SAYMA VERİSİ İÇİN REGRESYON MODELLERİ VE BİR UYGULAMA

COUNT DATA REGRESSION MODELS AND AN APPLICATION

GÖZDE NUR DİNARCAN

PROF. DR. MERAL ÇETİN
Tez Danışmanı

Hacettepe Üniversitesi
Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin
İstatistik Anabilim Dalı için Öngördüğü
YÜKSEK LİSANS TEZİ olarak hazırlanmıştır.

2018

GÖZDE NUR DİNARCAN' ın hazırladığı **"Sayma Verisi İçin Regresyon Modelleri ve Bir Uygulama"** adlı bu çalışma aşağıdaki jüri tarafından **İSTATİSTİK ANABİLİM DALI**' nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Prof. Dr. Serpil AKTAŞ ALTUNAY
Başkan ....

Prof. Dr. Meral ÇETİN
Danışman ....

Doç. Dr. Hülya OLMUŞ
Üye ....

Doç. Dr. Semra TÜRKAN
Üye ....

Doç. Dr. Ayten YİĞİTER
Üye ....

Bu tez Hacettepe Üniversitesi Fen Bilimleri Enstitüsü tarafından **YÜKSEK LİSANS TEZİ** olarak onaylanmıştır.

Prof. Dr. Menemşe GÜMÜŞDERELİOĞLU
Fen Bilimleri Enstitüsü Müdürü

YAYINLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanması zorunlu metinlerin yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

☐ **Tezimin/Raporumun tamamı dünya çapında erişime açılabilir ve bir kısmı veya tamamının fotokopisi alınabilir.**
(Bu seçenekle teziniz arama motorlarında indekslenebilecek, daha sonra tezinizin erişim statüsünün değiştirilmesini talep etseniz ve kütüphane bu talebinizi yerine getirse bile, tezinin arama motorlarının önbelleklerinde kalmaya devam edebilecektir.)

☐ **Tezimin/Raporumun ………….. tarihine kadar erişime açılmasını ve fotokopi alınmasını (İç Kapak, Özet, İçindekiler ve Kaynakça hariç) istemiyorum.**
(Bu sürenin sonunda uzatma için başvuruda bulunmadığım taktirde, tezimin/raporumun tamamı her yerden erişime açılabilir, kaynak gösterilmek şartıyla bir kısmı ve ya tamamının fotokopisi alınabilir)

☒ **Tezimin/Raporumun 29/05/2020 tarihine kadar erişime açılmasını istemiyorum, ancak kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisinin alınmasını onaylıyorum.**

☐ **Serbest Seçenek/Yazarın Seçimi**

29../05../.2018..

(İmza)

Öğrencinin Adı Soyadı

Gözde Nur DİNARCAN

ETİK

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

29/05/2018

Gözde Nur DİNARCAN

ÖZET

SAYMA VERİSİ İÇİN REGRESYON MODELLERİ VE BİR UYGULAMA

Gözde Nur DİNARCAN

Yüksek Lisans, İstatistik Bölümü

Tez Danışmanı: Prof. Dr. Meral ÇETİN

Mayıs 2018, 52 Sayfa

Regresyon analizi, bir bağımlı değişken ile bir ya da birden fazla açıklayıcı değişken arasındaki ilişkiyi ölçmek için kullanılır. Bağımlı değişken negatif olmayan kesikli değerlerden oluştuğunda, sayma verisi modellerinin kullanımı önerilmektedir. Sayma verisi analizinde kullanılan en yaygın modellerden biri Poisson regresyondur. Poisson regresyon analizinde ortalama ve varyans eşittir ancak bu eşitliğin sağlanmadığı durumda az yayılım ya da aşırı yayılım durumu gerçekleşmektedir. Bu durumlar için negatif binom ve genelleştirilmiş Poisson regresyon modelleri geliştirilmiştir.

Bu tez çalışmasında; genelleştirilmiş doğrusal modeller hakkında bilgi verilmiş, sayma verisi modellerinden Poisson regresyon modeli, negatif binom regresyon modeli ve sıfır yığılmalı modeller tanıtılmıştır. Türkiye İstatistik Kurumu'ndan alınan, 2016 yılı Hanehalkı İşgücü Araştırması verileri Poisson regresyon, negatif binom regresyon ve genelleştirilmiş Poisson regresyon analizleri kullanılarak modellenmiştir. Fertlerin bir haftada kaç saat çalıştığı bilgisini içeren çalışma değişkeni bağımlı değişken, sektör, yaş, cinsiyet, eğitim, meslek, kayıtlılık, süreklilik ise açıklayıcı değişkenler olarak alınmıştır.

Oluşturulan regresyon modelleri model seçim kriterleri ile karşılaştırılmıştır. Veri seti için en uygun model negatif binom regresyon modeli olarak seçilmiştir.

Uygulamada SPSS, SAS ve R programları kullanılmıştır.

Anahtar kelimeler: Sayma verisi, Genelleştirilmiş doğrusal modeller, Negatif binom regresyon, Poisson regresyon, Sıfır yığılmalı regresyon modelleri, Aşırı yayılım.

ABSTRACT

COUNT DATA REGRESSION MODELS AND AN APPLICATION

Gözde Nur DİNARCAN

Master, Department of Statistics

Supervisor: Prof. Dr. Meral ÇETİN

May 2018, 52 pages

Regression analysis is used to measure the relationship between a dependent variable and one or more explanatory variables. When the dependent variables consist of discrete non-negative values, the use of counting data models is recommended. Poisson regression which is one of the most common models used in count data analysis**.** In the Poisson regression analysis, mean and variance are assumed to be equal, but when this equality is not provided, underdispersion or overdispersion occurs. For these cases, negative binomial and generalized Poisson regression models have been developed.

In this study, generalized linear models were examined, Poisson regression model, negative binomial regression model and zero-inflated models were introduced. The data taken from Turkish Statistical Institute, Household Labour Force Survey for 2016 were modeled using negative binomial regression and generalized Poisson regression analysis. The “working” variable which contains knowledge of persons' how many hour working in a week is dependent variable. Sector, age, gender, education, occupation, registration, continuity are taken as explanatory variables.

The created regression models are compared with model selection criteria. The best model for the dataset applied to data as negative binomial regression model.

In practice, SPSS, SAS and R programs were used to obtain results.

Key Words: Count data, Generalized linear models, Negative binomial regression, Poisson regression, Zero - Inflated regression models, Overdispersion.

TEŞEKKÜR

Tez çalışmam sürecinde yol gösteren, yardımını esirgemeyen danışmanım Sayın Prof. Dr. Meral ÇETİN'e,

Tezimin uygulama kısmında veri seçimindeki yardımı ve desteği için Yosun KATI'ya,

Değerli mesai arkadaşlarım Hanehalkı İşgücü İstatistikleri Grubuna,

Destek ve arkadaşlığı için Arş. Gör. Gülin Zeynep ÖZTAŞ'a,

Emeklerini hiçbir zaman ödeyemeyeceğim, her zaman yanımda olan, en değerlilerim canım annem Seda, babam Akif ve kardeşim Berat BEŞLER'e,

Birlikte çıktığımız hayat yolunda desteğini en derinden hissettiğim, tez sürecinde umutsuzluğa kapıldığım her an sabırla yanımda olan, eşim, yol arkadaşım Sercan DİNARCAN'a,

teşekkürlerimi borç bilirim.

İÇİNDEKİLER

Sayfa

ÇİZELGELER

Sayfa

SİMGELER VE KISALTMALAR

Simgeler

$\boldsymbol{Y}$ Sayım verisi bağımlı değişken vektörü

$\boldsymbol{X}$ Açıklayıcı değişkenler matrisi

$\boldsymbol{\mu}$ İlgili dağılımın ortalaması

$\boldsymbol{\beta}$ Bilinmeyen parametreler vektörü

$\boldsymbol{\sigma^2}$ Varyans

$\boldsymbol{\theta}$ Üstel dağılım ailesi parametresi

$\emptyset$ Yayılım parametresi

$\boldsymbol{w}$ Önsel ağırlık

$\boldsymbol{n}$ Gözlem sayısı

$\boldsymbol{\hat{\sigma}^2}$ σ^2'nin kestirimi

$\boldsymbol{\eta}$ Doğrusal önkestirici

$\boldsymbol{\omega}$ Yapısal sıfır oranı

$\boldsymbol{v}$ Gamma dağılımı ölçek parametresi

$\boldsymbol{\alpha}$ Gamma dağılımı parametresi

$\boldsymbol{\chi^2}$ Pearson Ki-kare

$\boldsymbol{G^2}$ Sapma İstatistiği

Kısaltmalar

MLE Maximum Likelihood Estimation

ZIP Zero Inflated Poisson

ZIB Zero Inflated Binomial

ZINB Zero Inflated Negative Binomial

ZIGP Zero Inflated Generalized Poisson

AIC Akaike Information Criterion

BIC Bayesian Information Criterion

LR Likelihood Ratio

GKT Genel Kareler Toplamı

RKT Regresyon Kareler Toplamı

AKT Artık Kareler Toplamı

TÜİK Türkiye İstatistik Kurumu

HİA Hanehalkı İşgücü Araştırması

1. GİRİŞ

Regresyon analizi, bir bağımlı değişken ile bir ya da birden fazla açıklayıcı değişken arasındaki ilişkiyi ölçmek için kullanılır. Değişkenler arasındaki sebep - sonuç ilişkisini verir.

Klasik doğrusal regresyon analizinde bağımlı değişkenin normal dağılıma sahip olması gerekmektedir. Ancak, bağımlı değişken kesikli değerlerden oluşabilir. Bağımlı değişkenin negatif olmayan kesikli değerlerden oluştuğu durumda sayma verisi modelleri kullanılabilir.

Sayma verisi, herhangi bir olayın bir süre aralığında kaç kez gerçekleştiğini gösterir. Poisson regresyon modeli, sayım verisi analizinde kullanılan en yaygın modellerden biridir. Poisson regresyon analizi ortalama ve varyansın eşitliğine dayanır ancak uygulamada bu eşitlik sağlanmayabilir. Varyansın ortalamadan büyük olduğu aşırı yayılım durumunda, negatif binom regresyon analizi veya genelleştirilmiş Poisson regresyon analizi uygulanmaktadır.

Sayma verisi için regresyon modelleri ile ilgili yapılmış çalışmaların bazıları aşağıda verilmiştir:

Shankar ve diğ. (1997); kaza frekanslarını sıfır yığılmalı Poisson, sıfır yığılmalı negatif binom regresyon modellerini kullanarak bir uygulama yapmışlar ve sıfır yığılmalı sayım verisi modelleri ile negatif binom regresyon modelini karşılaştırmışlardır[1].

Shoukri ve diğ. (2004); kümelenmiş sayım verilerinin analizi için Poisson Ters Gauss regresyon modelini kullanmışlardır. Süt ineklerinde meme iltihabı hastalığının seyrinin ilişkisini Poisson Ters Gauss ve negatif binom regresyon modelleri ile incelemişlerdir. Sonuçları en çok olabilirlik yöntemi ile karşılaştırmışlardır[2].

Sezgin ve Deniz (2004); Türkiye'de yapılan grevler üzerine model çalışması yapmışlardır. 1964 - 2000 yılları arası grev sayılarına etki eden faktörleri bulmak amacıyla işsizlik oranı sendikalaşma oranı, çalışan başına milli gelirin değişimini açıklayıcı değişken olarak kullanarak Poisson regresyon analizi yapmışlardır. Aşırı yayılımın varlığından dolayı negatif binom regresyon modeli uygulamışlardır [3].

Saraçbaşı ve Aktaş (2005); 1988 Türkiye Nüfus ve Sağlık Araştırması'ndan elde edilen yeni doğan çocukların ölümlerine ilişkin verileri analiz etmişlerdir. Türkiye’de 5 bölge ve belirli yaş gruplarına göre annenin ölü doğum yapma riskini araştırmak için sayılabilir verilerden oluşan ölü doğum sayısı bağımlı değişkenine aşırı yayılımın varlığından dolayı negatif binom regresyon uygulanmıştır [4].

Yeşilova ve diğ. (2006); Norduz erkek kuzularının çeşitli yaş dönemlerinde ölçülen üreme davranışı özelliklerini Poisson ve negatif binom regresyon modelleri kullanarak analiz etmişlerdir [5].

Özmen ve Famoye (2007); Türkiye Dalyan'da 1991-1993 yılları arasında güneşe maruz kalarak ölen Caretta Caretta yavrularının ölüm hızlarını modellemişlerdir. Bölge, denizden uzaklık, dönem açıklayıcı değişken olarak kullanılmıştır. Çok sayıda sıfır ve aşırı yayılım içeren zoolojik veri seti Poisson, negatif binom, genelleştirilmiş Poisson, sıfır yığılmalı Poisson ve genelleştirilmiş sıfır yığılmalı Poisson regresyon modelleri ile incelenmiştir. Seçim kriterleri sonucunda veri seti için genelleştirilmiş Poisson regresyon modelinin diğer sayım verisi modellerine göre daha iyi olduğu sonucuna varılmıştır [6].

Tüzel (2011); Karayolları Motorlu Araçlar Zorunlu Mali Sorumluluk Sigortası yaptıran sigortalıların yaş, cinsiyet, kullanım tipi ve kullanılan aracın il plaka kodu açıklayıcı değişken; bir yıllık sürede gerçekleşen hasar sayısı ise bağımlı değişken olmak üzere Poisson, negatif binom, genelleştirilmiş Poisson regresyon modelleri ile bu modellerin sıfır yığılmalı karmalarının gözlemlenen veri setine uygunluğu araştırmıştır [7]

Arı ve Önder (2012); Farklı veri yapılarında kullanılabilecek regresyon yöntemlerinden; loğrusal regresyon, lojistik regresyon, Negatif binom regresyon, Poisson regresyon, temel bileşenler regresyonu, probit regresyon, ridge regresyonu ve Cox regresyon yöntemlerinin hangi durumlarda kullanılabileceğin incelemişlerdir [8].

Asrul ve Naing (2012), Malezya'da Kelantan Bölgesi'nde 2000-2008 yılları arasında yaş gruplarına göre AIDS hastalarının ölüm hızlarını modellemişlerdir. Cinsiyet, uyruk, medeni durum, meslek ve bulaşma yolu açıklayıcı değişken olarak kullanılmıştır. Sıfır yığılmalı negatif binom regresyon ve negatif binom regresyon kullanılmış, seçim kriterleri ile modeller karşılaştırılıp en uygun model seçilmiştir [9].

Pamukçu ve diğ. (2014); Türkiye'deki 2001-2009 yılları arasında boşanma sayılarını, boşanmış çiftlerin evli kalma sürelerine göre dört gruba ayırarak genelleştirilmiş Poisson ve negatif binom regresyon modellerini uygulamışlardır. Bağımlı değişken yıllık boşanma sayısı; açıklayıcı değişkenler olarak erkek ve kadınların ilk evlilik yaşı ortalamaları, evli kadının iş hayatına katılma oranı, erkek ve kadınlarda yüksek okul mezunu olma oranı kullanılmıştır [10].

Elmalı (2014); 2010-2013 yıllarında illerin ilaç tüketiminde etkili olan faktörleri kantil regresyon ve negatif binom regresyon ile incelemiştir. İllerin nüfusu, illerde bulunan hekim sayısı, eczacı sayısı, nüfus artış hızı, denize olan konumu değişkenleri açıklayıcı değişkenler; illerin ilaç tüketimi ise bağımlı değişken olarak alınmıştır [11].

Açıkyürek (2016); Sosyal Güvenlik Kurumu'ndan alınan 2000 ile 2014 yılları arasındaki malulen emekli sayılarını kapsayan verileri Poisson, negatif binom, sıfır yığılmalı ve engel regresyon modelleri kullanılarak modellemiştir. Verilerde malulen emekli sayısı bağımlı değişken; yıl, yaş, cinsiyet, bağlanan aylık, çalıştığı gün sayısı ve emekli olduğu bölge bağımsız değişkenler olarak kabul edilmiştir [12].

Bu tez çalışmasında genelleştirilmiş doğrusal modelleri ve parametre tahmini hakkında bilgi verilmiştir. Sayma verilerinde kullanılan regresyon modelleri açıklanmış ve bu kapsamda Poisson regresyon modeli anlatılmıştır. Bağımlı değişkende yayılım olması durumunda kullanılan negatif binom regresyon ve genelleştirilmiş Poisson regresyon modelleri hakkında bilgi verilmiştir. Bağımlı değişkende beklenenden fazla sıfır olması durumunda kullanılan sıfır yığılmalı regresyon modellerinden bahsedilmiştir. Model seçiminde uyum iyiliği ölçümü ve seçim kriterleri hakkında bilgi verilmiştir.

Uygulama bölümünde ise, Türkiye İstatistik Kurumu tarafından yürütülen Hanehalkı İşgücü Araştırması verileri kullanılmıştır. Fertlerin haftalık ortalama çalışma saati verilerine Poisson regresyon, genelleştirilmiş Poisson regresyon, negatif binom regresyon analizleri uygulanıp en uygun model seçilmiştir. Analiz aşamasında SPSS, SAS ve R programları kullanılmıştır.

2. GENELLEŞTİRİLMİŞ DOĞRUSAL MODELLER

Genelleştirilmiş doğrusal modeller, değişkenler arasındaki ilişkilerin modellenmesi için kullanılan bir yöntemdir. Bu modeller, klasik doğrusal modellerin bazı kısıtlayıcı varsayımlarını esneterek normal dağılmayan verilerin analiz edilmesini sağlar. Genelleştirilmiş doğrusal modeller, Nelder ve Weddeburn'un makalesine [13] dayanır ve o zamandan beri çok farklı uygulama alanlarında kullanılmış yaygın istatistiklerin bir parçası haline gelmiştir.

Genelleştirilmiş doğrusal modeller, açıklayıcı değişkenler ve bağımlı değişken arasındaki ilişkiyi nicelendirmede kullanılır. Model, klasik doğrusal regresyon modelinden iki önemli noktada ayrılır :

- Genelleştirilmiş doğrusal modelde bağımlı değişken üstel dağılım ailesinden olmalıdır. Böylece bağımlı değişkenin dağılımı normal dağılım olmak zorunda değildir. Normal olmayan bir dağılım da olabilir.
- Bağımlı değişkenin ortalamasının dönüşümü açıklayıcı değişkenlerle doğrusal olarak ilişkilidir.

Bağımlı değişkenin üstel dağılım ailesinden gelmesinin bir sonucu olarak genellikle değişen-varyanslılık (heteroskedasticity) oluşabilmektedir. Böylece varyans ortalamaya bağlı olarak değişebilmektedir. Bu durum klasik doğrusal regresyon varsayımıyla çelişmektedir [14].

Genelleştirilmiş doğrusal modeller, klasik doğrusal modelleri de içeren pek çok modelden oluşmaktadır.

Genelleştirilmiş doğrusal modeller, üç temel bileşenden oluşmaktadır:

1. *Rasgele bileşen*: Y'nin bileşenleri $E(Y) = \mu$ ortalama ve (σ^2) sabit varyans ile normal dağılıma sahiptir.
2. *Sistematik bileşen:* x_1, ..., x_p değişkenlerinin birleşimi sonucu η doğrusal önkestiricisi üretilir:

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

3. *Bağ fonksiyonu:* Rasgele bileşen ve sistematik bileşen arasındaki ilişki

$$\mu = \eta$$

genellemesi ile gösterilir.

Doğrusal önkestirici için gösterilen yeni sembol η ile μ aynıdır. Bu durumda aşağıdaki eşitlik;

$$g(\mu_i) = \eta$$

yazılabilir. $g(.)$ *bağ fonksiyonu* olarak isimlendirilir. Bu fonksiyon monoton türevlenebilir bir fonksiyondur.

2.1. Genelleştirilmiş Doğrusal Modellerin Yapısı

Y' nin bileşenlerinin üstel dağılım ailesinden geldiği varsayılsın, üstel dağılım ailesinin genel formu,

$$f_Y\left(y;\theta,\phi\right) = \exp\{(y\theta - b(\theta))/a(\phi) + c\left(y,\phi\right)\} \tag{2.1}$$

olarak ifade edilmektedir.

$a(.)$, $b(.)$, $c(.)$ bazı özel fonksiyonlardır, üstel ailelere göre farklılık göstermektedir. ϕ, ölçek veya yayılım parametresi; θ ise standart (kanonik) parametredir.

Eğer ϕ biliniyorsa, dağılım θ parametreli üstel dağılım ailesidir. Eğer ϕ bilinmiyorsa, dağılım iki parametreli üstel dağılım ailesi olabilir.

Normal dağılım olasılık yoğunluk fonksiyonu,

$$f_y(y;\theta,\phi) = \frac{1}{\sqrt{(2\pi\sigma^2)}}\exp\{-(y-\mu)^2 / (2\sigma^2)\}$$

$$= \exp\{(y\mu - \mu^2)/(\sigma^2) - 1/2(y^2/\sigma^2 + \log(2\pi\sigma^2))\} \tag{2.2}$$

şekllindedir.

Böylece $\theta = \mu$ ve $\phi = \sigma^2$ olur ve

buradan $a(\phi) = \phi,\ b(\theta) = \theta^2/2\ , c(y,\phi) = -\frac{1}{2}\left\{y^2/\sigma^2 + \log(2\pi\sigma^2)\right\}$ sonucuna ulaşılır [15].

2.2. Genelleştirilmiş Doğrusal Modellerde Parametre Tahmini

Bağımlı değişkenin normal dağılımdan farklı dağılıma sahip olduğu regresyon modellerinde parametre tahmini için yaygın olarak En Çok Olabilirlik Tahmin Yöntemi (*Maximum Likelihood Estimation-MLE*) kullanılmaktadır. Bu yöntemde olabilirlik fonksiyonu kullanarak parametre tahmini yapılmaktadır. Olabilirlik fonksiyonu; $f(y,\theta)$ olarak gösterilmektedir. Ancak uygulamadaki kolaylığı nedeniyle olabilirlik fonksiyonu yerine, log olabilirlik fonksiyonu kullanılmaktadır [16].

*Y'*nin verildiği θ ve ϕ için log olabilirlik fonksiyonu, $L(\theta,\phi;y)= logf_Y(y;\theta,\emptyset)$ olarak yazılabilir [15].

Genelleştirilmiş doğrusal modeller için varyansın genel formu *Fisher Skorlaması* yardımıyla elde edilir. *Fisher Skorlaması*, olabilirlik denklemlerini çözmek için kullanılan alternatif iterasyon yöntemlerinden biridir.

*Fisher Skorlaması'n*da *y*'nin ortalama ve varyansı aşağıdaki eşitliklerden kolayca türetilebilir [17]:

$$E\left(\frac{\partial L}{\partial \theta}\right)=0 \tag{2.3}$$

$$E\left(\frac{\partial^2 L}{\partial \theta^2}\right)=-\mathrm{E}\left(\frac{\partial L}{\partial \theta}\right)^2 \tag{2.4}$$

Eş. (2.1) 'de verilen *y* üstel dağılım ailesi için log olabilirlik fonksiyonu;

$$L(\theta;\, y) \;=\{y\theta-\; b\left(\theta\right)\}/\;\, a(\phi)+\; c(y,\phi)$$

şeklindedir. Log olabilirlik fonksiyonunun θ' ya göre birinci ve ikinci türevleri alınırsa;

$$\frac{\partial L}{\partial \theta}=\left\{y-b^{'}\left(\theta\right)\right\}/\,a(\phi) \tag{2.5}$$

$$\frac{\partial^2 L}{\partial \theta^2}=-b^{''}(\theta)\,/\,a(\phi) \tag{2.6}$$

elde edilir.

Eş. (2.5)' i Eş. (2.3)'te yerine yazıp sıfıra eşitlenince;

$$E\left(\frac{\partial L}{\partial \theta}\right)=0$$

$$E\left(\frac{Y-b^{'}\left(\theta\right)}{a\left(\phi\right)}\right)=0$$

elde edilir.

$a(\phi)$ terimi sabit katsayı olduğundan $a(\phi)=1$ kabul edilebilir. Buradan,

$$E(Y) - b'(\theta) = 0$$

ile genelleştirilmiş doğrusal modeller için beklenen değerin genel formu,

$$E(Y) = b'(\theta) = \mu$$

şeklinde elde edilir.

Eş. (2.5) ve (2.6), Eş. (2.4)'te yerine yazılır,

$$E\left(\frac{\partial^2 L}{\partial \theta^2}\right) + \mathrm{E}\left(\frac{\partial L}{\partial \theta}\right)^2 = 0$$

denklem çözülürse

$$E\left(-\frac{b''(\theta)}{a(\phi)}\right) + E\left(\frac{Y - b'(\theta)}{a(\phi)}\right)^2 = 0$$

elde edilir. Böylece,

$$-\frac{b''(\theta)}{a(\phi)} + \frac{V(Y)}{a^2(\phi)} = 0$$

ile genelleştirilmiş doğrusal modeller için varyansın genel formu,

$$V(Y) = b''(\theta)a(\phi)$$

biçiminde elde edilir. Buradan varyansın, ortalamanın bir fonksiyonu olduğu sonucuna ulaşılır.

$a(\phi)$ fonksiyonunun genel olarak gösterimi;

$$a(\phi) = \phi / w$$

şeklindedir. Burada ϕ parametresi, σ^2 olarak gösterilen yayılım parametresi; w ise gözlemden gözleme değişen önsel ağırlıktır [15] .

Üstel dağılım ailesinin bazı dağılımları ve fonksiyonları Çizelge 2.1'de verilmiştir.

Çizelge 2.1. Üstel Dağılım Ailesinin Bazı Dağılımları ve Fonksiyonları

Dağılım	$\boldsymbol{a(\phi)}$	$\boldsymbol{b(\theta)}$	$\boldsymbol{c(y,\phi)}$
Normal	ϕ / w	$\theta^2/2$	$-1/2(wy^2)/\phi + \ln(2\pi\phi/w)$
Poisson	ϕ / w	$\exp(\theta)$	$-lny!$
Binom	ϕ / w	$mln(1+e^{\theta})$	$\ln\binom{m}{y}$
Gamma	ϕ / w	$-\ln(\theta)$	$(w/\phi)\ln(wy/\phi) - lny - \ln(\Gamma(\mathrm{w}/\phi)$

Burada π parametresi, ilgili dağılımın ortalamasını (μ) ifade etmektedir [7].

2.3. Bağ Fonksiyonu

Bağ (*Link*) fonksiyonu, doğrusal önkestiricisi (η) ile bir y verisinin beklenen değerini (μ) ilişkilendirir.

Bağ fonksiyonu, $\eta = g(\mu)$ eşitliği ile gösterilir. Klasik doğrusal modellerde, bağımlı değişkenin ortalaması ve doğrusal tahmin edici eşittir, η ve µ her ikisi de reel düzlemde tüm değerleri alabilirler. Klasik doğrusal modellerde birim (*identity*) bağ fonksiyonu kullanılmaktadır. Binom dağılımında, $0 < \mu < 1$ olduğundan (*0,1*) aralığını bağ fonksiyonunu reel düzleme taşıması gerekmektedir. Bağ fonksiyonu monoton türevlenebilir bir fonksiyondur [15] .

Çizelge 2.2. Üstel Dağılım Ailesinin Bazı Dağılımları için Bağ Fonksiyonları

Dağılım	**Bağ Fonksiyonu**	$\boldsymbol{g(\mu)}$
Normal	Birim (identity)	μ
Poisson	Log	$\ln(\mu)$
Binom	Logit	$\ln(\frac{\pi}{1-\pi})$
Gamma	Ters	$1/\mu$

3. SAYMA VERİ REGRESYON MODELLERİ

Sayma verisi, herhangi bir olayın belirli bir süreçte kaç kez meydana geldiğini gösterir. Sayma modellerindeki ilk gelişmeler aktüerya, biyoistatistik ve demografide gözlemlenmiştir. Bu modeller son yıllarda, ayrıca ekonomi, siyaset bilimleri ve sosyolojide de geniş bir biçimde kullanılmıştır.

Sayma verisi için regresyon modellerinin geliştirilmesindeki dönüm noktası, Poisson regresyonun özel bir durumu olduğu "Genelleştirilmiş Doğrusal Modellerin" ortaya çıkmasıdır. İlk olarak Nelder ve Wedderburn tarafından tanımlanmış [13], McCullagh ve Nelder [15] tarafından detaylandırılmıştır.

Sayma veri regresyon modelleri kesikli regresyonun özel bir türüdür. Negatif olmayan kesikli değerlerden oluşur. Bu verilerde doğrusal regresyon modelinin uygulanması tahmin edilen katsayıların yanlı olmasına neden olmaktadır. Sayım verisi için kullanılan en yaygın regresyon modeli "***Poisson Regresyon***", varsayımların sağlanmadığı durumda ise "***Negatif Binom Regresyon***" modelidir [18] .

Bu bölümde sayma verisinde kullanılan çeşitli modeller ele alınacaktır.

3.1. Poisson Regresyon

Poisson regresyon, bağımsız değişkenler ile sayım verisinden oluşan bağımlı değişken arasındaki ilişkiyi açıklayan bir modeldir. Poisson regresyonda, bağımlı değişken Y_i, ($i=1,2,..,n$) kesikli-bağımsız-Poisson raslantı değişkenidir. Bağımlı değişkenin kesikli olması nedeniyle, klasik regresyon analizi varsayımları sağlanamaz. Bu durumda, uygun analiz yöntemlerinden biri de Poisson regresyon yöntemidir [3].

Poisson dağılımı belirli bir zamanda, belirli bir alanda veya hacimde gerçekleşebilecek olay sayısını tanımlayan bir dağılımdır. Poisson dağılımı çoğunlukla nadir olayların oluş sayısını modellemek için kullanılmaktadır. Bir saat aralığında belli bir hastaneye gelen hastaların sayısı, belli bir trafik kavşağından bir dakika içinde geçen otomobil sayısı örnek olarak verilebilir [12].

Poisson dağılımı ilk defa Siméon-Denis Poisson (1781–1840) tarafından diğer olasılık hakkındaki yazıları ile birlikte 1838'de yayınlanan *Recherches sur la probabilitédesjugements en matièrescriminelles et matièrecivile* ("Ceza hukuku ve medeni hukuk alanlarındaki hükümlerin olasılığı üzerinde araştırmalar") adındaki eserinde ortaya atılmıştır [19].

Kesikli Y rastlantı değişkeni $\mu > 0$ parametresi ile Poisson dağılımına sahip olduğunda, dağılımına ilişkin olasılık fonksiyonu;

$$P(y_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, \qquad y_i = 0,1,2,\ldots \tag{3.1}$$

şeklinde gösterilir.

Poisson dağılımının en önemli özelliği, dağılımın ortalama ve varyansının eşit olmasıdır [18] :

$$E(Y_i) = V(Y_i) = \mu \tag{3.2}$$

Poisson regresyon modeli, Poisson dağılımından türetilen doğrusal olmayan bir modeldir.

x_i ' ye bağlı y_i için Poisson regresyon modeli;

$$P(y_i \mid x_i) = \frac{e^{\mu_i}\mu_i^{y_i}}{y_i!}, \; y_i = 0,1,2,\ldots \tag{3.3}$$

biçimindedir. Bu model için ortalama parametresi;

$$E(Y_i \mid x_i) = \mu_i = \exp(x_i^{'}\beta) \tag{3.4}$$

şeklinde gösterilir ve "üstel ortalama fonksiyonu" olarak ifade edilir. İstatistik literatüründe bu fonksiyon ayrıca "*log-doğrusal fonksiyon*" olarak da ifade edilir. Çünkü koşullu ortalamanın logaritması, parametreleri doğrusal olarak vermektedir [20].

$$\ln E(Y_i \mid x_i) = \mu_i = x_i^{'}\beta \tag{3.5}$$

Poisson regresyon modelinde regresyon sürecindeki genel kestirimler en çok olabilirlik yöntemi (*MLE*) ile gerçekleştirilmektedir. Poisson en çok olabilirlik kestirimi için;

- Koşullu ortalamanın doğru tanımlanmasında bağımlılık şartı sağlanmalıdır. Ayrıca bağımlı değişken y 'nin Poisson dağılması gereklidir.
- En çok olabilirlik standart hataları ve t istatistikleri kullanarak hesaplanan istatistiksel sonuçlar, hem koşullu ortalama, hem varyansın doğru tanımlanmasını gerektirmektedir. Burada istenen koşul, koşullu varyans ve ortalamanın eşit olmasıdır.
- Veriler için koşullu varyans ve koşullu ortalamanın eşit olmaması durumunda, en çok olabilirlik yönteminin uygulanması ile elde edilmiş istatistiksel sonuçlar,

koşullu ortalamanın doğru tanımlandığının ispat edildiği durumlarda geçerli ve doğrudur.

- Veriler için koşullu varyans ve ortalamanın eşit olmaması durumunda, Poisson en çok olabilirlik tahmin edicisinden daha etkin tahmin ediciler kullanılabilir [20] .

Poisson regresyonda bağımlı değişkenlerin doğrusal yapısı logaritmik dönüşüm ile verilmektedir:

$$Log(\mu) = X'\beta + e \tag{3.6}$$

Poisson regresyonda ilgilenilen olayın sayısı olan Y bağımlı değişkeninin Poisson dağılışına sahip olduğu varsayılmaktadır. Bu durumda Poisson ortalaması olan μ'nün logaritmasının bağımsız değişkenlerin bir doğrusal fonksiyonu olduğu varsayılmaktadır.

Bu fonksiyon,

$$Log(\mu) = b_0 + b_1x_1 + \cdots + b_mx_m \tag{3.7}$$

olarak gösterilmektedir [11].

3.1.1. Aşırı Yayılım

Poisson regresyon modelinin en önemli özelliği bağımlı değişkenin beklenen değer ve varyansının eşit olduğu anlamına gelen eşit yayılım özelliğidir. Pratikte ortalama ve varyansın eşitliği nadiren gerçekleşen bir durumdur. Eğer Poisson modelde, varyans ortalamadan daha büyükse yani beklenenden daha fazla bir değişkenlik varsa bu duruma *aşırı yayılım (overdispersion)* adı verilir.

Aşırı yayılım varsa ve bu durum hesaba katılmadan parametre tahminleri yapılırsa bu durum parametre tahminlerinin standart hatalarının olduğundan daha düşük çıkmasına neden olur ve bunun sonucunda model için açıklayıcı değişkenlerin seçiminde yanlışlık yapılmış olur [21].

Bu durumda, veri kümesine Poisson regresyon yerine yayılım parametresini (*dispersion parameter*) içeren regresyon modelleri kullanılmalıdır. Poisson regresyonda oluşan aşırı yayılımı açıklayan iki yaklaşım vardır. Bu yaklaşımlardan biri, yarı olabilirlik yaklaşımı, diğer yaklaşım ise karma Poisson model yaklaşımıdır [22].

Negatif binom regresyon, karma Poisson yaklaşımlarından biridir.

3.1.2. Aşırı Yayılım Testleri

Poisson regresyon modelinin varsayımlarından biri eşit yayılımdır (equidispersion). Eşit yayılım sağlanmadığı durumda, yayılım durumu ortaya çıkmaktadır. Verinin aşırı yayılımlı mı az yayılımlı mı olduğuna karar vermek için kurulan Poisson modelinin varyansı ortalamasına oranlanır. Bu oran 1'den büyük ise aşırı yayılım, küçük ise az yayılım durumu söz konusudur.

Aşırı yayılımın varlığını belirleyebilmek için bazı test istatistikleri geliştirilmiştir. Skor Testleri, gözlemlenen verinin, Poisson regresyonuna göre aşırı yayılım gösterip göstermediğini belirlemede kullanılır [23].

Büyük örneklemler için önerilen *kısmi skor test istatistiği* (partial score);

$$T = \frac{1}{2}\sum_{i=1}^{n}\left\{(Y_i - \hat{\mu}_i)^2 - Y_i\right\} \tag{3.8}$$

biçimindedir.

T istatistiğinin standartlaştırılmış şekli olarak ifade edilen T_1 test istatistiği;

$$T_1 = \frac{\sum_{i=1}^{n}\left\{(Y_i - \hat{\mu}_i) - Y_i\right\}}{(2\sum_{i=1}^{n}\hat{\mu}_i^{\,2})^{\frac{1}{2}}} \tag{3.9}$$

şeklindedir.

Eşitlik (3.8) ve Eş.(3.9)'da $\hat{\mu}_i$ terimi, Poisson modeli altında μ_i'nin en çok olabilirlik tahmin edicisini göstermektedir [24].

T_1 test istatistiği, yeterince büyük gözlem sayısı için standart normal dağılıma yakınsadığından bu istatistik için standart normal dağılım tablo değeri ile karşılaştırılarak aşırı yayılımın varlığı test edilmektedir [7].

Negatif binom modelinin yayılım parametresinin (α) sıfıra ve genelleştirilmiş Poisson modelinin yayılım parametresinin (φ) bire eşit olması, gözlemlenen verinin aşırı yayılım göstermediğini belirtmektedir. Bu durumda Poisson regresyonunun kullanılması yeterlidir. Her iki modele ait yayılım parametrelerinin belirtilen değerlere eşitliği hipotez testleri kullanılarak test edilmektedir.

Negatif binom modelinin yayılım parametresi için kullanılan yokluk hipotezi ve seçenek hipotezler sırasıyla;

$H_o : \alpha = 0$

$H_1 : \alpha > 0$

şeklindedir.

Genelleştirilmiş Poisson modelinin yayılım parametresi için kullanılan yokluk hipotezi ve seçenek hipotez sırasıyla;

$H_o : \varphi = 1$

$H_1 : \varphi > 1$

şeklindedir.

Yokluk hipotezlerinin reddedilmesi, yayılımın olduğunu gösterir. Poisson modeli yerine negatif binom modeli veya genelleştirilmiş Poisson modelinin kullanılması uygundur [7].

Yokluk hipotezinin testinde *Olabilirlik Oran Testi, Wald Testi* kullanılabilir.

Olabilirlik Oran Testi

Olabilirlik Oran Testi, Poisson veya Negatif Binom regresyon modellerinde aşırı yayılımın varlığını test etmede kullanılır. Olabilirlik oran test istatistiği;

$$LR = 2\,(\,lnL_1 - lnL_0\,) \tag{3.10}$$

olarak hesaplanır.

Eş. (3.10)'da yer alan L_0: Yokluk hipotezi doğru olduğu varsayımı altında hesaplanan en çok olabilirlik fonksiyonu, L_1: Seçenek hipotezin doğru olduğu varsayımı altında hesaplanan en çok olabilirlik fonksiyonudur.

Olabilirlik oran test istatistiği, 1 serbestlik dereceli olan ki-kare dağılımına sahiptir.

Wald Testi

Wald testi, aşırı yayılımın varlığını test etmede kullanılır. Yayılım parametresi α ile ifade edildiğinde Wald test istatistiği;

$$W = \frac{\hat{\alpha}^2}{V(\hat{\alpha})} \tag{3.11}$$

olarak hesaplanır.

$\hat{\alpha}$: yayılım parametresinin tahmini, $V(\hat{\alpha})$: tahminin varyansını göstermektedir. Wald test istatistiği, 1 serbestlik dereceli olan ki-kare dağılımına sahiptir [25].

3.2. Negatif Binom Regresyon Modeli

Negatif binom dağılımı, ilk olarak 1679 yılında Pascal tarafından ele alınmıştır. Montmort negatif binom dağılımını bozuk paranın çevrilmesinde belli bir sayıya ulaşmak için kullanmıştır. 1907'de William Gosset adlı bir öğrenci, negatif binom dağılımını Poisson dağılımına alternatif olarak kullanmıştır. Daha sonra, Greenwood ve Yule ve Eggenberger ve Palyan negatif binom dağılımı uygulamalarının temelini atmışlardır [26].

1940'larda sayım modellerinin en orijinal çalışmaları George Beall, Anscombe ve Bartlett tarafından yapılmıştır. 1960 yılında Leroy Simon, Poisson ve negatif binom modellerini ayıran ilk çalışmasının ardından, negatif binom için en çok olabilirlik algoritmasını yayınlayan ilk kişidir.

Simon, Poisson ve negatif binom dağılımlarını sigorta verilerine uyarlamış aktüerya bilim adamıdır.

1972 yılında John Nelder ve R.W.M. Wedderburn'un [13] genelleştirilmiş doğrusal modelleri tanımlamasıyla model üzerinde durulan nokta, açıklayıcı değişkenleri içeren doğrusal olmayan modellerin kurulması olarak değişmiştir. 1982'de Nelder, Peter McCullagh ile bir araya gelerek Genelleştirilmiş Doğrusal Modeller'in (*Generalized Linear Models*) ilk baskısını yazmış, burada negatif binom regresyonu özetle tanımlamıştır. 1986 yılında Mullahy geometrik engel modellerini geliştirmiştir. Consul ve Felix Famoye [21], genelleştirilmiş en çok olabilirlik yöntemini kullanarak genelleştirilmiş negatif binom modellerinin çeşitli formlarını geliştirmişlerdir.

Negatif binom modeli, çoğu regresyon modelinde olduğu gibi bir olasılık dağılım fonksiyonu merkezlidir. Geleneksel negatif binom Poisson-gamma karma dağılımından türetilmiştir [27]. Negatif binom dağılımı, varyansın ortalamadan gözle görülür bir biçimde daha büyük olduğu durumda sayım verilerinin ele alınmasında Poisson dağılımına alternatif olarak yaygın biçimde kullanılır [28].

Negatif binom olasılık dağılım fonksiyonu, ortalaması 1 olan gamma karışımlı parametreli Poisson modeli ile aşağıdaki gibi türetilebilir:

$$f\left(y;\mu;u\right)=\frac{e^{-\lambda_i u_i}\left(\lambda_i u_i\right)^{y_i}}{y_i!} \tag{3.12}$$

Gamma karışımlı Poisson modeli, aşırı yayılım ya da ilişkili Poisson sayımlarını içerir . y’nin dağılımı Poisson, koşullu ortalama ve varyansı λ’dır. Böylece gamma karışımı altında *y’*nin koşullu ortalaması sadece λ değil, λu olarak ifade edilir. Sonuç olarak *y*’nin dağılımı aşağıdaki gibidir:

$$f\left(y;x;u\right)=\int_0^\infty \frac{e^{-\lambda_i u_i}\left(\lambda_i u_i\right)^{y_i}}{y_i!}g\left(u_i\right)\partial u_i \tag{3.13}$$

y’nin koşulsuz dağılımı $g(u_i)$’yi nasıl tanımladığımız ile belirlenir. Bu model için gamma dağılımında $u = exp(\varepsilon)$ olarak verilir, burada $ln(\mu) = x\beta + \varepsilon$ ‘dir.

Gamma dağılımında ortalamaya 1 verildiğinde,

$$f\left(y;x;u\right)=\int_0^\infty \frac{e^{-\lambda_i u_i}\left(\lambda_i u_i\right)^{y_i}}{y_i!}\frac{v^v}{\Gamma(v)}u_i^{\,v-1}e^{-vu_i}du_i$$

$$=\frac{\left(\lambda_i\right)^{y_i}}{\Gamma\left(y_i+1\right)}\frac{v^v}{\Gamma(v)}\int_0^\infty e^{-(\lambda_i+v)u}\,u_i^{(y_i+v)-1}du_i \tag{3.14}$$

eşitliği elde edilir.

Eş. (3.14)’teki integral çözümlendiğinde,

$$f\left(y;x;u\right)=\frac{\Gamma\left(y_i+v\right)}{\Gamma\left(y_i+v\right)\Gamma\left(v\right)}\left(\frac{1}{1+\lambda_i/v}\right)^v\left(1-\frac{1}{1+\lambda_i/v}\right)^{y_i}$$

sonucuna varılır.

v, gamma ölçek parametresidir. Negatif binom heterojenliğine ya da aşırı yayılım parametresi olan $\boldsymbol{\alpha}'$ nın tersine eşittir. ($v = 1/\alpha$)

λ ve μ eşitlendikten sonra oluşan negatif binom olasılık yoğunluk fonksiyonu

$$f\left(y;\mu;\alpha\right)=\frac{\Gamma\left(y_i+1/\alpha\right)}{\Gamma\left(y_i+1\right)\Gamma\left(1/\alpha\right)}\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\left(1-\frac{1}{1+\alpha\mu_i}\right)^{y_i} \tag{3.15}$$

biçiminde elde edilir.

Eş. (3.15)'te,

- $\Gamma\left(y+1\right)=y!$

* $\Gamma(y+1/\alpha)=(y+1/\alpha-1)!$
* $\Gamma(1/a)=(1/a-1)!$

olarak verildiğinde negatif binom dağılımının sık kullanılan bir formu aşağıdaki gibi elde edilir:

$$\frac{\Gamma(y_i+1/a)}{\Gamma(y_i+1)\Gamma(1/a)}=\frac{\left(y_i+\dfrac{1}{\alpha}\right)!}{y_i!(1/a-1)!}=\begin{pmatrix} y+\dfrac{1}{\alpha}-1 \\ \dfrac{1}{\alpha}-1 \end{pmatrix} \tag{3.16}$$

Eş. (3.15)'teki ilk terim yerine Eş. (3.16)'da yer alan en sağdaki terim kullanıldığında, negatif binom dağılımının yaygın kullanımı olan Eş. (3.17) elde edilir:

$$f(y;\mu,\alpha)=\begin{pmatrix} y+\dfrac{1}{\alpha}-1 \\ \dfrac{1}{\alpha}-1 \end{pmatrix}\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \tag{3.17}$$

Eş. (3.17)'deki dağılımın ortalaması ve varyansı sırasıyla;

$$\begin{aligned} E(Y_i) &= \mu_i \\ V(Y_i) &= \mu_i+\alpha\mu_i^2 \end{aligned} \tag{3.18}$$

biçimindedir [27].

3.2.1. Negatif Binom Regresyon Modeli Türleri

Sayım verilerini modellemek için pek çok negatif binom regresyon türü vardır. Geometrik model, α = 1 için negatif binomun bir çeşididir. Poisson da $\alpha \to 0$' a yaklaşırken negatif binomun bir türü olarak kabul edilebilir. Çizelge 3.1.'de negatif binom regresyonun bazı türleri listelenmiştir.

Çizelge 3.1. Negatif Binom Regresyon Türleri

1) NB2 V= $\mu + \alpha\mu^2$
2) NB1 V= $\mu + \alpha\mu$
3) NB-C < Kanonik>
4) Kesikli-budanmış- NB;NB1;NB-C
5) Sıfır – yığılmalı NB (ZINB)
6) Sansürlü NB
7) NB - Engel
8) NB-P
9) NB-H <heterojen>
10) Sonlu karma modeller
11) Koşullu sabit etkiler
12) Rasgele etkiler
13) Karma ve çok düzeyli etkiler
14) Endojen tabakalı NB
15) Basit seçim NB
16) Gizli sınıf NB
17) Kantil sayım
18) Momentlerin genelleştirilmiş metodu
19) Araç değişkenler
20) İki değişkenli NB

[27]

Cameron ve Trivedi tarafından 1986 'da negatif binomun genelleştirilmiş hali μ_i ortalamalı ve $\mu_i + \alpha{\mu_i}^p$ varyans fonksiyonlu olarak tanımlanmıştır [18] ve *NBp* olarak gösterilir. Genelleştirilmiş negatif binom dağılımı p = 2 için *NB2*, p=1 için *NB1* modeli olarak adlandırılmaktadır [29].

Negatif binom dağılımına sahip y_i'nin koşullu varyansının genel gösterimi,

$$\omega_i = V[Y_i \mid x_i] \tag{3.19}$$

biçimindedir [18] .

Standart negatif binom regresyon modeli, varyansın $\mu + \alpha\mu^2$ şeklinde yani kuadratik (2.dereceden) olmasından dolayı *NB2* modeli olarak da adlandırılır. *NB2* modeli, üstel dağılım ailesinin üyesi olan Poisson – gamma karışımlı bir modeldir. *NB2* modelini *NB1* modelinden ayıran özellik, *NB1* modelinin varyansının $\mu + \alpha\mu$ şeklinde olmasıdır. *NB1* modeli de Poisson – gamma karışımlı bir modeldir [28].

NB2 Modeli ve Parametre Tahmini

Negatif binom dağılımının yaygın kullanımı *NB2* modelidir. Ortalaması μ, varyansı $\mu + \alpha\mu^2$'dir. *NB2* modelinin varyansı kuadratiktir. y_i' nin varyansı için genel gösterimi $\omega_i = \mu_i + \alpha\mu_i{}^2$ biçimindedir.

NB2 modelinin olasılık yoğunluk fonksiyonu;

$$f(y;\mu,\alpha) = \frac{\Gamma(y+\alpha^{-1})}{\Gamma(y+1)\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{a^{-1}+\mu}\right)^{\alpha^{-1}}\left(\frac{\mu}{\alpha^{-1}+\mu}\right)^{y}$$

$$\alpha \geq 0,\; y = 0,1,2,\ldots \tag{3.20}$$

olarak ifade edilir.

Eş. (3.20)'de yer alan $\Gamma(.)$ fonksiyonu, y tam sayı ise;

$$\frac{\Gamma(y+\alpha)}{\Gamma(\alpha)} = \prod_{j=0}^{y-1}\ln(j+\alpha) \tag{3.21}$$

şeklinde yazılır.

Eş. (3.20)'de yer alan ifadenin logaritması aşağıdaki gibi yazılabilir:

$$\frac{\Gamma(y+\alpha^{-1})}{\Gamma(\alpha^{-1})} = \sum_{j=0}^{y-1}(j+\alpha^{-1}) \tag{3.22}$$

Eş. (3.22)'deki ifadenin Eş.(3.21)'deki gösterime uyarlanarak yazılmasıyla log-olabilirlik fonksiyonunun açık hali aşağıdaki gibi yazılabilir:

$$\ln L(\alpha,\beta) = \sum_{i=1}^{n}\left\{\left(\sum\nolimits_{j=0}^{y_i-1}\ln(j+a^{-1})\right) - \ln y_i! - (y_i+\alpha^{-1})\ln\left(1+\alpha\exp(x_i^{'}\beta)\right) + y_i\ln\alpha + y_i x_i^{'}\beta\right\} \tag{3.23}$$

Üstel ortalama $\mu_i = \exp(x_i' \; \beta)$ olarak yazılmıştır [18] .

Eş. (3.23)'ün çözümü için *En Çok Olabilirlik Yöntemi* kullanılır. Bu yöntem Bölüm 2.2'de verilmiştir.

NB2 modelinde $\hat{\beta}$, α parametrelerin tahmini için en çok olabilirlik fonksiyonu aşağıda verilmiştir [18] ;

$$\sum_{i=1}^{n} \frac{y_i - \mu_i}{1+\alpha\mu_i} x_i = 0 \tag{3.24}$$

$$\sum_{i=1}^{n} \left\{ \frac{1}{\alpha^2} \left(\ln\left(1+\alpha\mu_i\right) - \sum_{j=0}^{y_i - 1} \frac{1}{\left(j+\alpha^{-1}\right)} \right) + \frac{y_i - \mu_i}{\alpha\left(1+\alpha\mu_i\right)} \right\} = 0 \tag{3.25}$$

Ortalama parametresi üstel olması sonucu log-olabilirlik fonksiyonları doğrusal olmadığından, en çok olabilirlik yöntemi ile tek adımda sonuca ulaşılamaz. Bu nedenle parametre kestirimleri iteratif olarak elde edilmektedir. Eş. (3.24)ve (3.25)'in çözülmesi sonucu α ve β parametre tahminlerine ulaşılır [3].

3.3. Genelleştirilmiş Poisson Regresyon

Genelleştirilmiş Poisson regresyon (GPR), sayım verileri analizinde aşırı yayılım veya az yayılım durumlarında Poisson regresyona alternatif olarak Consul ve Famoye [21] tarafından önerilmiştir. Model, araştırmacılar tarafından ayrıntılı bir şekilde üzerinde çalışılan genelleştirilmiş Poisson dağılımına dayanmaktadır. Genelleştirilmiş Poisson regresyon modeli doğurganlık verilerini modellemede ve yaralanma verilerini modellemede kullanılmıştır [30].

$\boldsymbol{Y}$, (*nx1*) boyutlu, $y_1, y_2, \ldots, y_n$ 'den oluşan sayım verileri bağımlı değişken vektörü,

$\underline{\boldsymbol{X}}$, (*nxk*) boyutlu, $x_{i1} = 1, x_{i2}, x_{i3}, \ldots, x_{ik}$' dan oluşan açıklayıcı değişkenler girdi matrisi,

$\boldsymbol{\beta}$, (*kx1*) boyutlu regresyon parametreleri vektörü olsun.

Genelleştirilmiş Poisson dağılımında, ortalaması μ, $\mu = \theta(1-\lambda)^{-1} = \theta\varphi$ olarak verildiğinde genelleştirilmiş Poisson regresyon modeli:

$$P(Y_i = y_i) = \begin{cases} \mu_i \left[\mu_i + (\varphi - 1) y_i\right]^{y_i - 1} \varphi^{-y_i} e^{-\left[\left(\mu_i + (\varphi-1) y_i\right)/\varphi\right]} / y_i!, & y_i = 0,1,2,\ldots \\ 0, & \textit{diğer durumlarda} \end{cases} \tag{3.26}$$

şeklinde verilmektedir.

Genelleştirilmiş Poisson dağılımına ilişkin ortalama ve varyans,

$$E(Y_i \mid \underline{x}_i) = \mu_i \tag{3.27}$$

$$V(Y_i \mid \underline{x}_i) = \mu_i \varphi^2 \tag{3.28}$$

biçimindedir.

Eşitlik (3.26)'da φ genelleştirilmiş Poisson için yayılım parametresidir. $\varphi = 1$ olduğunda GPR modeli, Poisson regresyon modeline dönüşmektedir. $\varphi > 1$ olduğunda GPR modelinde sayım verileri aşırı yayılımlı olur. $\mu > 2,\ 1/2 \leq \varphi \leq 1$ durumunda ise sayım verilerinde az yayılım gözlemlenir [21] .

3.4. Katsayıların Yorumlanması

Regresyon katsayılarının yorumlanması önemli bir meseledir. Üstel dağılımlarda bu yorum doğrusal modellere göre değişmektedir.

Üstel dağılımın koşullu ortalaması $E[Y \mid x] = \exp(\boldsymbol{x'\beta})$ olarak gösterilebilir. (Gösterim için i alt indis düşürülmüştür.)

Üstel dağılım için, x_j, j. bağımsız değişkeni göstersin;

$$\frac{\partial E\left[Y \mid x\right]}{\partial x_j} = \beta_j \exp(\boldsymbol{x'\beta})$$

olarak yazılır.

Örneğin, $\widehat{\beta_j} = 0{,}2$ ve $\exp(x'\beta) = 2{,}5$ ise $\beta_j \exp(\boldsymbol{x_i'} \ \boldsymbol{\widehat{\beta}})$ çarpımından 0,2*2,5= 0,5 elde edilir. Yani, j. bağımsız değişkendeki bir birimlik değişim y' nin beklenen değerini 0,5 birim artırır [18] .

3.5. Sıfır Yığılmalı Regresyon Modelleri

Veride çok sayıda sıfır değerinin mevcut olması sıfır değer ağırlıklı yayılım (zero inflation) olarak tanımlanmaktadır. Gözlemlerin büyük bir kısmının sıfır olduğu veri kümelerinde, standart regresyon yöntemlerin uygulanması doğru olmayan sonuçların elde edilmesine, yanlı parametre tahminlerine yol açabilir.

Veri kümesinde beklenenden fazla sıfır değerin bulunması durumunda veri kümesinin sıfırları göz önünde bulunduran sıfır yığılmalı modeller (*zero-inflated models*) ile analiz edilmesi daha uygun olmaktadır [31].

Sıfır yığılmalı modeller, iki durumlu süreçler (dual state process) olarak da adlandırılmaktadır [32].

Sıfır yığılmalı modellerin incelenmesi için birçok regresyon modeli mevcuttur. Lambert [33] tarafından imalattaki kusur uygulamaları için sıfır yığılmalı Poisson regresyon modeli, Hall [34] tarafından sıfır yığılmalı binom regresyon modeli (*ZIB*), Lee [35] tarafından sıfır yığılmalı genelleştirilmiş Poisson regresyon modeli (*ZIGP*) önerilmiştir. Ridout ve diğ. [36] tarım, ekonometri, türlerin çeşitliliği, tıp gibi disiplinleri içeren sıfır yayılımlı veri örneklerinden bahsetmiştir [30]. Famoye ve Singh sıfır yayılımı ve Poisson dağılımını tek dağılımda ele alan sıfır yığılmalı genelleştirilmiş Poisson regresyon önermişlerdir [25].

Sıfır yığılmalı regresyon modeli,

$$P(Y_i = y_i) = \begin{cases} \omega_i + (1-\omega_i)P(J_i = 0), & y_i = 0 \\ (1-\omega_i)P(J_i = y_i), & y_i > 0 \end{cases} \tag{3.27}$$

şeklindedir.

J_i rastlantı değişkeni negatif olmayan kesikli dağılıma sahiptir. Sıfır yığılmalı modeller için beklenen değer ve varyans sırasıyla;

$$E(Y_i \mid x_i) = (1-\omega_i)E(J) \tag{3.28}$$

$$\begin{aligned} V(Y_i \mid x_i) &= E\left[J^2\right] + \left(E\left[J\right]\right)^2 \\ &= (1-\omega_i)\left\{V(J) + \omega_i\left(E\left[J\right]^2\right)\right\} \end{aligned} \tag{3.29}$$

biçimindedir [18] .

3.5.1. Sıfır Yığılmalı Poisson Regresyon Modeli

Sıfır yığılmalı Poisson regresyon (*ZIP*), araştırmacılar tarafından sıfır değerlerinin çok olduğu sayım verilerini incelemek için kullanılmıştır. Sıfır yığılmalı Poisson regresyon, hem sıfır sürecinin hem de Poisson dağılımındaki açıklayıcı değişkenlerin katılmasına izin vererek sıfır bozulumlu dağılım ile Poisson dağılımı karışımından elde edilebilir.

Sıfır yığılmalı Poisson regresyon için olasılık fonksiyonu;

$$P(Y_i = y_i) = \begin{cases} \omega_i + (1-\omega_i)\exp(-\mu_i), & y_i = 0 \\ (1-\omega_i)\dfrac{\mu_i^{y_i}}{y_i!}\exp(-\mu_i), & y_i > 0 \end{cases} \tag{3.30}$$

şeklindedir.

$0 \leq \omega_i < 1$ ve $\mu_i > 0$ varsayımı ile beklenen değer ve varyans sırasıyla;

$$E(Y_i) = (1-\omega_i)\mu_i \tag{3.31}$$

$$V(Y_i) = (1-\omega_i)\mu_i(1+\omega_i\mu_i) \tag{3.32}$$

olarak gösterilir.

Sıfır yığılmalı Poisson regresyonda,

- $\omega_i = 0$ olduğunda model Poisson regresyona indirgenir.
- $\omega_i > 0$ olduğunda ise aşırı yayılıma gider.

Sıfır yığılmalı modellerde, aşırı yayılımın da dahil edildiği dağılımlar geliştirilmiştir. Bu durumda bu dağılımlardan birinin kullanılması daha uygundur [25].

3.5.2. Sıfır Yığılmalı Negatif Binom Regresyon Modeli

Sıfır yığılmalı negatif binom modeli (*ZINB*), Greene [37] tarafından, uygulamada sayım verisinde sıfır yayılımı dışında aşırı yayılım olması durumunda sıfır yığılmalı Poisson regresyona alternatif olarak tanımlanmıştır.

Sıfır yığılmalı negatif binom regresyon, modelde iki ayrı veri üretim süreci olduğunu varsaymaktadır. Bernoulli denemesinin sonucunda hangi sürecin kullanılacağına karar verilir.

ω_i: *i*. gözlem için yapısal sıfır oranı, $(1-\omega_i)$: ikinci sürecin olasılığı, α: negatif binom yayılım parametresi, μ_i: ortalama olmak üzere sıfır yığılmalı negatif binom regresyon için olasılık fonksiyonu;

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

$$P(Y_i = y_i) = (1-\omega_i)\frac{\Gamma(y_i+\alpha^{-1})}{\Gamma(y_i+1)\Gamma(\alpha^{-1})}\frac{(\alpha\mu_i)^{y_i}}{(1+\alpha\mu)^{y_i+\alpha^{-1}}}, \qquad y_i = 1,2,... \tag{3.33}$$

şeklindedir.

Y_i raslantı değişkeni için beklenen değer ve varyans sırasıyla;

$$E(Y_i) = (1-\omega_i)\mu_i \tag{3.34}$$

$$V(Y_i) = (1-\omega_i)\mu_i(1+\mu_i(\omega_i+\alpha)) \tag{3.35}$$

biçimindedir. Yayılım parametresi $\alpha \to 0$'a yakınsıyorsa model sıfır yığılmalı Poisson regresyona (ZIP) indirgenir.

3.5.3. Sıfır Yığılmalı Genelleştirilmiş Poisson Regresyon Modeli

Son yıllarda, sayım verilerinde çok fazla sıfırın varlığına izin vermek için sıfır yığılmalı Poisson regresyon modeli kullanma konusunda hayli görüş olmuştur. Famoye ve Singh [30] sıfır yığılmalı Poisson regresyon modellerinin yetersiz olduğunu ve sıfır yığılmalı negatif binom regresyon modelinin gözlemlenen veri setine uyumsuz olduğu durumlarına dikkat çekmiştir. Çok fazla sıfır içeren aşırı yayılımlı sayım verisini modellemek için sıfır yığılmalı genelleştirilmiş Poisson regresyon (*ZIGP*) modelini geliştirmişlerdir. Örneklendirmek için şiddet suçunu içeren verilere *(Domestic Violence Data*) uygulamışlardır [6].

Sıfır yığılmalı negatif binom regresyon, parametre tahminleri için yetersiz olduğunda alternatif olarak sıfır yığılmalı genelleştirilmiş Poisson regresyon modeli kullanılmaktadır [30].

$Y_i,\ i = 1, 2, \ldots, n$' den oluşan genelleştirilmiş Poisson dağılan sayım verileri bağımlı değişkeni olsun. Y_i için olasılık fonksiyonu;

$$P(Y_i = 0) = \omega_i + (1-\omega_i)e^{\frac{-\mu_i}{\varphi}} \qquad y_i = 0$$

$$P(Y_i > 0) = (1-\omega_i)\frac{\mu_i\left(\mu_i + (\varphi - 1)y_i\right)^{y_i - 1}}{y_i!}\varphi^{-y_i}e^{\left(\frac{-1}{\varphi}(\mu_i + (\varphi-1))\right)y_i} \qquad y_i = 1, 2, \ldots$$

$$0 < \omega_i < 1 \tag{3.36}$$

biçimindedir.

Eşitlik (3.36)'da verilen Y_i raslantı değişkeni için beklenen değer ve varyans sırasıyla:

$$E(Y_i) = (1-\omega_i)\mu_i \tag{3.37}$$

$$V(Y_i) = E(Y_i)\left[\varphi^2 + \omega_i\mu_i\right] \qquad i = 1, 2, \ldots, n \tag{3.38}$$

şeklindedir.

Sıfır yığılmalı genelleştirilmiş Poisson regresyonda,

- Yapısal sıfır oranı sıfıra eşit $(\omega_i = 0)$ ve yayılım parametresi bire eşit ($\varphi^2 = 1$) olduğunda model Poisson regresyona indirgenir.

- Yapısal sıfır oranı sıfıra eşit $(\omega_i = 0)$ olduğunda model genelleştirilmiş Poisson regresyona indirgenir.
- Yayılım parametresi bire eşit (φ^2 = 1) olduğunda ise model sıfır yığılmalı Poisson regresyona indirgenir [38].

4. MODEL SEÇİMİ

Bir araştırmada modelin anlamlılığı için, bağımlı değişkenin dağılımına uygun bir regresyon (örneğin sayım verileri için Poisson modeli) modeli uygulanıp sadece açıklayıcı değişkenlerin p değerleri kontrol edilebilir. Bütün p değerlerinin 0.05'in altında olması, bütün açıklayıcı değişkenlerin modele uygun olduğunu ifade eder. Ancak bütün açıklayıcı değişkenlerin anlamlı olması, uygulanan regresyon modelinin veri için uygun olacağı anlamına gelmez [27] .

4.1. Uyum İyiliği Ölçümü

Bir modelden elde edilen sonuçlar istatistiksel olarak modelin geçerliliğine bağlıdır. Ele alınan veri kümesi için çeşitli regresyon modelleri mevcut olduğundan verileri en iyi açıklayan modele karar vermek gerekmektedir. Bu nedenle veri kümesi ile model arasındaki uygunluğun testi, istatistiksel modellemenin önemli bir parçasını oluşturmaktadır. Bu amaçla, kullanılan regresyon modellerinde normallik varsayımının sağlanmaması nedeniyle Pearson Ki-Kare, sapma istatistiği, yapay R^2 gibi uyum iyiliği ölçütlerinden yararlanılmaktadır [39].

4.1.1. Genelleştirilmiş Pearson Ki- Kare Testi

Pearson Ki-Kare Testi, belirlenen model ile veri kümesi arasındaki uyumu araştırmak için yaygın olarak kullanılan uyum iyiliği istatistiklerinden biridir. Genelleştirilmiş Pearson Ki-Kare istatistiği,

$$\chi^2 = \sum (y-\hat{\mu})^2 \Big/ V(\hat{\mu}) \qquad (4.1)$$

eşitliğinden elde edilir.

$V(\hat{\mu})$, ilgili dağılımdan tahmin edilen varyans fonksiyonudur [15] .

Veri setinin yeterince büyük olması halinde, n örneklem büyüklüğü, k sabitle birlikte açıklayıcı değişken sayısını göstermek üzere n-k serbestlik dereceli dağılımı göstermektedir. Hesaplanan Pearson Ki-Kare istatistiği n-k serbestlik dereceli χ^2_{n-k} değerinden büyükse hipotez reddedilir ve belirlenen modelin veri kümesine uyum sağlamadığı söylenir. Bu istatistiğin küçük örneklemler için kullanılması tercih edilmemektedir [16].

4.1.2. Sapma İstatistiği

Sapma istatistiği, genelleştirilmiş doğrusal modeller için uyum iyiliğinin ölçülmesinde kullanılan tekniklerden biridir. Bu istatistik değeri aynı zamanda "G^2 *istatistiği*" olarak da adlandırılmaktadır.

G^2 istatistiği,

$$G^2 = 2\sum_{i=1}^{n} y_i \ln\left(\frac{y_i}{\mu_i}\right) \tag{4.2}$$

şeklindedir. G^2 değeri;

- 0'a yakınsıyor ise model uyumunun arttığı,
- tam 0'a eşit ise model uyumunun mükemmel olduğu söylenebilir [20] .

4.1.3. Yapay R^2 ölçümü

Doğrusal olmayan modeller için evrensel bir R^2 tanımı yoktur. Birçok oran önerilebilir. Bu belirsizlikten dolayı yapay (*pseudo*) bir niteliyici kullanılır.

Doğrusal regresyon modelinde, R^2'yi elde etmek için başlangıç noktası genel kareler toplamının ayrıştırılmasıdır. Genel kareler toplamı,

$$\sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n}(y_i-\hat{\mu}_i)^2 + \sum_{i=1}^{n}(\hat{\mu}_i-\bar{y})^2 + 2\sum_{i=1}^{n}(y_i-\hat{\mu}_i)(\hat{\mu}_i-\bar{y}) \tag{4.3}$$

ile ifade edilir. İlk ifade genel kareler toplamı *(GKT)*, ikinci ifade artık kareler toplamı *(AKT)* ve üçüncü ifade açıklanmış kareler toplamı (*RKT*) olarak açıklanır. Bu eşitliğin 3 e ayrımına dikkat edelim. Son ifade ise model sabit terim içerdiğinde, doğrusal regresyon modelinin en küçük kareler kestirimine göre sıfıra eşit olacaktır. Ancak Poisson'u da içeren ve doğrusal olmayan en küçük kareler ve üstel koşullu ortalamaya sahip tüm kestiriciler ve modeller için sıfıra eşit olmayacaktır.

Bu durumda R^2'nin, $R^2 = 1 - {AKT}/{GKT}$ veya $R^2 = {RKT}/{GKT}$ yönteminden farklı bir yolla hesaplanması gerektiği Cameron ve Trivedi tarafından ortaya çıkarılmıştır. [18]

Normallik varsayımı gerektirmeyen Poisson regresyon modelinde R^2 ölçüsü olabilirlik oran yaklaşımına dayanmaktadır. Doğrusal regresyon modeline ilişkin en küçük kareler tahmini, artık kareler toplamının en çok olabilirlik tahmini ve sapma değeri ile benzer özellikler göstermesi nedeniyle önerilen R^2 ölçüsü,

$$R_p^2 = 1 - \frac{\log L(y) - \log L(\hat{\mu})}{\log L(y) - \log L(\bar{y})} \tag{4.4}$$

şeklinde tanımlanmaktadır. Burada $\log L(y)$ doygun modelin log olabilirliğini; $\log L(\bar{y})$ ilgilenilen modelin log-olabilirliğini ve $\log L(\hat{\mu})$ sadece sabit terimin bulunduğu minimal modelin log-olabilirliğini göstermektedir. Gözlenen değerler, $y_i \geq 0$ $\hat{\mu}_i = \exp(x_i\hat{\beta})$ ya da $\hat{\mu}_i = c_i \exp(x_i\hat{\beta})$ tahmin edilen değerler ve $\bar{y}_i = c_i \exp(\hat{\beta}_0)$ ya da $\bar{y}_i = c_i \exp(\hat{\beta}_0)$ ortalama değerler olmak üzere log- olabilirlik fonksiyonları,

$$\log L(y) = \sum_{i=1}^{n} \left(y_i \log(y_i) - y_i - \log(y_i!) \right) \tag{4.5}$$

$$\log L(\hat{\mu}) = \sum_{i=1}^{n} \left(y_i \log(\hat{\mu}_i) - \hat{\mu}_i - \log(y_i!) \right) \tag{4.6}$$

$$\log L(\bar{y}) = \sum_{i=1}^{n} \left(y_i \log(\bar{y}_i) - \bar{y}_i - \log(y_i!) \right) \tag{4.7}$$

biçiminde elde edilmektedir. Bu log-olabilirlik fonksiyonları düzenlenirse *yapay* R^2 ölçüsüne ulaşılmaktadır [20] .

4.2. Model Seçimi

Model seçiminde amaç, çok sayıda parametre arasında uzlaşan ve böylece gözlemlere uyan yakın bir model bulmaktır [14].

Model seçim kriterleri ilk olarak 60'ların başı ve 70'lerin sonlarında Akaike'nin FPE (Akaike, 1969) ve Mallow'un Cp testleri ile ortaya çıkmıştır. Kullback.-Leibler (Kullback ve Leibler, 1951) uyumsuzluğuna dayanan ve daha çok zaman serilerinde kullanılan Akaike Bilgi Kriterinin (*AIC*) ortaya çıkmasıyla başlayan bu dönüm noktasında 70'lerde bilgi teorileri yaklaşımı görülmeye başlamıştır. 70'lerin sonlarına gelindiğinde bilgi teorileri alanında çalışmalar artarak gelişmiş ve yeni kriterler doğmuştur [40].

4.2.1. Akaike Bilgi Kriteri (*AIC*)

Akaike bilgi kriteri, modelleri değerlendirmek için kullanılabilecek ölçütlerden biridir. Bu ölçüt, log olabilirlikleri kullanır. Bilindiği üzere, modele değişken eklemek modelin uyum iyiliğini artırabilir. Böylece, Akaike bilgi kriteri modele değişkenlerin dahil edilmesine karşın uyum iyiliğini dengelemeye çalışır. Akaike bilgi kriteri *(AIC)* aşağıdaki şekilde hesaplanır ;

$$AIC = -2\ln L + 2p \qquad (4.8)$$

Eş. (5.8)'de p modeldeki bilinmeyen parametrelerin sayısını gösterir. lnL ise log olabilirlik fonksiyonudur. Akaike bilgi kriterinin en küçük değer aldığı model, en uygun modeldir [41].

4.2.2. Bayes Bilgi Kriteri (*BIC*)

Bayes bilgi kriteri *(BIC)*, Akaike'ye benzer olarak örneklem genişliği *(n)* ve parametre sayısı*(p)* ile ilişkili bir ceza *(penalty)* terimi kullanır. Bu ölçüt aynı zamanda *Schwarz (bayesian) bilgi kriteri* olarak da adlandırılır. Bayes bilgi kriteri;

$$BIC = -2\ln L + p(\ln n) \qquad (4.9)$$

eşitliğinden hesaplanır.

AIC'den farklı olarak, BIC eşitliğin sağ kısmında yer alan örneklem büyüklüğüne bağlı bir kriterdir [41].

AIC daha fazla açıklayıcı değişkenle daha iyi model uyumunu araştırırken BIC daha az açıklayıcı değişken ile daha iyi model uyumunu sağlamaktadır. Ayrıca BIC, örneklem büyüklüğünü içerdiğinden farklı örneklem büyüklüklerine sahip modellerin karşılaştırılmasında daha uygundur [16].

Bayes bilgi kriterinin en küçük değer aldığı model, en uygun modeldir [41].

4.2.3. Mallow'un Cp Kriteri

İlk olarak Mallow tarafından 1973'te tanımlanan C_p kriteri yan ve varyansı birlikte içermektedir. Ölçüt,

$$C_p = \frac{(RSS)_p}{\hat{\sigma}^2} + 2p - n \qquad (4.10)$$

şeklinde verilir. Burada $\boldsymbol{\hat{\sigma}^2}$, σ^2'nin en iyi kestirimi; $\boldsymbol{p}$, parametre sayısı ve $(\boldsymbol{RSS})_{\boldsymbol{p}}$, p parametreli modelin artık kareler toplamıdır. Örneklem büyüklüğü C_p kriteri üzerinde etkilidir. Ölçütün kullanımında; **k**, değişken sayısı olmak üzere tüm 2^k - 1 olası alt küme üzerinden C_p hesaplanmakta ve en küçük C_p 'li alt küme ya da $C_p \leq p$ koşulunu sağlayan denklemler yeterli denklemler olarak kullanılabilir [42].

5. UYGULAMA

Bu çalışmada Türkiye İstatistik Kurumu tarafından yürütülen Hanehalkı İşgücü Araştırması verileri kullanılmıştır. Fertlerin haftalık ortalama çalışma saati verilerine Poisson regresyon, genelleştirilmiş Poisson regresyon, negatif binom regresyon analizleri uygulanıp en uygun model seçilmiştir.

Türkiye İstatistik Kurumu tarafından 1988 yılından itibaren düzenli olarak uygulanmakta olan Hanehalkı İşgücü Araştırması (HİA) istihdam edilenlerin; iktisadi faaliyet, meslek (ya da tuttuğu iş), işteki durum ve çalışma süresi, işsizlerin ise; iş arama süresi ve aradıkları meslek (ya da iş) ve benzer özellikleri hakkında bilgi derlemek amacıyla uygulanmakta olup, ülkedeki işgücü piyasasının özellikleri hakkında bilgi veren (arz yönüyle) temel veri kaynağıdır [43].

Referans döneminde; ücretli, maaşlı, yevmiyeli, kendi hesabına, işveren veya ücretsiz aile işçisi olarak en az bir saat bir iktisadi faaliyette bulunan veya bir işle bağlantısı olan çalışma çağındaki kişiler istihdam edilen nüfusu oluşturmaktadır [43].

Hanehalkı İşgücü Araştırmasında, çalışılan saatin ölçümü de yapılmaktadır. ILO, 18. Çalışma İstatistikçileri Konferansında (2008) kabul edilen "çalışma saatlerinin ölçümü" konulu düzenlemeye göre; genellikle çalışılan süre, uzun bir dönem süresince (kesintisiz olarak çalışılmış olan) gerçekleşen fiili çalışma sürelerinin en yaygın değeridir. Çalışma saatlerinin ölçümünde, kişinin normalde (genel olarak) bir haftada kaç saat çalıştığı bilgisi alınmaktadır [44].

Çalışmada, Hanehalkı İşgücü Araştırması sonucunda elde edilen *Mardin, Batman, Şırnak, Siirt* illerini kapsayan SİİRT Bölgesi verileri kullanılmıştır. Veriler, 2016 yılı Ocak ayından Aralık ayına dek 12 ayı kapsamaktadır. 2016 yılında Hanehalkı İşgücü Araştırması kapsamında SİİRT bölgesinden 14 523 kişi ile görüşülmüştür. Uygulama için, referans haftasında istihdamda olan 2488 kişi filtrelenmiş ve analizler yapılmıştır.

Uygulamada SPSS, SAS ve R programları kullanılmıştır.

5.1. TÜİK Verileri

Bağımlı değişken olan **Çalışma saati** değişkeni için SİİRT Bölgesinde, istihdamda olan 2488 kişinin normalde bir haftada kaç saat çalıştığı bilgisi alınmıştır. Veriler tam sayı olarak toplanmaktadır. Örneğin 10 saat, 25 saat, 40 saat gibi.

Gerçek dünyada orijinal değişkenlerin doğada sürekli olabileceği ancak gözlemle kesikli olabileceği çıkarımı yapılabilir. Bu nedenle, sürekli dağılımın bir ya da birden fazla önemli özelliğini koruyan sürekli temelli modelden uygun bir kesikli dağılımlı model türetmek uygundur [45]. Bu nedenle bağımlı değişkenimiz kesikli veri olarak alınmıştır.

Bağımsız değişkenler;

X_1 : Sektör,

X_2 : Yaş,

X_3 : Cinsiyet,

X_4 : Eğitim,

X_5 :Meslek,

X_6 : Kayıtlılık,

X_7 : Süreklilik

olarak belirlenmiştir.

Bağımsız değişkenler ile ilgili açıklamalar Çizelge 5.1'de verilmiştir.

Çizelge 5.1. Uygulamada Kullanılan Bağımsız Değişken Kategorileri

Bağımsız Değişken	Açıklama	Kategori
Sektör	Çalışılan sektör bilgisi	1.Tarım 2.Sanayi 3.İnşaat 4.Hizmet
Cinsiyet		1.Erkek 2.Kadın
Eğitim	Mezun olunan eğitim seviyesi	1.Okur yazar olmayanlar 2.Lise altı 3.Lise 4.Mesleki veya teknik lise 5.Yükseköğretim
Meslek	Meslek bilgisi	1.Yöneticiler 2. Profesyonel meslek mensupları 3. Teknisyenler, teknikerler ve yardımcı prof meslek mensupları 4. Büro hizmetleri 5. Hizmet ve satış elamanları 6. Nitelikli tarım, ormancılık ve su ürünlerinde çalışanlar 7. Sanatkarlar ve ilgili işlerde çalışanlar 8. Tesis ve makina operatörleri ve montajcılar 9. Nitelik gerektirmeyen işlerde çalışanlar
Kayıtlılık	Sosyal Güvenlik Kurumu'na kayıtlı olma durumu	1. Evet 2. Hayır
Süreklilik	Ferdin çalıştığı işteki süreklilik durumu	1. Sürekli iş 2. Geçici veya sınırlı süreli iş
Yaş	Yaş bilgisi	

Araştırmaya katılan istihdamdaki 2488 kişiye ait bağımsız değişkenlere ilişkin sıklık ve yüzde değerleri aşağıda tablolarda verilmiştir.

Çizelge 5.2. Sektör Değişkenine İlişkin Özet Bilgiler

	Sektör			
	1.Tarım	2.Sanayi	3.İnşaat	4.Hizmet
Sayı	332	279	256	1621
Yüzde	13,3	11,2	10,3	65,2

Çizelge 5.2'de görüldüğü üzere araştırmaya katılanların çalıştığı sektör incelendiğinde en fazla yüzdeye sahip olduğu kategori %65,2 ile hizmet sektörüdür. Bunu %13,3 ile tarım sektöründe çalışanlar takip etmektedir.

Çizelge 5.3. Cinsiyet Değişkenine İlişkin Özet Bilgiler

	Cinsiyet	
	1.Erkek	2.Kadın
Sayı	2034	454
Yüzde	81,8	18,2

Çizelge 5.3'te araştırmaya katılan istihdamdaki fertlerin %81,8'inin erkek, %18,2'sinin kadın olduğu görülmektedir.

Çizelge 5.4. Eğitim Değişkenine İlişkin Özet Bilgiler

	Eğitim				
	1. Okur yazar olmayanlar	2.Lise altı	3.Lise	4.Mesleki veya teknik lise	5.Yükseköğretim
Sayı	201	1298	350	124	515
Yüzde	8,1	52,2	14,1	5,0	20,7

Çizelge 5.4'te araştırmaya katılan istihdamdaki fertlerin yarısından fazlasının lise altı eğitimli, beşte birinin yüksek öğretim mezunu olduğu görülmektedir.

Çizelge 5.5. Meslek Değişkenine İlişkin Özet Bilgiler

	Meslek								
	1.Yöneticiler	2.Profesyonel meslek mensupları	3.Teknisyenler, teknikerler ve yardımcı prof meslek mensupları	4.Büro hizmetleri	5. Hizmet ve satış elamanları	6. Nitelikli tarım, ormancılık ve su ürünlerinde çalışanlar	7. Sanatkarlar ve ilgili işlerde çalışanlar	8. Tesis ve makina operatörleri ve montajcılar	9. Nitelik gerektirmeyen işlerde çalışanlar
Sayı	151	303	117	99	625	294	287	252	360
Yüzde	6,1	12,7	4,7	4,0	25,1	11,8	11,5	10,1	14,5

Çizelge 5.5'te görüldüğü üzere araştırmaya katılanların %25,1'i hizmet ve satış elemanı olarak çalışmaktadır. Bunu %14,5 ile nitelik gerektirmeyen işlerde çalışanlar takip etmektedir.

Çizelge 5.6. Kayıtlılık Değişkenine İlişkin Özet Bilgiler

	Kayıtlılık	
	1.Evet	2.Hayır
Sayı	1474	1014
Yüzde	59,2	40,8

Çizelge 5.6'da kayıtlılık durumu incelendiğinde, çalışanların %59,2'si Sosyal Güvenlik Kurumuna Kayıtlı olduğunu beyan ederken, %40,8'i kayıt dışı çalıştığını beyan etmiştir.

Çizelge 5.7. Süreklilik Değişkenine İlişkin Özet Bilgiler

	Süreklilik	
	1. Sürekli iş	2. Geçici veya sınırlı süreli iş
Sayı	1583	365
Yüzde	63,6	14,7

Çizelge 5.7'de fertlerin çalıştığı işte süreklilik durumu verilmiştir. Bu soru ücretli, maaşlı veya yevmiyeli olarak çalışan 1948 kişiye sorulmuştur. Fertlerin %63,6'sı sürekli bir işte çalışırken, %14,7'si geçici veya sınırlı bir süreli işte çalışmaktadır.

Yaş gruplarına göre fertlerin dağılımı incelendiğinde, 15 ile 80 yaşları arasında değişmekle birlikte yaş ortalaması 36,4'tür.

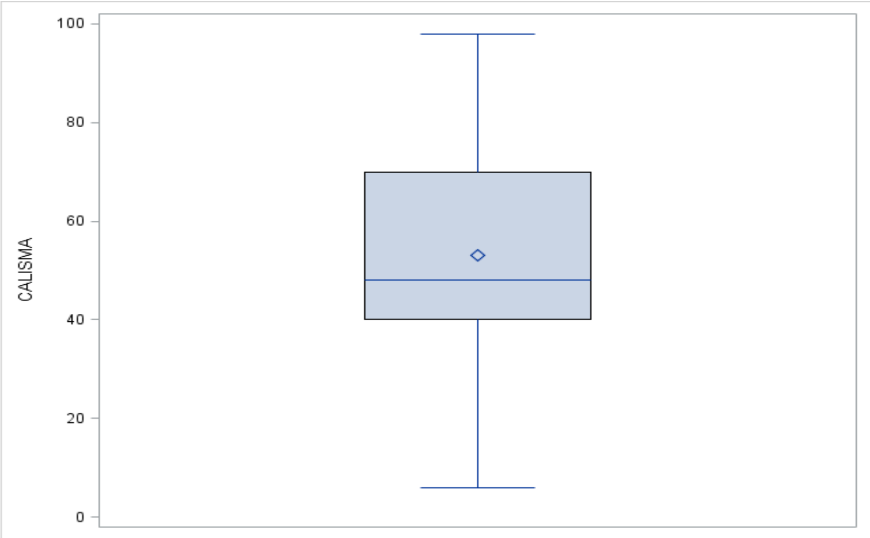
Fertlerin bir haftada kaç saat çalıştığı bilgisi olan *CALISMA* bağımlı değişkenine ilişkin tanımlayıcı istatistikler Çizelge 5.8'de verilmiştir.

Çizelge 5.8. Bağımlı Değişkene İlişkin Tanımlayıcı İstatistikler

Çalışma						
Minimum	**Maksimum**	**Ortalama**	**Varyans**	**Q1**	**Q2 (Medyan)**	**Q3**
6	98	52,9	341,5	40	48	70

Çizelge 5.8'de görüldüğü üzere araştırmaya katılan istihdamdaki fertlerin haftalık çalışma saatleri 6 saat ile 98 saat arasında değişmektedir. Varyansın (341,5) ortalamadan (52,9) büyük olması sebebiyle yayılım durumu söz konusudur.

Dağılımın çarpıklığı Q3-Q2 > Q2-Q1 formülünü sağladığından ***sağa çarpıktır.*** Şekil 1'deki Box-Plot grafiğinden de dağılımın sağa çarpık olduğu görülmektedir.



Şekil 1. Çalışma saati dağılımı

Sayma verisi bağımlı değişkenin sağa çarpık dağıldığı durumlarda kullanılan Poisson regresyon, genelleştirilmiş Poisson regresyon, negatif binom regresyon analizleri yapılmıştır. Bu analizler sonucunda anlamlı ve anlamsız bulunan açıklayıcı değişkenler belirlenmiştir.

İncelenen tüm modeller için uyum iyiliği testi yapılmış bu amaçla AIC, BIC, Log-olabilirlik ve Pearson Ki-kare seçim kriterleri uygulanmıştır. AIC, BIC ve Pearson Ki-kare değerlerinin en küçük olduğu, log-olabilirlik değerinin ise en büyük olduğu model; veri setine en uygun regresyon modeli olarak seçilmiştir.

Uygulanan regresyon modelinde, anlamlı olan bütün açıklayıcı değişkenlerin var olması veri için uygun olacağı anlamına gelmez. Son olarak, Tüm Olası Alt Kümeler tekniği uygulanmıştır. Bu çalışmada AIC, BIC, Mallow'un Cp kriterlerinin en düşük değer aldığı model veri setine en uygun model seçilerek değişkenleri belirlenmiştir Bu açıklayıcı değişkenler yorumlanmıştır.

5.2. Regresyon Analizi Sonuçları

Poisson Regresyon Analizi

Çalışma saatine etki eden sektör, cinsiyet, eğitim, meslek, kayıtlılık, süreklilik ve yaş açıklayıcı değişkenleri ile oluşturulan Poisson regresyon modeline ilişkin sonuçlar Çizelge 5.9'da verilmiştir.

Çizelge 5.9. Çalışma Saati Verisi için Poisson Regresyon Modeli

Açıklayıcı Değişkenler	**Tahmin (β)**	**St.Hata**	**Üstel Tahmin Exp(β)**	**p - değeri**
(Sabit)	3,8602	0,0191	47,4733	0,0000
Tarım	0,1651	0,0221	1,1795	0,0000
Sanayi	0,0569	0,0102	1,0585	0,0000
İnşaat	0,0544	0,0118	1,0559	0,0000
Erkek	-0,0206	0,0088	0,9796	0,0198
Okur yazar olmayanlar	0,2410	0,0168	1,2725	0,0000
Lise altı	0,1358	0,0115	1,1454	0,0000
Lise	0,1104	0,0123	1,1167	0,0000
Mesleki veya teknik lise	0,0609	0,0166	1,0627	0,0002
Yöneticiler	0,0163	0,0189	1,0165	0,3874
Profesyonel meslek mensupları	-0,1757	0,0160	0,8388	0,0000
Teknisyenler, teknikerler ve yardımcı prof meslek mensupları	0,0360	0,0173	1,0367	0,0375
Büro hizmetleri	-0,0531	0,0189	0,9482	0,0048
Hizmet ve satış elamanları	0,2117	0,0106	1,2357	0,0000
Nitelikli tarım, ormancılık ve su ürünlerinde çalışanlar	0,1313	0,0264	1,1403	0,0000
Sanatkarlar ve ilgili işlerde çalışanlar	0,0635	0,0121	1,0656	0,0000
Tesis ve makina operatörleri ve montajcılar	0,1749	0,0121	1,1911	0,0000
Kayıtlı	-0,0707	0,0078	0,9317	0,0000
Sürekli	0,1002	0,0091	1,1054	0,0000
Yaş	-0,0030	0,0003	0,9970	0,0000

Poisson regresyon modelindeki açıklayıcı değişkenlerin modele katkısının anlamlı olup olmadığını araştırmak için hipotezler kurulmuştur.

- H_0: $\beta_{1i} = 0$ (Sektör değişkeninin modele katkısı anlamlı değildir. $i = 1, 2, 3$)
  H_1: $\beta_{1i} \neq 0$ (Sektör değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_{2j} = 0$ (Cinsiyet değişkeninin modele katkısı anlamlı değildir. $j = 1$)
  H_1: $\beta_{2j} \neq 0$ (Cinsiyet değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_{3k} = 0$ (Eğitim değişkeninin modele katkısı anlamlı değildir. $k=1, 2, 3, 4$)
  H_1: $\beta_{3k} \neq 0$ (Eğitim değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_{4l} = 0$ (Meslek değişkeninin modele katkısı anlamlı değildir. $j = 1, 2, \ldots, 8$)
  H_1: $\beta_{4l} \neq 0$ (Meslek değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_{5m} = 0$ (Kayıtlılık değişkeninin modele katkısı anlamlı değildir. $m = 1$)
  H_1: $\beta_{5m} \neq 0$ (Kayıtlılık değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_{6n} = 0$ (Süreklilik değişkeninin modele katkısı anlamlı değildir. $n = 1$)
  H_1: $\beta_{6n} \neq 0$ (Süreklilik değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_7 = 0$ (Yaş değişkeninin modele katkısı anlamlı değildir.)
  H_1: $\beta_7 \neq 0$ (Yaş değişkeninin modele katkısı anlamlıdır.)

Çizelge 5.9'da belirtilen p-değerleri incelendiğinde; meslek gruplarından 'yöneticiler'in *p*-değeri 0.05'ten büyük olduğu için %5 yanılma düzeyinde istatistiksel açıdan modele anlamlı bir katkısı olmadığı, diğer açıklayıcı değişkenlerin %5 yanılma düzeyinde anlamlı olduğu bulunmuştur.

Yapılan analiz sonucu Poisson regresyon modeli için yayılım parametresi $\varphi = 3,83$ olarak hesaplanmıştır. Poisson modelinde yayılım parametresinin (φ) 1'den büyük olması gözlemlenen verinin aşırı yayılım gösterdiğini belirtmektedir.

Ayrıca, aşırı yayılımın varlığını test etmek için Olabilirlik Oran Test İstatistiği (LR)=3926,03 olarak hesaplanmıştır. Olabilirlik oran test istatistiğinin, ki-kare tablo değerinden büyük olması gözlemlenen veride aşırı yayılımın anlamlı olduğunu göstermektedir.

Aşırı yayılımın da dikkate alındığı Negatif Binom Regresyon modeli ve Genelleştirilmiş Poisson Regresyon modeli uygulanmıştır.

Negatif Binom Regresyon Analizi

Bir hafta içindeki çalışma saatine etki eden sektör, cinsiyet, eğitim, meslek, kayıtlılık, süreklilik ve yaş açıklayıcı değişkenleri ile oluşturulan negatif binom regresyon modeline ilişkin sonuçlar Çizelge 5.10'da verilmiştir.

Çizelge 5.10. Çalışma Saati Verisi için Negatif Binom Regresyon Modeli

Açıklayıcı Değişkenler	**Tahmin (β)**	**St.Hata**	**Üstel Tahmin Exp(β)**	**p - değeri**
(Sabit)	3,860	0,037	46,618	0,000
Tarım	0,167	0,046	1,181	0,000
Sanayi	0,061	0,020	1,063	0,003
İnşaat	0,064	0,023	1,066	0,006
Erkek	-0,003	0,017	0,997	0,885
Okur yazar olmayanlar	0,252	0,034	1,287	0,000
Lise altı	0,140	0,022	1,151	0,000
Lise	0,116	0,023	1,123	0,000
Mesleki veya teknik lise	0,070	0,031	1,073	0,024
Yöneticiler	0,018	0,035	1,018	0,621
Profesyonel meslek mensupları	-0,168	0,029	0,845	0,000
Teknisyenler, teknikerler ve yardımcı prof meslek mensupları	0,038	0,033	1,039	0,245
Büro hizmetleri	-0,048	0,035	0,953	0,168
Hizmet ve satış elamanları	0,215	0,021	1,240	0,000
Nitelikli tarım, ormancılık ve su ürünlerinde çalışanlar	0,128	0,055	1,137	0,020
Sanatkarlar ve ilgili işlerde çalışanlar	0,060	0,024	1,062	0,012
Tesis ve makina operatörleri ve montajcılar	0,177	0,024	1,194	0,000
Kayıtlı	-0,003	0,001	0,997	0,000
Sürekli	-0,071	0,016	0,931	0,000
Yaş	0,102	0,018	1,108	0,000
Yayılım Parametresi	0,053	0,002	0,997	0,000

- H_0: $\beta_{1i} = 0$ (Sektör değişkeninin modele katkısı anlamlı değildir. $i = 1, 2, 3$)
  H_1: $\beta_{1i} \neq 0$ (Sektör değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_{2j} = 0$ (Cinsiyet değişkeninin modele katkısı anlamlı değildir. $j = 1$)
  H_1: $\beta_{2j} \neq 0$ (Cinsiyet değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_{3k} = 0$ (Eğitim değişkeninin modele katkısı anlamlı değildir. $k=1, 2, 3, 4$)
  H_1: $\beta_{3k} \neq 0$ (Eğitim değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_{4l} = 0$ (Meslek değişkeninin modele katkısı anlamlı değildir. $j = 1, 2, \ldots, 8$)
  H_1: $\beta_{4l} \neq 0$ (Meslek değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_{5m} = 0$ (Kayıtlılık değişkeninin modele katkısı anlamlı değildir. $m = 1$)
  H_1: $\beta_{5m} \neq 0$ (Kayıtlılık değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_{6n} = 0$ (Süreklilik değişkeninin modele katkısı anlamlı değildir. $n = 1$)
  H_1: $\beta_{6n} \neq 0$ (Süreklilik değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_7 = 0$ (Yaş değişkeninin modele katkısı anlamlı değildir.)
  H_1: $\beta_7 \neq 0$ (Yaş değişkeninin modele katkısı anlamlıdır.)

Çizelge 5.10'da belirtilen *p*-değerleri incelendiğinde; cinsiyet değişkeni, meslek değişkeni gruplarından ‘yöneticiler, teknisyenler, büro hizmetleri’ *p*-değerleri 0.05’ten büyük olduğu için %5 yanılma düzeyinde istatistiksel açıdan modele anlamlı bir katkısı olmadığı, diğer açıklayıcı değişkenlerin %5 yanılma düzeyinde anlamlı olduğu bulunmuştur.

Negatif binom modeli yayılım parametresi olan $\boldsymbol{\alpha}$ değeri sıfırdan farklıdır ($\alpha = 0.05$). Bu nedenle veriler negatif binom modeline uygundur.

Genelleştirilmiş Poisson Regresyon Analizi

Çalışma saatine etki eden sektör, cinsiyet, eğitim, meslek, kayıtlılık, süreklilik ve yaş açıklayıcı değişkenleri ile oluşturulan ve genelleştirilmiş Poisson regresyon modeline ilişkin sonuçlar Çizelge 5.11'de verilmiştir.

Çizelge 5.11. Çalışma Saati Verisi için Genelleştirilmiş Poisson Regresyon Modeli

Açıklayıcı Değişkenler	**Tahmin (β)**	**St.Hata**	**Üstel Tahmin Exp(β)**	**p - değeri**
(Sabit)	3,6800	0,0200	39,6464	0,0000
Sektör	-0,0300	0,0700	0,9704	0,0000
Cinsiyet	0,0200	0,0166	1,0202	0,2200
Eğitim	-0,0900	0,0200	0,9139	0,0000
Meslek	0,0000	0,0068	1,0000	0,4270
Kayıtlılık	0,0900	0,0100	1,0942	0,0000
Süreklilik	-0,1200	0,0175	0,8869	0,0000
Yaş	0,0000	0,0000	1,0000	0,0022
Yayılım Parametresi	0,7500	0,0100		

- H_0: $\beta_1 = 0$ (Sektör değişkeninin modele katkısı anlamlı değildir.)
  H_1: $\beta_1 \neq 0$ (Sektör değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_2 = 0$ (Cinsiyet değişkeninin modele katkısı anlamlı değildir.)
  H_1: $\beta_2 \neq 0$ (Cinsiyet değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_3 = 0$ (Eğitim değişkeninin modele katkısı anlamlı değildir.)
  H_1: $\beta_3 \neq 0$ (Eğitim değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_4 = 0$ (Meslek değişkeninin modele katkısı anlamlı değildir.)
  H_1: $\beta_4 \neq 0$ (Meslek değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_5 = 0$ (Kayıtlılık değişkeninin modele katkısı anlamlı değildir.)
  H_1: $\beta_5 \neq 0$ (Kayıtlılık değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_6 = 0$ (Süreklilik değişkeninin modele katkısı anlamlı değildir.)
  H_1: $\beta_6 \neq 0$ (Süreklilik değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_7 = 0$ (Yaş değişkeninin modele katkısı anlamlı değildir.)
  H_1: $\beta_7 \neq 0$ (Yaş değişkeninin modele katkısı anlamlıdır.)

Çizelge 5.11'de belirtilen *p*-değerleri incelendiğinde; cinsiyet değişkeni ve meslek değişkeni *p*-değerleri 0.05'ten büyük olduğu için %5 yanılma düzeyinde istatistiksel açıdan modele anlamlı bir katkısı olmadığı, diğer açıklayıcı değişkenlerin %5 yanılma düzeyinde anlamlı olduğu bulunmuştur. Genelleştirilmiş Poisson modeli yayılım parametresi olan $\boldsymbol{\varphi}$ değeri sıfırdan farklıdır ($\boldsymbol{\varphi} = 0.75$). Veriler genelleştirilmiş Poisson modeline uygundur.

İncelenen tüm modeller için uyum iyiliği testi yapılmış bu amaçla hesaplanan Akaike, Bayes bilgi kriterleri ve log-olabilirlik ve Pearson Ki-kare değerleri Çizelge 5.12'de verilmiştir.

Çizelge 5.12. AIC, BIC, Log-olabilirlik ve Pearson Ki-kare Değerleri

	AIC	**BIC**	**Log-olabilirlik**	**Pearson Ki-Kare**
Poisson R.	18656	18767	-9308	7280,6
Negatif Binom R.	15793	15910	-7875	1876,4
Genelleştirilmiş Poisson R.	16111	16161	-8046	1937,4

En uygun modelin seçilmesinde kriter; AIC, BIC ve Pearson Ki-kare değerlerinin en küçük olduğu, log-olabilirlik değerinin ise en büyük olduğu modelin seçilmesidir.

Çizelge 5.12’de görüldüğü üzere çalışma saati verileri için en uygun model; AIC, BIC ve Pearson Ki-kare değerlerinin en küçük, log-olabilirlik değerinin ise en büyük olmasından dolayı ‘*Negatif Binom Regresyon*’ modeli olarak seçilmiştir.

Tüm Olası Alt Kümeler Tekniği ile Değişken Seçimi

Çalışmanın bu aşamasında en iyi modeli bulmak için değişken seçimi yapılmıştır. ***Tüm Olası Alt Kümeler Tekniği*** ile AIC, BIC, Mallow'un Cp kriteri kullanılarak modele alınacak değişkenler belirlenmiştir. Bu teknikte 2^k - 1 tane alt küme oluşturulmuştur. 7 tane açıklayıcı değişken için 127 (2^7 - 1) model kombinasyonu oluşturulmuştur.

1. Model : (*Sektör , Cinsiyet, Eğitim, Kayıtlılık, Süreklilik, Yaş)*

2. Model : (*Sektör , Cinsiyet, Eğitim, Meslek , Kayıtlılık, Süreklilik, Yaş)*

3. Model : (*Sektör , Eğitim, Kayıtlılık, Süreklilik, Yaş)*

4. Model : (*Sektör , Eğitim, Meslek , Kayıtlılık, Süreklilik, Yaş)*

5. Model : (*Sektör , Cinsiyet, Eğitim, Meslek , Kayıtlılık, Süreklilik)*

6. Model : (*Sektör , Cinsiyet, Eğitim, Kayıtlılık, Süreklilik)*

7. Model : (*Sektör , Eğitim, Kayıtlılık, Süreklilik)*

8. Model : (*Sektör , Eğitim, Meslek , Kayıtlılık, Süreklilik)*

9. Model : (*Sektör , Cinsiyet, Eğitim, Meslek , Kayıtlılık, Süreklilik, Yaş)*

10. Model : (*Eğitim, Meslek , Kayıtlılık, Süreklilik, Yaş)*

'

'

'

126. Model : (*Yaş)*

127. Model : (*Süreklilik)*

Seçim kriterleri olarak AIC, BIC ve Mallow'un Cp kriteri kullanılmıştır ve tüm bu alt modeller için bu kriterler hesaplanmıştır. Sonuçlar Çizelge 5.13'te verilmiştir. Modelde kullanılan değişkenlerin altında yer alan *0 değeri,* o değişkenin modele alınmadığını; 1 değeri ise o değişkenin modelde olduğu anlamına gelmektedir.

Çizelge 5.13. Çalışma Saati Verisi için Değişken Seçimi

Model	Sektör	Cinsiyet	Eğitim	Meslek	Kayıtlılık	Süreklilik	Yaş	AIC	BIC	Mallows' Cp
1	1	1	1	0	1	1	1	10693	10732	7.2
2	1	1	1	1	1	1	1	10694	10738	8.0
3	1	0	1	0	1	1	1	10694	10728	8.3
4	1	0	1	1	1	1	1	10696	10735	9.7
5	1	1	1	1	1	1	0	10700	10739	14.0
6	1	1	1	0	1	1	0	10700	10734	14.5
7	1	0	1	0	1	1	0	10702	10730	16.6
8	1	0	1	1	1	1	0	10703	10736	17.0
9	0	1	1	1	1	1	1	10707	10746	20.7
10	0	0	1	1	1	1	1	10707	10740	21.2
11	0	0	1	0	1	1	1	10708	10736	22.3
12	0	1	1	0	1	1	1	10709	10742	22.9
13	0	1	1	1	1	1	0	10716	10749	30.0
14	0	0	1	1	1	1	0	10717	10745	31.5
15	0	0	1	0	1	1	0	10721	10744	35.7
16	0	1	1	0	1	1	0	10721	10749	35.7
17	1	1	1	0	1	0	1	10743	10776	57.3
18	1	1	1	0	0	1	1	10743	10776	57.5
19	1	1	1	1	0	1	1	10745	10784	59.3
20	1	1	1	1	1	0	1	10745	10784	59.3
21	1	0	1	0	1	0	1	10745	10773	59.9
22	1	1	1	0	1	0	0	10747	10774	61.3
23	1	0	1	1	1	0	1	10747	10781	61.8
24	1	0	1	0	0	1	1	10748	10776	62.9
25	1	1	1	1	1	0	0	10748	10782	63.0
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
100	0	1	0	1	0	1	1	11036	11064	382.2
101	1	0	0	1	0	1	0	11037	11060	383.7
102	1	0	0	1	0	1	1	11039	11067	385.1
103	1	1	0	1	0	0	0	11040	11063	387.1
104	1	1	0	1	0	0	1	11042	11070	389.1
105	0	0	0	1	0	1	0	11055	11072	405.1
106	0	1	0	1	0	0	0	11055	11072	405.4
107	0	0	0	1	0	1	1	11055	11078	405.3
108	0	1	0	1	0	0	1	11057	11079	407.1
109	1	0	0	1	0	0	0	11063	11080	415.3
110	1	0	0	1	0	0	1	11065	11088	417.1
111	0	0	0	1	0	0	0	11075	11086	430.2
112	0	0	0	1	0	0	1	11077	11093	431.3
113	1	1	0	0	0	1	0	11126	11148	491.6
114	1	1	0	0	0	1	1	11127	11155	492.6
115	1	1	0	0	0	0	0	11134	11151	502.1
116	1	1	0	0	0	0	1	11135	11158	503.6
117	1	0	0	0	0	1	0	11138	11155	507.5
118	1	0	0	0	0	1	1	11139	11161	507.6
119	1	0	0	0	0	0	0	11147	11158	518.8
120	1	0	0	0	0	0	1	11148	11164	519.6
121	0	1	0	0	0	1	1	11191	11214	574.9
122	0	1	0	0	0	0	1	11192	11209	576.4
123	0	1	0	0	0	0	0	11194	11205	579.7
124	0	1	0	0	0	1	0	11194	11211	579.1
125	0	0	0	0	0	1	1	11196	11213	582.0
126	0	0	0	0	0	0	1	11198	11209	584.2
127	0	0	0	0	0	1	0	11200	11211	587.2

Değişken seçiminde kullanılan *Tüm Olası Alt Kümeler Tekniği* sonucu elde edilen 127 modelin AIC, BIC ve Mallow'un Cp kriteri tam sonuçları EK 1'de yer almaktadır.

Çizelge 5.13 incelendiğinde en küçük AIC, BIC ve Mallow'un Cp kriterinin yer aldığı 1.Model en iyi model olarak seçilmiştir. 1. Model'de *Meslek* değişkeni çıkartılmış; *Sektör, Cinsiyet, Eğitim, Kayıtlılık* ve *Süreklilik, Yaş* değişkenleri yer almıştır.

1. Model için Negatif Binom Regresyon modeli

En iyi model olarak seçilen 1. Model için Negatif Binom Regresyon modeli uygulanmış ve sonuçlar yorumlanmıştır.

Çizelge 5.14. 1.model için Negatif Binom Regresyon Modeli

Açıklayıcı Değişkenler	**Tahmin (β)**	**St.Hata**	**Üstel Tahmin Exp(β)**	**p - değeri**
(Sabit)	3,7919	0,0337	44,3405	0,0000
Tarım	0,1257	0,0358	1,1339	0,0004
Sanayi	0,0603	0,0204	1,0621	0,0031
İnşaat	0,0176	0,0228	1,0178	0,4395
Erkek	0,0046	0,0184	1,0046	0,8042
Okur yazar olmayanlar	0,4211	0,0330	1,5237	0,0000
Lise altı	0,2831	0,0187	1,3272	0,0000
Lise	0,2186	0,0226	1,2443	0,0000
Mesleki veya teknik lise	0,1681	0,0324	1,1830	0,0000
Kayıtlı	-0,1082	0,0165	0,8974	0,0000
Sürekli	0,1254	0,0187	1,1336	0,0000
Yaş	0,1254	0,0187	1,1336	0,0000
Yayılım Parametresi	0,0643	0,0027		

- H_0: $\beta_{1i} = 0$ (Sektör değişkeninin modele katkısı anlamlı değildir. $i = 1, 2, 3$)
  H_1: $\beta_{1i} \neq 0$ (Sektör değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_{2j} = 0$ (Cinsiyet değişkeninin modele katkısı anlamlı değildir. $j = 1$)
  H_1: $\beta_{2j} \neq 0$ (Cinsiyet değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_{3k} = 0$ (Eğitim değişkeninin modele katkısı anlamlı değildir. $k = 1, 2, 3, 4$)
  H_1: $\beta_{3k} \neq 0$ (Eğitim değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_{5m} = 0$ (Kayıtlılık değişkeninin modele katkısı anlamlı değildir. $m = 1$)
  H_1: $\beta_{5m} \neq 0$ (Kayıtlılık değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_{6n} = 0$ (Süreklilik değişkeninin modele katkısı anlamlı değildir. $n = 1$)
  H_1: $\beta_{6n} \neq 0$ (Süreklilik değişkeninin modele katkısı anlamlıdır.)

- H_0: $\beta_7 = 0$ (Yaş değişkeninin modele katkısı anlamlı değildir.)
  H_1: $\beta_7 \neq 0$ (Yaş değişkeninin modele katkısı anlamlıdır.)

Çizelge 5.14'te belirtilen p-değerleri incelendiğinde; cinsiyet değişkeni, sektör değişkeni gruplarından 'inşaat' p-değerleri 0.05'ten büyük olduğu için %5 yanılma düzeyinde istatistiksel açıdan modele anlamlı bir katkısı olmadığı, diğer açıklayıcı değişkenlerin %5 yanılma düzeyinde anlamlı olduğu bulunmuştur.

Negatif binom modeli yayılım parametresi olan $\boldsymbol{\alpha}$ değeri sıfırdan farklıdır. (α = 0.06) Veriler negatif binom modeline uygundur.

Çizelge 5.14'e göre,

- Beklenen haftalık çalışma saatine, tarım sektöründe çalışanların, hizmet sektöründe çalışanlara göre 1,13 kat daha fazla katkı yaptığı söylenebilir.
- Beklenen haftalık çalışma saatine, sanayi sektöründe çalışanların, hizmet sektöründe çalışanlara göre 1,06 kat daha fazla katkı yaptığı söylenebilir.
- Beklenen haftalık çalışma saatine, okur yazar olmayanların, yüksek öğretim mezunlarına göre 1,52 kat daha fazla katkı yaptığı söylenebilir.
- Beklenen haftalık çalışma saatine, lise altı eğitimlilerin, yüksek öğretim mezunlarına göre 1,33 kat daha fazla katkı yaptığı söylenebilir.
- Beklenen haftalık çalışma saatine, lise mezunlarının, yüksek öğretim mezunlarına göre 1,24 kat daha fazla katkı yaptığı söylenebilir.
- Beklenen haftalık çalışma saatine, mesleki veya teknik lise mezunlarının, yüksek öğretim mezunlarına göre 1,18 kat daha fazla katkı yaptığı; eğitim seviyesi arttıkça beklenen haftalık çalışma saatinin de azaldığı söylenebilir.
- Bir Sosyal Güvenlik Kurumu'na kayıtlı olduğunu beyan eden fertlerin, kayıtdışı olarak çalışan fertlere göre beklenen haftalık çalışma saatine daha az katkı yaptığı söylenebilir. Buradan kayıtlı istihdamın önemi anlaşılmaktadır.
- Beklenen haftalık çalışma saatine sürekli bir işe sahip olan fertlerin, geçici olarak çalışan fertlerden daha az katkı yaptığı söylenebilir.
- Yaş arttıkça beklenen haftalık çalışma saatinin arttığı söylenebilir.

6. SONUÇ

Bağımlı değişken, negatif olmayan kesikli değerlerden oluştuğunda sayma verisi modelleri kullanılabilir. Sayım verisi analizinde kullanılan en yaygın modellerden biri Poisson regresyon modelidir. Poisson regresyon modelinin en önemli özelliği varyansın ve ortalamanın eşit olmasıdır. Genellikle, bu özellik sağlanamamaktadır. Bu durumda negatif binom regresyon analizi veya genelleştirilmiş Poisson regresyon analizi kullanılmaktadır.

Bu çalışmada SPSS, R ve SAS programları ile analizler yapılmış ve sonuçlar elde edilmiştir.

Uygulamada, Türkiye İstatistik Kurumu'ndan alınan Hanehalkı İşgücü Araştırması verileri kullanılmıştır. 2016 yılında SİİRT bölgesinden araştırma kapsamında 14 523 kişi ile görüşülmüştür. Referans haftasında istihdamda olan 2488 kişi filtrelenmiş ve Poisson regresyon, negatif binom regresyon, genelleştirilmiş Poisson regresyon analizleri yapılmıştır.

Bağımlı değişken; fertlerin bir haftada kaç saat çalıştığı bilgisini içeren ***Çalışma*** değişkenidir. Bağımlı değişken üzerinde etkisi olduğu düşünülen açıklayıcı değişkenler ise ***Sektör, Yaş, Cinsiyet, Eğitim, Meslek, Kayıtlılık, Süreklilik*** olarak seçilmiştir. Uygulamada ilk olarak bu değişkenlere ilişkin analizler yapılmış ve özet bilgiler verilmiştir. Daha sonra, bağımlı değişkenin kesikli verilerden oluşması ve sağa çarpık olması nedeniyle Poisson regresyon analizi uygulanmıştır. Veride aşırı yayılımın var olmasından dolayı genelleştirilmiş Poisson regresyon ve negatif binom regresyon analizleri de uygulanmıştır. Seçim kriterleri ile en uygun model olarak negatif binom regresyon modeli seçilmiştir.

Bir sonraki adımda en iyi modeli bulmak için Tüm Olası Alt Kümeler Tekniği kullanılarak değişken seçimi yapılmıştır. Seçim kriterleri olarak AIC, BIC ve Mallow'un Cp kriteri kullanılmış ve en iyi model olarak *Sektör, Cinsiyet, Eğitim, Kayıtlılık* ve *Süreklilik, Yaş* değişkenlerinin yer aldığı model seçilmiştir. Bu modele negatif binom regresyon analizi uygulanmıştır.

Negatif binom regresyon analizi sonucunda bağımlı değişken olan çalışma saatine; cinsiyet değişkeni ve sektör değişkeni gruplarından 'inşaat'ın istatistiksel olarak modele anlamlı bir katkısı olmadığı görülmüştür.

Genel olarak sonuçlara bakıldığında; tarım ve sanayi sektöründe çalışanların, hizmet sektöründe çalışanlara göre beklenen çalışma saatine daha fazla katkıda bulunduğu, eğitim seviyesi arttıkça beklenen haftalık ortalama çalışma saatinin kısaldığı, kayıtlı çalışanların kayıt dışı çalışanlara göre haftalık daha az saat çalışması beklendiği, sürekli bir işte çalışanların geçici işte çalışanlara göre haftalık ortalama daha fazla çalışması beklendiği, yaş arttıkça beklenen haftalık ortalama çalışma saatinin de arttığı sonuçlarına ulaşılmıştır.

KAYNAKLAR

[1] Shankar, V., Milton, J., Mannering, F., Modeling Accident Frequencies As Zero-Altered Probability Processes: An Emprical Inquiry, *Accident Analysis and Prevention*, 29, 6, 829-837, **1997.**

[2] Shoukri et al., The Poisson Inverse Gaussian Regression Model in the Analysis of Clustered Counts Data, *Journal of Data Science,* 17-32, **2004**

[3] Sezgin, H., F., Deniz, E., Poisson Regresyon Modelinde Aşırı Yayılım Durumu ve Negatif Binomial Regresyon Analizinin Türkiye Grev Sayıları Üzerine Bir Uygulama, *Yönetim Dergisi*, 48, 17-25, **2004.**

[4] Aktaş, A., Saraçbaşı,O., Negatif Binom Regresyon Modeli, *VIII. Ulusal Biyoistatistik Kongresi,* 20-22 Eylül 2005, Bursa, 124 – 129, **2005.**

[5] Yeşilova, A., Yılmaz, A., Kaki B., Norduz Erkek Kuzularının Bazı Kesikli Üreme Davranış Özelliklerinin Analizinde Doğrusal Olmayan Regresyon Modellerin Kullanılması, *Yüzüncü Yıl Üniversitesi, Ziraat Fakültesi Tarım Bilimleri Dergisi,* 16(2):87-92, **2006.**

[6] Özmen, İ.,Famoye, F., Count Regression Modelswith an Application to Zoological Data Containing Structural Zeros, *Journal of Data Science,* 5, 491-502, **2007.**

[7] Tüzel, S., *Hasar Sıklıkları İçin Sıfır Yığılmalı Kesikli Modeller*, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, **2011.**

[8] Arı, A., Önder, H., Farklı Veri Yapılarında Kullanılabilecek Regresyon Yöntemleri, Ondokuz Mayıs Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, *Anadolu Tarım Bilim Dergisi*, Samsun, **2012.**

[9] Asrul, A. M., Naing, N. N., Analysis Death Rate of Age Model with Excess Zeros using Zero Inflated Negative Binomial and Negative Binomial Death Rate: Mortality AIDS Co-Infection Patients, *2nd Annual International Conference on Qualitative and Quantitative Economics Research,* Kelantan Malaysia, **2012.**

[10] Pamukçu, E., Çolak, C., Halisdemir, N., Modeling of The Number of Divorce in Turkey Using The Generalized Poisson, Quasi-Poisson and Negative Binomial Regression, *Turkish Journal of Science & Technology, Volume 9(1*), 89-96, **2014.**

[11] Elmalı, K., *Kantil Regresyon Ve Negatif Binomial Regresyon ile İllerde Kullanılan İlaç Sayısına Etki Eden Faktörlerin İncelenmesi*, Atatürk Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek lisans tezi, Erzurum, **2014.**

[12] Açıkyürek, G., *Poisson Regresyon Ve Bir Uygulama*, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, **2016.**

[13] Nelder J.A., Wedderburn, R.W.M., *Generalized Linear Models, Journal of the Royal Statistical Society Series* A, 135, 3, 370-384, **1972**.

[14] De Jong, P., Heller, G., Z., *Generalized Linear Models for Insurance Data* Cambridge University Press, London, 196, **2008.**

[15] McCullagh, P., Nelder, J.A., *Generalized Linear Models*, Chapman and Hall, London, 511, **1989.**

[16] Sarul, L. S., *Aktüerya Alanında Genelleştirilmiş Lineer Modeller Kullanılarak Prim Fiyatlama Ve Filo Verisi Üzerinde Bir Uygulama*, Doktora Tezi, Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul, **2012.**

[17] Agresti, A.. *Categorical Data Analysis*, A John Wiley & Sons, Ins, Canada, pp.710, **2002.**

[18] Cameron, C.- Trivedi, P., *Regression Analysis of Count Data*, Cambridge University Press, West Nyack, NY, USA, **1998.**

[19] Anonim, Wikipedia, http://www.wiki-zero.net/index.php?q=aHR0cHM6Ly90ci53aWtpcGVkaWEub3JnL3dpa2kvUG9pc3Nvbl9kYcSfxLFsxLFtxLE (Mayıs, **2018**)

[20] Deniz, Ö.,Poisson Regresyon Analizi, *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 7, 59-72, **2005.**

[21] Consul, P.C., Famoye, F., *Generalized Poisson Regression Model*, Communications in Statistics-Theory and Method, 21, 1, 89-109, **1992.**

[22] Yeşilova, A., Atlıhan, R., Farklı Sıcaklıkların Scymnussubvillosus'un Bıraktığı Yumurta Sayıları Üzerine Etkilerinin Karışımlı Poisson Regresyon ile Analiz Edilmesi, *Yüzüncü Yıl Üniversitesi, Ziraat Fakültesi, Tarım Bilimleri Dergisi,* 17(2): 73-79, **2007.**

[23] Üçer, B., *A Poisson Regression Model with Autocorrelation*, Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi Fen Bilimleri Enstitüsü, İzmir, **2003.**

[24] Dean, C., Lawless, J.F., Test for Defecting Overdispersion in Poisson Regression Models, *Journal of American Statistical Association*, 84, 467- 472, **1989.**

[25] Ismail, N.,Zamani, H., Estimation of ClaimCount Data Using Negative Binomial GeneralizedPoisson, Zero-Inflated Negative Binomia land Zero-Inflated Generalized Poisson Regression Models, *Casualty Actuarial Society E-Forum*, Spring, **2013.**

[26] Dodge, Y., *The Concise Encyclopedia of Statistics*, Springer, New York, **2008.**

[27] Hilbe J. M., *Negative Binomial Regression*, Second Edition, Cambridge University Press, **2011.**

[28] Denuit, M.,Marechal, X., Pitrebois, S., Walhin, J.F., 2007, *Actuarial Modelling of Claim Counts: Risk Classification, Credibilityand Bonus-Malus Systems*, John Wiley & Sons, 356p. 28, England, **2007.**

[29] Boucher, J.-P., Denuit, M., Guillen, M., Risk Classification for Claim Counts: Mixed Poisson, Zero-Inflated Mixed Poisson and Hurdle Models, *North American Actuarial Journal*, 11, 4, 110-131, **2007.**

[30] Famoye, F., Singh, K.P., Zero-Inflated Generalized PoissonRegression Model with an Application to Domestic Violence Data, *Journal of Data Science*, 4, 117-130, **2006.**

[31] Kaya, Y., Yeşilova, A., Çetinkaya İ., Akademik Yayın Başarısının Sıfır Değer Ağırlıklı Regresyon Yöntemler Kullanılarak Modellenmesi, *Akademik Bilişim,* Malatya, **2011.**

[32] Tüzel, S., Sucu, M., Hasar Sıklıkları için Sıfır Yığılmalı Kesikli Modeller, *İstatistikçiler Dergisi*, 5, 23-31, **2012**.

[33] Lambert, D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 34, 1-14, **1992.**

[34] Hall, D. B. *Zero-inflated Poisson and binomial regression with random effects: A case study*. Biometrics 56, 1030-1039, **2000.**

[35] Lee, A. H., Wang, K. and Yau, K. K. W., Analysis of zero-inflated Poisson data incorporating extent of exposure, *Biometrical Journal* 43, 963-975, **2001.**

[36] Ridout, M., Demetrio, C. G. B. and Hinde, J., Models for count data with many zeros. Invited paper presented at the *Nineteenth International Biometric Conference*, Cape Town, South Africa, 179-190, **1998.**

[37] Greene, W. H., *Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models*, Technical report, **1994.**

[38] Czado, C.,Erhardt, V., Min, A., Wagner, S., 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, **2007.**

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

[40] Ucal, M. Ş., Ekonometrik Model Seçim Kriterleri Üzerine Kısa Bir İnceleme, *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, Cilt 7, Sayı 2, **2006.**

[41] Park, B. J., Lord. D., Negative Binomial Regression Models and Estimation Methods, *CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations,* Appendıx-D, **2010.**

[42] Yardımcı, A., *Doğrusal Regresyonda Değişken Seçimine Bayesci Yaklaşımların Karşılaştırılması*, Doktora Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, **2014.**

[43] Anonim, Türkiye İstatistik Kurumu Tanım ve Kavramlar, http://www.tuik.gov.tr/MicroVeri/Hia_2011/turkce/metaveri/tanim/index.html, (Mayıs, **2018**)

[44] Anonim, 18. Çalışma İstatistikçileri Konferansı, http://www.ilo.org/global/statistics-and-databases/meetings-and-events/international-conference-of-labour-statisticians/WCMS_092024/lang--en/index.htm (Mayıs, **2018)**

[45] Chakraborty, S., Generating Discrete Analogues Of Continuous Probability Distributions-A Survey Of Methods And Constructions, Journal of Statistical Distributions and Applications, **2015.**

EKLER

EK 1. Çalışma Saati Verileri için Değişken Seçimi

Model	Sektör	Cinsiyet	Eğitim	Meslek	Kayıtlılık	Süreklilik	Yaş	AIC	BIC	Mallows' Cp
1	1	1	1	0	1	1	1	10693	10732	7.2
2	1	1	1	1	1	1	1	10694	10738	8.0
3	1	0	1	0	1	1	1	10694	10728	8.3
4	1	0	1	1	1	1	1	10696	10735	9.7
5	1	1	1	1	1	1	0	10700	10739	14.0
6	1	1	1	0	1	1	0	10700	10734	14.5
7	1	0	1	0	1	1	0	10702	10730	16.6
8	1	0	1	1	1	1	0	10703	10736	17.0
9	0	1	1	1	1	1	1	10707	10746	20.7
10	0	0	1	1	1	1	1	10707	10740	21.2
11	0	0	1	0	1	1	1	10708	10736	22.3
12	0	1	1	0	1	1	1	10709	10742	22.9
13	0	1	1	1	1	1	0	10716	10749	30.0
14	0	0	1	1	1	1	0	10717	10745	31.5
15	0	0	1	0	1	1	0	10721	10744	35.7
16	0	1	1	0	1	1	0	10721	10749	35.7
17	1	1	1	0	1	0	1	10743	10776	57.3
18	1	1	1	0	0	1	1	10743	10776	57.5
19	1	1	1	1	0	1	1	10745	10784	59.3
20	1	1	1	1	1	0	1	10745	10784	59.3
21	1	0	1	0	1	0	1	10745	10773	59.9
22	1	1	1	0	1	0	0	10747	10774	61.3
23	1	0	1	1	1	0	1	10747	10781	61.8
24	1	0	1	0	0	1	1	10748	10776	62.9
25	1	1	1	1	1	0	0	10748	10782	63.0
26	1	0	1	0	1	0	0	10750	10772	64.8
27	1	0	1	1	0	1	1	10750	10784	64.9
28	0	1	1	0	1	0	1	10750	10778	65.2
29	0	0	1	0	1	0	1	10751	10773	66.1
30	0	1	1	1	1	0	1	10752	10785	66.5
31	1	0	1	1	1	0	0	10752	10780	66.7
32	0	0	1	1	1	0	1	10753	10781	67.9
33	1	1	1	0	0	1	0	10755	10783	69.9
34	1	1	1	1	0	1	0	10756	10789	70.8
35	0	1	1	1	1	0	0	10757	10785	72.5
36	0	1	1	0	1	0	0	10758	10780	72.9
37	0	0	1	0	1	0	0	10759	10776	74.3
38	0	0	1	1	1	0	0	10760	10782	75.0
39	0	1	1	1	0	1	1	10760	10793	75.1
40	0	1	1	0	0	1	1	10760	10788	75.1
41	1	0	1	0	0	1	0	10762	10784	77.5
42	0	0	1	0	0	1	1	10762	10785	77.7
43	0	0	1	1	0	1	1	10763	10791	78.8
44	1	0	1	1	0	1	0	10764	10792	79.2
45	0	1	1	1	0	1	0	10776	10804	91.6
46	0	1	1	0	0	1	0	10779	10801	95.0
47	0	0	1	1	0	1	0	10782	10804	97.8
48	0	0	1	0	0	1	0	10783	10800	99.2
49	1	1	1	0	0	0	1	10784	10812	100.4

EK 1 devam: Çalışma Saati Verileri için Değişken Seçimi

Model	Sektör	Cinsiyet	Eğitim	Meslek	Kayıtlılık	Süreklilik	Yaş	AIC	BIC	Mallows' Cp
50	1	1	1	1	0	0	1	10786	10819	102.1
51	1	0	1	0	0	0	1	10791	10813	107.4
52	1	1	1	0	0	0	0	10792	10814	108.5
53	1	0	1	1	0	0	1	10792	10820	108.5
54	0	1	1	0	0	0	1	10794	10816	110.2
55	1	1	1	1	0	0	0	10794	10822	110.5
56	0	1	1	1	0	0	1	10795	10823	112.1
57	0	0	1	0	0	0	1	10798	10814	114.7
58	0	0	1	1	0	0	1	10800	10822	116.7
59	1	0	1	0	0	0	0	10800	10817	117.5
60	1	0	1	1	0	0	0	10802	10824	119.3
61	0	1	1	0	0	0	0	10806	10823	123.8
62	0	1	1	1	0	0	0	10807	10829	124.2
63	0	0	1	0	0	0	0	10812	10823	129.7
64	0	0	1	1	0	0	0	10813	10830	131.2
65	1	1	0	1	1	1	0	10838	10871	157.3
66	1	1	0	1	1	1	1	10840	10879	159.3
67	1	0	0	1	1	1	0	10844	10872	164.3
68	1	0	0	1	1	1	1	10846	10880	166.2
69	0	1	0	1	1	1	0	10852	10880	172.8
70	0	1	0	1	1	1	1	10854	10887	174.5
71	0	0	0	1	1	1	0	10857	10879	177.9
72	0	0	0	1	1	1	1	10858	10886	179.3
73	1	1	0	1	1	0	0	10883	10911	206.9
74	1	1	0	1	1	0	1	10885	10919	208.7
75	1	0	0	1	1	0	0	10890	10913	214.8
76	0	1	0	1	1	0	0	10892	10914	216.0
77	1	0	0	1	1	0	1	10892	10920	216.8
78	0	1	0	1	1	0	1	10894	10921	218.0
79	0	0	0	1	1	0	0	10897	10914	222.3
80	0	0	0	1	1	0	1	10899	10921	224.2
81	1	1	0	0	1	1	0	10909	10937	234.9
82	1	0	0	0	1	1	0	10910	10932	236.1
83	1	1	0	0	1	1	1	10910	10944	236.5
84	1	0	0	0	1	1	1	10911	10939	237.6
85	1	1	0	0	1	0	0	10936	10958	265.7
86	1	0	0	0	1	0	0	10938	10954	267.9
87	1	1	0	0	1	0	1	10938	10966	267.6
88	1	0	0	0	1	0	1	10940	10962	269.8
89	0	0	0	0	1	1	1	10951	10973	282.4
90	0	0	0	0	1	1	0	10951	10968	283.2
91	0	1	0	0	1	1	1	10952	10980	283.8
92	0	1	0	0	1	1	0	10952	10975	284.5
93	0	0	0	0	1	0	0	10967	10978	301.9
94	0	0	0	0	1	0	1	10968	10984	302.4
95	0	1	0	0	1	0	0	10968	10985	302.5
96	0	1	0	0	1	0	1	10969	10991	303.1
97	1	1	0	1	0	1	0	11014	11042	355.8
98	1	1	0	1	0	1	1	11016	11050	357.7
99	0	1	0	1	0	1	0	11035	11057	381.0
100	0	1	0	1	0	1	1	11036	11064	382.2

EK 1 devam: Çalışma Saati Verileri için Değişken Seçimi

Model	Sektör	Cinsiyet	Eğitim	Meslek	Kayıtlılık	Süreklilik	Yaş	AIC	BIC	Mallows' Cp
101	1	0	0	1	0	1	0	11037	11060	383.7
102	1	0	0	1	0	1	1	11039	11067	385.1
103	1	1	0	1	0	0	0	11040	11063	387.1
104	1	1	0	1	0	0	1	11042	11070	389.1
105	0	0	0	1	0	1	0	11055	11072	405.1
106	0	1	0	1	0	0	0	11055	11072	405.4
107	0	0	0	1	0	1	1	11055	11078	405.3
108	0	1	0	1	0	0	1	11057	11079	407.1
109	1	0	0	1	0	0	0	11063	11080	415.3
110	1	0	0	1	0	0	1	11065	11088	417.1
111	0	0	0	1	0	0	0	11075	11086	430.2
112	0	0	0	1	0	0	1	11077	11093	431.3
113	1	1	0	0	0	1	0	11126	11148	491.6
114	1	1	0	0	0	1	1	11127	11155	492.6
115	1	1	0	0	0	0	0	11134	11151	502.1
116	1	1	0	0	0	0	1	11135	11158	503.6
117	1	0	0	0	0	1	0	11138	11155	507.5
118	1	0	0	0	0	1	1	11139	11161	507.6
119	1	0	0	0	0	0	0	11147	11158	518.8
120	1	0	0	0	0	0	1	11148	11164	519.6
121	0	1	0	0	0	1	1	11191	11214	574.9
122	0	1	0	0	0	0	1	11192	11209	576.4
123	0	1	0	0	0	0	0	11194	11205	579.7
124	0	1	0	0	0	1	0	11194	11211	579.1
125	0	0	0	0	0	1	1	11196	11213	582.0
126	0	0	0	0	0	0	1	11198	11209	584.2
127	0	0	0	0	0	1	0	11200	11211	587.2

ÖZGEÇMİŞ

Kimlik Bilgileri

Adı Soyadı : Gözde Nur DİNARCAN

Doğum Yeri : Ankara

Medeni Hali : Evli

E-posta : gozdenur.besler@tuik.gov.tr

Adresi : Türkiye İstatistik Kurumu, Necatibey cad. No:114 Çankaya/ANKARA

Eğitim

Lise : Ümitköy Anadolu Lisesi

Lisans : Hacettepe Üniversitesi, İstatistik Bölümü

Yabancı Dil ve Düzeyi

İngilizce, İleri seviye

İş Deneyimi

2015- : Türkiye İstatistik Kurumu Uzman yardımcısı

Deneyim Alanları

-

Tezden Üretilmiş Projeler ve Bütçesi

-

Tezden Üretilmiş Yayınlar

-

Tezden Üretilmiş Tebliğ ve/veya Poster Sunumu ile Katıldığı Toplantılar

-

HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YÜKSEK LİSANS/~~DOKTORA~~ TEZ ÇALIŞMASI ORJİNALLİK RAPORU

HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ
......İSTATİSTİK...................... ANABİLİM DALI BAŞKANLIĞI'NA

Tarih: 22/06/2018

Tez Başlığı / Konusu: Sayma Verisi için Regresyon Modelleri ve Bir Uygulama

Yukarıda başlığı/konusu gösterilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler d) Sonuç kısımlarından oluşan toplam ...46... sayfalık kısmına ilişkin, 22../06../2018 tarihinde ~~şahsım~~/tez danışmanım tarafından *Turnitin* adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 9... 'tür.

Uygulanan filtrelemeler:

1- Kaynakça hariç
2- Alıntılar hariç/~~dâhil~~
3- 5 kelimeden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü Tez Çalışması Orjinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Tarih ve İmza
22.06.2018

Adı Soyadı: Gözde Nur DİNARCAN
Öğrenci No: N13120803
Anabilim Dalı: İstatistik
Programı: Tezli Yüksek Lisans
Statüsü: ☒ Y.Lisans ☐ Doktora ☐ Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.

Prof. Dr. Meral Çetin
(Unvan, Ad Soyad, İmza)