T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BÖLGELERARASI BOŞANMA NEDENLERİNİN İNCELENMESİ

YÜKSEK LİSANS TEZİ

İstatislik Ana Bilim Dalı

Hakan DEMİRELLİ

Danışman: Dr. Öğr. Üyesi Nurhan HALİSDEMİR

AĞUSTOS-2018

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BÖLGELERARASI BOŞANMA NEDENLERİNİN İNCELENMESİ

YÜKSEK LİSANS TEZİ

Hakan DEMİRELLİ
1521333101

Tezin Enstitüye Verildiği Tarih: 26 Temmuz 2018
Tezin Savunulduğu Tarih : 07 Ağustos 2018

Tez Danışmanı : Dr. Öğr. Üyesi Nurhan HALİSDEMİR (FÜ)

Diğer Jüri Üyeleri : Prof. Dr. Sinan ÇALIK (FÜ)

Doç.Dr. Yunus BULUT(İ.Ü)

AĞUSTOS-2018

ÖNSÖZ

Tez çalışmam için gerekli ortamı sağlayan, değerli katkı ve eleştirileriyle çalışmamın sonuca ulaşmasında ve karşılaşılan güçlüklerin aşılmasında yol gösterici olan, özverili yardımları ile her zaman yanımda olan danışman hocam Dr. Öğr. Üyesi Nurhan HALİSDEMİR hocama ve ayrıca yardımını esirgemeyen Arş. Gör. Yunus Güral'a teşekkür ederim.

İÇİNDEKİLER

ÖZET

Çalışma içerisinde, 2005 ile 2017 yılları arasında Türkiye'deki boşanma sayısı verileriyle ilgili bir Poisson regresyon uygulaması yapmak amaçlanmıştır. İlk olarak regresyon analizi ile ilgili bilgi verilmiştir. Regresyon, doğrusal regresyon ve doğrusal olmayan regresyon şeklinde ikiye ayrılmaktadır. Poisson regresyon bir doğrusal olmayan regresyondur. 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. 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 uygulanmaktadır.

Boşanma verileri üzerinde yapılan uygulamada bölgelerden alınan verilerin normal dağılıp dağılmadığı, ortalamaları, parametre tahminleri, bölgelerarası fark olup olmadığı SPSS paket programı kullanarak bakılmıştır. Uygulama bölümünde boşanma verileri poisson regresyon ile modellenmeye çalışılmıştır.

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

SUMMARY

Investıgations of Reasons For District Disasters

In this study, between 2005 and 2017 aimed at making a Poisson regression application data on the number of divorces in Turkey. Initially, information on regression analysis is given. Regression is divided into linear regression and non-linear regression. Poisson regression is a nonlinear regression. Poisson regression analysis is used when the dependent variable is a number of data (count). To date, the Poisson regression model has been investigated under the assumption that the mean and variance for the Poisson distribution are equal. In cases of overexploitation, two paths are followed. The first of these is estimating a weighing parameter, and then correcting the test statistics and remnants. The second is negative binomial regression.

In the implementation of the divorce data, whether the data in the regions are normally distributed, averages, parameter estimates, whether there are regional differences are examined using the SPSS package program. In practice, divorce data were tried to be modeled with poisson regression.

Key Words: Poisson regression, divorce, linear regression, non-linear regression, Negative Binomial Regression

ŞEKİLLER LİSTESİ

TABLOLAR LİSTESİ

SİMGELER ve KISALTMALAR

$\boldsymbol{b_0 \beta_0}$ **:** Tahmincisi

$\boldsymbol{b_1\ \beta_1}$ **:** Tahmincisi

X **:** Bağımsız değişken

Y **:** Bağımlı değişken

$\boldsymbol{X^2}$ **:** Pearson uyum test istatistiği

Y' **:** Y gözlemlerinin ortalaması

N **:** Gözlem sayısı

e **:** Kalıntı

G **:** Sapma

$\boldsymbol{k_i}$ **:** Doğrusal regresyonda kullanılan sabit

L(μ) **:** μ parametre değerinin olabilirlik değeri

$\boldsymbol{o_i}$ **:** Genelleştirilmiş doğrusal modeldeki sabit (offset)

Q **:** En küçük kareler yönteminde sapmaların karelerinin toplamı

r **:** Korelasyon katsayısı

$\boldsymbol{r_i}$ **:** Pearson kalıntıları

$\boldsymbol{R^2}$ **:** Belirginlik katsayısı

s{$\boldsymbol{b_0}$} $\boldsymbol{b_0}'$ın **:** Standart hatası

s{$\boldsymbol{b_1}$} **:** $b_1'in$ Standart hatası

$\boldsymbol{s^2}$ **:** Örnek varyansı

$\boldsymbol{s^2\{b_0\}}$ **:** b_0'ın varyansı

$\boldsymbol{s^2}\{\boldsymbol{b_1}\}$ **:** b_1'in varyansı

$\boldsymbol{s^2\{\widehat{Y_h}\}}$ **:** $\widehat{Y_h}$'ın tahmin edilen varyansı

W **:** Tartı matrisi

X' **:** X gözlemlerinin ortalaması

$\boldsymbol{\widehat{Y}}$ **:** Y'nin tahmin değeri

Z **:** Poisson regresyon modelindeki çalışan rastlantı değişkeni

α **:** Anlamlılık düzeyi

$\boldsymbol{\beta_0}$ **:** Doğrusal regresyon modelindeki sabit parametre

$\boldsymbol{\widehat{\beta_0}\beta_0}$**'ın** **:** Tahmincisi

$\boldsymbol{\beta_1}$ **:** Doğrusal regresyon modelindeki eğim parametresi

$\boldsymbol{\widehat{\beta_1}\ \beta_1}$**'in** **:** Tahmincisi

ε **:** Hata terimi

$\boldsymbol{\gamma_0}$ **:** Doğrusal olmayan regresyon modelinin parametresi

$\boldsymbol{\gamma_1}$ **:** Doğrusal olmayan regresyon modelinin parametresi

$\boldsymbol{\gamma_2}$ **:** Doğrusal olmayan regresyon modelinin parametresi

K : Negatif binom dağılımının parametresi

$\boldsymbol{\lambda_{ij}}$**(i, j)'inci :** Grup içindeki gerçek risk

$\boldsymbol{\lambda(X_i,\beta)}$ **:** Her i alt grubu için hata oranını gösteren X_i ve β'nın özel bir fonksiyonu

$\boldsymbol{\sigma\{b_0\}}$: $\boldsymbol{b_0}$'in standart sapması

$\boldsymbol{\sigma\{b_1\}}$ **:** b_1'in standart sapması

σ^2 : Varyans

$\widehat{\sigma^2}$: σ^2'nin tahmincisi

$\sigma^2\{b_0\}$: b_0'in varyansı

$\sigma^2\{b_1\}$: b_1'in varyansı

$\sigma^2\{\widehat{Y_h}\}$: Y_h'ın varyansı

ANOVA : Analysis of Variance

EM : Expectation Maximization Algorithm

IRWLS : Iteratively Reweighted Least Squares

ZIP : Zero Inflated Poisson

ZAP : ZeroAltered Poisson

SSE : Sum of SquaresDuetoError

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 şekilde açıklayan en sade yapıya sahip olan model, genel olarak en uygun model olarak değerlendirilmektedir [1].

Poisson regresyon, lojistik regresyondan sonra en genel olan ikinci genelleştirilmiş model olup doğrusaldır. Bağımlı değişken oluş sayısı (count) ile belirtilen bir veri olduğunda, yani belirli bir yerde ya da belirli bir zamanda 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 olup logaritmiktir [2].

Poisson regresyon genellikle sağlık çalışmalarında özellikle kanser araştırmalarında yoğunlukla kullanılmaktadır [1].

Bu zamana kadar yapılan çalışmalarda, Poisson dağılımı için ortalama ve varyansın eşit olması varsayımı altında poisson regresyon modeli incelenmiştir. Poisson dağılımında; ortalamanın varyanstan küçük olması haline aşırı yayılım denilmektedir. Böyle bir durum söz konusu olduğunda tartı parametresinin incelenmesi gerekmektedir. Tartı parametresi incelendikten sonra bu tarz bir durum varsa önce negatif binom regresyon modeli uygulanmalı ya da tartı parametresi ile test istatistikleri ve kalıntılar düzeltilmelidir [3].

Boşanma verileri üzerinde yapılan uygulamada bölgelerden alınan verilerin normal dağılıp dağılmadığı, ortalamaları, parametre tahminleri, bölgelerarası fark olup olmadığı SPSS paket programı kullanarak bakılmıştır. Uygulama bölümünde boşanma verileri poisson regresyon ile modellenmeye çalışılmıştır.

1.1. Türkiye'de Boşanmalar

Araştırma konusuna bağlı olarak literatür taraması yapıldığında Türkiye'de boşanma konusunun geniş şekilde ele alındığı ancak bölgelerarası farklılık ve boşanma sebepleri ile ilgili araştırma konularının ise kısıtlı olduğu görülmektedir. Bugüne kadar TUİK tarafından tutulan resmi istatistikler incelendiğinde, Türkiye'deki boşanma oranlarında her yıl değişim olduğu görülmektedir. Yıldırım N. tarafından 2000 yılında yapılan bir araştırma sonucunda elde edilen bulgular ile bu çalışmada elde edilen bulgular örtüşmektedir. Mesela boşanma oranlarının bölgesel kalkınmışlık kriterlerine göre bakıldığında; Sanayisi gelişmiş bölgelerde boşanma oranı, gelişmemiş bölgelerdeki orana göre daha yüksektir. 2000 yılı içinde Şırnak'ta 19 boşanma yaşanırken, aynı yıl İstanbul'da 6546, Ankara'da 2217, İzmir'de ise 4406 boşanma vakası meydana gelmiştir [4]. Bu çalışmada ise 2017 yılı baz alındığında; Doğu Anadolu Bölgesinde 1634 boşanma vakası, Güneydoğu Anadolu Bölgesinde 4227 boşanma vakası, İç Anadolu Bölgesinde 13222 boşanma vakası, Akdeniz Bölgesinde 9780 boşanma vakası, Karadeniz Bölgesinde 3331 boşanma vakası, Marmara Bölgesinde 10678 boşanma vakası, Ege Bölgesinde 15744 boşanma vakası yaşanmıştır [5]. Bu sayılardan hareketle, ülkemizin batı bölgeleri doğu bölgelere, ekonominin geliştiği bölgeler gelişmemiş bölgelere, kentsel kesimler kırsal kesimlere göre boşanma oranının daha yüksek olduğu yerler olarak düşünülebilir.

Diğer yandan Neşide Yıldırım tarafından 2010 yılında yapılan çalışma incelendiğinde Türkiye'de boşanma sebeplerini etkileyen en çarpıcı faktörler sırasıyla geçimsizlik, eğitim seviyesinin düşüklüğü, kadının özgür olmaması, erkek egemenliğinin özellikle Doğu, Güneydoğu, İç Anadolu ve Karadeniz bölgelerinde çok etkin olması, aşırı derecede alkol kullanma, hakaretler, kapalı aile gelenekleri gibi şeklinde tespit edilmiştir. Neşide Yıldırım'ın elde ettiği bulgular ile tarafımızdan yapılan çalışmanın sonuçları örtüşmektedir [6].

Kamu Yönetimi Uzmanı Sedat ERGENÇ 2010 yılında Türkiye'deki boşanma nedenlerini incelemiş ve ağırlıklı olarak en önemli faktörün "geçimsizlik" olduğunu tespit etmiştir. TÜİK boşanma nedenlerini gruplandırırken geçimsizlik başlığının altında özellikle "zina, cürüm, kötü muamele gibi" faktörlerin gizlendiğini öne sürülmektedir. Sedat Ergenç tarafından yapılan bu tespit tarafımızdan yapılan bu çalışmada da kendini göstermektedir. Çalışmamızda 2017 baz alındığında boşanma nedenlerinin neredeyse %98'i geçimsizlik olarak görülmektedir. Sedat Ergenç'in bu konudaki

tespitlerine katılmakla beraber geçimsizlik faktörü altında gizlenmiş olan faktörlerin oransal büyüklüğü ve etkisi hakkında kesin bir yargıda bulunmanın ise doğru olmayacağı söylenebilir [7].

Bölgelerarasında boşanmaların il olarak en fazla İstanbul'da gerçekleştiği ve haliyle Marmara bölgesinin en fazla boşanma görülen bölge olmasında İstanbul'un etkisi görülmektedir. Marmara bölgesini ise sırasıyla Ege ve Akdeniz bölgeleri takip etmektedir. Boşanmaların en az olduğu bölgeler ise sırasıyla Kuzeydoğu ve Orta Doğu Anadolu (Doğu Anadolu) ile Doğu Karadeniz bölgelerinin olduğu görülmektedir. Yine boşanma sebeplerini ve bölgesel dağılımlarını Türkiye Kamu-Sen gibi STK'larda araştırmışlardır. Ancak bu araştırmaların siyasi boyutu göz önüne alınarak kaynak olarak çalışmamızda kullanılmamıştır [8].

2. REGRESYON ANALİZİ

Regresyon analizi, araştırmak istediğimiz bağımlı değişkenin ya da değişkenlerin üzerinde bağımsız değişkenlerin etkisi olup olmadığını ve aralarındaki ilişkiyi araştıran bir yöntemdir. Verinin yapısına göre regresyon yöntemleri de değişmektedir. Araştırılacak bağımlı değişken kategorikte olabilir, aralıklı sayılardan da olabilir. ($Kanser = 0, Kanser\ değil = 1$) kategorik bir değişkendir. Mikrodizi çipinde üretilen aralıklı $(155{,}6;\ 152{,}4; \ldots)$ bir değişken gibi de olabilir [9].

Regresyon analizi, sebep-sonuç ilişkisi bulunan iki veya daha fazla değişken arasındaki ilişkiyi belirlemek ve bu ilişkiyi kullanarak o konu ile ilgili kestirimler ya da tahmin yapabilmek amacıyla kullanılır. Doğada birçok olayda sebep-sonuç ilişkisi bulunmaktadır. Örneğin; yaş-boy, gelir-harcama, gübre-verim, verilen yem miktarı-alınan süt miktarı vs.

Regresyon analizi, bir bağımlı değişken ile bir bağımsız değişken (basit regresyon) veya birden daha fazla bağımsız (çoklu regresyon) değişken arasındaki ilişkilerin matematiksel eşitlik ile açıklanması sürecidir. Bu ilişki aşağıdaki Şekil 2.1'de gösterilmiştir. Bu yöntemin matematiksel ifadesine regresyon denklemi diyoruz [10].

Regresyon denklemi genel ifadesi $Y' = a + bX$

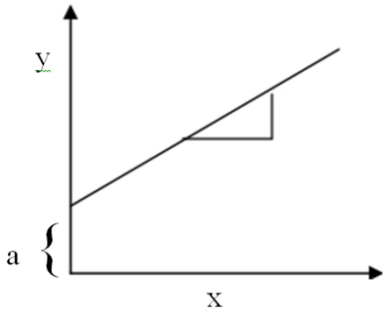
X : Seçilen bağımsız değişkenin değeri

Y' : Seçilmiş X değeri için tahmin edilen Y değeri

a : Doğrunun Y eksenini kestiği noktanın değeri

b : Doğrunun eğimi

$a\ ve\ b$: regresyon katsayıları



Şekil 2.1. Regresyon denklemindeki ilişki

Regresyon analizi ile bağımlı ve bağımsız değişkenler arasında bir ilişki var mıdır? Eğer bir ilişki varsa bunun gücü nedir? Değişkenler arasında ne tür bir ilişki vardır? Bağımlı değişkene ait ileriye dönük değerleri tahmin etmek mümkün müdür ve nasıl

tahmin edilmelidir? Belirli koşulların kontrol edilmesi durumunda özel bir değişken veya değişkenler grubunun diğer değişkenin veya değişkenler üzerindeki etkisi nedir ve nasıl değişir? gibi sorulara cevap aranmaya çalışılır.

Regresyon analizi, bilinen bulgulardan bilinmeyen gelecekteki olaylarla ilgili kestirimler yapılmasına izin verir. Regresyon, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi ve doğrusal eğri kavramını kullanarak bir tahmin eşitliği geliştirir. Değişkenler arasındaki ilişki belirlendikten sonra bağımsız değişkenlerin değeri bilindiğinde bağımlı değişkenin değeri tahmin edilebilir [11].

Bağımlı değişken (y)

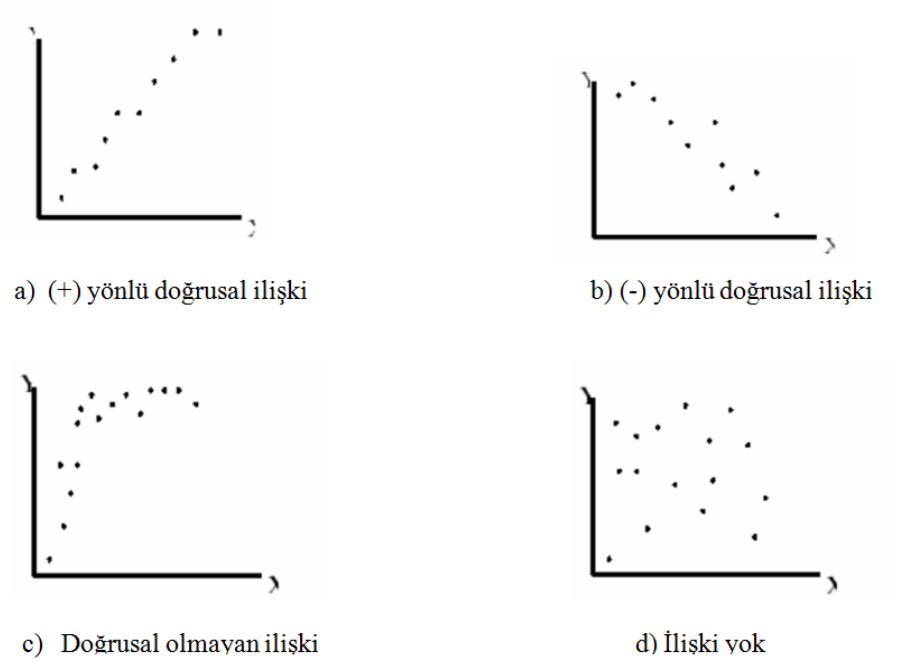
Bağımlı değişken, regresyon modelinde tahmin edilen ya da açıklayıcı değişkendir. Bu değişkenin bağımsız değişken ile ilişkili olduğu varsayılmaktadır.

Bağımsız değişken (x)

Bağımsız değişken, regresyon modelinde açıklayıcı değişken olup; bağımlı değişkenin değerini tahmin etmek için kullanılır.

- Değişkenler arasında doğrusal ilişki olabileceği gibi doğrusal olmayan bir ilişki de olabilir. Bu nedenle, değişkenler arasında korelasyon varlığına rastlanmadan ve saçılım grafiği yapılmadan regresyon modeli analizine karar verilmemesi gerekir.
- Bu bilgiler doğrultusunda tek/çok değişkenli doğrusal regresyon analizlerinin yanı sıra tek/çok değişkenli doğrusal olmayan regresyon analizleri de mevcuttur.

  Regresyon analizinde ilişki türleri aşağıdaki Şekil 2.2 ’de gösterilmiştir.



Şekil 2.2. İlişki türleri

Regresyon analizinde üç türlü amaç gözetilebilir:

- Teorinin test edilmesi: örneğin trafik kazaları ile alkol tüketimi arasında artan bir fonksiyonel ilişki ileri sürülüyorsa, bu iddianın testi regresyon analizi ile araştırılabilir.
- Politika tespiti: örneğin bir bölgede bir işletmenin yeni bir departman açmayı düşünüyorsa, o bölgede kendi malına olan talebi, talep fonksiyonu regresyonuyla araştırdıktan sonra karar verebilir.
- Geleceğe dönük ön tahmin: örneğin bir gazetenin aylık tiraj rakamları ile aylık reklam harcamaları arasında doğrusal artan bir regresyon bulunmuşsa, öngörülen daha büyük bir reklam harcaması karşılığında gazetenin muhtemel aylık tirajının ne olacağı bu regresyon yardımıyla tahmin edilebilir [10].

2.1. Tek Değişkenli Regresyon Analizi

Tek değişkenli regresyon analizi, bir bağımlı değişken ve bir bağımsız değişken arasındaki ilişkiyi inceleyen analiz yöntemidir. Bu analiz yöntemiyle bağımlı ve bağımsız değişkenler arasında doğrusal ilişkiyi ifade eden bir doğru denklemi formüle edilmektedir. Korelasyon analizinde olduğu gibi regresyon analizinde üzerinde durulan ilişki değişkenler

arasındaki ilişki doğrusal ilişkidir ve bu doğrunun hesaplanması ise en küçük kareler yöntemi yardımıyla yapılmaktadır. Basit regresyon denklemi

$$Y = a + bX + \varepsilon$$

şeklinde bir bağımlı değişken bir bağımsız değişken içeren bir modeldir [12].

Regresyon analizi sonuçlarının yorumlanmasında birçok araştırmacı ve öğrenci tarafından önemli hatalar yapılmaktadır. Bunlardan en yaygın olan hata, regresyon analizi sonuçlarının yorumlanmasında x bağımsız değişkeninin y bağımlı değişkenine neden olduğu şeklindeki yorumudur. Bağımsız değişkenlerin bağımlı değişkendeki değişimi açıklıyor olması sebepselliği gerekli kılmaz. Başka bir ifadeyle bağımlı ve bağımsız değişkenler arasında (pozitif ve negatif) bir ilişkinin olması her zaman bağımsız değişkenlerin bağımlı değişkenin neden olduğu sonucunu doğurmayacaktır [12].

İki değişken arasında bir ilişkinin olabilmesi için sebepsellik şart değildir. İlişkinin sebebi belki de iki değişkenin üçüncü bir değişkenle olan ilişkilerinden kaynaklanıyor olabilir ya da söz konusu ilişki tamamen tesadüfi olarak da ortaya çıkmış olabilir. Sebepsellik ile ilişkiselliğin aynı olmadığı unutulmamalıdır. Regresyon analizi değişkenler arasındaki ilişkinin yapısı ve değişkenler arasındaki ilişkinin derecesi ile ilgilenir [10].

2.2. En Küçük Kareler Yöntemi

Regresyon katsayılarının tahmini için en çok kullanılan yöntem en küçük kareler yöntemidir. Regresyon doğrusunun gözlem değerlerini en iyi bir şekilde temsil etmesi için, bu gözlem noktalarını tam olarak ortalaması gerekmektedir. Bu şekilde e_i kalıntıları minimize edilmiş olacağından ve bunun için En küçük kareler yönteminde gerçek ilişkiye yönelik eklenen bir u değişkeni hakkında şu varsayımları yapıyoruz:

- u bir rassal değişkendir.
- u rassal değişkeninin beklenen değeri sıfırdır. $E(u_i) = 0$.
- u rassal değişkeninin varyansı sabittir. $Var(u_i) = \sigma^2$
- u rassal değişkeni normal dağılıma sahiptir. u_i-$N(0, \sigma^2)$.
- u rassal değişkeninin farklı terimleri arasındaki korelasyon sıfırdır. $Kor(u_i u_j) = 0$.

- u rassal değişkeni açıklayıcı değişkenlerden bağımsızdır. $Kor(u_i X_i) = 0$.

Bu şartlar ardında, kalıntı kareleri toplamını minimize eden b_0 ve b_1 değerleri tesbit edilerek regresyon katsayılarının bulunması en küçük kareler kriteri olarak bilinmekte ve bu yöntem şöyle işlenmektedir.

$$Min \sum e_i^2 = Min \sum (Y_i - Y')^2 = Min \sum (Y_i - b_c - b_i X_i)^2 \qquad \textbf{(2.1)}$$

İkinci dereceden bir fonksiyonun minumunu olması için birinci türevlerinin sıfır olması gerekir. Buna göre yukarıdaki ifadenin b_0 ve b_1 için ayrı ayrı türevleri sıfıra eşitlenerek gerekli sadeleştirmeleri yaptığımızda normal denklemler dediğimiz,

$$\sum Y_i = b_0(n) + b_1 \sum X_i \sum Y_i X_i = b_0 \sum X_i + b_1 \sum X_i^2 \qquad \textbf{(2.2)}$$

denklem sistemi elde edilir. Burada b_0 ve b_1 katsayıları bilinmeyenleri oluştururken, bunların dışındaki diğer terimler bilinen değerlerdir. Bilinmeyenler denklemden çekildiğinde,

$$b_0 = \frac{\sum X^2 \sum Y - \sum X \sum XY}{n \sum X^2 - (\sum X)^2} \; b_1 = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum olX)^2} \qquad \textbf{(2.3)}$$

olarak elde edilir. Bulunan b_0 ve b_1 değerleri regresyon doğrusunda $Y'_i = b_0 + b_1 X_i$ yerine yazılmak suretiyle regresyon tahmini tamamlanmış olur.

Böyle bir regresyon tahmininde örneğin X işletmenin reklam harcamaları, Y satışlarını gösteriyorsa, reklam harcamaları önceden belirlenerek satışların buna göre gelecek bir dönemdeki ön tahmini elde edilebilir. Regresyon, bu şekilde işletme yöneticilerine bir karar alma aracı olarak işlev görebilir.

Regresyonda gözlem sayısı n ve tahmin edilen parametre sayısı k olmak üzere $n - k$ değerine serbestlik derecesi adı verilir. Bilindiği gibi, basit regresyonda b_0 ve b_1 olmak üzere iki katsayı mevcut olduğundan $k = 2$ dir. Tahmin edilen b_0 ve b_1 katsayılarının gerçek değerlere yakın olması için uygulamalarda serbestlik derecesi $(n - k)$ nın yüksek olması istenir. Bunun içinde gözlem sayısının arttırılmasına çalışılır [13].

2.3. Sapmalarda Tahmin

Yukarıda verilen EKK yöntemi ile b_0 ve b_1 katsayılarını bulan formüller orijinal değerler cinsinden ifade edilmiştir. X_i ve Y_i gözlem değerlerinden bunların ortalamalarını çıkarmak suretiyle sapmalar cinsinden ifade edersek x_i ve y_i değişkenlerini buluruz. Bunun için birinci normal denklemi yeniden yazalım,

$$\sum Y_i = b_0(n) + b_1 \sum X_i \qquad \textbf{(2.4)}$$

Denklemin her iki yanını n ile bölersek, Y'=$b_0 + b_1 X'$ olur. Regresyon denklemi ile bu ilişkiyi alt alta yazarak taraf tarafa çıkarma işlemi ile:

$$Y_i = b_0 + b_1 X_i + e_i$$

$$Y' = b_0 + b_1 X'$$

$$y_i = b_1 x_i + e_i \qquad \textbf{(2.5)}$$

Şeklinde ortalamalardan sapmalar cinsinden model elde edilir. Bu modelin özelliği kesme teriminin modelden düşmüş olmasıdır. Bu modeldeki kalıntıların kareleri

$\sum e_i^2 = \sum (y_i - b_1 x_i)^2$ toplamına minimize kriterini uyguladığımızda, yani $b_1{}'e$ göre 1. Türevi sıfıra eşitliğimizde,

$b_1 = \frac{\sum xy}{\sum x^2}$ şeklinde b_1 katsaysını sapmalar cinsinden veren ifadeyi buluruz. Öteyandan,

$Y' = b_0 + b_1 X'$ denkleminden b_0 çekilirse $b_0 = Y'$-$b_1 X'$ ifadesine ulaşırız. Daha önce sapmalar cinsinden bulunan b_1 değeri $b_0 = Y' - b_1 X'$ eşitliğinde yerine konularak b_0 kesme terimi hesaplanır **[13-1]**.

2.4. Katsayıların Hesaplama Formülü

Bir önceki kesimde sapmalar cinsinden veriler eğim katsayısı formülünün pay ve paydası açılarak sadeleştirilirse eğim katsayısı için hesaplama formülü adı verilen daha basit ifadeye ulaşılır. Sapmalar cinsinden eğim katsayısı,

$$b_i = \frac{\sum xy}{\sum x^2} = \frac{\sum (X - X')(Y - Y')}{\sum (X - X')^2} \qquad \textbf{(2.6)}$$

olarak verilmişti. Paydaki çarpma işlemi sadeleştirilirse ve paydada iki parantezin de karesi alınarak sadeleştirilirse,

$$b_1 = \frac{\sum XY - nX'Y'}{\sum X^2 - nX'^2} \quad \textbf{(2.7)}$$

bulunur. Zaten kesme terimi de eğim katsayısı yardımıyla b_0=Y'-b_1X'eşitliğinden bulunuyordu.

Bu şekilde basit regresyon katsayılarının hesaplanması için geliştirilmiş ifadeleri topluca yazacak olursak,

1. EKK kriteri ile elde edilen orijinal formüller:

$$b_0 = \frac{\sum X^2 \sum Y - \sum X \sum XY}{n\sum X^2 - nX'^2} \qquad b_1 = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2}$$

2. Sapmalar cinsinden EKK formülleri:

$$b_1 = \frac{\sum xy}{\sum x^2} \qquad b_0 = Y' - b_1X'$$

3. Hesaplama formülleri:

$$b_1 = \frac{\sum XY - nX'Y'}{\sum X^2 - nX'^2} \qquad b_0 = Y' - b_1X'$$

Her üç formül setinin de aynı sonuçları verceği tabiidir. Uygulamada çoğu kez hesaplama formülleri tercih edilmektedir [13].

2.5. Regresyonun Standart Hatası

Gözlem noktalarının regresyon doğrusundan sapmalarına regresyon kalıntıları adı verildiğini daha önce belirtmiştik. Gözlem noktalarına ilişkin model, $Y_i = b_0 + b_1X_i + e$ regresyon modeli $Y'_i = b_0 + b_1X_i$ buradan $Y_i = Y'_i + e_i$ ve e_i=$Y_i - Y'_i$. Kalıntıların bir kısmı pozitif bir kısmı negatif çıkar ve bunların toplamı sıfır olur. Dolayısıyla kalıntıların aritmetik ortalaması e'=0 çıkar. Kalıntıların ortalamadan sapmalarının karesi, $\sum(e_i - e')^2 - \sum e_i^2$ $olur$. Buradan kalıntıların varyansı,

$$\sigma^2 = \frac{\sum e_i^2}{n-k} \quad \textbf{(2.8)}$$

olarak bulunur. k regresyondaki katsayı sayısıdır. Kalıntıların varyansına kadar küçük olursa noktaların doğru etrafında kümelenmesi o ölçüde yakın olur. Kalıntıların varyansının pozitif karekökü regresyonun standart hatası olarak bilinir. Buna göre kalıntıların varyansını bulmak için sırasıyla şu işlemler yapılmalıdır:

- **Y'_i teorik değerlerinin ortalaması**: tahmin edilen b_0 ve b_1 katsayıları ve X_i değişkeninin değerleri $Y'_i = b_0 + b_1 X_i$ regresyonunda yerine konularak her bir X_i $için$ bir Y'_i üretilir. Böylece n tane Y'_i değeri bulunmuş olur. Bu Y'_i değerleri serpilme diyagramında regresyon doğrusu üzerinde yer alır. Gözlenen değerlerle teorik değerlerin toplamı eşit çıkmalıdır. Yani, $\sum Y_i = \sum Y'_i$.

- **e_i kalıntı değerlerinin bulunması**: daha önce de ifade edildiği gibi $e_{i=}Y_i - Y'_i$ eşitliği vardır. Birinci adımda bulunan her bir Y'_i teorik değeri karşı gelen Y_i gözlem değerlerinden çıkartılarak n tane e_i kalıntıları bulunur. bu kalıntıları cebirsel toplamı 0 çıkmalıdır. Yani, $\sum e_i = 0$.

- **Varyansın hesaplanması**: e_i kalıntılarının her birinin karesi alınarak bütün terimler toplamak suretiyle $\sum e_i^2$ bulunur. bu değeri varyans formülünde yerine yazarak

$\sigma^2 = \frac{\sum e_i^2}{n-k}$ kalıntıların varyansı elde edilir [13].

2.6. Determinasyon Katsayısı

Tahmin edilen bir regresyonun genel başarısı yüzdelik bir derece olarak determinasyon katsayısı ile ölçülür. R^2 ile gösterilen determinasyon katsayısı, basit regresyon için bağımlı değişken ve bağımsız değişkenler arasındaki basit korelasyon katsayısının karesinden başka bir şey değildir. Yani,

$$R^2 = \frac{(\sum xy)^2}{\sum x^2 \sum y^2}$$

Öte yandan eğim katsayısı formülünün

$$b_1 = \frac{\sum xy}{\sum x^2}$$

olduğu hatırlanırsa determinasyon katsayısını b_1 cinsinden

$R^2 = b_1 \frac{\sum xy}{\sum y^2}$ olarak da yazılabileceği görülür.

R^2 nin değişim aralığını bulmak zor değildir. Mademki determinasyon katsayısı korelasyon katsayısının karesidir, korelasyon katsayısının değişim aralığı olan [-1, +1] aralığındaki sayıların karesi alındığında, değişim aralığı, [0, +1] aralığındaki pozitif sayılardan ibaret olur. Determinasyon katsayısı,

R^2= 0 ise değişkenler arasında ilişki yoktur.

$R^2 < 0.50$ ise zayıf fonksiyonel ilişki,

R^2> 0.50 ise kuvvetli fonksiyonel ilişki,

$R^2 = 1$ ise iki tam (deterministik) bir ilişki vardır [14].

2.7. Anlamlılık Testleri

Tahmin edilen regresyonun genel başarısını determinasyon katsayısı ile ölçülebileceğini belirtmiştik. Bir regresyonda ayrıca tahmin edilen her bir katsayının istatistiksel anlamlılığı da test edilmelidir. b_0 ve b_1 katsayılarının örnekten örneğe tahmin edilen değerleri değiştiğinden bunlar birer rassal değişken olup bunların örneklem dağılımı, büyük örneklem hacimlerinde Z dağılımına, küçük örneklem hacimlerinde ise t dağılımına uyar. Test işlemi daha önceki bölümlerde olduğu gibi dört aşamadan oluşur: Hipotezlerin kurulması, Kritik bölgenin belirlenmesi, Test istatistiğinin hesaplanması ve Karar.

Hipotezler, bilindiği gibi anakütle değerleri hakkında kurulur. Burada da hipotezlerimiz β_0 veya β_1 hakkında olacaktır. Örneğin X ile Y arasında doğrusal bir ilişki varsa haliyle β_1 in sıfırdan farklı olduğu ($\beta_1 \neq 0$) iddia edilecektir. Bu iddia araştırma hipotezi olup, çift yanlı bir testi gerekilir. Buradaki sıfır hipotezi tabi atiyle ($\beta_1 = 0$) şeklinde olacaktır.

- Testin a anlamlılık düzeyi belirlenir. Tekyanlı "küçüktür" şeklindeki araştırma hipotezleri için red bölgesi olan a alanı sol tarafla "büyüktür" şeklindeki araştırma hipotezleri için red bölgesi olan a alanı sağ tarafta yer alır. "eşit değildir" şeklindeki iki yanlı araştırma hipotezleri için red bölgeleri olan $a/2$ alanları her iki uçta yer alır.

- Örneklem dağılımının gerçek varyansı biliniyorsa veya örnek çapı 30 dan büyükse kritik bölge Z tablosundan oluşturulur. Aksi halde yani hem örneklem

dağılımının gerçek varyansı bilinmiyor, hem de örnek çapı 30' dan küçükse kritik bölge t tablosundan oluşturulur.

Test istatistiğinin oluşturulması için tahmin edicinin beklenen değeri ve standart hatası bilinmelidir. EKK kriterine göre bulunan b_0 ve b_1 tahmin edicilerin beklenen değer ve varyanslarını ispatsız olarak verelim:

$$E(b_0) = \beta_0 \quad Var(b_0) = \sigma^2 \frac{\sum X^2}{n\sum x^2} \qquad sh(b_0) = \sqrt{\sigma^2 \frac{\sum X^2}{n\sum x^2}}$$

$$E(b_1) = \beta_1 \qquad Var(b_1) = \sigma^2 \frac{1}{\sum x^2} \quad sh(b_{1)} = \sqrt{\sigma^2 \frac{1}{\sum x^2}}$$

Ancak test işlemi için tahmin edicinin beklenen değeri, bilindiği gibi sıfır hipotezinde iddia edilen değerdir. Dolayısıyla beklenen değer hipotezden hipoteze değişir. Öte yandan standart hata formüllerinde bir benzerlik görülmektedir. b_1'in standart hata formülü daha basit olduğundan, b_0'ın standart hatasını b_1'in standart hatası cinsinden,

$$sh(b_0) = sh(b_1)\sqrt{\frac{\sum X^2}{n}} \qquad \textbf{(2.9)}$$

olarak yazabiliriz. Böylece bir defa $sh(b_1)$ bulunduktan sonra $sh(b_0)$ bunun aracılığıyla kısa yoldan hesaplanır [13].

3. ÇOKLU REGRESYON

Regresyon analizinde bağımsız değişken sayısı birden fazla olduğunda çoklu regresyon yöntemi durumu ortaya çıkar. Talep fonksiyonu örneğinde fiyatın yanı sıra fertlerin gelir düzeyleri de açıklayıcı değişken olarak modele dahil edilebilir. Y bağımlı değişken ve X_1 ve X_2 iki bağımsız değişken olmak üzere regresyon modeli $Y = b_0 + b_1X_1 + b_2X_2 + e$ olarak belirlenir. Bu şekilde üç değişkenli bir modeli grafik olarak üç boyutlu bir uzayda göstermek gerekir ki, bu üç boyutlu grafiğin gösterilmesini burada ihmal ederek, sadece katsayıların nasıl elde edileceğine değinilecektir. Çoklu regresyon denkleminde e kalıntı terimi bir tarafta yalnız bırakılır ve her iki tarafın karesi alınarak terimler boyunca toplama notasyonuna geçilirse $\sum e^2 = \sum(Y - b_0 - b_1X_1 - b_2X_2)^2$ EKK kriterini buluruz. Bu ifadenin sırasıyla $b_0, b_{1,,} b_2$ için kısmi türevleri sıfıra eşitlenirse aşağıdaki üç normal denklem ortaya çıkar.

$$\sum Y = b_0 n + b_1 \sum X_1 + b_2 \sum X_2$$

$$\sum YX_1 = b_0 \sum X_1 + b \sum X_1^2 + c \sum X_1X_2$$

$$\sum YX_2 = b_0 \sum X_2 + b_1 \sum X_1X_2 + b_2 \sum X_2^2,$$

Bu defa üç bilinmeyenli üç denklemli bir sistemle karşı karşıyayız. Sistemde toplama notasyonu ile ifade edilen değerler bilinenleri oluştururken b_0, b_1, b_2 katsayıları ise bilinmeyenlerdir. Bu şekilde çok denklemli sistemlerin çözümü için bazı matris yöntemleri mevcut olmakla birlikte burada bu yöntemlere girilmeyecektir [15].

3.1. Sapmalarda Tahmin

Modelin sapmalar cinsinden ifadesi çoklu regresyonda özellik tercih edebilir. Çünkü daha derli toplu bir model ve daha az sayıda denklem ortaya çıkacaktır. Yukarıda birinci normal denklem, $\sum Y = b_0 n + b_1 \sum X_1 + b_2 \sum X_2$ olarak bulunmuştur. Denklemin her iki yanını gözlem sayısı n ile bölersek $Y' = b_0 + b_1X'_1 + b_2X'_2$. Bu ortalamalar denklemini orijinal modelden taraf tarafa çıkarırsak,

$$Y = b_0 + b_1X_1 + b_2X_2 + e$$

$$Y' = b_0 + b_1 X'_1 + b_2 X'_2$$

$$y = b_1 x_1 + b_2 x_2 + e \quad \textbf{(3.1)}$$

bulunur. Bu son modelde e kalıntılarını bir tarafta yalnız bırakarak her iki tarafın karesi alındıktan sonra toplama notasyonuna geçelim:

$$\sum e^2 = \sum, (y - b_1 x_1 - b_2 x_2)^2 \quad \textbf{(3.2)}$$

Bu son ifadeye EKK kriteri uygulanarak $b_1 \; ve \; b_2$ için kısmi türev alınırsa,

$$\sum y x_1 = b_1 \sum x_1^2 + b_2 \sum x_1 x_2$$

$$\sum y x_2 = b_1 \sum x_1 x_2 + b_2 \sum x_2^2$$

İki denklemden oluşan sapmalar cinsinden normal denklemlere ulaşılır ki, bu sistemin b_1 ve b_2 için çözümü, üç denklemli sisteme nazaran daha kolaydır. Örneğin yok etme metoduyla ile: b_2 katsayısı birinci denklemden çekilir, ikinci denklemdeki yerine yazılırsa,

$$b_1 = \frac{(\sum x_1 y)(\sum x_2^2) - (\sum x_2 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2} \quad \textbf{(3.3)}$$

bulunur. $b_1 \; için$ bulunan bu ifade tekrar birinci denklemdeki yerine yazılırsa, $b_2 \; için,$

$$b_2 = \frac{(\sum x_2 y)(\sum x_1^2) - (\sum x_2 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2} \quad \textbf{(3.4)}$$

formülü bulunur. Bu şekilde bu iki formülden b_1 ve b_2 katsayıları hesaplanmış olur. Öte yandan, $Y' = b_0 + b_1 X'_1 + b_2 X'_2$ ifadesinden $b_0$çekilerek $b_0 = Y' - b_1 X'_1 - b_2 X'_2$ bulunur. Daha önce hesaplanan $b_1 \; ve \; b_2$ değerleri bu ifadede yerine konularak b_0 kesme terimi de hesaplanmış olur [15].

3.2. Regresyonun Standart Hatası

Çoklu regresyonun standart hatası basit regresyonda olduğu gibidir.

$$\sigma^2 = \frac{\sum e_i^2}{n-k} \quad \textbf{(3.5)}$$

Yalnız, burada tahmin edilen parametre sayısı k değişecektir. Örneğin, iki açıklayıcı değişkeni olan modelde $b_0, b_1, ve\ b_2$ olmak üzere üç parametrenin tahmini söz konusu olduğundan $k = 3$ olacağı aşikardır [15].

3.3. Çoklu Determinasyon Katsayısı

Çoklu regresyonda bağımlı ve bağımsız değişkenler arasındaki uyumun iyiliğini, yani çoklu regresyonun genel başarısını çoklu determinasyon katsayısı ile ölçüyoruz.

Basit determinasyon katsayısının basit korelasyon katsayısının karesi olduğunu belirtmiştik. Benzer bir şekilde, çoklu determinasyon katsayısı da çoklu korelasyon katsayısının karesinden ibarettir. $X_1, X_2\ ve\ X_3$değişkenleri verildiğinde, $X_1\ ile\ X_2, X_3$ arasındaki çoklu korelasyon katsayısı,

$$r_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}} \qquad \textbf{(3.6)}$$

ise, her iki tarafın karesi alınarak,

$$R_{1.23}^2 = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2} \qquad \textbf{(3.7)}$$

Çoklu determinasyon katsayısı bulunmuş olur. $Y, X_1\ ve\ X_2$ değişkenleri verildiğinde Y' nin $X_1\ ve\ X_2$ üzerine çoklu determinasyonu

$$R_{y.x_1x_2}^2 = \frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1}r_{yx_2}r_{x_1x_2}}{1 - r_{x_1x_2}^2} \qquad \textbf{(3.8)}$$

olarak yazılır. Bu çoklu determinasyon katsayısının hesabı için formülde görüleceği gibi, $r_{yx_1}, r_{yx_2}\ ve\ r_{x_1x_2}$ olmak üzere üç tane basit korelasyon katsayısı hesaplanmalıdır. Bu basit korelasyonların hesabı için ara değerler, çoklu regresyon hesabı ile zaten elde edilmiş olmaktadır **[16].**

Çoklu regresyon katsayılarını kullanarak çoklu determinasyon katsayısı, daha kısa ve basit olarak şu şekilde hesaplanabilir:

$$R_{y.x_1x_2}^2 = \frac{b_1 \sum yx_1 + b_2 \sum yx_2}{\sum y^2} \qquad \textbf{(3.9)}$$

3.4. Anlamlılık Testleri

Çoklu regresyonda tahmin edilen katsayıların anlamlılık testleri için, basit regresyondan farklı olarak sadece katsayıların standart hatalarının veren formüllerin değişmesi söz konusudur. Diğer işlemler basit regresyonda olduğu gibidir. İki açıklayıcı değişkeni olan bir modelinde b_2 katsayısını ihmal ederek b_1 ve b_2 katsayılarının beklenen değer ve varyansları:

$$E(b_1) = \beta_1 \qquad Var(b_1) = \sigma^2 \frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$$

$$E(b_2) = \beta_2 \qquad Var(b_2) = \sigma^2 \frac{\sum x_1^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$$

ile verilmektedir [16].

3.5. Regresyonda Dönüştürme

Regresyon katsayılarının tahmininde kullanılan en küçük kareler yöntemi doğrusal ilişkilere rahatça uygulanabilirken doğrusal dışı fonksiyonel ilişkilerde yöntemin uygulanması zorlaşmakta ve nümerik yöntemleri gerektirmektedir. Ancak araştırmaya konu olan fenomenler doğrusal dışı karakterde olabilir. Örneğin karesel, kübik, logaritmik biçimler söz konusu olabilir. Böyle doğrusal dışı durumları uygun dönüştürmelerle doğrusal hale getirip regresyon analizine dahil edebiliriz. Örneğin $Y = a + bX + cX^2$ ilişkisi $X^2 = Z$ dönüştürmesiyle $Y = a + bX + cZ$ çoklu regresyon olarak a,b,c katsayıları tahmin edilebilir [17].

Serpilme diyagramının incelenmesi ile verilere birden fazla model uygun gibi görünüyorsa uygun model hakkında karar vermek için regresyonun standart hatasına bakılabilir.

$$S = \sqrt{\frac{\sum (Y - Y')^2}{n - k}} \qquad \textbf{(3.10)}$$

İle verilen regresyonun standart hatası hangi model için minimum oluyorsa o model seçilir.

4. DOĞRUSAL OLMAYAN REGRESYON

4.1. Doğrusal Olmayan Regresyon Modelleri

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

$$Y_i=f(X_i,\gamma)+\varepsilon_i \quad \textbf{(4.1)}$$

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 dağılım hata terimli bir modelin, hata terimlerinin, sık sık sabit varyans ile bağımsız normal rastlantısal değişkenler olduğu varsayılarak kullanılmaktadır [18].

4.1.1. 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.

$$Y_i = \gamma_0 + \gamma_1 \exp(\gamma_2 X_i) + \varepsilon_i \quad \textbf{(4.2)}$$

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

$$f(X,\gamma) = \gamma_0 + \gamma_1 exp(\gamma_2 X) \quad \textbf{(4.3)}$$

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 [1].

4.1.2. 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.

$$Y_i = \frac{\gamma_0}{1+\gamma_1 \exp(\gamma_2, X_i)} + \varepsilon_i \qquad \textbf{(4.4)}$$

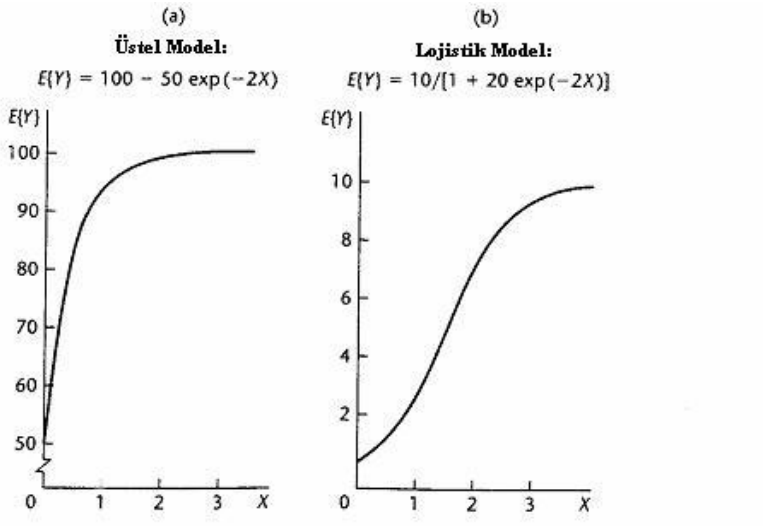
Buradan $£_i$hata terimleri sabit varyans σ^2 ile bağımsız normal dağılmaktadır. Bağımlı değişkenin fonksiyonu aşağıdaki gibidir:

$$f(X, \gamma) = \frac{\gamma_0}{1+\gamma_1 exp(\gamma_2, X_i)} \qquad \textbf{(4.5)}$$

yine bağımlı değişkenin fonksiyonu $\gamma_0, \gamma_1 \, ve \, \gamma_2$ parametrelerinde doğrusal değildir. Bu lojistik regresyon modeli, topluluk çalışmalarındaki ilişki için kullanılmaktadır. Örneğin zaman (X) içindeki türlerin sayısı (Y). Şekil (3.1) parametre değerleri $\gamma_0 = 10, \gamma_1 = 20$ ve $\gamma_2 = -2$ için (3.5) lojistik bağımlı değişkenin fonksiyonunu göstermektedir. Burada parametre $\gamma_0 = 10$ maksimum büyüme değerini göstermektedir [1]. Üstel ve Lojistik bağımlı değişken fonksiyonlarının çizimleri aşağıda Şekil 4.1'de gösterilmiştir.

Lojistik regresyon modeli (3.4) ayrıca bağımlı değişken kalitatif olduğu zaman da kullanılmaktadır.

.



Şekil 4.1. Üstel ve Lojistik bağımlı değişken fonksiyonlarının çizimleri

4.1.3. 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 $£_i nin$ toplamı olarak kabul edilmektedir. Ayrıca $£_i$ 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 p tane regresyon katsayısı vardır. 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 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 [19].

$$Y_i = f(X_i, \gamma) + \varepsilon_i \qquad \textbf{(4.6)}$$

Burada;X_i, γ matris biçiminde 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:

$$f(X, \gamma) = \gamma_0[exp(\gamma_1, X)] \qquad \textbf{(4.7)}$$

logaritmik dönüşümle doğrusallaştırılmaktadır.

$$\log_e f(X, \gamma) = \log_e \gamma_0 + \gamma_1 X \qquad \textbf{(4.8)}$$

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

$$g(X,\gamma) = \beta_0 + \beta_1 X \quad \textbf{(4.9)}$$

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 modelin uygun olduğunu 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 [20].

4.1.4. 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, doğrusal olmayan regresyon modelindeki hata terimleri sabit varyansla bağımsız normal dağıldıklarında bu yöntemlerin ikisi de aynı parametre tahminlerini vermektedir [1].

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 [21].

4.2. Doğrusal Olmayan Regresyonda En Küçük Kareler Yöntemi

Yalın doğrusal regresyon modeli için En küçük kareler yönteminde Q toplamını minimum yapan değerler alınmaktadır.

$$Q = \sum [Y_i - (\beta_0 + \beta_1 X_i, \gamma)]^2$$

Verilen (X_i, Y_i) örnek gözlemleri için Q'yu minimum yapan $\beta_0 \; ve \; \beta_1$ değerleri, en küçük kareler tahmincilerdir ve b_0, b_1 ile gösterilmektedir.

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 $\beta_1{}'$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:

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

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 tahminlerinin (6.12) denklemindeki en küçük kareler toplamı Q doğrusal olmayan regresyon parametreleri $\gamma_0, \gamma_1, \ldots, \gamma_{p-1}$'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 [22].

4.3. 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 [22].

4.3.1. 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: $£_i$ hata terimleri bağımsız $N(0,\sigma^2)$ ve örnek hacmi n yeterince büyük olduğunda, g'nin örnekleme dağılımı yaklaşık olarak normaldir. Beklenen değeri ise yaklaşık tahmini olarak:

$$E\{g\} = \gamma$$

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

$$s^2\{g\} = MSE(D'.D)^2 \qquad \textbf{(4.10)}$$

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

Böylece örnek hacmi büyük ve hata terimleri sabit varyans ile bağımsız normal dağıldığında, doğrusal olmayan regresyon için g'ye göre en küçük kareler tahminleri yaklaşık olarak normal dağılmaktadır ve hemen hemen yansızdır. Ayrıca (3.9) 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 $s^2\{g\}'den$ 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 [23].

4.3.2. Büyük Örneklem Teorisi Ne Zaman Uygulanır?

Verilmiş olan herhangi bir doğrusal olmayan regresyon uygulamasında örnek hacmi yeterince büyük olduğu zaman bu teorinin uygulanıp uygulanamayacağı 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 doğrusal yaklaşım modelinin doğrusal olmayan regresyon modelinin iyi bir yaklaşımı olduğuna işarettir ve bu yüzden regresyon tahminlerinin asimptotik özellikleri uygulanabilirdir. 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 örneklem 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 sınamasını sağlamaktadır. Eğer böyleyse doğrusal olmayan regresyon tahminlerinin örnekleme davranışı doğrusala yakın olarak ifade edilmekte ve büyük örneklem sonuç çıkarma işlemleri uygun bir şekilde kullanılmaktadır [24].

4.3.3. Bir $\gamma_k'nın$ Aralık Tahmini

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

$\frac{g_k-\gamma_k}{s\{g_k\}} \sim t(n-p) \qquad k = 0,1,\ldots.,p-1$

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

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

olmaktadır.

4.3.4. Bir γ_k'yla İlgili Test

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

Hipotezler:

$H_0: \gamma_k = \gamma_{k0}$

$H_1: \gamma_k \neq \gamma_{k0}$

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

$$t = \frac{g_k - \gamma_{k0}}{s\{g_k\}} \qquad \textbf{(4.11)}$$

Yaklaşık olarak a anlam seviyesinde karar kuralı:

Eğer $| t * | \leq t(1 - a/2 ; n - p)$ ise, H_0 reddedilmez. Eğer $| t * | > t(1 - a/2 ; n - p)$ ise H_0 reddedilir.

5. POİSSON REGRESYON ANALİZİ

Poisson regresyon, lojistik regresyondan sonra en genel olan ikinci genelleştirilmiş model olup doğrusaldır. Bağımlı değişken oluş sayısı (count) ile belirtilen bir veri olduğunda, yani belirli bir yerde ya da belirli bir zamanda olan olayların sayısı olduğunda poisson regresyon kullanılmaktadır.

$$\gamma_i \sim poisson(\mu_i) \quad i = 1, \dots, N \qquad \textbf{(5.1)}$$

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 [25].

5.1. Poisson Dağılımı

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

$$p(Y;\mu) = \frac{\mu^2 e^{-\mu}}{Y!} \quad Y = 0,1, \dots, \infty \qquad \textbf{(5.2)}$$

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

Poisson dağılımı çoğunlukla fazla görülmeye olayların oluş sayısını modellemek için kullanılmaktadır. Bir topluluk içinde kesin bir zaman aralığında cilt kanserine yakalananların sayısı ya da her yıl kalp krizinden ölenlerin sayısı gibi, poisson dağılımı ilginç bir istatistiksel özelliğe sahiptir [25].

$$E(Y) = Var(Y) = \mu \qquad \textbf{(5.3)}$$

5.2. 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 $\beta = (\beta_0, \beta_1, \ldots., \beta_k)$ ve $\lambda(X_i, \beta)$, her i alt grubu için hata oranını gösteren X_i ve β'nın özel bir fonksiyonudur. İ. Alt grup için beklenen hataların sayısı aşağıdaki gibidir:

$$E(Y_i) = \mu_i = l_i\lambda(X_i, \beta) \quad i = 1,2, \ldots, n \qquad \textbf{(5.4)}$$

Burada Y_i, bir poisson rastlantısal değişkenidir. Bu yüzden $\lambda(X_i, \beta) > 0$ olmalıdır.

$Y_i'nin\mu_i$ortalama ile poisson dağılımı gösterdiği varsayımı altında,

$$P(Y_i; \mu_i) = \frac{\mu_i^{Y_i}\exp(-\mu_i)}{Y_i!} \quad i = 1,2, \ldots n \qquad \textbf{(5.5)}$$

olmaktadır. (4.4) ve (4.5) , $i = 1,2, \ldots, n$ sonucunda,

$$\mathrm{p}(Y_i; \beta) = \frac{[l_i\lambda(X_i,\beta)]^{Y_i}\exp(-l_i\lambda(X_i,\beta))}{Y_i!} \qquad \textbf{(5.6)}$$

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

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

Y_i' nin aşağıdaki dağılıma sahip olduğu varsayışın;

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

ve $Y_1, Y_2, \ldots., Y_n' lerin$ bağımsız poisson rastlantısal değişkenlerinin bir seti olduğu varsayılsın. Böylece poisson regresyon analizi için olabilirlik fonksiyonu genel bir şekilde aşağıdaki gibi olmaktadır:

$E(Y_i) = \mu_i = l_i\lambda(X_i, \beta)$ formülü, binom rastlantısal değişkenin ortalaması μ=np için olan formüle benzerdir(fakat eşit değildir): burada l_i, n'e benzerdir ve $\lambda(X_i, \beta)$, p' ye benzerdir.

$$L(Y; \beta) = \prod_{i=1}^{n} \mathrm{p}(Y_i; \beta) = \prod_{i=1}^{n}\{\frac{[l_i\lambda(X_i,\beta)]^{Y_i}\exp(-l_i\lambda(X_i,\beta))}{Y_i!}\}$$

$$= \frac{\{\prod_{i=1}^{n}[l_i\lambda(X_i,\beta)]^{Y_i}\exp[-\sum_{i=1}^{n} l_i\lambda(X_i,\beta)]}{\prod_{i=1}^{n} Y_i!} \qquad \textbf{(5.7)}$$

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

Pratikte olabilirlik fonksiyonu (4.7)' 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:

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

kullanılabilir.

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

$$\frac{\partial}{\partial \beta_j}[lnL(Y;\beta)] = 0 \quad j = 0,1, \dots \dots, k \qquad \textbf{(5.8)}$$

Yukarıda verilen maksimum olabilirlik denklemlerinin çözümü genellikle iterasyon işlemleri sonucunda bulunmaktadır. Bu yönteme, iteratif olarak yeniden ağırlıklandırılmış en küçük kareler (IRWLS) denilmektedir [25].

5.2.1. Bağ(Link) Fonksiyonu

Poisson modellerinde, en çok kullanılan bağ (link) fonksiyon logaritmiktir ve aynı zamanda buna kanonik bağ da denilmektedir.

$$\log \mu_i = \eta = x_i^t \beta \qquad \textbf{(5.9)}$$

Bu logaritmik bağın kullanılması $\mu_i' nin$ tahmin edilen değerlerinin parametre uzayı $[0, \infty)$ arasında kalmasını sağlar. Logaritmik bağ kullanılan bir poisson modeline bazen logaritmik doğrusal model de denilmektedir [2].

5.2.2. Varyans Fonksiyonu

Poisson modeli altında,

$E\{y_i\} = \sigma^2\{y_i\} = \mu_i ! \, dir.$ bu yüzden varyans fonksiyonu:

$$\sigma^2\{\mu\} = \mu \qquad \textbf{(5.10)}$$

olmaktadır. Gerçek bir veri çoğu kez aşırı yayılım gösterir. Aşırı yayılım durumu poisson modelinin izin verdiğinden daha fazla bir değişim 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 modeli elde edilmiş olur.
- Bağımlı değişkenin dağılımı negatif binom (pascal) dağılımıyla değiştirilir. Böylece "Poisson" dağılımındakinden daha fazla yayılım olması sağlanır [2].

5.2.3. Uyum Ölçüleri

Poisson regresyon modellerinin uyum ölçüleri maksimum yapılmış olabilirlik değerlerinin karşılaştırılmasından bulunmaktadır. $Y_i'nin$ (4.5) şeklinde poisson regresyona sahip olduğu ve $Y_1, Y_2, \dots., Y_n$'in bağımsız olduğu varsayılsın; böylece $\mu_1, \mu_2, \dots., \mu_n$ 'in genel bir fonksiyon olarak olabilirlik fonksiyonu aşağıdaki şekilde ifade edilmektedir.

$$L(Y;\mu)=\prod_{i=1}^{n} \frac{\mu_i^{Yi}\exp(-\mu_i)}{Y_i} = \frac{(\prod_{i=1}^{n}\mu_i^{Yi})\exp(-\sum_{i=1}^{n}\mu_i)}{\prod_{i=1}^{n}Y_i!} \qquad \textbf{(5.11)}$$

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

$$\frac{\partial}{\partial \mu_i}[lnL(Y;\mu)] = 0 \qquad i = 1,2,\dots,n \qquad \textbf{(5.12)}$$

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

$$L(Y;\widehat{\mu_i}) = \frac{(\prod_{i=1}^{n}Y_i^{Yi})\exp(-\sum_{i=1}^{n}Y_i)}{\prod_{i=1}^{n}Y_i!} \qquad \textbf{(5.13)}$$

Burada, $\mu = (\mu_1, \mu_2, \dots, \mu_n) = (Y_1, Y_2, \dots., Y_n)$ olmaktadır.

$(k+1) < n$ olduğunda ve (5.7) olabilirlik fonksiyonu maksimum yapıldığında (5.11) 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 $\mu_i = l_i\lambda(X_i, \beta)$ şeklinde kısıtlama bulunmasıdır. Diğer bir deyişle, birinci denklem

$H_0: \mu_i = l_i\lambda(X_i, \beta),\ i = 1,2, \ldots., n$hipotezi altında olabilirlik fonksiyonu olarak düşünülürken ikinci denklem H_1: "μ_i' nin yapısında kısıtlama yoktur. $i = 1,2, \ldots, n$ hipotezi altında olabilirlik fonksiyondur.

Böylece eğer $L(Y; \hat{\beta})$ birinci denklem altında maksimum yapılmış olabilirlik değeri oluyorsa:

burada $\widehat{\beta}, \beta' nın$ maksimum olabilirlik tahminidir. Böylece

$$-2ln[\frac{L(Y;\widehat{\beta})}{L(Y;\widehat{\mu})}] \qquad \textbf{(5.14)}$$

μ_i üzerinde bir kısıtlama olmayan model ile ilişkili olarak $\mu_i = l_i\lambda(X_i, \beta)$ 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 $\mu_i = l_i\lambda(X_i, \beta)$ modeli, veri noktası kadar çok parametre içeren temel model ile bulunabilen maksimum yapılmış olabilirlik değeri kadar neredeyse büyük bir değer sağlayacaktır. Burada geçen neredeyse büyük sözüyle anlatılmak istenilen, $L(Y; \hat{\beta}$)'in (5.14) kullanılarak hesaplanan olabilirlik oranı testi temel alınarak bulunan $L(Y; \hat{\mu})$ değerinden anlamlı bir şekilde küçük olmadığıdır.

$$G(\hat{\beta}) = -2ln[\frac{\mathrm{L}(\mathrm{Y};\widehat{\beta})}{\mathrm{L}(Y;\widehat{\mu})}] \qquad \textbf{(5.15)}$$

miktarı, $L(Y; \hat{\beta})$'in $L(Y; \hat{\mu})$'den anlamlı şekilde daha küçük olup olmadığını değerlendirmek için kullanılan uyum test istatistiğidir ve böylece varsayılan regresyon modeli

$\mu_i = l_i\lambda(X_i, \beta)'$nın veriyle anlamlı bir uyumsuzluğu olup olmadığının anlaşılması için önerilmektedir. $G(\hat{\beta})\ miktarına$ ayrıca $\mu_i = l_i\lambda(X_i, \beta)$ poisson regresyon modeli için deviance da denilmektedir ve tahmin edilen model için kalıntı değişiminin bir ölçüsü olarak düşünülebilir.$H_0: \mu_i = l_i\lambda(X_i, \beta)\ hipotezi$altında $\mathrm{G}(\hat{\beta})$ sapmanın tipik olarak $n - k - 1$ serbestlik dereceli (büyük örnekler için) yaklaşık bir ki-kare dağılımı gösterdiği varsayılmaktadır. Burada n, (5.11) olabilirliği içindeki belirlenmiş parametrelerin sayısı (örneğin alt grupların, hücrelerin ya da kategorilerin sayısı) ve $k + 1$ (5.7) olabilirliği içindeki belirlenmiş parametrelerin (örneğin $\beta_j' ler$) sayısıdır. Böylece, verilen bir veri setinde $\mu_i = l_i\lambda(X_i, \beta)$ 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.

(5.5) modeli altında i. hücre içinde tahmin edilen bağımlı değişken değeri $\hat{Y}_i = l_i\hat{\lambda}(X_i, \beta)$ ile gösterilmektedir. (5.15) miktarı aşağıdaki gibi yazılabilir:

$$G(\hat{\beta}) = 2\sum_{i=1}^{n}[Y_i \ln\left(\frac{Y_i}{\hat{Y}_i}\right) - (Y_i - \hat{Y}_i)] \qquad \textbf{(5.16)}$$

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 $G(\hat{\beta}) = 0$ 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 $G(\hat{\beta})$, uygun bir şekilde, yaklaşık olarak pearson tipi gözlenen değerlere karşı ki-kare istatistiğinin tahminine daha benzer olmaktır.

$$X^2 = \sum_{i=1}^{n}\frac{(Y_i - \hat{Y}_i)}{\hat{Y}_i} \qquad \textbf{(5.17)}$$

Dikkat edilmesi gereken nokta, $\hat{Y}_i$ değerleri çok küçük olduğu zaman, (5.17) istatistiği yanıltıcı olarak büyük olabilmektedir.

Bir hiyerarşik sınıf içindeki çeşitli modeller için sapmalar, olabilirlik oran testlerini üretmek için kullanılmaktadır. Özellikle (5.15) eşitliğinde verilen $G(\hat{\beta})$ sapma değeri ile $\beta = (\beta_0, \beta_1, \ldots., \beta_k)$ parametreler setini içeren (5.7) olabilirliği tekrar düşünülsün $0 < r < k$ için β içindeki son k-r parametrenin sıfıra eşit olup olmadığının test edilmek istenildiği varsayılsın. Örneğin sıfır hipotezi $H_0: \beta_{r+1} = \beta_{r+2} = \cdots = \beta_k$=0 şeklinde olsun. $H_0\ hipotezi$altında olabilirlik β_r ile (5.7) içindeki β'nın yer değiştirmesiyle bulunabilir, burada

$\beta_r = (\beta_0, \beta_1, \ldots., \beta_r; 0,0, \ldots,0)'dır.$

Eğer bu olabilirlik fonksiyonu $L(Y; \beta_r)$ ile gösterilirse ve eğer $\beta_r, L(Y, \beta_r)$ kullanılarak β_r'nin maksimum olabilirlik tahmincisi ise, $H_0'ın$ olabilirlik oranı testi aşağıdaki test istatistiği kullanılarak hesaplanır.

$$-2ln[\frac{L(Y;\hat{\beta}_r)}{L(Y;\hat{\beta})}] \qquad \textbf{(5.18)}$$

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

Ayrıca (5.18) ifadesi tam olarak aşağıdaki gibi gösterilen sapma farkına eşittir.

$G(\widehat{\beta_r}) - G(\hat{\beta})$ **(5.19)**

Bunu görmek için (5.15) ifadesinde verilen $G(\hat{\beta})'$nin genel tanımına geri dönmek gereklidir. (5.15) ve (5.19) kullanılarak aşağıdaki sonuca ulaşılır.

$$G(\widehat{\beta_r}) - G(\hat{\beta}) = -2ln[\frac{L(Y;\widehat{\beta_r})}{L(Y;\hat{\mu})}] + 2ln[\frac{L(Y;\hat{\beta})}{L(Y;\hat{\mu})}] = -2ln[\frac{L(Y;\hat{\beta}_r)}{L(Y;\hat{\beta})}]$$

Bu ifade tam olarak (4.18) olabilirlik oranı (likelihoodratio test statistic) test istatistiğine eşittir. H_0: $\beta_{r+1} = \beta_{r+2} = \cdots = \beta_k$=0 hipotezi altında $G(\widehat{\beta_r}) - G(\hat{\beta})$ farkı, büyük örnekler için $k-r$ serbestlik derecesi yaklaşık olarak bir ki-kare dağılımı göstermektedir.

Böylece, bir veri setini analiz etmek için poissun regresyon kullanıldığında, aday modellerin bir seti içindeki üyelerin sapma çiftleri arasındaki farklar düşünülerek bu modeller karşılaştırılabilmektedir [2].

5.2.4. Model

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

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

$\beta = (X^T.W.X)^{-1}.X^T.W.z$ (5.20)

burada ağırlıkların matrisi aşağıdaki gibidir:

$W = Diag[\sigma^2\{Y_i\}\left(\frac{\partial_{\eta i}}{\partial_{\mu_i}}\right)^2]^{-1}$ w tartı matrisi **(5.21)**

$\eta_i = \log \mu_i$olduğu için,

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

sonucu bulunmaktadır.

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

$$z = \mathfrak{y} + (\frac{\partial \mathfrak{y}}{\partial \mu})(Y - \mu) \quad \textbf{(5.22)}$$

$z_i = \mathfrak{y}_i + (Y_i - \mu_i)/\mu_i$olarak elde edilmektedir. z'ye çalışan rastlantı değişkeni (workingvariate) denilmektedir.

Geçerli β tahmini bulmak için,

$\mathfrak{y}_i = X_i^T.\beta,\ \mu_i = \exp(\mathfrak{y}_i)\ ve\ z_i = \mathfrak{y}_i + (Y_i - \mu_i)/\mu_i$ hesaplanmaktadır. Daha sonra β' nın yeni tahminini elde etmek için $\mu_i' lerin$ 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:

- B tahminlerinin varyans-kovaryans matrisi

$$\sigma^2\{\beta\} = (X^T.W.X)^{-1} \quad \textbf{(5.23)}$$

- Pearson kalıntıları:

$$r_i = \frac{Y_i - \mu_i}{\sqrt{\mu_i}} \quad \textbf{(5.24)}$$

- Pearson uyum test istatistiği

$$X^2 = \sum_{i=1}^{n} r_i^2 \quad \textbf{(5.25)}$$

- Kaldıraç değerleri (Leveragevalues)

$$Diag(H) = Diag[W^{1/2}.X.(X^T.W.X)^{-1}.X^T.W^{1/2}] \quad \textbf{(5.26)}$$

Not: 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 olarak adlandırılır. Büyük kaldıraç değeri i. gözlem X gözlemlerinin merkezinden olan uzaklığını gösterir [2].

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

$$I=\sum_{i=1}^{N}(y_i \log \mu_i - \mu_i) \quad \textbf{(5.27)}$$

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 $\mu_i = y_i' dey_i \log \mu_i - \mu_i)$'nin maksimuma ulaşmasını göstermek oldukça kolaydır. Bu yüzden, sapma:

$$G = 2\sum_{i=1}^{n}[y_i \, log\left(\frac{y_i}{\mu_i}\right) - (y_i - \mu_i)] \quad \textbf{(5.28)}$$

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
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 (μ_i) şeklinde dağılan ölümlerin sayısını göstermekte ve m'de tedavi olanların sayısını gösteren bir çapraz tablo şeklinde ise model aşağıdaki şekilde kurulabilmektedir.

$$\mu_i = m_i \lambda_i \quad \textbf{(5.29)}$$

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,

$$\log \mu_i = \eta_i = o_i + X_i^T \beta \quad \textbf{(5.30)}$$

olmaktadır. Burada,

$$o_i = \log m_i \quad \textbf{(5.31)}$$

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 Fishers korlama algoritması aşağıdaki şekilde türetilmektedir:

$$\frac{\partial l}{\partial \beta_j} = (\frac{\partial l}{\partial \theta})(\frac{\partial \theta}{\partial \mu})(\frac{\partial \mu}{\partial \eta})(\frac{\partial \eta}{\partial \beta_j}) \quad \textbf{(5.32)}$$

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

$$\beta^{(r+1)} = \beta^{(1)} + (X^T W X)^{-1} X^T A(y - \mu) \quad \textbf{(5.33)}$$

Burada,

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

$$A = Diag[\sigma^2\{y_i\}\left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2]^{-1} = \left(\frac{\partial \eta}{\partial \mu}\right) W \quad \textbf{5.34)}$$

Bunlar IRWLS yönteminde kullanılarak aşağıdaki sonuçlar elde edilir:

$$\beta^{(r+1)} = (X^T W X)^{-1}[X^T W X \beta^{(t)} + X^T A(y - \mu)]$$

$$= (X^T W X)^{-1} X^T W \mathrm{z} \quad \textbf{(5.35)}$$

Burada çalışan rastlantı değişkeni (z_i) aşağıdaki şekilde değiştirilmiştir:

$$z_i = \eta_i - o_i + (\frac{\partial \eta_i}{\partial \mu_i})(y_i - \mu_i) \quad \textbf{(5.36)}$$

çünkü burada,

$$X_i^T \beta = \eta_i - o_i \quad \textbf{(5.37)}$$

Şeklindedir [2].

5.2.4.1. Aşırı Yayılım için Tartı Parametresi

Veride aşırı yayılım olduğunu göstermek olağan bir durumdur. Aşırı yayılım çözmenin bir yolu $\sigma^{2'}nin$ bazı sıfırdan büyük olduğu farz edilen değerleri için aşağıdaki varsayımı yapmaktadır:

$$\sigma^2\{y_i\} = \sigma^2\mu_i \quad \textbf{(5.38)}$$

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

$$\hat{\sigma}^2 = X^2/(n-p) \quad \textbf{(5.39)}$$

Burada,

$$X^2 = \sum_{i=1}^{N} \frac{(y_i-\mu_i)^2}{\mu_i} \quad \textbf{(5.40)}$$

olmaktadır. Bu da zaten pearson uyum test istatistiğidir.

Pek çok durumda modelin uyumsuzluğun 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

$$\hat{\sigma}^2\{\hat{\beta}\} = \hat{\sigma}^2(X^TWX)^{-1} \quad \textbf{(5.41)}$$

Olmaktadır [2].

- Test istatistikleri $X^2, G, \Delta X^2 \; ve \; \Delta G$; $\hat{\sigma}^{2'}e$ $bölünmelidir.$
- Pearson ve sapma kalıntıları σ'e bölünmelidir.

5.2.4.2. 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 ayrılsın 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 $\lambda_{ij}, (i,j)$'inci grup içindeki gerçek riski göstermektedir.

$$RR_i = \frac{\lambda_{i1}}{\lambda_{i0}} \quad \textbf{(5.42)}$$

Eğer $RR_i = 1$ ise topluluk riskleri i. yaş grubu içinde aynıdır; $RR_i > 1$ 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 hazardratio isimleriyle de kullanılmaktadır.

5.3. Negatif Binom Model

Aşırı yayılımı çözmenin bir başka yolu poissondan daha fazla yayılan parametrik bir model olan negatif binom'dur.

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

$$\gamma/\lambda \sim Poisson(\lambda) \quad \textbf{(5.43)}$$

$$\lambda \sim Gama(\alpha, \beta) \quad \textbf{(5.44)}$$

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

$$P(\lambda) = \frac{1}{\beta^{!!}\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\lambda/\beta), \lambda > 0 \quad \textbf{(5.45)}$$

$$0, \qquad \text{d.d.}$$

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

$$P(y) = \frac{\Gamma(\alpha+y)}{\Gamma(\alpha)y!} \left(\frac{\beta}{1+\beta}\right)^y \left(\frac{1}{1+\beta}\right)^\alpha, \ y = 0,1,2,\ldots, \quad \textbf{(5.46)}$$

Bu dağılımın ortalaması,

$$E\{y\} = E\{E\{y \mid \lambda\}\} = E\{\lambda\} = \alpha\beta \quad \textbf{(5.47)}$$

Ve varyansı,

$$\sigma^2\{y\} = E\{\sigma^2\{y \mid \lambda\}\} + \sigma^2\{E\{y \mid \lambda\}\}$$

$$= \sigma^2\{\lambda\} + E\{\lambda\}$$

$$= \alpha\beta + \alpha\beta^2 \quad \textbf{(5.48)}$$

olmaktadır.

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

$$P(y) = \frac{\Gamma(K^{-1}+y)}{\Gamma(K^{-1})y!}\left(\frac{K\mu}{1+K\mu}\right)^{y}\left(\frac{1}{1+K\mu}\right)^{1/K} \quad \textbf{(5.49)}$$

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

$y_i \sim Negbin(\mu_i, K)$

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 kullanılabilmektedir: [2].

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

5.4. Az Yayılımla İlgili Bir Açıklama

Daha önce negatif binom modeli aşağıdaki gibi düşünülmekteydi:

$y_i \sim Negatif \; binom \; (\mu_i, K)$

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

Yukarıdaki ifade $K \to 0$ olduğunda birpoisson modeline yaklaşmaktadır. Çünkü varyans aşağıdaki gibidir:

$$\sigma^2\{y_i\} = \mu_i + +\mu_i^2 \quad \textbf{(5.50)}$$

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 $\widehat{K}$=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 [25].

5.5. 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 düşünülsün. 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. 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
- Bağımlı değişken değerleri 0,1,2, ... kütlesi ile yayılan bir modeli (poiison gibi) takip etmektedir.

Sonraki kısım poisson olduğunda buna bol sıfırlı poisson (zero-inflated poisson kısaca ZİP) modeli denmektedir.

Literatürde bir ZİP 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.

ZİP model açıklayıcı değişkenleri kapsar şekilde yayılabilmektedir. Açıklayıcı değişkenler iki şekilde girilebilmektedir:

- $P(z_i = 1)$' i tahmin etmek için bir logit model içinde ve
- $E(y_i \mid z_i = 1)$'i tahmin etmek için bir poisson kısmı içinde

ZİP 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:

$$f(y_i \mid y_i > 0) = \frac{\mu_i^{y_i} exp(-\mu_i)}{y_i!(1-exp(\mu_i))} \quad , \quad y_i = 1,2, ... \qquad \textbf{(5.51)}$$

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 [25].

5.6. Çokterimli Oluş Sayısı ile Belirlenen Veri için Poisson Regresyonun Kullanımı

Poisson modeli için sapma aşağıdaki gibidir:

$$G = 2\sum_{i:y_{i\neq 1}}\{y_i \log\frac{y_i}{\mu_i} - (y_i - \mu_i)\} + 2\sum_{i:y_{i\neq}} \mu_i \qquad \textbf{(5.52)}$$

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

$$G = 2\sum_{i=1}^{N} y_i \, log\frac{y_i}{\mu_i} \qquad \textbf{(5.53)}$$

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

$y_i \sim Poisson(\mu_i),\ \ i = 1, \ldots, N$

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
- $y = (y_1, \ldots, y_N)^T$ için bir çokterimli dağılım.

$y \mid y_+ \sim Çokterimli\ (y_+, \pi)$

burada $\pi = (\pi_1, \ldots, \pi_N)^T\ ve\ \pi_i = \mu_i \mid \mu_+' dir.$

$y = (y_1, \ldots, y_N)^T$ örnek gözlemlerinin çokterimli bir model gösterdiği varsayılsın. 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 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 $\mu_i' lerden$ çok $\pi_i' ler\ için$ 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 [26].

5.7. Çokterimli Oluş Sayısı ile Belirlenen Veri için Logaritmik Doğrusal Modeller

$y = (y_1, \ldots, y_N)^{T'} nin$ çokterimli bir dağılım gösterdiği varsayılsın:

$y \sim$Çokterimli (n, π)

burada $n = y_+\ ve\ \pi = (\pi_1, \ldots, \pi_N)^T$ 'dir. Çok genel bir mantıkla bilinen bir açıklayıcı değişkenler matrisi $X_{(Nxp)}$ ve bilinmeyen katsayılar vektörü $\beta_{(px1)}$ için logaritmik doğrusal bir model aşağıdaki gibi olmalıdır:

$$\log \pi = X\beta \qquad \textbf{(5.54)}$$

Diğer bir deyişle π'nin tek yönlü

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

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

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

$A, 1,2, \ldots . I$ şeklinde muhtemel değerler

$B, 1,2, \ldots . , J$ şeklinde muhtemel değerler

$C, 1,2, \ldots . , K$ şeklinde muhtemel değerler

Veri aşağıdaki gibi özetlenmektedir

$$y = (y_{111}, y_{211}, \ldots, y_{I\,J\,K})^T \qquad \textbf{(5.55)}$$

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 $\pi = (\pi_{111}, \pi_{211}, \ldots, \pi_{I\,J\,K})$ 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 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 aşağıdaki Tablo 5.1'deki gibi tanımlanabilir:

Tablo 5.1 $Y_{nx1} = X_{nxp}\beta_{nxp} + \varepsilon$ Genel doğrusal regresyon model için ANOVA tablosu

ETKİLER				
İ	A_1	A_2	…	A_{I-1}
1	1	0	…	0
2	0	1	…	0
3	0	0	…	0
:	:	:	…	:
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 AXB terimlerinin yorumu, ANOVA 'daki $A\ X\ B$ terimlerinin yorumundan oldukça farklı olmaktadır. ANOVA'da $A\ X\ B$ terimlerinin anlamlı olması, bağımlı değişken üzerindeki A'nın etkisinin B'nin seviyeleri karşısında değiştiğini göstermektedir. Logaritmik doğrusal bir modelde $A\ X\ B$ terimlerinin anlamlı olması kategorik değişkenler A ve B'nin ilişkili olduğunu göstermektedir. Bu yüzden, logaritmik doğrusal modellerde, $A\ X\ B$ terimlerine etkileşimden daha çok ilişki 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 bulunmadığı bir model olarak düşünülebilir. Bu $A, B\ ve\ C$ arasında tamamen bağımsızlık olduğunu göstermektedir [26].

6. UYGULAMA

Bu çalışmada 2005 ile 2017 yılları arasında Türkiye'deki boşanma sayısı verileri bağımlı değişken olarak kullanılmıştır. Açıklayıcı değişkenlerden biri evliliği bitirme nedenleri, diğeri de boşanmaların yaşandığı bölgelerdir. Boşanma sayıları sekiz tane evliliği bitirme nedenleri kategorisine göre sınıflandırılmıştır. Bu evliliği bitirme nedenleri akıl hastalığı, bilinmeyen, cana kast ve pek çok muamele, cürüm ve haysiyetsizlik, diğer, geçimsizlik, terk ve zina şeklindedir. Veriler TÜİK ten alınarak analiz yapılmıştır.

Bu çalışmada kurduğumuz hipotezler;

H_0: Bölgelerarası boşanma nedenleri arasında fark yoktur.

H_1: Bölgelerarası boşanma nedenleri arasında fark vardır.

Burada;

Bağımlı değişken Y, boşanma sayısı

Bağımsız değişkenler X_1, evliliği bitirme nedenleri

X_1={1,….,8}

X_{11}: Akıl hastalığı

X_{12}: Bilinmeyen

X_{13}: Cana kast ve pek çok muamele

X_{14}: Cürum ve haysiyetsizlik

X_{15}: Diğer

X_{16}: Geçimsizlik

X_{17}. Terk

X_{18}: Zina

Diğer bağımsız değişken X_2, bölgeler

X_2 ={1,…..7} olarak tanımlanmıştır.

X_{21}: Doğu Anadolu Bölgesi

X_{22}: Güneydoğu Anadolu Bölgesi

X_{23}: İç Anadolu Bölgesi

X_{24}: Akdeniz Bölgesi

X_{25}: Karadeniz Bölgesi

X_{26}: Marmara Bölgesi

X_{27}: Ege Bölgesi

Bölgelerarası boşanma nedenlerini incelerken evliliği bitiren nedenlerin, boşanma sayısını ne kadar etkilediği bölgelerarasında fark olup olmadığını SPSS paket programında incelenerek ve aşağıdaki sonuçlar elde edilmiştir.

Veriler normal dağılmadığından parametrik olmayan test kullanılmıştır.

Bölgelere göre boşanma sayıları Kruskal Wallis testiyle incelenmiştir. $p<0,05$ olduğundan istatistiksel olarak fark vardır. (Tablo 6.1; Tablo 6.2)

Tablo 6.1. Tanımlayıcı İstatistikler

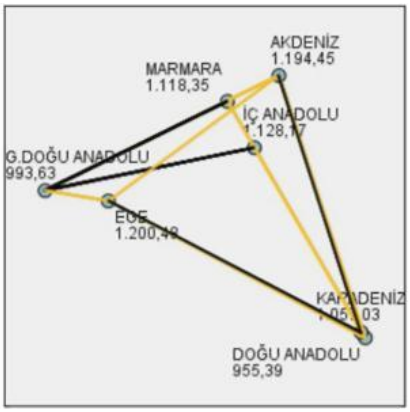
Tanımlayıcı İstatistikler								
	Gözlem sayısı	**Ortalama**	**Std. Hata**	**Minimum**	**Maximum**	**Percentiles**		
						25th	**50th (Median)**	**75th**
B.SAYILARI	2184	310,41	1236,474	0	11246	,00	2,00	14,00
BÖLGELER	2184	4,0000	2,00046	1,00	7,00	2,0000	4,0000	6,0000

Tabloya baktığımızda boşanma sayılarının ortalaması 310,41dır.

Tablo 6.2. Kruskal-Wallis Test Analizi

Test Statistics[a,b]	
	B.SAYILARI
Kruskal-Wallis H	43,546
df	6
Asymp. Sig.	,000
a. Kruskal Wallis Test	
b. Grouping Variable: BÖLGELER	

Kruskal Wallis testine göre istatistiksel olarak fark çıktığı için Çoklu karşılaştırma testi yapılmıştır. (Şekil 6.1; Tablo 6.3)



Şekil 6.1. Bölgelerin Çift Yönlü Karşılaştırmaları

Her düğüm, BÖLGELER'in örnek ortalama sırasını gösterir.

Tablo 6.3. Bölgelerin istatistiksel karşılaştırılması

Örnek 1- Örnek 2	İstatistik	Std Hata	Std. Test istatistik	p	Düzeltilmiş p
DOĞUANADOLU-G.DOĞU ANADOLU	-38,236	49.720	-.769	,442	1,000
DOĞU ANADOLU-KARADENİZ	-101,639	49,720	-2,044	,041	,860
DOĞU ANADOLU MARMARA	-162,955	49.720	-3,277	,001	,022
DOĞU ANADOLU-İÇ ANADOLU	-172,772	49,720	-3,475	,001	.011
DOĞU ANADOLU AKDENİZ	-239,054	49.720	-4,808	,000	,000
DOĞU ANADOLU-EGE	-245,083	49,720	-4,929	,000	,000
G.DOĞU ANADOLU-KARADENİZ	-63,404	49,720	-1,275	,202	1,000
G.DOĞU ANADOLU-MARMARA	-124,720	49,720	-2,508	.012	,255
G.DOĞU ANADOLU-İÇ ANADOLU	-134,537	49,720	-2,706	,007	.143
G.DOĞU ANADOLU-AKDENİZ	-200,819	49,720	-4,039	,000	.001
G.DOĞU ANADOLU-EGE	-206,848	49,720	-4,160	,000	.001
KARADENİZ-MARMARA	-61,316	49,720	-1,233	,217	1,000
KARADENİZ-İÇ ANADOLU	71,133	49,720	1,431	,153	1,000
KARADENİZ-AKDENİZ	137,415	49,720	2,764	,006	,120
KARADENİZ-EGE	-143,444	49,720	-2,885	,004	,082
MARMARA il. ANADOLU	9,817	49,720	,197	,843	1.000
MARMARA-AKDENİZ	76,099	49,720	1,531	,126	1,000
MARMARA-EGE	-82,128	49,720	-1,652	,099	1,000
İÇ ANADOLU-AKDENİZ	-66,282	49.720	-1,333	,182	1,000
İÇ ANADOLU-EGE	-72,311	49,720	-1,454	,146	1.000
AKDENİZ-EGE	-6,029	49,720	-.121	,903	1,000

Her satır, Örnek 1 ve Örnek 2 dağılımlarının sıfır hipotezini test eder.

Asimptotik anlamlılıklar (2-taraflı testler) görüntülenir. Anlamlılık düzeyi 05.

Anlamlı değerler, birden fazla test için Bonferroni düzeltmesi ile ayarlanmıştır.

Tablo 6.4. Boşanma Sayıları Ve Yüzde Olarak Oranları

EVLİLİĞİN B. NEDENLERİ	Akıl Hastalığı		Bilinmeyen		Cana Kast ve Pek fena muamele		Cürüm ve Haysiyetsizlik		Diğer		Geçimsizlik		Terk		Zina		Genel Ortalama
BÖLGELER	**Sayı ar**	**%**	**Sayılar**	**%**	**Sayılar**	**%**	**Sayılar**	**%**	**Sayılar**	**%**	**Sayılar**	**%**	**Sayı lar**	**%**	**Sayılar**	**%**	
Doğu Anadolu	16	**7**	618	**3**	13	**7**	2	**1**	607	**10**	17677	**3**	84	**6**	21	**5**	2379,75
Güneydoğu	32	**13**	2	**0**	22	**11**	11	**6**	635	**11**	45238	**7**	88	**6**	16	**4**	5755,50
İç Anadolu	56	**23**	3950	**17**	30	**16**	39	**20**	2544	**42**	151854	**23**	188	**13**	54	**13**	19839,38
Akdeniz	45	**18**	13142	**58**	28	**15**	44	**22**	1139	**19**	96539	**15**	296	**20**	66	**15**	13912,38
Karadeniz	21	**9**	1250	**6**	14	**7**	20	**10**	160	**3**	37789	**6**	196	**14**	81	**19**	4941,38
Marmara	47	**19**	1090	**5**	27	**14**	30	**15**	354	**6**	114070	**18**	296	**20**	75	**18**	14498,63
Ege	28	**11**	2665	**12**	57	**30**	51	**26**	541	**9**	182714	**28**	307	**21**	113	**26**	23309,50

Yukarıdaki çoklu karşılaştırma testine bakıldığında altı bölgede anlamlılık ilişkisine baktığımızda p değerinden küçük olduğundan boşanma nedenleri arasında fark olduğunu görebiliyoruz.

- Doğu Anadolu ile Marmara Bölgesi p<0,05 olduğu için boşanma nedenlerinde istatistiksel olarak fark vardır.

Boşanma sayısı ve yüzdelik oran tablosuna bakıldığında boşanma sebepleri arasından farklılığın akıl hastalığı Doğu Anadolu bölgesinde %7 iken Marmara bölgesinde %19, geçimsizlik Doğu Anadolu bölgesinde %3 iken Marmara bölgesinde %18, terk Doğu Anadolu bölgesinde %6 iken Marmara bölgesinde %20 olduğu görülmektedir

- Doğu Anadolu ile İç Anadolu Bölgesi p<0,05 olduğu için boşanma nedenlerinde istatistiksel olarak fark vardır.

Boşanma sayısı ve yüzdelik oran tablosuna bakıldığında boşanma sebepleri arasından farklılığın akıl hastalığı Doğu Anadolu bölgesinde %7 iken İç Anadolu bölgesinde %23, diğer Doğu Anadolu bölgesinde %10 iken İç Anadolu bölgesinde %42, geçimsizlik Doğu Anadolu bölgesinde %3 iken İç Anadolu bölgesinde %23 olduğu görülmektedir.

- Doğu Anadolu ile Akdeniz Bölgesi p<0,05 olduğu için boşanma nedenlerinde istatistiksel olarak fark vardır.

Boşanma sayısı ve yüzdelik oran tablosuna bakıldığında boşanma sebepleri arasından farklılığın bilinmeyen Doğu Anadolu bölgesinde %3 iken Akdeniz bölgesinde %58, cürüm ve haysiyetsizlik Doğu Anadolu bölgesinde %1 iken Akdeniz bölgesinde %22, terk Doğu Anadolu bölgesinde %6 iken Akdeniz bölgesinde %20 olduğu görülmektedir.

- Doğu Anadolu ile Ege Bölgesi p<0,05 olduğu için boşanma nedenlerinde istatistiksel olarak fark vardır.

Boşanma sayısı ve yüzdelik oran tablosuna bakıldığında boşanma sebepleri arasından farklılığın cana kast ve pek fena muamele Doğu Anadolu bölgesinde %7 iken Ege bölgesinde %30, cürüm ve haysiyetsizlik Doğu Anadolu bölgesinde %1 iken Ege bölgesinde %26, geçimsizlik Doğu Anadolu bölgesinde %3 iken Ege bölgesinde %28, zina Doğu Anadolu bölgesinde %5 iken Ege bölgesinde %26 olduğu görülmektedir.

- Güneydoğu ile Akdeniz Bölgesi $p<0,05$ olduğu için boşanma nedenlerinde istatistiksel olarak fark vardır.

Boşanma sayısı ve yüzdelik oran tablosuna bakıldığında boşanma sebepleri arasından farklılığın bilinmeyen Güneydoğu Anadolu bölgesinde %0 iken Akdeniz bölgesinde %58, cürüm ve haysiyetsizlik Güneydoğu Anadolu bölgesinde %6 iken Akdeniz bölgesinde %22, terk Güneydoğu Anadolu bölgesinde %6 iken Akdeniz bölgesinde %20 olduğu görülmektedir.

- Güneydoğu ile Ege Bölgesi $p<0,05$ olduğu için boşanma nedenlerinde istatistiksel olarak fark vardır.

Boşanma sayısı ve yüzdelik oran tablosuna bakıldığında boşanma sebepleri arasından farklılığın cana kast ve pek fena muamele Güneydoğu Anadolu bölgesinde %11 iken Ege bölgesinde %30, geçimsizlik Güneydoğu Anadolu bölgesinde %7 iken Ege bölgesinde %28, zina Güneydoğu Anadolu bölgesinde %4 iken Ege bölgesinde %26 olduğu görülmektedir.

- Tabloda diğerlerine bakıldığında $p<0,05$ olmadığından boşanma nedenlerinde istatistiksel olarak fark yoktur.

SONUÇLAR

Poisson regresyon, bağımlı değişken oluş sayısı ile belirtilen bir veri olduğunda, yani belirli bir yerde ya da belirli bir zamanda olan olayların sayısı olduğunda kullanılmaktadır. Buradaki regresyon fonksiyonu, bağımsız değişkenleri bağımlı değişkenin beklenen değerine bağlayan bir fonksiyondur. Bu fonksiyon, genellikle logaritmik olup doğrusaldır. Poisson regresyonda uygulamalar çoğu kez belli kategorilere ayrılarak oluş sayısı ile belirtilen veriler ile yapılmaktadır.

Poisson regresyonda, parametre tahminleri maksimum olabilirlik yöntemi ve IRWLS yöntemi kullanılarak yapılmaktadır. Çalışmada; regresyon analizi, poisson analizi ve poisson analizinde parametre tahminleri ve uyum ölçülerinin hesaplanmasıyla ile ilgili genel bilgiler bulunmaktadır.

Bu uygulamada yapılan çalışma parametre tahminleri ve uyum ölçülerinin hesaplanmasında SPSS paket 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 2005 ile 2017 yılları arasında Türkiye'deki boşanma sayısı verileri kullanılmıştır. Bağımsız değişkenler evliliği bitirme nedenleri ve bölgelerdir. Evliliği bitirme nedenleri grubu sekiz tane kategoriye ayrılarak sınıflandırılmıştır. Bu evliliği bitirme nedenleri grubu akıl hastalığı, bilinmeyen, cana kast ve pek fena muamele, cürüm ve haysiyetsizlik, diğer, geçimsizlik, terk ve zinadır. Yapılan çalışmada boşanma verilerini en iyi şekilde açıklayan modele ulaşmak amaçlanmaktadır.

Boşanma verileriyle ilgili yapılan çalışmalar sonucunda evliliği bitirme nedenleri grubunun modeli oldukça etkilediği görülmektedir. Bu sonuca varılmasının sebebi, modele evliliği bitirme nedenleri grubu eklenmediğinde boşanmaların nedenlerinin bilinmesi oldukça zordur.

Evliliği bitirme nedenleri grubu SPSS paket programına veri girişi yapıldıktan sonra normal dağılmadığı görülmüştür. Bunun için üç veya daha fazla sayıda grubun ortalamaları arasındaki farkın anlamlılığını test amacıyla Kruskal Wallis-H testini kullandık. Kruskal Wallis testine göre istatistiksel olarak fark çıktığı için Çoklu Karşılaştırma testi yapılmıştır.

Yapılan çoklu karşılaştırma testine göre Doğu Anadolu-Marmara bölgesi, Doğu Anadolu-İç Anadolu bölgesi, Doğu Anadolu-Akdeniz bölgesi, Doğu Anadolu-Ege bölgesi, Güneydoğu Anadolu-Akdeniz bölgesi ve Güneydoğu Anadolu-Ege bölgesinde $p<0,05$ olduğundan bölgelerarası boşanma nedenlerinde fark vardır diyoruz. Bu anlamlı farkların nelerden kaynaklandığını bulmaya çalıştık.

Boşanma sayısı ve yüzdelik oran tablosuna baktığımızda Doğu ve Güneydoğu Anadolu bölgesinden batı bölgelerine doğru gidildikçe boşanma sayılarında artış görülmektedir.

KAYNAKLAR

[1] **Yeşilyurt, H.** 2005, Poisson Regresyon Modeli ve Türkiye'deki Boşanma İstatistikleri,

[2] **Özmen, İ.,** 1998, Poisson Regresyon Çözümleme Teknikleri, Doktora Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü

[3] **Öztürkcan, M.,** 2010 Regresyon Analizi, Maltepe Üniversitesi

[4] **TÜİK** 2000, Boşanma İstatistikleri T.C. Başbakanlık DİE. Ankara.

[5] **TÜİK** 2017, Boşanma İstatistikleri T.C. Başbakanlık DİE. Ankara

[6] **Yıldırım N.** 2010Türkiye'deki Boşanmalar ve Nedenleri,

[7] **Ergenç S.** 2010 Kamu Yönetimi Uzmanı, Aile Bakanlığı İç Denetçi, Boşanma Sayıları ve Nedenleri

[8] **Türkiye Kamu Sen** 2002-2016 https://www.kamusaati.com/sendika/bosanma-oranlari-yillar-gectikce-artiyor-2002-2016-evlenme-ve-bosanma-h26888.html

[9] **Orhunbilge, N.** 2002, Uygulamalı Regresyon ve Korelasyon Analizi

[10] http://w3.balikesir.edu.tr/~bsentuna/wp-content/uploads/2013/03/Regresyon-Analizi.pdf

[11] **Işıkkara B**. Regresyon Yöntemleri ve Sorunları, İstanbul Yayınevi

[12] **Dursun, A.** 2014, Uygulamalı Korelasyon Analizi, Nobel Akademik Yayıncılık

[13] **Karagöz, M.** 2006, İstatistik Yöntemleri, Ekim Kitabevi Yayınları, Bursa

[14] **Albayrak Sait A.** 2006, Uygulamalı Çok Değişkenli İstatistik Teknikleri, Ankara

[15] **Alpar, R.** 2011, Çok Değişkenli İstatistiksel yöntemleri, Detay Yayıncılık

[16] **Sümbüloğlu, K. & Yrd. Doç. Dr. Akdağ, B.** 2007, Regresyon Yöntemleri ve Korelasyon Analizi Hatiboğlu yayıncılık

[17] **Ergun, M.** 1995, Bilimsel Araştırmalarda İstatistik Uygulamaları Spss For Windows

[18] **Fidell LS, Barbara T.** 2008 Çok Değişkenli İstatistikleri Kullanımı, Nobel Yayıncılık Çevirmen Aydın, M., Doğrusal Regresyon Analizine Giriş, Nobel Yayıncılık

[19] **Yolsal, H.,** 2010 Parametrik Olmayan Yoğunluk Tahmincileri ve Regresyon Analizi

[20] **Çağlayan E.** Nonparametrik Regresyon Modelleri

[21] **Weisberg, S.** 1985, Applied Linear Regression, John Willay & Sons, New York

[22] **Gamgam, H. ve Albayrak, B.** 2015, Uygulamalı Regresyon Analizi, Seçkin yayınevi

[23] **John Rowlinas, Sastry G. Pentula, David Dickey,** 2005Uygulamalı Regresyon Analizi

[24] **Long, S.** 1997, Regression Models for Categorical and Dependent Variables, London, Sage Publications

[25] **Lambert, D.** 1992 Zero-İnflated Poisson Regression With an Application to Defects in Mnaufacturn Technımetrics

[26] **Ridout M.**, **J Hinde & C.G.B. Demetrio**. 2001, A Score Test for a Zore İnflated Poisson Regression Model, Against Zero İnflated Megative Binominal A teratves Biometrics

ÖZGEÇMİŞ

1990 yılında Elazığ ili Kovancılar ilçesinde doğdum. İlk, orta, lise öğrenimi Elazığ'da tamamladım. 2009 yılında girdiğim Fırat Üniversitesi Fen Fakültesi İstatistik Bölümü'nden 2013 yılında mezun oldum. 2015 yılında Fırat Üniversitesi Fen Fakültesi İstatistik Bölümünde Yüksek Lisansa başladım.