

**T.C.
HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

**GENELLEŞTİRİLMİŞ EŞİTLİK
KESTİRİMİ (“GEE”)**

Anıl AKTAŞ

**Biyoistatistik Programı
YÜKSEK LİSANS TEZİ**

ANKARA

2005

**T.C.
HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

**GENELLEŞTİRİLMİŞ EŞİTLİK
KESTİRİMİ (“GEE”)**

Anıl AKTAŞ

**Biyoistatistik Programı
YÜKSEK LİSANS TEZİ**

**TEZ DANIŞMANI
Doç. Dr. Osman SARAÇBAŞI**

**ANKARA
2005**

TEŞEKKÜR

Çalışmamın başlangıcından sonuna kadar desteğini ve yardımını esirgemeyen, her konuda bilgi birikimiyle katkıda bulunan ve yol gösteren, öğrencisi olduğum süreç içinde kendisi ile çalışmaktan ve konuşmaktan zevk aldığım çok değerli danışman hocam Sayın Doç. Dr. Osman Saraçbaşı'na,

Yüksek lisans eğitimime başladığım andan itibaren her konuda destek olan ve benden hiçbir zaman yardımlarını esirgemeyen, bilimsel çalışmalarda, ortamlarda bulunmamız için bizi yönlendiren değerli hocam ve bölüm başkanımız Sayın Prof. Dr. Ergun Karaağaoğlu'na,

Her zaman öğrencisinin yanında olan ve destek veren hocam Sayın Prof. Dr. Reha Alpar'a,

Yardıma ihtiyaç duyduğum her konuda benden bilgilerini esirgemeyen Dr. Erdem Karabulut'a, Dr. Pınar Özdemir Geyik'e, birlikte olmaktan ve çalışmaktan mutluluk duyduğum adaşım Uzm. Anıl Barak'a, Uzm. Jale Karakaya'ya, Uzm. Umut Arslan'a, Uzm. Harika Gözde Gözükara Otlı'ya, Uzm. Sevilay Karahan'a, Uzm. Kerem Uludağ'a ve GATA'lı ağabeylerime, arkadaşlığı ve anlayışı için Selcen Pehlivan'a, bir abla gibi her zaman bana yardımcı olan bölüm sekreterimiz Menekşe Tarla'ya,

Çalışmamın uygulama kısmı için kendi çalışmasının verilerini kullanmamama izin veren Uzm. Çiğdem Civelek'e,

Her zaman yanımda olduğunu bildiğim sevgili büyüklerime ve tüm arkadaşlarıma,

Son olarak da, her zaman bana inanan, maddi ve manevi her konuda bana destek veren anneme, babama ve şu anda yanımda olamasa da desteğini benden hiç esirgemeyen canım kardeşime,

SONSUZ TEŞEKKÜRLER

ÖZET

Aktaş, A. Genelleştirilmiş Eşitlik Kestirimi (“GEE”). Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Programı Yüksek Lisans Tezi, Ankara, 2005.

Sağlık alanında en çok rastlanan durumlardan biri de bir denekten belli aralıklarla tekrarlı ölçümlerin alınmasıdır. Bu tekrarlı ölçümlerin analizlerinde farklı yöntemler kullanılmaktadır. Bunlardan biri de Genelleştirilmiş Eşitlik Kestirimidir (GEK).

Bir denekten farklı zamanlarda alınan ölçümler arasında ilişki olduğu düşünüldüğünde Genelleştirilmiş Eşitlik Kestirimi kullanılmaktadır. Bir denekten elde edilen tekrarlı ölçümlerin arasındaki ilişki, ilişki matrisi ile belirlenir. GEK, yanıt değişkenin ortak dağılımı ile ilgilenmez. Ölçümlerin sürekli olmasının yanı sıra genellikle sayılabilir ya da kategorik olması durumunda kullanılır. Bu çalışmada, bağımlı değişkenin tekrarlı ölçümlerden oluştuğu GEK çözümlemesi için, Hacettepe Üniversitesi Diş Hekimliği Fakültesi Diş Hastalıkları ve Tedavisi Anabilim Dalına ait veriler kullanılacaktır. Kliniğe gelen hastaların dişlerindeki renk uyumunu etkileyen değişkenler belirlenmiş, bu değişkenlere göre uygun ilişki yapısı bulunmuş ve GEK çözümlemesi uygulanmıştır. Ayrıca verilere Tekrarlı Ölçümlerde Varyans Çözümlemesi de uygulanmıştır. Çalışmada Varyans Çözümlemesinin gerekli varsayımları sağlanmadığından GEK çözümlemesinin sonuçlarının daha doğru olduğu sonucuna varılmıştır.

Anahtar kelimeler: İlişkili Veriler, Tekrarlı Ölçümler, İlişki Matrisi, Bağlantı Fonksiyonu, GEK.

ABSTRACT

Aktaş, A. Generalized Estimating Equations (“GEE”). Hacettepe University Institute of Health Sciences, Master Thesis in Biostatistics, Ankara, 2005.

One of the most utilized measures in health studies cover the usage of repeated measurements taken on same subjects. There are different methods developed for analysis of repeated measures. One of these methods is the Generalized Estimating Equation (GEE).

GEE is generally used when it is thought there is a relationship between repeated measurements of the same subject. This relationship is determined by “Working” Correlation Matrix. In GEE, no assumptions on the joint distribution of the subject’s response vector exist. GEE method is especially used for categorical and count outcomes as well as continuous outcomes. In this study, data was taken from Hacettepe University Faculty of Dentistry. In the first stage of the study, the variables of interest were determined and Analysis of Variance for Repeated Measures method was applied. In the second stage, different correlation structures were calculated for Working Correlation Matrix and after the selection of suitable correlation structure, GEE method was applied. At the end of the study, it has been concluded that the results obtained by GEE is more accurate than that of Analysis of Variance because of the violation of the assumptions.

Key words: Correlated Data, Repeated Measures, Working Correlation Matrix, Link Function, GEE

İÇİNDEKİLER

	Sayfa
ONAY SAYFASI	iii
TEŞEKKÜR	iv
ÖZET	v
ABSTRACT	vi
İÇİNDEKİLER	vii
SİMGELER VE KISALTMALAR DİZİNİ	ix
ŞEKİLLER DİZİNİ	x
TABLOLAR DİZİNİ	xi
1.GİRİŞ	1
1.1. En Küçük Kareler Kestirimi	3
1.2. En Çok Olabilirlik Kestirimi	4
1.3. Yarı - Olabilirlik Kestirimi	4
1.3.1. Yarı - Olabilirlik Fonksiyonu	5
1.4. EM Algoritması	6
1.4.1. Algoritma Yapısı	6
1.5. Bağımsız Eşitlik Kestirimleri	7
2.GENELLEŞTİRİLMİŞ DOĞRUSAL MODELLER	9
2.1. Model Yapısı	9
2.2. Bağntı Fonksiyonu	11
2.3. Uyum İyiği Testi	12
2.4. Genelleştirilmiş Doğrusal Model Fonksiyonları	13
2.5. Kestirimlerin Varyansının Elde Edilmesi	14
2.6. Boylamsal Veriler	16
2.7. Panel Veri	16
2.7.1. Bileşik Kestirim	17
2.7.2.Sabit Etkili ve Rastgele Etkili Modeller	17
2.7.2.1. Koşullu Olmayan Sabit Etkili Modeller	17
2.7.2.2. Koşullu Sabit Etkili Modeller	18
2.7.3. Rastgele Etkili Modeller	18
2.7.4. Kitle Ortalamalı ve Deneğe Özgü Modeller	19

İÇİNDEKİLER (DEVAM)

	Sayfa
3. GENELLEŞTİRİLMİŞ EŞİTLİK KESTİRİMİ	20
3.1. Genelleştirilmiş Eşitlik Kestirim Yaklaşımı	21
3.2. İlişki Matrisi	22
3.2.1. Bağımsız İlişki	22
3.2.2. Değiştirilebilir İlişki	23
3.2.3. Yapısal Olmayan İlişki	24
3.2.4. Sabit İlişki	24
3.2.6. Otoregresif İlişki	25
3.2.7. Durağan İlişki	25
3.2.8. Durağan Olmayan İlişki	26
3.3. En İyi İlişki Yapısını Belirleme	27
3.4. Yayılım Parametresi	28
3.5. Algoritma	29
3.6. Kovaryans Parametresinin Kestirimi	29
3.7. Değişen Lojistik Regresyon	31
3.8. Kayıp Veri	33
4. GEREÇ ve YÖNTEM	34
5. BULGULAR	38
5.1. Tanımlayıcı İstatistikler	39
5.2. Tekrarlı Ölçümlerde Varyans Çözümlemesi	40
5.3. Genelleştirilmiş Eşitlik Kestirimi	41
6. SONUÇ	45
KAYNAKLAR	48

SİMGELER VE KISALTMALAR

AIC	Akaike Bilgi Ölçütü
ALR	Alternatif Lojistik Regresyon
AR(1)	Otoregresif İlişki
BEK	Bağımsız Eşitlik Kestirimi
D	Sapma İstatistiği
EK	Eşitlik Kestirimi
EM	Expectation Maximization Algoritması
$g(.)$	Bağıntı fonksiyonu
GDM	Genelleştirilmiş Doğrusal Modeller
GEK	Genelleştirilmiş Eşitlik Kestirimi
$l(.)$	Olabilirlik Fonksiyonu
L	Log-Olabilirlik
MAR	Rasgele Kayıp Veri
MCAR	Tümüyle Rasgele Kayıp Veri
OR	Odds Oranı
$R(.)$	İlişki Çalışma Matrisi
QIC	Bağımsız Model Altında Yarı - Olabilirlik Bilgi Ölçütü
SBEO	Sınırlı Bilgiye Dayalı En çok Olabilirlik
TBEO	Tam Bilgiye Dayalı En çok Olabilirlik
$\hat{V}(.)$	Varyans Kestirimi
$\hat{V}_H(.)$	Sandwich Bilgi Matrisi
ϕ	Yayılım Parametresi
Σ_e	Deneysel Temelli Kestirici
Σ_i	Kovaryans Matrisi
Σ_m	Model Temelli Kestirici
χ^2	Pearson Ki-kare İstatistiği

ŞEKİLLER**Sayfa**

5.1. Hastaların Aylara Göre Renk Uyumunu Gösteren Grafik

40

TABLOLAR

	Sayfa
4.1 Çalışmada Kullanılan Değişkenler	35
4.2 Genelleştirilmiş Eşitlik Kestirimi Çözümlemesi İçin SAS Programında Komutlar	36
4.3 İlişki Tipinin Belirlenmesine İlişkin STATA Algoritma Yapısı	37
5.1 Yaş Değişkenine İlişkin Tanımlayıcı İstatistikler	38
5.2 Bağımlı Değişkene İlişkin Tanımlayıcı İstatistikler	38
5.3 Hastaların Dişlerinin Renk Uyumu, Yaş ve Cinsiyet Değişkenlerine İlişkin Veriler	39
5.4 Tekrarlı Ölçümlerde Varyans Çözümlemesine İlişkin Sonuçlar	40
5.5 Tekrarlı Ölçümlerde Varyans Çözümlemesi İçin Parametre Kestirim Değerleri	41
5.6 En İyi İlişki Yapısına İlişkin Sonuçlar	41
5.7 Uyum İyiliği Sonuçları	42
5.8 İlk Parametre Kestirim Değerleri	42
5.9 Değiştirilebilir İlişki Yapısı Altında İlişki Matrisi	42
5.10 Bağımsız İlişki Yapısı Altında İlişki Matrisi	43
5.11 Değiştirilebilir İlişki Yapısına Göre Model-Temelli Kovaryans Matrisi	43
5.12 Değiştirilebilir İlişki Yapısına Göre Deneysel Kovaryans Matrisi	43
5.13 Bağımsız İlişki Yapısına Göre Model-Temelli Kovaryans Matrisi	44
5.14 Bağımsız İlişki Yapısına Göre Deneysel Kovaryans Matrisi	44
5.15 Değiştirilebilir ve Bağımsız İlişki Yapısı İçin Genelleştirilmiş Eşitlik Kestirimi Çözümlmeleri	44

TABLÖLAR(DEVAM)

	Sayfa
6.1 Tekrarlı Ölçümlerde Varyans Çözümlemesinden ve Genelleştirilmiş Eşitlik Kestirim Çözümlemesinden Elde Edilen Parametre Kestirim Değerleri	45
6.2 Tekrarlı Ölçümlerde Varyans Çözümlemesinden Kestirilen Değerler	46
6.3 Varyans Çözümlemesi ve Genelleştirilmiş Eşitlik Kestirimi ile Elde Edilen Ortalama Değerlerinin Karşılaştırılması	47

BİRİNCİ BÖLÜM

1. GİRİŞ

Verilerin çözümlenmesi sırasında en önemli durum, değişkenlerin bağımlılığı ya da bağımsızlığıdır. Sağlık alanında en çok rastlanılan durumlardan biri, bir denekten belli zaman aralıklarında ölçümler alınmasıdır. Bu ölçümler, bağımlı ölçümlerdir. Çünkü tekrarlı ölçümler arasında bir ilişki bulunmaktadır.

Bir denekten alınan tekrarlı ölçümlerin ilişkili olduğu düşünüldüğünde, tekrarlı ölçümlere (verilere) özel yöntemlerden yararlanılır.

Genelleştirilmiş Eşitlik Kestirimi (GEK), bağımlı değişkenin tekrarlı ölçümlerden oluşması ve *gözlemlerin dağılımının bilinmediği* durumlarda kullanılır. GEK, ilişkili verilerde Genelleştirilmiş Doğrusal Modellerin (GDM) uzantısıdır. GEK, daha çok kategorik ya da sayılabilir verilerde kullanılmasına rağmen, sürekli verilerde de kullanılmaktadır (29).

GEK, bağımlı değişkenin ortak dağılımı ile ilgilenmez. Onun yerine zaman aralığındaki tekrarlı ölçümlerin marjinal dağılımını kullanır. GEK, GDM'in bir uzantısı olduğundan, GEK çözümlemesinde öncelikle doğrusal kestirici belirlenir sonra bir bağıntı fonksiyonu kullanılır. Örneğin varyans, ortalamanın bir fonksiyonu olarak tanımlanır. GEK'nin diğer eşitlik kestirim yöntemlerinden farklılığı, tekrarlı ölçümler arası ilişki yapısını (*Working Correlation Matrix*) göz önüne almasıdır. İlişki yapısı bu nedenle önem taşımaktadır. İlişki yapısının doğrudan kestirimi mümkün değildir. İlişkinin yapısı tam olarak belirlenemese de GEK etkin bir kestirime olanak sağlar (29).

Farklı şekillerde ilişki yapıları bulunmaktadır. Kestirimler için bu ilişki yapılarının bilinmesi önemlidir. Doğru ilişki yapısının belirlenmesi farklı şekillerde olabilmektedir.

GEK, ülkemizde çok yaygın olarak kullanılmamakla beraber, yurtdışında birçok araştırmada tercih edilen bir yöntemdir. Yalnızca sağlık alanında değil sosyal bilimler ve ekonomi alanlarında da kullanılmaktadır.

Araştırmacılar sıklıkla uzunlamasına (uzun dönemli) ya da kümelenmiş verilerle ilgili çözümlemelerle ilgilenmektedirler. Bu çözümlemelerde parametre kestirimleri olarak “*En Küçük Kareler Kestirimi*”, “*Yarı-olabilirlik*”, “*EM algoritması*”, “*En Çok Olabilirlik Kestirimi*”, “*Bağımsız Eşitlik Kestirimi*” ve “*Genelleştirilmiş Eşitlik Kestirimi*” yöntemlerinden yararlanılmaktadır.

Bu kestirim yöntemlerini kullanabilmek için bazı varsayımların sağlanması gereklidir. Bu varsayımlar genellikle dağılım üzerinedir. En küçük kareler yöntemi, yarı-olabilirlik, en çok olabilirlik kestirimi, EM algoritması için gözlemlerin dağılımlarının bilinmesi gerekir. Bağımsız eşitlik kestiriminde koşul, küme içi deneklerin bağımsız olmasıdır. GEK çözümlemesinde bu varsayımlara gerek duyulmaz.

Bazı uzun dönemli çalışmalarda, gözlemler arasında ilişkinin olduğu tekrarlı veriler kullanılır. Eğer veriler yaklaşık olarak normal dağılım gösteriyorlarsa, yukarıda adı geçen yöntemler oldukça yaygın kullanılmaktadır. Eğer veri tipi ikili yani “*binary*” ya da sayılabilir ise farklı yaklaşımlar kullanılabilir (1).

GEK’in incelendiği bu çalışmanın ilk bölümünde farklı parametre kestirimleri açıklanacak, ikinci bölümde, genelleştirilmiş doğrusal modellerin içeriği hakkında bilgi verilecek, üçüncü bölümde genelleştirilmiş eşitlik kestirimi açıklanacaktır. Dördüncü bölümde GEK için, sağlık alanında kullanılan veri yapısı hakkında bilgi verilecek, beşinci bölümde de verinin tanımlayıcı istatistikleri, gözlemlere ilişkin ilişki yapısı ve model için elde edilen katsayılar verilecektir. İlişki yapısı, bağımsız model altında yarı-olabilirlik bilgi ölçütü (QIC: *quasilikelihood under the indepedence model information criterion*) kullanılarak belirlenecek ve ilişki matrisinin elde edilmesiyle gerekli değerlere ulaşılabacaktır. Son bölümde sonuçlara ve tartışmalara yer verilecektir.

1.1. En Küçük Kareler Kestirimi

Parametre kestirim yöntemlerinden biri en küçük kareler kestirimidir. Regresyon çözümlmeleri için en küçük kareler kestirimi aşağıdaki şekildedir:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2) \quad (1.1)$$

$$D(\beta) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1.2)$$

Burada parametre vektörü $\beta = (\beta_0, \beta_1)'$ için ortalama aşağıdaki gibidir:

$$\hat{\mu}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (1.3)$$

En Küçük Kareler kestiriminde amaç, kareler toplamının en küçük olmasıdır. $D(\beta)$ 'nin en küçüklenmesi için gerekli adımlar aşağıdadır.

1. $\frac{\partial}{\partial \beta_0} D(\beta)$ ve $\frac{\partial}{\partial \beta_1} D(\beta)$ türevleri hesaplanır.
2. Bu eşitlikler sıfıra eşitlenir.
3. β_0 ve β_1 çözülür. Elde edilen kestirimler aşağıdadır.

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2} \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x}. \\ \hat{\mu}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i \end{aligned} \quad (1.4)$$

1.4'deki eşitlikleri farklı bir ifadeyle gösterirsek 1.5 eşitlikleri elde edilir.

$$\begin{aligned} \psi_0(\beta_0, \beta_1; y) &= \sum_i (y_i - (\beta_0 + \beta_1 x_i)) \\ \psi_1(\beta_0, \beta_1; y) &= \sum_i (y_i - (\beta_0 + \beta_1 x_i)) \cdot x_i \end{aligned} \quad (1.5)$$

$$\frac{\partial}{\partial \beta_0} D(\beta) = \frac{\partial}{\partial \beta_1} D(\beta) = 0 \quad \text{eşitliğinin çözümü ile } \psi_0(\beta_0, \beta_1; y) = 0 \quad \text{ve}$$

$\psi_1(\beta_0, \beta_1; y) = 0$ eşitliklerinin çözümü, β_0, β_1 katsayıları için aynı çözümü verir. ψ_0 ve ψ_1 eşitlik fonksiyonları olarak adlandırılır ve $\psi_0 = 0, \psi_1 = 0$ ile katsayılar kestirilir (3). Ancak modeli oluşturabilmek için normal dağılım koşulunun sağlanması gereklidir (3).

1.2. En Çok Olabilirlik Kestirimi

Parametre kestirim yöntemlerinden biri de en çok olabilirlik kestirimidir.

$X_1, \dots, X_n$, n tane rastgele değişken olmak üzere $f(\underline{x}, \theta)$, rastgele değişkenlerin yoğunluk fonksiyonu olsun. Rastgele değişkenler gözlendiğinde, θ 'nın bir fonksiyonu olan ortak yoğunluk fonksiyonu, olabilirlik fonksiyonu olarak adlandırılır ve $l(\theta, \underline{x})$ ile gösterilir. $\underline{x}$ vektörü verildiğinde; $l(\theta, \underline{x})$ 'nin en büyükleme için olabilirliğin logaritmasının $(L(\theta, \underline{x}))$ türevi sıfıra eşitlenir ve θ için çözüm elde edilir. En çok olabilirlik kestirim yöntemi için değişkenlerin dağılımının bilinmesi gerekir. Bu dağılımlar; Poisson, Binom, çok değişkenli dağılımlar olabilir. Üstel bir aile için, $[E(Y)=\mu]; \beta_1, \dots, \beta_p$ parametrelerine bağlı ise herhangi bir β parametresi için en çok olabilirlik kestirimi eşitliği Eşitlik 1.6'da verildiği gibidir (10):

$$\frac{\partial L_i}{\partial \beta_j} = \sum_{i=1}^n x_{ij} \frac{(y_i - \mu_i)}{\phi V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0 \quad (1.6)$$

1.3. Yarı - Olabilirlik Kestirimi “(Quasi-Likelihood)”

Bağımlı değişken Y 'nin dağılımı tam olarak belirlenemediğinde yalnızca ortalama ve varyans modellenemediğinde β 'nin kestirimi tutarlılık ve asimptotik normal özelliklerine sahip ise bu kestirim yöntemine yarı-olabilirlik denir (3).

1.3.1 Yarı-Olabilirlik Fonksiyonu:

$Z_1, \dots, Z_n; \mu_i$ beklenen değere ve $\phi V(\mu_i)$ varyansa sahip olsun. Burada V bilinen fonksiyon olmak üzere her bir gözlem için yarı-olabilirlik fonksiyonu $K(z_i; \mu_i)$ ile tanımlanırsa aşağıdaki eşitlik elde edilir:

$$\frac{\partial K(z_i, \mu_i)}{\partial \mu_i} = \frac{z_i - \mu_i}{\phi V(\mu_i)} \quad (1.7)$$

Kestirim yöntemlerinden biri olan yarı-olabilirlik fonksiyonu olabilirlik fonksiyonu ile benzerlik gösterir. $\mu; \beta_1, \dots, \beta_p$ parametrelerin bir fonksiyonu olsun. Bu durumda kestirimler aşağıdaki eşitlikler yardımıyla elde edilir:

$$E\left(\frac{\partial K}{\partial \beta_j}\right) = 0 \quad (1.8)$$

$$\hat{V}_H(\hat{\beta}) = E\left(\frac{\partial K}{\partial \beta_j} \frac{\partial K}{\partial \beta_k}\right) = -E\left(\frac{\partial^2 K}{\partial \beta_j \partial \beta_k}\right) = \frac{1}{\phi V(\mu)} \frac{\partial \mu}{\partial \beta_j} \frac{\partial \mu}{\partial \beta_k} \quad (1.9)$$

Bu iki eşitlik Bilgi “*Information*” matrisinin elemanlarıdır. Eşitlik kestirimi, en çok olabilirlik eşitlik kestirimlerine benzerlik gösterir:

$$\frac{\partial L_i}{\partial \beta_j} = \sum_{i=1}^n x_{ij} \frac{(y_i - \mu_i)}{\phi V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0 \quad (1.10)$$

Varyans, “***Sandwich***” kestirimi kullanılarak kestirilir ve aşağıdaki eşitliklerle ifade edilir (10):

$$B = \sum_{i=1}^n \frac{\partial \mu'_i}{\partial \beta} V_i^{-1} \text{Cov}(Y_i) V_i^{-1} \frac{\partial \mu_i}{\partial \beta} \quad (1.11)$$

$$V_s(\hat{\beta}) = V_H^{-1}(\hat{\beta}) B(\hat{\beta}) V_H^{-1}(\hat{\beta}) \quad (1.12)$$

Yarı-olabilirlik kestirimi, değişkenliğin fazla olduğu ve ortalamanın küçük olduğu durumlarda çok iyi çalışır. İlişkili veriler için de ikinci dereceden bir en iyileme sağlamaktadır (10).

1.4. EM Algoritması

Çözümlerler sırasında sıkça karşılaşılan bir durum eksik gözlemlerin varlığıdır. Eğer gözlemler tümüyle rastgele olarak eksik (MCAR) ise ortalama için yansız kestirim yöntemleri tanımlanmıştır (1, 14). Bu yaklaşımlardan biri de “EXPECTATION MAXIMIZATION” (EM) algoritmasıdır.

EM algoritması, veriler tamamlanmamış ya da eksik gözlemlere sahipse verilen veri setinin dağılımını oluşturan parametrelerinin, en çok olabilirlik kestirimlerini bulmada kullanılan genel bir yöntemdir ve oldukça ayrıntılı bir tekniktir (23, 24). EM algoritmasının hem üstün hem de eksik yönleri bulunmaktadır. Bunlar sırasıyla aşağıda belirtilmiştir:

- Sayısal olarak süreklilik sağlar. Her döngüde gözlemlenmiş verilerin olabilirliği artar.
- Zor da olsa parametrelerin bulunmasında yardımcı olmaktadır.
- En iyi noktaya yaklaştıkça algoritma çok yavaşlamaktadır.
- Bazı durumlarda eyer (semer) noktasına yaklaşım olabilmektedir.

1.4.1. Algoritma Yapısı:

$X=(Y,Z)$ için, olasılık model $f(Y,Z/\theta)$ olsun ve yalnızca Y değeri gözlemlensin. Bu durumda olabilirlik fonksiyonu aşağıdaki şekildedir:

$$l(\theta)=f(y/\theta)=\int f(y,z/\theta) dz \quad (1.13)$$

Amaç, $l(\theta)$ değerini en büyük yapmaktır. EM algoritması, olabilirliği en büyük yapan bir yöntemdir. EM algoritmasının E- adım ve M-adım olmak üzere iki adımı bulunmaktadır (11, 19). Bu adımlar aşağıdaki gibi sıralanmaktadır:

İlk olarak bir başlangıç noktası θ_0 belirlenir ve $i=0,1,2,\dots$, olmak üzere E-adımı;

$$\begin{aligned} Q(\theta/\theta_i) &= \int \log f(z,y/\theta) f(z/\theta_i,y) dz \\ &= E[\log(Z,Y/\theta) / Y, \theta_i] \end{aligned} \quad (1.14)$$

fonksiyonu elde edilir.

M-adımı;

$Q(\theta/\theta_i)$ fonksiyonu θ' 'ya göre en büyük yapılır. En büyük θ_{i+1} belirlenir

Bu adımlar tek bir noktaya ulaşınca kadar devam eder.

E-adım da amaç, sabit parametre değerleri ve gözlenen Y değeri verildiğinde tam verinin (X) yeterli istatistiklerini kestirmektir. M-adım da ise amaç, bu istatistikleri en büyük yapmaktır (23, 24). Bu iki adımın tekrarlanması her döngüde olabilirliğin artmasını ve olabilirlik fonksiyonun en büyüğe yaklaşmasını sağlar (11).

EM algoritması, bağımlı değişken y 'nin eksik gözlenmesi ve bağımsız değişken X 'in tam olarak gözlenmesi durumunda uygulanabilir. Normal dağılım gösteren veriler için EM algoritması, eğer veriler rastgele olarak eksikse (MAR) tutarlı parametre kestirimleri oluşturmada kullanılabilir (11, 19).

1.5. Bağımsız Eşitlik Kestirimleri

Bu kestirim yöntemi, genellikle kontrollü denemelerde tercih edilir. Geniş aralıklı modellerde ve eşit küme genişlikleri için kullanılan bağımsız eşitlik kestirimleri (BEK), aşağıdaki koşulların sağlanması durumunda etkilidir:

1. Küme içindeki yanıt değişkenleri bağımsız olmalı
2. Küme içindeki tüm ortak değişkenler sabit olmalı
3. Tüm ortak değişkenler ortalama dengeli olmalı

örneğin; p küme ortalaması aşağıdaki gibidir:

$$\frac{1}{T_i} \sum_{i=1}^{T_i} x_{ij}, j = 1, \dots, p \quad (1.15)$$

4. Ortak değişkenler, küme içindeki sabit olan ortak değişkenlerin ve ortalama dengeli ortak değişkenlerin bir karışımı olmalıdır (2).

BEK, en çok klinik denemelerde uygulanmaktadır. Paralel grup düzenlerinde deneğin tekrarlı ölçümlerinin bulunduğu küme için tedavi etkileri sabittir, çapraz düzenli çalışmalarda ise tedavi etkileri kümede çeşitlidir ancak ortalama dengelidir. 2x2 çapraz düzenlerde başlangıç (*baseline*) ölçümler, ilk tedavi periyodundan sonra bağımlı değişkenler için ortak değişken gibi alınabilir, ancak başlangıç ölçümler kümede sabittir. Paralel düzenlerde ise x değişkenleri yalnızca başlangıç değerleri ve tedavi etkilerini içerdiğinden kümede sabittir.

Diğer taraftan, çapraz düzenlerde taşıyıcı etki bulunabileceğinden yanıt değişkeni y sadece asıl etkiye değil ilk önceki tedaviye de bağımlıdır. Böyle bir durumda koşulların ilki çığnemiş olmaktadır. Bu gibi durumlarda, varsayımların sağlanmaması, BEK kullanılmasını engellemektedir (2).

İKİNCİ BÖLÜM

2. GENELLEŞTİRİLMİŞ DOĞRUSAL MODELLER

GDM, ilişkili verilerin modellenmesinde kullanılan bir algoritmadır. Olabilirlik temelli model olan GDM, bireylerin ya da gözlemlerin bağımsızlığı varsayımını temel alır. GDM'yi tanımlamak için gerekli birkaç ifade aşağıda belirtilmiştir (17,32):

- Bağımlı değişkenin dağılımı üstel aileden olmalı
- Bağlantı fonksiyonu tanımlanmalı
- Bağımsız değişkenler belirtilmelidir.

2.1. Model Yapısı

Genel olarak doğrusal yapı modeli Eşitlik 2.1'de verildiği gibidir:

$$y_i = x_i' \beta + \varepsilon_i \quad (2.1)$$

y_i ; i -inci gözlem için yanıt değişkenini, x_i ; i -inci gözlem için ortak değişkenlerin sütun vektörü ya da açıklayıcı değişkeni ifade etmektedir. Bilinmeyen katsayı vektörü β ; y verisine uygun olarak en küçük kareler yöntemi ile kestirilir. ε_i ; bağımsız olduğu varsayımı altında sıfır ortalamalı, sabit varyanslı normal dağılımlı rastgele değişkendir.

y_i 'nin beklenen değeri μ_i ile gösterilirse, doğrusal ifade Eşitlik 2.2'deki gibi tanımlanır (32).

$$\mu_i = x_i' \beta \quad (2.2)$$

Genel doğrusal modeller, istatistiksel veri çözümlemelerinde kullanılırken, doğrusal modeller ($\mu_i = x_i' \beta$) bazı sınırlamalar altında kalır (13). Bu durumlardan birkaçı aşağıda belirtilmiştir:

1. Verilerin normal dağılım varsayımı sağlanmayabilir. Örneğin, Normal dağılım varsayımı, gözlemlerin kesikli olması durumunda geçerli değildir.
2. Verilerin ortalaması sınırsızdır ancak bazı durumlarda örneğin; sağ-kalım ya da sayılabilir veriler için ortalama, negatif değer almazken, oranlar ya da ikili veriler için ortalama $[0,1]$ birim aralığında değer alabilir.
3. Tüm gözlemler için verilerin varyansının sabit olduğu varsayımı gerçekçi olmayabilir (13).

GDM, üstel aileden gelen tek değişkenli veriler için regresyon modeli oluşturmada kullanılan standart bir yöntem olarak tanımlanmıştır (5). GDM'deki bağımlı değişken Y için olasılık yoğunluk fonksiyonu:

$$f_y(y, \theta, \phi) = \exp\{(y\theta - b(\theta))/a(\phi) + c(y, \phi)\} \quad (2.3)$$

şeklinde ifade edilmektedir. $a(\cdot)$, $b(\cdot)$ ve $c(\cdot)$ verilmiş fonksiyonlar, θ ; yer parametresi, ϕ , yayılım parametresidir (5, 12, 19, 21).

Genelleştirilmiş doğrusal modeli oluşturan bileşenler aşağıda tanımlanmıştır:

- Doğrusal ifade Eşitlik 2.4'deki gibidir (21, 29):

$$\eta_i = x_i' \beta \quad (2.4)$$

- Doğrusal kestirici (η_i) ile ilişkili olarak y_i 'nin beklenen değerini tanımlayan g bağlantı "link" fonksiyonu aşağıdadır (13, 16, 18).

$$g(\mu_i) = g(E[Y_i]) = x_i' \beta \quad (2.5)$$

- Yanıt değişkeni y_i ($i=1,2,\dots$), bağımsız ve üstel aileden gelen bir dağılıma sahip olsun. μ ortalamaya bağlı olarak yanıt değişkeninin varyansı aşağıdadır.

$$\text{var}(y_i) = \phi V(\mu_i) / w_i \quad (2.6)$$

w_i her gözlem için bilinen ağırlıktır ve yayılım parametresi ϕ , bilinir ya da kestirilir (19).

GDM'nin özelliklerinden birisi, doğrusal olmayan modelleri bağıntı fonksiyonu yardımıyla doğrusal modellere dönüştürebilmesidir. Örneğin, $g(\cdot)$ bağlantı fonksiyonunun uygun seçimi ile negatif binom, geometrik, binom, poisson, gamma, ters-Gauss, normal dağılımlar doğrusal model biçiminde yazılabilir (5).

2.2. Bağıntı Fonksiyonu

Sonuç değişkenin beklenen değeri μ , açıklayıcı değişken x ve bilinmeyen parametre vektörü β olmak üzere, i -inci gözlemin beklenen değeri monoton bir bağıntı fonksiyonu olan g içinde bir doğrusal kestirim ile ilişkilidir. Bu ifade aşağıdadır (19).

$$g(\mu_i) = x_i' \beta \quad (2.7)$$

Her dağılım için kullanılan bağıntı fonksiyonları da farklılık gösterir. Dağılımlar ve dağılımlara karşılık gelen bağıntı fonksiyonları aşağıdaki gibidir: (20, 22, 28).

• Normal	Özdeşlik
• Poisson	Logaritmik
• Binom	Logit
• Bernoulli	Logit
• Negatif Binom	Log-ratio
• Geometrik	Log-ratio
• Üstel	Karşılıklı
• Gamma	Karşılıklı
• Ters Normal	Güç
• Çok değişkenli	Logit

GDM'deki varsayımlardan biri gözlemlerin bağımsız olarak dağılmasıdır. Fakat bazen bu varsayım gerçekleşmeyebilir. Bu durumda GDM yetersiz kalmaktadır.

2.3. Uyum İyiliği Testi

GDM'in uyum iyiliğinin değerlendirilmesinde yardımcı olacak iki istatistik; ölçeklendirilmiş olabilirlik sapma istatistiği (*scaled deviance*) ve Pearson Ki-kare istatistiğidir. Yayılım parametresi ϕ 'nin belli bir değeri için sapma istatistiği, regresyon parametrelerinin en çok olabilirlik kestirimlerinin log-olabilirliği ile elde edilen log-olabilirliğinin en büyüğü arasındaki farkın iki katı olarak tanımlanır (19).

Eğer $L(y, \mu)$, ortalama değerin (μ) ve yanıt değişkeninin (y) bir fonksiyonu olarak belirtilirse, ölçeklendirilmiş sapma istatistiği, aşağıdadır.

$$D^*(y, \mu) = 2(L(y, y) - L(y, \mu)) \quad (2.8)$$

Özel dağılımlar için, bu eşitlik

$$D^*(y, \mu) = D(y, \mu) / \phi \quad (2.9)$$

şeklinde ifade edilebilir. D , sapma istatistiğini göstermektedir. Farklı dağılımlar için farklı sapma istatistikler bulunmaktadır. Örneğin, w_i ; her gözlem için ağırlık olmak üzere normal dağılım için sapma istatistiği aşağıdadır.

$$D = \sum_i w_i (y_i - \mu_i)^2$$

Diğer istatistik, Pearson ki-kare istatistiği Eşitlik 2.10 ile gösterilmektedir:

$$X^2 = \sum_i \frac{w_i (y_i - \mu_i)^2}{V(\mu_i)} \quad (2.10)$$

Ölçeklendirilmiş Pearson ki-kare istatistiği, X^2 / ϕ ile ifade edilmektedir.

Bu iki ölçeklendirilmiş istatistik, belli koşullar altında, sınırlandırılmış ki-kare dağılımına sahiptir. Yayılım parametresi, bilinmediği durumlarda, sapma istatistiği ve Pearson ki-kare istatistiği kullanılarak aşağıdaki eşitliklerle kestirilir:

$$\begin{aligned}\hat{\phi} &= D/(n-p) \\ \hat{\phi} &= X^2/(n-p)\end{aligned}\tag{2.11}$$

Serbestlik derecesi $(n-p)$ de, parametre sayısı (p) ile gözlem sayısının (n) farkına eşittir (19).

2.4. Genelleştirilmiş Doğrusal Model Fonksiyonları

GDM’de genellikle bağımsız veriler üzerine model çıkarımlar kullanılır. Bu modeller, olabilirlik temelli regresyon modelleridir. Bağımsız verilere ilişkin model belirlendikten sonra, belirli parametreler yardımıyla log-olabilirlikler kestirilerek ya da ek parametreler kullanılarak çözümlemeler yapılabilir. Bu olabilirlik kestirim yapısının ilki Tam Bilgiye Dayalı En çok Olabilirlik (TBEO) (*full information Maximum Likelihood*), diğeri de Sınırlı Bilgiye Dayalı En çok Olabilirlik (SBEO) (*limited information maximum likelihood*) olarak adlandırılır. Kestirimler en iyileme yöntemi kullanılarak belirlenir. Bu konuda en çok kullanılan yöntem ise Newton-Rapshon yöntemidir (16).

TBEO kestiriminde, doğrusal, poisson ve bernoulli regresyon modelleri kullanılır. Bu üç regresyon modeli için üç dağılım kullanılır. Bu dağılımlar sırayla Normal, Poisson ve Bernoulli dağılımlarıdır. Bu dağılımların olasılık yoğunluk fonksiyonu ve/veya olasılık fonksiyonları kullanılarak olabilirlik fonksiyonları belirlenir. Olabilirlik temelli modellerde, kestirim eşitlikleri log-olabilirlik modelin türevinden elde edilir (16).

SBEO kestirim eşitliğinde üstel aile kullanılır. Üstel ailenin üstünlüğü birçok dağılımı içinde bulundurmasıdır. Bu dağılımlar Normal, Bernoulli, Binom, Poisson, Gamma, Ters-Normal, Geometrik ve Negatif Binom dağılımlarıdır. Model ve

kestirimlerin elde edilmesinde ki yöntem, TBEO kestirim eşitliklerinde kullanılan yöntemle benzerlik gösterir (16).

2.5. Kestirimlerin Varyansının Elde Edilmesi

Varyans kestirimleri, olabilirlik fonksiyonunun ikinci türevlerinin oluşturduğu matrisin (*Hessian*) tersi ya da ikinci türevinin beklenen değerinden oluşan Fisher skor matrisi ile hesaplanır. Bu hesaplamalar aynı kestirimleri verir. Varyans kestirimleri,

$$\hat{V}_H(\hat{\beta}) = \left\{ \left(-\frac{\partial^2 l}{\partial \beta_j \partial \beta_k} \right) \right\}_{p \times p}^{-1} \quad (2.12)$$

şeklindedir (16). $j, k = 1, \dots, p$ (p ; X 'in sütun boyutunu ifade etmektedir)

Hessian matrisi, olabilirlik fonksiyonun ikinci türevlerinden oluşmaktadır.

$$H = \left[\frac{\partial^2 l}{\partial \beta_j \partial \beta_k} \right]_{p \times p}$$

Fisher skor matrisi de, olabilirlik fonksiyonun ikinci türevlerinin beklenen değerlerinden oluşmaktadır.

$$I = -E \left[\frac{\partial^2 l}{\partial \beta_j \partial \beta_k} \right]_{p \times p}$$

Bu iki yaklaşım, bilgi matrisinin iki farklı yapısını temel alır. Olabilirlik temelli modellerde, dağılımların log-olabilirliğinin yani $\partial l / \partial \beta$ ikinci türevi ile eşitlikler kestirilir.

Varyansın “*Sandwich*” kestiriminin genel yapısı $A^{-1}BA^{-T}$ biçimindedir. A ; bilgi matrisinin varyans kestirimi, B ; kovaryans matrisidir. A ve B matrisinin elde edilmesi için gerekli eşitlikler aşağıdadır (16).

$$\begin{aligned}
\left[\psi = \sum_{i=1}^n \psi_i(x_i, \beta) \right]_{pxl} &= [0]_{pxl} \\
A &= E \left(\frac{\partial \psi(\beta)}{\partial \beta} \right)^{-l} \\
B &= \sum_{i=1}^n E \left[\psi_i(x_i, \beta) \psi_i^T(x_i, \beta) \right] + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n E \left[\psi_i(x_i, \beta) \psi_j^T(x_j, \beta) \right]
\end{aligned} \tag{2.13}$$

Gözlemlerin bağımsızlığı varsayımı altında yukarıdaki eşitliğin ikinci ifadesi sıfır olacağından, B matrisinin kestirimi aşağıdaki gibi olur:

$$\hat{B} = \sum_{i=1}^n \left[\hat{\psi}_i(x_i, \hat{\beta}) \hat{\psi}_i^T(x_i, \hat{\beta}) \right] \tag{2.14}$$

Bu durumda genel varyansın “*sandwich*” kestirimi de aşağıdadır.

$$\hat{V}_s(\hat{\beta}) = \hat{V}_H^{-l}(\hat{\beta}) \hat{B}(\hat{\beta}) \hat{V}_H^{-l}(\hat{\beta}) \tag{2.15}$$

Eğer ilişkili gözlemler, bağımsız kümeler altında toplanabilirse varyans, farklı bir şekilde kestirilir. Bu tür varyans “*modified sandwich*” varyans olarak adlandırılır.

$$\hat{V}_{MS}(\hat{\beta}) = \hat{V}_H^{-l}(\hat{\beta}) \hat{B}_{MS}(\hat{\beta}) \hat{V}_H^{-l}(\hat{\beta}) \tag{2.16}$$

Bu varyans kestiriminde kovaryans matrisi, her bağımsız kümedeki gözlemlerin toplamı olarak kestirilir. $\left[\sum_i (\sum_{it} \psi_{it}) (\sum_{it} \psi_{it})^T \right]$

“*Sandwich*” varyans kestirimi, örneklem genişliğine de bağlıdır. “*Sandwich*” varyans için kullanılan çarpan, $n / (n-p)$ dir. n , örneklem genişliğini; p , modeldeki ortak değişken sayısını göstermektedir. “*Modified Sandwich*” varyans için kullanılan çarpan, $n/(n-1)$ dir (16).

2.6. Boylamsal (*Longitudinal*) Veriler

Bireylerin belirli özelliklerinin, bir zaman aralığı içinde, ikiden çok kez ölçüldüğü çalışmalar oldukça yaygındır. Özellikle klinik çalışmalar, insan ve hayvanların gözlemlendiği çalışmalar bu tip verilere örnek gösterilebilir. Deneğe ilişkin ölçümler farklı zamanlarda olabilir. Ölçümlerin alındığı zamanlar her denek için farklılık gösterebilir. Bu nedenle verilerin toplanması oldukça zaman alıcıdır. Fakat boylamsal veriler, bilimsel olarak daha etkili ve daha güçlü sonuçlar ortaya koyar (31).

Boylamsal verilerin çözümlemesi için bazı yaklaşımlar bulunmaktadır. İlk olarak, kolay bir yaklaşım kullanılır. Bu yaklaşım için tekrarlı ölçümler arasındaki ilişki göz ardı edilir ve geleneksel doğrusal regresyon çözümlemesi kullanılır. Fakat burada bir tehlike bulunmaktadır. Varyansın çok büyük olması ve varsayımların sağlanamaması yanlış test sonuçlarına, etkisiz parametre kestirimlerine ve yanlış çıkarımlara neden olur. Tekrarlı ölçümler arasındaki ilişki değerlerini göz önüne alan yaklaşımlar gerekebilir. Bu nedenle daha genel yaklaşımlar belirtilmiştir. Bunlar, marjinal modeller (GEK) ve rastgele etkili modellerdir (30, 31).

2.7. Panel Veri

Panel veri, kümeli veri ya da tekrarlı ölçüm verisi olarak da adlandırılır. Panel veri, aynı kategoriye ilişkin farklı deneklerin, tekrarlı ölçümlerinin bulunduğu kümeye denir. Bu veriler ilişkili verilerdir. Örneğin, kadın hastalar paneli oluşturuyorsa, kadın hastalara ait kan değerlerinin tekrarlı ölçümleri panel veriyi oluşturur. Eğer küme içindeki veriler, aynı deney birimlerinden geliyor ve *zamana bağımlı* ise uzun dönemli veri olarak adlandırılır. Paneller içindeki tekrarlı ölçümler ilişkilidir. Sonuç değişkenleri arasında ilişki yokken gözlemlerin bağımsız olması durumunda kullanılan yöntem GDM'dir. Marjinal modeller (GEK) ve rastgele etkili modeller ilişkili veriler için kullanılır (16).

Panel veriler için iki durum söz konusudur. Gözlemler $i=1,2,\dots,n$ panele ayrılırsın. Her panelde $t=1,2,\dots,n_i$ gözlem bulunsun. Eğer her panelde bulunan gözlemler birbirine eşitse dengeli panel (*balanced panel*), paneller farklı örneklem genişliğine sahip ise dengeli olmayan panel (*unbalanced panels*) olarak açıklanmaktadır. Dengeli olmayan paneller, örneklem seçimi ile ortaya çıkmıştır. Paneller oluştuktan sonra da panel içinde bulunan gözlemlere benzer gözlemler dahil edilebilir (33). Panel verilerin modellenmesi değişik şekillerde olabilir. Bu modellerden ikisi, sabit etkili model ve rastgele etkili modeldir. (16)

2.7.1. Bileşik (*Pooled*) Kestirim

Panel verinin modellenmesinde basit bir yaklaşım; verideki panelin bağımlılığını yok saymaktır. Bu yaklaşım, bileşik kestirim olarak adlandırılır ve veriler bir kümede birleştirilir. Sonuçta kestirilen katsayı vektörü tutarlı iken, etkili değildir. GDM'den elde edilen varyans kestirimi $\hat{V}_H$, verinin bulunduğu panellerin bağımsız olması durumunda kullanılır. O nedenle “*sandwich*” varyans kestiriminin kullanılması daha uygun görülür. Fakat bu durumda ki varyans kestirimi de olabilirlik temeline uygun görülmemektedir (16).

2.7.2. Sabit Etkili ve Rastgele Etkili Modeller

Eşitlik kestirimlerinde her küme için bir etki bulunmaktadır. Bu etkiler sabit ve rastgele etkilerdir. Sabit etki de kendi içinde koşullu ve koşullu olmayan sabit etkiler olmak üzere ikiye ayrılır. Koşullu modeller, yalnızca küçük grup genişliklerine ilişkin durumlar için uygundur. Rastgele modeller, yanıt değişkeninin dağılımını, grup değişkenlerinin rastgele olması koşulu altında modeller (12).

2.7.2.1. Koşullu Olmayan Sabit Etkili Modeller

Koşullu olmayan sabit etkili modeller, panele ilişkin bir etki içerir. Koşullu olmayan sabit etkili modelde katsayılar ve etki, üstel dağılım ailesinden gelen

dağılımlar ve yarı-olabilirlikler kullanılarak kestirilir (16). Üstel aile için koşullu olmayan sabit etkili eşitlik kestiriminde doğrusal kestirici aşağıdadır.

$$\eta_{it} = x_{it}\beta + v_i \quad (2.17)$$

β ve v 'ye göre kestirimler eşitlikten ayrı ayrı elde edilir.

2.7.2.2. Koşullu Sabit Etkili Modeller

Bu tür kestirimler, koşullu olabilirlikler kullanılarak türetilir. Modeller, üstel dağılım ailesinden değil de belli dağılımlardan (Poisson, bernoulli, gibi) türetilir. Bu modeller için sonuç değişkeni, belirli bir dağılıma sahiptir. Bir kümeye ait tüm gözlemlere ilişkin ortak dağılım hesaplanır ve sabit etki için yeterli istatistik bulunarak yeterli istatistiğin dağılımı elde edilir. Yeterli istatistiğin dağılımına göre verilerin koşullu dağılımı bulunur. Koşullu dağılım kullanılarak katsayılar kestirilir. Tüm panellerin koşullu log-olabilirlik eşitliği aşağıdadır.

$$L = \ln \prod_{i=1}^n f_1(y_i) | f_2(\xi(y_i)) = \ln \prod_{i=1}^n f_3(y_i; \beta | \xi(y_i)) \quad (2.18)$$

$\xi(y_i)$; yeterli istatistiği, $f_3(y_i; \beta | \xi(y_i))$; koşullu dağılımı ifade etmektedir.

Bu dağılım sabit etkiden bağımsızdır (16).

2.7.3. Rastgele Etkili Modeller

Rastgele etkili modeller, parametreleri kestirilebilecek bir dağılım varsayımına göre rastgele etkiyi kestirir. Bu durum, hem sonuç değişkenin hem de rastgele etkinin dağılımına bağlıdır. Rastgele etkili model için log-olabilirlik aşağıdadır.

$$L = \ln \prod_{i=1}^n \int_{-\infty}^{\infty} f(v_i) \left\{ \prod_{t=1}^{n_i} f_y(x_{it}\beta + v_i) \right\} dv_i \quad (2.19)$$

Burada f_y ; tüm model için yoğunluk fonksiyonunu, f_i rastgele etki v_i 'nin yoğunluk fonksiyonunu göstermektedir. Bu integralin çözülebilmesi için uygun dağılım varsayımına göre kestirimler hesaplanır. Bu tip modeller, deneğe özgü modeller olarak adlandırılır ve kümenin marjinal dağılımı yerine bireye ilişkin gözlemlerin dağılımını kullanır. Bu model, koşullu sabit etkili modellerdeki varsayımları kullanır (16).

2.7.4. Kitle Ortalamalı ve Deneğe Özgü Modeller (Population-Averaged and Subject-Specific Models)

Verinin küme yapısı için iki durum bulunmaktadır. Bunlar kitle-ortalama model ve deneğe özgü modeldir. Kitle ortalamalı model, tüm panellerdeki ortalama etki ile panel içi bağımlılığı belirtir. Marjinal model olarak bilinen kitle-ortalama model, panel kovaryans yapısı ile elde edilir. Panel kovaryansı (ya da ilişkisi), panellerin tümüne ilişkin ortalama kullanılarak kestirilir.

Deneğe özgü model, belli bir panelin rastgele bileşenleri ile panel içi bağımlılığı belirtir. Deneğe özgü model, bir panel etkisinin belirtilmesi durumunda yani panel kovaryansı ile hesaplanır. Her bir panel etkisi yalnızca belirli bir panel içindeki bilgiler kullanılarak kestirilir (16).

ÜÇÜNCÜ BÖLÜM

3. GENELLEŞTİRİLMİŞ EŞİTLİK KESTİRİMİ

İncelenen verilerin çözümlemeleri için her zaman gerekli varsayımlar sağlanamayabilir. Birinci bölümde tanımlanan kestirimlerin uygulanabilmesi için bazı varsayımlar gereklidir. Bu kestirim yöntemleri için özellikle verilerin dağılımlarının bilinmesi gereklidir. Örneğin, en çok olabilirlik yöntemi yanıt değişkenin dağılımını gerektirir. En çok olabilirliğe alternatif olarak gösterilen yarı-olabilirlik, ilk iki momentin varlığını ve bağlantı fonksiyonunu gerektirir. Ayrıca farklı kümelerde bulunan gözlemler aynı örneklem genişliklerine sahip olmalıdır (27). GEK, bu tür varsayımlara gerek duymaz.

GEK ile ilgili bazı tanımlar verilmektedir. Bunlardan birkaçı aşağıda belirtilmiştir:

- GEK, ilişkili verilerin regresyon modelini oluşturmak için parametrelerin kestirim yöntemi olarak açıklanır.
- Uzun dönemli verilerin çözümlemesi için yarı-olabilirlik yönteminin çok daha genişletilmiş ifadesidir.
- İlişkili gözlemler için genelleştirilmiş doğrusal modellerin kullanılmasını sağlar ve oldukça çok kullanım alanına sahiptir.

GEK'nin kullanıldığı yerler üç başlıkta toplanabilir. Bunlar; uzun dönemli (boylamsal) veriler, klinik denemeler, epidemiolojik çalışmalardır. GEK çözümlemesi, yanıt değişkeninin kesikli, ikili ya da kategorik veri tipinde olması (mümkünse binom ya da poisson dağılım ailesinden) durumunda daha çok kullanılırken, yanıt değişkenlerin sürekli veri tipi için de uygulanabilmektedir (1, 10, 29).

GEK'e ilişkin olarak regresyon modelinin tanımlamasında gerekli birkaç ifade aşağıda belirtilmiştir (5, 17):

- Bağımlı değişkenin dağılımı (üstel aile üyesi olmalı) tanımlanmalı
- Bağlantı fonksiyonu tanımlanmalı
- Bağımsız değişkenler belirtilmeli
- Tekrarlı ölçümlerin kovaryans yapısı elde edilmelidir.

3.1. Genelleştirilmiş Eşitlik Kestirim Yaklaşımı

$Y_{ij}, j=1, \dots, n_i, i=1, \dots, K$ olmak üzere i -inci denekteki j -inci ölçümü gösterebilir. i -inci denek üzerinde n_i tane ölçüm alınsın ve bu durumda toplam ölçüm, K denek üzerinden $\sum_{i=1}^K n_i$ tanedir (5, 18, 27, 32).

İlişkili veriler, bağımsız gözlemlerde olduğu gibi aynı bağlantı fonksiyonu ve doğrusal kestirim bileşimi kullanılarak modellenir. Bağımsız durumlardaki gibi varyans fonksiyonu da tanımlanır ancak farklı olarak ilişkili ölçümlerin kovaryans yapısı modellenmelidir.

i -inci denekteki ölçüm vektörleri ve ortalamaları sırasıyla $Y = [Y_{i1}, \dots, Y_{in_i}]'$, $\mu = [\mu_{i1}, \dots, \mu_{in_i}]'$ olsun. V ; Y 'nin kovaryans matrisi olsun. i -inci denekteki j -inci ölçüm için bağımsız ya da açıklanan değişken vektörü, $X_{ij} = [x_{ij1}, \dots, x_{ijp}]'$ ile gösterilsin. İlişkili veriler için $p \times 1$ 'lik regresyon parametresi β 'nin kestirimi, aşağıdaki eşitlikten elde edilir:

$$S(\beta) = \sum_{i=1}^K \frac{\partial \mu_i'}{\partial \beta} V_i^{-1} (Y_i - \mu_i(\beta)) = 0 \quad (3.1)$$

$$g(\mu_{ij}) = x_{ij}' \beta$$

g bağlantı fonksiyonu olmak üzere i -inci denek için regresyon parametrelerinin ortalamaya göre kısmi türevlerini gösteren $p \times n_i$ 'lik matris aşağıdaki şekilde ifade edilmiştir (27):

$$\frac{\partial \mu'_i}{\partial \beta} = \begin{bmatrix} \frac{x_{i1l}}{g'(\mu_{i1})} & \dots & \frac{x_{in_i l}}{g'(\mu_{in_i})} \\ \vdots & & \vdots \\ \frac{x_{i1p}}{g'(\mu_{i1})} & \dots & \frac{x_{in_i p}}{g'(\mu_{in_i})} \end{bmatrix} \quad (3.2)$$

3.2. İlişki Matrisi (*Working Correlation Matrix*)

$R(\alpha)$; α parametre vektörleri tarafından belirlenen ilişki matrisi olsun. Y 'nin kovaryans matrisi aşağıdaki gibi modellenir (4, 6, 8, 16, 18, 26, 27, 29, 32):

$$V_i = \phi A_i^{1/2} R(\alpha) A_i^{1/2}$$

ϕ , yayılım parametresi, A_i ; j -inci köşegen elemanları $v(\mu_{ij})$ olan $n_i \times n_i$ köşegen matris olmak üzere eğer $R(\alpha)$, Y 'nin doğru ilişki matrisi ise, V 'de Y 'nin doğru kovaryans matrisidir. İlişki matrisi genellikle bilinmez, bu nedenle kestirim yoluna gidilir. Pearson artıklarının uygun fonksiyonları ile β parametre vektörünün değerleri kullanılarak kestirilir. Artıklar aşağıdaki gibi ifade edilir (7, 16, 18):

$$e_{ij} = \frac{y_{ij} - \mu_{ij}}{\sqrt{v(\mu_{ij})}} \quad (3.3)$$

İlişki matrisinin her zaman doğru ilişki yapısını göstermesi beklenemez, ancak doğru belirlenirse, kestirimlerin kesinliği artmaktadır. Eğer ilişki matrisi $R_0 = I$ yani *birim* matris olarak belirlenirse, GEK, bağımsız eşitlik kestirimlerine indirgenir (16, 18).

3.2.1. Bağımsız İlişki

Gözlemlerin farklı deneklerden alındığı ve gözlemler arasında hiçbir ilişki olmadığı durumdur. Yani bir deneğe ait tüm gözlemler bağımsızdır (5, 16, 26, 27, 29, 32). İlişki matrisi yapısı aşağıdaki gibidir:

$$R = I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \quad (3.4)$$

3.2.2. Değiştirilebilir İlişki

İlişki matrisinin en basit formudur ve bu yapı, bir küme içindeki gözlemler arasında sabit bir ilişkiyi göstermektedir. Bu ilişki de α ile gösterilmektedir. Değiştirilebilir ilişki matrisi yapısı aşağıdaki gibidir (5, 9, 16, 26, 27, 29):

$$R(\alpha) = \begin{bmatrix} 1 & \alpha & \alpha & \cdots & \alpha \\ \alpha & 1 & \alpha & \cdots & \alpha \\ \alpha & \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \alpha & \cdots & 1 \end{bmatrix} \quad (3.5)$$

$$R_{jk} = \begin{cases} 1 & j = k \\ \alpha & j \neq k \end{cases}$$

Bu ifade, aynı deneğe ilişkin gözlemlerin zamana bağlı olmadığı veri setleri için geçerlidir. Bu modelde gözlemler arası ilişki sabittir (α). Değiştirilebilir ilişki yapısı, Pearson artıklar kullanılarak hesaplanabilir. Artıkların kullanılmasıyla elde edilen α 'nın kestirimi ise aşağıdadır:

$$\hat{\alpha} = \frac{I}{(N^* - p)\phi} \sum_{i=1}^K \sum_{j \neq k} e_{ij} e_{jk} \quad (3.6)$$

$$N^* = \sum_{i=1}^K n_i (n_i - 1)$$

3.2.3. Yapısal Olmayan (*Unstructured*) İlişki

Bu ilişki yapısında varsayım; aynı deneğin tüm gözlemlerinin ilişkili olmasıdır. İki değişken arasındaki ilişki, α_{ij} ile gösterilmektedir ve bilinmeyen parametre sayısı $n(n-1)/2$ tanedir (16). Yapısal olmayan ilişki yapısı aşağıdaki gibidir (5, 9, 16, 26, 27):

$$R(\alpha) = \begin{bmatrix} 1 & \alpha_{1,2} & \alpha_{1,3} & \cdots & \alpha_{1,n} \\ \alpha_{1,2} & 1 & \alpha_{2,3} & \cdots & \alpha_{2,n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \alpha_{1,n-1} & \alpha_{2,n-1} & \cdots & 1 & \alpha_{n-1,n} \\ \alpha_{1,n} & \alpha_{2,n} & \cdots & \alpha_{n-1,n} & 1 \end{bmatrix} \quad (3.7)$$

α , Pearson artıkların (e_{ij}, e_{jk}) kullanımı ile kestirilebilir.

$$R_{jk} = \begin{cases} 1 & j = k \\ \alpha_{jk} & j \neq k \end{cases} \quad (3.8)$$

$$\hat{\alpha} = \frac{1}{(K-p)\phi} \sum_{i=1}^K e_{ij} e_{jk}$$

3.2.4. Sabit İlişki

Sabit ilişki matrisi aşağıda verildiği gibidir (5, 16, 18, 26):

$$R(\alpha) = \begin{bmatrix} 1 & r_{1,2} & \cdots & r_{1,n} \\ r_{1,2} & 1 & \cdots & r_{2,n} \\ \vdots & \vdots & \cdots & \vdots \\ r_{1,n} & r_{2,n} & \cdots & 1 \end{bmatrix} \quad (3.9)$$

$$R_{jk} = \begin{cases} 1 & j = k \\ r_{jk} & j \neq k \end{cases}$$

3.2.5. Otoregresif (*Autoregressive*) İlişki AR(1)

Küme içindeki tekrarlı gözlemler doğal bir sıra içindeyken gözlemler arasındaki ilişkinin zamana bağımlı olması halinde kullanılır (9, 16). Otoregresif ilişki matrisi (5, 16, 26, 27, 29):

$$R(\alpha) = \begin{bmatrix} 1 & \alpha^1 & \alpha^2 & \cdots & \alpha^{n-1} \\ \alpha^1 & 1 & \alpha^1 & \cdots & \alpha^{n-2} \\ \alpha^2 & \alpha^1 & 1 & \cdots & \alpha^{n-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \alpha^{n-1} & \alpha^{n-2} & \alpha^{n-3} & \cdots & 1 \end{bmatrix} \quad (3.10)$$

şeklinde belirtilmiştir. İlişki ve ilişki katsayısı aşağıdaki gibidir:

$$\begin{aligned} \text{Corr}(Y_{ij}, Y_{i,j+t}) &= \alpha^t \quad t = 0, 1, 2, \dots, n_i - j \\ \hat{\alpha} &= \frac{1}{(K_1 - p)\phi} \sum_{i=1}^K \sum_{j \leq n_i - 1} e_{ij} e_{i,j+1} \\ K_1 &= \sum_{i=1}^K (n_i - 1) \end{aligned} \quad (3.11)$$

3.2.6. Durağan (*Stationary*) İlişki

Durağan ilişki, M-bağımlı olarak da adlandırılmaktadır. Otoregresif zaman serisine alternatif olarak, bazı küçük zaman aralıklarında örneğin k gibi bir zaman aralığında ilişkinin varlığı için kullanılabilir. Bu durumda, ilişkili olabilecek gözlemler için en büyük zaman aralığı belirlenir (16). Durağan ilişki matrisi aşağıdaki gibidir (5, 16, 18, 26, 32):

$$R(\alpha) = \begin{bmatrix} 1 & \alpha_1 & \alpha_2 & \cdots & \alpha_{t-1} \\ \alpha_1 & 1 & \alpha_1 & \cdots & \alpha_{t-2} \\ \alpha_2 & \alpha_1 & 1 & \cdots & \alpha_{t-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \alpha_{t-1} & \alpha_{t-2} & \alpha_{t-3} & \cdots & 1 \end{bmatrix} \quad (3.12)$$

$$\begin{aligned}
Corr(Y_{ij}, Y_{i,j+t}) &= \begin{cases} 1 & t = 0 \\ \alpha_t & t = 1, 2, \dots, m \\ 0 & t > m \end{cases} \\
\hat{\alpha}_t &= \frac{1}{(K_t - p)\phi} \sum_{i=1}^K \sum_{j \leq n_i - t} e_{ij} e_{i,j+t} \\
K_t &= \sum_{i=1}^K (n_i - t)
\end{aligned} \tag{3.13}$$

3.2.7. Durağan Olmayan (*Nonstationary*) İlişki

Durağan ilişkide olduğu gibi bir g aralığında, ilişkinin varlığı için kullanılır. Durağan ilişkiden farkı ise köşegenlerin altındaki ilişki katsayılarının sabit olmadığı varsayımını taşır (16). Durağan olmayan ilişki matrisi aşağıda belirtildiği gibidir:

$$R(\alpha) = \begin{bmatrix} 1 & \alpha_1 & \alpha_2 & \cdots & \alpha_{t-1} \\ \alpha_1 & 1 & \alpha_1 & \cdots & \alpha_{t-2} \\ \alpha_2 & \alpha_1 & 1 & \cdots & \alpha_{t-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \alpha_{t-1} & \alpha_{t-2} & \alpha_{t-3} & \cdots & 1 \end{bmatrix} \tag{3.14}$$

$$Corr(Y_{ij}, Y_{i,j+t}) = \alpha_t \quad t = 0, 1, 2, \dots, n_i - j$$

$$\hat{\alpha} = \frac{\sum_{i=1}^n n_i}{\sum_{i=1}^n \sum_{t=1}^{n_i} e_{i,t}^2 / n_i} G \tag{3.15}$$

$$G = \begin{pmatrix} g_{1,1} e_{i,1}^2 & g_{1,2} e_{i,1} e_{i,2} & \cdots & g_{1,n_i} e_{i,1} e_{i,n_i} \\ g_{2,1} e_{i,2} e_{i,1} & g_{2,2} e_{i,2}^2 & \cdots & g_{2,n_i} e_{i,2} e_{i,n_i} \\ \vdots & \vdots & \ddots & \vdots \\ g_{n_i,1} e_{i,n_i} e_{i,1} & g_{n_i,2} e_{i,n_i} e_{i,2} & \cdots & g_{n_i,n_i} e_{i,n_i}^2 \end{pmatrix}$$

$$g_{u,v} = \left(\sum_{i=1}^n I(i, u, v) \right)^{-1} \tag{3.16}$$

$$I(i, u, v) = \begin{cases} 1 & u, v \in i - \text{inci küme} \\ 0 & \text{d.h} \end{cases}$$

3.3. En İyi İlişki Yapısını Belirleme

Genelleştirilmiş eşitlik kestiriminde değişkenlere ilişkin katsayıları kestirebilmek için ilişki yapısının belirlenmesi gerekmektedir. Farklı ilişki yapıları bulunmaktadır. Bu ilişki yapıları içinden en doğru ilişkinin belirlenmesi gerekir.

Olabilirlik temelli modeller için model yapısının uygunluğunda kullanılan ölçülerden biri de Akaike Bilgi Ölçütüdür (AIC). AIC ölçütü aşağıdaki gibidir:

$$AIC = -2L + 2p \quad (3.17)$$

L ; log-olabilirliği, p ; model parametrelerinin sayısını göstermektedir. Amaç, yarı-olabilirlik modeller için bu ölçüyü genelleştirmektir. Bu kesimde yarı-olabilirlik ifadesi Q ile ifade edilecektir. AIC, genelleştirilince yeni ölçü olarak QIC tanımlanmıştır. Modeller için yarı-olabilirlik aşağıdadır (16).

$$Q(y; \mu) = \int \frac{y - \mu^*}{V(\mu^*)} d\mu^* \quad (3.18)$$

Yarı-olabilirlik, bağımsızlık varsayımı altında ($R=I$) hesaplanır. $QIC(R)$ ifadesi aşağıdadır.

$$QIC(R) = -2Q(g^{-1}(x\beta_R)) + 2\text{trace}(A_I^{-1}V_{MS,R}) \quad (3.19)$$

Eşitlik 3.19'nin sağ tarafında bulunan;

- $Q(y; g^{-1}(x\beta_R))$, R ilişki yapısı için modelden hesaplanan yarı-olabilirliğin değerini göstermektedir. Yarı-olabilirlik için μ yerine $\hat{\mu} = g^{-1}(x\hat{\beta}_R)$ kullanılmaktadır. $g^{-1}()$ model için bağıntı fonksiyonunu göstermektedir.
- $V_{MS,R}$, R ilişki yapısı için modelden elde edilen “sandwich” varyans kestirimini belirtmektedir.
- A_I bağımsız modelden elde edilen varyans matrisini ifade etmektedir.

İlişki yapılarından uygun olanı QIC ölçütüne göre belirlenir. AIC ölçütünde en küçük değer anlamlı olduğundan QIC ölçütü için de en küçük değerli ilişki yapısı en uygun olanıdır.

İlişki yapısının belirlenmesinde göz önüne alınacak ifadeler aşağıda belirtilmiştir:

1. Eğer panellerin örneklem genişliği küçük ve veri tam ise, yapısal olmayan ilişki yapısı kullanılabilir.
2. Eğer panellerdeki gözlemler aynı örneklem birimlerinden çok fazla alınmışsa zamana bağlı bir yapı kullanılabilir.
3. Eğer gözlemler kümelenmiş (zamana bağlı değilse) ise değiştirilebilir ilişki yapısı kullanılabilir.
4. Eğer panellerin sayısı az ise, bağımsız model uygundur, ancak hipotez testi ve katsayıların yorumu için “*sandwich*” varyans kestirimini hesaplamak gerekir.
5. Eğer birden fazla ilişki yapısı uyumlu ise, QIC kullanılarak en iyi ilişki yapısı seçilebilir (16).

3.4. Yayılım Parametresi

Yayılım parametresi ϕ 'nin kestirimi aşağıdaki gibidir:

$$\hat{\phi} = \frac{1}{N - p} \sum_{i=1}^K \sum_{j=1}^{n_i} e_{ij}^2 \quad N = \sum_{i=1}^K n_i \quad (3.20)$$

N toplam ölçüm sayısı ve p 'de parametre sayısı olmak üzere GEK'nin uygulanmasında izlenen bir algoritma yapısı bulunmaktadır (16,18):

3.5 Algoritma

GEK’nde modelin elde edilebilmesi için belli bir algoritma bulunmaktadır. Bu algoritma olabilirlik kullanan bir yöntem değildir, bu nedenle olabilirlik çıkarımlar GEK’de mümkün değildir. Algoritma yapısı şu şekildedir:

1. Bağımsızlık varsayımı altında genelleştirilmiş doğrusal modeller kullanılarak ilk β ’lar hesaplanır.
2. İlişki matrisi R , hesaplanır. Sonra standartlaştırılmış artıklar hesaplanır.
3. Kovaryans, aşağıdaki eşitlik yardımıyla kestirilir:

$$V_i = \phi A_i^{-1/2} R(\alpha) A_i^{-1/2} \quad (3.21)$$

4. β ’lar

$$\beta_{r+1} = \beta_r + \left[\sum_{i=1}^K \frac{\partial \mu'_i}{\partial \beta} V_i^{-1} \frac{\partial \mu'_i}{\partial \beta} \right]^{-1} \left[\sum_{i=1}^K \frac{\partial \mu'_i}{\partial \beta} V_i^{-1} (Y_i - \mu_i) \right] \quad (3.22)$$

eşitliği ile yinelenir.

5. Adımlar 2-inci ve 4-inci adımlar arasında belli bir noktada birleşinceye kadar tekrarlanır (18, 29, 32).

3.6. Kovaryans Parametresinin Kestirimi

Zeger ve Liang (1986), ilişki matrisinin belirlenmesi sırasında β ’nın kestirim değerlerini bulmak için kovaryans parametresine ilişkin bazı yaklaşımlar önermişlerdir. Çünkü kovaryans parametresi tam olarak belirlenemeyebilir. Bu yaklaşımlar; deneysel ve model temelli kestirimlerdir (5, 26).

$Cov(\hat{\beta})$ 'nin model temelli kestirimi aşağıdadır (29):

$$\sum_m (\hat{\beta}) = I_0^{-1} \quad (3.23)$$

$$I_0^{-1} = \sum_{i=1}^K \frac{\partial \mu'_i}{\partial \beta} V_i^{-1} \frac{\partial \mu'_i}{\partial \beta}$$

Bu eşitlik, β 'nın en çok olabilirlik kestiricisinin kovaryans kestirimi olarak genelleştirilmiş doğrusal modellerde, çok sık kullanılan Fisher bilgi matrisinin tersidir. Eğer model ve ilişki matrisi doğru olarak belirtilirse $\hat{\beta}$ 'nın kovaryans matrisi tutarlı bir kestirimdir (18).

$\hat{\beta}$ 'nın kovaryans matrisinin deneysel (*robust, sandwich*) kestirimi aşağıdadır (18, 29):

$$\sum_e = I_0^{-1} I_1 I_0^{-1} \quad (3.24)$$

$$I_1 = \sum_{i=1}^K \frac{\partial \mu'_i}{\partial \beta} V_i^{-1} Cov(Y_i) V_i^{-1} \frac{\partial \mu'_i}{\partial \beta} \quad (3.25)$$

İlişki matrisi hatalı olarak belirtilse bile, $\hat{\beta}$ 'nın kovaryans matrisinin kestiricimi tutarlılık özelliğini korur (18). Bu nedenle uygulamalarda, deneysel kovaryans matrisi daha çok tercih edilir.

Biometri, epidemiyoloji, sosyal bilimler ve ekonomi alanlarında bazı varsayımlara, özellikle değişkenlerin bağımsızlığına ve değişkenlerin dağılımına ulaşmak mümkün olmayabilir. Örneğin, sayılabilir veriler ya da ikili veriler (hasta olan ya da olmayan insanların sayısı gibi) normal dağılım göstermezler. Sonuç değişkenlerin bağımsızlığı da verilmez. Birkaç tedavi gören tek bir hastadan alınan farklı ölçümler buna örnek olarak gösterilebilir (1).

Bağımsızlık varsayımları, aynı bireye ilişkin tekrarlı ölçümler kullanıldığında bozulur. Bağımlılığın ve tekrarlı gözlemler arasındaki ilişki yapısının belirlenebildiği durumlar için GENELLEŞTİRİLMİŞ EŞİTLİK KESTİRİMİ önerilmiştir.

GEK; marjinal modeller ailesindendir. GEK’de amaç, bağımsız değişkenlerin bir fonksiyonu olarak bağımlı değişkenlerin beklenen değerlerini modellemektir. Doğrusal modeller ortalama ve varyansın fonksiyonel bağımsızlığını gerektirirken, GDM’de kümelerde bağımlılığın önemi yoktur. GEK bu modellerin bir sentezidir (1, 32).

GEK, GEK1 ve GEK2 olarak iki sınıfa ayrılmaktadır. GEK1, regresyon parametrelerinin (β) ve ortak parametrelerinin (α) kestirimlerinin birbirinden bağımsız (ortogonal) olduğu varsayımını içerir ve yalnızca birinci momentlerin tutarlı kestirimini sağlar. GEK1’de amaç, bağımlı değişkenin beklenen değerini bağımsız değişkenin bir fonksiyonu gibi tanımlamaktır. Doğru ortalama yapısı belirlendiğinde GEK1, tutarlı kestirimler sağlar.

Çözümlemede amaç, ortak yapıyı belirlemek ise GEK2 kullanılır. GEK2’de amaç, çözümlemelerin ortalamanın dışında ilişki yardımıyla da yapılabilmesidir. GEK2, regresyon parametreleri ile ortak parametrelerin ikisini de içinde bulundurur böylece birinci momentler yanında ikinci momentlerin kestirimini de sağlar. GEK2’de bu parametrelerin kestiriminin ortogonal olduğu varsayımı yoktur. GEK2, ortak yapı doğru belirlenemese de ortalamanın tutarlı kestirimlerini sağlar. GEK2, GEK1’in daha genel halidir (1, 15, 25).

3.7. Değişen Lojistik Regresyon (ALR)

Yanıt değişkenler ikili olduğunda ölçümler arasında ilişkinin tanımlanması için farklı bir yöntem bulunmaktadır. Bu yöntem Değişen Lojistik Regresyon (*Alternating Logistic Regression*) olarak adlandırılmaktadır. ALR algoritması, log odds oranı ile yanıt değişken çiftleri arasındaki ilişkiyi modellemektedir (18). Pearson artıkları ile ilişkiyi kestirmek yerine ALR algoritması yardımıyla

karşılaştırılmak istenen çiftin odds oranları kestirilir ve ölçümler arasındaki ilişki elde edilir (16)

İkili veriler için j -inci ve k -ıncı yanıt değişkenleri arasındaki ilişki aşağıdaki gibi tanımlanmaktadır:

$$\begin{aligned} \text{Corr}(Y_{ij}, Y_{ik}) &= \frac{\Pr(Y_{ij} = 1, Y_{ik} = 1) - \mu_{ij}\mu_{ik}}{\sqrt{\mu_{ij}(1 - \mu_{ij})\mu_{ik}(1 - \mu_{ik})}} \\ \mu_{ij} &= \Pr(Y_{ij} = 1) \end{aligned} \quad (3.26)$$

$\max(0, \mu_{ij} + \mu_{ik} - 1) \leq \Pr(Y_{ij} = 1, Y_{ik} = 1) \leq \min(\mu_{ij}, \mu_{ik})$ kısıtlaması ile paydaki ortak olasılığa bir sınırlama getirilmiştir (16, 18).

Odds oranı ikili veriler için aşağıdaki gibi tanımlanabilir(4, 16, 18):

$$OR(Y_{ij}, Y_{ik}) = \left[\frac{(\Pr(Y_{ij} = 1, Y_{ik} = 1) \Pr(Y_{ij} = 0, Y_{ik} = 0))}{(\Pr(Y_{ij} = 1, Y_{ik} = 0) \Pr(Y_{ij} = 0, Y_{ik} = 1))} \right] \quad (3.27)$$

ALR algoritması, odds oranının logaritması şeklinde ifade edildiğinde aşağıdaki eşitlik ile tanımlanır:

$$\begin{aligned} \gamma_{ijk} &= \log(OR(Y_{ij}, Y_{ik})) \\ \gamma_{ijk} &= \mathbf{z}'_{ijk} \boldsymbol{\alpha} \end{aligned} \quad (3.28)$$

$\boldsymbol{\alpha}$; $q \times 1$ 'lik regresyon parametresi vektörü ve $\mathbf{z}_{ijk}$ katsayı vektörleri olmak üzere γ_{ijk} parametresi $(-\infty, \infty)$ aralığında herhangi bir değeri alabilir. $\gamma_{ijk}=0$ değeri ilişki olmadığını gösterir (18). İlişki matrisi, lojistik regresyonda log odds oranı ile kestirilebilir. İlişki yapısı ile parametrelerin her biri için odds oranı kestirimleri elde edilebilir (16).

Log odds oranını kullanarak ikili veriler için GEK modeli belirlenebilir. İkili veriler için ortalama model, uygun bağlantı fonksiyonlarından (lojistik, probit yada

tamamlayıcı log-log gibi) kullanılarak oluşturulabilir. ALR algoritması için α ve β 'nin değerleri, standart hataları ve kovaryansları gereklidir (16, 18).

3.8 Kayıp Veri

Kayıp veriler için bazı yöntemler tanımlanmıştır. Bu yöntemler genel olarak veri MCAR olduğunda yalnızca ortalama için yansız kestirimler vermektedir (26). Kestirimlerin yansızlığı için genel bir yaklaşım; tamamlanmamış verinin varlığında GEK'den yararlanmaktır (1, 5). GEK'de kayıp verilerle ilgili birkaç yaklaşım sunulmuştur.

Bir yaklaşım; eğer bağımlı değişken y_{it} eksik gözlenmişse GEK1 yaklaşımı uygulanabilir. Bu yaklaşımda bağımsız değişken, X_i 'lerin tam olarak gözlenmesi gereklidir. GEK1, ortalaması μ_i ve varyansı Σ_i olan çoklu normal dağılımın eşitlik kestirimleri gibi yorumlanabilir.

Diğer bir yaklaşım ise eğer veri MAR olarak gözlenmişse, bağımlı ve/veya bağımsız değişkenlerin eksikliğinde hem GEK1 hem de GEK2 nin ikisinin birden uygulanabilmesidir (1, 15)

DÖRDÜNCÜ BÖLÜM

4. GEREÇ ve YÖNTEM

Aynı denekten belirli zaman aralıklarında alınan ölçümler tekrarlı ölçümlerdir ve bu tekrarlı ölçümler arasında bir ilişki bulunmaktadır. Aynı denekten alınan tekrarlı ölçümlerinin dağılımlarının bilinmemesi, ölçümlere ilişkin varyans–kovaryans yapısının homojen olmaması ve kullanılan verinin tipi hakkında bilginin olmaması durumunda Tekrarlı Ölçümlerde Varyans Çözümlemesi uygulamak yerine GEK uygulamak daha doğru bir seçenektir.

Bu çalışmanın amacı, bir denek üzerinden belirli zaman aralıklarında alınan tekrarlı ölçümlerin birbirine bağımlı olmasından dolayı kullanılan GEK çözümlemesini incelemek, uygun ilişki yapısını ve veri yapısına göre uygun modeli belirlemektir.

Çalışmada kullanılan veriler, Hacettepe Üniversitesi Tıp Fakültesi Diş Hekimliği Diş Hastalıkları ve Tedavisi Anabilim Dalı tarafından sağlanmıştır. Çalışma 26 hasta üzerinde uygulanmıştır. Diş sağlığı kontrollerine gelen hastaların yaş, cinsiyet, dişin renk uyumunda kullanılan USPHS ölçütleri ile yenilenme (restorasyon) değişimi kaydedilmiştir. USPHS ölçütleri, iki seviyede ölçülmüştür. USPHS ölçütleri, dişin yenilenmesine gerek olmayan ve mükemmel yenilenme ile düşmüş ya da yenilenmesini gerektirecek düzeyde değişime uğramış yenilenme olarak alınmıştır.

$$USPHS = \begin{cases} 0 \text{ (Charlie):} & \text{Düşmüş ve yenilenmesine gerek duyulan} \\ 1 \text{ (Alfa ve Bravo):} & \text{Mükemmel ve yenilenmesine gerek olmayan} \end{cases}$$

Dişte zamanla meydana gelen renk değişikliği doğrudan klinik gözlemlere dayalı olarak incelenmektedir. Zaman içerisinde meydana gelen renk değişikliği ve translüsensinin analitik yöntemlerle kantitatif olarak ölçümünün zor olmasından

dolayı doğrudan klinik gözleme dayalı değerlendirme yapılmaktadır. Renk uyumunun belirlenmesinde USPHS ölçütleri başarılı bir şekilde yapılmaktadır. Araştırmacılar; yenilenmeyi çevreleyen diş dokusunu ya da komşu dişi dikkate alarak, yenilenmenin renginin koyulaşması ya da açılmasına göre değerlendirme yaparlar. Diş hekimliğine gelen hastaların renk uyumu değerleri 0., 6., 12., 18. ve 24. aylarda ölçülmüş ve kaydedilmiştir. Zamana göre değişim gösteren dişin renk uyumu değişkeni bağımlı değişken, yaş ve cinsiyet değişkenleri açıklayıcı değişken olarak alınmıştır. Açıklayıcı değişken olan cinsiyet kesikli değişken, yaş ise sürekli değişken olarak alınmıştır.

Tablo 4.1. Çalışmada Kullanılan Değişkenler

Değişkenler	Veri Tipi	Kod	
Yaş	Sürekli		(Bağımsız)
Cinsiyet	Kesikli	Erkek Kadın	(Bağımsız)
Renk uyumu (USPHS ölçütü)	Kesikli	0: Düşük 1: Mükemmel	(Bağımlı)

Uygulama bölümünde verilere öncelikle tekrarlı ölçümlerde varyans çözümlemesi uygulanmıştır. Bu uygulama için SPSS 11.5 paket programı kullanılmıştır. Daha sonra verilere GEK çözümlemesi uygulanmıştır. GEK çözümlemesinde iki istatistik paket programı kullanılmıştır. Değişkenlere ilişkin parametre kestirimlerinin ve parametre kestirimleri için gerekli olan ilişki matrisinin elde edilebilmesi için SAS istatistik paket programı kullanılırken, ilişki matrisinin seçimi için kullanılan program STATA istatistik paket programıdır. STATA paket programında ilişki yapısı belirlendikten sonra SAS paket programı kullanılarak GEK çözümlemesine ilişkin sonuçlar elde edilmiştir. SAS programında GEK çözümlemesi için gereken program GENMOD başlığı altında yazılmaktadır. Bu program Tablo 4.2’de verilmektedir.

Tablo 4.2. Genelleştirilmiş Eşitlik Kestirimi Çözümlemesi İçin SAS Programında Komutlar

<pre>proc genmod data=work.tezveri; class hasta cinsiyet; model ab=yas cinsiyet/d=bin; repeated subject=hasta / type=exch covb corrw; run;</pre>
<pre>proc genmod data=work.tezveri; class hasta cinsiyet; model ab=yas cinsiyet/d=bin; repeated subject=hasta / type=ind covb corrw; run;</pre>

Kullanılan komutlara ilişkin açıklamalar şu şekildedir:

Class: Sınıflandırılmış verileri tanımlar.

D/dist: Dağılımın türünü tanımlar. Ayrıca dağılıma ilişkin bağıntı fonksiyonunu belirler.

Repeated: Modelin belirlenmesinde kovaryans yapısını tanımlar.

Type: İlişki matrisinin yapısını belirtir.

GEK'ne ilişkin regresyon modelinin belirlenmesinde ilişki matrisinin tipinin belirlenmesi önemlidir. Uygun ilişki yapısının belirlenmesi için algoritma yapısı STATA paket programı kullanılarak elde edilmektedir. Algoritma Tablo 4.3'de verilmektedir.

Tablo 4.3. İlişki Tipinin Belirlenmesine İlişkin STATA Algoritma Yapısı

```

capture program drop qicm
program define qicm
  quietly {
    matrix V = e(V)
    matrix Vi= syminv(V)
    matrix T = Vi*V
    matrix B = trace(T)
    scalar cc =B[1,1]
    tempvar mu ql
    predict double `mu', mu
    gen double `ql' = (ab*log(`mu'/(1-`mu')))+log(1-`mu'))
    summ `ql',meanonly
  }
  display in green "QIC = " in yellow %8.4f 2*(cc-r(sum))
  display `ql'
end

xtgee ab yas cinsiyet, i(hasta) t(z) fam(bin) corr(exch) robust
qicm exch

xtgee ab yas cinsiyet, i(hasta) t(z) fam(bin) corr(ind) robust
qicm ind

xtgee ab yas cinsiyet, i(hasta) t(z) fam(bin) corr(ar 1) robust
qicm "ar 1"

xtgee ab yas cinsiyet, i(hasta) t(z) fam(bin) corr(ar 2) robust
qicm "ar 2"

xtgee ab yas cinsiyet, i(hasta) t(z) fam(bin) corr(unst) robust
qicm unst

xtgee ab yas cinsiyet, i(hasta) t(z) fam(bin) corr(sta 2) robust
qicm "sta 2"

```

İlişki matris tipinin belirlenmesi için algoritmada kullanılan komut, ‘qicm’ komutudur. En küçük qicm değeri en iyi ilişki yapısını vermektedir.

BEŞİNCİ BÖLÜM

5. BULGULAR

5.1. Tanımlayıcı İstatistikler

Hacettepe Üniversitesi Diş Hekimliği Fakültesine Mayıs 2003 ile Mayıs 2005 tarihleri arasında gelen 26 hastaya ilişkin tanımlayıcı bilgiler verilecektir. Araştırma kapsamında bulunan 26 hastaya 5 tekrarlı ölçüm yapıldığından araştırmada 130 gözlem bulunmaktadır. Bu 130 gözlemin yaşına ilişkin tanımlayıcı istatistikler Tablo 5.1’ de verilmiştir.

Tablo 5.1. Yaş Değişkenine İlişkin Tanımlayıcı İstatistikler

Değişken	N	Minimum	Maksimum	Ortalama	Standart sapma
Yaş	26	29	67	47.15	9.51

Çalışmaya katılan 26 hastanın %26.9’u (n=7) erkek ve %73.1’i (n=19) kadındır. Çalışmada bağımlı değişken olan dişteki renk uyumunun yenilenme değerleri Tablo 5.2’de verilmiştir.

Tablo 5.2. Bağımlı Değişkene İlişkin Tanımlayıcı İstatistikler

Renk uyumu	Düşük		Mükemmel	
	Sayı	%	Sayı	%
Başlangıç	0	0.0	26	100.0
6.ay	2	7.7	24	92.3
12.ay	3	11.5	23	88.5
18.ay	5	19.2	21	80.8
24.ay	8	30.8	18	69.2

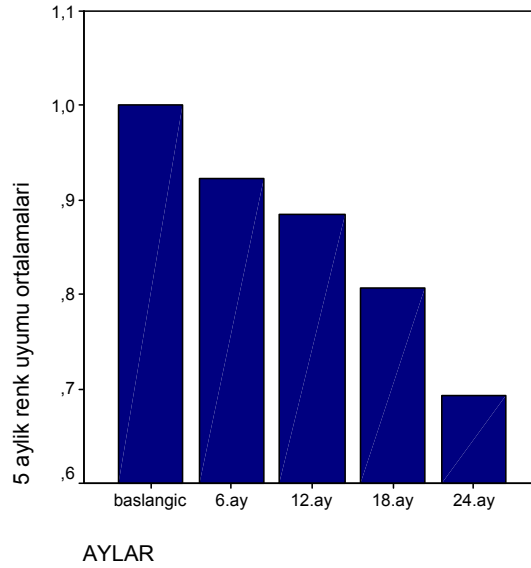
Çalışmada, dişin renk uyumu ile ilgili olarak yenilenme 5 farklı zamanda ölçülmektedir. 26 hastanın dişine ilişkin tekrarlı renk uyumu ölçümleri, hastaların yaşları ve cinsiyetlerini gösteren veriler Tablo 5.3’de verilmektedir.

Tablo 5.3. Hastaların Dişlerinin Renk Uyumu,Yaş ve Cinsiyet Değişkenlerine İlişkin Veriler

hasta	cinsiyet	yas	renk	zaman
1	e	49	1	1
1	e	49	1	2
1	e	49	1	3
1	e	49	1	4
1	e	49	1	5
2	e	54	1	1
2	e	54	1	2
2	e	54	1	3
2	e	54	1	4
2	e	54	1	5
3	k	37	1	1
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
9	e	53	1	3
9	e	53	1	4
9	e	53	1	5
10	k	33	1	1
10	k	33	1	2
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
15	k	53	1	3
15	k	53	1	4
15	k	53	1	5
16	k	47	1	1
16	k	47	1	2
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
25	k	35	1	4
25	k	35	1	5
26	k	38	1	1
26	k	38	1	2
26	k	38	1	3
26	k	38	1	4
26	k	38	0	5

5.2. Tekrarlı Ölçümlerde Varyans Çözümlemesi

26 hastanın her birinden dişlerinin renk uyumu ile ilgili 5 farklı ölçüm alınmıştır. Bu ölçümlere Tekrarlı Ölçümlerde Varyans Çözümlemesi uygulanırsa Normal dağılım varsayımı altında Şekil 5.1'e ulaşılmıştır.



Şekil 5.1. Hastaların Aylara Göre Renk Uyumunu Gösteren Grafik

Gözlemler Normal dağılım göstermemektedir. Normal dağılım gösterdiğini, varyans yapısının homojen olduğunu ve ölçümlerin sürekli olduğunu düşünerek Tekrarlı Ölçümlerde Varyans Çözümlemesi uygularsak sonuçlar Tablo 5.4'de ve Tablo 5.5'de verilmiştir.

Tablo 5.4. Tekrarlı Ölçümlerde Varyans Çözümlemesine İlişkin Sonuçlar

Değişken	Kareler Toplamı	Serbestlik derecesi	Kareler Ortalaması	F	p-değeri
Sabit	6,292	1	6,292	102,381	,000
YAS	,716	1	,716	11,647	,002
CINS	,794	1	,794	12,914	,002
Hata	1,414	23	,061		

Tablo 5.5. Tekrarlı Ölçümlerde Varyans Çözümlemesi İçin Parametre Kestirim Değerleri

Değişken	Kestirim	Standart		%95 Güven		t-değeri	p-değeri
		Hata		Aralığı			
Sabit	1.194	0.112	0.962	1.426		10.643	0.000
yas	-0.008	0.002	-0.013	-0.003		-3.413	0.002
erkek	0.179	0.050	0.076	0.281		3.594	0.002
kadın	0.000	0.000	0.000	0.000			

Varyans Analizi sonucunda değişkenler anlamlı bulunmuştur.

5.3 Genelleştirilmiş Eşitlik Kestirimi

Çalışmada 26 hasta bulunmaktadır. Bu hastaların her birinden 5 farklı zamanda ölçüm alınmıştır. Bu nedenle çalışmada 130 gözlem bulunmaktadır. SAS programında veri girişi Tablo 5.3’de olduğu gibidir. Verilere ilişkin uygun ilişki matrisinin sonuçları Tablo 5.6’da verilmiştir.

Tablo 5.6. En İyi İlişki Yapısına İlişkin Sonuçlar

İlişki Yapıları	QIC
Değiştirilebilir İlişki	99.4018
Bağımsız İlişki	99.4018
Otoregresif 1 İlişki	99.4047
Otoregresif 2 İlişki	99.4072
Yapısal olmayan	100.1857
Durağan İlişki	99.4065

Bu sonuçlara göre ilişki matrisi için en uygun yapı değiştirilebilir ya da bağımsız ilişki yapılarıdır. Değiştirilebilir ve bağımsız ilişki yapılarına göre verilere komut yapısı uygulandığında (Bkz.Tablo 4.2) uyum iyiliği testinin sonuçları Tablo 5.7’de verilmiştir.

Tablo 5.7. Uyum İyiliği Sonuçları

Ölçümler	SD	Değer	Değer/SD
Sapma	127	93.4018	0.7354
Ölçeklendirilmiş Sapma	127	93.4018	0.7354
Pearson Ki-Kare	127	134.0037	1.0551
Ölçeklendirilmiş Pearson Ki-kare	127	134.0037	1.0551
Log-olabilirlik		-46.7009	

Gözlemlerin birbirinden bağımsız olduğu varsayımı altında değişkenlere ilişkin ilk katsayı kestirim değerleri Tablo 5.8’de verilmiştir.

Tablo 5.8. İlk Parametre Kestirim Değerleri

			Standart	%95 Güven			
Değişken	SD	Kestirim	Hata	Aralığı		Ki-kare	p > Ki-kare
Sabit	1	4.2432	1.3358	1.6250	6.8614	10.09	0.0015
yaş	1	-0.0564	0.0261	-0.1076	-0.0053	4.67	0.0307
erkek	1	2.0992	1.0516	0.0381	4.1603	3.98	0.0459
kadın	0	0.0000	0.0000	0.0000	0.0000		

Yapılan çözümleme sonucunda yaş ve cinsiyet değişkenleri anlamlı bulunmuştur.

β katsayılarının kestirimi için ilişki matris tipine göre ilişki matrisi Tablo 5.9 ve Tablo 5.10’da verilmiştir.

Tablo 5.9. Değiştirilebilir İlişki Yapısı Altında İlişki Matrisi

	0.ay	6.ay	12.ay	18.ay	24.ay
0.ay	1.0000	-0.1037	-0.1037	-0.1037	-0.1037
6.ay	-0.1037	1.0000	-0.1037	-0.1037	-0.1037
12.ay	-0.1037	-0.1037	1.0000	-0.1037	-0.1037
18.ay	-0.1037	-0.1037	-0.1037	1.0000	-0.1037
24.ay	-0.1037	-0.1037	-0.1037	-0.1037	1.0000

Gözlemlere ilişkin tüm tekrarlı ölümler arasında zıt bir ilişki bulunmaktadır.

Tablo 5.10. Bağımsız İlişki Yapısı Altında İlişki Matrisi

	0.ay	6.ay	12.ay	18.ay	24.ay
0.ay	1.0000	0.0000	0.0000	0.0000	0.0000
6.ay	0.0000	1.0000	0.0000	0.0000	0.0000
12.ay	0.0000	0.0000	1.0000	0.0000	0.0000
18.ay	0.0000	0.0000	0.0000	1.0000	0.0000
24.ay	0.0000	0.0000	0.0000	0.0000	1.0000

İlişki matrisinin elde edilmesinden sonra β katsayısının kestirimi için gerekli olan değiştirilebilir ilişkiye ilişkin model-temelli ve deneysel kovaryans matrisi Tablo 5.11 ve Tablo 5.12’de verilmektedir.

Tablo 5.11. Değiştirilebilir İlişki Yapısına Göre Model-Temelli Kovaryans Matrisi

	Sabit	Yaş	Cinsiyet
sabit	1.04442	-0.01997	-0.04364
yaş	-0.01997	0.0003988	-0.000014
cinsiyet	-0.04364	-0.000014	0.64724

Tablo 5.12. Değiştirilebilir İlişki Yapısına Göre Deneysel Kovaryans Matrisi

	Sabit	Yaş	Cinsiyet
Sabit	1.02637	-0.02054	-0.08329
Yaş	-0.02054	0.0004243	-0.002336
Cinsiyet	-0.08329	-0.002336	0.99162

Bağımsız ilişkiye ilişkin model-temelli ve deneysel kovaryans matrisi Tablo 5.13 ve Tablo 5.14’de verilmiştir.

Tablo 5.13. Bağımsız İlişki Yapısına Göre Model-Temelli Kovaryans Matrisi

	Sabit	Yaş	Cinsiyet
sabit	1.78446	-0.03412	-0.07457
yaş	-0.03412	0.0006814	-0.000023
cinsiyet	-0.07457	-0.000023	1.10585

Tablo 5.14. Bağımsız İlişki Yapısına Göre Deneysel Kovaryans Matrisi

	Sabit	Yaş	Cinsiyet
Sabit	1.02637	-0.02054	-0.08329
Yaş	-0.02054	0.0004243	-0.002336
Cinsiyet	-0.08329	-0.002336	0.99162

Parametre kestirimlerinin GEK çözümlmeleri, değiştirilebilir ve bağımsız ilişki yapıları için aynıdır. Çünkü deneysel kovaryans matrisleri her iki ilişki yapısı için değişim göstermemektedir. Çözümleme sonuçları Tablo 5.15’de verilmiştir.

Tablo 5.15. Değiştirilebilir ve Bağımsız İlişki Yapısı İçin Genelleştirilmiş Eşitlik Kestirimi Çözümlmeleri

Değişken	Kestirim	Standart Hata	95% Güven Aralığı		Z	P > Z
Sabit	4.2432	1.0131	2.2575	6.2288	4.19	<.0001
yaş	-0.0564	0.0206	-0.0968	-0.0160	-2.74	0.0062
erkek	2.0992	0.9958	-4.0509	-0.1475	-2.11	0.0350
kadın	0.0000	0.0000	0.0000	0.0000		

GEK çözümleme sonucuna göre değiştirilebilir ve bağımsız ilişki yapısı alındığında değişkenlerin modele katkısı anlamlı bulunmuştur.

GEK çözümlemesi sonucunda, model aşağıdadır.

$$E(y)=4.2432-0.0564yas+2.0992cinsiyet_{erkek}$$

Bu modele göre, erkek hastaların dişlerinde yenilenmeye gerek duyulmaması olasılığı, bayan hastalara göre 2.09 kat daha fazladır. Ayrıca yaş arttıkça, dişin yenilenmeme olasılığı azalmaktadır.

SONUÇ

Sağlık, sosyal ve ekonomi alanlarında toplanan veriler genellikle tekrarlı ölçümlerden oluşmaktadır. Bu ölçümler için farklı yöntemler uygulanmaktadır. Bu yöntemlerden biri de, son yıllarda oldukça fazla kullanılan GEK'dir. Bu kestirim yöntemi, bir denek üzerinden alınan ölçümlerin arasında ilişki durumunu inceler. Çoğu kestirim yöntemi, ölçümlere ilişkin dağılımların bilinmesi durumunda kullanılır ve bazı varsayımları şart koşar. GEK'ni diğerlerinden ayıran en önemli özelliklerden biri de bu varsayımlara gerek duymamasıdır.

Çalışmada kullanılan verilere Tekrarlı Ölçümlerde Varyans Çözümlemesi ve Genelleştirilmiş Eşitlik Kestirimi uygulanmıştır. Her iki çözümleme sonucunda elde edilen parametre kestirim değerleri anlamlı bulunmuştur. Bu durum Tablo 6.1'de belirtilmiştir

Tablo 6.1. Tekrarlı Ölçümlerde Varyans Çözümlemesinden ve Genelleştirilmiş Eşitlik Kestirim Çözümlemesinden Elde Edilen Parametre Kestirim Değerleri

Tekrarlı Ölçümlerde Varyans Çözümlemesi		
Değişken	Kestirim	p-değeri
Sabit	1.194	0.000
Yaş	-0.008	0.002
Erkek	0.179	0.002
Kadın	0.000	-
Genelleştirilmiş Eşitlik Kestirimi Çözümlemesi		
Değişken	Kestirim	p-değeri
Sabit	4.2432	<.0001
Yaş	-0.0564	0.0062
Erkek	2.0992	0.0350
Kadın	0.0000	-

Varyans Çözümlemesi, ayrı ayrı her hastanın aylara göre ölçüm değerlerini kullanmaktadır. GEK çözümlemesinde ise her hastanın aylara göre ölçümleri bir küme gibi düşünülerek, kümeler için kestirimler hesaplanmaktadır. 26 hastanın dişlerindeki renk uyumuna ilişkin değerlerin kestirimleri Tablo 6.2’de verilmiştir.

Tablo 6.2. Tekrarlı Ölçümlerde Varyans Çözümlemesinden Kestirilen Değerler

HASTA	0.ay	6.ay	12.ay	18.ay	24.ay
1	1,00	1,00	1,01	0,87	1,01
2	1,00	1,00	0,93	0,80	0,95
3	1,00	0,88	1,00	0,91	0,69
4	1,00	0,92	0,49	0,52	0,34
5	1,00	0,88	1,13	1,01	0,78
6	1,00	1,00	1,00	0,85	1,00
7	1,00	0,88	1,08	0,97	0,74
8	1,00	1,00	1,05	0,89	1,03
9	1,00	1,00	0,94	0,81	0,96
10	1,00	0,88	1,07	0,96	0,73
11	1,00	0,90	0,74	0,71	0,51
12	1,00	1,00	1,01	0,87	1,01
13	1,00	0,90	0,79	0,75	0,55
14	1,00	0,92	0,51	0,53	0,35
15	1,00	0,90	0,73	0,70	0,50
16	1,00	0,90	0,83	0,78	0,57
17	1,00	0,91	0,68	0,66	0,47
18	1,00	0,90	0,81	0,77	0,56
19	1,00	0,89	0,93	0,86	0,64
20	1,00	0,91	0,66	0,65	0,45
21	1,00	0,90	0,76	0,73	0,52
22	1,00	0,89	0,86	0,81	0,59
23	1,00	1,00	1,06	0,91	1,04
24	1,00	0,89	0,91	0,84	0,63
25	1,00	0,88	1,03	0,93	0,71
26	1,00	0,89	0,98	0,90	0,67

Varyans çözümlemesinde, ayrı ayrı hastaların ölçümlerinin kestirilen değerlerin ortalaması alınır. GEK çözümlemesinde ise kümelerin ortalaması alınır. İki çözümleme yöntemi, aylara göre dişlerin renk uyumu değişiminin ortalamaları bakımından karşılaştırılmak istenirse elde edilen sonuçlar Tablo 6.3’de verilmiştir.

Tablo 6.3. Varyans Çözümlemesi ve Genelleştirilmiş Eşitlik Kestirimi ile Elde Edilen Ortalama Değerlerinin Karşılaştırılması

TÖVÇ	GEK
0,977	0,97282
0,937	0,96428
0,895	0,89620
0,653	0,61374
0,960	0,93131
0,969	0,97129
0,936	0,91966
0,993	0,97565
0,945	0,96617
0,928	0,91539
0,774	0,78740
0,977	0,97282
0,799	0,81436
0,661	0,62703
0,766	0,77781
0,815	0,83082
0,742	0,74719
0,807	0,82274
0,863	0,87325
0,734	0,73638
0,782	0,79670
0,831	0,84609
1,001	0,97696
0,855	0,86687
0,912	0,90623
0,887	0,89083

Her iki çözümleme sonunda elde edilen ortalama değerleri arasında büyük bir farklılık bulunmamaktadır

İki yönteme göre elde edilen kestirimler arasında fark olmaması, kullanılan veriye özgü olup karşılaştırma için kesin kanıt olamaz. Varyans çözümlemesinde bağımlı gözlemlere ilişkin ayrı ayrı kestirimler elde edilirken; GEK’de her denek için bağımlı gözlemlerin ortalamaları önceden belirlenen ilişki yapısına göre kestirilmektedir. Bunun en önemli nedeni, GEK’nin dağılım varsayımı olmayışıdır. Varyans çözümlemesinde ayrı kestirimler kullanılan bağımlı gözlemler arasındaki ilişki sonradan test edilir.

KAYNAKLAR

1. Ziegler, A., Kastner, C. ve Bleitner, M., The Generalized Estimating Equations : An Annotated Bibliography. Biometrical Journal, 40, 2, 115-139, 1998.
2. Dahmen, G., Ziegler, A., Generalized Estimating Equations in Controlled Clinical Trials: Hypotheses Testing. Biometrical Journal 46, 1, 214-232, 2004.
3. Hojsgaard, S., Quasi-likelihood and generalized estimating equations . Biometry research unit. Danish Institute of Agricultural Science <http://www.jbs.agrsci.dk/Biometri/Kurser/Kursusforloeb/BiostatKursusPHD-F04/day7/GEE04.pdf>. [16/11/2004]
4. O'Brien, L., M. ve Fitzmaurice, M.G , Analysis of longitudinal multiple-source binary data using generalized estimating equations. Application Statistical, 53, part1, 177-193, 2004.
5. Horton, N.J., Lipsitz, S., R. , Review of software to fit Generalized estimating equation regression Models. The American statistician, ,Vol.53,.160-169, 1999.
6. Park, P., Weisberg, S., Fisher consistency of GEE Models under link misspecification, Computational Statistics ve Data Analysis, 27, 229-235, 1998.
7. Ziegler, A. , Martus, P.,Stroux, A., Jünemann, M.A., Korth, M., Jonas, J.B., Horn, F.K., GEE approaches to marginal Regression Models for medical diagnostic tests, Statistics in medicine, 23, 1377-1398, 2004.
8. Horton, N.J., Bebhuck, J.D., Jones, C.L., Lipsitz, S.R., Catalona, P.J., Zahner, G.E.P., Fitzmaurice, G.M., Goodness of fit for GEE : An example with mental health service utilization, Statistics in Medicine, 18, 213-222, 1999.
9. Ziegler, A., GEE for Analysis of Clinical trials, http://www.kks-ms.de/Dateien/Ziegler_GEE.pdf. [08/03/2004]
10. Hamm, J., Unleashing the fury of Generalized Estimating equations, www.sfu.ca/~cdlinkle/gee.ppt. [13/02/2004]
11. Lee, J., The EM algoritm, <http://stats.snu.ac.kr/~leej/teaching/2004Fall/comp/EMAlgorithm/EM.pdf>. [15/06/2005]

12. Gadda, H.J., Commenges, D., Tests of homogeneity for Generalized linear Models, *Journal of the American Statistical Association*, 90 ,1237-1246, 1995.
13. Dunlop, D.D., Regression for Longitudinal Data : A bridge from Least Squares Regression, *The American Statistical Association*, 48 , 299-303, 1994.
14. Elashoff, M., Ryan, L., An EM algorithm for Estimating Equations, *Journal of Computational and Graphical Statistics*, 13 , 48-65, 2004.
15. Ziegler, A. , Kastner, C., Grömping, U., Blettner, M., The Generalized Estimating Equations in the past ten years: an overview and a biomedical application, *Biometrical Journal* 40 (2), 115-139, 1996.
16. Hardin, W.H., Hilbe, J., Generalized Estimating Equations, A. CRC Press Company, 2003.
17. Horrocks, J., Generalized Linear Models, Generalized Estimating Equations notes from workshop, <http://www.umanitoba.ca/centres/mchp/concept/dict/Statistics/gee.html>. [20/06/2004]
18. Sas Institute Inc., Cary, NC, USA, <http://www.id.unizh.ch/software/unix/statmath/sas/sasdoc/chap29/sec38.htm>, [21/06/2004]
19. Sas Institute Inc., Cary, NC, USA, <http://www.id.unizh.ch/software/unix/statmath/sas/sasdoc/chap29/sec27.htm>, [21/06/2004]
20. Sas Institute Inc., Cary, NC, USA, <http://www.id.unizh.ch/software/unix/statmath/sas/sasdoc/chap29/sec4.htm>, [21/06/2004]
21. Park, T., Davis, S.C., Li, N., Alternative GEE estimation procedures for discrete longitudinal data, *Computational statistics&Data analysis*, 28, 243-256, 1998.
22. Balan, R.M., Kratina-Schiopu, I., Asymptotic results with GEE for longitudinal data II, <http://lrsr.carleton.ca/trs/pdf/FILES/TRS398.pdf>, [09/03/2004]
23. Dellaert, F., The Expectation Maximization Algorithm, Georgia Institute of Technology, <http://www.cc.gatech.edu/~dellaert/em-paper.pdf>, [27/03/2005]
24. Lange, K., The EM algorithm, http://www.ipam.ucla.edu/publications/inv2003/inv2003_3986.pdf, [20/06/2005]

25. Ziegler, A. , Grömping, U., The Generalized estimating equations :A comparison of procedures available in commercial stastical software packages, Biometrical Journal 40, 3, 245-260, 1998.
26. Horton, N., Fitting Generalized estimating equations (GEE) regression models in Stata, <http://www.stata.com/meeting/1nasug/gee.pdf>, [09/05/2005]
27. Dong, J., GEE approach , Power Point sunumu, Department of mathematical and Statistical Sciences, University of Alberta, Kanada, (2004)
28. Genstat , 2002, <http://www.vsn-intl.com/genstat/index.htm>, [10/12/2004]
29. Chapter 8 Generalized Estimating Equations (GEE models), <http://www.uic.edu/classes/bstt/bstt513/chap8.pdf>, [14/06/2005]
30. Huang, G.H., What can Longitudinal data do?, http://www.stat.nctu.edu.tw/subhtml/source/teachers/ghuang/files/presentations/longitudinal_tran.pdf. [10/06/2005]
31. Longitudinal data, <http://www.public.iastate.edu/~dicook/ggobi-book/chap-long.pdf>, [11/06/2005]
32. Liang, B.K. ve Zeger, S.L., Longitudinal data analysis using generalized linear models. Biometrika, 73, 1, 13-22, 1986.
33. Methods for Panel data, http://www.economia.uniroma2.it/nuovo/didattica/MaterialeDidattico/43%5CPanel_data_-_Lecture_notes.pdf, [20/12/2005]