



EN İYİ DOĞRUSAL KARIŞIK MODELİN SEÇİMİ

Ebru BİLGİN

Doktora Tezi
Zootekni Ana Bilim Dalı
Prof. Dr. Memiş ÖZDEMİR
2022
(Her hakkı saklıdır.)

T.C.
ATATÜRK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ZOOTEKNİ ANA BİLİM DALI

EN İYİ DOĞRUSAL KARIŞIK MODELİN SEÇİMİ
(The Selection of the Best Linear Mixed Model)

DOKTORA TEZİ

Ebru BİLGİN

Danışman: Prof. Dr. Memiş ÖZDEMİR

Erzurum
Ekim, 2022

KABUL VE ONAY TUTANAĞI

Ebru BİLGİN tarafından hazırlanan “**EN İYİ DOĞRUSAL KARIŞIK MODELİN SEÇİMİ**” başlıklı çalışması 21/10/2022 tarihinde yapılan tez savunma sınavı sonucunda **başarılı** bulunarak jürimiz tarafından **Zootekni** Ana Bilim Dalı, **Biyometri ve Genetik** Bilim Dalında **Doktora** tezi olarak kabul edilmiştir.

Jüri Başkanı:	Prof. Dr. Mehmet TOPAL <i>Amasya Üniversitesi</i>	Aslı ıslak imzalıdır
Jüri Üyesi:	Doç. Dr. Hamit MİRTAĞIOĞLU <i>Bitlis Eren Üniversitesi</i>	Aslı ıslak imzalıdır
Jüri Üyesi:	Prof. Dr. Murat AYDIN <i>Atatürk Üniversitesi</i>	Aslı ıslak imzalıdır
Jüri Üyesi:	Doç. Dr. Sinan KOPUZLU <i>Atatürk Üniversitesi</i>	Aslı ıslak imzalıdır
Danışman:	Prof. Dr. Memiş ÖZDEMİR <i>Atatürk Üniversitesi</i>	Aslı ıslak imzalıdır

Enstitü Yönetim Kurulunun/.../.... tarih ve sayılı kararı.

Bu tezin Atatürk Üniversitesi Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili maddelerinde belirtilen şartları yerine getirdiğini onaylarım.

Prof. Dr. Saltuk Buğrahan CEYHUN

Enstitü Müdürü

Aslı ıslak imzalıdır

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

ETİK BİLDİRİM VE İNTİHAL BEYAN FORMU

Doktora Tezi olarak Prof. Dr. Memiş ÖZDEMİR danışmanlığında sunulan “ **EN İYİ DOĞRUSAL KARIŞIK MODELİN SEÇİMİ**” başlıklı çalışmanın tarafımızdan bilimsel etik ilkelere uyularak yazıldığını, yararlanılan eserlerin kaynakçada gösterildiğini, Fen Bilimleri Enstitüsü tarafından belirlenmiş olan Turnitin Programı benzerlik oranlarının aşılmadığını ve aşağıdaki oranlarda olduğunu beyan ederiz.

Tez Bölümleri	Tezin Benzerlik Oranı (%)	Maksimum Oran (%)
Giriş	4	30
Kuramsal Temeller	22	30
Materyal ve Yöntem	28	35
Bulgular ve Tartışma	11	20
Sonuç	0	20
Tezin Geneli	15	25

***Not:** Yedi kelimeye kadar benzerlikler ile Başlık, Kaynakça, İçindekiler, Teşekkür, Dizin ve Ekler kısımları tarama dışı bırakılabilir. Yukarıdaki azami benzerlik oranları yanında tek bir kaynaktan olan benzerlik oranlarının %5'den büyük olmaması gerekir.*

Beyan edilen bilgilerin doğru olduğunu, aksi halde doğacak hukuki sorumlulukları kabul ve beyan ederiz.

Tez Yazarı (Öğrenci)	Tez Danışmanı
Ebru BİLGİN	Prof. Dr. Memiş ÖZDEMİR
21.10.2022	21.10.2022
Aslı ıslak imzalıdır	Aslı ıslak imzalıdır

* Tez ile ilgili YÖKTEZ'de yayınlamasına ilişkin bir engelleme var ise aşağıdaki alanı doldurunuz.

☐ Tezle ilgili patent başvurusu yapılması / patent alma sürecinin devam etmesi sebebiyle Enstitü Yönetim Kurulunun/....../.... tarih ve sayılı kararı ile teze erişim 2 (iki) yıl süreyle engellenmiştir.

☐ Enstitü Yönetim Kurulunun/....../.... tarih ve sayılı kararı ile teze erişim 6 (altı) ay süreyle engellenmiştir.

TEŞEKKÜR

Doktora tez çalışmam süresince desteğini esirgemeyen, tez konumun seçilmesinde ve yürütülmesinde yol gösteren danışmanım Sayın Prof. Dr. Memiş ÖZDEMİR'e, ve tez izleme komitesinde yer alan Doç. Dr. Sinan KOPUZLU'ya ve Prof. Dr. Murat AYDIN'a değerli katkılarından dolayı teşekkürü bir borç bilirim. Ayrıca bu süreçte her zaman yanımda olan eşime ve kızım Naz'a gösterdikleri hoşgörü ve sabır için minnettarım.

Ebru BİLGİN



ÖZET

DOKTORA TEZİ EN İYİ DOĞRUSAL KARIŞIK MODELİN SEÇİMİ

Ebru BİLGİN

Danışman: Prof. Dr. Memiş ÖZDEMİR

Amaç: Doğrusal karışık model, boylamsal verinin analizinde yaygınca kullanılmaktadır. Doğrusal karışık model için, R^2_{adj} , PRESS ve CCC_{adj} kullanılmaktadır. Bu çalışma, simülasyon vasıtasıyla R^2_{adj} , PRESS, CCC_{adj} , F-testi ve iki enformasyon kriterinin (AIC ve BIC) uyum-iyiliği bakımından performanslarını doğrusal karışık model çerçevesi içerisinde incelemek amacıyla yürütülmüştür. Bu tahminsel istatistiklere ait marjinal ve koşullu yaklaşımlar, farklı varyans-kovaryans strüktürleri altında çalışılmıştır.

Yöntem: Çalışmada materyal olarak simülasyon çalışmasına ait üretilen ham veriler kullanılmıştır. Simülasyon çalışması için kullanılacak yöntem betimlenmiştir. Önce, aritmetik ortalamalar için spesifikleştirilen bir model ile ve aynı birey üzerinde bilinen bir kovaryans strüktürüne göre korelasyonlu yinelenen ölçümler ile veri üretilmiştir. Daha sonra, iki karışık model (bir tam model ve bir indirgenmiş model) veriye uydurulmuş ve modellerden birini seçmek için farklı prediktif yaklaşımlar kullanılmıştır. Doğru modelin seçildiği denemelerin oranısı kaydedilmiştir.

Bulgular: Simülasyon çalışması sonucunda, bileşik simetri strüktürü için marjinal ve koşullu R^2 , PRESS ve CCC ile beraber F-testi, AIC ve BIC istatistiklerinin on yedi versiyonunun başarı oranları yüksek bulunmuştur. R^2 , PRESS ve CCC için, marjinal kalıntıların kullanılmasının koşullu kalıntılara göre daha önerilebilir olduğu görülmüştür. AIC ve BIC için, $\frac{1}{2}\log|\sum_{i=1}^m X_i'X_i|$ terimli REML fonksiyonunun kullanılması istatistiklerin performansını artırmıştır. R^2 , CCC ve enformasyon kriterleri için, çeşitli parametre sayısı düzeltmelerinin bir etkisi görülmemiştir.

Birinci mertebeden otoregresif ve rastgele etki için, bu çalışmada görülmüştür ki R^2 ve CCC için marjinal kalıntıların, PRESS için koşullu kalıntıların ve AIC ve BIC için $\frac{1}{2}\log|\sum_{i=1}^m X_i'X_i|$ terimli REML fonksiyonunun kullanılması istatistiklerin performansını artırmıştır. Aynı zamanda, parametre sayısı için yapılan çeşitli düzeltmelerin bir etkisi gözlenmemiştir.

F-testi her kovaryans strüktürü için iyi performans göstermiştir. Bunun yanı sıra, verinin kovaryans strüktürü, parametre değerleri ve örnek hacmi gibi özelliklerinin bu çalışmadaki uyum-iyiliği istatistiklerinin başarısına büyük etkisi olduğu görülmüştür.

Sonuç: Bu simülasyon çalışmasında, bu kriterlerden hiç biri, her durumda, model seçimi performansı bakımından diğerlerine üstünlük sağlayabilmiş değildir.

Anahtar Kelimeler: Doğrusal karışık model, Model seçimi, Model seçimi kriterleri, Simülasyon

Ekim 2022, 94 sayfa

ABSTRACT

Ph. D. THESIS

THE SELECTION OF THE BEST LINEAR MIXED MODEL

Ebru BİLGİN

Supervisor: Prof. Dr. Memiş ÖZDEMİR

Purpose: Linear mixed models are broadly used to analyze longitudinal data. For linear mixed model, R^2_{adj} , PRESS and CCC_{adj} have been used. This study was carried out to investigate the performances of R^2_{adj} , PRESS, CCC_{adj} , F -test and two information criteria (AIC and BIC) with regard to goodness-of-fit via simulation in the frame of linear mixed model. Marginal and conditional approaches of this predictive statistics were studied under different variance-covariance structures.

Method: In the study, raw data produced from the simulation study were used as material. The method to be used for the simulation study is described. First, data was generated with a model specific for arithmetic means and correlated repeated measurements on the same individual against a known covariance structure. Then, two mixed models (a full model and a reduced model) were fitted to the data and different predictive approaches were used to select one of the models. The proportion of trials in which the correct model was selected was recorded.

Findings: In the result of the simulation study, for the compound symmetry structure, the performances of the seventeen versions of the statistics were found high, namely marginal and conditional R^2 , PRESS and CCC along with F -test, AIC and BIC statistics. For R^2 , PRESS and CCC, it was observed that to use marginal residuals are more suggestible than conditional residuals. For AIC and BIC, to use REML function with the term $\frac{1}{2}\log|\sum_{i=1}^m X_i'X_i|$ were increased the performance of the statistics. For R^2 , CCC and information criteria, there were no effect of various parameter number adjustments.

For the first order autoregressive and random effect, in this research, it was seen that to use marginal residuals for R^2 and CCC, to use conditional residuals for PRESS and to use the REML function with the term $\frac{1}{2}\log|\sum_{i=1}^m X_i'X_i|$ for AIC and BIC have increased the performances of the statistics. Also, various adjustments on the number of parameters have had no apparent effects.

F -test have had good performances with every covariance structures. On the other hand, it was observed that certain characteristics of the data such as covariance structure, values of parameters and sample size could be very effective on the performances of the goodness-of-fit statistics.

Results: this simulation study showed that none of the criterion investigated were superior to the others in every situation at all.

Keywords: Linear mixed model, Model selection, Model selection criteria, Simulation

October 2022, 94 pages

İÇİNDEKİLER

KABUL VE ONAY TUTANAĞI.....	i
ETİK BİLDİRİM VE İNTİHAL BEYAN FORMU	ii
TEŞEKKÜR	iii
ÖZET	iv
ABSTRACT	v
İÇİNDEKİLER.....	vi
TABLolar DİZİNİ.....	viii
ŞEKİLLER DİZİNİ	ix
SİMGELER VE KISALTMALAR DİZİNİ	x
GİRİŞ.....	1
Amaç	2
KURAMSAL TEMELLER.....	1
Tek-değişkeli Doğrusal Model.....	1
Yinelemeli Ölçümler İçin Doğrusal Karışık Model.....	1
Doğrusal Karışık Modellerde Model Seçimi	2
AIC ve BIC	2
Pseudo F-testleri ve olabilirlik oranı testleri	4
Tahminsel yaklaşımlar	5
Doğrusal Karışık Modeller.....	7
Model tanımı	7
Sabit etkiler modeli	8
Karışık model.....	9
Rastgele etkiler modeli kovaryans yapısı	10
Rastgele katsayılar modeli kovaryans yapısı	13
Kovaryans biçimi modeli kovaryans yapısı	15
Model Uydurma Metotları	17
Olabilirlik fonksiyonuna ve maksimizasyonuna dair yaklaşımlar	17
Sabit etkilerin kestirimi	21
Rastgele etkilerin ve katsayıların kestirimi (veya tahmini).....	22
Varyans parametrelerinin kestirimi.....	24
Kovaryans Strüktürleri	29
MATERYAL ve YÖNTEM	33
Materyal	33

Yöntem.....	33
Veri üretimi	33
Verinin Modellenmesi.....	34
ARAŞTIRMA BULGULARI ve TARTIŞMA	38
İstatistikler.....	38
Örnek Hacmi	39
Kovaryans Strüktürü	39
SONUÇ.....	52
KAYNAKLAR.....	54
EKLER	57
EK 1. Örnek SAS Kodu	57
EK 2. Örnek R Kodu.....	69
ÖZGEÇMİŞ.....	79

TABLÖLAR DİZİNİ

Tablo 1. Homojen Kovaryans Strüktürleri.....	30
Tablo 2. Heterojen Kovaryans Strüktürleri.....	31
Tablo 3. Simülasyonlarda Kullanılan Bileşik Simetri Kovaryans Strüktürleri için Parametre Değerleri.....	34
Tablo 4. Simülasyonlarda Kullanılan Birinci Mertebeden Otoregresif ve rastgele Etki Kovaryans Strüktürleri için Parametre Değerleri.....	34
Tablo 5. Simülasyonda Kullanılan β Değerleri	35
Tablo 6. Doğru Modelin Seçiminde İstatistiklerin Başarı Orantıları (Kovaryans Strüktürü = Bileşik Simetri, 0.5)	40
Tablo 7. Doğru Modelin Seçiminde İstatistiklerin Başarı Orantıları (Kovaryans Strüktürü = Bileşik Simetri, 0.9)	42
Tablo 8 Doğru Modelin Seçiminde İstatistiklerin Başarı Orantıları (Kovaryans Strüktürü = Birinci mertebeden Otoregresif ve Rastgele Etki, 0.5)	44
Tablo 9. Doğru Modelin Seçiminde İstatistiklerin Başarı Orantıları (Kovaryans Strüktürü = Birinci mertebeden Otoregresif ve Rastgele Etki, 0.9)	46

ŞEKİLLER DİZİNİ

Şekil 1. Doğru Modelin Seçiminde İstatistiklerin Başarı Orantıları (Kovaryans Strüktürü = Bileşik Simetri, 0.5).....	48
Şekil 2. Doğru Modelin Seçiminde İstatistiklerin Başarı Orantıları (Kovaryans Strüktürü = Bileşik Simetri, 0.9).....	49
Şekil 3. Doğru Modelin Seçiminde İstatistiklerin Başarı Orantıları (Kovaryans Strüktürü = Birinci mertebeden Otoregresif ve Rastgele Etki, 0.5)	50
Şekil 4. Doğru Modelin Seçiminde İstatistiklerin Başarı Orantıları (Kovaryans Strüktürü = Birinci mertebeden Otoregresif ve Rastgele Etki, 0.9)	51



SİMGELER VE KISALTMALAR DİZİNİ

AIC	: Akaike enformasyon kriteri
ARRE	: Birinci metrebeden otoregresif ve rastgele etki
BIC	: Bayesçi enformasyon kriteri
CCC	: Mutabakat korelasyon katsayısı
CCC_{adj}	: Düzeltilmiş mutabakat korelasyon katsayısı
MLE	: Maksimum olabilirlik kestirimi
PRESS	: tahmin edilen kalıntı karler toplamı
R^2	: Determinasyon katsayısı
R^2_{adj}	: Düzeltilmiş determinasyon katsayısı
REMLE	: Kısıtlı maksimum olabilirlik kestirimi
V	: Kovaryans matrisi
X	: Sabit etkiler dizayn matrisi
Z	: Rastgele etkiler dizayn matrisi

GİRİŞ

Doğrusal karışık model boylamsal verinin analizinde güçlü bir araçtır. Boylamsal veri aynı birey veya deney ünitesi üzerinde yapılan yinelemeli ölçümleri ihtiva eder. Kestirimsel ve çıkarımsal maksatlarla, doğrusal karışık model, bağımsızca ve özdeşçe dağılmış Gaussian hatalara münhasır sıradan doğrusal modelin korelasyonlu Gaussian veriyi de kapsamak üzere genişletilmiş halidir. Tek-değişkeli doğrusal model için çok farklı tipte frekansçı seçim metotları geliştirilmiş olmakla beraber, doğrusal karışık model için geliştirilmiş olan frekansçı seçim metotlarının nicelik ve nitelik bakımından daha iyileştirilmeleri devam etmektedir (Davidian, 2001; Damienko, 2004). Doğrusal karışık model hem bir aritmetik ortalama modelinin hem de kovaryans modelinin seçimini gerektirir ve bunlar birbirinden ayrı mülahaza edilmelidir. Günümüzde standart olarak uygulanan metotlar da dahil olmak üzere, bu seçim metotlarını değerlendiren çalışmalar tartışmasızca açık ve net sonuçları henüz ortaya koyabilmiş değildir.

Tek-değişkeli modelde, Mallow C_p kriteri tüm aday modelleri ve bunların içinde yuvalanmış olduğu bir tam modeli gerektirir (Mallows, 1973; Gilmour, 1996). Her bir aday modelin hata kareler aritmetik ortalaması (MSE) tam modelin MSE'si ile karşılaştırılır ve böylece aday modellerin de birbirleriyle karşılaştırılması sağlanır. Ancak, halihazırda, doğrusal karışık model için herhangi bir C_p istatistiği mevcut değildir.

Doğrusal karışık model kurulumunda, model seçimine ait uyum-iyiliği istatistikleri olarak, AIC (Akaike, 1969, 1973, 1974) ve BIC (Schwarz, 1978) gibi enformasyon kriterleri mevcuttur (Ferron, et al., 2002; Gomez et al., 2005). Enformasyon teorik kriterler model seçiminde öne çıkan bir rol oynamışlardır ve muhtemelen doğrusal karışık model için model seçiminde son zamanlarda yapılan araştırmaların en aktif alanı olmuşlardır. Mesela, Cavanaugh (1999), doğrusal tek-değişkeli model için, Kullback simetrik sapmasına dayanan bir yönsüz sapma kriteri (KIC) geliştirmiştir. Ancak KIC, doğrusal karışık model için henüz ele alınmamıştır (Cavanaugh, 2004; Hafidi and Mkhadri, 2006; Kim and Cavanaugh, 2005). İlaveten, AIC ve BIC doğrusal karışık model için mütenasip model seçimi kriterleri olarak kabul görmüş olmakla beraber, bunların seçim kriterleri olarak fiili performanslarına dair araştırmalar gerçek veri ve simülasyon yoluyla devam etmektedir.

Doğrusal karışık model seçiminde, çeşitli dar örnek yaklaşımlarını kullanan pseudo F -testleri mevcuttur. Bu testlerin kompleks doğrusal karışık modeller için performansları Schaalje et al. (2002) tarafından öğrenilmeye-çalışılmıştır.

Korelasyonlu hataları havi yinelemeli ölçümler verisini modellemek için, Liu et al. (1999) çapraz-değerlendirme modelini genelleştirmişler ve Tahmin-edilmiş Kalıntı Kareler Toplamı (PRESS) adını vermişlerdir. Allen (1974) PRESS'i tek-değişkeli doğrusal model için bir model seçimi kriteri olarak teklif etmiştir. PRESS istatistiği, uyum-iyiliğinin çapraz-değerlendirmeli bir ölçüsü, Schabenberger (2004) ve Christensen (1992) tarafından modifiye edilmiştir.

Lin (1989) tarafından, gözlem çiftlerinin özdeşlik hattına tekabül etme isabetliliğinin ve kesinliğinin alternatif bir ölçüsü olarak, Mutabakat Korelasyon Katsayısı (CCC) teklif edilmiştir ve Vonesh (1992) tarafından karışık-etkiler regresyon kurulumlarında kullanılmak üzere modifiye edilmiştir.

R^2 , PRESS ve CCC istatistikleri gibi tahminsel kriterler standart doğrusal modeller için model seçiminde güzelce tesis edilmiş olmakla beraber, doğrusal karışık modeller için yapılan öğrenme-çalışmalarının yeterli olduğu söylenemez (Edwards et al., 2005). Ancak, bu istatistiklerin pek çoğu karışık model kurulumuna uyarlanabilir (Vonesh et al., 1996). Mesela, $R^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 / \sum_{i=1}^n (y_i - \bar{y}_i)^2$, doğrusal regresyonda, y 'deki toplam varyasyonun model tarafından açıklanana oranı olarak tanımlanır. Boylamsal kurulumda, ya marjinal ya da koşullu tahmin edilmiş değerleri kullanmak suretiyle, R^2 modifiye edilir (Schabenberger, 2004; Hocking, 1976).

Yinelemeli ölçümler verisinin muhtelif bilim alanlarına ait boylamsal çalışmalarda yaygın olarak kullanılmasıyla, bu tip verinin modellenmesine dair daha fazla çalışma yapılması ihtiyacı da buna paralel olarak artmaktadır. Kesitsel veri analizinde doğrusal tek-değişkeli model hangi rolü oynuyorsa, boylamsal veri analizinde doğrusal karışık model aynı rolü oynamaktadır. Ancak, standart tek-değişkeli model ile kıyaslandığında, karışık doğrusal model için, daha az sayıda frekansçı model seçimi metotları geliştirilmiştir ve değerlendirilmiştir.

Amaç

Mevcut literatürde, yukarıda anılan tahminsel seçim kriterlerinin değerlendirilmesine dair çalışmalar mevcut olmakla beraber, bunların muhtelif koşullar altındaki veri ile performanslarına ait değerlendirmelere hala ihtiyaç vardır. Bu çalışmada, doğrusal karışık model çerçevesinde, AIC, BIC, pseudo- F , R^2 , ve CCC nin performanslarını soruşturmak için simülasyon kullanılmış ve bir gerçek veri uygulaması yapılmıştır. Bu istatistiklerin marjinal ve

koşullu versiyonları (Vonesh et al., 1996), farklı varyans-kovaryans s tr kt rleri altında,   renilmeye  alışılmıştır.



KURAMSAL TEMELLER

Bu bölümde, frekansçı model seçimi teknikleri ve bunlara dair yapılan çalışmalar oturdukları bağlam itibarıyla gözden geçirilmektedir. Önce tek-değişkeli doğrusal model tanımlanmakta ve bu tezin konusu olan model seçimi kriterlerinin benzerlerine dair çalışmalar özetlenmektedir. Daha sonra, karışık doğrusal model tanımlanmakta, ve kullanılacak notasyon sunularak, bu tezin esas konusu olan seçim kriterlerine dair mevcut literatür gözden geçirilmektedir.

Tek-değişkeli Doğrusal Model

Bu tezde ele alınan model seçimi metotları önce tek-değişkeli doğrusal model için geliştirilmiş olduklarından dolayı, evvela bu modeller hakkında kısaca bilgi vermenin yerinde olacağı düşünülmüştür. Tek-değişkeli doğrusal model, aynı zamanda, tek yanıt için genel doğrusal model olarak da bilinir ve n birey ve p regresyon parametresi için şu şekilde gösterilir:

$$\begin{aligned} \underset{(n \times 1)}{\mathbf{y}} &= \underset{(n \times p)}{\mathbf{X}} \underset{(p \times 1)}{\boldsymbol{\beta}} + \underset{(n \times 1)}{\mathbf{e}} \\ \mathbf{e} &\sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n) \end{aligned}$$

Bu modelde, $\mathbf{y} = \{y_i\}$ yanıt vektörünün elemanlarının birbirinden bağımsız oldukları varsayılır. $\mathbf{X}$ kovaryans değerlerinin bilinen bir matrisi ve $\boldsymbol{\beta}$ regresyon parametrelerinin bilinmeyen bir vektörüdür. $n \times 1$ boyutlu $\mathbf{X}\boldsymbol{\beta}$ sabit vektörü yanıtların aritmetik ortalamalarını ve $n \times 1$ boyutlu $\mathbf{e}$ rastgele vektörü yanıtların varyanslarını betimler (Neter et al., 1996).

Yinelemeli Ölçümler İçin Doğrusal Karışık Model

Yinelemeli ölçümler bağlamında, i^{inci} denekte gözlenen verinin $(n_i \times 1)$ vektörü y_i olsun ki burada $i \in \{1, \dots, m\}$ ve m bağımsız ünitelerin (denekler) sayısıdır. Karışık model,

$$y_i = X_i \boldsymbol{\beta} + Z_i u_i + \varepsilon_i$$

olarak yazılabilir ki burada X_i i^{inci} denek için bilinen $(n_i \times p)$ sabite dizayn matrisidir, $\boldsymbol{\beta}$ sabit etkilerin meçhul $(p \times 1)$ katsayılar vektörüdür, Z_i q ranklı i^{inci} denek için bilinen $(n_i \times q)$ sabite dizayn matrisidir, u_i meçhul $(q \times 1)$ birey-spesifik rastgele etkiler vektörüdür ve ε_i $(q \times 1)$ rastgele hatalar vektörüdür (Gurka, 2006). Rastgele etkilerin katsayıları vektörünün, u_i , 0 aritmetik ortalama vektörü ve G kovaryans matrisi ile bir normal dağılım gösterdiği varsayılır. ε_i nin, u_i , 0 aritmetik ortalama vektörü ve R_i kovaryans matrisi ile bir normal dağılım

gösterdiği varsayılır. u_i ve ε_i vektörlerinin bağımsız olduğu varsayılır; netice itibarıyla, $\text{cov}(u_i, \varepsilon_i) = 0$ ve $\text{var}(y_i) = Z_i G Z_i' + R_i = V_i$. Eğer $N = \sum_{i=1}^m n_i$ alınır ve R_i nin ve G nin yekta parametreleri $k \times 1$ lik θ vektörü ile gösterilirse, doğrusal karışık model için log-olabilirlik fonksiyonu (Laird, and Ware, 1982),

$$l_{ML}(\beta, \theta) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^m \log(V_i(\theta)) - \frac{1}{2} \sum_{i=1}^m (y_i - X_i \beta)' V_i(\theta)^{-1} (y_i - X_i \beta).$$

Maksimum olabilirlik kestirimcisi (MLE), $l_{ML}(\beta, \theta)$ meçhul parametrelere göre sayısal olarak maksimize edilerek üretilir. θ nın MLE si yanlı olduğu için, θ nın kısıtlanmış maksimum olabilirlik kestirimcisi (REMLE) tipik olarak kullanılır. $\text{rank}(X_1' X_2' \cdots X_m')$ ın p olduğu varsayılarak, θ nın REML kestirimcisi, orjinal N gözlemin yeni bir $N - 1$ lik gözlem kümesine transformasyonunun olabilirlik fonksiyonunu maksimize etmek suretiyle hesaplanır (Gurka, 2006). Bu transformasyon öyle yapılandırılır ki kısıtlanmış log-olabilirlik fonksiyonu,

$$l_{REML}(\theta) = \frac{N-p}{2} \log(2\pi) + \frac{1}{2} \log \left| \sum_{i=1}^m X_i' X_i \right| - \frac{1}{2} \sum_{i=1}^m \log |V_i(\theta)| \\ - \frac{1}{2} \log \left| \sum_{i=1}^m X_i' V_i(\theta)^{-1} X_i \right| - \frac{1}{2} \sum_{i=1}^m (y_i - X_i \hat{\beta})' V_i(\theta)^{-1} (y_i - X_i \hat{\beta})$$

olarak yazılabilir ki burada $\hat{\beta}$,

$$\hat{\beta} = \left(\sum_{i=1}^m X_i' \hat{V}_i(\hat{\theta})^{-1} X_i \right)^{-1} \sum_{i=1}^m X_i' \hat{V}_i(\hat{\theta})^{-1} y_i$$

ve $\hat{V}_i$ kestirilmiş kovaryans matrisi V_i dir. Pratikte, yukarıdaki log-olabilirlik fonksiyonlarının maksimize edilmesinde iteratif prosedürler kullanılır.

Şunu not etmek ilginçtir ki SAS PROC MIXED (2003) REML kestirimlerini yukarıdaki $l_{REML}(\theta)$ ifadesinin bir versiyonu ile hesaplar ki $\frac{1}{2} \log |\sum_{i=1}^m X_i' X_i|$ terimini dışarıda bırakır (Gurka, 2006). Aynı kestirimleri üretmekle beraber, enformasyon kriterleri farklıdır (Alt-bölüm 2.2 de tartışılacak). SAS tarafından kullanılan log-olabilirlik fonksiyonu $l_{REML2}(\theta)$ ile gösterilecektir (Muller and Fetterman, 2002; Muller and Stewart, 2006).

Doğrusal Karışık Modellerde Model Seçimi

AIC ve BIC

Doğrusal karışık modellerde kovaryansa yapısını (Z_i , G ve R_i) ve aritmetik ortalama modeli (X_i ve β) seçimi için AIC (Akaike, 1973), BIC (Schwarz, 1978), AICC (Hurvich Tsai,

1989; Hurvich et al., 1990) ve CAIC (Bozdogan, 1987) gibi enformasyon kriterleri mevcuttur. Genelde, bu enformasyon kriterleri verilen bir modelin (l) maksimize edilmiş log-olabilirliğinin ve bu modeldeki parametre sayısına (s) dayanan bir ceza teriminin her ikisinin fonksiyonlarıdır (Gurka, 2006). AIC ve BIC için mutlak formüller şunlardır:

$$AIC = l - s$$

$$BIC = l - s(\log N^*)/2$$

ki burada N^* N , p veya m 'nin bir fonksiyonudur (mutaden ML için $N^* = N$ ve REML için $N^* = N - p$), ve $s = p + k$. ML kestirimi kullanıldığı zaman, genelde, $s = p + k$ modeldeki parametrelerin toplam sayısıdır (Sugiura, 1978). Ancak, REML kestiriminde, $s = k$ yalnızca kovaryans parametrelerinin sayısıdır (Gurka, 2006). Verilen bir model için enformasyon kriterinin daha büyük bir değeri o modelin tercih edileceğini belirtir.

Keselman et al. (1998), AIC'i ve BIC'i, çeşitli popülasyon kovaryans strüktürlerini saptamaktaki etkinlikleri bakımından karşılaştırdılar. Bilhassa, eşit/eşit-olmayan grup hacimlerine, çeşitli kovaryans matrislerine ve normal/normal-olmayan dağılımlı yanıtlara sahip dengesiz küresel-olmayan yinelenen ölçümler dizaynlarında, bu kriterler karşılaştırıldı. Bu çalışmaların sonuçları şunu gösterdi ki, anılan yaklaşımların hiç biri doğru kovaryans strüktürünü istikrarlıca seçemedi. Aslında, bu çalışmada oluşturulan strüktürlerin çoğu için, AIC ve bilhassa BIC yanlış kovaryans strüktürünü, doğru kovaryans strüktürüne göre, daha yüksek bir frekansla seçti. Keselman'ın sonuçları göstermiştir ki AIC denemelerin %47'sinde doğru kovaryans strüktürünü seçerken, BIC denemelerin %35'inde haklıydı. BIC için elde edilen negatif sonuçlar, parametrelerin sayısı için koyulan şiddetli cezalardan dolayı olabilir (Keselman et al., 1998).

Ferron et al. (2002), yine AIC'in ve BIC'in bir kovaryans matrisini seçmekteki performanslarını öğrenmeye çalıştılar. Farklı örnek hacimleri, yinelenen ölçümlerin sayıları ve otokorelasyon düzeyleri ile, AR(1)'e uygun veri üretildi. Hakiki kovaryans strüktürü, AR(1) ve bağımsız strüktür, $\sigma^2 I$, için AIC ve BIC hesaplandı. Sonuçlar gösterdi ki her örnek hacmi, yinelenen ölçümlerin sayısı ve otokorelasyon kombinasyonu için, BIC'e göre AIC daha başarılıydı. AIC denemelerin %79'unda doğru kovaryans strüktürünü seçerken, BIC denemelerin %66'sında seçti. Örnek hacmi, yinelenen ölçümlerin sayısı ve otokorelasyon düzeyi daha yüksek olduğu zaman, hem AIC hem BIC daha iyi performans gösterdi. Yinelenen ölçümlerin sayısı en etkili faktördü ve yinelenen ölçümlerin sayısı daha küçük olduğu zaman, örnek hacminin etkisi daha büyüktü.

Hakiki kovaryans strüktürünü seçmede AIC ve BIC kriterlerinin performanslarını öğrenmeye çalışan Gomez et al. (2005), on beş kovaryans strüktürü kullanarak veri ürettiler. Sonuçlar gösterdi ki doğru kovaryans strüktürünü seçmede AIC'in ve BIC'in başarı oranı büyük ölçüde örnek hacmine ve kovaryans strüktürlerine bağlıdır. Başarı oranları genellikle düşük olmasına rağmen daha geniş örnek hacimleri ve daha basit kovaryans strüktürleri için daha yüksektir. AIC'in ve BIC'in performansına, eşleştirmenin (pozitif eşleştirme ile, en geniş örnek hacminin en çok değeri ihtiva eden kovaryans matrisi ile asosiyasyonu durumu kastedilir) çok küçük bir etkisi oldu.

Gurka (2006), AIC'in ve BIC'in performanslarını, l_{ML} , l_{REML} ve l_{REML2} için karşılaştırdı (Alt-bölüm 2'ye bak). Doğrusal karışık modelin seçimini üç senaryo altında değerlendirdi: (1) kovaryans strüktürünün bilindiği durumda sabit etkilerin seçimi (2) sabit etkilerin bilindiği durumda rastgele etkilerin seçimi ve (3) sabit ve rastgele etkilerin eş-zamanlı seçimi. Sonuçlar gösterdi ki iki kriterin tüm versiyonları denmelerin %90'ından çoğunda, hakiki varyans veya korelasyon ne olursa olsun, münasip rastgele etkiler strüktürünü seçmiştir. AIC'e göre, BIC biraz daha iyi performans göstermiştir. Sonuçlar, yine, göstermiştir ki l_{REML2} kullanıldığı zaman, BIC daha iyi performansa sahipken, AIC için l_{REML2} seçimi l_{REML} ve l_{ML} den daha üstündür.

Keselman et al. (1989), sosyal bilimlerde müşterek olan kovaryans strüktürleri bağlamında, AIC'in ve BIC'in performansını soruşturdu ve kendi sonuçlarını Gurka'nın (2006) sonuçları ile karşılaştırdı. İkisinin sonuçları arasındaki farklar gösterdi ki AIC'in ve BIC'in performansı soruşturma koşullarına bağlıdır.

Pseudo F-testleri ve olabilirlik oranı testleri

Eğer C q ranklı bir kontrast matrisi ise, $H_0: C\beta = 0$ için müştereken kullanılan bir test istatistiği Wald F istatistiğidir,

$$F_{q,ddf} = \frac{W}{q}$$

ki burada;

$$W = (C\hat{\beta})'(C(X'\hat{V}^{-1}X)^{-1}C')^{-1}(C\hat{\beta})$$

ve ddf payda serbestlik derecesidir. Sıfır hipotezi altında, Wald F istatistiği yaklaşık olarak q ve ddf ile F dağılımı gösterir (Schaalje et al., 2002). Pseudo F testlerinde ddf 'yi hesaplamak için müştereken kullanılan iki metot vardır. Fai ve Cernelius (FC) metodu ddf 'yi kontrastlar için yaklaşık kovaryans matrisinin spektral dekompozisyonunu kullanmak suretiye

hesaplar ki bu prosedür tek dereceli serbestlik testlerine ait bir metodun yinelenen uygulamasını ihtiva eder (Fai and Cornelius, 1996; Giesbrecht and Burns, 1985; Schaalje et al., 2002). Kenward and Roger (1997), pseudo F -testlerini hesaplamak için ıslah edilmiş bir metot (KR metodu) öne sürdüler. KR metodu parametre kestirimlerinin kestirilmiş kovaryans matrisini düzeltir, test istatistiğinin bir ölçek düzeltilmesini hesaplar ve pseudo F -testinin serbestlik derecesini hesaplar.

Bu testlerin kompleks doğrusal karışık modellerde aritmetik ortalamalara (“aritmetik ortalama strüktürü”) göre model seçimi performansları Schaalje et al. (2002) tarafından öğrenilmeye çalışıldı. Sonuçlar gösterdi ki her iki modele ait Tip I hata oranları kovaryans strüktürünün kompleksliği, örnek hacmi ve dengesizlik tarafından etkilenmiştir. KR metodu örnek hacimleri yeterli genişlikte olduğu ve dizayn dengeli olduğu zaman komplike kovaryans strüktürleri ile iyi performans göstermiştir. Kompleks varyans strüktürleri için, küçük örnek hacimleri ile beraber her ne kadar Tip I hata oranlarını şişiriyor olsa bile KR metodu FC metoduna tercih edilebilir.

Gomez et al. (2005), kovaryans strüktürlerini seçmek için AIC ve BIC metotlarını kullanarak, Tip I hata oranlarını inceledi. Sonuçlar gösterdi ki en iyi AIC ve en iyi BIC modelleri için KR metodu hipotez testlerinin Tip I hata oranları hedeflenen değerlerden her zaman daha yüksektir. Seçim kriteri tipi, muamele başına denek sayısı, yinelenen ölçümlerin sayısı, kovaryans strüktürü ve eşleştirme Tip I hata oranlarını etkilemiştir.

Tahminsel yaklaşımlar

Doğrusal karışık modeller ve doğrusal modeller birbirinden başlıca iki bakımdan ayrılır. Birincisi, karışık bir modelde veri koşullu ve marjinal manada mülahaza edilebilir. $y|u$ ’nun koşullu dağılımı rastgele etkilerin belirli bir realizasyonunu kullanır. y ’nin marjinal dağılımı rastgele etkilerin tüm olanaklı değerleri üzerinden ortalaması alınmış değerleri kullanır. Mütakabil olarak, karışık modeller için koşullu ve marjinal tahminsel yaklaşımlar mevcuttur. İkincisi, β sabit etkilerinin kestirimleri kovaryans parametrelerinin kestirimlerine bağlıdır (Schabenberger, 2004). Karışık modellerde model seçimi için tahminsel yaklaşımlar bu ayrımları yansıtır.

Kareli korelasyon katsayısı, R^2 , basit doğrusal modelde

$$R^2 = 1 - \frac{\sum_{i=1}^n r_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

olarak tanımlanır ki burada $r_i = y_i - \hat{y}_i$, y_i i inci gözlem, $\hat{y}_i$ i inci tahmin değeri ve $\bar{y}$ genel ortalamadır. Kareli korelasyon katsayısı, karışık model kurulumunda, yine, hesaplanabilir fakat marjinal kalıntıların, r_m , mı yoksa koşullu kalıntıların, r_c , mı kullanıldığına bağlı olarak farklılaşır (Vonesh and Chinchilli, 1996; Vonesh and Chinchilli, 1997). Bir marjinal kalıntı, $r_i = y_i - x_i' \hat{\beta}$, gözlenen veri ve kestirilen (marjinal) aritmetik ortalama arasındaki farktır. Bir koşullu kalıntı, $r_i = y_i - x_i' \hat{\beta} - z_i' \hat{u}$, gözlenen veri ve gözlemin tahmin edilen değeri arasındaki farktır. R^2 'nin karışık model versiyonları yapılacak parametre kestirimlerinin sayısına göre düzeltilir:

$$R_{adj}^2 = 1 - k^* (1 - R^2)$$

ki burada

$$k^* = \frac{N}{N - p} \text{ or } k^* = \frac{N}{N - (p + k)}.$$

PRESS kalıntısı (Allen, 1974) gözlenen değer ve tahmin edilen değer arasındaki farktır ki burada tahmin edilen değer ilgili gözlemin dışında elde edilir. Basit doğrusal model için PRESS istatistiği, modeli yeniden uydurmadan, güncelleme formülü uygulanarak hesaplanabilir (Schabenberger, 20004). Boylamsal kurulumda, böyle güncelleme formülleri elverişli değildir çünkü gözlemler korelasyonludur (Schabenberger, 20004). PRESS istatistiği yine karışık modeller için hesaplanabilir fakat marjinal kalıntıların, r_m , mı yoksa koşullu kalıntıların, r_c , mı kullanıldığına bağlı olarak farklılaşır (Liu et al., 1999).

Mutabakat korelasyon katsayısı (CCC) Lin (1989) tarafından betimlenmiştir ve Vonesh (1992) tarafından karışık-etkiler regresyon kurulumlarında kullanılmak üzere modifiye edilmiştir. İki değişkeli Y_1 ve Y_2 ölçümleri μ_1 ve μ_2 aritmetik ortalamaları, σ_1 ve σ_2 varyansları ve ρ korelasyonu ile verildiğinde, Y_1 ve Y_2 arasındaki korelasyonu, Lin (1989)

$$\rho_c = 1 - \frac{E[(Y_1 - Y_2)^2]}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2} = \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2}$$

olarak tanımlamıştır.

CCC gözlem çiftlerinin özdeşlik hattına tekabül etme isabetliliğinin ve kesinliğinin bir ölçüsüdür. ρ_c 'nin n gözlemlik bir rastgele örneğe dayalı bir nokta kestirimi,

$$\hat{\rho}_c = 1 - \frac{\sum_{i=1}^n (Y_{i1} - Y_{i2})^2}{\sum_{i=1}^n (Y_{i1} - \bar{Y}_1)^2 + \sum_{i=1}^n (Y_{i2} - \bar{Y}_2)^2 + n(\bar{Y}_1 - \bar{Y}_2)^2}.$$

Vonesh (1992) CCC'nin model uyumunu ölçmek için uyarlanmasını öne sürdü ve model CCC'yi

$$r_c = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)'(y_i - \hat{y}_i)}{\sum_{i=1}^n (y_i - \bar{y})'(y_i - \bar{y}) + \sum_{i=1}^n (\hat{y}_i - \bar{y})'(\hat{y}_i - \bar{y}) + N(\bar{y} - \bar{y})^2}$$

olarak tanımladı.

r_c 'yi model uyumunun bir ölçüsü olarak kullanmanın pek çok avantajı vardır. Birincisi, gözlenen ve tahmin edilen değerler arasında bir mutabakat korelasyon katsayısı olarak r_c 'nin doğrudan interpretasyonu yapılabilir. İkincisi, bir sıfır modelinin spesifikasyonunu gerektirmez çünkü özdeşlik hattı bir referans noktası olarak hizmet eder. Karışık model kurulumunda, marjinal veya koşullu kestirimler kullanılmak suretiyle, bir marjinal veya koşullu model CCC'si hesaplanabilir. Yine, CCC parametrelerin sayısı için,

$$CCC_{adj} = 1 - k^* (1 - CCC)$$

olarak düzeltilebilir ki burada

$$k^* = \frac{N}{N - p} \text{ or } k^* = \frac{N}{N - (p + k)}.$$

Doğrusal Karışık Modeller

Bu bölümde, normal dağılımlı hatalara sahip karışık model daha detaylıca tartışılmaktadır. Buna, “normal karışık model” de denilir ve bununla kastetilen yanıt değişkenine ait değerlerin normal dağılımlı olması değildir. Çünkü, gerçekte, yanıt değişkenlerine ait değerler farklı aritmetik-ortalamalarla etkilerin karışımlarıdır. Böyle olmakla beraber, uygulamada, eğer bir değişken normal dağılım gösteriyorsa, kalıntıların ve rastgele etkilerin normallliği varsayımı çoğu zaman geçerlidir.

Model tanımı

Bu bölümde, karışık model genel matris notasyonu kullanılarak tanımlanmaktadır ki bu notasyon her tipten karışık model için kullanılabilir (Demidenko, 2013).

Sabit etkiler modeli

Tüm sabit etkiler modelleri

$$y_i = \mu + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \cdots + \alpha_p x_{ip} + e_i$$

genel formunda spesifikleştirilebilir ki

$$\text{var}(e_i) = \sigma^2.$$

Mesela, yinelenen ölçümler dizaynı bağlamında,

$$y_{ij} = \mu + p_i + t_i + e_{ij}$$

olarak gösterilebilir ki burada i alt-indisi i inci bireyi ve j alt-indisi j inci ölçümü belirtir. Genel model notasyonunda, halbuki, her bir gözlem ayrı bir alt-indis ile belirtilir. Buna göre, y_1 ve y_2 1 nolu bireyin gözlemlerini temsil ederken, y_3 ve y_4 2 nolu bireyin gözlemlerini temsil etmekte ve bu şekilde devam etmektedir. Genel modeldeki α terimleri p_i terimlerine tekabül eden sabitelerdir ve birey ve ölçüm etkilerinin miktarlarını verirler. Bu model $p + 1$ sayıdaki sabit etki parametrelerini uydurur, yani $\alpha_1 - \alpha_p$ ve μ kesim terimi. Eğer n sayıda gözlem var ise, o zaman bu gözlemler

$$y_1 = \mu + \alpha_1 x_{11} + \alpha_2 x_{12} + \cdots + \alpha_p x_{1p} + e_1,$$

$$y_2 = \mu + \alpha_1 x_{21} + \alpha_2 x_{22} + \cdots + \alpha_p x_{2p} + e_2,$$

⋮

$$y_n = \mu + \alpha_1 x_{n1} + \alpha_2 x_{n2} + \cdots + \alpha_p x_{np} + e_n$$

olarak yazılır ve

$$\text{var}(e_1) = \sigma^2,$$

⋮

$$\text{var}(e_n) = \sigma^2.$$

Bunlar matris notasyonu ile daha kısa ve öz olarak şöyle ifade edilebilir:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{e},$$

$$\mathbf{V} = \text{var}(\mathbf{y}) = \sigma^2 \mathbf{I}$$

ki burada,

$\mathbf{y} = (y_1, y_2, y_3, \dots y_n)'$ = gözlenen değerler,

$\boldsymbol{\alpha} = (\mu, \alpha_1, \alpha_2, \dots \alpha_p)'$ = sabit etkiler parametreleri,

$\mathbf{e} = (e_1, e_2, e_3, \dots e_n)'$ = kalıntılar,

σ^2 = kalıntı varyansı,

$\mathbf{I}$ = $n \times n$ birim matris.

$\boldsymbol{\alpha}$ birden çok sayıda değişkene ait parametreleri barındırabilir. Yinelenen ölçümlerde, $\boldsymbol{\alpha}$ birey etkileri ve ölçüm etkilerine ait parametreleri barındırmaktadır. Bu değişkenlerin her ikisi de kategorik değerler alan nitel değişkenlerdir ve böylesi etkiler kategorik etkiler diye adlandırılır. Aynı zamanda bunlar bazen faktör etkileri diye de adlandırılır. Daha genel anlamda, kategorik etkiler öyle etkilerdir ki bunlara ait gözlemler belli sayıda sınıftan birine düşer. Aynı zamanda, $\boldsymbol{\alpha}$ bir veya daha çok sayıda kovaryete etkilerini de (mesela yaş veya başlangıç ölçümü gibi) barındırabilir. Bunlar, nicel bir ölçekte ölçülen değişkenlerdir. Bir kategorik etkiyi modele dahil etmek için birden çok sayıda parametre gerekli olabilir fakat bir kovaryete etkisini modele dahil etmek için yalnızca bir parametre yeterlidir.

$\mathbf{X}$ dizayn matrisi olarak bilinir ve $n \times p$ boyutludur (yani n satır ve p sütun). Bu matris her gözlem için her bir parametreye tekabül eden sabit etkilerin değerlerini spesifikleştirir. $\mathbf{X}$ 'te kategorik etkiler için, etki kategorilerinin yokluğunu ve varlığını belirtmek üzere 0 ve 1 değerleri kullanılırken, kovaryete etkileri için değişkenin aldığı değerlerin kendileri kullanılır.

$\mathbf{V}$ matrisi gözlemlerin varyanslarını ve kovaryanslarını içermektedir. Olağan bir sabit etkiler modelinde, tüm gözlemlerin varyansları eşittir ve gözlemler korelasyonlu değildir. Buna göre, $\mathbf{V}$ basitçe $\sigma^2 \mathbf{I}$ 'dir.

Karışık model

Karışık model, rastgele etkileri, rastgele katsayıları ve/veya kovaryans terimlerini kalıntı varyans matrisine dahil etmek suretiyle sabit etkiler modelini genişletir. Sabit etkiler modeli rastgele etkileri (veya katsayıları) içerecek şekilde genişletildiği zaman, karışık model p sayıda sabit etkiler parametrelerini ve q sayıda rastgele etkiler (veya katsayılar) parametrelerini uyduran bir model için

$$y_i = \mu + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \dots + \alpha_p x_{ip} + \beta_1 z_{i1} + \beta_2 z_{i2} + \dots + \beta_q z_{iq} + e_i$$

olarak spesifikleştirilebilir. Sabit etkiler sabit sabiteler olarak kabul edilirken rastgele etkilerin bir dağılıma sahip olduğu varsayılır. Bu model matris notasyonunda

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta} + \mathbf{e}$$

olarak ifade edilebilir ki burada $\mathbf{y}$, $\mathbf{X}$, $\boldsymbol{\alpha}$ ve $\mathbf{e}$ sabit etkiler modelinde tanımlandıkları üzeredir ve

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_q)' = \text{rastgele etkiler/katsayılar parametreleri.}$$

$\mathbf{Z}$ her bir gözleme tekabül eden rastgele etkilerin değerlerini veren $n \times q$ boyutlu bir ikinci dizayn matrisidir. Bir intersept teriminin dahil edilmemesi dışında, $\mathbf{Z}$ tamamen sabit etkiler için $\mathbf{X}$ 'in spesifikize edildiği şekilde spesifikleştirilir (Fitzmaurice et al., 2008.).

V kovaryans matrisi

Sabit etkiler modelinde tüm gözlemlerin eşit varyanslara sahip olduklarını ve gözlemlerin korelasyonsuz oldukları varsayılır. Bu, $\mathbf{V}$ matrisinin diyagonal olmasına yol açar. Oysa rastgele etkiler uydurulduğu zaman, gözlemler korelasyonludur ve bu durum matris notasyonu kullanılarak şöyle genelleştirilir. $\mathbf{y}$ 'nin kovaryansı, $\text{var}(\mathbf{y}) = \mathbf{V}$,

$$\mathbf{V} = \text{var}(\mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta} + \mathbf{e})$$

olarak yazılabilir. Rastgele etkilerin kalıntılarla korelasyonsuz olduğu varsayıldığında,

$$\mathbf{V} = \text{var}(\mathbf{X}\boldsymbol{\alpha}) + \text{var}(\mathbf{Z}\boldsymbol{\beta}) + \text{var}(\mathbf{e}).$$

$\boldsymbol{\alpha}$ sabit etkiler parametrelerini betimlediğine göre, $\text{var}(\mathbf{X}\boldsymbol{\alpha}) = \mathbf{0}$. Aynı zamanda, $\mathbf{Z}$ sabitelerin bir matrisidir. Buna göre,

$$\mathbf{V} = \mathbf{Z}\text{var}(\mathbf{Z}\boldsymbol{\beta})\mathbf{Z}' + \text{var}(\mathbf{e}).$$

$\mathbf{G}$ ile $\text{var}(\boldsymbol{\beta})$ 'yı belirtilirse, rastgele etkilerin normal dağılıma sahip olduğu varsayımından hareketle, $\boldsymbol{\beta} \sim \mathbf{N}(\mathbf{0}, \mathbf{G})$ yazılabilir. Benzer şekilde, kalıntı kovaryans matrisini $\text{var}(\mathbf{e}) = \mathbf{R}$ olarak yazılır ve $\mathbf{e} \sim \mathbf{N}(\mathbf{0}, \mathbf{G})$. Böylece,

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}.$$

Rastgele etkiler modeli kovaryans yapısı

G matrisi

$\mathbf{G}$ 'nin boyutu $q \times q$ 'dur, ki burada q rastgele etkiler parametrelerinin toplam sayısına eşittir. Rastgele etkiler modellerinde, $\mathbf{G}$ daima diyagonaldır (yani rastgele etkilerin korelasyonsuz olduğu varsayılır). Eğer üç ölçümün ve iki grubun olduğu boylamsal veride yalnızca birey etkileri rastgele olarak uydurulursa, o zaman $\mathbf{G}$

$$\mathbf{G} = \begin{pmatrix} \sigma_c^2 & 0 & 0 \\ 0 & \sigma_c^2 & 0 \\ 0 & 0 & \sigma_c^2 \end{pmatrix}$$

formuna sahip olur ki σ_c^2 birey varyans bileşenidir. Eğer hem birey hem de birey×grup etkileri rastgele olarak uydurulursa, o zaman **G**

$$\mathbf{G} = \begin{pmatrix} \sigma_c^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{ct}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{ct}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{ct}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{ct}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{ct}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{ct}^2 \end{pmatrix}$$

formuna sahip olur ki σ_{ct}^2 birey×grup varyans bileşenidir (Fujikoshi and Satoh, 1997).

R matrisi

Rastgele etkiler modellerinde kalıntılar korelasyonsuzdur ve $\mathbf{R} = \sigma^2 \mathbf{I}$:

$$\mathbf{R} = \begin{pmatrix} \sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 \end{pmatrix}$$

V matrisi

V varyans matrisinin $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$ formuna sahiptir. $\mathbf{ZGZ}'$ rastgele etkilerden kaynaklanan kovaryansı spesifikleştirir. Eğer yalnızca birey etkileri rastgele olarak uydurulursa, o zaman

$$\mathbf{ZGZ}' = \begin{pmatrix} \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 \end{pmatrix}$$

elde edilir. Bu matris zahmetli bir matris çarpımları süreciyle elde edilebilmekte fakat daima aynı forma sahip olmaktadır. Rastgele etkiler kategorilerinin her birindeki gözlem sayısına tekabül eden bloklardan müteşekkil bir *blok diyagonal* forma sahiptir. $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$ toplam varyans matrisi, bu halde,

$$\mathbf{V} = \begin{pmatrix} \sigma_c^2 + \sigma^2 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 + \sigma^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 + \sigma^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 + \sigma^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_c^2 + \sigma^2 & \sigma_c^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 + \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 + \sigma^2 & \sigma_c^2 & \sigma_c^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 + \sigma^2 & \sigma_c^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 + \sigma^2 \end{pmatrix}.$$

Bu matris te aynı zamanda bir blok diyagonal forma sahiptir. Bu blok diyagonal form σ_c^2 rastgele etkiler bileşenine eşit aynı bireylerdeki gözlemlerin kovaryanslarından ve $\sigma_c^2 + \sigma^2$ birey ve kalıntı varyans bileşenlerinin toplamına eşit diyagonaldeki varyans terimlerinden müteşekkildir. Eğer hem birey hem de birey×grup etkileri rastgele olarak uydurulursa, o zaman

$$\mathbf{ZGZ}' = \begin{pmatrix} \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 + \sigma_{ct}^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 + \sigma_{ct}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 + \sigma_{ct}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 + \sigma_{ct}^2 \end{pmatrix}$$

ve

$$\mathbf{V} = \begin{pmatrix} \theta & \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 + \sigma_{ct}^2 & \theta & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \theta & \sigma_c^2 + \sigma_{ct}^2 & 0 & 0 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 + \sigma_{ct}^2 & \theta & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \theta & \sigma_c^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_c^2 & \theta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta & \sigma_c^2 + \sigma_{ct}^2 & \sigma_c^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 + \sigma_{ct}^2 & \theta & \sigma_c^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 & \theta \end{pmatrix}$$

olur ki $\theta = \sigma_c^2 + \sigma_{ct}^2 + \sigma^2$. Buna göre, $\mathbf{V}$ yine bir blok diyagonal forma sahiptir ki bu diyagonal form biraz daha karmaşık bir yapıdadır. Birey×grup varyans bileşeni aynı gruptaki bireylerin ölçümlerine ait kovaryans terimlerine eklenir.

Rastgele katsayılar modeli kovaryans yapısı

Kovaryans yapısını misallendirmek için aşağıdaki veri kullanılacaktır. Veri iki muameleli bir yinelenen ölçümler denemesindeki ilk üç birey için ölçüm zamanlarını temsil ediyor.

Birey	Grup	Zaman
1	A	t_{11}
1	A	t_{12}
1	A	t_{13}
1	A	t_{14}
2	B	t_{21}
2	B	t_{22}
3	A	t_{31}
3	A	t_{32}
3	A	t_{33}

Eğer birey ve birey×zaman etkileri rastgele katsayılar olarak uydurulursa, o zaman altı adet rastgele katsayı olur. Okuma kolaylığı adına, öncekiki notasyonu değiştirerek bu katsayıları $\beta_{p,1}$, $\beta_{pt,1}$, $\beta_{p,2}$, $\beta_{pt,2}$, $\beta_{p,3}$ ve $\beta_{pt,3}$ olarak tanımlanırsa ve üç bireyin her biri için bir kesim (birey) ve eğim (birey×zaman) hesaplanabilmesine olanak sağlanır. Bu halde, $\mathbf{Z}$ matrisi

$$\mathbf{Z} = \begin{matrix} & \beta_{p,1} & \beta_{pt,1} & \beta_{p,2} & \beta_{pt,2} & \beta_{p,3} & \beta_{pt,3} \\ \begin{pmatrix} 1 & t_{11} & 0 & 0 & 0 & 0 \\ 1 & t_{12} & 0 & 0 & 0 & 0 \\ 1 & t_{13} & 0 & 0 & 0 & 0 \\ 1 & t_{14} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & t_{21} & 0 & 0 \\ 0 & 0 & 1 & t_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & t_{31} \\ 0 & 0 & 0 & 0 & 1 & t_{32} \\ 0 & 0 & 0 & 0 & 1 & t_{33} \end{pmatrix} \end{matrix}$$

olur.

R matrisi

Rastgele etkiler modellerinde olduğu gibi, kalıntılar korelasyonsuzdur ve kalıntı kovaryans matrisi

$$\mathbf{R} = \begin{pmatrix} \sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 \end{pmatrix}.$$

G matrisi

Bir rastgele katsayılar modelinde, birey etkileri (interseptler) rastgele birey×zaman etkileri (eğimler) ile korelasyonludur. Korelasyon yalnızca aynı bireye ait katsayılar için mevcuttur (yani $\beta_{p,i}$ ve $\beta_{pt,i}$ arasında) ve farklı bireylere ait katsayılar korelasyonsuzdur. Buna göre, **G** matrisi

$$\mathbf{G} = \begin{pmatrix} \sigma_p^2 & \sigma_{p,pt} & 0 & 0 & 0 & 0 \\ \sigma_{p,pt} & \sigma_{pt}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_p^2 & \sigma_{p,pt} & 0 & 0 \\ 0 & 0 & \sigma_{p,pt} & \sigma_{pt}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_p^2 & \sigma_{p,pt} \\ 0 & 0 & 0 & 0 & \sigma_{p,pt} & \sigma_{pt}^2 \end{pmatrix}$$

olur ki σ_p^2 ve σ_{pt}^2 birey ve birey×zaman varyans bileşenleri ve $\sigma_{p,pt}$ rastgele katsayılar arasındaki kovaryanstır. Görüyoruz ki **G** 6×6 boyutunda çünkü modelde altı adet rastgele katsayı var.

V matrisi

Yine, $\mathbf{V} \mathbf{V} = \mathbf{Z} \mathbf{G} \mathbf{Z}' + \mathbf{R}$ olarak elde edilir ki $\mathbf{Z} \mathbf{G} \mathbf{Z}'$ rastgele katsayılardan kaynaklanan kovaryansı spesifikize etmektedir ve misal alınan veri için

$$\mathbf{Z} \mathbf{G} \mathbf{Z}' = \begin{pmatrix} v_{1,11} & v_{1,12} & v_{1,13} & v_{1,14} & 0 & 0 & 0 & 0 & 0 \\ v_{1,12} & v_{1,22} & v_{1,23} & v_{1,24} & 0 & 0 & 0 & 0 & 0 \\ v_{1,13} & v_{1,23} & v_{1,33} & v_{1,34} & 0 & 0 & 0 & 0 & 0 \\ v_{1,14} & v_{1,24} & v_{1,34} & v_{1,44} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & v_{2,11} & v_{2,12} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & v_{2,12} & v_{2,22} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & v_{3,11} & v_{3,12} & v_{3,13} \\ 0 & 0 & 0 & 0 & 0 & 0 & v_{3,12} & v_{3,22} & v_{3,23} \\ 0 & 0 & 0 & 0 & 0 & 0 & v_{3,13} & v_{3,23} & v_{3,33} \end{pmatrix}$$

formuna sahiptir ki $v_{i,jk} = \sigma_p^2 + (t_{ij} + t_{ik})\sigma_{p,pt} + t_{ij}t_{ik}\sigma_{pt}$.

Buna göre, $\mathbf{Z} \mathbf{G} \mathbf{Z}'$ bir blok diyagonal forma sahiptir ve bloklar her bir bireydeki gözlemlerin sayısına tekabül eder. $\mathbf{V} = \mathbf{Z} \mathbf{G} \mathbf{Z}' + \mathbf{R}$ toplam kovaryans matrisini teşkil etmek için diyagonala $\mathbf{R} = \sigma^2 \mathbf{I}$ eklenir. $\mathbf{V}$ de aynı zamanda bir blok diyagonal forma sahiptir. Kovaryanslar zamanla birlikte artacaktır ve zaman için seçilebilecek farklı bir orijin farklı sonuçlar ortaya koyabilecektir. Bununla beraber, $\mathbf{V}$ zaman orijinine göre değişmezdir ve kovaryans parametreleri değişse bile yine aynı genel sonuçlar elde edilebilir.

Görülüyor ki rastgele katsayılar modellerinde kovaryans yapısı random katsayılar tarafından belirlenmektedir. Bu, kovaryans biçimi modellerinden farklıdır. Şöyle ki, orada, $\mathbf{R}$ (veya ara sıra $\mathbf{G}$) matrisindeki kovaryans parametreleri verinin belirli bir biçimini yansıtmak üzere seçilir.

Kovaryans biçimi modeli kovaryans yapısı

Kovaryans biçimi modellerde, verinin kovaryans yapısı, rastgele etkileri veya katsayıları spesifikize ederek değil doğrudan $\mathbf{R}$ (veya ara sıra $\mathbf{G}$) matrisindeki kovaryans terimlerinin biçimini spesifikize ederek tanımlanır. Bir *bloklama değişkenine* (mesela bireyler) ait gözlemlerin korelasyonlu olmasına izin verilir ve bunların kovaryansları için bir biçim spesifikize edilir. Bu biçim çoğunlukla zaman veya vizite numarası gibi bir değişkene bağlı olarak seçilir. $\mathbf{R}$ bir blok diyagonal forma sahip olur ve

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & . & . & . \\ \mathbf{0} & \mathbf{R}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & . & . & . \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & . & . & . \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_4 & \mathbf{0} & \mathbf{0} & \mathbf{0} & . & . & . \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_5 & \mathbf{0} & \mathbf{0} & . & . & . \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_6 & \mathbf{0} & . & . & . \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_7 & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \end{pmatrix}$$

olarak yazılır. $\mathbf{R}_i$ alt-matrisleri i inci bloklama etkisine (yani i inci birey) tekabül eden kovaryans bloklarıdır. Bunların boyutu her bir bireyde tekrarlanan ölçümlerin sayısına eşittir. $\mathbf{0}$ sıfırlı matris bloklarını temsil etmektedir ve farklı bireylere ait gözlemler için sıfır kovaryanslarını vermektedir. Farazi bir küçük veri kümesi kullanarak $\mathbf{R}$ matrisleri için şu iki misal verilebilir.

Birey	Ölçüm
1	1
1	2
1	3
2	1
2	2
2	3
2	4
3	1
3	2

Bireyleri bloklama etkisi olarak kullanmak suretiyle, her ölçüm çifti için ayrı bir korelasyona izin veren (bu “genel” bir kovaryans biçimi olarak betimlenebilir) bir $\mathbf{R}$ matrisi

$$\mathbf{R} = \begin{pmatrix} \sigma_1^2 & \theta_{12} & \theta_{13} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_{12} & \sigma_2^2 & \theta_{23} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_{13} & \theta_{23} & \sigma_3^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_1^2 & \theta_{12} & \theta_{13} & \theta_{14} & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_{12} & \sigma_2^2 & \theta_{23} & \theta_{24} & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_{13} & \theta_{23} & \sigma_3^2 & \theta_{34} & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_{14} & \theta_{24} & \theta_{34} & \sigma_4^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_1^2 & \theta_{12} & \theta_{13} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{12} & \sigma_2^2 & \theta_{23} \end{pmatrix}$$

olarak verilir. Alternatif olarak, her ölçüm çifti arasında sabit bir korelasyonu varsayan daha basit bir biçim (“bileşik simetri” biçimi olarak bilinir)

$$\mathbf{R} = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 & \rho\sigma^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \rho\sigma^2 & \sigma^2 \end{pmatrix}$$

olarak verilir ki ρ = aynı bireye ait gözlemler arasındaki korelasyon.

Yaygın olarak, tekrarlanan ölçümler verisinin analizinde bir rastgele etki uydurulmaz ki bu durumda varyans matrisi $\mathbf{V} = \mathbf{R}$. Aksi taktirde, $\mathbf{R}$, verinin $\mathbf{V}$ tam varyans matrisini teşkil etmek için $\mathbf{ZGZ}'$ 'ne eklenir.

G matrisinde kovaryans biçimleri

Her ne kadar az kullanılsa da, $\mathbf{G}$ matrisine, bir bloklama etkisi içinde rastgele etkilerin korelasyonlu olduğu bir kovaryans biçimi uydurmak olanaklıdır. Mesela, her bireyin bir kaç zaman ile değerlendirildiği ve her her zamanda bir kaç ölçümün alındığı bir tekrarlanan ölçümler denemesi düşünülebilir. Aynı zamandaki gözlemler arasındaki korelasyonu modellemek kadar, bireyler içinde zamanlar arasındaki korelasyon da modellenmek istenebilir. Bunu başarmak için, $\mathbf{R}$ matrisinde kovaryans biçimleri spesifikleştirildiği gibi $\mathbf{G}$ matrisinde de kovaryans biçimlerini spesifikleştirilmesi gerekir.

Model Uydurma Metotları

Model uydurma prosesi birbirinden başka üç bileşene sahiptir: sabit etkilerin kestirimi (yani, α), rastgele etkilerin kestirimi (yani, β) ve varyans parameterelerinin kestirimi (yani, varyans bileşenleri veya kovaryans terimleri).

Olabilirlik fonksiyonuna ve maksimizasyonuna dair yaklaşımlar

Karışık model, *olabilirlik fonksiyonunu*, veri değerleri için, maksimize edilmek suretiyle uydurulabilir. Olabilirlik fonksiyonu, L , mevcut veriye göre model parametrelerinin olabilirliğini ölçer ve gözlemlerin yoğunluk fonksiyonu kullanılarak tanımlanır. Gözlemlerin bağımsız olduğunun varsayıldığı modellerde (yani, sabit etkiler modelleri), olabilirlik fonksiyonu basitçe gözlemlere ait yoğunluk fonksiyonlarının çarpımıdır. Ancak, bir karışık modeldeki gözlemler bağımsız değildir ve bu yüzden olabilirlik fonksiyonunun gözlemlerin çokdeğişkeli bir yoğunluk fonksiyonuna dayandırılmasına ihtiyaç vardır. Varyans parameterelerinin ve sabit etkilerin olabilirliği $\mathbf{y}$ nin çokdeğişkeli normal dağılımı kullanılarak

tanımlanabilir (buradaki “varyans parametreleri” terimi $\mathbf{G}$ ve $\mathbf{R}$ matrislerindeki tüm parametrelere şamildir). Rastgele etkilerin beklenen değerleri sıfır olacağı ve bu yüzden aritmetik ortalamayı etkilemeyeceği için, bu dağılım bir $\mathbf{X}\boldsymbol{\alpha}$ aritmetik ortalama vektörüne ve bir $\mathbf{V}$ kovaryans matrisine sahiptir. Çokdeğişkeli normal yoğunluk fonksiyonuna dayanan olabilirlik fonksiyonu, o halde,

$$L = \frac{\exp \left[-\frac{1}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha})' \mathbf{V}' (\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha}) \right]}{(2\pi)^{(1/2)n} |\mathbf{V}|^{1/2}}$$

Pratikte, log olabilirlik fonksiyonu çoğu zaman olabilirlik fonksiyonu yerine kullanılır, çünkü onunla çalışmak daha basittir ve onun maksimum değeri olabilirliğinki ile çakışır. Log olabilirlik

$$\log(L) = K - \frac{1}{2} [\log |\mathbf{V}| + (\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha})' \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha})]$$

ile verilir ki burada

$$K = \frac{1}{2} n \log(2\pi) = (\text{maksimizasyon prosesinde ihmal edilebilir bir sabite})$$

n = gözlem sayısı.

Bu haliyle, log olabilirliği maksimize eden model parametrelerinin değerleri belirlenebilir.

Karışık modeli uydurmaya dair muhtelif yaklaşımlar mevcuttur ki bunların tümü olabilirlik fonksiyonunun maksimizasyonuna dayanmaktadır (doğrudan veya dolaylı olarak). Görüleceği üzere, metotların hepsi biribiriyle eşdeğer değildir ve model parametrelerinin farklı kestirimlerine yol açabilirler. Metotların takdiminden sonra, uydurma prosesinin ayrı hususlarına daha detaylıca bakılacaktır: rastgele etkilerin kestirimi ve varyans parametrelerinin kestirimi.

Maksimum olabilirlik (ML)

Bu metot, sabit etkilere, $\boldsymbol{\alpha}$, sabiteler olarak muamele etmek suretiyle, log olabilirliğin varyans parametrelerine göre maksimize edilmesi kavramına dayanır. Varyans parametrelerinin kestirimlerinin elde edilmesinden sonra, ancak sabit etkilerin kestirimleri elde edilir ki bunun için varyans parametreleri sabit olarak kabul edilir ve $\boldsymbol{\alpha}$ 'nın log olabilirliği maksimize eden değerleri bulunur. Bu metot varyans parametrelerini bir derece aşağıya yanlı kestirmeye meyillidir. Bu çok basit bir misalle gösterilebilir. Farzedelim ki $x_1, x_2, \dots, x_n$ basit rastgele örneğine sahibiz ve aritmetik ortalamayı ve varyansı kestirmek istiyoruz. Eğer $\hat{\mu}$ örnek aritmetik ortalaması ise, maksimum olabilirlik (ML) varyans kestirimcisi $\sum_i (x_i - \hat{\mu})^2 / n$ olurdu, yansız

kestirimci $\sum_i (x_i - \hat{\mu})^2 / (n - 1)$ yerine. Varyans parametrelerini kestirmek için ne kadar az sayıda serbestlik derecesi kullanılırsa, yanlışlık ta o kadar büyür.

Kalıntı maksimum olabilirlik (REML)

Kalıntı maksimum olabilirlik (REML) (bazen kısıtlanmış maksimum olabilirlik te denir) ilk defa Patterson ve Tohmppson (1971) tarafından öne sürüldü. Bu yaklaşımda, α parametresi elemine edilerek, log olabilirlik yalnızca varyans parametreleri cinsinden tanımlanmış olur. Bu metodun çerçevesi şöyle çizilebilir. İlk olarak, $\mathbf{y} - \mathbf{X}\hat{\alpha}$ kalıntı terimlerine dayanan bir olabilirlik fonksiyonunu alınır. Bu daha önce tanımlanan olabilirlik ile zıttır ki o doğrudan $\mathbf{y}$ gözlemlerine dayanıyordu. Dikkat edilmelidir ki bu kalıntılar $\mathbf{Z}\hat{\beta}$ 'nin çıkarılmamış olduğu sıradan $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\alpha} - \mathbf{Z}\hat{\beta}$ kalıntılarından farklıdır. Bazı yazarlar (REML ninkiler de dahil) aynı zamanda $\mathbf{y} - \mathbf{X}\hat{\alpha}$ ya da kalıntılar olarak atıfta bulunurlar. Bu mantıksız değil, çünkü onlar tüm rastgele varyasyonu (kalıntı ve rastgele etkiler) ihtiva eden hata terimleri olarak mülahaza edilebilirler. Burada, sıradan kalıntılardan ayırmak için, biz $\mathbf{y} - \mathbf{X}\hat{\alpha}$ 'yı *tam kalıntılar* olarak adlandırılacaktır. Tam kalıntılar, $\mathbf{y} - \mathbf{X}\hat{\alpha}$, olgusal olarak $\mathbf{y}$ 'nin bir doğrusal kombinasyonudur. Aynı zamanda gösterilebilir ki $\mathbf{y} - \mathbf{X}\hat{\alpha}$ ve $\hat{\alpha}$ bağımsızdır (Diggle *et al.*, 1994) ve bu yüzden α nın ve varyans parametreleri γ nın ortak olabilirlik fonksiyonu $\mathbf{y} - \mathbf{X}\hat{\alpha}$ 'ya ve $\hat{\alpha}$ 'ya dayanan olabilirliklerin çarpımı olarak ifade edilebilir:

$$L(\gamma, \alpha; \mathbf{y}) = L(\gamma; \mathbf{y} - \mathbf{X}\hat{\alpha})L(\alpha; \hat{\alpha}, \gamma).$$

Böylece, γ nın $\mathbf{y} - \mathbf{X}\hat{\alpha}$ 'ya dayanan olabilirliği

$$L(\gamma; \mathbf{y} - \mathbf{X}\hat{\alpha}) = L(\gamma, \alpha; \mathbf{y}) / L(\alpha; \hat{\alpha}, \gamma)$$

ile verilir. Şimdi, yukarıdaki eşitliklerden,

$$L(\gamma, \alpha; \mathbf{y}) \propto |\mathbf{V}|^{-1/2} \exp \left(-\frac{1}{2} (\mathbf{y} - \mathbf{X}\alpha)' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\alpha) \right)$$

ve $\hat{\alpha}$ aritmetik ortalaması ve varyansı ML kestirimlerince verilen çokdeğişkeli bir normal dağılıma sahiptir. Bundan ötürü,

$$L(\alpha; \hat{\alpha}, \gamma) \propto |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|^{-1/2} \exp \left(\frac{1}{2} (\hat{\alpha} - \alpha)' \mathbf{X}\mathbf{V}^{-1}\mathbf{X}(\hat{\alpha} - \alpha) \right).$$

Bu iki olabilirliğin oranını almak suretiyle, REML

$$L(\gamma; \mathbf{y} - \mathbf{X}\hat{\alpha}) \propto |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|^{-1/2} |\mathbf{V}|^{-1/2} \exp \left(-\frac{1}{2} (\mathbf{y} - \mathbf{X}\hat{\alpha})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\alpha}) \right)$$

olarak ve REML log olabilirliği

$$\log(L(\boldsymbol{\gamma}; \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})) = K - \frac{1}{2}[\log |\mathbf{V}| - \log |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|^{-1} + (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})]$$

olarak elde edilir.

Hala $\hat{\boldsymbol{\alpha}}$ görünüyorsa dahi, varyans parametrelerinin bir fonksiyonu suretindedir ($\hat{\boldsymbol{\alpha}}$ daha sonra $\boldsymbol{\alpha} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ olarak türetililecektir). Bu yerine koyma ile, $\boldsymbol{\alpha}$ parametresi çıkarılmış olur. Dikkat edilmeli ki REML log olabilirlik ve sıradan log olabilirlik arasındaki farkın nedeni $\log |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|^{-1}$ terimidir ki $\text{var}(\hat{\boldsymbol{\alpha}})$ 'nın determinantının logaritmasıdır. REML $\boldsymbol{\alpha}$ ve $\boldsymbol{\gamma}$ olabilirliğinin $\boldsymbol{\alpha}$ ile entegre edilmesine eşdeğerdir ve bu sebeple REML ye bazen “marjinal” metot ta denir. REML nin $\boldsymbol{\alpha}$ nın bir sabite değil bir parametre olduğu olgusunu hesaba katması hasebiyle, sonuçta elde edilen varyans parametresi kestirimleri yansızdır. ML ile olduğu üzere, böylece, varyans parametreleri sabit kabul edilerek ve $\boldsymbol{\alpha}$ 'nın REML log olabilirliğini maksimize eden değerleri bulunarak, $\boldsymbol{\alpha}$ kestirilir.

İteratif genelleştirilmiş en-küçük kareler (IGLS)

Bu metot bir karışık modeli uydurmak için iteratif olarak kullanılabilir ve sonuçları ML kullanılarak elde edilenlerle aynı olacaktır. Bu yaklaşım $\boldsymbol{\alpha}$ sabit etki parametrelerinin kestirimlerini, $\mathbf{V}^{-1}$ varyans matrisinin inversi tarafından ağırlıklandırılmış tam kalıntıların çarpımını minimize etmek suretiyle elde eder. Kalıntı çarpımı, $(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})$ tarafından verilir.

Varyans parametreleri, $(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})$ tam kalıntılar çarpımları matrisinin $\mathbf{V}$ varyans matrisine eşitlenmesiyle, varyans parametreleri cinsinden spesifikleştirilmiş olarak elde edilir. Bu,

$$(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}) = \mathbf{V}$$

yi verir ve $n \times n$ simultane eşitliklerinin bir kümesini ($n \times n$ matrislerindeki her bir eleman için bir tane) doğurur, öyle ki bunlar varyans parametreleri için iteratif olarak çözülebilir. Bu eşitlikler elbette $\boldsymbol{\alpha}$ 'nın kestirileceği ve bilinmediği olgusunu hesaba katmaz. Bu yüzden, elde edilen varyans parametresi kestirimleri aşağı-terafa yanlıdır ve ML kestirimleri ile aynıdır.

İteratif genelleştirilmiş en-küçük karelerin (IGLS) bir uyarlaması *kısıtlanmış iteratif genelleştirilmiş en-küçük karelerdir (RIGLS)* ki yansız REML varyans parametresi kestirimlerini verir. Goldstein (1989), $\hat{\boldsymbol{\alpha}}$ nın kestirildiğini ve bilinmediğini not ederek, bunu

$$\mathbf{E}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}) = \mathbf{V} - \text{var}(\mathbf{X}\hat{\boldsymbol{\alpha}}) = \mathbf{V} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'$$

ile betimledi ki bu da

$$(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}) = \mathbf{V} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'$$

varyans parametreleri için çözülecek alternatif bir $n \times n$ eşitlikleri kümesini doğurur. Gözlenen tam kalıntılar, $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}$, sabit etkiler parametre kestirimlerine bağlı olduğu için, hem IGLS hem de RIGLS çözümüne ulaşabilmek sabit etkiler ve varyans parametresi kestirimleri arasında iterasyonu gerektirir (Goldstein, 2010).

Varyans parametresi yanlılığı

İfade edildiği üzere, varyans parametrelerinin ML (ve IGSL) kestirimleri yanlıdır ve REML (ve RIGLS) kestirimleri yansızdır. Yanlılığın olmaması sahip olunan önemli bir özelliktir ve bu yüzden REML veri analizlerinin çoğunda tercihen kullanılmaktadır. Sabit etkilerin kestirimleri ML ve REML analizleri arasında çoğu zaman büyük ölçüde farketmeyecektir, fakat, eğer varyans parametreleri yanlıysa (varyans parametrelerinin ağırlıklandırılmış toplamları olarak hesaplandıkları için), bunların standart hataları daima aşağı- tarafa yanlı olacaktır. Ne zaman ki varyans parametrelerini kestirmek için kullanılan serbestlik derecesi (SD) küçük olursa, bu en çok dikkat çekecektir. Sabit etkilerin kestirimleri de aynı zamanda yanlılığın başka kaynaklarına maruzdur.

Sabit etkilerin kestirimi

Maksimum olabilirlik ve REML

Log olabilirliğin $\boldsymbol{\alpha}$ 'ya göre türevini alarak ve çıkan ifadeyi sıfıra eşitleyerek, olabilirliği maksimize etmek suretiyle, sabit etkiler çözümü elde edilebilir. Bu öyle bir çözüme götürür ki varyans parametreleri cinsinden ifade edilir:

$$\mathbf{X}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}) = \mathbf{0}$$

Yeniden düzenlendiğinde,

$$\hat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

eşitliğini verir ve $\hat{\boldsymbol{\alpha}}$ nın varyansı

$$\begin{aligned} \text{var}(\hat{\boldsymbol{\alpha}}) &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\text{var}(\mathbf{y})\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \end{aligned}$$

olarak elde edilir. Bu formül, $\mathbf{V}$ 'nin bilindiği varsayımına dayanır. Gerçekte, $\mathbf{V}$ kestirildiği için, $\text{var}(\hat{\boldsymbol{\alpha}})$ 'da bir miktar aşağı-tarafa yanlılık olacağı gösterilebilir. Ancak, bu çoğu zaman çok

küçüktür ve yanlışlık için yaklaşık düzeltmeler yapılabilir. Dikkat edilmelidir ki varyans bileşeni kestirimlerinin kendileri yanlış olduğu zaman, bu meydana gelir. Bayesçi yaklaşım bu varsayımda bulunmaktan kaçınır öyle ki yanlışlık problemi ortaya çıkmasın (Akaike, 1978).

İteratif genelleştirilmiş en-küçük kareler

Genelleştirilmiş en-küçük kareler kullanılarak, α 'nın ML ile aynı çözümü elde edilebilir. Bu yaklaşımla, varyansların inversi $(\mathbf{y} - \mathbf{X}\alpha)' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\alpha)$ tarafından ağırlıklandırılmış tam kalıntıların çarpımı, α 'ya göre türevi alınarak, minimize edilir:

Bu ifadeyi sıfıra eşitlemek suretiyle, bulunur ki kalıntı çarpımı, $\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \mathbf{X}'\mathbf{V}^{-1}\alpha$ ise, minimize edilir ki α için

$$\hat{\alpha} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1} \mathbf{y}$$

çözümünü verir ve yine varyansı da

$$\text{var}(\hat{\alpha}) = (\mathbf{X}'\mathbf{V}^{-1} \mathbf{X})^{-1}.$$

Bu çözüm bazen *genelleştirilmiş en-küçük kareler* çözümü olarak adlandırılır.

Ağırlıklandırılmamış en-küçük karelerde, ki $\mathbf{V} = \sigma^2 \mathbf{I}$, çözüm $\hat{\alpha} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$ olacaktır ve varyansı da $(\mathbf{X}'\mathbf{X})^{-1} \sigma^2$. Bu, sabit etkiler modelinin, *sıradan en-küçük kareleri* (OLS) kullanılarak uydurulmasından elde edilen çözümdür. Karışık model kestirimlerindeki farklılık $\hat{\alpha}$ 'nın formülünde $\mathbf{V}^{-1}$ invers varyans matrisinin kullanılmasından dolayıdır. Veri dengesiz olduğu zaman, sabit etkilerdeki enformasyonun farklı hata tabakalarından biraraya gelmesine neden olan bu farklılıktır.

Rastgele etkilerin ve katsayıların kestirimi (veya tahmini)

Genelde, rastgele etkiler ve katsayılar sıfır aritmetik ortalamalı normal dağılımlara sahip olarak tanımlanır ve aldıkları spesifik değerler bir dağılımdan gelen bir örneğin gerçekleştirilmesi olarak düşünülür. Nitekim, bunların beklenen değerleri, tanımı gereği, sıfırdır. Yine de, bir çapraz-geçişli denemede birey etkileri bağlamında görüldüğü üzere, bunların tahminlerini elde etmek olanaklıdır.

Maksimum olabilirlik ve REML

Rastgele etkileri, β , tahmin etmek için (basit olsun diye, hem rastgele etkilere hem de katsayılara atfen rastgele etkiler terimi kullanılıyor), α , β ve γ cinsinden bir olabilirlik fonksiyonu tanımlanır (γ varyans parametreleri vektörüdür). Bu, şu ikisinin olabilirliklerinin çarpımı olarak yazılabilir: (1) $\mathbf{R}$ matrisindeki (α ve $\gamma_{\mathbf{R}}$) sabit parametreler ve varyans

parametreleri şöyle ki β ile koşullandırılmış veri, bir an, sabit kabul edilerek $(y|\beta)$. (2) G matrisindeki (γ_G) varyans parametreleri öyle ki β rastgele etkileri verilerek.

$$L(\alpha, \beta, \gamma; y) = L(\alpha, \gamma_R; y|\beta)L(\gamma_G; \beta)$$

ki burada

$\gamma_R = R$ matrisindeki varyans parametreleri,

$\gamma_G = G$ matrisindeki varyans parametreleri.

$(y|\beta)$ ve β için çokdeğişkeli normal dağılımı kullanmak suretiyle,

$$L(\alpha, \beta, \gamma; y) \propto |R|^{-1/2} \exp\left(-\frac{1}{2}(y - X\alpha - Z\beta)'R^{-1}(y - X\alpha - Z\beta)\right) \\ \times |G|^{-1/2} \exp\left(-\frac{1}{2}\beta'G^{-1}\beta\right)$$

elde edilir ki mütakabil log olabilirliği,

$$\log(L) = -\frac{1}{2}[\log |R| + (y - X\alpha - Z\beta)'R^{-1}(y - X\alpha - Z\beta) \\ + \log |G| + \beta'G^{-1}\beta] + K.$$

olarak verir. Bu log olabilirlik, β ya göre türevi alınarak,

$$\delta \log(L) / \delta \beta = Z'R^{-1}(y - X\alpha - Z\beta) - G^{-1}\beta \\ = -(Z'R^{-1}Z + G^{-1})\beta + Z'R^{-1}(y - X\alpha),$$

sıfıra eşitlenirse,

$$\hat{\beta}(Z'R^{-1}Z + G^{-1}) = Z'R^{-1}(y - X\alpha), \\ \hat{\beta} = (Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}(y - X\alpha).$$

Eğer sabit olarak uydurulmuş olsalardı, elde edilecek kestirimlerle karşılaştırdığında, bu kestirimler küçüktür. Bu kestirimler sıfır etrafında merkezileştiğinden dolayı, aritmetik ortalama rastgele etkiler kestirimlerini elde etmek için, kesim kestiriminin eklenmiş olmasına ihtiyaç vardır. Rastgele etkiler modellerinde, R matrisi diyagonaldir ve alternatif olarak

$$\hat{\beta} = (Z'Z + G^{-1} / \sigma^2)^{-1}Z'(y - X\alpha)$$

yazılabilir. Bir sabit etkiler modelinin OLS çözümü $(\hat{\alpha} = (X'X)^{-1}X'y)$ ile karşılaştırıldığında, paydadaki G^{-1} / σ^2 terimi dikkat çekicidir ki kestirimlerin sıfıra doğru büzülmesine neden olan bu terimdir.

Aynı zamanda, matris manipölasyonu kullanılarak ve $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$ olduđu anımsanarak, $\hat{\boldsymbol{\beta}}$ 'nin daha kompakt bir alternatif formu, yukarıda verilen çözümden elde edilebilir:

$$\hat{\boldsymbol{\beta}} = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}).$$

$\hat{\boldsymbol{\beta}}$ nin varyansı,

$$\text{var}(\hat{\boldsymbol{\beta}}) = \mathbf{GZ}'\mathbf{V}^{-1}\mathbf{ZG} - \mathbf{GZ}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{ZG}$$

olarak elde edilir. $\text{var}(\hat{\boldsymbol{\alpha}})$ ile olduđu üzere, bu formöl $\mathbf{V}$ 'nin bilindiđi varsayımına dayanır. Ama, gerçekte $\mathbf{V}$ kestirildiđi için, $\text{var}(\hat{\boldsymbol{\beta}})$ bir miktar aşağı-tarafta yanlı olacaktır, gerçi bu çođu zaman küçüktür. Yine, Bayesçi yaklaşım, bu varsayımda bulunmaktan kaçınır ve bu yanlılık problemi de ortaya çıkmaz.

İteratif genelleştirilmiş en-küçük kareler

Şu açıktır ki $\hat{\boldsymbol{\beta}}$ her zamanki en-küçük kareler eşitliklerinden, IGLS kullanılarak, doğrudan elde edilemez. Ancak, varyansa parametrelerinin kestirimleri, IGLS kullanılarak, elde edildikten sonra, $\hat{\boldsymbol{\beta}}$ yı elde etmek için yukarıdaki formöl uygulanabilir.

Varyans parametrelerinin kestirimi

Bu alt-bölümde, ML yi uygulamak için kullanılan sayısal prosedürleri ve varyans parametrelerinin kestirimine ait en-küçük karelere dayalı metotları mülahaza edilmektedir. Bunlar, çođu zaman, analizin icrası için kullanılan istatistiksel paketlerin içinde mevcuttur ve bu sebeple bunların nasıl uygulanacağına dair detayların burada verilmesine lüzum görülmemiştir.

Maksimum olabilirlik ve REML

Bu metotların her ikisi de bir olabilirlik fonksiyonunu maksimize eden varyans parametresi kestirimlerini elde etmek suretiyle çalışır. Sabit ve rastgele etkilerde olduđu gibi tek bir eşitlik vasıtasıyla bir çözüm spesifikleştirilemez, çünkü log olabilirliğin varyans parametrelerine göre türevi doğrusal-olmayandır. Yaygın bir şekilde kullanılan *Newton-Raphson algoritması* gibi bir iteratif proses gereklidir. Bu, olabilirlik fonksiyonuna ait kauadratik yaklaşımları yineleyerek çözmek suretiyle çalışır. Şimdi bu algoritmanın nasıl çalıştığını göstermek için önce bir kuadratik fonksiyonun birinci ve ikinci türevleri cinsinden nasıl çözüldüğü gösterilecek ve sonra da bu çözümün bir olabilirlik fonksiyonunun maksimizasyonunda iteratif Newton-Raphson metodunu tanımlamak için nasıl kullanılacağı gösterilecektir.

Bir kuadratiğin çözümü

θ nın bir kuadratik fonksiyonunun birinci ve ikinci türevleri cinsinden çözümünü ifade etmek için, önce fonksiyon genel kuadratik formda matris notasyonu ile yazılır,

$$f(\theta) = a + \mathbf{b}'\theta + \frac{1}{2}\theta'\mathbf{C}\theta.$$

Şuna dikkat edilmelidir ki $f(\theta)$ bir vektör olacaktır ve ikinci türev bir kare matris olacaktır. Mesela, eğer $\theta = (\theta_1, \theta_2, \theta_3)$ ise, o zaman $f(\theta)$ nın birinci türevi 3 uzunluğunda bir vektördür ve log olabilirliğin ikinci türevi 3×3 lük bir matristir. Birinci türev

$$f'(\theta) = \mathbf{b} + \mathbf{C}\theta$$

tarafından verilir ve ikinci türev

$$f''(\theta) = \mathbf{C}$$

tarafından verilir. θ nın $f(\theta)$ yı maksimize eden $\hat{\theta}$ çözümüne, birinci türevi sıfıra eşitlemek suretiyle,

$$f'(\theta) = \mathbf{b} + \mathbf{C}\theta = \mathbf{0},$$

ulaşılır:

$$\hat{\theta} = -\mathbf{C}^{-1}\mathbf{b}.$$

Keyfi bir, diyelim ki, θ_i değeri eklemek ve çıkarmak suretiyle, kuadratik fonksiyonun bu çözümü o zaman

$$\hat{\theta} = \theta_i - \mathbf{C}^{-1}(\mathbf{b} + \mathbf{C}\theta_i)$$

olarak ifade edilebilir ve θ_i cinsinden yeniden yazılabilir, öyle ki $f(\theta)$ nın birinci ve ikinci türevleri θ_i de değerlendirilir:

$$\hat{\theta} = \theta_i - f''^{-1}(\theta_i)f'(\theta_i).$$

Şimdi gösterilebilir ki bu matematiksel manipülasyon Newton-Raphson algoritmasının kullanılmasının anahtarıdır.

Newton-Raphson iterasyonu $f(\theta)$ nın olabilirlik fonksiyonunu maksimize edecek θ değeri bulmak için şu yoldan gidilir. Eğer bir θ_1 başlangıç yaklaşık çözümü ile başlanırsa, şu varsayımda bulunur ki önceden elde edilmiş formülü kullanmak suretiyle ıslah edilmiş bir θ_2 yaklaşımı elde etmek için fonksiyon yaklaşık olarak kuadratik olacaktır. Bu halde, yaklaşık çözüm olarak θ_2 kullanılarak, ıslah edilmiş bir θ_3 yaklaşımı elde etmek için,

proses yinelenir. Gerçi fonksiyon, genel olarak, ML çözümü bölgesinde, kuadratik olmayacaksa da, kuadratik yaklaşım çoğu zaman hayli iyidir ve Newton-Raphson iteratif prosedürü çoğu zaman mütenasipçe yakınsayacaktır. Yakınsama, parametre değerleri ardışık iterasyonlar arasında çok az değiştiği zaman, elde edilir. Bu iteratif proses,

$$\theta_{i+1} = \theta_i - f''^{-1}(\theta_i) \times f'(\theta_i)$$

tarafından ifade edilebilir ki burada $f'(\theta_i)$ ve $f''(\theta_i)$, $f(\theta)$ nın θ_i de değerlendirilmiş gerçek (yakınlaşılmamış) türevleridir. İkinci türevlerin matrisi, $f''(\theta)$, çoğu zaman *Hessian matris* olarak adlandırılır. Karışık modellerde, $f(\theta)$, varyans parametreleri, θ , cinsinden ifade edilmiş log olabilirlik olarak alınır (Gorman and Toman, 1966).

Her iterasyonda türevleri değerlendirme ihtiyacı, Newton-Raphson algoritmasını hesaplamasal olarak yoğunlaştırabilir. İteratif proseste, $f''(\theta_i)$ yerine, enformasyon matrisi olarak bilinen bir matrisin kullanılması vasıtasıyla hesaplama daha kolaylaştırılabilir. Enformasyon matrisi Hessian matrisin beklenen değeridir ve hesaplanması Hessian matrisinden daha kolaydır çünkü korelasyon terimlerinin bazıları sıfırdır. Bu kullanıldığı zaman, proses *skorlandırma metodo* veya *Fisher skorlandırması* olarak adlandırılır. Newton-Raphson algoritmasına göre, bu metodun zayıf başlangıç değerlerine karşı daha sağlam olduğu gösterilmiştir. Mesela SAS ta, PROC MIXED default uydurma metodu olarak ilk iterasyon için Fisher skorlandırmasını ve daha sonrasındaki iterasyonlar için de Newton-Raphson algoritmasını kullanır.

Varyans parametrelerinin kovaryansları Varyans parametrelerinin kesinliklerinin bir belirtimi, onların varyansının bir kestiriminden elde edilebilir ve korelasyon derecesi kovaryanslardan elde edilebilir. Ancak, bu kestirim standart asimtotik (geniş örnek) teoriye dayanır. Varyans parametrelerinin asimtotik kovaryansları log olabilirliğin ikinci kısmi türevlerinin beklenen değerlerinin negatifi tarafından verilir (Searle et al., 1992):

$$\begin{aligned} \text{var}(\theta_i) &= -E\{\delta^2 \log(L)/\delta\theta_i\delta\theta_i\}, \\ \text{cov}(\theta_i, \theta_j) &= -E\{\delta^2 \log(L)/\delta\theta_i\delta\theta_j\}. \end{aligned}$$

Elde edilen kovaryanslar asimtotok teoriye dayandığı ve kestirilen varyans parametrelerinin kendileri ile ilişkili oldukları için, interpretasyonları dikkatlice yapılmalıdır. Aynı zamanda, unutmama ki varyans parametrelerinin dağılımı çoğu zaman simetrik değildir.

İteratif genelleştirilmiş en-küçük kareler

Bu metod, tam kalıntı çarpımlarını varyans parametreleri cinsinden spesifikleştirilmiş V varyans matrisine eşitlemek ve elde edilen eşitlikleri çözmek suretiyle, varyans parametrelerini kestirir. Bu $n \times n$ bir simültane eşitlikler kümesini verir ki varyans parametreleri için iteratif olarak çözülebilir (n = gözlemlerin sayısı). Aşağı-terafa yanlış varyans parametreleri veren sıradan IGLS de, eşitlikler şunlardır:

$$V = (y - X\alpha)(y - X\alpha)'$$

Eşitliklerin sütrüktürünü daha açıkça anlatmak için, küçük bir hipotetik verikümesine ait formlar gösterilecektir. Farzediliyor ki bir yinelenen ölçümler denemesindeki ilk iki bireye müteakip ölçümler yapıldı:

Birey	Ölçüm
1	1
1	2
1	3
2	1
2	2

Her ölçüm çiftine ait ayrı bir kovaryans terimi kullanılan bir modeldeki eşitlikler (yani, “genel” bir kovaryans sütrüktörü ile),

$$\begin{pmatrix} \sigma_1^2 & \theta_{12} & \theta_{13} & 0 & 0 \\ \theta_{12} & \sigma_2^2 & \theta_{23} & 0 & 0 \\ \theta_{13} & \theta_{23} & \sigma_3^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_1^2 & \theta_{12} \\ 0 & 0 & 0 & \theta_{12} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} (y_1 - \mu_1) & (y_1 - \mu_1) & (y_1 - \mu_1) & (y_1 - \mu_1) & (y_1 - \mu_1) \\ \times(y_1 - \mu_1) & \times(y_2 - \mu_2) & \times(y_3 - \mu_3) & \times(y_4 - \mu_4) & \times(y_5 - \mu_5) \\ (y_2 - \mu_2) & (y_2 - \mu_2) & (y_2 - \mu_2) & (y_2 - \mu_2) & (y_2 - \mu_2) \\ \times(y_1 - \mu_1) & \times(y_2 - \mu_2) & \times(y_3 - \mu_3) & \times(y_4 - \mu_4) & \times(y_5 - \mu_5) \\ (y_3 - \mu_3) & (y_3 - \mu_3) & (y_3 - \mu_3) & (y_3 - \mu_3) & (y_3 - \mu_3) \\ \times(y_1 - \mu_1) & \times(y_2 - \mu_2) & \times(y_3 - \mu_3) & \times(y_4 - \mu_4) & \times(y_5 - \mu_5) \\ (y_4 - \mu_4) & (y_4 - \mu_4) & (y_4 - \mu_4) & (y_4 - \mu_4) & (y_4 - \mu_4) \\ \times(y_1 - \mu_1) & \times(y_2 - \mu_2) & \times(y_3 - \mu_3) & \times(y_4 - \mu_4) & \times(y_5 - \mu_5) \\ (y_5 - \mu_5) & (y_5 - \mu_5) & (y_5 - \mu_5) & (y_5 - \mu_5) & (y_5 - \mu_5) \\ \times(y_1 - \mu_1) & \times(y_2 - \mu_2) & \times(y_3 - \mu_3) & \times(y_4 - \mu_4) & \times(y_5 - \mu_5) \end{pmatrix}.$$

tarfından verilir. Bu basit veri dizaynında, bu eşitliklerin sol-tarafındaki ve sağ-tarafındaki müteakabil terimleri eşitlemek şunu verir:

$$\theta_{13} = (y_1 - \mu_1)(y_3 - \mu_3),$$

$$\theta_{23} = (y_2 - \mu_2)(y_3 - \mu_3).$$

θ_{12} için iki eşitlik vardır ve onların ortalamalarından bir kestirim elde edebiliriz:

$$\theta_{12} = [(y_1 - \mu_1)(y_2 - \mu_2) + (y_5 - \mu_5)(y_4 - \mu_4)]/2.$$

Nitekim, bu basit yapay veri kümesinde, kovaryans terimleri, yalnızca bir veya iki bireyin gözlenen kovaryanslarının ortalaması üzereinden hesaplanır. Sahici bir verikümesinde, kovaryanslar daha çok sayıda bireyin gözlenen kovaryanslarının ortalamasından kestirilecektir. Varyans terimleri benzer bir yoldan kestirilebilir.

Bu yaklaşım başka kovaryans biçimlerine de uygulanabilir. Faraza aynı yapay veri kümesi ile, daha basit bir korelasyon biçimi, her ölçüm çifti arasında sabit bir kovaryans (yani, bileşik simetri biçimi) ile uydurulsun, o zaman,

$$\begin{pmatrix} \sigma^2 & \theta & \theta & 0 & 0 \\ \theta & \sigma^2 & \theta & 0 & 0 \\ \theta & \theta & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & \theta \\ 0 & 0 & 0 & \theta & \sigma^2 \end{pmatrix} = \begin{pmatrix} (y_1 - \mu_1) & (y_1 - \mu_1) & (y_1 - \mu_1) & (y_1 - \mu_1) & (y_1 - \mu_1) \\ \times(y_1 - \mu_1) & \times(y_2 - \mu_2) & \times(y_3 - \mu_3) & \times(y_4 - \mu_4) & \times(y_5 - \mu_5) \\ (y_2 - \mu_2) & (y_2 - \mu_2) & (y_2 - \mu_2) & (y_2 - \mu_2) & (y_2 - \mu_2) \\ \times(y_1 - \mu_1) & \times(y_2 - \mu_2) & \times(y_3 - \mu_3) & \times(y_4 - \mu_4) & \times(y_5 - \mu_5) \\ (y_3 - \mu_3) & (y_3 - \mu_3) & (y_3 - \mu_3) & (y_3 - \mu_3) & (y_3 - \mu_3) \\ \times(y_1 - \mu_1) & \times(y_2 - \mu_2) & \times(y_3 - \mu_3) & \times(y_4 - \mu_4) & \times(y_5 - \mu_5) \\ (y_4 - \mu_4) & (y_4 - \mu_4) & (y_4 - \mu_4) & (y_4 - \mu_4) & (y_4 - \mu_4) \\ \times(y_1 - \mu_1) & \times(y_2 - \mu_2) & \times(y_3 - \mu_3) & \times(y_4 - \mu_4) & \times(y_5 - \mu_5) \\ (y_5 - \mu_5) & (y_5 - \mu_5) & (y_5 - \mu_5) & (y_5 - \mu_5) & (y_5 - \mu_5) \\ \times(y_1 - \mu_1) & \times(y_2 - \mu_2) & \times(y_3 - \mu_3) & \times(y_4 - \mu_4) & \times(y_5 - \mu_5) \end{pmatrix}.$$

Bu halde, çözüm, tüm zaman noktası çiftleri üzerinde gözlenen kovaryansların ortalamasıyla, elde edilir, öyle ki

$$\theta = [(y_1 - \mu_1)(y_2 - \mu_2) + (y_4 - \mu_4)(y_5 - \mu_5) + (y_1 - \mu_1)(y_3 - \mu_3) + (y_2 - \mu_2)(y_3 - \mu_3)]/4$$

ve

$$\sigma^2 = \sum_{i=1}^5 (y_i - \mu_i)^2 / 5.$$

Rastgele etkiler ve katsayılar modellerinde, her doğrusal eşitlik birden çok sayıda parametre ihtiva edebilir. Basit ortalama, o zaman, parametre kestirimlerini elde etmek için yeterli olmayacaktır ve doğrusal eşitlik kümelerinin çözümü için standart metotlar

uygulanabilir. Yansız varyans parametreleri veren RIGLS (REML de olduđu üzere) için, eşitlikler şunlardır (Goldstein, 2003):

$$(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}) = \mathbf{V} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'.$$

Kovaryans Strüktürleri

Doğrusal karışık modellerde kovaryans matrisinin strüktürünün seçimi uygun model kurulumu için elzemdir. Yinelemeli ölçümler çalışmalarında temel varsayımlardan biri, farklı bireylerden alınan ölçümlerin birbirinden bağımsız, aynı bireyden alınan ölçümlerin ise birbiri ile korelasyonlu olmasıdır. Doğal olarak, aynı bireylerden alınan yinelemeli ölçümler için belirli bir kovaryans matrisi strüktürünün spesifikleştirilmesi gerekir. (Bagiella ve ark. 2000, Wolfinger 1996, Fitzmaurice ve ark. 2004).

Bu korelasyonu betimleyecek kovaryans yapılarının en yaygın kullanılanları literatürde on üç adettir. Bunlar kendi aralarında homojen kovaryans modelleri ve heterojen kovaryans modelleri olarak başlıca iki gruba ayrılır. Bunlar arasındaki temel fark diyagonal elemanlarıdır. Diyagonal elemanları sabit değerler alan kovaryans yapılarına homojen kovaryans strüktürleri, diyagonal elemanları sabit olmayanlara ise heterojen kovaryans strüktürleri denir.

Tablo 1. Homojen Kovaryans Strüktürleri.

Kovaryans Strüktürü	Notasyon	Matris Gösterimi	Parametre
Varyans Bileşenleri	VC	$\begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix}$	1
Bileşik Simetri	CS	$\begin{bmatrix} \sigma^2 + \sigma_u^2 & \sigma_u^2 & \sigma_u^2 & \sigma_u^2 \\ \sigma_u^2 & \sigma^2 + \sigma_u^2 & \sigma_u^2 & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 & \sigma^2 + \sigma_u^2 & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 & \sigma_u^2 & \sigma^2 + \sigma_u^2 \end{bmatrix}$	2
Toeplitz	TOEP	$\sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{bmatrix}$	n
Birinci mertebeden otoregresif	AR(1)	$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ & 1 & \rho & \rho^2 \\ & & 1 & \rho \\ & & & 1 \end{bmatrix}$	2

Tablo 2. Heterojen Kovaryans Strüktürleri

Kovaryans Strüktürü	Notasyon	Matris Gösterimi	Parametre
Yapılandırılmış	UN	$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$	$n(n+1)/2$
Köşegen	UN(1)	$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$	n
Heterojen bileşik simetri	CSH	$\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho & \sigma_1\sigma_4\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho \\ \sigma_3\sigma_1\rho & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \sigma_4\sigma_1\rho & \sigma_4\sigma_2\rho & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix}$	$n+1$
Heterojen Toeplitz	TOEPH	$\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_2 & \sigma_1\sigma_4\rho_3 \\ \sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_1 & \sigma_2\sigma_4\rho_2 \\ \sigma_3\sigma_1\rho_2 & \sigma_3\sigma_2\rho_1 & \sigma_3^2 & \sigma_3\sigma_4\rho_1 \\ \sigma_4\sigma_1\rho_3 & \sigma_4\sigma_2\rho_2 & \sigma_4\sigma_3\rho_1 & \sigma_4^2 \end{bmatrix}$	$2n-1$
Heterojen birinci mertebeden otoregresif	ARH (1)	$\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho & \sigma_1\sigma_4\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho \\ \sigma_3\sigma_1\rho & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \sigma_4\sigma_1\rho & \sigma_4\sigma_2\rho & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix}$	$n+1$

Tablo 2. Heterojen Kovaryans Strüktürleri (Devamı)

Kovaryans Strüktürü	Notasyon	Matris Gösterimi	Parametre
Huynh-Feldt	HF	$\begin{bmatrix} \sigma_1^2 & \frac{\sigma_1^2 + \sigma_2^2}{2} - \lambda & \frac{\sigma_1^2 + \sigma_3^2}{2} - \lambda & \frac{\sigma_1^2 + \sigma_4^2}{2} - \lambda \\ \frac{\sigma_2^2 + \sigma_1^2}{2} - \lambda & \sigma_2^2 & \frac{\sigma_2^2 + \sigma_3^2}{2} - \lambda & \frac{\sigma_2^2 + \sigma_4^2}{2} - \lambda \\ \frac{\sigma_3^2 + \sigma_1^2}{2} - \lambda & \frac{\sigma_3^2 + \sigma_2^2}{2} - \lambda & \sigma_3^2 & \frac{\sigma_3^2 + \sigma_4^2}{2} - \lambda \\ \frac{\sigma_4^2 + \sigma_1^2}{2} - \lambda & \frac{\sigma_4^2 + \sigma_2^2}{2} - \lambda & \frac{\sigma_3^2 + \sigma_3^2}{2} - \lambda & \sigma_4^2 \end{bmatrix}$	$n+1$
Birinci mertebeden anti-bağımlı	ANTE (1)	$\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_1\rho_2 & \sigma_1\sigma_4\rho_1\rho_2\rho_3 \\ \sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_2 & \sigma_2\sigma_4\rho_2\rho_3 \\ \sigma_3\sigma_1\rho_2\rho_1 & \sigma_3\sigma_2\rho_2 & \sigma_3^2 & \sigma_3\sigma_4\rho_3 \\ \sigma_4\sigma_1\rho_3\rho_2\rho_1 & \sigma_4\sigma_2\rho_3\rho_2 & \sigma_4\sigma_3\rho_3 & \sigma_4^2 \end{bmatrix}$	$2n-1$
Birinci mertebeden faktör analitik	FA(1)	$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} \begin{bmatrix} \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 \end{bmatrix} + \begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix}$	$2n$
Yapılandırılmamış korelasyonlu	UNR	$\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{21} & \sigma_1\sigma_3\rho_{31} & \sigma_1\sigma_4\rho_{41} \\ \sigma_2\sigma_1\rho_{21} & \sigma_2^2 & \sigma_2\sigma_3\rho_{32} & \sigma_2\sigma_4\rho_{42} \\ \sigma_3\sigma_1\rho_{31} & \sigma_3\sigma_2\rho_{32} & \sigma_3^2 & \sigma_3\sigma_4\rho_{43} \\ \sigma_4\sigma_1\rho_{41} & \sigma_4\sigma_2\rho_{42} & \sigma_4\sigma_3\rho_{43} & \sigma_4^2 \end{bmatrix}$	$n(n+1)/2$

MATERYAL ve YÖNTEM

Materyal

Çalışmada materyal olarak simülasyon çalışmasına ait üretilen ham veriler kullanılmıştır.

Yöntem

Bu bölümde simülasyon çalışması için kullanılacak yöntem betimlenmiştir. Önce, aritmetik ortalamalar için spesifikleştirilen bir model ile ve aynı birey üzerinde bilinen bir kovaryans strüktürüne göre korelasyonlu yinelenen ölçümler ile veri üretilmiştir. Daha sonra, iki karışık model (bir tam model ve bir indirgenmiş model) veriye uydurulmuş ve modellerden birini seçmek için farklı prediktif yaklaşımlar kullanılmıştır. Doğru modelin seçildiği denemelerin oranısı kaydedilmiştir.

Veri üretimi prosesinin bir açıklaması bu bölümün birinci alt-bölümünde ana hatlarıyla verilmektedir. İkinci alt-bölüm model seçimi için R^2_{adj} , PRESS, CCC_{adj} , enformasyon kriterleri ve F -testinin performanslarını sorgulamak için, simülasyonların kullanılmasını tartışılmıştır. Bu istatistikler için marjinal ve koşullu yaklaşımlar, farklı varyans-kovaryans strüktürleri altında, öğrenilmeye çalışılmıştır. Simülasyonlarda ve veri analizlerinde SAS, SPSS ve R yazılım paketlerine ait prosedürler yeri geldikçe kullanılmıştır (SAS Institute Inc., 2004, 2007; R Core Team, 2017; Baayen et al., 2008; Singmann and Kellen, 2018; SPSS, 2009).

Veri üretimi

Yinelenen ölçümler dizaynına göre, verinin simülasyonu yapılmıştır. Her veri kümesi 10 veya 30 birey ($n = 10, 30$) ve her denek için sabit bir sayıda ölçüm ($n_i = 3, 5; i = 1, \dots, n$) ihtiva etmektedir. Yinelenen ölçümler için dört kovaryans strüktürü ve sabit etkiler için altı olanaklı model tespit edildi. Böylece, $2 \times 2 \times 4 \times 6 = 96$ simülasyon çalışması yürütülmüştür.

Kovaryans strüktürleri her birey için aynı uygulanmıştır ($V_i = V, i = 1, \dots, n$). Dört strüktür CS(0.5), CS(0.9), ARRE(0.5) ve ARRE(0.9) olarak simgelenmiştir. Bu kovaryans strüktürlerinin, yinelenen ölçümlerin her iki durumu için detayları Tablo 3'te ve Tablo 4'te verilmiştir.

Tablo 3. Simülasyonlarda Kullanılan Bileşik Simetri Kovaryans Strüktürleri için Parametre Değerleri

n_i	Bileşik Simetri, CS(0.5)	Bileşik Simetri, CS(0.9)
3	$\begin{pmatrix} 1 & & \\ .5 & 1 & \\ .5 & .5 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & & \\ .9 & 1 & \\ .9 & .9 & 1 \end{pmatrix}$
5	$\begin{pmatrix} 1 & & & & \\ .5 & 1 & & & \\ .5 & .5 & 1 & & \\ .5 & .5 & .5 & 1 & \\ .5 & .5 & .5 & .5 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & & & & \\ .9 & 1 & & & \\ .9 & .9 & 1 & & \\ .9 & .9 & .9 & 1 & \\ .9 & .9 & .9 & .9 & 1 \end{pmatrix}$

Tablo 4. Simülasyonlarda Kullanılan Birinci Mertebeden Otoregresif ve rastgele Etki Kovaryans Strüktürleri için Parametre Değerleri

n_i	Birinci Mertebeden Otoregresif ve Rastgele Etki, ARRE(0.4)	Birinci Mertebeden Otoregresif ve Rastgele Etki, ARRE(0.8)
3	$\begin{pmatrix} 1 & & \\ .6 & 1 & \\ .4 & .6 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & & \\ .92 & 1 & \\ .848 & .92 & 1 \end{pmatrix}$
5	$\begin{pmatrix} 1 & & & & \\ .6 & 1 & & & \\ .4 & .6 & 1 & & \\ .3 & .4 & .6 & 1 & \\ .25 & .3 & .4 & .6 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & & & & \\ .92 & 1 & & & \\ .848 & .92 & 1 & & \\ .7832 & .848 & .92 & 1 & \\ .72488 & .7832 & .848 & .92 & 1 \end{pmatrix}$

Betimlenen sekiz kovaryans matrisi için simülasyon verisi Ripley'in (1987) iki adım metoduna göre üretilmiştir. Birinci adımda, bir rastgele çokdeğişkeli normal vektör üretilmiştir, öyle ki $E(y) = 0$ ve $Var(y) = I$, ki burada I özdeşlik matrisidir. İkinci adımda, bu vektör tespit edilen kovaryans strüktürüne tekabül eden kovaryans matrisinin Cholesky dekompozisyonu ile çarpılmıştır. Elde edilen rastgele vektör sıfırlı bir aritmetik ortalama vektörüne ve V kovaryans matrisine sahiptir (Tablo 3 ve 4).

Verinin Modellenmesi

Modellerin sabit etkiler parçası $X_i = [X_0 \ X_1 \ X_2]$ dizayn matrislerinin ve $\beta_i = [\beta_0 \ \beta_1 \ \beta_2]^T$ parametre vektörlerinin bir kombinasyonu ile tanımlanmıştır. $n = 5$ ile yapılan simülasyon çalışmalarında, X_i tüm bireyler için aynıdır ve

$$X_i = \begin{pmatrix} 1 & -3 & 1 \\ 1 & -2 & 2 \\ 1 & 0 & 3 \\ 1 & 2 & 1 \\ 1 & 3 & 2 \end{pmatrix}.$$

ile verilirken, $n = 3$ ile yapılan simülasyon çalışmalarında,

$$X_i = \begin{pmatrix} 1 & -3 & 0 \\ 1 & 0 & 1 \\ 1 & 3 & 0 \end{pmatrix}.$$

Bu dizayn matrisleri Tablo 5'te verilen parametre değerleri ile birleştirilmiştir. $\beta_2 = 0.6$ için, beklenen değerlerin bazıları 2 standart sapma kadar değişmektedir.

PROC MIXED'ta tatbik edildiği üzere, REML kullanılarak, her bir veri kümesine iki karışık model uydurulmuştur. Her iki model için, doğru kovaryans strüktürü RANDOM ve REPEATED beyanları kullanılarak spesifikleştirilmiştir. Yine, sabit etkiler için iki model, tam model ve indirgenmiş model, MODEL beyanı kullanılarak spesifikleştirilmiştir. Tam model için, MODEL beyanı $y = X_1 X_2$ olarak ve indirgenmiş model için $y = X_1$ olarak spesifikleştirilmiştir.

Tablo 5. Simülasyonda Kullanılan β Değerleri

$\beta = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$	$\beta = \begin{pmatrix} 1 \\ 1 \\ .05 \end{pmatrix}$	$\beta = \begin{pmatrix} 1 \\ 1 \\ .1 \end{pmatrix}$
$\beta = \begin{pmatrix} 1 \\ 1 \\ .2 \end{pmatrix}$	$\beta = \begin{pmatrix} 1 \\ 1 \\ .4 \end{pmatrix}$	$\beta = \begin{pmatrix} 1 \\ 1 \\ .6 \end{pmatrix}$

Her model için, PRESS, CCC_{adj} , R^2_{adj} , pseudo F -test ve enformasyon kriterleri (AIC ve BIC) hesaplanmıştır ve kaydedilmiştir. Enformasyon kriterlerinin ve F -testlerinin bazıları PROC MIXED tarafından default olarak verilmektedir. F -testi için Kenward-Roger metodu, MODEL beyanındaki `ddfm=kenwardroger` opsiyonu kullanılmak suretiyle spesifikleştirilmiştir. PRESS, CCC_{adj} ve R^2_{adj} için, hem marjinal hem de koşullu kalıntılar göz önüne alınmıştır. Marjinal kalıntı $r_m = y_i - x_i' \hat{\beta}$ gözlenen veri ve kestirilen (marginal)

aritmetik ortalama arasındaki farktır. Koşullu kalıntı $r_c = y_i - x_i' \hat{\beta} - z_i \hat{u}$ gözlenen veri ve gözlemin tahmin edilen değeri arasındaki farktır. Marjinal ve koşullu kalıntılar, PROC MIXED'in model beyanında *outpredm* ve *outpred* opsiyonları kullanılarak, verilmiştir. Uydurma istatistiklerini ve kalıntılarını almak için ODS (açıktı dağıtım sistemi) kullanılmıştır. Sonrasında, PROC ML vasıtasıyla, ve lme4 ve lmmfit R paketleri vasıtasıyla PRESS, CCC_{adj} ve R^2_{adj} hesaplanmıştır.

PRESS istatistiği gözlenen değerler ve tahmin edilen değerler arasındaki kareli farkların toplamıdır, öyle ki her tahmin edilen değere, ilgilenilen gözlem hariç tutularak, ulaşılır. PRESS'in iki versiyonu hesaplanmıştır, marjinal PRESS ve koşullu PRESS. marjinal PRESS'i hesaplamak için, PROC MIXED'in bir opsiyonu kullanıldı. Koşullu PRESS'i hesaplamak için lmmfit R paketi kullanılmıştır.

CCC_{adj} hesaplanırken,

$$CCC = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)' (y_i - \hat{y}_i)}{\sum_{i=1}^n (y_i - \bar{y})' (y_i - \bar{y}) + \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})' (\hat{y}_i - \bar{\hat{y}}) + N(\bar{y} - \bar{\hat{y}})^2}$$

ve

$$CCC_{adj} = 1 - k^* (1 - CCC)$$

ifadeleri kullanılmıştır ki burada

$$k^* = \frac{N}{N - s}$$

ve s modeldeki parametre sayısıdır. CCC_{adj} 'nin dört versiyonu hesaplanmıştır: marjinal CCC_{adj} ki $s = p$, koşullu CCC_{adj} ki $s = p$, marjinal CCC_{adj} ki $s = p + k$, koşullu CCC_{adj} ki $s = p + k$.

Düzeltilmiş R^2 için eşitlikler:

$$R^2 = 1 - \frac{\sum_{i=1}^n r_i^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$$

ve

$$R_{adj}^2 = 1 - k^* (1 - R^2)$$

ki burada

$$k^* = \frac{N}{N - s}.$$

R^2 'nin dört versiyonu hesaplanmıştır: marjinal R^2 ki $s = p$, koşullu R^2 ki $s = p$, marjinal R^2 ki $s = p + k$, koşullu R^2 ki $s = p + k$.

AIC ve BIC için, parsimonik formüller kullanılmıştır:

$$AIC = -2l + 2s \text{ ve } BIC = -2l + s(\log N^*)$$

ki burada $l \frac{1}{2} \log |\sum_{i=1}^m X_i' X_i|$ terimli veya terimsiz kısıtlanmış log-olabilirlik fonksiyonudur, $s = p + k$ ve $N^* = m$ veya $N^* = N - p$ dir. Bu yüzden, AIC'in ve BIC'in ikişer versiyonları göz önüne alınmıştır (Edwards et al., O., 2008).

Tam veya indirgenmiş modeli seçerken, her iterasyon için on yedi uyum iyiliği istatistiği hesaplanmıştır. Bu kriterlerin doğru modeli seçme (eğer $\beta_2 > 0$ ise tam model, eğer $\beta_2 = 0$ ise indirgenmiş model) sayısı her durum için kaydedilmiştir. Bu kriterlerin başarı oranları daha sonra karşılaştırılmıştır.

ARAŞTIRMA BULGULARI ve TARTIŞMA

Düzeltilmiş model seçim oranları, on yedi kriter için tablolandırıldı (Tablo 3, 4, 5 ve 6) ve grafiklerle gösterildi (Figür 1, 2, 3 ve 4). Tüm istatistikler, denek sayıları, ölçüm sayıları ve kovaryans strüktürleri bakımından, $\beta_2 = 0$ durumunda başarı oranı yüksek bulunmuştur. Genel olarak, denek sayısı ve/veya ölçüm sayısı artırıldığı zaman, istatistikler daha iyi performans göstermiştir. Koşullu istatistikler, marjinal istatistiklere göre, $\beta_2 = 0$ durumu hariç, her halde daha iyi performans göstermiştir (Tablo 7). Ortalamada, CCC'nin birinci tipi ve R^2 ($s = p$ ile hesaplanan), CCC'nin ikinci tipine ($s = p + k$ ile hesaplanan) göre biraz daha iyi performans göstermiştir. Daha iyi performans kriteri, $\beta_2 = 0$ durumunda, daha yüksek başarı oranına sahip olmuştur ve β_2 0.5'ten 0.6'ya çıkarıldığında daha hızlı bir artış oranına (daha dik bir eğime) sahip olmuştur.

İstatistikler

PRESS istatistiği ARRE için CS'ye göre daha iyi performans göstermiştir. Kovaryans strüktürü CS olduğu zaman, marjinal ve koşullu PRESS'in performansları benzer bulunmuştur. Fakat, kovaryans strüktürü ARRE olduğu zaman, koşullu PRESS marjinal PRESS'e göre çok daha yüksek performans göstermiştir.

CCC için, $s = p$ düzeltmesi ile $s = p + k$ düzeltmesi arasında bir fark görülmemiştir. ARRE(1) için, marjinal CCC koşullu CCC'den daha iyi performansa sahip olmuştur. CS için, marjinal CCC koşullu CCC'ye göre daha yüksek bir başarı oranına sahip olmuştur.

R^2 'nin performansı CCC'nin performansı ile benzer bulunmuştur. Hem $s = p$ hem de $s = p + k$ düzeltmeleri ile benzer sonuçlar vermiştir. Marjinal R^2 'nin performansı koşullu R^2 'ye göre daha yüksek bulunmuştur. Koşullu PRESS CCC'nin ve R^2 'nin her versiyonundan daha iyi performans göstermiştir. R^2 ve CCC (özellikle marjinal) β_2 'nin 0.1, denek sayısının 10 ve ölçüm sayısının 3 olduğu durumda (küçük örnek hacmi), gözle görülür bir sıçrama yapmıştır. R^2 ve CCC ölçüm sayısı 3 iken, ARRE(1) için çok kötü bir performansa sahip olmuştur.

AIC1 ($\frac{1}{2}\log|\sum_{i=1}^m X'X_i|$ terimli), β_2 'nin 0 olduğu durumda, özellikle daha düşük korelasyonlarla, daha iyi bir performans göstermiştir. AIC2 ($\frac{1}{2}\log|\sum_{i=1}^m X'X_i|$ terimsiz), β_2 'nin 0'dan büyük olduğu durumda, daha iyi bir performans göstermiştir. AIC2'nin performansı,

yüksek korelasyon durumunda PRESS ile benzer fakat düşük korelasyon durumunda daha kötü bulunmuştur.

KR ile F-testi CS ve $\beta_2 = 0$ durumunda ARRE(1) için %95'in biraz üzerinde bir başarı oranına sahip olmuştur. F-testinin performansı, CS için, PRESS ile benzer fakat ARRE(1) için daha düşük bulunurken, AIC1 ile benzer fakat $\beta_2 > 0$ durumunda AIC2'den daha düşük bulunmuştur.

BIC'in dört tipi de birbirinden farklı sonuçlar vermiştir. $\beta_2 = 0$ durumunda, BIC1 ($\frac{1}{2} \log|\sum_{i=1}^m X'X_i|$ ve $N^* = N - p$ terimli) ve BIC2 ($\frac{1}{2} \log|\sum_{i=1}^m X'X_i|$ ve $N^* = N - m$ terimli) en iyi performansı göstermişken (birbirleriyle benzer), $\beta_2 > 0$ durumunda, BIC3 ($\frac{1}{2} \log|\sum_{i=1}^m X'X_i|$ ve $N^* = N - p$ terimsiz) ve BIC2 ($\frac{1}{2} \log|\sum_{i=1}^m X'X_i|$ ve $N^* = N - m$ terimsiz) en iyi performansı göstermiştir (BIC4'ün performansı BIC3'e göre daha iyi).

Örnek Hacmi

Denek sayısının 10 ve denek başına ölçüm sayısının 3 olduğu duruma göre, denek sayısının 20 ve denek başına ölçüm sayısının 5 olduğu durumda, on yedi istatistiğin tümü çok daha iyi performans göstermiştir. Denek sayısının 20 ve denek başına ölçüm sayısının 5 olduğu durumdaki performanslar, denek sayısının 20 ve denek başına ölçüm sayısının 3 olduğu durumdaki performanslar ile benzer fakat denek sayısının 10 ve denek başına ölçüm sayısının 3 olduğu durumdan daha yüksek bulunmuştur. Tüm istatistikler denek sayısının ve ölçüm sayısının daha çok olduğu durumlarda daha iyi performans göstermişlerdir.

Kovaryans Strüktürü

R^2 , CCC ve PRESS bileşik simetrik kovaryans strüktürü için, birinci mertebeden otoregresif kovaryans strüktürüne göre, daha iyi performansa sahip olmuşlardır. İstatistiklerin diğerleri (AIC, BIC, F-test) kovaryans strüktüründen gözle görülür bir şekilde etkilenmemişlerdir. Yüksek korelasyonlar ile tüm istatistikler daha yüksek başarı yüzdelere ulaşmışlardır (Hannan and Quinn, 1979).

Tablo 6. Doğru Modelin Seçiminde İstatistiklerin Başarı Orantıları (Kovaryans Strüktürü = Bileşik Simetri, 0.5)

Kovaryans Strüktürü	Parametre	Birey Sayısı	Ölçüm Sayısı	β	Marjinal PRESS	Koşullu PRESS	Marjinal CCC1	Koşullu CCC1	Marjinal CCC2	Koşullu CCC2	F-testi	Marjinal R_1^2	Koşullu R_1^2
CS	0.5	10	3	0	0.87781	0.90417	0.76515	0.61750	0.88996	0.67885	0.85240	0.77955	0.71940
CS	0.5	10	3	0.05	0.23250	0.10163	0.23994	0.30081	0.12124	0.42335	0.01416	0.09721	0.32714
CS	0.5	10	3	0.1	0.05901	0.15727	0.29739	0.48948	0.17718	0.32783	0.06049	0.15476	0.49108
CS	0.5	10	3	0.2	0.31160	0.29373	0.23525	0.40311	0.31434	0.54647	0.07597	0.24407	0.54005
CS	0.5	10	3	0.4	0.50395	0.42662	0.61216	0.66294	0.48876	0.73212	0.34826	0.63090	0.79117
CS	0.5	10	3	0.6	0.73401	0.67030	0.71272	0.83183	0.70423	0.86707	0.61147	0.80132	0.90241
CS	0.5	10	5	0	0.94488	0.92759	0.78907	0.53140	0.87175	0.63019	0.94440	0.87735	0.57273
CS	0.5	10	5	0.05	0.18352	0.11352	0.12050	0.49250	0.18307	0.42944	0.04518	0.24995	0.35006
CS	0.5	10	5	0.1	0.17504	0.23687	0.22913	0.55878	0.22504	0.45879	0.10059	0.25397	0.43186
CS	0.5	10	5	0.2	0.49772	0.53311	0.57221	0.61830	0.56882	0.69899	0.36626	0.62809	0.73692
CS	0.5	10	5	0.4	0.83532	1.00000	0.87393	0.98120	0.96522	0.89171	0.83251	1.00000	1.00000
CS	0.5	10	5	0.6	0.97147	1.00000	1.00000	0.98889	0.99720	0.90981	0.99520	1.00000	1.00000
CS	0.5	20	3	0	0.86138	0.83740	0.72111	0.65834	0.83411	0.56424	1.00000	0.88475	0.64462
CS	0.5	20	3	0.05	0.13985	0.15668	0.20230	0.36516	0.11963	0.36026	0.09233	0.15207	0.34314
CS	0.5	20	3	0.1	0.11486	0.28566	0.30837	0.53909	0.30507	0.54090	0.11273	0.21669	0.48184
CS	0.5	20	3	0.2	0.36983	0.32085	0.45727	0.53484	0.45217	0.59130	0.08331	0.43536	0.58799
CS	0.5	20	3	0.4	0.72402	0.72469	0.84358	0.88601	0.65971	0.87760	0.55203	0.79856	0.97169
CS	0.5	20	3	0.6	0.99121	0.94259	0.90975	1.00000	1.00000	1.00000	0.76181	0.99204	1.00000
CS	0.5	20	5	0	0.80984	0.77134	0.76054	0.66423	0.90018	0.56333	1.00000	0.79876	0.67443
CS	0.5	20	5	0.05	0.15126	0.27515	0.28964	0.36062	0.32918	0.39766	0.16789	0.26984	0.39296
CS	0.5	20	5	0.1	0.28326	0.41775	0.32893	0.54686	0.44185	0.51093	0.14377	0.37219	0.63886
CS	0.5	20	5	0.2	0.77777	0.80486	0.86950	0.86378	0.78172	0.97690	0.59746	0.80005	0.85267
CS	0.5	20	5	0.4	0.92343	1.00000	1.00000	1.00000	0.92332	0.90440	1.00000	1.00000	0.89940
CS	0.5	20	5	0.6	1.00000	1.00000	1.00000	1.00000	0.94995	0.96697	0.98198	1.00000	1.00000

Tablo 6. Doğru Modelin Seçiminde İstatistiklerin Başarı Orantıları (Kovaryans Strüktürü = Bileşik Simetri, 0.5) (Devamı)

Kovaryans Strüktürü	Parametre	Birey Sayısı	Ölçüm Sayısı	β	Marjinal R_2^2	Koşullu R_2^2	AIC1	AIC2	BIC1	BIC2	BIC3	BIC4
CS	0.5	10	3	0	0.83009	0.64181	0.88499	0.67705	0.92077	0.83522	0.90306	0.70270
CS	0.5	10	3	0.05	0.15028	0.46639	0.09388	0.27850	0.00805	0.11883	0.11314	0.33083
CS	0.5	10	3	0.1	0.26238	0.42783	0.14672	0.39049	0.15980	0.12702	0.25246	0.31730
CS	0.5	10	3	0.2	0.24899	0.50553	0.26608	0.45307	0.04864	0.22476	0.34888	0.38327
CS	0.5	10	3	0.4	0.43022	0.72431	0.52180	0.67715	0.39620	0.32023	0.54453	0.67841
CS	0.5	10	3	0.6	0.73067	0.81192	0.79740	0.89530	0.57160	0.74898	0.83065	0.82295
CS	0.5	10	5	0	0.78485	0.57302	0.94769	0.59256	1.00000	0.98963	0.83554	0.62862
CS	0.5	10	5	0.05	0.12864	0.32805	-0.01547	0.30251	0.07235	0.00716	0.12769	0.23091
CS	0.5	10	5	0.1	0.25068	0.40034	0.09850	0.55408	-0.00996	0.00219	0.13923	0.47177
CS	0.5	10	5	0.2	0.54062	0.66836	0.27399	0.78386	0.08642	0.16370	0.54970	0.57731
CS	0.5	10	5	0.4	1.00000	0.96909	0.79528	1.00000	0.75758	0.75385	0.96305	0.87871
CS	0.5	10	5	0.6	0.94595	0.97187	0.99000	1.00000	0.88812	0.92704	0.96527	1.00000
CS	0.5	20	3	0	0.88374	0.66473	0.94962	0.72208	0.96218	0.94758	0.94904	0.86294
CS	0.5	20	3	0.05	0.27370	0.45857	0.14317	0.45138	0.08806	0.09876	0.17777	0.11042
CS	0.5	20	3	0.1	0.15964	0.39525	0.16506	0.42531	0.04048	0.03017	0.04961	0.30637
CS	0.5	20	3	0.2	0.26200	0.67088	0.18958	0.48140	0.04424	0.10276	0.31074	0.47198
CS	0.5	20	3	0.4	0.74831	0.87800	0.62745	0.81314	0.33966	0.43986	0.72968	0.74258
CS	0.5	20	3	0.6	0.92858	0.97649	0.87769	0.97349	0.78943	0.83491	0.85693	0.87321
CS	0.5	20	5	0	0.78565	0.59465	1.00000	0.64671	1.00000	0.88500	0.86233	0.91519
CS	0.5	20	5	0.05	0.18456	0.39247	0.12067	0.40337	0.04076	0.09747	0.17387	0.26583
CS	0.5	20	5	0.1	0.46289	0.52935	0.14432	0.63205	0.11267	0.05797	0.30318	0.37688
CS	0.5	20	5	0.2	0.73248	0.93346	0.43488	0.95998	0.16961	0.32066	0.57197	0.74307
CS	0.5	20	5	0.4	0.94895	0.99449	0.91004	1.00000	0.91910	0.92477	0.93975	1.00000
CS	0.5	20	5	0.6	0.93994	0.93794	0.94494	0.91692	0.96587	1.00000	0.98999	1.00000

Tablo 7. Doğru Modelin Seçiminde İstatistiklerin Başarı Orantıları (Kovaryans Strüktürü = Bileşik Simetri, 0.9)

Kovaryans Strüktürü	Parametre	Birey Sayısı	Ölçüm Sayısı	β	Marjinal PRESS	Koşullu PRESS	Marjinal CCC	Koşullu CCC	Marjinal CCC ₂	Koşullu CCC ₂	F-testi	Marjinal R ²	Koşullu R ²
CS	0.9	10	3	0	0.94519	0.91539	1.00000	0.58871	0.90311	0.54285	0.93408	0.92724	0.66698
CS	0.9	10	3	0.05	0.08912	0.14804	-0.03625	0.39664	0.01350	0.48875	0.00137	0.10499	0.43088
CS	0.9	10	3	0.1	0.26252	0.24825	-0.06229	0.58154	-0.01084	0.44802	0.06004	0.10258	0.44720
CS	0.9	10	3	0.2	0.60449	0.49320	0.11828	0.77537	0.11530	0.73534	0.36041	0.07686	0.77317
CS	0.9	10	3	0.4	0.99496	0.86811	0.63033	1.00000	0.60044	0.91152	0.88852	0.62016	1.00000
CS	0.9	10	3	0.6	1.00000	0.90701	0.93528	1.00000	0.99045	1.00000	1.00000	0.90577	0.93784
CS	0.9	10	5	0	0.77694	0.81439	1.00000	0.67505	1.00000	0.54591	0.95260	0.89410	0.54362
CS	0.9	10	5	0.05	0.19531	0.28838	0.03978	0.61159	0.05989	0.54113	0.16103	0.02275	0.59478
CS	0.9	10	5	0.1	0.46082	0.49801	0.21527	0.69731	0.22670	0.72224	0.26930	0.04622	0.83375
CS	0.9	10	5	0.2	0.89754	1.00000	0.72690	0.91272	0.68738	1.00000	0.94855	0.64425	1.00000
CS	0.9	10	5	0.4	0.92392	1.00000	0.96156	1.00000	0.93353	0.94795	0.99800	0.97197	1.00000
CS	0.9	10	5	0.6	1.00000	0.90290	0.99900	1.00000	0.91191	1.00000	1.00000	0.92893	0.97598
CS	0.9	20	3	0	0.88380	0.78145	0.92123	0.52404	0.93624	0.68409	0.85140	1.00000	0.66007
CS	0.9	20	3	0.05	0.24472	0.24371	0.03502	0.41794	-0.06448	0.45548	0.03214	-0.03295	0.49451
CS	0.9	20	3	0.1	0.39415	0.39338	0.11247	0.64909	0.05382	0.62688	0.28558	-0.00504	0.67172
CS	0.9	20	3	0.2	0.64687	0.86655	0.20445	1.00000	0.30026	0.83450	0.54263	0.18616	0.97845
CS	0.9	20	3	0.4	1.00000	1.00000	1.00000	0.92092	1.00000	1.00000	0.92993	1.00000	1.00000
CS	0.9	20	3	0.6	0.93462	1.00000	1.00000	0.98799	0.91381	1.00000	0.92993	1.00000	1.00000
CS	0.9	20	5	0	0.87161	0.74851	1.00000	0.60418	0.90781	0.71679	0.92908	0.97838	0.53050
CS	0.9	20	5	0.05	0.34548	0.39307	0.01506	0.59816	0.10406	0.60607	0.30329	0.08304	0.56083
CS	0.9	20	5	0.1	0.77880	0.73842	0.24186	0.96263	0.27120	1.00000	0.62728	0.20405	0.90297
CS	0.9	20	5	0.2	1.00000	0.92022	1.00000	1.00000	0.84320	0.96997	0.96987	0.84622	0.96096
CS	0.9	20	5	0.4	0.99199	1.00000	0.98799	1.00000	0.94595	0.90991	0.93994	1.00000	1.00000
CS	0.9	20	5	0.6	0.94995	1.00000	0.93293	0.91592	0.94995	1.00000	0.90390	1.00000	0.93794

Tablo 7. Doğru Modelin Seçiminde İstatistiklerin Başarı Orantıları (Kovaryans Strüktürü = Bileşik Simetri, 0.9) (Devamı)

[illegible]

Tablo 8 Doğru Modelin Seçiminde İstatistiklerin Başarı Orantıları (Kovaryans Strüktürü = Birinci mertebeden Otoregresif ve Rastgele Etki, 0.5)

Kovaryans Strüktürü	Parametre	Birey Sayısı	Ölçüm Sayısı	β	Marjinal PRESS	Koşullu PRESS	Marjinal CCC	Koşullu CCC	Marjinal CCC ₂	Koşullu CCC ₂	F-testi	Marjinal R ²	Koşullu R ²
ARRE(1)	0.5	10	3	0	0.74238	0.78726	0.88320	0.62358	0.87497	0.75778	0.87210	0.79690	0.72637
ARRE(1)	0.5	10	3	0.05	0.30927	0.10761	0.08916	0.40054	0.01711	0.23362	0.00714	0.05974	0.40496
ARRE(1)	0.5	10	3	0.1	0.35761	0.25894	0.24209	0.33348	0.19998	0.39178	0.13030	0.07364	0.38874
ARRE(1)	0.5	10	3	0.2	0.34254	0.28006	0.34579	0.30630	0.24152	0.39133	0.15888	0.16623	0.43625
ARRE(1)	0.5	10	3	0.4	0.50379	0.71930	0.61034	0.56457	0.44411	0.47681	0.26540	0.49915	0.43165
ARRE(1)	0.5	10	3	0.6	0.70910	0.87321	0.76041	0.60794	0.90928	0.51096	0.64058	0.89996	0.64689
ARRE(1)	0.5	10	5	0	0.95107	0.87495	0.80326	0.69590	0.88102	0.63082	0.89363	0.93157	0.71521
ARRE(1)	0.5	10	5	0.05	0.20897	0.23112	0.26459	0.29867	0.13978	0.24333	0.06501	0.15208	0.33351
ARRE(1)	0.5	10	5	0.1	0.24439	0.19860	0.15090	0.28161	0.19636	0.31896	0.20019	0.33498	0.32875
ARRE(1)	0.5	10	5	0.2	0.19086	0.67098	0.48947	0.50939	0.50520	0.47908	0.29768	0.62939	0.46063
ARRE(1)	0.5	10	5	0.4	0.32304	0.87061	0.92237	0.72728	0.86312	0.82268	0.84647	0.92087	0.67141
ARRE(1)	0.5	10	5	0.6	0.48939	1.00000	1.00000	0.75413	0.91121	0.88146	1.00000	1.00000	0.73771
ARRE(1)	0.5	20	3	0	0.53312	0.90848	0.89090	0.62883	0.89740	0.68296	0.88512	0.98799	0.66065
ARRE(1)	0.5	20	3	0.05	0.33524	0.21453	0.04131	0.31361	0.15724	0.44236	0.03750	0.14082	0.34876
ARRE(1)	0.5	20	3	0.1	0.31470	0.31488	0.26789	0.35885	0.24498	0.47439	0.02016	0.25139	0.42023
ARRE(1)	0.5	20	3	0.2	0.40222	0.46493	0.41255	0.49328	0.29515	0.47227	0.07771	0.44390	0.34744
ARRE(1)	0.5	20	3	0.4	0.57599	0.89264	0.87976	0.53596	0.85084	0.55909	0.65926	0.73413	0.38842
ARRE(1)	0.5	20	3	0.6	0.67343	1.00000	0.94466	0.36310	0.95087	0.54208	0.91659	0.89692	0.49043
ARRE(1)	0.5	20	5	0	0.81292	0.81629	0.92137	0.73848	0.93207	0.57621	0.94749	0.88913	0.60143
ARRE(1)	0.5	20	5	0.05	0.04033	0.32819	0.19772	0.28804	0.22986	0.35292	0.03695	0.26107	0.36391
ARRE(1)	0.5	20	5	0.1	0.22518	0.35183	0.33472	0.48909	0.27447	0.37649	0.13434	0.33028	0.45914
ARRE(1)	0.5	20	5	0.2	0.18244	0.82560	0.83000	0.48713	0.76144	0.50165	0.58737	0.71725	0.54897
ARRE(1)	0.5	20	5	0.4	0.31153	1.00000	1.00000	0.78309	1.00000	0.59310	0.97558	1.00000	0.66687
ARRE(1)	0.5	20	5	0.6	0.42252	0.95495	0.98899	0.77949	1.00000	0.71642	1.00000	0.92392	0.73084

Tablo 8. Doğru Modelin Seçiminde İstatistiklerin Başarı Orantıları (Kovaryans Strüktürü = Birinci mertebeden Otoregresif ve Rastgele Etki, 0.5)
(Devamı)

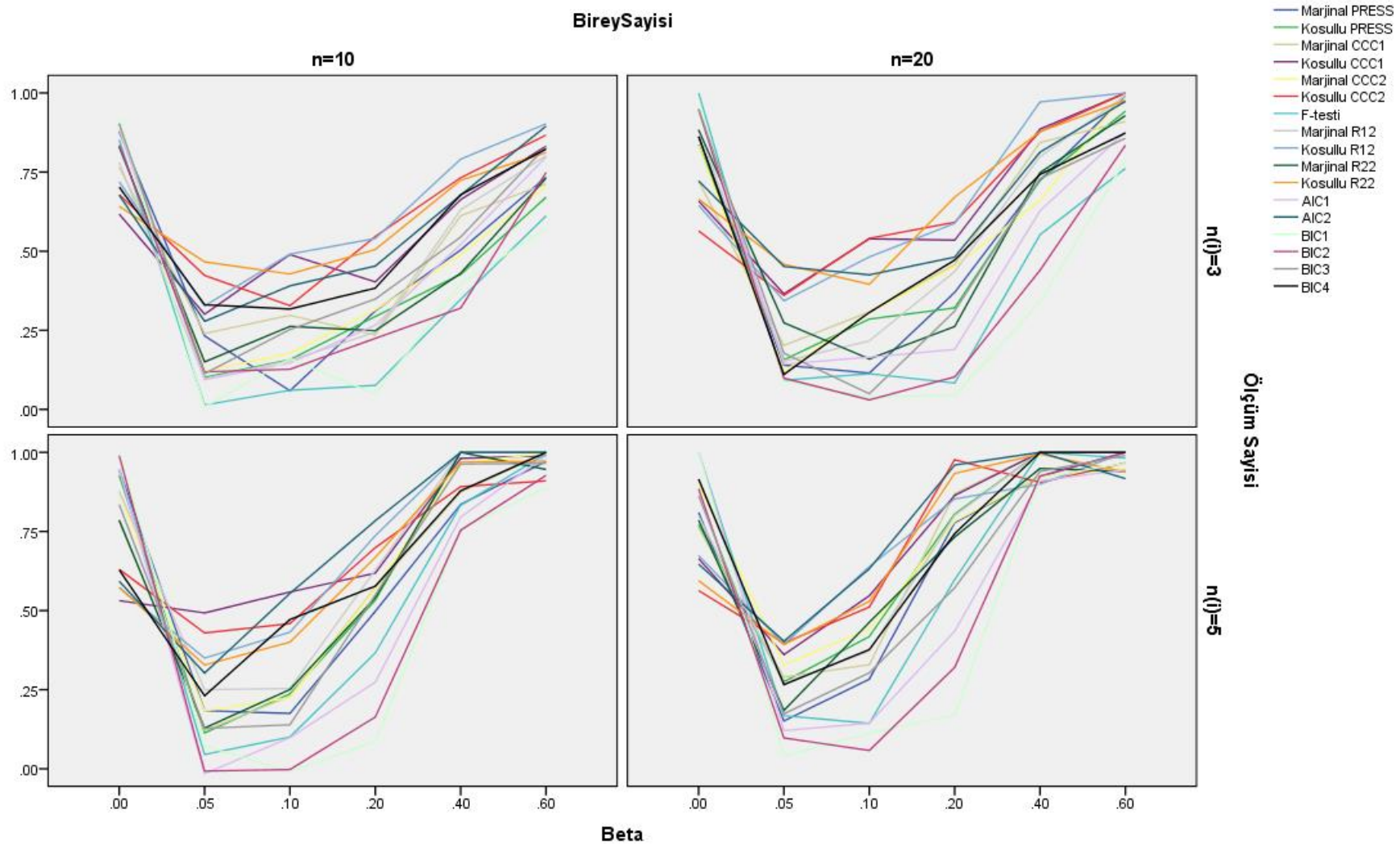
Kovaryans Strüktürü	Parametre	Birey Sayısı	Ölçüm Sayısı	β	Marjinal R_2^2	Koşullu R_2^2	AIC1	AIC2	BIC1	BIC2	BIC3	BIC4
ARRE(1)	0.5	10	3	0	0.81961	0.70722	0.95593	0.74672	0.97541	0.86332	0.82875	0.79151
ARRE(1)	0.5	10	3	0.05	0.11922	0.27057	0.17950	0.35217	0.05261	0.14218	0.07704	0.31819
ARRE(1)	0.5	10	3	0.1	0.06325	0.29909	0.19170	0.37436	0.12247	0.12705	0.15369	0.32977
ARRE(1)	0.5	10	3	0.2	0.16946	0.31336	0.21924	0.54800	0.04564	0.12647	0.33347	0.46438
ARRE(1)	0.5	10	3	0.4	0.48638	0.55120	0.51393	0.73287	0.27503	0.52308	0.57369	0.70780
ARRE(1)	0.5	10	3	0.6	0.81530	0.50436	0.79971	0.96442	0.76970	0.84178	0.90705	0.87935
ARRE(1)	0.5	10	5	0	0.83596	0.72251	0.92736	0.74364	1.00000	0.97300	0.99459	0.71686
ARRE(1)	0.5	10	5	0.05	0.23168	0.23753	0.09514	0.43290	0.10398	-0.04719	0.08896	0.28130
ARRE(1)	0.5	10	5	0.1	0.29886	0.30644	0.02812	0.43478	0.04680	0.08148	0.13333	0.42213
ARRE(1)	0.5	10	5	0.2	0.46675	0.58807	0.41439	0.77233	0.19369	0.31602	0.47417	0.79370
ARRE(1)	0.5	10	5	0.4	0.98594	0.68243	0.80682	1.00000	0.74246	0.95568	0.93869	0.90762
ARRE(1)	0.5	10	5	0.6	1.00000	0.91890	0.96166	1.00000	0.98590	1.00000	0.91081	0.93483
ARRE(1)	0.5	20	3	0	0.98858	0.55152	0.97703	0.71547	0.91072	1.00000	0.93681	0.94748
ARRE(1)	0.5	20	3	0.05	0.16295	0.34707	0.13137	0.21822	0.11320	0.10417	0.00519	0.25380
ARRE(1)	0.5	20	3	0.1	0.10936	0.34767	0.01933	0.31976	0.00644	0.15250	0.15963	0.13162
ARRE(1)	0.5	20	3	0.2	0.41529	0.44907	0.31219	0.64608	0.18868	0.07832	0.26758	0.45285
ARRE(1)	0.5	20	3	0.4	0.84986	0.45520	0.71572	0.86054	0.50321	0.68429	0.78284	0.84179
ARRE(1)	0.5	20	3	0.6	1.00000	0.42116	0.97101	0.97448	0.85196	0.90387	1.00000	0.89971
ARRE(1)	0.5	20	5	0	0.78461	0.58561	1.00000	0.72982	1.00000	0.91693	0.96172	0.82858
ARRE(1)	0.5	20	5	0.05	0.09401	0.32048	0.04569	0.27720	-0.04936	0.03427	0.08068	0.17736
ARRE(1)	0.5	20	5	0.1	0.26802	0.40509	0.09556	0.58943	-0.00886	0.07568	0.14701	0.43384
ARRE(1)	0.5	20	5	0.2	0.69964	0.50743	0.61298	0.87817	0.42650	0.49094	0.75237	0.73371
ARRE(1)	0.5	20	5	0.4	0.90200	0.63544	0.99821	0.93884	0.88181	1.00000	1.00000	0.96366
ARRE(1)	0.5	20	5	0.6	1.00000	0.86167	1.00000	1.00000	1.00000	0.93894	0.93794	1.00000

Tablo 9. Doğru Modelin Seçiminde İstatistiklerin Başarı Orantıları (Kovaryans Strüktürü = Birinci mertebeden Otoregresif ve Rastgele Etki, 0.9)

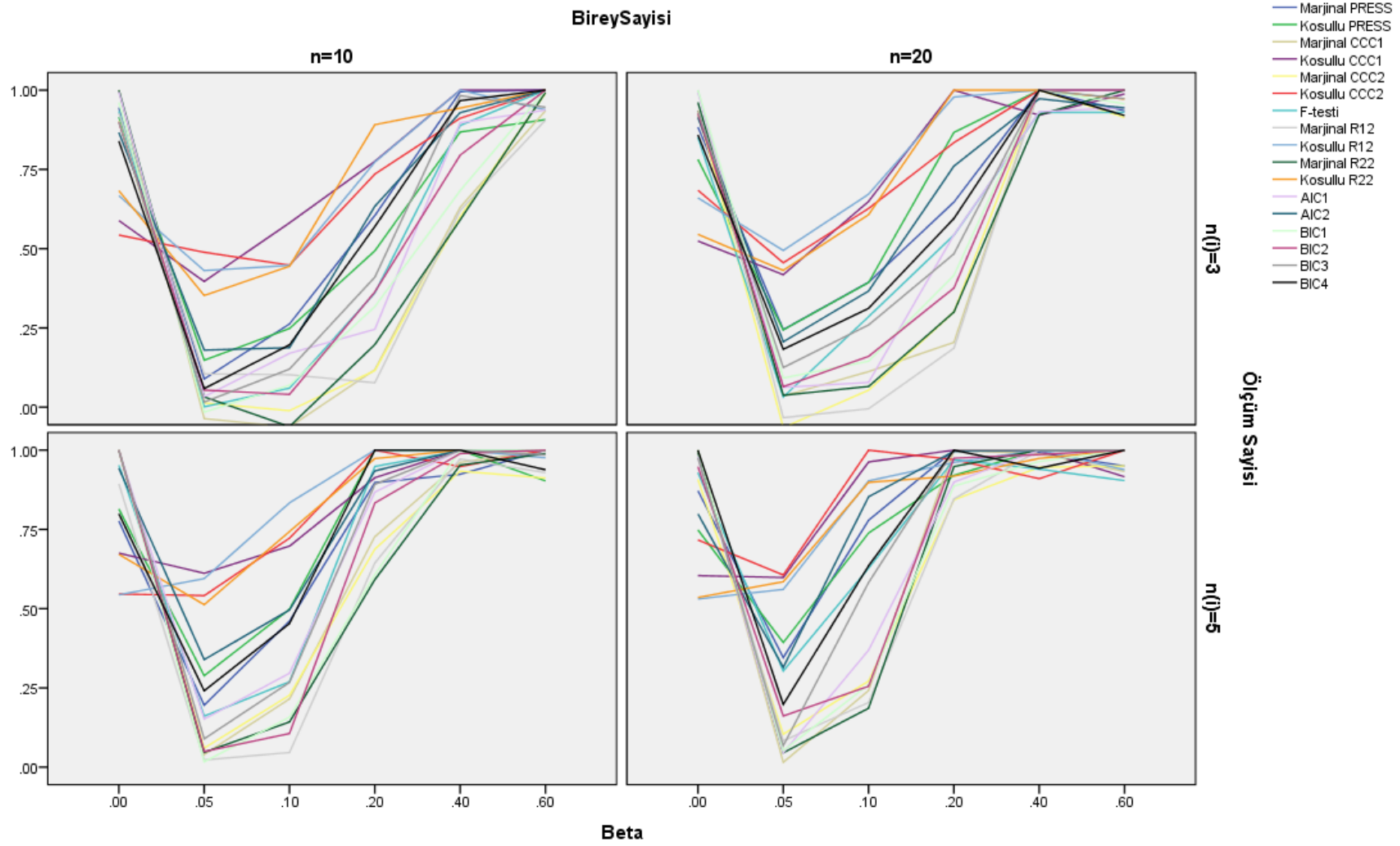
Kovaryans Strüktürü	Parametre	Birey Sayısı	Ölçüm Sayısı	β	Marjinal PRESS	Koşullu PRESS	Marjinal CCC	Koşullu CCC	Marjinal CCC ₂	Koşullu CCC ₂	F-testi	Marjinal R ²	Koşullu R ²
ARRE(1)	0.9	10	3	0	0.61911	0.85263	0.97347	0.68117	1.00000	0.83280	1.00000	0.93063	0.72001
ARRE(1)	0.9	10	3	0.05	0.33693	0.13161	0.32470	0.42851	0.40479	0.39669	0.06343	0.37736	0.35334
ARRE(1)	0.9	10	3	0.1	0.34923	0.43074	0.07466	0.35647	0.03415	0.21525	0.12124	-0.03995	0.21893
ARRE(1)	0.9	10	3	0.2	0.47077	0.68292	0.18169	0.35125	0.02405	0.36258	0.32022	0.01883	0.20871
ARRE(1)	0.9	10	3	0.4	0.68734	1.00000	0.70917	0.33486	0.55569	0.39693	0.75588	0.50601	0.28631
ARRE(1)	0.9	10	3	0.6	0.79885	1.00000	0.97169	0.44453	1.00000	0.36075	0.88807	1.00000	0.38277
ARRE(1)	0.9	10	5	0	0.62664	0.90678	0.91923	0.68277	1.00000	0.70358	1.00000	1.00000	0.70348
ARRE(1)	0.9	10	5	0.05	0.24610	0.42066	0.04827	0.29834	0.00589	0.32147	0.00622	0.06394	0.25349
ARRE(1)	0.9	10	5	0.1	0.42734	0.62120	0.16375	0.17052	0.05856	0.31998	0.39798	0.10119	0.19885
ARRE(1)	0.9	10	5	0.2	0.53398	0.91983	0.66636	0.38722	0.52205	0.38462	0.83568	0.58777	0.43557
ARRE(1)	0.9	10	5	0.4	0.66428	1.00000	1.00000	0.31472	1.00000	0.44365	1.00000	1.00000	0.46757
ARRE(1)	0.9	10	5	0.6	0.77449	1.00000	1.00000	0.44877	0.90691	0.50964	1.00000	1.00000	0.42485
ARRE(1)	0.9	20	3	0	0.64678	0.82058	1.00000	0.69152	1.00000	0.74896	1.00000	0.91592	0.68111
ARRE(1)	0.9	20	3	0.05	0.46612	0.23616	0.07277	0.33985	0.07637	0.37359	0.03259	-0.07158	0.23965
ARRE(1)	0.9	20	3	0.1	0.47260	0.48364	0.09757	0.20267	0.10859	0.23290	0.12354	-0.04947	0.16112
ARRE(1)	0.9	20	3	0.2	0.45622	0.87353	0.27172	0.12130	0.30649	0.25384	0.63254	0.14182	0.27956
ARRE(1)	0.9	20	3	0.4	0.61032	1.00000	1.00000	0.19810	0.90954	0.37168	0.84792	1.00000	0.34234
ARRE(1)	0.9	20	3	0.6	0.73819	1.00000	1.00000	0.42906	0.93594	0.40994	0.97219	1.00000	0.38421
ARRE(1)	0.9	20	5	0	0.54317	0.81779	0.98509	0.64597	0.95456	0.64817	0.95228	1.00000	0.60823
ARRE(1)	0.9	20	5	0.05	0.42047	0.49149	0.03067	0.15723	0.00654	0.22380	0.13665	0.04031	0.29247
ARRE(1)	0.9	20	5	0.1	0.42364	0.84714	0.25208	0.21810	0.18954	0.24754	0.64130	0.29204	0.13755
ARRE(1)	0.9	20	5	0.2	0.59746	1.00000	0.95833	0.29644	0.93302	0.21466	0.92874	1.00000	0.35450
ARRE(1)	0.9	20	5	0.4	0.66062	1.00000	1.00000	0.36026	0.92793	0.20310	1.00000	1.00000	0.38508
ARRE(1)	0.9	20	5	0.6	0.77185	0.92893	1.00000	0.49378	1.00000	0.42271	0.91792	1.00000	0.41230

Tablo 9. Doğru Modelin Seçiminde İstatistiklerin Başarı Orantıları (Kovaryans Strüktürü = Birinci mertebeden Otoregresif ve Rastgele Etki, 0.9)
(Devamı)

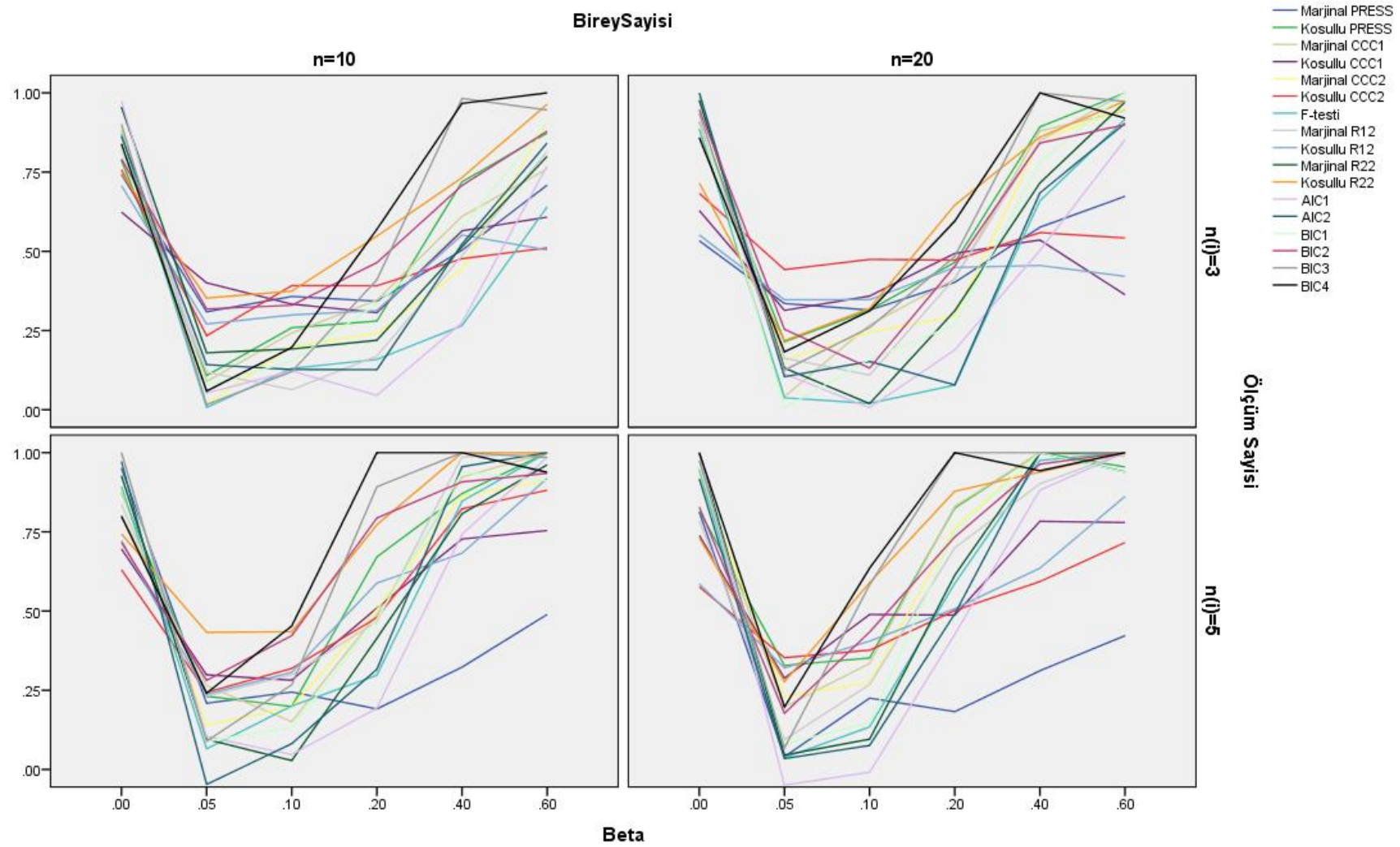
Kovaryans Strüktürü	Parametre	Birey Sayısı	Ölçüm Sayısı	β	Marjinal R_2^2	Koşullu R_2^2	AIC1	AIC2	BIC1	BIC2	BIC3	BIC4
ARRE(1)	0.9	10	3	0	0.90701	0.70947	1.00000	0.85423	1.00000	1.00000	1.00000	1.00000
ARRE(1)	0.9	10	3	0.05	0.22941	0.30470	0.09055	0.09318	0.04387	0.06420	0.02940	0.13994
ARRE(1)	0.9	10	3	0.1	0.06107	0.25920	0.04634	0.32899	0.00044	0.04375	0.06341	0.14073
ARRE(1)	0.9	10	3	0.2	0.01766	0.30993	0.33743	0.71979	0.34525	0.30704	0.59950	0.51072
ARRE(1)	0.9	10	3	0.4	0.61519	0.38702	0.97621	1.00000	0.87034	0.97552	0.92194	1.00000
ARRE(1)	0.9	10	3	0.6	0.99035	0.37266	1.00000	1.00000	1.00000	1.00000	0.90090	1.00000
ARRE(1)	0.9	10	5	0	1.00000	0.63681	1.00000	0.84324	1.00000	1.00000	0.94868	1.00000
ARRE(1)	0.9	10	5	0.05	0.05153	0.32047	0.01918	0.15891	0.00584	0.09644	0.22250	0.29848
ARRE(1)	0.9	10	5	0.1	0.18999	0.23580	0.20071	0.69777	0.13805	0.17821	0.33880	0.68860
ARRE(1)	0.9	10	5	0.2	0.49742	0.38733	0.89958	0.96117	0.80676	0.83172	1.00000	1.00000
ARRE(1)	0.9	10	5	0.4	1.00000	0.32002	1.00000	1.00000	0.98999	1.00000	1.00000	0.92893
ARRE(1)	0.9	10	5	0.6	1.00000	0.41654	0.96396	0.96797	1.00000	1.00000	1.00000	1.00000
ARRE(1)	0.9	20	3	0	1.00000	0.75256	1.00000	0.83211	0.94115	1.00000	1.00000	0.86350
ARRE(1)	0.9	20	3	0.05	0.05355	0.24566	0.03518	0.17172	0.06678	0.00117	0.00494	0.16066
ARRE(1)	0.9	20	3	0.1	0.09037	0.21699	0.20303	0.49279	0.15727	0.13862	0.18427	0.22753
ARRE(1)	0.9	20	3	0.2	0.10791	0.14913	0.73442	0.88668	0.45852	0.61840	0.73917	0.88495
ARRE(1)	0.9	20	3	0.4	0.99284	0.32362	0.96466	1.00000	1.00000	0.95696	0.97778	0.91481
ARRE(1)	0.9	20	3	0.6	1.00000	0.33797	0.97497	1.00000	1.00000	0.94695	1.00000	1.00000
ARRE(1)	0.9	20	5	0	0.97438	0.62314	1.00000	0.93562	0.94084	1.00000	1.00000	1.00000
ARRE(1)	0.9	20	5	0.05	0.12489	0.13652	0.02992	0.35856	0.05069	0.03386	0.21094	0.38319
ARRE(1)	0.9	20	5	0.1	0.26913	0.19100	0.65318	0.79181	0.44287	0.48278	0.66267	0.79048
ARRE(1)	0.9	20	5	0.2	1.00000	0.33458	1.00000	1.00000	0.92994	1.00000	0.92953	1.00000
ARRE(1)	0.9	20	5	0.4	0.90891	0.25975	0.94895	1.00000	1.00000	1.00000	1.00000	0.93894
ARRE(1)	0.9	20	5	0.6	1.00000	0.31620	1.00000	0.99600	1.00000	0.92793	1.00000	1.00000



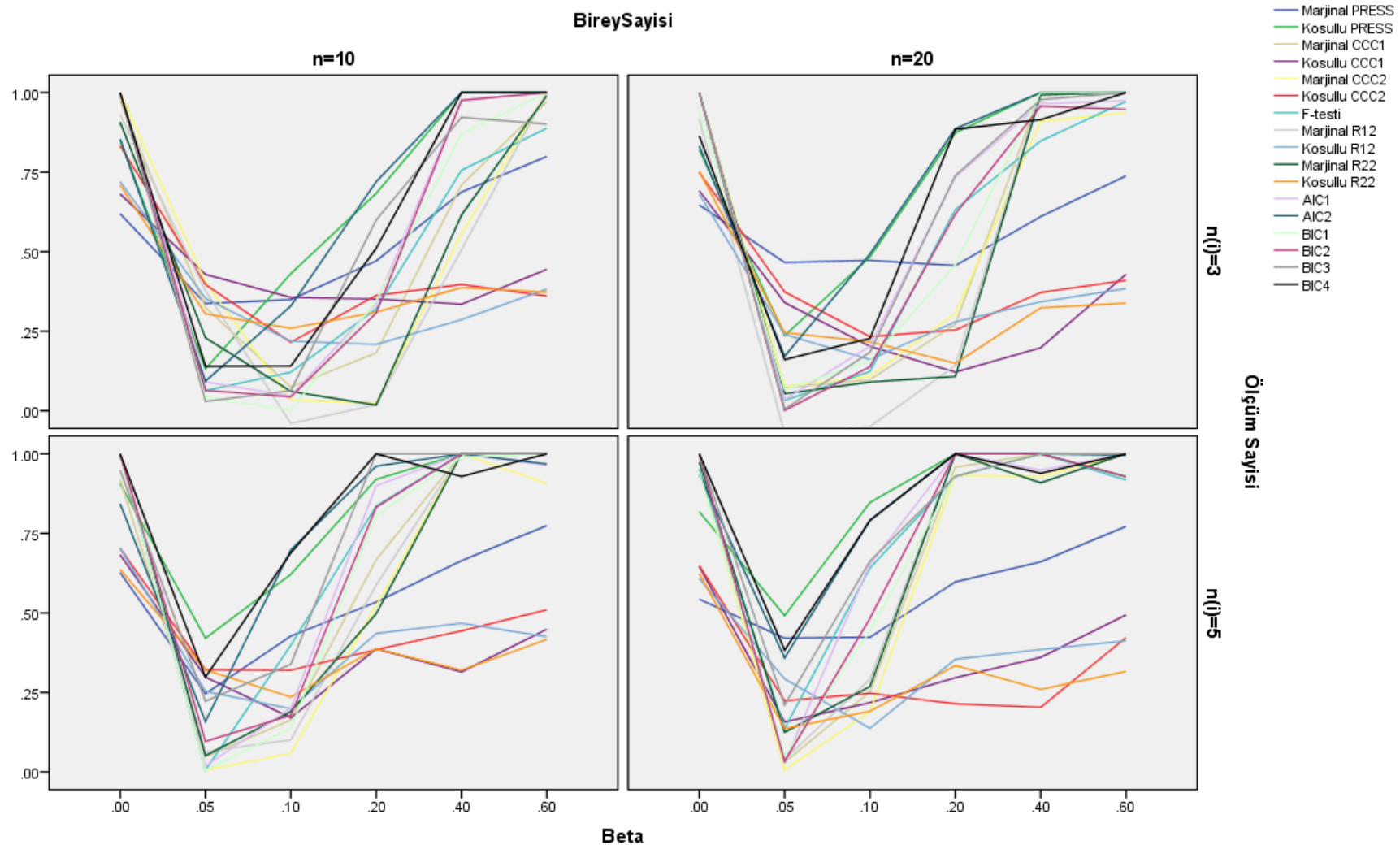
Şekil 1. Doğru Modelin Seçiminde İstatistiklerin Başarı Orantıları (Kovaryans Strüktürü = Bileşik Simetri, 0.5)



Şekil 2. Doğru Modelin Seçiminde İstatistiklerin Başarı Orantıları (Kovaryans Strüktürü = Bileşik Simetri, 0.9)



Şekil 3. Doğru Modelin Seçiminde İstatistiklerin Başarı Orantıları (Kovaryans Strüktürü = Birinci mertebeden Otoregresif ve Rastgele Etki, 0.5)



Şekil 4. Doğru Modelin Seçiminde İstatistiklerin Başarı Orantıları (Kovaryans Strüktürü = Birinci mertebeden Otoregresif ve Rastgele Etki, 0.9)

SONUÇ

Bu çalışma PRESS, CCC, R^2 , F -testi ve iki enformasyon kriterinin (AIC ve BIC) doğrusal karışık modellerde aritmetik ortalama için iki modelden daha iyi olanını seçmekteki performanslarını soruşturmak için yürütülmüştür. Aynı zamanda, bu çalışma, kalıntıların hesaplanmasında marjinal ve koşullu yaklaşımların etkisini, enformasyon kriterlerinde farklı REML fonksiyonlarının kullanılmasının etkisini ve CCC, R^2 ve enformasyon kriterlerinde yapılan farklı düzeltmelerin etkisini soruşturmak için yürütülmüştür.

Doğrusal karışık modelin sabit etkiler kısmı için model seçiminde on yedi istatistiğin gösterdiği performansı karşılaştırmak üzere yürütülen simülasyon çalışmasında, yaratılan farklı veri durumları için rastgele tekrarlanan senaryolarda belirli doğru modelin seçilmesi yüzdesi hesaplanmıştır.

Bileşik simetri strüktürü için, on yedi istatistiğin başarı yüzdeleri yüksek bulunmuştur. Bu çalışmanın bulgularına göre, PRESS, CCC ve R^2 için koşullu kalıntılara göre marjinal kalıntıların kullanılması en iyi modelin seçiminde genel olarak daha yüksek bir performans ortaya koymaktadır. Yine, AIC ve BIC için, $\frac{1}{2} \log |\sum_{i=1}^m X_i' X_i|$ terimli REML fonksiyonunun kullanılması ile genel olarak daha yüksek bir performans elde edilmiştir. CCC, R^2 ve iki enformasyon kriteri için, parametre sayılarındaki düzeltmelerin istatistiklerin performansları bakımından bir etkisi görülmemiştir.

Birinci mertebeden otoregresif kovaryans strüktürü ve rastgele etki için, bu çalışmada PRESS ile birlikte koşullu kalıntıların kullanılması, CCC ve R^2 ile birlikte marjinal kalıntıların kullanılması ve AIC ve BIC ile birlikte $\frac{1}{2} \log |\sum_{i=1}^m X_i' X_i|$ terimli REML fonksiyonunun kullanılması en iyi modelin seçiminde genel olarak daha yüksek bir performans temin etmiştir. Yine, parametre sayılarındaki düzeltmelerin bu istatistiklerin performansları bakımından bir etkisi görülmemiştir.

Tüm kovaryans strüktürleri için F -testi iyi bir performans göstermiştir. Bu çalışmada, aynı zamanda, verinin kovaryans strüktürü, parametre değerleri ve örnek hacmi gibi karakteristiklerinin çeşitli istatistiklerin performansları üzerinde büyük etkisi olduğu görülmüştür. Bu simülasyon çalışmasında ortaya çıkan bir sonuç şuudur ki model seçme kriterlerinden hiç biri seçim performansı bakımından diğerlerine göre her durumda daha iyi değildir. Daha çok sayıda model, daha başka kovaryans strüktürleri, örnek hacimleri ve hem

aritmetik ortalama modeli hem de kovaryans strüktürü için eş zamanlı model seçimi üzerine araştırmaların devam ettirilmesi yeni bulguların elde edilmesi ve pratiğe aktarılması bakımından yararlı olacaktır.

Bu tez çalışmasının sonuçlarının ışığında, doğrusal karışık modelde model seçimine dair gelecek araştırma için çeşitli alanlar belirlemiştir. Burada incelenen metotların bazılarının dağılımsal hususiyetlerinin yanısıra, daha ileri hesaplamalı çalışmalar, özellikle dar örnek hacmine sahip veri için önemli sonuçların elde edilmesini sağlayacaktır.

Tüm seçim kriterleri için, ileriki hesaplamalı çalışmalarda örnek hacminin dar olduğu, kayıp verinin mevcut olduğu ve veri üretimine çoklu interaksiyon terimlerinin dahil edildiği durumlar ele alınabilir. İleriki simülasyon çalışmalarını, altta yatan sabit etkinin ve kovaryans parametrelerinin gerçekliği bakımından, klasik veri kümelerinin parametrelerine dayandırmak yararlı olabilir. Gerçek veriye dayanan simülasyonlara bakılmak suretiyle, seçim kriterlerinin gerçek dünyada nasıl performans göstereceklerine dair daha iyi bir resim elde edilebilir. Yine, üretilen veri normal olmadığı zaman, seçim istatistiklerinin ne kadar sağlam olduklarını bilmek yararlı olacağından, seçim kriterlerinin muhtelif kovaryans strüktürleri ile ortaya koyacakları performansları incelenebilir.

Boylamsal çalışmaların popülerliği artarken, bu çalışmalardan elde edilecek veriyi isabetle betimleyecek modellerin seçiminde yararlı olacak ilave araçları keşfetmenin ve rafine etmenin de önemi artmaktadır.

- Akaike, H., 1969. Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics*, 21, 243.
- Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory*, B. N. Petrov and F. Csaki, eds. Akademiai Kiado, Budapest, 267-281.
- Akaike, H., 1974. A new look at the statistical model identification. *IEEE transactions on automatic control*, (19): 716.
- Akaike, H., 1978. A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics*, (30): 9.
- Allen, D.M., 1974. The Relationship between Variable Selection and Data Augmentation and a Method of Prediction. *Technometrics*, 16: 125-127.
- Baayen, R.H., Davidson, D.J. and Bates , D.M., 2008. Mixed-effects modeling with crossed random effects for subjects and items. *J. of Memory and Language*, (59): 390-412.
- Boozdogan, H., 1987. Model Selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical extensions. *Psychometrika*, 52: 345-370.
- Cavanaugh, J. E., 1999. A large-sample model selection criterion based on Kullback's symmetric divergence. *Statistics & Probability Letters*, (42): 333-343.
- Cavanaugh, J. E., 2004. Criteria for linear model selection based on Kullback's symmetric divergence. *Australian & New Zealand Journal of Statistics*, (46): 257-274.
- Christensen, R., Pearson, L. M., and Johnson, W., 1992. Case-Deletion Diagnostics for Mixed Models. *Technometrics*, 34(1): 38-45.
- Davidian, M., 2001. ST 732: Applied Longitudinal Data Analysis, Lecture Notes. North Carolina State University.
- Demidenko E., 2004. *Mixed Models—Theory and Application*. Wiley: New York.
- Edwards, L. J., Muller, K. E., Gurka, M. J. and Stewart, P. W., 2005. Selecting the Best Linear Mixed Model: The State of the Art for Longitudinal Data Analysis. *Joint Statistical Meetings*, Minneapolis, Minnesota.
- Edwards, L.J., Muller, K.E., Wolfinger, R.D., Qaqish, B.F. and Schabenberger, O., 2008. An R^2 statistic for fixed effects in the linear mixed model. *Statistics in Medicine*.
- Fai, A.H.T. and Cornelius, P.L., 1996. Approximate F -tests of Multiple Degree of Freedom Hypotheses in Generalized Least Squares Analyses of Unbalanced Split-plot Experiments. *Journal of Statistical Computing and Simulation*, 54: 363-378.
- Ferron, J., Dailey, R., and Yi, Q., 2002. Effects of misspecifying the first-level error structure in two-level models of change. *Multivariate Behavioral Research*, 37(3): 379-403.
- Fitzmaurice, G., Davidian, M., Molenberghs, G., and Verbeke, G., (eds), 2008. *Longitudinal Data Analysis: A Handbook of Modern Statistical Methods*. Chapman & Hall/CRC: BocaRaton, Florida.
- Fujikoshi, Y. and Satoh, K., 1997. Modified AIC and G: in multivariate linear regression. *Biometrika*, (84): 707-716.

- Giesbrecht, F.G. and Burns, J.C., 1985. Two-Stage Analysis Based on a Mixed Model: Large-sample Asymptotic Theory and Small-Sample Simulation Results. *Biometrics* 41: 853-862.
- Gilmour, S. G., 1996. The interpretation of Mallows's G -statistic. *The Statistician*, (45): 49-56.
- Gomez, E. V., Schaalje, G. B. and Fellingham, G. W., 2005. Performance of the Kenward-Roger Method when the Covariance Structure is Selected Using AIC and BIC. *Communications in Statistics-Simulation and Computation*, 34: 377-392.
- Gorman, J.W., and Toman, R.J., 1966. Selection of variables for fitting equations to data. *Technometrics*, 8(1): 27-51.
- Gurka, M. J., 2006. Selecting the Best Linear Mixed Model under REML. *The American Statistician*, 60(1): 19-26.
- Hafidi, B. and Mkhadri, A., 2006. A corrected Akaike criterion based on Kullback's symmetric divergence: Applications in time series, multiple and multivariate regression. *Computational Statistics & Data Analysis*, (50): 1524-1550.
- Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B (Methodological)*, (41): 190-195.
- Hocking, R.R., 1976. A biometrics invited paper. the analysis and selection of variables in linear regression. *Biometrics*, (32): 1-49.
- Hurvich, C. M., and Tsai, C. L., 1989. Regression and Time Series Model Selection in Small Samples. *Biometrika*, 76: 297-307.
- Hurvich, C.M., Shumway, R., and Tsai, C., 1990. Improved estimators of kullback-leibler information for autoregressive model selection in small samples. *Biometrika*, 77, 709-719.
- Kenward, M. G. and Roger, J. H., 1997. Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics*, 53: 983-997.
- Keselman, H. J., Algina, J., Kowalchuk, R. K., and Wolfinger, R. D., 1998. A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements. *Communications in Statistics: Simulation and Computation*, 27(3): 591-604.
- Kim, H., and Cavanaugh, J.E., 2005. Model selection criteria based on Kullback information measures for nonlinear regression. *Journal of Statistical Planning and Inference*, (134): 332-349.
- Laird, N.M., and Ware, J.H., 1982. Random-effects models for longitudinal data. *Biometrics*, (38): 963-974.
- Lin, L. I.-K., 1989. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics*, 45, 255-268.
- Liu, H., Weiss, R.E., Jennrich, R.I., and Wenger, N.S., 1999. PRESS model selection in repeated measures data. *Computational Statistics & Data Analysis*, (30): 169-184.
- Mallows, C.L., 1973. Some comments on C_p *Technometrics*. (15), 661-675.
- Muller, K. and Fetterman, B., 2002. *Regression and ANOVA: An integrated approach using SAS software*. Cary, NC, USA: SAS Institute Inc.
- Muller, K. and Stewart, P., 2006. *Linear Model Theory: Univariate, Multivariate, and Mixed Models*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

- Neter, J., Kutner, M.H., Nachtsheim, C.J. and Wasserman, W., 1996. Applied Linear Regression Models, 3rd Edition. USA: The McGraw-Hill Companies.
- R Core Team, 2017. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. (<http://www.R-project.org>)
- SAS Institute Inc., 2007. SAS (release 9.1.3), User's Guide. Cary, NC: SAS Institute Inc.
- Schaalje, G. B., McBride, J. B., and Fellingham, G. F., 2002. Adequacy of Approximations to Distributions of Test Statistics in Complex Mixed Linear Models. Journal of Agricultural, Biological, and Environmental Statistics, 7512-524.
- Schaalje, G. B., McBride, J. B., and Fellingham, G. W., 2001. Approximations to Distributions of Test Statistics in Complex Mixed Linear Models Using SAS Proc MIXED. Paper 262-26 in Proceedings of the SAS Users Group International 26th Annual Conference.
- Schabenberger, O., 2004. Mixed Model Influence Diagnostics. SUGI 29, Paper 189-29.
- Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics, 6, 461–464.
- Singmann, H. and Kellen, D., 2018. An introduction to linear mixed modeling in experimental psychology. Psychology Press.
- SPSS, 2009. Statistical Package for the Social Sciences for Windows (Version 17.0) SPSS Inc., Chicago, USA.
- Sugiura, N., 1978. Further analysts of the data by Akaike' s information criterion and the finite corrections. Communications in statistics. Theory and methods, (7): 13.
- Vonesh, E.F. and Chinchilli, V.M., 1997. Linear and Nonlinear Models for the Analysis of Repeated Measurements. New York: Marcel Dekker.
- Vonesh, E.F., 1992. Non-linear models for the analysis of longitudinal data. Statistics in Medicine, (11): 1929-1954
- Vonesh, E.F., Chinchilli, V. M. and Pu, K., 1996. Goodness-of-Fit in Generalized Nonlinear Mixed-Effects Models. Biometrics, 52(2): 575-587.

ÖZGEÇMİŞ

İlk, orta ve lise tahsilini aynı şehirde tamamladı. Yüzüncü Yıl Üniversitesi Veteriner Fakültesinden 2001 yılında mezun oldu. Sütçü İmam Üniversitesi Fen Bilimleri Enstitüsü Zootekni Anabilim Dalında 2005 yılında Yüksek Lisans öğrenimini tamamladı. 2006 yılından beri Tarım ve Orman Bakanlığı'nda çalışmaktadır.

