

**YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**POISSON REGRESYON MODELİ ve TÜRKİYE’DEKİ
BOŞANMA İSTATİSTİKLERİNE UYGULANMASI**

İstatistikçi Hande YEŞİLYURT

**FBE İstatistik Anabilim Dalında
Hazırlanan**

YÜKSEK LİSANS TEZİ

Tez Danışmanı : Yrd. Doç. Dr. Atıf EVREN

İSTANBUL, 2005

**YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**POISSON REGRESYON MODELİ ve TÜRKİYE’DEKİ
BOŞANMA İSTATİSTİKLERİNE UYGULANMASI**

İstatistikçi Hande YEŞİLYURT

**FBE İstatistik Anabilim Dalında
Hazırlanan**

YÜKSEK LİSANS TEZİ

Tez Danışmanı : Yrd. Doç. Dr. Atıf EVREN

İSTANBUL, 2005

İÇİNDEKİLER

	Sayfa
SİMGE LİSTESİ	vi
KISALTMA LİSTESİ	viii
ŞEKİL LİSTESİ	ix
ÇİZELGE LİSTESİ	x
ÖNSÖZ.....	xi
ÖZET	xii
ABSTRACT	xiii
1. GİRİŞ.....	1
2. TEK AÇIKLAYICI DEĞİŞKEN ile DOĞRUSAL REGRESYON	3
2.1 Değişkenler Arasındaki İlişkiler	3
2.1.1 İki Değişken Arasındaki Fonksiyonel İlişki	3
2.1.2 İki Değişken Arasındaki İstatistiksel İlişki	4
2.2 Regresyon Modelleri	4
2.2.1 Temel Kavramlar	4
2.2.1.1 Birden Fazla Açıklayıcı Değişken ile Regresyon Modelleri	4
2.2.2 Regresyon Modellerinin Oluşturulması.....	5
2.2.2.1 Açıklayıcı Değişkenlerin Seçimi	5
2.2.2.2 Regresyon İlişkisinin Fonksiyonel Formu.....	5
2.2.3 Regresyon Analizinin Kullanıldığı Durumlar	5
2.2.4 Regresyon ve Nedensellik	5
2.3 Dağılımı Belirlenmemiş Hata Terimleri ile Yalın Doğrusal Regresyon Modeli.....	6
2.3.1 Modelin Biçimlendirilmesi	6
2.3.2 Modelin Önemli Özellikleri.....	6
2.3.3 Regresyon Parametrelerinin Anlamı.....	7
2.4 En Küçük Kareler Yöntemi ile Parametrelerin Tahmini	8
2.4.1 En Küçük Kareler Tahmincileri.....	8
2.4.2 En Küçük Kareler Tahmincilerinin Özellikleri	9
2.5 Kalıntılar	10
2.6 Tahmin Edilen Regresyon Doğrusunun Özellikleri	10
2.7 Hata Terimlerinin Varyansının Tahmini	11
2.7.1 σ^2 'nin Nokta Tahmincisi	11
2.7.1.1 Açıklayıcı Değişkenin Bulunmadığı Durum	11
2.7.1.2 Regresyon Modeli.....	12
2.8 Normal Dağılan Hata Terimli Regresyon Modeli	12
2.8.1 Model.....	13
2.8.2 Maksimum Olabilirlik Yöntemi ile Parametrelerin Tahmini	13

2.8.2.1	Regresyon Modeli.....	14
3.	REGRESYON ANALİZİNDEN ÇIKARILAN SONUÇLAR	17
3.1	β_1 ile İlgili Aralık Tahminleri ve Hipotez Testleri	17
3.1.1	b_1 'in Örneklem Dağılımı	17
3.1.1.1	Normallik.....	17
3.1.1.2	Varyansın Tahmini	18
3.1.2	$(b_1 - \beta_1)/s\{b_1\}$ 'in Örneklem Dağılımı	18
3.1.3	β_1 için Güven Aralığı.....	18
3.1.4	β_1 ile İlgili Testler.....	19
3.1.4.1	Çift Taraflı Test	19
3.1.4.2	Tek Taraflı Test	19
3.2	β_0 ile İlgili Aralık Tahminleri ve Hipotez Testleri.....	20
3.2.1	b_0 'ın Örneklem Dağılımı.....	20
3.2.2	$(b_0 - \beta_0)/s\{b_0\}$ 'in Örneklem Dağılımı	20
3.2.3	β_0 için Güven Aralığı	21
3.3	Varyans Analizi Tablosu	21
3.4	X ve Y Arasındaki Doğrusal İlişkinin Tanımlayıcı Ölçüleri.....	21
3.4.1	Belirginlik Katsayısı	21
3.4.2	Korelasyon Katsayısı	22
4.	YALIN DOĞRUSAL REGRESYON ANALİZİNDE MATRİS YAKLAŞIMI ..	23
4.1	Matris Gösterimiyle Yalın Doğrusal Regresyon Modeli.....	23
4.2	Regresyon Parametrelerinin En Küçük Kareler Tahmini	25
4.2.1	Normal Denklemler	25
4.2.2	Tahmin Edilen Regresyon Katsayıları.....	25
4.3	Tahmin Değerleri.....	26
4.3.1	Projeksiyon Matrisi.....	26
4.4	Kalıntılar	27
4.4.1	Kalıntıların Varyans-Kovaryans Matrisi	27
4.4.2	Kareler Toplamının Kuadratik Formda Gösterilmesi.....	28
4.5	Regresyon Analiziyle İlgili Sonuçlar.....	28
4.5.1	Regresyon Katsayıları.....	28
4.5.2	Bağımlı Değişkenin Beklenen Değeri (Mean Response)	29
5.	ÇOKLU REGRESYON ANALİZİ	31
5.1	Matris Gösterimiyle Genel Doğrusal Regresyon Modeli	31
5.2	Regresyon Katsayılarının Tahmini	32
5.3	Tahmin Edilen Değerler ve Kalıntılar	33
5.4	Varyans Analizi Sonuçları.....	35
5.4.1	Kareler Toplamı ve Kareler Ortalaması	35
5.4.2	Regresyon İlişkisi için F Testi.....	36
5.4.3	Çoklu Belirginlik Katsayısı	36
5.4.4	Çoklu Korelasyon Katsayısı	37
5.5	Regresyon Parametreleriyle İlgili Elde Edilen Sonuçlar	37
5.5.1	β_k 'nın Tahmin Aralığı.....	38
5.5.2	β_k İçin Testler.....	38

5.6	Bağımlı Değişkenin Beklenen Değerinin Tahmini	39
5.6.1	$E\{Y_h\}$ 'nin Tahmin Aralığı.....	39
6.	DOĞRUSAL OLMAYAN REGRESYON	41
6.1.1	Doğrusal Olmayan Regresyon Modelleri	41
6.1.1.1	Üstel Regresyon Modelleri	41
6.1.1.2	Lojistik Regresyon Modelleri	42
6.1.1.3	Doğrusal Olmayan Regresyon Modellerinin Genel Biçimi.....	43
6.1.2	Regresyon Parametrelerinin Tahmini	45
6.2	Doğrusal Olmayan Regresyonda En Küçük Kareler Tahmini	45
6.2.1	Normal Denklemlerin Çözümü.....	47
6.2.2	Doğrudan Nümerik Arama (Gauss-Newton Metodu)	48
6.3	Modelin Oluşturulması ve Tanı (Diagnostics)	53
6.4	Doğrusal Olmayan Regresyon Parametreleri ile İlgili Çıkarılan Sonuçlar	53
6.4.1	Hata Terimleri Varyansının Tahmini.....	53
6.4.2	Büyük Örneklem Teorisi	54
6.4.3	Büyük Örneklem Teorisi Ne Zaman Uygulanır?.....	55
6.4.4	Bir γ_k 'nın Aralık Tahmini	56
6.4.5	Bir γ_k 'yla İlgili Test.....	56
7.	POISSON REGRESYON ANALİZİ	57
7.1	Poisson Dağılımı.....	57
7.2	Poisson Regresyon.....	57
7.2.1	Bağ (Link) Fonksiyonu.....	59
7.2.2	Varyans Fonksiyonu	60
7.2.3	Uyum Ölçüleri	60
7.2.4	Model.....	64
7.2.4.1	Aşırı Yayılım için Tartı Parametresi.....	68
7.2.4.2	Görel Risk ya da Risk Oranı (Relative Risk or Risk Ratio)	69
7.3	Negatif Binom Model	69
7.4	Az Yayılımla İlgili Bir Açıklama	71
7.5	Bol Sıfırlı Modeller (Zero-Inflated Models)	71
7.6	Çokterimli Oluş Sayısı ile Belirlenen Veri için Poisson Regresyonun Kullanımı	73
7.7	Çokterimli Oluş Sayısı ile Belirlenen Veri için Logaritmik Doğrusal Modeller...	74
8.	UYGULAMA	77
8.1	Boşanma Verileri ile İlgili Elde Edilen Sonuçlar	78
8.2	Boşanma Verileri2 ile İlgili Elde Edilen Sonuçlar	80
8.3	Boşanma Verileri3 ile İlgili Elde Edilen Sonuçlar	83
8.4	14.Modelin İncelenmesi	86
8.4.1	14.Modeldeki Aşırı Yayılımın İncelenmesi	87
9.	SONUÇLAR.....	89
KAYNAKLAR.....		91
EKLER		92
Ek 1 Binom Dağılımının Poisson Dağılımına Yaklaşımı.....		93

Ek 2 Poisson Dağılımına Normal Yaklaşım.....	94
Ek 3 Poisson Süreci	95
Ek 4 Yalın Doğrusal Regresyon Analizine Matris Yaklaşımı.....	98
Ek 5 Boşanma Verileri'nde İkinci Model ile İlgili Grafikler	101
Ek 6 Boşanma Verileri'nde Üçüncü Model ile İlgili Grafikler	103
Ek 7 Boşanma Verileri'nde Dördüncü Model ile İlgili Grafik.....	105
Ek 8 Boşanma Verileri'nde Beşinci Model ile İlgili Grafikler	106
Ek 9 Boşanma Verileri2'de Yedinci Model ile İlgili Grafikler.....	108
Ek 10 Boşanma Verileri2'de Sekizinci Model ile İlgili Grafikler.....	110
Ek 11 Boşanma Verileri2'de Dokuzuncu Model ile İlgili Grafik	112
Ek 12 Boşanma Verileri2'de Onuncu Model ile İlgili Grafik	113
Ek 13 Boşanma Verileri3'de On İkinci Model ile İlgili Grafikler	114
Ek 14 Boşanma Verileri3'de On Üçüncü Model ile İlgili Grafikler	116
Ek 15 Boşanma Verileri3'de On Dördüncü Model ile İlgili Grafik.....	118
Ek 16 Boşanma Verileri3'de On Beşinci Model ile İlgili Grafikler.....	119
Ek 17 İkinci Model için Elde Edilen Bilgisayar Çıktısı	120
Ek 18 Üçüncü Model için Elde Edilen Bilgisayar Çıktısı	121
Ek 19 Dördüncü Model için Elde Edilen Bilgisayar Çıktısı	122
Ek 20 Beşinci Model için Elde Edilen Bilgisayar Çıktısı	123
Ek 21 Yedinci Model için Elde Edilen Bilgisayar Çıktısı.....	124
Ek 22 Sekizinci Model için Elde Edilen Bilgisayar Çıktısı	125
Ek 23 Dokuzuncu Model için Elde Edilen Bilgisayar Çıktısı	126
Ek 24 Onuncu Model için Elde Edilen Bilgisayar Çıktısı.....	127
Ek 25 On İkinci Model için Elde Edilen Bilgisayar Çıktısı	128
Ek 26 On Üçüncü Model için Elde Edilen Bilgisayar Çıktısı	129
Ek 27 On Beşinci Model için Elde Edilen Bilgisayar Çıktısı.....	130
Ek 28 Boşanma Verileri	131
Ek 29 Boşanma Verileri2	136
Ek 30 Boşanma Verileri3	140
ÖZGEÇMİŞ.....	144

SİMGE LİSTESİ

b_0	β_0 'ın tahmincisi
b_1	β_1 'ın tahmincisi
$D_{ik}^{(0)}$	Başlangıç değerleriyle yer değiştiren parametreler ile i .durum için hesaplanan bağımlı değişkenin beklenen değerinin k .parametreye göre kısmi türevi
e	Kalıntı
G	Sapma (Deviance)
g_k	Doğrusal olmayan regresyonda γ_k parametresinin tahmini
$g^{(0)}$	Doğrusal olmayan regresyonda parametre başlangıç değerlerinin vektörü
H	Projeksiyon matrisi
k_i	Doğrusal regresyonda kullanılan sabit
$L(\mu)$	μ parametre değerinin olabilirlik değeri
n	Gözlem sayısı
o_i	Genelleştirilmiş doğrusal modeldeki sabit (Offset)
Q	En küçük kareler yönteminde sapmaların karelerinin toplamı
r	Korelasyon katsayısı
r_i	Pearson kalıntıları
R^2	Belirginlik katsayısı
R_a^2	Ayarlanmış çoklu belirginlik katsayısı
$s\{b_0\}$	b_0 'ın standart hatası
$s\{b_1\}$	b_1 'ın standart hatası
s^2	Örnek varyansı
$s^2\{b_0\}$	b_0 'ın varyansı
$s^2\{b_1\}$	b_1 'ın varyansı
$s^2\{\hat{Y}_h\}$	$\hat{Y}_h$ 'ın tahmin edilen varyansı
W	Tartı matrisi
X	Bağımsız değişken
$\bar{X}$	X gözlemlerinin ortalaması
X^2	Pearson uyum test istatistiği
Y	Bağımlı değişken
$\bar{Y}$	Y gözlemlerinin ortalaması
$\hat{Y}$	Y 'nin tahmin değeri
z	Poisson regresyon modelindeki çalışan (working variate) rastlantı değişkeni
α	Anlamlılık düzeyi
β_0	Doğrusal regresyon modelindeki sabit parametre
$\hat{\beta}_0$	β_0 'ın tahmincisi
β_1	Doğrusal regresyon modelindeki eğim parametresi
$\hat{\beta}_1$	β_1 'ın tahmincisi
ε	Hata terimi
γ_0	Doğrusal olmayan regresyon modelinin parametresi
γ_1	Doğrusal olmayan regresyon modelinin parametresi
γ_2	Doğrusal olmayan regresyon modelinin parametresi
κ	Negatif binom dağılımın parametresi

λ_{ij}	(i, j) 'inci grup içindeki gerçek risk
$\lambda(X_i, \beta)$	Her i alt grubu için hata oranını gösteren X_i ve β 'nin özel bir fonksiyonu
μ	Ortalama
$\sigma\{b_0\}$	b_0 'in standart sapması
$\sigma\{b_1\}$	b_1 'in standart sapması
σ^2	Varyans
$\hat{\sigma}^2$	σ^2 'nin tahmincisi
$\sigma^2\{b_0\}$	b_0 'in varyansı
$\sigma^2\{b_1\}$	b_1 'in varyansı
$\sigma^2\{\hat{Y}_h\}$	$\hat{Y}_h$ 'in varyansı

KISALTMA LİSTESİ

ANOVA	Analysis of variance
EM	Expectation maximization algorithm
IRWLS	Iteratively reweighted least squares
MS	Mean square
MSE	Error mean square
MSR	Regression mean square
RR	Risk Ratio
SSE	Sum of squares due to error
SSR	Sum of squares due to regression
SSTO	Total sum of squares
ZAP	Zero-Altered Poisson
ZIP	Zero-Inflated Poisson

ŞEKİL LİSTESİ

Şekil 2.1 Fonksiyonel İlişki	3
Şekil 2.2 İstatistiksel İlişki	4
Şekil 6.1 Üstel ve lojistik bağımlı değişken fonksiyonlarının çizimleri.....	43
Şekil 8.1 Boşanma Verileri'nde dördüncü model için Pearson kalıntılarının dağılımı.....	79
Şekil 8.2 Boşanma Verileri'nde dördüncü model için tahmin edilen değerlerle boşanma sayılarının grafiği	79
Şekil 8.3 Boşanma Verileri2'de dokuzuncu model için Pearson kalıntılarının dağılımı	81
Şekil 8.4 Boşanma Verileri2'de dokuzuncu model için tahmin edilen değerlerle boşanma sayılarının grafiği	81
Şekil 8.5 Boşanma Verileri2'de onuncu model için Pearson kalıntılarının dağılımı	82
Şekil 8.6 Boşanma Verileri2'de onuncu model için tahmin edilen değerlerle boşanma sayılarının grafiği	82
Şekil 8.7 Boşanma Verileri3'de on dördüncü model için Pearson kalıntılarının dağılımı.....	84
Şekil 8.8 Boşanma Verileri3'de on dördüncü model için tahmin edilen değerlerle boşanma sayılarının grafiği	84
Şekil 8.9 Boşanma Verileri3'de on beşinci model için Pearson kalıntılarının dağılımı.....	85
Şekil 8.10 Boşanma Verileri3'de on beşinci model için tahmin edilen değerlerle boşanma sayılarının grafiği	85
Şekil 8.11 BoşanmaVerileri3'de on dördüncü model için tahmin değerleriyle düzeltilmiş sapma kalıntılarının grafiği	88

ÇİZELGE LİSTESİ

Çizelge 2.1 Maksimum olabilirlik tahmincileri.....	15
Çizelge 3.1 Yalın doğrusal regresyon modeli için ANOVA tablosu	21
Çizelge 5.1 $Y = X \beta + \varepsilon$ Genel Doğrusal Regresyon Modeli için ANOVA Tablosu.....	36
$\begin{matrix} n \times 1 & n \times p & n \times p & n \times 1 \end{matrix}$	
Çizelge 7.1 Etkiler	75
Çizelge 8.1 Boşanma Verileri ile İlgili Poisson ANOVA Tablosu	78
Çizelge 8.2 Boşanma Verileri2 (Boşanma Verileri'nden 15-19 ve 20-24 Yaş Grupları çıkarılarak elde edilmiştir.) ile ilgili Poisson ANOVA Tablosu.....	80
Çizelge 8.3 Boşanma Verileri3 (Boşanma Verileri'nden 15-19, 20-24 ve 60 ve üzeri Yaş Grupları çıkarılarak elde edilmiştir.) ile ilgili Poisson ANOVA Tablosu	83
Çizelge 8.4 On dördüncü model ile ilgili S-PLUS 2000 programından elde edilen çıktı	86
Çizelge 8.5 On dördüncü model için hesaplanan tahminlerin korelasyon matrisi	87

ÖNSÖZ

Bu çalışma boşanma verileriyle ilgili bir Poisson regresyon uygulaması yapmak amacıyla hazırlanmıştır. Genel olarak konu bütünlüğü sağlanması amacıyla hem doğrusal hem de doğrusal olmayan regresyondan bahsedilmiştir. Özellikle doğrusal olmayan regresyonun bir çeşidi olan Poisson regresyon üzerinde durulmuştur. Poisson regresyon, bağımlı değişken oluş sayısı (count) şeklinde ele alındığında kullanılmaktadır. Bu çalışmada farklı bir alan olan boşanma verileriyle ilgili bir uygulama yapılmıştır. Ancak, Poisson regresyon genellikle sağlık alanıyla ilgili uygulamalarda hatta sıklıkla kanser araştırmalarında kullanılmaktadır. DİE'nin Boşanma İstatistikleri 1998 adlı kitapçığındaki veriler kullanılmıştır. Uygulama olarak boşanma verileri modellenmeye çalışılmıştır. Bu çalışmadaki katkılarından dolayı Yrd. Doç. Dr. Atıf Evren'e teşekkür ederim.

ÖZET

Çalışma içerisinde, 1989 ile 1998 yılları arasında Türkiye’deki boşanma sayısı verileriyle ilgili bir Poisson regresyon uygulaması yapmak amaçlanmıştır. İlk olarak regresyon analiziyle ilgili bilgi verilmiştir. Regresyon, doğrusal regresyon ve doğrusal olmayan regresyon şeklinde ikiye ayrılmaktadır. Poisson regresyon bir doğrusal olmayan regresyon şeklindedir. Bağımlı değişken oluş sayısı (count) ile belirtilen bir veri olduğunda, Poisson regresyon analizi kullanılmaktadır. Bugüne kadar yapılan çalışmalarda, Poisson dağılımı için ortalama ve varyansın eşit olması varsayımı altında Poisson regresyon modeli incelenmiştir. Ancak, uygulamada genellikle aşırı yayılım yani varyansın ortalamadan büyük olması haliyle karşılaşılmaktadır. Aşırı yayılım olan durumlarda ise iki yol izlenmektedir. Bunlardan birincisi bir tartı parametresi tahmin ederek, bununla test istatistikleri ve kalıntıların düzeltilmesidir. İkincisi ise negatif binom regresyon uygulamasıdır.

Poisson regresyon modeline ilişkin parametre tahminleri, iteratif olarak yeniden ağırlıklandırılmış en küçük kareler ve maksimum olabilirlik yöntemi kullanılarak yapılmaktadır. Veriyi en iyi açıklayan yapıyı bulabilmek için uyum ölçüleri ve kalıntılar incelenmektedir. Genellikle sapma (deviance) değeri en küçük olan model, veriyi en iyi açıklayan model olarak tercih edilmektedir.

Boşanma verileri üzerinde yapılan uygulamada parametre tahminleri, uyum ölçüleri ve kalıntıların hesaplanmasında S-PLUS 2000 programı kullanılmıştır. Uygulama bölümünde boşanma verileri Poisson regresyon ile modellenmeye çalışılmıştır.

Anahtar kelimeler: Poisson Regresyon, Boşanma, Doğrusal Regresyon, Doğrusal Olmayan Regresyon, Negatif Binom Regresyon

ABSTRACT

The aim of this study is to apply Poisson regression models to the data of divorce statistics in Turkey between the years of 1989 and 1998. Firstly, the basic theory underlying regression analysis has been given. Basically there are two types of regression models: Linear models and nonlinear models. Poisson regression can be accepted as a specific type of nonlinear regression model. Poisson regression analysis is a regression technique available for modeling dependent variables that describe count data. For Poisson distribution, the mean and the variance are equal. However, usually in this kind of applications the problem of overdispersion is faced, namely, the variance is larger than the mean. When overdispersion occurs, there are two strategies to overcome this problem: The first way is to estimate a scale parameter, and with the scale parameter adjusting the test statistics and residuals. The second way is to refer to negative binomial regression models.

The parameter estimators related to Poisson regression model are calculated by iteratively reweighted least squares and the maximum likelihood methods. To find out the best statistical model that fits data well, goodness of fit measures and residuals analysis techniques are employed. In Poisson regression models, typically, the model with the least deviance is the candidate to be chosen as the best model.

S-PLUS 2000 programme is used while calculating residuals, goodness of fit measures and parameter estimations for the application of divorce data. In the part of application, divorce data are tried to be modeled by the Poisson regression.

Keywords: Poisson Regression, Divorce, Linear Regression, Nonlinear regression, Negative Binomial Regression

1. GİRİŞ

Regresyon analizinde genellikle, bağımlı değişkenin alacağı değerler diğer açıklayıcı değişkenlerden yararlanılarak tahmin edilmektedir. Bu analiz yapılırken, kullanılan açıklayıcı değişkenlerle en uygun modelin bulunması amaçlanmaktadır. Bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi en iyi açıklayan en sade yapıya sahip olan model, genel olarak en uygun model olarak değerlendirilmektedir.

Poisson regresyon, lojistik regresyondan sonra en genel olan ikinci genelleştirilmiş doğrusal modeldir. Bağımlı değişken oluş sayısı (count) ile belirtilen bir veri olduğunda, yani belirli bir zaman ya da yerde olan olayların sayısı olduğunda Poisson regresyon kullanılmaktadır. Poisson regresyonda, veri olarak genellikle çeşitli araştırmalardan elde edilen sonuçların belli kategorilere göre gruplandırılması ile oluşan tablolardan faydalanılmaktadır. Bu tablolardaki hücre değerlerinin belli bir özellik gösteren ve oluş sayısı ile belirtilen bir veri olması gerekmektedir. Bu şekildeki tabloların oluşturduğu verilerin çözümlenmesinde logaritmik doğrusal modeller kullanılmaktadır. Poisson regresyon da bu logaritmik doğrusal modellerden biridir. Poisson regresyon modelinde, bağımsız değişkenlerin doğrusal yapısını bağımlı değişkenin beklenen değerine bağlayan bağ (link) fonksiyonu logaritmiktir.

Poisson regresyon, genellikle sağlık çalışmalarında özellikle kanser araştırmalarında yoğunlukla kullanılmaktadır.

Bu zamana kadar yapılan çalışmalarda, Poisson dağılımı için ortalama ve varyansın eşit olması (equidispersion) varsayımı altında Poisson regresyon modeli incelenmiştir. Poisson dağılımında; varyansın ortalamaдан büyük olması haline aşırı yayılım (overdispersion) ve varyansın ortalamaдан küçük olması haline az yayılım (underdispersion) denilmektedir. Böyle bir durum söz konusu olduğunda tartı parametresinin incelenmesi gerekmektedir. Tartı parametresi incelendikten sonra bu tarz bir durum varsa, negatif binom regresyon uygulanmalı ya da tartı parametresi ile test istatistikleri ve kalıntılar düzeltilmelidir.

Poisson regresyon modeline ilişkin parametre tahminleri, IRWLS (Iteratively reweighted least squares-İteratif olarak yeniden ağırlıklandırılmış en küçük kareler) ve maksimum olabilirlik yöntemi kullanılarak yapılmaktadır. Veriyi en iyi açıklayan yapıyı bulabilmek için uyum ölçüleri ve kalıntıların incelenmesi gerekmektedir. Genellikle sapma (deviance) değeri en küçük olan model, veriyi en iyi açıklayan model olarak tercih edilmektedir.

Boşanma verileri üzerine bir uygulama yapılmıştır. Bu çalışmada, parametre tahminleri, uyum

ölçüleri ve kalıntıların hesaplanmasında S-PLUS 2000 programı kullanılmıştır.

2. TEK AÇIKLAYICI DEĞİŞKEN ile DOĞRUSAL REGRESYON

Regresyon analizi, genellikle iki ya da daha çok nicel değişken arasındaki ilişkiyi inceleyen istatistiksel bir yöntemdir. Bu yöntemde bağımlı değişken ve bağımsız değişkenler bulunmaktadır. Bağımlı değişkenin alacağı değerler, diğer bağımsız değişken ya da değişkenlerden yararlanılarak tahmin edilmektedir. Bu yöntem, sosyal bilimlerde ve fen bilimlerinde kullanılmaktadır.

2.1 Değişkenler Arasındaki İlişkiler

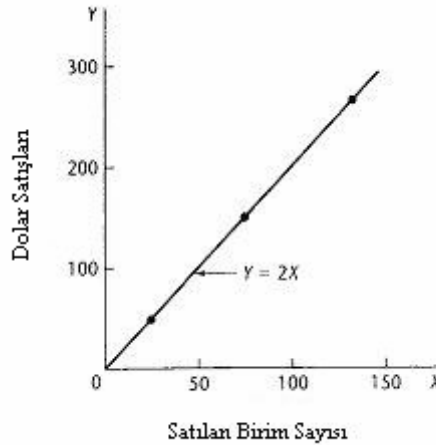
Fonksiyonel ilişki ve istatistiksel ilişki birbirinden ayrılmalıdır.

2.1.1 İki Değişken Arasındaki Fonksiyonel İlişki

İki değişken arasındaki fonksiyonel ilişki, matematiksel bir formülle ifade edilmektedir. Eğer X bağımsız değişkeni ve Y 'de bağımlı değişkeni gösterirse, fonksiyonel ilişkinin şekli aşağıdaki gibi olmaktadır:

$$Y = f(X)$$

X 'e özel bir değer verildiğinde, fonksiyon f , Y 'nin ilişkili değerini göstermektedir. Fonksiyonel ilişkide bütün gözlemler, bu ilişkinin oluşturduğu doğrunun ya da eğrinin tam olarak üzerine gelmektedir. Bu bütün fonksiyonel ilişkilerin bir özelliğidir.

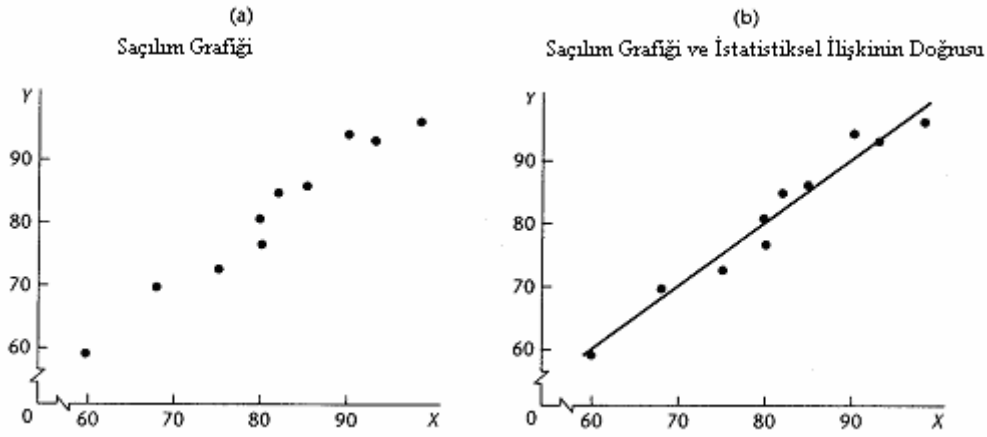


Şekil 2.1 Fonksiyonel İlişki

2.1.2 İki Değişken Arasındaki İstatistiksel İlişki

İstatistiksel ilişki, fonksiyonel ilişkiden farklı olarak determinist değildir. Genellikle, istatistiksel ilişkide gözlemler ilişkinin oluşturduğu eğrinin ya da doğrunun tam olarak üzerine gelmemektedir. İstatistiksel ilişkide genellikle, bağımlı değişken olarak adlandırılan bir değişkenle ilgili sonuçlar hakkında bilgi sahibi olmak istenilir. Bağımlı değişkenle ilgili bilgi sahibi olunmasını sağlayan değişken grubuna da bağımsız ya da açıklayıcı değişkenler denir.

İstatistiksel ilişkide diyagramdaki her nokta bir durum belirtmektedir. İstatistiksel ilişkiler fonksiyonel ilişki gibi kesin bir sonuç vermese de, oldukça faydalı olmaktadır.



Şekil 2.2 İstatistiksel İlişki

2.2 Regresyon Modelleri

2.2.1 Temel Kavramlar

Regresyon modelinde, dağılan noktalar regresyon eğrisi ya da doğrusunun çevresindedir. Regresyon modeli içinde aşağıdaki varsayımlar bulunmaktadır:

1. Her X seviyesi için Y 'nin bir olasılık dağılımı vardır.
2. Bu olasılık dağılımlarının ortalamaları X ile bazı sistematik şekillerde değişmektedir.

Y 'nin koşullu beklenen değerinin, X 'in seviyesi ile sistematik bir ilişkisi vardır. Bu sistematik ilişkiye Y 'nin X üzerinde regresyon fonksiyonu denir.

2.2.1.1 Birden Fazla Açıklayıcı Değişken ile Regresyon Modelleri

Regresyon modellerinin birden fazla açıklayıcı değişken içerdiği durumlar da olmaktadır.

Örneğin X_1 ve X_2 gibi iki açıklayıcı değişken varsa, her (X_1, X_2) kombinasyonu için Y 'nin olasılık dağılımı, regresyon modeli kullanılarak elde edilmektedir.

2.2.2 Regresyon Modellerinin Oluşturulması

2.2.2.1 Açıklayıcı Değişkenlerin Seçimi

Pek çok araştırma çalışmalarında temel problem, regresyon modeli için değişken seçimi yapmaktır. Açıklayıcı değişkenler arasından analizin amacı için en uygun sonucu verecek değişkenler seçilmelidir. Bu seçimi yaparken ana düşünce, seçilen değişkenin Y içindeki değişime olan katkısıdır.

2.2.2.2 Regresyon İlişkisinin Fonksiyonel Formu

Regresyon ilişkisinin fonksiyonel formunun seçimi, açıklayıcı değişkenin seçimi ile ilişkilidir. Bazen, teori ile ilgili uygun fonksiyonel form gösterilebilir. Eğer model oluşturulurken uygun fonksiyonel form mevcutsa, buradan hareket edilmektedir.

Ancak, çoğu kez regresyon ilişkisinin fonksiyonel formunun doğrusal ya da kuadratik (ikinci dereceden) olup olmadığı bilinmez ve toplanan veri değerlerinden karar verilmesi gerekir. Doğrusal ya da kuadratik regresyon fonksiyonları kullanıldığı zaman, bilinmeyen yapı için kullanılan regresyon fonksiyonlarının ilk yaklaşımları çoğu zaman tatmin edici sonuçlar vermektedir. Gerçekten, regresyon fonksiyonunun bu basit şekilleri, teori ilgili fonksiyonel formu sağladığı zaman, özellikle bilinen form çok kompleks olduğunda fakat doğrusal ya da kuadratik fonksiyona oldukça yaklaştığında kullanılmaktadır.

2.2.3 Regresyon Analizinin Kullanıldığı Durumlar

Regresyon analizi üç ana amacı sağlar: 1) tanımlama, 2) kontrol ve 3) tahmin. Regresyon analizi uygulamada sık sık kullanılmaktadır.

2.2.4 Regresyon ve Nedensellik

Bağımlı değişken Y ve açıklayıcı değişken X arasındaki istatistiksel ilişkinin varlığı hiçbir şekilde Y , X 'e bağlıdır anlamına gelmemektedir. X ile Y arasındaki istatistiksel ilişkinin ne kadar güçlü olduğunun bir önemi yoktur. Değişkenlerin bir regresyon modeli içerisinde birbirleri ile ilişkili bulunmaları, zorunlu olarak gerçek hayatta bu değişkenler arasında bir sebep-sonuç ilişkisi vardır anlamına gelmemektedir.

Hatta güçlü bir istatistiksel ilişki nedensel şartları yansıtsa bile, bu nedensel koşullar ters

yönde olabilir(Örneğin Y 'den X 'e doğru). Regresyon analizi nedensel ilişkiler hakkında bilgi vermez. Bu yüzden nedensel ilişkilerin varlığının saptanması ek analizler ve istatistik dışı teorik gerçeklerle bulunmalıdır. Örneğin, bir malın talebinin o malın fiyatı ile ters orantılı olduğunu istatistik teorisi değil iktisat kuramı söylemektedir.

2.3 Dağılımı Belirlenmemiş Hata Terimleri ile Yalın Doğrusal Regresyon Modeli

2.3.1 Modelin Biçimlendirilmesi

Sadece tek açıklayıcı değişken bulunduğunda ve regresyon fonksiyonu doğrusal olduğunda buna temel regresyon modeli denir. Bu modelde aşağıdaki varsayımlar yapılmaktadır:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (2.1)$$

Y_i , i .durumda bağımlı değişkenin değeri,

β_0 ve β_1 , parametreler,

X_i , bilinen sabit ve i .durumdaki bağımsız değişkenin değeri,

ε_i , ortalaması $E\{\varepsilon_i\} = 0$ ve varyansı $\sigma^2\{\varepsilon_i\} = \sigma^2$ olan rastlantısal hata terimidir; ε_i ve ε_j korelasyonsuzdur, bu yüzden kovaryansları sıfırdır.

$$\sigma(\varepsilon_i, \varepsilon_j) = 0 \quad i, j \text{ için } i \neq j, \quad i=1, \dots, n$$

Regresyon modeli (2.1)'in, yalın, parametrelerinde ve bağımsız değişkenlerinde doğrusal olduğu söylenmektedir*. Yalın denmesinin sebebi sadece bir açıklayıcı değişken olmasıdır. Parametrelerinde doğrusal denmesinin sebebi ise, modelde hiçbir parametrenin üs olarak bulunmaması ve başka parametrelerle çarpılmış ya da bölünmüş olmamasıdır. Bağımsız değişkeninde de doğrusaldır, çünkü değişkenin üssü birdir. Ayrıca, model parametrelerinde ve bağımsız değişkeninde doğrusal olduğu zaman buna birinci mertebeden model de denmektedir.

2.3.2 Modelin Önemli Özellikleri

1. i .durumdaki bağımlı değişken Y_i iki bileşenin toplamıdır: sabit terim $\beta_0 + \beta_1 X_i$ ve

* Kutner, M.H., Nachtsheim, C.J. ve Neter, J., (2004), Applied Linear Regression Models, McGraw-Hill, Boston.

rastlantısal terim ε_i . Bu yüzden, Y_i rastlantısal bir değişkendir.

2. $E\{\varepsilon_i\} = 0$ olduğu için,

$$E\{Y_i\} = E\{\beta_0 + \beta_1 X_i + \varepsilon_i\} = \beta_0 + \beta_1 X_i + E\{\varepsilon_i\} = \beta_0 + \beta_1 X_i$$

$$E\{Y_i\} = \beta_0 + \beta_1 X_i \quad (2.2)$$

Böylece (2.1) modeli için regresyon fonksiyonu aşağıdaki gibi olmaktadır:

$$E\{Y\} = \beta_0 + \beta_1 X \quad (2.3)$$

3. i .durumdaki bağımlı değişken Y_i , ε_i hata terimi ile regresyon fonksiyonunun değerini aşabilir ya da bunun gerisinde kalabilir.

4. Hata terimi ε_i 'nin sabit varyansa sahip olduğu varsayılmaktadır. Böylece bağımlı değişken Y_i 'nin sahip olduğu sabit varyans aşağıdaki gibi olmaktadır:

$$\sigma^2\{Y_i\} = \sigma^2 \quad (2.4)$$

$$\sigma^2\{\beta_0 + \beta_1 X_i + \varepsilon_i\} = \sigma^2\{\varepsilon_i\} = \sigma^2$$

Böylece, regresyon modeli (2.1), bağımsız değişken X 'in seviyesi ne olursa olsun, Y 'nin olasılık dağılımının aynı varyansa yani σ^2 'ye sahip olduğunu varsaymaktadır.

5. Hata terimlerinin korelasyonsuz olduğu varsayılmaktadır. Hata terimleri ε_i ve ε_j korelasyonsuz olduğu için, Y_i ve Y_j bağımlı değişkenleri de öyledir.

6. Özetle, X 'in her seviyesi için, regresyon modeli (2.1), bağımlı değişken Y_i 'nin ortalamasının $E\{Y_i\} = \beta_0 + \beta_1 X_i$ ve varyansının da σ^2 olduğunu göstermektedir. Ayrıca Y_i ve Y_j bağımlı değişkenleri de korelasyonsuzdurlar.

2.3.3 Regresyon Parametrelerinin Anlamı

Regresyon modeli (2.1)'deki, β_0 ve β_1 parametrelerine regresyon katsayıları denilmektedir. β_1 regresyon doğrusunun eğimidir. X 'teki her bir birim artışta Y 'deki değişimin ne kadar olacağını göstermektedir. Parametre β_0 ise, regresyon doğrusunun Y eksenini kestiği noktadır.

2.4 En Küçük Kareler Yöntemi ile Parametrelerin Tahmini

Regresyon parametreleri β_0 ve β_1 'in iyi tahmincilerini bulmak için, en küçük kareler yöntemi kullanılmaktadır. En küçük kareler yöntemi, her durumdaki (X_i, Y_i) gözlemleri için, Y_i 'nin beklenen değerinden sapmasını ifade etmektedir:

$$Y_i - (\beta_0 + \beta_1 \cdot X_i) \quad (2.5)$$

En küçük kareler yöntemi n 'e kadar olan sapmaların karelerinin toplamını kullanmaktadır. Bu toplam Q ile gösterilmektedir:

$$Q = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 \cdot X_i)^2 \quad (2.6)$$

En küçük kareler yöntemine göre, β_0 ve β_1 'in tahmincileri sırasıyla b_0 ve b_1 'dir. Bu değerler verilen örnek gözlemleri $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ için Q toplamını minimum yapmaktadırlar.

2.4.1 En Küçük Kareler Tahmincileri

b_0 ve b_1 tahmincileri, en küçük kareler yöntemi için iki temel şekilde bulunmaktadır:

1.Farklı b_0 ve b_1 tahmin çiftleri için, bu değer çiftlerinden biri Q toplamını minimum yapana kadar nümerik araştırma yöntemleri değerlendirilmektedir.

2.Eğer regresyon modeli matematiksel olarak karmaşık değilse, Q toplamını minimum yapan b_0 ve b_1 değerlerini bulmak için, analitik yöntemler sık sık kullanılmaktadır.

Analitik yaklaşım kullanılarak, (2.1) regresyon modelinde, herhangi bir örnek veri serisi için Q 'yu minimum yapan b_0 ve b_1 değerleri aşağıdaki gibi gösterilmektedir:

$$\sum Y_i = n \cdot b_0 + b_1 \cdot \sum X_i \quad (2.7)$$

$$\sum X_i \cdot Y_i = b_0 \sum X_i + b_1 \cdot \sum X_i^2 \quad (2.8)$$

(2.7) ve (2.8) denklemlerine, normal denklemler; b_0 ve b_1 'e de sırasıyla β_0 ve β_1 'in nokta tahmincileri denilmektedir.

(2.7) ve (2.8) normal denklemleri b_0 ve b_1 için eş zamanlı olarak çözülmektedir:

$$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \quad (2.9)$$

$$b_0 = \frac{1}{n} \cdot (\sum Y_i - b_1 \cdot \sum X_i) = \bar{Y} - b_1 \cdot \bar{X} \quad (2.10)$$

$\bar{X}$ ve $\bar{Y}$, sırasıyla X_i ve Y_i gözlemlerinin ortalamasıdır.

2.4.2 En Küçük Kareler Tahmincilerinin Özellikleri

Gauss-Markov Teoremi: (2.1) regresyon modelinin şartları altında, en küçük kareler tahmincileri b_0 ve b_1 , bütün yansız doğrusal tahminciler arasında yansız ve minimum varyansa sahip olanlardır.

İlk olarak, b_0 ve b_1 yansız tahmincilerdir. Bu yüzden;

$$E\{b_0\} = \beta_0 \quad E\{b_1\} = \beta_1 \text{ 'dir.}$$

İkinci olarak, teorem b_0 ve b_1 tahmincilerinin $Y_1, \dots, Y_n$ gözlemlerinin doğrusal fonksiyonlarının yansız tahminciler sınıfına ait olan diğer tahmincilerinin hepsinden daha kesin sonuç verdiğini göstermektedir.

$$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

Bu ifade aşağıdakine eşittir:

$$b_1 = \frac{\sum (X_i - \bar{X})Y_i}{\sum (X_i - \bar{X})^2} = \sum k_i Y_i \text{ ve } k_i = \frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2} \text{ 'dir.}$$

k_i 'ler bilinen sabitler olduğu için, b_1 , Y_i 'nin doğrusal bir kombinasyonudur, bu yüzden doğrusal bir tahmincidir.

Aynı şekilde b_0 'ın da doğrusal bir tahminci olduğu gösterilmektedir. Yansız olan bütün doğrusal tahminciler arasında, b_0 ve b_1 , X seviyeleri değişmeden kaldığında tekrar eden örnekler içinde en küçük değişkenliğe sahip olmaktadır.

2.5 Kalıntılar

i . kalıntı, gözlenen değer Y_i ile buna uyan tahmin değeri $\hat{Y}_i$ arasındaki farktır. Bu kalıntı e_i ile gösterilmektedir ve aşağıdaki şekilde tanımlanmaktadır:

$$e_i = Y_i - \hat{Y}_i \quad (2.11)$$

(2.1) regresyon modeli için, kalıntı e_i :

$$e_i = Y_i - (b_0 + b_1.X_i) = Y_i - b_0 - b_1.X_i \quad (2.12)$$

olmaktadır. Modelin hata terimi değeri $\varepsilon_i = Y_i - E\{Y_i\}$ ve kalıntı $e_i = Y_i - \hat{Y}_i$ birbirinden farklıdır. İlkinde, bilinmeyen gerçek regresyon doğrusundan Y_i 'nin düşey sapması gösterilmektedir ve dolayısıyla bilinmemektedir. Diğer yandan, kalıntı, tahmin edilen regresyon doğrusu üzerinde tahmin değeri $\hat{Y}_i$ 'den, Y_i 'nin düşey sapmasını göstermektedir ve bu hesaplanmaktadır.

2.6 Tahmin Edilen Regresyon Doğrusunun Özellikleri

En küçük kareler yöntemiyle tahmin edilen doğrusal regresyon modeli pek çok özelliğe sahiptir. Bu özellikler bütün regresyon modelleri için geçerli değildir.

1.Kalıntıların toplamı sıfırdır:

$$\sum_{i=1}^n e_i = 0 \quad (2.13)$$

2.Kalıntı karelerinin toplamı, $\sum e_i^2$ minimumdur. En küçük kareler tahminçileri b_0 ve b_1 , β_0 ve β_1 'i tahmin etmek için kullanıldığı zaman, b_0 ve b_1 denklemde yerine konduğunda (2.6) toplamının minimize olarak $\sum e_i^2$ 'ye eşit olması gerekmektedir.

3.Gözlenen değer Y_i 'lerin toplamı, tahmin değerleri $\hat{Y}_i$ 'lerin toplamına eşittir:

$$\sum_{i=1}^n Y_i = \sum_{i=1}^n \hat{Y}_i \quad (2.14)$$

Bu demektir ki tahmin değerleri $\hat{Y}_i$ 'lerin ortalaması ile gözlenen değerler Y_i 'lerin ortalaması birbirine eşittir. Bu da, $\bar{Y}$ 'dir.

4. i . durumdaki kalıntı, i . durumdaki açıklayıcı değişken ile ağırlıklandırılır. Bu durumda ağırlıklı kalıntıların toplamı sıfır olur.

$$\sum_{i=1}^n X_i \cdot e_i = 0 \quad (2.15)$$

$\sum_{i=1}^n X_i \cdot e_i = 0$ olması, X_i ve e_i arasındaki bağımsızlık, yani korelasyonsuzluğu da göstermektedir. Bu durum X ve e vektörlerinin dik olması anlamına gelmektedir.

5.(2.13) ve (2.15) özelliklerinin bir sonucu olarak, i . durumdaki kalıntı bağımlı değişkenin i . durumdaki tahmin değeri ile ağırlıklandırılır. Bu durumda ağırlıklı kalıntıların toplamı sıfır olmaktadır.

$$\sum_{i=1}^n \hat{Y}_i \cdot e_i = 0 \quad (2.16)$$

6.Regresyon doğrusu daima $(\bar{X}, \bar{Y})$ noktasından geçer.

2.7 Hata Terimlerinin Varyansının Tahmini

2.7.1 σ^2 'nin Nokta Tahmincisi

(2.1) regresyon modelindeki, σ^2 'nin tahmincisini bulabilmek için, ilk olarak tek bir topluluktan örnekleme yapmak düşünülmelidir.

2.7.1.1 Açıklayıcı Değişkenin Bulunmadığı Durum

Tek bir topluluğun varyansı σ^2 , örnek varyansı s^2 'den tahmin edilmektedir. Örnek varyansı s^2 'yi bulurken, şu yol izlenmektedir:

$$\sum_{i=1}^n (Y_i - \bar{Y})^2$$

Bu toplama kareler toplamı denir. Sonra kareler toplamı ilişkili olan serbestlik derecesine yani $n-1$ 'e bölünür. Sonuçta oluşan tahminci aşağıdaki gibidir:

$$s^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}$$

s^2 , sınırsız topluluğun varyansı σ^2 'nin yansız bir tahmincisidir.

2.7.1.2 Regresyon Modeli

$\sigma^2\{Y_i\} = \sigma^2$ eşitliğinden, her hata terimi ε_i 'ninki ile aynı olarak, (2.1) regresyon modeli için her Y_i gözleminin varyansı σ^2 'dir. Kalıntılar aşağıdaki gibidir:

$$Y_i - \hat{Y}_i = e_i$$

ve uygun kareler toplamı, SSE şeklinde gösterilmektedir:

$$SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n e_i^2 \quad (2.17)$$

SSE'ye de hata kareleri toplamı ya da kalıntı kareleri toplamı denilmektedir(SSE=Error sum of squares ya da residual sum of squares).

Kareler toplamı SSE, $n-2$ serbestlik derecesine sahiptir. $\hat{Y}_i$ 'lerin tahmin edilmesi için b_0 , b_1 bulunmalıdır. Bu yüzden iki serbestlik derecesi kaybedilmektedir. Uygun kare ortalaması, MSE ya da s^2 şeklinde gösterilmektedir:

$$s^2 = MSE = \frac{SSE}{n-2} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-2} = \frac{\sum e_i^2}{n-2} \quad (2.18)$$

MSE'ye ortalama hata kare ya da kalıntı kareleri ortalaması denilmektedir(MSE=Error mean square ya da residual mean square).

MSE, σ^2 'nin yansız bir tahmincisidir.

$$E\{MSE\} = \sigma^2 \quad (2.19)$$

Standart sapma σ 'nın bir tahmincisi de, $s = \sqrt{MSE}$ olarak, yani MSE'nin pozitif kare kökü şeklinde gösterilmektedir.

2.8 Normal Dağılan Hata Terimli Regresyon Modeli

Aralık tahminlerini ve testlerini yapabilmek için ε_i 'lerin dağılımının şekli hakkında bir varsayım yapılması gerekmektedir. Standart varsayım, hata terimleri ε_i 'lerin normal dağıldığıdır.

2.8.1 Model

Normal dağılan hata terimli regresyon modeli aşağıdaki gibidir:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (2.20)$$

Y_i , i .durumda gözlenen bağımlı değişken,

X_i , bilinen sabit ve i .durumdaki bağımsız değişken,

β_0 ve β_1 , parametreler,

ε_i , bağımsız $N(0, \sigma^2)$ ($i=1, \dots, n$)'dir.

2.8.2 Maksimum Olabilirlik Yöntemi ile Parametrelerin Tahmini

Hata terimlerinin olasılık dağılımlarının fonksiyonel formu belirlendiği zaman, β_0 , β_1 ve σ^2 parametre tahminleri, maksimum olabilirlik yöntemi ile bulunmaktadır. İlk olarak, tek bir topluluk, bir parametre ile örneklendiği zaman maksimum olabilirlik yöntemi açıklanacaktır.

Maksimum olabilirlik yöntemi, Y_i gözlemi için bir tutarlılık ölçüsü olarak, Y_i 'deki olasılık dağılımının yoğunluğunu kullanmaktadır. Normal bir rastlantısal değişken olan Y 'nin yoğunluk fonksiyonu aşağıdaki gibidir:

$$f(Y) = \frac{1}{\sqrt{2\pi\sigma}} \cdot \exp\left[-\frac{1}{2} \cdot \left(\frac{Y - \mu}{\sigma}\right)^2\right] \quad -\infty < Y < \infty \quad (2.21)$$

μ ve σ normal dağılımın iki parametresidir. Normal bir rastlantısal değişken olan Y 'nin ortalama ve varyansı aşağıdaki gibidir:

$$E\{Y\} = \mu \quad (2.22)$$

$$\sigma^2\{Y\} = \sigma^2 \quad (2.23)$$

Maksimum olabilirlik yöntemi, örnek veri ile parametre değerinin tutarlılığının ölçüsü olarak yoğunlukların çarpımını kullanmaktadır. Sonuca, μ parametre değerinin olabilirlik değeri denilmekte ve $L(\mu)$ ile gösterilmektedir.

Maksimum olabilirlik yöntemi, olabilirlik değeri en büyük olan μ değerini maksimum olabilirlik tahmini olarak seçmektedir. En küçük kareler yönteminde olduğu gibi, maksimum

olabilirlik tahminlerini bulmak için iki yöntem bulunmaktadır: sistematik nümerik arama ve analitik çözümü kullanma.

Bilinmeyen parametrelerin fonksiyonunda olduğu gibi, yoğunlukların çarpımına olabilirlik fonksiyonu denilmektedir.

2.8.2.1 Regresyon Modeli

(2.20) regresyon modeli için, her Y_i gözlemi, ortalaması $\beta_0 + \beta_1.X_i$ ve standart sapması σ olan normal dağılım göstermektedir.

$E\{Y_i\} = \beta_0 + \beta_1.X_i$ ve $\sigma^2\{Y_i\} = \sigma^2$ kullanılarak (2.20) regresyon modeli için Y_i gözleminin yoğunluğu aşağıdaki gibi olmaktadır:

$$f_i = \frac{1}{\sqrt{2\pi\sigma}} \cdot \exp\left[-\frac{1}{2} \cdot \left(\frac{Y_i - \beta_0 - \beta_1.X_i}{\sigma}\right)^2\right] \quad (2.24)$$

$Y_1, Y_2, \dots, Y_n$ şeklinde n gözlem için, olabilirlik fonksiyonu, (2.24)'teki gibi her birinin yoğunluğunun çarpımıdır. Hata terimlerinin varyansı σ^2 genellikle bilinmediği için, olabilirlik fonksiyonu β_0, β_1 ve σ^2 şeklinde üç parametrenin bir fonksiyonudur:

$$\begin{aligned} L(\beta_0, \beta_1, \sigma^2) &= \prod_{i=1}^n \frac{1}{(2\pi\sigma^2)^{1/2}} \cdot \exp\left[-\frac{1}{2\sigma^2} (Y_i - \beta_0 - \beta_1.X_i)^2\right] \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \cdot \exp\left[-\frac{1}{2\sigma^2} \cdot \sum_{i=1}^n (Y_i - \beta_0 - \beta_1.X_i)^2\right] \end{aligned} \quad (2.25)$$

β_0, β_1 ve σ^2 değerleri bu olabilirlik fonksiyonunu maksimum yapan, maksimum olabilirlik tahminçileridir ve sırasıyla $\hat{\beta}_0, \hat{\beta}_1$ ve $\hat{\sigma}^2$ olarak gösterilmektedirler. Bu tahminçiler analitik olarak bulunabilmektedir ve aşağıdaki gibidirler:

Çizelge 2.1 Maksimum olabilirlik tahmincileri

Parametre	Maksimum Olabilirlik Tahmincileri
β_0	$\hat{\beta}_0 = b_0$
β_1	$\hat{\beta}_1 = b_1$
σ^2	$\hat{\sigma}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n}$

Böylece, β_0 ve β_1 'in maksimum olabilirlik tahmincileri en küçük kareler yöntemiyle sağlanan tahmincilerle aynı olmaktadır. Maksimum olabilirlik tahmincisi $\hat{\sigma}^2$ yansızdır ve genellikle yansız tahminci MSE, (2.18)'deki gibidir.

Yansız tahminci MSE ya da s^2 , özellikle n küçük değilse, aşağıda gösterildiği gibi maksimum olabilirlik tahmincisi $\hat{\sigma}^2$ 'den biraz farklıdır:

$$s^2 = MSE = \frac{n}{n-2} \cdot \hat{\sigma}^2 \quad (2.26)$$

Açıklamalar

1. Maksimum olabilirlik tahmincileri $\hat{\beta}_0$ ve $\hat{\beta}_1$, en küçük kareler tahmincileri b_0 ve b_1 ile aynıdır ve bütün en küçük kareler tahmincilerinin bulundurdıkları özelliklere sahiptirler.

a) Yansızdırlar.

b) Bütün yansız doğrusal tahminciler arasında minimum varyansa sahiptirler. Ayrıca normal dağılan hata terimli regresyon modeli için maksimum olabilirlik tahmincileri b_0 ve b_1 istenen diğer özelliklere de sahiptirler.

c) Tutarlıdırlar.

d) Yeterlidirler.

e) Yansız minimum varyansa sahiptirler. Bu, doğrusal olan ya da olmayan bütün yansız tahminciler içinde minimum varyansa sahip oldukları anlamına gelmektedir.

Böylece, normal dağılan hata terimli regresyon modeli için, b_0 ve b_1 tahmincilerinin istenen özelliklere sahip oldukları görülmektedir.

2. β_0 , β_1 ve σ^2 'nin değerleri, L 'nin β_0 , β_1 ve σ^2 'ye göre kısmi türevleri alındıktan sonra her kısmi türev sıfıra eşitlenip denklem sistemi çözülerek, yani (2.25)'deki olabilirlik fonksiyonu L maksimize edilerek bulunmaktadır. L 'den çok, $\log_e L$ ile çalışılmaktadır. Çünkü L ve $\log_e L$ 'nin her ikisi de β_0 , β_1 ve σ^2 'nin aynı değerleri için maksimize olmaktadır:

$$\log_e L = -\frac{n}{2} \cdot \log_e 2\pi - \frac{n}{2} \cdot \log_e \sigma^2 - \frac{1}{2\sigma^2} \sum (Y_i - \beta_0 - \beta_1 \cdot X_i)^2 \quad (2.27)$$

3. REGRESYON ANALİZİNDEN ÇIKARILAN SONUÇLAR

Literatürde aksi belirtilmediği sürece, bu, normal dağılan hata terimli regresyon modelinin kullanıldığı anlamına gelmektedir.

Bu model aşağıdaki gibidir:

$$Y_i = \beta_0 + \beta_1 \cdot X_i + \varepsilon_i \quad (3.1)$$

Burada;

β_0 ve β_1 , parametreler,

X_i , bilinen sabit,

ε_i , bağımsız ve $N(0, \sigma^2)$ dağılmaktadır.

3.1 β_1 ile İlgili Aralık Tahminleri ve Hipotez Testleri

3.1.1 b_1 'in Örneklem Dağılımı

Nokta tahmincisi b_1 aşağıda gösterildiği gibidir:

$$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \quad (3.2)$$

Normal dağılan hata terimli regresyon modeli (3.1) için, b_1 'in örneklem dağılımı aşağıdaki ortalama ve varyans ile normaldir:

$$E\{b_1\} = \beta_1 \quad (3.3)$$

$$\sigma^2\{b_1\} = \frac{\sigma^2}{\sum (X_i - \bar{X})^2} \quad (3.4)$$

3.1.1.1 Normallik

b_1 'in örneklem dağılımının normalliği, b_1 , Y_i 'nin doğrusal bir kombinasyonu olduğu için şöyle açıklanmaktadır. Normal dağılan hata terimli regresyon modeli (3.1)'e göre, hata terimleri bağımsız normal dağıldığı için, Y_i 'de bağımsız normal dağılacaktır. Y_i bağımsız normal dağıldığından, normal rastlantısal bir değişkenin doğrusal bir kombinasyonu olan

b_1 'de normal dağılacaktır.

3.1.1.2 Varyansın Tahmini

b_1 'in örnekleme dağılımının varyansı şöyle tahmin edilmektedir:

$$\sigma^2\{b_1\} = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}$$

σ^2 parametresinin yerine MSE yazılırsa, σ^2 'nin yansız tahmincisi:

$$s^2\{b_1\} = \frac{MSE}{\sum (X_i - \bar{X})^2} \quad (3.5)$$

olmaktadır. Nokta tahmincisi $s^2\{b_1\}$, $\sigma^2\{b_1\}$ 'in yansız bir tahmincisidir. $s^2\{b_1\}$ 'in pozitif kare kökü alınır, $s\{b_1\}$ bulunmaktadır. Bu da, $\sigma\{b_1\}$ 'in nokta tahmincisi olmaktadır.

3.1.2 $(b_1 - \beta_1)/s\{b_1\}$ 'in Örnekleme Dağılımı

b_1 normal dağıldığı için, standardize istatistik $(b_1 - \beta_1)/s\{b_1\}$, bir standart normal değişken olmaktadır. Genellikle, $s\{b_1\}$ 'den, $\sigma\{b_1\}$ 'in tahmin edilmesi gerekmektedir ve bu yüzden $(b_1 - \beta_1)/s\{b_1\}$ istatistiğinin dağılımıyla ilgilenilir. Bir istatistik standardize edildiği zaman, paydadaki tahmin edilen standart sapma gerçek standart sapmadan daha büyük oluyorsa buna student istatistiği denilmektedir. $(b_1 - \beta_1)/s\{b_1\}$ student istatistiği hakkındaki teorem aşağıda gösterildiği gibidir:

(3.1) regresyon modeli için, $\frac{b_1 - \beta_1}{s\{b_1\}}$, $t(n-2)$ olarak dağılır.

Eğer Y_i gözlemleri aynı normal topluluktan geliyorsa, $(\bar{Y} - \mu)/s\{\bar{Y}\}$ 'nin $n-1$ serbestlik dereceli t dağılımı göstereceği bilinmektedir. Tahminci b_1 , $\bar{Y}$ gibi, Y_i gözlemlerinin doğrusal bir kombinasyonudur. Serbestlik dereceleri arasındaki farkın sebebi ise, regresyon modeli için iki parametrenin (β_0 ve β_1) tahmin edilmesinin gerekmesidir. Bu yüzden, burada iki serbestlik derecesi kaybedilmektedir.

3.1.3 β_1 için Güven Aralığı

$(b_1 - \beta_1)/s\{b_1\}$, t dağılımı göstermektedir. β_1 için, $1-\alpha$ anlamlılık düzeyindeki güven

sınırları:

$$b_1 \pm t(1 - \alpha/2; n - 2)s\{b_1\} \quad (3.6)$$

olmaktadır.

3.1.4 β_1 ile İlgili Testler

$(b_1 - \beta_1)/s\{b_1\}$, $n-2$ serbestlik dereceli t dağılımı gösterdiği için, β_1 ile ilgili testler t dağılımı kullanılarak kurulmaktadır:

3.1.4.1 Çift Taraflı Test

$\beta_1 = 0$ olup olmadığının test edilmesine çift taraflı test denilmektedir. Hipotez şöyle kurulmaktadır:

$$H_0 : \beta_1 = 0 \quad (3.7)$$

$$H_1 : \beta_1 \neq 0 \quad (3.8)$$

Test istatistiği aşağıdaki gibidir:

$$t^* = \frac{b_1}{s\{b_1\}} \quad (3.9)$$

Bu test istatistiğini kontrol etmek için kullanılan α anlamlılık derecesindeki karar kuralı:

$$\text{Eğer } |t^*| \leq t(1 - \alpha/2; n - 2) \text{ ise, } H_0 \text{ reddedilmez.} \quad (3.10)$$

$$\text{Eğer } |t^*| > t(1 - \alpha/2; n - 2) \text{ ise, } H_0 \text{ reddedilir.} \quad (3.11)$$

3.1.4.2 Tek Taraflı Test

β_1 'in pozitif olup olmadığı α anlamlılık seviyesine göre test edilmek istendiğinde, hipotez aşağıdaki gibi kurulmaktadır:

$$H_0 : \beta_1 \leq 0 \quad (3.12)$$

$$H_1 : \beta_1 > 0 \quad (3.13)$$

ve $t^* = \frac{b_1}{s\{b_1\}}$ istatistiği kullanılarak karar kuralı şöyle olur:

Eğer $t^* \leq t(1 - \alpha; n - 2)$ ise, H_0 reddedilmez. (3.14)

Eğer $t^* > t(1 - \alpha; n - 2)$ ise, H_0 reddedilir. (3.15)

3.2 β_0 ile İlgili Aralık Tahminleri ve Hipotez Testleri

3.2.1 b_0 'ın Örneklem Dağılımı

Nokta tahmini b_0 aşağıda belirtildiği gibi olmaktadır:

$$b_0 = \bar{Y} - b_1 \bar{X} \quad (3.16)$$

(3.1) regresyon modeli için, b_0 'ın örneklem dağılımı normaldir. Ortalama ve varyansı:

$$E\{b_0\} = \beta_0 \quad (3.17)$$

$$\sigma^2\{b_0\} = \sigma^2 \cdot \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right] \quad (3.18)$$

olmaktadır.

b_0 'ın örneklem dağılımı normaldir. Çünkü b_0 , Y_i gözlemlerinin doğrusal bir kombinasyonudur. b_0 'ın örneklem dağılımının ortalama ve varyansı b_1 'inkine benzer şekilde bulunmaktadır.

$\sigma^2\{b_0\}$ 'ın tahmincisi, σ^2 'nin yerine nokta tahmini MSE'nin getirilmesiyle bulunmaktadır.

$$s^2\{b_0\} = MSE \cdot \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right] \quad (3.19)$$

Pozitif karekök $s\{b_0\}$, $\sigma\{b_0\}$ 'ın tahmincisi olmaktadır.

3.2.2 $(b_0 - \beta_0)/s\{b_0\}$ 'ın Örneklem Dağılımı

b_1 'deki gibi, b_0 için de benzer bir teorem bulunmaktadır:

(3.1) regresyon modeli için, $\frac{b_0 - \beta_0}{s\{b_0\}}$, $t(n - 2)$ dağılmaktadır.

Bu yüzden, β_0 için güven aralığı ve β_0 ile ilgili testler, t dağılımı kullanılarak kurulmaktadır.

3.2.3 β_0 için Güven Aralığı

$(b_0 - \beta_0)/s\{b_0\}$, t dağılımı göstermektedir. β_0 için, $1 - \alpha$ 'lık güven aralıkları β_1 'inkine benzer şekilde bulunmaktadır:

$$b_0 \pm t(1 - \alpha/2; n - 2).s\{b_0\} \quad (3.20)$$

3.3 Varyans Analizi Tablosu

Kareler toplamının ve ilgili serbestlik derecesinin hesaplanması aşağıdaki gibi bir varyans analizi tablosu şeklinde gösterilmektedir. Aynı zamanda ilgili kareler ortalaması da gösterilmiştir. ANOVA tablosu, ayrıca kareler ortalamasının beklenen değerinin bulunduğu bir sütunu da içermektedir.

Çizelge 3.1 Yalın doğrusal regresyon modeli için ANOVA tablosu

Değişim Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	Kareler Ortalamasının Beklenen Değeri [E{MS}]
Regresyon	$SSR = \sum (\hat{Y}_i - \bar{Y})^2$	1	$MSR = \frac{SSR}{1}$	$\sigma^2 + \beta_1^2 \sum (X_i - \bar{X})^2$
Hata	$SSE = \sum (Y_i - \hat{Y}_i)^2$	n-2	$MSE = \frac{SSE}{n - 2}$	σ^2
Toplam	$SSTO = \sum (Y_i - \bar{Y})^2$	n-1		

3.4 X ve Y Arasındaki Doğrusal İlişkinin Tanımlayıcı Ölçüleri

3.4.1 Belirginlik Katsayısı

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO} \quad (3.21)$$

R^2 ölçüsüne belirginlik katsayısı denilmektedir. $0 \leq SSE \leq SSTO$ olacağı için;

$$0 \leq R^2 \leq 1 \quad (3.22)$$

elde edilmektedir.

3.4.2 Korelasyon Katsayısı

Y ve X 'in her ikisi de rastlantısal olduğunda, Y ve X arasındaki doğrusal ilişkinin ölçüsü korelasyon katsayısıdır. Bu ölçü, R^2 'nin karekökü alınarak bulunmaktadır:

$$r = \pm\sqrt{R^2} \quad (3.23)$$

Bu ölçünün alacağı artı ya da eksi işaret tahmin edilen regresyon doğrusunun eğiminin pozitif ya da negatif olmasına bağlıdır.

4. YALIN DOĞRUSAL REGRESYON ANALİZİNDE MATRİS YAKLAŞIMI

4.1 Matris Gösterimiyle Yalın Doğrusal Regresyon Modeli

Normal dağılan hata terimli regresyon modeli (3.1) için:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad i = 1, \dots, n \quad (4.1)$$

Bu,

$$Y_1 = \beta_0 + \beta_1 X_1 + \varepsilon_1$$

$$Y_2 = \beta_0 + \beta_1 X_2 + \varepsilon_2$$

$\vdots$

(4.2)

$$Y_n = \beta_0 + \beta_1 X_n + \varepsilon_n$$

anlamına gelmektedir. Gözlemler vektörü Y , X matrisi ve ε vektörü sırasıyla aşağıdaki gibi tanımlanmaktadır:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1}, \quad X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}_{n \times 2}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1} \quad (4.3)$$

Ayrıca regresyon katsayılarının vektörü de β olarak tanımlansın:

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}_{2 \times 1} \quad (4.4)$$

Böylece normal dağılan hata terimli regresyon modeli aşağıdaki gibi gösterilmektedir:

$$Y = X \cdot \beta + \varepsilon \quad (4.5)$$

$\begin{matrix} n \times 1 & n \times 2 & 2 \times 1 & n \times 1 \end{matrix}$

Çünkü:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$= \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \vdots \\ \beta_0 + \beta_1 X_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 + \varepsilon_1 \\ \beta_0 + \beta_1 X_2 + \varepsilon_2 \\ \vdots \\ \beta_0 + \beta_1 X_n + \varepsilon_n \end{bmatrix}$$

olmaktadır. $E\{Y_i\} = \beta_0 + \beta_1 X_i$ olduğu için, Y_i gözlemlerinin beklenen değerlerinin vektörü $X\beta$ vektörüdür; bu yüzden:

$$E\{Y\} = X\beta \quad (4.6)$$

$n \times 1 \quad \quad n \times 1$

elde edilmektedir.

(3.1) regresyon modeli, hata terimlerine göre, $E\{\varepsilon_i\} = 0$, $\sigma^2\{\varepsilon_i\} = \sigma^2$ ve ε_i 'lerin bağımsız normal rastlantısal değişkenler olduğunu varsaymaktadır. $E\{\varepsilon_i\} = 0$ şartı, matris biçimiyle şöyle gösterilmektedir:

$$E\{\varepsilon\} = 0 \quad (4.7)$$

$n \times 1 \quad \quad n \times 1$

Çünkü:

$$\begin{bmatrix} E\{\varepsilon_1\} \\ E\{\varepsilon_2\} \\ \vdots \\ E\{\varepsilon_n\} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \text{dır.}$$

Hata terimlerinin sabit varyans σ^2 'ye sahip olması ve $i \neq j$ için, $\sigma\{\varepsilon_i, \varepsilon_j\}$ bütün kovaryanslar $\sigma\{\varepsilon_i, \varepsilon_j\}$ 'lerin sıfır olması koşulu matris gösterimiyle hata terimlerinin varyans-kovaryans matrisine göre şöyle açıklanmaktadır:

$$\sigma^2\{\varepsilon\} = \begin{bmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{bmatrix} \quad (4.8)$$

$n \times n$

Bu bir skaler matris olduğu için, aşağıdaki gibi, daha kolay bir şekilde ifade edilmektedir:

$$\sigma^2\{\varepsilon\} = \sigma^2 I \quad (4.9)$$

$n \times n \quad \quad n \times n$

Böylece, matris gösterimiyle normal dağılan hata terimli regresyon modeli (3.1), aşağıdaki gibi yazılmaktadır:

$$Y = X\beta + \varepsilon \quad (4.10)$$

Burada, ε , $E\{\varepsilon_i\} = 0$ ve $\sigma^2\{\varepsilon\} = \sigma^2 I$ ile dağılan bağımsız normal rastlantısal değişkenlerin vektörüdür.

4.2 Regresyon Parametrelerinin En Küçük Kareler Tahmini

4.2.1 Normal Denklemler

Normal denklemler (2.7) ve (2.8):

$$nb_0 + b_1 \sum X_i = \sum Y_i \quad (4.11)$$

$$b_0 \sum X_i + b_1 \sum X_i^2 = \sum X_i Y_i \quad (4.12)$$

şeklindedir. Matris gösterimi ile:

$$\underset{2 \times 2}{(X'X)} \cdot \underset{2 \times 1}{b} = \underset{2 \times 1}{(X'Y)} \quad (4.13)$$

olarak ifade edilmektedir. b , en küçük kareler regresyon katsayılarının vektörüdür:

$$\underset{2 \times 1}{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \quad (4.14)$$

Ayrıca bunlar $(X'X)$ ve $(X'Y)$ çarpımlarından şu şekilde de gösterilmektedir:

$$\begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

$$\begin{bmatrix} nb_0 + b_1 \sum X_i \\ b_0 \sum X_i + b_1 \sum X_i^2 \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

4.2.2 Tahmin Edilen Regresyon Katsayıları

Matris yöntemiyle normal denklemlerden tahmin edilen regresyon katsayılarını bulmak için, denklemin her iki tarafı da baştan $(X'X)$ 'in tersi ile çarpılmaktadır:

$$(X'.X)^{-1}.(X'.X) = (X'.X)^{-1}(X'.Y)$$

Buradan, $(X'.X)^{-1}.(X'.X) = I$ ve $Ib = b$ olacağı için:

$$\underset{2 \times 1}{b} = \underset{2 \times 2}{(X'.X)^{-1}} . \underset{2 \times 1}{(X'.Y)} \quad (4.15)$$

elde edilmektedir. b içindeki, b_1 ve b_0 tahmincileri daha önce (2.9) ve (2.10) denklemlerinde bulunan değerlerle aynıdır.

4.3 Tahmin Değerleri

Tahmin değerleri $\hat{Y}_i$ 'lerin vektörü $\hat{Y}$ ile gösterilsin. Böylece:

$$\underset{n \times 1}{\hat{Y}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} \quad (4.16)$$

olmaktadır. Matris gösterimiyle ise aşağıdaki gibidir:

$$\underset{n \times 1}{\hat{Y}} = \underset{n \times 2}{X} . \underset{2 \times 1}{b} \quad (4.17)$$

Çünkü:

$$\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} b_0 + b_1 X_1 \\ b_0 + b_1 X_2 \\ \vdots \\ b_0 + b_1 X_n \end{bmatrix}$$

4.3.1 Projeksiyon Matrisi

(4.17)'deki $\hat{Y}$ için matris sonuçları, b yerine, (4.15) değeri yazılarak bulunmaktadır:

$\hat{Y} = X(X'.X)^{-1}.(X'.Y)$ ya da aşağıdaki gibi gösterilmektedir:

$$\underset{n \times 1}{\hat{Y}} = \underset{n \times n}{H} . \underset{n \times 1}{Y} \quad (4.18)$$

Burada aşağıdaki gibi ifade edilmektedir:

$$H = X(X'X)^{-1}X' \quad (4.19)$$

$n \times n$

(4.18)'den de anlaşılacağı gibi, tahmin değerleri $\hat{Y}_i$, gözlenen bağımlı değişken değerleri Y_i 'lerin ve H matrisinin elemanları olarak bulunan katsayıların doğrusal kombinasyonu olarak ifade edilmektedir.

$n \times n$ boyutlu bir kare matris olan H matrisine projeksiyon matrisi denir. H matrisi simetrik ve bazı özelliklere sahiptir. Bu özellikler H matrisinin eşgüçlü olmasından kaynaklanmaktadır:

$$H.H=H$$

4.4 Kalıntılar

Kalıntıların vektörü $e_i = Y_i - \hat{Y}_i$ olsun ve e ile gösterilsin:

$$e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \quad (4.20)$$

$n \times 1$

Matris gösterimi ile aşağıdaki gibi olmaktadır:

$$e = Y - \hat{Y} = Y - Xb \quad (4.21)$$

$n \times 1 \quad n \times 1 \quad n \times 1 \quad n \times 1 \quad n \times 1$

4.4.1 Kalıntıların Varyans-Kovaryans Matrisi

e_i kalıntıları, tahmin değerleri $\hat{Y}_i$ gibi, bağımlı değişkenin gözlem değerleri Y_i 'nin doğrusal bir kombinasyonu olarak gösterilmektedir. Burada $\hat{Y}$ için, (4.18) denklemi kullanılmaktadır:

$$e = Y - \hat{Y} = Y - HY = (I - H)Y$$

Buradan aşağıdaki sonuç elde edilmektedir:

$$e = (I - H)Y \quad (4.22)$$

$n \times 1 \quad n \times n \quad n \times n \quad n \times 1$

Buradaki H matrisi projeksiyon matrisidir. Böylece, $I-H$ matrisi de, H matrisi gibi simetrik ve eşgüçlüdür.

Kalıntıların vektörünün varyans-kovaryans matrisi ***I-H*** matrisini içermektedir:

$$\sigma^2_{n \times n}\{e\} = \sigma^2(I - H) \quad (4.23)$$

ve tahmini aşağıdaki gibidir:

$$s^2_{n \times n}\{e\} = MSE(I - H) \quad (4.24)$$

4.4.2 Kareler Toplamının Kuadratik Formda Gösterilmesi

Kareler toplamları kuadratik formlar şeklinde gösterilmektedir. $Y'.A.Y$ şeklindeki bir ifadeye kuadratik form denir. Burada A simetrik bir matristir. ANOVA kareler toplamı aşağıdaki gibi ifade edilmektedir:

$$SSTO = Y' \left[I - \left(\frac{1}{n} \right) J \right] Y \quad (4.25)$$

$$SSE = Y' (I - H) Y \quad (4.26)$$

$$SSR = Y' \left[H - \left(\frac{1}{n} \right) J \right] Y \quad (4.27)$$

4.5 Regresyon Analiziyle İlgili Sonuçlar

4.5.1 Regresyon Katsayıları

b 'nin varyans-kovaryans matrisi:

$$\sigma^2_{2 \times 2}\{b\} = \begin{bmatrix} \sigma^2\{b_0\} & \sigma\{b_0, b_1\} \\ \sigma\{b_0, b_1\} & \sigma^2\{b_1\} \end{bmatrix} \quad (4.28)$$

$$\sigma^2_{2 \times 2}\{b\} = \sigma^2 (X'.X)^{-1} \quad (4.29)$$

$$\text{ya da } (X'.X)^{-1}_{2 \times 2} = \begin{bmatrix} \frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} & \frac{-\bar{X}}{\sum (X_i - \bar{X})^2} \\ \frac{-\bar{X}}{\sum (X_i - \bar{X})^2} & \frac{1}{\sum (X_i - \bar{X})^2} \end{bmatrix} \quad \text{olduğu için aşağıdaki sonuç elde}$$

edilmektedir:

$$\sigma^2\{b\}_{2 \times 2} = \begin{bmatrix} \frac{\sigma^2}{n} + \frac{\sigma^2 \bar{X}^2}{\sum (X_i - \bar{X})^2} & \frac{-\bar{X}\sigma^2}{\sum (X_i - \bar{X})^2} \\ \frac{-\bar{X}\sigma^2}{\sum (X_i - \bar{X})^2} & \frac{\sigma^2}{\sum (X_i - \bar{X})^2} \end{bmatrix} \quad (4.30)$$

(4.30)'da, σ^2 yerine MSE konulduğu zaman, b 'nin tahmin edilen varyans-kovaryans matrisi bulunmaktadır ve $s^2\{b\}$ ile gösterilmektedir:

$$s^2\{b\}_{2 \times 2} = MSE(X'X)^{-1} = \begin{bmatrix} \frac{MSE}{n} + \frac{\bar{X}^2.MSE}{\sum (X_i - \bar{X})^2} & \frac{-\bar{X}.MSE}{\sum (X_i - \bar{X})^2} \\ \frac{-\bar{X}.MSE}{\sum (X_i - \bar{X})^2} & \frac{MSE}{\sum (X_i - \bar{X})^2} \end{bmatrix} \quad (4.31)$$

Tıpkı tahmin edilen varyans $s^2\{b\}_{2 \times 2}$ 'deki gibi, $\sigma^2\{b\}_{2 \times 2}$ ifadesinde de, b_0 'ın varyansı, b_1 'in varyansı ve ayrıca b_0 ve b_1 'in kovaryansı da bulunmaktadır.

4.5.2 Bağımlı Değişkenin Beklenen Değeri (Mean Response)

X_h değeri için, bağımlı değişkenin beklenen değeri tahmin edilmek istendiğinde, aşağıdaki vektörü tanımlamak gerekmektedir:

$$X_h = \begin{bmatrix} 1 \\ X_h \end{bmatrix}_{2 \times 1} \quad \text{ya da} \quad X_h' = [1 \quad X_h]_{1 \times 2} \quad (4.32)$$

Matris gösterimiyle tahmin edilen değer aşağıdaki gibidir:

$$\hat{Y}_h = X_h' b \quad (4.33)$$

Çünkü:

$$X_h' b = [1 \quad X_h] \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = [b_0 + b_1 X_h] = [\hat{Y}_h] = \hat{Y}_h$$

olmaktadır. Matris gösterimiyle $\hat{Y}_h$ 'ın varyansı ise:

$$\sigma^2\{\hat{Y}_h\} = \sigma^2 X_h' (X'X)^{-1} X_h \quad (4.34)$$

$\hat{Y}_h$ 'in varyansı, $\sigma^2\{b\}$ 'nin bir fonksiyonu olarak da gösterilmektedir. $\sigma^2\{b\} = \sigma^2_{2 \times 2} (X'X)^{-1}$

kullanılarak tahmin edilen regresyon katsayılarının varyans-kovaryans matrisi:

$$\sigma^2\{\hat{Y}_h\} = X_h' \sigma^2\{b\} X_h \quad (4.35)$$

olmaktadır. Matris gösterimiyle $\hat{Y}_h$ 'in tahmin edilen varyansı:

$$s^2\{\hat{Y}_h\} = MSE \left(X_h' (X'X)^{-1} X_h \right) \quad (4.36)$$

şeklinde elde edilmektedir.

5. ÇOKLU REGRESYON ANALİZİ

Çoklu regresyon analizinde birden fazla açıklayıcı değişken bulunmaktadır. Bunlar modele kareli ve eğrisel olarak da eklenebilmektedirler. Çoklu regresyon ile yalın doğrusal regresyon için matris gösterimleri birbirinin aynıdır.

5.1 Matris Gösterimiyle Genel Doğrusal Regresyon Modeli

Matris gösteriminde, genel doğrusal regresyon modeli ile yalın doğrusal regresyon modeli için elde edilen sonuçlar birbirinin aynıdır. Sadece serbestlik dereceleri, X değişkenlerinin sayısı ile ilgili diğer sabitler ve bazı matrislerin boyutları farklıdır.

Genel doğrusal regresyon modeli aşağıdaki gibi ifade edilmektedir:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad (5.1)$$

Bunu matris gösterimiyle ifade edebilmek için aşağıdaki matrisler tanımlanmalıdır:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1} \quad (5.2)$$

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,p-1} \end{bmatrix}_{n \times p} \quad (5.3)$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}_{p \times 1} \quad (5.4)$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1} \quad (5.5)$$

Yukarıda da görüldüğü gibi Y ve ε vektörleri yalın doğrusal regresyon modeli için de aynıdır. β vektörü toplamsal regresyon parametrelerini içermektedir. X matrisi, regresyon

modeli içinde $p-1$ tane X değişkeninin her biri için n gözlemin bir sütunu olarak 1'lerin bulunduğu sütunu içermektedir. X matrisi içindeki her X_{ik} elementi için satır indisi, bunun kaçınıcı deneme ya da durum olduğunu ve sütun indisi de, hangi X değişkeni olduğunu belirtmektedir.

Matris gösterimiyle, (5.1) genel doğrusal regresyon modeli aşağıdaki gibi ifade edilmektedir:

$$\underset{n \times 1}{Y} = \underset{n \times p}{X} \underset{n \times p}{\beta} + \underset{n \times 1}{\varepsilon} \quad (5.6)$$

Burada;

Y , bağımlı değişkenin vektörü,

β , parametrelerin vektörü,

X , sabitlerin matrisi ve

ε , bağımsız normal rastlantısal değişkenlerin vektörüdür.

$E\{\varepsilon\} = 0$ ve varyans-kovaryans matrisi:

$$\sigma^2\{\varepsilon\}_{n \times n} = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2 I$$

şeklindedir. Sonuç olarak, rastlantısal vektör Y 'nin beklenen değeri:

$$\underset{n \times 1}{E\{Y\}} = X\beta \quad (5.7)$$

ve Y 'nin varyans-kovaryans matrisi, ε 'nin varyans-kovaryans matrisinin aynısı olmaktadır:

$$\sigma^2\{Y\}_{n \times n} = \sigma^2 I \quad (5.8)$$

5.2 Regresyon Katsayılarının Tahmini

$Q = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2$ en küçük kareler kriteri, (5.1) genel doğrusal regresyon modeli için

aşağıdaki gibi genelleştirilmektedir:

$$Q = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{i1} - \dots - \beta_{p-1} X_{i,p-1})^2 \quad (5.9)$$

Q 'nun minimum yapılmasıyla elde edilen $\beta_0, \beta_1, \dots, \beta_{p-1}$ değerleri en küçük kareler tahmincileridir. Regresyon katsayılarının en küçük kareler tahmincileri $b_0, b_1, \dots, b_{p-1}$ 'in vektörü $\mathbf{b}$ ile gösterilmektedir:

$$\mathbf{b}_{p \times 1} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix} \quad (5.10)$$

(5.6) genel doğrusal regresyon modeli için en küçük kareler normal denklemleri:

$$(\mathbf{X}' \cdot \mathbf{X}) \cdot \mathbf{b} = (\mathbf{X}' \cdot \mathbf{Y}) \quad (5.11)$$

ve en küçük kareler tahmincileri aşağıdaki gibi olmaktadır:

$$\mathbf{b}_{p \times 1} = (\mathbf{X}' \cdot \mathbf{X})_{p \times p}^{-1} \cdot (\mathbf{X}' \cdot \mathbf{Y})_{p \times 1} \quad (5.12)$$

(5.6) normal dağılan hata terimli regresyon modeli için, maksimum olabilirlik yöntemi, en küçük kareler yöntemiyle bulunan tahmincilerle aynı sonucu vermektedir.

(2.25) olabilirlik fonksiyonu çoklu regresyon için aşağıdaki gibi genelleştirilmektedir:

$$L(\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \cdot \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{i1} - \dots - \beta_{p-1} X_{i,p-1})^2 \right] \quad (5.13)$$

$\beta_0, \beta_1, \dots, \beta_{p-1}$ 'e göre bu olabilirlik fonksiyonu maksimize edilirse, (5.12) denklemindeki tahmincilere ulaşılmaktadır. Bunlar, hem en küçük kareler hem de maksimum olabilirlik tahmincileridir ve yansız, tutarlı, yeterli, minimum varyansa sahiptirler.

5.3 Tahmin Edilen Değerler ve Kalıntılar

Tahmin edilen değerler $\hat{Y}_i$ 'lerin vektörü $\hat{\mathbf{Y}}$ ile ve kalıntılar $e_i = Y_i - \hat{Y}_i$ 'lerin vektörü $\mathbf{e}$ ile gösterilmektedir:

$$\hat{Y}_{n \times 1} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} \quad (5.14)$$

$$e_{n \times 1} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \quad (5.15)$$

Tahmin edilen değerler aşağıdaki gibi gösterilmektedir:

$$\hat{Y}_{n \times 1} = Xb \quad (5.16)$$

ve kalıntı terimleri aşağıdaki gibi olmaktadır:

$$e_{n \times 1} = Y - \hat{Y} = Y - Xb \quad (5.17)$$

Tahmin edilen değerlerin vektörü $\hat{Y}$, projeksiyon matrisi H ile aşağıdaki gibi gösterilmektedir:

$$\hat{Y}_{n \times 1} = HY \quad (5.18)$$

Burada H aşağıdaki gibi alınmıştır:

$$H_{n \times n} = X(X'.X)^{-1}X' \quad (5.19)$$

Benzer şekilde, kalıntıların vektörü de aşağıdaki gibi gösterilmektedir:

$$e_{n \times 1} = (I - H)Y \quad (5.20)$$

Kalıntıların varyans-kovaryans matrisi aşağıdaki gibi ifade edilmektedir:

$$\sigma^2\{e\}_{n \times n} = \sigma^2(I - H) \quad (5.21)$$

Tahmini ise:

$$s^2\{e\}_{n \times n} = MSE(I - H) \quad (5.22)$$

olmaktadır.

5.4 Varyans Analizi Sonuçları

5.4.1 Kareler Toplamı ve Kareler Ortalaması

Matris gösterimiyle varyans analizi için kareler toplamı:

$$SSTO = Y'Y - \left(\frac{1}{n}\right)Y'JY = Y'\left[I - \left(\frac{1}{n}\right)J\right]Y \quad (5.23)$$

$$SSE = e'e = (Y - Xb)'(Y - Xb) = Y'Y - b'X'Y = Y'(I - H)Y \quad (5.24)$$

$$SSR = b'X'Y - \left(\frac{1}{n}\right)Y'JY = Y'\left[H - \left(\frac{1}{n}\right)J\right]Y \quad (5.25)$$

şeklinde ifade edilmektedir. Burada J , $n \times n$ boyutunda 1'lerin oluşturduğu bir matris ve H 'de projeksiyon matrisidir.

SSTO, SSE ve SSR'ın serbestlik dereceleri sırasıyla $n-1$, $n-p$ ve $p-1$ 'dir.

Kareler ortalamaları MSR ve MSE aşağıdaki gibidir:

$$MSR = \frac{SSR}{p-1} \quad (5.26)$$

$$MSE = \frac{SSE}{n-p} \quad (5.27)$$

Yalın doğrusal regresyon modelinde olduğu gibi MSE'nin beklenen değeri σ^2 'dir. MSR'ın beklenen değeri ise, örneğin $p-1=2$ için şu şekilde olmaktadır:

$$E\{MSR\} = \sigma^2 + \frac{1}{2} \left[\beta_1^2 \sum (X_{i1} - \bar{X}_1)^2 + \beta_2^2 \sum (X_{i2} - \bar{X}_2)^2 + 2\beta_1\beta_2 \sum (X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2) \right]$$

Eğer β_1 ve β_2 'nin ikisi de sıfıra eşit olursa, $E\{MSR\} = \sigma^2$ olacaktır. Diğer durumlarda ise, $E\{MSR\} > \sigma^2$ olmaktadır.

Çizelge 5.1 $Y = X \beta + \varepsilon$ Genel Doğrusal Regresyon Modeli için ANOVA Tablosu

Değişim Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması
Regresyon	$SSR = b'.X'.Y - \left(\frac{1}{n}\right).Y'.J.Y$	p-1	$MSR = \frac{SSR}{p-1}$
Hata	$SSE = Y'.Y - b'.X'.Y$	n-p	$MSE = \frac{SSE}{n-p}$
Toplam	$SSTO = Y'.Y - \left(\frac{1}{n}\right).Y'.J.Y$	n-1	

5.4.2 Regresyon ilişkisi için F Testi

Bağımlı değişken Y ile $X_1, \dots, X_{p-1}$ değişkenleri arasında bir regresyon ilişkisi olup olmadığı test edilmek istenirse, aşağıdaki alternatifler arasında karar verilmeye çalışılır:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0 \quad (5.28)$$

$$H_1 : \beta_k \text{ 'lerden } (k=1, \dots, p-1) \text{ en az biri sıfırdan farklıdır.} \quad (5.29)$$

Test istatistiği aşağıdaki gibidir:

$$F^* = \frac{MSR}{MSE} \quad (5.30)$$

α anlamlılık seviyesinde karar kuralı:

$$F^* \leq F(1-\alpha; p-1, n-p) \text{ ise, } H_0 \text{ reddedilmez.} \quad (5.31)$$

$$F^* > F(1-\alpha; p-1, n-p) \text{ ise, } H_0 \text{ reddedilir.} \quad (5.32)$$

5.4.3 Çoklu Belirginlik Katsayısı

Çoklu belirginlik katsayısı R^2 ile gösterilmekte ve aşağıdaki gibi tanımlanmaktadır:

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO} \quad (5.33)$$

R^2 , $X_1, \dots, X_{p-1}$ değişkenlerinin kullanımı ile ilişkili Y içindeki toplam değişimi orantılı bir şekilde ölçmektedir. R^2 'nin sınırları aşağıdaki gibidir :

$$0 \leq R^2 \leq 1 \quad (5.34)$$

Regresyon modeline daha çok X değişkeni eklenirse; bu R^2 'nin artmasına sebep olmaktadır. Çünkü SSE daha çok X değişkeni ile asla daha büyük olmaz ve SSTO'da daima verilen bağımlı değişken seti için aynıdır. R^2 , açıklayıcı değişken sayısının arttırılmasıyla genellikle daha büyük yapılabileceği için, model içindeki X değişkenlerinin sayısının artmasından kaynaklanan yanlış anlamaları engellemek için bazen yeniden düzenlenmiş bir ölçü kullanılmaktadır. Buna ayarlanmış çoklu belirginlik katsayısı denilmekte ve R_a^2 ile gösterilmektedir. R_a^2 aşağıdaki gibi hesaplanmaktadır:

$$R_a^2 = 1 - \frac{\frac{SSE}{n-p}}{\frac{SSTO}{n-1}} = 1 - \left(\frac{n-1}{n-p} \right) \frac{SSE}{SSTO} \quad (5.35)$$

5.4.4 Çoklu Korelasyon Katsayısı

Çoklu korelasyon katsayısı R , R^2 'nin pozitif kareköküdür:

$$R = \sqrt{R^2} \quad (5.36)$$

(5.6) regresyon modelinde sadece bir X değişkeni olduğu zaman, yani $p-1=1$ olduğunda; çoklu korelasyon katsayısı R , yalın korelasyon için $r = \pm\sqrt{R^2}$ 'ye göre korelasyon katsayısı r 'ye eşit olmaktadır.

5.5 Regresyon Parametreleriyle İlgili Elde Edilen Sonuçlar

b 'nin en küçük kareler ve maksimum olabilirlik tahminçileri yansızdır:

$$E\{b\} = \beta \quad (5.37)$$

Varyans-kovaryans matrisi $\sigma^2\{b\}$:

$$\sigma^2\{b\}_{p \times p} = \begin{bmatrix} \sigma^2\{b_0\} & \sigma\{b_0, b_1\} & \cdots & \sigma\{b_0, b_{p-1}\} \\ \sigma\{b_1, b_0\} & \sigma^2\{b_1\} & \cdots & \sigma\{b_1, b_{p-1}\} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma\{b_{p-1}, b_0\} & \sigma\{b_{p-1}, b_1\} & \cdots & \sigma^2\{b_{p-1}\} \end{bmatrix} \quad (5.38)$$

Matris gösterimiyle şöyle ifade edilmektedir:

$$\sigma^2\{b\} = \sigma^2 (X'X)^{-1} \quad (5.39)$$

$p \times p$

Varyans-kovaryans matrisinin tahmini $s^2\{b\}$:

$$s^2\{b\} = \begin{bmatrix} s^2\{b_0\} & s\{b_0, b_1\} & \cdots & s\{b_0, b_{p-1}\} \\ s\{b_1, b_0\} & s^2\{b_1\} & \cdots & s\{b_1, b_{p-1}\} \\ \vdots & \vdots & \ddots & \vdots \\ s\{b_{p-1}, b_0\} & s\{b_{p-1}, b_1\} & \cdots & s^2\{b_{p-1}\} \end{bmatrix} \quad (5.40)$$

$p \times p$

Matris gösterimiyle şu şekilde ifade edilmektedir:

$$s^2\{b\} = MSE(X'X)^{-1} \quad (5.41)$$

$p \times p$

5.5.1 β_k 'nın Tahmin Aralığı

(5.6) normal dağılım hata terimli regresyon modeli için:

$$\frac{b_k - \beta_k}{s\{b_k\}} \sim t(n-p) \quad k = 0, 1, \dots, p-1 \quad (5.42)$$

olmaktadır. Bu yüzden, β_k için $1 - \alpha$ 'lık güven sınırları:

$$b_k \pm t(1 - \alpha/2; n-p) s\{b_k\} \quad (5.43)$$

şeklinde elde edilmektedir.

5.5.2 β_k İçin Testler

Hipotezler:

$$H_0 : \beta_k = 0 \quad (5.44)$$

$$H_1 : \beta_k \neq 0 \quad (5.45)$$

Test istatistiği:

$$t^* = \frac{b_k}{s\{b_k\}} \quad (5.46)$$

Karar kuralı:

Eğer $|t^*| \leq t(1 - \alpha/2; n - p)$ ise, H_0 reddedilmez. (5.47)

Tersi durumda, H_0 reddedilir.

Yalın doğrusal regresyonda olduğu gibi, ayrıca çoklu regresyon modelleri içinde $\beta_k = 0$ 'dır ya da değildir durumunun test edilmesinde F testi kullanılmaktadır.

5.6 Bağımlı Değişkenin Beklenen Değerinin Tahmini

5.6.1 $E\{Y_h\}$ 'nin Tahmin Aralığı

$X_1, \dots, X_{p-1}$ 'nin verilen değerleri için ki bunlar $X_{h1}, \dots, X_{h,p-1}$ ile, bağımlı değişkenin beklenen değeri de $E\{Y_h\}$ ile gösterilmektedir. X_h vektörü aşağıdaki gibi tanımlanmaktadır:

$$X_h = \begin{bmatrix} 1 \\ X_{h1} \\ \vdots \\ X_{h,p-1} \end{bmatrix} \quad (5.48)$$

Böylece bağımlı değişkenin beklenen değeri aşağıdaki gibi elde edilmektedir:

$$E\{Y_h\} = X_h' \cdot \beta \quad (5.49)$$

X_h 'a göre tahmin edilen bağımlı değişken, $\hat{Y}_h$ ile gösterilmektedir:

$$\hat{Y}_h = X_h' \cdot b \quad (5.50)$$

Bu tahminci yansızdır:

$$E\{\hat{Y}_h\} = X_h' \cdot \beta = E\{Y_h\} \quad (5.51)$$

ve varyansı:

$$\sigma^2\{\hat{Y}_h\} = \sigma^2 \cdot X_h' \cdot (X' \cdot X)^{-1} \cdot X_h \quad (5.52)$$

olmaktadır. Bu varyans tahmin edilen regresyon katsayılarının varyans-kovaryans matrisinin bir fonksiyonu olarak ifade edilmektedir:

$$\sigma^2\{\hat{Y}_h\} = X_h' \cdot \sigma^2\{b\} \cdot X_h \quad (5.53)$$

Burada varyans $\sigma^2\{\hat{Y}_h\}$, regresyon katsayılarının varyansının, yani $\sigma^2\{b_k\}$ 'nin bir fonksiyonudur ve regresyon katsayılarının çiftleri arasındaki kovaryanslar $\sigma\{b_k, b_{k'}\}$ yalın doğrusal regresyonda olduğu gibidir. Varyansın tahmini $s^2\{\hat{Y}_h\}$:

$$s^2\{\hat{Y}_h\} = MSE\left(X_h' (X'X)^{-1} X_h\right) = X_h' .s^2\{b\}.X_h \quad (5.54)$$

olmaktadır. $E\{Y_h\}$ için $1 - \alpha$ 'lık güven aralığı aşağıdaki gibidir:

$$\hat{Y}_h \pm t(1 - \alpha/2; n - p) s\{\hat{Y}_h\} \quad (5.55)$$

6. DOĞRUSAL OLMAYAN REGRESYON

6.1.1 Doğrusal Olmayan Regresyon Modelleri

Doğrusal olmayan regresyon modelleri, doğrusal regresyon modelleri için, $Y_i = f(X_i, \beta) + \varepsilon_i$ denklemindekiyle aynı şekilde gösterilmektedir:

$$Y_i = f(X_i, \gamma) + \varepsilon_i \quad (6.1)$$

Gözlem Y_i , verilen doğrusal olmayan bağımlı değişkenin fonksiyonu $f(X, \gamma)$ 'den oluşan bağımlı değişkenin beklenen değeri $f(X_i, \gamma)$ ve hata terimi ε_i 'nin toplamıdır. Hata terimlerinin genellikle doğrusal regresyon modellerinde olduğu gibi sıfır ortalamaya, sabit varyansa sahip oldukları ve korelasyonsuz oldukları varsayılmaktadır. Normal dağılan hata terimli bir modelin, hata terimlerinin, sık sık sabit varyans ile bağımsız normal rastlantısal değişkenler olduğu varsayılarak kullanılmaktadır.

Şimdi uygulamada çok kullanılan iki doğrusal olmayan regresyon modeline örnek verilecektir.

6.1.1.1 Üstel Regresyon Modelleri

Doğrusal olmayan regresyon modellerinde, çok kullanılan modellerden biri üstel regresyon modelidir. Sadece bir tek açıklayıcı değişken olduğunda, bu regresyon modeli normal hata terimleri ile şu şekilde gösterilmektedir:

$$Y_i = \gamma_0 \exp(\gamma_1 X_i) + \varepsilon_i \quad (6.2)$$

Burada;

γ_0 ve γ_1 , parametreler,

X_i , bilinen sabitler,

ε_i , bağımsız $N(0, \sigma^2)$ dağılmaktadır.

Bu model için bağımlı değişkenin fonksiyonu:

$$f(X, \gamma) = \gamma_0 \exp(\gamma_1 X) \quad (6.3)$$

olmaktadır. Yukarıda da görüldüğü gibi bu model, γ_0 ve γ_1 parametrelerinde doğrusal

değildir.

Normal dağılan hata terimleri ile tek açıklayıcı değişkenli daha genel bir doğrusal olmayan regresyon modeli ise şöyle gösterilmektedir:

$$Y_i = \gamma_0 + \gamma_1 \exp(\gamma_2 X_i) + \varepsilon_i \quad (6.4)$$

Burada hata terimleri, sabit varyans σ^2 ile bağımsız normal dağılmaktadır. Bu regresyon modeli için bağımlı değişkenin fonksiyonu aşağıdaki gibidir:

$$f(X, \gamma) = \gamma_0 + \gamma_1 \exp(\gamma_2 X) \quad (6.5)$$

(6.4) üstel regresyon modeli genellikle; verilen bir X zamanındaki büyüme oranının, zamanın artmasıyla meydana gelen büyüme miktarı ile γ_0 'la gösterilen maksimum büyüme değerinin oranı olduğu büyüme ile ilgili çalışmalarda kullanılmaktadır. Bu regresyon modelinin başka bir kullanımı ise geçen zamanla (X) bir maddenin toplanması (Y) arasındaki ilişkidir. Şekil 6.1.a parametre değerleri $\gamma_0 = 100$, $\gamma_1 = -50$ ve $\gamma_2 = -2$ için (6.5) bağımlı değişkenin fonksiyonunu göstermektedir.

6.1.1.2 Lojistik Regresyon Modelleri

Başka bir önemli doğrusal olmayan regresyon modeli de lojistik regresyon modelidir. Tek açıklayıcı değişken ve hata terimi ile bu model aşağıdaki gibidir:

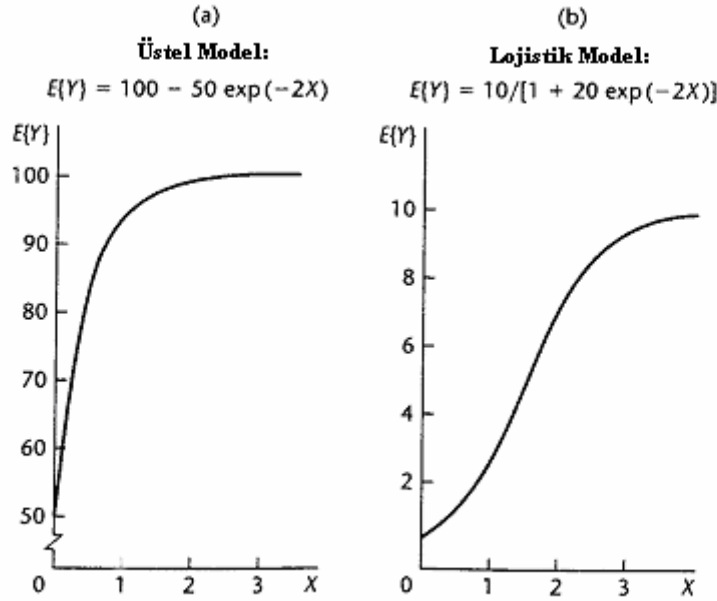
$$Y_i = \frac{\gamma_0}{1 + \gamma_1 \exp(\gamma_2 X_i)} + \varepsilon_i \quad (6.6)$$

Burada ε_i hata terimleri sabit varyans σ^2 ile bağımsız normal dağılmaktadır. Bağımlı değişkenin fonksiyonu aşağıdaki gibidir:

$$f(X, \gamma) = \frac{\gamma_0}{1 + \gamma_1 \exp(\gamma_2 X_i)} \quad (6.7)$$

Yine bağımlı değişkenin fonksiyonu γ_0 , γ_1 ve γ_2 parametrelerinde doğrusal değildir. Bu lojistik regresyon modeli, topluluk çalışmalarındaki ilişki için kullanılmaktadır. Örneğin, zaman (X) içindeki türlerin sayısı (Y). Şekil 6.1.b parametre değerleri $\gamma_0 = 10$, $\gamma_1 = 20$ ve $\gamma_2 = -2$ için (6.7) lojistik bağımlı değişkenin fonksiyonunu göstermektedir. Burada parametre $\gamma_0 = 10$ maksimum büyüme değerini göstermektedir.

Lojistik regresyon modeli (6.6), ayrıca bağımlı değişken kalitatif olduğu zaman da kullanılmaktadır.



Şekil 6.1 Üstel ve lojistik bağımlı değişken fonksiyonlarının çizimleri

6.1.1.3 Doğrusal Olmayan Regresyon Modellerinin Genel Biçimi

Bu modeller doğrusal regresyon modellerinin genel şekliyle benzerlik göstermektedir. Her Y_i gözlemi, verilen doğrusal olmayan bağımlı değişken fonksiyonu temel alınarak oluşturulan $f(X_i, \gamma)$ bağımlı değişkenin beklenen değeri ve rastlantısal hata terimi ε_i 'nin toplamı olarak kabul edilmektedir. Ayrıca, ε_i hata terimlerinin, sık sık sabit varyansla dağılan bağımsız normal rastlantısal değişkenler oldukları varsayılmaktadır.

Doğrusal olmayan regresyon modellerinin önemli bir farkı, modeldeki X değişkenlerinin sayısı ile regresyon parametrelerinin sayısının kesinlikle ilişkili olmak zorunda olmamasıdır. Doğrusal regresyon modellerinde, eğer modelde $p-1$ tane X değişkeni varsa, modelde p tane regresyon katsayısı vardır. (6.4) şeklindeki üstel regresyon modelinde, bir X değişkeni vardır, ama üç tane regresyon katsayısı bulunmaktadır. Aynı durum (6.6)'daki lojistik regresyon modeli içinde geçerlidir. Bu yüzden, doğrusal olmayan X değişkenlerinin sayısı doğrusal olmayan regresyon modeli için q ile gösterilecektir, fakat bağımlı değişken fonksiyonundaki regresyon parametrelerinin sayısı p ile gösterilmeye devam edilecektir. Örneğin, (6.2) üstel regresyon modeli, $p=2$ regresyon parametresi ve $q=1$ X değişkeni bulundurmaktadır.

Ayrıca, X değişkenleri üzerindeki gözlemlerin vektörü X_i ilk eleman 1 olmadan gösterilecektir. Bu yüzden doğrusal olmayan bir regresyon modelinin genel şekli aşağıdaki gibi ifade edilmektedir:

$$Y_i = f(X_i, \gamma) + \varepsilon_i \quad (6.8)$$

Burada;

$$X_i = \begin{bmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{iq} \end{bmatrix} \quad (6.9)$$

$$\gamma = \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{p-1} \end{bmatrix} \quad (6.10)$$

olarak gösterilmektedir.

Açıklama

Dönüşümle doğrusallaştırılabilen doğrusal olmayan bağımlı değişken fonksiyonlarına, bazen özünde doğrusal bağımlı değişken fonksiyonları denilmektedir. Üstel bağımlı değişken fonksiyonu ve lojistik bağımlı değişken fonksiyonu özünde doğrusal bağımlı değişken fonksiyonlarıdır. Örneğin, üstel bağımlı değişken fonksiyonu:

$$f(X, \gamma) = \gamma_0 [\exp(\gamma_1 X)]$$

logaritmik dönüşümle doğrusallaştırılmaktadır:

$$\log_e f(X, \gamma) = \log_e \gamma_0 + \gamma_1 X$$

Bu dönüştürülmüş bağımlı değişken fonksiyonu doğrusal model olarak aşağıdaki gibi gösterilmektedir:

$$g(X, \gamma) = \beta_0 + \beta_1 X$$

Burada $g(X, \gamma) = \log_e f(X, \gamma)$, $\beta_0 = \log_e \gamma_0$ ve $\beta_1 = \gamma_1$ 'dir.

Burada doğrusal olmayan bağımlı değişken fonksiyonunun sadece özünde doğrusal olması,

doğrusal regresyonun kesinlikle uygun olduğunu göstermez. Bunun sebebi, bağımlı değişken fonksiyonunu doğrusallaştıran dönüşümün modeldeki hata terimini etkileyecek olmasıdır. Örneğin, sabit varyansa sahip normal hata terimleri ile aşağıdaki üstel regresyon modelinin uygun olduğu varsayalım:

$$Y_i = \gamma_0 \exp(\gamma_1 X_i) + \varepsilon_i$$

Bağımlı değişkenin fonksiyonunu doğrusallaştırmak için yapılan Y 'nin logaritmik dönüşümü, ε_i normal hata terimini etkileyecektir. Böylece doğrusallaştırılmış model içindeki hata terimi artık normal dağılmayacaktır. Bu nedenle, herhangi bir doğrusal olmayan regresyon modeli doğrusallaştırılırken uygunluk önemlidir; doğrusal olmayan regresyon modeli doğrusallaştırılmış şekline göre tercih edilmektedir.

6.1.2 Regresyon Parametrelerinin Tahmini

Doğrusal regresyon modellerinde olduğu gibi, doğrusal olmayan bir regresyon modelinin parametrelerinin tahmini de genellikle en küçük kareler yöntemi ya da maksimum olabilirlik yöntemi ile gerçekleştirilmektedir. Ayrıca doğrusal regresyonda olduğu gibi, (6.8) doğrusal olmayan regresyon modelindeki hata terimleri sabit varyansla bağımsız normal dağıldıklarında, bu yöntemlerin ikisi de aynı parametre tahminlerini vermektedir.

Doğrusal regresyondan farklı olarak, doğrusal olmayan regresyon modellerine göre en küçük kareler ve maksimum olabilirlik tahmincileri için analitik ifadeler bulmak genellikle mümkün değildir. Onun yerine, bu iki tahmin yöntemi kullanılarak nümerik arama işlemleri uygulanmalıdır. Bu işlemler yoğun hesaplamalar gerektirmektedir. Doğrusal olmayan regresyon modellerinin analizi bu yüzden genellikle bilgisayar programları kullanılarak gerçekleştirilmektedir.

6.2 Doğrusal Olmayan Regresyonda En Küçük Kareler Tahmini

Yalın doğrusal regresyon modeli için en küçük kareler yönteminde Q toplamını minimum yapan değerler alınmaktadır:

$$Q = \sum_{i=1}^n [Y_i - (\beta_0 + \beta_1 X_i)]^2 \quad (6.11)$$

Verilen (X_i, Y_i) örnek gözlemleri için Q 'yu minimum yapan bu β_0 ve β_1 değerleri, en küçük kareler tahminleridir ve b_0 , b_1 ile gösterilmektedirler.

En küçük kareler tahminlerini bulmanın ikinci yolu, en küçük kareler normal denklemlerinin ortalamalarıdır. Burada, en küçük kareler normal denklemleri analitik olarak Q 'nun β_0 ve β_1 'e göre türevi alındıktan sonra türevlerin sıfıra eşitlenmesiyle bulunmaktadır. Normal denklemlerin çözümü en küçük kareler tahminlerini vermektedir.

Doğrusal regresyon için en küçük kareler tahmini, doğrusal olmayan regresyon modellerinde de benzer şekilde kullanılmaktadır. Doğrusal olmayan regresyon için en küçük kareler kriteri aşağıdaki gibidir:

$$Q = \sum_{i=1}^n [Y_i - f(X_i, \gamma)]^2 \quad (6.12)$$

Burada $f(X_i, \gamma)$, doğrusal olmayan bağımlı değişken fonksiyonu $f(X, \gamma)$ 'ya göre i .durum için bağımlı değişkenin beklenen değeridir. En küçük kareler tahminlerini bulmak için, (6.12) denklemindeki en küçük kareler toplamı Q doğrusal olmayan regresyon parametreleri $\gamma_0, \gamma_1, \dots, \gamma_{p-1}$ 'e göre minimum olmaktadır. En küçük kareler tahminlerini bulmak için kullanılan bu aynı iki yöntem nümerik arama ve normal denklemler, doğrusal olmayan regresyonda kullanılmaktadır. Doğrusal regresyondan farkı ise, normal denklemlerin çözümünün genellikle iteratif nümerik arama yöntemleriyle gerçekleştirilmesidir.

Örnek: Bağımlı değişkenin fonksiyonunun aşağıdaki gibi olduğu varsayalım:

$$f(X, \gamma) = \gamma_0 \exp(\gamma_1 X)$$

Bu yüzden, burada en küçük kareler toplamı Q :

$$Q = \sum_{i=1}^n [Y_i - \gamma_0 \exp(\gamma_1 X_i)]^2$$

Burada, ε_i hata terimleri sabit varyans ile bağımsız normal dağıldıkları zaman olabilirlik fonksiyonu düşünülürse, maksimum olabilirlik yöntemi aynı kritere ulaşmaktadır:

$$L(\gamma, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \cdot \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n [Y_i - \gamma_0 \exp(\gamma_1 X_i)]^2\right]$$

Doğrusal regresyon için olduğu gibi, regresyon parametreleri γ_0 ve γ_1 'e göre bu olabilirlik fonksiyonunun maksimum yapılması üs kısmında yer alan toplamın minimum yapılmasıyla sağlanmaktadır. Bu yüzden, maksimum olabilirlik tahminleri burada en küçük kareler

tahminleriyle aynıdır.

6.2.1 Normal Denklemlerin Çözümü

Doğrusal olmayan regresyon modeline göre normal denklemleri bulmak için:

$$Y_i = f(X_i, \gamma) + \varepsilon_i$$

kullanılmaktadır. En küçük kareler toplamı Q , $\gamma_0, \gamma_1, \dots, \gamma_{p-1}$ 'e göre minimum yapılmaktadır:

$$Q = \sum_{i=1}^n [Y_i - f(X_i, \gamma)]^2$$

γ_k 'ya göre Q 'nun kısmi türevi:

$$\frac{\partial Q}{\partial \gamma_k} = \sum_{i=1}^n -2[Y_i - f(X_i, \gamma)] \cdot \left[\frac{\partial f(X_i, \gamma)}{\partial \gamma_k} \right] \quad (6.13)$$

p tane kısmi türevin her biri sıfıra eşitlendiğinde ve parametreler yani γ_k 'lar yerine en küçük kareler tahminleri g_k 'lar yerleştirildiğinde, denklemler düzenlendikten sonra p tane normal denklem bulunmaktadır:

$$\sum_{i=1}^n Y_i \left[\frac{\partial f(X_i, \gamma)}{\partial \gamma_k} \right]_{\gamma=g} - \sum_{i=1}^n f(X_i, g) \left[\frac{\partial f(X_i, \gamma)}{\partial \gamma_k} \right]_{\gamma=g} = 0 \quad k = 0, 1, \dots, p-1 \quad (6.14)$$

Burada g , en küçük kareler tahminleri g_k 'ların vektörüdür:

$$g = \begin{bmatrix} g_0 \\ g_1 \\ \vdots \\ g_{p-1} \end{bmatrix}_{p \times 1} \quad (6.15)$$

Doğrusal olmayan regresyon modellerinde normal denklemler parametre tahminleri g_k 'lar için doğrusal değildir ve genellikle en basit durumlarda bile çözülmesi zordur.

Örnek: Daha önce verilen örnekte i . durum için bağımlı değişkenin beklenen değeri:

$$f(X_i, \gamma) = \gamma_0 \exp(\gamma_1 X_i) \quad (6.16)$$

Bu yüzden, $f(X_i, \gamma)$ 'nın kısmi türevleri:

$$\frac{\partial f(X_i, \gamma)}{\partial \gamma_0} = \exp(\gamma_1 X_i) \quad (6.17)$$

$$\frac{\partial f(X_i, \gamma)}{\partial \gamma_1} = \gamma_0 X_i \exp(\gamma_1 X_i) \quad (6.18)$$

(6.16), (6.17) ve (6.18) içinde γ_0 ve γ_1 yerine sırasıyla en küçük kareler tahminleri g_0 ve g_1 yazılırsa, normal denklemler aşağıdaki gibi olmaktadır:

$$\sum Y_i \exp(g_1 X_i) - \sum g_0 \exp(g_1 X_i) \exp(g_1 X_i) = 0$$

$$\sum Y_i g_0 X_i \exp(g_1 X_i) - \sum g_0 \exp(g_1 X_i) g_0 X_i \exp(g_1 X_i) = 0$$

Gerekli sadeleştirmeler yapıldıktan sonra, normal denklemler:

$$\sum Y_i \exp(g_1 X_i) - g_0 \sum \exp(2g_1 X_i) = 0$$

$$\sum Y_i X_i \exp(g_1 X_i) - g_0 \sum X_i \exp(2g_1 X_i) = 0$$

şeklinde elde edilmektedir. Bu normal denklemler g_0 ve g_1 'e göre doğrusal değildir.

6.2.2 Doğrudan Nümerik Arama (Gauss-Newton Metodu)

Birçok doğrusal olmayan regresyon probleminde, doğrudan nümerik arama yöntemleri kullanılmaktadır. Nümerik yöntemlerle çözüm bu denklemler için iteratif olarak hesaplanmaktadır. Bu yöntemlerden biri Gauss-Newton metodudur.

Gauss-Newton yönteminde, Taylor serisi açılımı kullanılarak, doğrusal terimlerle doğrusal olmayan regresyon modeline yaklaşılr ve sıradan en küçük kareler kullanılarak parametre tahminleri elde edilir. Bu adımların iterasyonu genellikle doğrusal olmayan regresyon probleminin çözümüne ulaşılmasını sağlamaktadır.

Gauss-Newton metodu bir başlangıç değeriyle ya da $\gamma_0, \gamma_1, \dots, \gamma_{p-1}$ regresyon parametreleri için alınacak olan bir başlangıç değeriyle başlar. Bunlar $g_0^{(0)}, g_1^{(0)}, \dots, g_{p-1}^{(0)}$ ile gösterilir. Burada parantez içindeki üstsimge iterasyon sayısını göstermektedir. $g_k^{(0)}$ başlangıç değeri, daha önceki ya da ilişkili çalışmalardan, teorik yaklaşımlar ya da Q toplamı ile bulunan parametre değerleri için yapılan ön hazırlık araştırmasından elde edilmektedir.

Parametreler için başlangıç değerleri bir kez bulunduktan sonra, başlangıç değerleri $g_k^{(0)}$ 'lar çevresindeki Taylor serisi açılımı içinde n durum için doğrusal terimlerle $f(X_i, \gamma)$ bağımlı değişkenin beklenen değerine yaklaşılr. i .durum için aşağıdaki sonuç bulunur:

$$f(X_i, \gamma) \approx f(X_i, g^{(0)}) + \sum_{k=0}^{p-1} \left[\frac{\partial f(X_i, \gamma)}{\partial \gamma_k} \right]_{\gamma=g^{(0)}} (\gamma_k - g_k^{(0)}) \quad (6.19)$$

Burada, $g^{(0)}$ aşağıdaki gibidir:

$$g_{p \times 1}^{(0)} = \begin{bmatrix} g_0^{(0)} \\ g_1^{(0)} \\ \vdots \\ g_{p-1}^{(0)} \end{bmatrix} \quad (6.20)$$

$g^{(0)}$, parametre başlangıç değerlerinin vektörüdür. (6.19)'da bulunan parantez içindeki terimler, daha önce normal denklemlerde karşılaşılan regresyon fonksiyonunun kısmi türevleriyle aynıdır. Fakat burada onlar $k = 0, 1, \dots, p-1$ için $\gamma_k = g_k^{(0)}$ 'da hesaplanmaktadır.

Gösterim aşağıdaki şekilde kolaylaştırılırsa:

$$f_i^{(0)} = f(X_i, g^{(0)}) \quad (6.21)$$

$$\beta_k^{(0)} = \gamma_k - g_k^{(0)} \quad (6.22)$$

$$D_{ik}^{(0)} = \left[\frac{\partial f(X_i, \gamma)}{\partial \gamma_k} \right]_{\gamma=g^{(0)}} \quad (6.23)$$

i .duruma göre bağımlı değişkenin beklenen değeri için, Taylor yaklaşımı (6.19), bu gösterimle şu şekilde olmaktadır:

$$f(X_i, \gamma) \approx f_i^{(0)} + \sum_{k=0}^{p-1} D_{ik}^{(0)} \beta_k^{(0)}$$

ve (6.1) doğrusal olmayan regresyon modeli için bu yaklaşım:

$$Y_i = f(X_i, \gamma) + \varepsilon_i$$

$$Y_i \approx f_i^{(0)} + \sum_{k=0}^{p-1} D_{ik}^{(0)} \beta_k^{(0)} + \varepsilon_i \quad (6.24)$$

olmaktadır. $f_i^{(0)}$ terimi sol tarafa alınırsa ve $Y_i - f_i^{(0)}$ farkı $Y_i^{(0)}$ ile gösterilirse, doğrusal regresyon modeli yaklaşımı bulunmaktadır:

$$Y_i^{(0)} \approx \sum_{k=0}^{p-1} D_{ik}^{(0)} \beta_k^{(0)} + \varepsilon_i \quad i = 1, \dots, n \quad (6.25)$$

Burada

$$Y_i^{(0)} = Y_i - f_i^{(0)} \quad (6.26)$$

olmaktadır. Bulunan doğrusal regresyon modeli yaklaşımı (6.25) aşağıdaki gibidir:

$$Y_i = \beta_0 X_{i0} + \beta_1 X_{i1} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

Doğrusal regresyon modeli yaklaşımındaki bağımlı değişkenler $Y_i^{(0)}$, kalıntılardır; başlangıç değerleri ile yer değiştiren parametrelerle oluşturulan doğrusal olmayan regresyon fonksiyonu etrafındaki gözlemlerin sapmalarıdır. X değişkenleri gözlemler $D_{ik}^{(0)}$, başlangıç değerleriyle yer değiştiren parametreler ile n durumun her biri için hesaplanan bağımlı değişkenin beklenen değerinin kısmi türevleridir. Her regresyon katsayısı $\beta_k^{(0)}$, gerçek regresyon parametreleri ile başlangıçtaki parametre tahminleri arasındaki farkı göstermektedir. Böylece, regresyon katsayıları, düzeltilen başlangıçtaki regresyon katsayıları tarafından ayarlama (düzeltme) miktarlarını göstermektedir.

Doğrusal regresyon modeli yaklaşımı matris gösterimiyle aşağıdaki gibi olmaktadır:

$$Y^{(0)} \approx D^{(0)} \beta^{(0)} + \varepsilon \quad (6.27)$$

Burada:

$$Y_{n \times 1}^{(0)} = \begin{bmatrix} Y_1 - f_1^{(0)} \\ \vdots \\ Y_n - f_n^{(0)} \end{bmatrix} \quad (6.28)$$

$$D^{(0)}_{n \times p} = \begin{bmatrix} D_{10}^{(0)} & \dots & D_{1,p-1}^{(0)} \\ \vdots & & \vdots \\ D_n^{(0)} & \dots & D_{n,p-1}^{(0)} \end{bmatrix} \quad (6.29)$$

$$\beta^{(0)}_{p \times 1} = \begin{bmatrix} \beta_0^{(0)} \\ \vdots \\ \beta_{p-1}^{(0)} \end{bmatrix} \quad (6.30)$$

$$\mathcal{E}_{n \times 1} = \begin{bmatrix} \mathcal{E}_1 \\ \vdots \\ \mathcal{E}_n \end{bmatrix} \quad (6.31)$$

şeklindedir. (6.27) yaklaşım modeli tam olarak, X matrisinin yerini alan kısmi türevlerin bulunduğu D matrisi ile (fakat sabit terim için 1'lerin bulunduğu sütun olmadan) genel doğrusal regresyon modeli şeklindedir. Bu yüzden, $\beta^{(0)}$ parametrelerinin tahmini sıradan en küçük kareler yöntemiyle yapılabilmekte ve böylece aşağıdaki sonuç bulunmaktadır:

$$b^{(0)} = \left(D^{(0)'} D^{(0)} \right)^{-1} D^{(0)'} Y^{(0)} \quad (6.32)$$

Burada $b^{(0)}$, en küçük karelerle tahmin edilen regresyon katsayılarının vektörüdür.

Daha sonra bulunan en küçük kareler tahminleri, (6.22)'un ortalamaları ile $g_k^{(1)}$ tahmin edilen regresyon katsayılarının düzeltilmiş halinin elde edilmesi için kullanılmaktadır:

$$g_k^{(1)} = g_k^{(0)} + b_k^{(0)}$$

Burada $g_k^{(1)}$, birinci iterasyonun sonunda γ_k 'nın düzeltilmiş tahminini göstermektedir.

Düzeltilme süreci matris gösterimiyle aşağıdaki gibi olmaktadır:

$$g^{(1)} = g^{(0)} + b^{(0)} \quad (6.33)$$

Böylece, düzeltilmiş regresyon katsayılarının üzerinde yapılacak ayarlamalar uygun şekilde gösterilmektedir. (6.12)'de bulunan Q , en küçük kareler kriteri ölçüsü, $SSE^{(0)}$ kullanılarak $g^{(0)}$ başlangıç regresyon katsayıları için hesaplanmaktadır:

$$SSE^{(0)} = \sum_{i=1}^n [Y_i - f(X_i, g^{(0)})]^2 = \sum_{i=1}^n (Y_i - f_i^{(0)})^2 \quad (6.34)$$

Birinci iterasyonun sonunda, düzeltilmiş tahmin edilen regresyon katsayıları $g^{(1)}$ ve bu durumda en küçük kareler kriteri ölçüsü hesaplanır. $SSE^{(1)}$ şu şekilde gösterilmektedir:

$$SSE^{(1)} = \sum_{i=1}^n [Y_i - f(X_i, g^{(1)})]^2 = \sum_{i=1}^n (Y_i - f_i^{(1)})^2 \quad (6.35)$$

Eğer Gauss-Newton metodu ilk iterasyonda etkin olarak çalışıyorsa; $g^{(1)}$ düzeltilmiş tahmin edilen regresyon katsayıları daha iyi tahminler olacağı için, $SSE^{(1)}$, $SSE^{(0)}$ 'dan daha küçük olacaktır.

$f(X_i, g^{(0)})$ ve $f(X_i, g^{(1)})$ doğrusal olmayan regresyon fonksiyonları, $SSE^{(0)}$ ve $SSE^{(1)}$ hesaplanırken kullanılmaktadır.

$g^{(1)}$ düzeltilmiş regresyon katsayıları, doğrusal olmayan regresyon problemi için en küçük kareler tahminleri değildir. Çünkü (6.27) tahmin edilen model, sadece doğrusal olmayan modelin bir yaklaşımıdır. Bu nedenle, Gauss-Newton metodu sadece yeni başlangıç değerleri için $g^{(1)}$ ile tanımlanan işlemi tekrar etmektedir. Bu, $g^{(2)}$ ile gösterilen, düzeltilmiş tahminlerin yeni bir setini oluşturmaktadır ve yeni en küçük kareler kriteri ölçüsü $SSE^{(2)}$ olmaktadır. İteratif işlem, $g^{(s+1)} - g^{(s)}$ ardışık katsayı tahminleri arasındaki fark ve $SSE^{(s+1)} - SSE^{(s)}$ ardışık en küçük kareler kriteri ölçüsü arasındaki fark önemsiz olana kadar devam etmektedir.

Gauss-Newton metodu birçok doğrusal olmayan regresyon uygulamalarında etkili olarak çalışmaktadır. Ancak bazı örnekler, çok fazla iterasyon gerektirebilir ve yakınsama gerçekleşmeyebilir.

Gauss-Newton ya da başka bir doğrudan nümerik arama yöntemi kullanılırken, sık sık kesin çözüm bulunduktan sonra başlangıç değerlerinin diğer setlerinden bulunan sonucun aynı olup olmayacağının araştırılması istenir.

Doğrusal regresyon için en küçük kareler ile doğrusal olmayan regresyon için en küçük karelerin birbirini tutmadığı bazı özellikler vardır. Örneğin, kalıntılar doğrusal olmayan en küçük kareler için her zaman sıfır toplamını vermez. Buna ek olarak, SSE ve SSR'ın toplamı her zaman SSTO'ya eşit değildir. Sonuç olarak, çoklu belirginlik katsayısı $R^2 = SSR/SSTO$ doğrusal olmayan regresyon için anlamlı tanımlayıcı bir istatistik değildir. Ancak, R^2 çok yüksek olmalıdır.

6.3 Modelin Oluşturulması ve Tanı (Diagnostics)

Doğrusal olmayan modellerin fonksiyonel formları genellikle, direkt bir şekilde açıklayıcı değişkenlerin, eğimin ve etkileşim etkisinin eklenmesi ya da silinmesine daha az uygundur. Seçilen doğrusal olmayan regresyon modelinin geçerliliği, doğrusal regresyon modellerinde olduğu gibi incelenmektedir.

Kalıntıların zamana karşı, tahmin değerlerine karşı ve açıklayıcı değişkenlerin her birine karşı grafiği doğrusal regresyon modellerinde olduğu gibi doğrusal olmayan regresyon modellerinde de varsayılan modelin incelenmesinde yardımcı olmaktadır.

Eğer eşit olmayan hata varyansları bulunursa, doğrusal olmayan regresyon modelinin tahmin edilmesinde ağırlıklı en küçük kareler kullanılmaktadır. Alternatif olarak, bağımlı değişkende yapılan dönüşümlerin hata terimlerinin varyansını dengeleyebileceği ve ayrıca doğrusal bir regresyon modelinin kullanımına izin verip vermeyeceği araştırılmaktadır.

6.4 Doğrusal Olmayan Regresyon Parametreleri ile İlgili Çıkarılan Sonuçlar

Örnek büyüklüğü her ne olursa olsun normal dağılan hata terimleri ile doğrusal regresyon modelleri için regresyon parametrelerinin incelenmesi gerekmektedir. Ancak, herhangi bir verilen örnek büyüklüğü için en küçük kareler ve maksimum olabilirlik tahmincilerinin normal dağılmadığı, yansız olmadığı ve minimum varyansa sahip olmadığı yerde, normal hata terimleri ile doğrusal olmayan regresyon modeli için bu durum söz konusu değildir.

Sonuç olarak, doğrusal olmayan regresyonda, regresyon parametreleri ile ilgili çıkarılan sonuçlarda genellikle büyük örneklem teorisi temel alınmaktadır. Bu teori, örnek hacmi büyük olduğu zaman, normal dağılan hata terimli doğrusal olmayan regresyon modelleri için en küçük kareler ve maksimum olabilirlik tahmincilerinin yaklaşık olarak normal dağıldığını, hemen hemen yansız olduğunu ve neredeyse minimum varyansa sahip olduğunu ifade etmektedir. Bu büyük örneklem teorisine, ayrıca hata terimleri normal dağılmadığı zaman da başvurulmaktadır.

6.4.1 Hata Terimleri Varyansının Tahmini

σ^2 'nin tahmini doğrusal regresyonda olduğu gibi yapılmaktadır:

$$MSE = \frac{SSE}{n - p} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n - p} = \frac{\sum [Y_i - f(X_i, g)]^2}{n - p} \quad (6.36)$$

Burada $\mathbf{g}$ en son bulunan parametre tahminlerinin vektörüdür. Bu yüzden, kalıntılar, $\mathbf{g}$ en sondaki regresyon katsayılarının tahminleri kullanılarak, tahmin edilen regresyon fonksiyonundan sapmalardır. Doğrusal olmayan regresyon için, MSE, σ^2 'nin yansız tahmincisi değildir, fakat örnek hacmi büyük olduğu zaman yanlılık küçüktür.

6.4.2 Büyük Örneklem Teorisi

Hata terimleri bağımsız ve normal dağıldıklarında ve örnek hacmi makul bir şekilde büyük olduğunda, aşağıdaki teorem doğrusal olmayan regresyon modelleri için çıkarılacak sonuçların temellerini oluşturmaktadır:

Teorem: ε_i hata terimleri bağımsız $N(0, \sigma^2)$ ve örnek hacmi n yeterince büyük olduğunda, $\mathbf{g}$ 'nin örnekleme dağılımı yaklaşık olarak normaldir. Beklenen değeri ise yaklaşık olarak:

$$E\{\mathbf{g}\} \approx \boldsymbol{\gamma} \quad (6.37)$$

şeklinde elde edilmektedir. Regresyon katsayılarının yaklaşık olarak varyans-kovaryans matrisi aşağıdaki gibi tahmin edilir:

$$s^2\{\mathbf{g}\} = \text{MSE}(\mathbf{D}'\mathbf{D})^{-1} \quad (6.38)$$

$\mathbf{D}^{(0)}$ 'ın $\mathbf{g}^{(0)}$ 'a göre hesaplanan kısmi türevlerin matrisi olması gibi, burada $\mathbf{D}$ 'de son olarak bulunan en küçük kareler tahminleri $\mathbf{g}$ 'ye göre kısmi türevlerin matrisidir. $s^2\{\mathbf{g}\}$ tahmin edilen yaklaşık varyans-kovaryans matrisi, doğrusal regresyonda olduğu gibidir; burada $\mathbf{D}$, $\mathbf{X}$ matrisinin yerine kullanılmaktadır.

Böylece, örnek hacmi büyük ve hata terimleri sabit varyans ile bağımsız normal dağıldığında, doğrusal olmayan regresyon için $\mathbf{g}$ 'ye göre en küçük kareler tahminleri yaklaşık olarak normal dağılmaktadır ve hemen hemen yansızdır. Ayrıca (6.38) varyans-kovaryans matrisi minimum varyansları tahmin ettiği için, en küçük kareler tahmincileri varyansı minimuma yakındır. Büyük örneklem teorisi hata terimleri normal dağılmadığı zaman da geçerlidir.

Bu teoremin bir sonucu olarak, örnek hacmi büyük olduğu zaman, doğrusal olmayan regresyon parametreleri ile ilgili çıkarılan sonuçlar doğrusal regresyon ile aynı şekilde elde edilmektedir. Böylece bir regresyon parametresi için aralık tahmini doğrusal regresyonda olduğu gibi bulunmakta ve test edilmektedir. İhtiyaç duyulan tahmin edilen varyans $s^2\{\mathbf{g}\}$ 'den elde edilmektedir. Sadece doğrusal olmayan regresyona yaklaşık olarak

başvurulduğu zaman bu sonuç çıkarma işlemleri kesindir ve yaklaşım çoğu zaman oldukça iyidir. Örnek hacmi büyük örneklem yaklaşımı için oldukça küçük olduğunda, bazı doğrusal olmayan regresyon modelleri için yaklaşım iyi olabilir. Diğer doğrusal olmayan regresyon modelleri için ise örnek hacminin büyük olmasına gerek duyulmaktadır.

6.4.3 Büyük Örneklem Teorisi Ne Zaman Uygulanır?

Verilmiş olan herhangi bir doğrusal olmayan regresyon uygulamasında örnek hacmi yeterince büyük olduğu zaman bu teorinin uygulanıp uygulanamayacağını belirten bir kural istenmektedir. Bu yüzden, asimptotik teoremi temel alınarak büyük örneklem sonuçlarını elde etmek daha uygundur. Fakat, büyük örneklem sonuç çıkarma yönteminin ne zaman uygun olup, ne zaman uygun olmadığını belirten basit bir kural yoktur. Ancak, büyük örneklem sonuç çıkarma işlemlerinin kullanımının uygunluğunun tahmin edilmesine yardımcı olacak birkaç ana nokta belirlenmiştir.

1.Doğrusal olmayan regresyon parametrelerinin tahminlerinin bulunmasındaki iteratif işlemlerin hızlı yakınsaması, sık sık (6.27) doğrusal yaklaşım modelinin doğrusal olmayan regresyon modelinin iyi bir yaklaşımı olduğuna işarettir ve bu yüzden regresyon tahminlerinin asimptotik özellikleri uygulanabilir. Yavaş yakınsama, büyük örneklem sonuç çıkarma yöntemleri kullanılmadan önce dikkat edilmesi ve diğer ana noktaların düşünülmesini önermektedir.

2.Büyük örneklem sonuç çıkarma işlemlerinin uygunluğu ile ilgili yol gösterecek çeşitli ölçüler geliştirilmiştir. Doğrusal olmayan modeller için eğrilik ölçüleri geliştirilmiştir. Bu ölçüler, eldeki veri doğrusal yaklaşım modeline yeterince yaklaşıyorsa, tahmin edilen doğrusal olmayan regresyon fonksiyonunun boyutunu göstermektedir. Ayrıca, tahmin edilen regresyon katsayılarının yanlılığının tahmin edilmesi için bir formül geliştirilmiştir. Küçük bir yanlılık büyük örneklem sonuç çıkarma işlemlerinin uygunluğunu desteklemektedir. Tahmin edilen regresyon katsayılarının örnekleme dağılımlarının çarpıklığının bir tahmini de geliştirilmiştir. Küçük bir çarpıklığın göstergesi, örnekleme dağılımlarının yaklaşık normalliğini ve büyük örneklem sonuç çıkarma işlemlerinin uygunluğunu desteklemektedir.

3.Bootstrap örnekleme, doğrusal olmayan regresyon parametrelerinin örnekleme dağılımlarının yaklaşık olarak normal olup olmadığını, örnekleme dağılımlarının varyanslarının doğrusal yaklaşım modeli için olan varyanslara yakın olup olmadığını ve parametre tahminlerinin her biri içindeki yanlılığın oldukça küçük olup olmadığını direkt bir şekilde sınanmasını sağlamaktadır. Eğer böyleyse, doğrusal olmayan regresyon tahminlerinin

örnekleme davranışı doğrusala yakın olarak ifade edilmekte ve büyük örneklem sonuç çıkarma işlemleri uygun bir şekilde kullanılmaktadır.

6.4.4 Bir γ_k 'nın Aralık Tahmini

Büyük örneklem teoremi temel alındığında, örnek hacmi büyük olduğunda ve hata terimleri normal dağıldığında aşağıdaki yaklaşım elde edilmektedir:

$$\frac{g_k - \gamma_k}{s\{g_k\}} \sim t(n-p) \quad k = 0, 1, \dots, p-1 \quad (6.39)$$

Burada $t(n-p)$, $n-p$ serbestlik dereceli bir t değişkenidir. Bu yüzden, her bir γ_k için $1-\alpha$ 'lık yaklaşık güven aralıkları:

$$g_k \pm t(1-\alpha/2; n-p)s\{g_k\} \quad (6.40)$$

olmaktadır.

6.4.5 Bir γ_k 'yla İlgili Test

Tek bir değişkenle yani γ_k 'yla ilgili büyük örneklem testi alışlagelmiş şekilde yapılmaktadır.

Hipotezler:

$$H_0 : \gamma_k = \gamma_{k0} \quad (6.41)$$

$$H_1 : \gamma_k \neq \gamma_{k0} \quad (6.42)$$

Burada γ_{k0} , γ_k 'nın özel olarak seçilmiş değeridir, n makul bir şekilde büyük olduğunda t^* test istatistiği kullanılmaktadır:

$$t^* = \frac{g_k - \gamma_{k0}}{s\{g_k\}} \quad (6.43)$$

Yaklaşık olarak α anlam seviyesinde karar kuralı:

$$\text{Eğer } |t^*| \leq t(1-\alpha/2; n-p) \text{ ise, } H_0 \text{ reddedilmez.} \quad (6.44)$$

$$\text{Eğer } |t^*| > t(1-\alpha/2; n-p) \text{ ise, } H_0 \text{ reddedilir.} \quad (6.45)$$

7. POISSON REGRESYON ANALİZİ

Poisson regresyon, lojistik regresyondan sonra en genel olan ikinci genelleştirilmiş doğrusal modeldir. Bağımlı değişken oluş sayısı (count) ile belirtilen bir veri olduğunda, yani belirli bir zaman ya da yerde olan olayların sayısı olduğunda Poisson regresyon kullanılmaktadır.

$$y_i \sim \text{Poisson}(\mu_i) \quad i = 1, \dots, N \quad (7.1)$$

olduğu varsayılmakta ve açıklayıcı değişkenlerin bir vektörü ile ortalama μ_i arasında bir ilişki kurulmaya çalışılmaktadır. Buradaki amaç, Poisson regresyon modelinin şeklini ve modelin çeşitli özelliklerini tanıtmaktır.

7.1 Poisson Dağılımı

Poisson regresyon analizi, bağımlı değişken Y 'nin Poisson dağılımı gösterdiğini varsaymaktadır. μ parametresi ile Poisson olasılık dağılımı aşağıdaki formülde verildiği gibidir:

$$p(Y; \mu) = \frac{\mu^Y e^{-\mu}}{Y!}, \quad Y = 0, 1, \dots, \infty \quad (7.2)$$

Teorik olarak, bir Poisson rastlantısal değişkeni herhangi bir negatif olmayan tamsayı değeri alabilmektedir. Buradaki olasılık, μ değerinin bir fonksiyonu olarak değişmektedir.

Poisson dağılımı çoğunlukla nadir olayların oluş sayısını modellemek için kullanılmaktadır. Bir topluluk içinde kesin bir zaman aralığında deri kanserine yakalananların sayısı ya da her yıl belirli bir yerde olan trafik kazalarının sayısı gibi. Poisson dağılımı ilginç bir istatistiksel özelliğe sahiptir:

$$E(Y) = \text{Var}(Y) = \mu \quad (7.3)$$

7.2 Poisson Regresyon

Alt gruplara sahip bağımlı değişken Y ve açıklayıcı değişkenler de $X_1, X_2, \dots, X_k$ olsun. Alt grup i için ($i = 1, 2, \dots, n$), Y_i gözlenen hataların sayısı ve ℓ_i 'de bu alt grup içindeki bütün kişiler için sonuna kadar devam eden sürelerin toplam uzunluğu olsun. i alt grubu için $X_1, X_2, \dots, X_k$ değerleri $X_i = (X_{i1}, X_{i2}, \dots, X_{ik})$ şeklinde; bilinmeyen parametreler $\beta = (\beta_0, \beta_1, \dots, \beta_k)$ ve $\lambda(X_i, \beta)$, her i alt grubu için hata oranını gösteren X_i ve β 'nin

(örneğin $\exp(\beta_0 + \sum_{j=1}^k \beta_j X_{ij})$ gibi) özel bir fonksiyonudur. i . alt grup için beklenen hataların sayısı aşağıdaki gibidir:

$$E(Y_i) = \mu_i = \ell_i \lambda(X_i, \beta) \quad i = 1, 2, \dots, n \quad (7.4)$$

Burada Y_i , bir Poisson rastlantısal değişkenidir. Bu yüzden $\lambda(X_i, \beta) > 0$ olmalıdır^{*}.

Y_i 'nin μ_i ortalama ile Poisson dağılımı gösterdiği varsayımı altında,

$$p(Y_i; \mu_i) = \frac{\mu_i^{Y_i} \exp(-\mu_i)}{Y_i!} \quad i = 1, 2, \dots, n \quad (7.5)$$

olmaktadır. (7.4) ve (7.5), $i = 1, 2, \dots, n$ sonucunda

$$p(Y_i; \beta) = \frac{[\ell_i \lambda(X_i, \beta)]^{Y_i} \exp(-\ell_i \lambda(X_i, \beta))}{Y_i!} \quad (7.6)$$

olmaktadır. Burada $Y_i = 0, 1, \dots, \infty$ ve $i = 1, 2, \dots, n$ 'dir.

Poisson regresyon ve standart çoklu regresyon arasındaki kavramsal fark, standart çoklu regresyonda normal dağılım kullanılırken diğerinde Poisson dağılımının kullanılmasıdır. Burada analizin amacı, açıklayıcı değişkenlerin bir fonksiyonu yardımı ile bağımlı değişkenin ortalamasının modellenmesidir. Regresyon katsayılarını tahmin etmek için, kullanılan olabilirlik fonksiyonunun şekline, bağımlı değişkenin dağılımının varsayımlarıyla ilgili olarak karar verilmektedir.

Y_i 'nin aşağıdaki dağılıma sahip olduğu varsayılın:

$$p(Y_i; \beta) = \frac{[\ell_i \lambda(X_i, \beta)]^{Y_i} \exp(-\ell_i \lambda(X_i, \beta))}{Y_i!}$$

ve $Y_1, Y_2, \dots, Y_n$ 'lerin bağımsız Poisson rastlantısal değişkenlerinin bir seti olduğu varsayılın.

Böylece, Poisson regresyon analizi için olabilirlik fonksiyonu genel bir şekilde aşağıdaki gibi olmaktadır:

^{*} $E(Y_i) = \mu_i = \ell_i \lambda(X_i, \beta)$ formülü, binom rastlantısal değişkenin ortalaması $\mu = np$ için olan formüle benzerdir (fakat eşit değildir); burada, ℓ_i , n 'e benzerdir ve $\lambda(X_i, \beta)$, p 'ye benzerdir.

$$\begin{aligned}
L(Y; \beta) &= \prod_{i=1}^n p(Y_i; \beta) = \prod_{i=1}^n \left\{ \frac{[\ell_i \lambda(X_i, \beta)]^{Y_i} \exp(-\ell_i \lambda(X_i, \beta))}{Y_i!} \right\} \\
&= \frac{\left\{ \prod_{i=1}^n [\ell_i \lambda(X_i, \beta)]^{Y_i} \right\} \exp\left[-\sum_{i=1}^n \ell_i \lambda(X_i, \beta)\right]}{\prod_{i=1}^n Y_i!}
\end{aligned} \tag{7.7}$$

Burada $E(Y_i) = \mu_i = \ell_i \lambda(X_i, \beta)$ ve $i = 1, 2, \dots, n$ 'dir.

Pratikte olabilirlik fonksiyonu (7.7)'yi hesaplamak için, $\lambda(X_i, \beta)$ fonksiyonunun özel bir şekli belirlenmelidir. Bu belirleme, çalışma sırasındaki süreç, daha önceki bilgiler ve değişkenler arasındaki ilişkilerle ilgili deneyimler temel alınarak yapılmaktadır. $\lambda(X_i, \beta)$ için, mümkün olabilecek seçimler şöyledir:

$$\lambda(X_i, \beta) = \begin{cases} \exp(\lambda_i^*), & \lambda_i^* = \beta_0 + \sum_{j=1}^k \beta_j X_{ij} & \text{ise} \\ \lambda_i^*, & \lambda_i^* > 0 & \text{ise} \\ \ln(\lambda_i^*), & \lambda_i^* > 1 & \text{ise} \end{cases}$$

kullanılabilir.

$\beta_0, \beta_1, \dots, \beta_k$ 'nın maksimum olabilirlik tahminleri $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$, olabilirlik fonksiyonundan $k+1$ denklemin çözümleri olarak aşağıdaki gibi bulunmaktadır:

$$\frac{\partial}{\partial \beta_j} [\ln L(Y; \beta)] = 0 \quad j = 0, 1, \dots, k \tag{7.8}$$

Yukarıda verilen maksimum olabilirlik denklemlerinin çözümü genellikle iterasyon işlemleri sonucunda bulunmaktadır. Bu yöntem, iteratif olarak yeniden ağırlıklandırılmış en küçük kareler (IRWLS) denilmektedir.

7.2.1 Bağ (Link) Fonksiyonu

Poisson modellerinde, en çok kullanılan bağ (link) fonksiyonu logaritmiktir, buna kanonik bağ da denilmektedir:

$$\log \mu_i = \eta = x_i^T \beta \tag{7.9}$$

Bir logaritmik bağın kullanılması μ_i 'nin tahmin edilen değerlerinin parametre uzayı $[0, \infty)$

arasında kalmasını sağlar. Logaritmik bağ kullanılan bir Poisson modeline bazen logaritmik doğrusal model de denilmektedir.

7.2.2 Varyans Fonksiyonu

Poisson modeli altında,

$E\{y_i\} = \sigma^2\{y_i\} = \mu_i$ 'dir. Bu yüzden varyans fonksiyonu:

$$\sigma^2\{\mu\} = \mu \quad (7.10)$$

olmaktadır. Gerçek bir veri çoğu kez aşırı yayılım* (overdispersed) gösterir. Aşırı yayılım durumu Poisson modelinin izin verdiğiinden daha fazla bir değişimin olduğunu ortaya koymaktadır. Aşırı yayılım söz konusu olduğu zaman, bu durum genellikle iki şekilde ele alınmaktadır:

- $\sigma^2\{y_i\} = \sigma^2\mu_i$ olduğu varsayılır ve tartı parametresi σ^2 tahmin edilir. Böylece bir yarı olabilirlik (quasilikelihood) modeli elde edilmiş olur.
- Bağımlı değişkenin dağılımı negatif binom (Pascal) dağılımıyla değiştirilir. Böylece “Poisson” dağılımındakinden daha fazla yayılım olması sağlanır.

7.2.3 Uyum Ölçüleri

Poisson regresyon modellerinin uyum ölçüleri maksimum yapılmış olabilirlik değerlerinin karşılaştırılmasından bulunmaktadır. Y_i 'nin (7.5) şeklinde Poisson regresyona sahip olduğu ve $Y_1, Y_2, \dots, Y_n$ 'in bağımsız olduğu varsayılın; böylece $\mu_1, \mu_2, \dots, \mu_n$ 'in genel bir fonksiyonu olarak olabilirlik fonksiyonu aşağıdaki şekilde ifade edilmektedir:

$$L(Y; \mu) = \prod_{i=1}^n \frac{\mu_i^{Y_i} \exp(-\mu_i)}{Y_i!} = \frac{\left(\prod_{i=1}^n \mu_i^{Y_i} \right) \exp\left(-\sum_{i=1}^n \mu_i \right)}{\prod_{i=1}^n Y_i!} \quad (7.11)$$

Burada $\mu = (\mu_1, \mu_2, \dots, \mu_n)$ 'dir. Maksimum olabilirlik denklemleri aşağıdaki gibidir:

* **Overdispersion:** Varyansın ortalamadan büyük olması haline overdispersion (aşırı yayılım) denir.

$$\frac{\partial}{\partial \mu_i} [\ln L(Y; \mu)] = 0 \quad i = 1, 2, \dots, n \quad (7.12)$$

Bu sistemin çözümü $\hat{\mu}_i = Y_i$, $i = 1, 2, \dots, n$ şeklinde olmaktadır. Böylece, yukarıdaki olabilirlik fonksiyonu için maksimum yapılmış olabilirlik değeri aşağıdaki gibidir:

$$L(Y; \hat{\mu}) = \frac{\left(\prod_{i=1}^n Y_i^{Y_i} \right) \exp\left(- \sum_{i=1}^n Y_i \right)}{\prod_{i=1}^n Y_i!} \quad (7.13)$$

Burada $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_n) = (Y_1, Y_2, \dots, Y_n)$ olmaktadır.

$(k+1) < n$ olduğunda ve (7.7) olabilirlik fonksiyonu maksimum yapıldığında, (7.11) olabilirlik fonksiyonu temel alınarak $L(Y; \hat{\mu})$ maksimum olabilirliğin değeri daha büyük olacaktır. Bunun sebebi, ikinci denklemde μ_i 'nin yapısı üzerinde kısıtlamalar bulunmazken, birinci denklemde $\mu_i = \ell_i \lambda(X_i, \beta)$ şeklinde kısıtlama bulunmasıdır. Diğer bir deyişle, birinci denklem $H_0 : \mu_i = \ell_i \lambda(X_i, \beta)$, $i = 1, 2, \dots, n$ hipotezi altında olabilirlik fonksiyonu olarak düşünülürken, ikinci denklem $H_1 : \mu_i$ 'nin yapısında kısıtlama yoktur, $i = 1, 2, \dots, n$ hipotezi altında olabilirlik fonksiyonudur.

Böylece, eğer $L(Y; \hat{\beta})$ birinci denklem altında maksimum yapılmış olabilirlik değeri oluyorsa; burada $\hat{\beta}$, β 'nin maksimum olabilirlik tahminidir. Böylece

$$-2 \ln \left[\frac{L(Y; \hat{\beta})}{L(Y; \hat{\mu})} \right] \quad (7.14)$$

μ_i üzerinde bir kısıtlama olmayan model ile ilişkili olarak $\mu_i = \ell_i \lambda(X_i, \beta)$ modelinin uyum değerini belirten bir olabilirlik oranı tipi (likelihood-ratio-type) istatistiktir. Herhangi bir regresyon analizinde veri üzerinde parsimoni* kuralı uygulanacağı için, $k+1$ parametre içeren $\mu_i = \ell_i \lambda(X_i, \beta)$ modeli, veri noktası kadar çok parametre içeren temel model ile bulunabilen maksimum yapılmış olabilirlik değeri kadar neredeyse büyük bir değer

* **Parsimoni:** Olası en az değişken sayısı ile en yalın makul model.

sağlayacaktır. Burada geçen neredeyse büyük sözüyle anlatılmak istenilen, $L(Y; \hat{\beta})$ 'in, (7.14) kullanılarak hesaplanan olabilirlik oranı testi temel alınarak bulunan $L(Y; \hat{\mu})$ değerinden anlamlı bir şekilde küçük olmadığıdır.

$$G(\hat{\beta}) = -2 \ln \left[\frac{L(Y; \hat{\beta})}{L(Y; \hat{\mu})} \right] \quad (7.15)$$

miktarı, $L(Y; \hat{\beta})$ 'nin $L(Y; \hat{\mu})$ 'den anlamlı şekilde daha küçük olup olmadığını değerlendirmek için kullanılan uyum test istatistiğidir ve böylece, varsayılan regresyon modeli $\mu_i = \ell_i \lambda(X_i, \beta)$ 'nin veriyle anlamlı bir uyumsuzluğu olup olmadığının anlaşılması için önerilmektedir. $G(\hat{\beta})$ miktarına ayrıca $\mu_i = \ell_i \lambda(X_i, \beta)$ Poisson regresyon modeli için deviance (Deviance bundan sonra sapma olarak kullanılacaktır) da denilmektedir ve tahmin edilen model için kalıntı değişiminin bir ölçüsü olarak düşünülebilir. $H_0 : \mu_i = \ell_i \lambda(X_i, \beta)$ hipotezi altında, $G(\hat{\beta})$ sapmanın tipik olarak, $n - k - 1$ serbestlik dereceli (büyük örnekler için) yaklaşık bir ki-kare dağılımı gösterdiği varsayılmaktadır. Burada n , (7.11) olabilirliği içindeki belirlenmiş parametrelerin sayısı (örneğin alt grupların, hücrelerin ya da kategorilerin sayısı) ve $k + 1$, (7.7) olabilirliği içindeki parametrelerin (örneğin β_j 'ler) sayısıdır. Böylece, verilen bir veri setinde $\mu_i = \ell_i \lambda(X_i, \beta)$ modelinin uyumu için yaklaşık bir test, $n - k - 1$ serbestlik dereceli ki-kare dağılımının uygun bir tablo değeri için $G(\hat{\beta})$ 'nin hesaplanan değerinin karşılaştırılması ile uygulanabilmektedir.

(7.4) modeli altında i .hücre içinde tahmin edilen bağımlı değişken değeri $\hat{Y}_i = \ell_i \lambda(X_i, \beta)$ ile gösterilmektedir, (7.15) miktarı aşağıdaki şekilde yazılabilir:

$$G(\hat{\beta}) = 2 \sum_{i=1}^n \left[Y_i \ln \left(\frac{Y_i}{\hat{Y}_i} \right) - (Y_i - \hat{Y}_i) \right] \quad (7.16)$$

Bu yüzden, $G(\hat{\beta})$, standart çoklu doğrusal regresyon analizindeki $SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$, ye benzer şekilde davranmaktadır. Tahmin edilen model gözlenen veriye tam olarak uyduğu zaman (örneğin $Y_i = \hat{Y}_i$, $i = 1, 2, \dots, n$ gibi), $G(\hat{\beta}) = 0$ olur ve bağımlı değişkenin gözlenen ve tahmin edilen değerleri arasındaki uyumsuzluk daha büyük olduğunda, $G(\hat{\beta})$ değeri de daha büyük olur.

Tahmin edilen değerler makul bir büyüklükte olduğu zaman (örneğin $\hat{Y}_i > 3$, $i = 1, 2, \dots, n$ gibi), $G(\hat{\beta})$, uygun bir şekilde, yaklaşık olarak Pearson tipi gözlenen değerlere karşı ki-kare istatistiğinin tahminine daha benzer olmaktadır:

$$X^2 = \sum_{i=1}^n \frac{(Y_i - \hat{Y}_i)^2}{\hat{Y}_i} \quad (7.17)$$

Dikkat edilmesi gereken nokta, $\hat{Y}_i$ değerleri çok küçük olduğu zaman, (7.17) istatistiği, yanıltıcı olarak büyük olabilmektedir.

Bir hiyerarşik sınıf içindeki çeşitli modeller için sapmalar, olabilirlik oranı testlerini üretmek için kullanılmaktadır. Özellikle, (7.15) eşitliğinde verilen $G(\hat{\beta})$ sapma değeri ile, $\beta = (\beta_0, \beta_1, \dots, \beta_k)$ parametreler setini içeren (7.7) olabilirliği tekrar düşünölsün. $0 < r < k$ için, β içindeki son $k - r$ parametrenin sıfıra eşit olup olmadığının test edilmek istenildiği varsayölsün. Örneğin sıfır hipotezi $H_0 : \beta_{r+1} = \beta_{r+2} = \dots = \beta_k = 0$ şeklinde olsun. H_0 hipotezi altında, olabilirlik β_r ile (7.7) içindeki β 'nın yer değıştirmesiyle bulunabilir, burada

$$\beta_r = (\beta_0, \beta_1, \dots, \beta_r; 0, 0, \dots, 0) \text{ 'dir.}$$

Eğer bu olabilirlik fonksiyonu $L(Y; \beta_r)$ ile gösterilirse ve eğer $\hat{\beta}_r$, $L(Y; \beta_r)$ kullanılarak β_r 'nin maksimum olabilirlik tahmindisi ise, H_0 'ın olabilirlik oranı testi aşğıdaki test istatistiğı kullanılarak hesaplanır:

$$-2 \ln \left[\frac{L(Y; \hat{\beta}_r)}{L(Y; \hat{\beta})} \right] \quad (7.18)$$

Bu değıer H_0 doğru olduğunda, büyük örnekler için $k - r$ serbestlik derecesi ile yaklaşık olarak bir ki-kare dağılımına sahiptir.

Ayrıca, (7.18) ifadesi tam olarak aşğıdaki gibi gösterilen sapma farkına eşittir:

$$G(\hat{\beta}_r) - G(\hat{\beta}) \quad (7.19)$$

Bunu görmek için, (7.15) ifadesinde verilen $G(\hat{\beta})$ 'in genel tanımına geri dönmek gereklidir. (7.15) ve (7.19) kullanılarak aşğıdaki sonuca ulaşılır:

$$\begin{aligned}
G(\hat{\beta}_r) - G(\hat{\beta}) &= -2 \ln \left[\frac{L(Y; \hat{\beta}_r)}{L(Y; \hat{\mu})} \right] + 2 \ln \left[\frac{L(Y; \hat{\beta})}{L(Y; \hat{\mu})} \right] \\
&= -2 \ln \left[\frac{L(Y; \hat{\beta}_r)}{L(Y; \hat{\beta})} \right]
\end{aligned}$$

Bu ifade tam olarak (7.18) olabilirlik oranı (likelihood ratio test statistic) test istatistiğine eşittir. $H_0 : \beta_{r+1} = \beta_{r+2} = \dots = \beta_k = 0$ hipotezi altında, $G(\hat{\beta}_r) - G(\hat{\beta})$ farkı, büyük örnekler için $k - r$ serbestlik derecesi yaklaşık olarak bir ki-kare dağılımı göstermektedir.

Böylece, bir veri setini analiz etmek için Poisson regresyon kullanıldığında, aday modellerin bir seti içindeki üyelerin sapma çiftleri arasındaki farklar düşünülerek bu modeller karşılaştırılabilmektedir.

7.2.4 Model

Parametrelerin tahmini, IRWLS yöntemi kullanılarak yapılmaktadır. Algoritmanın ilk iterasyonu şu şekilde gösterilebilir:

$$\log \mu_i = \eta_i = X_i' \cdot \beta$$

$$\beta = (X^T \cdot W \cdot X)^{-1} \cdot X^T \cdot W \cdot z \quad (7.20)$$

Burada ağırlıkların matrisi aşağıdaki gibidir:

$$W = \text{Diag} \left[\sigma^2 \{Y_i\} \left(\frac{\partial \eta_i}{\partial \mu_i} \right)^2 \right]^{-1} \quad \text{W tartı matrisi} \quad (7.21)$$

$\eta_i = \log \mu_i$ olduğu için,

$$\frac{\partial \eta_i}{\partial \mu_i} = \frac{1}{\mu_i}$$

sonucu bulunmaktadır.

$$W = \text{Diag} \left[\mu_i \left(\frac{1}{\mu_i} \right)^2 \right]^{-1} = \text{Diag}(\mu_i)$$

$$z = \eta + \left(\frac{\partial \eta}{\partial \mu} \right) (Y - \mu) \quad (7.22)$$

$z_i = \eta_i + (Y_i - \mu_i) / \mu_i$ olarak elde edilmektedir. z 'ye çalışan rastlantı değişkeni* (working variate) denilmektedir.

Geçerli β tahminini bulmak için,

$\eta_i = X_i^T \cdot \beta$, $\mu_i = \exp(\eta_i)$ ve $z_i = \eta_i + (Y_i - \mu_i) / \mu_i$ hesaplanmaktadır. Daha sonra β 'nın yeni tahminini elde etmek için μ_i 'lerin ağırlıklı değerleri kullanılarak z , X üzerinde regresyona sokulmaktadır. Bu işlem β 'nın değeri bir noktada yakınsayana kadar devam etmektedir.

β 'nın tahmini yakınsadıktan sonra, aşağıdaki değerler incelenmelidir:

- β tahminlerinin varyans-kovaryans matrisi:

$$\sigma^2 \{\hat{\beta}\} = (X^T \cdot W \cdot X)^{-1} \quad (7.23)$$

- Pearson kalıntıları:

$$r_i = \frac{Y_i - \mu_i}{\sqrt{\mu_i}} \quad (7.24)$$

- Pearson uyum test istatistiği:

$$X^2 = \sum_{i=1}^n r_i^2 \quad (7.25)$$

- Kaldıraç değerleri** (Leverage values):

* Bundan sonra working variate yerine, çalışan rastlantı değişkeni ifadesi kullanılacaktır.

** Regresyonda, bu terim, $(X(X'X)^{-1}X')$ projeksiyon matrisinin köşegen elemanları anlamına gelmektedir. Köşegen üzerindeki h_{ii} elemanı, i . gözlem X değerleri ve bütün X değerlerinin ortalamaları arasındaki farkı ifade etmektedir. Bu değerler verilen bir gözlem için X değerlerinin sapan değer olup olmadığını gösterir. Köşegen elemanı, kaldıraç (leverage) olarak adlandırılır. Büyük kaldıraç değeri, i . gözlemin X gözlemlerinin merkezinden olan uzaklığını gösterir (Neter, 1985).

$$Diag(H) = Diag[W^{1/2}.X.(X^T.W.X)^{-1}.X^T.W^{1/2}] \quad (7.26)$$

Sabitler ihmal edildiğinde, logaritmik olabilirlik fonksiyonu (loglikelihood) aşağıdaki gibidir:

$$l = \sum_{i=1}^N \{y_i \log \mu_i - \mu_i\} \quad (7.27)$$

Sapma G , $\log \mu_i = X_i^T \beta$ tam model ile geçerli modelin uyumunu karşılaştırmak için kullanılan olabilirlik oranı test istatistiğidir. Ortalamanın bağımsız değişkenlerle nasıl ilişkili olduğuna dair hiçbir varsayım yapılmadan, tam model her y_i değeri için ayrı bir ortalamaya sahiptir. Logaritmik olabilirliğin esası olan $\mu_i = y_i$ 'de, $y_i \log \mu_i - \mu_i$ 'nin maksimuma ulaşmasını göstermek oldukça kolaydır. Bu yüzden, sapma:

$$G = 2 \sum_{i=1}^n \left[y_i \cdot \log \left(\frac{y_i}{\mu_i} \right) - (y_i - \mu_i) \right] \quad (7.28)$$

olmaktadır. $y_i = 0$ olduğu herhangi bir durumda bu ifadenin tanımsız olduğu görülmektedir.

Eğer model uymuyorsa bu aşağıdaki sebeplerden kaynaklanabilir:

1. Eğer zamanla ilgili bir değişken modele dahil edilmiş ise, trend var olabilir,
2. Modele dahil edilmemiş değişkenler olabilir ya da
3. Aşırı yayılım olabilir.

Bu bilgiler ışığı altında ikinci ve üçüncü sebebi birbirinden ayırt etmek mümkün olmamaktadır. Ancak trendin var olup olmadığı daha yüksek dereceli bir denklem kullanılarak anlaşılabilmektedir.

Eğer eldeki veride y_i , $Poisson(\mu_i)$ şeklinde dağılan ölümünün sayısını göstermekte ve m 'de tedavi olanların sayısını gösteren bir çapraz tablo şeklinde ise, model aşağıdaki şekilde kurulabilmektedir:

$$\mu_i = m_i \lambda_i \quad (7.29)$$

Burada λ_i , her aydaki hastaların ortalama ölüm oranıdır. Tedavi olanlar değiştiği için, burada λ_i , μ_i değildir. λ_i bağımsız değişkenlerle ilgilidir. Logaritmik bağ altında,

$$\log \mu_i = \eta_i = o_i + X_i^T \beta \quad (7.30)$$

olmaktadır. Burada,

$$o_i = \log m_i \quad (7.31)$$

olmaktadır. Genelleştirilmiş doğrusal modeller terimleri içinde o_i 'ye offset denilmektedir. Offset bir sabittir, genelleştirilmiş bir doğrusal model içinde bulunan, zaten bilinmekte olan bir regresyon katsayısıdır. Bu yüzden offset tahmin edilmek zorunda değildir. Offset kullanılması Poisson regresyonda çok olağandır. Çünkü veri içinde maruz kalma durumu çoğu kez bir gözlemden sonrakine değişmektedir. Genelleştirilmiş doğrusal modeller içinde offset kullanımı oldukça kolaydır. Zincir kuralı kullanılarak Fisher skorlama algoritması aşağıdaki şekilde türetilmektedir:

$$\frac{\partial l}{\partial \beta_j} = \left(\frac{\partial l}{\partial \theta} \right) \left(\frac{\partial \theta}{\partial \mu} \right) \left(\frac{\partial \mu}{\partial \eta} \right) \left(\frac{\partial \eta}{\partial \beta_j} \right) \quad (7.32)$$

η 'nın yeni tanımı bu zincir kuralı içindeki terimlerin herhangi birinde değişmemektedir. β 'nin elemanlarına göre birinci ve ikinci türevler değişmemiştir. Algoritma hala aşağıdaki gibidir:

$$\beta^{(t+1)} = \beta^{(t)} + (X^T W X)^{-1} X^T A(y - \mu) \quad (7.33)$$

Burada,

$$W = \text{Diag} \left[\sigma^2 \{y_i\} \left(\frac{\partial \eta_i}{\partial \mu_i} \right)^2 \right]^{-1},$$

$$A = \text{Diag} \left[\sigma^2 \{y_i\} \left(\frac{\partial \eta_i}{\partial \mu_i} \right) \right]^{-1} = \left(\frac{\partial n}{\partial \mu} \right) W \quad (7.34)$$

Bunlar IRWLS yönteminde kullanılarak aşağıdaki sonuçlar elde edilir:

$$\begin{aligned} \beta^{(t+1)} &= (X^T W X)^{-1} [X^T W X \beta^{(t)} + X^T A(y - \mu)] \\ &= (X^T W X)^{-1} X^T W z \end{aligned} \quad (7.35)$$

Burada çalışan rastlantı değişkeni (z_i) aşağıdaki şekilde değiştirilmiştir:

$$z_i = \eta_i - o_i + \left(\frac{\partial \eta_i}{\partial \mu_i} \right) (y_i - \mu_i) \quad (7.36)$$

Çünkü burada,

$$X_i^T \beta = \eta_i - o_i \quad (7.37)$$

şeklindedir.

7.2.4.1 Aşırı Yayılım için Tartı Parametresi

Veride aşırı yayılım olduğunu göstermek olağan bir durumdur. Aşırı yayılımı çözmenin bir yolu σ^2 'nin bazı sıfırdan büyük olduğu farz edilen değerleri için aşağıdaki varsayımı yapmaktır:

$$\sigma^2\{y_i\} = \sigma^2 \mu_i \quad (7.38)$$

$\sigma^2 = 1$ olduğunda Poisson'dur; σ^2 'nin diğer değerleri için bir yarı olabilirlik modeli olmaktadır. Yarı olabilirlik yaklaşımı altında, $\hat{\beta}$, σ^2 'nin herhangi bir değeri için aynı olmaktadır ve Poisson modeli altında maksimum olabilirlik tahmini ile uymaktadır. Tartı parametresi aşağıdaki gibi tahmin edilmektedir:

$$\hat{\sigma}^2 = X^2 / (n - p) \quad (7.39)$$

Burada,

$$X^2 = \sum_{i=1}^N \frac{(y_i - \mu_i)^2}{\mu_i} \quad (7.40)$$

olmaktadır. Bu da zaten Pearson uyum test istatistiğidir.

Pek çok durumda modelin uyumsuzluğunun sebebi model dışında bırakılan değişkenler ya da aşırı yayılım olarak görülmektedir. Bu yüzden, ortalama yapısı için, tartı parametresinin tahmin edilmesi mantıklı olmaktadır. Tartı parametresi tahmin edildikten sonra, aşağıdaki düzeltmeler yapılmalıdır:

- $\hat{\beta}$ için tahmin edilen varyans-kovaryans matrisi σ^2 ile çarpılır, bu yüzden:

$$\hat{\sigma}^2\{\hat{\beta}\} = \hat{\sigma}^2 (X^T W X)^{-1} \quad (7.41)$$

olmaktadır.

- Test istatistikleri X^2 , G , ΔX^2 ve ΔG ; $\hat{\sigma}^2$ 'e bölünmelidir.

- Pearson ve sapma kalıntıları $\hat{\sigma}$ 'e bölünmelidir.

7.2.4.2 Göreli Risk ya da Risk Oranı (Relative Risk or Risk Ratio)

Bağımlı değişken Y oluş sayısı ile belirtilen bir veri olsun. Bu oluş sayıları bazı yaş kategorilerine ayrılın ve veriler farklı iki bölgeden toplanmış olsun. Bu iki bölgeyi karşılaştırmak için göreli risk ya da diğer adıyla risk oranı kullanılmaktadır. Buradaki risk terimi, aslında bir olayın oranı ile ilgili olasılık anlamına gelmektedir. i 'ler yaş gruplarını j 'ler ise bölgeleri belirtmektedir. Buradaki λ_{ij} , (i, j) 'inci grup içindeki gerçek riski (örneğin topluluğu) göstermektedir:

$$RR_i = \frac{\lambda_{i1}}{\lambda_{i0}} \quad (7.42)$$

Eğer $RR_i = 1$ ise, topluluk riskleri i . yaş grubu içinde aynıdır; eğer $RR_i > 1$ ise, bu yaş grubunda bir ile gösterilen bölge için risk sıfır ile gösterilen bölge için olan riskten daha büyük demektir. Bu oran aynı zamanda rate ratio, incidence density ratio ve hazard ratio isimleriyle de kullanılmaktadır.

7.3 Negatif Binom Model

Aşırı yayılımı çözenin bir başka yolu Poisson'dan daha fazla yayılan parametrik bir model olan negatif binom'dur.

$y \sim \text{Poisson}(\lambda)$ olduğu varsayılın, fakat λ 'nın kendisi de gama dağılımına sahip rastlantısal bir değişken olsun. Bu aşağıdaki şekilde gösterilmektedir:

$$y | \lambda \sim \text{Poisson}(\lambda) \quad (7.43)$$

$$\lambda \sim \text{Gama}(\alpha, \beta) \quad (7.44)$$

Burada $\text{Gama}(\alpha, \beta)$, $\alpha\beta$ ortalama ve $\alpha\beta^2$ varyansa sahip olan gama dağılımıdır. Yoğunluk aşağıda gösterildiği gibidir:

$$P(\lambda) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\lambda/\beta), & \lambda > 0 \\ 0 & , d.d. \end{cases} \quad (7.45)$$

Böylece, y 'nin şartsız dağılımının negatif binom olduğu aşağıdaki gibi gösterilir:

$$P(y) = \frac{\Gamma(\alpha + y)}{\Gamma(\alpha)y!} \left(\frac{\beta}{1 + \beta} \right)^y \left(\frac{1}{1 + \beta} \right)^\alpha, \quad y = 0, 1, 2, \dots \quad (7.46)$$

Bu dağılımın ortalaması,

$$E\{y\} = E\{E\{y \mid \lambda\}\} = E\{\lambda\} = \alpha\beta \quad (7.47)$$

ve varyansı,

$$\begin{aligned} \sigma^2\{y\} &= E\{\sigma^2\{y \mid \lambda\}\} + \sigma^2\{E\{y \mid \lambda\}\} \\ &= \sigma^2\{\lambda\} + E\{\lambda\} \\ &= \alpha\beta + \alpha\beta^2 \end{aligned} \quad (7.48)$$

olmaktadır.

Bir regresyon modeli oluşturmak için, $\mu = \alpha\beta$ ve $\kappa = 1/\alpha$ parametrelerinin terimleri ile negatif binom dağılımı oluşturulabilir. Bu yüzden $E\{y\} = \mu$ ve $\sigma^2\{y\} = \mu + \kappa\mu^2$ 'dir. Varyans fonksiyonunun ikinci dereceden olduğu görülmektedir. y'nin dağılımı:

$$P(y) = \frac{\Gamma(\kappa^{-1} + y)}{\Gamma(\kappa^{-1})y!} \left(\frac{\kappa\mu}{1 + \kappa\mu} \right)^y \left(\frac{1}{1 + \kappa\mu} \right)^{1/\kappa} \quad (7.49)$$

$\kappa \rightarrow 0$ 'a yaklaşırken bu ifade *Poisson*(μ)'ye yaklaşır. Negatif binom aşırı yayılım için uygun olmaktadır. Fakat Poisson modeline göre az yayılım* (underdispersed) durumuna uygun olmamaktadır. Böylece, aşırı yayılım sorununu çözmek için,

$$y_i \sim \text{Negbin}(\mu_i, \kappa)$$

olduğu varsayılır ve logaritmik bağ fonksiyonuna başvurulur. Bu yüzden,

$$\log \mu_i = \eta_i = X_i^T \beta$$

olmaktadır ya da bir offset'e ihtiyaç duyulduğunda aşağıdaki şekilde kullanılabilir:

$$\log \mu_i = \eta_i = o_i + X_i^T \beta$$

* **Underdispersion:** Varyansın ortalamadan küçük olması haline underdispersion (az yayılım) denir.

7.4 Az Yayılımla İlgili Bir Açıklama

Daha önce negatif binom modeli aşağıdaki gibi düşünülmekteydi:

$$y_i \sim \text{Negatif binom}(\mu_i, \kappa)$$

$$\log \mu_i = o_i + X_i^T \beta$$

Yukarıdaki ifade $\kappa \rightarrow 0$ olduğunda bir Poisson modeline yaklaşmaktadır. Çünkü varyans aşağıdaki gibidir:

$$\sigma^2\{y_i\} = \mu_i + \kappa \mu_i^2 \quad (7.50)$$

Bu model Poisson'daki aşırı yayılım sorununu çözebilir. Fakat az yayılım durumunun getirdiği sorunları aşamamaktadır. Eğer az yayılım gösteren bir veri ile bir model oluşturulursa, muhtemelen $\hat{\kappa} = 0$ sonucu elde edilecektir. Bu durumda model, bir Poisson modeline indirgenmiş olacaktır.

Eğer az yayılan bir veri ile bir Poisson modeli oluşturulursa, yalın sonuçlar kullanılarak, $\hat{\beta}$ katsayıları yansız olur. Fakat onların standart hataları büyük olabilir. Bu şu anlama gelmektedir:

- Güven sınırları kapsamın nominal oranlarından daha büyük olacaktır(Bu kabul edilebilir).
- Birinci tip hata yapma oranı normalden daha düşük olacaktır(Bu kabul edilebilir).
- Güç gerektiğinden daha az olacaktır(Bu belki kabul edilmeyebilir).

Oluş sayısı ile belirlenen bir veri az yayılım gösterdiğinde bu sorunu halletmenin en genel yolu bir Poisson modeli uygulamak ve bir tartı parametresi tahmin etmektir(Tahmin birden az olabilir). Bu bir yarı olabilirlik yaklaşımı olmaktadır.

7.5 Bol Sıfırlı Modeller (Zero-Inflated Models)

Poisson ya da negatif binom tarafından açıklanabilen, oluş sayısı ile belirlenen bir veride pek çok gözlenen değerin sıfır olması ($y_i = 0$) oldukça normal bir durumdur. Örneğin, üniversite öğrencileri arasındaki bir alkol araştırması düşünölsün. Bağımlı değişken geçmiş 30 gün içindeki alkol tüketme fırsatının sayısı olsun. Verinin pek çok sıfır içermesi olasıdır. Bazıları doğrudan sıfır diyeceklerdir. Çünkü onlar hiç alkol tüketmeyenlerdir(gerekli sıfırlar).

Diğerleri ise bazen alkol tüketmektedir fakat son 30 gün içinde bu olmamıştır(kazara sıfırlar). y_i 'yi iki dağılımın bir karışımı olarak modellemek mantıklı olabilir:

- Bağımlı değişken değerleri bir olasılığı ile sıfırlardan oluşmaktadır; ve
- Bağımlı değişken değerleri 0,1,2,... kütlesi ile yalın bir modeli (Poisson gibi) takip etmektedir.

Sonraki kısım Poisson olduğunda, buna bol sıfırlı Poisson (zero-inflated Poisson, kısaca ZIP) modeli denilmektedir.

Literatürde, bir ZIP* modeli, EM (Expectation Maximization Algorithm) algoritması kullanılarak çözülmektedir. Eğer y_i bazı sıfır parçalarından oluşuyorsa ve eğer y_i bazı Poisson bileşenlerinden oluşuyorsa, EM içinde sıfıra eşit olan bir z_i değişkeni tanımlanır. Eğer $y_i = 0$ ise, z_i kaybolur. EM algoritması z_i 'nin her iterasyondaki beklenen değerlerini tahmin etmektedir. Bu parametre tahmini kullanılarak, iterasyonlar yakınsayana kadar devam eder.

ZIP modeli açıklayıcı değişkenleri kapsar şekilde yayılabilmektedir. Açıklayıcı değişkenler iki şekilde girilebilmektedir:

- $P(z_i = 1)$ 'i tahmin etmek için bir logit model içinde ve
- $E(y_i | z_i = 1)$ 'i tahmin etmek için bir Poisson kısmı içinde.

ZIP modeli iyi bulunmayabilir. Çünkü ağırlıklı olarak, sıfır olmayan bileşenler için varsayılan bir dağılım şekline dayanmaktadır. Bu durumda aşağıdakilerin bir karışımı olan model tercih edilebilir:

- Bir olasılığı ile sıfır olan bağımlı değişkenler ve
- Kısaltılmış Poisson dağılımından elde edilen bağımlı değişkenler:

$$f(y_i | y_i > 0) = \frac{\mu_i^{y_i} \exp(-\mu_i)}{y_i! (1 - \exp(-\mu_i))}, \quad y_i = 1, 2, \dots \quad (7.51)$$

* Bu modelin nasıl tahmin edildiği ile ilgili detaylar Lambert tarafından sağlanmıştır (Technometrics, 1992).

Buna zero-altered Poisson (ZAP) modeli denilmektedir. Aynı zamanda engel (hurdle) modeli de denilmektedir. Bu model EM kullanılmadan doğrudan tahmin edilebilir. Çünkü her gözlemin ait olduğu bileşenler bilindiğinden burada kayıp veri bulunmamaktadır.

7.6 Çokterimli Oluş Sayısı ile Belirlenen Veri için Poisson Regresyonun Kullanımı

Poisson modeli için sapma aşağıdaki gibidir:

$$G = 2 \sum_{i: y_i \neq 0} \left\{ y_i \log \frac{y_i}{\mu_i} - (y_i - \mu_i) \right\} + 2 \sum_{i: y_i = 0} \mu_i \quad (7.52)$$

Bu ifade, logaritmik doğrusal bir model sabit olmadan tahmin edildiğinde aşağıdaki gibi olmaktadır:

$$G = 2 \sum_{i=1}^N y_i \log \frac{y_i}{\mu_i} \quad (7.53)$$

İkinci ifade çokterimli bir model için sapma istatistiğidir. Bağımsız Poisson gözlemleri için dağılım aşağıdaki gibi olmaktadır:

$$y_i \sim \text{Poisson}(\mu_i), \quad i = 1, \dots, N$$

Bu aşağıdaki şekilde ayrıştırılabilmektedir:

- $\mu_+ = \sum_{i=1}^N \mu_i$ ortalama ile $y_+ = \sum_{i=1}^N y_i$ için bir Poisson dağılımı ve
- $y = (y_1, \dots, y_N)^T$ için bir çokterimli dağılım.

$$y | y_+ \sim \text{Çokterimli}(y_+, \pi)$$

Burada, $\pi = (\pi_1, \dots, \pi_N)^T$ ve $\pi_i = \mu_i / \mu_+$ 'dır.

$y = (y_1, \dots, y_N)^T$ örnek gözlemlerinin çokterimli bir model gösterdiği varsayalım. Çünkü y_+ sabittir. Eğer y_i 'ler, $\mu_i = \mu_+ \pi_i$ ortalama ile bağımsız Poisson değişkenleri gibi davranıyorsa, modelin sabit terimi içerdiğinden emin olunduğu sürece çokterimli model logaritmik bağ fonksiyonu ile Poisson regresyon olarak tahmin edilebilir. Logaritmalar alındığı zaman, μ_+ terimi sabit terimi yutmaktadır.

Eğer veri Poisson'dan daha çok çokterimli ise, Poisson regresyon modelinin sabit terimi

serbest bir parametre olarak yorumlanamayabilir. Fakat sabit normalleştirilerek, $\sum_{i=1}^N \pi_i = 1$ ya da $\sum_{i=1}^N \mu_i = y_+$ olması sağlanabilmektedir. Fakat β 'nın kalan elemanları μ_i 'lerden çok π_i 'ler için bir logaritmik doğrusal modelin katsayıları olarak yorumlanabilmektedir. Poisson regresyon yazılımı sabit terim için, bir tahmin ve bir standart hata çıktısı verecektir. Fakat bunlar ihmal edilmelidir. $\hat{\beta}$ için varyans-kovaryans matrisinin satır ve sütunları sabitin ihmal edilmesine uymaktadır.

7.7 Çokterimli Oluş Sayısı ile Belirlenen Veri için Logaritmik Doğrusal Modeller

$y = (y_1, \dots, y_N)^T$ 'nin çokterimli bir dağılım gösterdiği varsayalım:

$$y \sim \text{Çokterimli}(n, \pi)$$

Burada $n = y_+$ ve $\pi = (\pi_1, \dots, \pi_N)^T$ 'dir. Çok genel bir mantıkla, bilinen bir açıklayıcı değişkenler matrisi $X_{(N \times p)}$ ve bilinmeyen katsayılar vektörü $\beta_{(p \times 1)}$ için, logaritmik doğrusal bir model, aşağıdaki gibi olmaktadır:

$$\log \pi = X\beta \quad (7.54)$$

Diğer bir deyişle, π 'nin tekyönlü

$$S = \{\pi : \pi_j > 0, \pi_1 + \pi_2 + \dots + \pi_N = 1\}$$

içinde bulunmasını gerektirmesine ek olarak, logaritmik doğrusal model ayrıca $\log \pi$ 'nin X 'in sütunlarının oluşturduğu doğrusal uzay içinde bulunmasını gerektirmektedir.

Pek çok durumda, logaritmik doğrusal modeller, oluş sayısı ile belirlenen verileri ya da bir çapraz tablonun oluşturduğu frekansları tanımlamak için kullanılmaktadırlar. Mesela, n gözlemden oluşan bir örneğin, A , B ve C şeklinde kategorik değişkenlerinin gözlemlendiği varsayalım. Burada;

A , $1, 2, \dots, I$ şeklinde muhtemel değerler,

B , $1, 2, \dots, J$ şeklinde muhtemel değerler,

C , $1, 2, \dots, K$ şeklinde muhtemel değerler alsın.

Veri aşağıdaki gibi özetlenmektedir:

$$y = (y_{111}, y_{211}, \dots, y_{IJK})^T \quad (7.55)$$

Burada y_{ijk} , $A=i$, $B=j$ ve $C=k$ olduğunda birimlerin sayısını göstermektedir. Eğer değişkenler ortak bir topluluktan bağımsız olarak örneklenmiş ise; y , $n = y_{+++}$ indeksi ve $\pi = (\pi_{111}, \pi_{211}, \dots, \pi_{IJK})$ parametresi ile çokterimli olmaktadır. Böylece, π 'yi aşağıdaki şekilde modellemek mantıklı olmaktadır:

$$\log \pi = X\beta$$

Burada X , üç yönlü faktöriyel bir deney için, ANOVA'daki tasarım (dizayn) matrisine benzemektedir. X 'in sütunları A, B ve C için ana etkileri ve bunlar arasındaki etkileşimleri içermektedir. Örneğin, A için ana etkiler $I-1$ etkinin kodları ile aşağıdaki gibi tanımlanabilir:

Çizelge 7.1 Etkiler

	Etkiler			
i	A_1	A_2	...	A_{I-1}
1	1	0	...	0
2	0	1	...	0
3	0	0	...	0
$\vdots$	$\vdots$	$\vdots$		$\vdots$
$I-1$	0	0	...	1
I	-1	-1	...	-1

B için ana etkiler $I-1$ etkinin kodları ile aşağıdaki gibi tanımlanabilir; A ve B arasındaki etkileşim, A ve B için kodların arasından elde edilen sonuçlar olarak tanımlanmaktadır. Bu şekilde devam etmektedir.

Logaritmik doğrusal modeller ile faktöriyel ANOVA arasında önemli bir fark bulunmaktadır. Üç yönlü faktöriyel bir deneyde, dördüncü bir değişken olan bağımlı değişken üzerinde A, B

ve C faktörlerin birleşik etkileri analiz ediliyor olsun. Fakat logaritmik doğrusal bir modelde, sadece üç değişken (A , B ve C) bulunmaktadır. Bağımlı değişken logaritmik olasılıktır. Logaritmik doğrusal model kategorik değişkenler arasındaki ilişkileri tanımlamaktadır. Bunların başka bir değişken üzerindeki etkilerini tanımlamamaktadır.

Logaritmik doğrusal bir model içindeki $A \times B$ terimlerinin yorumu, ANOVA'daki $A \times B$ terimlerinin yorumundan oldukça farklı olmaktadır. ANOVA'da, $A \times B$ terimlerinin anlamlı olması, bağımlı değişken üzerindeki A 'nın etkisinin B 'nin seviyeleri karşısında değiştiğini göstermektedir. Logaritmik doğrusal bir modelde, $A \times B$ terimlerinin anlamlı olması kategorik değişkenler A ve B 'nin ilişkili olduğunu göstermektedir. Bu yüzden, logaritmik doğrusal modellerde, $A \times B$ terimlerine etkileşimden daha çok ilişki (association) denilmesi daha mantıklı olmaktadır. Çünkü etkileşim, iki değil, üç değişken arasındaki bir ilişkidir.

ANOVA'da, eğer A faktörünün bağımlı değişken üzerindeki etkisi ihmal edilebilir ise, A faktörü için ana etkinin silinmesi mantıklı olmaktadır. Ancak, logaritmik doğrusal modellerde, A için ana etkinin silinmesi mantıklı değildir. Çünkü bunun yapılması logaritmik olasılığın A 'nın seviyelerine karşı sabit olduğunu göstermektedir. Üç değişken için en yalın logaritmik doğrusal model A , B ve C için ana etkilerin bulunduğu ancak ilişkilerin (associations) bulunmadığı bir model olarak düşünülebilir. Bu A , B ve C arasında tamamen bağımsızlık olduğunu göstermektedir.

8. UYGULAMA

Bu çalışmada 1989 ile 1998 yılları arasında Türkiye’deki boşanma sayısı verileri bağımlı değişken olarak kullanılmıştır. Açıklayıcı değişkenlerden biri yaş, diğeri de boşanmaların gerçekleştiği yıllardır. Boşanma sayıları on tane yaş grubu kategorisine göre sınıflandırılmıştır. Bu yaş grupları 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59 ve 60 ve üzeri şeklindedir.

Yaş grubu değişkeniyle ilgili yapılan çalışmalar sonucunda 15-19 ve 20-24 yaş grupları arasındaki deneklerin yaşlarının küçük olmasına bağlı olarak ekonomik özgürlüklerinin olmayışı, bir mesleği olsa bile genç yaşta mesleğinde kariyer sahibi olamaması, tecrübesizlik ve aile ya da toplum tarafından yanlış yönlendirmeler sonucu çabuk boşanabilecekleri ve bu yüzden genel dağılımı bozabilecekleri düşünülerek, daha iyi bir sonuca ulaşabilmek için, çalışmalar bu iki grup çıkarılarak tekrar yapılmıştır(Boşanma Verileri2). Çıkarılan iki yaş grubuna ek olarak, çalışmalar bir de 60 ve üzeri yaş grubu çıkarılarak denenmiştir(Boşanma Verileri3). 60 ve üzeri yaş grubunun çıkarılmasının sebebi ise bu yaş grubunun seçilen diğer yaş gruplarıyla arasında oldukça büyük bir fark bulunmasıyla birlikte, yaş ilerledikçe insan davranışlarının ve kişilerin sağlık durumlarının da farklılıklar gösterebileceğidir. Bu yaş grupları çıkarıldığında yapılan tahminlerden daha iyi sonuçlar elde edildiği görülmüştür.

Bu çalışmada yıllar X ve yaş grupları Y ile ifade edilmiştir. Boşanma Verileri için yaş grubu değişkeni:

$$Y_k = \begin{cases} 1, k = i & \text{ise} \\ 0, d.d. \end{cases} \quad k = 1, 2, \dots, 9$$

şeklinde alınmıştır. İlk iki yaş grubunun çıkarıldığı Boşanma Verileri2 için ise yaş grubu değişkeni,

$$Y_k = \begin{cases} 1, k = i & \text{ise} \\ 0, d.d. \end{cases} \quad k = 1, 2, \dots, 7$$

ve buna ilave olarak 60 ve üzeri yaş grubunun da çıkarıldığı Boşanma Verileri3 için ise,

$$Y_k = \begin{cases} 1, k = i & \text{ise} \\ 0, d.d. \end{cases} \quad k = 1, 2, \dots, 6$$

olarak alınmıştır.

Boşanma Verileri, Boşanma Verileri2 ve Boşanma Verileri3 için hesaplanan modellerle ilgili sapma değerleri tablo şeklinde incelenmiştir.

8.1 Boşanma Verileri ile İlgili Elde Edilen Sonuçlar

Boşanma Verileri için hesaplanan değerler Çizelge 8.1’de verilmiştir.

Çizelge 8.1 Boşanma Verileri ile İlgili Poisson ANOVA Tablosu

Sıra	Model	Parametre	Sapma	Serbestlik Derecesi
1	α	1	291914,1000	99
2	$\alpha + \gamma.X$	2	288499,9000	98
3	$\alpha + \sum_{k=1}^9 \beta_k Y_k$	10	5453,9360	90
4	$\alpha + \gamma.X + \sum_{k=1}^9 \beta_k Y_k$	11	2039,7480	89
5*	$\alpha + \gamma.X + \sum_{k=1}^9 \beta_k Y_k + \sum_{k=1}^9 \theta_k XY_k$	20	767,2275	80

*Bazı katsayıları anlamsız olan modeller

Boşanma Verileri için sapma değerleri en küçük olan üçüncü model ve dördüncü model karşılaştırıldığında etkileşim terimi olan modelin daha iyi sonuç verdiği görülmektedir.

$$H_0 : \gamma = 0$$

$$H_1 : \gamma \neq 0$$

$$\Delta G = G_3 - G_4 = 5453,936 - 2039,748$$

$$df_3 - df_4 = 90 - 89 = 1$$

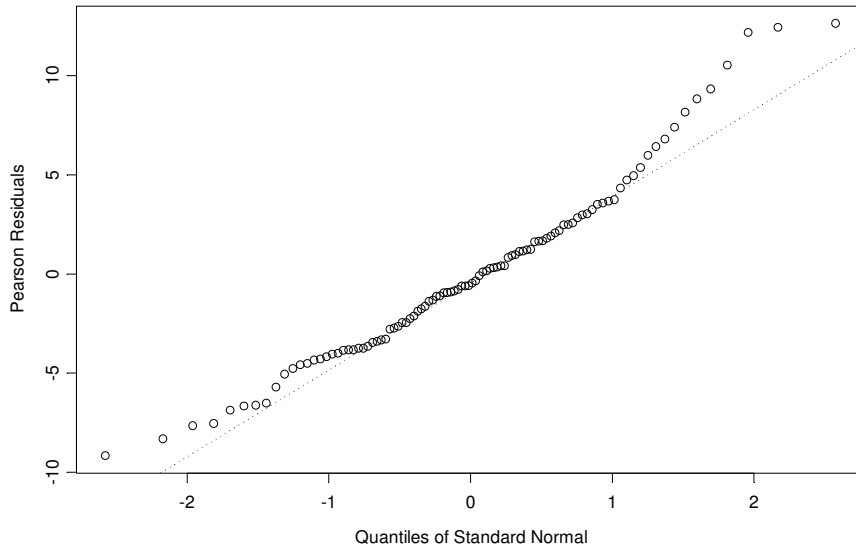
$$= 3414,188 \sim \chi^2_{(1;0,9995)}$$

$$\chi^2_{(1;0,9995)} = 12,116$$

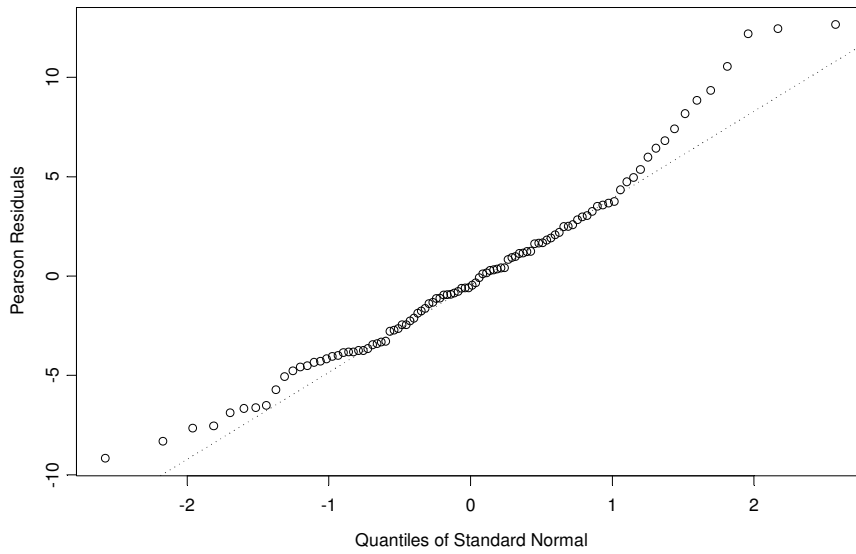
H_0 reddedilir. 4.model tercih edilir. γ sıfırdan farklıdır.

Yapılan işlemlerin ilk adımında dördüncü model Boşanma Verilerini açıklayan model olarak

tercih edilecektir. Çünkü dördüncü model için Pearson kalıntılarının dağılım grafiği ve tahmin edilen değerlerle elimizdeki verinin birbirine uyumu iyi sonuç vermektedir. Ancak bu noktada seçilen dördüncü model daha sonra Boşanma Verileri2 ve Boşanma Verileri3'ten seçilen modellerle de karşılaştırılacaktır.



Şekil 8.1 Boşanma Verileri'nde dördüncü model için Pearson kalıntılarının dağılımı



Şekil 8.2 Boşanma Verileri'nde dördüncü model için tahmin edilen değerlerle boşanma sayılarının grafiği

8.2 Boşanma Verileri2 ile İlgili Elde Edilen Sonuçlar

Boşanma Verileri2 ile ilgili elde edilen sonuçlar Çizelge 8.2’de bulunmaktadır.

Çizelge 8.2 Boşanma Verileri2 (Boşanma Verileri’nden 15-19 ve 20-24 Yaş Grupları çıkarılarak elde edilmiştir.) ile ilgili Poisson ANOVA Tablosu

Sıra	Model	Parametre	Sapma	Serbestlik Derecesi
6	α	1	243243,8000	79
7	$\alpha + \gamma.X$	2	239771,7000	78
8	$\alpha + \sum_{k=1}^7 \beta_k Y_k$	8	5172,7860	72
9	$\alpha + \gamma.X + \sum_{k=1}^7 \beta_k Y_k$	9	1700,7280	71
10	$\alpha + \gamma.X + \sum_{k=1}^7 \beta_k Y_k + \sum_{k=1}^7 \theta_k XY_k$	16	674,0899	64

Boşanma Verileri2 için, Çizelge 8.2’de de görüldüğü gibi en düşük sapma değerine sahip olan iki model bulunmaktadır. Bunlar dokuzuncu ve onuncu modellerdir. Bu iki model, aşağıda sapma değerlerine göre karşılaştırılmışlardır.

$$H_0 : \theta_k = 0$$

$$H_1 : \theta_k \neq 0$$

$$\Delta G = G_9 - G_{10} = 1700,728 - 674,0899$$

$$df_9 - df_{10} = 71 - 64 = 7$$

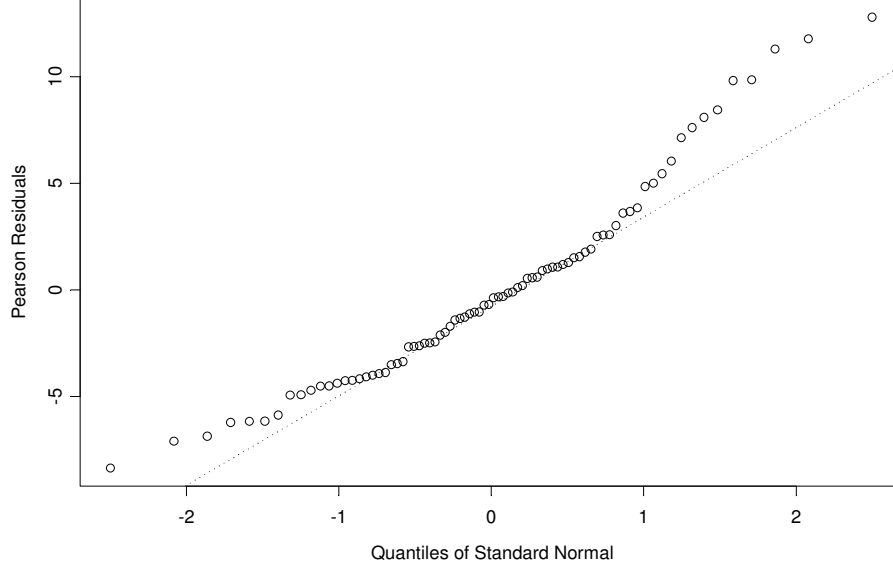
$$= 1026,6381 \sim \chi^2_{(7;0,9995)}$$

$$\chi^2_{(7;0,9995)} = 26,018$$

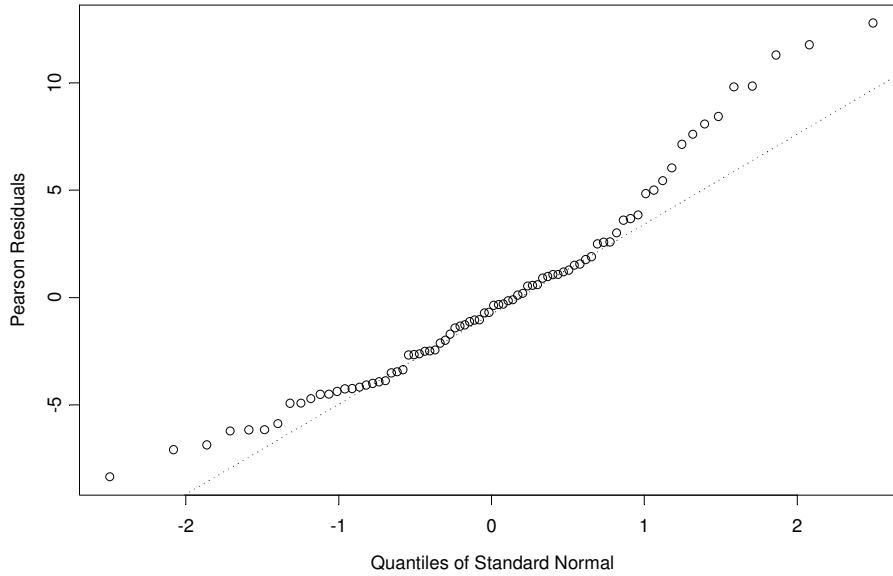
H_0 reddedilir. 10.model tercih edilir. θ_k sıfırdan farklıdır.

Ancak dokuzuncu model ile Boşanma Verileri2 için seçilen onuncu modelin Pearson kalıntıların dağılımı incelenirse (Şekil 8.3 ve Şekil 8.5), onuncu modelin uyumunun iyi olmasına rağmen Pearson kalıntıların dağılım grafiğinin normal olmadığı görülmektedir. Bu

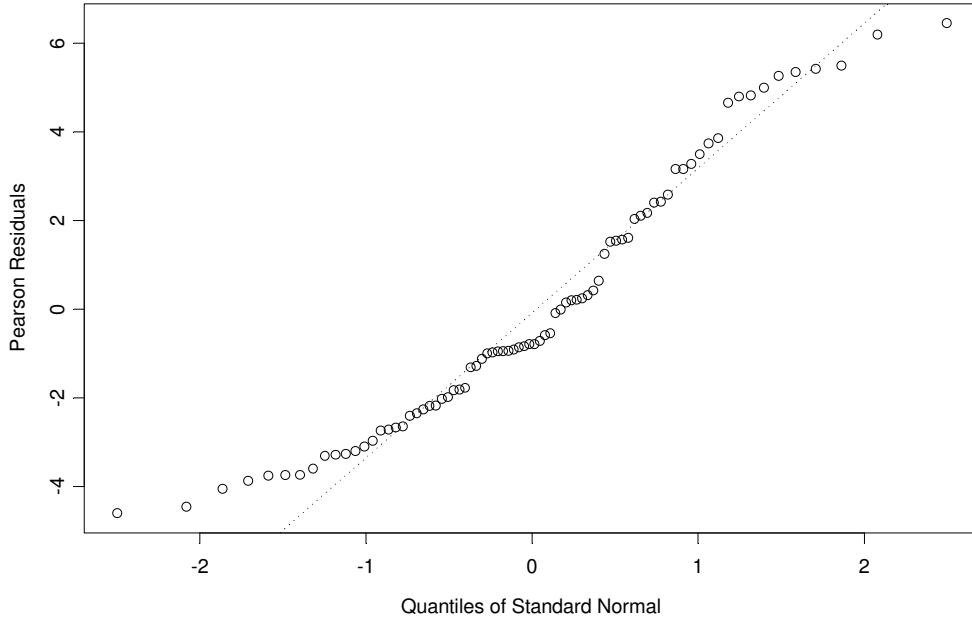
yüzden Boşanma Verileri2'yi açıklayan model olarak dokuzuncu model tercih edilmiştir. Ancak bu model de seçilen diğer modellerle karşılaştırılacaktır.



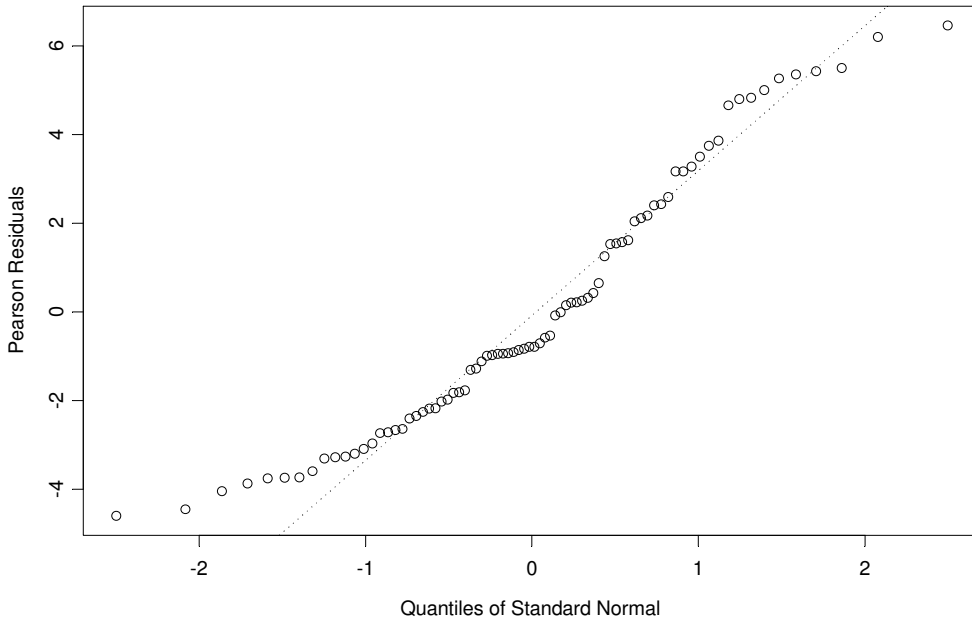
Şekil 8.3 Boşanma Verileri2'de dokuzuncu model için Pearson kalıntılarının dağılımı



Şekil 8.4 Boşanma Verileri2'de dokuzuncu model için tahmin edilen değerlerle boşanma sayılarının grafiği



Şekil 8.5 Boşanma Verileri2’de onuncu model için Pearson kalıntılarının dağılımı



Şekil 8.6 Boşanma Verileri2’de onuncu model için tahmin edilen değerlerle boşanma sayılarının grafiği

8.3 Boşanma Verileri3 ile İlgili Elde Edilen Sonuçlar

Boşanma Verileri3 ile ilgili sonuçlar Çizelge 8.3'te bulunmaktadır.

Çizelge 8.3 Boşanma Verileri3 (Boşanma Verileri'nden 15-19, 20-24 ve 60 ve üzeri Yaş Grupları çıkarılarak elde edilmiştir.) ile ilgili Poisson ANOVA Tablosu

Sıra	Model	Parametre	Sapma	Serbestlik Derecesi
11	α	1	205938,3000	69
12	$\alpha + \gamma.X$	2	202387,3000	68
13	$\alpha + \sum_{k=1}^6 \beta_k Y_k$	7	5107,6320	63
14	$\alpha + \gamma.X + \sum_{k=1}^6 \beta_k Y_k$	8	1556,6870	62
15	$\alpha + \gamma.X + \sum_{k=1}^6 \beta_k Y_k + \sum_{k=1}^6 \theta_k XY_k$	14	617,9975	56

Boşanma Verileri3 için, Çizelge 8.3'de de görüldüğü gibi en düşük sapma değerine sahip olan iki model bulunmaktadır. Bunlar on dördüncü ve on beşinci modellerdir. Bu iki model için, aşağıda uyum testi yapılmıştır.

$$H_0 : \theta_k = 0$$

$$H_1 : \theta_k \neq 0$$

$$\Delta G = G_{14} - G_{15} = 1556,687 - 617,9975$$

$$df_{14} - df_{15} = 62 - 56 = 6$$

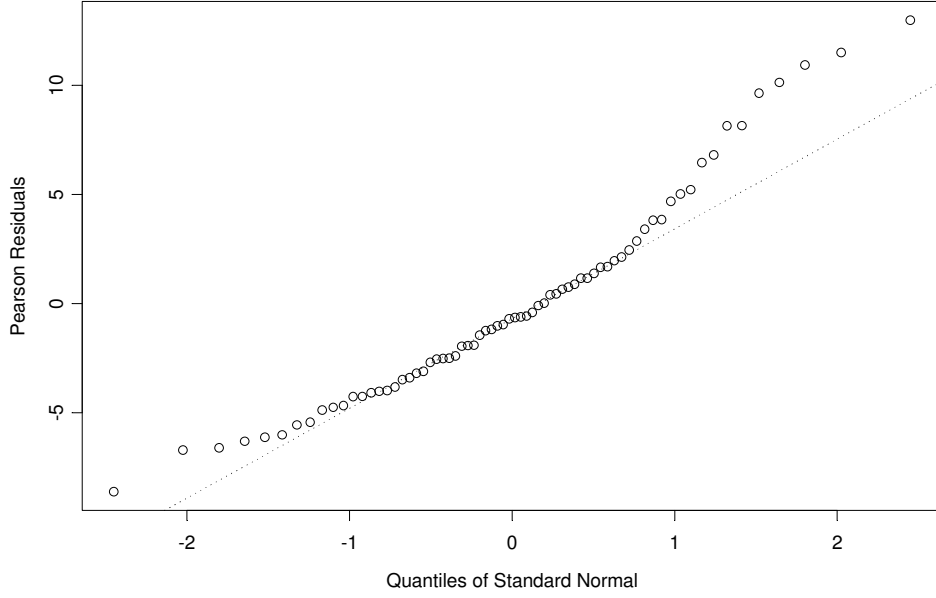
$$= 938,6895 \sim \chi^2_{(6;0,9995)}$$

$$\chi^2_{(6;0,9995)} = 24,103$$

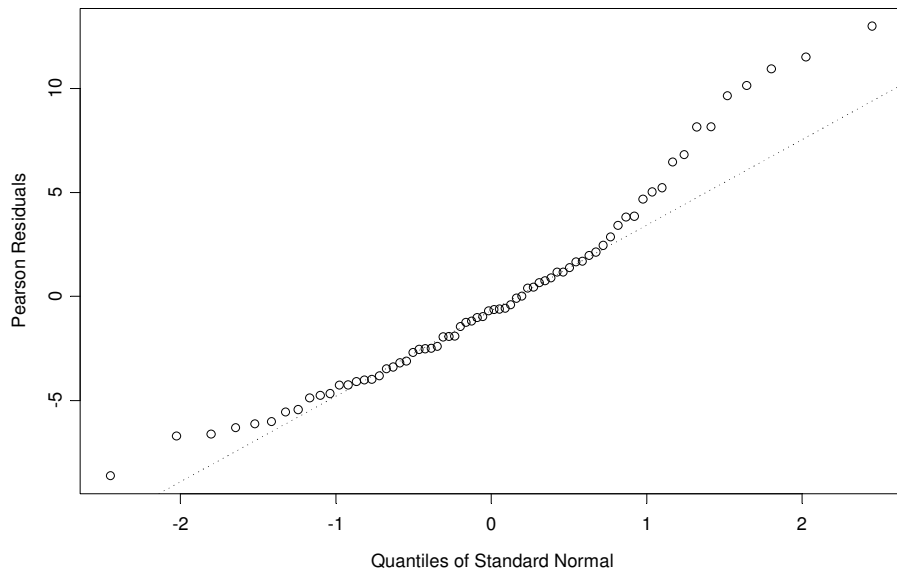
H_0 reddedilir. 15.model tercih edilir. θ_k sıfırdan farklıdır.

Ancak on dördüncü model ile Boşanma Verileri3 için seçilen on beşinci modelin Pearson kalıntıların dağılımı incelenirse (Şekil 8.7 ve Şekil 8.9), on beşinci modelin uyumunun iyi olmasına rağmen Pearson kalıntıların dağılım grafiğinin normal olmadığı görülmektedir. Bu

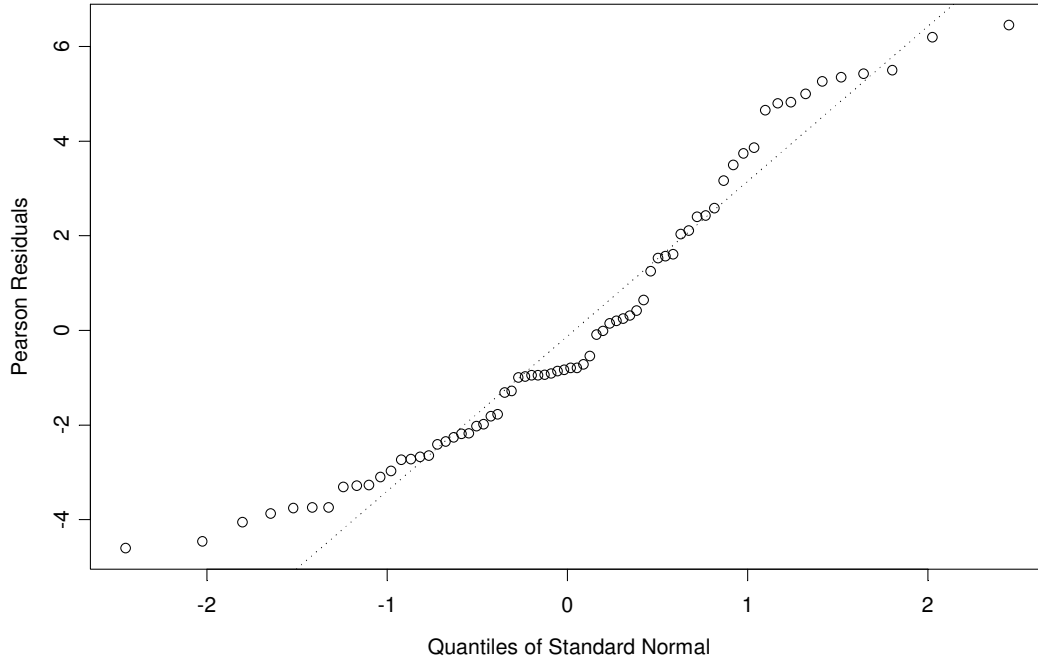
yüzden Boşanma Verileri3'ü açıklayan model olarak on dördüncü model tercih edilmiştir. Ancak bu model de seçilen diğer modellerle karşılaştırılacaktır.



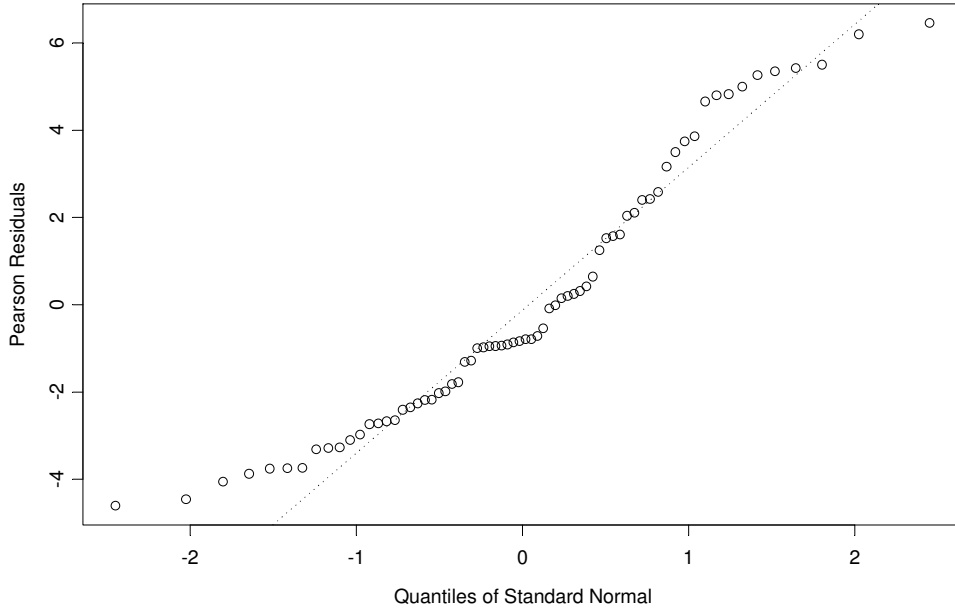
Şekil 8.7 Boşanma Verileri3'de on dördüncü model için Pearson kalıntılarının dağılımı



Şekil 8.8 Boşanma Verileri3'de on dördüncü model için tahmin edilen değerlerle boşanma sayılarının grafiği



Şekil 8.9 Boşanma Verileri3’de on beşinci model için Pearson kalıntılarının dağılımı



Şekil 8.10 Boşanma Verileri3’de on beşinci model için tahmin edilen değerlerle boşanma sayılarının grafiği

Boşanma Verileri'nden dördüncü model, Boşanma Verileri2'den dokuzuncu model ve Boşanma Verileri3'den de on dördüncü model seçilmiştir. Bu üç model de birbiriyle karşılaştırıldığında, Pearson kalıntılarının dağılımının normale en yakın ve en iyi uyumu sağlayan modelin on dördüncü model olduğu görülmektedir. Bu yüzden boşanma sayılarını en iyi açıklayan model olarak on dördüncü model tercih edilmiştir.

8.4 14.Modelin İncelenmesi

Bu modelde açıklayıcı değişken olarak boşanmaların gerçekleştiği yıl ve yaş grubu alınmıştır. Yaş grubu yedi kategoriye ayrılmış ve bunları göstermek için altı gölge değişken kullanılmıştır. On Dördüncü model ile ilgili bilgisayar çıktısı Çizelge 8.4'te bulunmaktadır.

Çizelge 8.4 On dördüncü model ile ilgili S-PLUS 2000 programından elde edilen çıktı

```
*** Generalized Linear Model ***

Call: glm(formula = Bosanma.Sayisi ~ Yil + Yas, family = poisson(link = log), data =
hande3, na.action = na.exclude, control = list(epsilon = 0.0001, maxit = 50, trace = F))
Deviance Residuals:
Min       1Q   Median       3Q      Max
-8.72756  -3.489572 -0.6720646  2.075031 12.27246

Coefficients:
Value      Std. Error    t value
(Intercept) -53.69280598 1.0446799229  -51.39642
Yil      0.03118533 0.0005239746   59.51688
Yas1    -0.06468870 0.0020467460  -31.60563
Yas2    -0.12634224 0.0013500288  -93.58485
Yas3    -0.16508304 0.0011613557 -142.14685
Yas4    -0.19673521 0.0011468562 -171.54305
Yas5    -0.21261803 0.0012004092 -177.12129
Yas6    -0.20942124 0.0012493732 -167.62103

(Dispersion Parameter for Poisson family taken to be 1 )

Null Deviance: 205938.3 on 69 degrees of freedom

Residual Deviance: 1556.687 on 62 degrees of freedom

Number of Fisher Scoring Iterations: 3

Analysis of Deviance Table

Poisson model

Response: Bosanma.Sayisi

Terms added sequentially (first to last)
Df Deviance Resid. Df Resid. Dev
NULL                69    205938.3
Yil      1    3550.9         68    202387.3
Yas     6  200830.7         62     1556.7
```

Elde edilen çıktıda da görüldüğü gibi katsayıların hepsi anlamlıdır. Boşanma sayılarının

ortalamasının doğal logaritmasını; Yıl 0,0312; Yaş Grubu1 -0,0647; Yaş Grubu2 -0,126; Yaş Grubu3 -0,165; Yaş Grubu4 -0,197; Yaş Grubu5 -0,213 ve Yaş Grubu6 -0,209 katsayılarıyla etkilemektedir. Yaş Grubu1'den Yaş Grubu6'ya doğru katsayılar giderek azalmaktadır. Buradan da yaş grubu ilerledikçe boşanma sayısının azaldığı sonucuna varılmaktadır. Bu duruma sebep olarak, düzenli hayatın insan üzerinde alışkanlık yapması ve bunu bırakmak istememesi gösterilebilir. Yıl değişkeni için elde edilen katsayı da 0,0312'dir. Buradan elde edilen sonuç da insanların her geçen yıl az bir farkla da olsa daha çok boşandıklarıdır. Bunun sebebi ise zamanla gelişen teknoloji, hayat şartları ve ekonomik özgürlük ile insanların bir başka kişiyle yaşamaktan kaynaklanan sorunlara dayanamayıp, yalnız ve sorunsuz yaşamayı tercih etmeleri olabilir.

Ayrıca on dördüncü model için hesaplanan tahminlerin korelasyon matrisi Çizelge 8.5'te bulunmaktadır.

Çizelge 8.5 On dördüncü model için hesaplanan tahminlerin korelasyon matrisi

Correlation of Coefficients:

	(Intercept)	Yil	Yas1	Yas2	Yas3	Yas4
Yil	-0.9999981					
Yas1	0.0000362	0.0000000				
Yas2	0.0001295	0.0000000	-0.0326457			
Yas3	0.0002562	0.0000000	-0.0189747	-0.0679568		
Yas4	0.0004314	0.0000000	-0.0115287	-0.0412896	-0.0816881	
Yas5	0.0006329	0.0000000	-0.0073429	-0.0262984	-0.0520292	-0.0876050
Yas6	0.0008127	0.0000000	-0.0050394	-0.0180484	-0.0357072	-0.0601226
	Yas5					
Yil						
Yas1						
Yas2						
Yas3						
Yas4						
Yas5						
Yas6						-0.0881995

8.4.1 14.Modeldeki Aşırı Yayılımın İncelenmesi

Aşırı yayılım olup olmadığını anlayabilmek için $Var(y_i) = \sigma^2 \mu_i$ olduğu varsayılmaktadır ve buradaki tartı parametresi tahmin edilmektedir. Eğer bu değer sıfırdan büyük çıkarsa aşırı yayılım var demektir. Bu durum da bazı değerlerin tartı parametresine bölünmesiyle düzeltilmektedir.

$$X^2 = \sum_{i=1}^N \frac{(y_i - \mu_i)^2}{\mu_i} \quad \text{Pearson uyum iyiliği istatistiği}$$

$X^2 = 1587,66$ olarak elde edilmiştir.

$$\hat{\sigma}^2 = \frac{X^2}{(n-p)} = \frac{1587,66}{62} = 25,61$$

$$\hat{\sigma} = 5,06$$

Yukarıda da görüldüğü gibi tartı parametresi birden büyük elde edilmiştir. Bu durum aşırı yayılımın olduğunu göstermektedir. Bu yüzden aşağıdaki düzeltmeler yapılmalıdır:

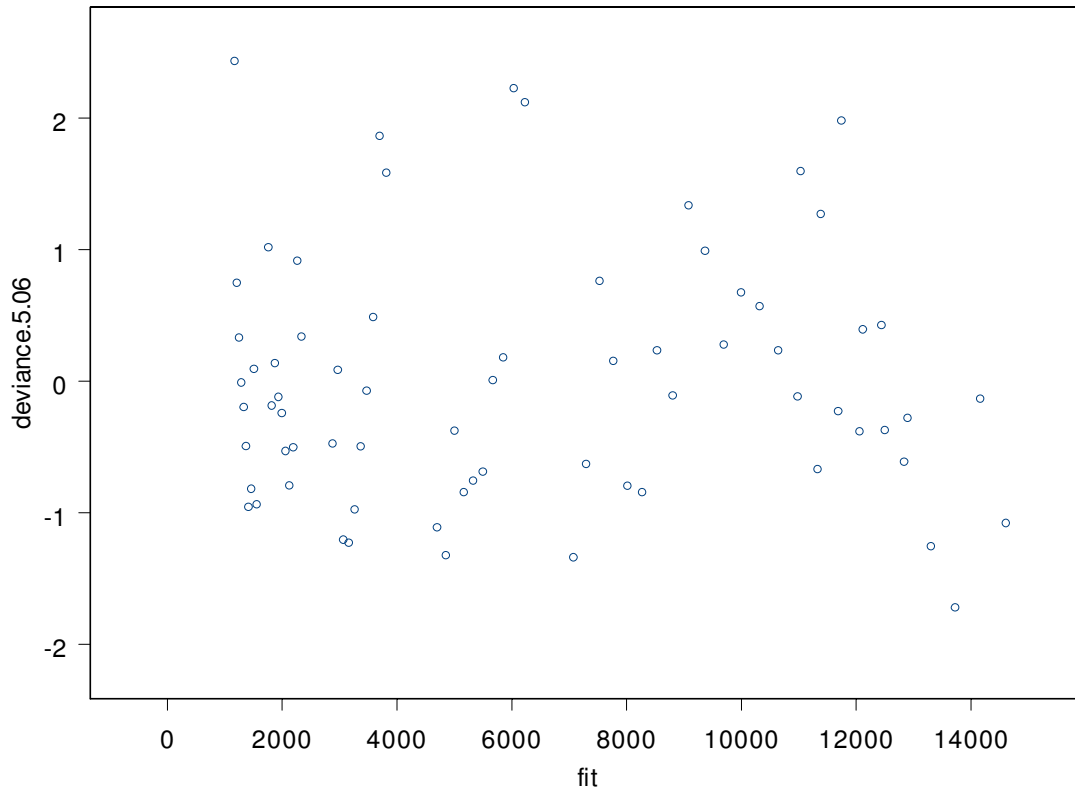
$$X^2 = 1587,66$$

$$X^2/\hat{\sigma}^2 = 1587,66/25,61 = 61,99$$

$$G^2 = 1556,687$$

$$G^2/\hat{\sigma}^2 = 1556,687/25,61 = 60,78$$

Aynı zamanda Pearson kalıntıları ve sapma kalıntıları da $\hat{\sigma} = 5,06$ 'ya bölünerek düzeltilmelidir. Bu işlem yapıldıktan sonra, tahmin değerleriyle düzeltilmiş kalıntıların grafiği incelenmiştir (Şekil 8.11). Şekil 8.11'de de görüldüğü gibi kalıntılar herhangi bir modele göre dağılmamaktadır.



Şekil 8.11 BoşanmaVerileri3'de on dördüncü model için tahmin değerleriyle düzeltilmiş sapma kalıntılarının grafiği

9. SONUÇLAR

Poisson regresyon, bağımlı değişken oluş sayısı (count) ile belirtilen bir veri olduğunda, yani belirli bir zaman ya da yerde olan olayların sayısı olduğunda kullanılmaktadır. Burada regresyon fonksiyonu, bağımsız değişkenleri bağımlı değişkenin beklenen değerine bağlayan bir fonksiyondur. Bu fonksiyon, genellikle logaritmik doğrusaldır. Poisson regresyonda uygulamalar çoğu zaman belli kategorilere ayrılan oluş sayısı ile belirtilen veriler ile yapılmaktadır.

Uygulamada genellikle Poisson regresyon modelindeki ortalama ve varyansın eşit olması sağlanamamaktadır. Daha çok varyansın ortalamadan büyük olduğu aşırı yayılım durumu ile karşılaşmaktadır. Böyle bir durum söz konusu olduğunda tartı parametresi kullanılmaktadır. Aşırı yayılım olduğu zaman, negatif binom regresyon uygulanmalı ya da tartı parametresi ile test istatistikleri ve kalıntılar düzeltilmelidir.

Poisson regresyonda, parametre tahminleri IRWLS yöntemi ve maksimum olabilirlik yöntemi kullanılarak yapılmaktadır. Çalışmada; Poisson regresyonda parametre tahminleri, uyum ölçüleri ve kalıntıların hesaplanmasıyla ilgili genel bilgiler bulunmaktadır.

Bu çalışmadaki uygulamada yapılan parametre tahminleri, uyum ölçüleri ve kalıntıların hesaplanmasında S-PLUS 2000 programı kullanılmıştır.

Yapılan literatür çalışmasında Poisson regresyonun daha çok sağlık alanında yapılan uygulamalarda kullanıldığı görülmüş, ancak burada boşanma verileriyle ilgili bir uygulama yapılmıştır. Uygulamada bağımlı değişken olarak 1989 ile 1998 yılları arasında Türkiye'deki boşanma sayısı verileri kullanılmıştır. Bağımsız değişkenler yıl ve yaştır. Yaş grubu değişkeni on tane kategoriye ayrılarak sınıflandırılmıştır. Bu yaş grupları 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59 ve 60 ve üzeri şeklindedir. Yapılan çalışmada boşanma verilerini en iyi açıklayan Poisson regresyon modeline ulaşmak amaçlanmıştır.

Boşanma verileriyle ilgili yapılan çalışmalar sonucunda yaş grubunun modeli oldukça etkilediği görülmektedir. Bu sonuca varılmasının sebebi, modele yaş grubu değişkeni eklenmediğinde hesaplanan sapma değerlerinin çok büyük olmasıdır.

Yaş grubu değişkeniyle ilgili yapılan çalışmalar sonucunda 15-19 ve 20-24 yaş grupları arasındaki deneklerin yaşlarının küçük olmasına bağlı olarak ekonomik özgürlüklerinin olmayışı, bir mesleği olsa bile genç yaşta mesleğinde kariyer sahibi olamaması, tecrübesizlik ve aile ya da toplum tarafından yanlış yönlendirmeler sonucu çabuk boşanabilecekleri ve bu

yüzden genel dağılımı bozabilecekleri düşünülerek, daha iyi bir sonuca ulaşabilmek için, çalışmalar bu iki grup çıkarılarak tekrar yapılmıştır(Boşanma Verileri2). Çıkarılan iki yaş grubuna ek olarak, çalışmalar bir de 60 ve üzeri yaş grubu çıkarılarak denenmiştir(Boşanma Verileri3). 60 ve üzeri yaş grubunun çıkarılmasının sebebi ise bu yaş grubunun seçilen diğer yaş gruplarıyla arasında oldukça büyük bir fark bulunmasıyla birlikte, yaş ilerledikçe insan davranışlarının ve kişilerin sağlık durumlarının da farklılıklar gösterebileceğidir. Bu yaş grupları çıkarıldığında yapılan tahminlerden daha iyi sonuçlar elde edildiği görülmüştür. Yapılan bu uygulamada pek çok model üzerinde çalışılmıştır. Tüm değişkenler modele dahil edildiğinde sapma (deviance) değerinin en küçük olduğu görülmüştür. Fakat bu modellerin kalıntıları ve veriye uyumları incelendiğinde on dördüncü modelin en iyi sonucu verdiği görülmüştür. Bu yüzden boşanma sayılarını en iyi açıklayan model olarak on dördüncü model tercih edilmiştir.

On dördüncü model incelendiğinde Yaş Grubu1'den Yaş Grubu6'ya doğru katsayıların giderek azaldığı görülmektedir. Buradan da yaş grubu ilerledikçe boşanma sayısının azaldığı sonucu elde edilmektedir. Bu duruma sebep olarak, düzenli hayatın insan üzerinde alışkanlık yapması ve bunu bırakmak istememesi gösterilebilir. Yıl değişkeni için elde edilen katsayı da 0,0312'dir. Buradan elde edilen sonuç da insanların her geçen yıl az bir farkla da olsa daha çok boşandıklarıdır. Bunun sebebi ise zamanla gelişen teknoloji, hayat şartları ve ekonomik özgürlük ile insanların bir başka kişiyle yaşamaktan kaynaklanan sorunlara dayanamayıp, yalnız ve sorunsuz yaşamayı tercih etmeleri olabilir.

KAYNAKLAR

Aiken, L.S., Cohen, J., Cohen, P. ve West, S.G., (2003), Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, Lawrence Erlbaum Associates, New Jersey.

Christensen, R., (1997), Log-Linear Models and Logistic Regression, Springer Text in Statistics, New York.

Jennrich, R.I., (1995), An Introduction to Computational Statistics-Regression Analysis, Prentice-Hall, New Jersey.

Kleinbaum, D.G., Kupper, L.L., Muller, K.E. ve Nizam, A., (1998), Applied Regression Analysis and Other Multivariable Methods, Duxbury Pres, Pacific Grove.

Kutner, M.H., Nachtsheim, C.J. ve Neter, J., (2004), Applied Linear Regression Models, McGraw-Hill, Boston.

Lee, A.J. ve Seber, G.A.F., (2003), Linear Regression Analysis, John Wiley & Sons, New Jersey.

Lloyd, C.J., (1999), Statistical Analysis of Categorical Data, John Wiley & Sons, New York.

Özmen, İ., (1998), Poisson Regresyon Çözümleme Teknikleri, Doktora Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü.

Weisberg, S., (1985), Applied Linear Regression, John Wiley & Sons, New York.

INTERNET KAYNAKLARI

www.stat.psu.edu

EKLER

- Ek 1 Binom Dağılımının Poisson Dağılımına Yaklaşımı
- Ek 2 Poisson Dağılımına Normal Yaklaşım
- Ek 3 Poisson Süreci
- Ek 4 Yalın Doğrusal Regresyon Analizine Matris Yaklaşımı
- Ek 5 Boşanma Verileri'nde İkinci Model ile İlgili Grafikler
- Ek 6 Boşanma Verileri'nde Üçüncü Model ile İlgili Grafikler
- Ek 7 Boşanma Verileri'nde Dördüncü Model ile İlgili Grafik
- Ek 8 Boşanma Verileri'nde Beşinci Model ile İlgili Grafikler
- Ek 9 Boşanma Verileri2'de Yedinci Model ile İlgili Grafikler
- Ek 10 Boşanma Verileri2'de Sekizinci Model ile İlgili Grafikler
- Ek 11 Boşanma Verileri2'de Dokuzuncu Model ile İlgili Grafik
- Ek 12 Boşanma Verileri2'de Onuncu Model ile İlgili Grafik
- Ek 13 Boşanma Verileri3'de On İkinci Model ile İlgili Grafikler
- Ek 14 Boşanma Verileri3'de On Üçüncü Model ile İlgili Grafikler
- Ek 15 Boşanma Verileri3'de On Dördüncü Model ile İlgili Grafik
- Ek 16 Boşanma Verileri3'de On Beşinci Model ile İlgili Grafikler
- Ek 17 İkinci Model için Elde Edilen Bilgisayar Çıktısı
- Ek 18 Üçüncü Model için Elde Edilen Bilgisayar Çıktısı
- Ek 19 Dördüncü Model için Elde Edilen Bilgisayar Çıktısı
- Ek 20 Beşinci Model için Elde Edilen Bilgisayar Çıktısı
- Ek 21 Yedinci Model için Elde Edilen Bilgisayar Çıktısı
- Ek 22 Sekizinci Model için Elde Edilen Bilgisayar Çıktısı
- Ek 23 Dokuzuncu Model için Elde Edilen Bilgisayar Çıktısı
- Ek 24 Onuncu Model için Elde Edilen Bilgisayar Çıktısı
- Ek 25 On İkinci Model için Elde Edilen Bilgisayar Çıktısı
- Ek 26 On Üçüncü Model için Elde Edilen Bilgisayar Çıktısı
- Ek 27 On Beşinci Model için Elde Edilen Bilgisayar Çıktısı
- Ek 28 Boşanma Verileri
- Ek 29 Boşanma Verileri2
- Ek 30 Boşanma Verileri3

Ek 1 Binom Dağılımının Poisson Dağılımına Yaklaşımı

$$f(x) = \binom{n}{x} p^x q^{n-x}, \quad E(x) = np, \quad V(x) = npq, \quad n \rightarrow \infty, \quad p \rightarrow 0, \quad np = \lambda \text{ sabit bir sayı}$$

$$\begin{aligned} \lim_{n \rightarrow \infty} B(n, p) &= \lim_{n \rightarrow \infty} \left[\binom{n}{x} p^x q^{n-x} \right] = \lim_{n \rightarrow \infty} \left[\frac{n!}{x!(n-x)!} \cdot \left(\frac{\lambda}{n} \right)^x \left(1 - \frac{\lambda}{n} \right)^{n-x} \right] \\ &= \lim_{n \rightarrow \infty} \left[\frac{n(n-1) \dots (n-x+1)}{x!} \cdot \frac{\lambda^x}{n^x} \cdot \left(1 - \frac{\lambda}{n} \right)^n \cdot \left(1 - \frac{\lambda}{n} \right)^{-x} \right] \\ &= \lim_{n \rightarrow \infty} \left[\frac{\underbrace{n(n-1) \dots (n-x+1)}_{\substack{x \text{ tane } e}}}{x!} \cdot \frac{\lambda^x}{n^x} \cdot \left(1 - \frac{\lambda}{n} \right)^n \cdot \left(1 - \frac{\lambda}{n} \right)^{-x} \right] \\ &= \lim_{n \rightarrow \infty} \left[\frac{1 \left(1 - \frac{1}{n} \right) \left(1 - \frac{2}{n} \right) \dots \left(1 - \frac{x-1}{n} \right)}{\left(1 - \frac{\lambda}{n} \right)^x} \cdot \frac{\lambda^x}{x!} \cdot \left(1 - \frac{\lambda}{n} \right)^n \right] \\ &= \frac{\lambda^x}{x!} \cdot \lim_{n \rightarrow \infty} \left[\frac{1 \left(1 - \frac{1}{n} \right) \dots \left(1 - \frac{x-1}{n} \right)}{\left(1 - \frac{\lambda}{n} \right)^x} \cdot \left(1 - \frac{\lambda}{n} \right)^n \right] \\ &= \frac{\lambda^x}{x!} \cdot \frac{1}{1} \cdot e^{-\lambda} = \frac{e^{-\lambda} \cdot \lambda^x}{x!} \end{aligned}$$

Not: $\sum_{x=0}^{\infty} \frac{\mu^x}{x!} = e^{\mu}$

Ek 2 Poisson Dağılımına Normal Yaklaşım

X rassal değişkeni bir Poisson değişkeni ise,

$$E(X) = V(X) = \lambda$$

olur.

Teorem: X , Poisson dağılımına sahip bir rassal değişken ise, λ 'nın çok büyük olması durumunda, standart hale getirilmiş Poisson rassal değişkeni Z , $E(X) = V(X) = \lambda$ olmak üzere, sıfır ortalama ve bir varyansı ile yaklaşık bir normal dağılıma sahiptir.

$$Z = \frac{X - E(X)}{\sqrt{V(X)}}$$

$$Z = \frac{X - \lambda}{\sqrt{\lambda}} \sim N(0,1)$$

Bir başka ifade ile Poisson dağılımından yararlanılarak elde edilecek olan olasılıklar, λ 'nın yeterli derecede büyük olması durumunda aşağıdaki eşitliklerle de bulunabilir:

$$P(a \leq X \leq b) \cong P\left(\frac{a - \lambda}{\sqrt{\lambda}} \leq Z \leq \frac{b - \lambda}{\sqrt{\lambda}}\right)$$

Eğer süreklilik düzeltmesi kullanılırsa, bu olasılık,

$$P(a \leq X \leq b) \cong P\left(\frac{a - 0,5 - \lambda}{\sqrt{\lambda}} \leq Z \leq \frac{b + 0,5 - \lambda}{\sqrt{\lambda}}\right)$$

olur.

Ek 3 Poisson Süreci

X rastlantı değişkeni birim zamanda bir olayın gerçekleşme sayısı (başarı sayısı) ve başarı olasılığı söz konusu zaman dilimi ile doğru orantılı olsun. Mesela geçen zaman dilimi Δt olsun. λ da söz konusu orantısallık sabiti olsun. Öyle ki X 'in sıfırdan farklı bir değer alma (ya da söz konusu olayın en azından bir kere gerçekleşme olasılığı) $\lambda \Delta t$ olacaktır. Dolayısıyla $P(X = 0) = 1 - \lambda \Delta t$ yazılabilir.

Şimdi de $t + \Delta t$ aralığı ele alınsın. t orijinden bu yana ölçülen zaman, Δt de t 'ye eklenen marjinal terim olsun. Yine $P_0(t)$, t uzunluğundaki zaman diliminde söz konusu olayın gerçekleşmeme (ya da sıfır kere başarı gözlenme) olasılığı olsun. Bu durumda da $P_0(t + \Delta t)$ de, $t + \Delta t$ aralığında X 'in sıfıra eşit olma olasılığı olacaktır.

Yine söz konusu olayın t zaman diliminde gerçekleşmesi ile Δt zaman dilimi içerisinde gerçekleşme olaylarının bağımsız olduğu varsayılın.

Yine Δt 'nin çok küçük olmasından dolayı $(\Delta t)^2$ (ve mertebesi daha yüksek olan terimler) sıfır kabul edilecektir.

$P_0(t + \Delta t)$ olayının gerçekleşme olasılığı şu şekilde hesaplanabilmektedir. Orijinden $t + \Delta t$ 'e kadar geçen zaman diliminde söz konusu olay hiç gerçekleşmediğine ($X=0$) göre orijinden t 'ye kadar geçen sürede söz konusu olay gerçekleşmemiş ve t 'den $t + \Delta t$ 'ye kadar geçen sürede de söz konusu olay gerçekleşmemiştir. O halde bu olayın gerçekleşme olasılığı

$$P_0(t + \Delta t) = P_0(t) \cdot (1 - \lambda \Delta t)$$

olmalıdır. Buradan hareketle

$$\lim_{\Delta t \rightarrow 0} \frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} = -\lambda P_0(t)$$

bulunur. Eşitliğin sol tarafı türevin limit tanımı olduğu için

$$\frac{dP_0}{dt} = -\lambda P_0 \Rightarrow \frac{dP_0}{P_0} = -\lambda dt \Rightarrow \ln P_0 = -\lambda t + c \Rightarrow P_0(t) = c e^{-\lambda t}$$

şeklinde bulunur. Burada c bir sabittir. c 'nin belirlenebilmesi için, $t=0$ için, $P_0(0) = c$

olacaktır. $t=0$ zamanında olayın gerçekleşme olasılığı sıfırdır. Dolayısıyla söz konusu olayın sıfır kere gerçekleşme olasılığı birdir. Buradan $c=1$ olarak bulunur. O yüzden de $P_0(t) = e^{-\lambda t}$ olur.

Şimdi de t zaman diliminde söz konusu olayın bir kez gerçekleşme olasılığı $P_1(t)$ ile gösterilsin. Yine $t + \Delta t$ zaman diliminde söz konusu olayın gerçekleşme olasılığı da $p_1(t + \Delta t)$ gösterimi ile ifade edilsin. Şimdi bu zaman dilimi içerisinde olayın bir kez gerçekleşmesi şu iki farklı şekilde olabilir:

- 1) t zaman diliminde olay bir kez gerçekleşir ve geri kalan Δt zaman diliminde gerçekleşmez. Bu durumun gerçekleşme olasılığı $P_1(t) \cdot (1 - \lambda \Delta t)$ 'ye eşit olmaktadır.
- 2) t zamanında olay gerçekleşmez ve geri kalan Δt zaman diliminde gerçekleşir. Bu durumun gerçekleşme olasılığı $P_0(t) \cdot \lambda \Delta t$ 'ye eşittir.

1. durumun olasılığı ile 2. durumun olasılığının toplamı da $p_1(t + \Delta t)$ olasılığına eşit olacaktır. Dolayısıyla

$$p_1(t + \Delta t) = P_1(t) \cdot (1 - \lambda \Delta t) + P_0(t) \cdot \lambda \Delta t$$

olacaktır. Bu denklemden yola çıkarak

$$\frac{p_1(t + \Delta t) - P_1(t)}{\Delta t} = \lambda P_0(t) - \lambda P_1(t)$$

olur. Yine eşitliğin her iki tarafının $\Delta t \rightarrow 0$ iken limiti alınacak (ve daha önce karşılığı bulunan $P_0(t)$ denklemden yerine konulacak) olursa

$$\lim_{\Delta t \rightarrow 0} \frac{p_1(t + \Delta t) - P_1(t)}{\Delta t} = \lambda e^{-\lambda t} - \lambda P_1(t) = \frac{d P_1}{dt}$$

Buradan da $\frac{d P_1}{dt} + \lambda P_1 = \lambda e^{-\lambda t}$ olur. Eşitliğin her iki yanını $e^{\lambda t}$ ile çarpılacak olursa

$$e^{\lambda t} \frac{dP_1}{dt} + \lambda e^{\lambda t} P_1 = \lambda \Rightarrow \frac{d(e^{\lambda t} P_1)}{dt} = \lambda$$

Her iki tarafın da integrali alınacak olursa

$$e^{\lambda t} P_1 = \lambda t + C_2 \Rightarrow P_1(t) = \lambda t e^{-\lambda t} + C_2 e^{-\lambda t}$$

bulunur. Yine $P_1(0) = 0$ olacağı için (sıfır uzunluğunda bir zaman diliminde bir başarı elde etme olasılığı sıfır olacaktır.)

$C_2 = 0$ bulunur.

Dolayısıyla $P_1(t) = \lambda t e^{-\lambda t}$ bulunur. Bu şekilde devam edilecek olursa t birimlik bir zaman

diliminde n başarı elde etme olasılığı $P_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$ bulunacaktır. Eğer $t=1$ (birim zaman) kabul edilecek olursa da bilinen Poisson formülüne ulaşılır:

$$P(X=n) = P_n(t) = \frac{\lambda^n e^{-\lambda}}{n!}$$

Bu formülasyondaki varsayımlar özetlenecek olursa

- i) Δt gibi son derece küçük bir zaman diliminde bir başarı elde etme olasılığı Δt 'nin büyüklüğü ile doğru orantılıdır. Zaman dilimi uzadıkça başarı elde etme olasılığı aralıklı olarak artar. $P_1(\Delta t) = \lambda \Delta t$
- ii) Δt yeterince küçük olduğu için söz konusu dilimde birden fazla başarı gözlenme olasılığı sıfır kabul edilebilir. Dolayısıyla $P_0(\Delta t) \cong 1 - P_1(\Delta t) = 1 - \lambda \Delta t$ olacaktır.
- iii) t zaman diliminde başarının ortaya çıkması olayı ile Δt zaman diliminde başarının ortaya çıkması olayı bağımsız kabul edilmektedir.

Ek 4 Yalın Doğrusal Regresyon Analizine Matris Yaklaşımı

1. Eşgüçlü (İdempotent) Matris

Eğer $P^2 = P$ ise, P eşgüçlü bir matristir. Simetrik bir eşgüçlü matrise ise projeksiyon matrisi denir. P gibi, $I-P$ de eşgüçlüdür.

2. Rastlantısal Vektörler ve Matrisler

Rastlantısal bir vektör ya da rastlantısal bir matris rastlantısal değişkenlerin elemanlarını içermektedir. Böylece, Y_i elemanları rastlantısal değişkenler olduğu için gözlemler vektörü Y rastlantısal bir vektör olmaktadır.

2.1 Rastlantısal Vektör ya da Matrislerin Beklenen Değeri

Gözlem vektörü Y içinde $n=3$ gözlem olduğu varsayalım:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}_{3 \times 1}$$

Y 'nin beklenen değeri bir vektördür, $E\{Y\}$ ile gösterilmekte ve aşağıdaki gibi tanımlanmaktadır:

$$E\{Y\} = \begin{bmatrix} E\{Y_1\} \\ E\{Y_2\} \\ E\{Y_3\} \end{bmatrix}_{3 \times 1}$$

Böylece, bir rastlantısal vektörün beklenen değeri, rastlantısal değişkenlerin beklenen değerlerinden oluşan bir vektördür. Buradaki rastlantısal değişkenler de rastlantısal vektörlerin elemanlarıdır. Benzer şekilde, rastlantısal bir matrisin beklenen değeri, asıl matristeki rastlantısal değişkenlerin beklenen değerlerine karşılık gelen elemanların oluşturduğu bir matristir.

Genellikle, rastlantısal bir vektör olan Y için beklenen değer:

$$E\{Y\} = [E\{Y_i\}]_{n \times 1} \quad i = 1, \dots, n$$

ve $n \times p$ boyutlu rastlantısal bir matris olan Y için beklenen değer:

$$E\{Y\} = \begin{bmatrix} E\{Y_{ij}\} \end{bmatrix}_{n \times p} \quad i = 1, \dots, n; \quad j = 1, \dots, p$$

olmaktadır.

2.2 Rastlantısal Vektörün Varyans-Kovaryans Matrisi

Rastlantısal vektör Y 'nin üç gözlem değeri Y_1, Y_2, Y_3 'ten oluştuğu varsayalım. Üç rastlantısal değişkenin varyansı, $\sigma^2\{Y_i\}$; ve bu rastlantısal değişkenlerin oluşturduğu ikililerin arasındaki kovaryanslar, $\sigma\{Y_i, Y_j\}$; Y 'nin varyans-kovaryans matrisi içinde bir araya getirilirse, bu $\sigma^2\{Y\}$ olarak gösterilmekte ve aşağıdaki gibi yazılmaktadır:

$$\sigma^2\{Y\} = \begin{bmatrix} \sigma^2\{Y_1\} & \sigma\{Y_1, Y_2\} & \sigma\{Y_1, Y_3\} \\ \sigma\{Y_2, Y_1\} & \sigma^2\{Y_2\} & \sigma\{Y_2, Y_3\} \\ \sigma\{Y_3, Y_1\} & \sigma\{Y_3, Y_2\} & \sigma^2\{Y_3\} \end{bmatrix}$$

Köşegen üzerinde varyanslar, matrisin i .satırı ve j .sütununda kovaryans $\sigma\{Y_i, Y_j\}$ bulunmaktadır. Böylece, $\sigma\{Y_2, Y_1\}$ ikinci satır birinci sütunda ve $\sigma\{Y_1, Y_2\}$ birinci satır ikinci sütunda bulunmaktadır. Fakat zaten $\sigma\{Y_2, Y_1\} = \sigma\{Y_1, Y_2\}$ birbirine eşittir. $i \neq j$ olduğu zaman $\sigma\{Y_i, Y_j\} = \sigma\{Y_j, Y_i\}$ olacağı için, $\sigma^2\{Y\}$ simetrik bir matris olmaktadır.

Varyans-kovaryans matrisi $n \times 1$ boyutlu rastlantısal bir vektör olan Y için genişletilirse:

$$\sigma^2\{Y\} = \begin{bmatrix} \sigma^2\{Y_1\} & \sigma\{Y_1, Y_2\} & \dots & \sigma\{Y_1, Y_n\} \\ \sigma\{Y_2, Y_1\} & \sigma^2\{Y_2\} & \dots & \sigma\{Y_2, Y_n\} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma\{Y_n, Y_1\} & \sigma\{Y_n, Y_2\} & \dots & \sigma^2\{Y_n\} \end{bmatrix}_{n \times n}$$

elde edilmektedir. Yukarıda da görüldüğü gibi $\sigma^2\{Y\}$ simetrik bir matristir.

2.3 Çok Değişkenli Normal Dağılım

2.3.1 Yoğunluk Fonksiyonu

Çoklu normal dağılım için, yoğunluk fonksiyonu, en iyi matris biçiminde gösterilmektedir. İlk olarak bazı vektör ve matrisleri tanımlamak gerekmektedir. Gözlem vektörü Y 'nin p elemandan oluştuğu kabul edilsin:

$$Y_{p \times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix}$$

Ortalama vektörü $E\{Y\}$, μ ile gösterilmektedir ve bunların her birinin beklenen değerinden meydana gelmektedir:

$$\mu_{p \times 1} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

Sonuç olarak, varyans-kovaryans matrisi $\sigma^2\{Y\}$, Σ ile gösterilmektedir ve daima p tane Y değişkeninin varyans ve kovaryanslarını içinde bulundurmaktadır:

$$\Sigma_{p \times p} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{bmatrix}$$

Burada, σ_1^2 , Y_1 'in varyansını ve σ_{12} , Y_1 ve Y_2 'nin kovaryansını göstermektedir.

Çok değişkenli normal dağılımın yoğunluk fonksiyonu aşağıdaki gibi gösterilmektedir:

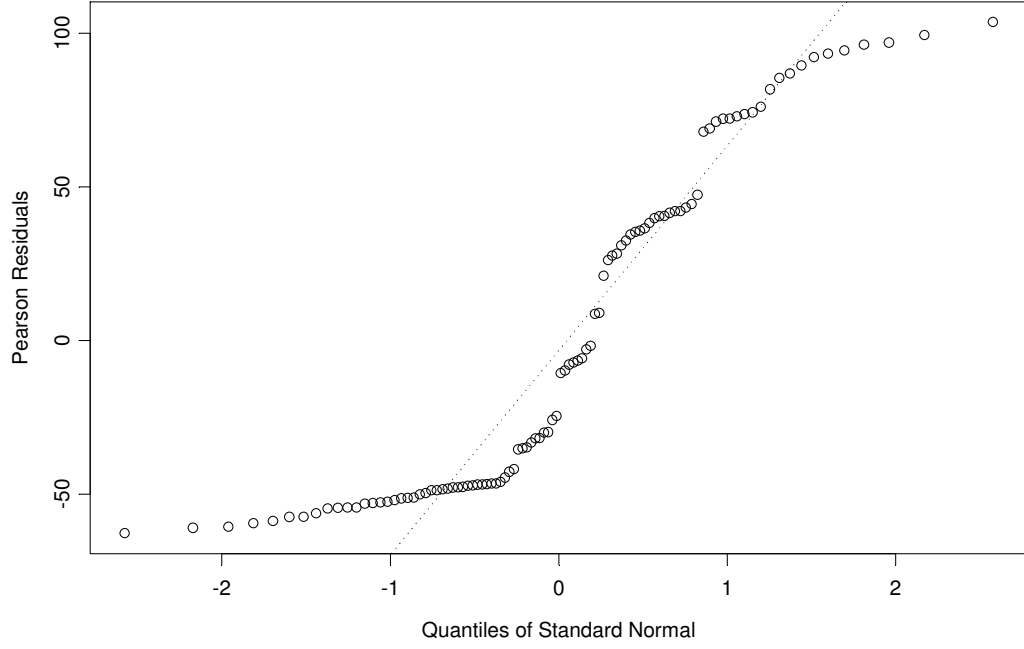
$$f(Y) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \cdot \exp \left[-\frac{1}{2} (Y - \mu)' \Sigma^{-1} (Y - \mu) \right]$$

Burada, $|\Sigma|$ varyans-kovaryans matrisi Σ 'nin determinantıdır. $p=2$ değişken olduğu zaman, çok değişkenli normal yoğunluk fonksiyonu iki değişkenli normal yoğunluk fonksiyonu halini almaktadır.

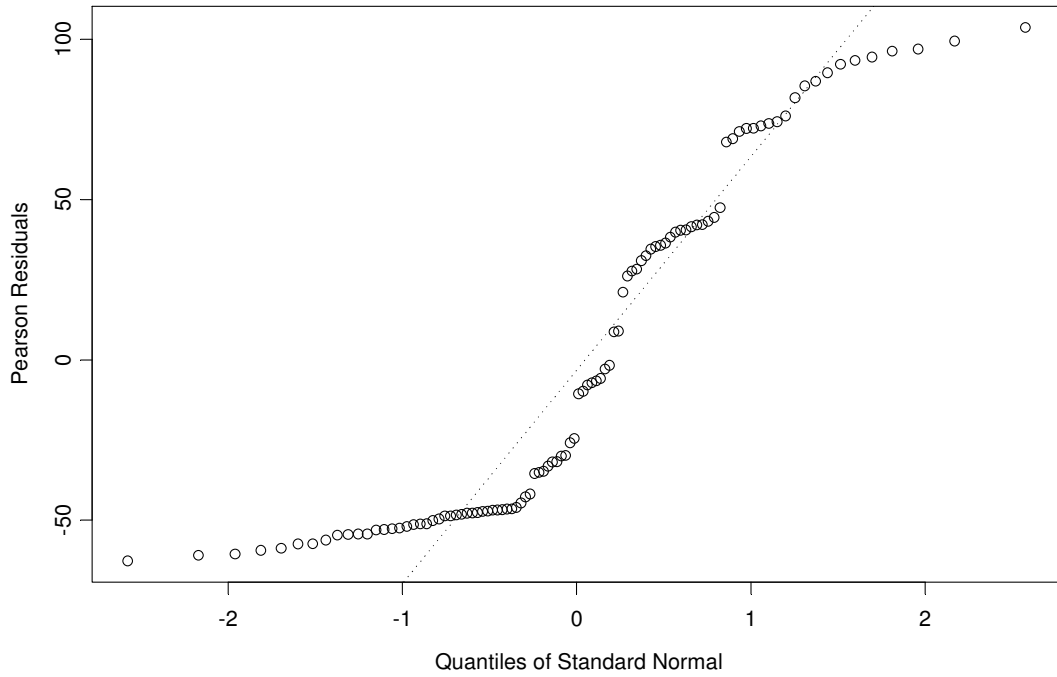
Çok değişkenli normal yoğunluk fonksiyonu iki değişkenli normal yoğunluk fonksiyonu için tanımlanan özelliklere uymaktadır. Örneğin, eğer $Y_1, \dots, Y_p$ birlikte normal dağılıyorsa, her bir Y_k değişkeninin marjinal olasılık dağılımı μ_k ortalama ve σ_k standart sapma etrafında normaldir.

Ek 5 Boşanma Verileri'nde İkinci Model ile İlgili Grafikler

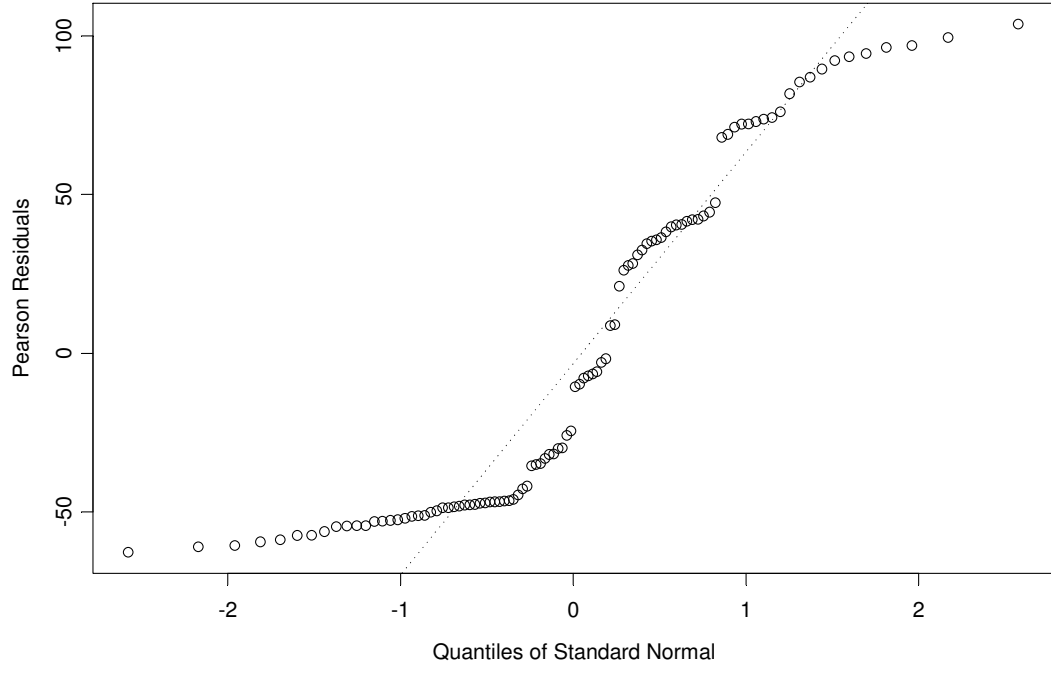
İkinci model için Pearson kalıntılarının dağılımı



İkinci model için tahmin edilen değerlerle boşanma sayılarının grafiği

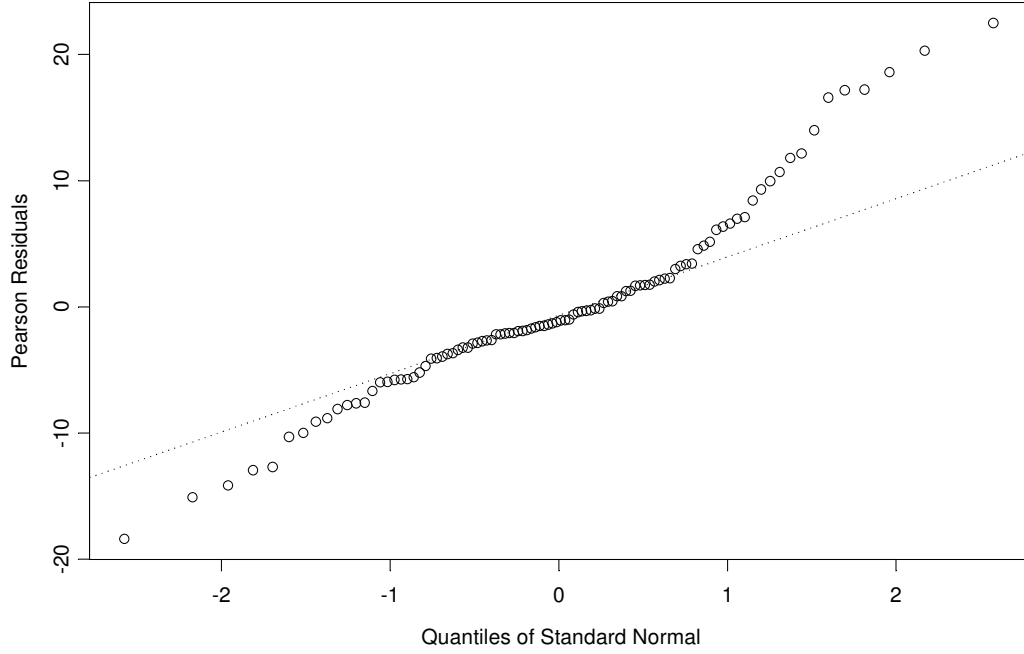


İkinci model için tahmin edilen değerlerle sapma kalıntılarının grafiği

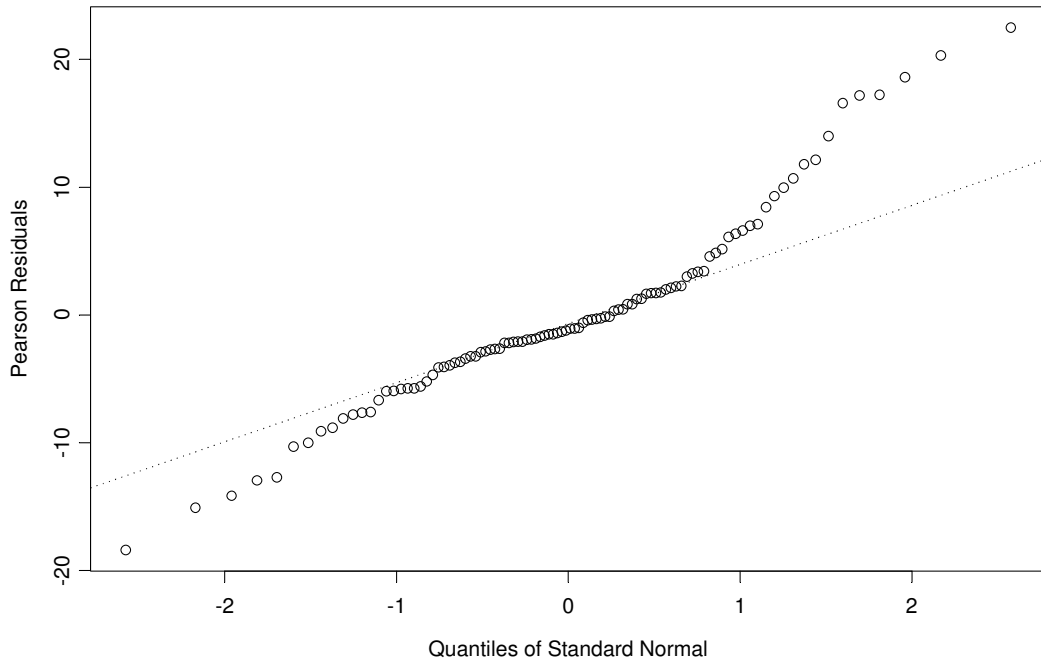


Ek 6 Boşanma Verileri'nde Üçüncü Model ile İlgili Grafikler

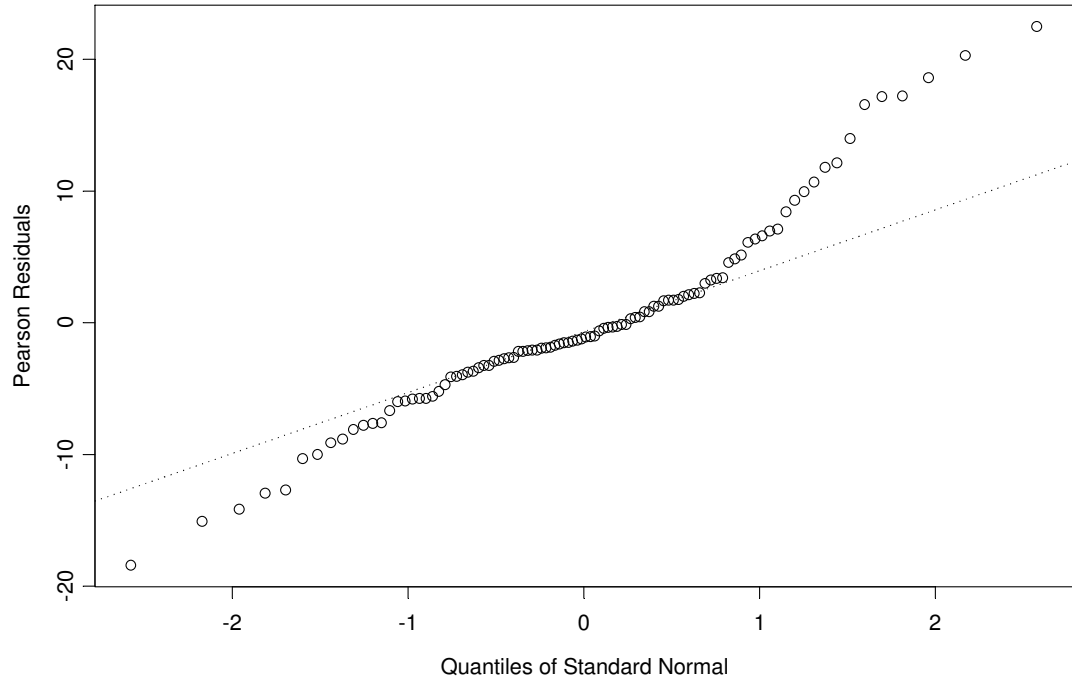
Üçüncü model için Pearson kalıntılarının dağılımı



Üçüncü model için tahmin edilen değerlerle boşanma sayılarının grafiği

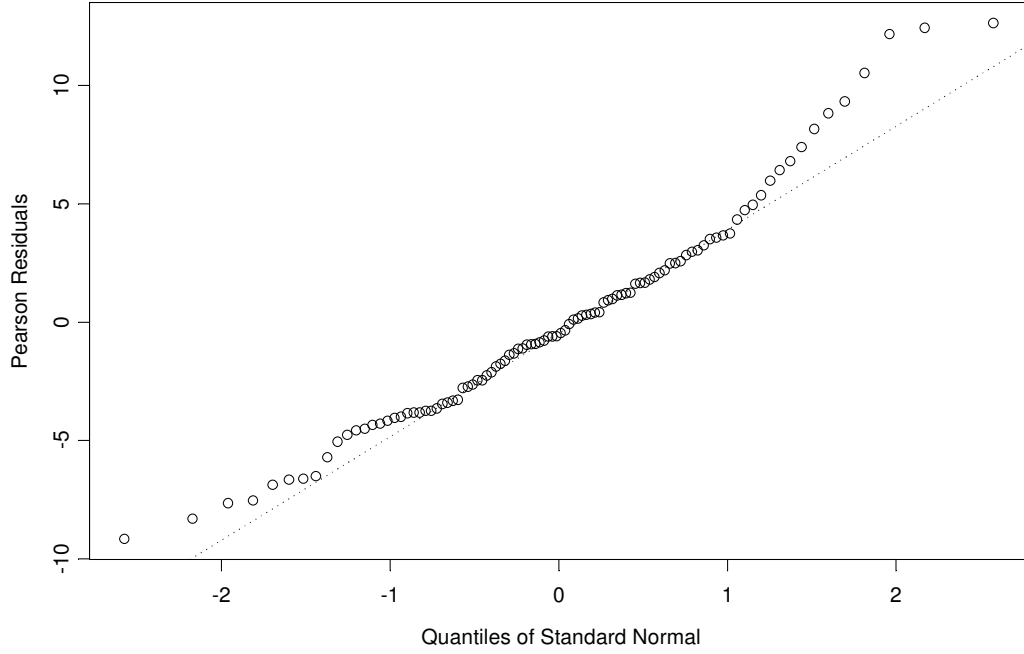


Üçüncü model için tahmin edilen değerlerle sapma kalıntılarının grafiği



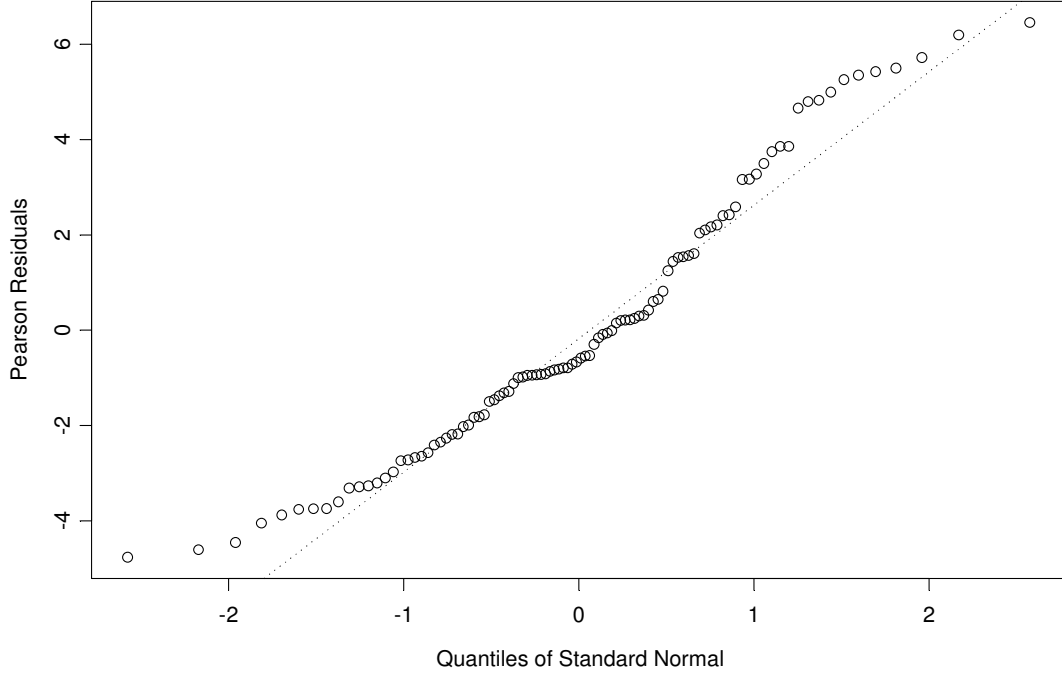
Ek 7 Boşanma Verileri'nde Dördüncü Model ile İlgili Grafik

Dördüncü model için tahmin edilen değerlerle sapma kalıntılarının grafiği

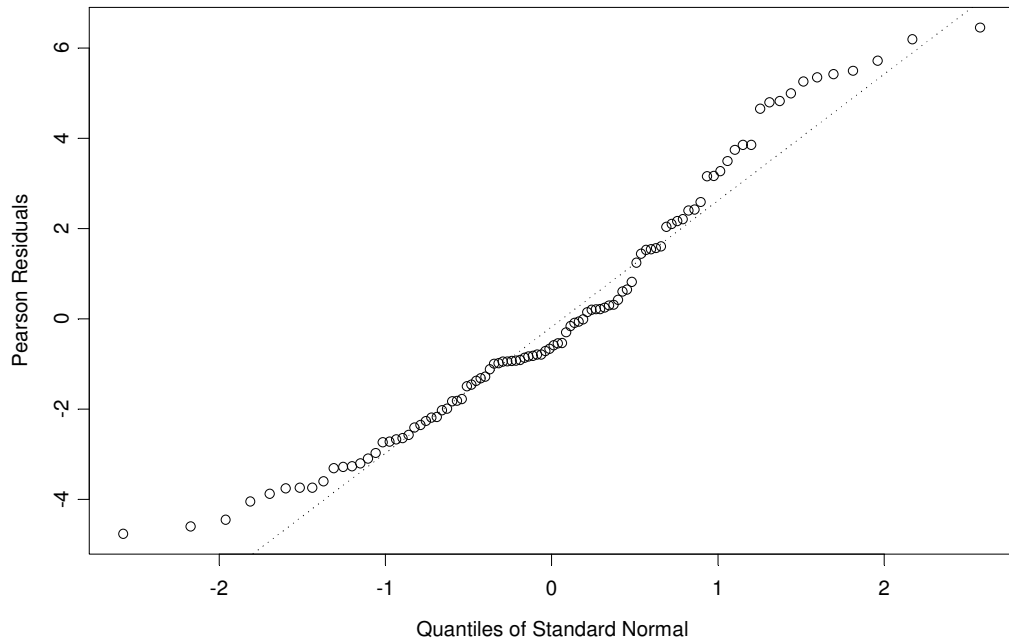


Ek 8 Boşanma Verileri'nde Beşinci Model ile İlgili Grafikler

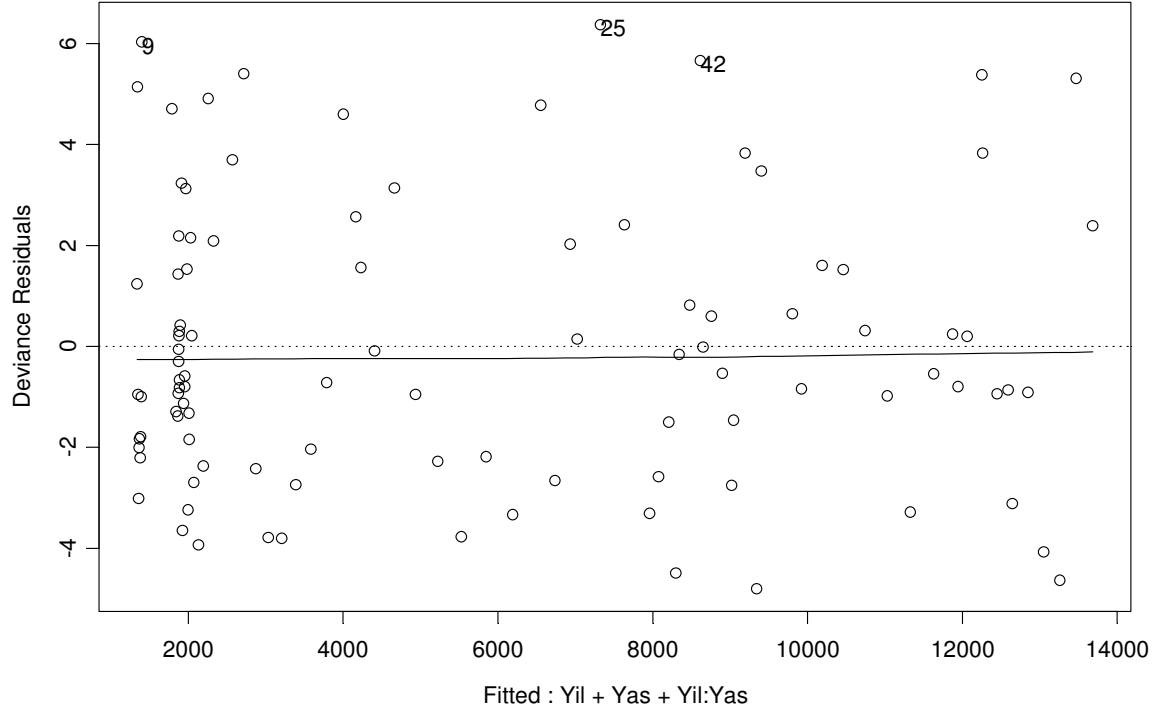
Beşinci model için Pearson kalıntılarının dağılımı



Beşinci model için tahmin edilen değerlerle boşanma sayılarının grafiği

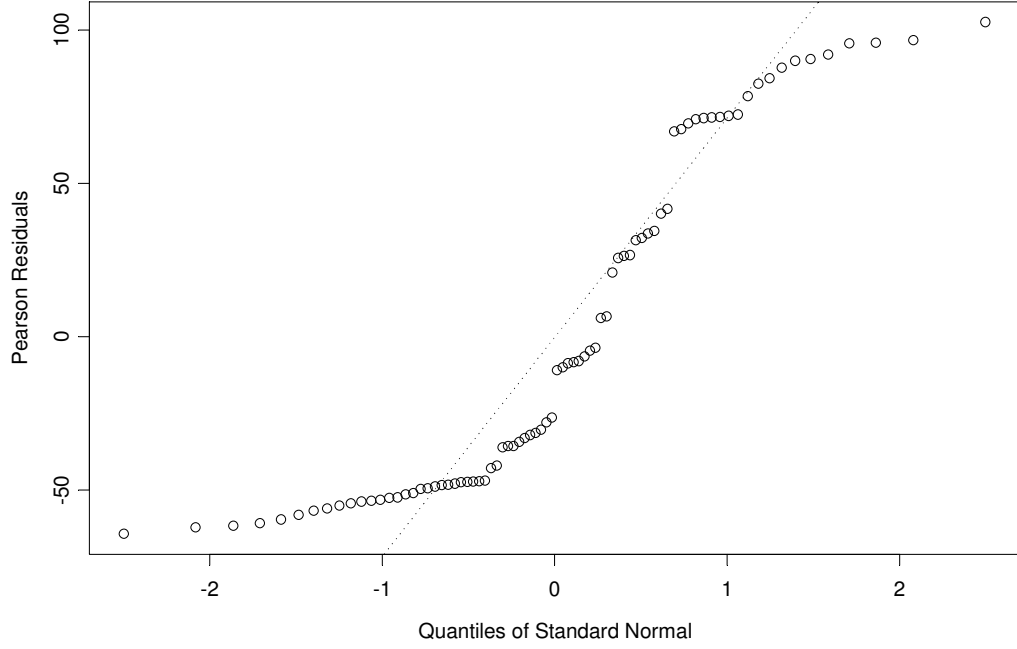


Beşinci model için tahmin edilen değerlerle sapma kalıntılarının grafiği

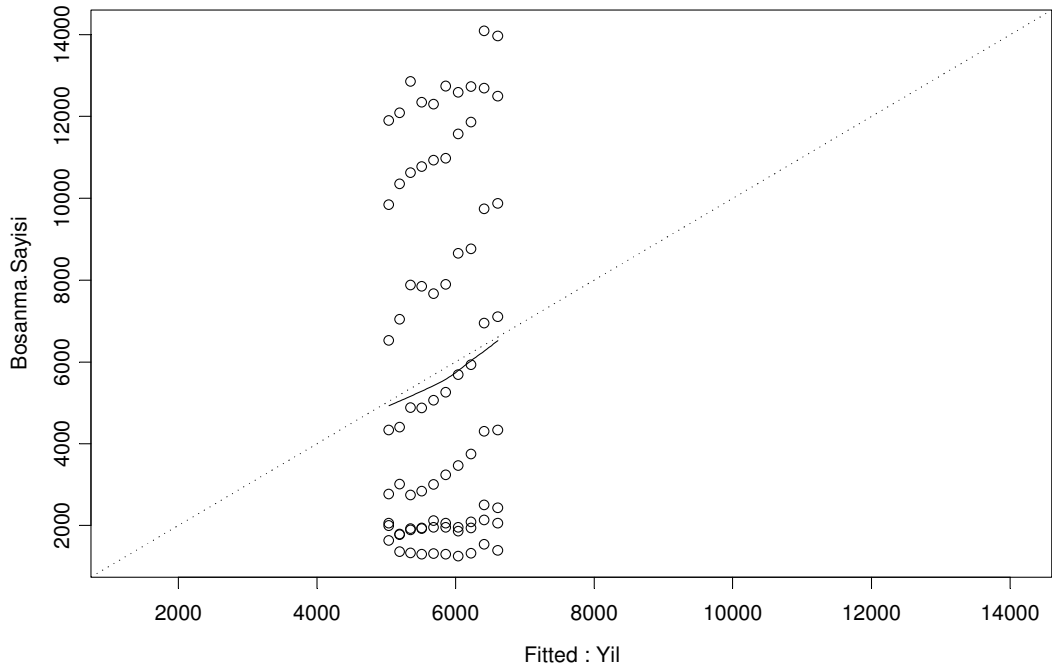


Ek 9 Boşanma Verileri2’de Yedinci Model ile İlgili Grafikler

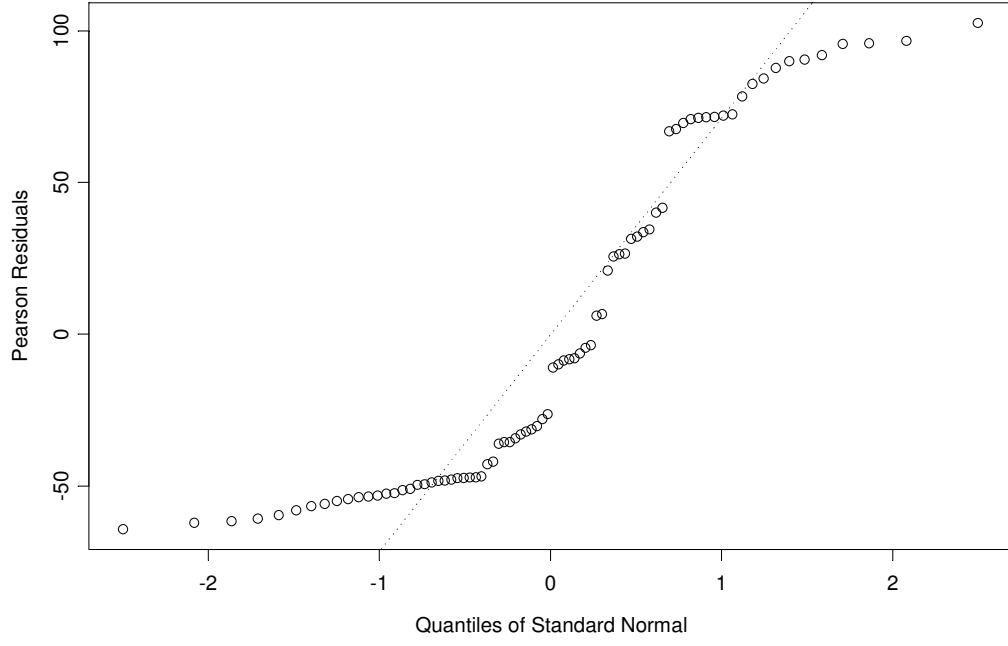
Yedinci model için Pearson kalıntılarının dağılımı



Yedinci model için tahmin edilen değerlerle boşanma sayılarının grafiği

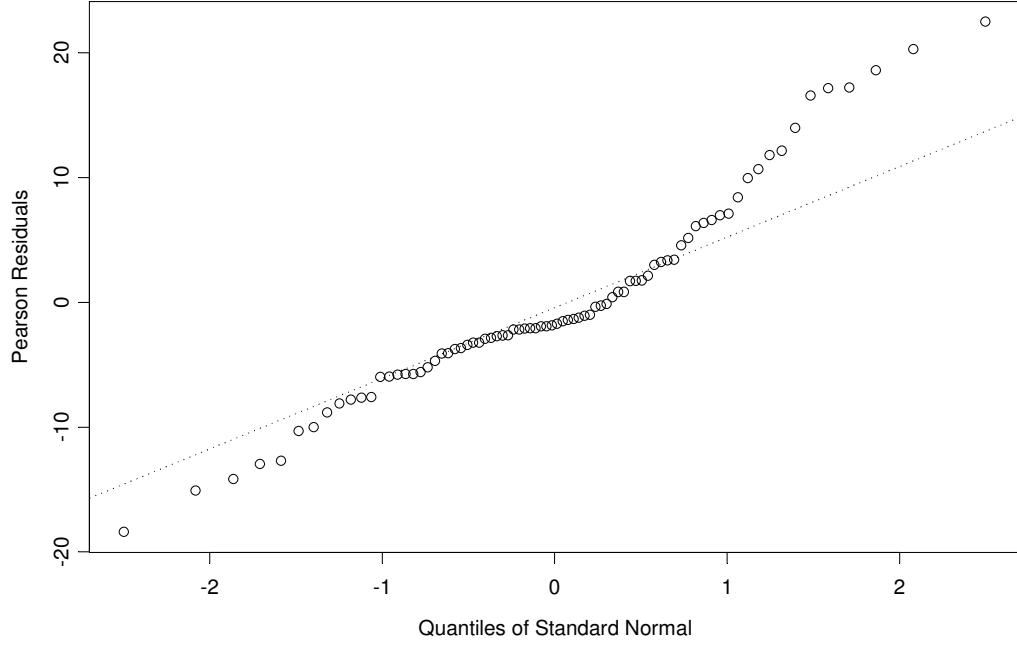


Yedinci model için tahmin edilen değerlerle sapma kalıntılarının grafiği

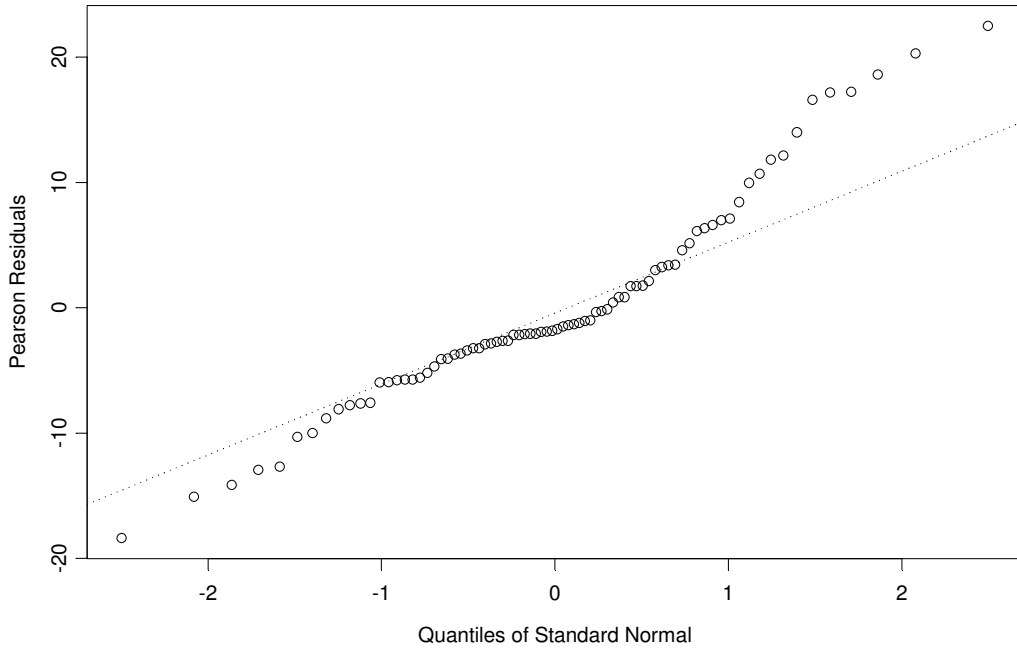


Ek 10 Boşanma Verileri2’de Sekizinci Model ile İlgili Grafikler

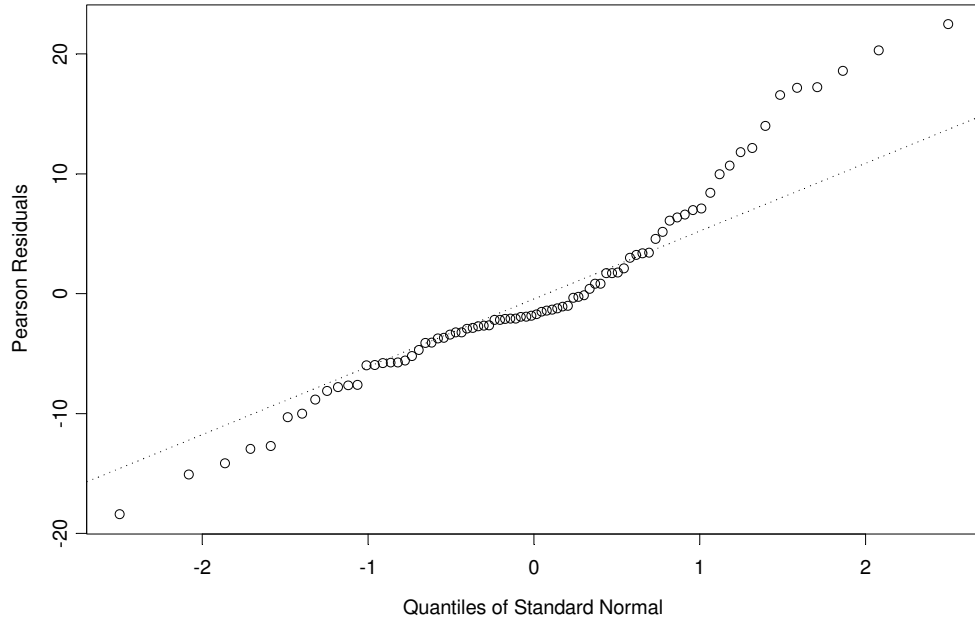
Sekizinci model için Pearson kalıntılarının dağılımı



Sekizinci model için tahmin edilen değerlerle boşanma sayılarının grafiği

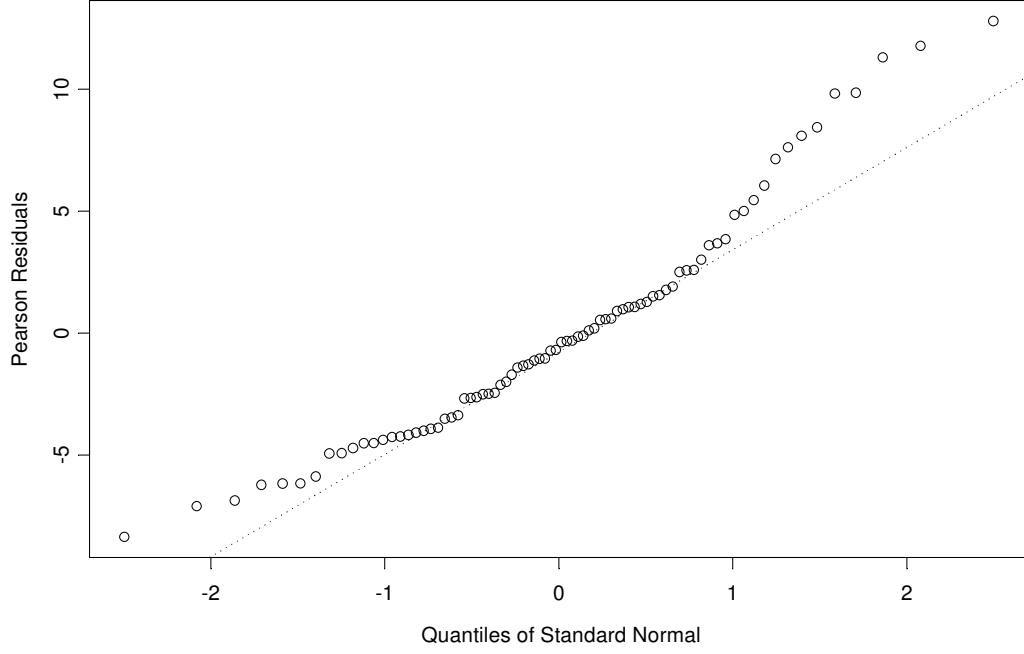


Sekizinci model için tahmin edilen değerlerle sapma kalıntılarının grafiği



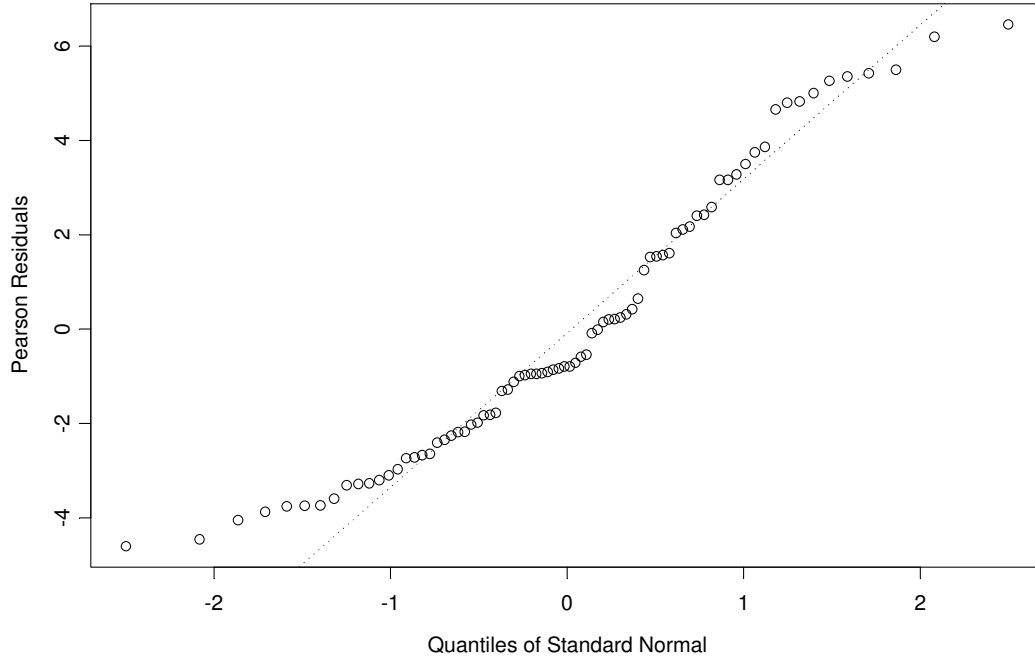
Ek 11 Boşanma Verileri2’de Dokuzuncu Model ile İlgili Grafik

Dokuzuncu model için tahmin edilen değerlerle sapma kalıntılarının grafiği



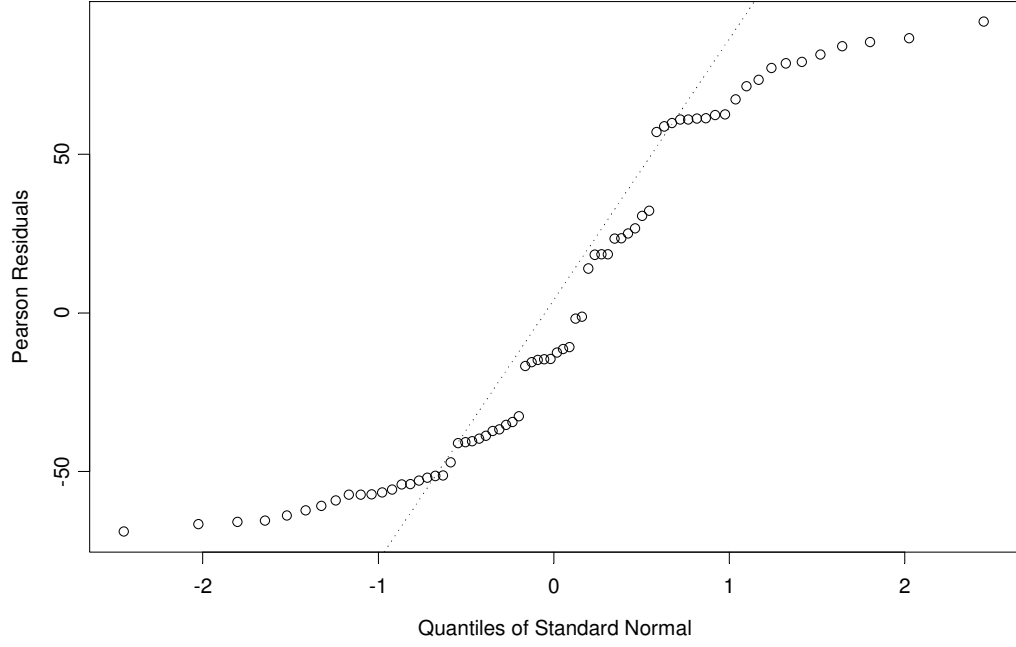
Ek 12 Boşanma Verileri2’de Onuncu Model ile İlgili Grafik

Onuncu model için tahmin edilen değerlerle sapma kalıntılarının grafiği

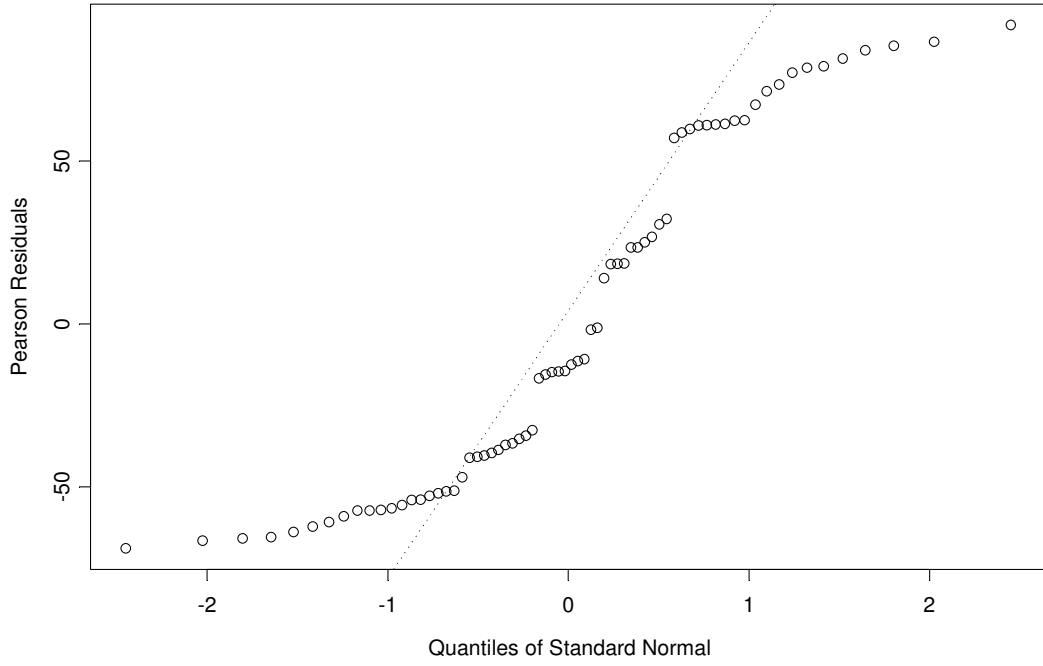


Ek 13 Boşanma Verileri3’de On İkinci Model ile İlgili Grafikler

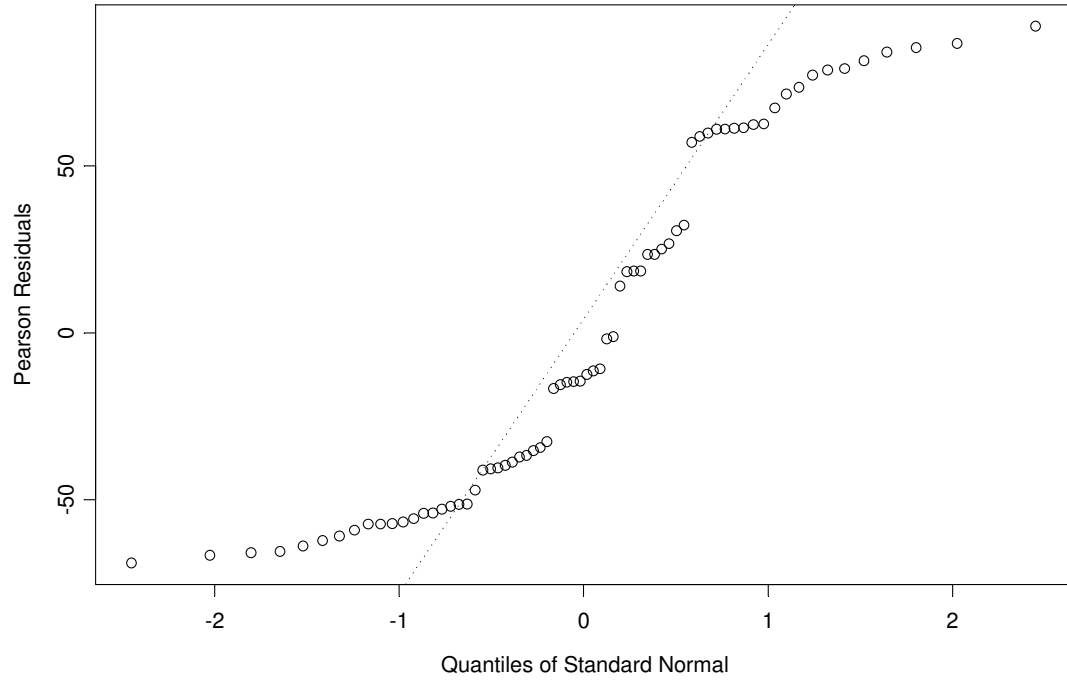
On ikinci model için Pearson kalıntılarının dağılımı



On ikinci model için tahmin edilen değerlerle boşanma sayılarının grafiği

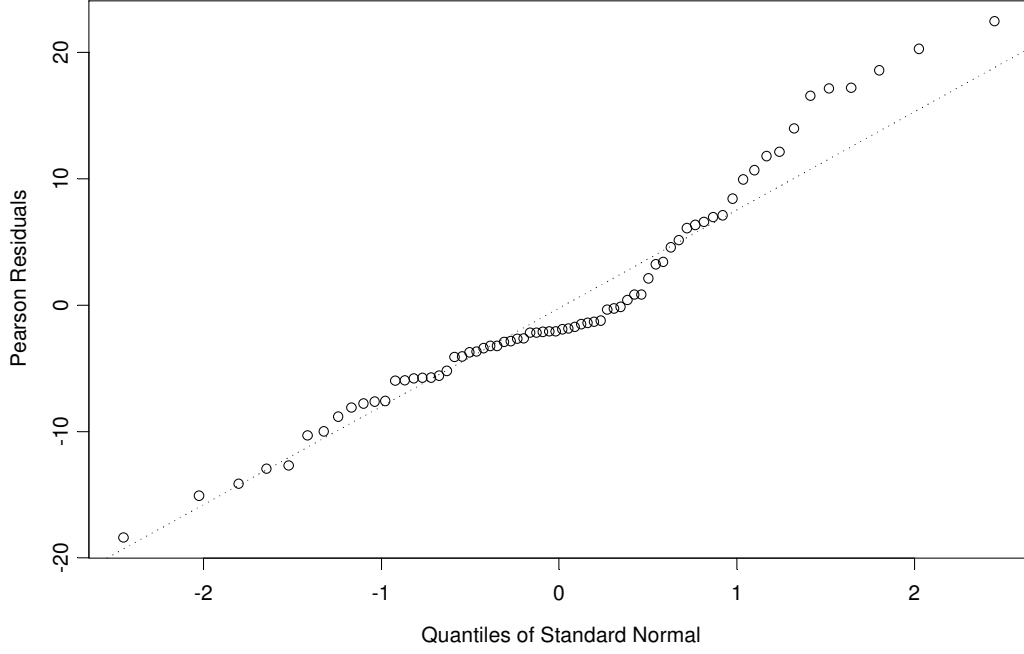


On ikinci model için tahmin edilen değerlerle sapma kalıntılarının grafiği

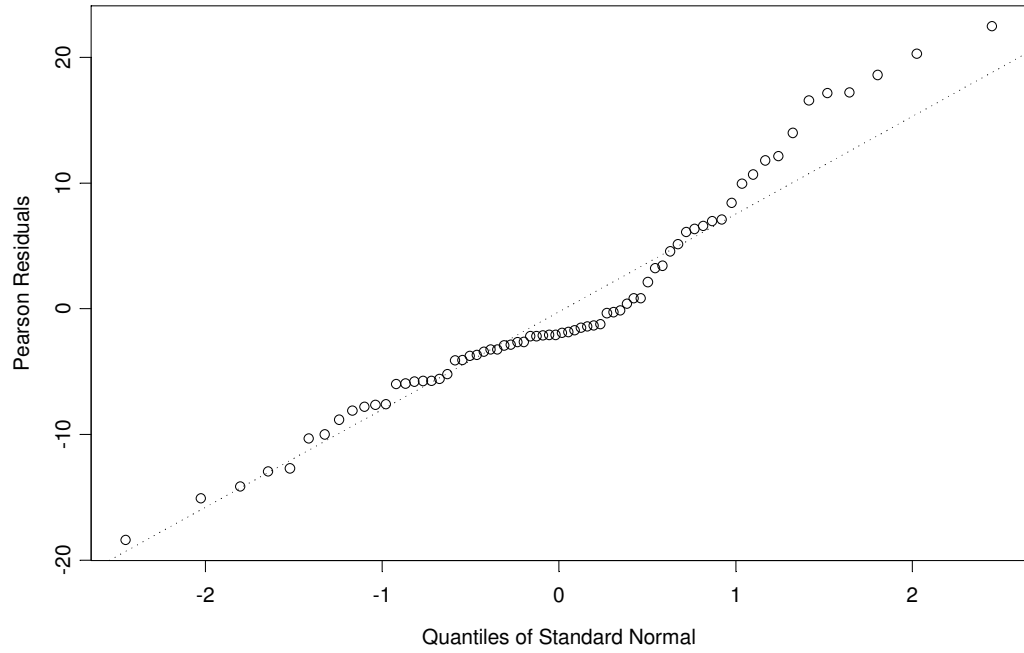


Ek 14 Boşanma Verileri3’de On Üçüncü Model ile İlgili Grafikler

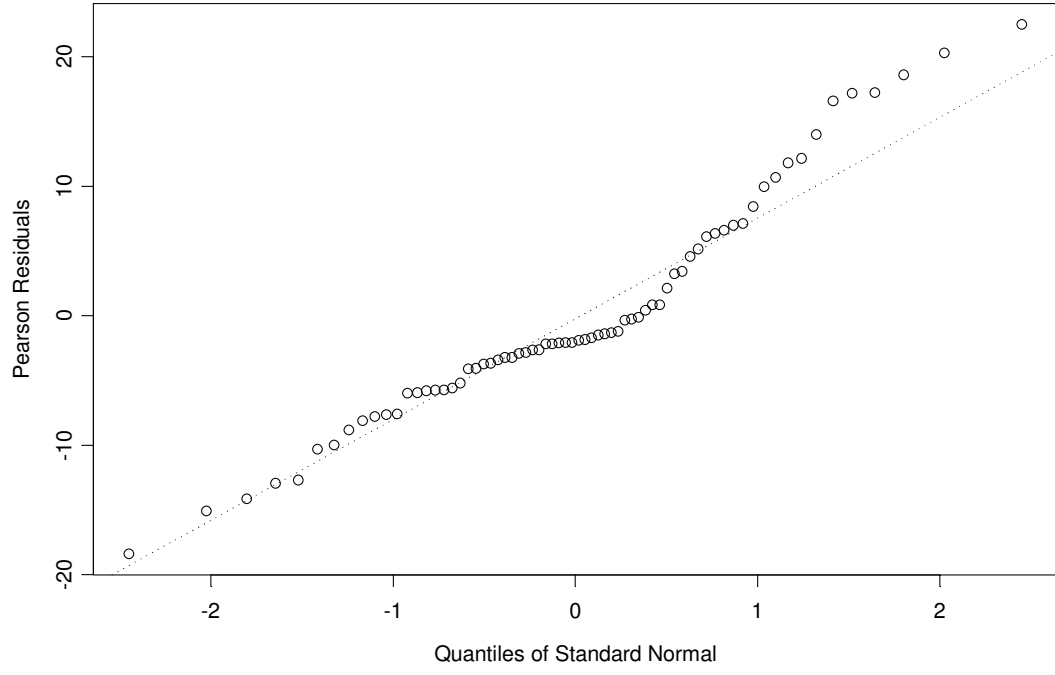
On üçüncü model için Pearson kalıntılarının dağılımı



On üçüncü model için tahmin edilen değerlerle boşanma sayılarının grafiği

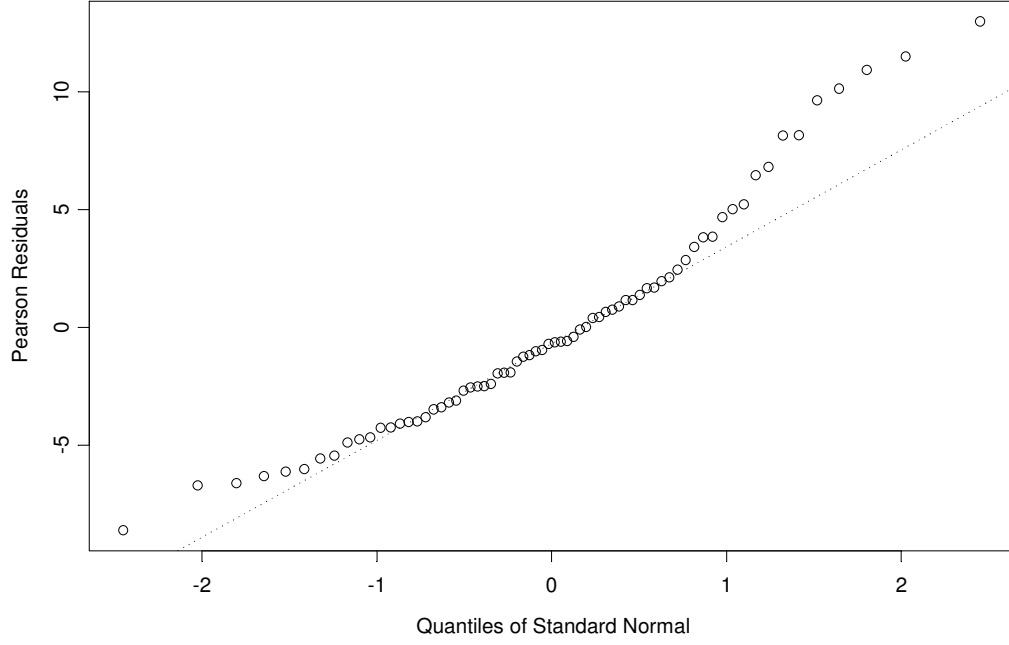


On üçüncü model için tahmin edilen değerlerle sapma kalıntılarının grafiği



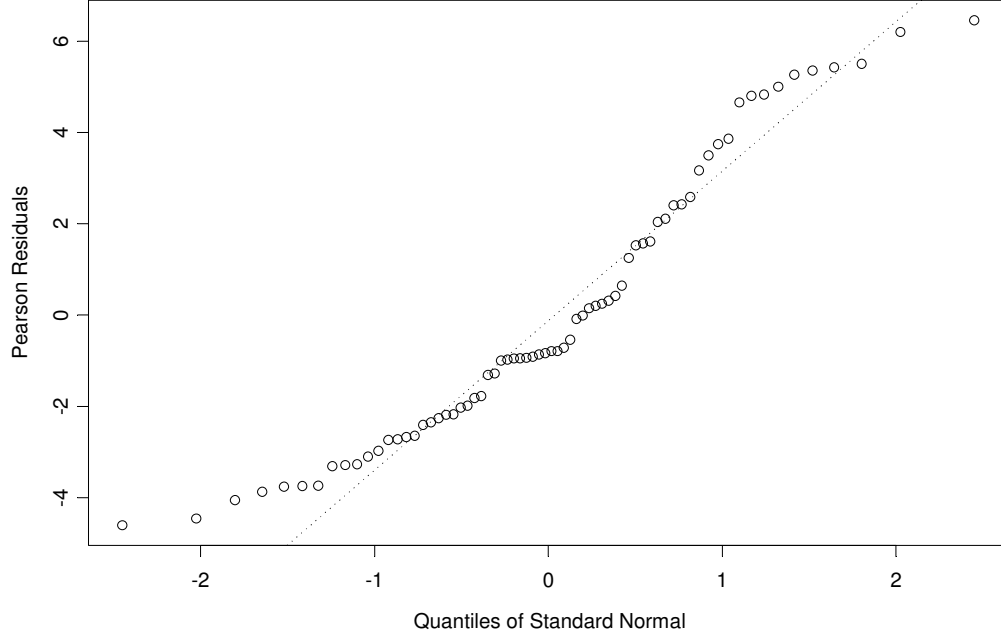
Ek 15 Boşanma Verileri3’de On Dördüncü Model ile İlgili Grafik

On dördüncü model için tahmin edilen değerlerle sapma kalıntılarının grafiği



Ek 16 Boşanma Verileri3’de On Beşinci Model ile İlgili Grafikler

On beşinci model için tahmin edilen değerlerle sapma kalıntılarının grafiği



Ek 17 İkinci Model için Elde Edilen Bilgisayar Çıktısı

*** Generalized Linear Model ***

Call: glm(formula = Bosanma.Sayisi ~ Yil, family = poisson(link = log), data =
hande,

na.action = na.exclude, control = list(epsilon = 0.0001, maxit = 50,
trace = F))

Deviance Residuals:

Min	1Q	Median	3Q	Max
-76.1764	-55.84321	-18.44645	38.4183	87.48669

Coefficients:

	Value	Std. Error	t value
(Intercept)	-45.15699608	0.9215083162	-49.00335
Yil	0.02698814	0.0004622045	58.39005

(Dispersion Parameter for Poisson family taken to be 1)

Null Deviance: 291914.1 on 99 degrees of freedom

Residual Deviance: 288499.9 on 98 degrees of freedom

Number of Fisher Scoring Iterations: 4

Analysis of Deviance Table

Poisson model

Response: Bosanma.Sayisi

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev
NULL			99	291914.1
Yil 1	3414.188		98	288499.9

Ek 18 Üçüncü Model için Elde Edilen Bilgisayar Çıktısı

*** Generalized Linear Model ***

Call: glm(formula = Bosanma.Sayisi ~ Yas, family = poisson(link = log), data =
handel, na.action = na.exclude, control = list(epsilon = 0.0001, maxit
= 50, trace = F))

Deviance Residuals:

Min	1Q	Median	3Q	Max
-19.08806	-3.827347	-1.166362	2.421714	21.45933

Coefficients:

	Value	Std. Error	t value
(Intercept)	8.35541684	0.0017727136	4713.34837
Yas1	0.76610049	0.0040221742	190.46924
Yas2	0.38311765	0.0016335031	234.53746
Yas3	0.15921448	0.0010301848	154.54943
Yas4	0.03266108	0.0008319960	39.25630
Yas5	-0.04616723	0.0007836277	-58.91475
Yas6	-0.10275186	0.0008210651	-125.14460
Yas7	-0.13815982	0.0008986247	-153.74585
Yas8	-0.15221969	0.0009684362	-157.18092
Yas9	-0.08497970	0.0007333056	-115.88579

(Dispersion Parameter for Poisson family taken to be 1)

Null Deviance: 291914.1 on 99 degrees of freedom

Residual Deviance: 5453.936 on 90 degrees of freedom

Number of Fisher Scoring Iterations: 3

Analysis of Deviance Table

Poisson model

Response: Bosanma.Sayisi

Terms added sequentially (first to last)

	Df	Deviance	Resid.	Df	Resid. Dev
NULL				99	291914.1
Yas	9	286460.1		90	5453.9

Ek 19 Dördüncü Model için Elde Edilen Bilgisayar Çıktısı

*** Generalized Linear Model ***

Call: glm(formula = Bosanma.Sayisi ~ Yil + Yas, family = poisson(link = log), data = handel, na.action = na.exclude, control = list(epsilon = 0.0001, maxit = 50, trace = F))

Deviance Residuals:

Min	1Q	Median	3Q	Max
-9.316255	-3.44854	-0.5303886	2.471756	12.31258

Coefficients:

	Value	Std. Error	t value
(Intercept)	-45.44844489	0.9217071960	-49.30898
Yil	0.02698814	0.0004623038	58.37750
Yas1	0.76610049	0.0040221639	190.46973
Yas2	0.38311765	0.0016335004	234.53784
Yas3	0.15921448	0.0010301849	154.54941
Yas4	0.03266108	0.0008320040	39.25592
Yas5	-0.04616723	0.0007836559	-58.91263
Yas6	-0.10275186	0.0008210984	-125.13952
Yas7	-0.13815982	0.0008986311	-153.74476
Yas8	-0.15221969	0.0009684262	-157.18254
Yas9	-0.08497970	0.0007333048	-115.88592

(Dispersion Parameter for Poisson family taken to be 1)

Null Deviance: 291914.1 on 99 degrees of freedom

Residual Deviance: 2039.748 on 89 degrees of freedom

Number of Fisher Scoring Iterations: 3

Analysis of Deviance Table

Poisson model

Response: Bosanma.Sayisi

Terms added sequentially (first to last)

	Df	Deviance	Resid.	Df	Resid. Dev
NULL				99	291914.1
Yil	1	3414.2		98	288499.9
Yas	9	286460.1		89	2039.7

Ek 20 Beşinci Model için Elde Edilen Bilgisayar Çıktısı

*** Generalized Linear Model ***

Call: glm(formula = Bosanma.Sayisi ~ Yil + Yas + Yil:Yas, family = poisson(link = log), data = handel, na.action = na.exclude, control = list(epsilon = 0.0001, maxit = 50, trace = F))

Deviance Residuals:

Min	1Q	Median	3Q	Max
-4.803384	-2.075824	-0.629111	1.708025	6.373225

Coefficients:

	Value	Std. Error	t value
(Intercept)	-39.9700746888	1.2327123558	-32.4244942
Yil	0.0242394350	0.0006183234	39.2018710
Yas1	-16.6299741651	2.7919421119	-5.9564180
Yas2	-5.1427827884	1.1340849703	-4.5347420
Yas3	-7.9364807652	0.7157690836	-11.0880463
Yas4	-10.8841008034	0.5793725736	-18.7860132
Yas5	-12.3057002066	0.5477494818	-22.4659276
Yas6	-8.4676116256	0.5740997278	-14.7493740
Yas7	0.0881840267	0.6251497358	0.1410606
Yas8	7.6023029502	0.6722510342	11.3087263
Yas9	3.6343684889	0.5091008914	7.1387981
YilYas1	0.0087261289	0.0014005107	6.2306765
YilYas2	0.0027718784	0.0005688758	4.8725549
YilYas3	0.0040607721	0.0003590281	11.3104585
YilYas4	0.0054755865	0.0002905931	18.8427946
YilYas5	0.0061488523	0.0002747121	22.3828953
YilYas6	0.0041954689	0.0002879241	14.5714414
YilYas7	-0.0001134215	0.0003135561	-0.3617265
YilYas8	-0.0038896205	0.0003372264	-11.5341526
YilYas9	-0.0018655176	0.0002553716	-7.3051096

(Dispersion Parameter for Poisson family taken to be 1)

Null Deviance: 291914.1 on 99 degrees of freedom

Residual Deviance: 767.2275 on 80 degrees of freedom

Number of Fisher Scoring Iterations: 3

Analysis of Deviance Table

Poisson model

Response: Bosanma.Sayisi

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev
NULL			99	291914.1
Yil	1	3414.2	98	288499.9
Yas	9	286460.1	89	2039.7
Yil:Yas	9	1272.5	80	767.2

Ek 21 Yedinci Model için Elde Edilen Bilgisayar Çıktısı

*** Generalized Linear Model ***

Call: glm(formula = Bosanma.Sayisi ~ Yil, family = poisson(link = log), data =
hande2, na.action = na.exclude, control = list(epsilon = 0.0001, maxit
= 50, trace = F))

Deviance Residuals:

Min	1Q	Median	3Q	Max
-78.24338	-55.97976	-19.69328	43.83679	86.71936

Coefficients:

	Value	Std. Error	t value
(Intercept)	-51.48097107	1.0217212187	-50.38651
Yil	0.03016888	0.0005124619	58.87048

(Dispersion Parameter for Poisson family taken to be 1)

Null Deviance: 243243.8 on 79 degrees of freedom

Residual Deviance: 239771.7 on 78 degrees of freedom

Number of Fisher Scoring Iterations: 4

Analysis of Deviance Table

Poisson model

Response: Bosanma.Sayisi

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev
NULL			79	243243.8
Yil 1	3472.058		78	239771.7

Ek 22 Sekizinci Model için Elde Edilen Bilgisayar Çıktısı

*** Generalized Linear Model ***

Call: glm(formula = Bosanma.Sayisi ~ Yas, family = poisson(link = log), data =
hande2, na.action = na.exclude, control = list(epsilon = 0.0001, maxit
= 50, trace = F))

Deviance Residuals:

Min	1Q	Median	3Q	Max
-19.08806	-4.302553	-1.810229	3.338851	21.45933

Coefficients:

	Value	Std. Error	t value
(Intercept)	8.3680956	0.0019746037	4237.86087
Yas1	-0.0646887	0.0020467435	-31.60567
Yas2	-0.1263422	0.0013500136	-93.58590
Yas3	-0.1650830	0.0011613100	-142.15243
Yas4	-0.1967352	0.0011468058	-171.55058
Yas5	-0.2126180	0.0012003995	-177.12272
Yas6	-0.2094212	0.0012493911	-167.61864
Yas7	-0.1110709	0.0009234735	-120.27509

(Dispersion Parameter for Poisson family taken to be 1)

Null Deviance: 243243.8 on 79 degrees of freedom

Residual Deviance: 5172.786 on 72 degrees of freedom

Number of Fisher Scoring Iterations: 3

Analysis of Deviance Table

Poisson model

Response: Bosanma.Sayisi

Terms added sequentially (first to last)

	Df	Deviance	Resid.	Df	Resid. Dev
NULL				79	243243.8
Yas	7	238071		72	5172.8

Ek 23 Dokuzuncu Model için Elde Edilen Bilgisayar Çıktısı

*** Generalized Linear Model ***

Call: glm(formula = Bosanma.Sayisi ~ Yil + Yas, family = poisson(link = log), data = hande2, na.action = na.exclude, control = list(epsilon = 0.0001, maxit = 50, trace = F))

Deviance Residuals:

Min	1Q	Median	3Q	Max
-8.463256	-3.634822	-0.5335848	2.037516	12.09635

Coefficients:

	Value	Std. Error	t value
(Intercept)	-51.77733269	1.0219617139	-50.66465
Yil	0.03016889	0.0005125821	58.85669
Yas1	-0.06468870	0.0020467461	-31.60563
Yas2	-0.12634224	0.0013500288	-93.58485
Yas3	-0.16508304	0.0011613553	-142.14689
Yas4	-0.19673521	0.0011468557	-171.54312
Yas5	-0.21261803	0.0012004092	-177.12129
Yas6	-0.20942124	0.0012493748	-167.62083
Yas7	-0.11107086	0.0009234720	-120.27528

(Dispersion Parameter for Poisson family taken to be 1)

Null Deviance: 243243.8 on 79 degrees of freedom

Residual Deviance: 1700.728 on 71 degrees of freedom

Number of Fisher Scoring Iterations: 3

Analysis of Deviance Table

Poisson model

Response: Bosanma.Sayisi

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev
NULL			79	243243.8
Yil 1	3472.1		78	239771.7
Yas 7	238071.0		71	1700.7

Ek 24 Onuncu Model için Elde Edilen Bilgisayar Çıktısı

*** Generalized Linear Model ***

Call: glm(formula = Bosanma.Sayisi ~ Yil + Yas + Yil:Yas, family = poisson(link = log), data = hande2, na.action = na.exclude, control = list(epsilon = 0.0001, maxit = 50, trace = F))

Deviance Residuals:

Min	1Q	Median	3Q	Max
-4.636821	-2.301101	-0.7952315	2.104357	6.373225

Coefficients:

	Value	Std. Error	t value
(Intercept)	-48.323029870	1.3737392640	-35.176275
Yil	0.028435435	0.0006890513	41.267517
Yas1	-10.730178742	1.4224223630	-7.543595
Yas2	-13.780413488	0.9404811606	-14.652514
Yas3	-14.464656250	0.8120802218	-17.811856
Yas4	-8.227749819	0.8020783371	-10.258038
Yas5	3.100023781	0.8352250408	3.711603
Yas6	11.900508182	0.8673570782	13.720426
Yas7	5.866038797	0.6411774422	9.148854
YilYas1	0.005349666	0.0007134661	7.498136
YilYas2	0.006848427	0.0004717022	14.518539
YilYas3	0.007171906	0.0004072752	17.609482
YilYas4	0.004027948	0.0004022570	10.013368
YilYas5	-0.001661399	0.0004189215	-3.965897
YilYas6	-0.006074233	0.0004350975	-13.960626
YilYas7	-0.002997951	0.0003216222	-9.321344

(Dispersion Parameter for Poisson family taken to be 1)

Null Deviance: 243243.8 on 79 degrees of freedom

Residual Deviance: 674.0899 on 64 degrees of freedom

Number of Fisher Scoring Iterations: 3

Analysis of Deviance Table

Poisson model

Response: Bosanma.Sayisi

Terms added sequentially (first to last)

	Df	Deviance	Resid.	Df	Resid. Dev
NULL			79		243243.8
Yil	1	3472.1	78		239771.7
Yas	7	238071.0	71		1700.7
Yil:Yas	7	1026.6	64		674.1

Ek 25 On İkinci Model için Elde Edilen Bilgisayar Çıktısı

*** Generalized Linear Model ***

Call: glm(formula = Bosanma.Sayisi ~ Yil, family = poisson(link = log), data =
hande3, na.action = na.exclude, control = list(epsilon = 0.0001, maxit
= 50, trace = F))

Deviance Residuals:

Min	1Q	Median	3Q	Max
-84.65865	-60.07628	-13.94042	53.66169	79.09166

Coefficients:

	Value	Std. Error	t value
(Intercept)	-53.41763597	1.0445258897	-51.14056
Yil	0.03118533	0.0005238978	59.52560

(Dispersion Parameter for Poisson family taken to be 1)

Null Deviance: 205938.3 on 69 degrees of freedom

Residual Deviance: 202387.3 on 68 degrees of freedom

Number of Fisher Scoring Iterations: 4

Analysis of Deviance Table

Poisson model

Response: Bosanma.Sayisi

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev
NULL			69	205938.3
Yil 1	3550.945		68	202387.3

Ek 26 On Üçüncü Model için Elde Edilen Bilgisayar Çıktısı

*** Generalized Linear Model ***

Call: glm(formula = Bosanma.Sayisi ~ Yas, family = poisson(link = log), data =
hande3, na.action = na.exclude, control = list(epsilon = 0.0001, maxit
= 50, trace = F))

Deviance Residuals:

Min	1Q	Median	3Q	Max
-19.08806	-5.55429	-2.020395	4.934645	21.45933

Coefficients:

	Value	Std. Error	t value
(Intercept)	8.4791664	0.002015360	4207.27184
Yas1	-0.0646887	0.002046744	-31.60567
Yas2	-0.1263422	0.001350014	-93.58590
Yas3	-0.1650830	0.001161310	-142.15243
Yas4	-0.1967352	0.001146806	-171.55058
Yas5	-0.2126180	0.001200400	-177.12272
Yas6	-0.2094212	0.001249391	-167.61864

(Dispersion Parameter for Poisson family taken to be 1)

Null Deviance: 205938.3 on 69 degrees of freedom

Residual Deviance: 5107.632 on 63 degrees of freedom

Number of Fisher Scoring Iterations: 3

Analysis of Deviance Table

Poisson model

Response: Bosanma.Sayisi

Terms added sequentially (first to last)

	Df	Deviance	Resid.	Df	Resid. Dev
NULL				69	205938.3
Yas	6	200830.7		63	5107.6

Ek 27 On Beşinci Model için Elde Edilen Bilgisayar Çıktısı

*** Generalized Linear Model ***

Call: glm(formula = Bosanma.Sayisi ~ Yil + Yas + Yil:Yas, family = poisson(link = log), data = hande3, na.action = na.exclude, control = list(epsilon = 0.0001, maxit = 50, trace = F))

Deviance Residuals:

Min	1Q	Median	3Q	Max
-4.636821	-2.349129	-0.8180281	2.072916	6.373225

Coefficients:

	Value	Std. Error	t value
(Intercept)	-54.189068667	1.4028784946	-38.627058
Yil	0.031433386	0.0007036579	44.671406
Yas1	-10.730178742	1.4224223630	-7.543595
Yas2	-13.780413488	0.9404811606	-14.652514
Yas3	-14.464656250	0.8120802218	-17.811856
Yas4	-8.227749819	0.8020783371	-10.258038
Yas5	3.100023781	0.8352250408	3.711603
Yas6	11.900508182	0.8673570782	13.720426
YilYas1	0.005349666	0.0007134661	7.498136
YilYas2	0.006848427	0.0004717022	14.518539
YilYas3	0.007171906	0.0004072752	17.609482
YilYas4	0.004027948	0.0004022570	10.013368
YilYas5	-0.001661399	0.0004189215	-3.965897
YilYas6	-0.006074233	0.0004350975	-13.960626

(Dispersion Parameter for Poisson family taken to be 1)

Null Deviance: 205938.3 on 69 degrees of freedom

Residual Deviance: 617.9975 on 56 degrees of freedom

Number of Fisher Scoring Iterations: 3

Analysis of Deviance Table

Poisson model

Response: Bosanma.Sayisi

Terms added sequentially (first to last)

	Df	Deviance	Resid.	Df	Resid. Dev
NULL				69	205938.3
Yil	1	3550.9		68	202387.3
Yas	6	200830.7		62	1556.7
Yil:Yas	6	938.7		56	618.0

Ek 28 Boşanma Verileri

Sıra	Boşanma Sayısı	Yıl	Yaş Grubu
1	1861	1989	15-19
2	7847	1989	20-24
3	11900	1989	25-29
4	9840	1989	30-34
5	6522	1989	35-39
6	4332	1989	40-44
7	2763	1989	45-49
8	1994	1989	50-54
9	1633	1989	55-59
10	2057	1989	60 ve üzeri
11	1852	1990	15-19
12	8074	1990	20-24
13	12084	1990	25-29
14	10352	1990	30-34
15	7038	1990	35-39
16	4402	1990	40-44
17	3006	1990	45-49
18	1789	1990	50-54
19	1357	1990	55-59
20	1770	1990	60 ve üzeri
21	1898	1991	15-19

22	8329	1991	20-24
23	12854	1991	25-29
24	10619	1991	30-34
25	7877	1991	35-39
26	4881	1991	40-44
27	2745	1991	45-49
28	1917	1991	50-54
29	1321	1991	55-59
30	1893	1991	60 ve üzeri
31	1892	1992	15-19
32	8555	1992	20-24
33	12344	1992	25-29
34	10776	1992	30-34
35	7848	1992	35-39
36	4870	1992	40-44
37	2831	1992	45-49
38	1920	1992	50-54
39	1299	1992	55-59
40	1931	1992	60 ve üzeri
41	1976	1993	15-19
42	9149	1993	20-24
43	12298	1993	25-29
44	10929	1993	30-34

45	7668	1993	35-39
46	5061	1993	40-44
47	2997	1993	45-49
48	1954	1993	50-54
49	1306	1993	55-59
50	2112	1993	60 ve üzeri
51	1865	1994	15-19
52	8815	1994	20-24
53	12744	1994	25-29
54	10980	1994	30-34
55	7894	1994	35-39
56	5251	1994	40-44
57	3234	1994	45-49
58	1951	1994	50-54
59	1293	1994	55-59
60	2055	1994	60 ve üzeri
61	1873	1995	15-19
62	8851	1995	20-24
63	12589	1995	25-29
64	11573	1995	30-34
65	8652	1995	35-39
66	5685	1995	40-44
67	3464	1995	45-49

68	1955	1995	50-54
69	1250	1995	55-59
70	1858	1995	60 ve üzeri
71	1833	1996	15-19
72	8908	1996	20-24
73	12729	1996	25-29
74	11857	1996	30-34
75	8761	1996	35-39
76	5932	1996	40-44
77	3745	1996	45-49
78	2087	1996	50-54
79	1318	1996	55-59
80	1934	1996	60 ve üzeri
81	1933	1997	15-19
82	9564	1997	20-24
83	14091	1997	25-29
84	12691	1997	30-34
85	9744	1997	35-39
86	6944	1997	40-44
87	4299	1997	45-49
88	2500	1997	50-54
89	1539	1997	55-59
90	2129	1997	60 ve üzeri

91	1809	1998	15-19
92	8884	1998	20-24
93	13965	1998	25-29
94	12496	1998	30-34
95	9869	1998	35-39
96	7105	1998	40-44
97	4334	1998	45-49
98	2431	1998	50-54
99	1385	1998	55-59
100	2056	1998	60 ve üzeri

Ek 29 Boşanma Verileri2

Sıra	Boşanma Sayısı	Yıl	Yaş Grubu
1	11900	1989	25-29
2	9840	1989	30-34
3	6522	1989	35-39
4	4332	1989	40-44
5	2763	1989	45-49
6	1994	1989	50-54
7	1633	1989	55-59
8	2057	1989	60 ve üzeri
9	12084	1990	25-29
10	10352	1990	30-34
11	7038	1990	35-39
12	4402	1990	40-44
13	3006	1990	45-49
14	1789	1990	50-54
15	1357	1990	55-59
16	1770	1990	60 ve üzeri
17	12854	1991	25-29
18	10619	1991	30-34
19	7877	1991	35-39
20	4881	1991	40-44
21	2745	1991	45-49

22	1917	1991	50-54
23	1321	1991	55-59
24	1893	1991	60 ve üzeri
25	12344	1992	25-29
26	10776	1992	30-34
27	7848	1992	35-39
28	4870	1992	40-44
29	2831	1992	45-49
30	1920	1992	50-54
31	1299	1992	55-59
32	1931	1992	60 ve üzeri
33	12298	1993	25-29
34	10929	1993	30-34
35	7668	1993	35-39
36	5061	1993	40-44
37	2997	1993	45-49
38	1954	1993	50-54
39	1306	1993	55-59
40	2112	1993	60 ve üzeri
41	12744	1994	25-29
42	10980	1994	30-34
43	7894	1994	35-39
44	5251	1994	40-44

45	3234	1994	45-49
46	1951	1994	50-54
47	1293	1994	55-59
48	2055	1994	60 ve üzeri
49	12589	1995	25-29
50	11573	1995	30-34
51	8652	1995	35-39
52	5685	1995	40-44
53	3464	1995	45-49
54	1955	1995	50-54
55	1250	1995	55-59
56	1858	1995	60 ve üzeri
57	12729	1996	25-29
58	11857	1996	30-34
59	8761	1996	35-39
60	5932	1996	40-44
61	3745	1996	45-49
62	2087	1996	50-54
63	1318	1996	55-59
64	1934	1996	60 ve üzeri
65	14091	1997	25-29
66	12691	1997	30-34
67	9744	1997	35-39

68	6944	1997	40-44
69	4299	1997	45-49
70	2500	1997	50-54
71	1539	1997	55-59
72	2129	1997	60 ve üzeri
73	13965	1998	25-29
74	12496	1998	30-34
75	9869	1998	35-39
76	7105	1998	40-44
77	4334	1998	45-49
78	2431	1998	50-54
79	1385	1998	55-59
80	2056	1998	60 ve üzeri

Ek 30 Boşanma Verileri3

Sıra	Boşanma Sayısı	Yıl	Yaş Grubu
1	11900	1989	25-29
2	9840	1989	30-34
3	6522	1989	35-39
4	4332	1989	40-44
5	2763	1989	45-49
6	1994	1989	50-54
7	1633	1989	55-59
8	12084	1990	25-29
9	10352	1990	30-34
10	7038	1990	35-39
11	4402	1990	40-44
12	3006	1990	45-49
13	1789	1990	50-54
14	1357	1990	55-59
15	12854	1991	25-29
16	10619	1991	30-34
17	7877	1991	35-39
18	4881	1991	40-44
19	2745	1991	45-49
20	1917	1991	50-54
21	1321	1991	55-59

22	12344	1992	25-29
23	10776	1992	30-34
24	7848	1992	35-39
25	4870	1992	40-44
26	2831	1992	45-49
27	1920	1992	50-54
28	1299	1992	55-59
29	12298	1993	25-29
30	10929	1993	30-34
31	7668	1993	35-39
32	5061	1993	40-44
33	2997	1993	45-49
34	1954	1993	50-54
35	1306	1993	55-59
36	12744	1994	25-29
37	10980	1994	30-34
38	7894	1994	35-39
39	5251	1994	40-44
40	3234	1994	45-49
41	1951	1994	50-54
42	1293	1994	55-59
43	12589	1995	25-29
44	11573	1995	30-34

45	8652	1995	35-39
46	5685	1995	40-44
47	3464	1995	45-49
48	1955	1995	50-54
49	1250	1995	55-59
50	12729	1996	25-29
51	11857	1996	30-34
52	8761	1996	35-39
53	5932	1996	40-44
54	3745	1996	45-49
55	2087	1996	50-54
56	1318	1996	55-59
57	14091	1997	25-29
58	12691	1997	30-34
59	9744	1997	35-39
60	6944	1997	40-44
61	4299	1997	45-49
62	2500	1997	50-54
63	1539	1997	55-59
64	13965	1998	25-29
65	12496	1998	30-34
66	9869	1998	35-39
67	7105	1998	40-44

68	4334	1998	45-49
69	2431	1998	50-54
70	1385	1998	55-59

ÖZGEÇMİŞ

Doğum tarihi 04.08.1980

Doğum yeri İstanbul

Lise 1994-1997 Kandilli Kız Lisesi

Lisans 1998-2003 Yıldız Teknik Üniversitesi Fen-Edebiyat Fakültesi
İstatistik Bölümü

Yüksek Lisans 2003- Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı