GENELLEŞTİRİLMİŞ LİNEER KARMA MODELLER

TUBA KOÇ

YÜKSEK LİSANS TEZİ İSTATİSTİK ANABİLİM DALI

T.C.

ONDOKUZ MAYIS ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

GENELLEŞTİRİLMİŞ LİNEER KARMA MODELLER

TUBA KOÇ

YÜKSEK LİSANS TEZİ

İSTATİSTİK ANABİLİM DALI

AKADEMİK DANIŞMAN

Doç. Dr. Mehmet Ali CENGİZ

SAMSUN – 2012

T.C.

ONDOKUZ MAYIS ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

Bu çalışma jürimiz tarafından .../.../..... tarihinde yapılan sınav ile **İSTATİSTİK** Anabilim dalında **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Başkan:

Üye:

Üye:

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.

.../..../........

Prof. Dr. Ümit SERDAR

Fen Bilimleri Enstitüsü Müdürü

GENELLEŞTİRİLMİŞ LİNEER KARMA MODELLER

ÖZET

İstatistikte, Genelleştirilmiş Lineer Karma Modeller (GLKM), lineer karma modellerin özel bir halidir. Genelleştirilmiş lineer karma modeller, lineer tahmin edicilerin sahip olduğu sabit etkilere ilave olarak rastgele etkilerin olduğu genelleştirilmiş lineer modellerin bir durumudur. Bu rastgele etkilerin genelde normal dağılıma sahip olduğu varsayılır. En çok olabilirlik yöntemi kullanılarak böyle modelleri uydurma, bu rastgele etkiler üzerinden integral alma işlemlerini içerir. Genelde bu integraller açık bir şekilde analitik formda ifade edilemezler. En çok olabilirlik yöntemi ile farklı yaklaşım yöntemleri geliştirilmiştir. Ancak bu yöntemlerin hiçbiri olası modeller ve veri setleri için iyi özelliklere sahip değildir. Bundan dolayı, Laplace, nümerik Quadrature veya MCMC gibi yöntemler bilgisayar kullanımının artmasıyla birlikte gelişmiştir ve bu ileri yöntemler uygulanabilir hale gelmişlerdir.

"Genelleştirilmiş Lineer Karma Modeller" adlı bu tez beş bölümden oluşmaktadır. İlk bölümde, lineer modeller, genelleştirilmiş lineer modeller, karma modeller ve genelleştirilmiş lineer karma modellere ilişkin literatür taraması verilmiştir. İkinci bölümde, genelleştirilmiş lineer modeller ve karma modeller ile ilgili temel bilgiler sunulmuş, üçüncü bölümde karma modellerin analizinde karşılaşılan genelleştirilmiş lineer karma modeller yapısı ve yöntemleri verilmiştir. Dördüncü bölümde, bağımlı değişkenin binary veya Poisson olduğu durumlarda parametre tahminlerinin analizi için farklı yaklaşımlar uygulanmıştır. Bu uygulamalar için " Genç Nüfusun Sorun Algılaması: Trabzon Örneği " adlı çalışma verileri kullanılmıştır. Beşinci bölümde ise yapılan tüm bu çalışmaların sonuçları verilmiştir.

Anahtar Kelimeler: Sabit etki, rastgele etki, genelleştirilmiş lineer karma modeller, SAS.

GENERALIZED LINEAR MIXED MODELS

ABSTRACT

In statistics, a generalized linear mixed model (GLMM) is a particular type of mixed model. It is an extension to the generalized linear model in which the linear predictor contains random effects in addition to the usual fixed effects. These random effects are usually assumed to have a normal distribution.

Fitting such models by maximum likelihood involves integrating over these random effects. In general, these integrals cannot be expressed in analytical form. Various approximate methods have been developed, but none has good properties for all possible models and data sets. For this reason, methods involving Laplace, numerical Quadrature or Markov Chain Monte Carlo have increased in use as increasing computing power and advances in methods have made them more practical.

The thesis entitled as "Generalized Linear Mixed Models" consists of five chapters. In the first chapter, a literature review for linear model, mixed model, generalized linear models and generalized linear mixed models are given. In the second chapter, general knowledge about mixture experiments and generalized linear models is presented. In the third chapter, the structure of the generalized linear mixed models that will be taken into account for analyzing the mixture experiments are given. In the fourth chapter, different approaches are applied for interpreting the parameters in mixture experiments and measuring the effects of the components in case the response has a Binary or a Poisson distribution with application to Trabzon Youth survey data. In the last chapter, conclusions are given.

Key Words: Fixed effect, random effect, generalized linear mixed models, SAS.

TEŞEKKÜR

Tez çalışmam boyunca benimle değerli zamanını paylaşan ve destekleyen sayın danışmanım Doç. Dr. Mehmet Ali CENGİZ' e saygı, sevgi ve teşekkürlerimi sunarım.

Her zaman yanımda olduklarını hissettiğim ve emeklerini hiçbir zaman ödeyemeyeceğim çok sevgili aileme, manevi ve bilimsel açıdan bana destek veren ve büyük bir sabırla yardımını esirgemeyen sevgili eşim Haydar KOÇ' a ve hayatıma renk katan değerli dostlarıma destekleri için çok teşekkür ederim.

İÇİNDEKİLER

SİMGE VE KISALTMALAR LİSTESİ

AIC :	Akaike Bilgi Kriteri
AICC :	Akaike Bilgi Kriterinin Küçük Örneklem Yanlı Düzeltilmiş Hali
ANOVA:	Varyans Analizi
BIC :	Bayesci Bilgi Kriteri
BLUP :	En İyi Lineer Yansız Tahmin Edici (Rastgele Etkiler İçin)
CAIC :	Tutarlı Akaike Bilgi Kriteri
GHQ :	Gauss Hermite Quadrature
GLIMMIX :	Sas Programında GLKM Süreci
GLM :	Genelleştirilmiş Lineer Modeller
GLMM :	Genelleştirilmiş Lineer Karma Modeller
HQIC :	Hannan ve Quinn Bilgi Kriteri
LMM :	Lineer Karma Modeller
MCMC :	Markov chain Monte Carlo
ML :	En Çok Olabilirlik Yöntemi
PQL :	Penalized Quasi Likelihood
REML :	Kısıtlandırılmış En Çok Olabilirlik Yöntemi
RSPL:	Residual Pseudo Likelihood
E(Y):	Y'nin Beklenen Değeri
V(Y):	Y'nin Varyansı
γ, Γ:	Gamma
$B(m, \pi)$:	Binom Dağılımı
$G(\mu, v)$:	Gamma Dağılımı
$IG(\mu, \sigma^2)$:	Invers Gaussian (Ters Normal)
$N(\mu, \sigma^2)$:	Normal Dağılım

$P(\mu)$: Poisson Dağılımı

α : Alpha

β: Beta

δ, Δ: Delta

ε : Epsilon

η: Nu

θ: Theta

λ: Lambda

μ: Mu

π, Π: Pi

σ: Sigma

ω, Ω : Omega

ϕ, Φ: Phi

ÇİZELGELER LİSTESİ

1. GİRİŞ

Verilerin analizinde yaygın olarak kullanılan istatistik değerlendirme araçlarından biri lineer regresyon modelidir. Lineer regresyon; bağımlı değişken ile bağımsız değişkeni ilişkilendiren istatistiksel bir tekniktir. Lineer regresyon modelinde bağımsız değişkenin değerinin artması veya azalması durumunda bağımlı değişkenin gerçek ortalamasının sabit bir oranda değiştiği ifade edilebilir. Lineer ve lineer olmayan regresyon modelinde normal dağılım varsayımı önemli bir rol oynamaktadır. Bağımlı değişkeninin sayı gibi kesikli değişken olduğu durumlarda normallik varsayımı sağlanmayabilir. Örneğin; "Samsun ilinde 2011 yılında ilçelerde görülen trafik kazalarının sayısı", "belirli bir kansere yakalanan insanların sayısı" gibi durumlar verilebilir. Bir başka durum olarak da bağımlı değişkenin ikili olduğu durumları düşünebiliriz. Bu durumlarda bağımlı değişken sürekli değildir. Ayrıca bağımlı değişkenin sürekli olduğu fakat normal dağılım göstermediği durumlar da olabilir. Bu tür verilerin analizine imkan sağlayacak modeller Genelleştirilmiş Lineer Modellerdir (GLM).

Genelleştirilmiş lineer model kavramı ilk olarak Nelder ve Wedderburn (1972) tarafından geliştirilmiştir. Daha sonraki yıllarda McCullagh ve Nelder (1989), Aitken ve ark. (1989), Lindleys (1997), McCulloch ve Searle (2001), Dobson (2002), Myers ve ark. (2001), Dunteman ve Ho (2006), Uusipaikka (2009) GLM teorisi hakkında çalışmaya devam etmişlerdir. Ayrıca Cengiz M.A. (1997, 2004), Cengiz ve Percy (2001) genelleştirilmiş lineer modellerin genel bir özetini vermektedir.

Genelleştirilmiş lineer modeller ile normal olmayan bağımlı değişkenler için lineer model yapısı genişletilmiştir. GLM' den önce normal olmayan veri modellemesi genellikle bağımlı değişken üzerine yapılan dönüşümlere bağlıydı. Normal olmayan bağımlı değişkenlerde, varyans değişken olup (Poisson, Gamma, Beta, v.b.) araştırmacıların çoğu varyansı sabitlemek için bağımlı değişken üzerine dönüşüm yapmaktadır. Ancak bağımlı değişkenin dönüşümüyle ilgili birçok problem öne çıkmıştır. Ayrıntılı olarak bu problemler Lewis tarafından (1998) ele alınmıştır. En çok karşılaşılan problem, normallik ve sabit varyansdır. Tek bir dönüşüm ile normallik ve sabit varyansın sağlanması zor olmaktadır. Ayrıca dönüşümler modeldeki hata terimi yapısını da etkiledikleri için dikkatle seçilmelidir.

GLM ile de veri dönüşümü uygulanmaktadır. Uygulanan bu dönüşüm " link fonksiyonu" olarak da bilinir. Fakat bu dönüşüm deterministik bir bileşene, verinin ortalamasına uygulanır (SAS Institute Inc., 2011).

Ayrıca normallik varsayımına yaklaşmak için veri dönüşümü yerine, dönüştürülmemiş bağımlı değişkenler için dağılımların sınıfı genişletilebilir. Bu geniş sınıf Üstel dağılımlar ailesi olarak adlandırılır. Genelleştirilmiş Lineer Modeller, bağımlı değişkenin Üstel dağılımlar ailesinden olması durumunda regresyon modelini uydurmaya izin vermektedir.

Son yıllarda, klasik istatistik modellerinin hemen hemen her formu, rastgele etkileri model içine yerleştirebilmek için geliştirilmiştir. Lineer modeller, lineer karma modellere; genelleştirilmiş lineer modeller, genelleştirilmiş lineer karma modellere genişletilmiştir.

Bir Karma model hem sabit hem de rastgele etkileri içeren bir modeldir. Karma model kavramını ilk veren kişi Eisenhart (1947)' dır. Eisenhart (1947), yaptığı çalışmada sabit etkileri içeren modelleri Model I, rastgele etkileri içeren modelleri Model II ve hem sabit hem de rastgele etkileri içeren modelleri ise Model III olarak tanımlamıştır.

Karma modeller bazen tekrarlı ölçümlü bir model bazen de bir hiyerarşi model olarak adlandırılabilirler. Karma modelleri, kesitsel verileri, kümelenmiş verileri veya panel verileri analiz etmede kullanabiliriz.

Karma modellerin genel istatistik yöntemlerine göre bazı avantajları vardır. İlki; karma modellerle birlikte kayıp gözlemli ve dengede olmayan verilerin analizinde daha etkili ve güvenilir sonuçlar elde edilir. İkincisi; karma modeller modelde hata varyanslarını da içerdiklerinden karma modeller ile daha güçlü testler ve daha iyi tahminler yapılabilir. Ve son olarak da; hem sabit hem de zamanla değişebilen tahmin edicilere bağlı modeller olup daha esnek bir yapıda oldukları söylenebilir (Bagiella ve ark., 2000).

Karma modeller biyoloji, genetik, tıp, psikoloji, ekonomi vb. alanlarda sıkça kullanılmaktadır.

Sürekli bir bağımlı değişken ile çeşitli bağımsız değişkenler arasındaki ilişkileri niceleştiren kümelenmiş, kesitsel veya tekrarlı ölçümlü veriler için parametrik bir lineer modele, lineer karma model denir (LKM).

LKM teorisinin literatürdeki yerinden bahsedecek olursak; Henderson (1950) karma model denklemlerinin sabit ve rastgele etkilerinin BLUP çözümlerini elde etmiştir. Anderson ve Bancroft (1952) yayınladıkları kitapta, dengeli veride varyans elemanının tahminini ayrıntılı bir şekilde açıklamışlardır ve iç içe rastgele etkili modellerde dengede olmayan veri analizini tanımlamışlardır. Henderson (1953), karma modellerde dengede olmayan veri de varyans elemanının tahminini kareler tahmin yöntemlerinden birini kullanarak, biyometri de önemli çalışmalar yapmıştır. Daha sonra 1950'lerin sonunda ve 1960'larda farklı karma modeller için başka tahmin yöntemlerini incelemiştir. Hartley ve Rao (1967) bir karma modelde denklemlerin matris gösterimleri kullanılarak en çok olabilirlik yöntemiyle varyans elemanının tek tahminleri olduğunu göstermişlerdir. Ancak kullanılan yöntemde sabit etkilerin bilinmesi varsayımıyla birlikte bulunan varyans elemanının tahminleri yanlı çıkmıştır. Townsend (1968) varyans bileşeninin minimum varyanslı kuadratik yansız tahmin edicilerini bulmuştur. Patterson ve Thomspson (1975), kısıtlandırılmış en çok olabilirlik tahminini (KEÇO) tanımlamışlardır. Bu yöntem ile dengede olmayan veriler ile genel lineer modelde (sabit etkilerin bilindiği varsayımı olmadan) varyans elemanının tahminleri elde edildi. Olabilirliğe dayalı bu yöntem, hesap kısmı oldukça zor olduğundan yavaş gelişmiştir. Searle (1971) bir rastgele faktör ile LKM de varyans elemanının tahminleri için güven aralıklarını tanımlamıştır. Khuri ve Sahai (1985), varyans elemanının tahminlerinin güven aralıkları için kapsamlı araştırma yapmışlardır. Pinheiro ve ark. (2000), Verbeke ve Molenbergs (2000), West ve ark. (2007), Zuur ve ark. (2009), Wu (2010) lineer karma modeller ile ilgili çalışmalar yapmışlardır.

Genelleştirilmiş Lineer KarmabModeller (GLKM), bir genelleştirilmiş lineer model ile lineer tahmin edicilerin rastgele etkilerinin birleştirilmesiyle elde edilmişlerdir ve uygulamada geniş bir alana sahiptirler. " Genelleştirme" kelimesi ile bağımlı değişkenin normal dağılmadığı, "Karma" kelimesiyle de modeldeki genel sabit etkilere rastgele etkilerinde eklenmesi anlatılmak istenmiştir (Işık, F., 2011).

GLKM yöntemi Breslow ve Clayton (1993), McGilchrist (1994) ve Lee ve Nelder (1996) tarafından bulunmuştur. Genelleştirilmiş lineer karma modeller çeşitli istatistik model sınıflarını içine alır. GLKM, lineer karma modellerin varsayımlardan kaynaklanan eksiklerinin giderilmesi için geliştirilmiş bir yöntemdir. Genelleştirilmiş lineer karma modeller birçok uygulama alanında kullanılır. Hastalık oranlarındaki eğilimleri tahmin etme, rastgele seçilmiş bloklarla veya rastgele seçilmiş denemeli bir tasarımdaki deneysel birimler üzerindeki enfekte bitkilerin oranının modellenmesi, sayımlardaki yüksek ozon seviyelerinin olasılığının tahmin etme, belirli bir zaman içindeki çarpık verinin modellenmesi, tüketici tercihlerinin analiz edilmesi gibi örnekler verilebilir (SAS Institute Inc., 2011).

Genelleştirilmiş lineer karma modeller ile ilgili, Mcculloch ve Searle (2001), Verbeke ve Molenbergs (2005), Littel ve ark. (2005), Jiang (2007), Sas/Stat 9.3 (2011) yapılan çalışmalardandır.

İstatistiksel bir modelin parametrelerini tahmin etme birçok istatistiksel analiz için anahtar bir adımdır. GLKM'ler için bu parametreler sabit etki ve rastgele etki parametreleridir. GLKM tahminlerini içeren birçok modern istatistiksel araç en çok olabilirlik yöntemi ile parametreleri tahmin eder. Bağımlı değişkenin normal dağıldığı birçok basit analiz için muameleler eşit örnek hacmine sahiptir ve tüm rastgele etkiler iç içe etkilerdir. Kareler toplamı farkını hesaplamaya dayalı klasik ANOVA metotları, en çok olabilirlik yaklaşımları ile aynı sonuçları verecektir. Ancak bu eşdeğerlik daha karmaşık olan LKM'ler için veya GLKM'ler için bozulur. En çok olabilirlik tahminlerini bulmak için rastgele etkilerin tüm olası değerleri üzerinde olabilirliklerin integrallenmesi gerekir. Genelde bu integraller açık bir şekilde analitik formda ifade edilemezler. Bundan dolayı, GLKM'de Laplace, nümerik quadrature veya MCMC gibi yöntemler bilgisayar kullanımının artmasıyla birlikte gelişmiştir ve bu ileri yöntemler uygulanabilir hale getirilmiştir.

2. GENEL BİLGİLER

2.1. Lineer Modeller

Gözlemlenmiş $x_1, x_2, \dots, x_p$ değerlerinin bir y değerini nasıl açıkladığı sorusuyla sıkça karşılaşılır. Bu soruya cevap olabilecek en basit model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \tag{2.1}$$

lineer modelidir. Burada $\beta_0, \beta_1, \dots, \beta_p$ 'ler modeldeki parametre katsayıları, ε modelin hata terimi, y bağımlı değişken, $x_1, x_2, \dots, x_p$ 'ler bağımsız (açıklayıcı) değişken olarak adlandırılırlar. n tane bağımsız değişken ve açıklayıcı değişken için (2.1) modeli;

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i \qquad i = 1,2, \dots n$$

şeklini alır.

$$\begin{aligned}
\mathrm{y}_1 &= \beta_0 + \mathrm{x}_{11}\beta_1 + \mathrm{x}_{12}\beta_2 + \cdots + \mathrm{x}_{1p}\beta_p + \varepsilon_1, \\
\mathrm{y}_2 &= \beta_0 + \mathrm{x}_{21}\beta_1 + \mathrm{x}_{22}\beta_2 + \cdots + \mathrm{x}_{2p}\beta_p + \varepsilon_2, \\
&\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \\
\mathrm{y}_\mathrm{n} &= \beta_0 + \mathrm{x}_{n1}\beta_1 + \mathrm{x}_{n2}\beta_2 + \cdots + \mathrm{x}_{np}\beta_p + \varepsilon_n.
\end{aligned}$$

Şeklinde açık olarak yazıldığında model

$$\underbrace{\begin{bmatrix} \mathrm{y}_1 \\ \mathrm{y}_2 \\ \vdots \\ \mathrm{y}_\mathrm{n} \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} 1 & \mathrm{x}_{11} & \cdots & \mathrm{x}_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \mathrm{x}_{n1} & \cdots & \mathrm{x}_{np} \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}}_{\varepsilon}$$

olmak üzere,

$$Y = X\beta + \varepsilon \tag{2.2}$$

formunu alır.

Burada hata terimi için $E(\varepsilon) = 0$, $\Sigma(\varepsilon) = \sigma^2 V$ varsayımları vardır. Hata teriminin beklenen değeri n elemanlı 0 vektörüdür. Varyans- kovaryans matrisindeki "V" elemanı $n \times n$ tipinde bilinen bir matristir. (2.1) ve (2.2) ile gösterilen lineer modeller için $(y, X\beta, \sigma^2 V)$ üçlü gösterimini kullanabiliriz.$\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ hata terimleri ilişkisiz olduklarında ve herbirinin varyansı σ^2 olduğunda V matrisi $n \times n$ tipinde birim matrise eşit olacaktır yani $V = I$ yazılır. Böyle bir durumda ise model $(y, X\beta, \sigma^2 I)$ üçlüsü ile gösterilir ve bu modellere eş değişkenli lineer modeller denir. Modelin parametre kümesi:

$$\Omega = \{(\beta, \sigma^2) : \beta \in \mathbb{R}^p, \sigma^2 > 0\}$$

şeklinde yazılırsa; $Y \sim N(X\beta, \sigma^2 I_n)$ olur ve bilinmeyen parametre tahmini

$$\hat{\beta} = (X'X)^{-1}X'Y \qquad \text{ve} \qquad \hat{\sigma}^2 = \frac{1}{n}(y - X\hat{\beta})'(y - X\hat{\beta})$$

şeklinde bulunur. Ayrıca parametrenin dağılımı da;

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

şeklindedir

2.2. Genelleştirilmiş Lineer Modeller

Lineer ve lineer olmayan regresyon modellerinde istatistiksel analiz yöntemleri bağımlı değişkenin normal dağıldığı varsayımına dayanmaktadır. Oysa her zaman veri normal dağılıma sahip olmayabilir. Ayrıca bağımlı değişkenin sürekli olmadığı durumlarda, bağımlı değişkenin sürekli olup da normal dağılım göstermediği durumlarda verilerin analizi için geliştirilmiş modeller Genelleştirilmiş Lineer Modellerdir.

Herhangi bir GLM üç bileşenden oluşur:

- Bağımlı değişkenin dağılımı
- Lineer tahmin edicilerin bulunduğu sistematik kısım
- Link fonksiyonu

2.2.1. Bağımlı Değişkenin Dağılımı

Genelleştirilmiş lineer modellerde bağımlı değişkenin dağılımının üstel dağılımlar ailesinden olduğu kabul edilir. Üstel dağılımlar ailesinin genel formu:

$$f(y_i/\theta_i, \phi) = exp\left\{\frac{\theta_i y_i - b(\theta_i)}{a(\phi)} + c(y_i/\ \phi)\right\} \qquad \textbf{(2.3)}$$

şeklindedir. Burada θ_i doğal konum parametresi, ϕ yayılım veya ölçek parametresidir. $a(.)$, $b(.)$ ve $c(.)$ üstel dağılımlar ailesi formunda elde edilen belirli fonksiyonlardır. $a(\phi) = \phi.w$, w bilinen bir önsel ağırlık olup gözlemden gözleme değişir. y_i bağımlı değişkeni üstel dağılımlar ailesinden bir dağılıma sahip ise y_i değişkeninin beklenen değeri ve varyansı aşağıdaki gibidir:

$$E(y_i) = \mu_i = b'(\theta_i)$$
$$Var(y_i) = {\sigma_i}^2 = b''(\theta_i)a(\phi)$$

Burada $b'(\theta_i)$ ve $b''(\theta_i)$ sırasıyla $b(\theta_i)$ birinci ve ikinci mertebeden türevleridir. Üstel dağılımlar ailesinin üyesi olan dağılımlar; Normal, Binom, Bernoulli, Negatif Binom, Poisson, Geometrik, Üstel, Gamma, Beta ve Ters Normal Dağılımdır.

y, μ ortalamalı σ^2 varyanslı normal dağılımlı olsun. Bu durumda y'nin olasılık yoğunluk fonksiyonu;

$$f(y\,/\,\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} = exp\left\{\frac{y\mu - \mu^2/2}{\sigma^2} - \frac{y^2/\sigma^2 + \log(2\pi\sigma^2)}{2}\right\}$$

olarak bulunur.

Burada $\theta = \mu$, $\phi = \sigma^2$, $a(\phi) = \phi$, $b(\theta) = \theta^2/2$ ve

$c(y/\phi) = -\{y^2/\phi + \log(2\pi\phi)\}/2$ şeklindedir.

Ayrıca $E(y) = b'(\theta) = \theta = \mu$ ve $Var(y) = b''(\theta)a(\phi) = \sigma^2$ olacaktır.

Benzer şekilde gamma dağılımı için olasılık yoğunluk fonksiyonunu üstel dağılımlar ailesi formunda yazmak istersek;

$$f(y/\mu, \phi) = exp\left\{\frac{y/\mu - (-ln\mu)}{-\phi} + \frac{\phi+1}{\phi} lny - \frac{ln\phi}{\phi} - ln\Gamma(\frac{1}{\phi})\right\}$$

şeklinde olur. Olasılık yoğunluk fonksiyonunda $\theta = 1/\mu$, $a(\phi) = -\phi$, $b(\theta) = \ln(\theta)$ ve

$$c(y/\phi) = \frac{\phi+1}{\phi} \ln y - \frac{ln\phi}{\phi} - ln\Gamma(\frac{1}{\phi})$$

dir. Bu durumda

$$E(y) = b'(\theta) = 1/\theta = \mu \text{ ve ve } Var(y) = b''(\theta)a(\phi) = \phi\mu^2$$

olur.

2.2.2. Lineer Tahmin Edici

Lineer tahmin edici modeldeki bağımsız değişkenler hakkındaki bilgiyi içeren bir niceliktir. Bir lineer tahmin edici $x_1, x_2, \dots, x_k$ bağımsız değişkenleri ile birlikte bilinmeyen β parametrelerinin bir lineer birleşimidir ve

$$\eta_i = x'\beta = \beta_0 + \sum_{i=1}^{k} \beta_i x_i \quad \textbf{(2.4)}$$

şeklinde ifade edilir.

2.2.3. Link Fonksiyonu

Link fonksiyonu $g(.)$, η_i lineer tahmin edicisi ile μ_i ortalaması arasında ilişki kuran bir fonksiyondur. Bu ilişki:

$$g(\mu_i) = \eta_i \qquad i = 1,2, \dots, n$$

şeklindedir. Link fonksiyonu terslenebilir bir fonksiyon olup $\mu_i = g^{-1}(\eta_i)$ yazılabilir. Aynı zamanda link fonksiyonu diferensiyellenebilen monoton bir fonksiyondur. İki çeşit link fonksiyonu vardır. Bunlar kanonik link ve kanonik olmayan link fonksiyonlarıdır. Eğer $g(\mu_i) = \theta_i$ ise $g(.)$ link fonksiyonuna kanonik link fonksiyonu denir. Link fonksiyonunun seçimi üstel dağılımlar ailesindeki dağılımın tipine bağlıdır. Üstel dağılımlar ailesinin her bir üyesi için birçok uygun link fonksiyonu vardır. Çizelge 2.1'de üstel dağılımlar ailesine ait yaygın olarak kullanılan bazı dağılımların özellikleri verilmiştir.

Çizelge 2.1. Üstel dağılımlar ailesi bazı yaygın tek değişkenli dağılımların özellikleri

	Normal	**Poisson**	**Binomial**	**Gamma**	**Invers Gaussian**
Dağılım ifadesi	$N(\mu,\sigma^2)$	$P(\mu)$	$B(m,\pi)$	$G(\mu,v)$	$IG(\mu,\sigma^2)$
y'nin değişim aralığı	$(-\infty,\infty)$	$(0,\infty)$	$\{0,1,\dots,n\}$	$(0,\infty)$	$(0,\infty)$
Uyum parametresi	$\phi=\sigma^2$	1	$1/m$	$\phi=v^{-1}$	$\phi=\sigma^2$
Birikimli fonksiyon	$\theta^2/2$	$\exp(\theta)$	$\log(1+e^{\theta})$	$-\log(-\theta)$	$-(-2\theta)^{1/2}$
$\boldsymbol{c(y;\phi)}$	$-\frac{1}{2}(\frac{y^2}{\phi}+\log(2\pi\theta))$	$-logy!$	$\log\binom{m}{my}$	$vlog(vy)-logy-log\Gamma(v)$	$-\frac{1}{2}(\log(2\pi\phi y^3)+\frac{1}{\phi y})$
$\boldsymbol{\mu(\theta)=E(Y;\theta)}$	θ	$\exp(\theta)$	$e^{\theta}/(1+e^{\theta})$	$-1/\theta$	$(-2\theta)^{1/2}$
Kanonik link:$\boldsymbol{\theta(\mu)}$	$\text{Özdeş}\ (Birim)$	log	$logit$	$1/\mu$	$1/\mu^2$
Varyans fonksiyonu:$\boldsymbol{V(\mu)}$	1	μ	$\mu(1-\mu)$	μ^2	μ^3

Genelleştirilmiş lineer modeller için diğer link fonksiyonları da aşağıdaki gibi verilebilir:

- Probit link $\eta_i = \Phi^{-1}\big(\mathrm{E}(\mathrm{y_i})\big)'$dir. Burada Φ, kümülatif standart normal dağılımı göstermektedir.
- Tamamlayıcı Log-log link, $\eta_i = \ln\{\ln[1-\mathrm{\mu_i}]\}$
- Güç aile link,

$$\eta_i = \begin{cases} {\mu_i}^{\lambda}, & \lambda \neq 0 \\ \ln(\mu_i), & \lambda = 0 \end{cases}$$

şeklindedir.

Genelleştirilmiş lineer modellerde, parametreleri tahmin etmek için en çok olabilirlik fonksiyonu kullanılır. Parametre tahminlerinin en çok olabilirlik tahminleri, iteratif olarak yeniden adlandırılmış en küçük kareler yöntemleriyle bulunmaktadır.

y_i bağımlı değişkeni üstel dağılımlar ailesinden bir dağılıma sahip ise log-olabilirlik fonksiyonu:

$$\ell = InL(y,\beta) = \textstyle\sum_{i=1}^{n}[y_i\theta_i - b(\theta_i)/a(\phi) + c(y_i|\phi)] \qquad \textbf{(2.5)}$$

şeklindedir. Bu ifadeyi maksimum yapan β parametrelerini bulmak için;

$$\begin{aligned}\frac{\partial\ell}{\partial\beta} &= \frac{1}{\big(a(\phi)\big)^2}\sum\left[y_i\frac{\partial\theta_i}{\partial\beta} - \frac{\partial b(\theta_i)}{\partial\theta_i}\frac{\partial\theta_i}{\partial\beta}\right] = \frac{1}{\big(a(\phi)\big)^2}\sum\left[(y_i-\mu_i)\frac{\partial\theta_i}{\partial\beta}\right] \\ &= \frac{1}{\big(a(\phi)\big)^2}\sum\left[(y_i-\mu_i)\frac{\partial\theta_i}{\partial\mu_i}\frac{\partial\mu_i}{\partial\beta}\right] \\ &= \frac{1}{\big(a(\phi)\big)^2}\sum\left[\frac{(y_i-\mu_i)}{v(\mu_i)g_\mu(\mu_i)}{x_i}'\right] \\ &= \frac{1}{\big(a(\phi)\big)^2}\cdot\sum\left[(y_i-\mu_i)w_i g_\mu(\mu_i)x_i'\right]\end{aligned}$$

elde edilir.

Burada

$\mathrm{w_i} = [v(\mu_i)g\mu^2(\mu_i)]^{-1}$, ${x_i}' = \dfrac{\partial\mu_i}{\partial\beta}$

şeklindedir.

$v(\mu_i) = \dfrac{\partial^2 b(\theta_i)}{\partial\theta_i^2}$ olup varyans fonksiyonu olarak adlandırılır. Yukarıdaki denklemi matris gösterimiyle yeniden yazarsak;

$$\frac{\partial \ell}{\partial \beta} = \frac{1}{\left(a(\phi)\right)^2} X'W\Delta(y-\mu)$$

elde edilir.

$W = \{_d w_i\} =$ ve $\Delta = \{_d g_\mu(\mu_i)\}$

Sonuç olarak en çok olabilirlik denklemleri;

$$X'W\Delta y = X'W\Delta\,\mu \qquad \textbf{(2.6)}$$

olarak bulunur. Burada $W, \Delta,$ ve μ bilinmeyen β parametreleri içermektedir. Genellikle β'nın lineer olmayan fonksiyonları (2.5) denklemleriyle çözülemez.

Bu durumda β çözümleri için yeniden ağırlıklandırılmış en küçük kareler yöntemleri uygulanır. Olabilirlik fonksiyonunu maksimum yapan çözümleri bulmak için iterasyon yöntemi olan Fisher – Scoring yöntemi kullanılabilir.

Fisher – Scoring yönteminin genel formu:

$$\theta^{m+1} = \theta^{(m)} + \mathrm{I}(\theta^{(m)})^{-1} \left.\frac{\partial \ell}{\partial \theta}\right|_{\theta=\theta^{(m)}}$$

şeklindedir. Burada (m), m.iterasyon sayısını $I(\theta)$ bilgi matrisini ve θ ise parametreler vektörünü gösterir. β parametresi için Fisher – Scoring yöntemiyle birlikte

$$\beta^{(m+1)} = \beta^{(m)} + (X'WX)^{-1}X'W\Delta(y-\mu)$$

elde edilir.

2.3. Lineer Karma Modeller

Bir lineer karma model parametrelere göre lineer olan ve hem sabit hem de rastgele etkileri içeren istatistiksel bir modeldir.

Lineer karma model, Wolfinger ve ark. (1991) tarafından aşağıdaki formda tanımlanmıştır.

$$Y = X\beta + Z\gamma + \varepsilon \qquad \textbf{(2.7)}$$

Burada;

Y : $nx1$ boyutlu bağımlı değişken vektörü

β : $px1$ boyutlu sabit etkiler vektörü

γ : $rx1$ boyutlu rastgele etkiler vektörü

ε : $nx1$ boyutlu bilinmeyen rastgele hatalar vektörü

X : nxp boyutlu modeldeki sabit etkilere ilişkin tasarı matrisi

Z : nxr boyutlu modeldeki rastgele etkilere ilişkin tasarı matrisidir.

Lineer karma model denkleminde yer alan γ rastgele etkiler vektörü ve ε rastgele hatalar vektörü için kurulan temel varsayımlar aşağıdaki gibidir.

- $E\,(\gamma\,) = 0$
- $E\,(\varepsilon\,) = 0$
- $Var\,(\gamma\,) = G$
- $Var\,(\varepsilon\,) = R$
- $Cov\,(\varepsilon,\gamma) = 0$

Böylece Y bağımlı değişkeni için beklenen değer ve varyans;

$$E\,(Y) = E(X\beta + Z\gamma + \varepsilon) = X\beta$$

$$\begin{aligned} Var(Y) = V &= Var\,(X\beta + Z\gamma + \varepsilon) \\ &= ZVar(\gamma)Z' + Var(\varepsilon) \\ &= ZGZ' + R \end{aligned}$$

Normal dağılımlı karma modellerde genel parametre tahmin yöntemleri; en çok olabilirlik yöntemi (EÇO) ve kısıtlandırılmış en çok olabilirlik yöntemidir. (KEÇO)

2.3.1. En Çok Olabilirlik Yöntemi

Lineer karma modellerde EÇO yöntemi Hartley ve Rao (1967) tarafından uygulanmıştır. y bağımlı değişkeni, $y{\sim}N(X\beta, V)$ dağılımına sahip iken y'nin olasılık yoğunluk fonksiyonu;

$$f(y) = \frac{1}{2\pi^{n/2}|V|^{1/2}} exp\left\{-\frac{1}{2}(y - x\beta)'V^{-1}(y - x\beta)\right\}$$

şeklindedir. Buradan log – olabilirlik fonksiyonu;

$$\ell(\beta,\theta) = c - \frac{1}{2}\log(|V|) - \frac{1}{2}[(y - x\beta)'V^{-1}(y - x\beta)]$$

şeklinde bulunur. Burada c bir sabittir ve varyans bileşenlerinin her biri V dahil θ'yı içerir. Sırasıyla β ve θ parametrelerine göre log – olabilirlik fonksiyonunun türevleri alınıp, sıfıra eşitlenirse;

$$\left.\frac{\partial\ell}{\partial\beta}\right|_{\beta=\hat{\beta}} = X'V^{-1}y - X'V^{-1}X\beta = 0$$

$$\left.\frac{\partial\ell}{\partial\theta_r}\right|_{\theta=\hat{\theta}} = \frac{1}{2}\{(y - x\beta)'V^{-1}\}\frac{\partial V}{\partial\theta_r}V^{-1}(y - X\beta) - tr\left(V'\frac{\partial V}{\partial\theta_r}\right) = 0, r = 1,2,\dots,q$$

olur.

β'nın $px1$ boyutlu sabit etkilerin bilinmeyen parametreleri ve $rank\,(X) = p$ olduğunu varsayalım.

θ_r : θ'nın r −bileşeni olmak üzere

elde edilen normal denklem sisteminin çözümü:

$$\hat{\beta} = \left(X'\hat{V}^{-1}X\right)^{-1}X'\hat{V}^{-1}y \qquad \text{(Searle ve ark.,1992)} \qquad \textbf{(2.8)}$$

$$tr\left(V'\frac{\partial V}{\partial \theta_r}\right) = \left\{\left(y - x\hat{\beta}\right)'V^{-1}\right\}\frac{\partial V}{\partial \theta_r}V^{-1}\left(y - X\hat{\beta}\right)$$

olarak bulunur.

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

olarak tanımlanırsa;

$$\hat{V}^{-1}\left(Y - X\hat{\beta}\right) = \hat{P}y$$

$$\hat{P}y = \hat{V}^{-1}y - \hat{V}^{-1}X(X'\hat{V}^{-1}X)^{-}X'\hat{V}^{-1}y \qquad \textbf{(2.9)}$$

olur.

Varyans denklemi ise;

$$iz(\hat{P}y) = \left(\hat{V}^{-1}\frac{\partial V}{\partial \theta}\right) = y'\hat{P}\frac{\partial V}{\partial \theta}\hat{P}y \qquad \textbf{(2.10)}$$

olarak bulunur.

$\hat{V}$ varyans bileşenleri matrisinin elemanları θ varyans bileşenlerine göre, lineer bir formda olmasına rağmen varyans denklemlerinde $\hat{V}^{-1}$ varyans bileşenleri matrisinin tersi kullanıldığı için elde edilen bu denklemlerin çözümü için analitik ifadelerin elde edilmesi mümkün değildir. Bu yüzden bu denklemler ancak nümerik olarak iteratif yöntemler yardımıyla bulunabilir.

2.3.2. Kısıtlandırılmış En Çok Olabilirlik Yöntemi

Kısıtlandırılmış en çok olabilirlik yöntemi karma modeller için Patterson ve Thompson (1971) tarafından geliştirilmiştir. KEÇO aynı zamanda artık en çok olabilirlik yöntemi olarak da bilinir.

Varyans bileşenlerinin en çok olabilirlik tahminleri genelde yanlıdır ve eğer sabit etkilerin sayısı örnek hacmi ile orantılı ise bu tür yanlılık örnek hacmi arttığında ortadan kalkmaz (Jiang, 1998).

Varyans bileşenlerinin en çok olabilirlik tahminlerinin yanlılığı kısıtlandırılmış en çok olabilirlik yöntemi ile ortadan kaldırılmaya çalışılır. KEÇO yönteminde en çok olabilirlik yönteminden farklı olarak y-nin lineer birleşimleri maksimize edilmeye çalışılır.

$Rank(X) = p$

varsayımıyla birlikte $nx(n-p)$ boyutlu bir A matrisini ele alalım.

$Rank(A) = n-p,$

$A'X = 0$ olsun.

$Z = A'y$ olacak şekilde y-nin lineer birleşimini tanımlayalım.

Burada $Z \sim N(0, A'VA)$ olur.

Z'nin ortak olasılık yoğunluk fonksiyonu;

$$f_R(Z) = \frac{1}{(2\pi)^{\frac{(n-p)}{2}}|A'VA|^{1/2}} exp\left\{-\frac{1}{2}Z'(A'VA)^{-1}Z\right\}$$

biçimindedir.

Burada alt indis R, " kısıtlandırılmış " anlamına gelir. Z-ye bağlı kısıtlandırılmış log-olabilirlik fonksiyonu;

$$\ell_R(\theta) = c - \tfrac{1}{2}log(|A'VA|) - \tfrac{1}{2}Z'(A'VA)^{-1}Z \qquad \textbf{(2.11)}$$

şeklindedir.

Burada c bir sabittir ve kısıtlandırılmış log-olabilirlik fonksiyonunun θ-ya göre türevi alınırsa;

$$\frac{\partial \ell_R}{\partial \theta_i} = \frac{1}{2}\left\{y'P\frac{\partial V}{\partial \theta_i}Py - tr\left(P\frac{\partial V}{\partial \theta_i}\right)\right\} \quad i = 1,2,\dots,q$$

Burada

$$P = A(A'VA)^{-1}A' = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$$

şeklindedir.

$$\frac{\partial \ell_R}{\partial \theta} = 0$$

eşitliğinden KEÇO tahmin edicileri bulunur. KEÇO tahmini A dönüşüm matrisiyle tanımlı olmasına rağmen, aslında KEÇO tahmini A matrisine bağlı değildir. Dolayısıyla dönüşüm matrisi değişse bile KEÇO tahmini değişmez (McCulloch ve Searle, 2001).

3. MATERYAL VE YÖNTEM

3.1. Genelleştirilmiş Lineer Karma Modeller

Bir Genelleştirilmiş Lineer Karma Model (GLKM), genelleştirilmiş lineer modellerde sabit etkileri içeren lineer tahmin edicilere rastgele etkilerin eklenmesiyle oluşturulur. Bir genelleştirilmiş lineer karma modelde rastgele etkiler $\eta = X\beta + Z\gamma$ lineer tahmin edicisinin bir parçasıdır ve verinin şartlı ortalaması ile lineer tahmin edicileri, bir lineer form ile bağlıdır.

y, $(n \times 1)$ boyutlu gözlem değerleri vektörü ve γ, $(r \times 1)$ boyutlu rastgele etkiler vektörü olmak üzere;

$$E[y|\gamma] = g^{-1}(X\beta + Z\gamma)$$

şeklinde yazılır.

Burada $g^{-1}(.)$ türevlenebilen monoton link fonksiyonunun tersi, X, $(n \times p)$ boyutlu modeldeki sabit etkilere ilişkin tasarı matrisi, Z, $(n \times r)$ boyutlu modeldeki rastgele etkilere ilişkin tasarı matrisidir ve rastgele etkiler için ortalaması sıfır, varyansı G olan normal dağılım gösterdikleri varsayımı söz konusudur.

3.1.1. Bağımlı Değişkenin Dağılımı

Modeli kurmak için öncelikle γ verildiğinde y'nin şartlı dağılımı belirlenmelidir. y bağımlı değişkeninin genel olarak şartlı bağımsız elemanlardan oluştuğu varsayımıyla birlikte, üstel dağılımlar ailesinden bir dağılım;

$$y_i|\gamma \sim bağımsız\ f_{y_i|\gamma}\ (y_i|\gamma\ )$$

$$f_{y_i|\gamma}\ (y_i|\gamma\ ) = exp\{[y_i\theta_i - b(\theta_i)]/\phi^2 - c(y_i,\phi)\} \qquad \textbf{(3.1)}$$

olur.

Genelleştirilmiş lineer modellerden bilindiği gibi y_i'nin şartlı ortalaması olan $\mu_i = \frac{\partial b(\theta_i)}{\partial \theta_i}$ özdeşliği θ_i ile ilişkilidir ve bu ortalamanın bir dönüşümü olup hem sabit hem de rastgele etkilerle model

$$E(y_i|\gamma) = \mu_i$$

$$g(\mu_i) = x_i'\beta + z_i{}'\gamma$$

şeklinde olacaktır. $g(.)$ bilinen bir link fonksiyonu, x_i' sabit etkiler için model matrisinin i.satırı ve β sabit etkili parametre vektörüdür. Bu tanımlamaya rastgele etkilerin model matrisinin i.satırı olan z_i' ve γ rastgele etkili faktörleri eklenir. Burada μ_i, γ verildiğinde y_i'nin şartlı ortalamasıdır.

3.1.2.Bağımlı Değişkenin Marjinal Dağılım Özellikleri

3.1.2.1. Bağımlı Değişkenin Ortalaması

y'nin ortalaması;

$$E[y_i] = E\big[E[y_i|\gamma\ ]\big] = E[\mu_i] = E[g^{-1}(x_i'\beta + z_i'\gamma)] \text{ dir.}$$

Özel olarak $g(\mu) = log\mu$ ve $g^{-1}(x) = exp\,\{x\}$ olacak şekilde bir log-link fonksiyonu alınırsa;

$$E[y_i] = E[exp\{x_i'\beta + z_i'\gamma\}] = exp\{x_i'\beta\}\, E[exp\{z_i'\gamma\}] = exp\{x_i'\beta\}\, M_\gamma(z_i)$$

bulunur. Burada $M_\gamma(z_i)$, γ'nın moment çıkaran fonksiyonudur. $\gamma_i \sim N\left(0, {\sigma_\gamma}^2\right)$ ve Z' nin her bir satırının tek bir girdiye sahip bire eşit ve geri kalan satırlarının sıfır olduğu varsayılsın. Bu durumda;

$$E(y_i) = exp\{x_i'\beta\}\, exp\,\{\sigma_\gamma^2/2\}$$

veya

$$logE[y_i] = x_i{}'\beta + \{\sigma_\gamma^2/2\} \quad \text{yazılır.} \tag{3.2}$$

3.1.2.2. Bağımlı Değişkenin Varyansı

y'nin marjinal varyansı

$$\begin{aligned} var(y_i) &= var[(E\{y_i|\gamma\})] + E[var\{y_i|\gamma\}\,] \\ &= var(\mu_i) + E[\phi^2 V(\mu_i)] \\ &= var(g^{-1}[x_i'\beta + z_i{}'\gamma]) + E\big[\phi^2 V\big(g^{-1}(x_i'\beta + z_i{}'\gamma)\big)\big] \end{aligned} \tag{3.3}$$

γ verildiğinde y_i'nin elemanlarının bağımsız bir Poisson dağılımlı olduğu kabul edilsin. Böylece γ verildiğinde y_i'nin şartlı varyansı $\phi^2 V(\mu_i) = \mu_i$ olacaktır. Bu durumda

$$var(y_i) = Var(\mu_i) + E(\mu_i)$$

$$= var[exp\{x_i'\beta + z_i'\gamma\}] + E[exp\{x_i'\beta + z_i'\gamma\}]$$

$$= E[exp\,2\{x_i'\beta + z_i'\gamma\}] - E[exp\{x_i'\beta + z_i'\gamma\}]^2 + E[exp\{x_i'\beta + z_i'\gamma\}]$$

$$= exp\,\{2x_i'\beta\}\left(M_\gamma(2z_i) - \left[M_\gamma(z_i)\right]^2 + exp\,\{-x_i'\beta\}M_\gamma(z_i)\right)$$

$\gamma_i \sim N\left(0, {\sigma_\gamma}^2\right)$ ve Z'nin her bir satırı tek bir girdiye sahip bire eşit ve geri kalan satırlarının sıfır olduğu varsayılsın. Bu varsayımlar altında yeniden düzenlemeler yapılarak;

$$var(y_i) = exp\{2x_i{}'\beta\}\left(exp\left\{2\sigma_\gamma^2\right\} - exp\left\{\sigma_\gamma^2\right\}\right) + exp\{x_i{}'\beta\}exp\left\{\sigma_\gamma^2/2\right\}$$

$$= \exp\left\{x_i'\beta + \frac{{\sigma_\gamma}^2}{2}\right\}.\left(\exp\{x_i'\beta\}\left[\exp\left\{3\frac{{\sigma_\gamma}^2}{2}\right\} - \exp\left\{\frac{{\sigma_\gamma}^2}{2}\right\}\right] + 1\right)$$

$$= E[y_i]\left(\exp\,\{x_i'\beta\}\left[\exp\left\{3\frac{{\sigma_\gamma}^2}{2}\right\} - \exp\left\{\frac{{\sigma_\gamma}^2}{2}\right\}\right] + 1\right)$$

elde edilir.

$\exp\,\{x_i'\beta\}\left[\exp\left\{3\frac{{\sigma_\gamma}^2}{2}\right\} - \exp\left\{\frac{{\sigma_\gamma}^2}{2}\right\}\right] > 1$ olduğundan varyansın ortalamadan daha büyük olduğu görülür. Bundan dolayı γ verildiğinde y_i'nin şartlı dağılımı Poisson olmasına rağmen marjinal dağılımı olmayacaktır. Gerçekte bu varsayımlar altında her zaman Poisson dağılımında aşırı yayılım ile karşılaşılacaktır (McCulloch ve Searle, 2001).

3.1.2.3. Bağımlı Değişkenin Kovaryans ve Korelasyonu

y'nin elemanlarının şartlı bağımsız oldukları varsayımıyla;

$$cov(y_i, y_j) = cov\left(E[y_i|\gamma], E\left[y_j|\gamma\right]\right) + E[cov(y_i, y_j|\gamma)]$$

$$= cov\left(\mu_i, \mu_j\right) + E[0]$$

$$= cov\left(g^{-1}(x_i'\beta + z_i{}'\gamma), g^{-1}\left(x_j'\beta + z_j{}'\gamma\right)\right) \tag{3.4}$$

şeklinde bulunur. Eğer link fonksiyonu log-link fonksiyonu ise;

$$
\begin{aligned}
cov(y_i, y_j) &= cov(\exp\{x_i'\beta + z_i'\gamma\}, \exp\{x_j'\beta + z_j{}'\gamma\}) \\
&= \exp\{x_i'\beta + x_j'\beta\}\, cov(exp\{z_i'\gamma\}, \exp\{z_j'\gamma\}) \\
&= \exp\{x_i'\beta + x_j'\beta\}\left[M_\gamma(z_i + z_j) - M_\gamma(z_i).M_\gamma(z_j)\right]
\end{aligned}
$$

elde edilir.

$\gamma \sim N(0, \mathrm{I}\sigma_\gamma{}^2)$ ve Z'nin her bir satırı tek bir girdiye sahip bire eşit ve geri kalan satırlarının sıfır olduğu varsayılsın. Bu durumda;

$$cov(y_i, y_j) = \exp\{x_i'\beta + x_j'\beta\}\left(\exp\{\sigma_\gamma{}^2\}(\exp\{z_i'z_j\sigma_\gamma{}^2\} - 1)\right) \text{ dir.}$$

$z_i'z_j$=0 olduğunda, eşitlik sıfır bulunur ve $z_i'z_j = 1$ olduğunda, bu eşitlik pozitif bir değer alacaktır. . $z_i'z_j = 1$ alındığında korelasyon hesaplanırsa;

$$
\begin{aligned}
cov(y_i, y_j) &= \frac{e^{2\sigma_\gamma{}^2} - e^{\sigma_\gamma{}^2}}{\sqrt{\left(e^{2\sigma_\gamma{}^2} - e^{\sigma_\gamma{}^2} + e^{-x_i'\beta + \sigma_\gamma{}^2/2}\right).\left(e^{2\sigma_\gamma{}^2} - e^{\sigma_\gamma{}^2} + e^{-x_j'\beta + \sigma_\gamma{}^2/2}\right)}} \\
&= \frac{1}{(1 + \eta e^{-x_i'\beta})(1 + \eta e^{-x_j'\beta})}
\end{aligned}
$$

şeklinde bulunur. Burada;

$$\eta = \frac{1}{e^{3\sigma_\gamma{}^2/2} - e^{\sigma_\gamma{}^2/2}}$$

ile tanımlanmıştır.

3.2. Genelleştirilmiş Lineer Karma Modellerde Tahminleme

GLKM' lerin parametrelerini tahmin etmede iki temel yaklaşım vardır. Bunlar; olabilirlik fonksiyonuna yaklaşım ve modele yaklaşımdır. Modele yaklaşımda algoritmalar genelde Taylor serileri ile ifade edilir. Bu yaklaşımlar aynı zamanda lineerleştirme yöntemleri olarak da bilinirler. Lineerleştirme yöntemiyle ilerleyen süreç durdurma kriteri sağlanıncaya kadar devam eder. Genelleştirilmiş lineer karma modellerde lineerleştirme yöntemi, rastgele etkileri içeren modeller için Pseudo ve Penalized yarıolabilirlik (PQL) (Wolfinger ve O'Connell, 1993) yöntemi kullanılır. Bu yöntemle bulunan parametre tahminleri, rastgele etkilerin en iyi yansız tahmin edicilerinin tahminleri olacaktır. GLKM parametrelerini tahmin etmek için olabilirlik fonksiyonuna çeşitli integral yaklaşım yöntemleri vardır. Bunlar, Laplace yaklaşımı, Gauss-Hermitte Quadrature (GHQ) ve Markov zinciri Monte Carlo algoritmasıdır (MCMC). Bu yaklaşımların hepsinde sabit etki tahminleri kesin olarak doğru varsayılarak, rastgele etkilerin standart sapmasını tahmin eden klasik en çok olabilirlik tahmini ve kısıtlandırılmış en çok olabilirlik tahmini arasındaki fark ayırt edilmelidir. En çok olabilirlik tahmini çok büyük veri setleri dışında rastgele etkilerin standart sapmasını göz ardı eder ancak farklı sabit etkilere sahip modelleri karşılaştırmada çok kullanışlı bir yöntemdir.

3.2.1. Lineerleştirmeye Dayalı Pseudo Olabilirlik Tahmini

$E(y|\gamma) = g^{-1}(X\beta + Z\gamma) = g^{-1}(\eta) = \mu$

ve

$\gamma \sim N(0, G)$ ve $Var[Y|\gamma] = A^{1/2}RA^{1/2}$ olsun.

Wolfinger ve O' Connell (1993) göre μ' nün $\tilde{\beta}$ ve $\tilde{\gamma}$ noktalarına göre 1. mertebeden Taylor açılımı

$g^{-1}(\eta) = g^{-1}(\tilde{\eta}) + \tilde{\Delta}X(\beta - \tilde{\beta}) + \tilde{\Delta}Z(\gamma - \tilde{\gamma})$ şeklinde olur.

Burada $\tilde{\Delta} = \left(\frac{\partial g^{-1}(\tilde{\eta})}{\partial \eta}\right)_{\tilde{\beta},\tilde{\gamma}}$, $x\tilde{\beta} + z\tilde{\gamma}$' nın bileşenlerindeki g^{-1}'in türevlerinin bir köşegen matrisidir.

Verilen bu ifade yeniden düzenlenirse:

$$\tilde{\Delta}^{-1}\left(\mu - g^{-1}(\tilde{\eta})\right) + X\tilde{\beta} + Z\tilde{\gamma} = X\beta + Z\gamma \qquad \textbf{(3.5)}$$

bulunur.

$Var[P|\gamma] = \tilde{\Delta}^{-1}A^{1/2}RA^{1/2}\tilde{\Delta}^{-1}$

şartlı varyansa sahip olan P,

$$\tilde{\Delta}^{-1}\left(Y - g^{-1}(\tilde{\eta})\right) + X\tilde{\beta} + Z\tilde{\gamma} \equiv P$$

ile tanımlansın.

Model, P pseudo bağımlı değişkeniyle birlikte $\mathrm{P} = \mathrm{X}\beta + \mathrm{Z}\gamma + \varepsilon$ şeklinde bir karma model olur. Burada β sabit etkilerin vektörü, γ rastgele etkilerdir. Böylece parametrelerin tahmin edilebileceği bir lineer modele dönüştürülmüş olunur.
$Var(\theta) = \mathrm{ZGZ}^{-1} + \widetilde{\Delta}^{-1}A^{1/2}RA^{1/2}\widetilde{\Delta}^{-1}$ ile lineer karma pseudo model için marjinal varyansı tanımlansın. Burada θ, $q \times 1$ boyutlu G ve R bilinmeyenlerini içeren parametre vektörüdür. Bu lineerleştirme modelinden yararlanarak P' nin dağılımının bilindiği varsayımı ile birlikte bir amaç fonksiyonu tanımlansın. ε' nun normal dağıldığını varsayımıyla P için log-pseudo olabilirlik fonksiyonu ve kısıtlandırılmış log-pseudo olabilirlik fonksiyonu aşağıdaki gibidir:

$$l(\theta, p) = -\frac{1}{2}log|V(\theta)| - \frac{1}{2}r'V(\theta)^{-1}r - \frac{f}{2}log\{2\pi\}$$

$$l_R(\theta, p) = -\frac{1}{2}log|V(\theta)| - \frac{1}{2}r'V(\theta)^{-1}r - \frac{1}{2}log|x'V(\theta)^{-1}x| - \frac{f-k}{2}log\{2\pi\}$$

Burada $r = p - X(X'V^{-1}X)^{-}X'V^{-1}p$, f analizde kullanılan frekansların toplamını ve k, x' in rankını göstermektedir. Minimizasyon için amaç fonksiyonlarını $-2l(\theta, p)$ ve $-2l_R(\theta, p)$ alındığında; sabit etkilerin ve rastgele etkilerin parametre tahminleri:

$$\hat{\beta} = X(X'V(\hat{\theta})^{-1}X)^{-}X'V(\hat{\theta})^{-1}p \qquad \textbf{(3.6)}$$

$$\hat{\gamma} = \hat{G}Z'V(\hat{\theta})^{-1}\hat{r} \qquad \textbf{(3.7)}$$

şeklinde bulunur.

3.2.2. Laplace Yaklaşımına Dayalı Olabilirlik Tahmini

β sabit etkilerin parametre vektörü ve θ da kovaryans parametrelerinin vektörü olsun. Laplace tahmini için θ, G-side parametrelerini ve olası bir ölçü parametresi ϕ' yi içerir ve verinin şartlı dağılımı böyle bir ölçü parametresini sağlar. θ^*, G-side parametrelerinin vektörüdür. Bir karma modeldeki verinin marjinal dağılımı:

$$p(y) = \int p(y|\gamma,\beta,\phi)p(\gamma|\theta^*)d\gamma$$

$$= \int \exp\{\log\{p(y|\gamma,\beta,\phi)\} + \log\{p(\gamma|\theta^*)\}\}d\gamma$$

$$= \int \exp\{c_1 f(y,\beta,\theta;\gamma)\}d\gamma \quad \textbf{(3.8)}$$

Eğer c_1 sabiti büyük bir değer ise bu integralin Laplace yaklaşımı:

$L(\beta,\theta,\hat{\gamma},y) = \left(\frac{2\pi}{c_1}\right)^{n_\gamma/2} |-f''(y,\beta,\theta;\hat{\gamma})|^{-1/2} e^{c_1 f(y,\beta,\theta;\hat{\gamma})}$ şeklindedir.

n_γ, γ' daki elemanların sayısıdır. f'', ikinci türev matrisi

$$f''(y,\beta,\theta;\hat{\gamma}) = \left.\frac{\partial^2 f(y,\beta,\theta;\gamma)}{\partial\gamma\partial\gamma'}\right|_{\hat{\gamma}}$$

ve $\hat{\gamma}$,

$$\frac{\partial f(y,\beta,\theta;\gamma)}{\partial\gamma} = 0$$

şartını sağlar.

Laplace parametre tahmini için amaç fonksiyonu:

$-2\log\{L(\beta,\theta,\hat{\gamma},y)\}$' dir.

3.2.3. Adaptive Gauss-Hermite Quadrature Dayalı Olabilirlik Tahmini

Quadrature yöntemi, Laplace yöntemi gibi bir integral yaklaşım yöntemidir. Genelleştirilmiş lineer karma modeller için Quadrature yöntemini seçildiğinde Adaptive Gauss-Hermite Quadrature kuralları ile birlikte marjinal log-olabilrlik fonksiyonuna yaklaşılır. Gaussian Quadrature yöntemi olabilirlik ölçümleri karşısında nümerik integrallerin çözümü için oldukça uygundur. Gauss-Hermite yöntemi yoğunluk fonksiyonu $\exp\{-x^2\}$ 'nin çekirdeği olduğu zaman ve integral reel sayılar boyunca alındığında uygun bir yöntem olur.

$p(x)$ bir olasılık yoğunluk fonksiyonu ve $f(x)$ fonksiyonu $p(x)$' e karşı integrallenebilir olsun. Quadrature kurallarından;

$$\int_{-\infty}^{\infty} f(x)p(x)dx \approx \sum_{n=1}^{N} w_i f(x_i)$$

yazılır. Burada N Quadrature noktalarının sayısı, w_i quadrature ağırlıklandırması ve x_i'lerde yatay eksenlerdir.

β sabit etkilerin parametre vektörü ve θ da kovaryans parametrelerinin vektörü olsun. Quadrature tahmini için θ, G-side parametrelerin ve olası bir ölçü parametresi ϕ' yi içerir ve verinin şartlı dağılımı böyle bir ölçü parametresini sağlar. θ^*, G-side parametrelerinin vektörüdür. Bir karma modeldeki i. grup için verinin marjinal dağılımı:

$$p(y_i) = \int \dots \int p(y_i|\gamma_i, \beta, \phi)p(\gamma_i|\theta^*)d\gamma_i$$

şeklinde ifade edilebilir.

N_q, her bir rastgele etkide Quadrature noktalarının sayısını ve γ rastgele etkilerin sayısını göstersin. Bu durumda;

$-\log\{p(y_i|\gamma_i, \beta, \phi)p(\gamma_i|\theta^*)\} = f(y_i, \beta, \theta; \gamma_i)$ fonksiyonunu minimize eden γ_i'nin $\widehat{\gamma_i}$ gibi Bayesci tahminleri elde edilir.

Eğer $z = [z_1, \dots, z_n]$ Gauss-Hermite Quadrature için standart apsisler ise ve

$z_j^* = [z_{j1}, \dots, z_{jr}]$

r- boyutlu Quadrature ağında bir nokta, merkezi ve ölçek apsisleri ise;

$a_j^* = \widehat{\gamma_i} + 2^{1/2} f''(y_i, \beta, \theta; \widehat{\gamma_i})^{-1/2} z_j^*$ şeklindedir.

Bu merkezi ve ölçek apsisleri, $w = [w_1, \dots, w_{N_q}]$ Gauss-Hermite Quadrature ağırlıklandırmalarıyla birlikte;

$$p(y_i) = \int \dots \int p(y_i|\gamma_i, \beta, \phi)p(\gamma_i|\theta^*)d\gamma_i \approx$$

$$2^{r/2} \left| f''(y_i, \beta, \theta; \widehat{\gamma_i})^{-1/2} \right| \sum_{j_1=1}^{N_q} \dots \sum_{j_r=1}^{N_q} \left[p(y_i|a_j^*, \beta, \phi)p(a_j^*|\theta^*) \prod_{k=1}^{r} w_{jk} exp z_{jk}^2 \right]$$

adaptive Quadrature tahminleri için olabilirlik fonksiyonu elde edilmiş olunur.

Yukarıda ifade edilen yöntemlerin avantaj ve dezavantajları ve genel yazılım isimleri Çizelge 3.1. ‘ de özetlenmiştir.

Çizelge 3.1. GLKM yöntemlerinin avantaj ve dezavantajları özet tablosu

Yöntem	Avantajı	Dezavantajı	Yazılım
Penalized Quasilikelihod	Esnektir ve yaygın olarak kullanılır.	Küçük ortalamalar veya büyük varyanslar için yanlı olduğundan olabilirlik çıkarımları için uygun değildir.	PROC GLIMMIX (SAS), GLMM(Genstat), glmmPQL, glmer ®
Laplace Yaklaşımı	PQL yönteminden daha doğru sonuçlar verir.	PQL yöntemine göre daha az esnek ve daha yavaştır.	PROC GLIMMIX, glmer ®, AD Model Builder HLM
Gauss-Hermite Quadrature	Laplace yönteminden daha doğru sonuçlar verir	Laplace yönteminden daha yavaş çalışır ve rastgele etkiler 2-3 ile sınırlıdır.	PROC GLIMMIX , PROC NLMIXED(SAS) , glmer ® , glmmML ®
Markov Chain Monte Carlo	Oldukça esnektir ve keyfi sayıda rastgele etkiler için doğru sonuçlar verir.	Çok yavaştır, Bayesci yaklaşım içersinde teknik açıdan hesaplanması zordur.	WinBUGS, JAGS, MCMCpack, ® , AD Model Builder

3.3. Uyum İstatistikleri

GLKM' de kullanılan çeşitli uyum istatistikleri; Akaike'nin Bilgi Kriteri (AIC), (Akaike, 1974), Akaike Bilgi Kriterinin Küçük Örneklem Yanlı Düzeltilmiş Hali (AICC) dir. AICC, Hurvich ve Tsai (1989) ve Burnham ve Anderson (1998) tarafından geliştirilmiştir. Bayesci Bilgi Kriteri (BIC), Schwarz (1978), Tutarlı Akaike Bilgi Kriteri (CAIC), Bozdoğan (1987) ve Hannan ve Quinn Bilgi Kriteri (HQIC), Hannan ve Quinn tarafından 1979 yılında geliştirilmiş bilgi kriterlerindendir. En küçük bilgi kriterlerini veren modeller en uygun modellerdir. Çizelge 3.2. de bilgi kriterleri için formüller verilmiştir(SAS Institute Inc., 2011).

Çizelge 3.2. Bilgi kriterleri

Kriter	**Formül**
AIC	$-2l + 2d$
AICC	$-2l + 2d\,n^*/(n^* - d - 1)$
HQIC	$-2l + 2dloglogn$
BIC	$-2l + dlogn$
CAIC	$-2l + d(logn + 1)$

l , log olabilirlik, log sözde olabilirlik veya log yarı (quasi) olabilirliğin maksimum değerini, d, modelin boyutunu ve n, n^* verinin boyutunu gösterir.d, n ve n^* modele bağlı büyüklüklerdir.

GLM de; d değeri optimizasyondaki parametrelerin sayısına, n değeri maksimum olabilirlik ve yarı olabilirlik için frekansların toplamına ve kısıtlandırılmış maksimum olabilirlik tahmini içinde $f - rank(X)$ ‘ e eşittir. n^* değeri, $n < d + 2$ olmadığı durumlarda $n^* = n$, diğer durumlarda $n^* = d + 2$ olacaktır.

GLKM de; d değeri etkili kovaryans parametrelerinin sayısına, n değeri etkili grupların sayısına eşittir. Eğer bu değer 1’ e eşit değilse bu durumda n değeri belirlenmiş ilk G- side rastgele etkilerin düzeylerinin sayısına eşit olacaktır. Etkili grupların sayısı 1’ e eşit ise G- side rastgele etki yoktur ve n değeri aşağıdaki şekilde tanımlanır:

$$n = \begin{cases} f - rank(X) \\ f \end{cases}$$

Burada f kullanılan frekansların toplamıdır. Ayrıca n^* değeri $n^* < d + 2$ olmadığı durumlarda; $n^* = f$ veya $n^* = f - rank(X)$, diğer durumlarda ise $n^* = d + 2$ olacaktır.

4. BULGULAR VE TARTIŞMA

Uygulamada, 2009 yılında ” Genç Nüfusun Sorun Algılaması; Trabzon” örneği adlı çalışmada (Murat, N. ve ark., 2009) Trabzon merkezde yaşayan 15-24 yaş arası 1286 gençle yüz yüze görüşülerek elde edilen veriler kullanılmıştır. Yapılan bu çalışmada Trabzon'da yaşayan gençlerin kendilerini nasıl tanımladıkları, aile, akraba ve arkadaş çevreleriyle olan ilişkilerinin tespit edilmesi, kötü alışkanlıklarının olup olmadığı, yakın çevreye ve dünyevi meselelere duyarlılıklarının ölçülmesi, beklentilerinin ve sorunlarının neler olduğunun tespit edilmesi, mesleki becerilerinin olup olmadığının araştırılması, harcama yapılan alanların tespiti, Trabzon'da toplanılan mekanların tespiti ve boş zamanlarının nasıl değerlendirdiklerinin belirlenmesi, ideal anne ve babadan alınması gereken önemli unsurların tespit edilmesi, kendilerini Trabzon'da nasıl hissettikleri ve gelecekte Trabzon'da yaşamayı düşünüp düşünmediklerinin tespiti, gelecekte Trabzon'un nasıl bir şehir olacağının gençler tarafından tahmin edilmesi gibi benzer hususlarda gençlerin görüşlerinin tespit edilmesi ve Trabzon belediyesine Gençlere yönelik neler yapılabileceğinin sunulması amaçlanmıştır. Çalışmada, cinsiyet yaş ve eğitim kotalarına dikkat edilmiştir. Bu çalışma için Çizelge 4.1. ‘ de tanımlanan değişkenler seçilmiştir.

Çizelge 4.1. Tanımlanan değişkenlerin özellikleri

Değişken	Açıklama	Değişken Türü	Değişken Düzey Değerleri
y1	"Kendinizi Trabzon da nasıl hissediyorsunuz?"	**Bağımlı**	**1:** Mutlu, **2:**Mutsuz
y2	"Günlük ortalama kaç saat internet kullanıyorsunuz?"	**Bağımlı**	0-15
x1	"Ailenizin aylık gelir ortalaması"	**Bağımsız**	**1:** 500-den az **4:**1001-1500 **7:**2500-... **2:** 500-75 **5:**1501-2000 **3:**751-1000 **6:**2001-2500
x2	"Yaş grubu"	**Bağımsız**	**1:** 1983-85 **2:**1986-88 **3:**1989-91
x3	"Cinsiyet"	**Bağımsız**	**1:**E **2:**B
x4	"Eğitim durumu"	**Bağımsız**	**1:**İlkokul mezunu **4:**Ünv. mezunu **2:** Ortaokul mezunu **5:** Lise öğrencisi **3:** Lise mezunu **6:**Ünv. Öğrencisi, yüksek lisans, doktora öğr.

İlk olarak " Kendinizi Trabzon da nasıl hissediyorsunuz?" soruyla ifade edilen bağımlı değişkene, açıklayıcı değişkenlerin etkisini modellemek için GLM ve GLKM modeller uygulanmıştır. Her iki modelin kullanılmasındaki amaç GLKM'nin gücünü ortaya koymaktır. GLM için Çizelge 4.2' de verilen SAS makrosu kullanılmıştır.

Çizelge 4.2. Sas 9.3' de veri girişi ve komutlar

```
data trabzon;
input y1 y2 x1 x2 x3 x4 ;
datalines;
1    8    7    2    1    6
2    8    2    3    1    5
1    2    2    3    1    5
1    3    5    2    1    6
...
...
...
1    5    6    3    2    3
1    0    4    3    2    6
1    6    3    1    2    6
;
proc genmod data=trabzon ;
class x1 x2 x3 x4 ;
model y1= x2 x3 x4 / dist=binary link=logit ;
run;
```

Çizelge 4.3. 'de GENMOD süreci model bilgisi sunulmuştur.

Çizelge 4.3. y1 bağımlı değişkeninin GLM analizinde model bilgisi

GENMOD Süreci	
Model Bilgisi	
Veri Seti	Trabzon çalışması
Dağılım	Binom
Link Fonksiyonu	Logit
Bağımlı Değişken	y1

Model bilgisi tablosu uygun model hakkındaki temel bilgileri verir. "Trabzon" adlı çalışma veri setinde y1 bağımlı değişkeninin olasılık dağılımı, binom ve link fonksiyonu, logit linktir.

Çizelge 4.4.'de açıklayıcı değişkenin seviye bilgileri verilmektedir

Çizelge 4.4. y1 bağımlı değişkeninin GLM analizinde bağımsız değişkenlere ait sınıf seviye bilgisi

Sınıf Düzey Bilgisi		
Sınıf	**Düzey**	**Değeri**
x1	7	1 2 3 4 5 6 7
x2	3	1 2 3
x3	2	1 2
x4	6	1 2 3 4 5 6

Çizelge 4.5. y1 ikili bağımlı değişkeninin profilini göstermektedir.

Çizelge 4.5. GLM analizinde bağımlı değişken bilgisi

Bağımlı Değişkenin Özellikleri		
Sıralı Değer	**y1**	**Toplam Frekans**
1	1	759
2	2	527

Çizelge 4.5.'e göre 1286 genç arasından bu soruyu, kendimi " mutlu" hissediyorum şeklinde yanıtlayan 759 kişi, kendimi " mutsuz" hissediyorum şeklinde yanıtlayan 527 kişidir. Çizelge 4.6'da model karşılaştırması için kullanılacak uyum iyiliği kriterleri verilmektedir.

Çizelge 4.6. y1 bağımlı değişkeninin GLM analizinde uyum iyiliği kriterleri

Uyum İyiliği Kriterleri	
Kriter	**Değer**
Log Olabilirlik	-8.479.475
Full Log Olabilirlik	-8.479.475
AIC	17.138.950
AICC	17.140.361
BIC	17.603.287

Farklı modeller karşılaştırılırken AIC, AICC, BIC kriterlerinin küçük olduğu model her zaman daha iyidir.

Çizelge 4.7'de GLM için parametre tahminleri verilmiştir.

Çizelge 4.7. y1 bağımlı değişkeninin GLM analizinde parametre tahminleri

En Çok Olabilirlik Parametre Tahminlerinin Analizi								
Parametre	**Serbestlik Derecesi**		**Tahmin**	**Standart Hata**	**%95 Wald Güven Aralıkları**		**Wald Ki-Kare**	**P Değeri**
Sabit		1	0.5331	0.1669	0.2060	0.8601	10.21	0.0014
x2	1	1	0.0243	0.1698	-0.3085	0.3572	0.02	0.8861
x2	2	1	-0.0012	0.1543	-0.3037	0.3012	0.00	0.9936
x3	1	1	0.3461	0.1241	0.1028	0.5894	7.77	0.0053
x4	1	1	-1.0844	0.3485	-1.7674	-0.4013	9.68	0.0019
x4	2	1	-0.8004	0.2089	-1.2098	-0.3909	14.68	0.0001
x4	3	1	-0.7184	0.1476	-1.0076	-0.4291	23.70	<.0001
x4	4	1	0.1322	0.2571	-0.3717	0.6362	0.26	0.6070
x4	5	1	-0.3210	0.2102	-0.7331	0.0911	2.33	0.1268

Çizelge 4.7.'den anlaşıldığı gibi x2, yaş grubu açıklayıcı değişkeninin tüm seviyeleri için %5 anlamlılık seviyesinde p değerleri 0.05' den büyük olduğundan model için anlamlı bir katkıya sahip değillerdir. x3, cinsiyet açıklayıcı değişkeninin birinci seviyesi için p değeri 0.05'den küçük olduğundan modelde anlamlı bir katkıya sahiptir ve x4, eğitim durumu açıklayıcı değişkenine bakıldığında bir, iki ve üçüncü seviyeleri için p değerlerinin 0.05' den küçük olduğu, dolayısıyla modelde anlamlı bir katkıya sahip olduğu fakat dört ve beşinci seviyelerinin p değerlerinin 0.05'den büyük olduğundan anlamlı bir katkıya sahip olmadıkları görülmektedir.

Benzer modellemeyi GLKM' yi kullanarak da yapabiliriz.

y1 bağımlı değişkeni için x2, x3 ve x4 bağımsız değişkenleri sabit etkili ve x1 bağımsız değişkeni rastgele etkili alınarak Çizelge 4.8' de gösterilen GLIMMIX makrosu kullanımıyla GLKM modelleri uygulanır.

Çizelge 4.8. GLIMMIX sürecinde veri girişi ve komutlar

```
data trabzon;
input y1 y2 x1 x2 x3 x4 ;
datalines;
1    8    7    2    1    6
2    8    2    3    1    5
1    2    2    3    1    5
1    3    5    2    1    6
2    0    4    3    1    3
1    6    5    3    1    4
...
...
...
2    8    1    3    2    6
1    0    2    2    2    1
1    5    6    3    2    3
1    0    4    3    2    6
1    6    3    1    2    6
;
proc glimmix data=trabzon method=laplace;
class x1 x2 x3 x4 ;
model y1= x2 x3 x4 / dist=binary link=logit solution ;
random intercept  /subject=x1;
run;
```

Çizelge 4.8.‘ de y1 bağımlı değişkeni için GLIMMIX sürecinde Laplace yöntemi seçildi.

Çizelge 4.8’de ifade edilen makro için model bilgisi Çizelge 4.9’da verilmiştir.

Çizelge 4.9. y1 bağımlı değişkeninin GLKM analizinde Laplace yöntemi için model bilgisi

GLIMMIX Süreci	
Model Bilgisi	
Veri Seti	Trabzon Çalışması
Bağımlı Değişken	y1
Dağılım	Binary
Link Fonksiyonu	Logit
Varyans Fonksiyonu	Tanımlı
Sınıflandırılan Varyans Matrisi	x1
Tahmin Yöntemi	En Çok Olabilirlik
Olabilirlik Yaklaşımı	Laplace
Degrees of Freedom Method	Containment

Çizelge 4.9.‘da ki “Model Bilgi” tablosuna göre; “Trabzon” adlı çalışma veri setinde y1 bağımlı değişkeninin olasılık dağılımı binary ve link fonksiyonu logit linktir. Varyans fonksiyonu tanımlıdır. Tahmin yöntemi en çok olabilirlik yöntemi ve olabilirlik yaklaşımı Laplace yaklaşımıdır.

GLKM için model boyut bilgisi Çizelge 4.10’ da ve optimizasyon bilgisi de Çizelge 4.11.’de verilmiştir.

Çizelge 4.10. y1 bağımlı değişkeninin GLKM analizinde Laplace yöntemi için model boyut bilgisi

Model Boyut Bilgisi	
G-side Kovaryans Parametresi	1
X’deki Sütunlar	12
Her Grup için Z’deki Sütunlar	1
V'deki Rastgele Gruplar	7
Her Grup için Maksimum Gözlem	391

Model için tek bir rastgele etki alındığından Z matrisinde 1 sütun elemanı vardır.

Çizelge 4.11. y1 bağımlı değişkeninin GLKM analizinde Laplace yöntemi için optimizasyon bilgisi

Optimizasyon Bilgisi	
Optimizasyon Tekniği	Dual Quasi-Newton
Optimizasyon Parametre Sayısı	10
Alt Kısıt	1
Üst Kısıt	0
Sabit Etkiler	Not Profiled
Starting From	GLM Tahminleri

Çizelge 4.11.’den; optimizasyon problemini çözmede Quasi-Newton yöntemi kullanılmaktadır. Newton-Raphson yöntemine karşıt Quasi-Newton yöntemi ikinci türevlere gerek duymaz. Bu çalışmada kovaryans parametreleri sınırsız olmadığından, süreç rastgele etkinin varyansı için bir alt sınır kısıtı (sıfır) uygular ve optimizasyon yöntemi bir dual Quasi- Newton yönteme dönüşür.

Çizelge 4.12’ de iterasyon süreci verilmiştir.

Çizelge 4.12. y1 bağımlı değişkeninin GLKM analizinde Laplace yöntemi için iterasyon süreci

İterasyon Süreci					
İterasyon	**Başlangıç**	**Fonksiyon**	**Amaç Fonksiyonu**	**Değişim**	**Maksimum Gradyant**
0	0	4	1694.5518127	-	4.814.008
1	0	5	1694.5423019	0.00951073	1.336.291
2	0	3	1694.5263395	0.01596245	0.684038
3	0	4	1694.5254159	0.00092364	0.377775
4	0	3	1694.5246215	0.00079437	0.456072
5	0	3	1694.523522	0.00109951	0.268803
6	0	3	1694.5231949	0.00032708	0.248203
7	0	3	1694.5230706	0.00012433	0.257201
8	0	3	1694.5229975	0.00007310	0.242158
9	0	2	1694.5229367	0.00006072	0.197807
10	0	4	1694.5223457	0.00059106	0.03137
11	0	4	1694.5223282	0.00001747	0.031901
12	0	2	1694.5223065	0.00002173	0.003588
13	0	3	1694.5223065	0.00000001	0.000051

Çizelge 4.12.’ ye göre, sürecin 13 iterasyon ve 46 fonksiyon değerlendirmesinden sonra yakınsadığı görülür. Değişim sütunu iterasyonlar arasındaki amaç fonksiyonlarının değişimini gösterir.

Çizelge 4.13’de Laplace yöntemi için uyum istatistikleri verilmiştir.

Çizelge 4.13. y1 bağımlı değişkeninin GLKM analizinde Laplace yöntemi için uyum istatistikleri

Uyum İstatistikleri	
-2Log Olabilirlik	1694.52
AIC	1714.52
AICC	1714.69
BIC	1713.98
CAIC	1723.98
HQIC	1707.84

“Uyum İstatistikleri “ tablosu uygun model hakkındaki bilgiyi listeler. -2 Log olabilirlik değerleri iç-içe modelleri karşılaştırmak için kullanışlıdır ve bilgi kriterleri; AIC, AICC, BIC, CAIC, HQIC iç-içe olmayan modelleri karşılaştırmada kullanışlıdır.

Çizelge 4.14’de şartlı dağılımın uyum istatistikleri verilmiştir.

Çizelge 4.14. y1 bağımlı değişkeninin GLKM analizinde Laplace yöntemi için şartlı dağılımın uyum istatistiği

Şartlı Dağılım için Uyum İstatistikleri	
-2 log L(y1 | r. etki)	1686.37
Pearson Ki-Kare	1279.37
Pearson Ki-Kare / Serbestlik Derecesi	0.99

Genelde, Pearson istatistiği ve serbestlik derecesi arasındaki oran 1’ e eşittir. Bu oran 1’den büyük olduğunda aşırı yayılım sorunu ortaya çıkar. Çizelge 4.14.’e göre bu oran 0,99 olduğundan aşırı yayılım sorunu bulunmamaktadır.

Çizelge 4.15’de kovaryans parametre tahmini verilmiştir.

Çizelge 4.15. y1 bağımlı değişkeninin GLKM analizinde Laplace yöntemi için kovaryans parametre tahmini

Kovaryans Parametre Tahminleri			
Kovaryans Parametresi	**Grup**	**Tahmin**	**Standart Hata**
Sabit	x1	0.02775	0.03860

Çizelge 4.16’da GLKM analizinde Laplace yöntemi kullanıldığında elde edilen parametre tahminleri verilmektedir.

Çizelge 4.16. y1 bağımlı değişkeninin GLKM analizinde Laplace yönteminde sabit etkiler için çözümleri

Sabit Etkiler için Çözümler								
Etkiler	**x2**	**x3**	**x4**	**Tahmin**	**Standart Hata**	**Serbestlik Derecesi**	**t Değeri**	**P değeri**
Sabit				0.5372	0.1835	6	2.93	0.0264
x2	1			0.02513	0.1703	1271	0.15	0.8827
x2	2			-0.00420	0.1548	1271	-0.03	0.9784
x3		1		0.3505	0.1252	1271	2.80	0.0052
x4			1	-10.913	0.3503	1271	-3.12	0.0019
x4			2	-0.7951	0.2098	1271	-3.79	0.0002
x4			3	-0.7283	0.1483	1271	-4.91	<.0001
x4			4	0.1254	0.2576	1271	0.49	0.6265
x4			5	-0.3443	0.2117	1271	-1.63	0.1041

GLM yöntemine benzer olarak; x2, yaş grubu açıklayıcı değişkeninin tüm seviyeleri için %5 anlamlılık seviyesinde p değerleri 0.05’ den büyük olduğundan model için anlamlı bir katkıya sahip değillerdir. x3, cinsiyet açıklayıcı değişkeninin birinci seviyesi için p değeri 0.05’den küçük olduğundan modelde anlamlı bir katkıya sahiptir ve x4, eğitim durumu açıklayıcı değişkenine bakıldığında bir, iki ve üçüncü seviyeleri için p değerlerinin 0.05’ den küçük olduğu, dolayısıyla modelde anlamlı bir katkıya sahip olduğu fakat dört ve beşinci seviyelerinin p değerlerinin 0.05’den büyük olduğundan anlamlı bir katkıya sahip olmadıkları görülmektedir.

GLKM analizinde, Laplace yöntemi için genel değerlendirme Çizelge 4.17’ye bakılarak yapılabilir.

Çizelge 4.17. y1 bağımlı değişkeninin GLKM analizinde Laplace yönteminde sabit etkilerin testleri

Sabit Etkilerin Tip-III Testleri				
Etki	**Payın S.Derecesi**	**Paydanın S.Derecesi**	**F Değeri**	**P Değeri**
x2	2	1271	0.02	0.9803
x3	1	1271	7.83	0.0052
x4	5	1271	7.74	<.0001

Çizelge 4.17'den, %5 anlamlılık seviyesinde x2, açıklayıcı değişkeninin model için anlamsız; x3 ve x4 açıklayıcı değişkenlerinin modelde anlamlı oldukları görülür.

Aynı bağımlı değişken için GLKM 'yi Quadrature yöntemi ile çözmede kullanılan SAS makrosu Çizelge 4.18'de verilmiştir.

Çizelge 4.18. GLIMMIX sürecinde Quadrature yönteminin seçimi

```
proc glimmix data=trabzon method=quad;
class x1 x2 x3 x4 ;
model y1= x2 x3 x4 / dist=binary link=logit solution ;
random intercept  /subject=x1;

run;
```

Quadrature yöntemi için uyum istatistikleri Çizelge 4.19'da verilmiştir.

Çizelge 4.19. y1 bağımlı değişkeninin GLKM analizinde Quadrature yöntemi için uyum istatistikleri

Uyum İstatistikleri	
-2Log Olabilirlik	1694.52
AIC	1714.52
AICC	1714.69
BIC	1713.98
CAIC	1723.98
HQIC	1707.84

Çizelge 4.20'de GLKM analizinde Quadrature yöntemi kullanıldığında elde edilen parametre tahminleri verilmektedir

Çizelge 4.20. y1 bağımlı değişkeninin GLKM analizinde Quadrature yönteminde sabit etkiler için çözümleri

Sabit Etkiler için Çözümler								
Etki	**x2**	**x3**	**x4**	**Tahmin**	**Standart Hata**	**Serbestlik Derecesi**	**t Değeri**	**P Değeri**
Sabit				0.5372	0.1835	6	2.93	0.0264
x2	1			0.02513	0.1703	1271	0.15	0.8827
x2	2			-0.00420	0.1548	1271	-0.03	0.9784
x3		1		0.3505	0.1252	1271	2.80	0.0052
x4			1	-10.913	0.3503	1271	-3.12	0.0019
x4			2	-0.7951	0.2098	1271	-3.79	0.0002
x4			3	-0.7283	0.1483	1271	-4.91	<.0001
x4			4	0.1254	0.2576	1271	0.49	0.6265
x4			5	-0.3443	0.2117	1271	-1.63	0.1041

x2 bağımsız değişkeninin tüm seviyeleri %5 anlamlılık seviyesinde model için anlamlı değildir. x3 bağımsız değişkeninin birinci seviyesi ve x4 bağımsız değişkeninin birinci, ikinci ve üçüncü seviyeleri %5 anlamlılık seviyesinde modele katkı sağlar.

GLKM analizinde, Quadrature yöntemi için genel değerlendirme Çizelge 4.21'e bakılarak yapılabilir.

Çizelge 4.21. y1 bağımlı değişkeninin GLKM analizinde Quadrature yönteminde sabit etkilerin testleri

Sabit Etkilerin Tip-III Testleri				
Etki	**Payın S.Derecesi**	**Paydanın S.Derecesi**	**F Değeri**	**P Değeri**
x2	2	1271	0.02	0.9803
x3	1	1271	7.83	0.0052
x4	5	1271	7.74	<.0001

Çizelge 4.21'den, %5 anlamlılık seviyesinde x2, açıklayıcı değişkeninin model için anlamsız; x3 ve x4 açıklayıcı değişkenlerinin modelde anlamlı oldukları görülür.

y1 bağımlı değişkeni için GLKM 'yi RSPL yöntemi ile çözmede kullanılan SAS makrosu Çizelge 4.22'de verilmiştir.

Çizelge 4.22. GLIMMIX sürecinde RSPL yönteminin seçimi

```
proc glimmix data=trabzon method=rspl;
class x1 x2 x3 x4 ;
model y1= x2 x3 x4 / dist=binary link=logit solution ;
random intercept  /subject=x1;

run;
```

Çizelge 4.23'de RSPL yöntemi için uyum istatistikleri verilmiştir.

Çizelge 4.23. y1 bağımlı değişkeninin GLKM analizinde RSPL yöntemi için uyum istatistikleri

Uyum İstatistikleri	
-2 Res Log Pseudo-Olabilirlik	5547.51
Genelleştirilmiş Ki-Kare	1282.07
Genelleştirilmiş Ki-Kare / Serbestlik Derecesi	1.00

Çizelge 4.24'de GLKM analizinde RSPL yöntemi kullanıldığında elde edilen parametre tahminleri verilmektedir

Çizelge 4.24. y1 bağımlı değişkeninin GLKM analizinde RSPL yönteminde sabit etkiler için çözümleri

Sabit Etkiler için Çözümler								
Etki	**x2**	**x3**	**x4**	**Tahmin**	**Standart Hata**	**Serbestlik Derecesi**	**t Değeri**	**P Değeri**
Sabit				0.5390	0.1914	6	2.82	0.0305
x2	1			0.02559	0.1702	1271	0.15	0.8805
x2	2			-0.00495	0.1547	1271	-0.03	0.9745
x3		1		0.3505	0.1253	1271	2.80	0.0052
x4			1	-1.0886	0.3502	1271	-3.11	0.0019
x4			2	-0.7908	0.2098	1271	-3.77	0.0002
x4			3	-0.7275	0.1482	1271	-4.91	<.0001
x4			4	0.1234	0.2574	1271	0.48	0.6317
x4			5	-0.3489	0.2111	1271	-1.65	0.0985

x2 bağımsız değişkeninin tüm seviyeleri %5 anlamlılık seviyesinde model için anlamlı değildir. x3 bağımsız değişkeninin birinci seviyesi ve x4 bağımsız değişkeninin birinci, ikinci ve üçüncü seviyeleri %5 anlamlılık seviyesinde modele katkı sağlar.

GLKM analizinde, RSPL yöntemi için genel değerlendirme Çizelge 4.25'e bakılarak yapılabilir.

Çizelge 4.25. y1 bağımlı değişkeninin GLKM analizinde RSPL yönteminde sabit etkilerin testleri

Sabit Etkilerin Tip-III Testleri				
Etki	**Payın S.Derecesi**	**Paydanın S.Derecesi**	**F Değeri**	**P Değeri**
x2	2	1271	0.02	0.9788
x3	1	1271	7.83	0.0052
x4	5	1271	7.68	<.0001

Çizelge 4.25'den, %5 anlamlılık seviyesinde x2, açıklayıcı değişkeninin model için anlamsız; x3 ve x4 açıklayıcı değişkenlerinin modelde anlamlı oldukları görülür.

"Günlük ortalama kaç saat internet kullanıyorsunuz?" soruyla ifade edilen bağımlı değişkene, açıklayıcı değişkenlerin etkisini modellemek için GLM ve GLKM modeller uygulanmıştır. GLM için Çizelge 4.26'da verilen SAS makrosu kullanılmıştır.

Çizelge 4.26. Sas 9.3. de GENMOD analizi seçimi

```
proc genmod data=trabzon;
class x1 x2 x3 x4 ;
model y2= x2 x3 x4 / dist=poisson link=log ;

run;
```

Çizelge 4.27'de y2 bağımlı değişkeni için GLM analizinde model bilgisi verilmiştir.

Çizelge 4.27. y2 bağımlı değişkeni için GLM analizinde model bilgisi

GENMOD Süreci	
Model Bilgileri	
Veri Seti	Trabzon Çalışması
Dağılım	Poisson
Link Fonksiyonu	Log
Varyans Fonksiyonu	Tanımlı
Bağımlı Değişken	y2

"Trabzon" adlı çalışma veri setinde y2 bağımlı değişkeninin olasılık dağılımı poisson ve link fonksiyonu log linktir.

Çizelge 4.28'de uyum iyiliği kriterleri verilmiştir.

Çizelge 4.28 y2 bağımlı değişkeni için GLM analizinde uyum iyiliği kriterleri

Uyum İyiliği Kriterleri			
Kriter	**Serbestlik Derecesi**	**Değer**	**Değer/Serbestlik Derecesi**
Sapma	1277	3529.4111	2.7638
Ölçeklendirilmiş Sapma	1277	3529.4111	2.7638
Pearson Ki-Kare	1277	3900.1683	3.0542
Ölçeklendirilmiş Ki-Kare	1277	3900.1683	3.0542
Log Olabilirlik		3689.001	
Full Log Olabilirlik		-3282.5343	
AIC		6583.0685	
AICC		6583.2096	
BIC		6629.5022	

Çizelge 4.29.'da GLM için parametre tahminleri verilmiştir.

Çizelge 4.29. y2 bağımlı değişkeni için GLM analizinde parametreler tahmini

En Çok Olabilirlik Parametre Tahminlerinin Analizi								
Parametre		**Serbestlik Derecesi**	**Tahmin**	**Standart Hata**	**%95 Wald Güven Aralıkları**		**Wald Ki-Kare**	**P Değeri**
Sabit		1	0.9801	0.0481	0.8858	10.745	414.68	<.0001
x2	1	1	-0.1558	0.0486	-0.2510	-0.0606	10.29	0.0013
x2	2	1	-0.0540	0.0431	-0.1385	0.0306	1.56	0.2111
x3	1	1	0.4287	0.0379	0.3545	0.5030	128.02	<.0001
x4	1	1	-0.7022	0.1313	-0.9595	-0.4448	28.60	<.0001
x4	2	1	-0.2803	0.0605	-0.3989	-0.1617	21.46	<.0001
x4	3	1	-0.2118	0.0407	-0.2917	-0.1320	5.12	<.0001
x4	4	1	-0.0597	0.0694	-0.1957	0.0763	0.74	0.3893
x4	5	1	-0.1855	0.0578	-0.2987	-0.0723	10.31	0.0013

x2, yaş grubu açıklayıcı değişkeninin birinci seviyesi p değeri 0.05’ den küçük olduğundan %5 anlamlılık seviyesinde model için anlamlı bir katkıya sahiptir. İkinci seviyesi ise 0.05’den büyük olduğundan anlamlı değildir. x3, açıklayıcı değişkeninin birinci seviyesi için p değeri 0.05’den küçük olduğundan modelde anlamlı bir katkıya sahiptir ve x4, açıklayıcı değişkenine bakıldığında bir, iki ve üçüncü seviyeleri için p değerlerinin 0.05’ den küçük olduğu, dolayısıyla modelde anlamlı bir katkıya sahip olduğu fakat dört ve beşinci seviyelerinin p değerlerinin 0.05’den büyük olduğundan anlamlı bir katkıya sahip olmadıkları görülmektedir.

Benzer modellemeyi GLKM’ yi kullanarak da yapabiliriz.

y2 bağımlı değişkeni için x2, x3 ve x4 bağımsız değişkenleri sabit etkili ve x1 bağımsız değişkeni rastgele etkili alınarak Çizelge 4.30’da gösterilen GLIMMIX makrosu kullanımıyla GLKM modelleri uygulanır.

Çizelge 4.30. y2 bağımlı değişkeni için GLIMMIX sürecinde Laplace yönteminin seçimi

```
proc glimmix data=trabzon method=laplace;
class x1 x2 x3 x4 ;
model y2= x2 x3 x4 / dist=poisson link=log solution ;
random intercept  /subject=x1;

run;
```

Çizelge 4.31'de Laplace yöntemi için uyum istatistikleri verilmiştir.

Çizelge 4.31. y2 bağımlı değişkeninin GLKM analizinde Laplace yöntemi için uyum istatistikleri

Uyum İstatistiği	
-2Log Olabilirlik	6421.95
AIC	6441.95
AICC	6442.13
BIC	6441.41
CAIC	6451.41
HQIC	6435.27

Çizelge 4.32'de Laplace yöntemi için şartlı dağılımın uyum istatistikleri verilmiştir.

Çizelge 4.32. y2 bağımlı değişkeninin GLKM analizinde Laplace yöntemi için şartlı dağılımın uyum istatistiği

Şartlı Dağılım için Uyum İstatistikleri	
-2 log L(y2 | r. etki)	6392.57
Pearson Ki-Kare	3744.49
Pearson Ki-Kare / Serbestlik Derecesi	2.91

Çizelge 4.33'de Laplace yöntemi için kovaryans parametre tahminleri verilmiştir.

Çizelge 4.33. y2 bağımlı değişkeninin GLKM analizinde Laplace yöntemi için kovaryans paramatere tahmini

Kovaryans Parametre Tahminleri			
Kovaryans Parametresi	**Grup**	**Tahmin**	**Standart Hata**
Sabit	x1	0.06381	0.03697

Çizelge 4.34'de GLKM analizinde Laplace yöntemi kullanıldığında elde edilen parametre tahminleri verilmektedir.

Çizelge 4.34. y2 bağımlı değişkeninin GLKM analizinde Laplace yönteminde sabit etkiler için çözümleri

Sabit Etkiler için Çözümler								
Etki	**x2**	**x3**	**x4**	**Tahmin**	**Standart Hata**	**Serbestlik Derecesi**	**t Değeri**	**P Değeri**
Sabit				1.0180	0.1085	6	9.38	<.0001
x2	1			-0.1677	0.04882	1271	-3.44	0.0006
x2	2			-0.06315	0.04330	1271	-1.46	0.1450
x3		1		0.3707	0.03839	1271	9.65	<.0001
x4			1	-0.5677	0.1319	1271	-4.30	<.0001
x4			2	-0.2391	0.06077	1271	-3.93	<.0001
x4			3	-0.1873	0.04087	1271	-4.58	<.0001
x4			4	-0.09113	0.06943	1271	-1.31	0.1895
x4			5	-0.2017	0.05811	1271	-3.47	0.0005

x2 bağımsız değişkeninin birinci seviyesi %5 anlamlılık seviyesinde model için anlamlı, ikinci seviyesi anlamlı değildir. x3 bağımsız değişkeninin birinci seviyesi ve x4 bağımsız değişkeninin birinci, ikinci, üçüncü ve beşinci seviyeleri %5 anlamlılık seviyesinde modele katkı sağlar.

GLKM analizinde, Laplace yöntemi için genel değerlendirme Çizelge 4.35'e bakılarak yapılabilir.

Çizelge 4.35. y2 bağımlı değişkeninin GLKM analizinde Laplace yönteminde sabit etkilerin testleri

Sabit Etkilerin Tip-III Testleri				
Etki	**Payın S.derecesi**	**Paydanın S.Derecesi**	**F Değeri**	**P Değeri**
x2	2	1271	6.08	0.0024
x3	1	1271	93.22	<.0001
x4	5	1271	8.41	<.0001

Çizelge 4.35’den, %5 anlamlılık seviyesinde x2, x3 ve x4 açıklayıcı değişkenlerinin modelde anlamlı oldukları görülür.

Aynı bağımlı değişken için GLKM ‘yi Quadrature yöntemi ile çözmede kullanılan SAS makrosu Çizelge 4.36’da verilmiştir.

Çizelge 4.36. y2 bağımlı değişkeni için GLIMMIX sürecinde Quadrature yönteminin seçimi

```
proc glimmix data=trabzon method=quad;
class x1 x2 x3 x4 ;
model y2= x2 x3 x4 / dist=poisson link=log solution ;
random intercept  /subject=x1;

run;
```

Çizelge 4.37’de Quadrature yöntemi için uyum istatistikleri verilmiştir.

Çizelge 4.37. y2 bağımlı değişkeninin GLKM analizinde Quadrature yöntemi için uyum istatistikleri

Uyum İstatistikleri	
-2Log Olabilirlik	6421.95
AIC	6441.95
AICC	6442.13
BIC	6441.41
CAIC	6451.41
HQIC	6435.27

Çizelge 4.38’de Quadrature yöntemi için şartlı dağılımın uyum istatistikleri verilmiştir.

Çizelge 4.38. y2 bağımlı değişkeninin GLKM analizinde Quadrature yöntemi için şartlı dağılımın uyum istatistikleri

Şartlı Dağılım için Uyum İstatistikleri	
-2 log L(y2 | r. etki)	6392.57
Pearson Ki-Kare	3744.49
Pearson Ki-Kare / Serbestlik Derecesi	2.91

Çizelge 4.39'da kovaryans parametre tahminleri verilmiştir.

Çizelge 4.39. y2 bağımlı değişkeninin GLKM analizinde Quadrature yöntemi için kovaryans parametre tahmini

Kovaryans Parametre Tahminleri			
Kovaryans Parametresi	**Grup**	**Tahmin**	**Standart Hata**
Sabit	x1	0.06381	0.03697

Çizelge 4.40'da GLKM analizinde Quadrature yöntemi kullanıldığında elde edilen parametre tahminleri verilmektedir.

Çizelge 4.40. y2 bağımlı değişkeninin GLKM analizinde Quadrature yönteminde sabit etkilerin çözümleri

Sabit Etkiler için Çözümler								
Etki	**x2**	**x3**	**x4**	**Tahmin**	**Standart Hata**	**Serbestlik Derecesi**	**t Değeri**	**P Değeri**
Sabit				1.0180	0.1085	6	9.38	<.0001
x2	1			-0.1677	0.04882	1271	-3.44	0.0006
x2	2			-0.06315	0.04330	1271	-1.46	0.1450
x3		1		0.3707	0.03839	1271	9.65	<.0001
x4			1	-0.5677	0.1319	1271	-4.30	<.0001
x4			2	-0.2391	0.06077	1271	-3.93	<.0001
x4			3	-0.1873	0.04087	1271	-4.58	<.0001
x4			4	-0.09113	0.06943	1271	-1.31	0.1895
x4			5	-0.2017	0.05811	1271	-3.47	0.0005

x2 bağımsız değişkeninin birinci seviyesi %5 anlamlılık seviyesinde model için anlamlı, ikinci seviyesi anlamlı değildir. x3 bağımsız değişkeninin birinci seviyesi ve x4 bağımsız değişkeninin birinci, ikinci, üçüncü ve beşinci seviyeleri %5 anlamlılık seviyesinde modele katkı sağlar.

GLKM analizinde, Quadrature yöntemi için genel değerlendirme Çizelge 4.41'e bakılarak yapılabilir.

Çizelge 4.41.y2 bağımlı değişkeninin GLKM analizinde Quadrature yöntemi için sabit etkilerin testleri

Sabit Etkilerin Tip-III Testleri				
Etki	**Payın S.derecesi**	**Paydanın S.Derecesi**	**F Değeri**	**P Değeri**
x2	2	1271	6.08	0.0024
x3	1	1271	93.22	<.0001
x4	5	1271	8.41	<.0001

Çizelge 4.41'den, %5 anlamlılık seviyesinde x2, x3 ve x4 açıklayıcı değişkenlerinin modelde anlamlı oldukları görülür.
y2 bağımlı değişken için GLKM 'yi RSPL yöntemi ile çözmede kullanılan SAS makrosu Çizelge 4.42'de verilmiştir.

Çizelge 4.42. y2 bağımlı değişkeni için GLIMMIX sürecinde RSPL yönteminin seçimi

```
proc glimmix data=trabzon method=rspl;
class x1 x2 x3 x4 ;
model y2= x2 x3 x4 / dist=poisson link=log solution ;
random intercept  /subject=x1;

run;
```

Çizelge 4.43'de RSPL yöntemi için uyum istatistikleri verilmiştir.

Çizelge 4.43. y2 bağımlı değişkeninin GLKM analizinde RSPL yöntemi için uyum istatistikleri

Uyum İstatistikleri	
-2 Res Log Pseudo-Olabilirlik	4857.75
Genelleştirilmiş Ki-Kare	3751.56
Genelleştirilmiş Ki-Kare / Serbestlik Derecesi	2.94

Çizelge 4.44'de kovaryans parametre tahminleri verilmiştir.

Çizelge 4.44 y2 bağımlı değişkeninin GLKM analizinde RSPL yöntemi için kovaryans parametre tahmini

Kovaryans Parametre Tahminleri			
Kovaryans Parametresi	**Grup**	**Tahmin**	**Standart Hata**
Sabit	x1	0.07546	0.04666

Çizelge 4.45'de GLKM analizinde RSPL yöntemi kullanıldığında elde edilen parametre tahminleri verilmektedir.

Çizelge 4.45. y2 bağımlı değişkeninin GLKM analizinde RSPL yönteminde sabit etkiler için çözümleri

Sabit Etkiler için Çözümler								
Etki	**x2**	**x3**	**x4**	**Tahmin**	**Standart Hata**	**Serbestlik Derecesi**	**t Değeri**	**P Değeri**
Sabit				1.0193	0.1159	6	8.80	0.0001
x2	1			-0.1677	0.04882	1271	-3.44	0.0006
x2	2			-0.06326	0.04330	1271	-1.46	0.1442
x3		1		0.3703	0.03837	1271	9.65	<.0001
x4			1	-0.5665	0.1318	1271	-4.30	<.0001
x4			2	-0.2386	0.06074	1271	-3.93	<.0001
x4			3	-0.1870	0.04086	1271	-4.58	<.0001
x4			4	-0.09145	0.06942	1271	-1.32	0.1880
x4			5	-0.2018	0.05811	1271	-3.47	0.0005

x2 bağımsız değişkeninin birinci seviyesi %5 anlamlılık seviyesinde model için anlamlı, ikinci seviyesi anlamlı değildir. x3 bağımsız değişkeninin birinci seviyesi ve x4 bağımsız değişkeninin birinci, ikinci, üçüncü ve beşinci seviyeleri %5 anlamlılık seviyesinde modele katkı sağlar.

GLKM analizinde, RSPL yöntemi için genel değerlendirme Çizelge 4.46'ya bakılarak yapılabilir.

Çizelge 4.46. y2 bağımlı değişkeninin GLKM analizinde RSPL yönteminde sabit etkiler için testler

Sabit Etkilerin Tip-III Testleri				
Etki	**Payın S.derecesi**	**Paydanın S.Derecesi**	**F Değeri**	**P Değeri**
x2	2	1271	6.08	0.0024
x3	1	1271	93.13	<.0001
x4	5	1271	8.39	<.0001

Çizelge 4.46'dan, %5 anlamlılık seviyesinde x2, x3 ve x4 açıklayıcı değişkenlerinin modelde anlamlı oldukları görülür.

Bütün bu analizlerden sonra model karşılaştırmaları için Çizelge 4.47 ve Çizelge 4.48; parametre tahminlerinde kullanılan tüm modellerin karşılaştırmaları içinde Çizelge 4.49 ve Çizelge 4.50' ye bakılır.

Çizelge 4.47. y1 bağımlı değişkeni için her bir yönteme göre model kriterleri

Y1	AIC	AICC	BIC	CAIC	HQIC	-2 Log Olabilirlik	Generalized Chi-Square
GLM	1713.8950	1714.0361	1760.3287	-	-	-	-
LAPLACE	1714.52	1714.69	1713.98	1723.98	1707.84	1694.52	1279.37
QUADRATURE	1714.52	1714.69	1713.98	1723.98	1707.84	1694.52	1279.37
RSPL	-	-	-	-	-	-	1282.07

Çizelge 4.48. y2 bağımlı değişkeni için her bir yönteme göre model kriterleri

Y2	AIC	AICC	BIC	CAIC	HQIC	-2 Log Olabilirlik	Generalized Chi-Square
GLM	6583.0685	6583.2096	6629.5022	-	-	-	-
LAPLACE	6441.95	6442,13	6441.41	6451.41	6435.27	6421.95	3744.49
QUADRATURE	6441.95	6442.13	6441.41	6451.41	6435.27	6421.95	3744.49
RSPL	-	-	-	-	-	-	3751.56

Çizelge 4.49. İki farklı model için bağımsız değişkenlerin her bir model yöntemine göre p değerleri

				YÖNTEMLER			
		Bağımsız Değişkenler		GLM	LAPLACE	QUADRATURE	RSPL
				P Değerleri	P Değerleri	P Değerleri	P Değerleri
Bağımlı Değişkenler	y1	x2	1	0.8861	0.0264	0.0264	0.8805
			2	0.9936	0.8827	0.8827	0.9745
		x3	1	0.0053	0.9784	0.9784	0.0052
		x4	1	0.0019	0.0052	0.0052	0.0019
			2	0.0001	0.0019	0.0019	0.002
			3	<.0001	0.0002	0.0002	<.0001
			4	0.6070	<.0001	<.0001	0.6317
			5	0.1268	0.6265	0.6265	0.0985
	y2	x2	1	0.0013	0.0006	0.0006	0.0006
			2	0.2111	0.1450	0.1450	0.1442
		x3	1	<.0001	<.0001	<.0001	<.0001
		x4	1	<.0001	<.0001	<.0001	<.0001
			2	<.0001	<.0001	<.0001	<.0001
			3	<.0001	<.0001	<.0001	<.0001
			4	0.3893	0.1895	0.1895	0.1880
			5	0.0013	0.0005	0.0005	0.0005

Çizelge 4.50. İki farklı model için bağımsız değişkenlerin her bir model yöntemine göre genel p değerleri

			YÖNTEMLER		
		Bağımsız Değişken	**LAPLACE**	**QUADRATURE**	**RSPL**
			P Değerleri	**P Değerleri**	**P Değerleri**
Bağımlı Değişkenler	**y1**	**x2**	0.9803	0.9803	0.9788
		x3	0.0052	0.0052	0.0052
		x4	<.0001	<.0001	<.0001
	y2	**x2**	0.0024	0.0024	0.0024
		x3	<.0001	<.0001	<.0001
		x4	<.0001	<.0001	<.0001

5. SONUÇ VE ÖNERİLER

Genelleştirilmiş lineer karma modeller birçok alanda uygulanmaktadır. Özellikle; rastgele seçilmiş bloklarla veya rastgele seçilmiş denemeli bir tasarımdaki deneysel birimlerin modellemesinde kullanılmaktadır. Genelleştirilmiş lineer karma modeller, genelleştirilmiş lineer modellerin, karma modellerle birleştirilmiş halidir. Genelleştirilmiş lineer modellerde, tüm etkiler sabit etki olarak modele katılırken genelleştirilmiş lineer karma modeller de modeldeki genel sabit etkilere rastgele etkilerde eklenmektedir. Bu durum, hem parametre tahminleri açısından hem de uyum iyiliği istatistikleri açısından farklılıklar ortaya koymaktadır.

Bu çalışmada GLKM' ler, Trabzon da yapılan genç nüfusun sorun algılaması çalışmasından elde edilen verilere uygulanmıştır. Bağımlı değişkenler; Trabzon da yaşamaktan mutlu olup olmadıkları şeklinde binary değişken ve günde ortalama kaç saat internet kullandıklarını gösteren Poisson değişkenlerdir. Bu bağımlı değişkenlerle, aylık gelir durumu, yaş, cinsiyet ve eğitim durumu gibi açıklayıcı değişkenler için modellemeler yapıldı. İlk olarak bu açıklayıcı değişkenler sabit etki olarak alındı ve GLM'ler uygulandı. Daha sonra anket örneklem tasarımında yaş grubu, cinsiyet ve eğitim durumu verileri için kota uygulaması yapıldığından bu değişkenler sabit etki; aylık gelir durumu için herhangi bir kota konulmadığından dolayı bu değişken rastgele etki olarak alınarak GLKM uygulanmıştır. Ayrıca GLKM' de tahmin yöntemleri olan; Laplace, Quadrature ve RSPL yöntemleri de bu veri setine uygulanmıştır.

Sonuç olarak, GLM ile GLKM ve GLKM' ler içinde kullanılan farklı tahmin yöntemlerinin karşılaştırmaları yapılmıştır. Parametre tahminleri anlamında GLM ile RSPL yöntemi uygulanarak yapılan GLKM modeli, benzer sonuçlar verirken Laplace ve Quadrature yöntemleri farklı sonuçlar vermektedir. Örneğin; y1 bağımlı değişkeni için x2 açıklayıcı değişkeninin birinci seviyesi, GLM ve RSPL yöntemleri için istatistik olarak anlamsız iken Laplace ve Quadrature yöntemleri anlamlı çıkmaktadır. Model karşılaştırması anlamında Laplace ve Quadrature yöntemlerinin daha iyi sonuçlar verdiği aşikardır. GLKM içinde kullandığımız; Laplace, Quadrature ve RSPL tahmin yöntemlerinin dışında MCMC tahmin yöntemi de bulunmaktadır. Önsel bilgilerin modele katılması olarak ifade ettiğimiz Bayesci yaklaşımlardan olan MCMC yöntemi gelecekteki çalışmalarımızda bize ışık tutacaktır. Ayrıca çok değişkenli GLKM yönteminin kullanımı da hem teorik hem de uygulama açısından çalışmalarımızda önemli yer alacaktır.

6. KAYNAKLAR

Aitken, M., Anderson, D., Francis, B., Hinde, J., 1989. *Statistical modelling in GLIM*, Oxford.

Anderson, R.L., Bancroft, T.A., 1952. *Statistical Theory in Research*, McGraw- Hill, NewYork

Bagiella, E., Sloan, R.P., Heitjan, D.F., 2000. Mixed-Effects Models in Psychophysiology, Psychophysiology, 37, 13-20.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., White, J.S., 2008**.** Generalized linear mixed models: a practical guide for ecology and evolution, Trends in Ecology & Evolution, Volume 24, Issue 3,127-135.

Bozdogan, H., 1987. Model selection and Akaike's Information Criterion (AIC): the general theory and its analytical extensions, Psychometrika, 52, 345-370.

Breslow, N.E., Clayton, D.G., 1993. Approximate Inference in Generalized Linear Mixed Models , Journal of the American Statistical Association, 88, 9−25.

Burnham, K.P., Anderson, D.R., 2002. *Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach,* Springer, Newyork.

Cengiz, M.A., 1997. *Bivariate Logistic Regression Analysis*. Thecnical report, the University of Salford, MCS-97-11.

Cengiz, M.A., 2005. Bayesian inference for bivariate generalized linear models in diagnosing renal arterial obstruction, Statistical Methodology 2, 168-174.

Cengiz, M.A., Percy, D.F., 2001. Mixed multivariate generalized linear models for assessing lower- limb arterial stenoses, Satatistics in Medicine 20, 1663-1679.

Dobson, A.J., 2002. *An Introduction to Generalized Linear Models*, 2nd ed. London: Chapman &Hall

Dunteman, G.H., Ho, Moon-Ho, R., 2006. *An Introduction to Generalized Linear Models,* Sage Publication 145, USA.

Eisenhart, C., 1947. The Assumptions Underlying the Analysis of Variance, Biometrics, 3, 1-21.

Hannan, E.J., Quinn, B.G., 1979. The Determination of The Order of an Autoregression, Journal of the Royal Statistical Society, Series B, 41, 190-195.

Hartley, H.O., Rao, J.N.K., 1967. Maximum Likelihood Estimation for The Mixed Analysis of Variance Model, Biometrika 54, 93-10.

Henderson, C.R., 1950. Estimation of genetic parameters, Ann. Math. Stat. 21, 309-310.

Henderson, C.R., 1953. Estimation of Variance and Covariance Components in Linear Models , Biometrics, Vol. 9, No. 2., 226-252.

Hurvich, C.M., Tsai, C.L., 1989. Regression and Time Series Model Selection in Small Samples, Biometrika 76, 297-307

Işık, F., 2011. Generalized Linear Mixed Models :An Introduction for Tree Breeders and Pathologists, Fourth International Workshop on the Genetics of Host-Parasite Interactions in Forestry, USA.

Jiang, J., 1998. Consistent Estimators in Generalized Linear Mixed Models, Journal of the American Statistical Association, 93, 720−729.

Jiang, J., 2007.*Linear and Generalized Linear Mixed Models and Their Aplications,* Sprinder Science, 257s, Newyork.

Lee, Y., Nelder, J.A., 1996. Hierarchical generalized linear models (with discussion) Journal of the Royal Statistical Society, Series B, 58, 619-678.

Lewis, S.L., 1998. *Analysis of Designed Experiments using Generalized Linear Model,* PhD Thesis, Arizona State University.

Lindsey, J.K., 1997. *Applying Generalized Linear Models*, Springer Verlag Newyork, 256s.

Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., 2005. *SAS System for Mixed Models,* SAS Institute Inc., Cary, NC, USA.

McCullagh, P., Nelder, J.A., 1989. *Generalized Linear Models,* Chapman & Hall, 2nd ed.

McCulloch, C.E., Searle, S.R., 2001. *Generalized Linear and Mixed Effects,* John Wiley& Sons, Inc., America, 325s.

Mcgilchrist, C.A., 1994. Estimation in Generalised Mixed Models. J. Roy. Statist. Soc. B., 56: 61-69.

Murat, N., Cengiz, M.A., Terzi, Y., 2009. Genç Nüfusun Sorun Algılaması: Trabzon Örneği, Journal of International Social Research, Volume 2, Issue 7, 175-184.

Myers, R.H., Montgomery, D. C., Vining, G.G., 2001. *Generalized Linear Models with Applications in Engineering and the Sciences* John Willey & Sons, NewYork.

Nelder, J.A., Wedderburn, R.W.M., 1972. Generalized Linear Models, Journal of the Royal Statistical Society,Series A, 135, 370-384.

Patterson, H.D., Thompson, R., 1971. Recovery of İnter-Block İnformation When Block Sizes are Unequal, Biometrika 58, 545-554.

Pinheiro, J.C., Bates, D.M., 2000. *Mixed Effects Models in S and S-Plus,* (Editörler: J. Chambers, L.Tierney), Springer Science, 528s, Newyork.

Sahai, H., Khuri, A.I., 1985. A Second Bibliography on variance components. Taylor & Francis, 14, 63-115.

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, 461464.

Searle, S.R., 1971. *Linear Models*. New York: John Wiley & Sons.

Searle, S.R., Casella, G., Mc Culloch, C.E., 1992. *Variance Components,* JohnWiley & Sons Inc., USA.

Townsend, E.C., 1968. Best Unbiased Estimation of Variance Components from Unbalanced Data in the 1-Way Classification. Biometrics Unit, Cornell University, 27, 643-557.

Uusipaikka, E., 2000. *Confidence intervals in generalized regressions models*, CRC Press.

Verbeke, G., Molenberghs, G., 2000.*Linear Mixed Models for Longitidunial Data,* Springer Science, 568s, Newyork.

Verbeke, G., Molenberghs, G., 2005. *Models for Discrete Longidutinal Data,* Springer, 683s, Newyork.

West, B.T., Welch, K.B., Galecki, A.T., 2007. *Linear Mixed Models:A Practical Guide Using Statistical Software,* Chapman& Hall/ CRC, 339s, Newyork.

Wolfinger, R.D., Tobias, R.D., Sall, J., 1991. *Mixed Models: A Future Direction*, Proceedings of the Sixteenth Annual SAS Users Group Conference, SAS Institute Inc., Cary, NC, 1380–1388.

Wu, L., 2010. *Mixed Effects Models for Complex Data,* Chapman& Hall/ CRC,419s, America.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., Smith, G.M., 2009.*Mixed Effect Models and Extensions in Ecology with R,* Springer Science, 574s, Newyork.

7. ÖZGEÇMİŞ

Adı Soyadı: Tuba KOÇ

Doğum Yeri: Bafra / SAMSUN

Doğum Tarihi: 11.12.1986

Medeni Hali: Evli

Bildiği Yabancı Diller: İngilizce

Lise: Bafra Anadolu Lisesi (2001-2005)

Lisans: Ondokuz Mayıs Üniversitesi (2005-2009)

Yüksek Lisans: Ondokuz Mayıs Üniverstesi (2009-2012)

Çalıştığı Kurum: Fen Bilimleri Enstitüsü (2011-...)

İletişim Bilgileri: Ondokuz Mayıs Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü SAMSUN
E-mail: tuba.koc@omu.edu.tr