

ONDOKUZ MAYIS ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

**UZUNLAMASINA KESİKLİ BAĞIMLI DEĞİŞKENLERE SAHİP
REGRESYON MODELLERİ İÇİN ANALİTİK METOTLARIN
İNCELENMESİ
VE UYGULANMASI**

135990

ELİF BULUT

YÜKSEK LİSANS TEZİ

İSTATİSTİK ANABİLİM DALI

DANIŞMAN

YRD.DOÇ.DR. MEHMET ALİ CENGİZ

SAMSUN-2003

135990

ONDOKUZ MAYIS ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Bu çalışma jürimiz tarafından 10 / 09 /2003 tarihinde yapılan sınav ile İSTATİSTİK Anabilim Dalı'nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Başkan : Prof. Dr. Yüksel BEK

Üye : Yrd. Doç. Dr. Mehmet Ali CENGİZ
(Danışman)

Üye : Yrd. Doç. Dr. Vedat SAĞLAM

ONAY:

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.

...../...../2003

Fen Bilimleri Enstitüsü Müdürü



UZUNLAMASINA KESİKLİ BAĞIMLI DEĞİŞKENLERE SAHİP REGRESYON MODELLERİ İÇİN ANALİTİK METOTLARIN İNCELENMESİ VE UYGULANMASI

ÖZET

Bu tezde kesikli sonuç değişkenlerinden oluşan ve tekrarlı olarak ölçülmüş veriler için istatistiksel süreç ele alınmıştır. Kesikli sonuç değişkeni için tekrarlı olarak ölçülen kesikli bağımlı değişkenler ile açıklayıcı değişkenler arasındaki ilişkiyi sağlayan regresyon modeli ve parametre tahmini üzerinde durulmuştur.

Tekrarlı ölçümün tanımı verilip, diğer veri setlerinden farkı ve tekrarlı ölçümün bir türü olan uzunlamasına veri tipinin tanımı yapılmıştır. Daha sonra bu veri kümesine özel bazı genel bilgiler yanında, çalışmamıza temel oluşturan metodun dayandığı genelleştirilmiş lineer modeller hakkında bilgi verilmiştir.

Genelleştirilmiş Lineer Modeller (GLM) basit ve çoklu lineer regresyon, varyans ve kovaryans analizi, lojistik regresyon gibi birçok istatistiksel analiz metodu için teorik ve kavramsal bir çatı oluşturmuştur. Aynı şekilde GLM, gözlemlenebilir faktörler ve açıklayıcı değişkenler kullanarak tesadüfi değişkenlerin davranışlarını belirlemede oldukça kullanışlıdır.

GLM' nin genel yapısı esas alınarak uzunlamasına veri için özel regresyon modelleri ve bu çalışmada kullanılacak olan genelleştirilmiş tahmin denklemleri (GTD) incelenmiştir.

GTD kullanılması anakütle ortalama etkilerinin daha güçlü tahminlerinin elde edilmesini ve özellikle grup farklılıkları hakkında sonuç çıkarmada da büyük kolaylıklar sağlar. Bu metot, yarı olabilirliğe dayanan regresyon parametreleri üzerinde odaklanan ve aynı bireyden elde edilen korelasyonlu gözlemleri de hesaba katarak parametre tahminlerinin yapılmasını temin eder.

Yöntem nörolojik hasarı olan ve olmayan çocuk hastalar üzerinde bazı ilaçların sedasyon etkisini belirlemek için elde edilen verilere uygulanarak uygulamanın sonuçları tartışılmıştır.

Anahtar Kelimeler: Uzunlamasına veri, Genelleştirilmiş Tahmin Denklemleri, GLM, Yarı olabilirlik fonksiyonu, Sedasyon

**A REVIEW ON ANALYTIC METHODS FOR REGRESSION MODELS FOR
LONGITUDINAL DISCRETE RESPONSES AND AN APPLICATION**

ABSTRACT

In this study, statistical processes for repeated discrete responses were investigated. Regression models which measures the relationship between discrete response, which is repeated and covariates were presented and parameter estimation method was given.

Definition of repeated measurements was given. The differences from the other data set were emphasized and longitudinal data, which is a specific case of repeated data was explained. Later giving some general definition used for this data set. Generalized Linear Models were presented.

GLM offer convenient and applicable tools for modelling and predicting the behaviour of random variables in terms of observable factors and covariates. These provide an unified theoretical and conceptual framework for many of the most commonly used statistical methods, including simple and multiple linear regression, analysis of variance, analysis of covariance and logistik regression.

GEE processes provides the robust estimation of effects of population averaged. This method focuses on regression parameters based on quasi-likelihood and give parameter estimations with correlation among observations from the same individual.

As an application data, regarding the sedation effects on patient with neurologic damage or without, were used. The results were given and discussed.

Key Words: Longitudinal data, Generalized Estimating Equations, GLM, Quasi-likelihood Functions, Sedation

TEŐEKKÜR

Yardımlarını esirgemeyen kıymetli hocam Yrd. Doç. Dr. Mehmet Ali Cengiz'e ve her zaman yanımda olan aileme saygılarımla.



İÇİNDEKİLER

	Sayfa
1. GİRİŞ	1
2. GENEL BİLGİLER	6
2.1. Lineer Model ve En Küçük Kareler Metodu	6
2.2. Çok Değişkenli Normal Teori	7
2.3. Olabilirlik Yöntemi	9
2.4. Genelleştirilmiş Lineer Modeller	10
2.4.1. Lojistik Regresyon	10
2.4.2. Poisson Regresyonu	10
2.4.3. Genel Sınıf	11
2.5. Yarı Olabilirlik Fonksiyonları	12
2.5.1. Bağımsız Gözlemler Durumu	13
2.5.1.1. Kovaryans Fonksiyonları	13
2.5.1.2. Yarı Olabilirlik Fonksiyonunun Yapısı	13
2.5.1.3. Parametre Tahmini	16
2.5.2. Bağımlı Gözlemler Durumu	17
2.5.2.1. Yarı Olabilirlik Tahmin Eşitlikleri	17
2.5.2.2. Yarı Olabilirlik Fonksiyonu	18
3. MATERYAL VE METOT	19
3.1. Uzunlamasına Veri için Genelleştirilmiş Lineer Modeller	19
3.1.1. Marjinal Modeller	19
3.1.2. Tesadüfi Etki Modelleri	20
3.2. Uzunlamasına Kesikli Yanıt Değişkenleri için Regresyon Modelleri	20
3.3. Genelleştirilmiş Tahmin Denklemleri	23
3.3.1. Korelasyon Matrisi R nin etkisi	24
3.3.2. Normal Durumda İteratif Prosedür ve Birim Link	27
3.3.3. σ^2 ölçek parametresi	28
3.3.4. Katsayıların Standart Hataları	28
3.4. Çok Terimli Dağılım	29
4. BULGULAR VE TARTIŞMA	31
4.1. Sedasyon	31

	Sayfa
4.2. Nörolojik Hastalara İlişkin Seadsyon Verisi için Genelleştirilmiş	
Tahmin Denklemlerinin Kullanımı	32
4.2.1. Değişkenlerin Tanımlanması	32
4.2.2. Analiz	33
4.2.3. Sonuçlar	34
5. SONUÇ VE ÖNERİLER	41
6. KAYNAKLAR	43



SİMGE VE KISALTMALAR

GLM	: Genelleştirilmiş lineer modeller
GTD	: Genelleştirilmiş tahmin denklemleri
EKK	: En küçük kareler
RSS	: Artık Kareler Toplamı
SEK	: Sıradan En Küçük Kareler
MLE	: Ençok olabilirlik tahmini
OR	: Odds ratio
RR	: Relative (görel) risk
exp	: Üstel
log, ln	: Logaritma
E(Y)	: Y' nin beklenen değeri
V(Y)	: Y' nin varyansı
$\phi(\cdot)$: N(0,1) için ihtimal yoğunluk fonksiyonu
$\Phi(\cdot)$: N(0,1) için yoğunluk fonksiyonu
$L(\beta(y))$: β ' nın olabilirlik fonksiyonu
$l(\beta(y))$: β ' nın ençok olabilirlik fonksiyonu
e_i	: i. birim için hata terimi
α	: Alpha
β	: Beta
ε	: Epsilon
ρ	: Rho
τ	: Tau
μ	: Mu
ν	: Nu
σ	: Sigma
ω	: Omega
δ, Δ	: Delta
ϕ, Φ	: Phi
ψ, Ψ	: Psi
π, Π	: Pi

θ, Θ	: Theta
\sim	: Denk
\perp	: Diktir
$>$: Büyük
$<$: Küçük
\sim	: Denk
\mathbb{R}	: Reel sayılar
\mathbb{R}^+	: Pozitif reel sayılar
\sum	: Toplam
$ \cdot $: Mutlak değer
$!$: Faktöriyel
\bar{x}	: Ortalama vektör
$\hat{\beta}$: Tahmini β vektörü
$\hat{\sigma}$: Tahmini σ değeri
$\beta^{(k)}$: β 'nin k .nci türevi
y'	: y fonksiyonunun birinci mertebeden türevi
$[a_{ij}]_{n \times m}$: a_{ij} elemanlarından oluşmuş $n \times m$ boyutlu matris; $1 \leq i \leq n, 1 \leq j \leq m$
$[a, b]$: Kapalı aralık
(a, b)	: Açık aralık
$B(\pi)$: Bernoulli dağılımı
$B(m, \pi)$: Binom dağılımı
$N(\mu, \sigma^2)$: Normal dağılım
$\text{Mult}(n, \pi)$: Multinomial Dağılım

ŞEKİLLER LİSTESİ

	Sayfa
Şekil 4.1 Zaman içerisinde midazolan ilaç grubunun sedasyona olan etkisinin değişimini gösteren grafik	34
Şekil 4.2 Zaman içerisinde kardioloji kokteyli ilaç grubunun sedasyona olan etkisinin değişimini gösteren grafik	34
Şekil 4.3 Zaman içerisinde diazem ilaç grubunun sedasyona olan etkisinin değişimini gösteren grafik	35
Şekil 4.4 Zaman içerisinde lüminal ilaç grubunun sedasyona olan etkisinin değişimini gösteren grafik	35
Şekil 4.5 Sedasyon verisi için ortalamaların testini gösteren grafik	36



TABLULAR LİSTESİ

	Sayfa
Tablo 4.1 Ramsey Sedasyon Skoru	32
Tablo 4.2 İlaç Türleri	33
Tablo 4.3 Cinsiyet değişkenini gösteren tablo	33
Tablo 4.4 Hastalık durumunu gösteren tablo	33
Tablo 4.5 SAS GENMOD Program yazılımı	36
Tablo 4.6 GENMOD Prosedürü Model Bilgisi	37
Tablo 4.7 Sınıf Seviye Bilgileri	37
Tablo 4.8 Bağımlı Değişken Profili	37
Tablo 4.9 Parametre Bilgileri	37
Tablo 4.10 Uyum İyiliğini Tahmin İçin Kriter	38
Tablo 4.11 Başlangıç Parametre Tahminlerinin Analizi	38
Tablo 4.12 Başlangıç Parametre Tahminlerinin Analizinin Devamı	38
Tablo 4.13 GTD Model Bilgisi	39
Tablo 4.14 GTD Parametre Tahminlerinin Analizi	40

1. GİRİŞ

Tekrarlı ölçümler; farklı zamanlarda veya farklı koşullar altında aynı deneysel birim üzerinde yapılan ölçümlerdir. Tekrarlanan ölçümlerde deneysel birim çoğunlukla 'bir canlı' dır. Yani insan, hayvan veya bitki olabilir.

Tekrarlanan ölçümler birçok farklı alanda ortaya çıkar ve yaygın olarak kullanılır. Genellikle sağlık alanında yapılan çalışmalarda, aynı deneklerin çeşitli zamanlardaki ölçümleri incelendiğinden tekrarlı ölçümlü deneyler çokça tercih edilmektedir. Bunun dışında biyolojik, psikolojik ve sosyolojik alanlarda da kullanımı mevcuttur. Bu alanlarda kullanılmasının yaygın bir nedeni bir sürecin birey üzerinde zaman içerisinde oluşturduğu değişimi incelemektir. Bu bilimlerin dışında güncel hayatta da kullanılabilir. Örneğin; bir otomobil tamircisi zaman içinde çeşitli model arabalarda oluşan problemlerle ilgileniyorsa, bunu çalışabilmek için belirli bir süre içinde farklı model arabalardan oluşan problemler yönünden bunları seçtiği aynı arabalarda takip ederek gözlemler yapabilir.

Deney tasarımı ve veri toplama şekline bağlı olarak tekrarlanan ölçümlerin analizlerinde belirgin farklılıklar vardır. Örneğin; 40 hastanın oluşturduğu bir grupta zaman içinde 4 farklı ilacın yarattığı durumun analizi ile her bir ilacın 10 hastadan oluşan 4 gruba verilerek belirli bir periyotta gözlemlenerek analiz edilmesi arasında gerek deney tasarımı gerekse verilerin analizi yönünden fark vardır. Tekrarlanan ölçümlerin varyans analizindeki genel amaç birey üzerinden farklı zamanlarda elde edilen iki veya daha fazla deney grupları arasındaki farkı belirlemek için kullanmaktır.

Tekrarlı ölçümlerin analizinde aynı birey üzerinden farklı zamanlarda alınan ölçümler kullanıldığından, bu ölçümlerin aynı birey üzerinden alınması nedeniyle korelasyonlu olması analizi karıştırmaktadır. Bu yüzden verinin bu özelliğini modelleme ve analizde belirtmek gereklidir.

Uzunlamasına veriler ise tekrarlı ölçümlerin genel bir halidir ve çoğu zaman eşdeğer anlamda kullanılır. Uzunlamasına çalışmalar, bağımlı değişkenin iki veya daha fazla durumu için her bir deneysel birimin gözlemlendiği çalışmalar olarak tanımlanabilir. Uzunlamasına veri kümesinin tanımlayıcı özelliği denek üzerindeki gözlemlerin tekrarlı oluşudur. Bu tip çalışmalarda veriler ya ileriye yönelik, bireyleri deneyin kuruluşundan ileri doğru takip ederek

ya da geçmişe dönük hasta kayıtlarından, her bir kişi üzerindeki çoklu ölçümlerinden elde edilir. Bu tip verilerde, bireyler üzerindeki tekrarlı ölçümler zamanla ortaya çıkan değişimin çalışılmasına imkan verirler.

Uzunlamasına çalışmalar özel istatistiksel metotlara ihtiyaç duyarlar. Çünkü; aynı kişi üzerinden elde edilen gözlemler kümesi kendi aralarında ilişkili olma eğilimindedir. Bu korelasyon geçerli bilimsel sonuç çıkarırken hesaba katılmalıdır. Korelasyonlu veriyi sanki bağımsızmış gibi analizde kullanmak bazı sakıncalar doğurur. Bunlar; standart hataların değerinin altında hesaplanmasından dolayı regresyon parametreleriyle ilgili olarak yanlış sonuç çıkarma ve büyük hata kareler ortalaması olması nedeniyle yani regresyon parametrelerinde tutarsız tahminlerin ortaya çıkması, dolayısıyla istatistiksel sonuç çıkarımının gücünün azalmasıdır.

Uzunlamasına çalışmaları diğer çalışmalardan ayıran özellikler de söz konusudur. Uzunlamasına çalışmanın kesitsel çalışmalara karşı en büyük avantajı bireyler arası farklılıklardan zamanla oluşan bireyler içi değişkenliğin ayıklanabilmesidir. Kesitsel çalışmalar buna imkan vermez. Diğer bir farksa ortaköz etkisinin yani bireylerin esas değerleri arasındaki farklılıkların uzunlamasına çalışmalarda ayırt edilebilir olup kesitsel çalışmalarda ayırt edilememesidir. Kesitsel çalışmalarda her bir deneysel birim için sadece tek bir yanıt değişkeni mevcuttur. Kesitsel çalışmalar tekbir anda, belli bir aralıkta, çalışma topluluğundaki her olguya ilişkin tek bir gözlemin saptanması ile gerçekleştirilir. Nedensel ilişkileri yansıtmada çok güçlü olmayabilir. Kohort (ortaköz) çalışmalarda ise belirli bir süre içerisinde izlenen araştırmanın özelliği açısından temelde benzer özellikleri paylaşan kişilerden oluşan topluluk söz konusudur. Topluluk kuramsal bir başlangıç noktasından (doğum ve tedavi başlangıcı gibi) yola çıkıp zaman içinde uzunlamasına izlenir ve bu süreçte beliren etkenler ve sonuçlar izleme sonunda yargılanır. Nedensel ilişkilerin ortaya konmasında en doğru araştırma tasarımı olmasına karşın zor, pahalı, uzun ve kayıplara çok açıktır. Uzunlamasına çalışma tasarımı ise daha iyi çalışan ve verimli olabilen, az pahalı ve model seçmede çok sağlam, artan istatistiksel güce sahip olabilen tasarımlardır.

Uzunlamasına veri analizi zaman serilerinin genelleştirilmesini ve regresyon modellerini kapsar. Uzunlamasına veriler ölçümlerin uzun serisi yerine kısa serilerin büyük yığını kapsadığından zaman serilerinden ayrılır. Verilerimizi çok değişkenli verilerden ayıran özelliklerde mevcuttur. Tekrarlı ölçümlerde ölçülen özellik aynı olup her bir zamanda

ölçülmüştür. Bu yüzden ölçümler oranlıdır. Bu genel çok değişkenli veride uygun değildir. (Orada bir değişken boy diğeri ağırlık v.b...) Ölçümler tipik olarak bahsedilen sürekli diziden seçilen durumlardan alınır. Bazen bireyler farklı durumlarda, durumların farklı sayılarında ölçülmektedir. Bu analizi pratikte karıştırabilir ama onu prensipte değiştirmez. Bununla birlikte, uzunlamasına veri analizinin temelini zamanın oluşturduğu sürekli bir dizi de vardır. Gözlemlerin ardışık doğal yapısı kovaryans yapısının özel çeşidini ifade eder.

Uzunlamasına veriler, günümüzde Biomedikal çalışmalarda ve değişik sağlık alanlarında yaygın olarak kullanılmaktadır. Kullanımındaki bu artış bilgisayar teknolojisindeki çok hızlı gelişimle birlikte istatistik teorisinin sağladığı bu avantajdan faydalanmak nedeniyledir. Bilgisayar teknolojisindeki gelişim istatistikçilere uzunlamasına veri analizi için istatistiksel teknikleri kolay uygulama imkanları sunmuştur. Klinik deneylerde uzunlamasına verilerin gözlemlenme nedeni tedavinin zaman boyunca etkisini değerlendirmektir.

Uzunlamasına veri analizinde;

1. Bireyler içi korelasyon dikkate alınır.
2. Eşit olmayan zaman aralıklarında ölçülen tekrarlı veriler çalışılabilir.
3. Kayıp gözlemleri dikkate almak mümkündür.

Eğer zaman aralıkları eşitse ve kayıp veri yoksa yani dengeli bir deney planının analizinde tekrarlanan ölçümlü varyans analizini kullanmak uygundur.

Farklı bireyler farklı sayıda tekrarlanan ölçüme sahipse GTD tekniği kullanılır. Burada modeller lojistik, log-lineer veya lineer olabilir.

Uzunlamasına çalışmaların amacı; birey yanıt değişkenleri modelini tanımlama ve bu modellerdeki önemli açıklayıcı değişkenlerin etkisini araştırmaktır. Bu tür çalışmalarda iki tip açıklayıcı değişken mevcuttur. Bunlar; zamanın fonksiyonu olmayan bireyler arası açıklayıcı değişkenler ve zamanın fonksiyonu olan bireyler içi açıklayıcı değişkenlerdir. Uzunlamasına çalışmalarda 4 tipte yanıt değişkenleriyle karşılaşılabilir. Bunlar; sürekli, kesikli, sayılabilen ve sağkalım türü değişkenlerdir. Uzunlamasına verilerinin en ilginç özelliklerinden biri büyük bir uğraş gerektirmeden hazır olarak yorumlanabilir grafiksel haritada gösterilebilmesidir.

Uzunlamasına veri analizi aşağıdaki şekilde özetlenerek, anlaşılması kolaylaştırılabilir.

- ✓ Birey davranışlarını tanımlamak; zamanın bazı noktalarında bireyler arasında verinin nasıl değiştiğini gözleme.

- ✓ Bağımlı değişkenin zamanla nasıl değiştiğini gözlemleme.
- ✓ Bağımlı değişkenin bir veya daha çok açıklayıcı değişkenle birlikte nasıl değiştiğini gözlemleme.
- ✓ Verideki olağan olmayan gözlemleri saptama.
- ✓ Bütün veri noktalarının kendilerini gözlemleme.

Uzunlamasına veri analizinde izlenecek dört adım vardır. Bunlar;

- I. Modelin seçimi
(kullanılacak model şeklini tanımlayan link fonksiyonunu belirleme)
- II. Varyans-kovaryans yapısının belirlenmesi.
(her bir birey için çalışma korelasyon yapısını belirleme)
- III. Modelin uyum iyiliğini tayin etme.
- IV. Varyans-kovaryans yapısının uyum iyiliğini tayin etme.

Uzunlamasına veri modelinin kullanılmasında amaçlanan hususlar;

- I. Değişimin sürecini (ilerleyişini) tanımlaması.
 - i. Değişimin bireysel modelini tanımlaması.
 - ii. Değişimin alt grup ve kitle ortalamalı modelini tanımlaması.
- II. İyi bir güçle önceki tahminlere kılavuzluk etmesi.
- III. Mevcut bilgileri kullanmasıdır.

Yakın zamanlarda uzunlamasına analiz ve tekrarlı ölçümlerle ilgili bazı çalışmalar yapılmıştır.

Hand ve Taylor (1987) istatistikçilerden ziyade tekrarlı ölçümler problemi ile sık karşılaşan davranış bilimleri araştırmacılarına yönelmiştir. Varyans analizini, varyansın çok değişkenli analizini ve matematiksel ayrıntılara inmeden tekrarlı ölçümlere karşılık gelen yaklaşımları anlatmaya çalışmışlardır. Bunu da global modeller oluşturmaktan ziyade gruplar arası ve ölçümler arası kontrastlara odaklanarak yapmışlardır. Çalışmalarının devamında ise uygun bilgisayar yazılımlarının kullanımını vermişlerdir.

Tekrarlı ölçümler analiz metoduna geniş yer veren diğer bir kitapsa Crowder ve Hand (1990) dir. Fakat bu kaynak sürekli normal dağılımlı verilere odaklanmıştır. Tek değişkenli ve çok değişkenli varyans analizi ve regresyon modellerini kapsamaktadır. Kategorik veri ve kesitsel tasarımları da incelemiştir. Ayrıca SAS; BMDP ve SPSS programlarında tekrarlı ölçümler probleminin analizi incelenmiştir.

Jones (1993) normal dağılımlı hataların varsayılabileceği durumları tanımlar ve tesadüfi gruplar arası etkilere ve grup içi korelasyonlara izin veren Laird-Ware modellerini oluşturmuştur. Lindsey (1993) ise analiz için GLIM ve MATLAB'ı da kapsayan çeşitli programlar kullanmıştır.

Diggle, Liang ve Zeger (1994) biyoloji ve sağlık bilimleri araştırmalarında önemli bir yere sahiptir. Çalışma, marjinal modelleri, tesadüfi etki modellerini ve dönüşüm modellerini kapsar.

Longford (1993) sadece tesadüfi katsayı modelleriyle sınırlanmıştır. Uzunlamasına veri analizi için literatürde birçok modeller geliştirilmiştir. Bu modellerin çoğu kesikli yapı gibi sürekli yapıya da sahip olabilen tek değişkenli bağımlı değişkenlerinin analizi ile ilgilidir. Bu projenin amacı farklı tipteki sonuç değişkenleri (sürekli, kesikli veya karışık) ortak analizine izin veren modelleri irdelemektir. Bu tip modellerin esas amacı farklı sonuç değişkenleri arasındaki ilişkiyi çalışmaya izin vermesi ve daha önemlisi zaman üzerinden bu ilişkinin nasıl değiştiğinin modele katılmasına olanak sağlamasıdır.

Uzunlamasına veriyi analizde bir yaklaşım Liang ve Zeger (1986) tarafından verilmiştir. Bu yaklaşımda, Genelleştirilmiş Tahmin denklemleri için GTD olarak terimlendirdikleri yarı olabilirlik metodolojisinin bir uzantısını geliştirmişlerdir. Bu yaklaşımda hem sürekli hem de kategorik bağımlı değişkenler analize dahil edilebilirken, ana etkilere ve etkileşim etkilerine ait genelleştirilmiş tahmin denklemi kullanılarak tekrarlı ölçümlerin regresyon analizi yapılmaktadır. GTD prosedürü kitle ortalama etkilerinin güçlü tahminlerini sağlar ve özellikle analizin amacı grup farklılıkları hakkında yorum yapma olduğunda bu yöntem çok avantaj sağlar.

Bu çalışmanın birinci bölümünde, uzunlamasına veriler tanımlanmış, benzer çalışmalardan farklılıkları ortaya konmuş, literatür çalışması verilmiş ve çalışmanın özeti sunulmuştur. İkinci bölümde, çalışmaya esas olacak temel bilgiler ve kavramlar verilerek, konunun iyi anlaşılması amaçlanmıştır. Üçüncü bölümde ise çalışmanın esas konusu olan kesikli uzunlamasına veriler ve bunlara ilişkin modeller sunulmuştur. Çalışmamızın orijinal kısmı olan uygulamada ise, dört farklı ilaç grubunun denendiği sedasyon verileri tanımlanmış, farklı gruplara göre grafiksel gösterimler ve analizler sunulmuş, sonuçlar elde edilmiştir. Bu sonuçlar, son bölümde (Sonuç ve Tartışma) tartışılmış ve gerekli karşılaştırmalar yapılmıştır. Ayrıca geleceğe ilişkin yorum ve önerilerde eklenmiştir.

2. GENEL BİLGİLER

2.1. Lineer Model ve En Küçük Kareler Metodu

Birçok bilimsel çalışmanın esas amacı, sonuç veya bağımlı değişkeni kestirici veya açıklayıcı değişken bilgilerini kullanarak tahmin etme veya tanımlamaktır. Regresyon analizi 19. yüzyılın başlarından beri bağımlı değişken Y nin bekleneni ile açıklayıcı değişkenler kümesi $x_{ij}; j=1, \dots, p$ arasındaki ilişkiyi açıklamada kullanılmaktadır. Lineer regresyon modeli, aşağıda gösterilen eşitlik ile bu ilişkiyi gösterir.

$$Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i = x_i' \beta + \varepsilon_i \quad i=1, \dots, m \quad (2.1)$$

Burada Y_i ; m bireyden i .nci birey için bağımlı değişken ve x_{ij} ise i .nci birey için j .nci açıklayıcı değişken değerini gösterir. Genellikle her birey için $x_{i1} = 1$ dir ve böylece β_1 regresyon modelinin kesişimi olarak isimlendirilir. ε_i ler birbirinden bağımsız varsayılan tesadüfi değişkenlerdir ve $E(\varepsilon_i) = 0$ ve $\text{Var}(\varepsilon_i) = \sigma^2$ dir. Y_i nin ilk iki momentleri, $E(Y_i) = x_i' \beta$ ve $\text{Var}(Y_i) = \sigma^2$ dir. Burada x_i ve β , p -elemanlı vektörlerdir. Y_1, \dots, Y_m ortak dağılımının çok değişkenli normal dağılım olması yaygın bir varsayımdır. Lineer modeller aşağıdaki gibi özel durumlar içerir;

- x_j lerin kukla değişkeni olduğu ve deneysel birimin tedaviye paylaşımını gösterdiği varyans analizinde.
- x_j lerin nicel değişkenler olduğu çoklu regresyonda.
- x_j lerin sürekli ve kukla değişkeninin birlikte kullanıldığını gösteren kovaryans analizinde.

EKK metodu (Legendre, 1805; Gauss, 1809) regresyon katsayıları β vektörünün tahmini için kullanılan bir metottur ve

$$\text{RSS} = \sum_{i=1}^m (Y_i - x_i' \beta)^2 \quad (2.2)$$

eşitliği ile tanımlanan kareler toplamını minimum yapan $\hat{\beta}$ değerinin bulunmasını amaçlar.

Bu işlemin β ya göre kısmi türevini alıp sıfıra eşitlersek

$$\frac{\partial \text{RSS}}{\partial \beta} = -2 \sum_{i=1}^m x_i (Y_i - x_i' \beta) = 0$$

buradan $\hat{\beta}$ değeri aşağıdaki gibi elde edilir.

$$\hat{\beta} = \left(\sum_{i=1}^m x_i x_i' \right)^{-1} \sum_{i=1}^m x_i Y_i$$

EKK tahmin edicisi $\hat{\beta}$ 'nin alternatif ve daha tanınan formu, m element vektörü $Y = (Y_1, \dots, Y_m)$ ve $m \times p$ boyutlu x_{ij} elemanlarından oluşan X matrisi ile tanımlanarak elde edilir. Şöyle ki;

$$\hat{\beta} = (X'X)^{-1} X'Y \quad (2.3)$$

EKK tahmini $\hat{\beta}$ birçok cazip istatistiksel özelliklere sahiptir. İlk olarak β 'nin yansız bir tahmin edicisidir. Yani $E(\hat{\beta}) = \beta$ dir. Varyans matrisi ise $\text{Var}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$ dir. İkinci olarak, bilinen katsayıların herhangi bir a vektörü için $\phi = a'\hat{\beta}$ alıp tahmin edicisini $\hat{\phi} = a'\hat{\beta}$, Y_i 'nin lineer kombinasyonları olan ϕ için bütün yansız tahminler arasında mümkün en küçük varyansa sahiptir. EKK tahmin edicisinin optimal özelliği Gauss-Markov Teoremi olarak bilinir.

ε_i 'nin sabit varyansı, σ^2 , genellikle bilinmez fakat aşağıdaki eşitlikle tahmin edilebilir.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^m (Y_i - x_i' \hat{\beta})^2}{(m-p)} \quad (2.4)$$

Seber (1977), Draper ve Smith (1981) EKK tahmin edicisi hakkında detaylı bilgi vermektedir.

2.2. Çok Değişkenli Normal Teori

Normal teori çok değişkenli normal gözlemler için önemli sonuçlardan bazılarını gösterir. Çok değişkenli normal teori hakkında daha detaylı bilgi için Graybill (1976) ve Rao (1973)'ya bakılabilir.

Bir tesadüfi vektör $Y = (Y_1, \dots, Y_m)$ için olasılık yoğunluk fonksiyonu aşağıdaki şekilde tanımlandığında, çok değişkenli bir normal dağılıma uyduğu söylenebilir.

O halde;

$$f(\mathbf{y}; \boldsymbol{\mu}, \mathbf{V}) = (2\pi)^{-n/2} |\mathbf{V}|^{-1/2} \exp\left\{-\frac{(\mathbf{y} - \boldsymbol{\mu})' \mathbf{V}^{-1} (\mathbf{y} - \boldsymbol{\mu})}{2}\right\} \quad \text{dir.} \quad (2.5)$$

Burada $-\infty < y_j < \infty$ ve $j=1, \dots, m$ dir. Tek deęişkenli durumdaki gibi bu daęılımın ilk iki momenti, $\boldsymbol{\mu} = E(\mathbf{Y})$ ve $\mathbf{V} = \text{Var}(\mathbf{Y})$ dir ve $\mathbf{Y} \sim \text{MVN}(\boldsymbol{\mu}, \mathbf{V})$ formunda gösterilir.

Çok deęişkenli daęılımın ařaęıdaki özelliklerini sıkça kullanacaęız.

- Her bir Y_j tek deęişkenli Normal daęılıma sahiptir.
- Genel olarak $m_1 < m$ olmak üzere $Z_1 = (Y_1, \dots, Y_{m_1})$ tesadüfi vektörü çok deęişkenli normal daęılıma ve $(\mu_1, \dots, \mu_{m_1})$ ortalama ile V_1 kovaryans matrisine sahiptir. Burada V_1 matrisi V matrisinin m_1 sayısına göre bölünmüş sol üst matrisidir.
- Ek olarak eęer $Z_2 = (Y_{m_1+1}, \dots, Y_m)$ ise o zaman $Z_2 = z_2$ şartı altında Z_1 tesadüfi vektörü çok deęişkenli normal daęılıma sahiptir. Ortalama vektörü ise;

$$\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_{m_1} \end{pmatrix} + V_{12} V_{22}^{-1} \left(z_2 - \begin{pmatrix} \mu_{m_1+1} \\ \vdots \\ \mu_m \end{pmatrix} \right)$$

ve koşullu varyans matrisi;

$$V_{11} - V_{12} V_{22}^{-1} V_{12}' \quad \text{öyle ki} \quad \mathbf{V} = \begin{pmatrix} V_{11} & V_{12} \\ V_{21}' & V_{22} \end{pmatrix} \quad \text{dir.}$$

- Eęer B $k \leq m$ ranklı $k \times m$ boyutlu matris ise o zaman $B\mathbf{Y}$, $B\boldsymbol{\mu}$ ortalamalı ve $B\mathbf{V}B'$ varyans matrisine sahip çok deęişkenli normal daęılım vektörüdür.
- $U = (\mathbf{Y} - \boldsymbol{\mu})' \mathbf{V}^{-1} (\mathbf{Y} - \boldsymbol{\mu})$, m serbestlik dereceli χ^2 daęılımına sahip tesadüfi vektördür ve $U \sim \chi_m^2$ ile ifade edilir.

2.3. Olabilirlik Yöntemi

Olabilirlik sonuç çıkarımı, gözlemlenen veri y için olasılık veya olasılık yoğunluğunun belirtilmesine dayanır. Bu anlatım, $f(y;\theta)$, bilinmeyen θ vektörü tarafından gösterilir. Veri gözlemlendiğinde araştırmacıya göre bilinmeyen nitelikler $f(\cdot)$ da ki θ dır. O halde θ için olabilirlik fonksiyonu aşağıdaki gibidir.

$$L(\theta|y) = f(y;\theta) \quad (2.6)$$

Olabilirlik gözlem değeri y sabit tutulduğunda θ nin fonksiyonu gibi yorumlanabilir. θ nin en çok olabilirlik tahmini; olabilirlik fonksiyonu veya eşdeğerce onun logaritmasını maksimize eden $\hat{\theta}$ değeridir. Yani, θ nin herhangi bir değeri için;

$$L(\theta|y) \leq L(\hat{\theta}|y)$$

Olabilirlik prensibine göre; $\hat{\theta}$ gözlemlenen veri tarafından en güçlü desteklenen θ değeri gibi görülür. Temelde, $\hat{\theta}$ ya direkt $\log L$ nin maksimizasyonu tarafından yada aşağıdaki denklemler kümesini çözerek sağlanır.

$$S(\theta) = \frac{\partial \log L}{\partial \theta} = 0$$

$S(\theta)$ fonksiyonu, θ için skor fonksiyonu olarak bilinir. Çoğu zaman maksimum olabilirlik tahminini değerlendirmede nümerik metotlar gereklidir. $\log L$ nin direkt maksimizasyonu için yaygın metotlar, Nelder ve Mead (1965) simplex algoritmalarını ya da skor fonksiyonu çözümü için Newton-Raphson iterasyonunu kapsar.

En çok olabilirlik tahmininin büyük örneklerde birçok uygun özelliklere sahip olduğu bilinir. Özellikle ılımlı düzenli koşullar altında $\hat{\theta}$ asimtotik olarak yansız ve θ elemanlarının herhangi bir asimtotik yansız tahmin edicinin en küçük mümkün asimtotik varyansıyla tahmin edildiği durumda asimtotik olarak yeterlidir.

$\hat{\theta}$ nin asimtotik kovaryans matrisi aşağıdaki gibidir.

$$V = \left\{ -E \left(\frac{\partial^2 \log L}{\partial \theta^2} \right) \right\}^{-1}$$

Burada V^{-1} matrisi aynı zamanda θ için Fisher Bilgi Matrisi olarak da ifade edilir.

2.4. Genelleştirilmiş Lineer Modeller

Bağımsız, kesikli ve sürekli yanıt değişkenleri için regresyon modelleri genelleştirilmiş lineer modeller sınıfı veya GLMs (McCullagh ve Nelder, 1989) altında toplanır ve farklı yanıt değişkenleri için istatistiksel metotların yaygın bir çatısını oluşturur. Burada bu modeller sınıfının bazı özelliklerinden bahsedilecektir.

Öncelikle iki önemli GLM tartışıldı; Lojistik ve Poisson regresyon modelleri ve daha sonra genel sınıf GLM tartışıldı.

2.4.1. Lojistik Regresyon

Bu model bir hastalığın varlığı veya yokluğu gibi ikili yanıt değişkenleri için kullanılır. Lojistik model, pozitif yanıt değişkeninin odds oranının logaritmasının açıklayıcı değişkenlerin lineer bir fonksiyonu olduğunu varsayar. Öyle ki;

$$\log \frac{\Pr(Y_i = 1)}{\Pr(Y_i = 0)} = \log \frac{\mu_i}{1 - \mu_i} = x_i' \beta \quad (2.7)$$

Lojistik regresyon ile lineer model arasındaki temel fark, doğrusallığın Y_i nin bekleneninin dönüştürülmüş haline uygulanması yani bekleneninin log odds dönüşümüne uygulanmasıdır. Böylece, regresyon katsayısı, β , x de ki her bir birim değişimde yanıt değişkeninin log odds unun değişimini temsil eder. İkili yanıt değişkenlerinin diğer bir özelliği ise Y_i nin varyansının tamamen onun ortalaması tarafından tanımlanmasıdır. Özellikle,

$$\text{Var}(Y_i) = E(Y_i) \{1 - E(Y_i)\} = \exp(x_i' \beta) / \{1 + \exp(x_i' \beta)\}^2$$

Bu, ortalamadan bağımsız olan, genellikle $\text{Var}(Y_i)$ nin bağımsız varsayıldığı, σ^2 , lineer modelle zıtlık halindedir.

2.4.2. Poisson Regresyonu

Poisson regresyonu veya log lineer modeller, yanıt değişkeninin zamanın sabit bir periyodunda meydana gelen olayların sayısını temsil ettiği problemlere uygulanır. Örnek; verilen zaman aralığında oluşan kriz sayısı gibi. Kesikli verinin negatif olmayan yapısından dolayı, makul varsayım kesikli verinin bekleneninin logaritmasının açıklayıcı değişkenlerin lineer fonksiyonu olmasıdır. Böylece

$$\log E(Y_i) = x_i' \beta \quad (2.8)$$

Burada, belirli bir açıklayıcı değişken için regresyon katsayısı, diğer bütün açıklayıcı değişkenlerin sabit tutulması ile, bu açıklayıcı değişkendeki bir birim artıştan önce veya sonra kesikli değişkenlerin bekleneninin logaritması gibi yorumlanabilmektedir. μ ortalamalı Y değişkeni için Poisson dağılımı,

$$p(y) = \frac{\exp(-\mu)\mu^y}{y!} \quad y = 0, 1, \dots$$

fonksiyonu ile ifade edilir. Lojistik regresyondaki gibi, Y_i nin poisson dağılımına sahip olması Y_i nin varyansının ortalama gibi tanımlandığını göstermektedir. Bu durumda varyans ve ortalama aynıdır.

$$\text{Var}(Y_i) = E(Y_i) = \exp(x_i'\beta)$$

2.4.3. Genel Sınıf

Lineer, lojistik ve poisson regresyon modelleri, aşağıda ki özellikleri taşıyan genelleştirilmiş lineer modellerin özel durumlarıdır.

İlk olarak, $\mu_i = E(Y_i)$ ortalama yanıt değişkeni, aşağıda ki eşitlik boyunca x_i açıklayıcı değişkenler vektörü ile ilgili olduğu varsayılır. Burada $h(\cdot)$ fonksiyonuna link fonksiyonu denir.

$$h(\mu_i) = x_i'\beta \quad \text{dır.} \quad (2.9)$$

- Lojistik regresyon için; $h(\mu_i) = \log\{\mu_i / (1 - \mu_i)\}$
- Poisson regresyon için $h(\mu_i) = \log(\mu_i)$

İkinci olarak, Y_i nin varyansı, ortalaması μ_i nin bilinen bir fonksiyonudur.

Yani;

$$\text{Var}(Y_i) = v_i = \varphi v(\mu_i)$$

Bu tanımda $v(\cdot)$ bilinen fonksiyonu, varyans fonksiyonu olarak isimlendirilir. φ ölçüm faktörü de GLM ailesinin bazı üyeleri için bilinen bir sabittir.

Üçüncü olarak, aşağıda ki olabilirlik fonksiyonu ile genelleştirilmiş lineer modellerin her bir sınıfın dağılımları üstel ailesinin bir üyesine karşılık gelir.

$$f(y_i) = \exp\left[\frac{y_i\theta_i - \psi(\theta_i)}{\varphi} + c(y_i, \varphi)\right] \quad (2.10)$$

θ_i parametresi doğal parametre olarak bilinir ve μ_i ile ilişkilidir. $\mu_i = \frac{\partial \psi(\theta_i)}{\partial \theta_i}$ dir.

Örneğin, poisson dağılımı aşağıda ki eşitlikler ile üstel ailenin özel bir durumudur.

$$\theta_i = \log \mu_i \quad \psi(\theta_i) = \exp(\theta_i) \quad c(y_i, \theta) = -\log(y_i!), \quad \theta = 1$$

Herhangi bir genelleştirilmiş lineer modelde, regresyon katsayısı, β , aynı tahmin denklemi tarafından tahmin edilebilir.

$$S(\beta) = \sum_{i=1}^m \left(\frac{\partial \mu_i}{\partial \beta} \right)' v_i^{-1} \{Y_i - \mu_i(\beta)\} = 0 \quad (2.11)$$

Burada $v_i = \text{Var}(Y_i)$ dir.

Dikkat edelim ki $S(\beta)$ olabilirlik fonksiyonunun logaritmasının türevidir. En çok olabilirlik tahmini $\hat{\beta}$ çözümü, iterasyonlu ağırlıklandırılmış en küçük karelerle elde edilebilir. Ayrıntı için Mc Cullagh ve Nelder (1989)' a bakılabilir.

Son olarak, geniş örneklerde $\hat{\beta}$, β ortalama ve (2.12) deki varyans ile Normal dağılıma sahiptir.

$$V = \left(\sum_{i=1}^m \left(\frac{\partial \mu_i}{\partial \beta} \right)' v_i^{-1} \frac{\partial \mu_i}{\partial \beta} \right)^{-1} \quad (2.12)$$

Bu varyans, (2.12) eşitliğinde ki β yı $\hat{\beta}$ ile değiştirerek elde edilecek \hat{V} tarafından tahmin edilebilir.

2.5. Yarı – Olabilirlik Fonksiyonları

Teorik çalışan istatistikçilerin birleştiği nokta, istatistiksel sonuç çıkarmada olabilirlik tarafından oynanan rolün önemi olduğudur. Bir olabilirlik fonksiyonu kurmada, parametre değerlerinin aralığı için, gözlemlenme olasılığı olan bütün ilişkili örneklem olasılıklarının genellikle ihtimal mekanizmasının belirtilme durumuna ihtiyacı vardır.

Böyle bir belirtme, ya mekanizmasının bilgisini yada önceki denemelerden benzer var olan verinin deneyimlerini içermektedir.

Bilinen olabilirlik fonksiyonlarının koşullarının hepsinin sağlanmadığı durumlarda yarı olabilirlik söz konusudur. Burada yarı olabilirlik fonksiyonu ve tahmini üzerinde durulacaktır.

2.5.1. Bağımsız Gözlemler Durumu

2.5.1.1. Kovaryans Fonksiyonları

Bağımlı değişkenler vektörü Y nin bileşenlerinin μ ortalama ve $\sigma^2 V(\mu)$ kovaryans matrisi ile bağımsız olduğunu varsayalım. σ^2 bilinmiyor ve $V(\mu)$ bilinen fonksiyonların matrisi olsun. İlgilenilen parametrelerin, β , x kovaryatları üzerinde μ nün bağımlılığına ilişkili olduğu varsayılır. Bu ilişkinin doğasıyla ilgilenmediğimiz için sadece $\mu(\beta)$ alınır. Böylece kovaryatları regresyon fonksiyonları içerisine dahil edilir ve σ^2 , β ya bağlı değildir. Y nin bileşenleri bağımsız olduğu için $V(\mu)$ matrisi köşegen matrisidir. Böylece;

$$V(\mu) = \text{diag}\{V_1(\mu), \dots, V_n(\mu)\}$$

$V_i(\mu)$ fonksiyonlarını ilgilendiren bir başka varsayım, $V_i(\mu)$ nin sadece μ nün i . bileşenine bağlı olmasıdır. Bağımsızlık altında, $V_i(\mu)$ için μ nün birkaç bileşenlerine bağlı olması mümkündür.

Yukarıdaki varsayım altında fonksiyonel bağımsızlık şöyle yazılabilir.

$$V(\mu) = \text{diag}\{V_1(\mu_1), \dots, V_n(\mu_n)\} \quad (2.13)$$

2.5.1.2. Yarı Olabilirlik Fonksiyonunun Yapısı

V veya y olarak alt indissiz yazdığımız Y bağımlı vektörünün tek bileşenini varsayalım. μ ve Y nin bir fonksiyonu olan U değişkenini

$$U = u(\mu; Y) = \frac{Y - \mu}{\sigma^2 V(\mu)}$$

biçiminde verilebilir ve aşağıdaki özelliklere sahiptir.

$$\begin{aligned} E(U) &= 0 \\ \text{Var}(U) &= \frac{1}{\{\sigma^2 V(\mu)\}} \\ -E\left(\frac{\partial U}{\partial \mu}\right) &= \frac{1}{\{\sigma^2 V(\mu)\}} \end{aligned} \quad (2.14)$$

(2.14) de verilen özelliklerin ispatı kolayca yapılabilir.

$E(U) = 0$ için;

$$E\left(\frac{Y-\mu}{\sigma^2 V(\mu)}\right) = \frac{1}{\sigma^2 V(\mu)} [E(Y) - \mu] = \frac{1}{\sigma^2 V(\mu)} [\mu - \mu] = 0$$

Var(U) için;

$$\begin{aligned} \text{Var}(U) &= E(U^2) - \underbrace{(E(U))^2}_0 \\ &= E\left(\left(\frac{Y-\mu}{\sigma^2 V(\mu)}\right)^2\right) = E\left(\frac{Y^2 - 2Y\mu + \mu^2}{\sigma^4 V^2(\mu)}\right) \\ &= \frac{1}{\sigma^4 V^2(\mu)} (E(Y^2) - 2E(Y)\mu + \mu^2) \\ &= \frac{1}{\sigma^4 V^2(\mu)} (E(Y^2) - 2\mu^2 + \mu^2) \\ &= \frac{1}{\sigma^4 V^2(\mu)} (E(Y^2) - \mu^2) = \frac{1}{\sigma^4 V^2(\mu)} (E(Y^2) - (E(Y))^2) \\ &= \frac{1}{\sigma^4 V^2(\mu)} \cdot V(Y) = \frac{\sigma^2 V(\mu)}{\sigma^4 V^2(\mu)} = \frac{1}{\sigma^2 V(\mu)} \\ \frac{\partial U}{\partial \mu} &= \frac{\partial \left(\frac{Y-\mu}{\sigma^2 V(\mu)}\right)}{\partial \mu} = \frac{-\sigma^2 V(\mu) - (Y-\mu)}{\sigma^4 V(\mu)^2} = -\frac{1}{\sigma^2 V(\mu)} \\ -E\left(\frac{\partial U}{\partial \mu}\right) &= -E\left(-\frac{1}{\sigma^2 V(\mu)}\right) = \frac{1}{\sigma^2 V(\mu)} \end{aligned}$$

Böylece yukarıdaki özelliklerle birlikte log yarı en çok olabilirlik fonksiyonu,

$$Q(\mu, y) = \int_y^\mu \frac{y-t}{\sigma^2 V(t)} dt \quad (2.15)$$

integrali ile ifade edilir. Y nin bileşenleri bağımsız varsayıldığından $Q(\mu; y) = \sum Q_i(\mu_i; y_i)$ dir.

Benzer şekilde, tek bir gözleme karşılık gelen yarı sapma fonksiyonu;

$$D(y; \mu) = -2\sigma^2 Q(\mu; y) = 2 \int_\mu^y \frac{y-t}{V(t)} dt \quad (2.16)$$

Burada $D(y; \mu)$, $y = \mu$ dışındaki bütün değerler için kesin pozitiftir. $D(y; \mu)$ sadece y ve μ ye bağlıdır, σ^2 ye bağlı değildir.

Şimdi bazı bilinen dağılımlar için log yarı olabilirlik fonksiyonlarını elde edeceğiz.

1. Normal Dağılım İçin Yarı Olabilirlik

$$f(Y = y) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty$$

$$\begin{aligned} L(y) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^n \left(\frac{y_i-\mu}{\sigma}\right)^2} \end{aligned}$$

$$\begin{aligned} \log L(y) &= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2} \sum_{i=1}^n \left(\frac{y_i-\mu}{\sigma}\right)^2 \\ &= -\frac{n}{2} \log 2\pi - n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \end{aligned}$$

$$\frac{\log \partial L(y)}{\partial(\mu)} = -\frac{1}{2\sigma^2} \cdot 2 \cdot (y - \mu) \cdot (-1) = \frac{y - \mu}{\sigma^2} = s(t)$$

Yarı olabilirlik fonksiyonu;

$$\begin{aligned} \int_y^\mu s(t) dt &= \int_y^\mu \frac{y-t}{\sigma^2} dt = \frac{1}{\sigma^2} \left[\left[yt - \frac{t^2}{2} \right] \Big|_y^\mu \right] \\ &= \frac{1}{\sigma^2} \left[y(\mu - y) - \left(\frac{\mu^2 - y^2}{2} \right) \right] \\ &= \frac{1}{\sigma^2} \left[\frac{2y\mu - 2y^2 - \mu^2 + y^2}{2} \right] = \frac{1}{\sigma^2} \left[-\frac{y^2 + 2y\mu - \mu^2}{2} \right] \\ &= \frac{1}{\sigma^2} \left[\frac{-(y - \mu)^2}{2} \right] \end{aligned}$$

2. Binom Dağılımı için Yarı Olabilirlik

$$V(\mu) = \mu(1 - \mu) \quad u = \frac{Y - \mu}{\sigma^2 \mu(1 - \mu)}$$

$$\begin{aligned}
Q(\mu) &= \int_y^{\mu} \frac{y-t}{\sigma^2 t(1-t)} dt = \frac{1}{\sigma^2} \left[\int_y^{\mu} \frac{y-t}{t(1-t)} dt \right] = \frac{1}{\sigma^2} \left[\int_y^{\mu} \frac{y}{t(1-t)} dt - \int_y^{\mu} \frac{t}{t(1-t)} dt \right] \\
&= \frac{1}{\sigma^2} \left[y \int_y^{\mu} \frac{1}{t(1-t)} dt - \int_y^{\mu} \frac{1}{1-t} dt \right] = \frac{y}{\sigma^2} \ln \left| \frac{t}{1-t} \right| + \frac{1}{\sigma^2} \ln |1-t| \\
&= \frac{y}{\sigma^2} \ln \left(\frac{\mu}{1-\mu} \right) + \frac{1}{\sigma^2} \ln(1-\mu)
\end{aligned}$$

3. Poisson Dağılımı için Yarı Olabilirlik

$$V(\mu) = \mu \quad \text{Var}(Y) = \sigma^2 \mu \quad U = \frac{Y - \mu}{\sigma^2 \mu}$$

$$\begin{aligned}
Q(\mu) &= \int_y^{\mu} \frac{y-t}{\sigma^2 t} dt = \frac{1}{\sigma^2} \left[\int_y^{\mu} \frac{y}{t} dt - \int_y^{\mu} 1 dt \right] \\
&= \frac{1}{\sigma^2} \left[y \log(t) \Big|_y^{\mu} - \mu + y \right] \\
&= \frac{1}{\sigma^2} [y \log \mu - \mu] + \frac{1}{\sigma^2} [y - y \log(y)] \\
&= y \log(\mu) - \mu
\end{aligned}$$

2.5.1.3. Parametre Tahmini

β regresyon parametreleri için $Q(\mu; y)$ nin diferansiyelleşmesiyle elde edilen yarı olabilirlik tahmin denklemleri $U(\hat{\beta}) = 0$ formunda yazılabilir. Böylece

$$U(\beta) = D^T V^{-1} \frac{(Y - \mu)}{\sigma^2} \quad (2.17)$$

dir ve yarı skor fonksiyonu olarak isimlendirilir. D nin bileşenleri $n \times p$ boyutlu $D_{ir} = \frac{\partial \mu_i}{\partial \beta_r}$ türevleridir.

$U(\beta)$ nin kovaryans matrisi $\left(\frac{\partial U(\beta)}{\partial \beta} \right)$ nin negatif beklenen değeri aşağıdaki gibidir.

$$\dot{I}_\beta = \frac{D^T V^{-1} D}{\sigma^2} \quad (2.18)$$

Yarı olabilirlik fonksiyonları için bu matris sıradan olabilirlik fonksiyonlarında Fisher bilgi matrisinin oynadığı rolü oynar ve $\hat{\beta}$ nın asimtotik kovaryans matrisi;

$$\text{cov}(\hat{\beta}) \cong \dot{I}_\beta^{-1} = \sigma^2 (D^T V^{-1} D)^{-1}$$

tahminidir. Böylece parametre $\hat{\beta}$ ya yeterince yakın keyfi bir $\hat{\beta}_0$ değeri ile başlanarak, Fischer skor ile Newton-Raphson tarafından geliştirilen parametre tahminleri ile

$$\hat{\beta}_1 = \hat{\beta}_0 + (\hat{D}_0^T \hat{V}_0^{-1} \hat{D}_0)^{-1} \hat{D}_0^T \hat{V}_0^{-1} (y - \hat{\mu}_0)$$

biçiminde ifade edilir. Gerçek değere yakınsayana kadar iteratif işlem devam eder.

$\hat{\beta}$ yarı olabilirlik tahmini tutarlı durum ortaya çıkana kadar iterasyonla sağlanabilir.

2.5.2. Bağımlı Gözlemler Durumu

2.5.2.1. Yarı Olabilirlik Tahmin Eşitlikleri

Kabul edelim ki $V(\mu)$, $V_{ij}(\mu)$ bilinen fonksiyonlarının bir simetrik pozitif tanımlı $n \times n$ matrisi olduğunda $\text{cov}(Y) = \sigma^2 V(\mu)$ diagonal olmasın. (2.17) de verilen skor fonksiyonunun $U_r(\beta)$ bileşenleri aşağıdaki özelliklere sahiptir.

- ✓ $E\{U_r(\beta)\} = 0$
- ✓ $\text{cov}\{U(\beta)\} = \frac{D^T V^{-1} D}{\sigma^2} = \dot{I}_\beta$
- ✓ $-E\left(\frac{\partial U_r(\beta)}{\partial \beta_s}\right) = \frac{D^T V^{-1} D}{\sigma^2}$

$U(\beta)$ ' ya sanki bir log olabilirlik fonksiyonunun β ya göre türeviymiş gibi davranılabilir.

Uygun sınırlayıcı koşullar altında, tahmin denkleminin $\hat{\beta}$ kökü;

$$U(\hat{\beta}) = \hat{D}^T \hat{V}^{-1} (Y - \hat{\mu}) = 0 \quad \text{dır.} \quad (2.19)$$

Yaklaşık olarak β için yansız ve sınırlayıcı varyans aşağıdaki gibi asimtotik olarak normal dağılımlıdır.

$$\text{cov}(\hat{\beta}) = \sigma^2 (D^T V^{-1} D)^{-1} = I_{\beta}^{-1} \quad (2.20)$$

$\hat{\beta}$ nin asimtotik normalliği ve tutarlılığı Liang ve Zeger (1986) ve Zeger & Liang (1986) tarafından gösterilmiştir.

2.5.2.2. Yarı Olabilirlik Fonksiyonu

Skor fonksiyonu $U(\beta)$ bir log olabilirliğin veya yarı olabilirliğin gradient vektörü ise $U(\beta)$ nın β ya göre türev matrisinin simetrik olması gerek ve yeterlidir. Genelde $r \neq s$ için

$$\frac{\partial U_r(\beta)}{\partial \beta_s} \neq \frac{\partial U_s(\beta)}{\partial \beta_r} \quad \text{dir.} \quad (2.21)$$

Bağımsız gözlemler durumuna benzer şekilde

$$Q(\mu; y, t(s)) = \sigma^{-2} \int_{t(s)=y}^{t(s)=\mu} (y - t)^T \{V(t)\}^{-1} dt(s) \quad (2.22)$$

elde edilir. Bu integral yola bağlı olan integraldir. Sadece başlangıç ve bitiş noktaları olan $t(s) = y$ ve $t(s) = \mu$ ye bağlıdır. Parametre tahminleri $Q(\mu; y, t(s))$ göz önüne alınarak ve bağımsız gözlemler durumuna benzetilerek yapılır.

3. MATERYAL VE METOT

3.1. Uzunlamasına Veri İçin Genelleştirilmiş Lineer Modeller

Bu bölümde Genelleştirilmiş Lineer Modelleri kullanarak kesikli uzunlamasına verilerin analizini sunulacaktır. Lineer modelleri kullanarak regresyon parametrelerinin tahmini; verideki korelasyonları hesaba katmasına rağmen, genel yapı korelasyondan bağımsızdır. Kesikli bağımlı değişken için; örneğin lojistik regresyon korelasyon hakkında farklı varsayımlar kabul eder. Araştırmalarda bu farklılık dikkatle göz önüne alınmalıdır.

3.1.1. Marjinal Modeller

Marjinal modelde; kitle ortalama yanıt değişkeni, açıklayıcı değişkenlerin fonksiyonu gibi modellenir. Bağımlı değişkenin açıklayıcı değişken üzerindeki regresyonu birey içi korelasyondan ayrı ayrı modellenir. Regresyon katsayıları bireylerden ziyade kitle için yorumlanır. Regresyonda $E(Y_{ij})$ marjinal bekleneni, açıklayıcı değişkenlerin fonksiyonu gibi modellenir. Marjinal model aşağıdaki varsayımlara sahiptir;

- Bağımlı değişkenin marjinal bekleneni, $E(Y_{ij}) = \mu_{ij}$, x_{ij} açıklayıcı değişkenleri için $h(\mu_{ij}) = x'_{ij}\beta$ dir. Burada h link fonksiyonudur.
- Marjinal varyans, $Var(Y_{ij}) = v(\mu_{ij})\phi$ dir. v bilinen varyans fonksiyonu ve ϕ ölçek parametresidir.
- Y_{ij} ve Y_{ik} arasındaki korelasyon marjinal ortalamaların ve muhtemelen parametrelerin fonksiyonudur ve α ile gösterilir. Örneğin; $Kor(Y_{ij}, Y_{ik}) = \rho(\mu_{ij}, \mu_{ik}; \alpha)$, $\rho(\cdot)$ bilinen bir fonksiyondur.

Yukarıdaki varsayımlar aşağıdaki gibi örneklenebilir.

- $logit(\mu_{ij}) = \log \frac{\mu_{ij}}{1 - \mu_{ij}} = \log \frac{Pr(Y_{ij} = 1)}{Pr(Y_{ij} = 0)} = \beta_0 + \beta_1 x_{ij}$
- $Var(Y_{ij}) = \mu_{ij}(1 - \mu_{ij})$
- $Kor(Y_{ij}, Y_{ik}) = \alpha$

3.1.2. Tesadüfi Etki Modelleri

Lineer tesadüfi etkiler modeli, bağımlı değişkenin bir bireyden diğerine değişen regresyon katsayıları ile açıklayıcı değişkenlerin lineer fonksiyonu olduğunu varsayar. Bu değişkenlik ölçülmemiş faktörlerden dolayı doğal heterojenliği etkiler. Lineer tesadüfi etkiler modeli bir birey için tekrarlı gözlemlerin bağımsız olduğunu varsayar. Varsayılır ki her bir birey için veriler Genelleştirilmiş Lineer Modellere uyan bağımsız gözlemlerdir. Kısaca tesadüfi etkilere ilişkin GLM nin genel tanımı aşağıdaki gibi verilebilir.

Y_{i1}, \dots, Y_{in_i} bağımlı değişkenleri karşılıklı bağımsızdır ve U_i açıklayıcı değişkenler yapısı ($i=1, \dots, m$) için $f(y_{ij}|U_i) = \exp\left\{\frac{y_{ij}\theta_{ij} - \Psi(\theta_{ij})}{\phi} + c(y_{ij}, \phi)\right\}$ yoğunluk fonksiyonu ile GLM' ye uygundur.

Burada koşullu momentleri, $\mu_{ij} = E(Y_{ij}|U_i) = \Psi'(\theta_{ij})\phi$ ve $v_{ij} = \text{Var}(Y_{ij}|U_i) = \Psi''(\theta_{ij})\phi$ dir. $h(\mu_{ij}) = x'_{ij}\beta^* + d'_{ij}U_i$, $v_{ij} = v(\mu_{ij})\phi$, $h(\mu_{ij})$ link fonksiyonu, $v(\mu_{ij})$ varyans fonksiyonu ve d_{ij} , x_{ij} nin alt kümesidir.

Özetlemek gerekirse, bir tesadüfi etkiler modelinde temel fikir bireyden bireye değişen regresyon katsayılarında doğal heterojenlik olmasıdır ve bu heterojenlik bir olasılık dağılımı tarafından temsil edilebilir. Tesadüfi etki modelleri amaç kitle ortalamasından ziyade bireyler hakkında sonuç çıkarımı yapmak olduğunda çok daha fazla yararlı olabilir.

Bir başka modelde geçmiş değerlerin etkisini göz önüne alan Geçiş (Markov) Modelleridir. Bu modeller Korn ve Whittemore (1979), Ware et al (1988), Wong (1986), Zeger ve Quagish (1988) ve Kaufmann (1987) tarafından incelenmiştir. Bu çalışmada üzerinde durulmamıştır.

3.2. Uzunlamasına Kesikli Yanıt Değişkenleri İçin Regresyon Modelleri

Uzunlamasına çalışmalarda her bir birey üzerindeki tekrarlı ölçümlerin zamanlarında sıralama olduğu kanısı vardır. Başlangıçta N bireyden her birinin aynı p durumda gözlemlendiği varsayılır. İlk olarak yanıt değişkenlerinin ikili olduğu varsayılırsa $Y_i = (Y_{i1}, \dots, Y_{ip})^T$ $p \times 1$ vektörü oluşturulabilir. Öyle ki, ikili tesadüfi değişken $Y_{ip} = 1$, eğer birey i, T=t anında bağımlı değişkene (başarı) sahipse 1 aksi takdirde 0. Burada ilk olarak ikili (binary) yanıt değişkenleri üzerinde durulacaktır. Her bir birey $T \times 1$ boyutlu p

durumunda x_{ip} açıklayıcı vektöre sahiptir ve $X_i = (x_{i1}, \dots, x_{ip})^T$, birey i için $P \times T$ boyutlu açıklayıcı matrisini temsil eder. Böylece i . birey için veri (Y_i, X_i) gözlemlerinden oluşur.

Y_{it} nin marjinal dağılımı Bernoulli kabul edilir. O halde;

$$f(y_{ip}|X_i) = \exp[y_{ip}\theta_{ip} - \log\{1 + \exp(\theta_{ip})\}] \quad (3.3)$$

$$\theta_{ip} = \log\left[\frac{\mu_{ip}}{1 - \mu_{ip}}\right] = x_{ip}^T \beta$$

Burada $\mu_{ip} = \mu_{ip}(\beta) = E(Y_{ip}) = \text{pr}(Y_{ip} = 1 | x_{ip}, \beta)$, t zamanında ki başarı olasılığı ve β ise $T \times 1$ boyutlu parametre vektörüdür. İkili yanıt değişkenleri için lojit link fonksiyonunun seçimi doğal bir durumdur. $\mu_{ip}(\beta)$ vektörünü aşağıdaki gibi ifade edebiliriz.

$$\mu_i(\beta) = E(Y_i) = (\mu_{i1}, \dots, \mu_{ip})^T$$

Eğer yanıt değişkenlerinin bağımsız olduğu varsayılıyorsa, yanıt değişkenlerinin ortak dağılımı;

$$f(y_i|X_i) = \exp\left(\sum_{t=1}^T y_{it}\theta_{it} - \sum_{t=1}^T \log\{1 + \exp(\theta_{it})\}\right) \text{ olur.} \quad (3.4)$$

Logaritmik olabilirlik fonksiyonunun β ' ya göre türevi alınıp, sıfıra eşitlenerek parametre tahminleri elde edilebilir. Bu amaçla türevde zincir kuralı uygulanırsa

$$\frac{\partial l_i}{\partial \beta} = \left(\frac{\partial \mu_i}{\partial \beta}\right)^T \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial l_i}{\partial \theta_i}$$

elde edilir. Üstel dağılım ailesini kullanarak, kanonik parametre θ_i ye göre türevi aşağıdaki gibidir.

$$\frac{\partial l_i}{\partial \theta_i} = y_i - E(Y_i) = y_i - \mu_i$$

Ayrıca $\frac{\partial \mu_i}{\partial \theta_i} = \text{cov}(Y_i)$ olduğu bilinmektedir. Bu sonuçlar kullanılarak

$$\frac{\partial l_i}{\partial \beta} = \left(\frac{\partial \mu_i}{\partial \beta}\right)^T \text{var}^{-1}(Y_i)(y_i - \mu_i)$$

elde edilir. $\theta_i = X_i \beta$ olduğunu varsayımı altında,

$$\left(\frac{\partial \mu_i}{\partial \beta} \right) = X_i^T \Delta_i$$

yazılabilir. Burada, $\Delta_i = \text{diag}[\text{var}(Y_{i1}), \dots, \text{var}(Y_{ip})]$, $P \times P$ boyutlu köşegen matrisidir.

Böylece en çok olabilirlik tahmin edicisi (MLE) $\hat{\beta}$, aşağıdaki eşitliğin çözümüdür.

$$\begin{aligned} \sum_{i=1}^N \frac{\partial l_i}{\partial \beta} &= \sum_{i=1}^N \left(\frac{\partial \mu_i}{\partial \beta} \right)^T \text{var}^{-1}(Y_i)(y_i - \mu_i) \\ &= \sum_{i=1}^N X_i^T (y_i - \mu_i) = 0 \end{aligned} \quad (3.5)$$

μ_i , $\hat{\beta}$ 'ya bağlı çözümünüyle bulunur.

Aynı bireydeki tekrarlı ölçümlerin ilişkili olması gerçeğine rağmen sıradan lojistik regresyon en çok olabilirlik tahmin edicisi (tekrarlı ölçümlerin bağımsız olduğunu varsayar) tutarlı ve asimtotikli normal tahmin edicileri üretir. Yine de bağımsızlık altındaki ortak olabilirlik ikili yanıt değişkenleri aralarında mümkün ilişkilere önem vermemektedir. Sonuç olarak, tahmin edilen bilgi matrisinin tersi tahmin edilen parametrelerin asimtotik varyansının tutarsız tahmin edicilerini verebilir.

Genel olarak, zamanın fonksiyonu olmayan etkilerin standart hatası düşük olarak, zamanın bir fonksiyonu olan etkilerin standart hatası ise olduğundan fazla tahmin edilme eğilimindedir. Bu problemden kaçınmak için Liang ve Zeger (1986) yanıt değişkenleri arasındaki ilişki ne olursa olsun tutarlı olan tahmin edilen parametrelerin varyansında 'robust' tahmini kullanmayı önermişlerdir. Bu 'robust' varyans 1967 yılında Huber tarafından ve yakın zamanda 1982 de White ve Royal tarafından ileri sürülmüştür. Bu sonuçla uzunlamasına ikili yanıt değişkenlerini analizde basit bir yaklaşım, robust(güçlü) varyans doğrulamasını takip ederek sıradan lojistik regresyonu kullanmaktır. Yanıt değişkenleri arasındaki korelasyon eğer çok yüksek değilse Zeger (1988) bu tahmin edicilerin güçlü ve etkili olduğunu önermiştir.

Yukarıda bahsedilen temel en çok olabilirlik fonksiyonlarının kullanımıyla parametre tahmini iteratif yöntemlerin zorluğu, uzunlamasına verilerin kompleksliği ve korelasyon yapısının zorluğu nedeniyle kolay değildir. Bu amaçla, bu zorlukları biraz daha kolaylaştıran ve korelasyon yapısı yanlış bile olsa β için tutarlı tahminler veren Genelleştirilmiş tahmin denklemleri (GTD) üzerinde durulacaktır. Bu metot, varyans-kovaryans yapısının tutarlı tahminlerini verir.

3.3. Genelleştirilmiş Tahmin Denklemleri (GTD)

Yakın zamanlarda gözlemlerin korelasyonlu olduğu durumlar için modeller geliştirilmiştir. Genelleştirilmiş lineer modellerin uygulamaları mevcut iken aynı birim üzerinde yapılan tekrarlı bağımlı ölçümlerin olduğu birçok durum vardır ve böylece bu tanım korelasyonlu gözlemlerin kümesini oluşturur. Zaman boyunca aynı birey üzerinde bağımlı değişkenin tekrarlı ölçümü ile yapılan uzunlamasına çalışma olduğunda klasik örnek olur. Bu tip çalışmanın avantajı zamanla bireyin kendi içindeki değişimleri sabit zamandaki farklılıklardan ayırt edebilmektir. Aynı birey üzerindeki gözlemler ilişkilidir. Korelasyonun varlığı maksimum olabilirliğin kullanımını şüpheli kılar, sonuç olarak yarı olabilirlik metodu kullanılacaktır.

Son zamanlarda gözlemlerin korelasyonlu olduğu durumlar için model geliştirme üzerine ilgi artmaktadır.

Bilinen veri yapısı

$$\begin{array}{cccccc}
 1 & x_{11} & x_{12} & \cdots & x_{1k} & y_1 \\
 2 & x_{21} & x_{22} & \cdots & x_{2k} & y_2 \\
 \vdots & \vdots & \vdots & & \vdots & \vdots \\
 m & x_{m1} & x_{m2} & \cdots & x_{mk} & y_m
 \end{array} \quad (3.6)$$

biçimindedir. Burada x_{ij} vektörü i . birey üzerinde açıklayıcı değişken veya j . değişken değerlerini kapsar. y_i vektörü ise i . bireydeki bağımlı değişkenleri kapsar.

Vektörler t boyutludur. Yani her bir bireyde t “regresyon” gözlemleri vardır. t gösterimi için mantık biomedikal uygulamalardan gelir öyle ki t zaman periyotlarının sayısından söz eder ve veriler her bir birey için zaman üzerinden alınır. Bu uygulamalarda x_1, x_2, \dots, x_m değişkenleri belirli bir ilacın doz değerleri veya bireyle ilgili bazı açıklayıcı karakteristikler olabilir. Örneğin; kalp atış hızı veya kan basıncı gibi. Tipik olarak bağımlı değişkenler üstel aileden dağılımlara uyacaktır. İlk olarak normal dağılım, birim link durumu göz önüne alınacaktır. Sonuç olarak model lineerdir ve çok değişkenli normal dağılımla ilgilenilmektedir. Aşağıdaki modeli göz önüne alalım.

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix} \quad (3.7)$$

Burada X_1, X_2, \dots, X_m matrisleri eşitlik (3.7) deki regresyon bilgilerini model formuna koyan model matrisleri böylece bu X_i ler ana etkilerin modele katkısını, herhangi bir etkileşimleri, kategorik değişkenleri, quadratik terimleri, vb,... tanımlar. Sonuç olarak X_i $j=1, 2, \dots, m$ $t \times m$ boyutludur. Eğer model x ler de lineerse, o zaman $p=m+1$ dir. Belirli birey için aşağıdaki model geçerlidir.

$$y_i = \beta_0 + X_i \beta + \varepsilon_i \quad i=1, 2, \dots, m$$

$\beta : m \times 1$ ve $y_i : X_i \beta$ ve V varyans-kovaryans matrisi ile çok değişkenli normal dağılıma sahiptir. V matrisi $t \times t$ boyutludur ve i . bireydeki gözlemlerin korelasyonu için hesaplardır. $N=t \times m$ olsun buda deneydeki toplam örneklem genişliğidir.

3.3.1. Korelasyon Matrisi R'nin Etkisi

Normal durumda her bir birey aynı V matrisine sahip olur çünkü her bir birey için korelasyon yapısı aynı varsayılır ve varyans homojen varsayılır, σ^2 ölçü parametresi tarafından tanımlanır. Böylece;

$$\begin{aligned} \text{var}(\varepsilon_i) = V &= \begin{bmatrix} \sigma^2 & \sigma_{12} & \dots & \dots & \sigma_{1s} \\ & \sigma^2 & \sigma_{23} & \dots & \sigma_{2s} \\ & & \ddots & & \vdots \\ & & & \ddots & \vdots \\ & & & & \sigma^2 \end{bmatrix} & \forall i, j \text{ için} \\ &= R\sigma^2 \end{aligned} \quad (3.8)$$

R $t \times t$ boyutlu, i . bireyde $j \neq j'$ için $\rho_{j,j'} = \text{corr}(y_{ij}, y_{ij'})$ elemanlarıyla korelasyon matrisidir. Tesadüfi hatalar için toplam varyans-kovaryans matrisi aşağıdaki gibidir.

$$\text{var}(\varepsilon) = V^* = \begin{bmatrix} V & & & 0 \\ & V & & \\ & & \ddots & \\ 0 & & & V \end{bmatrix}$$

Her bir birey için gösterilen eşit varyans-kovaryans matrisleri ile köşegen matrisidir. R nin öncelikli olarak tartışılması önemlidir.

Genelleştirilmiş Tahmin Denklemlerinde, kullanıcı sıkça çalışılan korelasyon matrisi denen korelasyon yapısını kullanır. Bunun nedeni korelasyonun ne olduğunun bilinmemesidir. Bu durumda çalışılan korelasyon matrisi kullanılır.

Yaygın korelasyon yapıları;

i. Herhangi iki yanıt değişkeni arasındaki korelasyonun bilinmediği durumdur. (Belirlenmemiş Durum)

$$\text{kor}(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j = k \\ \alpha_{jk} & j \neq k \end{cases}$$

$$R(\alpha) = \begin{pmatrix} 1 & \alpha_{12} & \dots & \alpha_{1k} \\ \alpha_{12} & 1 & \dots & \alpha_{2k} \\ \vdots & \vdots & & \vdots \\ \alpha_{1k} & \alpha_{2k} & \dots & 1 \end{pmatrix}$$

Tahmin edicisi aşağıdaki gibi bulunur.

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

ii. Değişebilir durum: Grup içi bireyler arası tüm korelasyonların eşit olduğu durumdur.

$$\text{kor}(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j = k \\ \alpha & j \neq k \end{cases}$$

$$R(\alpha) = \begin{pmatrix} 1 & \alpha & \alpha & \dots & \alpha \\ \alpha & 1 & \alpha & \dots & \alpha \\ \vdots & \vdots & & & \vdots \\ \alpha & \alpha & \alpha & \dots & 1 \end{pmatrix}$$

Bu durumda korelasyon matrisi tahmin edilmelidir.

$$\hat{\alpha} = \frac{1}{(N''-p)\phi} \sum_{i=1}^K \sum_{j \neq k} e_{ij} e_{ik}$$

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

iii. Bağımsız; Bütün korelasyonların 0 olduğu varsayıldığı durumdur.

$$\text{kor}(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases}$$

$$R(\alpha) = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

iv. M-Bağımlı; Sadece birbirine yakın m gözlem arasındaki korelasyonun sıfır olmadığı durumdur.

$$\text{kor}(Y_{ij}, Y_{i,j+t}) = \begin{cases} 1 & t=0 \\ \alpha_i & t=1,2,\dots,m \\ 0 & t>m \end{cases}$$

$$R(\alpha) = \begin{pmatrix} 1 & \alpha_1 & \dots & \alpha_{t-1} \\ \alpha_1 & 1 & \dots & \alpha_{t-2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{t-1} & \alpha_{t-2} & \dots & 1 \end{pmatrix}$$

v. Otoresif (AR); birbirine yakın gözlemler arasındaki korelasyonun, yakın olmayan korelasyonlara nazaran daha yüksek değere sahip olduğu durumdur. Bu gözlemler daha ileri zamanlarda küçük korelasyonlara sahiptir. İlişki şöyle verilebilir; $\rho_{ij} = \rho^{|i-j|}$. ρ veriden tahmin edilir.

$$\text{kor}(Y_{ij}, Y_{ik}) = \begin{cases} \alpha^{|j-k|} & j \neq k \\ 1 & j = k \end{cases}$$

Birinci dereceden otoresif süreç, herhangi iki gözlem arasındaki örnek aralık uzunluğunu aynı varsayar ve tahmin edicisi aşağıda verilmiştir.

$$\hat{\alpha} = \frac{1}{(K_1 - p)\phi} \sum_{i=1}^K \sum_{j \leq n_i - 1} e_{ij} e_{ik}$$

$$K_1 = \sum_{i=1}^K (n_i - 1)$$

$$R(\alpha) = \begin{pmatrix} 1 & \alpha & \dots & \alpha^{t-1} \\ \alpha & 1 & \dots & \alpha^{t-2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^{t-1} & \alpha^{t-2} & \dots & 1 \end{pmatrix}$$

vi. Sabit: R_i çalışma matrisini analizci belirler. Çalışma korelasyonu bu durumda ölçülemez.

$$\text{kor}(Y_{ij}, Y_{ik}) = r_{jk}$$

$$\begin{pmatrix} 1 & r_{12} & \cdots & r_{1t} \\ r_{12} & 1 & \cdots & r_{2t} \\ \vdots & \vdots & & \vdots \\ r_{1t} & r_{2t} & \cdots & 1 \end{pmatrix}$$

Yukarıdaki tanımlamalarda; $e_{ij} = \frac{y_{ij} - \mu_{ij}}{\sqrt{v(\mu_{ij})}}$ olan Pearson artığıdır. $\hat{\phi} = \frac{1}{N-p} \sum_{i=1}^K \sum_{j=1}^{n_i} e_{ij}^2$

dağılım parametresidir ve $N = \sum_{i=1}^K n_i$ ölçümlerin toplam sayısıdır.

GTD, korelasyon yapısı yanlış olsa bile β nın tutarlı tahmin edicisini verir çünkü prosedürlerimiz varyans-kovaryans yapısının tutarlı tahminlerini kullanır. (Myers, R.H, Montgomery, D.C)

3.3.2. Normal Durumda İteratif Prosedür ve Birim Link

R^* bütün deney boyunca korelasyon yapısını tanımlasın. Bu nedenle;

$$R^* = \begin{bmatrix} R & & & \\ & R & & \\ & & \ddots & \\ & & & R \end{bmatrix} \quad (3.9)$$

$R^* = ts \times ts$ boyutludur.

$V^* = \sigma^2 R$ olduğundan β için genelleştirilmiş en küçük kareler tahmin edicisinde V^* yı R^* ile değiştirilebilir. Sonuç olarak tahmin için iteratif düzen aşağıdaki gibidir.

1.adım: β nın tahmini için SEK i kullan. Ona b_0 de.

2.adım: b_0 'ı artıkları hesaplamada kullan ve buradan

3.adım: tahmin korelasyonlarını kullan ve b' yi hesapla.

$$b = (X' \hat{R}^{*-1} X)^{-1} X' \hat{R}^{*-1} y, \quad \hat{R} \text{ tahmin edilen korelasyonları kapsar.}$$

4.adım: Artıkları ve korelasyonları hesaplamaya adım 2'ye git.

5.adım: Yakınsama olana dek tekrarla.

3.3.3. σ^2 Ölçek Parametresi

Katsayıların standart hataları gibi korelasyonların hesabı, σ^2 ölçek parametresini gerektirir. Normal durum için ölçek parametresi aşağıdaki gibidir.

$$\hat{\sigma}^2 = \text{tr} \left[\frac{\sum_{i=1}^m (y_i - X_i b)(y_i - X_i b)'}{N - p} \right] \quad (3.10)$$

$X_i b$: tahmin edilen ortalama bağımlı değişken vektörü.

$(y_i - X_i b)(y_i - X_i b)'$ i. birey için artıkların kareler ve çarpımlar toplamı matrisi.

Eşitlik (3.10) da ki İz operatörü $N=tm$ veri noktası üzerinden artık kareler toplamını meydana getirir. $N-p$ 'hata serbestlik derecesi' rolünü üstlenir. σ^2 tahminin kolaylığı normal dağılımın durumunda homojen varyans varsayımından kaynaklanmaktadır. GLM' de diğer dağılımları tartışırken artıklar, sabit olmayan varyansı açıklamada standartlaştırılmış olacaktır.

3.3.4. Katsayıların Standart Hataları

Anlatılan iteratif prosedür, R de ki ağırlıkları her bir iterasyonda güncellenen, iterasyonlu yeniden ağırlıklandırılmış en küçük karelerle aynıdır. b nin β nin tahmininde varyans-kovaryansının hesaplanmasında iki yol vardır. Birincisi, katsayıların varyans-kovaryans matrisinin model tabanlı tahminini desteklemektedir. Bu prosedür, varsayılan korelasyon yapısının doğru olduğu varsayımına dayanır. β nin tahmini R doğru olmadığında bile tutarlı tahmin edicidir. İkincisi eğer R doğru ise o zaman standart genelleştirilmiş en küçük kareler prosedürü b nin model tabanlı varyans-kovaryansını şöyle önerir:

$$\text{var}(b) = (X' R^{*-1} X)^{-1} \sigma^2 \quad (3.11)$$

Böylece standart hatalar aşağıdaki matrisin köşegen elemanlarından gelecektir.

$$(X' R^{*-1} X)^{-1} \hat{\sigma}^2 = \left[\sum_{j=1}^s (X_j' \hat{R}^{-1} X_j) \right]^{-1} \hat{\sigma}^2$$

Katsayıların standart hataları için alternatif bir tahmin prosedürü çok daha güçlü (sağlam) tahmin edicilere izin verir yani biri R nin doğru olmayan belirtilmesine karşı duyarsızdır. Bu sağlam tahmin edicinin başlangıcı, genelleştirilmiş en küçük kareler tahmin edicisinde y -varyans operatörünün kullanımından gelir.

İdeal olarak

$$\begin{aligned} \text{var}(\mathbf{b}) &= \text{var} \left[\mathbf{X}' \mathbf{R}^{*-1} \mathbf{X} \right]^{-1} \mathbf{X}' \mathbf{R}^{*-1} \mathbf{y} \\ &= \left[\mathbf{X}' \mathbf{R}^{*-1} \mathbf{X} \right]^{-1} \mathbf{X}' \mathbf{R}^{*-1} [\text{var}(\mathbf{y})] \mathbf{R}^{*-1} \mathbf{X} \left[\mathbf{X}' \mathbf{R}^{*-1} \mathbf{X} \right]^{-1} \end{aligned} \quad (3.12)$$

Şimdi $\text{var}(\mathbf{y}) = \mathbf{V}^*$ doğru fakat bilinmeyen varyans-kovaryans matrisidir. Belirtelim ki eğer korelasyon yapısı \mathbf{R} doğru ise, o zaman $\mathbf{V}^* = \mathbf{R}^* \sigma^2$ olur ve denklem (3.12), denklem (3.11) e indirgenir. Bununla birlikte (3.12) deki güçlü yaklaşım tahminleri deneysel olarak veridendir. Bu nedenle denklem (3.12) den sağlanan tahmin ediciye aynı zamanda ampirik tahmin edici denir. Sonuç olarak (3.12) deki \mathbf{V}^* aşağıdaki ile yer değiştirir.

$$\hat{\mathbf{V}}^* = \begin{bmatrix} \hat{\mathbf{V}} & & 0 \\ & \hat{\mathbf{V}} & \\ 0 & & \ddots \\ & & & \hat{\mathbf{V}} \end{bmatrix}$$

$\hat{\mathbf{V}}$ denklem (3.8) den alınır. Sıkça ampirik veya güçlü tahmin ediciler ve model-temelli tahmin ediciler benzer sonuçları verir ki bu da korelasyon yapısının iyi bir seçim varsayıldığını göstermektedir.

3.4. Çokterimli Dağılım

n bağımsız gözleme sahip olan \mathbf{Y} tesadüfi değişkeni varsayalım. Bunlardan y_1 bağımlı değişken 1. nci kategoride, y_2 bağımlı değişken 2. nci kategoride v.b.... yer alıyorsa;

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \end{bmatrix} \quad \text{ve} \quad \sum_{i=1}^J y_i = n \quad (3.13)$$

olur. $i=1,2,\dots,J$ olarak indekslenen kesikli değerlerden birini alabilen Y_i tesadüfi değişkeninde i . bağımlı değişkenin j . kategoriye düşme ihtimali aşağıda verilmiştir.

$$\pi_{ij} = \Pr\{Y_i = j\}$$

Burada ihtimaller toplamı 1'e eşittir. Yani $\sum_{i=1}^J \pi_i = 1$ dir.

Multinomial dağılım ise aşağıdaki gibi gösterilebilir.

$$f(\mathbf{y}|\mathbf{n}) = \frac{n!}{y_1!y_2!\dots y_j!} \pi_1^{y_1} \pi_2^{y_2} \dots \pi_j^{y_j} \quad Y \sim \text{Mult}(n, \pi) \quad (3.14)$$

Eğer $j=2$ ise o zaman $\pi_2 = 1 - \pi_1$, $y_2 = n - y_1$ olur ki bu da binomial dağılıma karşılık gelir.

Bu dağılım için $E(Y_j) = n\pi_j$, $\text{Var}(Y_j) = n\pi_j(1 - \pi_j)$ ve $\text{cov}(Y_j, Y_k) = -n\pi_j\pi_k$ dır.

Gruplandırılmış veriler için, çeşitli kategorilerdeki bağımlı değişkenlerin sayısını temsil eden yardımcı bir tesadüfi değişken kullanmak uygun olacaktır. Yani, n_i 'nin i . nci gruptaki durumların sayısını göstermesi ve Y_{ij} 'nin y_{ij} gözlem değeri ile bağımlı değişkenin i . nci grupta j . nci kategorideki sayısını göstermesi gibi.

Bireysel veriler için $n_i=1$ ve Y_{ij} ise i . nci bağımlı değişken j . nci kategoride yer alıyorsa 1, aksi takdirde 0 değerini alan dummy değişkeni olur.

Çok terimli dağılıma uyan bağımlı değişkenler yapısı için GTD işlemi ikili (binary) bağımlı değişkenlere benzer şekilde elde edilir.

4. BULGULAR VE TARTIŞMA

4.1. Sedasyon

Bu bölümde uygulamanın temelini oluşturan sedasyon hakkında kısa bir bilgi vereceğiz. Sedasyon; stres faktörlerinin maksimuma ulaştığı yoğun bakım ünitelerindeki tıbbi uygulamaları hasta ve hekim açısından kolaylaştırmak için önemlidir. Yoğun bakım ünitesindeki hastada, invaziv tanı yöntemleri, endotrakeal tüp ve mekanik ventilasyon, trakeal aspirasyon, içinde bulunduğu durumun ciddiyetinin farkında olmanın neden olduğu sıkıntının giderilmesi için sedasyon gerekli olabilir. Sedasyon hastayı rahatlatmak ve sakinleştirmek için kullanılır. Hastanın çevreye ilgisinin ve dış uyaranlara karşı cevaplılığının azaltılması sağlanır. Genel ve lokal anesteziyi desteklemek için kullanılır.

Sedasyonda hedef; koopere olunabilen, kabul edilebilir sükunette, ağrıdan uzak, yoğun bakım uygulamalarına uyum gösteren hastadır. Bu hedefe ulaşmak için; ilaçların analjezik, anksiyolitik, amnezik, antidepressan, hipnotik ve bazen de kas gevşetici etkilerinden yararlanır. İdeal bir ilaçta istenilen özellikler; etkisinin hızlı başlaması, infüzyona uygun olması, özellikle kardiyovasküler ve respiratuar sisteme yönelik yan etkilerinin olmaması, birikici etkisinin olmaması, uygulamasının kolay olması, antagonistinin bulunması, ekonomik olmasıdır. Yoğun bakım ünitesinde uygulanan sedasyonun; kullanılan ilaçların yan etkilerinin olması, hasta takibinin zorlaşması, mekanik ventilasyon süresinin uzaması gibi olumsuz yönleri de bulunmaktadır. Bu nedenle sedasyon uygulanırken yararı ve zararı dikkatle gözden geçirilmelidir. 2 tip sedasyon vardır.

- *Bilinçli Sedasyon*

Hastanın çevreye karşı ilgisinin ve dış uyaranlara karşı cevaplılığının kontrollü olarak azaltılmasıdır. Vital bulgular değişmez.

Bilinçli (hafif) sedasyonda:

- Koruyucu refleksler kaybolmaz.
- Hasta havayolunu açık olarak koruyabilir, uygun oksijenizasyona devam eder, bağımsız ve sürekli ventilasyon mevcuttur.
- Dokunma şeklinde (hafif ağırlı) uyaranlara ve sözlü uyaranlara cevap verir.

- *Bilinçsiz (derin) sedasyon*

Bilinçli sedasyonun üç komponentinden birinin kaybolmasına neden olabilecek derinlikteki sedasyondur. Uygun havayolu, solunum yaklaşımı ve kan basıncı kontrolüne ileri derecede dikkat gerektirir.

Sedasyon dereceleri Ramsey Sedasyon Skoru olarak Tablo 4.1 deki gibi kullanılmaktadır.

Tablo 4.1. Ramsey Sedasyon Skoru

Puan	Sedasyon Derecesi
1	Anksiyetesi bulunan ve ajite hasta
2	Koopere, oryante ve sakin hasta
3	Basit emirleri yerine getiren hasta
4	Glabellaya uygulanan hafif uyarıya ve yüksek sesli uyarıya canlı cevap
5	Yukarıdaki uyarılara daha yavaş cevap

Uygulamada kullanılan veriler Ondokuz Mayıs Üniversitesi Tıp Fakültesinden elde edilmiştir.

4.2. Nörolojik Hastalara İlişkin Sedasyon Verisi İçin Genelleştirilmiş Tahmin Denklemlerinin Kullanımı

Bu bölümde bir grup nörolojik hasarı olan ve nörolojik hasarı olmayan spinal anestezi ile opere 127 hastada diazem, midazolan, lüminal, kardioloji kokteyli infüzyonlarının sedatif etkileri karşılaştırıldı. Her ilaç için uygun yükleme dozunu takiben bu ilaçlar için infüzyonları yapıldı. Postoperatif 90. dakikaya kadar sistolik kan basıncı, nabız ve solunum sayısı her bir hasta için tekrarlı olarak ölçüldü. Daha sonra hastalar tarafından sedasyon kalitesi değerlendirildi.

4.2.1. Değişkenlerin Tanımlanması

Bu çalışmamızda bağımlı değişken sedasyon olup 0., 15., 30., 60. ve 90. dakikalarda ölçülmüş ve bu tekrarlı ölçümlere karşılık gelen dereceleri Tablo 4.1 de verilen kategorik bir değişkendir.

Bu çalışmamızda bağımsız değişkenlerden birisi, dört çeşitten oluşan ilaç türüdür.

Tablo 4.2. İlaç Türleri

Kod	Değişken Adı
A	Midazolan
B	Kardioloji Kokteyli
C	Diazem
D	Lüminal

Açıklayıcı değişkenlerimizin ise bir kısmı durağan bir kısmı da zaman göre bağımlı değişkenlerdir. Bunlar; yaş, cinsiyet, hastalık ve kilodur. Burada cinsiyet ve hastalık kodlanmış verilerdir.

Tablo 4.3. Cinsiyet değişkenini gösterir.

Kod	Değişken Adı
1	Erkek
2	Kız

Tablo 4.4. Hastalık durumunu gösterir.

Kod	Değişken Adı
1	Nörolojik hasarlı
2	Nörolojik hasarı yok

Zamana göre bağımlı olan değişkenler ise her biri yine 0. , 15. , 30. , 60. ve 90. dakikalarda tekrarlı olarak ölçülmüş olan solunum sayısı, nabız ve sistolik kan basıncıdır.

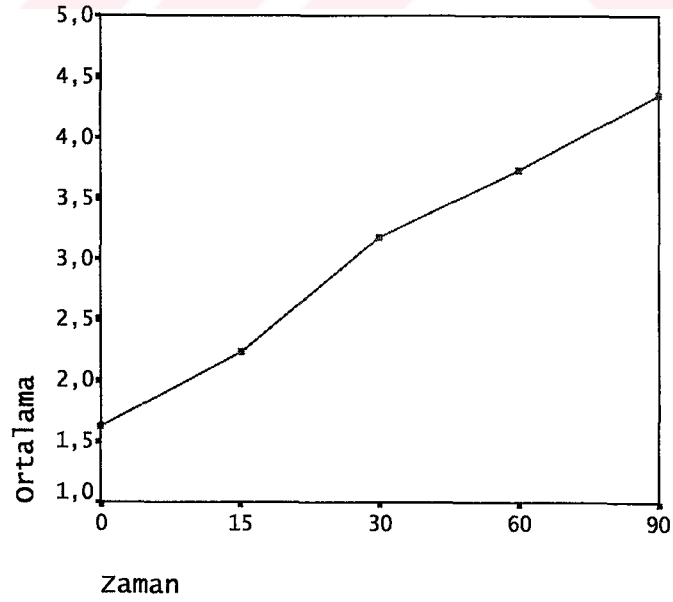
4.2.2 Analiz

Bu kısımda uygun ilişki yapısı belirlenip, genelleştirilmiş tahmin denklemleri ile regresyon katsayıları elde edilmiş, her katsayı için hem model-temelli hem de deneysel standart hatalar belirlenmiş ve sonuçlar yorumlanmıştır. Ayrıca ortalamaların karşılaştırılması grafiksel olarak gösterilmiştir. Analizde SAS paket programı, grafikler için ise SPSS programı kullanılmıştır.

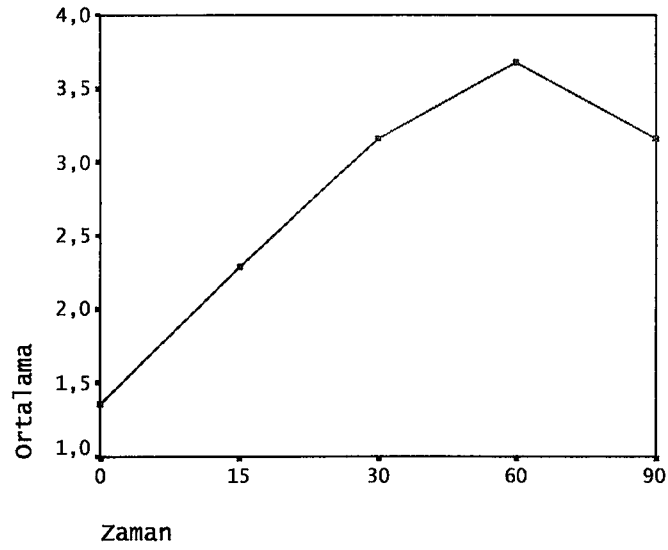
4.2.3. Sonular



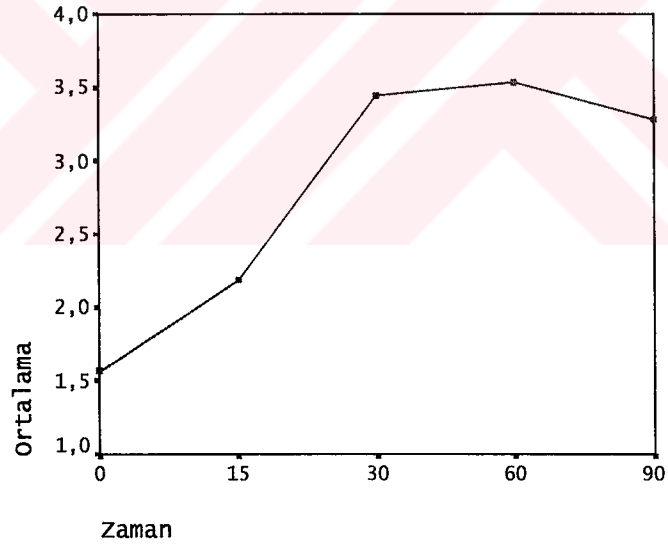
Şekil 4.1. Zaman içerisinde midazolam ilaç grubunun sedasyona olan etkisinin deęişimini gösteren grafik



Şekil 4.2. Zaman içerisinde kardioloji kokteyli ilaç grubunun sedasyona olan etkisinin deęişimini gösteren grafik



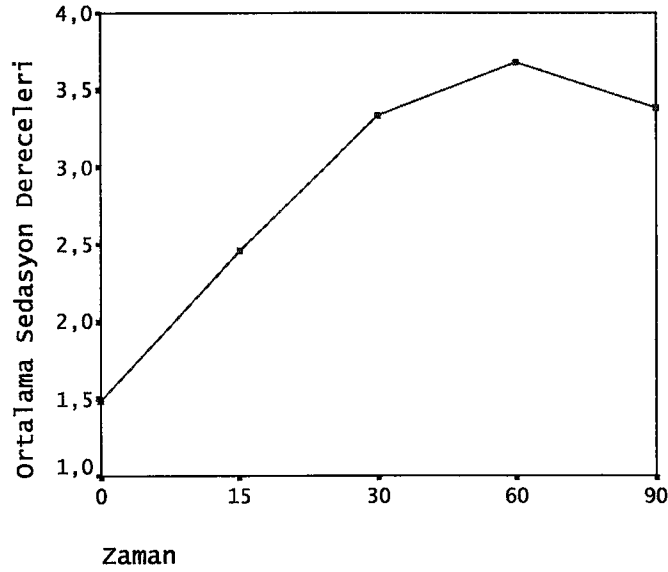
Şekil 4.3. Zaman içerisinde diazepam ilaç grubunun sedasyona olan etkisinin değişimini gösteren grafik



Şekil 4.4. Zaman içerisinde lüminal ilaç grubunun sedasyona olan etkisinin değişimini gösteren grafik

Yukarıdaki grafiklere bakılarak, her bir ilaç grubu için sedasyona olan etkilerinin zamanın belirli noktalarında nasıl değişiklik gösterdiği yorumlanabilir.

Farklı anlardaki sedasyon derecelerinin ortalamaları arasında fark olup olmadığı da Şekil 4.5. de verilen grafiğe bakarak yorumlanabilir.



Şekil 4.5. Sedasyon verisi için ortalamaların testini gösteren grafik

Yukarıdaki şekle bakılarak sedasyonlar arasında fark olduğu söylenebilir. Eğer sedasyon şiddetleri aynı doğrultu üzerinde olsaydı o zaman farklarının sıfır olduğunu ve ortalamalar arasında fark bulunmadığını söyleyebiliriz.

GTD analizi için SAS, GENMOD prosedürü uygulanmıştır. Genel bilgiler Tablo 4.6 da sunulmuştur. Bağımlı değişkenler çok terimli (multinomial) olduğundan dağılım multinomial dir. Link fonksiyonu kümülatif lojittir. İkili (binary) bağımlı değişkenden farklı olarak korelasyon yapısı bağımsız olarak seçilmiştir. Program yapısı Tablo 4.5 de verilmiştir.

Tablo 4.5. SAS GENMOD Program yazılımı

```

Data children;
input case group$ @; do i=1 to 5;
input sedasyon yaş cinsiyet kilo hastalık kb ss nab @@;
output; end; datalines;
1 a 1 5 1 15 1 90 24 104 2 5 1 15 1 90 24 102 5 5 1 15 1 90 22 98 4 5 1 15 1 90 22 96 3 5
1 15 1 90 22 96
2 b 1 4 1 12 1 100 36 105 3 4 1 12 1 90 32 109 4 4 1 12 1 90 30 97 5 4 1 12 1 90 30 95 5 4 1
12 1 90 30 93
3 b 1 6 2 25 1 100 27 97 2 6 2 25 1 100 27 97 4 6 2 25 1 90 28 94 4 6 2 25 1 90 28 94 5 6 2
25 1 90 27 95
:
proc genmod data=children; class case group ;
model sedasyon=group yas cinsiyet kilo hastalik kb ss nab/dist=mult;
repeated subject=group/type=ind covb corrw; run;

```

Tablo 4.6. GENMOD Prosedürü Model Bilgisi

Veri Kümesi	WORK.CHILDREN
Dağılım	Multinomial
Link Fonksiyonu	Kümülatif Logit
Bağımlı Değişken	Sedasyon
Kullanılan Gözlem	635

Tablo 4.7. Sınıf Seviye Bilgileri

Sınıf	Seviyeleri	Değerler
Durum	127	1 2 3 4 ...127
Grup	4	a b c d
Sedasyon	5	1 2 3 4 5

Tablo 4.8. Bağımlı Değişken Profili

Sıralanmış Seviye	Sıralanmış Değer	Miktar
1	1	82
2	2	172
3	3	206
4	4	96
5	5	79

Tablo 4.9. Parametre Bilgileri

Parametre	Etki	Grup
Prm1	grup	a
Prm2	grup	b
Prm3	grup	c
Prm4	grup	d
Prm5	yaş	
Prm6	cinsiyet	
Prm7	kilo	
Prm8	hastalık	
Prm9	kb	
Prm10	ss	
Prm11	nab	

Tablo 4.10. Uyum İyiliğini Tahmin İçin Kriter

Kriter	SD	Değer	Değer/SD
Log Olabilirlik			-942.2210
Algoritma Yakınsak			

Tablo 4.11. Başlangıç Parametre Tahminlerinin Analizi

Wald 95% Güven								
Parametre	SD	Tahmin	Standard		Limitler		Pr > Ki- Kare	
			Hata		Alt	Üst		Ki-Kare
Kesişim 1	1	-6.6013	0.8958		-8.3569	-4.8456	54.31	<.0001
Kesişim 2	1	-4.9937	0.8750		-6.7086	-3.2787	32.57	<.0001
Kesişim 3	1	-3.5535	0.8662		-5.2512	-1.8558	16.83	<.0001
Kesişim 4	1	-2.5433	0.8676		-4.2437	-0.8429	8.59	0.0034
Grup a	1	0.1513	0.2753		-0.3882	0.6908	0.30	0.5826
Grup b	1	0.0714	0.2514		-0.4214	0.5641	0.08	0.7765
Grup c	1	0.3652	0.2660		-0.1562	0.8865	1.88	0.1698
Yaş	1	0.1560	0.0734		0.0121	0.3000	4.52	0.0336
Cinsiyet	1	-0.4457	0.1509		-0.7416	-0.1499	8.72	0.0031
Kilo	1	-0.0242	0.0311		-0.0851	0.0367	0.61	0.4362
Hastalık	1	-0.0407	0.1565		-0.3475	0.2660	0.07	0.7946

Tablo 4.12. Başlangıç Parametre Tahminlerinin Analizinin Devam

Wald %95 Güven								
Parametre	SD	Tahmin	Standart		Sınırlar		Ki-Kare değeri	Pr > Ki-Kare
			Hata		Alt	Üst		
KB	1	0.0414	0.0080		0.0256	0.0572	26.51	<.0001
SS	1	0.0420	0.0157		0.0112	0.0728	7.15	0.0075
NAB	1	0.0020	0.0024		-0.0027	0.0066	0.69	0.4049
Ölçü	0	1.0000	0.0000		1.0000	1.0000		

NOT: Ölçüm parametresi sabit tutuldu.

Wald testi, Polit (1996) ve Agresti (1990) tarafından açıklanmıştır. Bu test, bir grup açıklayıcı değişkenle ilgili parametrelerin sıfır olup olmadığını test etme yollarından birisidir. Yani modeldeki her bir β katsayısı için istatistiksel anlamlığı test etmede kullanılır. Lineer regresyondaki bireysel katsayılar için t-testine benzerlik gösterir.

Hipotezimiz; $H_0 : \beta_i = 0$
 $H_1 : \beta_i \neq 0$

Wald Testi; $Z = \frac{\hat{\beta}}{SE_{\beta}}$ olan Z istatistiğini hesaplar.

Eğer özel bir açıklayıcı değişken veya açıklayıcı değişkenler grubu için Wald testi anlamlıysa o zaman bu değişkenlerle ilgili parametrelerin sıfırdan farklı olduğu ve böylece bu değişkenlerin modelde yer alabileceği söylenir. Eğer test anlamlı değilse o takdirde açıklayıcı değişkenler modelden çıkarılabilir. Tek bir açıklayıcı değişken için Altman (1991) parametrelerin anlamlı olup olmadığını testte t-testini kullandı. Tek bir parametre için Wald testi sadece t istatistiğinin kareköküdür ve eş değer sonuçları verir.

Tablo 4.13. GEE Model Bilgisi

Korelasyon Yapısı	Bağımsız
Birey Etkisi	grup (4 seviye)
Küme Sayısı	4
Korelasyon Matris Boyutu	170
Maksimum Küme Genişliği	170
Minimum Küme Genişliği	150

Tablo 4.14. GEE Parametre Tahminlerinin Analizi

Parametre	Tahmin	Deneysel Standart Hata Tahminleri		Z	Pr > Z
		Standart Hata	Sınırlar		
			Alt Üst		
Kesişim 1	-6.6013	1.6812	-9.8964 -3.3061	-3.93	<.0001
Kesişim 2	-4.9937	1.6688	-8.2645 -1.7229	-2.99	0.0028
Kesişim 3	-3.5535	1.7037	-6.8927 -0.2143	-2.09	0.0370
Kesişim 4	-2.5433	1.6409	-5.7593 0.6728	-1.55	0.1212
Grup a	0.1513	0.1843	-0.2100 0.5125	0.82	0.4118
Grup b	0.0714	0.1709	-0.2635 0.4063	0.42	0.6762
Grup c	0.3652	0.1495	0.0721 0.6582	2.44	0.0146
Yaş	0.1560	0.0888	-0.0180 0.3301	1.76	0.0789
Cinsiyet	-0.4457	0.2867	-1.0076 0.1162	-1.55	0.1200
Kilo	-0.0242	0.0313	-0.0855 0.0371	-0.77	0.4394
Hastalık	-0.0407	0.1941	-0.4211 0.3396	-0.21	0.8338
KB	0.0414	0.0093	0.0232 0.0596	4.45	<.0001
SS	0.0420	0.0246	-0.0062 0.0901	1.71	0.0874
NAB	0.0020	0.0014	-0.0007 0.0047	1.42	0.1543

5. SONUÇ VE ÖNERİLER

Uzunlamasına verilerle ilgili bu çalışmada konu ile ilgili bilgiler ve diğer istatistiksel veri yapılarından farkı verildi. Uzunlamasına verilerin, sağlık ve benzeri alanlarda yaygın olarak kullanımı verilmiş, gelişimindeki temel faktörler vurgulandı. Kesikli bağımlı değişken yapısı için esas olan ve son 30 yılda bilgisayar teknolojisindeki inanılmaz gelişimle beraber kullanılmaya başlanan Genelleştirilmiş Lineer Modeller verildi.

Çalışmanın analiz aşamasında ise GTD yöntemi anlatılıp, bu metotla ilgili formüllerden bahsedildi. Kitle ortalama değerlerinin tahmininde anlamlı sonuçlar veren GTD sürecinin aynı zamanda grup farklılıklarını ortaya koyduğu da vurgulandı.

Tezdeki asıl amaç; tekrarlı olarak ölçülmüş, kesikli bağımlı değişkenler ile açıklayıcı değişkenler arasındaki istatistiksel modellemeyi oluşturmaktır. Yanıt değişkeni normal dağılıdıysa uygun analiz basit doğrusal regresyonun genelleştirilmesidir. Uzunlamasına veri analizinde ise, bağımlı değişkenin sürekli olduğu durumlar için istatistiksel yöntemler daha geniştir. Yanıt değişkeninin normal dağılım göstermediği durumlarda GTD yönteminin uygun olduğu belirtilmiştir. Bu yöntem tekrarlanan bağımlı değişkenin ortak dağılımın belirtilmesine ihtiyaç duymaz.

Tezin uygulama bölümünde tıbbın değişik alanlarında özellikle anesteziye sıkça kullanılan, hastanın çevreye ilgisinin azaltılması veya ağrıyı azaltma gibi rahatlatıcı ve sakinleştirici etkisi olan sedasyon verisi incelenmiştir. Bu veriyi analizde GTD yöntemi kullanılmıştır.

Analiz sonucuna göre sedasyonu açıklayan değişkenlerden kan basıncı ve ilaç gruplarından diazem istatistiksel olarak anlamlı bulunmuştur. Bu açıklayıcı değişkenlerin tahmin değerleri pozitif çıktığından dolayı sedasyon şiddeti ile kan basıncı ve diazem ilaç grubu arasında doğru yönde ilişki vardır denilebilir. Ayrıca grup etkilerin toplamı sıfır olacak şekilde sınırlama getirildiğinden dolayı grup d nin etkisi tabloda yer almamaktadır.

Çok terimli değişkenlerin analizi yapılırken ikili (binary) değişkenlerin analizinden farklı olarak yeni bir link fonksiyonu ve korelasyon yapısı tanımlanmıştır. İkili değişkenler için korelasyon yapısı değiştirilebilir ve link fonksiyonu lojit iken bu yapılar çok terimli için sırasıyla bağımsız ve kümülatif lojit olarak verilmiştir.

Bu çalışmada, bahsedilen metotlar farklı alanlarda başarıyla uygulanabilir. Ayrıca parametre tahminleri için son yıllarda istatistiksel bir çıkarım tekniği olarak kabul edilen ve

basit olarak önsel bilgilerin analize katılması olarak yorumlanabilen Bayesci yaklaşımda kullanılabilir. Özellikle bu yaklaşım, gelecekle ilgili çalışmalarımızda bize yön verecektir.



6. KAYNAKLAR

- 1) **Alexander von Eye.** Statistical Methods in Longitudinal Research. Academic Pres.
- 2) **Crowder, M.J., Hand, D.J., 1990.** Analysis of Repeated Measures, London: Chapman and Hall.
- 3) **Diggle, P.J., 1990.** Time Series: A Biostatistical Introduction, Oxford: Clarendon Press.
- 4) **Diggle, P.J., Liang, K.-Y., Zeger, S.L., 1994** Analysis of Longitudinal Data, Oxford: University Press.
- 5) **Dr. Pahwa, P., 2002,** “Statistical Models for The Analysis of Longitudinal Data”, www.cacr.ca/news/2002/0204pahwa.htm (3.03.2003).
- 6) **Edwards, L.Y., 2000.** Modern Statistical Techniques for The Analysis of Longitudinal Data in Biomedical Research. Pediatric Pulmonology, 30, 330-344.
- 7) **Fitzmaurice, G.M., Laird, N.M., Rotnizky, A.G., 1993.** Regression Models for Discrete Longitudinal Responses. Statistical Science, 8, 284-309.
- 8) **“Generalized Estimating Equations”. The GENMOD Procedure.** www.id.unizh.ch/software/unix/statmath/sas/sasdoc/stat/chap29/sect38.htm (12.05.2003)
- 9) **Goldstein, H., 1979.** The Design and Analysis of Longitudinal Studies. Academic Press.
- 10) **Griswold, M., 2000.** “Likelihood&Quasi-Likelihood Marginal Models for Analyzing Longitudinal Binary Data”. http://biosun01.biostat.jhsph.edu/~mgriswol/students/oralpapers/2000/griswold_dept_oral.pdf (3.03.2003)
- 11) **Hand, D., Crowder, M., 1995.** Practical Longitudinal Data Analysis, United Kingdom: Chapman and Hall.
- 12) **Heitjan, D.F., Sharma, D: 1997.** Modelling Repeated-Series Longitudinal Data, Statistic in Medicine, 16, 347-355.
- 13) **Horton, N.J., Lipsitz, S.R., 1999.** Review of Software to Fit Generalized Estimating Equation Regression Models. The American Statistician, 53. <http://www.biostat.harvard.edu/horton/geereview.pdf> (3.03.2003)
- 14) **Johnston, G., Stokes, M.** “Repeated Measures Analysis with DiscreteData Using The SAS SYSTEM. SAS Institue Inc., Cary, NC.

- <http://www2.sas.com/proceedings/sugi22/STATS/PAPER278.PDF> (3-03-2003)
- 15) **Laird, N.M., Fitzmaurice, G.M., 1993.** A Likelihood-based Method for Analysing Longitudinal Binary Responses. *Biometrika*, 80, 141-151.
 - 16) **Liang, K.-Y., Zeger, S.L., 1986.** Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, 73, 13-22
 - 17) **Liang, K.-Y., Zeger, S.L., 1986.** Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics*, 42, 121-130.
 - 18) **Liang, K.-Y., Zeger, S.L., Albert, P.S., 1988.** Models for Longitudinal Data: A Generalized Estimating Equation Approach. *Biometrics*, 44, 1049-1060.
 - 19) **Littell, R.C., Pendergast, J., Natarajan R., 2000.** Modelling Covariance Structure in the Analysis of Repeated Measures Data. *Statistics in Medicine*, 19, 1973-1819.
 - 20) **“Longitudinal Data Analysis, Chapter 4”.**
<http://mrs.umn.edu/anderson/math4601/notes/ch4longdata.pdf>. (3.03.2003)
 - 21) **McCullagh, p. And Nelder, J.A.,1989.** Generalized Linear Models, Second Edition, Chapman and Hall.
 - 22) **Myers, R.H., Montgomery, D.C., Vining, G.G., 1994.** Generalized Linear Models with Applications in Engineering and the Sciences, NewYork: John Wiley and Sons.
 - 23) **Prentice, R.L., Zhao, L.P., 1990.** Correlated Binary Regression Using a Quadratic Exponential Model. *Biometrika*, 77, 642-648.
 - 24) **Ravindra, K., Dayanand, N.N., 1999.** Applied Multivariate Statistics with SAS Software, Second Edition. SAS Institue.
 - 25) **Ruspini, E., 2000.** Longitudinal Research in the Social Sciences, Social Research Update. Issue 28.
<http://www.soc.surrey.ac.uk/sru/SRU28.html> (3.03.2003)
 - 26) **Schmid, C.H., 2001.** Marginal and Dynamic Regression Models for Longitudinal Data. *Statistics in Medicine*, 20, 3295-3311.
 - 27) Sedasyon Nedir? www.mustafayildiz.8m.com (5.03.2003)

ÖZGEÇMİŞ

1977 yılında Samsun'da doğdum. İlk ve orta öğrenimimi Samsun'da tamamladım. 1996 yılında Ondokuz Mayıs Üniversitesi İstatistik bölümüne başladım. 2000 yılında lisans diploması aldım ve aynı yılın Eylül ayında istatistik bölümünde yüksek lisansa başladım. 2000 yılının sonlarına doğru Araştırma Görevlisi olarak atandım.

