

**GENELLEŐTİRİLMİŐ LİNEER KARMA MODELLERDE
PARAMETRE TAHMİNİ VE SONUÇ ÇIKARIM**

ŐIDA SEÇKİN KURT

TEMMUZ 2024

DİYARBAKIR

DİCLE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**GENELLEŞTİRİLMİŞ LİNEER KARMA MODELLERDE
PARAMETRE TAHMİNİ VE SONUÇ ÇIKARIM**

ŞİDA SEÇKİN KURT

DİCLE ÜNİVERSİTESİ LİSANSÜSTÜ EĞİTİM-ÖĞRETİM VE SINAV
YÖNETMELİĞİNİN BİR PARÇASI OLARAK
İSTATİSTİK ANA BİLİM DALINDA
YÜKSEK LİSANS TEZİ
OLARAK HAZIRLANMIŞTIR

TEMMUZ 2024

DİYARBAKIR

GENELLEŐTİRİLMİŐ LİNEER KARMA MODELLERDE PARAMETRE TAHMİNİ VE SONUÇ ÇIKARIM

Őida S. KURT tarafından Dicle Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğı'nin bir parçası olarak hazırlanan bu çalışma, aőağıda bilgileri yazılı jüri üyeleri tarafından deęerlendirilerek **İstatistik Ana Bilim Dalı'nda Yüksek Lisans Tezi** olarak kabul edilmiőtir.

Prof. Dr. Neslihan DALKILIÇ
Müdür, **Fen Bilimleri Enstitüsü**

Doç. Dr. Özge KURAN
Danışman, **İstatistik Bölümü,**
Dicle Üniversitesi

Sınav Jürisi:

Prof. Dr. Mahmude Revan ÖZKALE ATICIOĞLU (*)
İstatistik Bölümü, Çukurova Üniversitesi

Doç. Dr. Özge KURAN (**)
İstatistik Bölümü, Dicle Üniversitesi

Doç. Dr. Seçil YALAZ
İstatistik Bölümü, Dicle Üniversitesi

ONAY

Savunma Tarihi: 22/07/2024

(*) Jüri Başkanı.

(**) Tez Danışmanı.



Aileme...

Dicle Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırlanan bu tez çalışmasında yer alan tüm bilgilerin akademik kurallara ve etik ilkelere uygun olarak elde edildiğini ve sunulduğunu beyan ederim. Ayrıca, bahse konu bu kural ve ilkelerin gerektirdiği üzere, bu çalışmada özgün olmayan tüm bilimsel içerikleri kurallara uygun biçimde alıntılıyıp kaynak gösterdiğimi beyan ederim. Beyanımınla çelişen herhangi bir delil bulunduğu takdirde tüm sorumluluğu üstleneceğimi kabul ederim.

Ad, Soyad: Şida Seçkin KURT

İmza:

TEŞEKKÜR

Bu tezin hazırlanmasında bana destek olan, katkı sağlayan ve yol gösteren birçok kişiye teşekkür etmek isterim.

Öncelikle, tez danışmanım Doç. Dr. Özge KURAN 'a, beni her aşamada yönlendiren, bilgi ve deneyimlerini benimle paylaşan, sabır ve anlayışla yaklaştığı için minnettarım. Onun değerli önerileri ve eleştirileri olmasaydı, bu çalışma bu seviyeye ulaşamazdı.

Tez savunmamda, değerli zamanlarını ayıran ve çalışmama önemli katkılarda bulunan jüri üyelerime teşekkür ederim. Onların yapıcı eleştirileri ve önerileri, tezimin kalitesini yükseltmede büyük önem taşımaktadır.

Ayrıca, çalışmalarım süresince moral desteğini hiç esirgemeyen, beni her zaman cesaretlendiren ve yanımda olan eşime, anneme, babama ve kardeşime sonsuz teşekkür ederim. Onların varlığı ve desteği, bu sürecin en değerli parçası olmuştur.

Yazar, Dicle Üniversitesi Bilimsel Araştırma Projeleri Koordinatörlüğü'ne (DÜBAP), bu tez çalışması için sağladıkları finansal destek için (Proje No: DÜBAP-FEN.23.029) için teşekkürü bir borç bilir.

Herkese içtenlikle teşekkür eder, saygılarımı sunarım.

İÇİNDEKİLER

TEŞEKKÜR.....	v
İÇİNDEKİLER	vi
ŞEKİLLER LİSTESİ	viii
TABLolar LİSTESİ.....	ix
SİMGELER VE KISALTMALAR LİSTESİ	x
ÖZET.....	xi
ABSTRACT.....	xii
1. GİRİŞ	1
2. LİNEER MODELLER VE GENELLEŞTİRİLMİŞ LİNEER MODELLER	6
2.1 Lineer Modeller	6
2.2 Genelleştirilmiş Lineer Modeller	7
2.2.1 Bağımlı değişkenin dağılımı (rastgele bileşen)	7
2.2.1.1 Poisson dağılımı.....	8
2.2.1.2 Binom dağılımı	8
2.2.1.3 Normal dağılım.....	9
2.2.1.4 Gamma dağılımı	9
2.2.1.5 Üstel dağılım ailesinin olabilirlik, log-olabilirlik fonksiyonları ve özellikleri.....	10
2.2.2 Sistematiik bileşen.....	14
2.2.3 Bağlantı (link) fonksiyonu.....	14
3. GENELLEŞTİRİLMİŞ LİNEER KARMA MODELLER VE TAHMİN YÖNTEMLERİ.....	17
3.1 Lineer Karma Modeller	17
3.2 Genelleştirilmiş Lineer Karma Modeller.....	18
3.2.1 Genelleştirilmiş lineer karma modellerin genel yapısı	18
3.2.1.1 Bağımlı değişkenin koşullu dağılımı	18

3.2.1.2 Bağımlı değişkenin (marjinal) ortalaması, varyansı, kovaryansı ve korelasyonu	19
3.2.2 Genelleştirilmiş lineer karma modellerde olabilirlik fonksiyonu.....	22
3.3 Genelleştirilmiş Lineer Karma Modellerde Parametre Tahmini	22
3.3.1 Lineerleştirme yöntemleri	23
3.3.1.1 Penalized yarı-olabilirlik ve Pseudo yarı-olabilirlik yöntemleri	23
3.3.2 İntegral yaklaşımları	27
3.3.2.1 Laplace yaklaşımı	28
3.3.2.2 Gauss-Hermite Quadrature ve Adaptive Gauss-Hermite Quadrature yaklaşım yöntemleri	29
3.4 Genelleştirilmiş Lineer Karma Modellerde Sonuç Çıkarımı	33
4. BULGULAR VE TARTIŞMA	36
4.1 Bellbird Veri Seti.....	36
4.1.1 Veri seti ve değişkenler	36
4.1.2 Analiz	37
4.1.3 Model tanımlama ve uygulama	38
4.1.4 Sonuçlar ve yorum.....	38
4.2 Owl Veri Seti.....	41
4.2.1 Veri seti ve değişkenler	42
4.2.2 Analiz	45
4.2.3 Model tanımlama ve uygulama	45
4.2.4 Sonuçlar ve yorum.....	46
5. SONUÇLAR VE ÖNERİLER	50
KAYNAKLAR	52
ÖZGEÇMİŞ	57

ŞEKİLLER LİSTESİ

Şekil 4.1 Yuvaya göre gruplandırılmış, varış zamanına göre ebeveynin varışından önceki 30 saniyelik aralıktaki çağrı sayılarının grafiği	44
--	----



TABLÖLAR LİSTESİ

Tablo 2.1 Üstel dağılım ailesi üye dağılımlarının uyarlanmış halleri, fonksiyonları, $a(\phi)$, $b(\theta)$ ve $c(y; \phi)$, θ kanonik parametreleri, $\mu = b'(\theta)$ ortalamaları ve $V(y) = b''(\theta)a(\phi)$ varyansları	13
Tablo 2.2 GLM’de yaygın olarak kullanılan kanonik bağlantı fonksiyonları	15
Tablo 3.1 GLMM’de parametre tahmin yöntemlerinin avantajları, dezavantajları ve yazılımlar	32
Tablo 3.2 Bilgi kriterlerinin isimleri, referansları ve formülleri	35
Tablo 4.1 LMM’de sabit etkilerin katsayıları, standart hataları, t ve p değerleri	39
Tablo 4.2 LMM’de rastgele etkilerin standart sapmaları	39
Tablo 4.3 GLMM’de sabit etkilerin katsayıları, standart hataları, Z ve p değerleri ..	39
Tablo 4.4 GLMM’de rastgele etkilerin standart sapmaları	39
Tablo 4.5 LMM’nin ve GLMM’nin karşılaştırılması	40
Tablo 4.6 Laplace yaklaşımı için R çıktısı	46
Tablo 4.7 PQL yöntemi için R çıktısı	47
Tablo 4.8 Tahmin yöntemlerinin karşılaştırılması	47

SİMGELER VE KISALTMALAR LİSTESİ

Simge

Açıklama

ε	Hata Terimi
β	Bilinmeyen Parametre
σ^2	Varyans
θ	Konum Parametresi
ϕ	Yayılm Parametresi
μ	Ortalama
ψ	Ölçek Parametresi
ω	Biçim Parametresi

Kısaltma

Açıklama

AGHQ	Adaptive Gaussian-Hermite Quadrature
AIC	Akaike Information Criterion
AICc	Akaike Information Criterion corrected
BIC	Bayesian Information Criterion
CAIC	Consistent Akaike Information Criterion
EKK	En Küçük Kareler
GHQ	Gaussian-Hermite Quadrature
GLM	Generalized Linear Model
GLMM	Generalized Linear Mixed Model
NB	Negatif Binom
PQL	Penalized Quasi-Likelihood
QAIC	Quasi-Akaike Information Criterion

ÖZET

GENELLEŞTİRİLMİŞ LINEER KARMA MODELLERDE PARAMETRE TAHMİNİ VE SONUÇ ÇIKARIM

KURT, Şida Seçkin

Yüksek Lisans, İstatistik Bölümü

Danışman: Doç. Dr. Özge KURAN

Temmuz 2024, 70 sayfa

Genelleştirilmiş lineer karma modeller, istatistiksel modelleme alanında geniş bir uygulama yelpazesine sahip olan ve hem sabit hem de rastgele etkileri içerebilen modellerdir. Bu modeller, farklı dağılım türlerini ve bağlantı fonksiyonlarını kullanarak, geleneksel lineer modellerin ötesinde esneklik sağlar ve böylece çok çeşitli veri türlerine uyum gösterebilir. Özellikle tekrarlı ölçümler, hiyerarşik yapılar ve iç-içe geçmiş veri setleri gibi karmaşık veri yapılarını analiz etmek için idealdirler. Karmaşık ve hiyerarşik yapıdaki verilerin analizinde yaygın olarak kullanılan bu modeller, doğru parametre tahmini ve elde edilen sonuçların güvenilirliği açısından büyük önem taşımaktadırlar.

Bu çalışmanın amacı, genelleştirilmiş lineer karma modellerde parametre tahmin yöntemlerini ve sonuç çıkarımını incelemektir. Bu amaçla ilk olarak, hem lineer modellerin hem de genelleştirilmiş lineer modellerin temel kavramları ve uygulamaları incelenmiştir. Modellerin yapısı ve çeşitli dağılım türleri, sistematik bileşenler ve bağlantı fonksiyonları ayrıntılı bir şekilde ele alınmıştır. Özellikle Poisson, Binom, Normal ve Gamma dağılımları gibi yaygın kullanılan dağılımlar üzerinde durulmuş ve üstel dağılım ailesinin olabilirlik, log-olabilirlik fonksiyonları ve özellikleri incelenmiştir. İkinci olarak, genelleştirilmiş lineer karma modellerin yapısı ve bu modellerde parametre tahmin yöntemleri üzerinde durulmuştur. Ayrıca, genelleştirilmiş lineer karma modellerde olabilirlik fonksiyonu ve bu fonksiyonun tahmini için kullanılan çeşitli yöntemler incelenmiştir. Parametre tahmini bölümünde, Penalized yarı-olabilirlik ve Pseudo yarı-olabilirlik yöntemlerinin yanı sıra, Laplace yaklaşımı, Gauss-Hermite Quadrature ve Adaptive Gauss-Hermite Quadrature yöntemleri üzerinde durulmuştur. Sonuç çıkarımı bölümünde ise, genelleştirilmiş lineer karma modellerde tahmin edilen parametreler ve model uygunluğu konusunda yapılan analizler ve yorumlar paylaşılmıştır. Çalışmanın bulgular ve tartışma kısmında, Bellbird ve Owl veri setleri kullanılarak yapılan uygulamalar yer almaktadır. Bu uygulamalar R programı kullanılarak incelenmiştir.

Anahtar Kelimeler: Lineer modeller, Lineer karma modeller, Genelleştirilmiş lineer karma modeller, Olabilirlik fonksiyonu, İntegral yaklaşımları

ABSTRACT

PARAMETER ESTIMATION AND INFERENCE IN GENERALIZED LINEAR MIXED MODELS

KURT, Şida Seçkin

Master of Science in Department of Statistics

Supervisor: Doç. Dr. Özge KURAN

July 2024, 70 pages

Generalized linear mixed models are models in statistical modeling that have a wide range of applications and can incorporate both fixed and random effects. These models provide flexibility beyond traditional linear models by accommodating various distribution types and link functions, allowing them to adapt to diverse types of data. They are particularly well-suited for analyzing complex data structures such as repeated measures, hierarchical data, and nested data sets. Widely used in the analysis of complex and hierarchical data, these models are crucial for accurate parameter estimation and the reliability of the obtained results.

The aim of this study is to examine parameter estimation methods and inference in generalized linear mixed models. To this end, the fundamental concepts and applications of both linear models and generalized linear models have been initially reviewed. The structure of the models, various types of distributions, systematic components, and link functions are discussed in detail. Special emphasis is placed on commonly used distributions such as Poisson, Binomial, Normal, and Gamma distributions, and the likelihood and log-likelihood functions and properties of the exponential family of distributions are examined. Secondly, the structure of generalized linear mixed models and parameter estimation methods for these models are addressed. Additionally, various methods used to estimate the likelihood function in GLMMs are examined. In the parameter estimation section, methods such as Penalized Quasi-Likelihood and Pseudo Quasi-Likelihood, as well as the Laplace approximation, Gauss-Hermite Quadrature, and Adaptive Gauss-Hermite Quadrature methods, are highlighted. In the inference section, analyses and interpretations regarding the estimated parameters and model fit in GLMMs are shared. The findings and discussion part of the study includes applications conducted using the Bellbird and Owl datasets. These applications were examined using the R programming language.

Keywords: Linear models, Linear mixed models, Generalize linear mixed models, Likelihood function, Integral approaches

1. GİRİŞ

Model, gerçek olan bir durumun onu temsil eden değişkenlerle özetlenmiş şeklidir. Modeller deterministik veya stokastik olabilirler. Fen ve mühendislik bilimlerinde sık rastlanan deterministik modellerde elde edilen sonuçlar kesindir. Stokastik modellerde, bilinmeyen rastgele değişkenlerle bağlantılı olarak sonuçlar değişebilir. İçinde olasılık bileşeni bulunan bu modellere istatistiksel modeller denmektedir (Lindsey, 1997).

Verilerin analizinde sıklıkla kullanılan istatistik değerlendirme araçlarından biri lineer regresyon modelidir. Lineer regresyon; bağımlı değişken ile bağımsız değişkeni ilişkilendiren istatistiksel bir analiz tekniğidir. Basit bir lineer model; örneğin gelir ve eğitim yılı sayısı, insanların boyu ve kilosu, zarfların uzunluğu ve genişliği, endüstriyel bir işlemin sıcaklığı ve çıktısı, suyun yüksekliği ve kaynama noktası veya bir ilacın dozu ve yanıtı gibi değişkenler arasında birisi yanıt değişken, bir ya da birden çok değişkeni ise açıklayıcı değişken olmak üzere bu değişkenler arasındaki ilişkiyi modellemek için kurulur (Montgomery vd., 2021). Lineer modellerdeki yanıt değişkeninin normal dağılımlı olduğu varsayımı her zaman gerçekleşmeyebilir. Örnek olarak, yanıt değişkeninin başarı ve başarısızlık olmak üzere 0 ve 1 sayılarından oluşması, yaralanma sayıları, hastalık sayıları, depremler ya da doğal olayların meydana gelme sıklıkları gibi verilerden oluşması halinde lineer modeller hataları bir hayli büyük olan yanlış ve hatalı tahminler verecektir (Myers vd., 2012).

Yanıt değişkeninin normal dağılmadığı durumlarda verilerin analizini yapabilmek için, istatistikte kullanılan birçok dağılımı içine alan üstel dağılım ailesinden gelmek suretiyle genelleştirilmiş lineer model (GLM) kavramı ortaya atılmıştır. GLM kavramı ilk olarak Nelder ve Wedderburn (1972) tarafından geliştirilmiştir. Sonraki yıllarda Aitken vd. (1989), McCullagh ve Nelder (1989), Lindsey (1997), McCulloch ve Searle (2001), Myers vd. (2001), Dobson (2002), Duntelman ve Ho (2006), Uusipaikka (2009) GLM'ler hakkında çalışmaya devam etmişlerdir. Ayrıca Cengiz (1997, 2005), Cengiz ve Percy (2001) GLM'lerin genel bir özetini oluşturmuşlardır. GLM'ler ile normal dağılıma sahip olmayan bağımlı değişkenler için lineer model yapısı genişletilmiştir. GLM'lerden önce, normal dağılmayan veri modellemesinin

genellikle bağımlı değişken üzerine yapılan dönüşümlere bağlı olduğu bilinmektedir. Normal dağılıma sahip olmayan bağımlı değişkenlerde, varyansın dağılımı değişken olup (Poisson, Gamma, Beta, vb.) araştırmacılar varyansı sabitlemek için bağımlı değişken üzerine dönüşüm yapmak için çalışmıştır. Ancak bağımlı değişkenin dönüşümüyle ilgili birçok problem ortaya çıkmıştır. Ayrıntılı olarak bu problemler Lewis (1998) tarafından ele alınmaya başlanmıştır. En sık rastlanan problemler, normallik ve sabit varyanslılık varsayımlarının sağlanmamasıdır. Tek bir dönüşüm ile normallik ve sabit varyansın sağlanması zordur. Ayrıca dönüşümler modeldeki hata terimi yapısını da etkiledikleri için yapılacak dönüşümlerin çok dikkatli seçilmeleri gerekmektedir. GLM'ler ile veri dönüşümü yapılmaktadır. Yapılan bu dönüşüm “bağlantı fonksiyonu” olarak da bilinir. Ancak bu dönüşüm deterministik bir bileşene, verinin ortalamasına uygulanır (SAS Institute Inc., 2011). Öte yandan normallik varsayımına yaklaşmak için veri dönüşümü yerine, dönüştürülmemiş bağımlı değişkenler için dağılımların sınıfı genişletilebilir. Bu geniş sınıf “üstel dağılımlar ailesi” olarak isimlendirilir. GLM'ler, bağımlı değişkenin üstel dağılımlar ailesi'nden olması halinde regresyon modelini uydurmaya izin vermektedir.

Yanıt değişkeni normal dağılıma sahip lineer regresyon modellerinin, lineer tahmin edici kısmına rastgele etki eklenerek oluşturulan modeller lineer karma modellerdir (LMM) (Stroup, 2012). LMM'nin sıklıkla kullanılan ilk formülasyonları 1970'lerin sonlarına ve 1980'lerin başlarına dayanmaktadır. Laird ve Ware (1982), LMM'nin geliştirilmesine önemli katkıda bulunmuşlardır. Bu makale, LMM'nin gelişiminde ve uygulamasında önemli bir dönüm noktası olarak kabul edilir. Son yıllarda, klasik istatistik modellerinin neredeyse her formu, rastgele etkileri model içine yerleştirebilmek için geliştirilmiştir. Genelleştirilmiş lineer karma modeller (GLMM) ise yanıt değişkenin üstel dağılım ailesinden gelmesi yönüyle GLM'nin, lineer tahmin edicisine rastgele etki eklenmesi yönü ile de LMM'nin genişletilmiş, çok kapsamlı bir model grubudur. GLMM'ler, rastgele etkiler var olduğunda normal dağılıma sahip olmayan verileri analiz etmede daha esnek bir yaklaşım sağlar (Bolker vd., 2009; Gbur vd., 2020; Jiang ve Nguyen, 2007; McCulloch ve Searle, 2001; Stroup, 2012). GLMM, LMM'lerin varsayımlardan kaynaklanan eksiklerinin giderilmesi için geliştirilmiş bir yöntemdir. Bu yöntem Breslow ve Clayton (1993), McGilchrist (1994), Lee ve Nelder (1996) tarafından kullanılmıştır, McCulloch ve

Searle (2001), Verbeke ve Molenberghs (2005), Littell vd. (2005) ve Jiang (2007) tarafından detaylı incelenmiştir.

GLM için kullanılabilir ilk yazılım 1980'lerin ortalarında, modelde sabit ve rastgele etkili parametreleri bir arada içeren LMM için ilk yazılım 1990'larda, GLMM için gerçekten kullanılabilir ilk yazılım ise 2000'lerin ortalarında ortaya çıkmıştır (Gbur vd., 2012; Lindsey, 1997; McCullagh ve Nelder, 1989).

İstatistikte bir modelin parametrelerini tahmin etmek birçok istatistiksel analiz için anahtar bir adımdır. GLMM'ler için bu parametreler sabit etki ve rastgele etki parametrelerinden oluşur. GLMM'lerin parametrelerini tahmin etmek için iki temel yaklaşım vardır. Bu yaklaşımlar; olabilirlik fonksiyonuna yaklaşım ve modele yaklaşımdır. Modele yaklaşımda algoritmalar genelde Taylor serileri ile gösterilir. Bu yaklaşımlar lineerleştirme yöntemleri olarak da bilinirler. Lineerleştirme yöntemiyle ilerleyen tahmin süreci durdurma kriteri sağlanıncaya kadar devam eder. GLMM'lerde lineerleştirme yöntemi, rastgele etkileri kapsayan modeller için Pseudo yarı-olabilirlik (Diggle vd., 1989) ve Penalized yarı-olabilirlik (PQL) (Wolfinger ve O'Connell, 1993) yöntemi kullanılır. Pseudo yarı-olabilirlik yöntemi çoklu bağımlı değişkenlerin olduğu bir modelde maksimum olabilirlik tahmininin yapıldığı fakat bu tahminin aslında gerçek yarı-olabilirlik olmadığı durumu ifade eder. Bir başka deyişle doğrudan verilerin dağılımını modellemek yerine, bazı parametrik veya yarı-parametrik bir dağılımın taklit edilmesine dayanır. Bu sebeple elde edilen tahminler, gerçek yarı-olabilirlik değerlerinden farklı olabilir. PQL yöntemi ise, büyük veri setlerinde parametrelerin tahmin edilmesinde kullanılan bir yöntemdir. Bu yöntemle bulunan parametre tahminleri, rastgele etkilerin en iyi yansız tahmin edicilerinin tahminleri olacaktır.

GLMM parametrelerini tahmin etmek için olabilirlik fonksiyonuna dayanan çeşitli integral yaklaşım yöntemleri mevcuttur. Bunlardan en yaygın kullanılanları, Laplace yaklaşımı, Gauss-Hermitte Quadrature (GHQ) (Stroud ve Secrest, 1966) ve Adaptive Gauss-Hermitte Quadrature (AGHQ) (Naylor ve Smith, 1982) yöntemleridir. Laplace yaklaşımının istatistiksel alanda uygulamaları, 20. yüzyılın başlarından itibaren görülmeye başlanmıştır. Olasılık dağılımlarının yaklaşık olarak

çözülmesinde ve parametrik tahminlerin elde edilmesinde kullanılmıştır. Bu yaklaşım, karmaşık olabilirlik fonksiyonlarının elde edilmesinde veya karmaşık integrallerin hesaplanmasında faydalıdır. GHQ yöntemi Hermite polinomlarına dayanmaktadır. Bu polinomlar, 19. yüzyılın ortalarında Charles Hermite tarafından ortaya çıkmıştır. İstatistik alanında yaygın bir şekilde kullanılmaları GHQ yönteminin 20. yüzyılın başlarına, AGHQ yönteminin ise 20. yüzyılın ortalarına dayanmaktadır. Bu iki yöntem de integrallerin analitik olarak çözülememesi halinde, özellikle olasılık dağılımlarının yaklaşık olarak hesaplanması gereken durumlarda veya karmaşık integral hesaplamalarında kullanılır.

GLMM'lerin uygunluğunu değerlendirmek ve farklı modeller arasından en iyi modeli seçmek için AIC (Akaike Information Criterion) (Akaike, 1974) ve BIC (Bayesian Information Criterion) (Schwarz, 1978) bilgi kriterleri kullanılır. Bunların arasında en yaygın olarak kullanılan AIC ve BIC, farklı model seçimi yaklaşımları sunarlar ve bu yaklaşımlar arasında bir denge bulunur. AIC, daha esnek bir model seçimi yaklaşımı sunar, çünkü daha geniş bir model yelpazesini değerlendirmeye alır (Akaike, 1974). Bu, daha karmaşık modelleri, daha basit modellere kıyasla daha az bir kısıtlama ile değerlendirebilir. AIC, daha karmaşık modellerin daha fazla değişkenliği açıklama yeteneğine ve daha düşük bir hataya sahip olabileceğini kabul eder. Buna karşılık, BIC daha katı bir model seçimi yaklaşımı sunar (Schwarz, 1978). BIC, genellikle daha az değişken içeren ve daha az karmaşık modelleri teşvik eder.

Bu çalışmanın temel amacı, GLMM'ler çerçevesinde parametre tahmin yöntemlerini incelemek ve bu yöntemlerin doğruluğunu sonuç çıkarımı ile değerlendirmektir. GLMM'ler, karmaşık ve hiyerarşik yapıdaki verilerin analizinde yaygın olarak kullanılmakta olup, bu modellerde doğru parametre tahmini, elde edilen sonuçların güvenilirliği açısından büyük önem taşımaktadır. Bu çalışma, farklı parametre tahmin yöntemlerini karşılaştırarak, en uygun yöntemin belirlenmesini amaçlamaktadır. Bu tez çalışması beş bölüm olarak oluşmaktadır. Birinci bölüm tez çalışmasına giriş kısmıdır. İkinci bölümde konunun daha iyi anlaşılması açısından model kavramı ve lineer modeller özet olarak verilecektir. GLM kavramı ve üstel dağılım ailesi üye dağılımları incelenecektir. GLM'nin sistematik bileşeni, bağlantı

fonksiyonu ele alınacaktır. Üçüncü bölümde GLMM kavramı detaylı olarak incelenmiş olup normal, Poisson, geometrik ve negatif binom (NB) dağılımlarına sahip GLMM'ler verilecektir. GLMM'deki tahmin yöntemlerinden PQL yöntemi, Laplace yaklaşımı, Gauss-Hermite Quadrature yöntemi, Adaptive Gauss-Hermite Quadrature yöntemi incelenecektir. Dördüncü bölümde ise Bellbird ve Owl veri setlerine ait tanımlayıcı istatistiklere yer verilecektir. Veri setlerine uygun tahmin yöntemleri araştırılarak parametre tahminleri yapılacaktır. GLMM'deki tahmin yöntemleriyle elde edilen tahmin sonuçlarını karşılaştırmak ve sonuç çıkarımı yapabilmek için Bellbird ve Owl veri setleri üzerinde analizler yapılacaktır. Bu tez çalışmasının son kısmı olan beşinci bölüm de ise araştırma sonuçları ve tartışma kısmına yer verilecektir.

2. LİNEER MODELLER VE GENELLEŞTİRİLMİŞ LİNEER MODELLER

2.1 Lineer Modeller

Modelleme, bir veri setindeki açıklayıcı olarak belirtilen gözlemler ile yanıt değişkeni olarak belirtilen gözlemlerin arasındaki ilişkiyi matematiksel olarak tanımlamanın bir yoludur (Stroup, 2012).

Modeller genel olarak, deterministik ve olasılıksal olmak üzere iki geniş sınıfa ayrılırlar. Deterministik modellerde değişkenler kesin olarak tanımlanmış ve kesin sonuçlara sahip iken olasılığa dayalı modellerde değişkenler rastgele olup sonuçları değişkenlik göstermektedir (Myers vd., 2012).

İstatistiksel bir model, veriyi açıklamak, veri yapısını en başarılı şekilde temsil eden matematiksel formu belirlemek ve değişkenler arasındaki ilişkiyi yaklaşık olarak tanımlamak amacıyla kullanılır. İstatistiksel modellerde yanıt değişkeni y 'nin gözlemlenmiş $x_1, x_2, \dots, x_p$ değerleri tarafından açıklandığı en temel model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon \quad (2.1)$$

şeklinde yazılan lineer regresyon modelidir. " $\beta_0, \beta_1, \dots, \beta_p$ " ler modeldeki parametre katsayılarını; " ε " modelin hata terimini; " y " bağımlı (yanıt) değişkeni; " $x_1, x_2, \dots, x_p$ " ler bağımsız (açıklayıcı) değişkeni temsil eder. Eşitlik (2.1) modeli, bağımlı ve bağımsız değişkenlerin n tane olması halinde

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i \quad i = 1, 2, \dots, n \quad (2.2)$$

şeklini alır ve çoklu lineer regresyon modeline dönüşür. Çoklu lineer regresyon modelinin matris gösterimi

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

olmak üzere, modelin kapalı formu

$$y = X\beta + \varepsilon \quad (2.3)$$

şeklinde yazılır. " y "; $(n \times 1)$ boyutlu gözlem vektörü, " X "; $n \times (p + 1)$ boyutlu tasarım matrisi, " β "; $(p + 1) \times 1$ boyutlu bilinmeyen parametre vektörü, " ε "; $(n \times 1)$ boyutlu hata vektörü olarak tanımlanır.

Çoklu lineer regresyon modellerinde lineerlik (bağımlı değişkenin bağımsız değişkenlerin lineer kombinasyonu olarak yazılabilir olması), sabit varyanslılık ($E(\varepsilon_i) = 0$ ve $Var(\varepsilon_i) = \sigma^2$) ve hata terimlerinin birbirinden bağımsız olması ($Cov(\varepsilon_i, \varepsilon_j) = 0; i \neq j$) varsayımları bulunur. Beklenen değer $E(y) = X\beta$ ve varyans $Var(y) = \sigma^2 I$ olmak üzere En Küçük Kareler (EKK) yöntemiyle yapılan parametre tahmini $\hat{\beta} = (X'X)^{-1}X'y$ ve varyans $\hat{\sigma}^2 = \frac{1}{n-p}(y - X\hat{\beta})'(y - X\hat{\beta})$ şeklinde olup dağılımları $y \sim N(X\beta, \sigma^2 I)$ ve $\hat{\beta} \sim (\beta, \sigma^2(X'X)^{-1})$ şeklindedir.

2.2 Genelleştirilmiş Lineer Modeller

Lineer veya lineer olmayan regresyon modelleri, bağımlı değişkenin normal dağılıma sahip ve hataların sabit varyanslı olduğu varsayımına dayanmaktadır. Fakat veri her zaman normal dağılıma sahip olmayabilir. Nelder ve Wedderburn (1972) GLM kavramı ile lineer modellerdeki varsayımları gevşetmiş, başlangıçta lineer değilmiş gibi görünen ilişkilerin lineer hale gelebildiğini göstermiştir (Hardin ve Hilbe, 2007). GLM'ler bağımlı değişkenin normal dağılımlı olmadığı durumlara, değişen varyanslılık durumlarına ve bağımlı değişkenin ortalaması ile bağımsız değişkenler arasında doğrusal olmayan ilişkilere kısıtlama getirmemektedir (Anderson vd., 2012).

GLM'ler, lineer modellerin genişletilmiş halidir ve 3 bileşenden oluşur (Myers vd., 2012; İyit vd., 2016):

1. Bağımlı değişkenin dağılımı
2. Lineer tahmin edicilerin olduğu sistematik kısım
3. Bağlantı (link) fonksiyonu

2.2.1 Bağımlı değişkenin dağılımı (rastgele bileşen)

Fisher (1935), modern istatistik uygulamalarında yer alan dağılımların büyük çoğunluğunun eşitlik (2.4)'deki üstel aile olarak adlandırıldığı bir formda yazılabileceği fikrini ortaya atmıştır (Fox, 2015; Gill ve Torres, 2019). GLM'ler üstel dağılım ailesinden gelmektedir ve buna bağlı olarak bağımlı değişkenin dağılımının üstel dağılım ailesinden olduğu kabul edilir. y değişkeni üstel dağılıma sahip rastgele değişken olmak üzere üstel dağılım ailesinin olasılık yoğunluk fonksiyonu:

$$f(y; \theta, \phi) = \exp \left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right] \quad (2.4)$$

şeklindedir. “a(.), b(.) ve c(.)” fonksiyonları üstel dağılım ailesi olasılık yoğunluk fonksiyonu için gerekli olan özel olarak belirlenen fonksiyonlardır, “ θ ” dağılımın konum parametresidir ve “ ϕ ” dağılımın yayılım parametresidir.

Üstel dağılım ailesinin üyesi olan dağılımlardan; Poisson, Binom, Normal, Gamma dağılımları ve a(.), b(.) ve c(.) fonksiyonları aşağıda verilmiştir (Gbur vd., 2020; Gill ve Torres, 2019; İyit vd., 2016; İyit, 2021; McCulloch ve Searle, 2001; Olsson, 2002).

2.2.1.1 Poisson dağılımı

Poisson dağılımı kesikli bir dağılıma sahip olup bağımlı değişkenin sayılabilir olduğu verilerden oluşan GLM için kullanılır. $y = 0, 1, 2, \dots$ için μ (ortalama) parametrelili olasılık yoğunluk fonksiyonunun üstel dağılım ailesine uyarlanmış hali aşağıda gösterilmiştir (Gbur vd., 2020):

$$\begin{aligned} f(y; \mu) &= \frac{e^{-\mu} \mu^y}{y!} = \exp\{-\mu + y \ln(\mu) - \ln(y!)\} \\ &= \exp\left\{\frac{y \ln(\mu) - \mu}{1} - \ln(y!)\right\}. \end{aligned} \quad (2.5)$$

Eşitlik (2.5)’teki olasılık yoğunluk fonksiyonundan $\theta = \ln \mu$, $b(\theta) = \mu = e^\theta$, $a(\phi) = 1$ ve $c(y; \phi) = -\ln(y!)$ eşitlikleri elde edilir.

2.2.1.2 Binom dağılımı

Binom dağılımı, kesikli bir dağılım olan Bernoulli dağılımının n denemeyle genişletilmiş halidir. Genellikle bağımlı değişkenin 0 ve 1’lerden oluştuğu GLM için kullanılır. $y = 0, 1, 2, \dots, n$ için π bir denemedeki başarı olasılığı olmak üzere binom dağılımının olasılık yoğunluk fonksiyonunun üstel dağılım ailesine uyarlanmış hali aşağıda gösterilmiştir (İyit vd., 2023; Myers vd., 2012):

$$\begin{aligned} f(y; \pi) &= \binom{n}{y} \pi^y (1 - \pi)^{n-y} \\ &= \exp\left\{y \ln \pi + (n - y) \ln(1 - \pi) + \ln \binom{n}{y}\right\} \end{aligned}$$

$$\begin{aligned}
&= \exp \left\{ y \ln \pi - y \ln(1 - \pi) + n \ln(1 - \pi) + \ln \binom{n}{y} \right\} \\
&= \exp \left\{ \frac{y \ln \left(\frac{\pi}{1-\pi} \right) - n \ln \left(\frac{1}{1-\pi} \right)}{1} + \ln \binom{n}{y} \right\}.
\end{aligned} \tag{2.6}$$

Eşitlik (2.6)'daki olasılık yoğunluk fonksiyonundan $\theta = \ln \left(\frac{\pi}{1-\pi} \right)$, $\pi = \frac{e^\theta}{1+e^\theta}$, $b(\theta) = n \ln(1 + e^\theta)$, $a(\phi) = 1$ ve $c(y; \phi) = \ln \binom{n}{y}$ eşitlikleri elde edilir.

2.2.1.3 Normal dağılım

Normal dağılım, μ ve σ^2 (varyans) olmak üzere iki parametrelili, sürekli bir dağılımdır. $-\infty < y < \infty$, $-\infty < \mu < \infty$ ve $0 < \sigma^2 < \infty$ olmak üzere normal dağılımın olasılık yoğunluk fonksiyonunun üstel dağılım ailesine uyarlanmış hali aşağıda gösterilmiştir (Hilbe, 2011):

$$\begin{aligned}
f(y; \mu, \sigma^2) &= \frac{1}{(2\pi\sigma^2)^{1/2}} \exp \left\{ -\frac{1}{2\sigma^2} (y - \mu)^2 \right\} \\
&= \exp \left\{ -\frac{y^2 - 2\mu y + \mu^2}{2\sigma^2} - \frac{1}{2} \ln(2\pi\sigma^2) \right\} \\
&= \exp \left\{ \frac{\mu y - \mu^2/2}{\sigma^2} - \frac{1}{2} \left(\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2) \right) \right\}.
\end{aligned} \tag{2.7}$$

Eşitlik (2.7)'deki olasılık yoğunluk fonksiyonundan $\theta = \mu$, $b(\theta) = \theta^2/2$, $a(\phi) = \sigma^2 = \phi$ ve $c(y; \phi) = -\frac{1}{2} \left(\frac{y^2}{\phi} + \ln(2\pi\phi) \right)$ eşitlikleri elde edilir.

2.2.1.4 Gamma dağılımı

Gamma dağılımı bağımlı değişkenin sürekli olduğu GLM'de kullanılan esnek bir dağılımdır. Birçok dağılım yapısını parametrelerine bağlı olarak içinde bulundurmaktadır. Bağımlı değişkenin sabit varyasyon katsayılı, negatif olmayan ve çarpık dağılıma sahip olduğu durumlarda sıklıkla kullanılır (Gbur vd., 2020).

$\psi > 0$ ölçek (scale) parametresi, $\omega > 0$ biçim (shape) parametresi ve $y > 0$ olmak üzere gamma dağılımının olasılık yoğunluk fonksiyonu aşağıda gösterilmiştir (Fox, 2015):

$$f(y; \psi, \omega) = \frac{1}{\Gamma(\psi)\omega^\psi} y^{\psi-1} e^{-\frac{y}{\omega}}. \tag{2.8}$$

Eşitlik (2.8)'de verilen gamma dağılımının olasılık yoğunluk fonksiyonu, $\psi = \alpha$ ve $\omega = \beta/\alpha$ parametre değişimleri yapılarak $\alpha > 0, \beta > 0$ ve $y > 0$ olmak üzere;

$$f(y; \alpha, \beta) = \frac{1}{\Gamma(\alpha) \left(\frac{\beta}{\alpha}\right)^\alpha} y^{\alpha-1} e^{-\frac{\alpha}{\beta}y} \quad (2.9)$$

şeklinde yazılabilir. Eşitlik (2.9)'da verilen gamma dağılımının olasılık yoğunluk fonksiyonunun üstel dağılım ailesine uyarlanmış hali aşağıda gösterilmiştir:

$$\begin{aligned} f(y; \alpha, \beta) &= \frac{1}{\Gamma(\alpha) \left(\frac{\beta}{\alpha}\right)^\alpha} y^{\alpha-1} e^{-\frac{\alpha}{\beta}y} \\ &= \exp \left\{ \ln \left[\Gamma(\alpha) \left(\frac{\beta}{\alpha}\right)^\alpha \right]^{-1} + (\alpha - 1) \ln y - \frac{\alpha}{\beta} y \right\} \\ &= \exp \left\{ \frac{-\frac{1}{\beta}y - \ln \beta}{1/\alpha} + \alpha \ln(\alpha y) - \ln[y\Gamma(\alpha)] \right\}. \end{aligned} \quad (2.10)$$

Eşitlik (2.10)'daki olasılık yoğunluk fonksiyonundan $\theta = -1/\beta$, $b(\theta) = \ln \beta = -\ln(-\theta)$, $a(\phi) = \phi = 1/\alpha$ ve $c(y; \phi) = \alpha \ln(\alpha y) - \ln[y\Gamma(\alpha)] = -\frac{1}{2} \ln \left(\frac{y}{\phi} \right) - \ln \left[y\Gamma \left(\frac{1}{\phi} \right) \right]$ eşitlikleri elde edilir.

2.2.1.5 Üstel dağılım ailesinin olabilirlik, log-olabilirlik fonksiyonları ve özellikleri

GLM'de parametre tahmininin teorik kısmı üstel dağılım ailesine ait dağılımların olabilirlik ve log-olabilirlik fonksiyonlarına dayanmaktadır (Myers vd., 2012). Üstel dağılım ailesi için yazılan olabilirlik fonksiyonunun çarpım halindeki formu Barndorff-Nielsen (1976) tarafından ortaya atılmıştır. $y_1, y_2, \dots, y_n$ üstel dağılım ailesinin herhangi bir dağılımdan alınmış birbirinden bağımsız n örneklem olmak üzere olabilirlik fonksiyonu ve logaritması alınmış hali olan log-olabilirlik fonksiyonları;

$$L(\theta, \phi; y) = \prod_{i=1}^n \exp \left[\frac{y_i \theta_i - b_i(\theta_i)}{a(\phi)} + c(y_i, \phi) \right] \quad (2.11)$$

$$l(\theta, \phi; y) = \log \left[\prod_{i=1}^n \exp \left[\frac{y_i \theta_i - b_i(\theta_i)}{a(\phi)} + c(y_i, \phi) \right] \right] = \sum_{i=1}^n \left\{ \frac{y_i \theta_i - b_i(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\} \quad (2.12)$$

şeklinde yazılır (Barndorff-Nielsen, 1976; Hardin ve Hilbe, 2007).

Fisher bilgi kriterleri, istatistiksel bir modelin parametrelerinin doğruluğunu ve kesinliğini değerlendirmeye yarayan bir kıstastır. En iyi modeli seçmek için Fisher bilgi kriterlerinden faydalanılır. Ayrıca parametre tahminlerinin güvenilirliğini gösterir. (2.11) ve (2.12)'deki eşitliklerle alakalı Fisher bilgi kriterlerine dayanan önemli özellikler vardır. Bu özellikler (Gould vd., 2006; Mihoc ve Fātu, 2020):

$$E \left[\frac{\partial l(\theta, \phi; y)}{\partial \theta} \right] = 0 \quad (2.13)$$

$$E \left[\frac{\partial^2 l(\theta, \phi; y)}{\partial \theta^2} \right] + E \left\{ \left[\frac{\partial^2 l(\theta, \phi; y)}{\partial \theta^2} \right]^2 \right\} = 0 \quad (2.14)$$

şeklinde. (2.13) eşitliği, log-olabilirlik fonksiyonunun konum parametresine göre birinci türevinin beklenen değerinin sıfır olduğunu belirtmektedir. Doğru konum parametrelerini tahmin etmek için birinci türevin beklenen değerinin sıfır olması gerekmektedir. (2.14) eşitliği, log-olabilirlik fonksiyonunun konum parametresine göre ikinci türevinin beklenen değeriyle, ikinci türevin karesinin beklenen değerinin toplamının sıfır olduğunu belirtmektedir. Doğru konum parametrelerini tahmin etmek için ikinci türevin beklenen değerinin de sıfır olması gerekmektedir.

Eşitlik (2.14)'teki özellikten yola çıkarak;

$$E \left[\frac{\partial l(\theta, \phi; y)}{\partial \theta} \right] = E \left[\frac{y - b'(\theta)}{a(\phi)} \right] = \int_R \frac{y - b'(\theta)}{a(\phi)} f(y; \theta, \phi) y dy = 0 \quad (2.15)$$

$$\underbrace{\int_R f(y; \theta, \phi) y dy}_{=E[y]} - b'(\theta) \underbrace{\int_R f(y; \theta, \phi) dy}_{=1} = 0 \quad (2.16)$$

denklemleri yazılır. Dolayısıyla eşitlik (2.16)'dan;

$$E[y] = \mu = b'(\theta) \quad (2.17)$$

ve eşitlik (2.13)'teki özellikten;

$$E \left[\frac{\partial^2 l(\theta, \phi; y)}{\partial \theta^2} \right] = E \left[\frac{-b''(\theta)}{a(\phi)} \right] = \frac{-b''(\theta)}{a(\phi)} \quad (2.18)$$

$$\begin{aligned} E \left\{ \left[\frac{\partial^2 l(\theta, \phi; y)}{\partial \theta^2} \right]^2 \right\} &= E \left[\frac{y^2 - 2yb'(\theta) + \{b'(\theta)\}^2}{\{a(\phi)\}^2} \right] \\ &= \frac{1}{\{a(\phi)\}^2} \left\{ E[y^2] - \frac{2b'(\theta)E[y] + \{b'(\theta)\}^2}{(E[y])^2} \right\} \\ &= \frac{E[y^2] - (E[y])^2}{\{a(\phi)\}^2} = \frac{Var(y)}{\{a(\phi)\}^2} \end{aligned} \quad (2.19)$$

elde edilir. Eşitlik (2.14)'teki özellik eşitlik (2.18) ve eşitlik (2.19) ile birleştirildiğinde;

$$\frac{-b''(\theta)}{a(\phi)} + \frac{V(y)}{\{a(\phi)\}^2} = 0 \quad (2.20)$$

veya

$$V(y) = b''(\theta)a(\phi) \quad (2.21)$$

elde edilir (Gbur vd., 2020; Gill ve Torres, 2019; Olsson, 2002). (2.21) eşitliği;

$$V(y) = a(\phi) \frac{\partial^2 b(\theta)}{\partial \theta^2} = a(\phi)v(\mu) \quad (2.22)$$

şeklinde yazıldığında $v(\mu) = \frac{\partial^2 b(\theta)}{\partial \theta^2}$ fonksiyonuna varyans fonksiyonu denilmektedir (McCulloch ve Searle, 2001; Yonar, 2019).

(2.17)'deki ve (2.21)'deki eşitliklerden yola çıkarak, istatistikte belli öneme sahip olan birçok dağılımın bulunduğu üstel dağılım ailesinde hem ortalama hem de varyans hesaplanırken θ kanonik parametresinin bir fonksiyonu olan $b(\theta)$ fonksiyonu önemli bir göreve sahiptir.

İncelenen üstel dağılım ailesi üye dağılımlarının uyarlanmış halleri, $a(\phi)$, $b(\theta)$ ve $c(y; \phi)$ fonksiyonları, θ kanonik parametreleri, $\mu = b'(\theta)$ ortalamaları ve $V(y) = b''(\theta)a(\phi)$ varyansları özetlenerek Tablo 2.1'de verilmiştir (Fox, 2015; Gbur vd., 2020; Gill ve Torres, 2019).

Tablo 2.1 Üstel dağılım ailesi üye dağılımlarının uyarlanmış halleri, fonksiyonları, $a(\phi)$, $b(\theta)$ ve $c(y; \phi)$, θ kanonik parametreleri, $\mu = b'(\theta)$ ortalamaları ve $V(y) = b^*(\theta)a(\phi)$ varyansları

Dağılım	$f(y; \theta, \phi)$	$a(\phi)$	$b(\theta)$	$c(y; \phi)$	θ	$\mu = b'(\theta)$	$V(y) = b^*(\theta)a(\phi)$
Poisson	$\exp\left\{\frac{y\ln(\mu) - \mu}{1} - \ln(y!)\right\}$	1	e^θ	$-\ln(y!)$	$\ln\mu$	e^θ	e^θ
Binom	$\exp\left\{\frac{y\ln\left(\frac{\pi}{1-\pi}\right) - n\ln\left(\frac{1}{1-\pi}\right)}{1} + \ln\binom{n}{y}\right\}$	1	$n\ln(1 + e^\theta)$	$\ln\binom{n}{y}$	$\ln\left(\frac{\pi}{1-\pi}\right)$	$\frac{ne^\theta}{1 + e^\theta}$	$\frac{ne^\theta}{(1 + e^\theta)^2}$
Normal	$\exp\left\{\frac{\mu y - \mu^2/2}{\sigma^2} - \frac{1}{2}\left(\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)\right)\right\}$	$\phi = \sigma^2$	$\frac{\theta^2}{2}$	$-\frac{1}{2}\left(\frac{y^2}{\phi} + \ln(2\pi\phi)\right)$	μ	θ	ϕ
Gamma	$\exp\left\{\frac{-\frac{1}{\beta}y - \ln\beta}{1/\alpha} + \alpha\ln(\alpha y) - \ln[y\Gamma(\alpha)]\right\}$	$\phi = 1/\alpha$	$-\ln(-\theta)$	$-\frac{1}{2}\ln(y/\phi) - \ln[y\Gamma(1/\phi)]$	$-1/\mu$	$-1/\theta$	ϕ/θ^2

2.2.2 Sistematik bileşen

GLM’de lineer tahmin edicilerin olduğu sistematik kısımdır. $i = 1, 2, \dots, n$ gözlem sırası ve $j = 1, 2, \dots, p$ bağımsız değişken sırası olmak üzere bir GLM’in lineer tahmin edicisi, $\{n_i\}$ kümesine karşılık gelen $x_i = (x_{i1}, \dots, x_{ip})$ olacak biçimde $x_1, x_2, \dots, x_p$ bağımsız değişkenlerinin $\beta_0, \beta_1, \dots, \beta_p$ parametreleri ile oluşturduğu lineer bir kombinasyondur;

$$\eta_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} \quad (2.23)$$

“ $\hat{\eta}$ ”; $(n \times 1)$ boyutlu lineer tahmin edici vektörü, “ $\hat{\beta}$ ”; $(p + 1) \times 1$ boyutlu model parametre vektörü ve model matrisi olarak isimlendirilen “ X ” ise $n \times (p + 1)$ boyutlu tasarım matrisi olmak üzere eşitlik (2.23)’ün matris halinde gösterimi;

$$\hat{\eta} = X\hat{\beta} \quad (2.24)$$

şeklindedir (Agresti, 2015; Gbur vd., 2020; Gill ve Torres, 2019).

Modellemelerde çoğunlukla parametre sayısı gözlem sayısından olabildiğince küçük tutularak ($p \leq n$) daha az parametre sayısı ile daha iyi sonuçlar elde edilmesi amaçlanır (Agresti, 2015; İyit, 2021; Yonar ve İyit, 2021).

2.2.3 Bağlantı (link) fonksiyonu

GLM’de bağlantı fonksiyonu $g(\cdot)$, yanıt değişkeninin beklenen değeri (ortalama) ile lineer tahmin edici arasındaki bağlantıyı kuran bir fonksiyondur. $E[y_i] = \mu_i$ ve eşitlik (2.23)’te verilen sistematik kısımla arasındaki bağlantıyı kuran bağlantı fonksiyonu;

$$\eta_i = g(\mu_i) = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} \quad (2.25)$$

şeklinde yazılır. “ $\hat{\mu} = E[\hat{y}]$ ”; $n \times 1$ boyutlu $\hat{y}$ bağımlı değişken vektörünün beklenen değer vektörü, “ $g(\hat{\mu})' = [g(\mu_1), g(\mu_2), \dots, g(\mu_n)]'$ ”; $n \times 1$ boyutlu bağlantı fonksiyon vektörü, “ $\hat{\eta}$ ”; $n \times 1$ boyutlu lineer tahmin edici vektörü, “ $\hat{\beta}$ ”; $(p + 1) \times 1$ boyutlu model parametre vektörü ve model matrisi olarak isimlendirilen “ X ” ise $n \times (p + 1)$ boyutlu tasarım matrisi olmak üzere eşitlik (2.25)’in matris halinde gösterimi;

$$\hat{\eta} = g(\hat{\mu}) = X\hat{\beta} \quad (2.26)$$

şeklinde olur (Agresti, 2015; Gbur vd., 2020; Gill ve Torres, 2019; İyit vd., 2023).

Bağlantı fonksiyonu monoton ve türevlenebilir bir fonksiyondur. GLM’de bağlantı fonksiyonu bağımlı değişkenin kendisinden çok beklenen değerini sistematik kısım ile ilişkilendirir, yani $g(y_i) \neq g(\mu_i)$ şeklinde gösterilir. Böylece hem bağımlı değişken orijinal ölçeğinde kalır hem de model, bağımlı değişkenin ortalama bir ölçeği ile lineer tahmin edicilerin lineer kombinasyonu (sistematik kısım) arasında kurulmuş olur (Gbur vd., 2020; Olsson, 2002; Yonar, 2019).

Kanonik bağlantı ve kanonik olmayan bağlantı fonksiyonları olmak üzere iki tür bağlantı fonksiyonu vardır. Eğer $g(\mu_i) = \theta_i$ eşitliği varsa $g(\cdot)$ bağlantı fonksiyonuna kanonik bağlantı fonksiyonu denir. Bağlantı fonksiyonu, üstel dağılım ailesindeki dağılımın türüne göre ortaya çıkar. GLM’de yaygın olarak kullanılan bazı kanonik bağlantı fonksiyonları, ters bağlantı fonksiyonları, ortaya çıktıkları dağılımlar ve θ parametreleri Tablo 2.2’de gösterilmiştir (Fox, 2015; Gbur vd., 2020; İyit vd., 2016; Myers vd., 2012; Urgancı Tekin vd., 2021; Yonar ve Neslihan, 2018).

Tablo 2.2 GLM’de yaygın olarak kullanılan kanonik bağlantı fonksiyonları

Dağılım	θ	Kanonik Bağlantı Fonksiyonu	Ters Bağlantı Fonksiyonu
		$\eta = g(\mu)$	$\mu = g^{-1}(\eta)$
Normal	Birim	μ	η
Binom	Logit	$\ln\left(\frac{\pi}{1-\pi}\right)$	$\frac{1}{1 + \exp(-\eta)}$
Poisson	Log	$\ln \mu$	$\exp(\eta)$
Gamma	Ters	μ^{-1}	η^{-1}

GLM’de kullanılan diğer bağlantı fonksiyonları aşağıda verilmiştir (Fox, 2015; Myers vd., 2012).

Probit bağlantı fonksiyonu: Probit bağlantı ve ters bağlantı fonksiyonları;

$$\eta_i = \Phi^{-1}[E(y_i)] \text{ veya } \mu_i = \Phi^{-1}[\eta_i] \quad (2.27)$$

şeklinde yazılır (Olsson, 2002; Yonar, 2019). “ Φ ”, birikimli standart normal dağılım fonksiyonudur.

Tamamlayıcı log-log bağlantı fonksiyonu: Tamamlayıcı log-log bağlantı fonksiyonu ve ters bağlantı fonksiyonları;

$$\eta = g(\mu) = \ln[-\ln(1 - \mu)] \quad (2.28)$$

$$\mu = g^{-1}(\eta) = 1 - \exp[-\exp(\eta)] \quad (2.29)$$

şeklindedir (Olsson, 2002; Fox, 2015).

Güç (Power) bağlantı fonksiyonlar ailesi: Genel güç bağlantı fonksiyon ailesi ve tersi;

$$Power(\alpha) = \begin{cases} \mu^\alpha, & \alpha \neq 0 \\ \ln \mu, & \alpha = 0 \end{cases} \quad (2.30)$$

$$Power(\alpha) = \begin{cases} \eta^{-\alpha}, & \alpha \neq 0 \\ \exp(\eta), & \alpha = 0 \end{cases} \quad (2.31)$$

şeklindedir.

3. GENELLEŞTİRİLMİŞ LİNEER KARMA MODELLER VE TAHMİN YÖNTEMLERİ

3.1 Lineer Karma Modeller

GLMM'ler incelenmeden önce LMM'lere kısaca değinilmiştir. Bir modelde parametreler sabit etkili ise lineer model, parametrelerin bazıları sabit etkili bazıları rastgele etkili olduklarında ise LMM kullanılmaktadır. Lineer modelleri ve LMM'leri birbirinden ayıran en önemli özellik model kurma amacının farklı olmasıdır. Lineer modellerin amacı; bağımlı değişken vektörünün ortalama vektör yapısını parametre vektörünü kullanarak modellemek iken, LMM'lerin amacı; bağımlı değişken vektörünün varyans-kovaryans matris yapısını rastgele etki vektörüne bağlı olarak modelleyebilmektir.

LMM'ler, yanıt değişkeninin normal olduğu ve bağımlı değişkenin ortalaması ile sabit ve rastgele etkiler arasındaki ilişkinin doğrusal bir fonksiyon olarak modellenebileceği, varyansın ortalamaya bağlı olmadığı, rastgele etkilerin normal bir şekilde dağıldığı varsayımlarını içeren modellerdir.

LMM'lerin matris halinde gösterimi;

$$y = X\beta + Zu + \varepsilon \quad (3.1)$$

şeklinde. Burada “ y ”; $(n \times 1)$ boyutlu bağımlı değişken vektörü, “ X ”; sabit etkilerin $n \times (p + 1)$ boyutlu tasarım matrisi, “ β ”; $(p + 1) \times 1$ boyutlu bilinmeyen parametre vektörü, “ Z ”; $n \times q$ boyutlu tasarım matrisi, “ u ”; $q \times 1$ boyutlu rastgele etkilerin bilinmeyen parametre vektörü, “ ε ”; $(n \times 1)$ boyutlu hata vektörü olarak tanımlanır. Rastgele etkiler vektörü u , ortalama değeri 0 ve varyans matrisi G olan normal dağılıma sahiptir ($u \sim N(0, G)$). Hata vektörü ε , ortalama değeri 0 ve varyans-kovaryans matrisi R olan normal dağılıma sahiptir ($\varepsilon \sim N(0, R)$); u ve ε vektörlerinin birbirinden bağımsız olduğu varsayılmaktadır ($Cov(\varepsilon, u) = 0$).

Eşitlik (3.1)'deki model, olasılık dağılımı açısından iki şekilde tanımlanabilir. İlk formu marjinal model $y \sim N(X\beta, V = ZGZ' + R)$ olmak üzere, ortalama sadece sabit etkilere dayanmaktadır ve rastgele etkileri tanımlayan parametreler varyans-kovaryans matrisi V 'de bulunur (Littell vd., 2006). İkinci formu ise koşullu model $y|u \sim N(X\beta + Zu, R)$ 'dir. Normallik varsayımı altında, her iki model de tam olarak

aynıdır ve aynı çözümü üretir. Normallik varsayımı sağlanmadığında ise modeller farklı çözümler üretir (Stroup, 2012).

3.2 Genelleştirilmiş Lineer Karma Modeller

İstatistiksel modellerde, modeli etkileyen faktörler sabit etki ve rastgele etki olmak üzere ikiye ayrılırlar. Sabit etkiler, modelde bağımlı değişkenlerin ya da ortak değişkenlerin gözlenen değerleri için yapılan istatistiksel çıkarımla alakalıdır. Rastgele etkiler ise daha çok modele alınan değişkenlerin gözlem değerlerinin veriye rastgele olarak daha önceden belirlenen seviyelerde örneklenme şekliyle alakalıdır. Bu sebeple sabit etkiler modelin bağımlı değişkeni için bir model belirlerken rastgele etkiler ise modelin varyans-kovaryans yapısı için bir model belirlemektedir (Gbur vd., 2020; İyit ve Genç, 2009; İyit ve Genç, 2011).

GLMM'ler, GLM'de lineer tahmin ediciye rastgele etki ya da etkiler eklenmesiyle oluşan modellerdir. GLMM, LMM ve GLM'nin bir araya gelmesiyle oluşmuştur ve geniş bir model sınıfını içermektedir. GLMM, GLM'de olduğu gibi bağımlı değişkenin üstel dağılım ailesi gibi istatistikte kullanılan geniş bir dağılım sınıfından gelmesine imkân sunarken aynı zamanda modelin lineer tahmin edici kısmına rastgele etki ya da etkiler ekleyerek modele varyans-kovaryans katkısı da sağlamaktadır (İyit ve Genç, 2009; İyit ve Genç, 2011; Jiang ve Nguyen, 2007; Myers vd., 2012; Stroup, 2012).

3.2.1 Genelleştirilmiş lineer karma modellerin genel yapısı

3.2.1.1 Bağımlı değişkenin koşullu dağılımı

GLMM'lerde rastgele etkiler verildiğinde bağımlı değişkenin koşullu dağılımı üstel dağılım ailesi üyesinden gelmektedir (Gbur vd., 2020; Jiang ve Nguyen, 2007; Stroup, 2012). Koşullu dağılım;

$$f_{y_i|u}(y_i|u; \theta_i, \phi) = \exp \left[\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right] \quad (3.2)$$

şeklindedir. “ u ” rastgele etkiler verildiğinde “ y_i ” i 'nci ($i = 1, 2, \dots, n$) bağımlı değişkeninin koşullu dağılımı üstel dağılım ailesinden gelmek üzere, “ $x_i, \dots, x_p$ ” ; p adet sabit etki diye adlandırılan bağımsız değişken, “ x_{ij} ” ; j 'nci ($j = 1, 2, \dots, p$) sabit

etkili bağımsız değişkenin i 'nci gözlenen değeri, " $u_1, u_2, \dots, u_q$ "; q adet rastgele etki, " z_{kj} "; k 'nci ($k = 1, 2, \dots, q$) u_k rastgele etkisinin j 'nci gözlenen değeri için rastgele etki katsayısı ve " β_j " j 'nci bağımsız değişken için model parametresi olmak üzere GLMM'de bağımlı değişkenin koşullu beklenen değeri ve bağımlı değişkenin koşullu ortalamasını lineer tahmin ediciye bağlayan bağlantı fonksiyonu " $g(\cdot)$ " ile oluşturulan modelin açık halde gösterimi;

$$\begin{aligned}\eta_1 &= g(E[y_1|u_1, u_2, \dots, u_q]) = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_p x_{1p} + u_1 z_{11} + \\ &u_2 z_{12} + \dots + u_q z_{1q} \\ \eta_2 &= g(E[y_2|u_1, u_2, \dots, u_q]) = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_p x_{2p} + u_1 z_{21} + \\ &u_2 z_{22} + \dots + u_q z_{2q} \\ &\vdots \\ \eta_n &= g(E[y_n|u_1, u_2, \dots, u_q]) = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_p x_{np} + u_1 z_{n1} + \\ &u_2 z_{n2} + \dots + u_q z_{nq}\end{aligned}\quad (3.3)$$

ve genel gösterimi;

$$\eta_i = g(E[y_i|u]) = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + \sum_{k=1}^q u_k z_{ik} \quad (3.4)$$

şeklindedir. Eşitlik (3.4) GLMM model denkleminin vektörel halde gösterimi;

$$x_i = (1 \ x_{i1} \ x_{i2} \ \dots \ x_{ip}), \quad z_i = (1 \ z_{i1} \ z_{i2} \ \dots \ z_{iq}), \quad \beta = (\beta_0 \ \beta_1 \ \dots \ \beta_p)' \quad \text{ve} \quad u = (u_1 \ u_2 \ \dots \ u_q)' \text{ olmak üzere}$$

$$g(\mu_i) = x_i' \beta + z_i' u \quad (3.5)$$

şeklindedir (Gbur vd., 2020). (2.22)'de $a(\phi) = \sigma^2$ olarak alınırsa $v(\cdot)$ varyans fonksiyonu olmak üzere, u rastgele etki vektörü verildiğinde bağımlı değişkenin varyansı;

$$Var[y_i|u] = \sigma^2 v(g^{-1}(x_i' \beta + z_i' u)) = \sigma^2 v(\mu_i) \quad (3.6)$$

şeklinde yazılır.

3.2.1.2 Bağımlı değişkenin (marjinal) ortalaması, varyansı, kovaryansı ve korelasyonu

y herhangi bir rastgele değişken, $y|u$ ise u verildiğinde koşullu rastgele değişken olmak üzere, GLMM'de bağımlı değişkenin (marjinal) ortalaması;

$$E[y_i] = E[g^{-1}(x_i' \beta + z_i' u)] = E[y_i|u] = E[\mu_i] \quad (3.7)$$

şeklindedir. Bağlantı fonksiyonu olarak $g(\mu_i) = \log(\mu_i)$ ya da $g^{-1}(\mu_i) = \exp(\mu_i)$ kullanılarak

$$E[y_i] = E[g^{-1}(x_i'\beta + z_i'u)] = E[\exp\{x_i'\beta + z_i'u\}] = \exp\{x_i'\beta\}E[\exp\{z_i'u\}] = \exp\{x_i'\beta\}M_u(z_i) \quad (3.8)$$

elde edilir. “ $M_u(z_i)$ ”; u rastgele etki vektörünün z_i noktasında moment çıkaran fonksiyonudur (Öztürk, 2011). GLMM’de $u \sim N(0, \sigma_u^2)$ ve “ Z ” rastgele etki matrisinin her satırında sadece bir elemanın 1 ve diğerlerinin 0 olması sebebiyle;

$$M_u(z_i) = \exp\{\sigma_u^2/2\} \quad (3.9)$$

olarak yazılacağından;

$$E[y_i] = \exp\{x_i'\beta\}M_u(z_i) = \exp\{x_i'\beta\}\exp\{\sigma_u^2/2\} \quad (3.10)$$

$$\log(E[y_i]) = x_i'\beta + \sigma_u^2/2 \quad (3.11)$$

elde edilir (McCulloch ve Searle, 2001; Jiang ve Nguyen, 2007).

$Var(y)$ (marjinal) varyansı hesaplamak için koşullu varyans ile koşullu ortalamaya ait eşitlik (Gueorguieva, 2001; McCulloch ve Searle, 2001; Vangeneugden vd., 2010);

$$Var(y) = Var(E[y|u]) + E[Var(y|u)] \quad (3.11)$$

şeklindedir.

$$\begin{aligned} \sigma_y^2 &= Var(y_i) = Var(E[y_i|u]) + E[Var(y_i|u)] \\ &= Var(\mu_i) + E[\sigma^2 v(\mu_i)] \\ &= Var(g^{-1}(x_i'\beta + z_i'u)) + E[\sigma^2 v(g^{-1}(x_i'\beta + z_i'u))]. \end{aligned} \quad (3.12)$$

Eşitlik (3.11)’de (Carrasco, 2010; McCulloch ve Searle, 2001; Searle vd., 1992) g^{-1} ters bağlantı fonksiyonu lineer olmayan bir fonksiyon olduğu için daha açık bir halde belirtilemez.

Bağlantı fonksiyonu olarak $g(\mu_i) = \log(\mu_i)$ ya da $g^{-1}(\mu_i) = \exp(\mu_i)$ kullanılarak bağımlı değişkenin varyansı;

$$\begin{aligned} Var(y_i) &= E[(\exp\{x_i'\beta + z_i'u\})^2] - (E[\exp\{x_i'\beta + z_i'u\}])^2 \\ &\quad + E[\sigma^2 v(\exp\{x_i'\beta + z_i'u\})] \\ &= \exp\{2x_i'\beta\}E[\exp\{z_i'u\}] - (\exp\{x_i'\beta\})^2(E[\exp\{z_i'u\}])^2 \\ &\quad + E[\sigma^2 v(\exp\{x_i'\beta + z_i'u\})] \\ &= \exp\{2x_i'\beta\}[M_u(2z_i) - [M_u(z_i)]^2] + E[\sigma^2 v(\exp\{x_i'\beta + z_i'u\})] \end{aligned} \quad (3.13)$$

şeklinde elde edilir ve eşitlik (3.9)'daki gibi u rastgele değişkeninin moment çıkaran fonksiyonu $M_u(z_i) = \exp\{\sigma_u^2/2\}$ yerine yazılarak eşitlik (3.13);

$$Var(y_i) = \exp\{2x_i'\beta\}[\exp\{2\sigma_u^2\} - \exp\{\sigma_u^2\}] + E[\sigma^2 v(\exp\{x_i'\beta + z_i'u\})] \quad (3.14)$$

şeklinde düzenlenir.

y 'nin terimlerinin koşullu bağımsız varsayımıyla;

$$\begin{aligned} cov(y_i, y_j) &= cov(E[y_i|u], E[y_j|u]) + E[cov(y_i, y_j|u)] \\ &= cov(\mu_i, \mu_j) + E[0] \\ &= cov(g^{-1}(x_i'\beta + z_i'u), g^{-1}(x_j'\beta + z_j'u)) \end{aligned} \quad (3.15)$$

şeklinde ifade edilir. Bağlantı fonksiyonu olarak $g(\mu_i) = \log(\mu_i)$ ya da $g^{-1}(\mu_i) = \exp(\mu_i)$ kullanılarak kovaryans;

$$\begin{aligned} cov(y_i, y_j) &= cov(\exp\{x_i'\beta + z_i'u\}, \exp\{x_j'\beta + z_j'u\}) \\ &= \exp\{x_i'\beta + x_j'\beta\} cov(\exp\{z_i'u\}, \exp\{z_j'u\}) \\ &= \exp\{x_i'\beta + x_j'\beta\} [M_u(z_i + z_j) - M_u(z_i)M_u(z_j)] \end{aligned} \quad (3.16)$$

şeklinde elde edilir. GLMM'de $u \sim N(0, \sigma_u^2)$ ve “Z” rastgele etki matrisinin her satırında sadece bir elemanın 1 ve diğerlerinin 0 olması sebebiyle;

$$cov(y_i, y_j) = \exp\{x_i'\beta + x_j'\beta\} (\exp\{\sigma_u^2\} (\exp\{z_i'z_j'\sigma_u^2\} - 1)) \quad (3.17)$$

olarak elde edilir. Eğer herhangi iki gözlem aynı rastgele etkiye sahip değilse $z_i'z_j' = 0$, sebebiyle $cov(y_i, y_j) = 0$ olacağı bellidir. Diğer durumlarda ise $z_i'z_j' = 1$ olacağı için eşitlik (3.17);

$$cov(y_i, y_j) = \exp\{(x_i', x_j')\beta\} [\exp\{\sigma_u^2\} (\exp\{\sigma_u^2\} - 1)] \quad (3.18)$$

şekline dönüşür.

$z_i'z_j' = 1$ alındığında $\eta = \frac{1}{e^{3\sigma_u^2/2} - e^{\sigma_u^2/2}}$ olmak üzere korelasyon;

$$\begin{aligned} \rho_{y_i, y_j} &= \frac{e^{2\sigma_u^2} - e^{\sigma_u^2}}{\sqrt{(e^{2\sigma_u^2} - e^{\sigma_u^2} + e^{-x_i'\beta + \sigma_u^2/2})(e^{2\sigma_u^2} - e^{\sigma_u^2} + e^{-x_j'\beta + \sigma_u^2/2})}} \\ &= \frac{1}{(1 + \eta e^{-x_i'\beta})(1 + \eta e^{-x_j'\beta})} \end{aligned} \quad (3.19)$$

şeklinde bulunur.

3.2.2 Genelleştirilmiş lineer karma modellerde olabilirlik fonksiyonu

GLMM’de olabilirlik fonksiyonu, modelin parametrelerine bağlı olarak gözlemlenen verilerin olasılığını ifade eder. Bu tür modeller, bağımlı değişkenin bir lineer kombinasyonu ve bir bağımsız değişkenin genellikle belirli bir dağılıma göre modellenmiş hata teriminden meydana gelir. Olabilirlik fonksiyonu analitik olarak hesaplanamayan çok boyutlu integralleri kapsayabilir, bu yüzden kapalı formda ifade edilemeyebilir (Jiang, 2007).

Bir GLMM’nin olabilirlik fonksiyonu, veri noktalarının dağılıma göre olasılığının çarpımı olarak ifade edilir. Olabilirlik fonksiyonu maksimize edilmeye çalışılır, yani en yüksek olasılığı sağlayan model parametrelerini bulmak amaçlanır. GLMM’de marjinal olabilirlik fonksiyonunun en genel hali:

$$L(\beta, \theta; y) = \int \prod_{i=1}^n f_{Y_i|u}(y_i|u; \theta_i, \phi) f_u(u) du \quad (3.20)$$

şeklinde yazılır.

$f_{Y_i|u}(y_i|u)$ üstel dağılım ailesinden bir dağılıma sahipse ve $u_i \sim N(0, \sigma_u^2)$ ise marjinal olabilirlik fonksiyonu;

$$\begin{aligned} L(\beta, \theta; y) &= \int \prod_{i=1}^n \prod_{j=1}^p f_{Y_{ij}|u}(y_{ij}|u; \theta_{ij}, \phi) f_u(u) du \\ &= \int \prod_{i=1}^n \prod_{j=1}^p \exp \left[\frac{y_{ij} \theta_{ij} - b(\theta_{ij})}{a(\phi)} + c(y_{ij}, \phi) \right] f_u(u) du \end{aligned} \quad (3.21)$$

şeklinde ifade edilir.

3.3 Genelleştirilmiş Lineer Karma Modellerde Parametre Tahmini

GLMM’de her bir gözlemin ayrı ayrı rastgele etki içermesi, yani bütün değişkenlerin ilişkili olarak modellenmesi halinde ve log bağlantı fonksiyonu kullanılması durumunda marjinal olabilirlik fonksiyonları çok boyutlu integralleri kapsayabilir. Bu marjinal olabilirlik fonksiyonlarını lineer modellerde ya da GLM’de olduğu gibi maksimize ederek bir yaklaşım elde etmek oldukça zordur.

GLMM’de parametre tahmini yaparken iki temel yaklaşım vardır. Birinci yaklaşım lineerleştirme yöntemlerine dayalı yaklaşımlardır, ikincisi ise integral yaklaşımlardır (Gbur vd., 2020; Kim vd., 2015; Stroup, 2012).

3.3.1 Lineerleştirme yöntemleri

Lineerleştirme yöntemlerinde Taylor serileriyle log-olabilirlik fonksiyonuna lineer bir yaklaşım söz konusudur. Lineerleştirme yöntemi ile GLMM'deki bağımlı değişken, LMM'deki bir bağımlı değişken gibi davranacak pseudo (yalancı) değişkene dönüştürülür.

GLMM'de lineerleştirme yöntemleri, PQL (Breslow ve Clayton, 1993) ve Pseudo yarı-olabilirlik (Wolfinger ve O'connell, 1993) yaklaşımlarıdır.

3.3.1.1 Penalized yarı-olabilirlik ve Pseudo yarı-olabilirlik yöntemleri

Maksimum olabilirlik tahmini, veri dağılımı hakkında tam bilgi gerektirir. Başka bir deyişle, maksimum olabilirlik tahmini tam bir olasılık fonksiyonuna dayanmalıdır. Yalnızca ortalamalar veya ortalamalar ile varyanslar arasındaki ilişki biliniyorsa, parametre tahmininde yarı-olabilirlik yöntemleri kullanılmaktadır. Yarı-olabilirlik yöntemi, modeldeki hatalara karşı daha dirençli olabilir. Bu nedenle, model yanlış belirtildiğinde veya model normal dağılıma sahip değilse, yarı-olabilirlik yöntemi daha güvenilir tahminler sunabilir.

GLMM'de, verilerin dağılımı bilinmediğinde, varyans bileşenlerinin tahmini için rastgele etkilere bir ceza terimi ekleyerek yarı-olabilirlik yöntemi kullanılmaktadır. Bu, PQL yöntemi olarak bilinmektedir. Ceza eklemenin amacı, rastgele etkinin bazı keyfi değerlerini önlemek ve rastgele etkinin sifıra yakın olmasını sağlamaktır. PQL yöntemi, verilerin dağılımının belirsiz olduğu veya doğru bir şekilde modellenmesi mümkün olmadığı durumlarda esneklik ve dayanıklılık sağlamaktadır.

PQL, GLMM için parametre tahmin yöntemleri arasında yaygın olarak kullanılan ilk yöntemidir. Bu yöntem, olasılık fonksiyonunun yaklaşık olarak bulunması üzerine çalışır. PQL, Breslow ve Clayton (1993) tarafından değiştirilmiş bir Laplace yaklaşımı olarak tanıtılmıştır. Normal dağılımlı yaklaşık bir olabilirlik sağlamıştır.

Verilere Laplace yaklaşımı genellikle PQL'de uygulanır. GLMM'de, koşullu beklenen değer $E(y_i|u) = h(x_i'\beta + z_i'u)$ şeklinde yazılır. Bağımlı değişken y_i için yaklaşım formülü;

$$y_i \approx \mu_i + \varepsilon_i = E(y_i|u) + \varepsilon_i = h(x_i'\beta + z_i'u) + \varepsilon_i \quad (3.22)$$

şeklindedir. Eşitlik (3.5)'te verilen genel model denkleminde dayanarak $h(\cdot) = g^{-1}(\cdot)$ ters bağlantı fonksiyonu olarak belirtilsin. Eşitlik (3.22)'deki y_i için $\hat{\beta}$ ve $\hat{u}$ etrafında bir Taylor serisi açılımı;

$$\begin{aligned} y_i &\approx h(x_i'\hat{\beta} + z_i'\hat{u}) + h'(x_i'\hat{\beta} + z_i'\hat{u})x_i'(\beta - \hat{\beta}) + h'(x_i'\hat{\beta} + z_i'\hat{u})z_i'(\beta - \hat{\beta}) + \varepsilon_i \\ &= \hat{\mu}_i + V(\hat{\mu}_i)x_i'(\beta - \hat{\beta}) + V(\hat{\mu}_i)z_i'(\beta - \hat{\beta}) + \varepsilon_i \end{aligned} \quad (3.23)$$

şeklindedir ve eşitlik (3.23)

$$\begin{aligned} y_i^* &\equiv \hat{V}_i^{-1}(y_i - \hat{\mu}_i) + x_i\hat{\beta} + z_i\hat{u}_i \\ &\approx x_i\beta + z_iu_i + \varepsilon_i^* \end{aligned} \quad (3.24)$$

şeklinde de yazılabilir. “ y_i^* ” sözde yanıt olarak bilinmektedir. Model uyumu, y_i^* 'nin güncellenmesi ve LMM uyumu arasında sürecin tekrarlanmasıyla gerçekleştirilir.

Model uyumunu sağlamak için β ve varyans bileşeni θ değerleri ayarlanır. u_i ve y_i^* için Bayes tahmin edicisi hesaplanır. y_i^* 'ye dayanarak model kurulur ve β , θ 'ya göre güncellenir. Yakınsama sağlanana kadar bu adımlar tekrarlanır.

u_i verildiğinde y_i 'nin koşullu varyansı;

$$var(y_i|u) = a_i'(\phi)v(\mu_i) \quad (3.25)$$

şeklindedir. “ $a_i(\cdot)$ ” genellikle ϕ/w_i 'ye eşit belirtilen bilinen bir fonksiyondur; “ w_i ” değeri belirtilen bir ağırlıktır ve “ $v(\cdot)$ ” bilinen bir varyans fonksiyonudur.

PQL yöntemiyle parametre tahmini, yarı-olabilirlik fonksiyonunun türetilmesiyle de açıklanabilir. Bağımlı değişken $y = (y_1, y_2, \dots, y_n)'$ için $y_1, y_2, \dots, y_n$ değişkenleri birbirlerinden bağımsız ve $u \sim N(0, G)$ iken, yarı-olabilirlik fonksiyonu;

$$L_Q \propto |G|^{-1/2} \int \exp \left\{ -\frac{1}{2} \sum_{i=1}^n d_i - \frac{1}{2} u' G^{-1} u \right\} du \quad (3.26)$$

şeklindedir. Burada, $d_i = -2 \int_{y_i}^{\mu_i} \frac{y_i - u}{a_i(\phi)v(u)} du$ sapma olarak adlandırılır.

Laplace yaklaşımı kullanılarak, L_Q 'nin logaritması olan l_Q ;

$$l_Q \approx c - \frac{1}{2} \log |G| - \frac{1}{2} \log |q''(\tilde{u})| - q(\tilde{u}) \quad (3.27)$$

elde edilir. “ c ” parametreye bağılı olmayan bir değerdir, $q(u) = \frac{1}{2}(\sum_{i=1}^n d_i + u'G^{-1}u)$ ve “ $\tilde{u}$ ”, $q(u)$ ’yu minimize eden bir değerdir. Özellikle, u , $q'(u) = 0$ için bir çözümdür.

$$G^{-1}u - \sum_{i=1}^n \frac{y_i - \mu_i}{a_i(\phi)v(\mu_i)g'(\mu_i)} z_i = 0 \quad (3.28)$$

burada $\mu_i = x_i'\beta + z_i'u$ ‘dur. $q(u)$ ’nun ikinci türevi;

$$q''(u) = G^{-1} \sum_{i=1}^n \frac{z_i - z_i'}{a_i(\phi)v(\mu_i)\{g'(\mu_i)\}^2} + r \quad (3.29)$$

şeklinde elde edilir. “ r ” beklenen değeri sıfır olan bir artıktır. Aşağıda verilecek olan (3.33)’teki eşitliğin paydası w_i^{-1} ile gösterilirse ve r artık terimi göz ardı edilirse, $q''(u)$ yaklaşımı;

$$q''(u) \approx Z'WZ + G^{-1} \quad (3.30)$$

şeklinde yazılır. “ Z ”, i ’nci satırda z_i' elemanına sahip bir matris ve w_i , GLM döngü ağırlığı olmak üzere $W = \text{diag}(w_1, w_2, \dots, w_n)$ ’dir. Eşitlik (3.29) ve (3.30) birleştirilerek, yarı-olabilirlik fonksiyonunun logaritması;

$$l_Q \approx c - \frac{1}{2}(\log|I + Z'WZG| + \sum_{i=1}^n \tilde{d}_i + \tilde{u}'G^{-1}\tilde{u}) \quad (3.31)$$

şeklinde elde edilir. “ $\tilde{d}_i$ ”, d_i ’nin α yerine $\tilde{u}$ ile değiştirilmiş halidir.

Yarı-olabilirlik fonksiyonunun logaritması için başka bir yaklaşım, GLM ağırlıklarının sabit olmadıkları, fakat ortalamanın fonksiyonları olduklarını varsayarak elde edilebilir. Eşitlik (3.31)’de parantez içindeki ilk terim β ’ya yalnızca W ile bağlıdır, bu yüzden bu terim göz ardı edilebilir. Dolayısıyla yarı-olabilirlik fonksiyonunun logaritması için yaklaşım;

$$l_{PQ} \approx c - \frac{1}{2}(\sum_{i=1}^n \tilde{d}_i + \tilde{u}'G^{-1}\tilde{u}) \quad (3.32)$$

şeklinde yazılır. Eşitlik (3.32), $q(u)$ ’yu minimize eden $\tilde{u}$ ’ya ceza terimi eklenmiş yarı-olabilirlik şeklidir. Eğer β verilirse $\tilde{u}$, l_{PQ} ’yu maksimize edecektir. Çünkü $\tilde{u}$, β ’ya bağılı olduğundan $\tilde{u} = \tilde{u}(\beta)$ şeklinde yazılabilir. Sabit bir θ için, $\hat{\beta}$ ’nın l_{PQ} ’yu maksimize eden bir değer olarak düşünülmesi gerekir. Sonuç olarak, $\hat{\beta}$ ve $\tilde{u}$;

$$l_{PQ}(\beta, u) = -\frac{1}{2}(\sum_{i=1}^n d_i + u'G^{-1}u) \quad (3.33)$$

fonksiyonunu maksimize eder. $l_{PQ}(\beta, u)$, $\tilde{u} = \tilde{u}(\hat{\beta})$ olmak üzere β ve u ’nun bir fonksiyonudur. Eşitlik (3.33)’teki fonksiyonu maksimize etmek için standart

yöntem, lineer olmayan bir denklem sisteminin çözülmesidir, $\partial l_{PQ}/\partial \beta = 0$ ve $\partial l_{PQ}/\partial u = 0$, veya

$$\sum_{i=1}^n \frac{(y_i - \mu_i)x_i}{a_i(\phi)v(\mu_i)g'(\mu_i)} G^{-1}u = 0 \quad (3.34)$$

$$\sum_{i=1}^n \frac{(y_i - \mu_i)z_i}{a_i(\phi)v(\mu_i)g'(\mu_i)} G^{-1}u = 0 \quad (3.35)$$

formülleri çözümle eşdeğerdir.

Uygulamada, GLMM'deki rastgele etkiler genellikle yeterince büyüktür, bu nedenle eşitlik (3.34) ve (3.35)'in çözümleri oldukça karmaşıktır. Breslow ve Clayton (1993), eşitlik (3.34) ve (3.35)'i çözmek için genelleştirilmiş bir tahmin edici momentler (GEE) yöntemi önermişlerdir. Bu yöntem, doğru dağılım modeli belirleme zorunluluğunu ortadan kaldırarak, marjinal model parametrelerinin doğru tahmin edilmesini sağlar. Bu yöntemin ilginç bir özelliği, Henderson vd. (1959) tarafından geliştirilen lineer karma model eşitlikleri ile yakın bir ilişkiyi kullanmasıdır. İlk önce çalışma vektörü $\tilde{y} = (\tilde{y}_i)$; $i = 1, 2, \dots, n$ tanımlanır, $\tilde{y}_i = \eta_i + g'(\mu_i)(y_i - \mu_i)$ şeklindedir, η_i ve μ_i mevcut β ve u tahminleri üzerinde değerlendirilir. Eşitlik (3.34) ve (3.35) için çözüm adımları aşağıdaki gibidir:

$$\begin{pmatrix} X'WX & X'WZ \\ Z'WX & G^{-1} + Z'WZ \end{pmatrix} \begin{pmatrix} \beta \\ u \end{pmatrix} = \begin{pmatrix} X'W \\ Z'W \end{pmatrix} \tilde{y}. \quad (3.36)$$

W , hem β 'ya hem de u 'ya bağlı olduğu için, her adımda W 'yu güncellemek gerekmektedir. Güncellenen W 'lar için eşitlik (3.36)'ya eşdeğer çözüm olarak;

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}\tilde{y} \quad (3.37)$$

$$\hat{u} = GZ'V^{-1}(\tilde{y} - X\hat{\beta}) \quad (3.38)$$

eşitlikleri verilebilir. Her güncelleme için bu eşitlikler tekrar hesaplanır (Handayani vd., 2017). W^{-1} 'in var olduğu varsayılırsa $V = W^{-1} + ZGZ'$ şeklinde yazılır. Büyük veri setleri için V^{-1} 'in hesaplanması, V blok köşegen gibi belirli bir özel yapıya sahip değilse kolay olmayabilir.

PQL tahminçileri tutarsızdır. Bir başka deyişle, yaklaşımdan kaynaklanan sapma sıfır olmayacaktır. Eğer Laplace yaklaşımı kullanılıyorsa, yaklaşımın derecesi büyüdükçe indirgeme yanlılığı daha büyük olacaktır. PQL yöntemi hesaplama avantajlarına sahiptir çünkü genellikle uygulaması kolaydır, ancak yüksek dereceli yaklaşım durumunda zor olabilir. PQL, genellikle varyans bileşenleri küçük olduğunda iyi

çalışır. Bunun nedeni, varyans bileşenleri sıfıra yakın olduğunda Laplace yaklaşımının doğru hale gelmesidir.

Diğer yöntemlerle karşılaştırıldığında, PQL, GLMM'lerin tahmin edilmesi için en hızlı ve en esnek yaklaşımdır. Rastgele etkilerin sayısı ve yapısı, AGHQ kadar sınırlı değildir ve örneklem büyüklüğü arttıkça asimptotik özelliklere sahip olmalıdır (Breslow ve Clayton, 1993). En hızlı olmasına rağmen, özellikle en küçük veri kümeleri söz konusu olduğunda diğer yöntemler arasında en az doğru olanıdır (Bolker, 2013). Breslow ve Clayton, PQL tahmininin, küçük küme boyutları için binom dağılımlı GLMM'lerde sabit etkilere ilişkin tutarsız tahminler verebileceğini belirtmişlerdir ve bir simülasyon çalışmasında bu tutarsızlığı doğrulamışlardır (Breslow ve Clayton, 1993; Tuerlinckx vd., 2006).

Literatürde GLMM'ler ile ilgili çalışmalarda PQL ve Pseudo yarı-olabilirlik arasında bir sınırlama olmaksızın sıklıkla birbirlerinin yerine kullanılabilen terimler olduğu görülmektedir. Her iki yöntem de farklıdır, ancak farklı genişlemelere dayandıkları için aynı parametre tahminlerini vereceklerdir. Aradaki tek fark Pseudo yarı-olabilirlik yönteminde ek bir sabitin olmasıdır (Capanu vd., 2013).

3.3.2 İntegral yaklaşımları

GLMM'de marjinal olabilirlik fonksiyonuna yaklaşımlardan ikincisi integral yaklaşımlarıdır. Marjinal olabilirlik fonksiyonları ya da marjinal log-olabilirlik fonksiyonları hesaplanması zor olan karmaşık integraller içerdiğinden, bu fonksiyonların tahmininde daha hızlı ve yaklaşık sonuçlar veren Laplace yaklaşımı, GHQ ve AGHQ gibi integral yaklaşım yöntemleri kullanılır. Laplace ve Quadrature yaklaşım yöntemlerinin PQL yöntemine göre avantajı, elde edilen tahminlerin daha iyi asimptotik bir davranışa sahip olması ve daha küçük örneklem yanlılığına sahip olma eğiliminde olmasıdır (Gbur vd., 2020; Jiang ve Nguyen, 2007; McCulloch ve Searle, 2001; Stroup, 2012).

Bu çalışmada GLMM'de parametre tahmin yöntemlerinden integral yaklaşımı olarak Laplace yaklaşımı, GHQ ve AGHQ integral yaklaşımları ele alınacaktır.

3.3.2.1 Laplace yaklaşımı

Liu (2016) ve Stroup ve Claasen (2020), GLMM marjinal olabilirlik fonksiyonunun Laplace yaklaşımını bir integral alma işlemi olarak tartışırken, Tuerlinckx vd. (2006) ve Fitzmaurice vd. (2011) ise integrale yaklaşım ve lineerleştirme stratejisi olarak tartışmayı tercih etmektedirler. Her iki yönde yapılan tartışmalar da uygun olabilir. GLMM’de, rastgele etki ile i ’nci kümenin j ’nci birimi için bağımlı değişken Y ’nin koşullu dağılımının bağımsız olduğu ve üstel aile dağılımlarını takip ettiği varsayılır:

$$f(y_{ij}|u_i) = \exp\{\phi^{-1}[y_{ij}\eta_{ij} - \psi(\eta_{ij})] + c(y_{ij}, \phi)\}. \quad (3.39)$$

Burada, $\mu_{ij} = E(Y_{ij}|u_i) = h(\eta_{ij})$ koşullu ortalama ve $\eta_{ij} = x'_{ij}\beta + z'_{ij}u_i$ lineer tahmin edicidir. Rastgele etki u_i ’nin olasılık yoğunluk fonksiyonu $f(u_i)$ ile gösterilir ve $E(u_i) = 0$ ve $Var(u_i) = G$ olmak üzere $u_i \sim N(0, G)$ olduğu varsayılır. Bu nedenle GLMM’de $f(y_{ij}|u_i)$ ve $f(u_i)$ belirtmek gereklidir. $f(y_{ij}|u_i)$ ve $f(u_i)$ özelliklerine dayanarak, u_i verildiğinde y ’nin marjinal yoğunluğu;

$$f(y_{ij}) = \int \prod_{j=1}^{n_i} f(y_{ij}|u_i) f(u_i) du_i \quad (3.40)$$

şeklindedir, burada n_i i ’nci kümenin örneklem büyüklüğüdür. LMM’de marjinal yoğunluk $f(y_{ij})$ normal yoğunluk olduğu için eşitlik (3.40)’ta olan integralin analitik veya kapalı bir formda çözümü elde edilebilir. Genelde $f(y_{ij})$ ’yi hesaplamak kolay değildir çünkü $f(y_{ij}|u_i)f(u_i)$ karmaşık bir fonksiyon olabilir ve integraller yüksek boyutlu olabilir. Bu yüzden integrali çözmek için yaklaşım yöntemleri gereklidir.

GLMM’de olabilirlik fonksiyonunun yapısı $f(y_{ij})$ marjinal yoğunluğuna dayanmaktadır. Varsayalım m küme ve her bir küme için birim sayısı n_i ($i = 1, 2, \dots, m$) olsun. Sonra, m küme için olabilirlik fonksiyonu elde edilir;

$$\begin{aligned} L(\theta) &= \prod_{i=1}^m f(y_{ij}) \\ &= \prod_{i=1}^m \int \prod_{j=1}^{n_i} f(y_{ij}|u_i) f(u_i) du_i \end{aligned} \quad (3.41)$$

ve Laplace yaklaşımı, eşitlik (3.41)’deki integral terimini tekrar yazarak integralin yaklaşık değerini bulmak için kullanılır;

$$\int \prod_{j=1}^{n_i} f(y_{ij}|u_i) f(u_i) du_i = \int f(y|u) f(u) du$$

$$= \int e^{\log f(y|u)f(u)} du = \int e^{g(u)} du. \quad (3.42)$$

Burada $g(u) = \log f(y|u)f(u)$ eşitliği elde edilmektedir. Yaklaşımında amaç, gerekli ve yeterli koşullar $g'(u) = 0$ ve $g''(\hat{u}) < 0$ yerine getirilerek $g(u)$ fonksiyonunu maksimize edecek şekilde $\hat{u}$ 'yı seçmektir. $g(u)$ için $\hat{u}$ etrafındaki ikinci dereceden Taylor açılımı;

$$\begin{aligned} g(u) &\approx \tilde{g}(u) = g(\hat{u}) + (u - \hat{u})g'(\hat{u}) + \frac{1}{2}(u - \hat{u})^2 g''(\hat{u}) \\ &= g(\hat{u}) - \frac{1}{2}(u - \hat{u})^2 (-g''(\hat{u})) \end{aligned} \quad (3.43)$$

şeklindedir. $e^{\tilde{g}(u)}$ 'nin normal yoğunlukla (μ_L, σ_L^2) orantılı olduğu görülmektedir; burada $\mu_L = \hat{u}$ ve $\sigma_L^2 = -\frac{1}{g''(\hat{u})}$ eşitlikleri elde edilir. Ardından, $L(\theta)$ olabilirlik fonksiyonu için Laplace yaklaşımı;

$$\begin{aligned} L(\theta) &= \int e^{g(u)} du \approx \int e^{\tilde{g}(u)} du \\ &= \exp\left(g(\hat{u})\right) \int \exp\left(-\frac{1}{2\sigma_L^2}(u - \mu_L)^2\right) du = \exp\left(g(\hat{u})\sqrt{2\pi\sigma_L^2}\right) \end{aligned} \quad (3.44)$$

şeklindedir. Ayrıca,

$$(u|y) = \frac{f(y|u)f(u)}{f(y)} \propto \exp(g(u)) \approx \text{const} \exp\left(-\frac{1}{2\sigma_L^2}(u - \mu_L)^2\right)$$

şeklinde gösterilmektedir ve burada “const”, normalizasyon sabiti anlamına gelmektedir. Dolayısıyla, $U|Y = y \approx N(\mu_L, \sigma_L^2)$ elde edilir.

Laplace yaklaşımı, en hızlı yaklaşım olarak bilinmektedir. PQL ve Adaptive Gauss yöntemlerinin matematik temelini oluşturmaktadır. Laplace yaklaşımı ve Gauss-Hermite yöntemi sadece normal dağılıma sahip rastgele etkiler için dikkate alınmaktadır. Bu yaklaşım, daha büyük kümelerle sahip n_i 'ler için daha iyi performans göstermektedir. Ayrıca, bu yaklaşımın doğruluğu, daha yüksek dereceli Taylor serisi kullanıldığında artar. Ancak, rastgele etkilerin varyansı büyükse bu yaklaşımın doğruluğu azalır.

3.3.2.2 Gauss-Hermite Quadrature ve Adaptive Gauss-Hermite Quadrature yaklaşım yöntemleri

GHQ, temelde Laplace yaklaşımının bir genellemesidir ve yaklaşımı sadece t_0 noktasında değerlendirmek yerine, integralin değerlendirilmesi için farklı bir dizi

nokta seçilebilir (Tuerlinckx vd., 2006). t_0 noktası, genellikle fonksiyonun etrafında yapılan bir Taylor serisi için seçilen bir referans noktasını belirtir. GHQ’da, integralin hesaplanması ve tahmin edilebilmesi için farklı bir dizi nokta seçilebilir. Bu noktalar, genellikle belirli bir dağılıma göre ağırlıklandırılan noktalardır ve fonksiyonun daha geniş bir bölgesinin hesaba katılmasını, integralin daha doğru sonuçlar vermesini sağlar. Genel olarak, yaklaşımın değerlendirildiği noktalar, yani düğümler, en uygun şekilde aralıklandırılmışsa, neredeyse kesin yaklaşımlar elde edebilir. Yaklaşım için kullanılacak kareleme sayısı literatür kaynağına bağlı olarak değişmektedir, fakat düğümlerin konumları algoritmik olarak belirlenir.

GHQ yanlılığa duyarlıdır, küme boyutları veya varyansları büyükse GHQ’nun yanlılığı artar, bu nedenle GHQ’nun bir varyasyonu olan AGHQ genellikle uygulanır. AGHQ’da, düğümlerin konumları, rastgele etkilerin en iyi tahminlerine ve sabit etkilerin ve varyans-kovaryans matrisinin tahminlerine göre seçilir (Tuerlinckx vd., 2006).

Kareleme, aşağıdaki gibi ağırlıklandırılmış toplamı içeren sayısal integralin başka bir adıdır;

$$I = \int_a^b f(x)dx \approx \sum_{i=1}^n w_i f(x_i) \quad (3.45)$$

burada “ x_i ” düğüm veya kareleme noktası, “ w_i ” ağırlıktır. Eğer eşitlik (3.45)’teki $f(x)$ normal yoğunluksa veya log-quadratic faktörü varsa, o zaman $f(x)$ w_i ’nin bir Gaussian ağırlığı ile Hermite polinomu olarak genişletilebilir.

GLMM’de her bir gözlem için olasılığın katkısı;

$$\int f(x)\phi(x)dx \quad (3.46)$$

şeklinde ifade edilebilir. Burada “ $f(x)$ ” bilinen bir fonksiyon ve “ $\phi(x)$ ” standart normal yoğunluktur. GHQ, eşitlik (3.46)’daki integrale ağırlıklı bir toplam ile yaklaşır;

$$\int f(x)\phi(x)dx \approx \sum_{l=1}^L w_l f(x_l). \quad (3.47)$$

Burada “ L ”, “ x_l ”nin kareleme noktaları (veya yaklaşımın derecesi), “ w_l ” ağırlık sabiti ve “ x_l ” L ’nci dereceden Hermite polinomu için çözüm olan bir değerlendirme noktasıdır.

Kareleme noktası x_l ve ağırlık w_l aşağıdaki gibi hesaplanabilir;

$$x_l = i^{th} \text{ zero of } H_n(x) \quad (3.48)$$

$$w_l = \frac{2^{n-1} n! \sqrt{\pi}}{n^2 [H_{n-1}(x_k)]^2} \quad (3.49)$$

burada “ $H_n(x)$ ” n dereceli Hermite polinomudur. Eşitlik (3.48), $H_n(x)$ polinomunun köklerinden i 'nci kökü temsil eder. Bu kökler, Gauss-Hermite polinomlarının sıfır yaptığı değerlerdir. $H_n(x)$ Hermite polinomlarının $x = 10$ dereceye kadar hesaplanmış hali;

$$H_0(x) = 1$$

$$H_1(x) = 2x$$

$$H_2(x) = 4x^2 - 2$$

$$H_3(x) = 8x^3 - 12x$$

$$H_4(x) = 16x^4 - 48x^2 + 12$$

$$H_5(x) = 32x^5 - 160x^3 + 120x$$

$$H_6(x) = 64x^6 - 480x^4 + 720x^2 - 120$$

$$H_7(x) = 128x^7 - 1344x^5 + 3360x^3 - 1680x$$

$$H_8(x) = 256x^8 - 3584x^6 + 13440x^4 - 13440x^2 + 1680$$

$$H_9(x) = 512x^9 - 9216x^7 + 48384x^5 - 80640x^3 - 30240x$$

$$H_{10}(x) = 1024x^{10} - 23040x^8 + 161280x^6 - 403200x^4 + 302400x^2 - 30240$$

şeklindedir (Nia, 2006).

GHQ'da x_l kareleme noktaları ve w_l ağırlıkları sabittir ve $f(x)\phi(x)$ 'ten bağımsızdır.

AGHQ'da ise x_l kareleme noktaları ve w_l ağırlıkları $f(x)\phi(x)$ 'in aralığına uyacak şekilde ayarlanır. AGHQ,

$$\begin{aligned} \int (y|u)f(u)du &= \int \frac{f(y|u)f(u)du}{\phi(u; \mu, \sigma^2)} \phi(u; \mu, \sigma^2)du \\ &= \int \frac{f(y|\sigma x + \mu)f(\sigma x + \mu)}{\phi(x)} \sigma \phi(x)dx; x = \frac{(u-\mu)}{\sigma} \end{aligned} \quad (3.50)$$

integrali tahmin etmek için kullanılan yöntemdir. AGHQ'daki “uyarlanabilir” terimi, Laplace yöntemine benzer şekilde optimum noktadaki Hessian (2. dereceden kısmi türev matrisi) kullanılarak ölçeklendirilen integrali ifade eder.

AGHQ x_l 'yi daha uygun bir şekilde seçer ve GHQ'dan çok daha az kareleme noktasına ihtiyaç duyar. Ancak ilk adımda sabit etki β ve varyans bileşeni θ 'ya bağlı

olan x_l ve w_l hesaplandığından dolayı çok fazla zaman alır. Daha sonra, her adımda parametrenin güncellenmesi gerekir. Rastgele etkinin boyutu arttığında da AGHQ'nun karmaşıklığı artacaktır. AGHQ'da rastgele etkilerin boyutu 2'den büyükse yöntem hatalıdır. Bunun nedeni, yöntemin yüksek boyutlu integralleri düşük boyutlu bazı integrallere ayırmada sınırlı olmasıdır. McCullagh (1989), doğruluğu arttırmak için 10'dan az noktaya sahip karelemenin yanlış cevaplar verdiğini, 20 noktalı karelemenin ise iyi bir yaklaşım için genellikle yeterli olduğunu belirtmiştir. İlginçtir ki, $L = 1$ için AGHQ, Laplace yaklaşımına indirgenir.

AGHQ, modelde yalnızca bir tane rastgele etki varsa veya iki tane iç içe rastgele etki için hala iyi sonuçlar veriyorsa iyi bir yaklaşım sağlayacaktır. Ancak, çapraz rastgele etki varsa veya ikiden fazla iç içe etkinin olduğu durumlarda AGHQ'nun kullanılmasıyla elde edilen sonuçlara izin verilmez.

R paketinde, AGHQ yöntemi özellikle “lme4” paketindeki “glmer” fonksiyonunun bir seçeneğidir. Eğer yalnızca bir tane rastgele etki varsa, “glmer” fonksiyonu kullanılarak AGHQ yaklaşımı 25 kareleme noktasına kadar hesaplanabilir. Öte yandan, birden fazla rastgele etki varsa, o zaman “glmer” yalnızca bir karelemeyi tahmin edebilir.

Açıklanan parametre tahmin yöntemlerinin avantajları, dezavantajları ve uygun yazılımlar Tablo 3.1’de verilmiştir.

Tablo 3.1 GLMM’de parametre tahmin yöntemlerinin avantajları, dezavantajları ve yazılımlar

Yöntem	Avantaj	Dezavantaj	Yazılım
Penalized Yarı-olabilirlik (PQL)	Hesaplama avantajı vardır. En hızlı ve en esnek yöntemdir.	En hızlı yöntem olmasına rağmen en az doğru olanıdır. Karmaşık modellerde veya yüksek boyutlu veri setlerinde yakınsama problemlerine bağlı olarak R programında AIC, BIC ve Log-Likelihood değerleri hesaplanamayabilir.	R SAS Stata

Tablo 3.1 (Devam)

Laplace	AGHQ'dan daha		
	az hesaplama		
	yüküyle tarafsız	Küçük küme boyutlarına sahip	R
	tahminler verir.	veriler için daha az doğru	SAS
	PQL'den daha iyi	tahminler sağlar.	Stata
	performans		
	gösterir.		
Gauss-Hermite Quadrature (GHQ)	Laplace		R
	yöntemine göre	Küme boyutları ve varyansları	MATLAB
	daha doğru	büyükse yanlılık artar.	SAS
	tahminler verir.		
Adaptive Gauss-Hermite Quadrature (AGHQ)		Hesaplama yükü kareleme	
		noktası sayısı ile birlikte artar	
		ve hızı azalır.	
	Laplace	Kullanımında sınırlamalar	R
	yöntemine göre	vardır. Karmaşık modellerde	MATLAB
	daha doğru	veya yüksek boyutlu veri	SAS
	tahminler verir.	setlerinde yakınsama	
	GHQ'ya göre	problemlerine bağlı olarak R	
	daha kesindir.	programında AIC, BIC ve Log-	
		Likelihood değerleri	
		hesaplanamayabilir.	

3.4 Genelleştirilmiş Linear Karma Modellerde Sonuç Çıkarımı

GLMM'de parametre tahmininden sonra veri setinden bilgi ve istatistiksel sonuçlar elde etmek için hipotez testleri ve güven aralıkları ile analiz yapılır. Veri setindeki değişkenliği en iyi şekilde tanımlayan veya açıklayan modeli seçmek amaçlanır. Sonuç çıkarımı genellikle hipotez testi, model karşılaştırması veya Bayes yaklaşımları ile yapılır. Hipotez testi, test istatistiklerini (ANOVA' da F-Testi gibi) H_0 hipotezi altında beklenen dağılımlarıyla karşılaştırarak, H_0 'ın reddedilip reddedilemeyeceğini belirlemek için p değerini (p-value) tahmin eder.

Model seçimi aday modellerin uyumlarını karşılaştırır. Modeller, iç içe geçmiş ve daha karmaşık modelleri test eden hipotez testi veya bilgi teorisi yaklaşımları olan Wald testleri (Z, χ^2 , t ve F testleri) kullanılarak seçilebilirler. Model seçiminde, olasılık oranı (LR) testleri faktörlerin anlamlılığını sağlayabilirler veya aday modellerin bir çifti arasından en iyisini seçebilirler.

Model seçimindeki temel amaç verilere en iyi uyum sağlayan modeli seçmek ve aynı zamanda bu modelin aşırı uyum (overfitting) göstermesini önlemektir. Bilgi kriterleri, model seçiminde kullanılan istatistiksel ölçütlerdir. GLMM’de, aynı bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi incelemek üzere oluşturulan ancak bağlantı fonksiyonu ya da rastgele etki gibi değişimlerle oluşturulan modeller arasındaki kıyaslama bilgi kriterlerine göre yapılır (Gbur vd., 2020; İyit vd., 2023; Stroup, 2012).

Bilgi kriterleri, genellikle bir modelin uygunluğunu ölçmek için karmaşıklık ve uygunluk arasındaki dengeyi sağlamaya çalışırlar ve çoklu karşılaştırmaları ve iç içe olmayan modellerin seçimini sağlarlar. Daha karmaşık bir model genellikle daha iyi uyum sağlayabilir, ancak aşırı uyum gösterme eğilimindedir. Bu sebeple, bilgi kriteri daha basit bir modelin tercih edilmesini ve en küçük değere sahip olan modelin en iyi uyum sağlayan model olduğunu savunur. En küçük bilgi kriteri değerine sahip modeller en uygun modellerdir. Bu kriterler arasında Akaike Bilgi Kriteri (Akaike Information Criterion, AIC) ve uyum ölçüsü olarak sapmayı kullanan ilgili bilgi kriterleri bulunur; daha karmaşık modelleri cezalandırmak için bir terim eklerler. Bilgi kriterleri daha iyi tahminler sağlayabilirler. Örneğin, örneklem büyüklüklerinin büyük olmadığı (AICc), veride aşırı dağılım olduğu (Quasi-AIC, QAIC) veya bir modeldeki parametre sayısı belirlenmek istendiği (Bayesian Information Criterion, BIC) durumlarda AIC’nin varyasyonları oldukça yaygın kullanılır.

Yaygın olarak kullanılan bilgi kriterleri istatistikleri Tablo 3.2’de gösterilmiştir.

Tablo 3.2 Bilgi kriterlerinin isimleri ve formülleri (İyit vd., 2023)

Bilgi Kriterleri	Formüller
Akaike Bilgi Kriteri (AIC) (Akaike, 1974)	$-2l(\hat{\theta}) + 2p$
Düzeltilmiş Akaike Bilgi Kriteri (AICc) (Cavanaugh, 1997)	$-2l(\hat{\theta}) + \frac{2pN}{N - p - 1}$
Bayesçi Bilgi Kriteri (BIC) (Schwarz, 1978)	$-2l(\hat{\theta}) + p \ln(N)$
Tutarlı Akaike Bilgi Kriteri (CAIC) (Bozdoğan, 1987)	$-2l(\hat{\theta}) + (p \ln(N) + 1)$

“ $l(\hat{\theta})$ ” log-olabilirlik istatistiği; “ p ” modeldeki parametre sayısı ve “ N ” gözlem sayısıdır. Bilgi kriterleri istatistikleri, $l(\hat{\theta})$ log-olabilirlik fonksiyonunun maksimize olduğu noktada $\hat{\theta}$ tahmin edicisinde hesaplanan değer ile bağlantılıdır.

4. BULGULAR VE TARTIŞMA

4.1 Bellbird Veri Seti

Bellbird (*Anthornis melanura*), Yeni Zelanda'ya özgü bir kuş türüdür ve çan sesine benzeyen ötüşü ile bilinir. Bu kuşlar, Yeni Zelanda'nın yerli ekosisteminde önemli bir rol oynar ve ekolojik süreçlere katkıda bulunurlar (Anderson, 2003). Bellbird'lerin popülasyon dinamikleri ve çevresel faktörlerin etkilerini incelemek, bu özel türün ekolojik davranışlarını anlamak ve koruma stratejilerini geliştirmek açısından önemlidir. Bellbird popülasyonların izlenmesi ve analiz edilmesi, ekolojik dengeyi koruma ve doğru yönetim stratejileri geliştirme açısından büyük bir gereklilik teşkil eder. Bu yüzden araştırmacılar için önemli bir çalışma alanıdır.

GLMM'ler, özellikle ekolojik ve biyolojik çalışmalarda yaygındır çünkü farklı gruplar ve zaman noktaları arasındaki değişimleri modellemeye imkan tanır. GLMM'nin uygulanmasında kullanılan en önemli araçlardan biri R programıdır. R, istatistiksel analiz için yaygın olarak kullanılan bir programlama dilidir. R'ın geniş paket kütüphanesi, karmaşık istatistiksel modellerin uygulanmasını kolaylaştırır, ayrıca veri analizi süreçlerinde esneklik ve güçlü analiz yetenekleri sunduğundan, bu çalışmanın gerektirdiği istatistiksel analizler için idealdir. R, görselleştirme konularında da güçlü araçlara sahip olduğundan, analiz sonuçlarının etkili bir şekilde sunulmasını sağlar. Bu uygulama, R programı aracılığıyla bilimsel araştırmacılara ve koruma uzmanlarına GLMM'nin Bellbirds'in yaşam alanlarını etkileyen faktörleri belirlemede ve koruma stratejileri geliştirmede yol gösterebilir. Bu veri analizinde, özellikle “nlme” ve “lme4” paketleri kullanılmıştır. LMM'ler için “nlme” paketinde bulunan “lme” fonksiyonu kullanılırken, GLMM'ler için “lme4” paketinde bulunan “lmer” fonksiyonu kullanılmıştır.

4.1.1 Veri seti ve değişkenler

Analiz için kullanılan veri seti Dawber'in (2009) çalışmasından elde edilen Bellbird popülasyon verilerinden oluşmaktadır. Bellbird veri seti, çeşitli yıllar boyunca farklı uygulamaların Bellbird popülasyonu üzerindeki etkisini içermektedir. Analiz için seçilen değişkenler şunlardır:

- **Bağımlı Değişken:**
 - **BellbirdTotal:** Bellbird sayısı

- **Bağımsız Değişkenler (Sabit Etkiler):**
 - **I(Year-2003):** Referans alınan 2003 yılı
 - **Treatment:** Uygulama alanı (True/False)
 - **I(Year-2003)*Treatment:** Yıl ve Tedavi arasındaki etkileşim etkisi
- **Rastgele Etkiler:**
 - **fTransect:** Farklı transektler (Her uygulama alanında gözlem yapılan belirli hatlar)
 - **fRep:** Farklı tekrarlar (Aynı transekt boyunca yapılan gözlemlerin farklı zamanlardaki tekrarları)
 - **fCount:** Farklı sayımlar (Belirli bir transekt boyunca belirli bir zaman noktasında gözlemlenen Bellbird sayısı)

Bellbird veri setiyle ilgili bilgiler aşağıda belirtilmiştir:

1. Farklı uygulama alanlarında, farklı transektlerde ve farklı zamanlarda gözlemlenen kuşların sayımlarını içermektedir. Sayım verilerinden oluştuğu için bağımlı değişkenlerin Poisson dağılımını takip ettiği varsayılabilir.
2. Bireysel transektler incelenmez; bunun yerine tüm alan araştırılır. Odak noktası belirli beş sayım arasında değil, tüm gün içindeki değişimdir. Benzer şekilde odak noktası, seçilen üç tekrar arasındaki değil, hafta içindeki değişimdir.
3. Her bir transekt, üç günün her birinde ölçülür ve üç günün her birinde beş sayım gerçekleştirilir.
4. Gözlemler hiyerarşik bir yapıya sahiptir: uygulama alanları, transektler ve zamanlar arasında bağımlılıklar olabilir. Bu bağımlılıkları doğru bir şekilde modellemek için rastgele etkiler kullanılır.
5. Tüm rastgele etkilerin normal dağılıma uyduğu varsayılmaktadır.

4.1.2 Analiz

İstatistiksel analiz, Bellbird veri setini kullanarak LMM ve GLMM'i uygulamayı ve bu modelleri karşılaştırmayı içermektedir. LMM ve GLMM tanımlanmıştır. LMM, bağımlı değişkenin normal dağıldığı varsayımı ile; GLMM ise bağımlı değişkenin farklı dağılım yapısına (bu veri seti için Poisson) sahip olabileceği varsayımı ile kullanılmıştır. Bellbird veri seti kullanılarak modeller tahmin edilmiştir. LMM ve

GLMM karşılaştırılmıştır. Elde edilen sonuçlar yorumlanmış ve modelin uyumu değerlendirilmiştir.

4.1.3 Model tanımlama ve uygulama

LMM R kodu

LMM R koduyla aşağıdaki gibi tanımlanmıştır:

```
lmm <- lme(BellbirdTotal ~ I(Year-2003)*Treatment, random=~1 | fCount/fRep/fTransect))
```

GLMM R kodu

GLMM R koduyla aşağıdaki gibi tanımlanmıştır:

```
glmm <- lmer(BellbirdTotal ~ I(Year-2003)*Treatment + (1 | fCount/fRep/fTransect), family = poisson)
```

“lme4” paketindeki “lmer” fonksiyonu ile Poisson dağılımına sahip verilerle çalışırken p-değerleri hesaplanabilir ve bu fonksiyonun varsayılan olarak Laplace yaklaşımı yöntemini kullanması, tahminlerin doğruluğunu artırır (Dawber, 2009).

LMM ve GLMM için R kodlarındaki değişkenler ve faktörler:

- **BellbirdTotal:** Bellbird sayısı (bağımlı değişken)
- **Year:** Veri toplama yılı (bağımsız değişken)
- **Treatment:** Uygulama alanı (bağımsız değişken)
- **fTransect:** Farklı transektler (birinci düzey rastgele etki)
- **fRep:** Farklı tekrarlar (ikinci düzey rastgele etki, fTransect içinde)
- **fCount:** Farklı sayımlar (üçüncü düzey rastgele etki, fRep içinde)

şeklindedir.

4.1.4 Sonuçlar ve yorum

Tablo 4.1’de LMM’de sabit etkilerin katsayıları, standart hataları, t ve p değerleri Tablo 4.2’de LMM’de rastgele etkilerin standart sapmaları verilmiştir.

Tablo 4.1 LMM’de sabit etkilerin katsayıları, standart hataları, t ve p değerleri

Değişken	Katsayı	Standart Hata	t Değeri	p Değeri
Sabit Terim	1.5955556	0.15299952	10.428500	0.0000
I(Year-2003)	0.0244444	0.02583565	0.946152	0.3442
Treatment	-0.5361111	0.12117995	-4.424091	0.0000
I(Year-2003) *Treatment	0.2183333	0.03653713	5.975656	0.0000

Tablo 4.2 LMM’de rastgele etkilerin standart sapmaları

Rastgele Etki	Standart Sapma
1 fTransect	0.4333816
1 fRep/fTransect	0.06749594
1 fCount/fRep/fTransect	0.2278353

Tablo 4.3’te GLMM’de sabit etkilerin katsayıları, standart hataları, Z ve p değerleri, Tablo 4.4’te GLMM’de rastgele etkilerin standart sapmaları verilmiştir.

Tablo 4.3 GLMM’de sabit etkilerin katsayıları, standart hataları, Z ve p değerleri

Değişken	Katsayı	Standart Hata	Z Değeri	p Değeri
Sabit Terim	0.43416	0.10312	4.210	0.0000255
I(Year-2003)	0.01427	0.01825	0.782	0.434050
Treatment	-0.31800	0.08803	-3.612	0.000303
I(Year-2003) *Treatment	0.12269	0.02552	4.807	0.00000153

Tablo 4.4 GLMM’de rastgele etkilerin standart sapmaları

Rastgele Etki	Standart Sapma
fTransect	0.28667
fRep/fTransect	0.038733
fCount/fRep/fTransect	0.000022361

Her iki modelde de $I(\text{Year}-2003) * \text{Treatment}$ değişkeninin Bellbird sayısı üzerindeki etkisi istatistiksel olarak anlamlıdır ($p < 0.05$). Bu, yıl ve tedavi arasındaki etkileşimin Bellbird popülasyonunu önemli ölçüde etkilediğini gösterir.

Her iki modelde de Treatment değişkeninin etkisi anlamlı bulunmuştur ($p < 0.05$). Bu, farklı uygulamaların Bellbird popülasyonu üzerinde önemli bir etkisi olduğunu gösterir.

Rastgele etkilerin standart sapmaları, farklı transektler, tekrarlar ve sayımlar arasında önemli değişimler olduğunu gösterir. Bu durum, modele rastgele etkilerin dahil edilmesinin gerekli olduğunu göstermektedir.

Tablo 4.5'te LMM ve GLMM'lere ait AIC, BIC ve log-likelihood değerleri verilmiştir.

Tablo 4.5 LMM'in ve GLMM'in Karşılaştırılması

Model	AIC	BIC	Log-Likelihood
LMM	5558.9	5586.3	-2774.4
GLMM	1667.8	1706.3	-826.9

Bellbird veri setindeki modellerin karşılaştırılması sonucunda, GLMM'nin hem AIC hem de BIC değerleri LMM'ye göre çok daha düşüktür. AIC ve BIC, modelin uyum kalitesini ve karmaşıklığını dengeleyen kriterlerdir. Daha düşük AIC ve BIC değerleri, modelin veriye daha iyi uyum sağladığını ve daha az karmaşık olduğunu gösterir. Dolayısıyla GLMM tercih edilmelidir.

GLMM'nin Log-Likelihood değeri, LMM'ye göre çok daha yüksektir. Bu, GLMM'nin veriye daha iyi uyum sağladığını gösterir. Log-Likelihood'un daha yüksek (veya mutlak değerin daha düşük) olması, modelin gözlemlenen veriyi daha iyi açıkladığını ve daha yüksek olasılıkla veriyi ürettiğini gösterir. Dolayısıyla GLMM tercih edilmelidir.

Analizler sonucunda, Year ve Treatment deęiřkenlerinin Bellbird sayısı üzerindeki etkisinin her iki modelde de istatistiksel olarak anlamlı olduęu g r lmektedir. Year deęiřkeninin etkisi, Bellbird pop lasyonunun zamanla deęiřtięini, Treatment deęiřkeninin etkisi ise uygulamaların pop lasyon üzerindeki  nemli etkilerini ortaya koymaktadır.

Rastgele etkilerin standart sapmaları, farklı transektler, tekrarlar ve sayımlar arasında  nemli deęiřimler olduęunu g stermektedir. Bu durum, g zlemler arasındaki baęımlılıkları doęru bir řekilde modellemek i in rastgele etkilerin dahil edilmesinin gerekli olduęunu vurgulamaktadır.

AIC, BIC ve Log-Likelihood deęerlerinin analizine dayanarak, GLMM'nin Bellbird veri seti i in daha uygun bir model olduęu sonucuna varılmıřtır. GLMM'nin tercih edilmesi, modelin veriye daha iyi uyum saęladığını, daha az karmařık olduęunu ve g zlemleri daha doęru řekilde a ıkladıęını g stermektedir. Bu sonu lar, Bellbird pop lasyon dinamiklerinin ve  evresel fakt rlerin etkilerinin anlařılması ve bu  zel t r n korunması i in geliřtirilecek stratejilerin belirlenmesinde  nemli bir rol oynayacaktır.

4.2 Owl Veri Seti

Owl veri seti, kulu ka baykuřu yavrularının anne ve babalarının varlıęına baęlı olarak farklı bir dilenme davranıřı sergileyip sergilemediklerini belirlemek i in kullanılmıřtır (Roulin ve Bersier, 2007). Bu veri setinde, her bir kulu ka i in yavruların davranıřlarını kaydeden ses ve video kayıtları bulunmaktadır. Veri seti i in farklı deney kořulları uygulanmıřtır.  rneęin, sabah saatlerinde yavrular ya doyurulmuř ya da a  bırakılmıřtır. Gece saatlerinde ise anne ve babaların yavrlara olan ilgisi g zlemlenmiř ve kaydedilmiřtir. Veri setindeki baęımlı deęiřken, yavruların anne veya babalarının varlıęında sergiledikleri dilenme davranıřının sayısal bir  l  s  olarak kaydedilmiřtir. Baęımsız deęiřkenler arasında ise anne veya babanın varlıęı gibi fakt rler bulunmaktadır.

GLMM'ler, g zlemlerin baęımlı olduęu ve normal daęılıma uymayan yanıtların bulunduęu ekoloji ve biyoloji gibi alanlarda yaygın olarak kullanılır (McCullagh ve

Nelder, 1989). GLMM'lerin olasılık fonksiyonları için kapalı form çözümleri bulunmadığından, bu modellerin parametre tahminleri için çeşitli olasılık temelli yaklaşımlar geliştirilmiştir.

GLMM tahmin yöntemlerinin performansına ilişkin önceki çalışmalar genellikle binom dağılımlı GLMM'ler gibi belirli model sınıflarına odaklanmış ya da klinik denemeler gibi çok yapılandırılmış veriler üzerinde gerçekleştirilmiştir (Zhang vd., 2011). Ancak, ekolojik ve biyolojik çalışmalarda, küçük ve karmaşık veri setleri yaygındır ve bu yönden tahmin yöntemlerinin performans özellikleri yeterince araştırılmamıştır. Bu alanlarda, yanıt değişkeni genellikle kesikli sayılar veya kategorik veriler gibi sürekli veya normal dağılımlı olmayan verilerden oluşur. Bağımsız gözlemler için GLM'ler kullanılabilirken, gözlemlerin bağımlı veya kümelenmiş olduğu durumlarda GLMM'ler tercih edilmelidir (Pinheiro ve Bates, 2000).

4.2.1 Veri seti ve değişkenler

Owl veri setiyle ilgili bilgiler aşağıda belirtilmiştir:

1. Toplamda 599 gözlem bulunmaktadır. Her bir gözlem, farklı bir yuvadaki gözlemleri temsil etmektedir. Yuvalar arasında gözlemler arasında bağımlılık olduğu varsayılarak, yuva etkisi rastgele bir etki olarak modele dahil edilmiştir.
2. Yirmi yedi kuluçka yuvası, ebeveynlerin geliş zamanlarını ve yavruların çağrılarını kaydetmek için ses ve video kaydedicilerle donatılmıştır.
3. Ebeveynler, tanınmaları ve beslenme oranlarının belirlenmesi için bacak bantları ile işaretlenmiştir (dişiler sol bacakta, erkekler sağ bacakta).
4. Sabah saatlerinde (09:00 civarı), yavrular ya deneyciler tarafından beslenmiş (doyurulmuş) ya da yuvada bulunan yiyecekler çıkarılmıştır (aç bırakılmış).
5. Gece saatlerinde (21:30'dan 05:30'a kadar), ebeveynlerin beslemeleri gözlemlenmiş, ebeveynlerin geliş zamanı ve kalış süreleri kaydedilmiş ve yavruların çağrı sayıları not edilmiştir.
6. Ebeveynin gelişinden önceki 30 saniye boyunca, yavruların çağrı sayısı (Calls) (kardeş müzakeresi) kaydedilmiştir. Bu analizde bağımlı değişken olarak kullanılmıştır.

7. Aynı kuluka iindeki gzlemlerin birbirine baėlı olabileceėi (kuluka ii baėlılık) durumlarından dolayı veri setindeki baėımlı deėiřken Poisson daėılımına uymaktadır. Bu yzden GLMM tercih edilmiřtir.

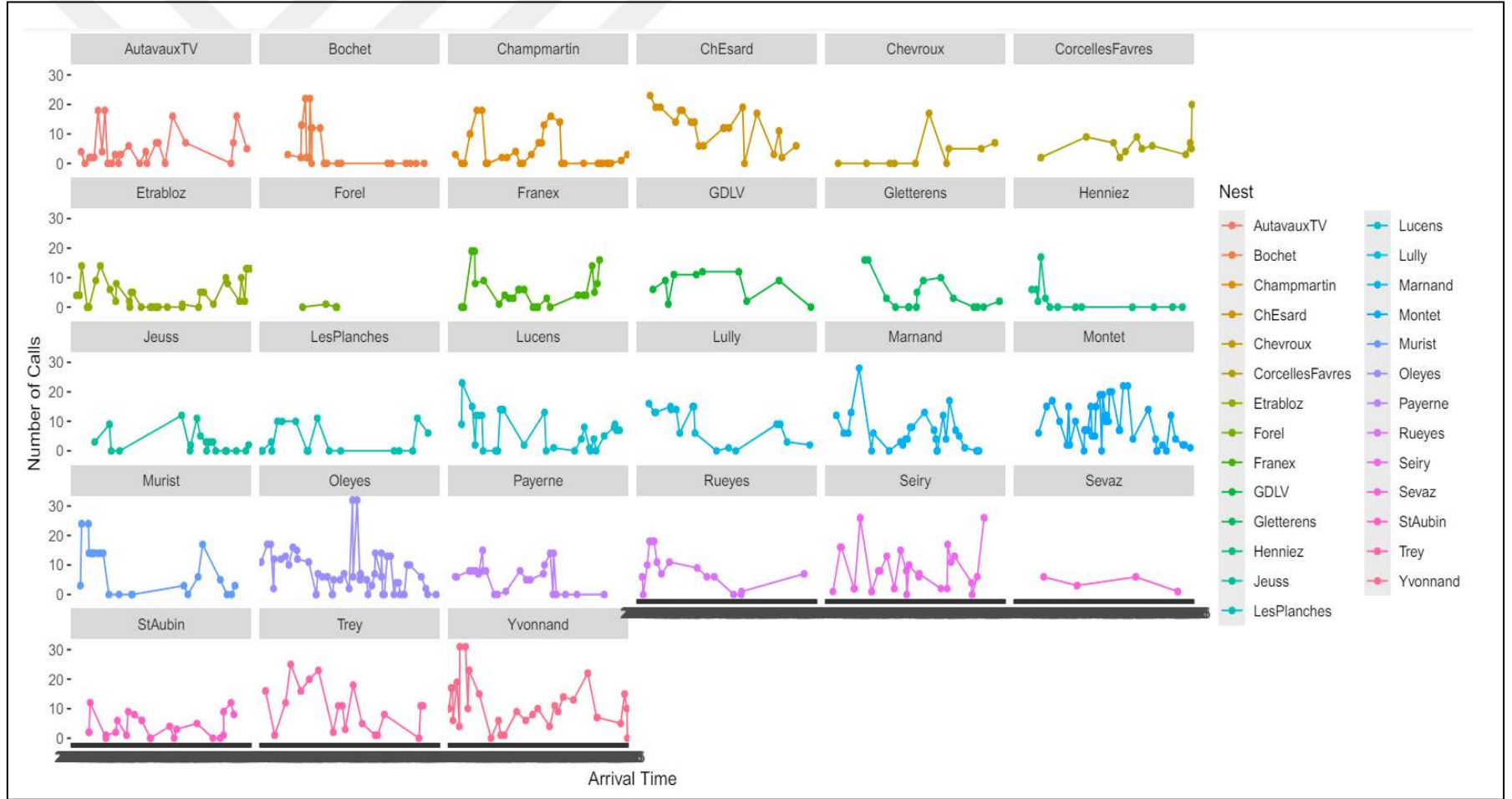
Analiz iin seilen deėiřkenler řunlardır:

- **Baėımlı Deėiřken:**
 - μ_{ij} : (Ortalama aėrı sayısı/yavruların ebeveynin geliřinden nceki 30 saniye boyunca sergiledikleri aėrı sayısı)
- **Baėımsız Deėiřkenler (Sabit Etkiler):**
 - **log(BroodSize_{ij})**: Logaritmik dnřm alınmıř kuluka boyutu
 - **FoodTreatment_{ij}**: Beslenme durumu
 - **ArrivalTime_{ij}**: Ebeveyn varıř zamanı
- **Rastgele Etkiler:**
 - **Nest_i**: Yuva etkisi

Owl veri seti iin model:

$$\log(\mu_{ij}) = \beta_0 + \beta_1 \times \log(\text{BroodSize}_{ij}) + \beta_2 \times \text{FoodTreatment}_{ij} + \beta_3 \times \text{ArrivalTime}_{ij} + \text{Nest}_i$$

řeklindedir. řekil 4.1 ile yuvaya gre gruplandırılmıř, varıř zamanına gre ebeveynin varıřından nceki 30 saniyelik aralıktaki aėrı sayılarının grafiėi verilmiřtir.



Şekil 4.1 Yuvaya göre gruplandırılmış, varış zamanına göre ebeveynin varışından önceki 30 saniyelik aralıktaki çağrı sayılarının grafiği

4.2.2 Analiz

Owl verisiyle, kuluçka baykuşu yavrularının beslenme davranışlarının incelemek için Jacob (2018) tarafından toplanan veri seti kullanılarak GLMM oluşturmak ve Laplace yaklaşımı, PQL, AGHQ tahmin yöntemlerini karşılaştırmak amaçlanmıştır.

GLMM tahmin yöntemleri ile elde edilen sonuçlar yanlılık ve hassasiyet açısından karşılaştırılmıştır. Farklı tahmin yöntemlerinin GLMM performansı değerlendirilerek, hangi yöntemin daha doğru ve güvenilir sonuçlar verdiği belirlenmek istenmiştir.

R programında GLMM tahmininde kullanılan tahmin yöntemlerinin paketleri ve fonksiyonları aşağıda belirtilmiştir:

- **Laplace Yaklaşımı:** “lme4” paketindeki “glmmTMB” fonksiyonu ile uygulanmıştır.
- **PQL:** “MASS” paketindeki “glmmPQL” fonksiyonu ile uygulanmıştır.
- **AGHQ:** “lme4” paketindeki “glmer” fonksiyonu ile 25 kareleme noktası kullanılarak uygulanmıştır.

4.2.3 Model tanımlama ve uygulama

Laplace yaklaşımı için R kodu

Laplace yaklaşımı R koduyla aşağıdaki gibi tanımlanmıştır:

```
owl_glmer_LA <- glmmTMB(Calls ~ offset(BroodSize) + ArrivalTime+  
FoodTreatment + (1|Nest),data = owl_dat, family = poisson)
```

PQL yöntemi için R kodu

PQL yöntemi R koduyla aşağıdaki gibi tanımlanmıştır:

```
owl_glmmPQL<- glmmPQL(Calls ~ offset(BroodSize)+ ArrivalTime+  
FoodTreatment, random = ~1|Nest, data = owl_dat, family = poisson,  
verbose=FALSE)
```

AGHQ yöntemi için R kodu

AGHQ yöntemi R koduyla aşağıdaki gibi tanımlanmıştır:

```
owl_glmmer_AGHQ<- glmer(Calls ~ offset(BroodSize) +
ArrivalTime+FoodTreatment + (1|Nest), data = owl_dat, family = poisson,nAGQ =
25)
```

AGHQ yönteminde kullanılan 25 kareleme noktası, hesaplamanın daha hassas olmasını sağlar çünkü daha fazla noktadan veri toplandıkça integral tahmininin doğruluğu artar. Bu sebeple, daha yüksek kareleme noktası sayısı genellikle daha doğru sonuçlar verir, fakat hesaplama süresi ve kaynakları açısından daha maliyetli olabilir (McCulloch ve Searle, 2001; Raudenbush ve Bryk, 2002).

R’da “offset” terimi, modelde sabit bir katsayı olarak kullanılan ve tahmin edilen etkileri diğer değişkenlere göre ayarlamak için kullanılır. Bu modelde “BroodSize” değişkeni sabit bir katsayı olarak kullanılır ve “Calls” değişkenine etkisi bu sabitlikle ilişkilendirilir. Özellikle Poisson regresyon modellerinde, gözlem birimlerinin farklı boyutları veya maruz kalma süreleri gibi faktörlerin dikkate alınmasına yardımcı olabilir.

4.2.4 Sonuçlar ve yorum

Laplace yaklaşımı için R çıktısı Tablo 4.6’da verilmiştir.

Tablo 4.6 Laplace yaklaşımı için R çıktısı

```
> owl_glmmer_LA <- glmmTMB(Calls ~ offset(BroodSize) + ArrivalTime+ FoodTreatment + (1|Nest),data = owl_dat, family = poisson)
> summary(owl_glmmer_LA)
Family: poisson ( log )
Formula: Calls ~ offset(BroodSize) + ArrivalTime + FoodTreatment + (1 | Nest)
Data: owl_dat

          AIC      BIC    logLik deviance df.resid
 3912.8    5314.9   -1637.4   3274.8      280

Random effects:
Conditional model:
 Groups Name      Variance Std.Dev.
 Nest   (Intercept) 1.281    1.132
Number of obs: 599, groups: Nest, 27

Conditional model:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.548e+00  4.008e-01  -3.862 0.000112 ***
ArrivalTime21.73 -1.838e-05  4.472e-01   0.000 0.999967
ArrivalTime21.76  7.983e-02  4.539e-01   0.176 0.860397
ArrivalTime21.78 -1.619e+01  1.643e+03  -0.010 0.992139

      :
ArrivalTime25.61 -2.508e+00  1.061e+00  -2.363 0.018141 *
ArrivalTime25.63 -3.593e-01  4.156e-01  -0.865 0.387305
ArrivalTime25.68 -2.226e+00  6.694e-01  -3.326 0.000882 ***
ArrivalTime25.7  -1.002e+00  6.704e-01  -1.494 0.135130
ArrivalTime25.71 -1.618e+01  1.549e+03  -0.010 0.991668
[ reached getOption("max.print") -- omitted 118 rows ]
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

PQL yöntemi için R çıktısı Tablo 4.7’de verilmiştir.

Tablo 4.7 PQL yöntemi için R çıktısı

```
> owl_glmPQL<- glmPQL(Calls ~ offset(BroodSize)+ ArrivalTime+ FoodTreatment, random = ~1|Nest, data = owl_dat, family
= poisson, verbose=FALSE)
> summary(owl_glmPQL)
Linear mixed-effects model fit by maximum likelihood
Data: owl_dat
AIC BIC logLik
NA NA NA

Random effects:
Formula: ~1 | Nest
(Intercept) Residual
StdDev: 1.076458 1.554173

Variance function:
Structure: fixed weights
Formula: ~invwt
Fixed effects: Calls ~ offset(BroodSize) + ArrivalTime + FoodTreatment
Value Std.Error DF t-value p-value
(Intercept) -2.304820 0.8 256 -2.8593805 0.0046
ArrivalTime21.73 0.000000 1.0 256 0.0000000 1.0000
ArrivalTime21.76 0.171599 1.0 256 0.1671608 0.8674
ArrivalTime21.78 -27.620684 707090.2 256 -0.0000391 1.0000
ArrivalTime21.81 -27.620684 707090.2 256 -0.0000391 1.0000
: : : : : :
ArrivalTime25.7 -0.876922 1.5 256 -0.5778038 0.5639
ArrivalTime25.71 -27.812136 707090.2 256 -0.0000393 1.0000
[ reached 'max' / getOption("max.print") -- omitted 117 rows ]
Correlation:
(Intr) AT21.73 AT21.76 AT21.78 AT21.8 AT21.91 AT21.93 AT21.95 AT21.96 AT22.01 AT22.05 AT22.06
ArrivalTime21.73 -0.628
ArrivalTime21.76 -0.669 0.493
ArrivalTime21.78 0.000 0.000 0.000
AT22.08 AT22.13 AT22.15 AT22.18 AT22.23 AT22.25 AT22.26 AT22.28 ArT22.3 AT22.31 AT22.35 AT22.36
ArrivalTime21.73
ArrivalTime21.76
ArrivalTime21.78 AT22.38 ArT22.4 AT22.43 AT22.45 AT22.46 AT22.48 ArT22.5 AT22.51 AT22.53 AT22.55 AT22.56 AT22.58
ArrivalTime21.73
ArrivalTime21.76
ArrivalTime21.78 ArT22.6 AT22.61 AT22.62 AT22.63 AT22.65 AT22.66 AT22.68 ArT22.7 AT22.71 AT22.73 AT22.75 AT22.76
:
AT28.38 ArT28.4 AT28.43 AT28.45 AT28.5 ArT28.7 AT28.78 ArT28.8 AT28.81 AT28.82 AT28.86 ArT28.9
ArrivalTime21.73
ArrivalTime21.76
ArrivalTime21.78 AT28.98 AT29.1 AT29.21 AT29.23
ArrivalTime21.73
ArrivalTime21.76
ArrivalTime21.78
[ reached getOption("max.print") -- omitted 313 rows ]

Standardized Within-Group Residuals:
Min Q1 Med Q3 Max
-2.0854927 -0.6499013 0.0000000 0.3916097 4.5993191

Number of Observations: 599
Number of Groups: 27
```

Tahmin yöntemlerinin karşılaştırılması Tablo 4.8’de verilmiştir.

Tablo 4.8 Tahmin yöntemlerinin karşılaştırılması

Tahmin Yöntemi	AIC	BIC	Log-Likelihood
Laplace	3912.8	5314.9	-1637.4
PQL	Hesaplanamadı	Hesaplanamadı	Hesaplanamadı
AGHQ	Hesaplanamadı	Hesaplanamadı	Hesaplanamadı

Laplace yaklaşımı, GLMM tahminlerinde R’da varsayılan olarak kullanılan yöntemdir. Laplace yöntemi ile hesaplanan AIC, BIC ve Log-Likelihood değerleri, modelin veri seti ile iyi bir uyum sağladığını ve modelin parametrelerinin tutarlı bir şekilde tahmin edildiğini göstermektedir.

PQL, yalnızca birinci ve ikinci dereceden koşullu momentleri içeren bir yarı-olabilirlik fonksiyonunu tahmin eder. (Molenberghs ve Verbeke, 2005). PQL yöntemi ile ilgili olarak, R programında AIC, BIC ve Log-Likelihood değerleri hesaplanmamaktadır (Dawber, 2009). Bu durum, PQL yönteminin bazı durumlarda asimptotik olarak yanlış tahminler üretebileceğini ve diğer yöntemlerle karşılaştırıldığında daha az güvenilir olabileceğini göstermektedir (Dawber, 2009; McCulloch ve Searle, 2001).

AGHQ yöntemi, yüksek doğruluk sağlayan ancak hesaplama açısından zor bir yaklaşımdır. Ancak, bu yöntem bazen yakınsama sorunları yaşayabilir. Özellikle karmaşık modellerde veya yüksek boyutlu veri setlerinde bu sorunlar daha sık ortaya çıkar. AGHQ yöntemi çözüm bulmada zorlanabilir ve bu durum daha fazla hesaplama kaynağı ve zaman gerektirir. Bu yöntem R’da tam olarak geliştirilmemiştir (Dawber, 2009). AGHQ yöntemi ile AIC, BIC ve Log-Likelihood değerleri hesaplanamamıştır. AGHQ yönteminin R’da tam olarak geliştirilmemiş olması, bazı sınırlamalara yol açabilir ve sonuçların elde edilmesini zorlaştırabilir (Ju vd., 2020).

Sonuç olarak, Laplace yöntemi için AIC, BIC ve Log-Likelihood değerleri sağlanmış ve hesaplanabilmiştir. PQL ve AGHQ yöntemlerinde bazı veri setlerinde yakınsama sorunlarına bağlı olarak R’da AIC, BIC ve Log-Likelihood değerlerinin hesaplanamadığı hem yukarıdaki paragraflarda hem de Tablo 3.1’in dezavantaj sütununda belirtilmiştir. Owl veri seti için de yakınsama problemleri nedeniyle R’da PQL ve AGHQ yöntemleri için AIC, BIC ve Log-Likelihood değerleri hesaplanamamıştır. Böylece, GLMM’de R programı kullanılırken AIC, BIC ve Log-Likelihood değerlerinin veri setindeki yakınsama durumlarına bağlı olarak hesaplanamama dezavantajının ortaya çıkabileceği gösterilmiştir. Bu durumda, Owl veri seti için GLMM’deki tahmin yöntemleri arasından Laplace yöntemi daha

güvenilir ve uygun bir yöntem olarak öne çıkmaktadır. PQL ve AGHQ yöntemleri ise bu veri seti için kullanışlı olmadığını göstermiştir. Bu nedenle, Owl veri seti için GLMM modellemesinde Laplace yaklaşımını kullanmak daha iyi bir seçim olacaktır.

Yapılan analizler, ebeveynlerin varış zamanının yavruların ortalama çağrı sayısı üzerinde önemli bir etkisi olduğunu göstermiştir. Ebeveyn varış süresinin uzamasıyla birlikte dilenme sayısının arttığı belirlenmiştir. GLMM analizleri, ortalama çağrı sayısının çevresel faktörlere bağlı olarak nasıl değiştiğini ve bu değişimlerin belirli dağılımlara nasıl uyduğunu ortaya koymuştur. Model sonuçları, bağımlı değişken olan ortalama çağrı sayısının, bağımsız değişkenler olan çevresel faktörler ve deney koşullarına bağlı olarak anlamlı farklılıklar gösterdiğini belirtmektedir. GLMM'ler, verinin karmaşıklığını ve tekrarlayan yapısını dikkate alarak daha doğru tahminler sağlamıştır. Bu yöntemler, kuluçka baykuşu yavrularının davranışlarının daha iyi anlaşılmasına ve bu türün ekolojik gereksinimlerinin daha ayrıntılı bir şekilde belirlenmesine olanak tanımıştır.

5. SONUÇLAR VE ÖNERİLER

Çalışmamız, GLMM’lerde parametre tahmini ve sonuç çıkarımı üzerine odaklanmıştır.

İlk bölümde, lineer modellerden başlayarak, GLM’lerin yapısı, detaylandırılmıştır. Farklı dağılım türleri, sistematik bileşenler ve bağlantı fonksiyonları ayrıntılı olarak incelenmiştir. Özellikle, Poisson, Binom, Normal ve Gamma gibi yaygın dağılımlar ile üstel dağılım ailesine ait olabilirlik, log-olabilirlik fonksiyonları ve bu fonksiyonların özellikleri kapsamlı bir şekilde ele alınmıştır. İkinci bölümde, GLMM’lerin yapısı ve bu modellerde parametre tahmini yöntemleri üzerinde durulmuştur. GLMM’lerde bağımlı değişkenin koşullu dağılımı, marjinal ortalaması, varyansı, kovaryansı ve korelasyonu gibi temel özellikler derinlemesine açıklanmıştır. Ayrıca, GLMM’lerde kullanılan olabilirlik fonksiyonunun tahmini için çeşitli yöntemler incelenmiştir. Parametre tahminine ilişkin bölümde, lineerleştirme teknikleri ve integral yaklaşımlar gibi farklı tahmin yöntemleri ele alınmıştır. Pseudo yarı-olabilirlik yöntemleri, PQL, Laplace yaklaşımı, GHQ ve AGHQ yöntemleri üzerinde ayrıntılı olarak durulmuştur. Sonuç çıkarımı kısmında, genelleştirilmiş lineer karma modellerde tahmin edilen parametrelerin ve model uygunluğunun analizi ve yorumlanması yapılmıştır. Bulgular ve tartışma bölümünde ise, Bellbird ve Owl veri setleri kullanılarak yapılan uygulamalara yer verilmiştir. Her iki veri seti için değişkenlerin tanımlanması, analizler, model tanımlama ve uygulama süreçleri detaylandırılmış, elde edilen bulgular yorumlanmıştır.

GLMM’leri kullanarak yaptığımız çalışmalar, parametre tahmini açısından başarılı sonuçlar vermiştir. Modelimizi oluşturan değişkenlerin etkileri doğru bir şekilde tahmin edilmiştir. Kullandığımız GLMM yöntemi, verilerdeki heterojenliği ve grup düzeyindeki ilişkileri başarılı bir şekilde ele almıştır. Modelimiz, özellikle grup yapısının ve gözlemlerin bağımlılığı göz önüne alındığında daha doğru sonuçlar üretmiştir. Çalışmamız, GLMM kullanımının hangi koşullarda ve nasıl etkili olduğunu göstermiştir. Dolayısıyla bu çalışma, GLMM’nin birçok araştırma alanında uygulanabilirliği ve güvenilirliği üzerine önemli bir katkı sağlamaktadır.

GLMM'lerin daha geniş bir veri yelpazesinde nasıl performans gösterebileceğini anlamak için daha fazla araştırma gereklidir. Özellikle büyük ölçekli veri setlerinde modelin hesaplama ve yorumlama süreçlerinin optimize edilmesi önemlidir. GLMM tahminlerini doğrulamak ve kararlılıklarını artırmak için çeşitli yaklaşımların karşılaştırmalı analizini yapmak faydalı olabilir. GLMM'lerin karmaşık veri yapılarıyla nasıl başa çıkabileceğini ve bu analizlerin uygulamalı araştırmalar üzerindeki etkisini vurgulamak önemlidir. Özellikle GLMM'lerin, sosyal bilimler ve doğa bilimlerindeki karar verme süreçlerinde nasıl kullanılabileceği değerlendirilmelidir.

Çalışmamız gelecekteki araştırmalar için önemli bir başlangıç noktası sağlayacaktır. Ayrıca, çoklu iç ilişki altında GLMM'de parametre tahmini ve sonuç çıkarım gelecek çalışmaların konusunu oluşturabilir.

KAYNAKLAR

- Agresti, A. (2015). *Foundations of linear and generalized linear models*. John Wiley & Sons.
- Aitken, M., Anderson, D., Francis, B., Hinde, J. (1989). *Statistical modelling in GLIM*. Oxford University Press.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.
- Anderson, C. J., Verkuilen, J., Johnson, T. (2010). *Applied generalized linear mixed models: Continuous and discrete data*. Springer.
- Anderson, S. H. (2003). The relative importance of birds and insects as pollinators of the New Zealand flora. *New Zealand Journal of Ecology*, 27(2), 83-94.
- Barndorff-Nielsen, O. (1976). Factorization of Likelihood Functions for Full Exponential Families. *Journal of the Royal Statistical Society. Series B (Methodological)*, 38(1), 37-44.
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H. White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution, *Trends in Ecology & Evolution*, 24(3), 127-135.
- Bolker, B. M. (2013). Generalized linear mixed models. In S. E. Frontiers in Ecology and the Environment (Ed.), *Handbook of Ecological Models* (pp. 1-20). Wiley.
- Breslow, N. E. ve Clayton, D. G. (1993), Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421), 9-25.
- Capanu, M., Gönen, M., Begg, C. B. (2013). An assessment of estimation methods for generalized linear mixed models with binary outcomes. *Statistics in Medicine*, 32(26):4550-4566.
- Carrasco, J. L. (2010). A generalized concordance correlation coefficient based on the variance components generalized linear mixed models for overdispersed count data. *Biometrics*, 66(3), 897-904.
- Cengiz, M. A. (1997). Bivariate logistic regression analysis (Technical report No. MCS-97-11). The University of Salford.
- Cengiz, M. A. (2005). Bayesian inference for bivariate generalized linear models in diagnosing renal arterial obstruction. *Statistical Methodology*, 2, 168-174.
- Cengiz, M. A. ve Percy, D. F. (2001). Mixed multivariate generalized linear models for assessing lower-limb arterial stenoses. *Statistics in Medicine*, 20(11), 1663-1679.
- Dawber, J. (2009). Generalised linear mixed models and its application in R. *Summer Research Project, Department of Mathematics and Statistics, College of Engineering*.

- Diggle, P. J., Heagerty, P., Liang, K. Y., Zeger, S. L. (2002). *The analysis of longitudinal data* (2nd ed.). Oxford: Oxford University Press.
- Dunn, P. K. ve Smyth, G. K. (2005). Series evaluation of Tweedie exponential dispersion model densities. *Statistics and Computing*, 15(4), 267-280.
- Dunteman, G. H. ve Ho, M. R. (2006). *An introduction to generalized linear models*. Sage Publications, 72, USA.
- Dobson, A. J. (2002). *An introduction to generalized linear models* (2nd ed.). Chapman & Hall.
- Fisher, R. A. (1935). *The design of experiments*. Oliver and Boyd.
- Fitzmaurice, G. M., Laird, N. M., Ware, J. H. (2011). *Applied longitudinal analysis* (2nd ed.). John Wiley & Sons.
- Fox, J. (2015). *Applied regression analysis and generalized linear models*. Sage Publications.
- Gbur, E. E., Stroup, W. W., McCarter, K. S., Durham, S., Young, L. J., Christman, M., West, M., Kramer, M. (2012). *Analysis of generalized linear mixed models in the agricultural and natural resources sciences*. American Society of Agronomy, Soil Science Society of America, Crop Science Society of America.
- Gbur, E. E., Stroup, W. W., McCarter, K. S., Durham, S., Young, L. J., Christman, M., West, M., Kramer, M. (2020). *Analysis of generalized linear mixed models in the agricultural and natural resources sciences*. John Wiley & Sons.
- Gill, J. ve Torres, M. (2019). *Generalized linear models: A unified approach*. SAGE Publications.
- Gould, W., Pitblado, J., Sribney, W. (2006). *Maximum likelihood estimation with Stata*. Stata Press.
- Gueorguieva, R. (2001). A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. *Statistical Modelling*, 1(3), 177- 193.
- Handayani, D., Notodiputro, K.A., Sadik, K., Kurnia, A. (2017). A comparative study of approximation methods for maximum likelihood estimation in generalized linear mixed models (GLMM). *AIP Conference Proceedings*, 1827 (1), 020033.
- Hardin, J. W. ve Hilbe, J. M. (2007). *Generalized linear models and extensions*. Stata press.
- Henderson, C.R., Kempthorne, O., Searle, S. R., von Krosigk, C.N. (1959). Estimation of environmental and genetic trends from records subject to culling. *Biometrics*, 15, 192-218.
- Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press.
- İyit, N. ve Genç, A. (2009). Constitution of linear mixed models (LMMs) in the

- analysis of correlated data: random intercept model (RIM) for repeated measurements data, *Journal of Modern Mathematics and Statistics*, 3(3), 60-68.
- İyit, N. ve Genç, A. (2011). Constitution of random intercept and slope model (RISM) as a special case of linear mixed models (LMMs) for repeated measurements data. *Applied Mathematics and Computation*, 218(3), 827-831.
- İyit, N., Yonar, H., Genç, A. (2016). Generalized Linear Models for European Union Countries Energy Data. *Acta Physica Polonica A*, 130, 397-400.
- İyit, N. (2021). An application of generalized linear model approach on econometric studies. *Research & Reviews in Science and Mathematics-II*. Publisher.
- İyit, N., Sarı, E., Sevim, F. (2023). Modeling COVID-19 Binary Data in the Aspect of Neoplasms as a Potential Indicator of Cancer by Logit and Probit Regression Models. *International Journal of Advanced Natural Sciences and Engineering Researches*, 7(4), 400–407.
- İyit, N., Sevim, F., Kahraman, Ü. M. (2023). Investigating the impact of CO2 emissions on the COVID-19 pandemic by generalized linear mixed model approach with inverse Gaussian and gamma distributions, *Open Chemistry*, 21(1).
- Jiang, J. ve Nguyen, T. (2007). *Linear and generalized linear mixed models and their applications*. Springer.
- Ju, K., Lin, L., Chu, H., Cheng, L. L., Xu, C. (2020). Laplace approximation penalized quasi-likelihood, and adaptive Gauss–Hermite quadrature for generalized linear mixed models: Towards meta-analysis of binary outcome with sparse data. *BMC Medical Research Methodology*, 20, 152.
- Kim, S.-H., Bao, Y., Horan, E., Kim, M., Cohen, A. S. (2015). Gauss–Hermite quadrature in marginal maximum likelihood estimation of item parameters, In: *Quantitative psychology research*, Springer, 43-58.
- Lee, Y. ve Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(4), 619-678.
- Lewis, S. L. (1998). Analysis of designed experiments using generalized linear model (Doctoral dissertation). Arizona State University.
- Lindsey, J. K. (1997). *Applying generalized linear models*. Springer.
- Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D. (2005). *SAS system for mixed models*. SAS Institute.
- Liu, X. (2016). *Methods and applications of longitudinal data analysis*. Academic Press.
- McCullagh, P. ve Nelder, J.A. (1989). *Generalized linear models* (2nd ed.). Chapman & Hall.
- McCulloch, C. E. ve Searle, S. R. (2001). *Generalized linear and mixed effects*. Open

Journal of Psychiatry, 5(1).

- Mcgilchrist, C. A. (1994). Estimation in generalized mixed models. *Journal of the Royal Statistical Society. Series B*, 56(1), 61-69.
- Mihoc, I. ve Fătu, C. (2020). About some properties of the exponential families and Fisher information. *Acta Mathematica Universitatis Comenianae*, 19, 221-234.
- Montgomery, D. C., Peck, E. A., Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.
- Myers, R. H., Montgomery, D. C., Vining, G. G. (2012). *Generalized Linear Models with Applications in Engineering and the Sciences*. John Wiley & Sons.
- Naylor, J. C. ve Smith, A. F. M. (1982). Application of a method for the efficient computation of posterior distributions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 31(3), 214-225.
- Nia, V. P. (2006). Gauss-Hermite Quadrature: Numerical or Statistical Method? Web Sitesi: <https://datawisdom.ca/paper/2006-GHQ.pdf> .
- Nelder, J. A. ve Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A*, 135(3), 370-384.
- Olsson, U. (2002). Generalized linear models. *An applied approach*. Studentlitteratur.
- Öztürk, F. (2011). *Olasılık ve istatistiğe giriş I: Ders notları*. Gazi Kitabevi.
- Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer.
- Raudenbush, S. W. ve Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods (2nd ed.)*. Sage Publications.
- Roulin, A. ve Bersier, L.-F. (2007). Nestling barn owls beg more intensely in the presence of their mother than in the presence of their father. *Animal Behaviour*, 74(4), 1099-1106.
- Sas Institute Inc. (2011). *SAS/STAT 9.3 User's Guide*. Cary, NC, USA.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.
- Searle, S., Casella, G., McCulloch, C. (1992). *Variance components*. John Wiley & Sons.
- Stroud, A. H. ve Secrest, D. (1966). *Gaussian quadrature formulas*. Prentice-Hall.
- Stroup, W. W. (2012). *Generalized linear mixed models: modern concepts, methods and applications*. CRC Press.
- Stroup, W. ve Claassen, E. (2020). Pseudo-Likelihood or Quadrature? What We Thought We Knew, What We Think We Know, and What We Are Still Trying

- to Figure Out. *Journal of Agricultural, Biological and Environmental Statistics*, 25(4), 639–656.
- Tuerlinckx, F., Rijmen, F., Verbeke, G., De Boeck, P. (2006). Statistical Inference in Generalized Linear Mixed Models: A Review. *British Journal of Mathematical and Statistical Psychology*, 59(2), 225–255.
- Urgancı Tekin, K., Mestav, B., İyit, N. (2021). Robust Logistic Modelling for Datasets with Unusual Points. *Journal of New Theory*, (36), 49-63.
- Uusipaikka, E. (2009). *Confidence intervals in generalized regressions models*. CRC Press.
- Vangeneugden, T., Molenberghs, G., Laenen, A., Geys, H., Beunckens, C., Sotto, C. (2010). Marginal correlation in longitudinal binary data based on generalized linear mixed models. *Communications in Statistics—Theory and Methods*, 39(19), 3540-3557.
- Wolfinger, R. ve O’Connell, M. (1993). Generalized linear mixed models: A pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, 48, 233-243.
- Yonar, H. ve İyit, N. (2018). Modeling The Causality Relationships Between Gdp/Gni and Electricity Consumption According to Income Levels of Countries By Generalized Estimating Equations. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*(39), 191-200.
- Yonar, H. (2019). *Ülkelerin ekonomik büyüme ve risk değerlendirmelerinin genelleştirilmiş lineer modeller ile incelenmesi*. Selçuk Üniversitesi Fen Bilimleri Enstitüsü Doktora Tezi.
- Yonar, H. ve İyit, N. (2021). Some Generalized Estimating Equations Models Based on Causality Tests for Investigation of The Economic Growth of The Country Groups. *Foundations of Computing and Decision Sciences*, 46(3), 297-315.
- Zhang, D., Lin, X., Raz, J., Sowers, M. (2011). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*, 96(456), 906-914.
- Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., Smith, G. M. (2009). *Mixed Effects Models and Extensions in Ecology with R*. Springer.

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyad, Ad	KURT, Şida Seçkin
Web sayfası (ResearchGate, Academia, vs.)	ResearchGate: https://www.researchgate.net/profile/Sida-Kurt OrcID: 0000-0003-2880-2710

Eğitim Bilgileri

Derece	Kurum	Mezuniyet Yılı
Yüksek Lisans	Dicle Üniversitesi Fen Bilimleri Enstitüsü İstatistik Bölümü	2024
Lisans	Dokuz Eylül Üniversitesi Fen Fakültesi İstatistik Bölümü	2014
Lise	Özel Amid Lisesi	2009

İş Deneyimi

Dönem (Yıl)	Şirket, Kurum	Görev
-		

Yabancı Dil

İngilizce - Orta Düzey

Yayınlar

- Şida Seçkin KURT, Seçil YALAZ “The Effect of Smoking on Blood Pressure, Diabetes, Cholesterol, Body Mass Index”, 2nd Silk Road International Scientific Research Congress, 1369-1376, 26-27 September 2023, Iğdır/Türkiye (Sözlü Tam Metin).
- Şida Seçkin KURT, Özge KURAN “Application of Generalized Linear Mixed Models in R: A Case Study with Bellbirds”, V. Uluslararası Uygulamalı İstatistik Kongresi (UYIK-2024), 168-174, 21-23 May 2024, İstanbul/Türkiye (Sözlü Tam Metin).
- Özge KURAN, Şida Seçkin KURT “Genelleştirilmiş Karma Modellerde Parametre Tahmini ve Sonuç Çıkarım”, 21.11.2023-20.11.2024 (DÜBAP-FEN.23.029).

Özel İlgiler

Doğa Gezisi, Yüzme, Kitap Okuma, Bulmaca Çözme.

DİCLE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
TEZ BENZERLİK BİLDİRİMİ FORMU

Öğrencinin Adı, Soyadı	Şida Seçkin KURT		
Öğrenci No	22830011		
Ana Bilim Dalı	İstatistik		
Program Türü	Proje ☐	Yüksek Lisans ☒	Doktora ☐
Tez Danışmanı (Ünvanı, Adı, Soyadı)	Doç. Dr. Özge KURAN		
(Varsa) II. Tez Danışmanı (Ünvanı, Adı, Soyadı)			
Tez Başlığı	Genelleştirilmiş Lineer Karma Modellerde Parametre Tahmini ve Sonuç Çıkarım		
RAPOR BİLGİLERİ			
Raporlama Aşaması	Tez Savunma Sınavı Sonrası		
Sayfa Sayısı	70		
Raporlama Tarihi	23/07/2024		
Benzerlik Oranı (%)	14		

Yukarıda bilgileri verilen tez çalışmamın toplam 70 sayfalık kısmına ilişkin, 23/07/2024 tarihinde tez danışmanım tarafından Turnitin isimli intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan intihal raporuna göre, tezimin benzerlik oranı % 14 olarak tespit edilmiştir.

Uygulanan filtrelemeler:

- ☐Başlangıç Bölümleri (Kabul ve Onay sayfası, Teşekkür sayfası, Özet/Abstract) hariç
☒Kaynaklar hariç
☐Alıntılar hariç/dâhil
☐Diğer (Açıklayınız)

Tezimin benzerlik oranı, Dicle Üniversitesi Fen Bilimleri Enstitüsü İntihal Raporu Uygulama Esaslarında belirtilen üst sınır benzerlik oranını aşmamaktadır. Tez benzerlik oranı üst sınır benzerlik oranının altında olsa dahi aksinin tespit edilmesi durumunda her türlü yasal sorumluluğu kabul ettiğimi ve hukuki sonuçlarına razı olduğumu bildirir, gereğini arz ederim.

Öğrencinin Adı, Soyadı: Şida Seçkin KURT

Tarih:23/07/2024

İmza:

Danışman Adı, Soyadı: Doç. Dr. Özge KURAN

İmza:

Tarih:23/07/2024

Ana Bilim Dalı Başkanı Adı, Soyadı: Doç. Dr. Seçil YALAZ

İmza:

Tarih:23/07/2024