T.C. İSTANBUL TİCARET ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

POISSON REGRESYON ANALİZİ VE TÜRKİYE'DEKİ EVLİLİK İSTATİSTİKLERİNE UYGULANMASI

Ozan ALTUNBAŞ

Danışman
Prof. Dr. Özlem DENİZ BAŞAR

YÜKSEK LİSANS TEZİ
İSTATİSTİK ANABİLİM DALI
İSTANBUL - 2022

KABUL VE ONAY SAYFASI

Ozan ALTUNBAŞ tarafından hazırlanan **"Poisson Regresyon Analizi ve Türkiye'deki Evlilik İstatistiklerine Uygulanması"** adlı tez çalışması 27/04/2022 tarihinde aşağıdaki jüri üyeleri önünde başarı ile savunularak, İstanbul Ticaret Üniversitesi Fen Bilimleri Enstitüsü **İstatistik Anabilim Dalı**'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Danışman	**Prof. Dr. Özlem DENİZ BAŞAR** İstanbul Ticaret Üniversitesi	.....
Jüri Üyesi	**Doç. Dr. Seda BAĞDATLI KALKAN** İstanbul Ticaret Üniversitesi	.....
Jüri Üyesi	**Doç. Dr. Selay GİRAY YAKUT** Marmara Üniversitesi	.....

Onay Tarihi : 27/05/2022

İstanbul Ticaret Üniversitesi, Fen Bilimleri Enstitüsünün 27.05.2022 tarih ve 2022/347 numaralı Yönetim Kurulu Kararının 5. Maddesi gereğince, ders yüklerini ve tez yükümlülüğünü yerine getirdiği belirlenen "Ozan ALTUNBAŞ"
adlı öğrencinin mezun olmasına oy birliğiyle karar verilmiştir.

Prof. Dr. Necip ŞİMŞEK
Enstitü Müdürü

AKADEMİK VE ETİK KURALLARA
UYGUNLUK BEYANI

İstanbul Ticaret 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.

Ozan ALTUNBAŞ

İÇİNDEKİLER

Sayfa

İÇİNDEKİLER i
ÖZET iii
ABSTRACT iv
TEŞEKKÜRLER v
ŞEKİLLER DİZİNİ vi
ÇİZELGELER DİZİNİ vii
SİMGELER VE KISALTMALAR DİZİNİ viii
1. GİRİŞ 1
2. LİTERATÜR ÖZETİ 3
2.1. Genelleştirilmiş Doğrusal Modeller ile İlgili Yapılan Çalışmalar 3
2.2. Poisson Regresyon ile İlgili Yapılan Çalışmalar 4
3. GENELLEŞTİRİLMİŞ DOĞRUSAL MODELLER 5
3.1. Genelleştirilmiş Doğrusal Modellerin Biçimleri 5
3.2. Genelleştirilmiş Doğrusal Modeller ve Tahmin 10
3.3. Genelleştirilmiş Doğrusal Modeller ve Sonuçlar 14
3.3.1. β adına hipotez testi ve güven aralıkları 14
3.3.2. Sapma ve uyum iyiliği 16
3.4. Model Tercihi 17
3.5. Artıklar ve Analizleri 18
3.5.1.Sapma artığı 18
3.5.2.Pearson artığı 19
4. POISSON REGRESYON 20
4.1. Poisson Regresyon Modeli 20
4.2. Çok Sıfırlı Poisson Regresyon Modeli 21
4.3. Genelleştirilmiş Çok Sıfırlı Poisson Regresyon Modeli 21
4.4. En Çok Olabilirlik Tahmin Edici 22
4.5. Poisson Regresyon Katsayıları ve Değerlendirilmesi 25
4.6. Poisson Regresyonu ve Yanlı Tahmin Ediciler 25
4.6.1. Poisson modellerinde ridge tahmin edici 25
4.6.1.1. Önerilen ridge parametresi tahmin yaklaşımları 27
4.6.2. Poisson ridge tahmin edici ve çok sıfırlı modeller 29
4.6.2.1. Önerilen ridge parametresi tahmin yaklaşımları 31
4.6.3. Modifiye edilmiş jackknifed poisson ridge tahmin edici 33
4.6.3.1. Önerilen ridge parametresi tahmin yaklaşımları 36
4.6.4. Poisson regresyon modellerinde liu tahmin edici 37
4.6.5. Liu tahmin edicide yanlılık kriterinin belirlenmesi 37
4.7. Log-Normal Dağılımlı Poisson Regresyon Modeli 38
4.8. Poisson Engel Regresyon Modeli 39
4.9. Poisson Regresyon Modellerinde Kullanılan Uyum Testleri 41
4.9.1. Sapma istatistiği 41
4.9.2. Pearson ki-kare istatistiği 42
4.9.3. Yapay r2 istatistiği 43
4.9.4. Akaike bilgi parametresi 44
4.9.5. Pearson istatistiği 44
4.10. Tahmin Edicilerin Performanslarının Değerlendirilmesi 45

5. TÜRKİYE'DEKİ EVLİLİK İSTATİSTİKLERİ ÜZERİNE POISSON REGRESYON ANALİZİ UYGULAMASI 46
5.1. Evlilik Verileri 48
5.2. Evlilik Verileri ile Elde Edilen Sonuçlar 52
6. SONUÇ VE ÖNERİLER 63
KAYNAKLAR 65
ÖZGEÇMİŞ 69

ÖZET

Yüksek Lisans Tezi

POISSON REGRESYON ANALİZİ VE TÜRKİYE'DEKİ EVLİLİK İSTATİSTİKLERİNE UYGULANMASI

Ozan ALTUNBAŞ

İstanbul Ticaret Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Prof. Dr. Özlem DENİZ BAŞAR
2022, 69 sayfa

Bağımlı değişkenler, regresyon analizlerinde normal dağılımlı, sayısal ya da iki değere sahip olabilmektedir. Veri setleri bağımlı değişkenlere ilişkin durumların ölçümlenmesiyle elde edilmekle birlikte bu verilerin modellenmesi ve değerlendirilmesi önem arz etmektedir.

Poisson regresyon modelleri, genelleştirilmiş doğrusal modeller kapsamında yer almakla birlikte aşırı dağılım, yetersiz dağılım ya çok sayıda sıfır mevcut olduğunda işlevi sınırlanmaktadır. Bu durumlarda mevcut verileri çözümleyebilmek için çok sıfırlı, engel (hurdle) ve negatif binom regresyon modelleri kullanılabilmektedir.

Çalışma kapsamında poisson regresyon modelleri kullanılarak Türkiye genelindeki evlilik sayılarına etken unsurlar araştırılmaktadır. Bu bağlamda evlilik sayısı bağımlı değişken olarak belirlenmekle birlikte cinsiyet, evlilik yaşı ve evlilik yılları bağımsız değişken olarak kabul edilmiştir. Elde edilen verilerin analiz edilmesi sonucunda genel olarak kadınlarda en yüksek evlilik yaşının 20-24 yaş arasında gerçekleştiği görülmektedir. Erkeklerde ise evliliklerin en çok 25-29 yaş aralığında gerçekleştiği gözlemlenmiştir.

Anahtar Kelimeler: Genelleştirilmiş doğrusal modeller, negatif binom regresyon, poisson regresyon modeli.

ABSTRACT

M.Sc. Thesis

POISSON REGRESSION ANALYSIS AND APPLICATION TO MARRIAGE STATISTICS IN TURKEY

Ozan ALTUNBAŞ

İstanbul Commerce University
Graduate School of Applied and Natural Sciences
Department of Statistics

Supervisor: Prof. Dr. Özlem DENİZ BAŞAR
2022, 71 pages

Dependent variables can be normally distributed, numerical or have two values in regression analysis. Modeling of selling and measuring for different types of variables other than data is important.

Poisson regression models are included in generalized enclosed models, but are limited by overdistribution, underdistribution, or when large numbers are present. For this compression, the multizero, barrier and negative binomial regression models can be used for the current version.

Working experience will be presented with a poisson regression model, and the factors that affect the marriage numbers around Turkey are being investigated. In this context, marriage numbers was determined as the dependent variable, gender, marriage age and years was accepted as the independent variable. In the analysis obtained, it is seen that the highest marriage ages for women in general are between the ages of 20-24. It has been observed that marriages mostly take place between the ages of 25-29 for men.

Keywords: Generalized linear models, negative binomial regression, poisson regression models.

TEŞEKKÜRLER

Lisans ve yüksek lisans öğrenimim ve tez çalışmam süresince bilgi ve tecrübeleriyle bana yol gösteren ve her türlü desteği sağlayan, saygı değer danışman hocam sayın Prof. Dr. Özlem DENİZ BAŞAR'a tüm katkılarından dolayı teşekkür ve saygılarımı sunarım.

Eğitim hayatım boyunca desteklerini esirgemeyen bölüm hocalarım Prof. Dr. Münevver TURANLI, Prof. Dr. Ünal Halit ÖZDEN ve Doç. Dr. Seda BAĞDATLI KALKAN'a teşekkürlerimi sunarım.

Tez çalışmam boyunca bana desteklerini ve yardımlarını esirgemeyen değerli dostum Mehmet YALÇIN'a teşekkürü borç bilirim.

Son olarak her zaman yanımda olan, eğitim hayatım boyunca ve her konuda beni destekleyen aileme ve yakın arkdaşlarıma teşekkürlerimi sunarım.

Ozan ALTUNBAŞ
İSTANBUL, 2022

ŞEKİLLER DİZİNİ

Sayfa

Şekil 5.1. Yaşlara göre ortalama evlilik sayısı 50
Şekil 5.2. Yaşlara göre kadın ve erkeklerin evlilik sayıları 50
Şekil 5.3. Yaşlara göre cinsiyetlerin evlilik farkları 51
Şekil 5.4. Kütüphanelerin çağrılması ve verinin düzenlenmesi 52
Şekil 5.5. Yaşa göre evliliğin poisson regresyonu 52
Şekil 5.6. Yıllara göre evliliğin poisson regresyonu 54
Şekil 5.7. Evliliğin yaş ve yıllar ile birlikte regresyonu 55
Şekil 5.8. Evliliğin yaş ve cinsiyet ile birlikte regresyonu 56
Şekil 5.9. Evliliğin cinsiyet ve yıllar ile birlikte regresyonu 57
Şekil 5.10. Evliliğin cinsiyet, yaş ve yıllar ile birlikte regresyonu 58
Şekil 5.11. Yıllara göre toplam evlilikler 61
Şekil 5.12. Yaşlara göre z değeri 61

ÇİZELGELER DİZİNİ

Sayfa

Çizelge 3.1. Bazı ortak bağ fonksiyonları ve ters fonksiyonları 8
Çizelge 3.2. Kanonik bağ, bağımlı değişken aralığı ve üstel aileler için koşullu varyans fonksiyonu. 9
Çizelge 5.1. Dünya Çapında Evlilik Oranları 46
Çizelge 5.2. Analizde kullanılan yaş aralıkları 48
Çizelge 5.3. 2010-2020 Arası Yaş ve Cinsiyete Göre Evlenme Sayıları. 48
Çizelge 5.4. Yaşa Göre Evlilik Sayılarının Regresyon Çizelgesi 53
Çizelge 5.5. Yıllara Göre Evlilik Sayılarının Regresyon Çizelgesi 54
Çizelge 5.6. Evliliğin Yaş ve Yıllar ile Birlikte Regresyonu Çizelgesi 55
Çizelge 5.7. Evliliğin Yaş ve Cinsiyet ile Birlikte Regresyonu Çizelgesi 56
Çizelge 5.8. Evliliğin Cinsiyet ve Yıllar ile Birlikte Regresyonu Çizelgesi 57
Çizelge 5.9. Evliliğin Cinsiyet, Yaş ve Yıllar ile Birlikte Regresyonu Çizelgesi..... 58
Çizelge 5.10. Pearson Ki-Kare Uyum İyiliği Çizelgesi 59

SİMGELER VE KISALTMALAR DİZİNİ

AHKT	Ağırlıklı Hata Kareler Toplamı
AIC	Akaike Bilgi Ölçütü (Akaike Information Criterion)
BIC	Bayes Bilgi Ölçütü (Bayesian Information Criterion)
EKK	En Küçük Kareler
IWLS	Adımsal Ağırlıklı En Küçük Kareler (Iteratively Weighted Least Squares)
JRR	Jackknifed Ridge Regresyon
LGNB	Lognormal Gamma Mixed Negatif Binom Regresyon Modeli
LIU	Lineer Birleştirilmiş Regresyon Tahmin Edici
MAE	Ortalama Mutlak Hata (Mean Absolute Error)
MJPR	Modifiye Edilmiş Jackknifed Poisson Regresyon
MJRR	Modifiye Edilmiş Jackknifed Ridge Regresyon
MLE	En Çok Olabilirlik Tahmin Edici (Maksimum Likelihood Estimator)
MSE	Hata Kareler Ortalaması (Mean Square Error)
NBHR	Negatif Binom Hurdle (Engel) Regresyon
NBLR	Negatif Binom Liu Regresyon
NBR	Negatif Binom Regresyon
NBRR	Negatif Binom Ridge Regresyon
PHR	Poisson Hurdle (Engel) Regresyon
PLR	Poisson Liu Regresyon
PR	Poisson Regresyon
PRR	Poisson Ridge Regresyon
RR	Ridge Regresyon
VIF	Varyans Şişirme Faktörü (Variance Inflation Factor)
WMSE	Ağırlıklandırılmış Hata Kareler Ortalaması (Weighted Mean Square Error)
ZIGPR	Çok Sıfırlı Genelleştirilmiş Poisson Regresyon (Zero-Inflated Generalized Poisson Regression)
ZINBR	Çok Sıfırlı Negatif Binom Regresyon (Zero-Inflated Negatif Binom Regression)
ZIPR	Çok Sıfırlı Poisson Regresyon (Zero-Inflated Poisson Regression)
ZIPRR	Çok Sıfırlı Poisson Ridge (Regresyon Zero-Inflated Poisson Ridge Regression)
X	Bağımsız değişkenler kümesi
Y	Negatif değerli olmayan tamsayı değerli bağımlı değişken
ε	Hata terimi
β	Bilinmeyen parametre vektörü
η	Doğrusal ön kestirici
θ	Yayılım parametresi
μ	Bağımlı değişken ortalama değeri
σ 2	Varyans
n	Örneklem büyüklüğü
N	Kitle büyüklüğü
W	Ağırlık matrisi
z	β'nın bir fonksiyonu
L	Olabilirlik değeri $S\beta$

j	β_j 'nin standart hatası
β_j	β_j 'nin tahmin edici
G^2	Sapma istatistiği
$\hat{\mu}$	μ'nün tahmin edici
p	Modelin parametre sayısı
q	Modeldeki açıklayıcı değişken sayısı
β_0	Doğrusal regresyon modelindeki sabit parametre
β_1	Doğrusal regresyon modelindeki eğim parametresi
d_i	i.gözleme ait sapma artığı
R_2	Belirleme katsayısı
X	Bağımsız değişkenler kümesi
Y	Bağımlı, negatif olmayan tamsayı değişkeni
β	Bilinmeyen parametre vektörü
η	Doğrusal önkestirici
ϕ	Yayılım parametresi
μ	Bağımsız değişkenlerin tahmin edilen değerleri
σ^2	Varyans
Φ(.)	Birikimli dağılım fonksiyonu
n	Gözlem sayısı
W	Köşegen matris
z	β 'nın bir fonksiyonu
Λ	Olabilirlik oranı testi
L	Olabilirlik değeri
$G^2 = D$	Sapma

1. GİRİŞ

Regresyon analizi sayesinde bağımlı değişken ile bağımsız değişkenler arasındaki bağlantı kolaylıkla incelenebilir. Regresyon ile birlikte bağımlı ve bağımsız değişkenler arasında neden-sonuç ilişkisi kurulabilir.

Kitle regresyon modelinde var olan hataların olasılık dağılım fonksiyonlarında standart normal kabul edilen dağılım söz konusu değilse genelleştirilmiş doğrusal modellerin kullanımı doğru olacaktır. Üstsel dağılım, Gamma dağılımı, Poisson dağılımı ve Binom dağılımı hataların olasılık dağılım fonksiyonları arasındadır.

Sayma verileri söz konusu olduğunda hem en basit hem de en kullanışlı yöntemlerden biri poisson regresyon modelidir. Poisson regresyon modelinde verilerin olasılığı için öncelikle poisson dağılımından yararlanılır. Poisson dağılımının en önemli özelliklerinden biri sonucun ortalamasının varyansına eşit olmasıdır. Ancak uygulamada çoğu zaman varyans ortalama değerini aştığı görülmektedir. Bu duruma aşırı yayılım (overdispersion) durumu denilmektedir ve bu tür durumlarda negatif binom regresyon modelleri kullanılmaktadır.

Poisson regresyonun klasik regresyon modelinden farkı; bağımlı değişkenin, olasılık dağılımının negatif olmayan tamsayılardan oluşan kesikli bir dağılım olmasıdır. Poisson regresyon modelinin genellikle sağlık çalışmalarında, aktüeryal bilimlerde, biyoistatistik ve demografide kullanıldığı gözlemlenmiştir.

Poisson regresyon modeline ilişkin parametre tahminleri en çok olabilirlik yöntemi kullanılarak yapılmaktadır. Veriyi en iyi açıklayan yapıyı bulabilmek için uyum ölçüleri ve artıkların incelenmesi gerekmektedir.

Bu çalışmada, ilk olarak genelleştirilmiş doğrusal modeller için modelin şekli, model seçimi, tahmini, güven aralıkları, hipotez testleri, artıkların analizi, sapma ve uyum iyiliği testi açıklanmıştır. Daha sonra sayım verilerinde en sık kullanılan yöntem olan Poisson regresyon modeline ait literatür araştırmalarına, katsayıların tahminine, yorumlanmasına, uyum iyiliği testlerine yer verilmiştir. Ayrıca Poisson regresyonda

varyansın ortalamadan büyük olması durumu olan aşırı yayılım durumu ve buna bağlı olarak da negatif binom regresyon modeli anlatılmıştır. Aşırı yayılım veya az yayılım gösteren veriler hem sayıma dayalı hem de sıfır değeri çok fazla olduğu için sağa çarpıktır. Bu tür verilerin modellenmesinde kullanılan çok sıfırlı regresyon modelleri anlatılmıştır.

Uygulama bölümünde ise, TÜİK üzerinden alınan 2010-2020 yılları arasındaki evlilik sayıları üzerine bir uygulama yapılmıştır. Bu çalışmada verilerin analizi için R programı kullanılmıştır.

2. LİTERATÜR ÖZETİ

2.1. Genelleştirilmiş Doğrusal Modeller ile İlgili Yapılan Çalışmalar

Çodur ve diğerleri (2013) tarafından gerçekleştirilen "Genelleştirilmiş Lineer Regresyon ile Erzurum Kuzey Çevre Yolu Kaza Tahmin Modeli" adlı çalışmada 2005-2010 yılları arasında Erzurum Kuzey Çevre Yolunda gerçekleşen kazalara ilişkin elde edilen istatistikler değerlendirilmiştir. Erzurum Kuzey Çevre Yolunda gerçekleşen kazaların sebeplerinin değerlendirilebilmesi amacı ile genelleştirilmiş doğrusal regresyon modeller oluşturulmuştur. Yapılan analiz ile düşey kurp sayısı, yol kesim kilometresi ve yıllık ortalama günlük trafiğin artması neticesinde bu bölgede kaza sayıları ile kaza sayısında artış arasındaki ilişki anlamlı bulunmuştur.

Karadağ (2014) tarafından gerçekleştirilen "Genelleştirilmiş Doğrusal Modeller için Sınırlı Dalgalanmalı Kredibilite Yaklaşımı" adlı çalışmada hayat dışı sigorta ürünlerinin fiyatlarına etken değişkenlerin değerlendirilebilmesi amacıyla genelleştirilmiş doğrusal modeller ve kredibilite kuramını kullanmıştır. Elde edilen verilerin dağılımları genelleştirilmiş doğrusal modellerle değerlendirilmiş ve Tam Kredibilite Standartı kapsamında genelleştirilmiş doğrusal modeller ve Sınırlı Dalgalanmalı Kredibilite arasında anlamlı bir ilişki tespit edilmiştir. Çalışma kapsamında risk sınıflarının güvenilirliğini tespit edebilmek amacıyla genelleştirilmiş doğrusal modeller karşılaştırma parametresi olarak kullanılmıştır. Çalışma kapsamında rezerv riskleri değerlendirilmesi bağlamında hasar rezerv tespitine ilişkin dağılımlar değerlendirilmiştir. Bununla birlikte kârlılık için pearson artıkları tekrar düzenlenmiş ve bootstrap yöntemi oluşturulmuştur.

Arslan ve Altaş (2019) tarafından gerçekleştirilen "Genelleştirilmiş Doğrusal Modeller İle Sigorta Şirketlerinde Hasar Rezervinin Tahmini" adlı çalışmada hasar rezervlerinin tespit edilmesinde genelleştirilmiş doğrusal modellerin kullanımı değerlendirilmektedir. Gerçek hayattaki veriler kullanılarak modellemelerin performansları analiz edilmiştir.

2.2. Poisson Regresyon ile İlgili Yapılan Çalışmalar

Sezgin vd. (2004) tarafından gerçekleştirilen "Poisson Regresyon Modelinde Aşırı Yayılım Durumu ve Negatif Binomial Regresyon Analizinin Türkiye Grev Sayıları Üzerine Bir Uygulaması" adlı çalışmada 1964-2000 yılları arasında Türkiye'de gerçekleşen grev sayılarına ilişkin etken unsurların değerlendirilebilmesi amacıyla poisson regresyon modeli kullanılmıştır. Değerlendirilen verilerin fazla dağılımlı olması nedeniyle tahminlerin kuvvetlendirilmesi amacıyla negatif binom regresyon modeli kullanılmıştır.

Yeşilyurt (2005) tarafından gerçekleştirilen "Poisson regresyon modeli ve Türkiye'deki boşanma istatistiklerine uygulanması" adlı çalışmada 1989-1998 arasında Türkiye'de boşanma verilerinin değerlendirilebilmesi için poisson regresyon modeli kullanılmıştır. 1989-1998 yılları arasına gerçekleşen boşanma verilerine ilişkin kriter tahminleri, uyum testleri ve kalıntıların hesaplanması için S-PLUS 2000 programını kullanmış ve verileri Poisson regresyon modellemesi ile modellemiştir.
Mouatassim ve Ezzahid (2012) tarafından gerçekleştirilen "Poisson regression and Zero-inflated Poisson regression: application to private health insurance data" adlı çalışmada özel sağlık sigortası sonuçlarını değerlendirebilmek amacıyla yaş, medeni durum ve cinsiyet gibi bağımsız değişken olarak belirlenmiş, çok sıfırlı Poisson ve Poisson regresyon modelleri kullanılmıştır.

Yeşilova vd. (2017) tarafından yapılan "Van Gölü Sahil Şeridindeki Zooplankton Populasyon Yoğunluğu Değişiminin Doğrusal Olmayan Regresyon Yöntemleri Kullanılarak İncelenmesi" adlı çalışmada Van Gölü'nde yaşayan zooplanktonların sayımlarının modellenmesi amacıyla poisson regresyonları kullanılmış ve aşırı dağılımlı verilerin değerlendirilebilmesi için negatif binomial regresyon kullanılmıştır. Negatif binomial regresyon ile değerlendirilen aşırı dağılımlı verilerde zooplankton sucul böcek sayımları ile bağımsız değişkenler arasındaki ilişki anlamlı bulunmuştur.
Famoye ve Singh (2006) tarafından gerçekleştirilen "Zero-Inflated Generalized Poisson Regression Model with an Application to Domestic Violence Data" adlı çalışmada aile içi şiddete ilişkin elde edilen dağınık veri setlerini modellemek amacı ile genelleştirilmiş poisson regresyon modeli kullanılmış ve verilerin değerlendirilmesi amacıyla uyum testi yapılmıştır.

3. GENELLEŞTİRİLMİŞ DOĞRUSAL MODELLER

3.1. Genelleştirilmiş Doğrusal Modellerin Biçimleri

Genelleştirilmiş doğrusal modeller, bağımlı değişken ve doğrusal model arasındaki ilişki kapsamında ölçülen değerlerin genelleştirilmesi amacıyla kullanılmaktadır. Genelleştirilmiş doğrusal modellerde, bağımlı değişkenin dağılımının üstel bir aile dağılımına sahip olduğu kabul edilir. $\mathrm{E}(\varepsilon_i) = 0$ olmak üzere model;

$$y_i = x_i'\beta + \varepsilon_i \quad i = 1,2,\dots,n$$

şeklinde yazılabilir. Burada $x_i = [1, x_{i1}, x_{i2}, \dots, x_{ip}]$, $(p+1) \times n$ boyutlu *i.* kümeye ilişkin satır vektörünü $\beta = [\beta_0, \beta_1, \dots, \beta_p]$, $1 \times (p+1)$ boyutlu bilinmeyen parametreler ve *Y* ise (n×1) boyutlu bağımlı değişken vektörü olmaktadır (Arıcan, 2010).

Genelleştirilmiş doğrusal modellerin üç bölümü vardır:

1) Rasgele bileşen (random component), veri setindeki bağımsız değişkenlerin birbirinden bağımsız olduğu ve üstel dağılım ailesinin her bir dağılımından geldiği kabul edilir. Sonuç olarak, sürekli rastgele değişken Y için olasılık yoğunluk fonksiyonu veya Y rastgele değişkeni için olasılık fonksiyonu,

$$f(y_i; n_i, \phi) = exp\left(\frac{y_i n_i - b(n_i)}{a(\phi)} + c(y_i, \phi)\right) \tag{3.1}$$

biçiminde verilir. Bu kısımda;

- *a*(.), *b*(.) ve *c* (.) üstel olan bir aileden bir diğerine değişen fonksiyonlardır.
- ϕ, yayılım parametresidir.
- n_i, doğrusal önkestiricidir.

Genelleştirilmiş doğrusal modellerde olabilirlik fonksiyonu,

$$L(\beta) = \prod_{i=1}^{n} \left(\frac{y_i n_i - b(n_i)}{a(\phi)} + c(y_i, \phi) \right) \tag{3.2}$$

şeklinde gösterilirken, n tane birim içeren örnek fonksiyonlar için $\ell = \ell(y_i; n_i, \phi) = \ell n L(y_i, n_i, \phi)$ olmak üzere log-olabilirlik fonksiyonu ise,

$$\ell(\beta) = \sum_{i=1}^{n} \ell_i(\beta) = \sum_{i=1}^{n} \left(\frac{y_i n_i - b(n_i)}{a(\phi)} + c(y_i, \phi) \right) \quad i = 1,2, \dots, n \tag{3.3}$$

biçiminde ifade edilir.

Y üstel bir aile dağılımına sahip olduğundan, ortalama, n_i'ye göre ℓ 'nin ilk türevi alınarak belirlenir.

$$\frac{\partial \ell}{\partial \eta_i} = \frac{y_i - b(\eta_i)}{a(\phi)} \tag{3.4}$$

ve ikinci türevinin alınmasıyla da varyans,

$$\frac{\partial^2 \ell}{\partial {\eta_i}^2} = \frac{-b^{"}(\eta_i)}{a(\phi)} \tag{3.5}$$

Sonucuna ulaşılır.

Denklem (3.4)'ü sıfıra eşitleyerek, olabilirlik fonksiyonunun eşitliğinin $a(\phi)$'dan bağımsız bir denklem olduğu saptanır. $\eta = x'\beta$'deki η ve β'nın tahmin uygunluğu ile ilgilenildiğinden dolayı $a(\phi) = 1$ olarak kabul edilir (Rao, 1999). Fisher skorlama ile;

$$E\left(\frac{\partial \ell}{\partial \eta_i}\right) = 0 \tag{3.6}$$

$$-E\left(\frac{\partial^2 \ell}{\partial {\eta_i}^2}\right) = E\left(\frac{\partial \ell}{\partial \eta_i}\right)^2 \tag{3.7}$$

ve Denklem (3.4) ile Denklem (3.6) bir araya getirilirse,

$$E\left(\frac{Y_i - b(\eta_i)}{a(\phi)}\right) = 0$$

$$E(Y_i) - b'(\eta_i) = 0 \tag{3.8}$$

elde edilir. Denklem (3.8)'den,

$$E(Y_i) = b'(\eta_i) = \mu_i$$

Tahmin edilen değerin genel şekli, genelleştirilmiş doğrusal modeller için bulunur. Benzer olarak Denklem (3.5) ve Denklem (3.7) birleştirilirse,

$$\frac{b^{"}(\eta_i)}{a(\phi)} = E\left(\frac{Y_i - b'(\eta_i)}{a(\phi)}\right)^2$$

$$\frac{b^{"}(\eta_i)}{a(\phi)}) \frac{V(Y_i)}{a(\phi)^2}$$

$$V(Y_i) = a(\phi)b^{"}(\eta_i) = \frac{\partial \mathrm{E}(Y_i)}{\partial \eta_i} = \frac{\partial \mu_i}{\partial \eta_i} \tag{3.9}$$

Söz konusu genelleştirilmiş doğrusal modellerin varyansan genel formuna ulaşılır. Varyans, Denklem (3.9) 'de öngörülen değerin bir fonksiyonudur, ancak doğrusal regresyon varsayımlarından biri olan sabit bir varyansa sahip olmadığı açıktır (Arıcan, 2010).

2) Sistematik bileşen (Systematic component), modelin ortalaması alınacak olan parametresinin hesaplamasında kullanılan doğrusal önkestirim (Linear Predictor), bağımsız değişkenlerin doğrusal kombinasyonu,

$$\eta_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} = \sum_{i=1}^{n} X_i \beta$$

şeklinde belirtilmektedir.

3) Bağ fonksiyonu (link function) g(.) , bağımlı değişkenin beklentisini rastgele ve sistematik bileşen arasındaki bağlantıyı belirtir. $\mu_i = E(Y_i)$ için doğrusal önkestirici,

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 X_{1i} + \beta_2 + X_{2i} + \cdots + \beta_p X_{pi}$$

şeklinde belirtilir. Bundan dolayı bağ fonksiyonu *Y*'nin ortalaması ile η_i 'yi ilişkilendirir.

Genel anlamda η_i , μ_i 'nin bir fonksiyonudur:

$$g(\mu_i) = \eta_i$$

Bağ fonksiyonu tersinir olduğu için de,

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi})$$

Biçiminde gösterilebilir.

Doğrusal olmayan bir regresyon modeli ya da tahmini bir bağımlı değişkenin geçişi bağımlı değişkenler için genelleştirilmiş doğrusal model olarak ifade edilebilir. Bu ters bağ aynı zamanda $g^{-1}(.)$ ortalama fonksiyonu (mean function) olarak da ifade edilmektedir. Çizelge 3.1'de bağ fonksiyonları ve ters bağ fonksiyonları gösterilmektedir.

Çizelge 3.1. Bazı ortak bağ fonksiyonları ve ters fonksiyonları (De Jong ve Heller, 2008; Ölmez ve Ata, 2018)

Bağ	$\eta_i = g(\mu_i)$	$\mu_i = g^{-1}(\eta_i)$
Özdeş (Identity)	μ_i	η_i
Log	$\log_e \mu_i$	e^{η_i}
Ters (Inverse)	μ_i^{-1}	η_i^{-1}

Ters-kare (Inverse-square)	μ_i^{-2}	$\eta_i^{-1/2}$
Karekök (Square-root)	$\sqrt{\mu_i}$	η_i^2
Logit	$\log_e \frac{\mu_i}{1-\mu_i}$	$\frac{1}{1+e^{-n_i}}$
Probit	$\Phi^{-1}(\mu_i)$	$\Phi(\eta_i)$
Log-Log	$-\log_e[-\log_e(\mu_i)]$	$\exp[-\exp(\eta_i)]$
Tamamlayıcı log-log (Complementary log-log)	$\log_e[-\log_e(1-\mu_i)]$	$1-\exp[-\exp(\eta_i)]$

μ_i, bağımlı değişkenin tahmin edilen değeri olduğunda, η_i doğrusal göstergedir ve Φ(.), normal normal dağılımın kümülatif dağılım işlevidir. Çizelge 3.2, her olasılık ailesindeki bağımlı değişkenlerin kayma döngülerinin yanı sıra, kanonik bağlantı işlevleri olarak bilinen işlevlerin her bir aile ile ilişkisini göstermektedir.

Çizelge 3.2. Kanonik bağ, bağımlı değişken aralığı ve üstel aileler için koşullu varyans fonksiyonu (Erdemir ve Sucu, 2016).

Aile	**Kanonik Bağ**	**Y_i'nin Aralığı**	**$V(\gamma_i\|\mu_i)$**
Normal	Özdeş	(-∞,+∞)	1
Binomal	Logit	$\frac{0,1,\ldots,n_i}{n_i}$	$\frac{u_i(1-u_i)}{n_i}$
Poisson	Log	1,2,…	u_i
Gamma	Ters	(0,∞)	μ_i^2
Ters-Normal	Ters-Kare	(0,∞)	μ_i^3

Diğer ilişki fonksiyonları da kullanılsa da, genelleştirilmiş doğrusal modeller kanonik bağlantılar için daha basit hale getirilmiştir. Örneğin, Poisson dağılımı oranı parametresi $\mu > 0$ ile olasılık fonksiyonu,

$$f(y_i) = \frac{\mu_i^{y_i}}{y_i!} \qquad i = 1,2, \ldots$$

üstel bir ailedir.

Poisson rastgele değişkeninin ortalaması ve varyansı μ'ye eşittir. Sayım veri modellemesi için Poisson dağılımı kullanışlıdır. Poisson dağılımı, μ büyüdükçe daha simetrik hale gelir ve nihayetinde normal dağılıma yakınsar (Çelik ve Avcı, 2019).

Poisson dağılımının olasılık fonksiyonu üstel biçimde yazılırsa (Arslan ve Altaş, 2019):

$$f(y_i) = \exp\left(y_i \ell n \mu_i - \ell n(y_i!)\right) \tag{3.10}$$

elde edilir. Üstel ailenin genel biçiminin verildiği Denklem (3.1)'deki ifadede yer alan terimlerin Denklem (3.10)'daki karşılıkları,

$$\eta_i = \ell n \mu_i$$

$$\mu_i = e^{\eta_i} = b(\eta_i)$$

$$a(\phi) = 1$$

$$c(y_i, \phi) = -\ell n(y_i!)$$

şeklindedir.

3.2. Genelleştirilmiş Doğrusal Modeller ve Tahmin

Olasılık ilkesi, genelleştirilmiş doğrusal modellerde tahmin yapmak için kullanılır. Denklem (3.3), $y_i, \ldots, y_n$ gözlemlenen değerler için log-olabilirlik fonksiyonunu verir.

Log-olabilirlik fonksiyonunun türevi β 'ya göre alınır ve zincir yasası kullanılarak sıfıra eşitlenir.

$$\frac{\partial l_i(\beta)}{\partial \beta} = \frac{\partial l_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta} = 0 \tag{3.11}$$

elde edilir. Denklem (3.11)'daki türevler,

$$\frac{\partial l_i}{\partial \eta_i} = \frac{\partial}{\partial \eta_i}\left(\sum_{i=1}^{n}\left(\frac{y_i\,\eta_i - b(\eta_i)}{a(\emptyset)} + c(y_i, \emptyset)\right)\right) = \sum_{i=1}^{n}\left(\left(y_i - \frac{\partial b(\eta_i)}{\partial \eta_i}\right)\frac{1}{a(\emptyset)} + \frac{\partial c(y_i,\emptyset)}{\partial \eta_i}\right) =$$

$$\frac{1}{a(\emptyset)}\sum_{i=1}^{n}\left(y_i - \frac{\partial b(\eta_i)}{\partial \eta_i}\right) \tag{3.12}$$

Ve

$$\frac{\partial \eta_i}{\partial \beta} = \frac{\partial}{\partial \beta}(x_i\beta) = x_i \tag{3.13}$$

şeklinde bulunur. Denklem (3.12) ve Denklem (3.13)'ü Denklem (3.10)'de yerine yazılırsa:

$$\frac{\partial \ell}{\partial \beta} = \frac{\partial \ell}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta} = \frac{1}{a(\phi)}\sum_{i=1}^{n}\left(y_i\frac{\partial \mathrm{b}(n_i)}{\partial \eta_i}\right)x_i = 0 \tag{3.14}$$

elde edilir. Denklem (3.14)'ün β için çözülmesi gereklidir. Fakat $\ell(\beta)$, β'nın doğrusal şekli olmadığından dolayı doğrudan çözüme kavuşturulamaz ve Newton-Rahpson metodu gibi yinelemeli (iteratif) yöntemler yardımı sayesinde çözülebilmektedir. Maksimum $\ell(\beta)$, fonksiyonu bulmak $\frac{\partial \ell(\beta)}{\partial \beta}$ ifadesi Taylor serisi ile β^0 etrafında açılırsa (Koç ve Cengiz, 2012):

$$\frac{\partial l(\beta)}{\partial \beta} = l'(\beta) = l'(\beta^0) + \frac{\partial^2 l(\beta)}{\partial \beta \partial \beta'}(\beta - \beta^0) \tag{3.15}$$

elde edilir. Denklem (3.15)'deki ifade sıfıra eşitlenerek, kök için çözüm:

$$l'(\beta^0) + \frac{\partial^2 l(\beta)}{\partial\beta\partial\beta'}(\beta - \beta^0) = 0$$

$$\beta = \beta^0 - \left[\frac{\partial^2 l(\beta)}{\partial\beta\partial\ '}\right]^{-1} l'(\beta^0) \tag{3.16}$$

olması üzerine kök için bir çözüm bulunabilir. Kökün tahminini geliştirmek için *Denklem* (3.15) tekrar etmekte ve böylelikle gerçek köke yakınsamada,

$$\beta^{(k+1)} = \beta^{(k)} - \left[\frac{\partial^2 l(\beta)}{\partial\beta\partial\beta'}\right]^{-1} l'(\beta^k) \tag{3.17}$$

eşitliği kullanılabilir. Genelleştirilmiş doğrusal modeller durumunda, Newton-Raphson yaklaşımındaki $\frac{\partial\ell}{\partial\beta}$ ifadesi β etrafında,

$$\beta_{EÇO} = \beta - \left[\frac{\partial^2 l}{\partial\beta\partial\beta'}\right]^{-1} \frac{\partial l}{\partial\beta} \tag{3.18}$$

biçiminde elde edilir. Denklem (3.11)'de verilen zincir kuralı genişletilirse (Toutounian ve Nasabzadeh, 2014):

$$\frac{\partial l(\beta)}{\partial\beta} = \sum_{i=1}^{n} \frac{\partial l_i}{\partial\eta_i}\frac{\partial\eta_i}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_i}\frac{\partial\eta_i}{\partial\beta} = \sum_{i=1}^{n} x_i \frac{\partial\mu_i}{\partial\eta_i}\frac{(y_i-\mu_i)}{V(Y_i)} \tag{3.19}$$

$$\frac{\partial l(\beta)}{\partial\beta'} = \sum_{i=1}^{n} \frac{\partial\mu_i}{\partial\eta_i}\frac{(y_i-\mu_i)}{V(Y_i)} x_i' \tag{3.20}$$

bulunur. β bileşenlerini bulmak için olabilirlik denklemi ise,

$$\sum_{i=1}^{n} x_i \frac{\partial\mu_i}{\partial\eta_i}\frac{(y_i-\mu_i)}{V(Y_i)} = 0 \tag{3.21}$$

biçimindedir.

Denklem (3.17)'deki $\frac{\partial^2 \ell}{\partial \beta \partial \beta'}$ ifadesi, Denklem (3.7)'de verilen Fisher Skorlama ile,

$$E\left[\frac{\partial^2 \ell}{\partial \beta \partial \beta'}\right] = E\left[\frac{\partial \ell}{\partial \beta}\frac{\partial \ell}{\partial \beta'}\right] = E\left[\sum_{i=1}^{n} x_i \frac{\partial \mu_i}{\partial \eta_i}\frac{(y_i-\mu_i)}{V(Y_i)}\frac{(y_i-\mu_i)}{V(Y_i)}\frac{\partial \mu_i}{\partial \eta_i}x_i'\right]$$

$$= -\sum_{i=1}^{n} x_i x_i' \left[\frac{\partial \mu_i}{\partial \eta_i}\right]^2 \frac{1}{V(Y_i)} = -\phi \qquad (3.22)$$

olarak hesaplanır (Arıcan,2010). $W = diag\left[\frac{\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2}{V(Y_i)}\right]$ olmakta beraber, $\phi = XW\,X$ bilgi matrisi olarak adlandırılmaktadır. Denklem (3.19) ve Denklem (3.22), Denklem (3.18)'de yerine yazılıp her iki tarafı $\mathbf{XWX}$ ile çarpılırsa,

$$\beta^{(k+1)}(X'WX) = \beta^{(k)}(X'WX) + \frac{\partial l}{\partial \beta}|_{\beta=\beta^{(k)}} \qquad (3.23)$$

biçiminde olur.

Denklem (3.23)'ün sağ tarafı Denklem (3.19) ve Denklem (3.22)'nin birleşimi olarak yazılırsa,

$$\left[\sum_{i=1}^{n} x_i x_i' \left[\frac{\partial \mu_i}{\partial \eta_i}\right]^2 \frac{1}{V(Y_i)}\beta^{(k)}\right] + \sum_{i=1}^{n} x_i \frac{\partial \mu_i}{\partial \eta_i}\frac{(y_i-\mu_i)}{V(Y_i)} \qquad (3.24)$$

elde edilir ve Denklem (3.24) aşağıdaki gibi yazıldığında,

$$X'Wz \qquad (3.25)$$

olur. Denklem (3.25)'deki z değeri,

$$\sum_{i=1}^{n} x_i \beta^{(k)} + (y_i-\mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right) = \eta_i + (y_i-\mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right) \qquad (3.26)$$

olur. Denklem (3.23)'e de Denklem (3.26)'daki değer yazılırsa genelleştirilmiş doğrusal model için en çok olabilirlik iteratif tahmin edicisi,

$$\beta'^{(k+1)} = (X'W^{(R)}X)^{-1}X'W^{(R)}\left[\eta_i + (y_i - \mu_i)\frac{\partial \eta_i}{\partial \mu_i}\right]^{(k)} \tag{3.27}$$

Biçimine girer (Rao, 1999).

$$\beta'^{(k+1)} = (X'W^{(k)}X)^{-1}X'W^{(k)} \tag{3.28}$$

şeklinde gösterilir.

3.3. Genelleştirilmiş Doğrusal Modeller ve Sonuçlar

İstatistiksel çıkarımın iki temel yöntemi, güven aralıkları ve hipotez testleridir. Kalıntılar, varsayımların gerçekleştirilmesi açısından modelin yeterliliği ve seçilen ilişki fonksiyonunun uygunluğu hakkında ayrıntılar sağlar.

3.3.1. β adına hipotez testi ve güven aralıkları

Genelleştirilmiş doğrusal modellerde, Olabilirlik Oranı (Likelihood Ratio) ve Wald istatistik testi en temel çıkarım yöntemlerinden ikisidir.

Olabilirlik oranı testi, H_0 'ın H_1 'e karşı testi için kullanılan genel bir yöntemdir. $M_2 \subseteq M_1$ parametre kümelerine karşılık gelen olabilirlik oranı test istatistiği,

$$\Lambda = \frac{L(M_2)}{L(M_1)} \tag{3.29}$$

olarak bulunur. n $\rightarrow\infty$ yaklaştığında ise,

$$G^2 = -2ln\Lambda \tag{3.30}$$

olur. Denklem (3.30)'dan da doymuş model için log olabilirlik fonksiyonu l_s olduğunda,

$$G^2(M_2|M_1) = -2\big(l(M_2) - l(M_1)\big) = -2(l(M_2) - l_s) - [-2(l(M_1) - l_s] = G^2(M_2) - G^2(M_1) \quad (3.31)$$

biçiminde gösterilir.

Bu iki modelin karşılaştırılması için $G^2(M_2|M_1)$ istatistiği iki modelin uyum iyiliği test istatistiklerinin farkıyla aynıdır (Rao, 1999).

Regresyon modeli için hipotezler,

$$\mathrm{H}_0: \beta_1 = \cdots = \beta_\mathrm{p} = 0, \quad (3.32)$$

H_1: En az biri β,

$$j = 1,2,\ldots,p \quad (3.33)$$

şeklinde gösterilir ve χ^2_{sd} istatistik değeri ile karşılaştırılır. Serbestlik derecesi ise parametre sayılarının arasındaki değişkene eşittir.

Rastgele seçilmiş bir gözlemin regresyon katsayısı β_j için önem testleri,

$$H_0: \beta_j = 0 \quad (3.34)$$

$$H_1: \beta_j \neq 0 \quad (3.35)$$

biçiminde kurulur ve Denklem (3.30)'daki olabilirlik oranı kullanılarak test edilebilir.

Olabilirlik oranı testi, t-testine karşılık gelen Wald testi ile eşdeğerdir ve $\hat{\beta} \sim N(\beta, (X'WX)^{-1})$. H_0 hipotezinin asimptotik testi ise Wald testidir. Test istatistiği

$$Z_j = \frac{\widehat{\beta_j}}{s_{\beta_j}} \quad (3.36)$$

biçimindedir ve β_j için $100\times(1-\alpha)$ güven aralığı, $\hat{\beta}_j \pm Z_{\alpha/2} S_{\hat{\beta}_j}$ ile bulunur.

Wald testlerinin ve olabilirlik oranı testlerinin hipotezlerine bakıldığında asimptotik olarak eşdeğer olduğu görülür, aynı zamanda H_0 hipotezi doğru olduğunda daha küçük örnekler için benzerlik gösterir.

3.3.2. Sapma ve uyum iyiliği

Normal doğrusal regresyonda, sapma (deviance) kalıntıları standart kalıntılara yakındır. Sonuç olarak, genelleştirilmiş doğrusal modeller için, doğrusal regresyonda beklenen değerlere karşı kalıntıların grafiği sıklıkla kullanılır. Normal kalıntılar yerine sapma kalıntıları kullanılır çünkü standart kalıntılar basitleştirilmiş doğrusal modeller için uygun değildir.

Genelleştirilmiş doğrusal modelde X doymuş modelin matrisi gözlem sayısı olarak parametrelerin sayılarıyla aynı sayıyı içerdiğinde, $\hat{\eta}$ doymuş modeldeki $\hat{\mu}_i = y_i (i = 1,2,\dots,n)$ tahminlerine sahip olan η 'nın bir tahmini olarak tanımlanmaktadır. Doymuş model olmayan her alt model X_i için,

$$G^2(X_i|X) = 2\sum_{i=1}^{n} y_i(\hat{\eta}_i - \eta_i) - b(\hat{\eta}_i) + b(\eta_i) = D(y;\hat{\mu}_i) \tag{3.37}$$

olur. Burada yazılana göre doymuş model tarafından elde edilen en iyi uyuma karşı X_i modelinin uyum iyiliğindeki kaybı için bir ölçüm niteliği taşır. $D(y;\hat{\mu}_i)$ istatistiği X_i modelinin sapması adını alır:

$$G^2(X_1|X_2) = G^2(X_1|X) - G^2(X_2|X) = D(y;\hat{\mu}_1) - D(y;\hat{\mu}_2) \tag{3.38}$$

Yani test istatistiği, X_2 geniş modelinin X_1 modeline karşın iki modelin uyum iyiliği istatistiğinin farkına eşittir (Rao,1999).

Doymamış ve doymuş modellerin parametre numaraları arasındaki uyumsuzluk, serbestlik derecelerine eşittir.

3.4. Model Tercihi

Model seçimi, sonuçlar için en iyi eşleşmeyi sağlayan alt (basit) bağımsız değişken kümesini bulmak için kullanılır. Genel model *p* parametresine sahip olduğunda, indirgenmiş model bu fonksiyon için daha az parametreye sahiptir. Varyans açısından indirgenmiş form ise daha yüksek bir değere sahiptir. Buna göre form karşılaştırmaları yapılabilir ve modelde kullanılacak bileşenin üretilen model üzerinde önemli bir etkisinin olup olmadığı belirlenebilir.

$$\eta = X\beta$$
$$\eta = X_1\beta_1 + X_2\beta_2 \tag{3.39}$$

modeli ele alınacak olursa, genel model *p* tane değişkene, β_1 ise p-r tane değişkene ve β_2 ise *r* tane değişkene sahip olur. Model testinde,

$$H_0: \beta_2 = 0 \tag{3.40}$$
$$H_1: \beta_2 \neq 0$$

hipotezleri test edilmek istenir ve H_0 hipotezi kabul edilirse indirgenmiş model $\eta_{\text{indirgenmiş}} = X_1\beta_1$ olur. Bu hipotezler sapma istatistiği,

$$D(\beta_2|\beta_1) = D(\beta_2) - D(\beta_1) \tag{3.41}$$

kullanılarak test edilir. Denklem (3.41)'de verilen indirgenmiş modelin sapma değerinden genel modelin sapma değerinin çıkarılması, test edilen modelin sapma değerini verir. Sonuç olarak, hipotez testi, bu varyans istatistiğinin değeri, r serbestlik derecesinin ki-kare tablosunun değeri ile karşılaştırılarak gerçekleştirilir. Eğer $D(\beta_2|\beta_1) \geq X^2_{\alpha,r}$ ise, α anlamlılık düzeyinde H_0 reddedilir ve test edilen modelin sahip olduğu değişkenlerin oluşturduğu bir alt kümenin genel modele önemli bir katkısının olduğu kararına varılır (Arıcan, 2010).

3.5. Artıklar ve Analizleri

Kalıntıları incelemek, örnek uyumsuzluklarını araştırmak ve regresyon analizinde aykırı değerleri belirlemek için en güçlü yaklaşımlardan biridir. "Artıklar" terimi, hataların mevcut veya gözlemlenen değerlerini ifade eder. Sonuç olarak, hatalar için sonuçlardaki çarpıklıkları test etmek için kalıntılar kullanılacaktır (Köleoğlu, 2006).

Genel anlamda artıklar,

$$r_i = y_i - \hat{\mu}_i \tag{3.42}$$

şeklide ifade edilmektedir. Doğrusal regresyon modelinde hata, normal dağılımlıdır ve homoskedastik $(Y-\mu) \sim N[0, \sigma^2]$ olarak tanımlanır. Fakat, sayma verileri için ise $(Y-\mu)$, heteroskedastik yani varyansı ortalamadan büyük olması durumudur ve asimetriktir. Sonuçların sayılmasında sıfır ortalamalı, sabit varyanslı ve simetrik dağılımlı kalıntı yoktur (Deniz, 2005). Arıcan (2010) bunun yerine önyargı kalıntısının ve Pearson kalıntısının kullanılmasını önermektedir.

3.5.1.Sapma artığı

Sapma artığı (Deviance Residual), her bir gözlemin sebep olduğu $-2\ell nL$ değerinin ölçüsüdür. Yani, her bir gözlemin modele anlamlı katkısının bir ölçüsüdür. i -inci gözleme ait sapma artığı ise,

$$d_i = sing(y_i - \hat{\mu}_i)\sqrt{2\left[y_i ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i)\right]} \tag{3.43}$$

ile verilmektedir. Artığın sahip olduğu işaret, $(y_i - \hat{\mu}_i)$ 'nin işareti ile aynıdır. Sapma artıkları Denklem (3.29) ile verilen G^2 istatistiği kullanılarak test edilmektedir (Köleoğlu, 2006).

3.5.2.Pearson artığı

Pearson artıkları (Pearson Residual) genelleştirilmiş doğrusal modeller için her gözlemde model uyumunu kontrol etmek için kullanılabilir. Pearson artıkları için denklem,

$$e_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\phi \hat{\mu}_i}} \qquad i = 1,2,\dots,n \tag{3.44}$$

olarak ifade edilmektedir. Pearson artıkları χ^2 testi ile test edilmektedir (Köleoğlu, 2006).

4. POISSON REGRESYON

Bağımlı değişkenin sayılabilir olduğu durumlarda Poisson regresyon modeli kullanılabilmektedir. Poisson regresyonu, amacın analiz edilen zaman aralığında meydana gelen olaylar ile tanımlanan bağımsız değişkenler arasında bir korelasyon bulmak olduğu problemleri çözmek için kullanılabilmektedir. Model süreksiz ve negatif olmayan sayılabilir sonuçlar içerdiğinden, söz konusu tahmin edilen sayıların logaritması genel olarak bağımsız değişkenlerin bir fonksiyonu olarak kabul edilir.

4.1. Poisson Regresyon Modeli

Regresyon modeli Poisson, bağımsız değişken $X' = [1, X_{i1}, X_{i2}, \dots, X_{ip}]_{nx(p+1)}$ ile parametre μ arasındaki bağlantıyı ileten Poisson dağılımı tarafından sağlanır. Poisson'un regresyon modeli, Poisson dağılımının ortalaması ile tanımlanır ve formül şu şekildedir:

$$f(y_i|x_i) = \begin{cases} \dfrac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, & y_i = 0, 1, 2 \dots \end{cases} \tag{4.1}$$

biçiminde gösterilmektedir (Cameron ve Trivedi, 2013). Bu kısımda μ_i,

$$E(y_i|x_i) = \mu(x_i) = c_i f(x_i, \beta) = \mu_i \qquad i = 1,2, \dots, n \tag{4.2}$$

biçimindedir.

Maksimum olabilirlik yaklaşımı, Poisson regresyon modelindeki parametreleri tahmin etmek için yaygın olarak kullanılır. Bağımlı değişken yi'nin dağılımına dayalı olarak β tahmin edicilerini tahmin etmenin en yaygın yöntemi, en olasılık yöntemidir. Bu yaklaşımda β tahminleri, log olasılık fonksiyonu maksimize edilecek şekilde seçilebilir.

4.2. Çok Sıfırlı Poisson Regresyon Modeli

Poisson regresyon modelinde, aşırı dağılımın yalnızca sayılardaki sıfır değerlerdeki yoğunluktan kaynaklandığı, çok miktarda sıfır ve genelleştirilmiş poisson regresyonu içeren negatif binom regresyon modelinde bir varsayım olduğu, çok miktarda sıfır içeren bir varsayım vardır. Sonuç olarak, tüm çoklu sıfır regresyon modelleri aynı ortalamaya sahipken, varyansları değişir. Kesinlikle sıfır olan sayım verilerini incelemek için çok sayıda sıfır içeren Poisson regresyon modelleri oluşturulmuştur. Çok sayıda sıfır içeren Poisson regresyon modeli:

$$P(Y_i = 0) = \pi_i + (1 - \pi_i)e^{\mu_i} y_i = 0 \tag{4.3}$$

$$P(Y_i = y_i) = (1 - \pi_i)\frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} y_i > 0 \tag{4.4}$$

biçimindedir (Cameron ve Trivedi, 2013; Agresti, 2015).

Aşağıda birçok sıfır içeren poisson regresyon modelinin tahmin edilen değeri ve varyansı verilmiştir:

$$E(Y_i) = (1 - \pi_i)\mu_i \tag{4.5}$$

$$V(Y_i) = \mu_i(1 - \pi_i)(1 + \mu_i\pi_i) \tag{4.6}$$

π_i = 0, çok sayıda sıfır içeren bir poisson regresyonunda gerçekleştiğinde, poisson regresyona girer ve $\pi_i > 0$ olduğunda, regresyon aşırı yayılım moduna geçer.

4.3. Genelleştirilmiş Çok Sıfırlı Poisson Regresyon Modeli

Consul ve Famoye (1992), sayım verilerindeki sıfırların çokluğunu araştırmak ve sıfırların çokluğunu içeren negatif iki terimli regresyona alternatif olarak aşırı yayılma sorununu önlemek için bu modeli önermişlerdir. Önerilen modelde genelleştirilmiş bir poisson dağılımı kullanılmıştır.

Y_i , i = 1,2, … , n bağımlı değişkenlerin ölçümünü ifade etmektedir. Bu bağlamda Y_i olasılık fonksiyonu şu şekilde ifade edilmektedir:

$$f(Y_i > y_i) = \left(\frac{\mu_i}{1+\theta\mu_i}\right)^{y_i} \left(\frac{1+\theta y_i}{y_i!}\right)^{y_i-1} exp\left\{\frac{-\mu_i(1+\theta y_i)}{1+\theta\mu_i}\right\}. y_i > 0 \quad (4.7)$$

Genelleştirilmiş poisson dağılımı kapsamında ortalama ve varyans sırasıyla şu şekilde gösterilmektedir:

$$E(Y_i) = \mu_i \quad (4.8)$$

$$V(Y_i) = \mu_i(1 + 0\mu_i)^2 \quad (4.9)$$

şeklinde gösterilir. Bu modelin θ parametresi, çok fazla sıfır içeren poisson regresyon modelinin aksine, aşırı veya zayıf bir yayılma koşuluna sahiptir. Bu model, $\theta > 0$ olduğunda aşırı dağılma durumuna ve $\theta < 0$ olduğunda zayıf bir dağılma durumuna yol açar (Famoye ve Singh, 2006). Genelleştirilmiş poisson regresyon modeli, beklenenden daha fazla sıfır gözlem olduğunda pek uygun değildir (Cameron ve Trivedi, 2013; Agresti, 2015).

4.4. En Çok Olabilirlik Tahmin Edici

$\hat{\beta}$ tahminleri, log-olabilirlik fonksiyonunu optimize etmek için maksimum olasılık yaklaşımında seçilir. Poisson regresyon modelinde ortalama parametre şöyledir:

$$E(y_i|x_i) = \mu_i = \exp(x_i'\beta) \quad (4.10)$$

biçimindedir.

Bu denklem olasılık fonksiyonunda yerine yerleştirilirse:

$$f(y_i|x_i) = \left\{\frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} = 0,1,2, \ldots \right. \quad (4.11)$$

Bağımsız gözlemler için maksimum olabilirlik işlevi:

$$L(y,\beta) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} \frac{e^{-\mu_i}\mu_i{}^{y_i}}{y_i!} = \frac{(\prod_{i=1}^{n}\mu_i{}^{y_i})e^{\sum_{i=1}^{n}\mu_i}}{\prod_{i=1}^{n}(y_i!)} \quad (4.12)$$

şeklinde yazılabilir.

Log-olabilirlik fonksiyonu:

$$l(\beta) = \sum_{i=1}^{n} y_i \ln(\mu_i) - \sum_{i=1}^{n}(\mu_i) - \sum_{i=1}^{n} ln(y_i!) \quad (4.13)$$

biçimindedir. $\mu_i = \exp(x_i'\beta)$ Denklemde yerine yazılırsa:

$$l(\beta) = \sum_{i=1}^{n} y_i\, x_i'\beta - \sum_{i=1}^{n} e^{x_i'\beta} - \sum_{i=1}^{n} \ln(y_i!) \quad (4.14)$$

sonucuna ulaşılır.

Poisson regresyon modeli kapsamında yer alan parametrelerden biri olan en çok olabilirlik tahmin edici için log-olabilirlik fonksiyonunun β'ya göre birinci dereceden kısmi türevi alınmakla birlikte sıfıra eşitlenmektedir.

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^{n} y_i x_i' - \sum_{i=1}^{n} x_i e^{-x_i'\beta} = \sum_{i=1}^{n} x_i(y_i - \mu_i) = 0 \quad (4.15)$$

$$\sum_{i=1}^{n} x_i' = (y_i - \exp(x_i'\beta) = 0 \quad (4.16)$$

$$\sum_{i=1}^{n} x_i'(y_i - \mu_i) = 0 \quad (4.17)$$

biçiminde yazılır (Cameron ve Trivedi, 2013).

Bu denklemde görüldüğü gibi, açıklayıcı değişkenlerin sabit bir ifadeye sahip olması durumunda ($y_i - \mu_i$) artıkların toplamı sıfır olacaktır. Maksimum olabilirlik yaklaşımı, üstel ortalama parametresi nedeniyle log-olabilirlik fonksiyonları doğrusal olmadığından tek adımlı bir çözüm elde edemez.

Bu durumda $\hat{\beta}_p$ değerini ölçmenin ampirik bir yolu yoktur. Çözüm, aşamalı ağırlıklı en küçük kareler (IWLS) işlemi kullanılarak aranabilir. İteratif ağırlıklandırılmış en küçük kareler yöntemi algoritması:

$$\beta_{ML} = (x'\widehat{W}\hat{z})^{-1}x'Wz \tag{4.18}$$

şeklindedir. Burada $\widehat{W} = \text{diag}[\hat{\mu_i}]$ ve $\hat{z}$,

$$\hat{z}_l = \log(\hat{\mu}_i) + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \tag{4.19}$$

İkinci türevler matrisinin tersi olan asimptotik biçimde normal dağılımlı kovarayans matrisi en çok olabilirlik tahmin edici için şu şekilde gösterilmektedir:

$$Var(\hat{\beta}_{ML}) = [-E\left(\frac{\partial^2 l}{\partial\beta_j\partial\beta_k'}\right)]^{-1} = (x'\widehat{W}x)^{-1} \tag{4.20}$$

biçimindedir. Ayrıyeten MSE eşitliği

$$E(L_{ML}^2) = E(\hat{\beta}_{ML} - \beta)'(\hat{\beta}_{ML} - \beta) = tr\left[(x'\widehat{W}x)^{-1}\right] = \sum_{j=1}^{j}\frac{1}{\boldsymbol{\lambda}_j} \tag{4.21}$$

olarak tanımlanır. $\lambda_j, x'\widehat{W}x$, matrisin j.inci özdeğeridir. Açıklayıcı değişkenler güçlü bir şekilde ilişkilendirildiğinde, ağırlıklı çapraz-çarpım matrisi $x'\widehat{W}x$ zayıf bir şekilde koşullu hale gelir, bu da yüksek oranda varyans ve vektör en olasılık tahminiyle sonuçlanır. Kestirilen katsayıların ortalama vektörü çok uzun olduğu için bu durumda

kestirilen parametrelerin okunması zordur. Verilen veri ve modellere göre varyans değeri için:

$$\hat{\beta}_{PR} \sim N\left[\beta, V_{ML}[\beta_{PR}]\right]$$

$$V_{ML}\left[\hat{\beta}_{PR}\right] = \left(\sum_{i=1}^{n} \mu_i x_i x_i'\right)^{-1} \tag{4.22}$$

sonuçlarına ulaşılır (Cameron ve Trivedi, 2013; Agresti, 2015).

4.5. Poisson Regresyon Katsayıları ve Değerlendirilmesi

Bağımlı değişken y_i değerleri için tahminler, Poisson regresyon yönteminden elde edilen katsayılar kullanılarak üretilebilir. i. bağımsız değişkenlerindeki bir birim vektörün etkisi, doğrusal modellerde $E(Y|x_i) = x_i'\beta$ ve β katsayıları olarak temsil edilir. i. bağımsız değişkeni anlamı için üstel koşullu ortalama şudur:

$$\frac{\partial E(y|x_i)}{\partial x_i} = \beta_i \exp(x_i'\beta) \tag{4.23}$$

biçiminde ifade edilir. Burada $\hat{\beta}_i$ için x_i'deki bir birimlik değişime bağlı olan $E(Y|x_i)$ 'deki nispi değişimi ölçtüğü yorumu yapılabilir.

4.6. Poisson Regresyonu ve Yanlı Tahmin Ediciler

4.6.1. Poisson modellerinde ridge tahmin edici

Månsson ve Shukur (2011), çoklu ilişkiler sorununa bir çözüm olarak frekans verileri için poisson ridge regresyon yaklaşımını önermiştir. Bu prosedürün çözümüne ulaşmak için maksimum olasılık stratejisinin ağırlıklı hata kareleri sayısını (AHKT) kabaca en aza indireceği bilgisi kullanılır. Sonuç olarak, kare sayısı açısından, $\hat{\beta}_{MLE}$ ağırlıklı hata optimal bir tahminci olarak kabul edilebilir. Bu yöntem aşağıdaki tahminciyi vermiştir:

$$\hat{\beta}_{PRR} = (X'\widehat{W}X + kl)^{-1}X'\widehat{W}X\hat{\beta}_{ML} \tag{4.24}$$

$$E(L^2_{PRR}) = E\big(\hat{\beta}_{PRR} - \beta\big)'\big(\hat{\beta}_{PRR} - \beta\big)$$
$$= E\left[\big(\hat{\beta}_{ML} - \beta\big)'Z'Z\big(\hat{\beta}_{ML} - \beta\big)\right] + (Z\beta - \beta)'(Z\beta - \beta) \tag{4.25}$$

$$= \sum_{j=1}^{j} \frac{\boldsymbol{\lambda}_j}{(\boldsymbol{\lambda}_j + k)^2} + \beta'^{k^2\left(X'\widehat{W}X + kI\right)^{-2}}\beta = \gamma_{1(k)} + \gamma_{2(k)} \tag{4.26}$$

olarak tanımlanır.

Varyans $\gamma_{1(k)}$ ile temsil edilir ve yanlılık ifadesinin karesi $\gamma_{2(k)}$ ile temsil edilir. İki durumda, PRR yaklaşımı öne çıkar. Birincisi ve en önemlisi, herhangi bir yazılımsal değişikliği gerektirmeyen çok basit bir yaklaşımdır. İkinci olarak, varyans terimindeki azalmanın yanlılık teriminin karesindeki artıştan daha büyük olacağı şekilde yeterli bir k değerinin belirlenmesi, en fazla olasılık tahmininden daha düşük bir MSE değeri verir.

Hoerl ve Kennard'a (1970a, 1970b) göre MSE genellikle EKK'dan daha küçüktür ve sıfırdan büyük bir k değeri vardır. Poisson tahmincisinin MSE'si de bir k değeridir ve burada görüldüğü gibi en olasılık tahmincisinin MSE'si daha düşüktür. $\gamma_{1(k)}$için k sürekli azalan bir fonksiyon olduğu ve $Y_{2(k)}$için ise sürekli artan bir fonksiyon olduğu belirtilmesi gerekmektedir.

$$\gamma_{2(k)} = k^2\beta' k^2(x'^{\widehat{W}}x + kl)^{-2}\beta = k^2\sum_{i=1}^{p}\frac{a_i^2}{(\lambda_i + k)^2} \tag{4.27}$$

Burada $\alpha_i^2 = \gamma\hat{\beta}_{ML}$'dir ve $\gamma, x'\widehat{W}x$ matrisinin özvektörüdür.

$k \geq 0$ ve $\lambda_i > 0$ tüm i değerleri için olduğundan, $\gamma_{2(k)}$'in k 'nın monoton artan bir fonksiyonu olduğunu varsaymalıyız. $E(L^2_{PRR}) < E(L^2_{ML})$ olduğunu kanıtlamak için, $E(L^2_{PRR})$'nin k 'ya göre ilk türevi alınırsa:

$$\frac{\delta E(L_{PRR}^2)}{\delta k} = \frac{\delta Y_{1(k)}}{\delta k} + \frac{\delta Y_{2(k)}}{\delta k} = -2\sum_{i=1}^{p} \frac{\boldsymbol{\lambda}_i}{(\boldsymbol{\lambda}_i + k)^3} + 2k\sum_{i=1}^{p} \frac{\boldsymbol{\lambda}_i \boldsymbol{a}_i^2}{(\boldsymbol{\lambda}_i + k)^3} \quad (4.28)$$

Öte yandan, $\gamma_{1(k)}$, k 'nın monoton olarak azalan bir fonksiyonu, $\gamma_{2(k)}$ ise k 'nın monoton olarak artan bir fonksiyonudur. Ayrıca $\gamma_{1(k)}$ ve $\gamma_{2(k)}$'nin ilk türevlerinin her zaman pozitif ve non- sırasıyla negatif. Sonuç olarak, $\frac{\delta E(L_{PRR}^2)}{\delta k} < 0$ denkleminde olduğu gibi, $E(L_{PRR}^2) < E(L_{ML}^2)$'nin her zaman sıfırdan büyük bir k değerine sahip olduğunu kanıtlamak önemlidir.

$$k < \frac{1}{a_{maksimum}^2} \quad (4.29)$$

4.6.1.1. Önerilen ridge parametresi tahmin yaklaşımları

Ridge parametresi önerilerinden biri Hoerl ve Kennard (1970a, 1970b) tarafından yapılmıştır. Bu parametre şu şekilde gösterilmektedir:

$$K1 = k_{HK1} = \frac{\sigma^2}{\hat{a}_{maksimum}^2} \quad (4.30)$$

tahmin edicidir. Öte yandan;

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2}{n - p - 1} \quad (4.31)$$

Yukarıda verilen tahmin edicinin değiştirilmiş bir başka hali;

$$K2 = \hat{k}_{HKM} = \frac{1}{\hat{a}_{maksimum}^2} \quad (4.32)$$

olarak önerilmiştir. PRR eşitliğindeki $\frac{1}{\hat{a}_{maksimum}}$ için Denklem (3.30)'da bulunan optimum k değerine karşılık gelmektedir.

Bu tahmin edicilerin çok sayıda çalışmada (Alkhamisi ve Shukur, 2008) k'nin en iyi değerine yaklaştığı görülmesine rağmen, Kibria'dan (2003) iki öngörücü daha eklenecektir:

$$K3 = \hat{k}_{GM} = \frac{\sigma^2}{(\prod_{i=1}^{p} \hat{a}_i^2)^{\frac{1}{p}}} \tag{4.33}$$

Ve

$$K4 = \hat{k}_{MED} = Medyan\{m_i^2\} \tag{4.34}$$

Bu Denklemde

$$m_i = \sqrt{\frac{\hat{\sigma}^2}{\hat{a}_i^2}} \tag{4.35}$$

şeklinde ilerleyecektir.

Alkhamisi vd. (2006) tarafından önerilen k tahmin edici ise şu şekilde gösterimektedir:

$$K5 = \hat{k}_{maksimum}^{KS} = maksimum(s_i) \tag{4.36}$$

Burada

$$s_i = \frac{t_i \hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + t_i \hat{a}_i^2} \tag{4.37}$$

dir ve ti , $X'X$ matrisinin özdeğeridir.

Kibria (2003)'ün ve aynı zamanda Hoerl ve Kennard (1970a, 1970b)'in önermiş olduğu regresyon tahmin edicileri için beklenen katsayı vektörü ortalaması bazında

daha uzundur. Bu nedenle korelasyon (p) değerinin artmasıyla birlikte tahmin edilen değerlerde düşüş olmaktadır.

Bazı durumlarda, bu fonksiyon etkisiz olabilir, çünkü $x'\widehat{W}x$ matrisinin neredeyse tekillik problemini artan korelasyon (ρ) değerlerine karşı çözmek için daha geniş bir k'nın değerler spektrumuna ihtiyaç vardır. Muniz ve Kibria (2009), Alkhamisi ve Shukur'un (2008) önceki tahmin edicilerine ve karekök dönüşümleri kavramına dayanarak aşağıdaki tahminleyicileri önermiştir:

$$\text{K6}=\hat{k}_{KM2} = maksimum\left(\frac{1}{m_i}\right) \tag{4.38}$$

$$K7 = \hat{k}_{KM4} = \left(\prod_{i=1}^{p}\frac{1}{m_i}\right)^{1/p} \tag{4.39}$$

$$K8 = \hat{k}_{KM6} = medyan\left(\frac{1}{m_i}\right) \tag{4.40}$$

Bu tahmin ediciler, m_i 'nin karşılığına dayalıdır, bu nedenle benzerliğin (ρ) değeri ne kadar yüksekse, tahmin ediciler o kadar fazla olur. Sonuç olarak, çoklu ilişki problemi için en dayanıklı tahmin edicilerin bu tahmin ediciler olduğu düşünülmektedir.

4.6.2. Poisson ridge tahmin edici ve çok sıfırlı modeller

Bağımlı değişkenlerin orantısız dağılımını ve çok sıfırlı poisson regresyon modeli, negatif olmayan ekonomik sonuçların çalışılmasında sıklıkla kullanılır. Bu model söz konusu olduğunda patent sayısı, ipoteklerin ve kredi kartlarının ödenme sıklığı, işletmelere verilen tekliflerin hacmi ve çeşitli ihracatların sayısı gibi farklı ekonomik değişkenlerin nasıl olduğunu görmek için inovasyon araştırmasında kullanılmaktadır (Kibria vd., 2013).

Çoklu dahili ilişkiler çok fazla belirsizlik yaratır ve katsayı hesaplamalarını görüntülemeyi imkansız hale getirir. Sonuç olarak, birkaç karşılıklı ilişki güvenilir nedensel çıkarımlar yapar ve çeşitli ekonomik değişkenlerin bağımlı değişken

üzerindeki etkisini araştırmayı zorlaştırır. Hoerl ve Kennard (1970a, 1970b) doğrusal regresyon modelleri için bu problemi çözmek için sıradan KKK tahmincisi yerine çarpık bir tahminci olan ridge regresyonunun kullanılmasını önermiştir.

Hoerl ve Kennard (1970a, 1970b), az miktarda önyargının, tahmin edilen katsayıların varyansını büyük ölçüde azaltacağını gösterdi. Schaffer vd. (1984), logit ve poisson modelleri de dahil olmak üzere makine öğrenimi beklenen modellerinde kullanılmak üzere bu modeli geliştirmiştir (Kibria ve diğerleri, 2013).

Çok sıfırlı poisson regresyon modeli için ilk ridge regresyon tahmin edicisi Kibria vd. (2013) tarafından önerilmiş ve ilk ölçüm gerçekleştirilmiştir. Önerilen bu tahmin edici daha sonrasında Kibria vd. (2011) tarafından genelleştirilmiştir. Bu bağlamda ZIPRR artık bu genelleştirilmiş tahmin edicilere erişebilir. Çeşitli tahmin edicilerin sonuçlarını karşılaştırmak için MSE ve MAE işlemeyi ölçmek için Monte Carlo Simülasyonu kullanılmıştır.

Örnek boyutu, açıklayıcı değişkenler arasındaki ilişki derecesi ve logit modellerinin kesişim değerleri gibi faktörler simülasyon analizinde değiştirildi. Simülasyon çalışmasının bulguları, özellikle ZIPR modeli için yeni bir RR tahmincisinin ML tahmincisinden daha iyi performans gösterdiğini göstermektedir.

β'yı tahmin etmek için en yaygın metot

$$l = \sum_{i=1}^{N} \gamma_i \log(\pi_i) + \sum_{i=1}^{N} (1 - \gamma_i) \log(1 - \pi_i) \tag{4.41}$$

ML yöntemi, ile gösterilen tipik log-olabilirlik denklemini maksimize etmektedir.

Nelder ve Mead (1965) tarafından önerilen simpleks yaklaşımı ile karmaşık olasılık fonksiyonu hesaplanabilmektedir. Bu bağlamda ilk değerlerin poisson ve logit tahminleri baz alınarak tekrarlayan ağırlıklı en küçük kareler (IWLS) algoritması kullanılmaktadır. Kibria vd. (2013), bu yönteme alternatif olarak poisson modeli için aşağıdaki tahmin ediciyi önermiştir:

$$\hat{\beta}_{RR} = (X'WX + kl)^{-1}X'\widehat{W}X\hat{\beta}_{ML} \quad (4.42)$$

Tek yönlü işlem kullanılarak yapılan tahminler, $\hat{\beta}_{ML}$ tahminleri olarak adlandırılır. $\widehat{W}$, sıfır diyagonal olmayan elemanlara ve i. eşit bir $\hat{\mu_i}$ diyagonal elemana sahip bir matristir. Doğrusal, Logit ve poisson modelleriyle aynı mantığın yanlılık tahmin edicisi Hoerl ve Kennard (1970a, 1970b) tarafından keşfedilmiştir. Sapma parametresi k, 0 ile sonsuz arasında bir değer aralığına sahiptir. Sıfıra eşitse $\hat{\beta}_{RR} = \hat{\beta}_{ML}$ olur. Sıfırdan büyükse $||\hat{\beta}_{RR}|| \leq ||\hat{\beta}_{ML}||$ olur. Birden fazla dahili ortaklık durumunda en olası öngörücü olduğu tahmin edilmektedir ve $\hat{\beta}_{RR}$'nin $\hat{\beta}_{ML}$'den daha iyi yapması beklenmektedir çünkü $\hat{\beta}_{ML}$ ortalamada çok uzun olacaktır (normu büyük) (Kibria vd. 2013).

4.6.2.1. Önerilen ridge parametresi tahmin yaklaşımları

Ridge parametresini hesaplamak için ayarlanmış bir yöntem yoktur. Bununla birlikte, doğrusal RR modeli için birkaç yöntem geliştirilmiştir ve bu yöntemler ZIPR modeline genişletilmiştir (Kibria ve diğerleri, 2013). İlk tahmin edici,

$$K1 = \hat{k}_{HK1} = \frac{1}{\hat{a}^2_{maksimum}} \quad (4.43)$$

Hoerl ve Kennard bunu (1970a) ve (1970b')de önermiştir. Bu, $\hat{\alpha}^2_{maksimum}$ $\gamma\hat{\beta}_{ML}$'in en önemli işlevidir. $\gamma, X'\widehat{W}X$ ve $\sigma^2 = \frac{\sum_{i=1}^{n}(y_i-\mu_i)^2}{n-p-1}$ matrislerinin özvektörüdür.

Schaeffer vd. (1984) ikinci bir k tahtarafından önerilen tahmin edicidir. Lojistik RR ile ilgili olarak:

$$K2 = \hat{k}_{HKM} = \frac{1}{\hat{\sigma}^2_{maksimum}} \ (Schaefer\ vd. 1984) \quad (4.44)$$

Sonraki iki belirleyici, Kibria (2003) tarafından önerilen doğrusal regresyon tahmin edicileridir.

$$K3 = k_{GM} = \frac{\hat{\sigma}^2}{\left(\prod_{i=1}^{p} \hat{a}_i^2\right)^{1/p}} \tag{4.45}$$

$$K4 = \hat{k}_{MED} = medyan\{m_j^2\} \tag{4.46}$$

Burada;

$$m_i = \sqrt{\frac{\hat{\sigma}^2}{\hat{a}_i^2}} \tag{4.47}$$

şeklinde vurgulanır. k parametresinin seçimi için Denklem (4.48) – (4.59) ile verilen tahmin ediciler de kullanılabilir.

$$\text{K5} = \hat{k}_{m5} = \text{Maksimum}\left(\frac{1}{m_j}\right) \tag{4.48}$$

$$K6 = \hat{k}_{m6} = Maksimum(m_j) \tag{4.49}$$

$$K7 = \hat{k}_{m7} = \prod_{j=1}^{p} (\frac{1}{m_j})^p \tag{4.50}$$

$$K8 = \hat{k}_{m8} = \prod_{j=1}^{p} (m_j)^{\frac{1}{p}} \tag{4.51}$$

$$K9 = \hat{k}_{m9} = medyan\left(\frac{1}{m_j}\right) \tag{4.52}$$

$$K10 = \hat{k}_{m10} = medyan(m_j) \tag{4.53}$$

$$K11 = \hat{k}_{m11} = Maksimum\left(\frac{1}{q_j}\right) \tag{4.54}$$

$$K12 = \hat{k}_{m12} = Maksimum(q_j) \tag{4.55}$$

$$K13 = \hat{k}_{m13} = \prod_{j=1}^{p} (\frac{1}{q_j})^{\frac{1}{p}} \quad (4.56)$$

$$K14 = \hat{k}_{m14} = \prod_{j=1}^{p} (q_j)^{\frac{1}{p}} \quad (4.57)$$

$$K15 = \hat{k}_{m15} = medyan\left(\frac{1}{q_j}\right) \quad (4.58)$$

$$K16 = \hat{k}_{m1} = medyan(q_j) \quad (4.59)$$

Muniz ve Kibria (2009) yapmış olduğu çalışmasında ile Kibria ve ark. (2012) yapmış olduğu çalışmasında bunu $q_j = \frac{\lambda_{maksimum}}{(n-p)\hat{\sigma}^2 + \lambda_{maksimum}\hat{\alpha}_j^2}$ olarak tanımlar. $X'\widehat{W}X$ matrisinin en büyük öz değeri $\lambda_{maksimum}$ değeridir (Kibria vd., 2013).

4.6.3. Modifiye edilmiş jackknifed poisson ridge tahmin edici

Poisson regresyonu, uygulamalı bilimlerde sayılabilir veri işlemede yaygın olarak kullanılmaktadır. Bununla birlikte, bağımsız değişkenler güçlü bir şekilde ilişkilendirildiğinde, Poisson regresyon modeli çoklu doğrusal bağlantı sorunuyla karşı karşıya kalır. Yanlılık tahmin edicilerinin kullanımı, çoklu doğrusal ilişkilerden kaynaklanan problemleri çözmek için kullanılan standart bir çözümdür. Poisson regresyon modelinde ridge parametresini k'yı tahmin etmek için birkaç yöntem yakın zamanda önerilmiştir. Bu tekniklere dayanan herhangi bir tahmin edicinin, en yaygın olarak kullanılan olasılık tahmin edicisinden daha üstün olduğu gösterilmiştir.

Çoklu doğrusal ilişki problemini çözmek için Türkan Özel (2016), analizlerinde modifiye edilmiş jackknifed poisson regresyon (MJPR) tahmin ediciyi önermiştir. Tahmin edicilerin verimliliği, bir simülasyon analizi ve gerçek bir veri seti kullanılarak değerlendirilmiştir. Hata kareler ortalaması ve karşılaştırmalı hata yüzdesi performansı tahmin edicileri kriter olarak kabul edilmştir. Analizde test edilen tüm durumlar için,

modifiye edilmiş jackknifed poisson regresyon (MJPR) yaklaşımı, simülasyonlarda ve gerçek veri örneklerinde poisson ridge regresyonu, jackknifed poisson ridge regresyonu ve MLE yöntemlerinden daha iyi performans göstermiştir. Singh vd. (1986) tarafından önerilen önerilen jackknifed ridge regresyon tahmin edicisi şu şekildedir:

$$\hat{Y}_{JRR} = (I + kA^{-1})^{-1}\hat{Y}_{RR} = (I - k^2A^{-2})\hat{Y}_{OLS} \tag{4.60}$$

Batah vd. (2008) tarafından önerilen modifiye edilmiş jaccknifed ridge regresyon (MJRR) tahmin edicisinde ise Singh vd. (1986) tarafından önerilen jackknifed ridge regresyon tahmin edicisi referans alınmıştır. Önerilen jackknifed ridge regresyon şu şekildedir:

$$\hat{Y}_{MJRR} = (I - k^2A^{-2})\hat{Y}_{RR} \tag{4.61}$$
$$= (I - k^2A^{-2})(I - kA^{-1})\hat{Y}_{OLS}$$

şeklinde tanımlanmıştır. Burada $A = (\Lambda + kl)$ olup $\Lambda = diag(\Lambda_1, \Lambda_2, \dots, \Lambda_p)$ özdeğerler matrisidir.

$G = (g_1, g_2, \dots, g_p)$ pxp boyutlu, $X'\widehat{W}X$ matrisinin normalleştirilmiş öz vektörlerinden oluşan kolon matrisi, $\Lambda_{PR} = diag(\Lambda_{1P}, \Lambda_{2PR}, \dots, \Lambda_{pPR})$ öyle ki $G'X'\widehat{W}XG = Z'\widehat{W}Z = \Lambda_{PR}$ ve $Z = XG$ olmak üzere, ^{ML}'nin poisson tahmin edici aşağıda gösterildiği gibi yazılabilir:

$$\hat{\Upsilon}^{ML} = \Lambda^{-1}{}_{PR}Z'\widehat{W}\hat{s} \tag{4.62}$$

$$\hat{\beta}_{ML} = G_{\hat{\gamma}^{ML}} \tag{4.63}$$

Tekrardan poisson ridge regresyon tahmin edici,

$$\hat{\gamma}^{PRR} = (\Lambda_{PR} + kI)^{-1}Z'\hat{W}\hat{s} = (I - kB^{-1})\hat{\gamma}^{ML} \tag{4.64}$$

bu kısımda $B = (\Lambda_{PR} + kl) ve\ K > 0$'dır.

$$\hat{\gamma}^{PRR} = (Z'\hat{W}Z + kl)^{-1}Z'\hat{W}\hat{s} \tag{4.65}$$

Hinkley (1977) tarafından uyarlanmış JPR tahmin edici

$$\hat{\gamma}^{JPR} = \hat{\gamma}^{PRR} + (Z'\hat{W}Z + kI)^{-1}\sum_{i=1}^{n} z_i'\mu_i(\hat{s}_i - z_i'\hat{\gamma}^{PRR}) \tag{4.66}$$

Şeklinde belirtilir ve bu Denklem üzerinde sadeleştirme yapılırsa

$$\hat{\gamma}^{JPR} = (I + B^{-1}k)\hat{\gamma}^{PRR} = (I - (B^{-1}k)^2)\hat{\gamma}^{ML} \tag{4.67}$$

Eşitliği sağlanmış olur. Bu kısımda

$$B = \left(Z'\widehat{W}Z + kI\right) \tag{4.68}$$

dir ve yukarıdaki denklemden modifiye edilmiş jackknifed poisson regresyon (MJPR) tahmin edicisi

$$\hat{\gamma}^{MJPR} = (I - B^{-2}k^2)(I - B^{-1}k)\hat{\gamma}^{ML} \tag{4.69}$$

biçiminde ulaşılmış olur.

$\hat{\gamma}^{PRR}, \hat{\gamma}^{JPR}\ ve\ \hat{\gamma}^{MJPR}$'nin MSE ölçütü aşağıda belirtildiği gibi elde edilebilir:

$$MSE(\hat{\gamma}^{PRR}) = Var(\hat{\gamma}^{PRR}) + [bias(\hat{\gamma}^{PRR})][bias(\hat{\gamma}^{PRR})]'$$
$$= (I - kB^{-1})\Lambda^{-1}{}_{PR}(I - kB^{-1}) + k^2B^{-1}\gamma\gamma'B^{-1} \tag{4.70}$$

$$MSE(\hat{\gamma}^{JPR}) = Var(\hat{\gamma}^{JPR}) + [bias(\hat{\gamma}^{JPR})][bias(\hat{\gamma}^{JPR})]'$$
$$= (I - k^2B^{-2})\Lambda^{-1}{}_{PR}(I - k^2B^{-2}) + k^4B^{-2}\gamma\gamma'B^{-2} \tag{4.71}$$

$$MSE(\hat{\gamma}^{MJPR}) = Var(\hat{\gamma}^{MJPR}) + [bias(\hat{\gamma}^{MJPR})][bias(\hat{\gamma}^{MJPR})]'$$
$$= (I - k^2B^{-2})(I - kB^{-1})\Lambda^{-1}{}_{PR}(I - kB^{-1})(I - k^2B^{-2}) + k^2w_pB^{-1}\gamma\gamma'w_p \tag{4.72}$$

Burada

$$w_p = (I + kB^{-1} - k^2B^{-2}) \tag{4.73}$$

dir.

4.6.3.1. Önerilen ridge parametresi tahmin yaklaşımları

Ridge parametresini k'yı tahmin edebilmek adına herhangi açık bir yasa söz konusu değildir. Ridge regresyon modeli için önerilen yöntemler, modifiye edilmiş jackknifed poisson regresyonu (MJPR) için de kullanılabilir. Ridge parametrelerini tahmin etmek için en yaygın olarak kullanılan beş tahmin ediciden aşağıda bahsedilmektedir. Hoerl ve Kennard (1970a, 1970b) bunlardan ilki olarak aşağıdaki tahminciyi önermiştir:

$$MJPR1 = \hat{k}_1 = \frac{\partial^2}{\hat{\alpha}^2_{maksimum}} \tag{4.74}$$

Yukarıda $\hat{\alpha}^2_{maksimum}$, $\delta^T\hat{\beta}^{ML}$ 'nin büyük elemanıdır. k için ikinci tahmin edici Schaeffer ve ark. (1984) tarafından geliştirilmiştir.

$$MJPR2 = \hat{k}_2 = \frac{1}{\hat{\alpha}^2_{maksimum}} \tag{4.75}$$

Aşağıda belirtilen diğer tahmin ediciler ise 2011 yılında Kibria ve ark tarafından önerilmiştir.

$$MJPR3 = \hat{k}_3 = medyan(m_i^2) \tag{4.76}$$

$$MJPR4 = \hat{k}_4 = maksimum(m_i) \tag{4.77}$$

$$MJPR5 = \hat{k}_5 = maksimum\left(\frac{1}{m_i}\right) \tag{4.78}$$

Burada;

$m_i = \sqrt{\hat{\sigma}^2/\hat{a}_i^2}$ ve $\hat{a}_i^2 = \delta^T \hat{\beta}^{ML}$

yukarıdaki denklemin elemanıdır.

4.6.4. Poisson regresyon modellerinde liu tahmin edici

Månsson vd. (2012) Poisson Liu tahmincisi kullanılarak elde edilen maksimum olasılık çözümlerinden elde edilen ortalama kare hata distorsiyonu sorununa bir çözüm olarak verilerin sayılmasında daha başarılı sonuçlar sunmak için yinelemeli ağırlıklı en küçük kareler yaklaşımı kullanmıştır. $\hat{\beta}_{ML}$'nin klasik Poisson regresyon tahminleri olduğu durumlarda, Poisson Liu regresyon (PLR) tahmincisi kullanılır:

$$\hat{\beta}_{PLR} = \left(X'\widehat{W}X + 1\right)^{-1}\left(X'^{\widehat{W}}X + dI\right)\hat{B}_{ML} \tag{4.79}$$

Şeklinde tanımlanır.

4.6.5. Liu tahmin edicide yanlılık kriterinin belirlenmesi

Månsson vd. (2012) tarafından önerilen tahmin ediciler, Liu parametresinin MSE'sini d parametresinin MSE'sinden çıkararak ve ardından her parametre için sonuç fonksiyonunu sıfıra eşitleyerek oluşturmaktadır.
Månsson ve ark. (2012) tarafından önerilen bazı tahmin ediciler aşağıdaki gibidir:

$$d2 = maksimum\left(0, \frac{1}{p}\sum_{j=1}^{p}\frac{\hat{a}_j^2 - 1}{\frac{1}{\boldsymbol{\lambda}_j} + \hat{a}_j^2}\right) \tag{4.80}$$

$$d3 = maksimum\left(0, medyan\left(\frac{\hat{a}_j^2 - 1}{\frac{1}{\lambda_j} + \hat{a}_j^2}\right)\right) \tag{4.81}$$

$$d4 = maksimum\left(0, maksimum\left(\frac{\hat{a}_j^2-1}{\frac{1}{\lambda_j}+\hat{a}_j^2}\right)\right) \tag{4.82}$$

$$d5 = maksimum\left(0, minimum\left(\frac{\hat{a}_j^2-1}{\frac{1}{\lambda_j}+\hat{a}_j^2}\right)\right) \tag{4.83}$$

Bu kısımda;

$\hat{a}_j, \hat{a} = \phi\hat{\beta}_{MLE}$ matrisinin j. bileşenidir.
Φ ise $X'\widehat{W}X$ matsinin öz vektörlerinin sütunudur.

Açıklayıcı değişkenler ilişkiliyse, optimal parametreler, klasik MLE, ridge ve Liu tahmin edicilerinin MSE çıktılarını karşılaştırarak bir simülasyonda belirlenebilir.

4.7. Log-Normal Dağılımlı Poisson Regresyon Modeli

Log-normal poisson regresyon modelinin ortaya konulabilmesi için log-normal dağılımının değerlendirilmesi gerekmektedir (Breslow, 1984). Bu bağlamda $E(\epsilon_i) = e^{\sigma^2/2}$ ve $Var(\epsilon_i) = e^{\sigma^2}(e^{\sigma^2}-1)$ şeklinde ϵ_i log-normal dağılıma sahip bir değişkendir:

$$\in_i \sim \ln N(0, \sigma^2) \tag{4.84}$$

şeklinde gösterilir. Bu değişkenin beklenen değeri ve varyansı ise,

$$E(y_i|x_i) = \exp\left(x_i^T\beta + \frac{\sigma^2}{2}\right) \tag{4.85}$$

$$Var(y_i|x_i) = E(y_i|x_i) + \left(e^{\sigma^2}-1\right)E^2(y_i|x_i)$$

biçiminde tanımlanır (Breslow, 1984; Winkelmann, 2008; Agresti, 2015).

e_i, NBR modelinden çıkarıldığında (ayrı bir şekilde), y_i dağılımının analitik formu yoktur ve MLE doğrudan ölçülemez. Winkelmann (2008) ise log-normal poisson modelinin teknik olarak daha avantajlı olduğu ve sonuçlarla en iyi şekilde eşleşebileceği için yeniden değerlendirilmesini önermiştir (Zhou vd., 2012).

4.8. Poisson Engel Regresyon Modeli

Yanıt verileri sayım (frekans) değişkenlerinden oluştuğunda ve gözlemlenen yanıt değerleri önemli sayıda sıfır değeri içerdiğinde, model olumsuz etkilenecek ve modelleme sürecine yeni bir yaklaşım gerekli olacaktır. Birkaç sıfır içeren modellere benzer şekilde, engelli regresyon modelleri bu sorunu çözmek için kullanılabilir. Cragg (1971) sansürlenmiş veriler için engel (hurdle) regresyon yöntemini önermesine rağmen, Mullahy (1986) genel çerçevesini sayım verilerindeki uygulamalarıyla paylaşmıştır.

Hurdle regresyon modelinde iki adımlı bir yöntem kullanılır. Cevap değişkeninin değerinin sıfır olup olmamasına bağlı olarak, bu nedenle çeşitli denklemler yapılır. Sonuç değişkeni, bu nedenle sıfır ve pozitif değerlere yanıt olarak ikili bir yanıt olarak görülür. Yanıt değişkeninin sıfır değerini aldığı durumlar için bir olasılığı belirleyen iki aşamalı bir prosedürün ilk aşamasıdır. Sıfır dışındaki değerler için kesilmiş bir dağılım kullanır, bu durum operasyonunun ikinci aşamasıdır (Agresti, 2015). Engel modelleri uyarlanabilir ve hem aşırı hem de yetersiz dağılma durumlarında kullanışlıdır (Saffari vd, 2012).

Çok fazla sıfır olan modeller ve engel modelleri benzer uygulamalara sahipken, kavramsal olarak farklılık gösterirler. Her iki gruptaki modellerin iki aşamalı bir yöntemi vardır, ancak çok fazla sıfır olan modeller hariç, tahmin formülü sıfır olmayan değerlerdir, oysa yanıt yine sıfırdır.

Hurdle modelinde sıfır olmayan yanıt değerlerini modellerken, yanıtın pozitif olması şartı vardır. Yapı açısından, her iki model de sonucu ikili değişkenler olarak ele alırken sıfır değerlerinin tahmin modelini oluşturmak için klasik ayrık dağılımlar kullanır, çok

fazla sıfıra sahip sıfır olmayan değerler modelleri kesilmiş ayrık dağılımları kullanırken, Hurdle modelleri kesilmiş ayrık dağılımları kullanmaktadır. Engel regresyon modellerinin günlük olabilirlik fonksiyonu genel bir forma sahiptir;

$$L(\delta_1, \delta_2) = L_1(\delta_1) + L_2(\delta_2)$$
$$= \sum_{i=1}^{n} [(1 - c_i) ln f_1(0|x_i, \delta_1) + c_i \ln\,(1 - f_1(\,0|x_i, \delta_1))]$$
$$+ \sum_{i=1}^{n} c_i [ln f_2(y_i|x_i, \delta_2) - \ln\big(1 - f_2(0|x_i, \delta_2)\big)] \quad (4.86)$$

şeklindedir. Burada c_i,

$$c_i = \{1, y_i > 0\}$$
$$c_i = \{0, y_i = 0\} \quad (4.87)$$

olarak tanımlanan bir gösterge değişkendir. Sıfırların olasılığı $f_1(0|x_i, \delta_1)$ ile temsil edilir ve sıfır olmayan cevapların olasılıkları $f_2(y_i|x_i, \delta_2)$ ile temsil edilir. Sonuç olarak, $L_1(\delta_1)$ sıfır ve sıfır olmayan değerli birimleri birbirinden ayıran ikili bir süreci gösterirken, $L_2(\delta_2)$ aşağıdakilerden türetilen log olabilirlik fonksiyonunu gösterir (Cameron ve Trivedi, 2013).

Hurdle regresyon modelleri oluşturmak için, f_1 ve f_2 fonksiyonlarının (dağılımlarının) işlem sırasında belirlenmesi gerekir. Genel olarak, ilk aşamayı oluşturan ikili faz için f_1 fonksiyonunda logit veya probit modelleri kullanılırken, ikinci aşamada sıfır olmayan yanıt değerlerini tahmin etmek için hem poisson hem de negatif binom dağılımları tercih edilir (Cameron ve Trivedi, 2013). İkinci aşamada kullanılan dağıtımlar, klasik dağıtımların kesilmiş kopyalarıdır.

4.9. Poisson Regresyon Modellerinde Kullanılan Uyum Testleri

Elde edilen bulguların değerlendirilebilmesi için değerlendirilen veri grubuna uygun regresyon modelinin seçilmesi gerekmektedir. Bu bağlamda uyumu değerlendirebilmek için kullanılabilecek testler şu şekilde sıralanmaktadır:

- Sapma İstatistiği (G^2)
- Pearson Ki-Kare χ^2
- Yapay (R^2) İstatistiği
- Akaike Bilgi Kriteria
- Pearson İstatistiği

4.9.1. Sapma istatistiği

Uyum iyiliğinin ölçülmesinde kullanılan bir istatistiktir ve

$$D = 2\sum_{i=1}^{n}\left[\gamma_i \ln\left(\frac{y_i}{\mu_i}\right) - (y_i - \mu_i)\right] \tag{4.88}$$

şeklinde tanımlanır.

Eğer modelde

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{\mu}_i \tag{4.89}$$

ise olabilirlik oranı uyum iyiliği istatistiği

$$D = 2\sum_{i=1}^{n} y_i \ln\left(\frac{y_i}{\mu_i}\right) \tag{4.90}$$

biçiminde olacaktır.

Elde edilen varyans istatistiği sıfıra eşit olduğundan, model uyumunun geliştiği söylenir. Gözlenen varyans istatistik değeri 0 olduğunda, modelin uyumu optimal düzeydedir (Deniz, 2005).

4.9.2. Pearson ki-kare istatistiği

Pearson χ^2 istatistiği, araştırılan model için tahmin edilen ve gözlemlenen frekanslar kullanılarak belirlenir. Aşağıdakiler, uyumun iyiliğini belirlemek için kullanılacak teorilerdir:

H_0 = Veriler f(x)dağılımına uygunluk göstermektedir.
H_1= Veriler f(x)dağılımına uygunluk göstermemektedir.

şeklinde gösterilecektir. Pearson ki-kare istatistiği,

$$x^2 = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\widehat{\omega}_i} \tag{4.91}$$

biçimindedir. Bu değer, dizinin çok fazla yayılmasına sahip olup olmadığına karar vermek için kullanılır. Tahmin edilen μ_i ve ω_i değerleri, $\hat{\mu}_i$ ve $\widehat{\omega}_i$ değerleri ile ifade edilir. $\hat{\mu}$ için tanımlanan serbestlik derecesi (n-k), ölçülen χ^2 değeri ile karşılaştırılır.

Oluşturulan formül poisson regresyon modeline uygulandığında $\omega_i = \mu_i$ olacaktır ve

$$x^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu})^2}{\hat{\mu}_i} \tag{4.92}$$

biçiminde elde edilebilir.

Ölçülen χ^2 miktarının serbestlik derecesi oranına oranı 1'den büyükse, seri çok fazla yayılım olduğunu gösterir ve veriler model için uygun değildir; oran 1'den küçükse serinin yayılması çok azdır ve veriler modele uygundur (Deniz, 2005).

4.9.3. Yapay r2 istatistiği

Poisson regresyon modellerinin tahmininde gösterge olarak yapay R^2 istatistiği kullanılmıştır. Doğrusal regresyon modellerinde R^2'yi ölçmek için önce genel kare sayısı ayrıştırılmalıdır. Bu toplam (Deniz, 2005);

$$\sum_{i=1}^{n}(\gamma_i - \bar{\gamma})^2 = \sum_{i=1}^{n}(y_i - \mu_i)^2 + \sum_{i=1}^{n}(\hat{\mu}_i - \bar{\gamma})^2$$
$$+2\sum_{i=1}^{n}(y_i - \hat{\mu}_i)\,(\hat{\mu}_i - \bar{y}) \qquad (4.93)$$

şeklinde gösterilir. Poissons gibi doğrusal olmayan EKK ve üstel koşullu ortalamayı içeren tüm tahmin ediciler ve modeller için bu ifade sıfırdan farklı olacaktır. Bu durumda R^2, doğrusal regresyon dışında bir yöntem kullanılarak belirlenebilir.

R^2 ölçüsü, Poisson regresyon modelinde normallik varsayımı gerektirmeyen olasılık oranı yöntemine bağlıdır. AKT'nin olasılık tahmini ve sapma değeri ile doğrusal regresyon modeliyle ilgili tahmini aynı olduğundan, R_p^2 ölçüsü (Deniz, 2005):

$$R_p^2 = 1 - \frac{l(y) - l(\hat{\mu})}{l(y) - l(\bar{y})} \qquad (4.94)$$

l(y): doymuş modelin en çok olabilirliğini

l($\hat{\mu}$): iligili modelin en çok olabilirliğini

l($\bar{y}$): sabit terimin bulunduğu modelin en çok olabilirliğini ifade etmektedir.

$y_i \geq 0$ gözlemlenen değerler, $\hat{\mu}_i = \exp(x_i'\hat{\beta})$ tahmin edilen değerleri ve $\bar{y}_i = \exp(\hat{\beta}_0)$ ortalama değerleri olarak sıralandığı zaman log-olabilirlik fonksiyonları aşağıda sırası ile belirtilmektedir.

$$l(y) = \sum_{i=1}^{n} (y_i \ln(y_i) - y_i - \ln(y_i!)) \tag{4.95}$$

$$l(\hat{\mu}) = \sum_{i=1}^{n} (y_i \ln(\mu_i) - \hat{\mu}_i - \ln(y_i!)) \tag{4.96}$$

$$l(\bar{y}) = \sum_{i=1}^{n} (y_i \ln(\bar{y}_i) - \bar{y}_i - \ln(y_i!)) \tag{4.97}$$

biçiminde elde edilmektedir. Yukarıdaki log-olabilirlik fonksiyonları düzenlendiği takdirde yapay R^2 ölçüsüne ulaşılabilmektedir.

Serbestlik derecesine $n - p \geq 0$ kısıtı uygulandıktan sonra varyans tutarsızlığı test edilerek Poisson regresyon analizinde deneydeki değişkenlerin modele katkısı kontrol edilir. Bu fark değerini ilgili ki-kare tablo değerine uygulayarak test edilen değişkenin modele anlamlı bir katkı yapıp yapmadığını belirlemek mümkündür.

4.9.4. Akaike bilgi parametresi

1973'te Akaike, Akaike Bilgi Ölçütü'nü geliştirmiştir. Bu kriter, verilen sonucun diğer modellerle karşılaştırılabilmesi için oluşturulmuştur. Basit model ile gerçek model kıyaslandığında basit modelden daha karmaşık bir yapıya sahiptir. Ancak, gerçek modelin bazı işlevlerini daha iyi tahmin ettiği için seçilebilir. Akaike bilgi ölçütü:

$$AIC = 2(q - l) \tag{4.98}$$

şeklinde verilir. q sembolü tahmin edilen parametre sayısı olarak kabul edilmektedir (Cameron ve Trivedi, 2013; Deniz, 2005).

4.9.5. Pearson istatistiği

y_i'ye ait olan herhangi bir modelde μ_i ortalamalı ve ω_i varyanslı bağımlı değişken olduğunda standart uyum iyiliği ölçüm yöntemi olarak genellikle Pearson istatistiği kullanılır.

$$P = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu})^2}{\widehat{\omega}_i} \tag{4.99}$$

Bu sayı, dizinin dağılımının çok geniş olup olmadığına karar vermek için kullanılır. $\hat{\mu}_i$ ve $\widehat{\omega}_i$ değerleri, μ_i ve ω_i tahmin değerleridir.

Belirlenen P değeri, $\hat{\mu}$ değeri için hesaplanan (n-q-1) serbestlik derecesi ile karşılaştırılır. Burada;

$P_p > n - q - 1$ ise seride aşırı yayılım,
$P_p < n - q - 1$ ise seride eksik yayılım olduğundan bahsedilebilir.

Aşırı yayılım söz konusu olduğunda ise daha çok negatif binom regresyon modeli tercih edilir (Cameron ve Trivedi, 2013; Deniz, 2005).

4.10. Tahmin Edicilerin Performanslarının Değerlendirilmesi

Tahmin yöntemlerinin performanslarının ortaya çıkarılabilmesi için MSE değeri şu şekilde hesaplanır.

$$MSE = \frac{\sum_{i=1}^{R} (\hat{\beta} - \beta)'(\hat{\beta} - \beta)}{R} \tag{4.100}$$

MAE değeri ise aşağıda verilmiştir.

$$MAE = \frac{\sum_{i=1}^{R} |\hat{\beta} - \beta|}{R} \tag{4.101}$$

R: değeri tekrar sayısını belirtmektedir.
$\hat{\beta}$: β 'nin göstergesini ifade eder.

5. TÜRKİYE'DEKİ EVLİLİK İSTATİSTİKLERİ ÜZERİNE POISSON REGRESYON ANALİZİ UYGULAMASI

Küresel alanda yapılan pek çok araştırma gelişmiş ülkelerde bile erken evlilik sorunu ile karşı karşıya kalındığını ispatlamaktadır. Her ne kadar ülkeden ülkeye farklılıklar söz konusu olsa da evlilik yaşının giderek düşüş göstermesi toplum açısından büyük sorun teşkil etmektedir.

UNICEF'in 'Çocuk Evliliğini Sona Erdirme' raporuna göre, küresel alanda 700 milyondan fazla kadının erken evlilik yaptığı, özellikle 250 milyon kadının 15 yaşından küçükken evlendiği görülmektedir (UNICEF, 2013; 2). Söz konusu rapora göre dünya çapında erken yaşta evlenen kadınların %42'lik bölüm Güney Asya'da yaşamaktadır. Sanayisi gelişmiş olan ülkelerde bu oran %2, Orta ve Doğu Avrupa'da söz konusu oran %4'tür.

Çizelge 5.1. Dünya Çapında Evlilik Oranları (OECD, 2019)

	Erkek ortalama ilk evlenme yaşı			**Kadın ortalama ilk evlenme yaşı**		
	1990	**2000**	**2017**	**1990**	**2000**	**2017**
İsveç	30,3	33,0	36,6	27,7	30,4	33,8
ispanya	27,8	30,2	35,4	25,6	28,1	33,2
Danimarka	30,5	32,5	34,8	27,8	29,9	32,4
İzlanda	29,4	33,3	34,4	26,9	30,6	32,4
İtalya	28,9	30,9	35,0	25,9	27,8	32,3
Norveç	29,0	30,9	35,0	26,4	28,4	32,3
Fransa	..	30,7	34,4	..	28,4	32,2
İrlanda	28,7	..	33,8	26,6	..	32,0
Finlandiya	28,4	30,5	33,9	26,3	28,3	31,6
Lüksemburg	27,7	30,3	34,0	25,6	27,4	31,6
Avusturya	27,7	30,0	34,2	25,2	27,4	31,5
Şili	..	27,7	32,7	..	25,6	31,5
Hollanda	28,5	30,7	33,9	26,1	28,0	31,4
Belçika	26,5	29,1	33,5	24,4	26,9	31,2
Almanya	28,2	30,5	34,0	25,5	27,7	31,2
Birleşik Krallık	27,2	30,5	33,2	25,2	28,2	31,2
OECD25 ortalaması	27,7	29,8	33,2	25,2	27,3	30,8
Portekiz	26,6	27,4	32,5	24,6	25,2	30,7
Slovenya	26,9	29,9	32,9	23,9	27,0	30,4
İsviçre	29,5	30,8	32,8	27,0	28,2	30,4
AB ortalaması	-	-	32,9	-	-	30,4

Yunanistan	29,0	31,1	33,3	24,9	27,2	30,3
Macaristan	24,7	27,6	32,3	22,0	24,8	30,3
OECD ortalaması	-	-	32,5	-	-	30,2
Kore	27,8	29,3	32,9	24,8	26,5	30,2
Estonya	..	28,1	32,2	..	25,2	29,7
Japonya	28,4	28,8	31,1	25,9	27,0	29,4
Letonya	..	..	31,5	..	..	29,4
Malta	27,2	..	31,6	24,6	..	29,2
Çek Cumhuriyeti	24,3	27,6	31,8	21,6	24,6	29,1
Yeni Zelanda	26,8	29,2	30,2	24,5	27,4	29,0
Avustralya	26,5	28,5	30,4	24,3	26,7	28,8
Slovak cumhuriyeti	..	26,9	31,2	..	24,1	28,6
Hırvatistan	..	..	31,0	..	..	28,2
Litvanya	24,6	26,1	30,5	22,7	23,8	27,8
Meksika	..	..	30,4	..	..	27,5
Romanya	25,6	27,3	31,0	22,4	23,7	27,5
Amerika Birleşik Devletleri	26,1	26,8	29,5	23,9	25,1	27,4
Bulgaristan	24,6	28,5	30,6	21,5	24,7	27,3
Polonya	..	26,5	29,6	..	24,1	27,3
İsrail (a)	..	..	27,5	..	..	25,1
Türkiye	..	..	27,8	..	..	24,8

Yukarıda OECD'den alınan bilgilerle oluşturulan tabloya bakıldığında son 27 yıl içerisinde ilk evlenme yaşının artış eğiliminde olduğu görülmektedir. 1990 yılında Danimarka'da erkeklerde ilk evlilik yaşı 30,5, kadınlarda ilk evlilik yaşı ise 27,8 olarak tespit edilmiş, bu yaşlar en yüksek yaş olarak kabul edilmiştir. 2017 yılı verilerine bakıldığında ise erkeklerde 36,6 yaş, kadınlarda ise 33,8 ile İsveç ilk sırayı almaktadır. Türkiye ise 2017 yılı verilerine bakıldığında kadınlarda 24,8 yaş ile en düşük evlenme yaşına sahip olan ülke olmuştur. Erkeklerde ise bu yaş 27,8'dir ve bu ortalama ile İsrail'in 27,5 yaş oranını geçmiştir.

Evliliği ve evlilik yaşını etkileyen en önemli unsurlardan biri sosyo-demografik özelliklerdir. Erken evlilikler söz konusu olduğunda ekonomik yetersizliklerden ve eğitim eksikliğinden bahsedilebileceği gibi, geleneksellikten, dini pratiklerden de söz edilmelidir. Aynı zamanda felaketler, aile içi şiddet durumu ve toplumsal baskı da ön plana çıkmaktadır. Genellikle derin hasar bırakan bu durumlar eğitimsizliği giderek arttırmakta, yoksulluğa, anne ya da çocuk ölümlerine sebebiyet vermektedir (Jensen ve Thornton, 2003; 17; Aydemir, 2011: 3).

Yaş ve yıllara göre evlilik sayılarının değişiminin regresyon ile incelenmesi Türkiye'de yapılan evliliklerinin nedenlerini ve sonuçlarını anlamak adına önemli bir adımdır. Bu nedenle Poisson Regresyon uygulaması için Türkiye'deki 2010-2020 yılları arasındaki evlilik sayılarından, yaşlardan ve yıllardan yararlanılmıştır.

5.1. Evlilik Verileri

Çalışma kapsamında 2010 ile 2020 yılı arasında Türkiye'deki evlilik sayısı verileri bağımlı değişken olarak kullanılmıştır. Yaş, cinsiyet ve evliliğin gerçekleşmiş olduğu yıllar ise bağımsız değişkenlerdir. Evlilik sayıları 13 farklı yaş grubu ve yaş belirtilmemiş kategorisine göre sınıflandırılmıştır. Kullanılan yaş grupları şu şekildedir:

Çizelge 5.2. Analizde kullanılan yaş aralıkları

16-17 yaş,	50-54 yaş,
18-19 yaş,	55-59 yaş
20-24 yaş,	60-64 yaş,
25-29 yaş,	65-69 yaş,
30-34 yaş,	70-74 yaş,
35-39 yaş,	75+ yaş,
40-44 yaş,	Yaş belirtilmemiş.
45-49 yaş,	

Yıllara göre hangi yaş aralığında ne kadar evlilik yapıldığını gösteren düzenlenmiş veri seti ise TÜİK'ten alınmıştır.

Çizelge 5.3. 2010-2020 Arası Yaş ve Cinsiyete Göre Evlenme Sayıları (TÜİK 2016).

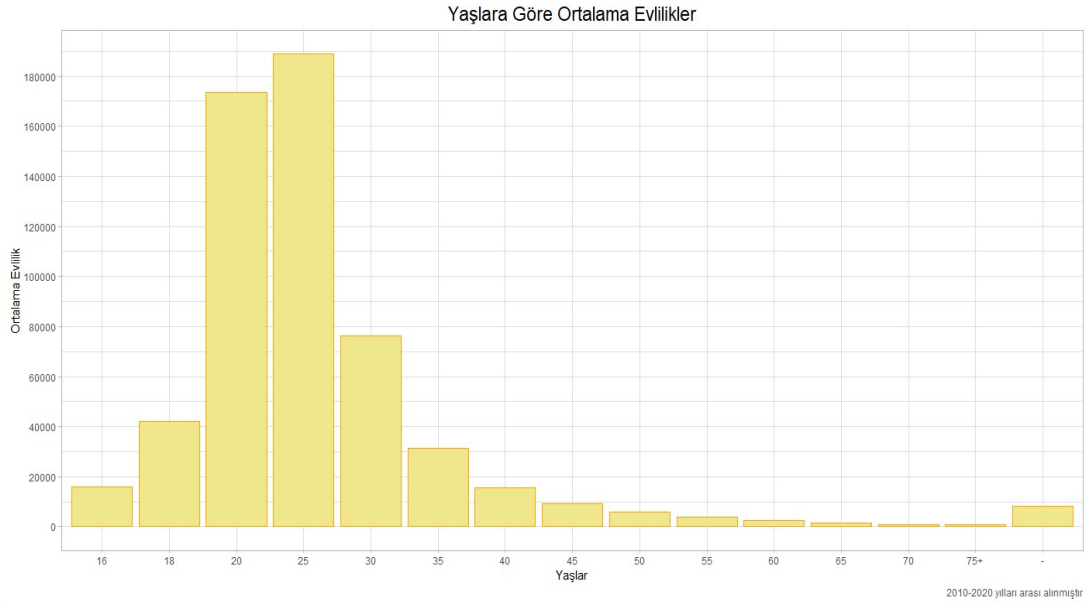
Yıl	Cinsiyet	16-17	18-19	20-24	25-29	30-34	35-39	40-44	45-49
2020	Erkek	726	6 231	106 047	195 290	89 341	37 296	19 591	11 429
	Kadın	13 014	46 948	169 412	145 841	49 850	26 036	15 909	9 512
2019	Erkek	940	7 481	118 724	218 767	99 335	41 331	20 528	12 121
	Kadın	17 058	56 135	188 376	159 689	54 422	28 767	16 379	9 959
2018	Erkek	1 029	8 327	127 288	220 732	100 271	42 271	19 932	11 799
	Kadın	20 809	62 053	191 940	151 336	51 148	27 213	14 289	8 588
2017	Erkek	1 081	8 786	135 927	227 475	101 409	42 108	19 440	11 196
	Kadın	23 906	65 946	200 225	152 574	51 801	27 118	14 256	7 922
2016	Erkek	1 319	9 689	144 925	235 986	104 725	43 323	19 724	11 258
	Kadın	27 637	72 596	208 954	153 567	54 259	28 050	14 161	7 618
2015	Erkek	1 483	10 173	149 767	240 881	106 709	40 223	18 656	10 634
	Kadın	31 337	77 002	213 596	153 297	54 871	25 842	13 459	6 996

2014	Erkek	1 670	10 949	152 863	239 621	105 095	37 504	17 958	10 576
	Kadın	34 629	81 646	213 112	147 355	54 787	23 916	12 851	6 710
2013	Erkek	1 866	11 680	155 555	241 462	103 064	35 982	16 537	10 291
	Kadın	37 481	85 639	213 019	145 037	53 463	22 989	11 632	6 804
2012	Erkek	1 903	12 304	160 034	240 428	102 385	34 436	16 312	10 869
	Kadın	40 428	88 702	215 487	139 695	53 450	22 309	11 383	7 479
2011	Erkek	1 860	12 357	162 204	238 572	97 512	31 881	14 600	10 300
	Kadın	42 700	87 947	214 283	136 540	49 768	20 800	10 039	6 848
2010	Erkek	2 000	12 824	163 791	237 474	91 054	30 276	13 729	9 237
	Kadın	45 738	89 136	212 923	132 952	45 817	19 409	9 326	6 057

Yıl	**Cinsiyet**	**50-54**	**55-59**	**60-64**	**65-69**	**70-74**	**75+**
2020	Erkek	7 369	5 405	3 697	2 251	1 437	1 160
	Kadın	5 123	3 040	1 575	679	224	101
2019	Erkek	8 069	5 637	3 881	2 620	1 525	1 354
	Kadın	5 528	3 152	1 609	699	259	118
2018	Erkek	7 946	5 610	3 813	2 474	1 441	1 455
	Kadın	4 810	2 712	1 425	616	294	129
2017	Erkek	7 756	5 289	3 795	2 351	1 469	1 377
	Kadın	4 589	2 525	1 439	678	262	146
2016	Erkek	8 431	5 331	4 001	2 472	1 440	1 401
	Kadın	4 847	2 536	1 560	659	234	138
2015	Erkek	7 348	5 071	3 336	2 388	1 338	1 409
	Kadın	4 301	2 374	1 317	662	238	136
2014	Erkek	7 068	4 881	3 315	2 085	1 307	1 405
	Kadın	4 048	2 307	1 174	580	243	167
2013	Erkek	6 915	5 005	3 355	2 213	1 444	1 409
	Kadın	3 962	2 371	1 326	625	272	154
2012	Erkek	7 274	5 489	3 455	2 301	1 539	1 539
	Kadın	4 320	2 630	1 400	594	290	192
2011	Erkek	6 616	5 283	3 271	2 128	1 501	1 399
	Kadın	3 941	2 465	1 275	563	295	153
2010	Erkek	6 326	4 404	3 089	2 136	1 498	1 552
	Kadın	3 675	2 204	1 250	542	295	179

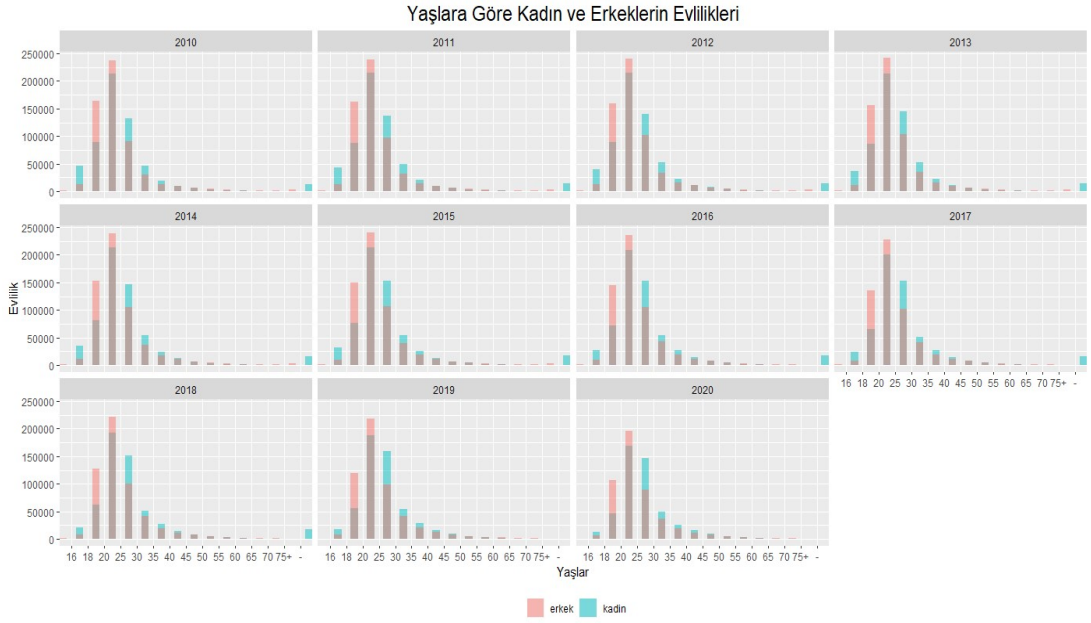
2010 ile 2020 yılları arasında evlilik sayılarına bakıldığında hem kadınlarda hem de erkeklerde evlenme oranlarının artışta olduğu görülmektedir.

Yaşlara göre 2010-2020 yılları arasında evlilik sayıları aşağıdaki grafikte verilmiştir:



Şekil 5.1. Yaşlara göre ortalama evlilik sayısı

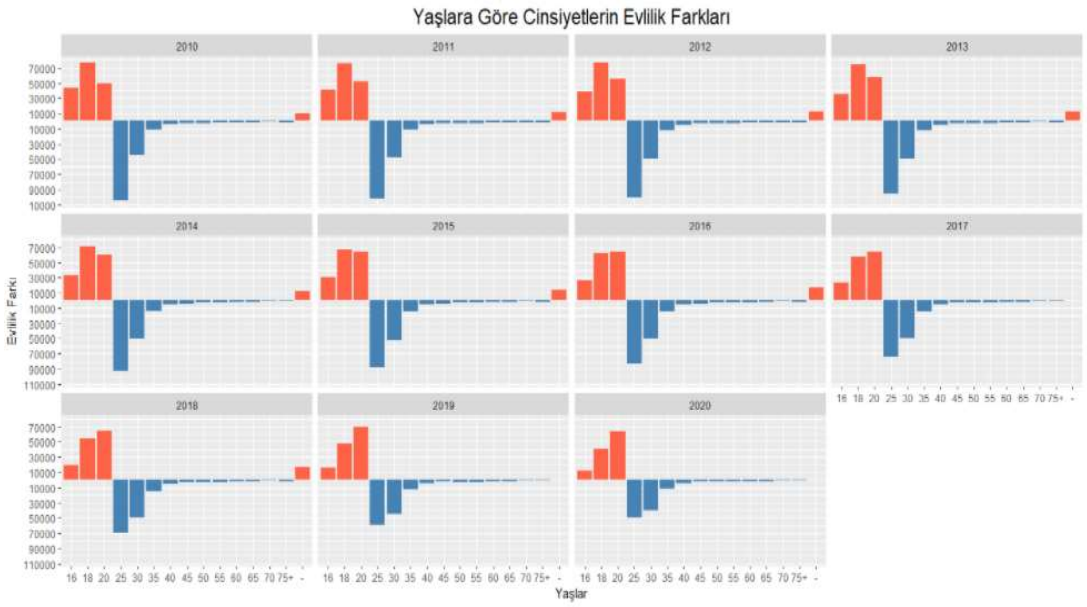
Yaşlara göre ortalama evlilik sayısına bakıldığında en yüksek evlilik oranlarının 25-29 yaş aralığında olduğu görülmektedir. Onu takip eden yaş aralığı ise 20-24'tür. Buradan hareket ile eğitim hayatını bitiren gençlerin kendini evlilik için daha hazır hissettiği, üniversite çağında evlenenlerin de azımsanamayacak bir çoğunlukta olduğu söylenebilir. 45-49 yaş aralığından sonra ise evlenme oranların giderek azaldığı 69 yaş ve sonrasında da sayının oldukça az olduğu görülmektedir.



Şekil 5.2. Yaşlara göre kadın ve erkeklerin evlilik sayıları

2010-2020 yılı aralığında verilere bakıldığında 25-29 yaş ve 30-34 yaş aralıklarında erkeklerin daha fazla evlendiği, 20-24 yaş aralığında da kadınların evliliği daha fazla tercih ettiği görülmektedir. Aynı zamanda 50-54 ve 55-59 yaş aralığında erkeklerin kadınlardan daha fazla evlilik gerçekleştirdiği tespit edilmiştir.

Yukarıdaki verilere bakıldığında erkeklerin büyük bir kısmının eğitim hayatlarını bitirdikten, askerlik yaptıktan ve iş bulduktan sonra evlenmeye açık olduğu, kadınların ise eğitim hayatları sırasında ya da sonrasında hemen evlenmeye meyilli oldukları anlaşılmaktadır. 50-54 yaş ve 55-59 yaş aralığında da daha sık erkeklerin evlilik yapması, genel olarak ikinci bahar olarak adlandırılan ikinci evlilikten kaynaklı olduğu görülmektedir. Emeklilik yaşamına başlayan erkekler, daha rahat bir hayat sürebilmek, yalnız yaşlanmamak adına da bu yaş aralıklarında evliliğe olumlu bakmaktadır.



Şekil 5.3. Yaşlara göre cinsiyetlerin evlilik farkları

Yaşlara göre cinsiyetlerin evlilik farklarına bakıldığı zaman özellikle 2010 yılında 30-34 yaş aralığında evlenen erkek sayısının 2020 yılında düşüşe geçtiği, 2010 yılında 25-29 yaş aralığında evlenen kadınların sayısının da 2020 yılında artış gösterdiği anlaşılmaktadır. Günümüz şartlarında maddi geçim sağlamanın giderek zorlaşması erkeklerin 30-34 yaş aralığında evlilik için hazır olamamasının en önemli nedenlerinden biridir. Kadınların 25-29 yaş aralığında evlilik oranlarının artışının da yine en önemli etkenlerinden biri maddi konum olabilir.

5.2. Evlilik Verileri ile Elde Edilen Sonuçlar

Verilerin düzenlenmesi aşamasında en geniş kütüphanelerden biri olan "Tidyverse" kütüphanesinden ve Excel formatlarını okuyabilmek adına da "readxl" kütüphanesinden yararlanılmıştır. TÜİK üzerinden elde edilen 2010-2020 yılı evlilik verileri "read_excel" fonksiyonu ile okutulmuştur. Okutulmuş olan veri "pipe operatörü" yardımı ile "gather fonksiyonu" içerisinde veri seti olarak kullanılmıştır.

```
library(readxl)


veri <- read_excel("veri_duz.xlsx") %>%
  gather("16-17": "Unknown" , key = "Yas", value = "Evlilik")

veri$cinsiyet <- as.factor(veri$cinsiyet)
levels(veri$cinsiyet) = c("erkek" , "kadin")

veri$Yas <- as.factor(veri$Yas)
levels(veri$Yas) = c("16-17" , "18-19" , "20-24" , "25-29" , "30-34" ,
                     "35-39" , "40-44" , "45-49" , "50-54", "55-59" ,
                     "60-64" , "65-69" , "70-74" , "75+" , "belirtilmemiş")

veri$Evlilik <- suppressWarnings(as.numeric(veri$Evlilik , na.rm = T))

veri$Year <- as.factor(veri$Year)
```

Şekil 5.4. Kütüphanelerin çağrılması ve verinin düzenlenmesi

Düzenlenmiş olan veriler üzerinden ilk Poisson Regresyon modeli oluşturulmuş ve bu regresyon modeli tablo formatına dönüştürülmüştür.

```
Yas_Regresyon <- summary(glm(Evlilik~Yas , veri , family = poisson))

Yas_Regresyon_Tablosu <- data.frame(Yas_Regresyon$coefficients)

colnames(Yas_Regresyon_Tablosu) <- c("Tahmin" , "Standart_Error" , "Z_Degeri" , "Pr(>
|z|)")

print(Yas_Regresyon_Tablosu)
```

Şekil 5.5. Yaşa göre evliliğin poisson regresyonu

Poisson Regresyon uygulaması ile elde edilen tahmin, standart hata, z-değeri ve p-değeri aşağıdaki tabloda verilmiştir.

Çizelge 5.4. Yaşa Göre Evlilik Sayılarının Regresyon Çizelgesi

	Tahmin	**Standart Hata**	**z-değeri**	**p-değeri**
(Intercept)	9.6763987	0.001688796	5729.76312	0.000000
Yaş18-19	0.9696223	0.001983340	488.88368	0.000000
Yaş20-24	2.3879145	0.001764629	1353.21021	0.000000
Yaş25-29	2.4722785	0.001758616	1405.80941	0.000000
Yaş30-34	1.5636055	0.001857205	841.91318	0.000000
Yaş35-39	0.6756715	0.002074425	325.71500	0.000000
Yaş40-44	-0.0287100	0.002405669	-11.93431	0.000000
Yaş45-49	-0.5405713	0.002783724	-194.18999	0.000000
Yaş50-54	-0.9901381	0.003244821	-305.14409	0.000000
Yaş55-59	-1.4085881	0.003810221	-369.68672	0.000000
Yaş60-64	-1.8640941	0.004609619	-404.39224	0.000000
Yaş65-69	-2.3841234	0.005813472	-410.10321	0.000000
Yaş70-74	-2.9234383	0.007477733	-390.95246	0.000000
Yaş75+	-3.0221876	0.007837341	-385.61389	0.000000
Yaş Belirtilmemiş	-0.6602076	0.002987201	-221.01211	0.000000

Yukarıda verilen tabloya bakıldığında p-değerlerinin 0.05'den küçük olduğu görülmektedir. İstatistiksel açıdan bu katsayılar önemli bir katkı sağlamakta, bu nedenle de modele dahil edilmektedir.

Yıllara göre toplam evliliğin Poisson Regresyonunun yapılabilmesi için veriler düzenlenmiş, ikinci regresyon modeli oluşturulmuş ve tablo formatına dönüştürülmüştür.

```
veri_yillik_reg <- read_excel("veri_duz.xlsx")

veri_yillik_reg$cinsiyet <- as.factor(veri_yillik_reg$cinsiyet)
levels(veri_yillik_reg$cinsiyet) = c("erkek" , "kadın")

veri_yillik_reg$Total <- as.numeric(veri_yillik_reg$Total , na.rm = T)

veri_yillik_reg$Year <- as.factor(veri_yillik_reg$Year)


Yillik_Regresyon <- summary(glm(Total~Year , veri_yillik_reg , family = poisson))

Yillik_Regresyon_Tablosu <- data.frame(Yillik_Regresyon$coefficients)

colnames(Yillik_Regresyon_Tablosu) <- c("Tahmin" , "Standart_Error" , "Z_Degeri" , "P
r(>|z|)")

print(Yillik_Regresyon_Tablosu)
```

Şekil 5.6. Yıllara göre evliliğin poisson regresyonu

Yıllara göre evlilik sayılarının regresyon tablosu ise aşağıda verilmiştir.

Çizelge 5.5. Yıllara Göre Evlilik Sayılarının Regresyon Çizelgesi

	Tahmin	**Standart Hata**	**z-değeri**	**p-değeri**
(Intercept)	13.27545350	0.0009263112	14331.52687	0.000000
2011	0.01711668	0.0013044320	13.12194	0.000000
2012	0.03546365	0.0012985409	27.31038	0.000000
2013	0.02946141	0.0013004592	22.65462	0.000000
2014	0.02873798	0.0013006910	22.09440	0.000000
2015	0.03418913	0.0012989475	26.32064	0.000000
2016	0.02001073	0.0013034973	15.35157	0.000000
2017	-0.02301143	0.0013176034	-17.46461	0.000000
2018	-0.04983161	0.0013266296	-37.56256	0.000000
2019	-0.07185305	0.0013341767	-53.85572	0.000000
2020	-0.17887983	0.0013726535	-130.31682	0.000000

Yukarıda verilen tabloya bakıldığında p-değerlerinin 0.05'den küçük olduğu görülmektedir. İstatistiksel açıdan bu katsayılar önemli bir katkı sağlamakta, bu nedenle de modele dahil edilmektedir.

Birleştirilmiş regresyon tablosu kullanılarak üçüncü regresyon modeli oluşturulmuş olup regresyon tablosu ise aşağıda verilmiştir.

```
Birlesik_Regresyon_1 <- summary(glm(Evlilik~Yas+Year , veri , family = poisson))

Birlesik_Regresyon_Tablosu_1 <- data.frame(Birlesik_Regresyon_1$coefficients)

colnames(Birlesik_Regresyon_Tablosu_1) <- c("Tahmin" , "Standart_Error" , "Z_Degeri" , "Pr(>|z|)")

print(Birlesik_Regresyon_Tablosu_1)
```

Şekil 5.7. Evliliğin yaş ve yıllar ile birlikte regresyonu

Yaş ve yıllara göre evlilik sayılarının regresyon tablosu oluşturulmuş, elde edilen veriler tablo halinde aşağıda verilmiştir.

Çizelge 5.6. Evliliğin Yaş ve Yıllar ile Birlikte Regresyonu Çizelgesi

	Tahmin	**Standart Hata**	**z-değeri**	**p-değeri**
(Intercept)	9.6763987	0.001688796	5729.76312	0.000000
Yaş18-19	0.9696223	0.001983340	488.88368	0.000000
Yaş20-24	2.3879145	0.001764629	1353.21021	0.000000
Yaş25-29	2.4722785	0.001758616	1405.80941	0.000000
Yaş30-34	1.5636055	0.001857205	841.91318	0.000000
Yaş35-39	0.6756715	0.002074425	325.71500	0.000000
Yaş40-44	-0.0287100	0.002405669	-11.93431	0.000000
Yaş45-49	-0.5405713	0.002783724	-194.18999	0.000000
Yaş50-54	-0.9901381	0.003244821	-305.14409	0.000000
Yaş55-59	-1.4085881	0.003810221	-369.68672	0.000000
Yaş60-64	-1.8640941	0.004609619	-404.39224	0.000000
Yaş65-69	-2.3841234	0.005813472	-410.10321	0.000000
Yaş70-74	-2.9234383	0.007477733	-390.95246	0.000000
Yaş75+	-3.0221876	0.007837341	-385.61389	0.000000
Yaş Belirtilmemiş	-0.6602076	0.002987201	-221.01211	0.000000
2011	0.01711668	0.0013044320	13.12194	0.000000
2012	0.03546365	0.0012985409	27.31038	0.000000
2013	0.02946141	0.0013004592	22.65462	0.000000
2014	0.02873798	0.0013006910	22.09440	0.000000
2015	0.03418913	0.0012989475	26.32064	0.000000
2016	0.02001073	0.0013034973	15.35157	0.000000
2017	-0.02301143	0.0013176034	-17.46461	0.000000
2018	-0.04983161	0.0013266296	-37.56256	0.000000
2019	-0.07185305	0.0013341767	-53.85572	0.000000
2020	-0.17887983	0.0013726535	-130.31682	0.000000

Yukarıdaki tabloda yaşa göre evlilik oranlarında anlamlı bir fark bulunmuştur. P değerinin 0,05'ten küçük olması yıllar arasında anlamlı bir farklılık olduğunu göstermektedir.

Birleştirilmiş regresyon tablosu kullanılarak dördüncü regresyon modeli oluşturulmuş olup regresyon tablosu ise aşağıda verilmiştir.

```
Birlesik_Regresyon_2 <- summary(glm(Evlilik~cinsiyet+Yas , veri , family = poisson))

Birlesik_Regresyon_Tablosu_2 <- data.frame(Birlesik_Regresyon_2$coefficients)

colnames(Birlesik_Regresyon_Tablosu_2) <- c("Tahmin" , "Standart_Error" , "Z_Degeri" , "Pr(>|z|)")

print(Birlesik_Regresyon_Tablosu_2)
```

Şekil 5.8. Evliliğin yaş ve cinsiyet ile birlikte regresyonu

Yaş ve cinsiyete göre evlilik sayılarının regresyon tablosu oluşturulmuş, elde edilen veriler tablo halinde aşağıda verilmiştir.

Çizelge 5.7. Evliliğin Yaş ve Cinsiyet ile Birlikte Regresyonu Çizelgesi

	Tahmin	**Standart Hata**	**z-değeri**	**p-değeri**
(Intercept)	9.67769905	0.001711941	5653.0559	0.000000
Erkek	-0.0023420	0.000515236	-4.245255	0.025452
Kadın	-0.0026023	0.000562135	-4.629394	0.036673
Yaş18-19	0.9696223	0.001983340	488.88368	0.000000
Yaş20-24	2.3879145	0.001764629	1353.21021	0.000000
Yaş25-29	2.4722785	0.001758616	1405.80941	0.000000
Yaş30-34	1.5636055	0.001857205	841.91318	0.000000
Yaş35-39	0.6756715	0.002074425	325.71500	0.000000
Yaş40-44	-0.0287100	0.002405669	-11.93431	0.000000
Yaş45-49	-0.5405713	0.002783724	-194.18999	0.000000
Yaş50-54	-0.9901381	0.003244821	-305.14409	0.000000
Yaş55-59	-1.4085881	0.003810221	-369.68672	0.000000
Yaş60-64	-1.8640941	0.004609619	-404.39224	0.000000
Yaş65-69	-2.3841234	0.005813472	-410.10321	0.000000
Yaş70-74	-2.9234383	0.007477733	-390.95246	0.000000
Yaş75+	-3.0221876	0.007837341	-385.61389	0.000000
Yaş Belirtilmemiş	-0.6602076	0.002987201	-221.01211	0.000000

Yukarıdaki tabloda yaşa göre cinsiyetler arasındaki evlilik oranlarında anlamlı bir fark bulunmuştur. P değerinin 0,05'ten küçük olması yaşlar ve cinsiyetler arasında anlamlı bir farklılık olduğunu göstermektedir.

Birleştirilmiş regresyon tablosu kullanılarak beşinci regresyon modeli oluşturulmuş olup regresyon tablosu ise aşağıda verilmiştir.

```
Birlesik_Regresyon_3 <- summary(glm(Evlilik~cinsiyet+Year , veri , family = poisson))

Birlesik_Regresyon_Tablosu_3 <- data.frame(Birlesik_Regresyon_2$coefficients)

colnames(Birlesik_Regresyon_Tablosu_3) <- c("Tahmin" , "Standart_Error" , "Z_Degeri" , "Pr(>|z|)")

print(Birlesik_Regresyon_Tablosu_3)
```

Şekil 5.9. Evliliğin cinsiyet ve yıllar ile birlikte regresyonu

Cinsiyete ve yıllara göre evlilik sayılarının regresyon tablosu oluşturulmuş, elde edilen veriler tablo halinde aşağıda verilmiştir.

Çizelge 5.8. Evliliğin Cinsiyet ve Yıllar ile Birlikte Regresyonu Çizelgesi

	Tahmin	**Standart Hata**	**z-değeri**	**p-değeri**
(Intercept)	10.5831442	0.000096754	10927.794	0.000000
Erkek	-0.00966412	0.000441025	-18.689522	0.000000
Kadın	-0.01151514	0.000562145	-20.484254	0.000000
2011	0.01711668	0.0013044320	13.12194	0.000000
2012	0.03546365	0.0012985409	27.31038	0.000000
2013	0.02946141	0.0013004592	22.65462	0.000000
2014	0.02873798	0.0013006910	22.09440	0.000000
2015	0.03418913	0.0012989475	26.32064	0.000000
2016	0.02001073	0.0013034973	15.35157	0.000000
2017	-0.02301143	0.0013176034	-17.46461	0.000000
2018	-0.04983161	0.0013266296	-37.56256	0.000000
2019	-0.07185305	0.0013341767	-53.85572	0.000000
2020	-0.17887983	0.0013726535	-130.31682	0.000000

Yukarıdaki tabloda yıllara göre cinsiyetler arasındaki evlilik oranlarında anlamlı bir fark bulunmuştur. P değerinin 0,05'ten küçük olması yıllar ve cinsiyetler arasında anlamlı bir farklılık olduğunu göstermektedir.

Birleştirilmiş regresyon tablosu kullanılarak beşinci regresyon modeli oluşturulmuş olup regresyon tablosu ise aşağıda verilmiştir.

```
Birlesik_Regresyon_4 <- summary(glm(Evlilik~cinsiyet+Yas+Year , veri , family = poisson))

Birlesik_Regresyon_Tablosu_4 <- data.frame(Birlesik_Regresyon_4$coefficients)

colnames(Birlesik_Regresyon_Tablosu_4) <- c("Tahmin" , "Standart_Error" , "Z_Degeri" , "Pr(>|z|)")

print(Birlesik_Regresyon_Tablosu_4)
```

Şekil 5.10. Evliliğin cinsiyet, yaş ve yıllar ile birlikte regresyonu

Cinsiyete, yaş ve yıllara göre evlilik sayılarının regresyon tablosu oluşturulmuş, elde edilen veriler tablo halinde aşağıda verilmiştir.

Çizelge 5.9. Evliliğin Cinsiyet, Yaş ve Yıllar ile Birlikte Regresyonu Çizelgesi

	Tahmin	**Standart Hata**	**z-değeri**	**p-değeri**
(Intercept)	9.6763987	0.001688796	5729.76312	0.000000
Erkek	-0.0020147	0.000441025	-18.689522	0.019554
Kadın	-0.0023853	0.000562145	-20.484254	0.022015
Yaş18-19	0.9696223	0.001983340	488.88368	0.000000
Yaş20-24	2.3879145	0.001764629	1353.21021	0.000000
Yaş25-29	2.4722785	0.001758616	1405.80941	0.000000
Yaş30-34	1.5636055	0.001857205	841.91318	0.000000
Yaş35-39	0.6756715	0.002074425	325.71500	0.000000
Yaş40-44	-0.0287100	0.002405669	-11.93431	0.000000
Yaş45-49	-0.5405713	0.002783724	-194.18999	0.000000
Yaş50-54	-0.9901381	0.003244821	-305.14409	0.000000
Yaş55-59	-1.4085881	0.003810221	-369.68672	0.000000
Yaş60-64	-1.8640941	0.004609619	-404.39224	0.000000
Yaş65-69	-2.3841234	0.005813472	-410.10321	0.000000
Yaş70-74	-2.9234383	0.007477733	-390.95246	0.000000
Yaş75+	-3.0221876	0.007837341	-385.61389	0.000000
Yaş Belirtilmemiş	-0.6602076	0.002987201	-221.01211	0.000000
2011	0.01711668	0.0013044320	13.12194	0.000000
2012	0.03546365	0.0012985409	27.31038	0.000000
2013	0.02946141	0.0013004592	22.65462	0.000000
2014	0.02873798	0.0013006910	22.09440	0.000000
2015	0.03418913	0.0012989475	26.32064	0.000000
2016	0.02001073	0.0013034973	15.35157	0.000000
2017	-0.02301143	0.0013176034	-17.46461	0.000000
2018	-0.04983161	0.0013266296	-37.56256	0.000000
2019	-0.07185305	0.0013341767	-53.85572	0.000000
2020	-0.17887983	0.0013726535	-130.31682	0.000000

Yukarıdaki tabloda yaşlar, yıllar ve cinsiyetler arasındaki evlilik oranlarında anlamlı bir fark bulunmuştur. P değerinin 0,05'ten küçük olması yıllar ve cinsiyetler arasında anlamlı bir farklılık olduğunu göstermektedir.

2010-2020 yılları arasındaki evlilik sayılarının poisson regresyon analizine uygun olup olmadığını anlamak için Pearson ki-kare testi kullanılmıştır. Evlilik bağımlı değişkeninin, poisson dağılımına uygunluğu pearson ki-kare istatistiğiyle kontrol edilmiştir. Pearson ki-kare testi bağımlı değişkende aşırı yayılım olup olmadığını ölçmede kullanılan testlerden biri olup, test sonucunda çıkan p-değerinin 0,05'ten küçük olduğu durumlarda aşırı yayılım olduğu söylenir. Buna göre hipotezler aşağıdaki gibi kurulmuştur.

H_0 = Veriler poisson dağılımına uygunluk göstermektedir.
H_1= Veriler poisson dağılımına uygunluk göstermemektedir.

Pearson Ki-Kare uyum iyiliği tablosu oluşturulmuş, elde edilen veriler tablo halinde aşağıda verilmiştir.

Çizelge 5.10. Pearson Ki-Kare Uyum İyiliği Çizelgesi

Yıllar	Evlilik Sayısı	Beklenen Yıllar	Beklenen Evlilik Sayıları	İstatistik Değeri	P-Değeri
2020	487.270	1.957	559.883	15,9540	0,1009
2019	542.314	1.991	569.679	15,9540	0,1009
2018	554.389	2.000	572.142	15,9540	0,1009
2017	569.459	2.011	575.272	15,9540	0,1009
2016	594.493	2.029	580.617	15,9540	0,1009
2015	602.982	2.035	582.283	15,9540	0,1009
2014	599.704	2.032	581.331	15,9540	0,1009
2013	600.138	2.031	581.205	15,9540	0,1009
2012	603.751	2.033	581.787	15,9540	0,1009
2011	592.775	2.024	579.123	15,9540	0,1009

2010	582.715	2.016	576.663	15,9540	0,1009

Pearson Ki-kare istatistiği testinin sonucu Tablo 5.7'de de gösterildiği üzere;

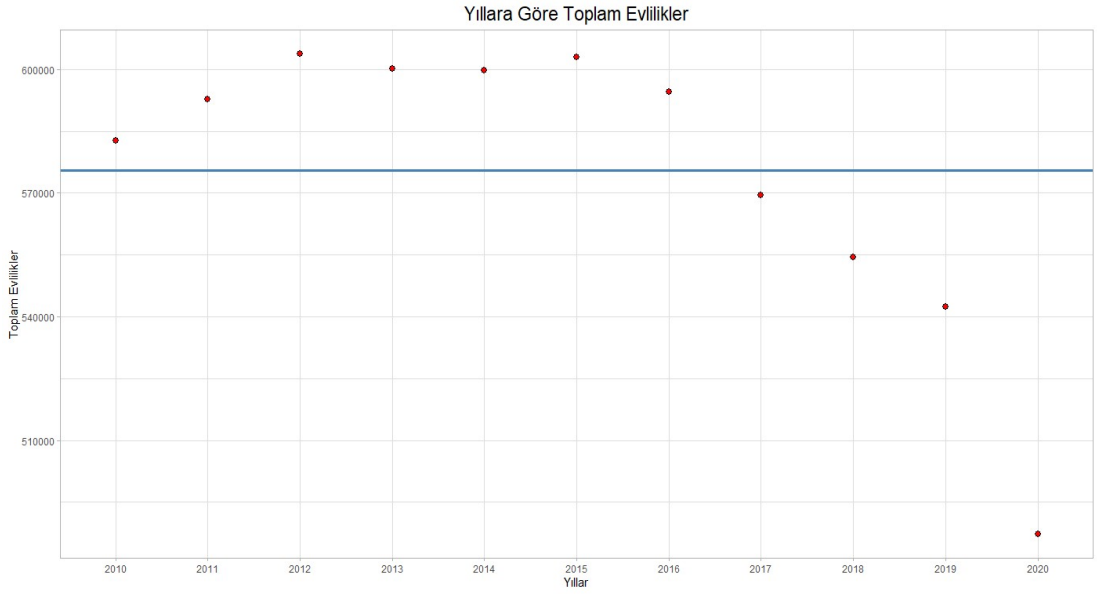
P-değeri 0,101 şeklinde bulunmuştur ve p-değerinin 0,05'ten büyük olduğu görülmüştür. Bu durumda P>0,05 olduğundan H_0 hipotezi kabul edilir ve verilerin poisson dağıldığı görülmektedir.

Adebowale vd. (2012) tarafından yapılan araştırmada Nijerya'da ilk evlilik süresini etkileyen en önemli unsurlardan birinin bölgesel farklılıklar olduğu tespit edilmiştir. Özellikle kırsal alanda yaşayan ve dini baskı hisseden kişilerin daha genç yaşta evlilik birliği kurduğu görülmektedir.

Sanizah vd. (2014) tarafından yapılan araştırmaya bakıldığında Malezya'da erkeklerin erken yaşta evlenmesinin boşanmalara neden olduğu görülmüştür. Özellikle evlilik danışmanlığından yardım almanın ise boşanma oranlarını azalttığı tespit edilmiştir. Yaptığımız incelemeye bakıldığında kadınların 50 yaş sonrasında evlilik oranlarının artmış olması ile erken yaşta evliliğin boşanma getirmesi arasında bağlantı kurulabilir. Erken yaşta evlilik sonucu evlilik birliğinin bozulması ile kadınlar ikinci bahar adı altında 50 yaşından sonra yeniden evlilik birliği içerisine girmek istemektedir.

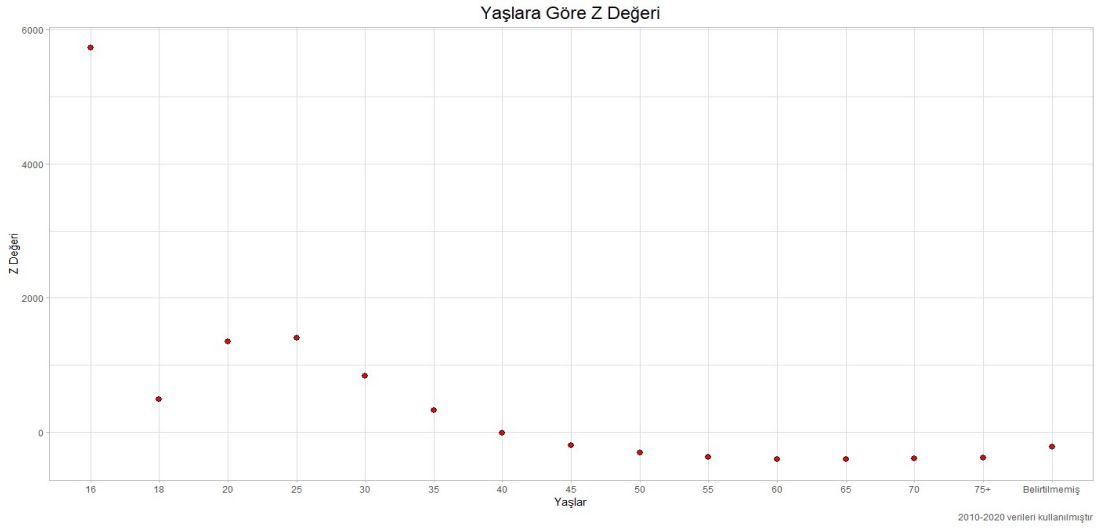
Türkiye'de yapılan araştırmalara bakıldığında Eryavuz ve Birecikli (2018) çalışmasında boşanmayı etkileyen unsurları ele almıştır. Buradaki incelemeler sonucunda erkek ile kadınların yaşlarındaki farkın, erkeğin eğitim durumunun boşanmayı etkilediği görülmüştür. Yaptığımız çalışmada kadınların en çok evlendiği yaş aralığı ile erkeklerin en çok evlendiği yaş aralığının farklı olduğu tespit edilmiştir. Bu da kadın-erkek arasında belirli bir yaş farkının bulunduğunu, temelde evliliklerin sağlamlık kazanması adına bunun gerekli olduğu görüşü ileri sürülebilir.

Yapılan Poisson Regresyon analizine göre yıllara göre gerçekleştirilen toplam evlilik sayıları aşağıda verilmiştir.



Şekil 5.11. Yıllara göre toplam evlilikler

Yaşlara göre Poisson Regresyondan elde edilen Z değeri ise şekilde verilmiştir.



Şekil 5.12. Yaşlara göre z değeri

Evlilikler genellikle 40 yaş ve öncesinde meydana geldiği için 40 yaş ve öncesinde z değerinin pozitif olarak bulunduğu ve ortalamadan pozitif saptığı görülmektedir.

Evlilik yaşlarının ve yıllarının ele alındığı bu çalışma boşanma oranlarını yorumlamak için de avantaj sağlamaktadır. Toplumsal değerlerin değişmesi ile birlikte genç nesil yeni fikirlere açık hale gelirken yaşlı yetişkinler geleneksel değerlerine bağlı kalmaktadır (Thornton, 1985: 859). Bu nedenle boşanma ile ilgili araştırma yaparken

evlenme yaşından feyz almak önemlidir. Evlenme yaşının boşanma üzerinde negatif bir etkisinin olması nedeni ile evlilik yaşı ile ilgili pek çok çalışma yapılmıştır (Bumpass ve Sweet, 1972; Booth ve Edwards, 1985; Sanizah vd., 2014). Evlilik yaşının düşmesi ile birlikte boşanma eğiliminde bir artış görülmektedir. Yapılan çalışmalara göre kadınların erken evlenme yaşı ile erkeklerin yaşı genç olan kadınlarla ile evlenmesi evliliklerin sürdürülebilirliğini olumsuz yönde etkilemektedir. Çalışmamız kapsamında da evlilik yaşının kadınlarda erkeklerden daha az olduğu saptanmıştır. Bu bağlamda Türkiye özelinde erken evlenme yaşının cinsiyet temelli bir özellik taşıdığı söylenebilir.

6. SONUÇ VE ÖNERİLER

Bu çalışma kapsamında Poisson regresyon ile uygulamalı bir çalışmanın yürütülmesi amaçlanmıştır. Bu doğrultuda Poisson regresyon hakkında literatür taramları üç bölüm altında değerlendirilmiştir. Buna göre Poisson regresyonun kullanım alanlarına bakılması gerekliliği ön plana çıkmaktadır. Bağımlı değişkenin sayım verisi bulunduğu takdirde Poisson regresyonun kullanıldığı görülmüştür. Buradan hareketle Poisson regresyon, bağımlı değişkenin sayım verisi olduğunda kullanılmaktadır. Yapılan işlem sürecinde bağımsız değişkenlerin diğer önemli olan bağımlı değişkenin beklenmekte olan değerine bağlı bir fonksiyon şeklinde regresyon fonksiyonu ortaya çıkmaktadır. Uygulama süreçlerinde diğer bir teorik ifade ise genel olarak belirlenmiş olan kategorilerin sayımlar sonucunda meydana gelen verilerle yapılıyor olmasıdır. Dolayısı ile Poisson regresyon uygulamalarında model içindeki ortalama ve varyansın eşit olmaları sağlanmaktadır. Bu uygulama süreçlerinde aşırı yayılımlarıyla karşılaşıldığı görülmüştür. Bu tür durumlar da ise negatif binom regresyon uygulamalarının yapılması gerekliliğinin olduğu sonucuna varılmıştır.

Çalışmada bağımlı değişkenin poisson regresyona uygunluğu pearson ki-kare testi yapılarak kontrol edilmiş olup, testin sonucunda verilerin poisson dağıldığı görülmüştür.

Yapılan bu çalışma kapsamında, uygulama sürecinde altı farklı regresyon modeli kurulmuş olup, regresyon modelleri oluştururken ($p<0,05$) tüm değişkenlerin model için uygun olduğu, bu nedenle değerlendirme yaparken herhangi bir değişkenin analizden çıkarılmaya ihtiyacı olmadığı belirlenmiştir. Bu bağlamda tüm bağımsız değişkenlerin evlilik sayısı üzerinde etkisi olduğu görülebilir. Tüm değişkenler anlamlı bulunduğu için kurulan modellerden tüm değişkenleri içeren modelin en anlamlı model olduğu söylenebilir.

Poisson regresyon analizi sonuçlarına bakıldığında 25-29 yaş aralığında evlilik oranının erkeklerde daha yüksek olduğu, 20-24 yaş aralığında evlenen kadın sayısının da yüksek olduğu anlaşılmaktadır. Eğitim hayatını bitiren kadınların kendilerini evlilik için hazır hissettikleri, okul bittikten sonra hemen aile kurma telaşına girdikleri

söylenebilir. 45-49 yaş aralığından sonra evlilik oranlarının düşmesi hem bekar sayısının azlığı hem de bekarlığa alışma durumu ile özetlenebilir.

Erkeklerin eğitim hayatını bitirdikten sonra hemen evliliğe hazır olmadığı, askerlik ve iş telaşına düştükleri, evliliği daha sonraki bir zamana bıraktığı tespit edilmiştir. 50-54 yaş ile 55-59 yaş aralığında yapılan evliliklerin ise ikinci bahar olarak adlandırılan süreç olduğu anlaşılmaktadır.

Kadınların evliliğe hazır hissettikleri yaş ile erkeklerin evliliğe hazır hissettikleri yaş farklıdır. Bu durum evliliklerde yaş farkının ortaya çıktığını, evliliklerin sağlamlığı aşamasında yaş farkının önemli bir unsur olduğunu söylemek mümkündür.

Yıllara göre evlenme oranlarına bakıldığında 2010 yılındaki evlilik oranları ile 2020 yılındaki evlilik oranlar arasında farklılık bulunmaktadır. Evlilik sayılarında 2015 yılından itibaren günümüze kadar olan sürece baktığımızda ciddi sayıda diyebileceğimiz düşüşler olduğu da dikkat çekmektedir.

KAYNAKLAR

Adebowale, S. A., Fagbamigbe, F. A., Okareh, O. T., Lawal, O. G., 2012. Survival Analysis of Timing of First Marriage Among Women of Reproductive Age in Nigeria: Regional Differences, African Journal of Reproductive Health, 16(4), ss. 95-107.

Agresti, A., 2015. Foundations of Linear and Generalized Linear Models. John Wiley & Sons.

Alkhamisi, M. A., Shukur, G., 2008. Developing Ridge Parameters for SUR Model. Communications in Statistics-Theory and Methods, 37(4), 544-564.

Alkhamisi, M. A., Khalaf, G., Shukur, G., 2006. Some Modifications for Choosing Ridge Parameters. Communications in Statistics-Theory and Methods, 35(11), 2005-2020.

Arıcan, E., 2010. Nitel Yanıt Değişkene Sahip Regresyon Modellerin Tahmin Yöntemleri, Yüksek Lisans Tezi, Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Adana.

Arslan, Y., Altaş, D., 2019. Genelleştirilmiş Doğrusal Modeller İle Sigorta Şirketlerinde Hasar Rezervinin Tahmini. Finansal Araştırmalar Ve Çalışmalar Dergisi, 11(21), 185-196.

Aydemir, E., 2011. Evlilik mi Evcilik mi? Erken ve Zorla Evlilikler: Çocuk Gelinler. USAK.

Batah, F. S. M., Ramanathan, T.V., Gore, S.D., 2008. The Efficiency of Modifed Jackknife and Ridge Type Regression Rstimators: A Comparison. Surv. Math., pp. 111–122.

Booth, A., Edwards, J. N., 1985. Age at Marriage and Marital Instability, Journal of Marriage and Family, 47(1), ss. 67-75.

Breslow, N. E., 1984. Extra-Poisson Variation in Log-linear Models. Journal of the Royal Statistical Society: Series C (Applied Statistics), 33(1), 38-44.

Bumpass, L. L., Sweet, J. A., 1972. Differentails in Marital Instability: 1970, American Sociological Review, 37(6), ss. 754-766.

Cameron, A. C., Trivedi, P. K., 2013. Regression Analysis of Count Data (Vol. 53).Cambridge University Press, 598p.

Consul, P., Famoye, F., 1992. Generalized Poisson Regression Model. Communications in Statistics-Theory and Methods, 21(1), 89-109.

Cragg, J. G., 1971. Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. Econometrica (pre-1986), 39(5), 829.

Çelik, B., Avci, E., 2019. Conway-Maxwell-Poisson (COM-Poisson) Regresyon Modeli: Aşırı ve Az Yayılımlı Verilere Uygulanması. Türkiye Klinikleri Biyoistatistik, 11(3), 213-223.

Çodur, M., Tortum, A., Çodur, M., 2013. Genelleştirilmiş Lineer Regresyon ile Erzurum Kuzey Çevre Yolu Kaza Tahmin Modeli . Journal of the Institute of Science and Technology , 3 (1) , 79-84 .

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

De Jong, P., Heller Z. H., 2008. Generalized Linear Models for Insurance Data. Cambridge University Press.

Erdemir, Ö. K., Sucu, M., 2016. Genelleştirilmiş Doğrusal Model Bileşenlerinin Kredibilite Üzerindeki Etkilerinin Araştırılması. İstatistikçiler Dergisi: İstatistik ve Aktüerya, 9(1), 37-45.

Eryavuz, A. K., Birecikli, Ş. Ü., 2018. Türkiye'de Kadının Boşanması: Yaşam Analizi Modelleri. Çalışma ve Toplum, 3, 1227-1251.

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

Hinkley, D. V., 1977. Jackknifing in Unbalanced Situations. Technometrics, 19(3), 285-292.

Hoerl, A. E., Kennard, R. W., 1970a. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 12(1), 55-67.

Hoerl, A. E., Kennard, R. W., 1970b. Ridge regression: Applications to Nonorthogonal Problems. Technometrics, 12(1), 69-82.

Jensen, R.A., Thornton, R., 2003. Early Female Marriage in the Developing World. Gender and Development, 11(2), 9-19.

Karadağ, Ö., 2014. Genelleştirilmiş Doğrusal Modeller için Sınırlı Dalgalanmalı Kredibilite Yaklaşımı, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

Kibria, B. G., 2003. Performance of Some New Ridge Regression Estimators. Communications in Statistics-Simulation and Computation, 32(2), 419- 435.

Kibria, B. G., Månsson, K., Shukur, G., 2011. A Ridge Regression Estimator For The Zero-İnflated Poisson Model. Royal Institute of Technology, CESISCentre of Excellence for Science and Innovation Studies, Stockholm.

Kibria, B. G., Månsson, K., Shukur, G., 2012. Performance of Some Logistic Ridge Regression Estimators. Computational economics, 40(4), 401-414.

Kibria, B. G., Månsson, K., Shukur, G., 2013. Some Ridge Regression Estimators For The Zero-İnflated Poisson Model. Journal of Applied Statistics, 40(4), 721-735.

Köleoğlu, N., 2006. Olay Zamanı Analizinde Tesadüfi Etkiler Poisson Regresyon Modeli ile Gözlemlenemeyen Heterojenliğin İncelenmesi, Doktora Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.

Månsson, K., Shukur, G., 2011. A Poisson Ridge Regression Estimator. Economic Modelling, 28(4), 1475-1481.

Månsson, K., Kibria, B. G., Sjolander, P., Shukur, G., 2012. Improved Liu Estimators For The Poisson Regression Model. International Journal of Statistics and Probability, 1(1), 2.

Marquardt, D. W., Snee, R. D., 1975. Ridge Regression in Practice. The American Statistician, 29(1), 3-20.

Mouatassim, Y., Ezzahid, E. H., 2008. Poisson Regression And Zero-inflated Poissonregression: Application To Private Health İnsurance Data. European Actuarial Journal 2(2)

Mullahy, J., 1986. Specification and Testing of Some Modified Count Data Models. Journal of Econometrics, 33(3), 341-365.

Muniz, G., Kibria, B. G., 2009. On Some Ridge Regression Estimators: An Empirical Comparisons. Communications in Statistics - Simulation and Computation, 38(3), 621-630.

Toutounian, F., Nasabzadeh, H., 2014. A New Method Based on Generalized Taylor Expansion For Computing A Series Solution Of The Linear Systems. Applied Mathematics and Computation, 248, 602-609.

Nelder, J. A., Mead, R., 1965. A Simplex Method For Function Minimization. The Computer Journal, 7(4), 377-394.

OECD, 2019. Erişim Tarihi 31.03.2021 https://www.oecd.org

Ölmez, Hosta, H., Ata, Tutkun, N., 2018. Kesikli Yaşam Süresi Modelleri: Evlilik Süreleri Üzerine Bir Uygulama. Alphanumeric Journal, 7(4), 308-313.

Rao, C., R., Toutenburg, H. 1999. Linear Models: Least Squares and Alternatives, 2.Baskı, Springer, Germany.

Saffari, S. E., Adnan, R., Greene, W., 2012. Hurdle Negative Binomial Regression Model with Right Censored Count Data. SORT: Statistics and Operations Research Transactions, 36(2), 0181-194.

Sanizah, A., Hasfariza, F., Rahayu, S. N., Naslina, A. N. N., 2014. Determinants of Marital Dissolution: A Survival Analysis Approach, International Journal of Economics and Statistics, 2, 348-354.

Sezgin, F., Deniz, E., 2004. Genelleştirilmiş Doğrusal Modeller ile Sigorta Şirketlerinde Hasar Rezervinin Tahmini. İstanbul Üniversitesi İşletme Fakültesi İşletme İktisadı Enstitüsü Dergisi, Haziran 2004, Issue 48, p. 17-25.

Singh, B., Chaubey, Y. P., Dwivedi, T. D., 1986. An Almost Unbiased Ridge Estimator. Sankhyā: The Indian Journal of Statistics, Series B, 342-346.

Thornton, A., 1985. Changing Attitudes Towards Separation and Divorce: Causes and Consequences, American Journal of Sociology, 90(4), 856-872.

Türkan, S., Özel, G., 2016. A New Modified Jackknifed Estimator For The Poisson Regression Model. Journal of Applied Statistics, 43(10), 1892-1905.

Türkiye İstatistik Kurumu, 2021. Erişim Tarihi 31.03.2021 https://www.tuik.gov.tr

UNICEF, 2013. Dünya Çocuklarının Durumu 2013 Raporu

Winkelmann, R., 2008. Econometric Analysis of Count Data. Springer Science & Business Media, 320p.

Yeşilova, A., Erdinç, S., Ser, G., 2017. Van Gölü Sahil Şeridindeki Zooplankton Populasyon Yoğunluğu Değişiminin Doğrusal Olmayan Regresyon Yöntemleri Kullanılarak İncelenmesi. Yüzüncü Yıl Üniversitesi Tarım Bilimleri Dergisi, 27(1), 58-64.

Yeşilyurt, H., 2014. Poisson Regresyon Modeli ve Türkiye'deki Boşanma İstatistiklerine Uygulanması, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

Zhou, M., Li, L., Dunson, D., Carin, L., 2012. Lognormal And Gamma Mixed Negative Binomial Regression. In Proceedings of the International Conference on Machine Learning.International Conference on Machine Learning (Vol. 2012, p. 1343).NIH Public Access.

ÖZGEÇMİŞ

Adı Soyadı : Ozan Altunbaş

Eğitim Durumu

Lise : ECA Elginkan Anadolu Lisesi, 2009

Lisans : İstanbul Ticaret Üniversitesi,
Fen Edebiyat Fakültesi, İstatistik Bölümü

Yüksek Lisans : İstanbul Ticaret Üniversitesi,
Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı

Yayınları

Altunbaş, O., Deniz Başar, Ö., 2022. Poisson Regresyon ve Evlilik İstatistiklerine Uygulanması. İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, Basımda.