan Arslan, 10.03.1977 tarihinde Ankara 'da doğdu. İlkokulu Ankara Balgat İlkokulu 'nda, orta öğrenimini ise Ankara Cumhuriyet Lisesi 'nde tamamladı. 24.06.2002 tarihinde Gazi Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü 'nden mezun oldu. 19.09.2002 tarihinde Gazi Üniversitesi, Fen Bilimleri Enstitüsü Yüksek Lisans Programına kaydoldu.

