T.C.
AYDIN ADNAN MENDERES ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI
2022-YL-179

SAYMA VERİLERİNİN MODELLENMESİ VE BİR UYGULAMA

HAZIRLAYAN
Afet SÖZEN ÖZDEN

TEZ DANIŞMANI
Dr. Öğr. Üyesi Elvan HAYAT

AYDIN-2022

T.C.

AYDIN ADNAN MENDERES ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ

AYDIN

Bu tezde sunulan tüm bilgi ve sonuçların, bilimsel yöntemlerle yürütülen gerçek deney ve gözlemler çerçevesinde tarafımdan elde edildiğini, çalışmada bana ait olmayan tüm veri, düşünce, sonuç ve bilgilere bilimsel etik kuralların gereği olarak eksiksiz şekilde uygun atıf yaptığımı ve kaynak göstererek belirttiğimi beyan ederim.

Afet SÖZEN ÖZDEN

20/12/2021

ÖZET

SAYMA VERİLERİNİN MODELLENMESİ VE BİR UYGULAMA

Afet SÖZEN ÖZDEN

Yüksek Lisans Tezi, Ekonometri Anabilim Dalı

Tez Danışmanı: Dr. Öğr. Üyesi Elvan HAYAT

2021, XIV + 65 sayfa

Ekonometrik ve istatistiksel analizlerde elde edilen verilerin yapısına uygun regresyon modelinin belirlenmesi en önemli aşamalardan biridir. Bağımlı değişkenin sayıma dayalı elde edildiği durumlarda güvenilir tahminler yapabilmek için Sayma Verisi Regresyon Modellerini kullanmak gerekmektedir. Sayıma dayalı veriler kesikli bir yapıda olduğu için sayma veri regresyon modelleri kesikli dağılımlardan yararlanılarak geliştirilmektedir.

Bu tez çalışmasında Türkiye İstatistik Kurumu (TÜİK) tarafından 2019 yılı Gelir ve Yaşam Koşulları Araştırması (GYKA) mikro veri seti kullanılarak Türkiye'de bireylerin işsiz kaldıkları ay sayısını etkileyen faktörlerin sayma veri modelleri ile incelenmesi amaçlanmıştır. Çalışmada bağımlı değişken bireylerin işsiz kaldıkları ay sayısı olarak belirlenmiş bir sayma verisidir. İşsiz kalan bireylerin işsiz kaldıkları sürenin uzaması hem bireysel hem de toplumsal ve ekonomik yönden olumsuz durumların ortaya çıkmasına neden olmaktadır. Bireylerin işsiz kaldıkları süreyi etkileyen birçok faktör bulunmaktadır ve bu faktörlerin belirlenmesi bireylerin işsiz kaldıkları sürenin kısaltılabilmesi açısından önem arz etmektedir. Türkiye'de bireylerin işsiz kaldıkları ay sayısını sayma veri regresyon modelleri kullanarak ampirik olarak inceleyen ilk çalışma olmasının yanında işsiz kalınan süreyi etkileyen faktörleri modellemesi açısından da çalışmanın literatüre katkı sağlayacağı düşünülmektedir.

Çalışmanın analizinde kullanılan sayma verisi regresyon modelleri Poisson, Negatif Binom, Sıfır Değer Ağırlıklı Negatif Binom ve Genelleştirilmiş Poisson regresyon modelleridir. Her bir sayma verisi modeli için değişken seçiminde tüm olası alt küme yönetimi kullanılarak en uygun değişkenlerin medeni durum, eğitim durumu, genel sağlık durumu ve kronik hastalık durumu değişkenleri olduğu belirlenmiş ve bu değişkenler kullanılarak sayma verisi regresyon modelleri uygulanmıştır. Belirlenen bu model için en uygun tahmin yönteminin Sıfır Değer Ağırlıklı Negatif Binom Regresyon Modeli olduğu sonucuna ulaşılmıştır.

ANAHTAR SÖZCÜKLER: Poisson Regresyon Modeli, Sayma Verisi, Sıfır Değer Ağırlıklı Negatif Binom Regresyon Modeli.

ABSTRACT

COUNT DATA MODELLING AND AN APPLICATION

Afet SÖZEN ÖZDEN

MSc Thesis at Econometrics

Supervisor: Assist. Prof. Elvan HAYAT

2021, XIV+ 65 pages

Determining the regression model suitable for the structure of the data obtained in econometric and statistical analyzes is one of the most important steps. In cases where the dependent variable is obtained based on counting, it is necessary to use Count Data Regression Models in order to make reliable estimates. Since count-based data has a discrete structure, counting data regression models are developed using discrete distributions.

In this thesis, it is aimed to examine the factors affecting the number of months that individuals are unemployed in Turkey with count data models, using the 2019 Income and Living Conditions Survey (ILCS) micro data set by the Turkish Statistical Institute (TURKSTAT). The dependent variable in the study is a count data determined as the number of months that individuals are unemployed. The prolongation of the unemployment period of the unemployed leads to the emergence of negative situations in terms of both individual, social and economic aspects. There are many factors that affect the time that individuals are unemployed, and the determination of these factors is important in terms of shortening the time that individuals are unemployed.

Count data regression models used in the analysis of the study are Poisson, Negative Binomial, Zero Inflated Negative Binomial and Generalized Poisson regression models. In the selection of variables for each count data model, using all possible subset regression approach, it was determined that the most appropriate variables were marital status, education status, general health status and chronic disease status, and count data regression models were applied using these variables. It has been concluded that the most appropriate estimation method for this determined model is the Zero Inflated Negative Binomial Regression Model.

KEY WORDS: Count Data, Poisson Regression Model, Zero Inflated Negative Binomial Regression Model.

TEŞEKKÜR

Tez çalışmamın her aşamasında bilgi ve deneyimiyle bana daima yol gösteren, desteğini hep hissettiğim ve beni daima çalışmaya teşvik eden değerli danışman hocam Dr. Öğr. Üyesi Elvan HAYAT' a sonsuz teşekkürlerimi sunarım.

Eğitim hayatım boyunca maddi ve manevi desteklerini esirgemeyen başta sevgili annem Yasemin YEŞİL' e, sevgili babam Ahmet SÖZEN' e, sevgili kardeşim Güliz SÖZEN' e ve şu an hayatta olmasalar da varlıklarını hep hissettiğim sevgili dedem Mustafa Fikret SÖZEN' e, sevgili babaannem Afet SÖZEN' e ve sevgili anneannem Şenay YEŞİL' e çok teşekkür ederim. Her daim destekleri ile bana güç veren sevgili ikinci annem Nesrin ÖZDEN, sevgili ikinci babam Hasan ÖZDEN, sevgili ablam Ebru AŞIK ve sevgili abim Koray AŞIK' a teşekkürü bir borç bilirim. Evdeki çalışmalarım boyunca beni asla yalnız bırakmayan yeğenlerim sevgili Tuğçe Naz AŞIK' a, sevgili Elif Naz AŞIK' a ve güzel kızım can dostum Linda' ya teşekkür ederim.

Son olarak; sadece bu süreçte değil yedi yıldır her anımda yanımda olan, birlikte yürüdüğümüz bu hayat yolunda sevgisini ve desteğini en derinden hissettiğim, tez çalışmam boyunca da göstermiş olduğu sabır ve özveri için eşim, yol arkadaşım Serkan ÖZDEN' e çok teşekkür ederim.

Afet SÖZEN ÖZDEN

İÇİNDEKİLER

KABUL ONAY SAYFASI iii

BİLİMSEL ETİK BİLDİRİM SAYFASI iv

ÖZET v

ABSTRACT vi

TEŞEKKÜR vii

SİMGELER DİZİNİ .... **xi**

ŞEKİLLER DİZİNİ xii

TABLOLAR DİZİNİ xiii

KISALTMALAR DİZİNİ xiv

GİRİŞ 1

1. BÖLÜM 3

1. LİTERATÜR ÖZETİ 3

1.1. Türkiye Verileri Kullanılarak Yapılan Önceki Çalışmalar 3

1.2. Yabancı Kaynaklı Veriler ile Yapılan Önceki Çalışmalar 11

2. BÖLÜM 14

2. GENELLEŞTİRİLMİŞ DOĞRUSAL MODELLER 14

2.1. Genelleştirilmiş Doğrusal Modellerin Bileşenleri 14

2.2. Genelleştirilmiş Doğrusal Modeller İçin Üstel Dağılım Ailesi 15

2.2.1. Üstel Dağılım Ailesinde Poisson, Binom ve Normal Dağılım 17

2.3. Genelleştirilmiş Doğrusal Modellerde Katsayı Tahmini 18

2.4. Genelleştirilmiş Doğrusal Modellerde Uyum İyiliği 19

2.4.1. Sapma İstatistiği 19

2.4.2. Pearson χ^2 İstatistiği 19

2.4.3. Akaike Bilgi Kriteri 20

2.5. Veri Türüne Göre Genelleştirilmiş Doğrusal Modeller 20

3. BÖLÜM ... 22
3. SAYMA VERİSİ REGRESYON MODELLERİ ... 22
3.1. Poisson Regresyon Modeli ... 23
3.1.1. Poisson Dağılımı ... 23
3.1.2. Poisson Regresyon Modeli ... 23
3.2. Negatif Binom Regresyon Modeli ... 25
3.3. Sıfır Değer Ağırlıklı Regresyon Modelleri ... 28
3.3.1. Sıfır Değer Ağırlıklı Poisson Regresyon Modeli ... 29
3.3.2. Sıfır Değer Ağırlıklı Negatif Binom Regresyon Modeli ... 31
3.4. Genelleştirilmiş Poisson Regresyon Modeli ... 33
3.5. Engel (Hurdle) Regresyon Modeli ... 34
3.6. Modelin Değerlendirilmesi ve Seçimi ... 35
3.6.1. Katsayı Tahminlerinin Değerlendirilmesi ... 36
3.6.1.1. Wald $\chi 2$ testi ... 36
3.6.2. Modelin Uyum İyiliğinin Değerlendirilmesi ... 36
3.6.2.1. Pseudo (Yapay) R^2 istatistiği ... 37
3.6.2.2. Pearson χ^2 istatistiği ... 37
3.6.2.3. Sapma istatistiği ... 38
3.6.3. Model Seçimi ... 38
3.6.3.1. Akaike Bilgi Kriteri (AIC) ... 38
3.6.3.2. Bayes Bilgi Kriteri (BIC) ... 39
3.6.3.3. Mallow'un C_p Kriteri ... 39
4. BÖLÜM ... 40
4. UYGULAMA ... 40
4.1. Veri seti ve Tanımlayıcı İstatistikler ... 40
4.2. İşsiz Kalınan Ay Sayısı Verileri İçin Modeller ve Tahmin Sonuçları ... 44

5. SONUÇ VE ÖNERİLER 57

6. KAYNAKLAR 59

ÖZGEÇMİŞ 64

SİMGELER DİZİNİ

N	:	Gözlem Sayısı
Y	:	Bağımlı Değişken Vektörü
X	:	Bağımsız Değişken Vektörü
k	:	Bağımsız Değişken Sayısı
η	:	Doğrusal Kestirim Parametresi
E(.)	:	Beklenen Değer
g(.)	:	Link (Bağ) Fonksiyonu
μ	:	Ortalama
θ	:	Doğal (Kanonik) Parametre
f(y)	:	Olasılık Yoğunluk Fonksiyonu
β	:	Bilinmeyen Parametre Vektörü
V(.)	:	Varyans
$\Gamma(.)$	:	Gamma Fonksiyonu
ω	:	Yapısal sıfırların olma olasılığı
B	:	Bağımlı Değişkendeki Poisson Sonucunun Ortak Değişken Matrisi
G	:	Bağımlı Değişkendeki Fazla Sıfırların Ortak Değişken Matrisi
χ_w^2	:	Wald Ki-Kare Test İstatistiği
SE	:	Standart Hata
R_p^2	:	Pseudo (Yapay) R^2
χ_p^2	:	Pearson Ki-Kare Test İstatistiği
D	:	Sapma İstatistiği
C_p	:	Mallow'un C_p İstatistiği
L(.)	:	Olabilirlik Fonksiyonu
lnL(.)	:	Log-Olabilirlik Fonksiyonu
Φ	:	GDM için Yayılım Parametresi
α	:	Yayılım Parametresi
υ	:	Negatif Binom için Yayılım Parametresinin Tersi

ŞEKİLLER DİZİNİ

Şekil 4.1. İşsiz kalınan ay sayısı değişkenine ait dağılım grafiği 41

Şekil 4.2. 2018 yılında 1 ay ve üzeri işsiz kalınan ay sayısı için dağılım grafiği.................. 44

TABLOLAR DİZİNİ

Tablo 2.1. Üstel dağılım ailesindeki bazı dağılımların özellikleri 18

Tablo 2.2. Üstel dağılım ailesinin bazı üyeleri için sapma istatistikleri............................. 19

Tablo 4.1. Analizde kullanılan bağımsız değişkenler ve düzeyleri 40

Tablo 4.2. Bağımsız değişkenlere ait frekans tablosu 42

Tablo 4.3. İşsiz kalınan ay ve yaş değişkenine ait tanımlayıcı istatistikler 43

Tablo 4.4. İşsiz kalınan ay sayısı frekans tablosu.. 43

Tablo 4.5. Tüm olası altküme yöntemine göre Poisson regresyon sonuçları 45

Tablo 4.6. Tüm olası alküme yöntemine göre NBRM sonuçları.. 46

Tablo 4.7. NBRM tahmin sonuçları 48

Tablo 4.8. Tüm olası altküme yöntemine göre Sıfır Değer Ağırlıklı Negatif Binom regresyon sonuçları 49

Tablo 4.9. Sıfır Değer Ağırlıklı Negatif Binom regresyon modeli tahmin sonuçları.......... 51

Tablo 4.10.Tüm olası altküme yöntemine göre Genelleştirilmiş Poisson regresyon sonuçları... 52

Tablo 4.11. Genelleştirilmiş Poisson regresyon modeli tahmin sonuçları 53

Tablo 4.12.Model seçim kriterlerine ait sonuçlar 54

KISALTMALAR DİZİNİ

AIC	:	Akaike Bilgi Kriteri
AIDS	:	Edinilmiş Bağışıklık Yetmezliği Sendromu
AKT_k	:	k Parametreli Modelin Artık Kareler Toplamı
Ar-Ge	:	Araştırma Geliştirme
Bağ-Kur	:	Esnaf ve Sanatkârlar ve Diğer Bağımsız Çalışanlar Sosyal Sigortalar Kurumu
BIC	:	Bayes Bilgi Kriteri
COM-Poison	:	Conway-Maxwell-Poisson
DİE	:	Devlet İstatistik Enstitüsü
GDM	:	Genelleştirilmiş Doğrusal Model
GSOEP	:	Alman Sosyoekonomik Paneli
KDRM	:	Klasik Doğrusal Regresyon Modeli
MLE	:	En Çok Olabilirlik Tahmini
NBRM 1	:	Negatif Binom Regresyon Modeli 1
NBRM 2	:	Negatif Binom Regresyon Modeli 2
ÖSYM	:	Öğrenci Seçme ve Yerleştirme Merkezi
PSID	:	Gelir Dinamiği Paneli Çalışması
SGK	:	Sosyal Güvenlik Kurumu
SSK	:	Sosyal Sigortalar Kurumu
TÜİK	:	Türkiye İstatistik Kurumu
YGS	:	Yükseköğretime Geçiş Sınavı
ZINB	:	Sıfır Değer Ağırlıklı Negatif Binom
ZINB 1	:	Sıfır Değer Ağırlıklı Negatif Binom Modeli 1
ZINB 2	:	Sıfır Değer Ağırlıklı Negatif Binom Modeli 2
ZIP	:	Sıfır Değer Ağırlıklı Poisson

GİRİŞ

Ekonometrik araştırmalarda yaygın olarak kullanılan yöntemler arasında yer alan regresyon analizi, bağımlı değişken ile bağımsız değişken ya da değişkenler arasındaki ortalama ilişkinin matematiksel bir fonksiyonla incelenmesidir. Literatürde bağımlı değişken; açıklanan, etkilenen, tepki, öngörülen olarak da ifade edilmektedir. Bağımsız değişken ise; açıklayıcı, etkileyici, kontrol ve öngören olarak ifade edilmektedir. Regresyon analizinin amacı; bağımsız değişkenlerin değerleri ile bağımlı değişkenin ortalama değerini tahmin ederek bağımsız değişken ya da değişkenlerin bağımlı değişken üzerinde bir etkisi olup olmadığını incelemek ve bağımlı değişkenin gelecekte alacağı değerle ilgili öngörüde bulunmaktır (Tarı, 2018: 15).

Bağımlı değişkenin türüne ve dağılımına göre kullanılması gereken regresyon modelleri de farklılık göstermektedir. Bağımlı değişken Y, sürekli olduğu durumda klasik en küçük kareler (EKK) yöntemi ile kategorik olduğunda genelleştirilmiş ağırlıklı en küçük kareler yöntemi ile iteratif çözüm yapan lojistik regresyon analizi ile modellenmektedir. Bağımlı değişken Y, kesikli yapıda olduğu durumda ise sayma verisi regresyon modelleri ile model tahmin edilmektedir.

Sayıma dayalı olarak elde edilen veri "sayma verisi" olarak ifade edilmektedir. Sayma verisi regresyon modelleri, Genelleştirilmiş Doğrusal Modellere dayanmaktadır. Bu kapsamda çalışmada, Genelleştirilmiş Doğrusal Modellerden olan Poisson ve Negatif Binom Regresyon modelleri ele alınmıştır. Bu modellere dayanan Sıfır Değer Ağırlıklı Poisson Regresyon, Sıfır Değer Ağırlıklı Negatif Binom Regresyon ve Genelleştirilmiş Poisson Regresyon modelleri incelenmiştir.

Bu tez çalışmasının birinci bölümünde literatür araştırmasına yer verilmiştir. Literatür araştırması iki kısma ayrılarak incelenmiştir. İlk kısım Türkiye verileri kullanılarak yapılan önceki çalışmaları, ikinci kısım ise yabancı kaynaklardan elde edilen veriler kullanılarak yapılan önceki çalışmaları kapsamaktadır.

Çalışmanın ikinci bölümünde Genelleştirilmiş Doğrusal Modeller hakkında bilgi verilmiştir. Bu bölümde, Genelleştirilmiş Doğrusal Modellerin bileşenleri, parametre tahmini, üstel dağılım ailesi, üstel dağılım ailesindeki bazı dağılımların özellikleri ve veri türüne göre Genelleştirilmiş Doğrusal Modeller konularına değinilmiştir.

Üçüncü bölümde Genelleştirilmiş Doğrusal Modeller içerisinde yer alan sayma verisi regresyon modellerinin tanımı, özellikleri ve çeşitlerine detaylı olarak yer verilmektedir. Bu modeller Poisson, Negatif Binom, Sıfır Değer Ağırlıklı Poisson, Sıfır Değer Ağırlıklı Negatif Binom, Genelleştirilmiş Poisson ve Engel (Hurdle) regresyon modelleri olarak ele alınmıştır. Her bir model için ortalama, varyans, olasılık yoğunluk fonksiyonu, parametre tahmini ve bu tahminde kullanılan yöntemle ilgili bilgiler ayrı ayrı incelenmiştir. Sayma verisi regresyon modellerinin tahmininden sonra modelin genel olarak değerlendirmesini yapabilmek için kullanılması gereken hipotez testlerine yer verilmiştir. Bu bağlamda katsayı tahminlerinin anlamlılığı için kullanılması gereken testlere ve modelin uyum iyiliğinin ölçülmesi için kullanılması gereken istatistiklere bu bölümde değinilmiştir. Ayrıca tahmin edilen birden fazla sayma verisi regresyon modeli için hangi modelin veri setini en iyi açıklayan model olduğuna karar verirken kullanılması gereken model seçim kriterlerine de yine bu bölüm içerisinde yer verilmiştir.

Çalışmanın dördüncü bölümünde ise uygulama kısmına yer verilmiştir. Uygulamada oluşturulan modelde kullanılan değişkenler için tanımlayıcı istatistikler verilmiştir. Veri setine uygun olan model belirlenerek, belirlenen modele ait tahmin sonuçları da bu bölüm içerisinde verilmiştir.

Çalışmanın beşinci bölümünde ise çalışmayla ilgili genel bir değerlendirme yapılarak ulaşılan sonuçlar ortaya konmuştur.

1. BÖLÜM

1. LİTERATÜR ÖZETİ

1.1. Türkiye Verileri Kullanılarak Yapılan Önceki Çalışmalar

Selim ve Üçdoğruk (2003), ‘Sayma Veri Modelleri ile Çocuk Sayısı Belirleyicileri: Türkiye’deki Seçilmiş İller İçin Sosyoekonomik Analizler’ adlı çalışmalarında Poisson Quasi Maksimum Olabilirlik tahmin yöntemini kullanmışlardır. Çalışmada bağımlı değişkeni canlı doğan çocuk sayısı olarak belirlemişlerdir. Bağımsız değişkenleri yaş, eğitim, meslek, SGK kayıtlılık, çalışma durumu, işteki durum, yıllık maaş-ücret geliri olarak kadınlar ve erkekler için ayrı ayrı belirlemişlerdir. Bunlar dışında iller, hanenin elde ettiği toplam gelir ve kadınlar için ölüm oranı bağımsız değişkenleri de analizde kullanılmıştır. Çalışmada kullanılan veriler Devlet İstatistik Enstitüsü (DİE) tarafından 1994 yılında yapılan Hanehalkı Gelir Dağılımı Araştırması’ndan elde edilmiştir. Analiz sonuçlarına göre; ailenin çocuk sahibi olma kararını almasında kadının eğitim düzeyinin oldukça önemli bir etkisinin olduğu tespit edilmiştir. Ayrıca kadının eğitim durumu, çalışma durumu ve ücreti ile erkeğin çalışma durumu, eğitimi ve ücreti gibi değişkenlerin doğum kararını etkileyen önemli faktörler arasında olduğunu tespit etmişlerdir.

Yeşilyurt (2005), ‘Poisson Regresyon Modeli ve Türkiye’deki Boşanma İstatistiklerine Uygulanması’ adlı yüksek lisans tezinde 1989-1998 yılları arasında Türkiye’deki boşanma sayısı verileriyle ilgili olarak bir Poisson regresyon modeli tahmin etmiştir. Analizde bağımlı değişken Türkiye’deki boşanma sayıları, bağımsız değişkenler ise yaş grubu ve boşanmanın gerçekleştiği yıl olarak belirlemiştir. Çalışmada yaş grubu verisi 10 gruptan oluşmaktadır. Analizde 3 ayrı model tahmin edilmiştir. İlk model bütün yaş gruplarının yer aldığı veri seti ile tahmin edilmiştir İkinci modeli, 15-19 ve 20-24 yaş gruplarının dahil edilmediği veri seti ile tahmin etmiştir. Üçüncü model ise, 15-19, 20-24, 60 ve üzeri yaş gruplarının dahil edilmediği veri seti ile tahmin edilmiştir. Analiz sonucuna göre; yaş grubu değişkeninin boşanma sayısı üzerinde önemli derecede etkisi olduğu sonucuna ulaşmıştır.

Özmen ve Famoye (2007), ‘Count Regression Models with an Application to Zoological Data Containing Structural Zeros’ adlı çalışmalarında Dalyan Plajı’nda 1991-1993 yılları arasında güneşe maruz kalarak ölen caretta caretta yavrularının sayısı için bazı model karşılaştırmaları yapmışlardır. Uygun modelin saptanması için yapılan analizlerde bağımlı değişkeni güneşe maruz kalarak ölen caretta caretta yavrularının sayısı olarak belirlemişlerdir. Bağımsız değişkenleri ise; alan, denizden uzaklık ve yıl olarak belirlemişlerdir. Analiz sonuçlarına göre Genelleştirilmiş Poisson Regresyon modelinin veriye diğer sayma verisi regresyon modellerine göre daha iyi uyum sağladığını tespit etmişlerdir.

Kaya ve Yeşilova (2012), ‘E-Posta Trafiğinin Sıfır Değer Ağırlıklı Regresyon Yöntemleri Kullanılarak İncelenmesi’ adlı çalışmalarında Yüzüncü Yıl Üniversitesi personelinin e-posta trafiğini incelemeyi amaçlamıştır. Analizde bağımlı değişken gönderilen e-posta sayısı olarak belirlenmiştir. Bağımsız değişkenleri ise; personel tipi, birim kodu, unvan kodu, cinsiyet, medeni hal, yaş, sahip olunan e-posta adresi sayısı, e-posta YYU uzantılı kullanımı, sohbet programı kullanımı, bilimsel toplantı (YYU adresi ile), ortalama günlük internet kullanım saati, website sahipliği, internet bağlantısı durumu ve dizüstü bilgisayar sahipliği olarak belirlemişlerdir. Yüzüncü Yıl Üniversitesi 2009 Bahar eğitim öğretim döneminde e-posta sunucusundan ve anket yoluyla elde ettikleri verileri Sıfır Değer Ağırlıklı Regresyon modelleri ile analiz etmişlerdir. Analiz sonuçlarına göre; gönderilen e-posta sayısı üzerinde E-Posta YYU, Sohbet Programı, Bilimsel Toplantı, Ortalama İnternet Kullanım Saati, Web Site değişkenleri istatistiksel olarak anlamlı bulunmuştur. E-Posta gönderme olasılığı için yapılan analiz sonucuna göre ise; Birim Kodu, Unvan Kodu, Personel Tipi, E-Posta Adet, E-Posta YYU ve Bilimsel Toplantı değişkenlerinin önemli bir etkiye sahip oldukları görülmüştür.

Tamar (2013), ‘Poisson Regresyonu’ adlı yüksek lisans tezinde trafik kazalarına etki eden faktörleri belirlemeyi amaçlamıştır. Çalışmanın veri setini Aksaray Emniyet Müdürlüğü Trafik Şube Müdürlüğü’nden elde etmiştir. Çalışmada Aksaray ilinde meydana gelen trafik kazaları sayısını bağımlı değişken olarak belirlemiştir. Bağımsız değişkenleri; sürücünün cinsiyeti, karayolundaki buzlanma durumu, yağış miktarı ve ehliyet kullanım yılı olarak belirlemiştir. Analizi 2011, 2012 ve 2013 yılları Aralık, Ocak, Şubat ayları için Poisson regresyonu ile yapmıştır. Analiz sonuçlarına göre; kaza sayısına etki eden iki önemli faktörün sürücünün cinsiyeti ve karayolundaki buzlanma durumu olduğunu tespit etmiştir. Tüm bağımsız değişkenlerin kullanıldığı modelin Akaike Bilgi Kriteri (AIC) değeri 399 olarak hesaplamıştır. Kadın sürücü sayısı ve buzlanma durumu değişkenleri ile oluşturulan modelin AIC değerinin 400; kadın sürücü sayısı, buzlanma durumu ve yağış miktarı değişkenleri ile oluşturulan modelin AIC değerinin 398 olarak hesaplamıştır. Bu 3 model arasında kadın sürücü sayısı, buzlanma durumu ve yağış miktarı değişkenlerinin dahil edildiği modelin en uygun model olduğu sonucuna ulaşmıştır.

Açıkyürek (2016), 'Poisson Regresyon ve Bir Uygulama' adlı yüksek lisans tezinde Türkiye'deki malulen emekli birey sayısını etkileyen faktörleri belirlemeyi amaçlamıştır. Sosyal Güvenlik Kurumu'ndan (SGK) alınan veri setini Poisson, Negatif Binom, Sıfır Değer Ağırlıklı Poisson, Sıfır Değer Ağırlıklı Negatif Binom ve Hurdle (Engel) regresyon modelleri ile tahmin etmiştir. Analizde kullanılan bağımlı değişkenleri; SSK malulen emekli sayısı, Bağ-Kur malulen emekli sayısı ve Emekli Sandığı malulen emekli sayısı olarak belirlemiştir. Bağımsız değişkenleri ise; yıl, yaş, cinsiyet, hizmet, aylık ve bölge olarak belirlemiştir. Analizde SSK için 2000-2014 dönemi, Bağ-Kur için 2006-2014 dönemi ve Emekli Sandığı için 2000-2014 dönemi verilerini kullanmıştır. Analizde bağımlı değişkenler için 3 model tahmin etmiştir. Analiz sonuca göre; Poisson regresyonun hiçbir modele uyum sağlamadığını tespit etmiştir. SSK malulen emekli sayısı için AIC değerine göre Sıfır Değer Ağırlıklı Negatif Binom Regresyon modelini tercih etmiştir. Bağ-Kur malulen emekli sayısı için AIC değerine göre Hurdle (Engel) Negatif Binom Regresyon modelini tercih etmiştir. Emekli Sandığı malulen emekli sayısı için AIC değerine göre Sıfır Değer Ağırlıklı Negatif Binom Regresyon modelini tercih etmiştir. Bu modellerin analiz sonuçlarına göre; malulen emekli olma yaşı için en fazla riskin SSK ve Bağ-Kur'da 40-49 yaş aralığı iken Emekli Sandığı'nda 30-39 yaş aralığında olduğunu tespit etmiştir.

Balyaner (2016), 'Türkiye'de Hanehalkının Sahip Olduğu Bilişim Teknolojileri Ürünleri Sayısının Araştırılması: Bir Sayma Veri Modeli' adlı yüksek lisans tezinde Hanehalkı Bilişim Teknolojileri Kullanım Araştırması 2013 mikro veri setini kullanmıştır. Türkiye geneli, kentsel ve kırsal kesim için hem yetişkin hem de 6-15 yaş arası çocukların sahip olduğu bilişim teknolojileri ürünleri sayısını belirleyen faktörleri sayma veri modeli kullanarak incelemeyi amaçlamıştır. Analiz için bağımlı değişkenleri; 6-15 yaş arası çocukların sahip olduğu bilişim teknolojileri ürünleri sayısı ve yetişkinlerin sahip olduğu bilişim teknolojileri ürünleri sayısı olarak belirlemiştir. Bağımsız değişkenleri ise; çocuk ve yetişkinler için ayrı ayrı belirlemiştir. Analizde ortak kullanılan bağımsız değişkenleri; yaş, cinsiyet, yerleşim yeri, bölgeler, gelir ve eğitim olarak belirlemiştir. Çocuklar için belirlenen bağımsız değişkenler ise; en çok izlenilen program türleri (çocuk), cep telefonu kullanma amacı (çocuk), internet kullanım yeri, bilgisayar kullanma amacı (çocuk), internete bağlanma araçları, daha az zaman ayrılan faaliyetler, internet kullanım sıklığı, internet kullanım yılı, bilgisayar kullanım yılı olarak belirlemiştir. Yetişkinler için belirlenen bağımsız değişkenler ise; internet kullanım yeri, meslek, işteki durum, ev ve işyeri dışında internete bağlanmak için kullanılan cihaz, iş dışında internet kullanım amacı, kamu kurumu sitelerinden bilgi edinme amacı, e-ticaret amacı, yeterli bilgisayar bilgisi olanlar, bilgisayar kullanım sıklığı olarak belirlemiştir. Analizde Poisson Quasi Maksimum Olabilirlik Tahmin yöntemini kullanmış ve Robust Poisson regresyon ve Bootstrap tekniğinden faydalanmıştır. Analiz sonuçlarına göre; Türkiye'nin batı bölgelerinde yaşayan çocuklara kıyasla doğu bölgelerinde yaşayan çocukların daha az bilişim teknolojileri ürünlerine sahip olduğunu saptamıştır. Kentsel kesimde yaşayan yetişkin ve çocukların kırsal kesimde yaşayan yetişkin ve çocuklara göre daha fazla bilişim teknoloji ürünlerine sahip olduğunu tespit etmiştir. Hanehalkı geliri ve sahip olunan bilişim teknolojileri ürün sayısı arasında pozitif ve anlamlı bir ilişki olduğunu ortaya koymuştur. Yetişkinlerde meslek gruplarına göre bilişim teknolojileri ürün sayılarının farklılık gösterdiğini belirlemiştir. Yetişkinlerin yaşı ile sahip olduğu teknoloji ürünleri adedi arasında anlamlı ve pozitif ilişki olduğunu tespit etmiştir.

Soycan (2017), 'Sayıma Dayalı Olarak Elde Edilen Veri Kümelerinde Bağımlı Değişkenin Modellenmesinde Uygun Doğrusal Olmayan Regresyon Modelinin Belirlenmesi' adlı yüksek lisans tezinde, rüşvet suçlaması ile hüküm giyen kişilerin yıllara ve bölgelere göre durumunu belirmeyi amaçlamıştır. Çalışmada 2009-2012 dönemi için TÜİK arşivinden elde edilen verilere hem Poisson hem de Exact Poisson regresyonunu uygulamıştır. Analiz için bağımlı değişkeni rüşvet suçlaması ile hüküm giyen kişilerin sayısı olarak belirlemiştir. Bağımsız değişkenleri ise; yıllar ve bölgeler olarak belirlemiştir. Yıllar değişkenini analize 3 grup olarak dahil etmiştir. İlk olarak bölgeler ve bütün yılları dahil edip Poisson ve Exact Poisson uygulamıştır. İkinci olarak 2012 yılını analize dahil etmeden Poisson ve Exact Poisson uygulamıştır. Üçüncü olarak ise hem 2012 hem de 2011 yıllarını analize dahil etmeden Poisson ve exact Poisson uygulamıştır. Analiz sonucuna göre; bütün yıllar modelde iken yıllar değişkenini istatistiksel olarak anlamlı bulmuştur. Bölge değişkeni için ise referans kategorisi seçilen Marmara bölgesine göre Doğu Anadolu, İç Anadolu, Karadeniz ve Ege bölgeleri istatistiksel olarak anlamlı iken, Akdeniz ve Güneydoğu Anadolu bölgelerinin katsayıları istatistiksel olarak anlamlı olmadığını saptamıştır.

Şimşek (2017), ‘Poisson Regresyon Analizinde Log-Doğrusal Modellerin Kullanılması’ adlı yüksek lisans tez çalışmasında YGS’de 180 puan altı alan öğrenciler ve 180 puan ve üstü alan ancak herhangi bir programa yerleşemeyen öğrencilerin rölatif risklerini bölgelere ve yıllara göre belirlemeyi amaçlamıştır. Çalışmada 2010-2013 dönemi için TÜİK ve ÖSYM arşivinden elde edilen verilerin rölatif riskleri dikkate alınarak Poisson log-doğrusal regresyon analizini uygulamıştır. Analiz için bağımlı değişkenleri; YGS’de 180 puan altı alan öğrenci sayısı ve herhangi bir lisans programına yerleşemeyen öğrenci sayısı olarak belirlemiştir. Bağımsız değişkenleri ise; yıllar ve bölgeler olarak belirlemiştir. Analiz sonucuna göre; YGS’de öğrencilerin 180 puan ve altı alma durumu bakımından Güneydoğu Anadolu’daki öğrencilerin rölatif riskinin en fazla olduğu bölge İç Anadolu Bölgesi, en az olduğu bölge ise Doğu Anadolu Bölgesi’nin olduğunu tespit etmiştir. Yıllar açısından bakıldığında, 2010 yılındaki rölatif riskin en fazla olduğu yıl 2011, en az olduğu yıl 2013 yılları olarak tespit etmiştir. 180 ve üstü puan alıp herhangi bir programa yerleşemeyen öğrenciler için bölgelere göre bakıldığında; Güneydoğu Anadolu’daki öğrencilerin rölatif riskinin en fazla olduğu bölge Marmara, en az olduğu bölge ise Doğu Anadolu olduğunu saptamıştır. Yıllar için bakıldığında ise, 2010 yılındaki rölatif riskin en fazla olduğu yıl 2013, en az olduğu yıl 2011 yılı olarak saptamıştır.

Dinarcan (2018), ‘Sayma Verisi İçin Regresyon Modelleri ve Bir Uygulama’ adlı yüksek lisans tez çalışmasında bireylerin haftalık çalışma saatlerine etki eden faktörleri belirlemeyi amaçlamıştır. Çalışmada kullanılan veri setini 2016 yılı TÜİK tarafından yapılan Hanehalkı İşgücü Araştırması’ndan elde etmiştir. Veri setine en uygun regresyon modelini belirlemek için Poisson, Negatif Binom ve Genelleştirilmiş Poisson regresyon ile model tahmini yapmıştır. Analiz için bağımlı değişken olarak, bireylerin bir haftadaki çalışma saatlerini kullanmıştır. Bağımsız değişkenleri ise; sektör, yaş, cinsiyet, eğitim düzeyi, meslek, SGK kayıtlılık ve süreklilik olarak belirlemiştir. Analizden elde ettiği sonuçlara göre; veri setine en uygun modelin Negatif Binom Regresyon modeli olduğunu tespit etmiştir. Negatif Binom Regresyon modeli sonuçlarına göre; cinsiyet ve sektör değişkeni gruplarından inşaat sektörünün çalışma saati üzerinde bir etkisinin olmadığını ortaya koymuştur. Eğitim düzeyinin artmasının haftalık ortalama çalışma saatini kısalttığını tespit etmiştir. Tarım ve sanayi sektöründe çalışanların hizmet sektöründe çalışanlara göre haftalık ortalama çalışma saatine daha fazla etki ettiklerini tespit etmiştir. Yaş arttıkça haftalık ortalama çalışma saatinin de arttığı sonucuna ulaşmıştır.

Karaca (2018), ‘Sayma Verileri İçin Regresyon Modellerinin Karşılaştırılması Üzerine Bir Uygulama’ adlı yüksek lisans tezinde hizmet sektöründeki personellerin şikayet alma sayılarına etki eden faktörleri belirlemeyi amaçlamıştır. Çalışmada kullanılan veriler 2016 yılı için aylık olarak bir hizmet sektöründeki bir firmadan elde edilmiştir. Bu verilere en uygun regresyon modelini belirlemek için Poisson Regresyon, Negatif Binom Regresyon, Sıfır Değer Ağırlıklı Poisson Regresyon ve Sıfır Değer Ağırlıklı Negatif Binom Regresyon modellerini kullanmıştır. Bağımlı değişkeni, hizmet sektöründeki personellerin şikayet alma sayıları olarak belirlemiştir. Bağımsız değişkenleri ise; cinsiyet, yaş, eğitim durumu ve tecrübe süresi (ay) olarak belirlemiştir. Analiz sonuçlarına göre Ocak, Mart, Haziran ayları için Sıfır Değer Ağırlıklı Poisson; Şubat, Nisan, Ağustos ayları için Negatif Binom; Temmuz, Eylül, Ekim, Aralık ayları için Poisson ve Mayıs ayı için ise Sıfır Değer Ağırlıklı Negatif Binom Regresyon modellerinin veri setine en uygun modeller olduğu sonucuna ulaşmıştır. 12 aylık birikimli veri için ise en uygun modelin Sıfır Değer Ağırlıklı Negatif Binom Regresyon modeli olduğunu tespit etmiştir. 12 aylık veri için yapılan analiz sonucuna göre; cinsiyet, yaş, eğitim durumu ve tecrübe değişkenlerinin şikayet sayısı üzerinde istatistiksel olarak anlamlı bir etkisinin olduğunu tespit etmiştir.

Avcı (2018), ‘Using Count Regression Models to Determine the Factors which Effects the Hospitalization Number of People Schizophrenia’ adlı çalışmasında 2011-2014 yılları arasında Giresun’daki Prof. Dr. A. İlhan Özdemir Devlet Hastanesi Toplum ve Ruh Sağlığı Merkezi’nde bulunan şizofren hastalarına ait elde ettiği verilere sayma verisi regresyon modellerini uygulamıştır. Analizde bağımlı değişkeni, şizofren hastalarını hastaneye yatış sayısı olarak belirlemiştir. Bağımsız değişkenleri ise; yaş, cinsiyet, eğitim durumu, medeni durum, gelir durumu, kentsel statü, yalnız yaşama durumu, aile hastalığı, madde bağımlılığı, aile ve çevre ile ilişki ve etkinlik durumu olarak belirlemiştir. Poisson, Negatif Binom ve Conway-Maxwell-Poisson (COM-Poisson) Regresyon ile model tahmini yapılmıştır. Analiz sonuçlarına göre; AIC ve Log olasılık değerlerine bakıldığında veri setine en iyi uyum sağlayan modelin Negatif Binom Regresyon modeli olduğunu tespit etmiştir.

Çelik (2019), ‘Conway-Maxwell-Poisson Regresyon Modeli’ adlı yüksek lisans tezinde COM-Poisson regresyon modelini ele almıştır. Çalışmanın uygulama kısmında 2 ayrı veri seti kullanmıştır. Birinci veri setini, 2011-2014 dönemi için Giresun ili Toplum Ruh Sağlığı Merkezi’nden elde etmiştir. İkinci veri setini ise, 2015-2016 dönemi için Giresun Üniversitesi Bilimsel Araştırma Proje desteği ile gerçekleştirilen projeye ait verilerden elde etmiştir. Bağımlı değişkeni; birinci veri seti için Giresun ili Toplum Ruh Sağlığı Merkezi'ne kayıtlı olan hastaların merkeze gelme sayıları, ikinci veri seti için ise öğretmenlerin sosyal medya hesap sayıları olarak belirlemiştir. Bağımsız değişkenleri birinci veri seti için cinsiyet, yaş, medeni durum, eğitim düzeyi, sosyal güvenceye sahip olma, madde bağımlılık durumları olarak belirlemiştir. İkinci veri seti için bağımsız değişkenleri okul türü, cinsiyet ve yaş grubu olarak belirlemiştir. Analiz sonucuna göre; 2011-2014 dönemini içeren veri setine en uygun modelin AIC değerine göre COM-Poisson Regresyon modeli olduğu sonucuna ulaşmıştır. 2015-2016 dönemini içeren veri setinin az yayılım gösterdiğini tespit etmiştir. COM-Poisson regresyon modelinin ikinci veri setine en uygun model olduğu belirlenmiştir. Birinci veri seti için analiz sonucuna göre, kadınların erkeklerden daha fazla Toplum ve Ruh Sağlığı Merkezi'ne gittiğini saptamıştır. Eğitim düzeyi yükseldikçe Toplum ve Ruh Sağlığı Merkezi'ne gitme durumunun da arttığı sonucuna ulaşmıştır. İkinci veri seti için analiz sonucuna göre; okul türünün sosyal medya hesap sayısı üzerinde etkili olduğunu tespit etmiştir. Yaş grubuna göre yaş arttıkça sosyal medya hesap sayısının azaldığı sonucuna ulaşmıştır.

Kartal (2019), 'Kırpılmış Poisson Regresyon Analizi ve Bir Uygulama' adlı yüksek lisans tez çalışmasında, Kırpılmış Poisson Regresyon modelini ele almıştır. Çalışmanın uygulama kısmında üç farklı veri seti kullanmıştır. Bağımlı değişkeni birinci veri seti için hastaların hastanede yatma süreleri, ikinci veri seti için trafikte meydana gelen kaza sayıları ve üçüncü veri seti için akciğer kanseri teşhisi konan hasta sayısı olarak belirlemiştir. Bağımsız değişkenleri birinci veri seti için bağımsız değişkenler yaş, sağlık sigortası durumu, hastanın hastanede ölüm durumu; ikinci veri seti için kaza yapan kadın sürücüler, buzlanma, yağış, ehliyet; üçüncü veri seti için takip süresi, teşhis, yaş olarak belirlemiştir. Analiz sonuçlarına göre; birinci veri seti için en uygun modeli Negatif Binom Regresyon modeli ile elde etmiştir. İkinci ve üçüncü veri setleri için en uygun modeli Kırpılmış Poisson Regresyon modeli ile elde etmiştir. Birinci veri seti analiz sonuçlarına göre; sağlık sigortasının, ölüm durumunun ve sabit terimin istatistiksel olarak anlamlı olduğu sonucuna ulaşmıştır. İkinci veri seti analiz sonuçlarına göre; kaza yapan kadın sürücüler ve buzlanma değişkenlerinin istatistiksel olarak anlamlı olduğu sonucunu elde etmiştir. Yağışın, ehliyetin ve sabit terimin anlamlı olmadığı sonucuna ulaşmıştır. Üçüncü veri seti analiz sonuçlarına göre; tüm değişkenlerin katsayılarının istatistiksel olarak anlamlı olduğunu tespit etmiştir.

Kıran (2019), ‘Sayma Veri Modelleri ve Bir Uygulama’ adlı yüksek lisans tez çalışmasında hanehalkının sahip olduğu konut sayısını modellemiştir. Analizde kullandığı veri setini Türkiye İstatistik Kurumu 2016 yılı Hanehalkı Bütçe Anketi’nden elde etmiştir. Konut sayısı için en uygun modeli ve konut sayısı üzerinde etkisi olan faktörleri belirlemeyi amaçlamıştır. Analizde bağımlı değişkeni konut sayısı olarak belirlemiştir. Bağımsız değişkenleri ise; yaş, cinsiyet, medeni durum, hanehalkı geliri, tasarruf, şans oyunu, işteki durum ve çalışma saati olarak belirlemiştir. Veri setine en uygun modeli belirlemek için Poisson, Negatif Binom ve Sıfır Değer Ağırlıklı Poisson regresyon modellerini kullanmıştır. Analizden elde ettiği sonuçlara göre; veride aşırı yayılım durumu olmadığı için Negatif Binom Regresyon modelinin veri setine uygun olmadığını tespit etmiştir. AIC değerine ve maksimum log-olabilirlik sonuçlarını dikkate aldığında veri için en uygun modelin Sıfır Değer Ağırlıklı Poisson Regresyon modeli olduğu sonucuna ulaşmıştır. Seçilen modele göre; çalışma saati ile konut sahipliği sayısı arasında negatif yönlü bir ilişki olduğunu saptamıştır. Hanehalkı sorumlusu yaş artışının, evli, erkek veya işveren-kendi hesabına çalışıyor olması konut sayısını arttırdığı sonucuna ulaşmıştır. Şans oyunu oynayan bireyin hanehalkı içerisinde olması konut sayısını azaltıcı faktörlerden olduğunu tespit etmiştir.

Giray (2019), ‘Poisson ve Negatif Binom Regresyon Modelleri’ adlı yüksek lisans tez çalışmasında bir hastanede doktorların hastaları ziyaret sayılarını modellemiştir. Analizde Deb ve Trivedi (1997) çalışmasından alınan veri setini kullanmıştır. Veri Setine en uygun modeli belirlemek için Poisson, Negatif Binom, Sıfır Değer Ağırlıklı Poisson, Sıfır Değer Ağırlıklı Negatif Binom, Poisson Hurdle, Negatif Binom Hurdle, Poisson Ridge, Negatif Binom Ridge, Poisson Liu ve Negatif Binom Liu regresyon modelleri ile analiz etmiştir. Bağımlı değişkeni doktorların vizite sayısı olarak belirlemiştir. Bağımsız değişkenleri ise; hastanede kalınan gün sayısı, bireyin kendi hakkındaki sağlık algısı, kronik rahatsızlık sayısı, günlük yaşamdaki kısıtlılık durumu, yaş, cinsiyet, medeni durum, eğitim düzeyi, ailenin geliri, iş durumu ve özel sağlık sigortası durumu olarak belirlemiştir. Analiz sonucuna göre; veri seti için en uygun modelin Sıfır Değer Ağırlıklı Negatif Binom Regresyon modeli olduğunu tespit etmiştir. Bu modele ait analiz sonucuna göre; cinsiyet ve aile gelirinin vizite sayısı üzerinde etkisi olmadığını görmüştür. Doktorun vizite sayısı evli bireylerde evli olmayan bireylere göre yaklaşık %11 daha az olduğunu saptamıştır. Özel sağlık sigortası olan bireylerin olmayanlara göre doktorun vizite sayısı yaklaşık %14 daha fazla olduğu sonucuna ulaşmıştır.

1.2. Yabancı Kaynaklı Veriler ile Yapılan Önceki Çalışmalar

Wang ve Famoye (1997), 'Modeling Household Fertility Decisions with Generalized Poisson Regression' adlı çalışmada veri setine en iyi uyum sağlayan modeli belirlemeyi amaçlamışlardır. 1989 yılında yapılan Michigan Panel Gelir Dinamiği Çalışması (PSID) doğurganlık veri setini Poisson ve Genelleştirilmiş Poisson Regresyon modelleri ile tahmin etmişlerdir. Analiz için bağımlı değişkeni ailedeki çocuk sayısı olarak belirlemişlerdir. Bağımsız değişkenleri ise; istihdam durumu, eğitim durumu, yaş, ırk, aile geliri ve ailenin şehirde oturma durumu olarak belirlemişlerdir. Analiz sonuçlarına göre; veri seti için en uygun modelin Genelleştirilmiş Poisson Regresyon modeli olduğunu tespit etmişlerdir. Kadının eğitim durumunun doğurganlık üzerinde önemli bir etkisinin olduğu sonucuna ulaşmışlardır. Kadının ırkı, çalışma durumu ve geliri hanehalkının doğurganlık kararını etkileyen önemli faktörler arasında olduğunu tespit etmişlerdir.

Famoye ve Singh (2006), 'Zero-Inflated Generalized Poisson Regression Model with an Application to Domestic Violence Data' adlı çalışmalarında Portland'da 1996-1997 yılları arasında aile içi şiddet sayısını Sıfır Değer Ağırlıklı Genelleştirilmiş Poisson Regresyon ile modellemişlerdir. Bağımlı değişkeni, aile şiddet sayısı olarak belirlemişlerdir. Bağımsız değişkenleri ise mağdur ve saldırgan için eğitim düzeyi, tam zamanlı iş, gelir düzeyi, aile etkileşimi, bir kulübe ait olma ve uyuşturucu sorunu olarak ayrı ayrı belirlemişlerdir. 1996-1997 yılları arasında Portland Polis Bürosu'ndan alınan saldırganlara yönelik resmi kayıtlar ve mağdurlara yönelik uygulanan anketlerden elde edilen verileri Sıfır Değer Ağırlıklı Genelleştirilmiş Poisson Regresyon Regresyon modeli ile analiz etmişlerdir. Analiz sonuçlarına göre; mağdurun gelir düzeyi arttıkça daha az şiddet gördüğünü tespit etmişlerdir. Mağdurun eğitim düzeyi arttıkça şiddet sayısının daha az olduğu sonucuna ulaşmışlardır. Uyuşturucu sorunu olan saldırganın daha fazla şiddet uyguladığını analiz sonuçlarından elde etmişlerdir. Mağdurun bir kulübe ait olması şiddet sayısı ile pozitif ilişki içerinde olduğunu tespit etmişlerdir.

Czado vd. (2007), 'Zero-Inflated Generalized Poisson Models with Regression Effects On The Mean, Dispersion and Zero-Inflation Level Applied to Patent Outsourcing Rates' adlı çalışmada Avrupa Patent Bürosu'ndan elde ettikleri veriler ile patent sayısı için en uygun tahmin modelini belirlemeyi amaçlamışlardır. Çalışmada bağımlı değişkeni patent sayısı olarak belirlemişlerdir. Bağımsız değişkenleri ise; başvuru prosedüründen bağımsız olarak bir şirketin yıllık toplam başvuru sayısı, bir şirketin son 5 yıla atıfta bulunan uygulama sayısının varyasyonu, bir şirketin patent başvurusu yaptığı bilimsel alanların sayısı, bir şirketin çalışan sayısı, patent başına ortalama Ar-Ge maliyeti (Euro), çalışan başına 1000 Euro cinsinden Ar-Ge maliyeti, Ar-Ge verilerinin eksikliği, endüstri kolları ve yıllar olarak belirlemişlerdir. 1993-2000 yılları arasındaki 8 yıllık dönem için 107 Avrupa şirketine ait verileri Avrupa Patent Bürosu'ndan elde ederek bu veri setini Poisson, Genelleştirilmiş Poisson, Sıfır Değer Ağırlıklı Poisson ve Sıfır Değer Ağırlıklı Genelleştirilmiş Poisson Regresyon modelleri ile tahmin etmişlerdir. Analiz sonucuna göre; tahmin edilen modellerin AIC değerleri incelendiğinde veri setine en iyi uyum sağlayan modelin Sıfır Değer Ağırlıklı Genelleştirilmiş Poisson Regresyon modeli olduğunu tespit etmişlerdir.

Aa ve Naing (2012), 'Analysis Death Rate of Age Model with Excess Zeros using Zero Inflated Negative Binomial and Negative Binomial Death Rate: Mortality AIDS Co-Infection Patients, Kelantan Malaysia' adlı çalışmalarında Malezya'nın Kelantan bölgesindeki AIDS hastalarının ölüm sayısını modellemişlerdir. Veri setine en iyi uyum sağlayan tahmin modelini belirlemeyi amaçlamışlardır. Çalışmada bağımlı değişkeni AIDS hastalarının yaş kategorilerine göre ölüm sayısı olarak belirlemişlerdir. Bağımsız değişkenleri ise; cinsiyet, uyruk, ırk, medeni hal, meslek ve AIDS bulaşma yolu olarak belirlemişlerdir. Analiz için Malezya'nın Kelantan bölgesinde 2000-2008 yılları arasında Kota Bharu'nun ikincil ölüm verilerini (AIDS) Negatif Binom ve Sıfır Değer Ağırlıklı Negatif Binom regresyon modelleri ile tahmin etmişlerdir. Analiz sonuçlarına göre; elde edilen Log-olabilirlik, AIC ve Bayes Bilgi Kriteri (BIC) değerlerine bakıldığında veri setine en iyi uyum sağlayan modelin Sıfır Değer Ağırlıklı Negatif Binom Regresyon modelinin olduğunu tespit etmişlerdir.

Mouatassim ve Ezzahid (2012), 'Poisson Regression and Zero-Inflated Poisson Regression: Application to Private Health Insurance Data' adlı çalışmada Fas'ta özel sağlık sigortasına sahip bireylerden alınan yıllık hasar talep sayısı için en uygun tahmin modelini belirlemeyi amaçlamışlardır. Fas'ta özel sağlık sigortasına sahip bireylerden alınan yıllık hasar talep sayısı için oluşturulan veri setini Poisson ve Sıfır Değer Ağırlıklı Poisson Regresyon modelleri ile tahmin etmişlerdir. Bağımlı değişkeni özel sağlık sigortası sahiplerinden alınan yıllık hasar talebi sayısı olarak belirlemişlerdir. Bağımsız değişkenleri ise; sigortalının sanayi kentinde yaşama durumu, cinsiyet, sigortalının sanayi sektöründe çalışma durumu, sigortalının hizmet sektöründe çalışma durumu, sigortalının 30 yaşından küçük olma durumu, sigortalının 30-40 yaş aralığında olma durumu, sigortalının 40-60 yaş aralığında olma durumu, sigortalının evli olma durumu, sigortalını bekar olma durumu, ailedeki birey sayısı ve yıl içinde sigortalının teminat süresi olarak belirlemişlerdir. Analiz sonuçlarına göre; bağımlı değişkenin aşırı yayılım gösterdiğini tespit etmişlerdir. Sıfırların aşılması durumunda en iyi modeli seçmek için Voung testi hesaplanmış ve sonuç olarak veri setine en iyi uyum sağlayan modelin Sıfır Değer Ağırlıklı Poisson Regresyon modeli olduğunu tespit etmişlerdir.

Ismail ve Zamani (2013), 'Estimation of Claim Count Data using Negative Binomial, Generalized Poisson, Zero-Inflated Negative Binomial and Zero-Inflated Generalized Poisson Regression Models' adlı çalışmalarında Malezya OD verilerine ve GSOEP 2003 verilerine uyum sağlayan en iyi modeli belirlemeyi amaçlamışlardır. Ismail ve Zamani (2012) çalışmalarında kullandıkları veri setini 2001-2003 yılları arası 3 yıllık dönem için Özel Araç Kasko Hasar (OD) talep sayıları veri seti Negatif Binom ve Genelleştirilmiş Poisson Regresyon modelleri kullanarak tahmin etmişlerdir. 2003 yılı için Alman Sosyoekonomik Panel (GSOEP) verilerini Negatif Binom, Sıfır Değer Ağırlıklı Negatif Binom ve Sıfır Değer Ağırlıklı Genelleştirilmiş Poisson Regresyon modelleri ile tahmin etmişlerdir. Malezya OD verileri için bağımlı değişkeni özel araç kasko hasar talep sayısı ve GSOEP verileri için bağımlı değişkeni son 3 aydaki doktor vizite sayısı olarak belirlemişlerdir. Malezya OD verisi için bağımsız değişkenleri aracın yaşı, aracın motor gücü, aracın yapım türü ve lokasyon olarak belirlemişlerdir. GSOEP veri seti için bağımsız değişkenleri cinsiyet, yaş, sağlıktan memnuniyet, medeni hal, çalışma durumu ve eğitim alınan süre (yıl) olarak belirlemişlerdir. Malezya OD verileri için elde edilen analiz sonucuna göre; AIC ve BIC değerlerine dayanarak her iki kriter için de veri setine en iyi uyum sağlayan modelin Genelleştirilmiş Poisson Regresyon modeli olduğu tespit edilmiştir. GSOEP verileri için elde edilen Analiz sonucuna göre; AIC ve BIC değerlerine dayanarak Sıfır Değer Ağırlıklı Genelleştirilmiş Poisson Regresyon modelinin en düşük AIC değerine sahip olduğunu, fakat en düşük BIC değerine sahip olan modelin Sıfır Değer Ağırlıklı Negatif Binom Regresyon modeli olduğunu tespit etmişlerdir. Bu iki model arasındaki Vuong testine bakıldığında ise hiçbir modelin diğerine tercih edilemeyeceğini tespit etmişlerdir.

2. BÖLÜM

2. GENELLEŞTİRİLMİŞ DOĞRUSAL MODELLER

Klasik Doğrusal Regresyon Modelleri, bağımlı değişkenin dağılımının normal olduğu varsayımıyla, bağımlı değişkenin ortalaması ile bir dizi bağımsız değişken arasındaki ilişkiyi tanımlamak için doğrusal olma varsayımını kullanmaktadır. Bu varsayımların geçerli olmadığı durumda veri dönüşümü teknikleri kullanılmaktadır. Veri dönüşümü tekniklerine alternatif olarak Nelder ve Wedderburn (1972) tarafından ortaya atılan, McCullagh ve Nelder (1989) tarafından geliştirilen Genelleştirilmiş Doğrusal Modeller (GDM) kullanılmaktadır (Cameron ve Trivedi, 1998: 2). Genelleştirilmiş Doğrusal Modeller; normal dağılım göstermeyen bağımlı değişkenin dağılımını ve genelde ortalamanın doğrusal olmayan fonksiyonlarını kapsayacak biçimde Klasik Doğrusal Modelin varsayımlarını genişleten bir yöntemdir (Agresti, 2015: 2). GDM, klasik doğrusal regresyon modellerinden iki açıdan önemli olarak farklılık göstermektedir. Bunlar; bağımlı değişkenin üstel dağılım ailesinden olması ve bağımlı değişkenin ortalamasının dönüşümü, açıklayıcı değişkenlerle doğrusal olarak ilişkili olmasıdır. Bağımlı değişkenin üstel dağılım ailesinden olması bağımlı değişkenin normal veya normale yakın bir dağılım göstermesini gerektirmemektedir. Yani bağımlı değişken normal dağılım göstermeyebilir. Bağımlı değişkenin üstel dağılım ailesinden olması bağımlı değişkenin sabit varyanslı olmama durumunu ortaya koymaktadır. Yani varyans ortalama ile değişkenlik göstermektedir. Bu durum klasik doğrusal regresyon modelindeki sabit varyanslılık özelliği ile çelişmektedir (Jong ve Heller, 2008: 64).

GDM, klasik doğrusal regresyon modellerinin genelleştirilmiş halidir. GDM, normal dağılım gösterdiği varsayılan doğrusal modellerin üstel dağılımlarını tanımlamak için kullanılmaktadır.

2.1. Genelleştirilmiş Doğrusal Modellerin Bileşenleri

GDM üç bileşenden oluşmaktadır. Bunlar rasgele bileşen, doğrusal kestirici (sistematik) bileşen ve link (bağ) fonksiyonudur (McCullagh ve Nelder,1989: 27).

i) *Rasgele bileşen*

Rasgele bileşen; bağımlı değişken Y ve bağımlı değişkenin olasılık dağılımını belirtmektedir. $y=(y_1, y_2, \ldots, y_n)$ gözlemleri birbirinden bağımsız olarak kabul edilmektedir.

ii) *Sistematik bileşen*

$x_1, x_2, \ldots, x_p$ bağımsız değişkenlerinin birleşmesinden elde edilen doğrusal kestirici olan η sistematik bileşeni ifade eder. Eşitlik 1'deki gibi gösterilmektedir (McCullagh ve Nelder, 1989: 27).

$$\eta = \sum_{j=1}^{p} \beta_j x_j \tag{1}$$

iii) *Link (Bağ) fonksiyonu*

Link fonskiyonu, E(Y)'nin her bileşenine uygulanan ve onu doğrusal kestirici ile ilişkilendiren bir g(.) fonksiyonudur. Rasgele ve sistematik bileşen arasındaki ilişki olan link fonksiyonu Eşitlik 2'de ifade edilmektedir (McCullagh ve Nelder, 1989: 27):

$$E(Y) = \mu = \eta \tag{2}$$

Böylece link (bağ) fonksiyonu Eşitlik 3'teki gibi yazılabilmektedir.

$$g(\mu_i) = \eta = \sum_{j=1}^{p} \beta_j x_j \tag{3}$$

Genelleştirilmiş doğrusal modellerde link fonksiyonunun belirlenmesi önemlidir. Link fonksiyonu, bağımlı değişkenin gösterdiği dağılıma uygun olarak belirlenmektedir. Üstel dağılım ailesinin üyesi olan dağılımlar kanonik link olarak tanımlanan doğal link fonksiyonuna sahiptir (McCullagh ve Nelder, 1989: 27).

Bağımlı değişken Bernoulli, Poisson, Gamma veya başka bir dağılım gösterebilmektedir. Bu dağılımlar genel olarak üstel dağılım olarak adlandırılmaktadır. Üstel dağılım gösteren bir bağımlı değişkenin modellenmesinde GDM kullanılmaktadır.

2.2. Genelleştirilmiş Doğrusal Modeller İçin Üstel Dağılım Ailesi

Literatürde üstel dağılım ailesinin olasılık yoğunluk fonksiyonu Eşitlik 4'te gösterilmektedir (Agresti, 2015: 121):

$$f(y;\theta,\phi) = exp\left[\left(\frac{y\theta - b(\theta)}{\alpha(\phi)}\right) + c(y,\phi)\right] \tag{4}$$

Eşitlik 4'te θ doğal (kanonik) parametreyi ve ϕ ise yayılım parametresini göstermektedir. a(.), b(.) ve c(.) bilinen bir fonksiyonu ifade etmektedir. Literatürde ϕ yayılım parametresi gruplanmış ve gruplanmamış veriler için iki türlü tanımlanabilmektedir. $\alpha(\phi) = \phi/\omega$ olarak tanımlanan yayılım parametresini gruplanmış veya sıklık verilerinin analiz edildiği durumlarda kullanmak uygundur. Burada ω, bir ağırlıktır. $\omega = 1$ olduğunda yayılım parametresi $\alpha(\phi) = \phi$ olur ve bu durum gruplanmamış verilerin analiz edildiği durumda tercih edilmektedir. Gruplandırılmış verilerin analizi için ise $\omega = n$ olarak alınmaktadır (Jong ve Heller, 2008: 35).

Üstel dağılım ailesinde beklenen değer ve varyans Üstel dağılım ailesine ait log-olabilirlik fonksiyonu kullanılarak türetilmektedir. Üstel dağılım ailesinin log-olabilirlik fonksiyonu Eşitlik 5'te verilmektedir (Olsson, 2002: 39):

$$l(\theta, \phi; y) = logf(y; \theta, \phi) = \left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right) \tag{5}$$

θ'ya göre birinci mertebeden türev ve log-olabilirlik fonksiyonunun özelliği kullanıldığında Eşitlik 6 elde edilmektedir:

$$E\left(\frac{\partial l}{\partial \theta}\right) = 0 \tag{6}$$

Eşitlik 6 kullanılarak elde edilen üstel dağılım ailesinin ortalaması Eşitlik 7'de gösterilmektedir:

$$E\left(\frac{\partial l}{\partial \theta}\right) = E\left(\frac{y - b\prime(\theta)}{a(\phi)}\right) = 0 \qquad E(y) = \mu = b'(\theta) \tag{7}$$

Üstel dağılım ailesinin Eşitlik 5'teki log-olabilirlik fonksiyonunun özelliği kullanıldığında ve θ'ya göre ikinci mertebeden türevi alındığında ise Eşitlik 8'deki eşitlikler elde edilmektedir:

$$-E\left(\frac{\partial^2 l}{\partial \theta^2}\right) = E\left(\frac{\partial l}{\partial \theta}\right)^2 \qquad \frac{\partial^2 l}{\partial \theta^2} = -\frac{b\prime\prime(\theta)}{a(\phi)} \tag{8}$$

Eşitlik 8 kullanılarak elde edilen üstel dağılım ailesinin varyansı Eşitlik 9'da gösterilmektedir:

$$0 = \frac{-b\prime\prime(\theta)}{a(\phi)} + \frac{Var(Y)}{a^2(\phi)} \qquad Var(Y) = b''(\theta)a(\phi) \tag{9}$$

2.2.1. Üstel Dağılım Ailesinde Poisson, Binom ve Normal Dağılım

Üstel dağılım ailesinden olan Poisson, Binom ve normal dağılımlar genelleştirilmiş doğrusal modellemede yaygın olarak kullanılan istatistiksel dağılımlardandır. Üstel dağılım ailesinin bir üyesi olan Poisson dağılımında olasılık yoğunluk fonksiyonu Eşitlik 10'da gösterilmektedir (Agresti, 2015: 122):

$$f(y;\mu) = \frac{e^{-\mu}\mu^{y}}{y!} = exp[y\ln\mu - \mu - ln(y!)] \tag{10}$$

Eşitlik 10'da $\ln\mu = \theta$ doğal parametredir. Eşitlik 5'teki $b(\theta) = \exp(\theta)$, $\alpha(\phi) = 1$ ve $c(y,\phi) = -\ln(y!)$ olmaktadır. Poisson dağılımında ortalama ve varyans eşitlikleri Eşitlik 11 ve Eşitlik 12'de gösterilmektedir:

$$E(y) = b'(\theta) = \exp(\theta) = \mu \tag{11}$$

$$Var(y) = b''(\theta) = \exp(\theta) = \mu \tag{12}$$

Üstel dağılım ailesinin bir diğer üyesi olan Binom dağılımında olasılık yoğunluk fonksiyonu Eşitlik 13'te verilmektedir (Jong ve Heller, 2008: 22):

$$f(y;\pi,n) = \binom{n}{y}\pi^{y}(1-\pi)^{n-y} \tag{13}$$

Binom dağılımında ortalama ve varyans Eşitlik 14 ve Eşitlik 15'te gösterilmektedir:

$$E(y) = \pi \tag{14}$$

$$Var(y) = \pi(1-\pi)/n \tag{15}$$

Normal dağılımın olasılık yoğunluk fonksiyonu Eşitlik 16'da verilmektedir:

$$f(y;\ \mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right] \tag{16}$$

Eşitlik 16'da $\theta = \mu$ doğal parametredir. Eşitlik 5'teki $b(\theta) = \frac{1}{2}\mu^2 = \frac{1}{2}\theta^2$, $\alpha(\phi) = \sigma^2$ ve $c(y,\phi) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{y^2}{2\sigma^2}$ olmaktadır. Normal dağılımda ortalama ve varyans Eşitlik 17 ve Eşitlik 18'de gösterilmektedir:

$$E(y) = b'(\theta) = \theta = \mu \tag{17}$$

$$Var(y) = b''(\theta)\ \alpha(\phi) = \sigma^2 \tag{18}$$

Üstel dağılım ailesindeki bazı dağılımlar ve özellikleri Tablo 2.1'de gösterilmektedir (Olsson, 2002: 41):

Tablo 2.1. Üstel dağılım ailesindeki bazı dağılımların özellikleri

	Poisson	**Binom**	**Normal**	**Gamma**	**Ters Gauss**	**Negatif Binom**
α(ϕ)	1	1	σ^2	v^{-1}	σ^2	1
b(θ)	e^θ	$n\log(1+e^\theta)$	$\sigma^2/2$	$-\ln(-\theta)$	$-(-2\theta)^{1/2}$	$-\ln(1-e^\theta)/\alpha$
c(y;ϕ)	-ln(y!)	$\ln\binom{n}{y}$	$-\frac{1}{2}\left(\frac{y^2}{\phi}\right) + \ln(2\pi\phi)$	vln(vy)	$-\frac{1}{2}\ln\left(2\pi\phi y^3 + \frac{1}{\phi y}\right)$	
μ(θ)=E(Y)	e^θ	$ne^\theta/(1+e^\theta)$	θ	-1/θ	$-1/\theta^2$	$-e^\theta/(-1+e^\theta)\alpha$
V(μ)	μ	np(1-p)	σ^2	μ^2	μ^3	$\mu+\alpha\mu^2$
V(Y)	μ	nμ(1-μ)	σ^2	μ^2/v	$\sigma^2\ \mu^3$	$\mu+\alpha\mu^2$
Kanonik Link	Log	Logit	Birim	-1/μ	$-1/\mu^2$	ln(αμ/1+αμ)

Kaynak: Olsson, 2002

2.3. Genelleştirilmiş Doğrusal Modellerde Katsayı Tahmini

Katsayı tahmini; oluşturulan modelin katsayılarını tahmin etmek için örnek verilerin kullanılmasını sürecini ifade etmektedir. Birçok katsayı tahmin yöntemi vardır. GDM için katsayı tahmini En Çok Olabilirlik fonksiyonuna dayanmaktadır. En Çok Olabilirlik yönteminin temel fikri, belirli bir dağılım için verileri en iyi açıklayacak parametrelerin en olası değerlerini elde etmektir. Eşitlik 5'teki olabilirlik fonksiyonu β için düzenlendiğinde Eşitlik 19 elde edilmektedir:

$$L(\beta) = \sum_{i=1}^{n} L_i(\beta) \tag{19}$$

Eşitlik 19'un β'ya göre türevi alınarak sıfıra eşitlendiğinde Eşitlik 20 elde edilmektedir:

$$\frac{\partial L(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial L_i}{\partial \beta_j} = 0 \qquad j = 1,2,\dots,p \tag{20}$$

β parametresinin en çok olabilirlik tahminleri Eşitlik 20 çözülerek elde edilmektedir. Olasılık eşitliklerinin doğrusal olmaması katsayı tahminini zorlaştırmaktadır. Bu nedenle Newton-Raphson ve Fisher's Skoru gibi iteratif (yinelemeli) yöntemler kullanılarak katsayı tahmini yapılmaktadır (Jong ve Heller, 2008: 69). Newton-Raphson algoritması Hessian matrisinden, Fisher Skoru algoritması ise Hessian matrisinin beklenen değerini kullanarak çözümleme yapmaktadır (Çelik, 2019: 12).

2.4. Genelleştirilmiş Doğrusal Modellerde Uyum İyiliği

2.4.1. Sapma İstatistiği

Genelleştirilmiş doğrusal modellerde en iyi uyum, gözlem sayısı kadar parametre olduğunda elde edilmektedir. Böyle bir model doymuş model olarak tanımlanmaktadır. (Jong ve Heller, 2008:72). Doymuş olmayan model ise ilgilenilen modeldir. Sapma istatistiği, doymuş ve doymuş olmayan modeller arasındaki mesafe ölçüsü olarak tanımlanmaktadır. Sapma istatistiği "D" ile gösterilmekte ve Eşitlik 21'de verilmektedir:

$$D = 2(\check{l} - \hat{l}) \tag{21}$$

Eşitlik 21'de $\check{l}$ doymuş modelin maksimum log-olabilirliğini ve ise $\hat{l}$ doymuş olmayan modelin log-olabilirliğini göstermektedir. Üstel dağılım ailesinin bazı üyeleri için sapma istatistikleri Tablo 2.2'de verilmektedir.

Tablo 2.2. Üstel dağılım ailesinin bazı üyeleri için sapma istatistikleri

Dağılım	**Sapma İstatistiği**
Normal	$\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2$
Poisson	$2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i)\right]$
Binom	$2\sum_{i=1}^{n} n_i \left[y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) + (n_i - y_i) \ln\left(\frac{n_i - y_i}{1 - \hat{\mu}_i}\right)\right]$
Gamma	$2v\sum_{i=1}^{n}\left[-\ln\left(\frac{y_i}{\hat{\mu}_i}\right) + \frac{n_i - y_i}{\hat{\mu}_i}\right]$
Ters Gauss	$\frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i{}^2 y_i}$
Negatif Binom	$2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) - \left(y_i + \frac{1}{\kappa}\right) \ln\left(\frac{y_i + 1/\kappa}{\hat{\mu}_i + 1/\kappa}\right)\right]$

Kaynak: Jong ve Heller, 2008: 73

2.4.2. Pearson χ^2 İstatistiği

Genelleştirilmiş doğrusal modellerde uyum iyiliğini değerlendirmek için bir diğer istatistik Pearson χ^2 istatistiğidir. Pearson χ^2 istatistiği Eşitlik 22'de gösterilmektedir:

$$\chi^2 = \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \tag{22}$$

Eşitlik 22'de $\hat{\mu}_i$, μ_i'nin En Çok Olabilirlik tahmini; $V(\hat{\mu}_i)$ ise, $\hat{\mu}_i$'nin varyansını ifade etmektedir. Pearson χ^2 istatistiği yaklaşık olarak n gözlem değere ve k parametreye sahip χ^2_{n-k} dağılımı göstermektedir ve asimptotik olarak sapmaya eşdeğerdir (McCullagh ve Nelder, 1989: 34).

2.4.3. Akaike Bilgi Kriteri

Genelleştirilmiş doğrusal modellerde olabilirlik fonksiyonlarını karşılaştırmak için bilgi kriterleri kullanılan uyum iyiliği istatistikleri arasındadır. Akaike bilgi kriterine göre bir modelin uyum iyiliğini ölçmek için kullanılan formül Eşitlik 23'te verilmektedir:

$$D_c = D - \alpha k \phi \tag{23}$$

Eşitlik 23'te gösterilen D sapma, k modeldeki parametre sayısı ve ϕ dağılım parametresini göstermektedir. ϕ sabitse $\alpha \approx 4$'ün bir parametreyi %5 düzeyinde test etmeye eşdeğer olduğu söylenebilmektedir. $\alpha \approx 2$ ise minimuma yakın tahmin hatalarına neden olacağını gösterebilmektedir (Olsson, 2002: 46). Akaike bilgi kriteri Akaike (1973) tarafından önerilen ve model seçiminde sıklıkla kullanılan bilgi kriteridir. Eşitlik 23'te D_c değeri en küçük olan model sonuç olarak tercih edilen model olmalıdır. Log-olabilirlik ve parametre sayısına dayanan AIC formülü ise Eşitlik 24'te verilmektedir:

$$AIC = -2l + 2k \tag{24}$$

Eşitlik 24'te l log-olabilirlik fonksiyonu ve k ise modeldeki parametre sayısını ifade etmektedir. Yine tercih edilen model en küçük AIC değerine sahip olan model şeklindedir (Jong ve Heller, 2008: 62).

2.5. Veri Türüne Göre Genelleştirilmiş Doğrusal Modeller

GDM veri türüne göre farklılık göstermektedir. Sayma verileri, sürekli veriler ve kategorik veriler için Genelleştirilmiş Doğrusal Model tahminleri farklılaşmaktadır.

Sayma verileri için Poisson Regresyon, Negatif Binom Regresyon, Sıfır Değer Ağırlıklı Regresyon, Genelleştirilmiş Poison, Robust Poisson gibi sayma veri regresyon modelleri kullanılmaktadır. Sürekli veriler için Gamma Regresyon, Ters Gauss Regresyon, Twedie Regresyon, Çoklu Regresyon, Varyans Analizi gibi regresyon modelleri kullanılmaktadır. Kategorik veriler için ise; Lojistik Regresyon, Probit Modeller gibi tahmin modelleri kullanılmaktadır.

Bu tez çalışmasında Genelleştirilmiş Doğrusal Modellerden Sayma Verisi Regresyon Modelleri ayrıntılı olarak incelenmektedir.

3. BÖLÜM

3. SAYMA VERİSİ REGRESYON MODELLERİ

Saymak, bir şeyin kaç tane olduğunu anlamak için birer birer elden veya gözden geçirme eylemidir. Sayma verisi, sayıma dayalı olarak elde edilen veri olarak ifade edilmektedir. Sayma verisi genellikle olay sayılarını göstermektedir. Başka bir deyişle sayma verisi bir olayın kaç kez gerçekleştiğini ifade etmektedir. Sayma verisi negatif olmayan tamsayılardan oluşmaktadır. Bir ailedeki çocuk sayısı, bir doktorun bir günde baktığı hasta sayısı, belirli bir bölgede gerçekleşen trafik kazalarının sayısı, bireylerin sahip olduğu konut sayısı veya araç sayısı gibi değişkenler sayma verisine örnek olarak verilebilmektedir.

Sayıma dayalı olarak elde edilen veriler bazı özellikleri taşımaktadırlar. Bu özelliklerden ilki, değerleri her zaman tamsayıdır. İkinci özellik, olası en küçük değerleri sıfırdır ve bu sebeple asla negatif değer içermemektedir. Üçüncü özellik ise, değerlerin çoğu düşük ve nispeten yüksek değerler olmakla birlikte sıklıkla pozitif olarak çarpık görünmektedir.

Sayma verisi ilk olarak aktüerya ve biyoistatistik alanlarında kullanılmaya başlanmış olsa da zaman içerisinde ekonomi, sağlık, siyaset bilimi ve sosyal bilimler gibi alanlarda da sıkça kullanılmaktadır. Sayıma dayalı olarak elde edilen veri türünün günlük hayatta kullanılabilirliği açısından ekonometrik analizlerde de geniş yer tuttuğu görülmektedir.

Ekonometrik analizlerde kullanılması gereken regresyon modeli verinin yapısına göre farklılık göstermektedir. Bağımlı değişkenin sayıma dayalı olarak elde edilen veri setinden oluşması durumunda Sayma Verisi Regresyon Modelleri kullanılmaktadır. Sayma Verisi Regresyon Modelleri, Tamsayılı Regresyon Modelleri olarak da adlandırılmaktadır. Bu modellerde bağımsız değişken ya da değişkenler için herhangi bir kısıt bulunmamaktadır (Mert,2016;240).

Sayma verisi regresyon modellerinde bağımlı değişken kesikli değerler almaktadır. Sayıma dayalı olarak elde edilen sayma verisinin yapısı kesikli olduğu için bu veri türüne ait regresyon modeli oluşturulurken kesikli dağılımlar kullanılmıştır. Sayma verisi regresyon modelleri, kesikli dağılımlardan Poisson ve Negatif Binom dağılımları kullanılarak geliştirilmiştir. Sayma verisi regresyon modelleri bağımlı değişkenin kesikli yapıda olmasının yanında kullanıldığı fonksiyonel form gereği doğrusal olmayan modellerdendir (Cameron ve Trivedi, 1998;3).

Sayma verisi regresyon modellerinde dikkat edilmesi gereken en önemli hususlardan biri bağımlı değişkenin ortalama ve varyansıdır. Kullanılması gereken yöntem bağımlı değişkenin ortalama ve varyansına göre değişmektedir. Bir diğer önemli husus ise, bağımlı değişkene ait veri setinde sıfır (0) frekansının büyüklüğüdür. Bağımlı değişkene ait veri setinde çok fazla sıfır değeri varsa değişkenin ortalaması azalacağı için kullanılması gereken yöntem bu durumda da değişmektedir (Mert, 2016; 240).

3.1. Poisson Regresyon Modeli

Poisson Regresyon Modeli, sayma verisi modellerinin temelini oluşturan bir modeldir. Bu model doğrusal olmayan bir regresyon modelidir. Poisson regresyon modelinde bağımlı değişkene ait verilerin dağılımı Poisson dağılım göstermektedir.

3.1.1. Poisson Dağılımı

Poisson dağılımı gerçekte nadiren karşılaşılan olayların belirli bir zaman aralığında oluş sayısını belirlemek için kullanılmaktadır. Kesikli Y rassal değişkeni, μ parametresi $\mu > 0$ ile Poisson dağılımının olasılık yoğunluk fonksiyonu Eşitlik 25'te gösterilmektedir (Cameron ve Trivedi, 1998:3):

$$f(Y = y) = \frac{\mu^{y} e^{-\mu}}{y!}, \qquad y = 0,1,2,3,\dots \tag{25}$$

Poisson dağılımı, ortalamasının ve varyansının eşit olduğu bir özelliğe sahiptir. Bu durum Eşitlik 26'da gösterilmektedir:

$$E(Y) = Var(Y) = \mu \tag{26}$$

3.1.2. Poisson Regresyon Modeli

Poisson regresyon modeli, μ parametresi ile bağımsız değişken X arasındaki ilişkiyi gösteren Poisson dağılımından türetilen bir modeldir. x_i'ye bağlı y_i için Poisson regresyon modeli Eşitlik 27'de gösterilmektedir:

$$f(y_i|x_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}, \quad y_i = 0,1,2,\dots. \tag{27}$$

Modelin koşullu ortalama parametresi üstel fonksiyon olarak Eşitlik 28'de gösterildiği gibi tanımlanmaktadır:

$$E(y_i|x_i) = \mu_i = \exp(x_i'\beta) \qquad i = 1,2,3,\ldots \tag{28}$$

Eşitlik 28'de exp(.) daima pozitif olduğu için y'nin de tahmin edilen değerlerinin pozitif olacağı görülmektedir. Sayma verisine sahip olan bağımlı değişken (y) sıfır değerini alabileceği için Eşitlik 28'e logaritmik dönüşüm uygulanarak Eşitlik 29'daki gibi doğrusal bir yapı elde edilmektedir:

$$\ln E\,(y_i|x_i) = x_i'\beta \qquad i = 1,2,3,\ldots \tag{29}$$

Bu eşitlik istatistik literatüründe 'Log-Doğrusal Fonksiyon' olarak da bilinmektedir.

Poisson regresyon modelinin tahmininde En Çok Olabilirlik (MLE) yöntemi kullanılmaktadır. $X_1, X_2, \ldots, X_n$, θ parametresi ile f(x,θ) olasılık yoğunluk fonksiyonuna sahip rastgele bir örneklem olmak üzere bu değişkenlerin bileşik olasılık yoğunluk fonksiyonu Eşitlik 30'da gösterilmektedir (Erdem, 2013; 272):

$$f_{x_1,x_2,\ldots,x_n}(x_1, x_2, \ldots, x_n) = f_{x_1}(x_1) f_{x_2}(x_2) \ldots f_{x_n}(x_n) = \prod_{i=1}^{n} f_{x_i}(x_i;\theta) \tag{30}$$

Olabilirlik fonksiyonunda x_i'ler değişmez olup, L(θ) parametresinin bir fonksiyonudur. L(θ) olabilirlik fonksiyonunu maksimum yapan $\hat{\theta}$ parametresi En Çok Olabilirlik tahmin edicisi olarak ifade edilmektedir. L(θ) olabilirlik fonksiyonunu maksimum yapan $\hat{\theta}$ parametresi bulunurken logaritmik dönüşüm kullanılmaktadır. Logaritmik dönüşümden faydalanarak ln L(θ) elde edilir. Bu fonksiyonu maksimum yapan θ değerini elde edebilmek için θ'ya göre birinci dereceden kısmi türevi alınarak sıfıra eşitlenir:

$$\frac{d \ln L(\theta)}{d\theta} = 0 \tag{31}$$

En Çok Olabilirlik yöntemine dayanarak verilen bilgilerden faydalanarak Poisson regresyonuna ait log-olabilirlik fonksiyonu Eşitlik 32'de verilmektedir (Cameron ve Trivedi,1998:62):

$$\ln L(\beta) = \sum_{i=1}^{n} [y_i x_i'\beta - \exp(x_i'\beta) - \ln y_i!] \tag{32}$$

Poisson regresyon modelinin β parametresinin en çok olabilirlik tahmini için Eşitlik 19'daki log-olabilirlik fonksiyonunun β'ya göre 1. dereceden kısmi türevi alınıp sıfıra eşitlenerek Eşitlik 33 elde edilmektedir (Cameron ve Trivedi, 1998:62).

$$\frac{d\ln L(\beta)}{d\beta} = \sum_{i=1}^{n} [y_i - \exp(x_i'\beta)] x_i = 0 \tag{33}$$

Eşitlik 33'teki ortalama parametresi üstel olduğu için model doğrusal bir model değildir. Bu durumda β tahmini için bir çözüm yoktur. Poisson regresyonda β tahminlerini en çok olabilirlik yöntemi ile bulabilmek için iteratif (tekrarlı) yöntemlerden olan Newton-Raphson ya da Gauss-Newton yöntemleri kullanılmaktadır. Buna göre Eşitlik 33'teki eşitliğin β'ya göre ikinci türevi alınırsa Eşitlik 34 elde edilmektedir (Cameron ve Trivedi,1998:62).

$$\frac{d^2\ln L(\beta)}{d\beta d\beta\prime} = -\sum_{i}^{n} \exp(x_i'\beta) x_i x_i' \tag{34}$$

Tekrarlı yöntemler kullanılarak elde edilen β tahminleri, x_i'deki bir birimlik değişimin $E(Y_i|x_i)$'deki neden olduğu nispi değişimi ölçmektedir. Örneğin; $\hat{\beta}_i = 0.25$ ve $\exp(x_i'\hat{\beta}) = 3$ ise i. bağımsız değişkendeki bir birimlik değişim Y'nin beklentisini 0.75 birim arttırmaktadır (Cameron ve Trivedi,2003:334).

Koşullu ortalama ve varyansın birbirine eşit olduğu Poisson regresyon modelinde bu durum eşit yayılım (equidispersion) olarak ifade edilmektedir. Uygulamalarda genellikle bu eşitlik sağlanamamaktadır. Bu durumda aşırı yayılım (overdispersion) veya az yayılım (underdispersion) durumları ortaya çıkmaktadır.

3.2. Negatif Binom Regresyon Modeli

Poisson dağılımı, bağımlı değişkenin ortalama ve varyansının eşit olduğu bir özelliğe sahiptir. Bu varsayıma eşit yayılım (equidispersion) denir. Sayıma dayalı elde edilen veri seti içerisinde sıfır değerlerinin fazla olması ya da daha yüksek değerlerin olması eşit yayılım varsayımını ihlal etmektedir. Bağımlı değişkenin varyansının ortalamadan büyük olduğu durumlarda aşırım yayılım ortaya çıkmaktadır. Aşırı yayılımın geçerli olduğu durumlarda Poisson regresyon modeli tutarlıdır fakat yeterli değildir. Bu durum standart hataların yanlı olmasına ve araştırmacı tarafından katsayıların yanlış yorumlanmasına sebep olabilmektedir. Bağımlı değişkenin aşırı yayılım gösterdiği durum için alternatif yöntem olarak Negatif Binom Regresyon Modeli (NBRM) kullanılmaktadır (Beaujean ve Grant , 2016: 4).

NBRM'de bağımlı değişkenin negatif binom dağılım gösterdiği varsayılmaktadır. Negatif Binom (NB) dağılımı, Poisson-Gamma karışımı bir dağılıma dayanmaktadır. Aşırı yayılımın ortaya çıktığı durumda tutarlı tahminciler elde etmek için bağımlı değişkene ait dağılım $Y \sim \mathrm{Poisson}(Y|\mu\upsilon)$ şeklinde yazılmaktadır. Burada υ Gamma dağılımına sahiptir ve

$\upsilon \sim Gamma(1, \alpha)$ şeklinde gösterilmektedir (Mert,2016:250). Başka bir ifadeyle υ aşırı yayılım parametresinin (α) tersine eşittir. Böylelikle Y bağımlı değişkeninin dağılımı, Poisson-Gamma karışık dağılımı olan NB dağılımıdır. NBRM'nin olasılık yoğunluk fonksiyonu Eşitlik 35'te gösterilmektedir (Lawless, 1987: 210):

$$f(y_i) = \frac{\Gamma(y_i+\alpha^{-1})}{y_i!\Gamma(\alpha^{-1})} \left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \left(\frac{1}{1+\alpha\mu_i}\right)^{\alpha^{-1}} \quad , \quad y_i = 0,1,2,\ldots \quad \alpha > 0 \tag{35}$$

Eşitlik 35'te $\Gamma(.)$, Gamma fonksiyonunu göstermektedir. Yayılım parametresi ise α ile gösterilmektedir. $\alpha>0$ durumunda aşırı yayılım vardır.

NBRM, Poisson Regresyon modelinin özel bir durumudur. NBRM için parametre tahminleri iteratif (tekrarlı) yöntemler yardımıyla En Çok Olabilirlik yöntemi ile elde edilmektedir.

NBRM'nin birçok türü vardır. NBRM'nin genelleştirilmiş hali, ortalaması μ ve varyansı $\mu + \alpha\mu^t$ olarak tanımlanmaktadır (Cameron ve Trivedi, 1998:63). Burada t sabit bir değerdir. t=2 durumu Negatif Binom Regresyon Modeli 2 (NBRM-2) modeli, t=1 durumu Negatif Binom Regresyon Modeli 1 (NBRM-1) modeli olarak adlandırılmaktadır. NBRM'de sıklıkla kullanılan model NBRM-2 modelidir. NBRM-2 modelinin varyansı $\mu(1 + \mu\alpha) = \mu + \alpha\mu^2$ olarak tanımlanmaktadır ve karesel yapıda olduğu için NBRM-2 olarak adlandırılmaktadır (Ismail ve Zamani, 2013: 9). NBRM'nin bir diğer türü NBRM-1 modelidir. NBRM-1 modelinin NBRM-2 modelinden farkı, NBRM-1 modelinin varyansının $\mu + \alpha\mu$ şeklinde olmasıdır. NBRM-1 modelinin varyansı doğrusal yapıdadır.

NBRM-2 modelinin ortalaması μ, varyansı $\mu(1 + \mu\alpha) = \mu + \alpha\mu^2$ şeklinde gösterilmektedir. NBRM-2 modeli, karesel varyans fonksiyonu olarak da adlandırılmaktadır. NBRM-2 modelinin olasılık yoğunluk fonksiyonu Eşitlik 36'da verilmektedir.

$$f(y_i|\mu_i, \alpha) = \frac{\Gamma(y_i+\alpha^{-1})}{y_i!\Gamma(\alpha^{-1})} \left(\frac{\mu_i}{\alpha^{-1}+\mu_i}\right)^{y_i} \left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu_i}\right)^{\alpha^{-1}} , \quad y_i = 0,1,2,\ldots \quad \alpha \geq 0 \tag{36}$$

Eşitlik 36'daki gamma fonksiyonu logaritmik dönüşüm yapıldığında Eşitlik 37'de gösterildiği gibi düzenlenmektedir.

$$\ln\left(\frac{\Gamma(y_i+\alpha^{-1})}{\Gamma(\alpha^{-1})}\right) = \sum_{j=0}^{y_i-1} \ln(j + \alpha^{-1}) \tag{37}$$

NBRM-2 modeline ait log-olabilirlik fonksiyonu için üstel formda ortalama $\mu_i = \exp(x_i'\beta)$ olarak yazıldığında Eşitlik 38 elde edilmektedir.

$$\ln L(\alpha,\beta) = \sum_{i=1}^{n} \left\{ \begin{matrix} \left(\sum_{j=0}^{y_i-1} \ln(j+\alpha^{-1})\right) - \ln y_i! - \\ (y_i+\alpha^{-1})\ln\left(1+\alpha\exp(x_i{'}\beta)\right) + y_i\ln\alpha + y_i x_i'\beta \end{matrix} \right\} \tag{38}$$

NBRM-2 modeli için $\hat{\beta}$, α parametrelerinin tahmini için En Çok Olabilirlik fonksiyonu Eşitlik 39'da ve Eşitlik 40'ta verilmektedir (Winkelmann, 2008: 135):

$$\sum_{i=1}^{n} \frac{y_i - \exp(x_i'\beta)}{1+\alpha\exp(x_i'\beta)} x_i = 0 \tag{39}$$

$$\sum_{i=1}^{n} \left[\frac{1}{\alpha^2} \left(\ln\left(1+\alpha\exp(x_i'\beta)\right) - \sum_{j=0}^{y_i-1} \frac{1}{j+\alpha^{-1}} \right) - \frac{(y_i+\alpha^{-1})\exp(x_i'\beta)}{1+\alpha^{-1}\exp(x_i'\beta)} + \frac{y_i}{\alpha} \right] = 0 \tag{40}$$

NBRM-2 modeline ait ortalama parametresi doğrusal yapıda olmadığı için En Çok Olabilirlik yöntemiyle tek adımda çözüme ulaşılamamaktadır. Bu nedenle NB2 modeli için parametre tahminleri iteratif (tekrarlı) yöntemler kullanılarak yapılmaktadır. Eşitlik 39 ve Eşitlik 40'taki eşitliklerin çözülmesiyle parametre tahminleri elde edilmektedir.

NBRM-2 modeli için tahmin edilen β parametresine ait varyans ise Eşitlik 41'de gösterilmektedir (Winkelmann, 2008: 135).

$$\mathrm{Var}(\hat{\beta}) = \left(\sum_{i=1}^{n} \frac{\exp(x_i'\beta)}{1+\alpha\exp(x_i'\beta)} x_i x_i' \right)^{-1} \tag{41}$$

NBRM-1 modelinin ortalaması μ ve varyansı μ+αμ şeklinde gösterilmektedir. NBRM-1 modeli, doğrusal varyans fonksiyonu olarak da adlandırılmaktadır. NBRM-1 modeline ait varyans fonksiyonu klasik doğrusal regresyon modelinde kullanılan varyans fonksiyonu ile benzerlik göstermektedir. NBRM-1 modeline ait log-olabilirlik fonksiyonu Eşitlik 42'de verilmektedir (Winkelmann, 2008: 136):

$$\ln L(\alpha,\beta) = \sum_{i=1}^{n} \left[\begin{matrix} \left(\sum_{j=0}^{y_i-1} \ln(j+\alpha^{-1}\exp(x_i'\beta)\right) - \ln y_i! - \\ \left(y_i+\alpha^{-1}\exp(x_i'\beta)\right)\ln(1+\alpha) + y_i\ln\alpha \end{matrix} \right] \tag{42}$$

NBRM-1 modeli için En Çok Olabilirlik tahminin ilk sıra koşulları çözümü Eşitlik 43 ve Eşitlik 44'te verilmektedir (Cameron ve Trivedi,1998: 74):

$$\sum_{i=1}^{n}\left[\left(\sum_{j=0}^{y_i-1}\frac{\alpha^{-1}\exp(x_i'\beta)}{j+\alpha^{-1}\exp(x_i'\beta)}\right)x_i+\alpha^{-1}\exp(x_i'\beta)x_i\right]=0 \tag{43}$$

$$\sum_{i=1}^{n}\frac{1}{\alpha^2}\left[-\left(\sum_{j=0}^{y_i-1}\frac{\exp(x_i'\beta)}{j+\alpha^{-1}}\right)-\alpha^{-2}\exp(x_i'\beta)\ln(1+\alpha)-\frac{\alpha}{1+\alpha}+y_i\alpha\right]=0 \tag{44}$$

NBRM-1 olasılık yoğunluk fonksiyonunun ilk iki momentine dayanan tahmin, aynı zamanda NBRM-1 tahmincisi olarak da adlandırılan Poisson tahmincisini vermektedir (Cameron ve Trivedi, 1998:74).

3.3. Sıfır Değer Ağırlıklı Regresyon Modelleri

Sayıma dayalı olarak elde edilen veri seti içerisinde sıfır değerinin ağırlıklı olarak fazla olduğu durumla karşılaşılabilmektedir. Veri setindeki sıfır değerleri ortaya çıkış durumlarına göre iki şekildedir. Bunlar yapısal sıfır ve örnekleme sıfırı olarak adlandırılmaktadır. Değerlendirilen veri setinde elde etme ihtimali olmadığı durumda yani gözlem yapmanın olanaksız olduğu durumlardaki sıfır değeri yapısal sıfır; elde edilmesi mümkün olduğu halde veri setinde gözlenmeyen ve sıfır değeri alan değerlere de örnekleme sıfırı denir. Örneğin; bir sahil kasabasında yaşayan kişilerden bir örneklem seçilerek belirli bir gün içerisinde kaç balık tuttukları sorulmaktadır. O gün balık tutmaya gitmeyenlerin alacağı değer sıfır olacaktır ve bu sıfır yapısal sıfır olarak adlandırılmaktadır. Örnekleme giren bireyler o gün balık tutmaya gitmiş fakat hiç balık tutamamışlarsa da bağımlı değişkenin alacağı değer yine sıfır olacaktır ve bu sıfır ise örnekleme sıfırıdır (Mert, 2016: 256).

Gözlemlerin büyük bir kısmının sıfır değerini aldığı veri setinde sıfır değerlerinin analiz dışında bırakılması tutarlı olmayan sonuçların elde edilmesine neden olmaktadır. Veri setinin çok fazla sıfır değerine sahip olması durumunda analiz için Sıfır Değer Ağırlıklı Regresyon modellerinin kullanılması gerekir (Kaya ve Yeşilova, 2012: 52). Değerlendirilen bağımlı değişkene ait veri setinin gösterdiği eşit veya aşırı yayılım durumuna göre kullanılması gereken sıfır değer ağırlıklı modeller, Sıfır Değer Ağırlıklı Poisson ve Sıfır Değer Ağırlıklı Negatif Binom regresyon modelleridir.

Sıfır Değer Ağırlıklı regresyon modelleri, bağımlı değişkenin sahip olduğu sıfır değerleri için Bernoulli denemesi yapılarak belirlenen bir yoğunluk ile sıfır değerlerinin de dahil olduğu kesikli bir dağılımın karmasından oluşmaktadır (Tüzel ve Sucu, 2012: 25).

Sıfır Değer Ağırlıklı regresyon modelleri, iki bölümde tahmin edilmektedir. Bağımlı değişkenin 0 ve 1 değerlerini aldığı lojistik (logit) bölüm ve tamsayılı bölüm olmak üzere tahmin edilen modelde bağımlı değişkenin iki türlü 0 (sıfır) değerini alacağı düşünülerek parametre tahminleri yapılmaktadır.

3.3.1. Sıfır Değer Ağırlıklı Poisson Regresyon Modeli

Bağımlı değişkene ait veri seti eşit yayılıma sahip olduğunda, yapısal sıfırın önemli bir etkisinin olup olmadığını gösteren model Sıfır Değer Ağırlıklı Poisson Regresyon (ZIP) modelidir. Sıfır Değer Ağırlıklı Poisson regresyon modeli, fazla sıfırlarla ilgili heterojen olma durumuyla ilgilenmektedir (Akinpelu vd., 2016: 223).

Sıfır Değer Ağırlıklı Poisson regresyon modelinde, veride aşırı yayılımın sadece sıfır değerlerinin fazla olmasından kaynaklandığı varsayılmaktadır. Veri setindeki aşırı yayılım durumu sıfır değerlerinden kaynaklanmıyorsa, Sıfır Değer Ağırlıklı Poisson regresyon modeli ile elde edilen sonuçlar tutarlı olmamaktadır.

Sıfır Değer Ağırlıklı Poisson Regresyon modeli için olasılık yoğunluk fonksiyonu $\mathrm{Y_i \sim 0 \rightarrow \omega_i}$ ve $\mathrm{Y_i \sim Poisson(\mu_i)}$ olasılıkla Eşitlik 45 ve Eşitlik 46'da gösterilmektedir (Lambert, 1992; 3).

$$\mathrm{f(Y_i = 0) = \omega_i + (1 - \omega_i)exp(-\mu_i), \ \ y_i = 0} \tag{45}$$

$$\mathrm{f(Y_i = y_i) = (1 - \omega_i)exp(-\mu_i)\frac{\mu_i^{y_i}}{y_i!}, \ \ y_i > 0} \tag{46}$$

Eşitlik 45 ve Eşitlik 46'da verilen ω_i yapısal sıfırların olma olasılığını göstermektedir. Sıfır Değer Ağırlıklı Poisson regresyon modeli için $0 \leq \omega_i < 1$ ve $\mu_i > 0$ varsayımı altında ortalama ve varyans Eşitlik 47 ve Eşitlik 48'de gösterilmektedir (Zulkifli vd., 2011: 515).

$$\mathrm{E(Y_i) = (1 - \omega_i)\mu_i} \tag{47}$$

$$\mathrm{V(Y_i) = (1 - \omega_i)\mu_i(1 + \omega_i\mu_i)} \tag{48}$$

ω_i değeri sıfıra eşit olduğunda Sıfır Değer Ağırlıklı Poisson regresyon modeli, Poisson regresyon modeline dönüşmektedir. Sıfır Değer Ağırlıklı Poisson regresyon modelinin tahmini için Eşitlik 49'daki fonksiyonlar kullanılmaktadır (Lambert, 1992: 3).

$\mu_i = (\mu_1, \mu_2, \dots, \mu_n)$ ve $\omega_i = (\omega_1, \omega_2, \dots, \omega_n)$ olmak üzere,

$$\log(\mu_i) = B_i\beta$$

$$\operatorname{logit}(\omega_i) = \log\left(\frac{\omega_i}{1-\omega_i}\right) = G_i\gamma \tag{49}$$

$$\omega_i = \frac{e^{G_i\gamma}}{1+e^{G_i\gamma}}$$

$$(1-\omega_i) = \frac{1}{1+e^{G_i\gamma}}$$

Eşitlik 49'daki B ve G ortak değişken matrislerini göstermektedir. β ve γ ise bilinmeyen parametre vektörlerini göstermektedir. B ortak değişken matrisi, bağımlı değişkendeki Poisson sonucunu göstermektedir. G ortak değişken matrisi ise, bağımlı değişkendeki fazla sıfırlardan sorumludur (Mamun, 2014: 6).

Verilen Eşitlikler kullanılarak Eşitlik 50'deki olabilirlik fonksiyonu elde edilmektedir (Mamun, 2014: 6):

$$L(Y|\gamma,\beta) = \prod\left(\left(\frac{e^{G_i\gamma}}{1+e^{G_i\gamma}} + \frac{e^{-e^{B_i\beta}}}{1+e^{G_i\gamma}}\right)\left(\frac{e^{-e^{B_i\beta}}\left(e^{B_i\beta}\right)^{y_i}}{(1+e^{G_i\gamma})y_i!}\right)\right) \tag{50}$$

Log-olabilirlik fonksiyonu ise Eşitlik 51'de gösterilmektedir (Lambert, 1992; 4):

$$\ln L(Y|\gamma,\beta) = \sum_{y_i=0} \log\left(e^{G_i\gamma} + e^{-e^{B_i\beta}}\right) + \sum_{y_i>0}\left(y_iB_i\beta - e^{B_i\beta}\right) - \sum_{i=1}^{n}\log(1+e^{G_i\gamma}) - \sum_{y_i>0}\log(y_i!) \tag{51}$$

G_i ve B_i, G ve B matrislerinin i. sırasını göstermektedir. Log-olabilirlik fonksiyonundaki üstel terimlerin maksimize edilmesi oldukça karmaşıktır. Bu yüzden En Çok Olabilirlik fonksiyonu ile parametre tahmini için, EM algoritması veya Newton-Raphson algoritması kullanılmaktadır (Agresti, 2015: 252).

3.3.2. Sıfır Değer Ağırlıklı Negatif Binom Regresyon Modeli

Sıfır değerlerinin çok fazla olduğu ve aşırı yayılım özelliği gösteren bağımlı değişkenin modellenmesinde kullanılan regresyon yöntemi Sıfır Değer Ağırlıklı Negatif Binom regresyon modelidir.

Sıfır Değer Ağırlıklı Negatif Binom (ZINB) regresyon modelinin birçok türü vardır. Sıfır Değer Ağırlıklı Negatif Binom regresyon modelinin genelleştirilmiş hali, ortalaması $(1-\omega_i)\mu_i$ ve varyansı $(1-\omega_i)\mu_i(1+\alpha\mu_i^{t-1}+\omega_i\mu_i)$ olarak tanımlanmaktadır. Burada t sabit bir değerdir. t=1 durumu Sıfır Değer Ağırlıklı Negatif Binom 1 (ZINB-1) modelini ve t=2 durumu Sıfır Değer Ağırlıklı Negatif Binom 2 (ZINB-2) modelini göstermektedir (Ismail ve Zamani, 2013: 10). ZINB-1 modeline ait olasılık yoğunluk fonksiyonu Eşitlik 52'de verilmektedir (Zulkifli vd., 2012: 4).

$$f(Y_i=0)=\omega_i+(1-\omega_i)(1+\alpha)^{-\mu_i\alpha^{-1}},\ \ y_i=0$$

$$f(Y_i=y_i)=(1-\omega_i)\frac{\Gamma(y_i+\mu_i\alpha^{-1})}{y_i!\Gamma(\mu_i\alpha^{-1})}(1+\alpha)^{-\mu_i\alpha^{-1}}(1+\alpha^{-1})^{-y_i},y_i>0 \tag{52}$$

$0\le\omega_i<1$ ve $\mu_i>0$ olmak üzere ZINB-1 modelinin ortalama ve varyansı Eşitlik 53 ve Eşitlik 54'te verilmektedir (Zulkifli vd., 2012; 4).

$$E(Y_i)=(1-\omega_i)\mu_i \tag{53}$$

$$\mathrm{Var}(Y_i)=(1-\omega_i)\mu_i(1+\alpha+\omega_i\mu_i) \tag{54}$$

ω_i sıfıra eşit olduğunda ZINB-1 modeli, Negatif Binom regresyon modeline dönüşmektedir. ω_i ile birlikte α sıfıra eşit olduğunda ise ZINB-1, Sıfır Değer Ağırlıklı Poisson (ZIP) regresyon modeline dönüşmektedir (Ismail ve Zamani, 2013: 10).

ZINB-1 modeline ait log-olabilirlik fonksiyonu Eşitlik 55'te verilmektedir (Zulkifli vd., 2012; 5).

$$\begin{aligned}\ln L(Y|\alpha,\omega)=&\sum_{y_i=0}\log\left(\omega_i+(1-\omega_i)(1+\alpha)^{-\mu_i\alpha^{-1}}\right)\\&+\sum_{y_i>0}\Big(\log(1-\omega_i)+\log\big(\Gamma(y_i+\mu_i+\alpha^{-1})\big)-\log(y_{i!})-\log\big(\Gamma(\mu_i\alpha^{-1})\big)\\&-\mu_i\alpha^{-1}\log(1+\alpha)-y_i\log(1+\alpha)\Big)\end{aligned} \tag{55}$$

Eşitlik 55'te verilen log-olabilirlik fonksiyonundaki üstel terimlerin maksimize edilmesi oldukça karmaşıktır. Bu yüzden En Çok Olabilirlik fonksiyonu ile parametre tahmini için, EM algoritması veya Newton-Raphson algoritması kullanılmaktadır (Agresti, 2015: 252).

ZINB-2 modeline ait olasılık yoğunluk fonksiyonu Eşitlik 56'da verilmektedir (Peng vd., 2014: 2299).

$$f(Y_i = 0) = \omega_i + (1 - \omega_i)\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu_i}\right)^{\alpha^{-1}}, \ y_i = 0$$

$$f(Y_i = y_i) = (1 - \omega_i)\frac{\Gamma(y_i+\alpha^{-1})}{\Gamma(y_i+1)\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu_i}\right)^{\alpha^{-1}}\left(\frac{\mu_i}{\mu_i+\alpha^{-1}}\right)^{y_i}, \ y_i > 0 \tag{56}$$

Eşitlik 56'da ω_i yapısal sıfırların olma olasılığını, α yayılım parametresini ve μ_i ise negatif binom dağılımının ortalamasını göstermektedir. $0 \leq \omega_i < 1$ ve $\mu_i > 0$ olmak üzere Sıfır Değer Ağırlıklı Negatif Binom regresyon modelinin ortalaması ve varyansı Eşitlik 57 ve Eşitlik 58'de gösterilmektedir (Tüzel ve Sucu, 2012: 26).

$$E(Y_i) = (1 - \omega_i)\mu_i \tag{57}$$

$$V(Y_i) = (1 - \omega_i)\mu_i(1 + \alpha\mu_i + \omega_i\mu_i) \tag{58}$$

ω_i sıfıra eşit olduğunda ZINB-2 modeli, Negatif Binom regresyon modeline dönüşmektedir. α sıfıra eşit olduğunda ise ZINB-2 modeli, Sıfır Değer Ağırlıklı Poisson regresyon modeline dönüşmektedir (Ismail ve Zamani, 2013: 10).

ZINB-2 modeline ait log-olabilirlik fonksiyonu Eşitlik 59'da verilmektedir (Zulkifli vd., 2012: 6).

$$\begin{aligned} \ln L(Y|\omega,\alpha) = & \sum_{y_i=0} \log\left(\omega_i + (1-\omega_i)(1+\alpha\mu_i)^{-\alpha^{-1}}\right) \\ & + \sum_{y_i>0} \Big(\log(1-\omega_i) + \log\big(\Gamma(y_i+\alpha^{-1})\big) - \log(y_i!) - \log\big(\Gamma(\alpha^{-1})\big) \\ & - \alpha^{-1}\log(1+\alpha\mu_i) - y_i\log(1+\alpha^{-1}\mu_i^{-1})\Big) \end{aligned} \tag{59}$$

ZINB-2 modelinin En Çok Olabilirlik fonksiyonu ile parametre tahmini için, EM algoritması veya Newton-Raphson algoritması kullanılmaktadır.

3.4. Genelleştirilmiş Poisson Regresyon Modeli

Genelleştirilmiş Poisson dağılımı hem aşırı hem de az yayılım durumlarına izin veren bir dağılım türüdür. Genelleştirilmiş Poisson Regresyon modeli, Poisson Regresyon modelini iç içe geçiren özel bir durumdur. Rasgele bir Y değişkeni Genelleştirilmiş Poisson dağılımına sahip olduğunda olasılık dağılımı Eşitlik 60'daki gibi gösterilmektedir (Wang ve Famoye, 1997: 277).

$$f(y_i;\mu_i,\alpha) = \left(\frac{\mu_i}{1+\alpha\mu_i}\right)^{y_i} \frac{(1+\alpha y_i)^{y_i-1}}{y_i!} exp\left(-\frac{\mu_i(1+\alpha y_i)}{1+\alpha\mu_i}\right), \quad y_i = 0,1,2,\ldots \tag{60}$$

Genelleştirilmiş Poisson Regresyon modelinin ortalaması ve varyansı Eşitlik 61 ve Eşitlik 62'de gösterilmektedir.

$$E(Y_i) = \mu_i \tag{61}$$

$$V(Y_i) = \mu_i(1+\alpha\mu_i)^2 \tag{62}$$

Eşitlik 61 ve Eşitlik 62'de $\alpha=0$ olduğunda $E(Y_i)=V(Y_i)$ durumu oluşmaktadır. Bu da eşit yayılım özelliğini göstermektedir. Ortalama ve varyansın birbirine eşit olduğu bu durumda Eşitlik 59 Poisson Regresyon modeline dönüşmektedir. $\alpha>0$ olduğunda $V(Y_i)>E(Y_i)$ durumu oluşmaktadır. Bu durumda aşırım yayılım özelliği ortaya çıkmaktadır. Varyansın ortalamadan büyük olduğu bu durumda Genelleştirilmiş Poisson Regresyon modeli aşırı yayılım ile sayma verilerini temsil etmektedir. $\alpha<0$ olduğunda $V(Y_i)<E(Y_i)$ durumu oluşmaktadır. Bu durumda az yayılım özelliği ortaya çıkmaktadır. Varyansın ortalamadan küçük olduğu bu durumda Genelleştirilmiş Poisson Regresyon modeli az yayılım ile sayma verilerini temsil etmektedir.

Genelleştirilmiş Poisson Regresyon Modelinin parametre tahmini için log-olabilirlik fonksiyonu Eşitlik 63'te verilmektedir (Wang ve Famoye, 1997: 277).

$$lnL(\alpha,\beta;y_i) = \sum_{i=1}^{n}\left\{y_i ln\left(\frac{\mu_i}{1+\alpha\mu_i}\right) + (y_i-1)ln(1+\alpha y_i) - \frac{\mu_i(1+\alpha y_i)}{1+\alpha\mu_i} - ln(y_i!)\right\} \tag{63}$$

Genelleştirilmiş Poisson Regresyon Modelinin parametre tahmini Eşitlik 63'teki log-olabilir fonksiyonun kısmi türevi alınarak sıfıra eşitlenmesiyle Eşitlik 64 ve Eşitlik 65 elde edilmektedir (Singh ve Famoye, 1993: 919).

$$\frac{\partial lnL}{\partial \alpha} = \sum_{i=1}^{n} \left\{ \frac{-y_i \mu_i}{1+\alpha \mu_i} + \frac{y_i(y_i-1)}{1+\alpha \mu_i} - \frac{\mu_i(y_i-\mu_i)}{(1+\alpha \mu_i)^2} \right\} = 0 \qquad (64)$$

$$\frac{\partial lnL}{\partial \beta} = \sum_{i=1}^{n} \left\{ \frac{y_i-\mu_i}{\mu_i(1+\alpha \mu_i)^2} \frac{\partial \mu_i}{\partial \beta_j} \right\} = 0 \ , \quad j = 1,2,\dots,k \qquad (65)$$

Eşitlik 65 $\mu_i = \exp(x_i \beta)$ için yeniden düzenlendiğinde Eşitlik 66 ve Eşitlik 67 elde edilmektedir.

$$\frac{\partial lnL}{\partial \beta_j} = \frac{\partial lnL}{\partial \beta_1} = \sum_{i=1}^{n} \frac{y_i-\mu_i}{(1+\alpha \mu_i)^2} = 0 \ , \ j = 1 \qquad (66)$$

$$\frac{\partial lnL}{\partial \beta_j} = \sum_{i=1}^{n} \frac{(y_i-\mu_i)x_i}{(1+\alpha \mu_i)^2} = 0 \ , \ j = 2,3,\dots,k \qquad (67)$$

Eşitlik 66 ve Eşitlik 67, tekrarlı (iteratif) yöntemlerden Newton-Raphson iterasyon yöntemi kullanılarak aynı anda çözülebilmektedir (Singh ve Famoye, 1993: 919).

3.5. Engel (Hurdle) Regresyon Modeli

Engel regresyon modelleri, bağımlı değişkenin sıfır değerini aldığı durumlarda kullanılan sayma verisi regresyon modellerindendir. Mullahy tarafından 1986 yılında, çeşitli içeceklerin günlük tüketimine yönelik yapılan bir çalışmada Engel regresyon modelinin genel hali ortaya konmaktadır (Winkelmann, 2008: 179).

Engel regresyonu iki bölümden oluşan modellerdendir. İlk bölüm bağımlı değişkenin sıfır değerlerini alma olasılıklarının dağılımına ve ikinci bölüm ise sıfır olmayan pozitif değerleri alma olasılıklarının kesilmiş dağılımına dayanmaktadır (Mert, 2016: 256). Engel regresyon modelinin olasılık fonksiyonu Eşitlik 68'de verilmektedir (Winkelmann, 2008: 179).

$$\mathrm{f}(\mathrm{Y}=0) = \mathrm{f}_1(0)$$

$$\mathrm{f}(\mathrm{Y}=\mathrm{k}) = \frac{1-\mathrm{f}_1(0)}{1-\mathrm{f}_2(0)}\mathrm{f}_2(\mathrm{k}) = \theta \mathrm{f}_2(\mathrm{k}),\ \mathrm{k} > 0 \qquad (68)$$

Eşitlik 68'de θ parametresi engeli geçme olasılığını vermektedir. Engel regresyon modelinin beklenen değeri ve varyansı Eşitlik 69 ve Eşitlik 70'de verilmektedir (Saffari vd., 2012: 183).

$$E(Y) = \frac{1-f_1(0)}{1-f_2(0)} \sum_{j=1}^{\infty} k f_2(k) \tag{69}$$

$$V(Y) = \sum_{k=1}^{\infty} k^2 f_2(k)\theta - \left[\theta \sum_{k=1}^{\infty} k f(k)\right]^2 \tag{70}$$

Engel regresyon modelinde parametre tahmini için En Çok Olabilirlik yöntemi kullanılmaktadır. Bu modelin log-olabilirlik fonksiyonu Eşitlik 71'de verilmektedir (Winkelmann, 2008: 181).

$$\mathrm{lnL}(\theta_1, \theta_2) = \sum_{i=1}^{n} \mathrm{d_i} \mathrm{lnf_1}(0; \theta_1) + (1 - \mathrm{d_i})\mathrm{ln}[1 - \mathrm{f_1}(0, \theta_1)] + (1 - \mathrm{d_i})\mathrm{ln}\left[\frac{\mathrm{f_2}(\mathrm{y_i}; \theta_2)}{1 - \mathrm{f_2}(0; \theta_2)}\right] \tag{71}$$

Eşitlik 70'te $d_i = 1 - min(y_i, 1)$ olarak tanımlanmaktadır.

Engel regresyon modelinde bulunan iki bölüm fonksiyonel olarak birbirinden bağımsızdır. İlk bölümde bağımlı değişken sıfır ve diğer değerler olmak üzere iki kategorili olarak lojistik regresyon modeli ile modellenmektedir. İkinci bölümde ise bağımlı değişken, sadece pozitif değerler kesilmiş poisson veya kesilmiş negatif binom dağılımı kullanılarak modellenmektedir. İkinci bölümde kullanılan regresyona göre Engel regresyon modeli, Engel Poisson Regresyonu veya Engel Negatif Binom Regresyonu olmaktadır. Bu iki regresyona karar vermek için bağımlı değişkene ait veri setindeki eşit yayılım, aşırı yayılım ve az yayılım durumlarına bakılmaktadır (Mert, 2016: 257).

3.6. Modelin Değerlendirilmesi ve Seçimi

Katsayı tahmini yapıldıktan sonra ortaya konulan model uyum iyiliği ve katsayı anlamlılığı açısından değerlendirilmelidir. Modelin katsayılarının anlamlılığının ve modelin uyum iyiliğinin değerlendirilmesi süreçleri modelin kullanılabilirliği ve geçerliliği hakkında bilgi vermektedir. Modelin katsayılarının anlamlı olması modelin genel anlamda kullanılabilir bir model olduğu düşüncesini bazen karşılayamamaktadır. Bu sebeple modelin genel anlamlılığı, katsayı anlamlılığı ve uyum iyiliği sınaması belirlenen model için önemlidir. Belirli bir veri seti için birden fazla regresyon modeli mevcut olduğunda bazı uyum iyiliği testleri kullanılmaktadır.

3.6.1. Katsayı Tahminlerinin Değerlendirilmesi

3.6.1.1. Wald χ2 testi

Sayma Verisi Regresyon modellerinde katsayı tahminlerinin anlamlılığının değerlendirilmesi için Wald χ^2 testi kullanılmaktadır. Wald χ^2 test istatistiği, model parametreleriyle ilgili hipotezleri test etmek için kullanılan bir hipotez testidir. Wald χ^2 testi için kurulan hipotezler Eşitlik 72'de verilmektedir:

$$H_0: \beta_i = 0 \qquad H_1: \beta_i \neq 0, \quad i = 1,2,\ldots,n \tag{72}$$

Wald χ^2 test istatistiği genel olarak Eşitlik 73'teki eşitlik ile elde edilmektedir.

$$\chi_w^2 = \left[\frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}\right]^2 \tag{73}$$

Wald testi, χ^2 dağılımına uyduğu için elde edilen Wald test istatistiği, k parametre sayısı ve α anlamlılık düzeyi olmak üzere $\chi^2_{k,1-\alpha}$ tablo değeri ile karşılaştırılmaktadır. Katsayıların anlamlılığı için karar verirken Eşitlik 74 dikkate alınmaktadır (Agresti, 2015: 131).

$$\chi_w^2 > \chi^2_{k,1-\alpha} \tag{74}$$

Eşitlik 74'te hesaplanan Wald Ki-Kare test istatistiğinin Ki-Kare tablo değerinden büyük olduğu durumda karar; sıfır hipotezinin reddi olarak verilmektedir. Kısacası Wald test istatistiğinin tablo değerinden büyük olması durumunda test edilen β katsayısının anlamlı olduğu ifade edilmektedir.

3.6.2. Modelin Uyum İyiliğinin Değerlendirilmesi

İlgilenilen veri seti için birçok regresyon modeli kullanılabilir olduğunda veri setine en iyi uyum sağlayan regresyon modeline karar vermek için bazı uyum iyiliği ölçütleri dikkate alınmaktadır. Sayma verisi regresyon modellerinde normallik varsayımının ihlal edilmesi sebebiyle Pseudo (Yapay) R^2, Pearson χ^2 ve Sapma İstatistiği gibi uyum iyiliği ölçütleri kullanılmaktadır (Winkelmann, 2008: 118).

3.6.2.1. Pseudo (Yapay) R^2 istatistiği

Sayma verisi regresyon modelleri arasında uyum iyiliğini belirlemek için Pseudo R^2 istatistiği kullanılmaktadır. Pseudo R^2 istatistiği, normallik varsayımının ihlal edildiği modeller için kullanılmaktadır. Pseudo R^2 istatistiğini elde etmek için Eşitlik 75 kullanılmaktadır (Cameron ve Trivedi,2013: 191).

$$R_p^2 = 1 - \frac{lnL(y) - \ln(\mu)}{lnL(y) - lnL(\overline{y})} \tag{75}$$

Eşitlik 75’te lnL(y) doymuş modelin log-olabilirliğini, lnL(μ) seçilen modelin log-olabilirliğini ve lnL($\overline{y}$) sadece sabit terimin olduğu modelin log-olabilirliğini ifade etmektedir.

3.6.2.2. Pearson χ^2 istatistiği

Ortalaması μ_i ve varyansı φ_i olan herhangi bir y_i modeli için standart bir uyum iyiliği ölçütü Pearson χ^2 istatistiğidir. Pearson test istatistiği için kurulan hipotezler;

H_0: Model veri setine uyum sağlamaktadır.

H_1: Model veri setine uyum sağlamamaktadır.

Pearson χ^2 test istatistiğini elde etmek için Eşitlik 76 kullanılmaktadır (Madsen ve Thyregod, 2010: 115).

$$\chi_P^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2}{\hat{\varphi}_i} \tag{76}$$

Eşitlik 76 kullanılarak hesaplanan Pearson χ^2 test istatistiği χ^2 tablo değeri ile karşılaştırılmaktadır. Hipotezler için karar verirken Eşitlik 77 dikkate alınmaktadır.

$$\chi_P^2 > \chi_{n-k}^2 \tag{77}$$

Eşitlik 77’deki durum dikkate alındığında hesaplanan test istatistiği değeri tablo değerinden büyük olduğunda sıfır hipotezinin reddedileceği yönünde karar verilmektedir. Sıfır hipotezinin reddedilmesi modelin veri setine uyum sağlamadığını göstermektedir.

3.6.2.3. Sapma istatistiği

Uyum iyiliği ölçülmesinde kullanılan yöntemlerden biri sapma istatistiğidir. Sapma istatistiği, doymuş ve seçilen model arasındaki log-olabilirlik farkının iki katı olarak tanımlanmaktadır. Doymuş model ise, ilgilenilen veri seti ile mümkün olan en karmaşık model olarak tanımlanmaktadır. Seçilen model ise, ilgilenilen modeldir. Sapma istatistiği için Eşitlik 78 kullanılmaktadır (Olsson, 2002: 45).

$$D(y, \hat{\mu}) = 2[lnL(y) - lnL(\hat{\mu})] \tag{78}$$

Eşitlik 78'de lnL(y) doymuş modelin log-olabilirliğini ve $lnL(\hat{\mu})$ ise seçilen modelin log-olabilirliğini göstermektedir. Model doğruysa sapma, örneklem büyüklüğü arttıkça asimptotik olarak χ^2 dağılımına yönelmektedir. $\frac{D(y,\hat{\mu})}{n-p}$ ifadesi ortalama sapma değerini göstermektedir. Ortalama sapma değerinin 1'i aşması durumu, modelin uyumunun iyi olmadığının göstergesidir (Myers vd., 2010: 218).

3.6.3. Model Seçimi

Modelde yer alması gereken bağımsız değişkenlerin belirlenmesi model seçim sürecinin bir yönüdür. Bu yönüyle model seçimi, farklı sayıda bağımsız değişken içeren alternatif modeller arasından birini seçmektir. Model seçiminin bir diğer yönü ise; ilişkinin fonksiyonel şeklinin belirlenmesi üzerinedir. Bu aşamada doğru model farklı fonksiyonel biçimlere sahip olan modellerden seçilmektedir. (Tarı, 2018: 333). Model seçimi için; AIC, BIC, Mallow'un CP Kriteri ve Voung Testi gibi birçok ölçüt bulunmaktadır.

3.6.3.1. Akaike Bilgi Kriteri (AIC)

Akaike Bilgi Kriteri, modele eklenen bağımsız değişkenlerin sebep olduğu uyum iyiliğindeki artışa sınırlama koyarak uyum iyiliğini dengelemeye çalışmaktadır. Akaike Bilgi Kriteri (AIC) değeri Eşitlik 79 kullanılarak elde edilmektedir (Hilbe, 2014: 116).

$$\mathrm{AIC} = -2\ln(\mathrm{L}) + 2\mathrm{k} \tag{79}$$

Eşitlik 79'da ln(L) modelin log-olabilirliğini, k ise sabit terim dahil modeldeki toplam parametre sayısını göstermektedir. Model karşılaştırmalarında daima en küçük AIC değerine sahip model tercih edilmelidir.

3.6.3.2. Bayes Bilgi Kriteri (BIC)

Model seçim ölçütlerinden bir diğeri Bayes Bilgi Kriteridir. Bayes Bilgi Kriteri, Akaike Bilgi Kriteri ile benzerlik göstermektedir. Bayes Bilgi Kriteri değeri Eşitlik 80 kullanılarak elde edilmektedir (Hilbe, 2014: 119).

$$BIC = -2\ln(L) + kln(n) \tag{80}$$

Eşitlik 80'de ln(L) modelin log-olabilirliğini, k sabit terim dahil modeldeki toplam parametre sayısını ve n gözlem sayısını göstermektedir. Model karşılaştırmalarında en küçük BIC değerine sahip model tercih edilmelidir.

3.6.3.3. Mallow'un C_p Kriteri

Mallow (1973) tarafından önerilen C_p kriteri, farklı parametre sayılarına sahip alternatif modelleri karşılaştırırken en uygun modelin seçimi için kullanılmaktadır. Eşitlik 81'de C_p istatistiğinin formülü verilmektedir (Claeskens ve Hjort, 2008: 107).

$$C_p = \frac{AKT_k}{\hat{\sigma}^2} - n + 2k \tag{81}$$

Eşitlik 81'de AKT_k k parametreli modelin artık kareler toplamını, $\hat{\sigma}^2$ tam modelden hesaplanan artık kareler varyansını, n gözlem sayısını ve k modeldeki parametre sayısını göstermektedir. Seçim yapılacak modeller arasında hangisinin C_p istatistiği o modelin parametre sayısına yakınsa o model en iyi model olarak karar verilmektedir.

4. BÖLÜM

4. UYGULAMA

4.1. Veri seti ve Tanımlayıcı İstatistikler

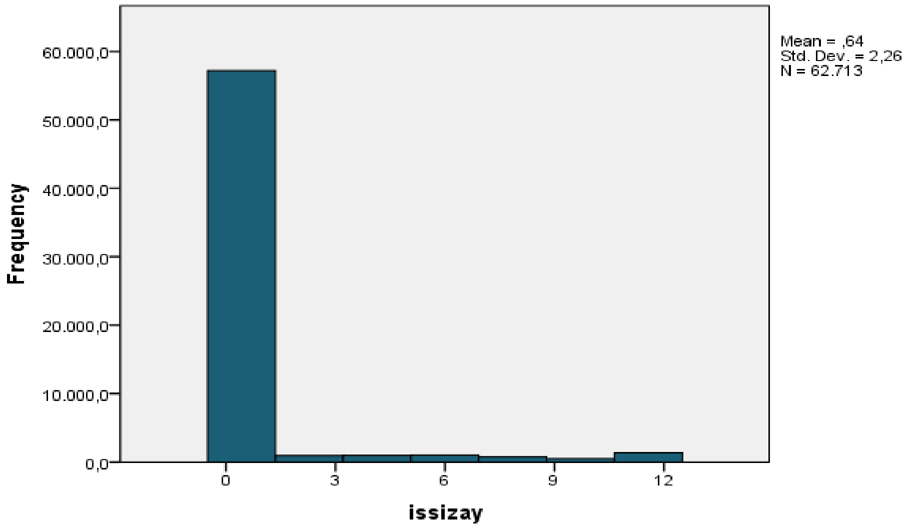
Bu tez çalışmasında Türkiye İstatistik Kurumu'ndan alınan 2019 Gelir ve Yaşam Koşulları Araştırması'na (GYKA) ait anket verileri kullanılmaktadır. Çalışmada birincil amaç olarak bireylerin 2018 yılında işsiz kaldıkları ay sayısı verilerine Poisson regresyon, Negatif Binom regresyon, Genelleştirilmiş Poisson regresyon, Sıfır Değer Ağırlıklı Negatif Binom regresyon modelleri kullanılarak en uygun model belirlenmeye çalışılmıştır.

İkincil olarak, 2019 yılı GYKA'da işsiz kalınan ay sayısına etki eden sosyo-demografik faktörlerin neler olduğunu belirlemek amaçlanmıştır. GYKA'da 'Ferdin 2018 yılında işsiz olarak geçirdiği ay sayısı kaçtır?' sorusu 15 yaş ve üzeri 62713 kişi tarafından cevaplanmıştır. Bu bireylerin 2018 yılında işsiz kaldıkları ay sayısına etki ettiği düşünülen değişkenler Tablo 4.1'de verilmektedir.

Tablo 4.1. Analizde kullanılan bağımsız değişkenler ve düzeyleri

Bağımsız Değişkenler	**Açıklamalar**
X_1: Cinsiyet	1: Erkek 2: Kadın
X_2:Medeni Durum	1: Evli 2: Hiç Evlenmedi 3: Eşi Öldü 4: Boşandı
X_3:Eğitim Durumu	0: Okuryazar olmayan 1: Bir okul bitirmedi 2: İlkokul 3: İlköğretim 4: Ortaokul ve dengi 5: Genel lise 6: Mesleki veya teknik lise 7: Yüksekokul 8: Fakülte 9: Yüksek Lisans 10: Doktora
X_4:Genel Sağlık	1: Çok İyi 2: İyi 3: Orta 4: Kötü 5: Çok Kötü
X_5: Kronik Hastalık	1: Evet 2: Hayır
X_6: Yaş	Ferdin yaşı

Analizde kullanılan yaş değişkeni dışındaki diğer bağımsız değişkenler kategoriktir. İşsiz kalınan ay sayısı değişkenine ait dağılım grafiği Şekil 4.1'de verilmektedir.



Şekil 4.1. İşsiz kalınan ay sayısı değişkenine ait dağılım grafiği

Şekil 4.1'de verilen birikimli dağılım grafiği incelendiğinde dağılımın sağa çarpık olduğu görülmektedir. Bağımlı değişkenin sayıma dayalı olarak elde edildiği ve bu değişkenin sağa çarpık bir dağılım gösterdiği durumda sayma verisi regresyon modellerinden uygun olan model kullanılmaktadır.

Analizde kullanılan bağımsız değişkenlerden yaş değişkeni dışındaki değişkenler kategorik yapıdadır. Tablo 4.2.'de bu değişkenlere ait frekans ve yüzde değerleri verilmektedir.

Tablo 4.2. Bağımsız değişkenlere ait frekans tablosu

Bağımsız Değişken	Frekans	Yüzde (%)
Cinsiyet		
1: Erkek	30215	48,18
2: Kadın	32498	51,82
Medeni Durum		
1: Evli	41943	66,88
2: Hiç Evlenmedi	14942	23,83
3: Eşi Öldü	4035	6,43
4: Boşandı	1793	2,86
Eğitim Durumu		
0: Okuryazar olmayan	6472	10,32
1: Bir okul bitirmedi	4131	6,59
2: İlkokul	18880	30,11
3: İlköğretim	3760	6,00
4: Ortaokul ve dengi	8908	14,20
5: Genel lise	6239	9,95
6: Mesleki veya teknik lise	4921	7,85
7: Yüksekokul	3082	4,91
8: Fakülte	5684	9,06
9: Yüksek Lisans	495	0,79
10: Doktora	141	0,22
Genel Sağlık		
1: Çok İyi	4737	7,55
2: İyi	35988	57,39
3: Orta	14653	23,37
4: Kötü	6415	10,23
5: Çok Kötü	920	1,47
Kronik Hastalık		
1: Evet	23329	37,20
2: Hayır	39384	62,80

Tablo 4.2 incelendiğinde GYKA'ya katılan bireylerin %48,18'inin erkek, %51,82'sinin kadın olduğu tespit edilmiştir. Araştırmaya katılan bireylerin %66,88'i evli, %23,83'ü hiç evlenmemiş, %6,43'ünün eşi ölmüş ve %2,86'sı boşanmış oldukları görülmektedir. Analizde kullanılan eğitim durumu değişkeni incelendiğinde araştırmaya katılan bireylerin %10,32'sinin okuryazar olmadığı, %6,59'unun bir okul bitirmediği, %30,11'inin ilkokul mezunu, %6'sının ilköğretim mezunu, %14,2'sinin ortaokul ve dengi bir okul mezunu, %9,95'inin genel lise mezunu, %7,85'inin mesleki veya teknik lise mezunu, %4,91'inin yüksekokul mezunu, %9,06'sının fakülte mezunu, %0,79'unun yüksek lisans mezunu ve %0,22'sinin ise doktora mezunu oldukları tespit edilmiştir. Araştırmaya katılan bireylere genel sağlık durumlarına yönelik soru sorulduğunda bireylerin %7,55 çok iyi olduğunu, %57,39'u iyi olduğunu, %23,37'si orta düzeyde olduğunu, %10,23'ü kötü olduğunu ve %1,47'si ise çok kötü olduğunu belirmiştir. Kronik hastalık değişkeni incelendiğinde "Kronik bir hastalığınız var mı?" sorusuna araştırmaya katılan bireylerin %37,2'si evet, %62,8'i hayır yanıtını verdiği tespit edilmiştir. Bağımlı değişken olan işsiz

kalınan ay sayısı ve bağımsız değişken olan yaş değişkenine ait tanımlayıcı istatistikler Tablo 4.3'te verilmektedir.

Tablo 4.3. İşsiz kalınan ay ve yaş değişkenine ait tanımlayıcı istatistikler

Değişkenler	Gözlem Sayısı	Ortalama	Standart Sapma	Min-Max	Medyan
İşsiz kalınan ay	62713	0,6357055	2,260344	0-12	0
Yaş	62713	42,32534	17,78856	15-109	41

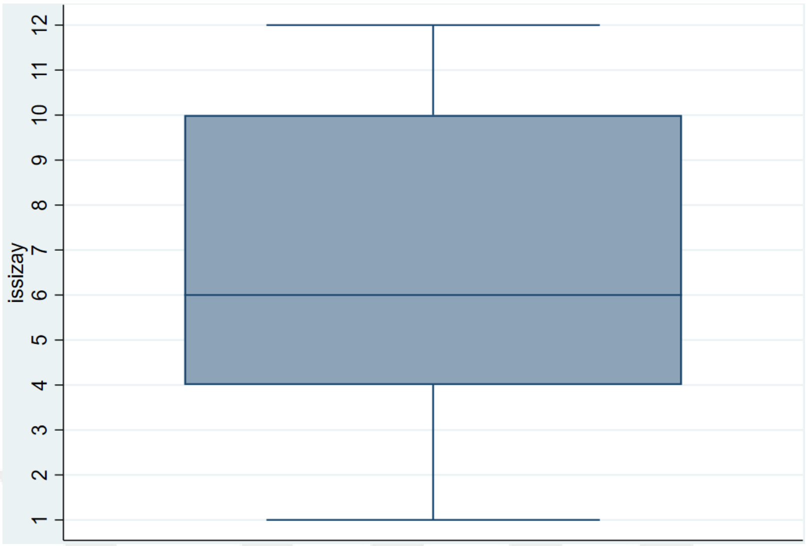
Tablo 4.3 incelendiğinde işsiz kalınan ay sayılarının ortalaması 0,636 ve standart sapması 2,260 olarak hesaplanmıştır. İşsiz kalınan ay sayısı değişkeninin medyan (ortanca) değeri 0 ve aralığı ise 0 ile 12 arasında değişim gösterdiği tespit edilmiştir. Yaş değişkenine ait istatistikler incelendiğinde ortalamasının 42,325 ve standart sapmasının 17,789 olarak hesaplandığı görülmektedir. Yaş değişkeninin medyan değeri 41 ve aralığı 15 ile 109 yaş arasında değişiklik gösterdiği tespit edilmiştir. Bağımlı değişken olan işsiz kalınan ay sayısı değişkenine ait varyans değeri 5,109 olarak hesaplanmıştır. İşsiz kalınan ay sayısı değişkeninin varyansının ortalamasından büyük olması nedeniyle verilerde aşırı yayılım söz konusudur. Ayrıca bağımlı değişkende sıfır değerini alan gözlemlerin çok fazla olması ortalamayı çok düşüreceği için Poisson regresyonda bozulmalara neden olabilmektedir. Bağımlı değişkene ait frekans tablosu Tablo 4.4'te verilmektedir.

Tablo 4. 4. İşsiz kalınan ay sayısı frekans tablosu

İşsiz Kalınan Ay Sayısı	Frekans	Yüzde
0	56824	90,61
1	365	0,58
2	455	0,73
3	460	0,73
4	515	0,82
5	438	0,70
6	1004	1,60
7	397	0,63
8	365	0,58
9	308	0,49
10	206	0,33
11	105	0,17
12	1271	2,03

Tablo 4.4 incelendiğinde işsiz kalınan ay sayısı sıfır olan yani 2018 yılında işsiz kalmayan bireyler %90,61 oranındadır. Sıfır değerini alan gözlemler çok fazladır. Araştırmaya katılan bireylerin %90,61'inin son bir yıl içerinde işsiz kalmadıkları sonucuna ulaşılmıştır. Ankete katılan bireylerden 5889 kişinin ise 2018 yılında 1 ay ve üzeri işsiz

kaldıkları belirlenmiştir. Şekil 4.2'deki grafik dağılımın sağa çarpık olduğunun bir göstergesidir.



Şekil 4.2. 2018 yılında 1 ay ve üzeri işsiz kalınan ay sayısı için dağılım grafiği

Şekil 4.2 incelendiğinde en küçük değerin 1 ve en büyük değerin ise 12 olduğu tespit edilmiştir. İlk %25'lik kısmı kapsayan Q_1 yani 1. çeyrekliğin 4, %50'lik kısmı kapsayan Q_2 yani 2. çeyrekliğin 6 ve %75'lik kısmı kapsayan Q_3 yani 3. çeyrekliğin 10 olduğu sonucuna ulaşılmıştır. 2. çeyreklik aynı zamanda medyan değerini vermektedir. 2018 yılında 1 ay ve üzeri işsiz kalan bireylere ilişkin 1. çeyreklik (Q_1) 4, 2. çeyreklik (Q_2) 6 ve 3. çeyreklik (Q_3) ise 10 olarak hesaplanmıştır. Burada dağılımın çarpıklığı için $Q_3 - Q_2 > Q_2 - Q_1$ formülü sağlanmaktadır. Bu durumda hem $Q_3 - Q_2 > Q_2 - Q_1$ formülü ile hem de box-plot grafiği ile dağılımın sağa çarpık olduğu sonucuna ulaşılmıştır.

4.2. İşsiz Kalınan Ay Sayısı Verileri İçin Modeller ve Tahmin Sonuçları

Bireylerin işsiz kaldıkları ay sayıları için uygulamada kullanılan bağımsız değişkenler belirlenerek model tüm olası altküme yöntemi ile belirlenmiştir. Tüm olası altküme yöntemine göre Poisson regresyon modeli sonuçları Tablo 4.5'te verilmektedir.

Tablo 4.5. Tüm olası altküme yöntemine göre Poisson regresyon sonuçları

Model	AIC	LLModel	Sapma	Sapma SD	Yayılım Parametresi
X_1	221911,5	-110953,7	200721,4	62711	3,200737
X_2	202594,8	-101293,4	181400,7	62709	2,892738
X_3	212685,8	-106331,9	191477,8	62702	3,053775
X_4	217878,9	-108934,5	196682,9	62708	3,136488
X_5	217632,2	-108814,1	196442,1	62711	3,132498
X_6	221918,9	-110957,4	200728,8	62711	3,200855
X_1,X_2	202594,6	-101292,3	181398,6	62708	2,892751
X_1,X_3	212683,8	-106329,9	191473,8	62701	3,05376
X_1,X_4	217876,6	-108932,3	196678,6	62707	3,13647
X_1,X_5	217629,2	-108811,6	196437,1	62710	3,132468
X_1,X_6	221913,5	-110953,7	200721,4	62710	3,200788
X_2,X_3	197825,1	-98898,54	176611	62699	2,816807
X_2,X_4	201953,8	-100968,9	180751,8	62705	2,882574
X_2,X_5	202005	-100997,5	180809	62708	2,883348
X_2,X_6	202596,8	-101293,4	181400,7	62708	2,892784
X_3,X_4	211842,8	-105906,4	190626,8	62698	3,040397
X_3,X_5	211416,2	-105696,1	190206,2	62701	3,033543
X_3,X_6	212687,8	-106331,9	191477,8	62701	3,053824
X_4,X_5	217053,5	-108520,7	195855,4	62707	3,123342
X_4,X_6	217880,8	-108934,4	196682,8	62707	3,136537
X_5,X_6	217634,1	-108814,1	196442,1	62710	3,132548
X_1,X_2,X_3	199894,1	-99933,06	178680,1	62699	2,849808
X_1,X_2,X_4	201953,8	-100967,9	180749,8	62704	2,882588
X_1,X_2,X_5	202004,7	-100996,3	180806,6	62707	2,883356
X_1,X_2,X_6	202596,6	-101292,3	181398,6	62707	2,892797
X_1,X_3,X_4	211841,9	-105905	190623,9	62697	3,040399
X_1,X_3,X_5	211414,9	-105694,5	190202,9	62700	3,033539
X_1,X_3,X_6	212685,8	-106329,9	191473,8	62700	3,053808
X_1,X_4,X_5	217051,1	-108518,6	195851,1	62706	3,123323
X_1,X_4,X_6	217878,5	-108932,2	196678,5	62706	3,136518
X_1,X_5,X_6	217631,1	-108811,6	196437,1	62709	3,132518
X_2,X_3,X_4	197662,5	-98813,25	176440,5	62695	2,814267
X_2,X_3,X_5	197595	-98782,48	176378,9	62698	2,81315
X_2,X_3,X_6	197827,1	-98898,53	176611	62698	2,816852
X_2,X_4,X_5	201735	-100858,5	180530,9	62704	2,879097
X_2,X_4,X_6	201955,8	-100968,9	180751,7	62704	2,882618
X_2,X_5,X_6	202007	-100997,5	180809	62707	2,883394
X_3,X_4,X_5	211314,3	-105641,1	190096,3	62697	3,031984
X_3,X_4,X_6	211844,8	-105906,4	190626,7	62697	3,040444
X_3,X_5,X_6	211418,2	-105696,1	190206,1	62700	3,03359
X_4,X_5,X_6	217055,4	-108520,7	195855,4	62706	3,123392
X_1,X_2,X_3,X_4	197663,2	-98812,58	176439,1	62694	2,81429
X_1,X_2,X_3,X_5	197595,5	-98781,73	176377,4	62697	2,813171
X_1,X_2,X_3,X_6	197827,6	-98897,78	176609,5	62697	2,816873
X_1,X_2,X_4,X_5	201734,6	-100857,3	180528,5	62703	2,879105
X_1,X_2,X_4,X_6	201955,8	-100967,9	180749,8	62703	2,882634
X_1,X_2,X_5,X_6	202006,7	-100996,3	180806,6	62706	2,883402
X_1,X_3,X_4,X_5	211313,3	-105639,6	190093,2	62696	3,031983
X_1,X_3,X_4,X_6	211843,8	-105904,9	190623,8	62696	3,040446

Tablo 4.6. Tüm olası altküme yöntemine göre Poisson regresyon sonuçları (devamı)

X_1,X_3,X_5,X_6	211416,9	-105694,4	190202,8	62699	3,033586
X_1,X_4,X_5,X_6	217053	-108518,5	195851	62705	3,123371
X_2,X_3,X_4,X_5	197467,7	-98714,83	176243,6	62694	2,811172
X_2,X_3,X_4,X_6	197664,5	-98813,23	176440,4	62694	2,814311
X_2,X_3,X_5,X_6	197597	-98782,48	176378,9	62697	2,813195
X_2,X_4,X_5,X_6	201736,9	-100858,4	180530,8	62703	2,879141
X_3,X_4,X_5,X_6	211316,2	-105641,1	190096,2	62696	3,032031
X_1,X_2,X_3,X_4,X_5	197468,1	-98714,03	176242	62693	2,811191
X_1,X_2,X_3,X_4,X_6	197665,1	-98812,57	176439,1	62693	2,814335
X_1,X_2,X_3,X_5,X_6	197597,5	-98781,73	176377,4	62696	2,813216
X_1,X_2,X_4,X_5,X_6	201736,5	-100857,3	180528,5	62702	2,879151
X_1,X_3,X_4,X_5,X_6	211315,2	-105639,6	190093,1	62695	3,03203
X_2,X_3,X_4,X_5,X_6	197469,6	-98714,81	176243,6	62693	2,811216
X_1,X_2,X_3,X_4,X_5,X_6	197470	-98714,01	176242	62692	2,811236

Tablo 4.5 ve Tablo 4.6'daki modeller Poisson regresyon yöntemiyle tahmine edilmiştir. Poisson regresyon ile tahmin edilen tüm modellerde yayılım parametresinin 1'den büyük olduğu sonucuna ulaşılmıştır. Analiz için en uygun model AIC değeri en küçük olan modeldir. AIC değeri en küçük olan model X_2,X_3,X_4,X_5 bağımsız değişkenlerinden oluşan modelin olduğu tespit edilmiştir. Fakat modele ait yayılım parametresi 1'den büyük olduğu için bu modelin aşırı yayılım özelliğine sahip olduğu sonucuna ulaşılmıştır. Bu durumda Poisson regresyon modelinin veri setine uyum sağlamadığı tespit edilmiştir.

Aşırı yayılım nedeniyle analizde kullanılabilecek regresyon modellerinden biri NBRM'dir. Tüm olası altküme yöntemi ile NBRM için tahmin sonuçları Tablo 4.7 ve Tablo 4.8'de verilmektedir.

Tablo 4.7. Tüm olası alküme yöntemine göre NBRM sonuçları

Model	**AIC**	**LLModel**	**Sapma**	**Sapma SD**	**Yayılım Parametresi**
X_1	76098,31	-38046,16	76092,312	62710	1,2134
X_2	74862,75	-37426,38	74852,750	62708	1,193671
X_3	75536,53	-37756,26	75512,528	62701	1,204327
X_4	75869,78	-37928,89	75857,781	62707	1,209718
X_5	75858,19	-37926,09	75852,189	62710	1,209571
X_6	76098,68	-38046,34	76092,683	62710	1,213406
X_1,X_2	74863,2	-37425,6	74851,200	62707	1,193666
X_1,X_3	75538,31	-37756,16	75512,311	62700	1,204343
X_1,X_4	75871,49	-37928,74	75857,488	62706	1,209732
X_1,X_5	75859,83	-37925,92	75851,834	62709	1,209584
X_1,X_6	76100,31	-38046,16	76092,311	62709	1,213419
X_2,X_3	74656,94	-37313,47	74626,940	62698	1,19026
X_2,X_4	74808,4	-37395,2	74790,405	62704	1,192753

Tablo 4.8. Tüm olası altküme yöntemine göre NBRM sonuçları (devamı)

X_2,X_5	74791,56	-37389,78	74779,557	62707	1,192523
X_2,X_6	74864,41	-37426,21	74852,414	62707	1,193685
X_3,X_4	75453,19	-37710,59	75421,187	62697	1,202947
X_3,X_5	75418,95	-37696,48	75392,950	62700	1,202439
X_3,X_6	75538,25	-37756,13	75512,252	62700	1,204342
X_4,X_5	75823,5	-37904,75	75809,502	62706	1,208967
X_4,X_6	75871,74	-37928,87	75857,736	62706	1,209736
X_5,X_6	75860,18	-37926,09	75852,183	62709	1,20959
X_1,X_2,X_3	74836,44	-37403,22	74806,441	62698	1,193123
X_1,X_2,X_4	74809,2	-37394,6	74789,199	62703	1,192753
X_1,X_2,X_5	74792,36	-37389,18	74778,356	62706	1,192523
X_1,X_2,X_6	74864,81	-37425,4	74850,805	62706	1,193678
X_1,X_3,X_4	75455,03	-37710,51	75421,026	62696	1,202964
X_1,X_3,X_5	75420,74	-37696,37	75392,740	62699	1,202455
X_1,X_3,X_6	75540,06	-37756,03	75512,058	62699	1,204358
X_1,X_4,X_5	75825,19	-37904,59	75809,189	62705	1,208982
X_1,X_4,X_6	75873,43	-37928,72	75857,431	62705	1,209751
X_1,X_5,X_6	75861,83	-37925,92	75851,832	62708	1,209604
X_2,X_3,X_4	74631,32	-37296,66	74593,316	62694	1,1898
X_2,X_3,X_5	74608,6	-37288,3	74576,603	62697	1,189476
X_2,X_3,X_6	74658,91	-37313,46	74626,912	62697	1,190279
X_2,X_4,X_5	74781,87	-37380,94	74761,875	62703	1,192317
X_2,X_4,X_6	74809,96	-37394,98	74789,959	62703	1,192765
X_2,X_5,X_6	74793,21	-37389,61	74779,213	62706	1,192537
X_3,X_4,X_5	75411,98	-37688,99	75377,983	62696	1,202277
X_3,X_4,X_6	75455,15	-37710,58	75421,150	62696	1,202966
X_3,X_5,X_6	75420,83	-37696,41	75392,826	62699	1,202457
X_4,X_5,X_6	75825,49	-37904,75	75809,491	62705	1,208986
X_1,X_2,X_3,X_4	74632,51	-37296,25	74592,508	62693	1,189806
X_1,X_2,X_3,X_5	74609,79	-37287,9	74575,792	62696	1,189482
X_1,X_2,X_3,X_6	74659,91	-37312,96	74625,910	62696	1,190282
X_1,X_2,X_4,X_5	74782,73	-37380,37	74760,733	62702	1,192318
X_1,X_2,X_4,X_6	74810,69	-37394,35	74788,691	62702	1,192764
X_1,X_2,X_5,X_6	74793,96	-37388,98	74777,956	62705	1,192536
X_1,X_3,X_4,X_5	75413,78	-37688,89	75377,784	62695	1,202293
X_1,X_3,X_4,X_6	75457	-37710,5	75420,996	62695	1,202983
X_1,X_3,X_5,X_6	75422,63	-37696,32	75392,632	62698	1,202473
X_1,X_4,X_5,X_6	75827,17	-37904,59	75809,171	62704	1,209001
X_2,X_3,X_4,X_5	74605,79	-37282,9	74565,790	62693	1,18938
X_2,X_3,X_4,X_6	74633,23	-37296,62	74593,234	62693	1,189818
X_2,X_3,X_5,X_6	74610,55	-37288,27	74576,548	62696	1,189494
X_2,X_4,X_5,X_6	74783,51	-37380,75	74761,508	62702	1,19233
X_3,X_4,X_5,X_6	75413,91	-37688,95	75377,908	62695	1,202295
X_1,X_2,X_3,X_4,X_5	74606,99	-37282,5	74564,995	62692	1,189386
X_1,X_2,X_3,X_4,X_6	74634,4	-37296,2	74592,401	62692	1,189823
X_1,X_2,X_3,X_5,X_6	74611,72	-37287,86	74575,716	62695	1,1895
X_1,X_2,X_4,X_5,X_6	74784,31	-37380,15	74760,309	62701	1,19233
X_1,X_3,X_4,X_5,X_6	75415,72	-37688,86	75377,722	62694	1,202312
X_2,X_3,X_4,X_5,X_6	74607,74	-37282,87	74565,739	62692	1,189398
X_1,X_2,X_3,X_4,X_5,X_6	74608,92	-37282,46	74564,923	62691	1,189404

Tablo 4.7 ve Tablo 4.8 incelendiğinde AIC değeri en küçük olan model X_2,X_3,X_4,X_5 bağımsız değişkenlerinden oluşan modelin olduğu tespit edilmiştir. Yayılım parametresi ise tüm modeller için 1'e yakın olarak hesaplanmıştır. NBRM için en uygun model tahmin edilmiş ve elde edilen sonuçlar Tablo 4.9'de gösterilmiştir.

Tablo 4.9. NBRM tahmin sonuçları

Log likelihood = -37282,895		Gözlem Sayısı = 62,713 LR $\chi^2(19)$ = 1526.90 Prob > χ^2 = 0.0000		
Bağımsız Değişkenler	**Katsayı (β)**	**Standart Hata**	**IRR**	**Prob Değeri**
Sabit	-2,214037	0,2158581	,1238757	0,000***
Medenihal_2	1,140635	0,0542094	3,128754	0,000***
Medenihal_3	-2,412633	0,1346922	0,0895791	0,000***
Medenihal_4	0,8684653	0,1239154	2,38325	0,000***
Egitim_1	1,059566	0,1138548	2,886612	0,000***
Egitim_2	1,006765	0,0872917	2,74314	0,000***
Egitim_3	1,451455	0,1164089	4,2845	0,000***
Egitim_4	0,7781729	0,0986022	2,183891	0,000***
Egitim_5	0,9714161	0,1042844	2,647729	0,000***
Egitim_6	0,8843744	0,1105027	2,424227	0,000***
Egitim_7	1,145019	0,1247469	3,145577	0,000***
Egitim_8	1,240686	0,1073654	3,457982	0,000***
Egitim_9	0,8534772	0,2479107	2,350807	0,001***
Egitim_10	0,9103607	0,446613	2,50139	0,040**
Genelsaglik_2	-0,0383133	0,0804513	0,9624114	0,634
Genelsaglik_3	0,0846468	0,1021619	1,088333	0,407
Genelsaglik_4	-0,1641684	0,1234567	0,8485991	0,184
Genelsaglik_5	-0,2149804	0,2269772	0,8065573	0,343
Kronikhastalik_2	0,348638	0,0657703		0,000***
α(Yayılım parametresi)	1,1893367			
α=0 için LR test: chibar2(01) = 1,2e+05 Prob > chibar2 = 0,000				
*** %1 hata payıyla istatistiksel olarak anlamlılığı göstermektedir.				

Tablo 4.9 incelendiğinde genel sağlık değişkenleri dışındaki tüm bağımsız değişkenlerin anlamlı olduğu sonucuna ulaşılmıştır. α değeri yaklaşık 1,19 olarak hesaplanmıştır. LR test istatistiği eşit yayılım hipotezini aşırı yayılım alternatif hipotezine karşı test etmektedir. Tablo 4.9'da LR test istatistiğine ait olasılık değeri $p<0,05$ olduğu için sıfır hipotezi reddedilmektedir. Bu durumda aşırı yayılım olduğu tespit edilmiştir. NBRM, kurulan model için uygun modeller arasındadır. Ancak, bağımlı değişkene ait frekans tablosu incelendiğinde sıfır değerlerinin çoğunlukta olduğu görülmektedir. Bağımlı değişkenin sıfır değerler aldığı durumda kullanılan Sıfır Değer Ağırlıklı Negatif Binom ve Genelleştirilmiş Poisson regresyon modelleri de tahmin edilip her üç model için uyum iyiliği göstergelerine göre karar vermek gerekmektedir. Bu aşamada analize Sıfır Değer Ağırlıklı Negatif Binom Regresyon modeli tahmin edilerek devam edilmiştir. Tablo 4.10'da tüm olası altküme

yöntemi ile Sıfır Değer Ağırlıklı Negatif Binom tahmini için uyum iyiliği sonuçları verilmektedir.

Tablo 4.10. Tüm olası altküme yöntemine göre Sıfır Değer Ağırlıklı Negatif Binom regresyon sonuçları

Model	AIC	LLModel	Sapma	Sapma SD	Yayılım Parametresi
X_1	70530,987	-35260,493	70520,987	62708	1,124593
X_2	67932,322	-33957,161	67914,322	62704	1,083094
X_3	69131,340	-34542,670	69085,340	62690	1,102015
X_4	69854,124	-34916,062	69832,124	62702	1,113714
X_5	69842,356	-34916,178	69832,356	62708	1,113612
X_6	-	-	-	-	-
X_1,X_2	67935,715	-33956,857	67913,715	62702	1,083119
X_1,X_3	69134,587	-34542,294	69084,587	62688	1,102039
X_1,X_4	69857,595	-34915,798	69831,595	62700	1,113742
X_1,X_5	69845,738	-34915,869	69831,738	62706	1,113637
X_1,X_6	70532,909	-35260,455	70520,909	62707	1,12461
X_2,X_3	67142,951	-33542,475	67084,951	62684	1,070209
X_2,X_4	67772,229	-33869,115	67738,229	62696	1,080423
X_2,X_5	67786,494	-33882,247	67764,494	62702	1,080739
X_2,X_6	67934,265	-33957,132	67914,265	62703	1,08311
X_3,X_4	68947,650	-34442,825	68885,650	62682	1,09897
X_3,X_5	68898,751	-34424,375	68848,751	62688	1,098276
X_3,X_6	69133,337	-34542,668	69085,337	62689	1,102033
X_4,X_5	69733,852	-34853,926	69707,852	62700	1,111768
X_4,X_6	69856,054	-34916,027	69832,054	62701	1,113731
X_5,X_6	69844,284	-34916,142	69832,284	62707	1,113628
X_1,X_2,X_3	67569,707	-33755,854	67511,707	62684	1,077016
X_1,X_2,X_4	67775,625	-33868,812	67737,625	62694	1,080448
X_1,X_2,X_5	67789,864	-33881,932	67763,864	62700	1,080763
X_1,X_2,X_6	67937,644	-33956,822	67913,644	62701	1,083135
X_1,X_3,X_4	68951,002	-34442,501	68885,002	62680	1,098995
X_1,X_3,X_5	68902,077	-34424,039	68848,077	62686	1,098301
X_1,X_3,X_6	69136,586	-34542,293	69084,586	62687	1,102056
X_1,X_4,X_5	69737,317	-34853,659	69707,317	62698	1,111795
X_1,X_4,X_6	69859,518	-34915,759	69831,518	62699	1,113758
X_1,X_5,X_6	69847,659	-34915,829	69831,659	62705	1,113654
X_2,X_3,X_4	67076,819	-33501,410	67002,819	62676	1,069035
X_2,X_3,X_5	67071,340	-33504,670	67009,340	62682	1,069036
X_2,X_3,X_6	67144,943	-33542,472	67084,943	62683	1,070225
X_2,X_4,X_5	67734,642	-33848,321	67696,642	62694	1,079795
X_2,X_4,X_6	67774,194	-33869,097	67738,194	62695	1,08044

Tablo 4.11. Tüm olası altküme yöntemine göre Sıfır Değer Ağırlıklı Negatif Binom regresyon sonuçları (devamı)

X_2,X_5,X_6	67788,427	-33882,214	67764,427	62701	1,080755
X_3,X_4,X_5	68869,133	-34401,566	68803,133	62680	1,097689
X_3,X_4,X_6	68949,636	-34442,818	68885,636	62681	1,098987
X_3,X_5,X_6	68900,747	-34424,374	68848,747	62687	1,098294
X_4,X_5,X_6	69735,791	-34853,895	69707,791	62699	1,111785
X_1,X_2,X_3,X_4	67079,956	-33500,978	67001,956	62674	1,069055
X_1,X_2,X_3,X_5	67074,499	-33504,250	67008,499	62680	1,069057
X_1,X_2,X_3,X_6	67148,121	-33542,061	67084,121	62681	1,070246
X_1,X_2,X_4,X_5	67737,996	-33847,998	67695,996	62692	1,079819
X_1,X_2,X_4,X_6	67777,578	-33868,789	67737,578	62693	1,080465
X_1,X_2,X_5,X_6	67791,784	-33881,892	67763,784	62699	1,080779
X_1,X_3,X_4,X_5	68872,467	-34401,234	68802,467	62678	1,097713
X_1,X_3,X_4,X_6	68952,994	-34442,497	68884,994	62679	1,099012
X_1,X_3,X_5,X_6	68904,076	-34424,038	68848,076	62685	1,098318
X_1,X_4,X_5,X_6	69739,248	-34853,624	69707,248	62697	1,111812
X_2,X_3,X_4,X_5	67044,926	-33483,463	66966,926	62674	1,068496
X_2,X_3,X_4,X_6	67078,819	-33501,410	67002,819	62675	1,069052
X_2,X_3,X_5,X_6	67073,332	-33504,666	67009,332	62681	1,069053
X_2,X_4,X_5,X_6	67736,607	-33848,304	67696,607	62693	1,079811
X_3,X_4,X_5,X_6	68871,116	-34401,558	68803,116	62679	1,097706
X_1,X_2,X_3,X_4,X_5	67048,039	-33483,019	66966,039	62672	1,068516
X_1,X_2,X_3,X_4,X_6	67081,955	-33500,978	67001,955	62673	1,069072
X_1,X_2,X_3,X_5,X_6	67076,484	-33504,242	67008,484	62679	1,069074
X_1,X_2,X_4,X_5,X_6	67739,950	-33847,975	67695,950	62691	1,079835
X_1,X_3,X_4,X_5,X_6	68874,457	-34401,228	68802,457	62677	1,097731
X_2,X_3,X_4,X_5,X_6	67046,926	-33483,463	66966,926	62673	1,068513
X_1,X_2,X_3,X_4,X_5,X_6	67050,038	-33483,019	66966,038	62671	1,068533

Tablo 4.10 ve Tablo 4.11 incelendiğinde tüm olası altküme yöntemine dayalı olarak belirlenen modellerden AIC değeri en düşük olan X_2,X_3,X_4,X_5 bağımsız değişkenlerini içeren modeldir. Bu model, Sıfır Değer Ağırlıklı Negatif Binom regresyon ile tahmin edilmiş ve Tablo 4.12'deki sonuçlar elde edilmiştir.

Tablo 4.12. Sıfır Değer Ağırlıklı Negatif Binom regresyon modeli tahmin sonuçları

Log likelihood = -33483,46		Gözlem Sayısı = 62,713 LR $\chi^2(18)$ = 253,69 Prob > χ^2 = 0,0000		
Bağımsız Değişkenler	**Katsayı (β)**	**Standart Hata**	**IRR**	**Prob Değeri**
Sabit	1,80139	0,0617103	6,058064	0,000***
Medenihal_2	0,2115631	0,0174319	1,235608	0,000***
Medenihal_3	-0,1012576	0,139773	0,9037002	0,000***
Medenihal_4	0,1832079	0,0386586	1,201064	0,000***
Egitim_1	-0,0456602	0,0600643	0,9553665	0,000***
Egitim_2	-0,0960596	0,0538543	0,9084099	0,000***
Egitim_3	-0,1317381	0,0565943	0,8765705	0,000***
Egitim_4	-0,0734755	0,0556908	0,9291589	0,000***
Egitim_5	-0,0311467	0,0566235	0,9693334	0,000***
Egitim_6	-0,139555	0,0575976	0,8697452	0,000***
Egitim_7	-0,0929263	0,0585946	0,9112607	0,000***
Egitim_8	0,0248711	0,0559315	1,025183	0,000***
Egitim_9	0,050757	0,1034335	1,052067	0,000***
Egitim_10	0,2919409	0,1945425	1,339024	0,096*
Genelsaglik_2	0,0392496	0,0249321	1,04003	0,102
Genelsaglik_3	0,0911595	0,0376423	1,095444	0,008***
Genelsaglik_4	0,1981903	0,0599089	1,219194	0,000***
Genelsaglik_5	0,0683765	0,1114381	1,070768	0,185
Kronikhastalik_2	0,0034677	0,0254876	1,003474	0,000***
α (Yayılım parametresi)	1,068496			
α=0 için LR test: chibar2(01) = 1562,78 Prob > chibar2 = 0,000		Vuong test z=42,90 Prob >z = 0,000		
*** %1 hata payıyla istatistiksel olarak anlamlılığı gösterir. *%10 hata payıyla istatistiksel olarak anlamlılığı gösterir.				

Tablo 4.12 incelendiğinde LR test istatistiğine ait olasılık değeri p<0,05 olduğu için modelde aşırı yayılım olduğu sonucuna ulaşılmıştır. Vuong testi ise, oluşturulan model için sıfır değer ağırlıklı negatif binom modelinin mi yoksa negatif binom modelinin mi uygun olduğunu test etmektedir. Vuong test istatistiğine ait olasılık değeri p<0,05 olduğu için modelde yapısal sıfırın önemli olduğu ve sıfır değer ağırlıklı negatif binom regresyon modelinin belirlenen model için uygun olan tahmin yöntemlerinden biri olduğu sonucuna ulaşılmıştır.

Bağımlı değişkendeki sıfır değerini dikkate alan bir diğer model olan Genelleştirilmiş Poisson regresyon modeli ile tahmin edilmiş modellere ait uyum iyiliği göstergeleri Tablo 4.13'te gösterilmektedir.

Tablo 4.13.Tüm olası altküme yöntemine göre Genelleştirilmiş Poisson regresyon sonuçları

Model	AIC	LLModel	Sapma	Sapma SD	Yayılım Parametresi
X_1	79146,54	-39570,27	131842,6	62710	2,102418
X_2	78292,57	-39141,28	124147,8	62708	1,979776
X_3	78788,73	-39382,36	128076,2	62701	2,04265
X_4	79015,03	-39501,51	130395,1	62707	2,079435
X_5	79010,83	-39502,42	130399	62710	2,079398
X_6	79146,73	-39570,36	131844,7	62710	2,102451
X_1,X_2	78292,54	-39140,27	124133,8	62707	1,979584
X_1,X_3	78790,69	-39382,35	128073	62700	2,042632
X_1,X_4	79016,83	-39501,42	130391,1	62706	2,079404
X_1,X_5	79012,58	-39502,29	130394,5	62709	2,079359
X_1,X_6	79148,54	-39570,27	131840,5	62709	2,102418
X_2,X_3	78152,9	-39061,45	122808,7	62698	1,958734
X_2,X_4	78244,15	-39113,08	123822,5	62704	1,974714
X_2,X_5	78227,02	-39107,51	123697,5	62707	1,972627
X_2,X_6	78293,67	-39140,84	124153,5	62707	1,979899
X_3,X_4	78706,93	-39337,47	127419	62697	2,032298
X_3,X_5	78683,74	-39328,87	127208,5	62700	2,028844
X_3,X_6	78790,34	-39382,17	128070,5	62700	2,042591
X_4,X_5	78988,8	-39487,4	130093,9	62706	2,074664
X_4,X_6	79016,95	-39501,48	130392,5	62706	2,079427
X_5,X_6	79012,81	-39502,41	130396,8	62709	2,079396
X_1,X_2,X_3	78293,87	-39131,94	123992,2	62698	1,97761
X_1,X_2,X_4	78244,7	-39112,35	123804,2	62700	1,974548
X_1,X_2,X_5	78227,68	-39106,84	123686,2	62706	1,972478
X_1,X_2,X_6	78293,55	-39139,78	124139,6	62706	1,979709
X_1,X_3,X_4	78708,93	-39337,46	127416,3	62696	2,032288
X_1,X_3,X_5	78685,7	-39328,85	127205,2	62699	2,028824
X_1,X_3,X_6	78792,32	-39382,16	128067,5	62699	2,042577
X_1,X_4,X_5	78990,59	-39487,3	130089,7	62705	2,07463
X_1,X_4,X_6	79018,74	-39501,37	130388,4	62705	2,079394
X_1,X_5,X_6	79014,57	-39502,29	130392,4	62708	2,079358
X_2,X_3,X_4	78116,1	-39039,05	122607,4	62694	1,955648
X_2,X_3,X_5	78093,46	-39030,73	122487,7	62697	1,953645
X_2,X_3,X_6	78154,7	-39061,35	122811,3	62697	1,958807
X_2,X_4,X_5	78221,04	-39100,52	123598	62703	1,971165
X_2,X_4,X_6	78244,99	-39112,49	123827,8	62703	1,97483
X_2,X_5,X_6	78228,09	-39107,05	123702,6	62706	1,97274
X_3,X_4,X_5	78675,13	-39320,57	127100,6	62696	2,027253
X_3,X_4,X_6	78708,93	-39337,47	127416,8	62696	2,032295
X_3,X_5,X_6	78685,66	-39328,83	127205	62699	2,028821
X_4,X_5,X_6	78990,78	-39487,39	130091,7	62705	2,074662
X_1,X_2,X_3,X_4	78117,69	-39038,84	122599,7	62693	1,955556
X_1,X_2,X_3,X_5	78095,05	-39030,52	122480,6	62696	1,953563
X_1,X_2,X_3,X_6	78155,93	-39060,96	122802,1	62696	1,958691
X_1,X_2,X_4,X_5	78221,79	-39099,89	123587,1	62702	1,971024
X_1,X_2,X_4,X_6	78245,42	-39111,71	123815,4	62702	1,974664
X_1,X_2,X_5,X_6	78228,66	-39106,33	123691,4	62705	1,972593
X_1,X_3,X_4,X_5	78677,11	-39320,55	127097,5	62695	2,027235
X_1,X_3,X_4,X_6	78710,93	-39337,46	127414,2	62695	2,032286
X_1,X_3,X_5,X_6	78687,63	-39328,82	127202	62698	2,028804

Tablo 4.14.Tüm olası altküme yöntemine göre Genelleştirilmiş Poisson regresyon sonuçları (devamı)

X_1,X_4,X_5,X_6	78992,56	-39487,28	130087,4	62704	2,074627
X_2,X_3,X_4,X_5	78091,16	-39025,58	122401,6	62693	1,952396
X_2,X_3,X_4,X_6	78117,59	-39038,79	122612,1	62693	1,955754
X_2,X_3,X_5,X_6	78095,11	-39030,56	122490,9	62696	1,953728
X_2,X_4,X_5,X_6	78222,02	-39100,01	123603,3	62702	1,971281
X_3,X_4,X_5,X_6	78677,11	-39320,56	127097,9	62695	2,027241
X_1,X_2,X_3,X_4,X_5	78092,81	-39025,4	122394,9	62692	1,952321
X_1,X_2,X_3,X_4,X_6	78119,13	-39038,56	122604,4	62692	1,955663
X_1,X_2,X_3,X_5,X_6	78096,67	-39030,33	122484	62695	1,953648
X_1,X_2,X_4,X_5,X_6	78222,68	-39099,34	123592,6	62701	1,971142
X_1,X_3,X_4,X_5,X_6	78679,09	-39320,54	127094,8	62694	2,027225
X_2,X_3,X_4,X_5,X_6	78092,77	-39025,38	122405,3	62692	1,952486
X_1,X_2,X_3,X_4,X_5,X_6	78094,38	-39025,19	122398,7	62691	1,952413

Tablo 4.13 ve Tablo 4.14 incelendiğinde AIC değeri en düşük olan model X_2,X_3,X_4,X_5 bağımsız değişkenlerinden oluşan modeldir. Bu model, aşırı yayılımın olduğu durumda kullanılan modellerden bir diğeri olan Genelleştirilmiş Poisson ile tahmin edilmiştir. Genelleştirilmiş Poisson modeline ait regresyon sonuçları Tablo 4.15'te verilmektedir.

Tablo 4.15. Genelleştirilmiş Poisson regresyon modeli tahmin sonuçları

Log likelihood = -39025,581		Gözlem Sayısı = 62713 LR $\chi^2(18)$ = 1039,30 Prob > χ^2 = 0,0000		
Bağımsız Değişkenler	**Katsayı (β)**	**Standart Hata**	**IRR**	**Prob Değeri**
Sabit	-2,024206	0,1575235	0,013209	0,000***
Medenihal_2	1,105916	0,0866502	3,021992	0,000***
Medenihal_3	-2,377735	0,1249096	0,0927604	0,000***
Medenihal_4	0,938165	0,1950158	2,555288	0,000***
Egitim_1	0,9855992	0,1190057	2,679417	0,000***
Egitim_2	1,026756	0,0870631	2,791993	0,000***
Egitim_3	1,543129	0,1613289	4,679209	0,000***
Egitim_4	0,944724	0,1100065	2,572103	0,000***
Egitim_5	1,02482	0,117501	2,786594	0,000***
Egitim_6	0,8387514	0,1235054	2,313477	0,000***
Egitim_7	0,9763867	0,1464065	2,654846	0,000***
Egitim_8	1,122802	0,1234417	3,073453	0,000***
Egitim_9	0,8474008	0,2981713	2,333573	0,004***
Egitim_10	0,8595097	0,5231404	2,362002	0,100
Genelsaglik_2	-0,1272161	0,1240689	0,8805434	0,305
Genelsaglik_3	-0,0381354	0,1428511	0,9625826	0,790
Genelsaglik_4	-0,2921774	0,1580734	0,746636	0,065*
Genelsaglik_5	-0,367133	0,2523634	0,6927175	0,146
Kronikhastalik_2	0,4099255	0,0775345	1,506706	0,000***
α (Yayılım parametresi)	1,952396			
*** %1 hata payıyla istatistiksel olarak anlamlılığı gösterir.				

Tablo 4.15 incelendiğinde yayılım parametresinin 1'den büyük olduğu ve modelde aşırı yayılımın olduğu sonucuna ulaşılmıştır.

Analiz boyunca tahmin edilen Negatif Binom, Sıfır Değer Ağırlıklı Negatif Binom ve Genelleştirilmiş Poisson regresyon modellerinden en iyi uyum sağlayan modele karar verebilmek için çalışmada bilgi kriterleri ve olabilirlik değeri kullanılmıştır. Tahmin edilen üç modele ait bilgi kriterleri ve olabilirlik değeri Tablo 4.16'da verilmektedir.

Tablo 4.16.Model seçim kriterlerine ait sonuçlar

Model	**Olabilirlik Değeri**	**AIC**	**BIC**
Negatif Binom	-37282,9	74605,79	74786,72
Sıfır Değer Ağırlıklı Negatif Binom	-33483,46	67044,93	67397,73
Genelleştirilmiş Poisson	-39025,58	78091,16	78272,09

Tablo 4.16 incelendiğinde veri setine en iyi uyum sağlayan modelin Sıfır Değer Ağırlıklı Negatif Binom regresyon modeli olduğu görülmektedir. Bu sebeple çalışmada tahmin edilen modellerden sadece Sıfır Değer Ağırlıklı Negatif Binom regresyon modeli yorumlanmıştır.

Sıfır Değer Ağırlıklı Negatif Binom regresyon modeline ait katsayı, standart hata, vaka hız oranları (IRR)[1] ve olasılık değerleri yukarıda Tablo 4.12'de verilmiştir. Tahmin edilen modele ait ki-kare testi olasılık değeri $p<0,01$ olduğu için model genel olarak anlamlıdır. Tahmin edilen modeldeki bağımsız değişkenlere ait olasılık değerlerine bakıldığında Medenihal_2, Medenihal_3, Medenihal_4, Egitim_1, Egitim_2, Egitim_3, Egitim_4, Egitim_5, Egitim_6, Egitim_7, Egitim_8, Egitim_9, Egitim_10, Genelsaglik_3 ve Genelsaglik_4, Kronikhastalık_2 adlı değişkenlerin istatistiksel olarak anlamlı olduğu görülmektedir.

Tablo 4.12'de Medenihal_2 değişkeni "Hiç evlenmedi", Medenihal_3 değişkeni "Eşi öldü", Medenihal_4 değişkeni ise "Boşandı" kategorilerini temsil etmektedir. Medeni hal değişkeni için dışlanan kategori ise "Evli" kategorisidir.

Egitim_2 değişkeni "İlkokul", Egitim_3 değişkeni "İlköğretim", Egitim_4 değişkeni "Ortaokul", Egitim_5 "Genel lise", Egitim_6 değişkeni "Mesleki veya Teknik Lise",

[1] Vaka hız oranları (IRR-Incident Rate Ratios): Regresyon katsayılarının eksponansiyelidir.

Egitim_7 değişkeni “Yüksekokul”, Egitim_8 değişkeni “Fakülte”, Egitim_9 değişkeni “Yüksek Lisans” ve Egitim_10 değişkeni “Doktora” kategorilerini temsil etmektedir. Eğitim kategorisinde dışlanan değişken ise “Okuryazar olmayan” kategorisidir.

Genelsaglik_3 değişkeni “Orta” ve Genelsaglik_4 değişkeni “Kötü” kategorilerini temsil etmektedir. Genel sağlık değişkeninde dışlanan kategori ise “Çok İyi” kategorisidir.

Kronikhastalik_2 değişkeni “Hayır” kategorisini temsil etmektedir. Referans kategorisi “Evet” kategorisidir.

Tablo 4.12’daki Sıfır Değer Ağırlıklı Negatif Binom regresyon modeline ait tahmin edilen vaka hız oranlarının (IRR) yorumu aşağıda verilmektedir.

- **Medenihal_2:** Hiç evlenmemiş bir bireyin işsiz kaldığı ay sayısı evli bir bireyin işsiz kaldığı ay sayısından yaklaşık olarak 1,24 kat daha fazla olmaktadır.

- **Medenihal_3:** Eşi ölmüş bir bireyin işsiz kaldığı ay sayısı evli bir bireyin işsiz kaldığı ay sayısından 0,9 kat daha az olmaktadır.

- **Medenihal_4:** Boşanmış bir bireyin işsiz kaldığı ay sayısı evli bir bireyin işsiz kaldığı ay sayısından 1,2 kat daha fazla olmaktadır.

- **Egitim_1:** Bir okul bitirmeyen bireyin işsiz kaldığı ay sayısı okuryazar olamayan bir bireyin işsiz kaldığı ay sayısından 0,96 kat daha az olmaktadır.

- **Egitim_2:** İlkokul mezunu bir bireyin işsiz kaldığı ay sayısı okuryazar olmayan bir bireyin işsiz kaldığı ay sayısından 0,91 kat daha az olmaktadır.

- **Egitim_3:** İlköğretim mezunu bir bireyin işsiz kaldığı ay sayısı okuryazar olmayan bir bireyin işsiz kaldığı ay sayısından 0,88 kat daha az olmaktadır.

- **Egitim_4:** Ortaokul mezunu bir bireyin işsiz kaldığı ay sayısı okuryazar olmayan bir bireyin işsiz kaldığı ay sayısından yaklaşık olarak 0,93 kat daha az olmaktadır.

- **Egitim_5:** Genel lise mezunu bir bireyin işsiz kaldığı ay sayısı okuryazar olmayan bir bireyin işsiz kaldığı ay sayısından yaklaşık olarak 0,97 kat daha az olmaktadır.

- **Egitim_6:** Mesleki veya teknik lise mezunu bir bireyin işsiz kaldığı ay sayısı okuryazar olmayan bir bireyin işsiz kaldığı ay sayısından 0,87 kat daha az olmaktadır.

- **Egitim_7:** Yüksekokul mezunu bir bireyin işsiz kaldığı ay sayısı okuryazar olmayan bir bireyin işsiz kaldığı ay sayısından yaklaşık olarak 0,91 kat daha az olmaktadır.

- **Egitim_8:** Fakülte mezunu bir bireyin işsiz kaldığı ay sayısı okuryazar olmayan bir bireyin işsiz kaldığı ay sayısından yaklaşık olarak 1,03 kat daha fazla olmaktadır.

- **Eğitim_9:** Yüksek lisans mezunu bir bireyin işsiz kaldığı ay sayısı okuryazar olmayan bir bireyin işsiz kaldığı ay sayısından yaklaşık olarak 1,05 kat daha fazla olmaktadır.

- **Egitim_10:** Doktora mezunu bir bireyin işsiz kaldığı ay sayısı okuryazar olmayan bir bireyin işsiz kaldığı ay sayısından yaklaşık olarak 1,34 kat daha fazla olmaktadır.

- **Genelsaglik_3:** Genel sağlık durumu orta düzeyde olan bir bireyin işsiz kaldığı ay sayısı genel sağlık durumu çok iyi olan bir bireyin işsiz kaldığı ay sayısından 1,1 kat daha fazla olmaktadır.

- **Genelsaglik_4:** Genel sağlık durumu kötü olan bir bireyin işsiz kaldığı ay sayısı genel sağlık durumu çok iyi olan bir bireyin işsiz kaldığı ay sayısından 1,22 kat daha fazla olmaktadır.

- **Kronikhastalık_2:** Kronik hastalığı olmayan bir bireyin işsiz kaldığı ay sayısı kronik hastalığı olan bir bireyin işsiz kaldığı ay sayısından yaklaşık olarak 1,003 kat daha fazla olmaktadır.

5. SONUÇ VE ÖNERİLER

Bu çalışmada sayma verisi regresyon modelleri incelenmiş ve bir uygulama gerçekleştirilmiştir. Sayma verisi regresyon modelleri Genelleştirilmiş Doğrusal Modellere dayanmaktadır. Bu bağlamda çalışmada öncelikle GDM konusuna değinilmiştir. Veri türüne göre genelleştirilmiş doğrusal modellerden olan sayma verisi regresyon modelleri üçüncü bölüm içerisinde detaylı olarak tanıtılmıştır. Çalışmada modelin değerlendirilmesi ve seçimine de yer verilmiştir.

Çalışmada kullanılan veri seti TÜİK tarafından uygulanan 2019 yılı Gelir ve Yaşam Koşulları Araştırması'ndan elde edilmiştir. Analizde Stata 15.0 istatistik paket programı kullanılmıştır. Bu tez çalışmasının uygulama kısmında bağımlı değişken "işsiz kalınan ay sayısı" olarak belirlenmiştir. "İşsiz kalınan ay sayısı" değişkeni sayıma dayalı olarak elde edilen bir veri olduğu için analiz yöntemi olarak sayma verisi regresyon modelleri kullanılarak bu modellerden en uygun olan modelin belirlenmesi amaçlanmıştır. Analizde bağımlı değişken üzerinde etkili olabileceği düşünülen cinsiyet, medeni durum, eğitim durumu, genel sağlık durumu, kronik hastalık durumu ve yaş değişkenleri bağımsız değişken olarak belirlenmiştir. Bağımsız değişkenler ile bağımlı değişkeni en iyi açıklayan modeli seçebilmek için tüm olası altküme yöntemi ile modeller arasında seçim yapılmıştır. Tüm olası altküme yöntemi ile oluşturulan modeller sayma verisi regresyon modellerinden Poisson, Negatif Binom, Sıfır Değer Ağırlıklı Negatif Binom ve Genelleştirilmiş Poisson modelleri ile tahmin edilmiştir. Analizde kullanılan tüm sayma verisi regresyon modelleri için uygun modelin, tüm olası altküme yöntemi ile "Medeni hal", "Eğitim durumu", "Genel sağlık durumu" ve " Kronik hastalık durumu" bağımsız değişkenlerinden oluşan model olarak belirlenmiştir. Tüm olası altküme yöntemi ile belirlenen 4 bağımsız değişkenli model, Poisson regresyon modeli ile tahmin edildiğinde modelin aşırı yayılım gösterdiği bu nedenle Poisson regresyonun modele uygun olmadığı sonucuna ulaşılmıştır. Modelin aşırı yayılım gösterdiği durumda kullanılan modeller olan Negatif Binom, Sıfır Değer Ağırlıklı Negatif Binom ve Genelleştirilmiş Poisson tahmin yöntemleri ile belirlenen model tahmin edilmiştir. Tahmin edilen modellere ait model seçim kriterleri incelendiğinde 4 bağımsız değişkenli model için en uygun modelin Sıfır Değer Ağırlıklı Negatif Binom regresyon modeli olduğu sonucuna ulaşılmıştır. Sıfır Değer Ağırlıklı Negatif Binom regresyon modeli ile tahmin edilen model medeni durum değişkenine ait hiç evlenmemiş, eşi öldü ve boşanmış kategorileri; eğitim durum değişkenine ait bir okul bitirmedi, ilkokul mezunu, ilköğretim mezunu, ortaokul mezunu, genel lise mezunu, mesleki veya teknik lise mezunu, yüksekokul mezunu, fakülte mezunu, yüksek lisans mezunu ve doktora mezunu kategorileri; genel sağlık durumu değişkeni için orta ve kötü kategorileri ile kronik hastalık değişkeninin hayır kategorisi istatistiksel olarak anlamlı olduğu tespit edilmiştir. Analiz sonucuna göre; hiç evlenmemiş ve boşanmış bireylerin işsiz kaldığı ay sayısının evli bireylerin işsiz kaldığı ay sayısından daha fazla olduğu görülmüştür. Eşi ölmüş bireylerin evli bireylere göre işsiz kaldıkları ay sayısının daha az olduğu tespit edilmiştir. İlkokul mezunu, ilköğretim mezunu, ortaokul mezunu, genel lise mezunu, mesleki veya teknik lise mezunu ve yüksekokul mezunu bireylerin okuryazar olmayan bireylere göre işsiz kaldıkları ay sayısının daha az olduğu tespit edilmiştir. Fakülte mezunu, yüksek lisans mezunu ve doktora mezunu bireylerin okuryazar olmayan bireylere göre işsiz kaldıkları ay sayısının daha fazla olduğu sonucuna ulaşılmıştır. Bu durum eğitim düzeyi ile işsiz kalınan ay sayısı arasında belirli bir eğitim seviyesine kadar negatif yönlü bir ilişkiyi ortaya koyarken belirli bir eğitim seviyesinden sonra bu ilişkinin pozitif yönlü olduğunu ortaya koymaktadır. Eğitim durumu belirli bir düzeye kadar işsiz kalınan ay sayısını azalmakta fakat belirli bir düzeyden sonra işsiz kalınan ay sayısı arttırmaktadır. Bu durum fakülte, yüksek lisans ve doktora mezunlarının eğitim düzeylerine uygun olarak iş tercihlerinde daha seçici davranmaları sonucunda işsiz kaldıkları ay sayısının artmasına neden olabilmektedir. Genel sağlık durumu orta düzeyde ve kötü olan bireylerin işsiz kaldıkları ay sayısının sağlık durumu çok iyi olan bireylerin işsiz kaldıkları ay sayısından daha fazla olduğu sonucuna ulaşılmıştır. Bu durum bireyin sağlık durumunun kötüye gitmesinin işsiz kalınan ay sayısını arttırdığını ortaya koymaktadır. Kronik hastalığı olan bireylerin kronik hastalığı olmayan bireylere göre işsiz kaldıkları ay sayısının daha az olduğu da ulaşılan sonuçlar arasında yer almaktadır.

Bu tez çalışmasında yapılan uygulama ve analizlerin gelecekte yapılacak olan çalışmalarla zenginleştirilmesi mümkündür. Literatürde, bağımlı değişken olan işsiz kalınan ay sayısına ait herhangi bir çalışmaya rastlanmamıştır. Bu nedenle bu çalışmanın bireylerin işsiz kaldıkları ay sayısını etkileyen faktörlerin belirlenmesine yönelik gelecekte yapılacak olan çalışmalar için yol gösterici nitelikte olduğu düşünülmektedir.

6. KAYNAKLAR

Aa, M. A. & Naing, N. N. (2012). Analysis Death Rate of Age Model with Excess Zeros using Zero Inflated Negative Binomial and Negative Binomial Death Rate: Mortality AIDS Co-Infection Patients, Kelantan Malaysia. *Procedia Economics and Finance, 2, 275-283.*

Açıkyürek, G. (2016). *Poisson Regresyon ve Bir Uygulama.* Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. John Wiley & Sons Inc., Hoboken, New Jersey.

Akinpelu, K. P., Yusuf, O. B., Akpa, O. M. & Gbolahan, A. O. (2016). Zero Inflated Regressions Models with Application to Malaria Surveillance Data. *International Journal of Statistics and Applications, 6(4), 223-234.*

Avcı, E. (2018). Using Count Regression Models to Determine the Factors which Effects the Hospitalization Number of People with Schizophrenia. *Journal of Data Science, 16(3), 511-528.*

Balyaner, İ. (2016). *Türkiye'de Hanehalkının Sahip Olduğu Bilişim Teknolojileri Ürünleri Sayısının Araştırılması: Bir Sayma Veri Modeli.* Yüksek Lisans Tezi, Celal Bayar Üniversitesi Sosyal Bilimler Enstitüsü, Manisa.

Beaujean, A. A. & Grant, M. B. (2016). Tutorial on Using Regression Models with Count Outcomes Using R. *Practical Assessment, Research, and Evaluation, 21(2).*

Cameron, A. C. & Trivedi, P. K. (2003). Essentials of Count Data Regression. Baltagi, B. H., (Editör), *A Companion To Theoretical Econometrcis* içinde (331-348). Blackwell Publishing.

Cameron, A. C. & Trivedi, P. K. (2013). *Regression Analysis of Count Data.* (Second Edition). New York: Cambridge University Press.

Cameron, A. C., Trivedi, P. K. (1998). *Regression Analysis of Count Data.* New York: Cambridge University Press.

Claeskens, G. & Hjort, N. L. (2008). *Models Selection and Model Averaging.* New York: Cambridge University Press.

Czado, C., Erhardt, V., Min, A. & Wagner, S. (2007). Zero-Inflated Generalized Poisson Models with Regression Effects On The Mean, Dispersion and Zero-Inflation Level Applied to Patent Outsourcing Rates. *Statistical Modelling, 7(2), 125-153.*

Çelik, B. (2019). *Conway-Maxwell-Poisson Regresyon Modeli.* Yüksek Lisans Tezi, Giresun Üniversitesi Fen Bilimleri Enstitüsü, Giresun.

Dinarcan, G. N. (2018). *Sayma Verisi İçin Regresyon Modelleri Ve Bir Uygulama.* Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

Erdem, İ. (2013). *Matematiksel İstatistik: Problemler ve Çözümleri.* (2. Baskı). Ankara: Seçkin Yayıncılık.

Famoye, F. & Singh, K. D. (2006). Zero-Inflated Generalized Poisson Regression Model with an Application to Domestic Violence Data. *Journal of Data Science, 4(1), 117-130.*

Giray, G. (2019). *Poisson ve Negatif Binom Regresyon Modelleri.* Yüksek Lisans Tezi, Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Adana.

Hilbe, J. M. (2014). *Modeling Count Data.* New York: Cambridge University Press.

Ismail, N., Zamani, H. (2013). Estimation of Claim Count Data using Negative Binomial, Generalized Poisson, Zero-Inflated Negative Binomial and Zero-Inflated Generalized Poisson Regression Models. *Casualty Actuarial Society E-Forum.*

Jong, P. de & Heller, G. Z. (2008). *Generalized Linear Models for Insurance Data.* New York: Cambridge University Press.

Karaca, A. G. (2018). *Sayma Verileri İçin Regresyon Modellerinin Karşılaştırılması Üzerine Bir Uygulama.* Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

Kartal, S. (2019). *Kırpılmış Poisson Regresyon Analizi ve Bir Uygulama.* Yüksek Lisans Tezi, Sivas Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü, Sivas.

Kaya, Y., Yeşilova, A. (2012). E-Posta Trafiğinin Sıfır Değer Ağırlıklı Regresyon Yöntemleri Kullanılarak İncelenmesi. *Anadolu Üniversitesi Bilim ve Teknoloji Dergisi-A Uygulamalı Bilimler ve Mühendislik,13(1),51-63.*

Kıran, G. (2019). *Sayma Veri Modelleri ve Bir Uygulama.* Yüksek Lisans Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.

Lambert, D. (1992). Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing. *Technometrics, 34(1).*

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

Madsen, H. & Thyregod, P. (2010). *Introduction to General and Generalized Linear Models.* Boca Raton: CRC Press.

Mamun, A. A. (2014). *Zero-Inflated Regression Models for Count Data An Application to Under-5 Deaths*. M. Sc. Thesis, Muncie, Ball State University.

McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models.* (Second Edition). London: Chapman and Hall.

Mert, M. (2016). *Yatay Kesit Veri Analizi Bilgisayar Uygulamaları.* Ankara: Detay Yayıncılık.

Mouatassim, Y. & Ezzahid, E. H. (2012). Poisson Regression and Zero-inflated Poisson Regression: Application to Private Health İnsurance Data. *European Actuarial Journal, 2(2), 187-204.*

Myers, R. H., Montgomery, D. C., Vining, G. G. & Robinson, T. J. (2010). *Generalized Linear Models with Applications in Engineering and the Sciences.* (Second Edition). Hoboken, New Jersey: John Wiley & Sons, Inc.

Olsson, U. (2002). *Generalized Linear Models An Applied Approach.* Sweden: Studentlitteratur.

Özmen, İ. ve Famoye, F. (2007). Count Regression Models with an Application to Zoological Data Containing Structural Zeros. *Journal of Data Science,5(4), 491-502.*

Peng, J., Lyu, T., Shi, J., Nagaraja, H. N. & Xiang, H. (2014). Models for Injury Count Data in the U.S. National Health Interview Survey. *Journal of Scientific Research & Reports,3(17), 2286-2302.*

Saffari, E. S., Adnan, R. & Greene, W. (2012). Hurdle Negative Binomial Regression Model with Right Cencored Count Data. *SORT, 36(2), 181-194.*

Selim, S. ve Üçdoğruk, Ş. (2003). Sayma Veri Modelleri ile Çocuk Sayısı Belirleyicileri: Türkiye'deki Seçilmiş İller İçin Sosyoekonomik Analizler. *Dokuz Eylül Üniversitesi İ.İ.B.F. Dergisi, 18(2), 13-31.*

Singh, K. P. & Famoye, F. (1993). Analysis of Rates Using a Generalized Poisson Regression Model. *Biometrical Journal, 35(8), 917-923.*

Soycan, G. (2017). *Sayıma Dayalı Olarak Elde Edilen Veri Kümelerinde Bağımlı Değişkenin Modellenmesinde Uygun Doğrusal Olmayan Regresyon Modelinin Belirlenmesi.* Yüksek Lisans Tezi, Van Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü, Van.

Şimşek, E. (2017). *Poisson Regresyon Analizinde Log-Doğrusal Modellerin Kullanılması.* Yüksek Lisans Tezi, Van Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü, Van.

Tamar, M. (2013). *Poisson Regresyonu.* Yüksek Lisans Tezi, Fırat Üniversitesi Fen Bilimleri Enstitüsü, Elazığ.

Tarı, R. (2018). *Ekonometri*. (13. Baskı). Kocaeli: Umuttepe Yayınları.

Tüzel, S. ve Sucu, M. (2012). Hasar Sıklıkları İçin Sıfır Yığılmalı Kesikli Modeller. *İstatistikçiler Dergisi, 5(1), 23-31.*

Wang, W. & Famoye, F. (1997). Modeling Household Fertility Decisions with Generalized Poisson Regression. *Journal of Population Economics, 10(3), 273-283.*

Winkelmann, R. (2008). *Econometric Analysis of Count Data.* (5th Edition). Heildberg: Springer.

Yeşilyurt, H. (2005). *Poisson Regresyon Modeli ve Türkiye'deki Boşanma İstatistiklerine Uygulanması.* Yüksek Lisans Tezi. Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

Zulkifli, M., Ismail, N., Razali, A. M. (2011). Zero-Inflated Poisson Versus Zero-Inflated Negative Binomial: Application to Theft Insurance Data. *The 7th IMT-GT International Conference on Mathematics, Statistics and its Application, 511-524.*

Zulkifli, M., Ismail, N., Razali, A. M. (2012). Zero-Inflated Regression Models With an Application to Vehicle Theft Count Data. *Technology, Science, Social Science and Humanities International Conference.*