SİVAS CUMHURİYET ÜNİVERSİTESİ

Sosyal Bilimler Enstitüsü

Ekonometri Ana Bilim Dalı

KIRPILMIŞ POİSSON REGRESYON ANALİZİ VE BİR UYGULAMA

Yüksek Lisans Tezi

Seçil KARTAL

Sivas

Ağustos 2019

SİVAS CUMHURİYET ÜNİVERSİTESİ

Sosyal Bilimler Enstitüsü

Ekonometri Ana Bilim Dalı

KIRPILMIŞ POİSSON REGRESYON ANALİZİ VE BİR UYGULAMA

Yüksek Lisans Tezi

Seçil KARTAL

Tez Danışmanı

Doç. Dr. Necati Alp ERİLLİ

Sivas

Ağustos 2019

KABUL VE ONAY

Üniversite	**: Sivas Cumhuriyet Üniversitesi**
Enstitü	**: Sosyal Bilimler Enstitüsü**
Ana Bilim Dalı	**: Ekonometri**
Tezin Başlığı	**: Kırpılmış Poisson Regresyon Analizi Ve Bir Uygulama**
Savunma Tarihi	**: 12/07/2019**
Danışmanı	**: Doç. Dr. Necati Alp ERİLLİ**

	Unvanı - Adı Soyadı	İmza
Jüri Başkanı:	**Dr. Öğr. Üyesi Özge GÜNDOĞDU**	
Üye:	**Doç. Dr. Necati Alp ERİLLİ**	
Üye:	**Dr. Öğr. Üyesi Engin KARAKIŞ**	

Oy Birliği ☒

Oy Çokluğu ☐

Seçil KARTAL tarafından hazırlanan "Kırpılmış Poisson Regresyon Analizi Ve Bir Uygulama" başlıklı tez, kabul edilmiştir. …/…/……

Prof. Dr. Ahmet ŞENGÖNÜL

Enstitü Müdürü

ETİK İLKELERE UYGUNLUK BEYANI

Sivas Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü bünyesinde hazırladığım bu Yüksek Lisans tezinin bizzat tarafımdan ve kendi sözcüklerimle yazılmış orijinal bir çalışma olduğunu ve bu tezde;

1- Çeşitli yazarların çalışmalarından faydalandığımda bu çalışmaların ilgili bölümlerini doğru ve net biçimde göstererek yazarlara açık biçimde atıfta bulunduğumu;
2- Yazdığım metinlerin tamamı ya da sadece bir kısmı, daha önce herhangi bir yerde yayımlanmışsa bunu da açıkça ifade ederek gösterdiğimi;
3- Başkalarına ait alıntılanan tüm verileri (tablo, grafik, şekil vb. de dahil olmak üzere) atıflarla belirttiğimi;
4- Başka yazarların kendi kelimeleriyle alıntıladığım metinlerini, tırnak içerisinde veya farklı dizerek verdiğim yine başka yazarlara ait olup fakat kendi sözcüklerimle ifade ettiğim hususları da istisnasız olarak kaynak göstererek belirttiğimi,

beyan ve bu etik ilkeleri ihlal etmiş olmam halinde bütün sonuçlarına katlanacağımı kabul ederim.

2./08/2019
Seçil KARTAL

TEŞEKKÜR

Çalışmamın gerçekleştirilmesinde, değerli bilgilerini benimle paylaşan, bana faydalı olabilmek için elinden gelenin fazlasını sunan, yapıcı ve öğretici eleştirileriyle bana yol gösteren saygıdeğer danışman hocam; Sayın Doç. Dr. Necati Alp ERİLLİ’ ye teşekkürü bir borç biliyor ve şükranlarımı sunuyorum. Çalışma süresince tüm zorlukları benimle göğüsleyen ve hayatımın her evresinde bana destek olan aileme sonsuz teşekkürlerimi sunarım.

Seçil KARTAL

İÇİNDEKİLER

İÇİNDEKİLER ... i

KISALTMALAR ... iii

TABLO LİSTESİ ... v

ÖZET ... vii

ABSTRACT ... ix

GİRİŞ ... 1

BÖLÜM 1 ... 3

LİTERATÜR TARAMASI ... 3

BÖLÜM 2 ... 7

2. REGRESYON ANALİZİ ... 7

2.1. Regresyon Analizi ... 7

2.2. Bağımlı Değişken Sayısına Göre Regresyon Modelleri ... 9

2.2.1. Tek Değişkenli Regresyon Analizi ... 9

2.2.2. Çok Değişkenli Regresyon Analizi ... 9

2.3. Matematiksel Biçimine Göre Regresyon Modelleri ... 10

2.3.1. Doğrusal Regresyon Modeli ... 10

2.3.1.1 Parametre Tahmini ... 11

2.3.2. Doğrusal Olmayan Regresyon Modeli ... 14

2.3.2.1. Doğrusal Olmayan Regresyon Modellerinde Parametre Tahmini ... 14

2.4. Varsayımların Sağlanma Durumuna Göre Regresyon Modelleri ... 17

2.4.1. Parametrik Regresyon Analizi ... 17

2.4.2 Parametrik Olmayan Regresyon Analizi ... 17

2.4.3. Parametrik Regresyon ile Parametrik Olmayan Regresyonun Karşılaştırılması ... 18

2.4.4. Yarı Parametrik Regresyon Analizi ... 19

BÖLÜM 3 ... 21

3. SAYMA VERİLERİ İÇİN REGRESYON MODELLERİ 21

3.1 Poisson Regresyon Analizi 21

3.1.1. Poisson Dağılımı 23

3.1.2. Aşırı Yayılım Durumu 23

3.1.3. Aşırı Yayılım Testleri 24

3.2.Negatif Binom Regresyon 26

3.3. Poisson Regresyon Modelleri 27

3.3.1. Genelleştirilmiş ve Kısıtlanarak Genelleştirilmiş Poisson Regresyon Modelleri (GPR) 27

3.3.2. Birleşik Poisson Regresyon Modelleri 28

3.3.3. Tekrarlı Verilerde Poisson Regresyon Modeli 30

3.3.4. Poisson Regresyon Analizinde Katsayıların Tahmini 33

3.3.4.1. Poisson En Çok Olabilirlik Tahmin Yöntemi 33

3.3.4.2. Genelleştirilmiş Doğrusal Model (GLM) ile Tahmin Yöntemi 34

3.3.5. Model Anlamlılığının Sınanması 35

3.3.5.1. Parametre Tahminlerinin Sınanması 36

3.3.5.2. Modelin Uyum İyiliğinin Sınanması 37

3.4. Kırpılmış Poisson Regresyon Analizi 38

3.4.1. Sıfır Kırpılmış Poisson Regresyon Modeli 39

BÖLÜM 4 41

UYGULAMA 41

SONUÇ 55

KAYNAKÇA 57

ÖZGEÇMİŞ 65

KISALTMALAR

EÇO	: En Çok Olabilirlik
KT_x	: Düzeltilmiş Kareler Toplamı
$ÇT_{xy}$	: Düzeltilmiş Çarpımlar Toplamı
EKK	: En Küçük Kareler Yöntemi
ÇDB	: Çoklu Doğrusal Bağlantı
GKT	: Genel Kareler Toplamı
AKT	: Artık Kareler Toplamı
RAKT	: Regresyonla Açıklanan Kareler Toplamı
GP	: Genelleştirilmiş Poisson
GPR	: Genelleştirilmiş Poisson Regresyon
KGP	: Kısıtlanarak Genelleştirilmiş Poisson
KGPR	: Kısıtlanarak Genelleştirilmiş Poisson Regresyon
KPR	: Karışık Poisson Regresyon
NBR	: Negatif Binom Regresyon
PTGR	: Poisson Ters Gaussian Regresyon
IEKK	: Yeniden Ağırlıklandırılmış En Küçük Kareler
IRWLS	: İteratif Olarak Yeniden Ağırlıklandırılmış En Küçük Kareler
GLM	: Genelleştirilmiş Doğrusal Model
AIC	: Akaike Bilgi Ölçütü
BIC	: Bayes Bilgi Ölçütü
NB	: Negatif Binom Regresyon Analizi

TABLO LİSTESİ

Tablo 2.1 Bazı Regresyon Analizi Modelleri Çeşitleri 8
Tablo 4.1. 1. Veri Seti Tanımlayıcı İstatistikler 41
Tablo 4.2. 1.Veri Seti İçin Kırpılmış Poisson Regresyonu Sonuçları 42
Tablo 4.3 1. Veri Setinde Y=75 İçin Sağdan Kırpılmış Poisson Regresyonu Sonuçları 43
Tablo 4.4. 1. Veri Setinde Y=80 İçin Sağdan Kırpılmış Poisson Regresyonu Sonuçları 44
Tablo 4.5. 1. Veri Seti İçin Poisson Regresyonu Sonuçları 44
Tablo 4.6. 1. Veri Seti İçin Negatif Binom Regresyonu Sonuçları 45
Tablo 4.7. 1. Veri Seti İçin Yöntemlerin Karşılaştırılması 45
Tablo 4.8. 2.Veri Seti Tanımlayıcı İstatistikler 46
Tablo 4.9. 2. Veri Seti İçin Kırpılmış Poisson Regresyonu Sonuçları 46
Tablo 4.10. 2. Veri Setinde Y=12 için sağdan Kırpılmış Poisson Regresyonu Sonuçları 47
Tablo 4.11. 2. Veri Setinde Y=13 için sağdan Kırpılmış Poisson Regresyonu Sonuçları 48
Tablo 4.12. 2. Veri Seti İçin Poisson Regresyonu Sonuçları 49
Tablo 4.13. 2. Veri Seti İçin Negatif Binom Regresyonu Sonuçları 50
Tablo 4.14. 2. Veri Seti İçin Yöntemlerin Karşılaştırılması.................................... 51
Tablo 4.15 3.Veri Seti İçin Tanımlayıcı İstatistikler 51
Tablo 4.16. 3.Veri Seti İçin Kırpılmış Poisson Regresyonu Sonuçları 52
Tablo 4.17. 3.Veri Setinde Y=4 için sağdan Kırpılmış Poisson Regresyonu Sonuçları 52
Tablo 4.18. 3.Veri Seti İçin Poisson Regresyonu Sonuçları 53
Tablo 4.19. 3.Veri Seti İçin Negatif Binom Regresyonu Sonuçları 54
Tablo 4.20. 3. Veri Seti İçin Yöntemlerin Karşılaştırılması.................................... 54

ÖZET

Regresyon analizi aralarında neden sonuç ilişkisi olan iki veya daha fazla değişkenin aralarındaki ilişkiyi gözlemek ve konuya ait öngörüde bulunabilmek için elde edilen matematiksel bir modelle belirtilen istatistiksel bir yöntemdir. Basit doğrusal regresyon modeli bir bağımlı bir bağımsız değişken olmak üzere aralarındaki fonksiyonel ilişkiyi gözlemlerken, çoklu doğrusal regresyon modeli bir bağımlı ve birden fazla bağımsız değişken arasındaki fonksiyonel ilişkiyi incelemektedir. Literatürde verilerin veya kurulacak modellerin farklı yapıları için farklı regresyon modelleri tanıtılmıştır. Bunlardan biri de Poisson regresyon modelidir. Poisson regresyonu sayıma dayalı olarak elde edilen bağımlı değişkenin modellenmesinde kullanılır. Bununla birlikte, bağımlı değişkenin sayıma dayalı olarak elde edildiğinde, bağımsız değişken kategorileri için relatif risk değerini de hesaplamaktadır. Negatif Binom dağılımı, Poisson regresyon modeline göre aşırı yayılım gösteren verilerin modellenmesi amacıyla kullanılır. Veri kümesinde aşırı yayılım olması durumunda Poisson regresyon modelini kullanmak yanlı parametre tahminlerinin elde edilmesine yol açacağından yayılımı dikkate alan Negatif Binom regresyonunun kullanılması daha uygun olmaktadır. Aşırı uçlu verilerde Poisson regresyon analizi yanlı tahminler vermektedir. Bu yanlılığı ortadan kaldırmada Kırpılmış Poisson Regresyon Analizi kullanımı tavsiye edilmektedir.

Bu çalışmada Kırpılmış Poisson Regresyon Analizi, üç farklı veri seti üzerinde test edilmiştir. Birinci veri seti için en iyi sonuçlar Negatif Binom Regresyonu ile elde edilirken, ikinci ve üçüncü veri setlerinde Kırpılmış Poisson Regresyonu en iyi model sonuçlarını vermiştir.

Bu çalışma ile Kırpılmış Poisson Regresyon Analizinin Kategorik bağımlı değişken yapılarındaki regresyon modellerinde başarılı sonuçlar verebileceği bu yüzden de karşılaştırmalar yaparken bu teknikten de yararlanılması gerektiği vurgulanmıştır.

Anahtar Kelimeler: Regresyon Analizi, Poisson Regresyon Analizi, Kırpılmış Poisson Regresyon Analizi, Negatif Binom Regresyon Analizi

ABSTRACT

Regression analysis is a statistical method which is expressed by a mathematical model obtained to observe the relationship between two or more variables with cause-effect relationship and to make predictions about the subject. While the simple linear regression model observes the functional relationship between one dependent and one independent variable, the multiple linear regression model examines the functional relationship between one dependent and multiple independent variables. Various regression models have been introduced in the literature for the models to be established or different structures of the data. One of them is the Poisson regression model. Poisson regression is used to model the dependent variable obtained based on the counting. It also calculates the relative risk value for the independent variable categories when the dependent variable is obtained based on the counting. Negative binomial distribution is used to model the data with excessive spread compare to the Poisson regression model. In case of excessive spread in the data set, it is more appropriate to use Negative Binomial regression considering the spread, since using Poisson regression model leads to biased parameter estimation. In extreme pointed data Poisson regression analysis gives biased estimates. In order to eliminate this bias, the use of Clipped Poisson Regression Analysis is recommended.

In this study, Clipped Poisson Regression Analysis was tested on three different data sets. The best results for the first data set were obtained by Negative Binomial Regression, while Clipped Poisson Regression gave the best model results in the second and third data sets.

In this study, it is emphasized that Cropped Poisson Regression Analysis can give successful results in regression models in categorically dependent variable structures therefore this technique should be used when making comparisons.

Keywords: Regression Analysis, Poisson Regression Analysis, Clipped Poisson Regression Analysis, Negative Binomial Regression Analysis, Log-likelihood

GİRİŞ

Regresyon analizi bağımlı bir değişkenin bir veya birden fazla bağımsız değişkenle arasındaki ilişkinin matematiksel fonksiyon şeklinde ifade edilmesidir. Elde edilen bu fonksiyona regresyon denklemi denilmektedir (Orhunbilge 2000: 12).

Regresyon analizinin amacı yanıt değişken ile açıklayıcı değişken arasındaki ilişkiyi matematiksel modellerle belirterek, geleceğe yönelik öngörüde bulunmaktır. Bu konuda geliştirilmiş birçok istatistiksel yöntemden bahsetmek mümkündür (Karadavut vd. 2005:328-333).

Bu yöntemlerin uygulanacağı veri setlerinde yanıt değişkenin normal dağılım içerikli olması, açıklayıcı değişkenlerin normal dağılım gösteren değişken ya da değişkenlerden meydana gelmesi ve hata terimlerinin varyansının normal dağılımlı olması gerekmektedir (Çokluk 2010:1357-1407).

Yanıt değişkenlerin negatif olmayan kesikli sayılardan meydana geldiği ve belirli bir zaman aralığında oluşan olay sayısının küçük olduğu durumda, bu değişken için belirlenen regresyon modellerinden biri Poisson modelidir (Tüzel vd. 2012:23-31).

Bu tez çalışması ile; ele alınan verilere Poisson Regresyon Analizi, Kırpılmış Poisson Regresyon Analizi, Sağdan Kırpılmış Poisson Regresyon Analizi ve Negatif Binom Regresyon Analizi yöntemi uygulanarak analiz sonuçları karşılaştırılmış ve hangi yöntemin daha iyi sonuçlar verdiği incelenmiştir.

Bu çalışmada Kırpılmış Poisson Regresyon Analizinden faydalanılacaktır. Çalışmamızın ilk bölümünde Regresyon Analizi ve Poisson Regresyon Analizi yöntemleri hakkında literatür taraması yapılmıştır. İkinci bölümde Regresyon Analizi ve Modelleri, Parametrik Regresyon Analizi ve Parametrik Olmayan Regresyon Analizi'nden bahsedilmiştir. Üçüncü bölümde Poisson Regresyon Analizi, Negatif Binom Regresyon Analizi, Poisson Regresyon Modelleri'nden bahsedilmiştir. Dördüncü bölümde ise Kırpılmış Poisson Regresyon Analizinden söz edilmiştir. Beşinci bölümde 3 veri setinden yararlanılarak STATA.15 paket programı yardımı ile uygulama yapılmış ve yorumlanmıştır. Altıncı bölümde ise sonuçlar özetlenmiştir.

BÖLÜM 1

LİTERATÜR TARAMASI

Wang ve Fameo (1997), “Modeling household fertility decisions with generalized Poisson regression” adlı çalışmasında hane halkı doğurganlık kararının modellenmesinde standart Poisson regresyon ve genelleştirilmiş Poisson regresyon modellerinden faydalanarak uygun modelin saptanması için karşılaştırmalar yapmışlardır. Bu çalışma sonucunda doğurganlık verilerinde az yayılım gözlenmesinden dolayı genelleştirilmiş poisson regresyon modelinin standart Poisson regresyon modeline göre daha uygun olduğunu belirtmişlerdir.

Selim ve Üçdoğruk (2003), “Sayma veri modelleri ile çocuk sayısı belirleyicileri: Türkiye’deki seçilmiş iller için sosyoekonomik analizler” adlı çalışmasında Poisson quasi maksimum olabilirlik yönteminden yararlanılarak çocuk sayısı belirleyicilerini modellemişlerdir.

Yeşilova ve Atlıhan (2007), “Farklı sıcaklıkların Scymnus Subvillosus’un bıraktığı yumurta sayıları üzerine etkilerinin karışımlı Poisson regresyon ile analiz edilmesi” adlı çalışmalarında Scymnus Subvillosus dişilerinin oviposizyon sürecinde bıraktıkları yumurta sayılarını Poisson regresyon modelinden faydalanarak yorumlamışlardır.

Ghitany ve diğ. (2012), “An EM algorithm for multivariate mixed Poisson regression models and its application” adlı çalışmalarında çok değişkenli karışımlı Poisson regresyon modellerini gruplandırmak için maksimum olabilirlik tahminini kolaylaştırmada genel bir EM algoritması önermişlerdir.

Lambert (1992), çalışmasında AT&T laboratuarında alan başına düşen bozuk lehim sayısını incelemiştir. Çalışmada ayrıca gemi yüzeyine yapılacak lehim türleri de takip edilmiştir. Modellemeyi sıfır değerlerinin çokluğundan dolayı Sıfır Değer Ağırlıklı Poisson Regresyon ile yapmıştır.

Deniz (2005), çalışmasında Cameron ve Trivedi’nin “Categorical Data Analysis” kitabından yararlanmış ve Poisson regresyonun tanımı, tahmin edicileri, artıkların incelenmesi, uyum iyiliği kontrolü gibi belli başlı konuları ele almıştır.

Heinzl ve Mittlböck (2003), çalışmalarında aşırı ya da az yayılım gösteren Poisson regresyon modeli için yapay R^2 ölçütlerini önermiş ve küçük örnekler ya da fazla sayıda kovaryetin olduğu durumlar için yanlılık düzeltmeleri tavsiye etmişlerdir.

Köleoğlu (2006), doktora tezinde olay zamanı analizinde kullanılan Poisson regresyon modeli, Poisson regresyon modelinde aşırı yayılımla karşılaşıldığında kullanılan Poisson-gamma regresyon modeli ve olay zamanı verilerinde gözlemlenemeyen heterojenlikle karşılaşıldığında kullanılan tesadüfi etkiler Poisson-gamma regresyon modelini incelemiştir.

Yeşilyurt (2005), yüksek lisans tezinde yanıt değişken olarak 1989 ile 1998 yılları arasında Türkiye'deki boşanma sayısı verileri kullanarak, poisson regresyon analizini uygulamıştır.

Yeşilova ve diğ. (2006), "Norduz erkek kuzularının bazı kesikli üreme davranış özelliklerinin analizinde doğrusal olmayan regresyon modellerin kullanılması" adlı çalışmalarında Poisson ve negatif binom regresyon modellerinden faydalanarak genelleştirilmiş tahmin modellerini oluşturmuşlardır.

Saffari ve Adnan (2010), "Zero-inflated Poisson models with right censored count data" adlı çalışmalarında fazla sayıda sıfır değerlerinden elde edilen Poisson regresyon modeli için yanıt değişkenindeki bazı değerlerde sınırlamalar meydana getirmişlerdir.

Memiş ve Önder (2012), Bu çalışma Poisson regresyon analizi ve açıklayıcı değişkenler ile sayıma dayalı olarak bulunan cevap değişkeni arasındaki ilişkiyi açıklamamıza yardımcı olmaktadır. Yapılan bu çalışmada Poisson regresyonu yardımı ile ortaya çıkan açıklayıcı değişkenlerin doğrusal yapısını cevap değişkenin beklenen değerine bağlayan bağlantın fonksiyonu logaritmik dönüşüm ile ifade edilmiştir.

Şahin (2002), Bu çalışma Poisson regresyon modelini 1964-1998 döneminde Türkiye'de yapılan grevlere uygulamaktadır. Grevde ekonomik ve sosyal durumu yansıtan değişkenler ele alınmıştır.

Kılıç (2002), Bu çalışma İngiltere'de sigara içme yoğunluğunun belirleyicilerini 1991-2008 dönemi için panel veri metotlarından faydalanarak incelemeyi amaçlamıştır ve Sayma veri bağımlı değişkeninin ortalama ve varyans karşılaştırması yapılarak kullanılacak en uygun metodun Poisson modeli olduğu belirtilmiştir.

Erdinç, Yeşilova ve Ser (2017), Çalışmada, sayıma dayalı olarak oluşturulan zooplankton sucul böcek sayımlarının elde edilmesinde Poisson Regresyon ve Negatif Binom Regresyonlarının uygulaması ele alınmıştır.

BÖLÜM 2

2. REGRESYON ANALİZİ

2.1. Regresyon Analizi

Regresyon Analizi; son yıllarda hemen her alanda uygulanabilen, paket programların yaygınlaşması ile de çok tercih edilen bir tahmin modelidir. Regresyon terimini ilk kez Francis Galton kullanmıştır. Galton ünlü bir çalışmasında, uzun boylu ana babaların çocuklarının uzun, kısa boylu ana babaların çocuklarının kısa boylu olması yatkınlığına karşın, belli bir boydaki ana babaların çocuklarının ortalama boyunun, genel nüfustaki ortalama boya doğru yaklaşma yatkınlığında olduğunu tespit etmiştir. Galton'un evrensel bağlanım yasası, aile bireylerinin boylarına dair bini aşkın veri elde eden arkadaşı Karl Pearson tarafından doğrulanmıştır. Pearson, bir grup uzun boylu babanın oğullarının boy ortalamasının babalarınınkinden kısa, bir grup kısa boylu babanın oğullarının boy ortalamasının babalarınınkinden uzun olduğunu, böylece hem uzun hem kısa oğullarının boylarının tüm erkeklerin ortalama boyuna doğru çekildiğini göstermiştir. Galton'un sözleriyle bu "sıradanlığa doğru çekilme" olarak ifade edilmiştir (Galton 1886:42-72; Pearson 1903:357-462).

İstatistiksel kuram çalışmalarında sıkça ele alınan metotlardan biri regresyon analizi metodudur. Regresyon analizi; bir bağımlı değişkenin bir veya birden fazla bağımsız değişkenle arasındaki bağın matematiksel bir denklem formunda ele alınması olarak ifade edilebilir. Bu denkleme regresyon denklemi denilmektedir (Orhunbilge 2000:12).

Regresyon analizinin amaçları şu şekilde sıralanabilir (Maddala 2001:60);

1) Bağımsız değişkenlerin bağımlı değişken üstünde önemli bir etki meydana getirip getirmediğini belirlemek

2) Bağımlı değişkenin gelecekte alabileceği değerleri öngörmek

3) Bağımsız değişkenlerdeki değişimin bağımlı değişken üstünde meydana getireceği etkiyi ölçmek

Basit bir regresyon modeli Eşitlik 2.1'de verildiği gibidir.

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad (2.1)$$

Burada X bağımsız değişkeni, Y bağımlı değişkeni ve ε hata terimini belirtmektedir. Eşitlik (2.1)'de yer alan parametreler β_0, modelde sabit değişkendir. β_1, doğrusal fonksiyonun eğimidir ve β_1 parametresi regresyon analizinde X' deki bir birimlik değişmenin Y'de ne kadarlık bir değişim göstereceğini belirten regresyon katsayısını ifade etmektedir.

Regresyon Analizi doğru model seçiminde kesin kurallar olmamasına karşın, Tablo.1 uygun modeller için özet bilgi sunmaktadır (Erilli 2015:175);

Tablo 2.1 Bazı Regresyon Analizi Modelleri Çeşitleri

Regresyon Analizi Ayrımları	**Regresyon Modelleri**	**Açıklamalar**
Bağımlı Değişken Sayısına Göre	Tek Değişkenli Regresyon	Tek bir nicel bağımlı değişken olmalıdır.
	Çok Değişkenli Regresyon	İki veya daha fazla nicel bağımlı değişken olmalıdır.
Bağımsız Değişken Sayısına Göre	Basit Regresyon	Bir bağımlı ve sadece bir tane bağımsız değişken olmalıdır.
	Çoklu Regresyon	Bir bağımlı ve iki veya daha fazla bağımsız değişken olmalıdır.
Matematiksel Biçimine Göre	Doğrusal Regresyon	Tüm parametreler doğrusal olmalıdır. Değişken dönüştürme ile doğrusal hale getirilen yapılar da bu kısımda yer almaktadır.
	Doğrusal Olmayan Regresyon	Bazı parametrelerin doğrusal olmadığı durumlar. Bağımlı değişken ile bazı bağımsız değişkenlerin doğrusal olmayan ilişkide olmaları. Fakat değişken dönüştürme ile doğrusal hale getirilemeyen yapılar olmalıdırlar.
Bağımsız Gölge Değişkenli	Varyans Analizi Modelleri	Tüm bağımsız değişkenler nitel yapıda olmalıdırlar.
	Kovaryans Analizi Modelleri	Bazı bağımsız değişkenler nitel, bazıları ise nicel yapıda olmalıdırlar.
Varsayımların Sağlanma Durumuna Göre	Parametrik Regresyon	Parametrik regresyon analizinin başarılı bir şekilde uygulanabilmesi için verinin normal dağılması, eşit varyanslı olması, otokorelasyonlu olmaması gibi varsayımların sağlanması gerekmektedir.
	Parametrik Olmayan Regresyon	Parametrik regresyon yöntemler için geçerli olan bazı varsayımların sağlanamaması durumlarında kullanılan yöntemlerdir.

2.2. Bağımlı Değişken Sayısına Göre Regresyon Modelleri

2.2.1. Tek Değişkenli Regresyon Analizi

Tek değişkenli regresyon analizi bir bağımlı değişken ve bir veya birden fazla bağımsız değişken arasındaki bağı gözlemleyen analiz yöntemidir. Bu analiz ile bağımlı ve bağımsız değişken(ler) arasındaki doğrusal bağ elde edilir. Korelasyon analizinde olduğu gibi, regresyon analizinde de üzerinde durulan ilişki, parametreler arasındaki doğrusal ilişkidir. Bu doğrunun hesaplanmasında en küçük kareler yönteminden yararlanılır. Regresyon analizi sonuçlarının yorumlanmasında çoğu araştırmacı ve öğrenci tarafından önemli hatalar yapılmaktadır. Çoğunlukla yapılan hata ise, regresyon analizi sonuçlarının yorumlanmasında, X değişkeninin Y değişkenine sebep olması yorumudur. Bağımlı ve bağımsız parametreler arasında pozitif ve negatif bir ilişkinin var olması daima bağımsız değişkenlerin bağımlı değişken üzerinde etkili olduğu sonucunu meydana getirmeyecektir. Birden fazla parametre arasında bir bağ olabilmesi için nedensellik durumu aranmaz. İlişkinin nedeni belki iki değişkenin üçüncü bir değişkenle olan ilişkilerinden kaynaklanıyor olabileceği gibi, belki de ele alınan ilişki tamamen tesadüf olarak da ortaya çıkmış olabilir. Regresyon analizi parametreler arasındaki ilişkinin yapısı ve derecesi ile ilgilenmektedir (Wooldridge 2013; Gujarati 1999:74).

2.2.2. Çok Değişkenli Regresyon Analizi

Birden fazla bağımlı ve birden fazla bağımsız parametrenin bulunduğu regresyon modellerine çok değişkenli regresyon analizi adı verilir. Çok değişkenli regresyon analizinde bağımsız değişkenler aynı zamanda bağımlı değişkendeki farklılığı tanımlamaya çalışır. Hesaplama ve yorumlama açısından çok değişkenli regresyon analizinin yorumu tek değişkenli regresyon analizinin yorumuna benzer. Ancak birtakım farklılıklar göstermektedir. Örneğin, tek değişkenli regresyon analizindeki karşılığı çoklu regresyon katsayısı R olarak gösterilir. Çoklu regresyon katsayısı R, bir bağımlı değişkendeki değişim ile birden fazla bağımsız değişkendeki değişim arasındaki ilişkiyi açıklamaktadır. Çok değişkenli regresyon analizi sosyal bilimlerin çoğu dalında kullanılmaktadır. Pazarlama, sosyoloji ve psikoloji gibi davranışsal hareketlerin belirlenmesinde, ekonomide zaman serisi türü ekonomik değişkenleri etkileyen faktörlerin tespiti ve geleceğe yönelik tahmininde kullanım alanı bulmaktadır (Gujarati 1999:74).

2.3. Matematiksel Biçimine Göre Regresyon Modelleri

2.3.1. Doğrusal Regresyon Modeli

Doğrusal regresyon Y olarak adlandırılan sayısal bir bağımlı değişkenle X olarak belirtilen bir veya birden fazla bağımsız değişken arasındaki doğrusal yapıdaki ilişkiyi inceler. Regresyon modelindeki bağımsız değişken sayısı bir ise model basit doğrusal regresyon olarak ifade edilir. Modelde yer alan açıklayıcı bağımsız değişken sayısı birden fazla ise çoklu doğrusal regresyon olarak adlandırılır. Doğrusal regresyon modelinde bağımlı değişken sayısal bir değişken olma mecburiyetindedir. İstatistiksel olarak anlamlı bir regresyon analizi bağımlı ve bağımsız parametreler arasında nedensel ilişkiyi göstermez (Wooldridge 2013).

Kullanılan modeller parametrelere göre doğrusaldır. Değişkenlere göre doğrusallık modeldeki bütün değişkenlerin basit olmasıdır.

Doğrusal regresyonda belirtilen bağımlı değişken sürekli veya sıralı sayısal veriler olması gerekir; kategorik değişken olmamalıdır. Buna karşılık bağımsız değişkenlerin de tercihen sayısal olması gerekir ancak bazı durumlarda cinsiyet gibi kategorik de olabilir (Alpar 2010: 285-304; Kirkwood 2003: 315-342; Dawson 2001: 236-242).

Basit doğrusal regresyon modeli Eşitlik (2.2)'de verildiği gibidir.

$$Y_i = \beta_0 + \beta_1 X_i + e_i \tag{2.2}$$

Bu modelden elde edilen sonuçların başarılı olabilmesi için şu varsayımlar sağlanmalıdır:

1) Tüm bağımsız parametrelerin nicel ya da nitel olması, bağımlı parametre Y 'nin nicel ve sürekli olması gerekir.

2) Bütün bağımsız parametrelerin varyansının sıfırdan farklı değerler alması gerekir.

3) Bağımsız parametreler arasında doğrusal bir bağ olmamalıdır.

4) Hata terimleri ortalaması sıfırdır. $E\left(\varepsilon_j\right)=0$

5) Bağımsız parametreler ve hata terimi arasında korelasyon aranmamalıdır.

6) Hata terimlerinin varyansı sabit olmalıdır. $E\left(\varepsilon_i^2\right)=\sigma^2$

7) Hata terimleri arasında otokorelasyon olmamalıdır. $E\left(\varepsilon_j\varepsilon_i\right)=0,(i\neq j)$

8) Hata terimleri ε_i, normal dağılmalıdır (Gujarati 1999:61-68).

2.3.1.1 Parametre Tahmini

Regresyon analizinde birçok parametre tahmin yöntemi kullanılmaktadır. Burada çoğunlukla kullanılan tahmin yöntemleri En Küçük Kareler (EKK) ve En Çok Olabilirlik (EÇO) tahmin yöntemleri kısaca tanıtılacaktır.

En Küçük Kareler Yöntemi

$(x_1,y_1)(x_2,y_2),...,(x_n,y_n)$ gibi n veri çifti olduğu zaman, kitle regresyon doğrusunun bilinmeyen katsayıları β_0 ile β_1'in tahminlerini elde etmek gerekir. β_0 ve β_1'in tahmin değerleri b_0 ve b_1 olsun, böylece tahmin edilen doğru aşağıdaki denklem (2.3) ile ifade edilmiş olacaktır.

$$y=b_0+b_1x \tag{2.3}$$

Bunun doğru bir tahmin olup olmadığını öğrenebilmek için, (x_i,y_i) noktalarının bu doğrudan uzaklıklarını ölçecek bir ölçüye gereksinim duyulur. Bağımlı değişkenin bilinen değeri y_i 'dir. Bilinen değer ile öngörülen değer arasındaki fark $e_i=y_i-(b_0+b_1x_i)=y_i-\hat{y}_i$ şeklinde ifade edilir. e_i farkı, bağımlı değişkenin önerilen doğru üzerindeki $b_0+b_1x_i$ değerinden sapmasını göstermektedir (Newbold 1995:567).

En küçük kareler yöntemi, bilinmeyen β_0 ve β_1 değişkenlerinin tahmin edilmesinde yararlanılan metotlardan biridir. Bu metot ile, hataların karelerine ait toplamın en küçük yapılması amaçlanır. Bu amaçla $y_i-\overline{y}_i=y_i-(b_0-b_1x_i)$

eşitliğinin her iki tarafının karesi alınıp n gözlem için yeniden yazıldığı zaman aşağıdaki Eşitlik (2.4) ile elde edilmiş olunur;

$$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}(y_i-b_0-b_1x_{i1})^2 \tag{2.4}$$

Eşitliğin sağ tarafını sıfıra eşitleyip, b_0 ve b_1 'e göre türev alındığı zaman, hata kareler toplamı $\sum_{i=1}^{n}e_i^2$ 'yi en küçük yapan b_0 ve b_1 tahmin değerleri aşağıdaki Eşitlik (2.5) ve (2.6) ile gösterilmiştir;

$$b_0=\bar{y}-b_1\bar{x} \tag{2.5}$$

$$b_1=\frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_{i1}\bar{x}_1)}{\sum_{i=1}^{n}(X_i-\bar{X}_1)^2}=\frac{\sum_{i=1}^{n}x_iy_i-\frac{\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n}}{\sum_{i=1}^{n}X_{i1}{}^2-\frac{\left(\sum_{i=1}^{n}X_i\right)^2}{n}}=\frac{\sum_{i=1}^{n}x_iy_i-n\overline{xy}}{\sum_{i=1}^{n}x_i{}^2-n\bar{x}}=\frac{ÇT_{XY}}{KT_X} \tag{2.6}$$

KTx'e, x_i 'lerin ortalamaya göre düzeltilmiş kareler toplamı denir. ÇTxy: x ve y değişkenlerinin kendi ortalamalarına göre düzeltilmiş değerlerinin toplanması ile elde edilir. Düzeltilmiş çarpımlar toplamı olarak da adlandırılır. Burada, $\bar{y}=\sum_{i=1}^{n}y_i/n-b_1\bar{x}$ ve $\bar{x}=\sum_{i=1}^{n}x_i/n$ olarak tanımlanır. Buradan, örnekleme ilişkin regresyon tahmin denklemi; $\hat{y}_i=b_0-b_1\bar{x}_i$ olarak yazılır. Bu denklemde, her biri x değerinin yerine koyulması ile elde edilen $\hat{y}_i$ değerleri regresyon doğrusu üzerinde olacaktır (Alpar 2003: 259).

En Küçük Kareler Tahmin Edicilerinin Özellikleri şu şekilde sıralanabilir:

1-Doğrusaldır, yani regresyon modelindeki bağımlı değişken Y gibi tesadüfi bir değişkenin doğrusal fonksiyondur.

2-Yansızdır, yani ortalaması ya da beklenen değeri E (β_2) $=\beta_2$ 'dir.

3-Doğrusal yansız tahmin ediciler içinde en küçük varyanslı olanıdır; en küçük varyanslı yansız bir tahmin edici etkin bir tahmin edici olarak adlandırılır (Gujarati 1999:71-72).

En Çok Olabilirlik Yöntemi

Hataların normallik varsayımı altında regresyon modelinin matris gösterimi Eşitlik (2.7) ile belirtilir,

$$Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \varepsilon_{n\times 1}\,,\, \varepsilon \sim N\left(0,\sigma^2 I\right) \tag{2.7}$$

Bu denklemde $\beta \in \Re^P$ ve σ^2 parametrelerini tahmin etmek için en çok olabilirlik yönteminden faydalanılır ve Eşitlik (2.8) ile gösterilir:

$$L\left(\beta,\sigma^2;Y\right) = \frac{1}{\left(\sqrt{2\pi}\right)^{n/2}\left(\sigma^2\right)^{n/2}} e^{\frac{-1}{2\sigma^2}(Y-X\beta)'(Y-X\beta)} \tag{2.8}$$

Olabilirlik fonksiyonunun logaritması Eşitlik (2.9), Eşitlik (2.10), Eşitlik (2.11) ve Eşitlik (2.12) ile gösterilir;

$$\ln L\left(\beta,\sigma^2;Y\right) = \ell\left(\beta,\sigma^2;Y\right) - \frac{n}{2}\ln\left(2\pi\right) - \frac{n}{2}\ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\left(Y - X\beta\right)'\left(Y - X\beta\right) \tag{2.9}$$

$$= -\frac{n}{2}\ln\left(2\pi\right) - \frac{n}{2}\ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\left(Y'Y - 2\beta' X'Y + \beta' X' X\beta\right) \tag{2.10}$$

$$\frac{\ell\left(\beta,\sigma^2;Y\right)}{\partial\beta} = -\frac{1}{2\sigma^2}\left(-2X'Y + 2X'X\beta\right) \tag{2.11}$$

$$\frac{\ell\left(\beta,\sigma^2;Y\right)}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\left(\left(Y - X\beta\right)'\left(Y - X\beta\right)\right) \tag{2.12}$$

Yukarıda gösterilen Eşitlikten yararlanılarak $\beta \in \Re^P$ ve σ^2 değişkenlerinin En Çok Olabilirlik Tahmin Edicilerini Eşitlik (2.13) ve Eşitlik (2.14) ile ifade edilir (Genç 1997:1-103).

$$\hat{\beta} = \left(X'X\right)^{-1} X'Y \tag{2.13}$$

$$\hat{\sigma}^2 = \frac{1}{n}\left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right) \tag{2.14}$$

2.3.2. Doğrusal Olmayan Regresyon Modeli

Bağımlı ve bağımsız değişkenler arasındaki ilişki denkleminin doğrusal olmadığında yapılan regresyon analizine doğrusal olmayan regresyon analizi veya kısaca eğrisel regresyon analizi adı verilmektedir. Sıklıkla yararlanılan eğrisel regresyon modelleri polinomsal, üstel ve logaritmik regresyon modelleridir. Eğrisel regresyon modelinde regresyon katsayılarını tahmin edebilmek için en küçük kareler yönteminden yararlanılırken hata kareler toplamından katsayılara göre türevler alınarak elde edilen normal denklemleri çoğunlukla çözülememektedir. Bundan dolayı eğrisel model hata kareler toplamına yerleştirilirken eğrisel modelin sadece iki terimli Taylor açılımından yararlanılır. Böylece eğrisel model birinci dereceli doğrusal bir denklem ile temsil edilmeli ve en küçük kareler yönteminin çözümü bulunabilmektedir. Taylor açılımı yapılırken başta eğrisel model tahmini katsayı değerleri civarında seriye açılarak işlem yapılır. En küçük kareler çözümü elde edildikten sonra işlemler bir kere daha tekrarlanırken Taylor açılımı elde edilen bu çözüm etrafında yapılmalıdır. Çözümler birbirine yaklaştıkça optimal regresyon katsayı değerlerine ulaşılmaktadır (Gujarati 1999:525-530). Uygulamada en çok kullanılan doğrusal olmayan regresyon modelleri Üstel Regresyon ve Lojistik Regresyon Modelleri olarak karşımıza çıkmaktadır (Maddala 2001).

2.3.2.1. Doğrusal Olmayan Regresyon Modellerinde Parametre Tahmini

Parametreler açısından doğrusal olmayan regresyon modeli Eşitlik (2.15) ile ifade edilir;

$$y_i = f\left(x_i, \theta_*\right) + \varepsilon_i\ ,\ \ i = 1, 2, \ldots, n \tag{2.15}$$

Bu modelde ε_i rassal hata terimi olup, bu rassal değişkenlerin bağımsız ve $\varepsilon_i \sim \left(0, \sigma^2\right)$ oldukları varsayılmaktadır. Modelde f fonksiyonu beklenen fonksiyonu, x_i vektörü bağımsız değişkenleri ve θ vektörü p tane bilinmeyen doğrusal olmayan parametreleri belirtmektedir. Parametre vektörünün gerçek değeri

θ_* şeklinde ifade edilmiştir. Doğrusal olmayan regresyon modellerinde $E(y_i)$ parametrelerin doğrusal olmayan bir fonksiyonu olmak üzere parametrelerine göre türevlerinden en az bir tanesi en az bir parametreye bağımlıdır (Bates, Watts 1988:32-167; Bierens, Gallant 1997). Bu özellik parametrelere göre doğrusal olmamanın bir göstergesidir. Varsayımları bu şekilde verilen modelin parametrelerinin tahminleri En Küçük Kareler veya Genelleştirilmiş En Küçük Kareler Yöntemi yöntemiyle bulunabilir.

En Küçük Kareler Yöntemi

Rastgele elde edilen bir θ için $r_i = y_i - f(x_i, \theta)$ doğrusal olmayan model için artık kareler toplamı fonksiyonu Eşitlik (2.16) ile gösterilmiştir;

$$S(\theta) = \sum_{i=1}^{n} \left[y_i - f(x_i, \theta) \right]^2 \tag{2.16}$$

Buna bağlı olarak en küçük kareler tahmini $\hat{\theta}$ ile ifade edilir ve $S(\theta)$ kareler toplamı fonksiyonunu minimum eder (Denison vd. 2003:1-5).

Genelleştirilmiş En Küçük Kareler Yöntemi

Bu metot ağırlıklandırılmış en küçük kareler metodu olarak da bilinmektedir. Genelleştirilmiş ya da ağırlıklandırılmış en küçük kareler metodu adı altında birçok yaklaşım toplanmaktadır (Denison vd. 2003:1-5). Sunulan regresyon modelinin hataları homojenlik (bağımsız ve sabit varyanslı) koşuluna uymadığında en küçük kareler metoduyla oluşan parametre tahminleri gerçek değerleri ile ilişkisiz olmaktadır (Bates, Watts 1988:32-167). Hataların varyanslarının değişkenlikleri ve hatalar arasında belirli korelasyonlar varsa hatalar doğru biçimde ağırlıklandırılarak yeni oluşturulan hataların homojenliği sağlanmaktadır.

Tanım olarak $y_i = f(x_i, \theta_*) + \varepsilon_i$, $i = 1, 2, ..., n$ modelinin hatalarının $E(\varepsilon) = 0$ ve $V(\varepsilon) = E(\varepsilon\varepsilon') = \sigma^2 V$ niteliği bulunuyorsa, bu durumda V matrisi bilinmemekle birlikte hataların temsili olarak önerilen regresyon modelinin artıkları ele alınarak V matrisinin tahmini yapılabilmektedir. Bu açıdan V matrisi, tahmini regresyon modeline bağımlı olduğu gibi tahmin için yararlanılan metotlara da bağımlı hale gelir. Ağırlıklandırılmış en küçük kareler metodunun bütünlüğünü sağlayabilmek için V matrisinin bilindiği varsayılarak $V = K^2$ olacak biçimde simetrik $K^t = K$ matrisi elde edilir. Sabit bir matris olarak alınan K matrisi aracılığıyla hatalar homojenlik koşulunu sağlar hale getirilir. Bu şekilde ι (yota) hataları $E(\iota) = 0$ ve $V(\iota) = \sigma^2 I$ ile homojenlik koşulunu Eşitlik (2.17) ve Eşitlik (2.18) sağlamış olur.

$$K^{-1}\varepsilon = \iota \rightarrow E(\iota) = K^{-1}E(\varepsilon) = 0 \tag{2.17}$$

$$\rightarrow V(\iota) = K^{-1}V(\varepsilon)K^{-1} = K^{-1}\sigma^2 VK^{-1} = \sigma^2 I \tag{2.18}$$

K^{-1} dönüşümü ile $y = f(\theta_*) + \varepsilon$ modelinin yeni formu aşağıdaki Eşitlik (2.19) ve (2.20) ile gösterilir;

$$y = f(\theta_*) + \varepsilon \rightarrow K^{-1}y = K^{-1}f(\theta_*) + K^{-1}\varepsilon \Rightarrow z = g(\theta_*) + \iota \tag{2.19}$$

$$K^{-1}y = z\,, K^{-1}f(\theta_*) = g(\theta_*)\,, K^{-1}\varepsilon = \iota \tag{2.20}$$

Bu şekildeki bir dönüştürmeden sonra ortaya çıkan modele en küçük kareler yöntemi uygulanarak genelleştirilmiş en küçük kareler yöntemi tamamlanır. Daha sonra homojenlik şartlarını sağlayan hatalara sahip olan dönüştürülmüş modelin parametre tahmini aşağıda verilen Eşitlik (2.21) ile gösterilir (Seber, Wild 1989:151-160).

$$R = \frac{\partial g(\theta)}{\partial \theta^t} = K^{-1}\frac{\partial f(\theta)}{\partial \theta^t} = K^{-1}Z \tag{2.21}$$

Kurulan bu Eşitlikten yararlanarak $R_* = R(\theta_*)$, $\hat{R} = (\hat{\theta})$ ve $R_k = R(\theta_K)$ dönüşümleri kullanılarak ağırlıklandırılmış en küçük kareler yöntemi sonuçları benzer şekilde oluşturulur. Ağırlıklandırılmış $z = g(\theta_*) + \iota$ modeli doğrusallaştırılarak en küçük kareler yöntemi ile Gauss-Newton iterasyonu bulunur. $R_k = R(\theta_k)$ ile θ_k civarında $g(\theta) \cong g(\theta_k) + R_k(\theta - \theta_k)$ açılımı yardımıyla kareler toplamı Eşitlik (2.22) ile gösterilir.

$$S(\theta) = \left\| z - g(\theta) \right\|^2 \cong \left\| z - g(\theta_k) - R_k(\theta - \theta_k) \right\|^2 \tag{2.22}$$

fonksiyonunu minimize eden çözüm (2.23) ile ifade edilir ve buradan $R_k = K^{-1} Z_k$ ile birlikte Gauss-Newton algoritması denklem (2.24) ile elde edilir.

$$\theta_{k+1} = \theta_k + \left(R_k^t R_k \right)^{-1} R_k^t \left(z - g(\theta_k) \right) \tag{2.23}$$

$$\theta_{k+1} = \theta_k + \left(Z_k^t V^{-1} Z_k \right)^{-1} Z_k^t V^{-1} \left[y - f(\theta_k) \right] \tag{2.24}$$

Yardımıyla $\hat{\theta}_G$ genelleştirilmiş en küçük kareler tahminine ulaşılır.

2.4. Varsayımların Sağlanma Durumuna Göre Regresyon Modelleri

2.4.1. Parametrik Regresyon Analizi

Parametrik regresyon analizinin önemli bir niteliği regresyon fonksiyonunun şeklinin önceden biliniyor olmasıdır. Ayrıca çoğunlukla yararlanılan En Küçük Kareler Metoduna dayanan parametrik regresyon analizinde; bağımsız parametre değerlerinin tümü için hata varyanslarının sabit olması, hata terimleri arasında otokorelasyonun bulunmaması, hata terimlerinin normal dağılımlı olması ve eğer çoklu regresyon analizi varsa, bağımsız parametreler arasında çoklu doğrusal bağlantı olmaması gibi önemli hipotezlerin sağlanması gerekir. Eğer hipotezler sağlanamazsa, regresyon fonksiyonu hakkında yapılan tahminler gerçeğe uymaz ve sonuçlar yanlış yorumlanır. Bununla birlikte, hipotezlerin sağlanmadığı anlaşılırsa, bazı düzeltici önlemler alınarak hipotezler sağlanabilir. Böylece parametrik regresyon analizi ile elde edilen model gerekli varsayımları sağladığı için öngörülerde yararlanılabilmektedir (Hardle 1994:6-11).

2.4.2 Parametrik Olmayan Regresyon Analizi

Parametrik olmayan regresyon analizinde fonksiyonun şekli önceden belirlenemez ve parametrik regresyon analizinde olduğu gibi önemli hipotezler söz konusu değildir. Yapılan önerme, hata terimlerinin ortalamasının sıfır ve varyansının da sonlu bir sayı olması yönündedir. Böylece değişkenler arasındaki bağın ortaya çıkamasında esneklik sağlanılır. Bu nedenle parametrik olmayan regresyon analizi “verinin kendi kendisini ifade etmesine izin veren” metot olarak da söylenebilir (Ullah 1988:625-658).

Parametrik regresyon yöntemler için gerekli olan bazı varsayımların sağlanamadığı durumlarda kullanılan yöntemlerdir. Örneklem hacmi oldukça düşük veya aykırı gözlemlerin etkilerini farklı biçimlerde inceleyen güçlü (Robust) parametrik yöntemler bulunmaktadır. Bununla birlikte, aykırı gözlemlerden dolayı parametreler bozulduğu için bu güçlü yöntemler bile benzer çözümler üretemeyebilir ve verinin gerçek yapısı modele yansıtılamayabilir. Böyle bir durumda parametrik olmayan regresyon, ön bilgi oluşturmaktadır (Hardle 1994:11).

Parametrik olmayan regresyon yönteminde öngörü yapılırken kısıtlayıcı varsayımları olmamasına karşın bazı sakıncaları vardır. Bağımsız değişken sayısı çok olduğu durumda öngörüde bulunmak zor olmakla birlikte elde edilen grafikler karmaşık bir yapıda karşımıza çıkmaktadır. Parametrik olmayan yöntemlerin sakıncalarından yarı parametrik regresyon modelinden yararlanılarak giderilebilmektedir (Horowitz 1993:49-70).

2.4.3. Parametrik Regresyon ile Parametrik Olmayan Regresyonun Karşılaştırılması

Parametrik olmayan yöntemler, parametrik yöntemlere göre birimlerin büyüklüklerinden ziyade onların sırası ile ilgilendikleri için normal dağılım varsayımı gerektiren çözümlemelerin gerçekleştirilemediği çoğu koşulda yararlanılabilir (Erilli, Alakuş 2016:30). Bilgisayar teknolojisinin ilerlemesi, sadece yaklaşık p-değerlerini içeren sonuçlar yerine, koşullu parametrik olmayan yöntemler için fazlaca işlem gerektiren kesin p-değerlerinin hızlı bir şekilde elde edilmesini sağlamıştır. Buna ek olarak, jackknife ve yinelemeli örnekleme (bootstrap) gibi metotlar ile parametrik olmayan metotlar bir arada ele alınmıştır (Gibbons 1999:3).

Parametrik regresyon analizinde, biçimi önceden oluşturulmuş modelin parametreleri öngörülürken, parametrik olmayan regresyonda ise amaç, regresyon fonksiyonu olan m(x)'i doğrudan tahmin etmektir (Hardle vd. 2004:115). Bunun yanı sıra, parametrik olmayan regresyon için model elde edilirken bilinmeyen regresyon eğrisini kapsayan doğru fonksiyon uzayı belirlenmekte ve bu seçim oluşturulurken regresyon fonksiyonunun süreklilik ve türevlenebilme gibi düzgünlük kriterlerine sahip olması istenilir (Eubank 1990: 4).

Parametrik regresyon analizinde; regresyon fonksiyonunun biçimi ile birlikte X veri iken Y ’nin koşullu varyansı ve hata terimi u ile X arasındaki otokorelasyon gibi diğer fonksiyonlar gösterilmiştir. Bunun yanı sıra X ile Y ’nin parametrik bileşik yoğunluğunun çoğunlukla normal dağıldığı öngörülür. Parametrik olmayan regresyon analizinde ise söz edilen bu kuramlar yapılmaz. Yapılan en önemli varsayım hata terimlerinin sıfır ortalama ve sonlu bir varyansa sahip bir dağılımdan geldiğidir (Ullah 1988:625-658). Parametrik olmayan regresyon modeli seçilirken bilinmeyen regresyon eğrisini kapsayan doğru fonksiyon uzayı seçilir. Bu seçim yapılırken regresyon fonksiyonunun süreklilik ve türevlenebilme gibi düzgünlük niteliğine sahip olduğu noktasından hareket edilir (Eubank 1990:4).

Parametrik olmayan regresyonun güçlü kuramlarda bulunmaması ve sabit parametrik bir modele bağlı kalmaması dışında; iki parametre arasında bilinen ilişkinin ortaya çıkmasında esnek olması ve ardışık X değerleri arasında interpolasyon ve kayıp gözlemlerin yerine konulması için esnek metotlar elde etmesi gibi avantajlarından da söz edilir (Hardle 1994: 6).

Parametrik olmayan regresyonda, yakın gözlemlerin incelenmesi ile elde edilen değişkenler arasındaki ilişki ortaya çıktığından çok sayıda veriye ihtiyaç duyulur (Yatchew 1998: 672).

2.4.4. Yarı Parametrik Regresyon Analizi

Yarı parametrik regresyon modeli, hem parametrik hem de parametrik olmayan regresyon modelinden faydalanır. Böylece yarı parametrik regresyon modeli parametrik modellerin kısıtlayıcı varsayımlarından etkilenmemekte ve parametrik olmayan yöntemlerin cazip yönlerini bir araya getirmektedir (Cox, Kernel Regresyon gibi) (Erilli, Alakuş 2016:31).

Yarı parametrik regresyon modelleri, parametrik regresyon modeli ve parametrik olmayan regresyon modellerinin birleşimi şeklinde ifade edilebilir. Yarı parametrik regresyon modelleri; parametrik olmayan regresyon yöntemlerinin iyi yöntemler oluşturamadığı şartlarda veya araştırmacının parametrik yöntemlerden yararlanmak istemesine karşın hataların dağılımlarının bilinemediği şartlarda kullanılmaktadır. Bu modellerde parametre tahmini yapılırken normallik varsayımına ihtiyaç duyulmamaktadır (Takezawa 2005).

Parametrik olmayan model tahmininde yorumlanabilir sonuçlar elde etmek amacıyla en fazla iki açıklayıcı değişkenle çalışmak mümkün iken, yarı parametrik yöntemde k tane açıklayıcı değişkenin bağımlı etkisini incelemek mümkündür. Ayrıca parametrik modeldeki kadar varsayım yapılamaması nedeniyle bu yaklaşımdan uygulamalı çalışmalarda yararlanılması ileri sürülmektedir (Horowitz 1993:49-70).

BÖLÜM 3

3. SAYMA VERİLERİ İÇİN REGRESYON MODELLERİ

Doğrusal regresyon ve varyans analizi gibi doğrusal modellerin temel varsayımlarından biri, hataların normal bir dağılıma uymasıdır. Sürekli bir bağımlı değişken çarpık hale geldiğinde bu varsayımı karşılamak için, bağımlı değişkenin dönüşümü ile elde edilen yeni bir model yapısı ile yaklaşık normal olan hatalar üretebilir. Bununla birlikte, çoğu zaman, ilgilenilen yanıt değişkeni sürekli değildir, kategorik ya da kesiklidir. Bu durumda, basit bir dönüşüm normal dağılmış hataları üretemeyecektir.

Sayma verisi, herhangi bir olayın belirli bir süre içerisinde kaç kez meydana geldiğini gösterir. Sayma modellerindeki ilk gelişmeler aktüerya, biyoistatistik ve demografide gözlemlenmiştir. Bu modeller son zamanlarda ekonomi, siyaset bilimleri ve sosyolojide de geniş bir biçimde kullanılmaktadır (Dinarcan 2018:9).

Sayma verisi için regresyon modellerinin geliştirilmesindeki dönüm noktası, Poisson Regresyonun özel bir durumu olduğu "Genelleştirilmiş Doğrusal Modellerin" ortaya çıkmasıdır. İlk olarak Nelder ve Wedderburn tarafından tanımlanmış, McCullagh ve Nelder tarafından detaylandırılmıştır (Nelder, Wedderburn 1972:370-384; McCullagh, Nelder 1989).

Sayma veri regresyon modelleri kesikli regresyonun özel bir türüdür. Negatif olmayan kesikli değerlerden elde edilir. Bu verilerde doğrusal regresyon modelinin uygulanması öngörülen katsayıların yanlı olmasına neden olur. Sayım verisini elde etmek için yararlanılan en yaygın regresyon modeli "Poisson Regresyon", varsayımların sağlanamadığı durumda ise "Negatif Binom Regresyon" modelidir (Cameron, Trivedi 1998:59-139).

3.1 Poisson Regresyon Analizi

Poisson Regresyon analizi, bağımsız değişkenler ile sayımla ulaşılan bağımlı değişken arasındaki bağlantıyı gösteren bir metottur. Poisson regresyon analizindeki baz alınan yapı, Y_i yanıt değişkeninin kesikli bağımsız Poisson rastlantı değişkeni olmasıdır. Kesiklilik sebebiyle normallik varsayımı sonucu elde edilememesiyle klasik doğrusal regresyon analizine ek olarak gösterilen metotlardan biridir (Frome vd. 1973:935-940; Frome 1983:665-674).

Poisson Regresyon modeli, Poisson dağılımının ortalamasına göre belirlenir ve model Eşitlik (3.1) ile ifade edilmiştir (Cameron, Trivedi 1986:29-53);

$$p(y_i / x_i) = \begin{cases} \dfrac{e^{-\mu_i} \mu_i^{yi}}{y_i!}, y = 0,1,2,... \end{cases} \tag{3.1}$$

Burada μ_i Eşitlik (3.2)'de verildiği gibi belirtilmiştir;

$$E(y_i / x_i) = \mu(x_i) = c_i f(x_i, \beta) = \mu_i, i = 1,...,n \tag{3.2}$$

Eşitlik (3.1) ve Eşitlik (3.2)'de, $x_i = (x_{i0,...,}x_{im}), 1 \times (m+1)$ boyutlu i'nci kümeye ait satır vektörünü, $\beta = (\beta_{i0,...,}\beta_{ik}), (k+1) \times 1$ boyutlu bilinmeyen parametrelerin meydana getirdiği sütun vektörünü belirtmektedir. Poisson rassal değişkeninin aldığı y_i değerleri, çoğunlukla bir denemedeki başarısızlık sayısını gösterir (Bu başarısızlık kimi zaman ölüm sayısı, kimi zaman trafik kazası sayısıdır). μ_i'ler denemelerdeki ortaya çıkan durumun ortalama oluş sayısını, c_i'ler üzerinde durulan olay için riskli toplam kişileri veya kitle çoğunluğunu, $f(x_i, \beta)$'da regresyon hız fonksiyonunu belirtmektedir.

Poisson Regresyon analizinde çoğunlukla yararlanılan regresyon fonksiyonu log-doğrusal model olmakla beraber doğrusal ve doğrusal olmayan modellerde de uygulanabilmektedir. Log-doğrusal modelden yararlanarak Poisson Regresyon modelinin ortalama parametresi Eşitlik (3.3) gibidir;

$$E(y_i / x_i) = \mu_i = \exp(x_i\beta) = \exp(x_{i0}\beta_0 + ... + x_{im}\beta_k), i = 1,...,n \tag{3.3}$$

Doğrusal ve doğrusal olmayan model için bu denklem Eşitlik (3.4) ve Eşitlik (3.5) ile ifade edilir (Özmen 1998:1-77).

- Doğrusal model

$$E(y_i / x_i) = (x_i\beta) = (x_{i0}\beta_0 + ... + x_{im}\beta_k), i = 1,...,n \tag{3.4}$$

- Doğrusal olmayan model; (*k=3*, *m=2*)

$$E\left(y_i / x_i\right) = x_{i0}\beta_0 + x_{i1}\beta_1\left\{1-\left[1-\exp\left(x_{i1}\beta_2\right)\right]^{\beta_3}\right\}, i = 1,...,n \quad (3.5)$$

3.1.1. Poisson Dağılımı

Poisson dağılımı belirli bir zaman ve belirli bir alanda meydana gelebilecek olay sayısını ifade eden bir dağılımdır. Dağılım 18. yüzyılda yaşamış olan Fransız matematikçi Poisson tarafından ortaya atılmıştır (Gürsakal 1997:238).

Poisson dağılımı Eşitlik (3.6) ile ifade edilmiştir;

$$P(y;\mu) = \frac{\mu^y e^{-\mu}}{y!}, \quad y = 0,1,2,... \quad (3.6)$$

Bu eşitlikte y_i istenen olayların meydana gelme sayısını belirtmektedir. Poisson dağılımı tek parametreli bir dağılımdır ve dağılımın parametresi μ_i'dir. Poisson dağılımının şartlı beklenen değeri $E(y_i) = \mu_i$'dir. μ_i'nün açıklayıcı değişkenlere bağlı olarak tanımlanması durumunda Poisson regresyon modeli elde edilmektedir. μ_i çoğunlukla $\mu_i = e^{x'\beta}$ şeklinde gösterilmektedir. Burada x açıklayıcı değişken vektörünü gösterirken, β de tahmin edilecek parametre vektörünü belirtmektedir (Memiş 2013:9).

3.1.2. Aşırı Yayılım Durumu

Poisson Regresyon'da aşırı yayılım durumu genellikle model kurma sorunlarını ortaya çıkarmaktadır. Aşırı yayılım sorununu ortaya çıkaran birkaç önemli durumlar şunlardır (Miaou 1994:471-482):

- İncelenen bir Poisson sürecinin aralık uzunluğunun önceden belirlenmesi
- Verinin, her bir olayın rastlantıyla kümelenmiş Poisson sürecinden elde edilmesi
- Genel hipotez olan bağımsızlığın elde edilememesi
- Oluşturulan model için önemli açıklayıcı değişkenlerin belirlenememesi
- Önemli açıklayıcı değişkenlerin modele dahil edilmemesi

Aşırı yayılım durumu, modelde istenmeyen problemleri meydana getirebilmektedir. Meydana gelen problemlerden bir tanesi modelin açıklayıcılık yönünün yeterli olmamasıdır. Eğer model açıklayıcılık yönünden yeterli değilse aşırı yayılım durumundan kuşku duyulması gerekmektedir. Poisson regresyon modelinde aşırı yayılımın göz ardı edilmesi, regresyon parametrelerine ait en çok olabilirlik yöntemi gibi metotlardan yararlanılarak ulaşılan öngörülerin tutarlı olmasına karşın, varyanslarının olması gereken değerinden düşük tahmin edilmesine sebep olmaktadır. Diğer bir gösterimle, gözlem sayısı büyükken, Poisson modeline ilişkin regresyon parametrelerine ait beklenen değerler ile gerçek parametre değerleri arasındaki farkın fazla olmamasına rağmen, öngörülen parametrelerin anlamlılık düzeyleri olması gerekenden daha fazladır. (Yip, Yau 2005:153-163).

3.1.3. Aşırı Yayılım Testleri

Verinin aşırı yayılımlı mı az yayılımlı mı olduğunu anlamak için oluşturulan Poisson modelinin varyansı ortalamaya göre hesaplanır. Hesaplanan bu değer 1'den fazla ise aşırı yayılım, 1'den küçük ise az yayılım durumundan bahsetmek mümkündür.

Aşırı yayılım durumunu belirleyebilmek için bazı test istatistiklerinden faydalanılabilmektedir. Skor Testleri, gözlemlenen verinin Poisson regresyonuna göre aşırı yayılım olup olmadığını belirlemede yol gösterir.

Büyük örneklemler için önerilen kısmi skor test istatistiği Eşitlik (3.7) ile ifade edilmektedir;

$$T = \frac{1}{2}\sum_{i=1}^{n}\left\{\left(Y_i - \mu_i\right)^2 - Y_i\right\} \tag{3.7}$$

T istatistiğinin standartlaştırılmış şekli olarak ifade edilen T_1 test istatistiği Eşitlik (3.8)' de verilmiştir;

$$T_1 = \frac{\sum_{i=1}^{n}\left\{\left(Y_i - \mu_i\right) - Y_i\right\}}{\left(2\sum_{i=1}^{n}\mu_i^2\right)^{\frac{1}{2}}} \tag{3.8}$$

T_1 test istatistiği, büyük gözlem sayısı için standart normal dağılıma yakınlaştığından istatistik için standart normal dağılım tablo değeriyle kıyaslanarak aşırı yayılım durumu ölçülmektedir (Tüzel 2011:35).

Negatif binom modelinin yayılım parametresinin (α) sıfıra ve genelleştirilmiş Poisson modelinin yayılım parametresinin (φ) bire eşit olduğu durumda, incelenen verinin aşırı yayılım durumu içermediğini göstermektedir. Böylece Poisson regresyonundan yararlanılması öngörülmektedir. Her iki modele bağlı yayılım parametrelerinin gösterilen değerlere eşitliği hipotez testleri uygulanarak belirlenmektedir.

Negatif binom modelinin yayılım parametresi için kullanılan yokluk hipotezi ve alternatif hipotezler sırasıyla aşağıda verildiği gibidir;

$$H_0 : \alpha = 0$$
$$H_1 : \alpha \rangle 0$$

Genelleştirilmiş Poisson modelinin yayılım parametresinde uygulanan yokluk hipotezi ve alternatif hipotezi ise şu şekildedir;

$$H_0 : \varphi = 1$$
$$H_1 : \varphi \rangle 1$$

Yokluk hipotezlerinin reddedilmesi, yayılımın varlığını belirtir (Tüzel 2011:36). Yokluk hipotezinin testinde Olabilirlik Oran Testi ile Wald Testi uygulanabilir.

Olabilirlik Oran Testi

Olabilirlik Oran Testi, Poisson veya Negatif Binom regresyon modellerinde aşırı yayılımın olup olmadığını belirlemede kullanılır. Olabilirlik oran test istatistiği Eşitlik (3.9) ile ifade edilir;

$$LR = 2\left(\ln L_1 - \ln L_0\right) \tag{3.9}$$

Burada;

L_0 : Yokluk hipotezinin doğruluğu tahmin edilen en çok olabilirlik fonksiyonu,

L_1 : Seçenek hipotezinin doğruluğu tahmin edilen en çok olabilirlik fonksiyonudur (Ismail N. Zamani 2013:1-18).

Wald Testi

Wald testi, aşırı yayılım durumunu ölçmede yararlanılır. Yayılım parametresi α ile belirtildiğinde Wald test istatistiği Eşitlik (3.10)' da belirtilmiştir;

$$W = \frac{\hat{\alpha}^2}{V(\hat{\alpha})} \tag{3.10}$$

$\hat{\alpha}$: yayılım parametresinin tahmini,

$V(\hat{\alpha})$: tahminin varyansını belirtmektedir. (Ismail N. Zamani 2013:1-18)

3.2.Negatif Binom Regresyon

Negatif Binom Regresyon Analizi, Poisson Regresyon Modeline göre aşırı yayılım durumu gösteren verilerin modellenmesi için yararlanılan dağılımlardan biridir. Veri kümesinde aşırı yayılım meydana gelmesinde Poisson regresyon modelinden yararlanmak yanlı parametre tahminlerinin elde edilmesine ve böylece geçersiz sonuçların ortaya çıkmasına sebep olmaktadır. Veri kümesinde aşırı yayılım meydana geldiği zaman, yayılımı dikkate alan Negatif Binom (NB) Regresyon Analizinin kullanılması daha uygun olmaktadır. NB, aşırı yayılım durumundan kaynaklanan etki için modele bir değişken ekleyerek analiz yapmaktadır (Boucher vd. 2007:110-131; Denuit, Marechal, Pitrebois, Walhin 2007: 28-29). NB regresyon modeli aşağıda ifade edilen Eşitlik (3.11) ile gösterilir,

$$P(Y=y) = \frac{T\left(Y+\frac{1}{k}\right)}{T(Y+1)T\left(\frac{1}{k}\right)}\left(\frac{1}{1+k\mu}\right)^{1/k}\left(\frac{k\mu}{1+k\mu}\right)^{Y} \tag{3.11}$$

Burada k>0 yayılım parametresini belirtmektedir. Modelin ortalaması ve varyansı Eşitlik (3.12) ve Eşitlik (3.13) ile ifade edilmektedir:

$$E(Y) = \mu \tag{3.12}$$

$$V(Y) = \mu + k\mu^2 \tag{3.13}$$

Ortalama ve varyans eşitliklerinden de görüldüğü gibi yayılım parametresinin modele dahil edilmesi ile oluşturulan Negatif Binom Regresyon Modeli, varyansın ortalamadan yüksek hesaplanmasına neden olarak aşırı yayılımın modele eklenmesini sağlamaktadır. k'nın sıfıra yaklaştığı durumda Negatif Binom Regresyon Modelinin sınırlandırılmış ya da özel bir durumu olan Poisson regresyon modeli oluşturulmaktadır. Bu da iki modelin iç içe olduğunu belirtmektedir (Xie vd. 2013:99-108).

3.3. Poisson Regresyon Modelleri

3.3.1. Genelleştirilmiş ve Kısıtlanarak Genelleştirilmiş Poisson Regresyon Modelleri (GPR)

GPR modeli iki bölümden meydana gelmektedir. Birinci bölüm sayım sonucu meydana gelen sıfır değerlerini, ikinci bölüm ise sıfırdan büyük olan değerleri kapsar. GP dağılımına ait ortalama aşırı yayılımdan dolayı $\mu = \theta(1-\lambda)^{-1} = \theta\varphi$ şeklinde ifade edildiğinde x_i için GPR modeli sıfırdan büyük değerler edinildiğinde Eşitlik (3.14)' de ifade edilir. (Consul, Famoye 1992:89-109; Singh, Famoye 1993:917-923).

$$p(y_i > 0 / x_i) = \{(1-w)\frac{\mu(x_i)\left[\mu(x_i)+(\varphi-1)y_i\right]^{y_i-1}\varphi^{-y_i}\exp\{-\left[\mu(x_i)+(\varphi-1)y_i\right]/\varphi\}}{y_i!}, \quad (3.14)$$

$y_i = 0,1,...$

İkinci bölüm ise sıfır değerini aldığında Eşitlik (3.15)'de ifade edilir;

$$p(y_i = 0 / x_i) = \left[w + (1-w)e^{-\frac{\mu(x_i)}{\varphi}} \right] \quad (3.15)$$

$W = (1, w_1, w_2, ..., w_n)'$dan belirlenen aşırı yayılım için bağımsız değişkenlerdir.

Y_i'nin ortalaması ve varyansı Eşitlik (3.16) ile gösterilir,

$$E(Y_i / x_i) = \mu(x_i), Var(Y_i / x_i) = \varphi^2\mu(x_i) \quad (3.16)$$

Yukarıdaki gibi ifade edilen $\mu(x_i)$, log-doğrusal şeklinde gösterilmekte ve φ , yayılım parametresi olarak nitelendirilmektedir. Eşitlik (3.13)'deki GPR modeli

$\varphi = 1$ olduğu durumda Poisson Regresyon Modeline dönüşür. $\varphi > 1$ olduğunda aşırı yayılım ve $\mu_i > 2, \frac{1}{2} \leq \varphi < 1$ olduğunda da az yayılımdan bahsedilir.

KGP dağılımına bağlı ortalama $\mu = \theta(1-\alpha\theta)^{-1}$ olmak üzere verilen bir x_i için KGPR modeli aşağıda yer alan Eşitlik (3.17)'de gösterildiği gibidir (Famoye 1993:1335-1354).

$$p(y_i / x_i) = \begin{cases} \dfrac{(1+ay_i)^{y_i-1}\left(\dfrac{\mu(x_i)}{1+\alpha\mu(x_i)}\right)\exp\left(\dfrac{-\mu(x_i)[1+ay_i]}{1+\alpha\mu(x_i)}\right)}{y_i!}, y_i = 0,1,... \end{cases} \quad (3.17)$$

Y_i 'nin ortalaması ve varyansı Eşitlik (3.18) ile ifade edilir,

$$E(Y_i / x_i) = \mu(x_i), Var(Y_i / x_i) = \mu(x_i)\left[1+\alpha\mu(x_i)\right]^2 \quad (3.18)$$

$\alpha = 0$ olduğunda KGPR modeli Poisson Regresyon modeline dönüşür. $\alpha > 0$ olduğunda aşırı yayılım ve $\alpha < 0$ olduğunda da az yayılımdan bahsedilir.

3.3.2. Birleşik Poisson Regresyon Modelleri

Karışık Poisson Regresyon modellerini Negatif Binom ve Poisson Ters Gaussian Regresyon Modeli olmak üzere iki gruptan oluşturabilir. Karışık Poisson Regresyon (KPR) modellerinden çoğunlukla faydalanılan model Negatif Binom Regresyon (NBR) modelidir. Aşırı yayılım için uygulanan başka bir KPR modeli Poisson Ters Gaussian Regresyon (PTGR) modelidir (Özmen 1998:1-77).

KPR modelleri, verilen bir x_i açıklayıcı değişken vektörü ve v_i rassal etkisi denklem (3.19) ile gösterilir (Dean, Lawless 1989:467-472).

$$p(y_i / x_i) = \int_0^\infty \frac{e^{-v_i\mu(x_i)}\left[v_i\mu(x_i)\right]^{y_i}}{y_i!} g(v_i)\,dv_i, y_i = 0,1,... \quad (3.19)$$

Yukarıdaki denklemde gösterilen $g\left(v_i\right)$, v_i etkisine ait olasılık yoğunluk fonksiyonunu belirtmektedir. Benzer durumda $\mu\left(x_i\right)$ ise x_i ve β'ların fonksiyonu olup log-doğrusal şekilde gösterilmektedir. KPR modeline, rastgele etkili çarpımsal Poisson model olarak da ifade edilmektedir (Dean 1992:451-457).

KPR modellerinde Y_i 'nin dağılımı $v_i\mu_i$ ortalaması ile Poisson dağılımı olup v_i rastgele etkisinin de $E\left(v_i\right)=1$ ortalaması ve $Var\left(v_i\right)=\tau$ varyansıyla pozitif değerler alan bir dağılıma sahip olduğu tahmin edilmektedir. Böylece Y_i'nin marjinal dağılımına ait ortalaması ile varyansı Eşitlik (3.20) ile ifade edilir (Lawless 1987:209-225; Dean, Lawless 1989:467-472).

$$E\left(Y_i / x_i\right)=\mu\left(x_i\right), Var\left(Y_i / x_i\right)=\mu\left(x_i\right)\left[1+\tau\mu\left(x_i\right)\right] \qquad (3.20)$$

Poisson sayımlarına ait verilerin regresyon analizinde, aşırı yayılım görüldüğünde çoğunlukla uygulanan KPR modeli, Negatif Binom Regresyon modeli olarak karşımıza çıkmaktadır. NBR modelinde, v_i rastgele etkisi $E\left(v_i\right)=1$ ortalaması ve $Var\left(v_i\right)=\tau$ varyansı ile gamma dağılımına sahiptir. v_i 'ye ilişkin gamma dağılımının olasılık yoğunluk fonksiyonu Eşitlik (3.21)'de belirtildiği gibidir;

$$g\left(v_i\right)=\begin{cases} \dfrac{1}{\Gamma\left(\tau^{-1}\right)\tau^{\tau^{-1}}} v_i^{\tau^{-1}-1} e^{-\frac{v_i}{\tau}}, v_i>0 \\ 0, v_i \leq 0 \end{cases} \qquad (3.21)$$

Buna göre Eşitlik (3.20)' deki ortalama ve varyans ile NBR modeli aşağıda ifade edilen Eşitlik (3.22)' deki gibidir; (Lawless 1987:209-225; Xue, Deddens 1992:2215-2226)

$$p\left(y_i / x_i\right)=\begin{cases} \dfrac{\Gamma\left(y_i+\tau^{-1}\right)}{y_i!\Gamma\left(\tau^{-1}\right)}\left(\dfrac{\tau\mu\left(x_i\right)}{1+\tau\mu\left(x_i\right)}\right)^{y_i}\left(\dfrac{1}{1+\tau\mu\left(x_i\right)}\right)^{\tau^{-1}}, y_i=0,1,\ldots \\ 0, diğerdurumlarda \end{cases} \qquad (3.22)$$

Aşırı yayılım meydana geldiğinde yararlanılan başka bir KPR modeli ise Poisson Ters Gaussian Regresyon modelidir. PTGR modelinde, v_i rastgele etkisi $E(v_i)=1$ ortalaması ve $Var(v_i)=\tau$ varyansı ile Ters Gaussian modelinin diğer bir durumu olan Wald dağılımıdır. Böylece v_i 'nin olasılık yoğunluk fonksiyonu Eşitlik (3.23) ile gösterilmştir;

$$g(v_i)=\begin{cases}\left(2\pi\tau v_i^3\right)^{-\frac{1}{2}} e^{\frac{-(v_i-1)^2}{2\tau v_i}}, v_i>0 \\ 0, v_i \le 0\end{cases} \tag{3.23}$$

Denklem (3.14)'de gösterilen Karışık Poisson Regresyon formu kullanılarak elde edilen PTGR modeli Eşitlik (3.24) ile ifade edilmiştir;

$$p(y_i / x_i)=\int_0^\infty \frac{e^{-v_i\mu(x_i)}\left[v_i\mu(x_i)\right]^{y_i}}{y_i!}\left(2\pi\tau v_i^3\right)^{-\frac{1}{2}} e^{\frac{-(v_i-1)^2}{2\tau v_i}} dv_i, y_i=0,1,... \tag{3.24}$$

Karışık Poisson Regresyon dağılımında rastgele etki ve yayılım durumu göz önünde bulundurulduğunda Poisson Regresyon dağılımına oranla doğru tahminler meydana getirildiği belirlenmiştir (Özmen 1998:1-77).

3.3.3. Tekrarlı Verilerde Poisson Regresyon Modeli

Veri kütlesinde g grup ve her grupta $n_i\,(i=1,...,g)$ değişken meydana geldiğinde Y_{ij} Poisson yanıt değerlerine ait regresyon fonksiyonu Eşitlik (3.25)'de gösterildiği gibidir (Frome vd. 1973:935-940).

$$E(Y_{ij} / x_i)= f(x_i,\beta); i=1,...,n \tag{3.25}$$

Eşitlik (3.24)'de gösterilen n_i, x_i 'nin tekrarlamasını belirtir. $f(x_i,\beta)$ fonksiyonundaki β parametre tahminleri, En Çok Olabilirlik (EÇO) ve İteratif olarak Yeniden Ağırlıklandırılmış En Küçük Kareler (IEKK) metotları ile meydana gelmektedir. $Y_i=\sum_{j=1}^{n_i} Y_{ij}$ olmak üzere log-olabilirlik fonksiyonu Eşitlik (3.26) ile gösterilmiştir,

$$\ln L = \sum_{i=1}^{g} \left\{ Y_i \ln f\left(x_i, \beta\right) - n_i f\left(x_i, \beta\right) \right\} \tag{3.26}$$

β parametrelerinin EÇO tahminleri, log-olabilirlik fonksiyonunun parametreye uygun türevleri alınarak sıfır değeri elde edildikten sonra olabilirlik fonksiyonlarının çözümünden meydana gelmektedir. IEKK metoduyla β parametrelerinin tahminleri, $w_i = \dfrac{n_i}{f\left(x_i, \beta\right)}$ Poisson ağırlıkları ve $z_i = \dfrac{Y_i}{n_i}$ olduğu durumda, Eşitlik (3.27) ile ifade edilen ağırlıklı kareler toplamının en küçüklenmesi ile meydana gelmektedir.

$$S = \sum_{i=1}^{g} w_i \left\{ z_i - f\left(x_i, \beta\right) \right\}^2 \tag{3.27}$$

Açıklayıcı değişkenlere ait incelemelerde bir tekrarlama meydana geldiğinde belirlenen regresyon fonksiyonunun yararlılığı ve yayılım şekli ki-kare istatistiği ile ölçülebilmektedir. EÇO tahminlerini göstermek üzere ki-kare istatistiği Eşitlik (3.28) ile gösterilmektedir. (Frome vd. 1973:935-940; Consul, Famoye 1992:89-109).

$$Q_1 = \sum_{i=1}^{g} \sum_{j=1}^{n_i} \frac{\left(Y_{ij} - \hat{\mu}_i\right)^2}{\hat{\mu}_i} \tag{3.28}$$

Q_1 istatistiği, $D_1 = \sum_{i=1}^{g} n_i - p$ serbestlik dereceli ki-kare dağılımına sahiptir. Q_1 Eşitlik (3.29) ile belirtilen ki-kare rastlantı değişkeni bağımsız olarak iki ki-kare rastlantı değişkenine bölünebilmektedir ($Q_1 = Q_{11} + Q_{12}$)

$$\sum_{i=1}^{g} \sum_{j=1}^{n_i} \frac{\left(Y_{ij} - \hat{\mu}_i\right)^2}{\hat{\mu}_i} = \sum_{i=1}^{g} \sum_{j=1}^{n_i} \frac{\left(Y_{ij} - z_i\right)^2}{\hat{\mu}_i} + \sum_{i=1}^{g} n_i \frac{\left(z_i - \hat{\mu}_i\right)^2}{\hat{\mu}_i} \tag{3.29}$$

Eşitlikte gösterilen Q_{11} istatistiği $D_{11} = \sum_{i=1}^{g} (n_i - g)$ ve Q_{12} istatistiği $D_{12} = (g - p)$ serbestlik dereceli ki-kare dağılımını belirtmektedir. Eğer Q_1 istatistiği anlamlı ölçüde büyük çıkıyorsa ya varyansın aynı olmasından ya da regresyon modelinin uyum yetersizliğinden kuşku duyulur. Q_{11} değeri D_{11} serbestlik dereceli ki-kare tablo değeri ile kıyaslandığında büyük çıkıyorsa bu durum ya aşırı yayılım ya da az yayılımın belirtisidir (Özmen 1998:1-77).

Regresyon modelinin uyum yetersizliğini ölçmek amacıyla Eşitlik (3.30)'da ifade edilen F oranı kullanılmaktadır.

$$F = \frac{Q_{12} D_{11}}{Q_{11} D_{12}} \tag{3.30}$$

F oranı anlamlı ölçüde büyük çıkıyorsa bulunan regresyon modelinde, uyum yetersizliğinden bahsedilir. Eğer bulunan model Poisson Regresyon modeli ise ve bu model reddedilemiyor sadece varyansların aynı olmasından kuşkulanılıyorsa bu durumda kararlaştırılan kovaryans matrisi $\frac{Q_{11}}{D_{11}}$ faktörü ile çarpılarak bulunmaktadır. (Frome vd. 1973:935-940).

Karışık Poisson Regresyon modellerinde de $H_0 : \tau = 0$ yokluk hipotezini $H_1 : \tau >| 0$ alternatif hipotezine karşı ölçmek için Q_{11} istatistiğinden yararlanılmaktadır. Eğer n_i 'ler küçük $\mu_i, f(x_i, \beta)$ regresyon fonksiyonundan bulunabiliyorsa Q_{11} 'den faydalanılabilir. Eğer n_i 'ler büyük ve μ_i 'lerin meydana gelmesinden kuşku duyuluyor ise, $\bar{Y} = \frac{\sum_i n_i z_i}{\sum_i n_i}$ olmak üzere Eşitlik (3.31) ile ifade edilen test istatistiğinden yararlanılabilir (Collings, Margolin 1985:411-418).

$$Q_{11}^{*} = \sum_{i=1}^{g} \sum_{j=1}^{n_i} \frac{\left(Y_{ij} - z_i\right)^2}{\bar{Y}} \tag{3.31}$$

3.3.4. Poisson Regresyon Analizinde Katsayıların Tahmini

Poisson regresyon analizinde yanıt değişkeninin dağılışına uygun, tahmin edicilerini ölçme metotları farklılık bildirmektedir. En çok olabilirlik yöntemi (EÇO) ve genelleştirilmiş doğrusal modeller bu metotlardan çoğunlukla yararlanılan ve uygulananlardandır.

3.3.4.1. Poisson En Çok Olabilirlik Tahmin Yöntemi

Poisson modeli, gerçekte doğrusal olmayan bir regresyondur. Ancak maksimum olabilirlik yöntemleri ile parametre kestirimi oluşturmak olabildiğince basittir. Poisson regresyon modelinin standart tahmin edicisi Maksimum Olabilirlik tahmin edicisidir. Bağımsız incelemeler için, maksimum olabilirlik fonksiyonu Eşitlik (3.32) ile ifade edilir;

$$L(y,\beta)=\prod_{i=1}^{n} f(y_i)=\prod_{i=1}^{n}\frac{e^{-\mu_i}\mu_i^{y_i}}{(y_i)!}=\frac{\left(\prod_{i=1}^{n}\mu_i^{y_i}\right)e^{-\sum_{i=1}^{n}\mu_i}}{\prod_{i=1}^{n}(y_i)!} \tag{3.32}$$

Log-olabilirlik fonksiyonu ise Eşitlik (3.33) ile gösterilir;

$$\ell(\beta)=\sum_{i=1}^{n} y_i\ln(\mu_i)-\sum_{i=1}^{n}\mu_i-\sum_{i=1}^{n}\ln(y_i!) \tag{3.33}$$

Log-link yerine konulursa Eşitlik (3.34) elde edilir;

$$\ell(\beta)=\sum_{i=1}^{n} y_i x_i'\beta-\sum_{i=1}^{n} e^{x_i'\beta}-\sum_{i=1}^{n}\ln(y_i!) \tag{3.34}$$

Burada MLE bulunmak istendiğinden, β 'ya göre birinci dereceden kısmı türevi alındığı zaman Eşitlik (3.35) elde edilir (Memiş 2013:15).

$$\frac{\partial\ell(\beta)}{\partial\beta}=\sum_{i=1}^{n} y_i x_i-\sum_{i=1}^{n} x_i e^{x_i'\beta}=\sum_{i=1}^{n} x_i(y_i-\mu_i)=0 \tag{3.35}$$

Böylece elde edilmiş Poisson MLE $\hat{\beta}_p$ değeri Eşitlik (3.36) ile bulunur;

$$\sum_{i=1}^{n} x_i\left(y_i-\exp(x_i\beta)\right)=0 \tag{3.36}$$

$\hat{\beta}$ değeri birinci derecen türev alınarak elde edilir. $\hat{\beta}_p$ ‘yı elde etmek için analitik çözüm bulunmamaktadır. Çözüme ulaşmak için çoğunlukla iterasyon metodundan yararlanılmaktadır. Bu metoda, iteratif olarak yeniden ağırlıklandırılmış en küçük kareler (IRWLS) adı verilmektedir. Uygulamada en fazla 10 tane iterasyon uygulamak kafidir. Elde edilen bilgiler ile kullanılan modeller yönünde varyans değeri Eşitlik (3.37) ile gösterilir (Deniz 2005:59-72).

$$V_{ML}\left[\hat{\beta}_P\right] = \left(\sum_{i=1}^{n} \mu_i x_i x_i^{'}\right)^{-1} \tag{3.37}$$

3.3.4.2. Genelleştirilmiş Doğrusal Model (GLM) ile Tahmin Yöntemi

GLM metodunda verilerin orijinal dağılışını üslü biçimde ifade ederek, parametre tahminleri en çok olabilirlik veya yarı olabilirlik metotları ile oluşturulmaktadır. Birtakım durumlarda gözlem değerleri y_i normal dağılmayabilir. GLM, standart doğrusal modeller ile verilerin orijinal dağılımını temel alan en çok olabilirlik metodu yardımıyla parametreleri tahmin eder. (Yeşilova vd. 2006:87-92). GLM’ de gözlem değerlerinin olasılık yoğunluk fonksiyonu Eşitlik (3.38) ile gösterilmiştir;

$$f\left(y_i, \theta_i, \phi\right) = \exp\left\{\left[\frac{y_i\theta_i - b\left(\theta_i\right)}{\alpha\left(\phi\right)}\right] + c\left(y_i, \phi\right)\right\} \tag{3.38}$$

Bu Eşitlikte $c\left(y_i, \phi\right)$, normalleştirme katsayısıdır. Parametre tahminini elde etmek için olabilirlik fonksiyonu Eşitlik (3.39) ile ifade edilir;

$$L\left(y_i, \beta\right) = \prod_{i=1}^{n} \exp\left\{\left[\frac{y_i\theta_i - b\left(\theta_i\right)}{\alpha\left(\phi\right)}\right] + c\left(y_i, \phi\right)\right\} \tag{3.39}$$

Bu formülden yola çıkarak log olabilirlik fonksiyonunu Eşitlik (3.40) ile gösterebiliriz;

$$\ell\left(y_i, \beta\right) = \sum_{i=1}^{n} \left(\left[\frac{y_i\theta_i - b\left(\theta_i\right)}{\alpha\left(\phi\right)}\right] + c\left(y_i, \phi\right)\right) \tag{3.40}$$

Log-olabilirlik fonksiyonu β’ ya göre zincir kuralı uygulanarak kısmi türevi alınıp sıfıra eşitlenirse Eşitlik (3.40), (3.41), (3.42), (3.43), (3.44) ve (3.45) elde edilir,

$$\frac{\partial \ell}{\partial \beta} = \frac{\partial \ell}{\partial \theta_i}\frac{\partial \theta_i}{\partial \beta} = 0 \tag{3.41}$$

$$\frac{\partial \ell}{\partial \theta_i} = \frac{\partial}{\partial \theta_i}\left(\sum_{i=1}^{n}\left(\frac{y_i\theta_i - b(\theta_i)}{\alpha(\phi)} + h(y_i,\phi)\right)\right) \tag{3.42}$$

$$= \sum_{i=1}^{n}\left(\left(y_i - \frac{\partial b(\theta_i)}{\partial \theta_i}\right)\frac{1}{\alpha(\phi)} + \frac{\partial h(y_i,\phi)}{\partial \theta_i}\right) \tag{3.43}$$

$$= \frac{1}{\alpha(\phi)}\sum_{i=1}^{n}\left(y_i - \frac{\partial b(\theta_i)}{\partial \theta_i}\right) \tag{3.44}$$

$$\frac{\partial \theta_i}{\partial \beta} = \frac{\partial}{\partial \beta}\left(x_i^{'}\beta\right) = x_i \tag{3.45}$$

Eşitlik (3.41) ve (3.44) birlikte, (3.40)’da yerine yazılırsa aşağıdaki formül Eşitlik (3.46) ile gösterilir (Arıcan 2010:5);

$$\frac{\partial \ell}{\partial \beta} = \frac{\partial \ell}{\partial \theta_i}\frac{\partial \theta}{\partial \beta} = \frac{1}{\alpha(\phi)}\sum_{i=1}^{n}\left(y_i - \frac{\partial b(\theta_i)}{\partial \theta_i}\right)x_i = 0 \tag{3.46}$$

Genelleştirilmiş doğrusal modeller yardımıyla bulunan Poisson tahmin edicisi $\hat{\beta}_{GLM}$ Eşitlik (3.47) ile hesaplanır (Cameron, Trivedi 1998:411);

$$\sum_{i=1}^{n}\frac{1}{\phi}\left(y_i - \exp\left(x_i^{'}\beta\right)\right)x_i = 0 \tag{3.47}$$

3.3.5. Model Anlamlılığının Sınanması

Regresyon analizi neticesinde elde edilen modelin doğruluğu ispatlanmışsa, daha ileri bir analize ihtiyaç duymadan çalışma bitirilebilir. Fakat ortaya çıkan sonuçlar üzerinde özenli bir incelemede bulunmadan sonuç modelinden yararlanmamak gerekir. Bu inceleme süresinde çoğunlukla modelin yeterliliğinin belirlenmesi süreci olarak bilinir (Alpar 2003). Modelin yeterliliğinin belirlenmesi, parametre tahmincilerinin ölçülmesi ve modelin uyum iyiliğinin denenme aşamasından elde edilmektedir. Parametre tahmincilerinin ve model uyum iyiliğinin ölçülmesinde başka testlerde yer almaktadır. Bu testler sayesinde verilerin modele uygun olduğu ve modelde anlamlı parametreler olduğu söylenebilir.

3.3.5.1. Parametre Tahminlerinin Sınanması

Poisson regresyon analizinde, parametre tahminlerinin anlamlılıklarının değerlendirilmesinde, doğrusal modeller için yapılan hipotez testleri burada da geçerlidir. Yararlanılan sınamalar; Wald testi, t-testi ve benzerlik oran testi formunda olduğu söylenebilir. Parametre tahmincilerin ölçülmesinde yararlanılan varsayımları aşağıdaki gibi gösterilir (Memiş 2013:18):

$H_0 : \beta_i = 0, \left(i = 1, ..., n\right)$ (β_i parametresi anlamsızdır)

$H_0 : \beta_i \neq 0, \left(i = 1, ..., n\right)$ (β_i parametresi anlamlıdır)

Wald Test

Wald testi en az bir modelin ölçülmesinde yararlanılan bir testtir. Wald χ^2 istatistiği Eşitlik (3.48) ile oranlanır;

$$\chi_w^2 = \left(\frac{b_i}{s_{bi}^{'}}\right)^2 \tag{3.48}$$

Burada gösterilen b_i regresyon katsayısıdır. $s_{bi}^{'}$ standart hata terimidir ve $s_{bi}^{'} = s_{bi}^{'}\sqrt{\phi}$ eşitliğinden bulunur. ϕ ise, k tahmin edilecek parametre sayısı Eşitlik (3.49) ile ifade edilmektedir;

$$\phi = \frac{1}{n-k}\sum_{i=1}^{n}\frac{\left(y_i - \mu_i\right)^2}{\mu_i} \tag{3.49}$$

Bulunan Wald χ^2 istatistik değeri, 1 serbestlik dereceli χ^2 cetvel değeriyle kıyaslanır, elde edilen değer tablo değerinden büyükse H_0 hipotezi kabul edilemez (Deniz 2005:59-72).

3.3.5.2. Modelin Uyum İyiliğinin Sınanması

Doğrusal regresyon modellerinde regresyon doğrusunun verilere ne ölçüde uyumlu olduğunun ölçütü, bir veri kümesine uygulamak regresyon doğrusunun uyum iyiliği olarak ifade edilebilir (Gujaratti 1999). Parametre tahminleri gerçekleştirildikten sonra gözlemlerin modelin şekli etrafındaki dağılımlarının hesaplanması gerekmektedir. Çünkü gözlemler tahmin edilen modelin çizilen şekline ne kadar yakın olursa modelin uyum iyiliği de bir o kadar yüksek olacaktır. Bir başka ifade ile, Y' deki değişimim açıklayıcı değişkenlerdeki değişmelerle ifade edilmesi o kadar iyi olacaktır (Koutsoyiannis 1989:69-70).

Poisson regresyon modelinin uyum iyiliğinin ölçülmesinde; Pearson χ^2 , sapmalar istatistiği, yapay R^2 ölçümü, Akaike Bilgi Ölçütü (AIC) ve Bayes Bilgi Ölçütü (BIC)'nden çoğunlukla faydalanılır.

Pearson İstatistiği

μ_i ortalamalı ve ω_i varyanslı yanıt değişkeni y_i' ye bağlı rastgele seçilen model için standart uyum iyiliği ölçüm metodu Pearson istatistiğidir ve bu formül Eşitlik (3.50) ile gösterilir;

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(y_i - \hat{\mu}\right)^2}{\hat{\omega}_i} \qquad (3.50)$$

Bu değer seriden yayılımın aşırı olup olmadığını belirlemede yararlanılır. Burada $\hat{\mu}_i$ ve $\hat{\omega}_i$ değerleri μ_i ve ω_i 'nin tahmin değerlerini ifade eder. Hesaplanan χ^2 değeri, $\hat{\mu}$ için belirlenmiş serbestlik derecesi (n-k) ile karşılaştırılır. Bu formül Poisson regresyon için uygulandığında, $\omega_i = \mu_i$ olur ve Eşitlik (3.51) ile ifade edilir;

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(y_i - \hat{\mu}\right)^2}{\hat{\mu}_i} \qquad (3.51)$$

Hesaplanan χ^2 değerinin serbestlik derecesine oranının 1'den büyük değer alması verilerin modele uygun olmadığını ve aşırı yayılım durumunun varlığını göstermektedir (Deniz 2005:59-72).

Sapma İstatistiği

Uyum iyiliğinin ölçülmesinde yararlanılan bir diğer teknik de sapma istatistiğidir. Bu istatistik değerine aynı zamanda “G^2 istatistiği” adı da verilmektedir ve Eşitlik (3.52) ile ifade edilir;

$$G^2 = 2\sum_{i=1}^{n} y_i \ln\left(\frac{y_i}{\mu_i}\right) \tag{3.52}$$

Bu istatistik değeri 0’a yaklaşıyor ise model uyumu artıyor eğer bu istatistik değeri tam 0’a eşit ise model uyumunun mükemmel olduğu söylenebilir (Deniz 2005: 59-72).

Akaiki Bilgi Ölçütü

Akaike tarafından öne atılan ve farklı modellerin karşılaştırılmasında yaygın olarak kullanılan ölçüt, Akaike bilgi ölçütü (Akaike Information Criterion: AIC) olarak ifade edilir. Akaike bilgi ölçütü Eşitlik (3.53) ile gösterilir;

$$AIC = -2\log L + 2q \tag{3.53}$$

Bu Eşitlikte L; log olabilirlik fonksiyonunun maksimum değerini, q; açıklayıcı değişken sayısını ifade etmektedir. En küçük AIC değerine sahip model en iyi model olarak bilinmektedir (Ercanlı vd. 2012:75-84).

Bayes Bilgi Ölçütü

Bayes bilgi ölçütü Akaike bilgi ölçütü gibi veriler ve model arasında uygunluğu ölçmeye yarayan metotlardan biridir. Bayes bilgi ölçütü Eşitlik (3.54) ile ifade edilir;

$$BIC = -2\ln(L) + q\ln(N) \tag{3.54}$$

Eşitlikte L; log olabilirlik fonksiyonunun maksimum değerini, q; açıklayıcı değişken sayısını N; örnek büyüklüğünü ifade etmektedir (Ercanlı vd. 2012:75-84).

3.4. Kırpılmış Poisson Regresyon Analizi

Kırpılmış Poisson Regresyon Analizi, Poisson Regresyon Analizinde uç değerlerin analizlerde hesaplama veya yorum zorlukları çıkarması sebebiyle tanıtılan bir yöntemdir. Kırpılmış Poisson Regresyon analizi temelde üç ana grupta incelenebilir: Soldan Kırpılmış Poisson Regresyon analizi, Sağdan Kırpılmış Poisson Regresyon analizi ve çift yönlü kırpılmış Poisson Regresyon Analizi.

Sayımla elde edilen verilerde $k = c_L$ noktası için $y < c_L$ değerleri gözlenmemişse bu tarz verilere soldan kırpılmış veriler, $k = c_R$ noktası için değerleri gözlenmemişse bu tarz verilere ise sağdan kırpılmış veriler adı verilmektedir.

Soldan kırpılmış poisson regresyonu genellikle tıbbi çalışmalarda görülür. Mesela, hastaların hastalık başladıktan sonra hayatta kalma süresi üzerinde çalışmak isteyen birileri için uygun yöntem olarak karşımıza çıkmaktadır. Benzer şekilde aktüeryal araştırmalarda, astronomik, demografik, epidemiyolojik, güvenilirlik testlerinde ve diğer çalışmalarda da yaygındır.

Sağdan kırpılma genellikle, her bir bireyi ayırt etmedeki yetersizlik nedeniyle yüksek sayma zorluğu olduğunda veya yüksek değerli sayımlar için yardımcı malzemeler sorun çıkardığında ortaya çıkar. İki yönlü kırpılmış Poisson regresyon analizi ise, soldan ve sağdan kırpılmış Poisson verisinin birleşimi ile yapılan modellerdir.

Kırpılmış Poisson analiz yöntemleri içerisinde kullanılan tekniklerin biri sıfır kırpılmış poisson regresyon modelidir.

3.4.1. Sıfır Kırpılmış Poisson Regresyon Modeli

Sıfır Kırpılmış Poisson Dağılımı, y> 0 koşullu bir olasılık fonksiyonu Eşitlik (3.55) ile gösterilmiştir,

$$P\left(y_i / y_i > 0, \lambda\right) = \frac{P\left(y_i / \lambda\right)}{P\left(y_i > 0 / \lambda\right)} = \frac{\exp\left(-\lambda\right)\lambda^{y_i}}{y_i!\left(1-\exp\left(-\lambda\right)\right)}, y_i = 1, 2, ..., \tag{3.55}$$

$Y_1, ..., Y_{N_{obs}}$, λ_i, $i = 1, ..., N_{obs}$ parametresiyle Sıfır Kırpılmış Poisson Dağılımından rastgele bir örnek olsun. $\beta = \left(\alpha, \beta_1, ..., \beta_p\right)^T$ ve x_i 'nin i konusu için değişken değerlerin bir vektörü olduğu regresyon modelini düşünelim, yani $x_i = \left(1, x_{i1}, ..., x_{ip}\right)^T$ olmaktadır. Günlük olasılığı, Eşitlik (3.56) ile ifade edebiliriz. (Cameron, Trivedi 1998:59-139)

$$\log\left(\lambda_i\right) = \beta^T x_i \tag{3.56}$$

BÖLÜM 4

UYGULAMA

Uygulamada 3 farklı veri seti kullanılmıştır. 1. Veri seti 1493 gözlemden oluşmaktadır. Bağımlı değişken hastaların hastanede yatma süreleri (stay), bağımsız değişkenler yaş, sağlık sigortası ve ölüm parametreleridir (stats. idre. ucla. ed.). 2. Veri seti 120 gözlemden oluşmaktadır. Bağımlı değişken trafikte meydana gelen kaza sayılarını, bağımsız değişkenler ise kaza yapan kadın sürücüler, buzlanma, yağış, ehliyet gibi değişkenlerden oluşmaktadır (Tamar 2013:19). 3. Veri seti 535 gözlemden oluşmaktadır. Bağımlı değişken akciğer kanseri (durum) bağımsız değişken takip süresi, teşhis ve yaş parametreleridir (Zorlutuna vd. 2016:13-22).

Analizler STATA.15 paket programı ile yapılmıştır. Katsayı anlamlılıkları önem seviyesi 0,05 olarak alınmıştır.

1.Veri seti; 1493 hastanın hastanede kalış süresi, yaş, sağlık sigortası türü (var-yok) ve hastanın hastanede iken ölüp ölmediği değişkenlerine sahip bir veri setidir. Hastanede kalış süresi en az bir gün olarak kaydedilmiştir. Veriye ait değişkenlerin tanımlayıcı istatistikleri Tablo.4.1'de verilmiştir.

Tablo 4.1. 1.Veri Seti Tanımlayıcı İstatistikler

Variable	Obs	Mean	Std. Dev.	Min	Max
stay	1,493	9.728734	8.132908	1	74
age	1,493	5.233758	1.669273	1	9
hmo	1,493	.1600804	.3668034	0	1
died	1,493	.3429337	.4748486	0	1

Tablo 4.1'e bakıldığında, ortalama kalma süresi 9,72 gün, ortalama hasta yaşı 5,23 yıl olduğu görülmektedir.1.493 hastanın 239'u sağlık sigortasına sahip iken, 512'si de tedavi sırasında hayatını kaybetmiştir.

Tablo.4.2'de 1. veri seti için kırpılmış poisson regresyon analizi sonuçları verilmiştir.

Tablo 4.2. 1.Veri Seti İçin Kırpılmış Poisson Regresyonu Sonuçları

```
Iteration 0:   log likelihood = -6908.7992
Iteration 1:   log likelihood = -6908.7991

Truncated Poisson regression                    Number of obs     =      1,493
Limits:  lower =          0                     LR chi2(3)        =     181.13
         upper =       +inf                     Prob > chi2       =     0.0000
Log likelihood = -6908.7991                     Pseudo R2         =     0.0129
```

stay	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	-.014442	.0050347	-2.87	0.004	-.0243099	-.0045742
hmo	-.1359033	.0237419	-5.72	0.000	-.1824365	-.0893701
died	-.2037709	.0183728	-11.09	0.000	-.239781	-.1677608
_cons	2.435808	.0273324	89.12	0.000	2.382238	2.489379

Tablo 4.2’deki analiz sonuçlarına göre tüm katsayılar anlamlı bulunmuştur ($p<0,05$). Buna göre Kırpılmış Poisson Regresyonu modeli için katsayı yorumları şu şekilde olacaktır:

Yaş katsayısı -0,014442; hastanede günlük kalma sayısının yaştaki her yıl artışı için 0,14 azaldığını göstermektedir. HMO katsayısı -0,1359033; sağlık sigortası olan hastalar için hastanede kalma süresinin sağlık sigortası olmayan hastalara kıyasla 0,13 daha az olduğunu gösterir. Ölme katsayısı -0,2037709; hastanedeyken ölen hastaların kalış süreleri, ölmeyen hastalardan 0,203 gün daha azdır. Sabit katsayısı; tüm tahmincilerin sıfıra eşit olması durumunda ortalama günlük kalış sayısı 2,43’tür.

Tablo 4.3’de hastanede kalma süresi üst limit değeri 75 olmak üzere sağdan kırpılmış poisson regresyon analizi sonuçları verilmiştir.

Tablo 4.3. 1.Veri Setinde Y=75 İçin Sağdan Kırpılmış Poisson Regresyonu Sonuçları

```
Iteration 0:   log likelihood = -6908.9611
Iteration 1:   log likelihood = -6908.9611

Truncated Poisson regression                    Number of obs     =      1,493
Limits:  lower =       (none)                   LR chi2(3)        =     180.99
         upper =           75                   Prob > chi2       =     0.0000
Log likelihood = -6908.9611                     Pseudo R2         =     0.0129
```

stay	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	-.0144305	.0050326	-2.87	0.004	-.0242943	-.0045668
hmo	-.1357457	.023722	-5.72	0.000	-.1822399	-.0892516
died	-.2035726	.0183595	-11.09	0.000	-.2395567	-.1675886
_cons	2.435763	.0273224	89.15	0.000	2.382212	2.489314

Tablo 4.3'deki analiz sonuçlarına göre tüm katsayılar anlamlı bulunmuştur ($p<0,05$). Buna göre Kırpılmış Poisson Regresyonu modeli için katsayı yorumları şu şekilde olacaktır:

Yaş katsayısı; hastanede günlük kalma sayısının yaştaki her yıl artışı için 0,14 gün azaldığını göstermektedir. HMO katsayısı; sağlık sigortası olan hastalar için hastanede kalma süresinin sağlık sigortası olmayan hastalara kıyasla 0,13 gün daha az olduğunu gösterir. Ölme katsayısı; hastanedeyken ölen hastaların kalış süreleri, ölmeyen hastalardan 0,203 gün daha azdır. Sabit katsayısı; tüm tahmincilerin sıfıra eşit olması durumunda ortalama günlük kalış sayısı 2,43'tür.

Tablo 4.3'e göre sağdan kırpılmış poisson regresyon uyguladığımız zaman katsayılarda azalma meydana geldiği görülmektedir. Buna karşılık katsayı anlamlılıklarında bir farklılık olmamıştır.

Tablo 4.4. 1.Veri Setinde Y=80 İçin Sağdan Kırpılmış Poisson Regresyonu Sonuçları

```
Iteration 0:   log likelihood = -6908.9611
Iteration 1:   log likelihood = -6908.9611

Truncated Poisson regression                    Number of obs     =      1,493
Limits:  lower =       (none)                   LR chi2(3)        =     180.99
         upper =           80                   Prob > chi2       =     0.0000
Log likelihood = -6908.9611                     Pseudo R2         =     0.0129
```

stay	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	-.0144305	.0050326	-2.87	0.004	-.0242943	-.0045668
hmo	-.1357457	.023722	-5.72	0.000	-.1822399	-.0892516
died	-.2035726	.0183595	-11.09	0.000	-.2395567	-.1675886
_cons	2.435763	.0273224	89.15	0.000	2.382212	2.489314

Y=80 ile yapılan analizler, Y=75 ile verilen sonuçlar ile aynı çıktığı görülmüştür. Buna göre hastanede kalma üst limiti 75 ve daha fazlası olduğunda sonuç değişmeyecektir.

Birinci veri setine uygulanan Kırpılmış Poisson Regresyonu sonuçlarını karşılaştırmak için veriye Poisson ve Negatif Binom Regresyon yöntemleri de uygulanarak karşılaştırılmıştır.

Tablo 4.5'de aynı verilere Poisson Regresyon Analizi uygulanmıştır.

Tablo 4.5. 1.Veri Seti İçin Poisson Regresyonu Sonuçları

```
Iteration 0:   log likelihood = -6908.9611
Iteration 1:   log likelihood = -6908.9611

Poisson regression                              Number of obs     =      1,493
                                                LR chi2(3)        =     180.99
                                                Prob > chi2       =     0.0000
Log likelihood = -6908.9611                     Pseudo R2         =     0.0129
```

stay	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	-.0144305	.0050326	-2.87	0.004	-.0242943	-.0045668
hmo	-.1357457	.023722	-5.72	0.000	-.1822399	-.0892516
died	-.2035726	.0183595	-11.09	0.000	-.2395567	-.1675886
_cons	2.435763	.0273224	89.15	0.000	2.382212	2.489314

Tablo 4.5'de verilen Poisson Regresyonu Sonuçları, Y=75 ile sağdan kırpılmış poisson regresyonu ile elde edilen sonuçlarla aynı çıktığı görülmektedir. Tablo 4.6'da aynı verilere Negatif Binom Regresyonu uygulanmıştır.

Tablo 4.6. 1. Veri Seti İçin Negatif Binom Regresyonu Sonuçları

```
Iteration 0:   log likelihood = -4802.6959
Iteration 1:   log likelihood = -4802.4619
Iteration 2:   log likelihood = -4802.4619

Negative binomial regression                    Number of obs     =      1,493
                                                LR chi2(3)        =      33.65
Dispersion     = mean                           Prob > chi2       =     0.0000
Log likelihood = -4802.4619                     Pseudo R2         =     0.0035

------------------------------------------------------------------------------
        stay |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0147032   .0118136    -1.24   0.213    -.0378575    .0084512
         hmo |  -.1372682     .05335    -2.57   0.010    -.2418323   -.0327041
        died |  -.2037797    .041596    -4.90   0.000    -.2853063   -.1222531
       _cons |   2.437479   .0648092    37.61   0.000     2.310456    2.564503
-------------+----------------------------------------------------------------
    /lnalpha |   -.782414   .0441951                     -.8690348   -.6957933
-------------+----------------------------------------------------------------
       alpha |   .4573007   .0202104                      .4193561    .4986787
------------------------------------------------------------------------------
LR test of alpha=0: chibar2(01) = 4213.00              Prob >= chibar2 = 0.000
```

Tablo 4.6'daki analiz sonuçlarına göre Sağlık sigortası, ölüm ve sabit katsayıları anlamlı bulunmuştur ($p<0,05$). Yaş katsayısı anlamsız bulunmuştur ($p>0,05$). Buna göre Negatif Binom Regresyonu modeli için katsayı yorumları şu şekilde olacaktır:

Yaş katsayısı; hastanede günlük kalma sayısının yaştaki her yıl artışı için 0,01 azaldığını göstermektedir. HMO katsayısı; sağlık sigortası olan hastalar için hastanede kalma süresinin sağlık sigortası olmayan hastalara kıyasla 0,13 daha az olduğunu gösterir. Ölme katsayısı; hastanedeyken ölen hastaların kalış süreleri, ölmeyen hastalardan 0,203 daha azdır. Sabit katsayısı; tüm tahmincilerin sıfıra eşit olması durumunda ortalama günlük kalış sayısı 2,43'tür.

Yukarıda analizleri yapılan yöntemleri karşılaştırmak için her modelden elde edilen Log.Likelihood değerleri Tablo 4.7'de karşılaştırılmıştır.

Tablo 4.7. 1.Veri Seti İçin Yöntemlerin Karşılaştırılması

Yöntem	**Log Likelihood**
Poisson Regresyon	-6908,9611
Kırpılmış Poisson Regresyon	-6908,7991
Negatif Binom Regresyon	-4802,4619
Y=75 için Sağdan Kırpılmış Poisson R.	-6908,9611

Tablo 4.7'ye göre mutlak değerce minimum olan Negatif Binom Regresyon, Hastane verisine göre en iyi sonuçları vermiştir. En kötü sonuç ise Poisson Regresyon ve Y=75 için Kırpılmış Poisson Regresyon sonuçları olmuştur.

2.Veri seti; Kullanmış olduğumuz verilerde trafik kaza sayılarının belirleyicileri olarak, cinsiyet, hava şartları, sürücünün ehliyet kullanma yılı seçilmiştir. Bağımsız değişken olarak cinsiyet; kaza yapan kadın sürücü sayısı olarak alınmıştır. Buna ilaveten bağımsız değişken olarak karayolundaki buzlanma durumu, havanın yağış miktarı ve sürücünün ehliyetli olarak araba kullanma yılı alınmıştır. Tablo.4.8'de veriye ait değişkenlerin tanımlayıcı istatistikleri verilmiştir.

Tablo 4.8. 2.Veri Seti Tanımlayıcı İstatistikler

Variable	Obs	Mean	Std. Dev.	Min	Max
Kaza_Sayı	120	3.908333	2.881397	1	11
Kaza_Kadın	120	2.733333	1.912952	1	8
Buzlanma	120	1.666667	.4733811	1	2
Yagis	120	2.15	1.081937	1	4
Ehliyet	120	9.858333	7.944446	1	45

Tablo 4.8'e bakıldığında, ortalama kaza sayısı 3.90, ortalama kaza yapan kadın sayısı 2.73, yollarda meydana gelen ortalama buzlanma sayısı 1.66, ortalama yağış sayısı 2.15, ve ortalama ehliyetli araç kullanma sayısı 9.85 olarak görülmektedir.

Tablo 4.9'da 2. Veri seti için Kırpılmış Poisson Regresyon Analizi Sonuçları verilmiştir.

Tablo 4.9. 2.Veri Seti İçin Kırpılmış Poisson Regresyonu Sonuçları

```
Iteration 0:   log likelihood = -199.39415
Iteration 1:   log likelihood = -197.75055
Iteration 2:   log likelihood = -197.74791
Iteration 3:   log likelihood = -197.74791

Truncated Poisson regression                      Number of obs     =        120
Limits:  lower =          0                       LR chi2(4)        =     191.41
         upper =       +inf                       Prob > chi2       =     0.0000
Log likelihood = -197.74791                       Pseudo R2         =     0.3261
```

Kaza_Sayı	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Kaza_Kadın	.2541357	.0268678	9.46	0.000	.2014757	.3067956
Buzlanma	.3525811	.1549728	2.28	0.023	.04884	.6563222
Yagis	.0812222	.0490753	1.66	0.098	-.0149635	.177408
Ehliyet	.0010086	.0078509	0.13	0.898	-.0143789	.0163962
_cons	-.3638257	.2845958	-1.28	0.201	-.9216232	.1939718

Tablo 4.9’daki analiz sonuçlarına göre Kaza yapan kadın sürücüler ve buzlanma katsayıları anlamlı bulunmuştur ($p<0,05$). Yağış, ehliyet ve sabit katsayıları anlamsız bulunmuştur ($p>0,05$). Buna göre Kırpılmış Poisson Regresyonu modeli için katsayı yorumları şu şekilde olacaktır:

Kaza yapan kadın katsayısı; Trafikte kaza yapan kadın sayısı her kazada 0,25 oranında artmaktadır. Buzlanma katsayısı; yollarda meydana gelen buzlanma nedeni ile gerçekleşen kaza sayıları 0,35 oranla artmaktadır. Yağış katsayısı; 0,081 artan yağış miktarları ile ortaya çıkan kaza sayısı 0,08 oranında artmaktadır. Ehliyet katsayısı; ehliyet almaya hak kazanmış kadın sürücülerin trafikte ehliyetsiz araç kullanarak kaza yapan kadın sürücülere oranla 0,001 artmaktadır. Sabit katsayısı; tüm tahmincilerin sıfıra eşit olması durumunda ortalama kaza sayısı -0,36’dır.

Tablo 4.10’de üst limit değeri 12 olmak üzere sağdan kırpılmış poisson regresyon analizi sonuçları verilmiştir.

Tablo 4.10. 2. Veri Setinde Y=12 için sağdan Kırpılmış Poisson Regresyonu Sonuçları

```
Iteration 0:   log likelihood = -197.24139
Iteration 1:   log likelihood = -191.52714
Iteration 2:   log likelihood = -191.49099
Iteration 3:   log likelihood = -191.49098

Truncated Poisson regression                      Number of obs     =        120
Limits:  lower =      (none)                      LR chi2(4)        =     208.61
         upper =          12                      Prob > chi2       =     0.0000
Log likelihood = -191.49098                       Pseudo R2         =     0.3526

------------------------------------------------------------------------------
    Kaza_Sayı |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
   Kaza_Kadın |   .3388595   .0346714     9.77   0.000     .2709048    .4068141
     Buzlanma |   .2648092   .1353393     1.96   0.050     -.000451    .5300695
        Yagis |   .1308236   .0516932     2.53   0.011     .0295067    .2321404
      Ehliyet |  -.0031997   .0077323    -0.41   0.679    -.0183547    .0119553
        _cons |  -.3840296   .2618801    -1.47   0.143    -.8973052     .129246
------------------------------------------------------------------------------
```

Tablo 4.10’daki analiz sonuçlarına göre Kaza yapan kadın sürücüler, buzlanma ve yağış katsayıları anlamlı bulunmuştur ($p<0,05$). Ehliyet ve sabit katsayımız anlamsız bulunmuştur ($p>0,05$). Buna göre Kırpılmış Poisson Regresyonu modeli için katsayı yorumları şu şekilde olacaktır:

Kaza yapan kadın katsayısı; trafikte kaza yapan kadın sayısı kaza yapmayan kadın sayısı 0,33 oranında artmaktadır. Buzlanma katsayısı; yollarda meydana gelen buzlanma nedeni ile gerçekleşen kaza sayısı 0,26 oranında artmaktadır. Yağış katsayısı; artan yağış miktarları ile ortaya çıkan kaza sayısı 0,13 oranında artmaktadır. Ehliyet katsayısı; ehliyet almaya hak kazanmış kadın sürücülerin trafikte ehliyetsiz araç kullanarak kaza yapan kadın sürücülere oranı 0,003 oranında azalmaktadır. Sabit katsayısı; tüm tahmincilerin sıfıra eşit olması durumunda ortalama kaza sayısı -0,38'dir.

Tablo 4.11'de üst limit değeri 13 olmak üzere sağdan kırpılmış poisson regresyon analizi sonuçları verilmiştir.

Tablo 4.11. 2.Veri Setinde Y=13 için sağdan Kırpılmış Poisson Regresyonu Sonuçları

```
Truncated Poisson regression                      Number of obs     =        120
Limits:  lower =       (none)                     LR chi2(4)        =     196.72
         upper =           13                     Prob > chi2       =     0.0000
Log likelihood = -197.49715                       Pseudo R2         =     0.3325
```

Kaza_Sayı	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Kaza_Kadın	.3051441	.0320429	9.52	0.000	.2423411	.3679471
Buzlanma	.2587887	.134121	1.93	0.054	-.0040836	.521661
Yagis	.1130618	.0494358	2.29	0.022	.0161694	.2099543
Ehliyet	-.0022542	.0075217	-0.30	0.764	-.0169965	.0124882
_cons	-.2692183	.2538042	-1.06	0.289	-.7666654	.2282289

Tablo 4.11'deki analiz sonuçlarına göre Kaza yapan kadın sürücüler, buzlanma ve yağış katsayıları anlamlı bulunmuştur ($p<0,05$). Ehliyet ve sabit katsayımız anlamsız bulunmuştur ($p>0,05$). Buna göre Kırpılmış Poisson Regresyonu modeli için katsayı yorumları şu şekilde olacaktır:

Kaza yapan kadın katsayısı; trafikte kaza yapan kadın sayısı kaza yapmayan kadın sayısı 0,30 oranında artmaktadır. Buzlanma katsayısı; yollarda meydana gelen buzlanma nedeni ile gerçekleşen kaza sayısı 0,25 oranında artmaktadır. Yağış katsayısı; artan yağış miktarları ile ortaya çıkan kaza sayısı 0,11 oranında artmaktadır. Ehliyet katsayısı; ehliyet almaya hak kazanmış kadın sürücülerin trafikte ehliyetsiz araç kullanarak kaza yapan kadın sürücülere oranı 0,002 oranında azalmaktadır. Sabit katsayısı; tüm tahmincilerin sıfıra eşit olması durumunda ortalama kaza sayısı -0,26'dır.

Tablo 4.12'de Poisson Regresyonu Sonuçları verilmiştir.

Tablo 4.12. 2.Veri Seti İçin Poisson Regresyonu Sonuçları

```
Iteration 0:   log likelihood = -207.78803
Iteration 1:   log likelihood = -207.71767
Iteration 2:   log likelihood = -207.71766

Poisson regression                              Number of obs     =        120
                                                LR chi2(4)        =     176.34
                                                Prob > chi2       =     0.0000
Log likelihood = -207.71766                     Pseudo R2         =     0.2980
```

Kaza_Sayı	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Kaza_Kadın	.235541	.0249888	9.43	0.000	.1865638	.2845182
Buzlanma	.2696443	.1331659	2.02	0.043	.0086439	.5306447
Yagis	.0730315	.0461197	1.58	0.113	-.0173613	.1634244
Ehliyet	-.0002633	.0071511	-0.04	0.971	-.0142792	.0137525
_cons	-.0516263	.2412063	-0.21	0.831	-.524382	.4211294

Tablo 4.12'deki analiz sonuçlarına göre Kaza yapan kadın sürücüler ve buzlanma katsayıları anlamlı bulunmuştur (p<0,05). Yağış, ehliyet ve sabit katsayımız anlamsız bulunmuştur (p>0,05). Buna göre Poisson Regresyonu modeli için katsayı yorumları şu şekilde olmaktadır:

Kaza yapan kadın katsayısı; Trafikte kaza yapan kadın sayısı her kazada 0,23 oranında artmaktadır. Buzlanma katsayısı; yollarda meydana gelen buzlanma nedeni ile gerçekleşen kaza sayıları 0,26 artış göstermektedir. Yağış katsayısı; Artan yağış miktarları ile ortaya çıkan kaza sayısı 0,07 artmaktadır. Ehliyet katsayısı; Ehliyet almaya hak kazanmış kadın sürücülerin trafikte ehliyetsiz araç kullanarak kaza yapan kadın sürücülere kıyasla kaza yapma oranı 0,0002 daha azdır. Sabit katsayısı; tüm tahmincilerin sıfıra eşit olması durumunda ortalama kaza sayısı -0,05'dir.

Tablo 4.13'de Negatif Binom Regresyon Sonuçları verilmiştir.

Tablo 4.13. 2.Veri Seti İçin Negatif Binom Regresyonu Sonuçları

```
Negative binomial regression                    Number of obs     =        120
                                                LR chi2(4)        =     138.05
Dispersion     = mean                           Prob > chi2       =     0.0000
Log likelihood = -207.71766                     Pseudo R2         =     0.2494
```

Kaza_Sayı	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Kaza_Kadın	.235541	.0249889	9.43	0.000	.1865638	.2845183
Buzlanma	.2696443	.1331659	2.02	0.043	.0086439	.5306447
Yagis	.0730315	.0461197	1.58	0.113	-.0173614	.1634244
Ehliyet	-.0002633	.0071511	-0.04	0.971	-.0142792	.0137525
_cons	-.0516263	.2412063	-0.21	0.831	-.524382	.4211294
/lnalpha	-17.08	587.0404			-1167.658	1133.498
alpha	3.82e-08	.0000224			0	.

```
LR test of alpha=0: chibar2(01) = 4.1e-06              Prob >= chibar2 = 0.499
```

Tablo 4.13'deki analiz sonuçlarına göre Kaza yapan kadın sürücüler, buzlanma katsayıları anlamlı bulunmuştur ($p<0,05$). Ehliyet, yağış ve sabit katsayımız anlamsız bulunmuştur ($p>0,05$). Buna göre Negatif Binom Regresyonu modeli için katsayı yorumları şu şekilde olacaktır:

Kaza yapan kadın katsayısı; trafikte kaza yapan kadın sayısı kaza yapmayan kadın sayısı 0,23 oranında artmaktadır. Buzlanma katsayısı; yollarda meydana gelen buzlanma nedeni ile gerçekleşen kaza sayısı 0,26 oranında artmaktadır. Yağış katsayısı; artan yağış miktarları ile ortaya çıkan kaza sayısı 0,07 oranında artmaktadır. Ehliyet katsayısı; ehliyet almaya hak kazanmış kadın sürücülerin trafikte ehliyetsiz araç kullanarak kaza yapan kadın sürücülere oranı 0,002 oranında azalmaktadır. Sabit katsayısı; tüm tahmincilerin sıfıra eşit olması durumunda ortalama kaza sayısı-0,051'dir.

Yukarıda analizleri yapılan yöntemleri karşılaştırmak için her modelden elde edilen Log.Likelihood değerleri Tablo 4.14'de verilmiştir.

Tablo 4.14. 2.Veri Seti İçin Yöntemlerin Karşılaştırılması

Yöntem	**Log Likelihood**
Poisson Regresyon	-207,71766
Kırpılmış Poisson Regresyon	-197,74791
Negatif Binom Regresyon	-207,71766
Y=12 için Sağdan Kırpılmış Poisson R.	-191,49098
Y=13 için Sağdan Kırpılmış Poisson R.	-197,49715

Tablo 4.14'e göre mutlak değerce minimum olan Y=12 için Kırpılmış Poisson Regresyon, trafik kazası verisine göre en iyi sonuçları vermiştir. Kırpılmış Poisson Regresyon sonuçları; Poisson Regresyon ve Negatif Binom Regresyonuna göre daha iyi olduğu görülmektedir.

3.Veri seti; Akciğer kanseri teşhisi konmuş 535 hastaya aittir. Hastanın durumu (yaşıyor, vefat etmiş, haber alınamıyor), takip süreci (ay), teşhis konma süresi (yıl) ve hastanın yaşı değişkenleri kullanılmıştır. Tablo 4.15'te veriye ait tanımlayıcı istatistikler verilmiştir.

Tablo 4.15. 3.Veri Seti İçin Tanımlayıcı İstatistikler

Variable	Obs	Mean	Std. Dev.	Min	Max
durum	535	1.642991	.9249063	1	3
takip	535	14.97196	16.18112	0	134
teşhis	535	4.927103	1.674974	3	14
yaş	535	64.8243	10.32192	30	88

Tablo 5.15'e bakıldığında, ortalama durum sayısı 1.642, ortalama akciğer kanseri takip süresi; 14.97, ortalama akciğer kanseri teşhis süresi; 4.92 yıl ve ortalama yaş sayısı 64.82 olarak görülmektedir. En genç hasta 30 yaşında, en yaşlı hasta 88 yaşındadır.

Tablo 4.16'da veri seti için Kırpılmış Poisson Regresyon Analizi sonuçları verilmiştir.

Tablo 4.16. 3.Veri Seti İçin Kırpılmış Poisson Regresyonu Sonuçları

```
Iteration 0:   log likelihood = -534.17176
Iteration 1:   log likelihood = -409.13491
Iteration 2:   log likelihood = -401.67978
Iteration 3:   log likelihood = -401.44912
Iteration 4:   log likelihood = -401.44897
Iteration 5:   log likelihood = -401.44897

Truncated Poisson regression                      Number of obs     =        535
Limits:  lower =          0                       LR chi2(3)        =     464.56
         upper =       +inf                       Prob > chi2       =     0.0000
Log likelihood = -401.44897                       Pseudo R2         =     0.3665
```

durum	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
takip	.1077154	.0078407	13.74	0.000	.0923479	.1230828
teşhis	-1.288745	.0966472	-13.33	0.000	-1.47817	-1.09932
yaş	-.0113409	.0050292	-2.26	0.024	-.021198	-.0014838
_cons	4.5006	.4097263	10.98	0.000	3.697551	5.303649

Tablo 4.16'daki analiz sonuçlarına göre tüm katsayılar anlamlı bulunmuştur ($p<0,05$). Buna göre Kırpılmış Poisson Regresyonu modeli için katsayı yorumları şu şekilde olacaktır:

Takip katsayısı; akciğer kanseri takip süresindeki artış hastanın durumunu 0,10 olarak arttıracaktır. Teşhis katsayısı; akciğer kanseri erken teşhis geç kalınmış tedavilere oranla hastanın durumunu -1,28 azaltacaktır. Yaş katsayısı; akciğer kanserinin yaştaki her yıl artışı için hastanın durumunun-0,011 azaldığını görmekteyiz. Sabit katsayısı; tüm tahmincilerin sıfıra eşit olması durumunda ortalama hasta durumu 4,50'dir.

Tablo 4.17'de akciğer kanserinde üst limit değeri 4 olmak üzere sağdan Kırpılmış Poisson Regresyon Analizi sonuçları verilmiştir.

Tablo 4.17. 3.Veri Setinde Y=4 için sağdan Kırpılmış Poisson Regresyonu Sonuçları

```
Iteration 0:   log likelihood = -613.26147
Iteration 1:   log likelihood = -576.00906
Iteration 2:   log likelihood = -575.92032
Iteration 3:   log likelihood = -575.92031

Truncated Poisson regression                      Number of obs     =        535
Limits:  lower =     (none)                       LR chi2(3)        =     250.72
         upper =          4                       Prob > chi2       =     0.0000
Log likelihood = -575.92031                       Pseudo R2         =     0.1788
```

durum	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
takip	.0505754	.0045003	11.24	0.000	.041755	.0593957
teşhis	-.5035377	.041578	-12.11	0.000	-.5850292	-.4220463
yaş	-.0111734	.0047581	-2.35	0.019	-.0204992	-.0018477
_cons	3.213129	.3635542	8.84	0.000	2.500576	3.925682

Tablo 4.17'deki analiz sonuçlarına göre tüm katsayılar anlamlı bulunmuştur (p<0,05). Buna göre Kırpılmış Poisson Regresyonu modeli için katsayı yorumları şu şekilde olacaktır:

Takip katsayısı; akciğer kanseri takip süresindeki artışın, hastanın durumunda 0,05 oranında artış göstermektedir. Teşhis katsayısı; akciğer rahatsızlığında erken teşhis geç kalınmış tedavilere kıyasla hastanın durumunu 0,50 oranla daha azaltmaktadır. Yaş katsayısı; akciğer rahatsızlığının yaştaki her yıl artışı için 0,11 azaldığını görmekteyiz. Sabit katsayısı; tüm tahmincilerin sıfıra eşit olması durumunda ortalama hasta durumu 3,21'dir.

Tablo 4.18'de ise Akciğer kanser verisi için Poisson Regresyonu Sonuçları verilmiştir.

Tablo 4.18. 3.Veri Seti İçin Poisson Regresyonu Sonuçları

```
Iteration 0:   log likelihood = -684.54043
Iteration 1:   log likelihood = -684.53859
Iteration 2:   log likelihood = -684.53859

Poisson regression                                Number of obs   =        535
                                                  LR chi2(3)      =     128.35
                                                  Prob > chi2     =     0.0000
Log likelihood = -684.53859                       Pseudo R2       =     0.0857
```

durum	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
takip	.0279125	.002844	9.81	0.000	.0223383	.0334867
teşhis	-.3045164	.0310039	-9.82	0.000	-.3652829	-.2437499
yaş	-.0054494	.0033517	-1.63	0.104	-.0120186	.0011198
_cons	1.851292	.2441593	7.58	0.000	1.372748	2.329835

Tablo 4.18'deki analiz sonuçlarına göre takip süresi ve teşhis katsayıları anlamlı bulunmuştur (p<0,05). Yaş katsayısı ise anlamsız bulunmuştur (p>0,05). Buna göre Poisson Regresyonu modeli için katsayı yorumları şu şekilde olacaktır:

Takip katsayısı; akciğer kanseri takip süresindeki artış, hastanın durumunda 0,02 oranında artış göstermektedir. Teşhis katsayısı; akciğer kanserinde erken teşhis geç kalınmış tedavilere oranla hastanın durumu 0,30 daha azaltmaktadır. Yaş katsayısı; akciğer kanserinde yaştaki her yıl artışı için hastanın durumunun 0,005 azaldığını görmekteyiz. Sabit katsayısı; tüm tahmincilerin sıfıra eşit olması durumunda ortalama hasta durumu 1,85'dir.

Tablo 4.19'da Negatif Binom Regresyon sonuçları verilmiştir.

Tablo 4.19. 3.Veri Seti İçin Negatif Binom Regresyonu Sonuçları

```
Negative binomial regression                      Number of obs     =        535
                                                  LR chi2(3)        =     128.35
Dispersion     = mean                             Prob > chi2       =     0.0000
Log likelihood = -684.53859                       Pseudo R2         =     0.0857

------------------------------------------------------------------------------
       durum |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       takip |   .0279125    .002844     9.81   0.000     .0223383    .0334867
      teşhis |  -.3045164   .0310039    -9.82   0.000    -.3652829   -.2437499
         yaş |  -.0054494   .0033517    -1.63   0.104    -.0120186    .0011198
       _cons |   1.851292   .2441593     7.58   0.000     1.372748    2.329835
-------------+----------------------------------------------------------------
    /lnalpha |  -25.33541          .                             .           .
-------------+----------------------------------------------------------------
       alpha |   9.93e-12          .                             .           .
------------------------------------------------------------------------------
LR test of alpha=0: chibar2(01) = 0.00                 Prob >= chibar2 = 1.000
```

Tablo 4.19'daki analiz sonuçlarına göre takip süresi ve teşhis katsayıları anlamlı bulunmuştur (p<0,05). Yaş katsayısı anlamsız bulunmuştur (p>0,05). Buna göre Negatif Binom Regresyonu modeli için katsayı yorumları şu şekilde olacaktır:

Takip katsayısı; akciğer kanseri takip süresindeki artış, hastanın durumunda 0,02 oranında artış göstermektedir. Teşhis katsayısı; akciğer kanserinde erken teşhis geç kalınmış tedavilere oranla hastanın durumunu 0,30 daha azaltmaktadır. Yaş katsayısı; akciğer kanserinde yaştaki her yıl artışı için hastanın durumunu 0,005 azaltmaktadır. Sabit katsayısı; tüm tahmincilerin sıfıra eşit olması durumunda ortalama hasta durumu 1,85'dir.

Yukarıdaki yöntemleri Likelihood değerleri Tablo 4.20'de karşılaştırılmıştır;

Tablo 4.20. 3.Veri Seti İçin Yöntemlerin Karşılaştırılması

Yöntem	**Log Likelihood**
Poisson Regresyon	-684,53859
Kırpılmış Poisson Regresyon	-401,44897
Negatif Binom Regresyon	-684,53859
Y=4 için Sağdan Kırpılmış Poisson R.	-575,92031

Tablo 4.20'ye göre mutlak değerce minimum olan Kırpılmış Poisson Regresyon, akciğer kanseri verisine göre en iyi sonuçları vermiştir. En kötü sonuçlar ise Poisson Regresyon ve Negatif Binom Regresyon sonuçlarında gözlemlenmiştir.

SONUÇ

Regresyon analizi, istatistiksel tahmin yöntemlerinin en önemlisi ve belki de en çok tercih edilen yöntemidir. Burada amaç; eldeki verileri temsil eden ve anlamlı katsayılara sahip bir model kurmak, verinin gelecek değerlerini başarılı bir şekilde tahmin etmektir. Regresyon analizi, kullanım amacına, verinin yapısına ve uygulama türüne göre çeşitlilik göstermektedir. Poisson regresyon analizi, sayımla elde edilen bağımlı değişken ile bağımsız değişkenler arasındaki bağlantıyı çözümleyen bir analiz metodudur.

Poisson Regresyon modeli, ölümle sonuçlanan kaza analizlerinde, evli çiftlerin çeşitli sebeplerden dolayı boşanma oranları ile ilgili analizlerde ve akademik alanda erkek ve kadın akademisyenlerin unvan açısında yükselmeleri dahil olmak üzere insanların hayatlarına dair sosyal, ekonomik, sağlık ve eğitim gibi pek çok alanda kullanılmaktadır.

Poisson Regresyon Analizinin farklı veri yapılarında kullanılmak üzere farklı çeşitleri vardır. Bunlardan biri de aşırı değere sahip verilerde daha başarılı sonuçlar verdiği düşünülen Kırpılmış Poisson Regresyon Analizidir.

Bu çalışmada Kırpılmış Poisson Regresyon Analizi; Poisson Regresyon Analizi ve Negatif Binom Regresyon Analiz teknikleri karşılaştırılmıştır. Buna göre bu analiz teknikleri içerisinden Kırpılmış Poisson Regresyon yöntemiyle elde edilen sonuçların Poisson Regresyon ve Negatif Binom Regresyon yöntemiyle elde edilen sonuçların yanında başarılı sonuçlar verdiği görülmüştür.

Çeşitli parametre değerlerine göre de farklı yapılardaki Kırpılmış Poisson Regresyon Analizi sonuçları da yorumlanmıştır.

Bu tez çalışmasında 3 farklı veri yapısında Kırpılmış Poisson Regresyon uygulaması yapılmıştır.

Birinci veri için elde edilen sonuçlara göre Negatif Binom Regresyonu en iyi tahmin modeli olarak görülmektedir. Poisson Regresyonu ve Y=75 için Kırpılmış Poisson Regresyon aynı sonuca sahip olduğu hesaplanmıştır.

İkinci ve üçüncü veri için elde edilen sonuçlarda Kırpılmış Poisson Regresyon sonuçları, Negatif Binom ve Poisson Regresyonu sonuçlarına göre daha başarılı oldukları görülmektedir.

Aykırı değerlere sahip kategorik değişkenleri içeren model yapılarında, Kırpılmış Poisson Regresyon Analizi kullanımının başarılı sonuçlar vereceği bu çalışma ile de gösterilmiştir. İleriki çalışmalarda verilerin sağdan veya soldan sansürlü biçimleri ile Kırpılmış Regresyon Analizi sonuçlarının karşılaştırılması planlanmaktadır.

KAYNAKÇA

Alpar, Reha (2003). Uygulamalı Çok Değişkenli İstatiksel Yöntemlere Giriş 1, Nobel Yayın Dağıtım, s.259, Ankara.

Alpar, Reha (2010). Basit Doğrusal Regresyon Çözümlemesi. Spor, Sağlık ve Eğitim Bilimlerinden Örneklerle Uygulamalı İstatistik ve Geçerlik-Güvenirlik. Detay Yayıncılık, s.285-304, Ankara.

Arıcan, Engin (2010). Nitel Yanıt Değişkene Sahip Regresyon Modellerinde Tahmin Yöntemleri, Yayınlanmamış Yüksek Lisans Tezi, Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, s.5, Adana.

Bates Douglas M. ve Watts Donald G. (1988). "Nonlinear Regression Analysis and Its Applications", Wiley, s.32-167, New York.

Bierens Herman J. ve Gallant Ronald (1997). "Nonlinear Models" (2 cilt), The international library of critical writings in econometrics, Edward Elgar Publishing Limited, Cheltenham, UK.

Boucher Jean Philippe, Denuit Michel ve Guillen Montserrat (2007). Risk Classification for Claim Counts: Mixed Poisson, Zero-Inflated Mixed Poisson and Hurdle Models. North American Actuarial Journal, 11(4), 110-131.

Cameron A. Colin ve Trivedi Pravin (1986). "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators," Journal of Applied Econometrics, 1, 29–53.

Cameron A. Colin ve Trivedi Pravin (1998). Regression Analysis of Count Data, Cambridge University Press, West Nyack, NY, USA.

Collings Bruce J. ve Margolin Barry H. (1985). Testing Goodness-of-fit fort he Poisson Assumption when Observations are not Identically Distributed, Journal of the American Statistical Association, 80, 411-418.

Consul Perm C. ve Famoye Felix (1992). Generalized Poisson Regression Model, Comm. Statist. – Theor. and Meth., 21(1), 89-109.

Cox, Warburton Robert (1983). Some Remarks on Overdispersion, Biometrika, 70, s.269-274.

Çokluk, Ömay. (2010). Lojistik Regresyon Analizi: Kavram ve Uygulama. Kavram ve Uygulamada Eğitim Bilimleri/ Educational Scienses: Theory & Practice 10 (3) Yaz / Summer, s: 1357-1407.

Dawson Beth ve Trapp Robert G. (2001). Statistical Methods for Multiple Variables. Basic & Clinical Biostatistics. Lange Medical books/McGraw Hill Medical Publishing Division, s.236-242, USA.

Dean, Charmaine B. (1992). Testing for Overdispersion in Poisson and Binomial Regression Model, Journal of the American Statistical Association, 87, 451-457.

Dean Charmaine B. ve Lawless Jerald F. (1989). Tests for Detecting Overdispersion in Poisson Regression Model, Journal of the American Statistical Association, 84, 467-472.

Deniz, Özlem (2005). Poisson Regresyon Analizi, İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 4(7): 59-72.

Denison David D., Hansen Mark H., Holmes Christopher C., Mallick Bani ve Yu Bin (2003). “Nonlinear Estimation and Classification”, Springer-Verlag New York İnc. s.1-5.

Denuit Michel, Marechal Xavier, Pitrebois Sandra ve Walhin Jean François (2007). Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems (First Edition), England: John Wiley & Sons, s.28-29, England.

Dinarcan Gözde Nur (2018). Sayma Verisi için Regresyon Modelleri ve Bir Uygulama. Yayınlanmamış Yüksek Lisans Tezi, İstatistik Ana Bilim Dalı, Fen Bilimleri Enstitüsü, Hacettepe Üniversitesi, s.9, Ankara.

Ercanlı İlker, Kahriman Aydın ve Yavuz Hakkı (2012). Trabzon Orman Bölge Müdürlüğü Doğu Ladini-Sarıçam Karışık Meşcereleri İçin Karışık Etkili Doğrusal Olmayan Regresyon Denklemleri ile Doğu Ladini Çap-Boy Modellerinin Geliştirilmesi, SDÜ Orman Fakültesi Dergisi, 13: 75-84.

Erdinç Suna, Yeşilova Abdullah ve Ser Gazel (2017). Van Gölü Sahil Şeridindeki Zooplankton Populasyon Yoğunluğu Değişiminin Doğrusal Olmayan Regresyon Yöntemleri Kullanılarak İncelenmesi. Yüzüncü Yıl Üniversitesi Tarım Bilimleri Dergisi, c.27, s.58-64

Erilli Necati Alp (2015). İstatistik 2, Seçkin Yayıncılık, s.175, Ankara.

Erilli Necati Alp ve Alakuş Kamil (2016). Parameter Estimation In Theil-Sen Regression Analysis With Jackknife Method. Eurasian Econometrics, Statistics & Emprical Economics Journal, s.28-41.

Eubank Randall L. (1990). Nonparametric Regression And Spline Smoothing, Second Edition, s.4.

Famoye, Felix (1993). Resticted Generalized Poisson Regression Model, Comm. Statist.- Theor. and Meth., 22(5), 1335-1354.

Galton, Francis (1886).''Family Likeness in Stature,'' Proceedings of Royal Society, Londra, c.40, s.42-72

Frome, Edward L. (1983). The Analysis of Rates using Poisson Regression Models, Biometrics, 39, 665-674.

Frome Edward L., Kutner Michael H. ve Beauchamp John J. (1973) Regression Analysis of Poisson Distributed Data, Journal of American Statistical Association, 68, 935- 940

Genç, Aşır (1997). Çok Değişkenli Lineer Olmayan Modeller: Parametre Tahmini ve Hipotez Testi, Ankara Üniversitesi, Yayınlanmamış Doktora Tezi, s.1-103.

Ghitany Mohamed E., Karlis Dimitris, Al-Mutairi Dhaif K., Al-Awadhi Fahimah A. (2012). An EM Algorithm for Multivariate Mixed Poisson Regression Models and its Application. Applied Mathematical Sciences, Cilt:6, no:137, s.6843-6856.

Gibbons, Jean D. (1999). Nonparametric Statistics: An Introduction, Sage Publications, s.3.

Gujarati, Damodar N. (1999). Temel Ekonometri, Çev. Ümit Şenesen ve Gülay Günlük Şenesen, Literatür Yayıncılık, İstanbul.

Gürsakal, Necmi (1997). Bilgisayar Uygulamalı 1, Bursa: Marmara Kitabevi, s. 238.

Hardle Wolfgang, Müller Marlene, Sperlich Stefan ve Werwatz Axel. (2004). Nonparametric and Semiparametric Models: An Introduction, Springer Series in Statistics, s.115, USA.

Hardle, Wolfgang (1994). Applied Nonparametric Regression. Cambridge University, s.6-11, UK.

Heinzl Harald ve Mittlböck Martina (2003). Pseudo R-squared Measures for Poisson Regression Models with Over-or Underdispersion. Computation Statistic & Tata Analysis ,44, 253-271.

Horowitz, Joel L. (1993). Semiparametric Estimation of a Work-Trip Mode Choice Model, Journal of Econometrics, 58, 49-70.

Ismail Noriszura, Zamani Hossein (2013). Estimation of Claim Count Data Using Negative Binomial Generalized Poisson, Zero-Inflated Negative Binomia land Zero-Inflated Generalized Poisson Regression Models, Casualty Actuarial Society E-Forum, Spring, s.1-18.

Karadavut Ufuk, Genç Aşır, Tozluca Abdurrahman, Kınacı İsmail, Aksoyak Şeref, Palta Çetin, Pekgör Ahmet (2005). Nohut (Cicer Arietinum L.) Bitkisinde Verime Etki Eden Bazı Karakterlerin Alternatif Regresyon Yöntemleriyle Karşılaştırılması. Tarım Bilimleri Dergisi 11, (3), 328-333.

Kirkwood Betty R. ve Sterne Jonathan A.C. (2003). Regression Modelling. Medical Statistics. Blackwell Science, s.315-342, Australia.

Koutsoyiannis, Anna (1989). Ekonometri Kuramı, (çev. Ümit Şenesen ve Gülay Günlük Şenesen), Verso yayıncılık, s.69-70, Ankara.

Köleoğlu, Nilay (2006). Olay Zamanı Analizinde Tesadüfi Etkiler Poisson Regresyon Modeli ile Gözlemlenemeyen Heterojenliğin İncelenmesi, Marmara Üniversitesi, Sosyal Bilimler Enstitüsü, Yayınlanmamış Doktora Tezi.

Kılıç, Dilek (2014). Sigara Tüketiminin Panel Veri Analizi: İngiltere İçin Bir Uygulama. Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 16 (3), 128-142.

Lambert, Diane (1992). Zero-Inflated Poisson Regression with an Application to Defects in Manufacturing, Technometrics,34(1), 1-13.

Lawless, Jerald F. (1987). Negative Binomial and Mixed Poisson Regression, The Canadian Journal of Statistics, 15(3), 209-225.

Maddala, Gangadharrao Soundalyarao (2001). Introduction to Econometrics, Third Edition. John Wiley and Sons, s:60, USA.

McCullagh Peter ve Nelder John A. (1989). Generalized Linear Models, Chapman and Hall, London, 511, UK.

Memiş Müslüme ve Önder Hasan (2012). Poisson Regresyon Analizi, Ondokuz Mayıs Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, Samsun.

Memiş, Müslüme (2013). Poisson Regresyon Tahmin Yöntemlerinin Karşılaştırılması. On Dokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Zootekni Anabilim Dalı, s.9-18, Samsun.

Miaou, Shaw Pin (1994). The Relationship Between Truck Accidents and Geometric Design of Road Sections: Poisson Versus Negative Binomial Regressions, Accident Analysis and Prevention, 26(4), 471-482.

Nelder John Ashworth ve Wedderburn Robert William Maclagan (1972). Generalized Linear Models, Journal of the Royal Statistical Society Series A, 135, 3, 370-384.

Newbold, Paul (1995). Statistics for Business and Economics, Fourth Edition, Prentice-Hall, s.567.

Orhunbilge, A. Neyran (2000). Uygulamalı Regresyon ve Korelasyon Analizi. 2. Baskı, Nobel Akademik Yayıncılık, s.12, İstanbul.

Özmen, İlknur (1998). Poisson Regresyon Çözümleme Teknikleri, Hacettepe Üniversitesi, Yayınlanmamış Doktora Tezi, s.1-77.

Pearson Karl ve Lee Alice (1903). “On the Laws of Inheritance”, Biometrika, c. 2, s. 357-462. s.60.

Saffari Seyed Ehsan ve Adnan Robiah (2010). Zero-Inflated Poisson Regression Models With Right Censored Count Data. Matematika, c: 27, no:1, s. 21-29.

SAS, (2007). SAS/Stat, Software, Hangen and Enhanced, SAS, Institute, Incorporation, USA.

Seber George Arthur Frederick ve Wild, C.J. (1989) “Nonlinear Regression”, John Wiley and Sons, s.151-160, New York, USA.

Selim Sibel ve Üçdoğruk Şenay (2003). Sayma Veri Modelleri ile Çocuk Sayısı Belirleyicileri: Türkiye’deki Seçilmiş İller İçin Sosyoekonomik Analizler. D.E.Ü.İ.İ.B.F. Dergisi, Cilt: 18, sayı: 2, s:13-31.

Singh, Karan P. ve Famoye Felix (1993). Analysis of Rates Using a Generalized Poisson Regression Model, Biom., J., 8, 917-923.

Şahin, Hasan (2002). Poisson Regresyon Uygulaması: Türkiye'deki Grevlerin Belirleyicileri 1964-1998, Doğuş Üniversitesi Dergisi, (5), 173-180.

Takezawa, Kunio (2005). Introduction to Nonparametric Regression. Wiley-Interscience, Canada.

Tamar, Mehmet (2013). Poisson Regresyonu, Fırat Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, s.19.

Tüzel, Sema (2011). Hasar Sıklıkları İçin Sıfır Yığılmalı Kesikli Modeller, Yayınlanmamış Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, s.35, Ankara.

Tüzel, Sema (2012). Sucu, M., 2012. Hasar Sıklıkları için Yığılmalı Kesikli Modeller, İstatikçiler Dergisi 5, 23-31.

Ullah, Aman (1988). “Non-Parametric Estimation of Econometric Functionals”, The Canadian Journal of Economics, Vo. 21, No.3, s. 625-658

Wang Weiren ve Famoye Felix (1997). Modeling Household Fertility Decisions With Generalized Poisson Regression. J Popul Econ, 10:273-283.

Wooldridge Jeffrey Marc (2013). Ekonometriye Giriş. (çev. Ebru Çağlayan). Nobel Yayınları, Ankara.

Xie Haiyi, Tao Jill, McHugo Gregory J. ve Drake Robert E. (2013). Comparing statistical methods for analyzing skewed longitudinal count data with many zeros: An example of smoking cessation. Journal of Substance Abuse Treatment, 45 (1), 99–108.

Xue Dixi ve Deddens James A. (1992). Overdispered Negative Binomial Regression Models, Comm. Statist.-Theor and Meth., 21(8), 2215-2226.

Yatchew, Adonis (1998). “Nonparametric Regression Techniques in Economics”, Journal of Economic Literature, 36, 2, s. 672.

Yeşilova, Abdullah (2009). Sıfır Değer Ağırlıklı Sayıma Dayalı Verilerin Analizinde Hurdle Modelin Kullanılması, Anadolu Üniversitesi Bilim ve Teknoloji Dergisi, Cilt:10, Sayı:2, 467-475.

Yeşilova Abdullah ve Atlıhan Remzi (2007). Farklı Sıcaklıkların Scymnus Subvillosus’un Bıraktığı Yumurta Sayıları Üzerine Etkilerinin Karışımlı Poisson Regresyon ile Analiz Edilmesi. Yüzüncü Yıl Üniversitesi, Ziraat Fakültesi, Tarım Bilimleri Dergisi (J. Agric. Sci), 17(2):73-79.

Yeşilova Abdullah, Yılmaz Ayhan ve Kaki Barış (2006). Norduz Erkek Kuzularının Bazı Kesikli Üreme Davranış Özelliklerinin Analizinde Doğrusal Olmayan Regresyon Modellerin Kullanılması, Yüzüncü Yıl Üniversitesi, Ziraat Fakültesi, Tarım Bilimleri Dergisi (J. Agric. Sci), 16(2):87-92

Yeşilyurt, Hande (2005). Poisson Regresyon Modeli ve Türkiye’deki Boşanma İstatistiklerine Uygulanması, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi.

Yip Karen C.H. ve Yau Kelvin K.W. (2005). On Modeling Claim Frequency Data In General Insurance With Extra Zeros, Insurance: Mathematics and Economics, 36 (2), 153- 163.

Zorlutuna Şebnem, Erilli Necati Alp, Yücel Birsen (2016). Tobit Regresyon Analizi ile Akciğer Kanseri Çalışması: Sivas İli Örneği. Eurasian Econometrics, Statistics and Emprical Economics Journal, s.13-22.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı	**:** Seçil KARTAL
Uyruğu	**:** T.C.
Doğum Tarihi ve Yeri	**:** 26/03/1992 SİVAS
e-posta	**:** krtlscl@gmail.com

EĞİTİM

Derece	**Kurum**	**Mezuniyet Yılı**
Lisans	Sivas Cumhuriyet Üniversitesi	2012-2016
Yüksek Lisans	Sivas Cumhuriyet Üniversitesi	2016- …

İŞ TECRÜBESİ

Tarih	**Kurum**	**Görev**

YABANCI DİL BİLGİSİ

KPDS ( ) **ÜDS ( )** **TOEFL ( )** **EILTS ( )**