

T.C.
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ZOOTEKNİ ANABİLİM DALI

**DOĞRUSAL OLMAYAN REGRESYONDA BAĞIMLI DEĞİŞKENDEKİ
FARKLI SIFIR GÖZLEM ORANLARINA GÖRE UYUM ÖLÇÜTLERİ
KULLANILARAK UYGUN MODELİN BELİRLENMESİ**

YÜKSEK LİSANS TEZİ

HAZIRLAYAN: Şerife ÇARDAK
DANIŞMAN: Prof. Dr. Abdullah YEŞİLOVA

VAN - 2022

T.C.
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ZOOTEKNİ ANABİLİM DALI

**DOĞRUSAL OLMAYAN REGRESYONDA BAĞIMLI DEĞİŞKENDEKİ
FARKLI SIFIR GÖZLEM ORANLARINA GÖRE UYUM ÖLÇÜTLERİ
KULLANILARAK UYGUN MODELİN BELİRLENMESİ**

YÜKSEK LİSANS TEZİ

HAZIRLAYAN: Şerife ÇARDAK

VAN-2022

KABUL VE ONAY SAYFASI

Biyometri ve Genetik Anabilim Dalı'nda Prof. Dr. Abdullah YEŞİLOVA danışmanlığında, Şerife ÇARDAK tarafından sunulan “Doğrusal olmayan regresyonda bağımlı değişkendeki farklı sıfır gözlem oranlarına göre uyum ölçütleri kullanılarak uygun modelin belirlenmesi (Determining the appropriate model by using fit criteria according to different zero value ratios in the dependent variable for nonlinear regression)” isimli bu çalışma Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili hükümleri gereğince/...../..... tarihinde aşağıdaki jüri tarafından oy birliği / oy çokluğu ile başarılı bulunmuş ve Yüksek Lisans tezi olarak kabul edilmiştir.

Başkan:..... İmza:

.

Üye:..... İmza:

Üye:..... İmza:

Üye:..... İmza:

Üye:..... İmza:

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun/..../..... tarih ve sayılı kararı ile onaylanmıştır.

İmza

.....

Enstitü Müdürü

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

Şerife ÇARDAK

ÖZET

DOĞRUSAL OLMAYAN REGRESYONDA BAĞIMLI DEĞİŞKENDEKİ FARKLI SIFIR GÖZLEM ORANLARINA GÖRE UYUM ÖLÇÜTLERİ KULLANILARAK UYGUN MODELİN BELİRLENMESİ

ÇARDAK, Şerife

Yüksek Lisans Tezi, Zootečni Anabilim Dalı

Tez Danışmanı: Prof. Dr. Abdullah YEŞİLOVA

Temmuz 2022, 45 sayfa

Bu tez çalışmasında, sayıma dayalı olarak elde edilen bağımlı değişkendeki sıfır değerlerinin %5, %25, %50, %75 ve %100 oranlarında silinmesi sonucunda, model uyum ölçütlerine bağlı olarak, hangi doğrusal olmayan regresyon modelinin uygun olacağının belirlenmesi amaçlanmıştır. Ayrıca her bir farklı oranlarda kesilmiş sıfır gözlemler için oluşturulan regresyon modellerin birbirleriyle olan etkinlikleri ele alınmıştır. Sıfır gözlemlerin tamamı modeldeyken ve sıfır gözlemlerin %50'sine kadar silindiğinde, sıfır yayımlı hurdle negatif binomial regresyon en iyi model olarak seçilmiştir. Sıfır gözlemlerin tamamı silindiğinde ise sıfır değer kesilmiş negatif binomial regresyon en iyi model seçilmiştir. Genellikle negatif binomial esaslı regresyonlar en iyi model olarak seçilmiştir. Çünkü Poisson regresyon ve türevleri aşırı yayılımdan olumsuz anlamda daha fazla etkilenmektedirler. Sonuç olarak, bağımlı değişkende var olan sıfır gözlemlerin, sayısal oranlarına göre uygun model belirlenmiştir.

Anahtar Kelimeler: Aşırı yayılım, Doğrusal olmayan regresyon yöntemleri, Model uyum ölçütleri, Sıfır gözlem.

ABSTRACT

DETERMINING THE APPROPRIATE MODEL BY USING FIT CRITERIA ACCORDING TO DIFFERENT ZERO VALUE RATIOS IN THE DEPENDENT VARIABLE FOR NONLINEAR REGRESSION

ÇARDAK, Şerife

M.Sc.Thesis, Department of Animal Science

Supervisor: Prof. Dr. Abdullah YEŞİLOVA

July 2022, 45 pages

In this study, it was aimed to determine which nonlinear regression model would be appropriate, depending on the model fit criteria, as a result of the deletion of the zero values in the dependent variable obtained based on counting at the rates of 5%, 25%, 50%, 75% and 100%. Zero-inflated hurdle negative binomial regression was chosen as the best model when all of the zero observations were in the model and up to 50% of the zero observations were deleted. When all zero observations were deleted, the truncated negative binomial regression was chosen as the best model. Generally, negative binomial-based regressions were chosen as the best model. Because Poisson regression and its derivatives are more negatively affected by overdispersion. As a result, the appropriate model was determined according to the numerical ratios of the zero observations in the dependent variable.

Keywords: Fit criteria, Nonlinear regression models, Overdispersion, Zero value.



ÖN SÖZ

Bu tez çalışmasında, her türlü ilgi ve yardımlarını esirgemeyen, yol gösterici ve yapıcı eleştirileri, hassas ve titiz çalışmaları ile yardımlarını esirgemeyen danışmanım Sayın Prof. Dr. Abdullah YEŞİLOVA'ya ve ayrıca birçok konuda yardım ve desteklerini esirgemeyen Sayın Dr. Öğr. Üyesi Hasan ÇELİKYÜREK'e teşekkürlerimi sunarım.

2022

Şerife ÇARDAK



İÇİNDEKİLER

	Sayfa
ÖZET	i
ABSTRACT	iii
ÖN SÖZ.....	v
İÇİNDEKİLER.....	vii
ÇİZELGELER LİSTESİ	ix
ŞEKİLLER LİSTESİ.....	xi
SİMGELER VE KISALTMALAR	xiii
1. GİRİŞ.....	1
2. KAYNAK BİLDİRİŞLERİ	3
3. MATERYAL VE YÖNTEM.....	13
3.1. Materyal	13
3.2. Yöntem.....	13
3.2.1. Poisson regresyonu	13
3.2.2. Negatif binomial regresyonu	14
3.2.3. Kesilmiş modeller	14
3.2.4. Sıfır değer kesilmiş Poisson regresyonu	15
3.2.5. Sıfır değer kesilmiş negatif binomial regresyonu	16
3.2.6. Poisson hurdle regresyon modeli	16
3.2.7. Negatif binomial hurdle regresyon modeli	17
3.2.8. Model seçimi	18
4. BULGULAR	21
4.1. Tanıtıcı İstatistikler	21
4.2. Parametre Tahminleri	21
4.2.1. Ham veriler için elde edilen parametre tahminleri	22
4.2.2. Veri setindeki sıfır gözlemlerinin %25'i silindiğinde elde edilen parametre tahminleri.....	26
4.2.3. Veri setindeki sıfır gözlemlerinin %50'si silindiğinde elde edilen parametre tahminleri.....	30

4.2.4. Veri setindeki sıfır gözlemlerinin %75'i silindiğinde elde edilen parametre tahminleri.....	34
4.2.5. Veri setindeki sıfır gözlemlerinin tamamı silindiğinde elde edilen parametre tahminleri.....	37
5. TARTIŞMA VE SONUÇ.....	41
KAYNAKLAR.....	43
ÖZ GEÇMİŞ.....	47



ÇİZELGELER LİSTESİ

Çizelge	Sayfa
Çizelge 4.1. Tüm sıfır gözlemler modelde olduğunda model uyum ölçütleri.....	23
Çizelge 4.2. Sıfır yayımlı negatif binomial hurdle regresyonu için parametre tahminleri.....	23
Çizelge 4.3. Gözlemlerin %25'i silindiğinde model uyum ölçütleri.....	27
Çizelge 4.4. Sıfır yayımlı negatif binomial hurdle regresyonu için parametre tahminleri.....	27
Çizelge 4.5. Gözlemlerin %50'si silindiğinde model uyum ölçütleri	30
Çizelge 4.6. Sıfır yayımlı negatif binomial hurdle regresyonu için parametre tahminleri.....	31
Çizelge 4.7. Gözlemlerin %75'i silindiğinde model uyum ölçütleri.....	34
Çizelge 4.8. Sıfır yayımlı negatif binomial regresyonu için parametre tahminleri.....	35
Çizelge 4.9. Sıfır gözlemlerin tamamı silindiğinde model uyum ölçütleri	37
Çizelge 4.10. Sıfır kesilmiş negatif binomial regresyonu için parametre tahminleri	38
Çizelge 5.11. Model uyum ölçütlerinin regresyon modellerine göre dağılımı.....	42

ŞEKİLLER LİSTESİ

Şekil	Sayfa
Şekil 4.1. Bağımlı değişkenin dağılımı.	21
Şekil 4.2. Bağımlı değişkenin dağılımı.	22
Şekil 4.3. Bağımlı değişkenin box-plot dağılımı.....	22
Şekil 4.4. Sıfır yayılımlı Poisson regresyonu için rootogram grafiği.....	24
Şekil 4.5. Sıfır yayılımlı negatif binomial regresyonu için rootogram grafiği.....	24
Şekil 4.6. Poisson regresyonu için rootogram grafiği	25
Şekil 4.7. Negatif binomial regresyonu için rootogram grafiği.....	25
Şekil 4.8. Sıfır yayılımlı hurdle Poisson regresyonu için rootogram grafiği.....	26
Şekil 4.9. Sıfır yayılımlı hurdle negatif binomial regresyonu için rootogram grafiği....	26
Şekil 4.10. Negatif binomial regresyonu için rootogram grafiği.....	28
Şekil 4.11. Sıfır yayılımlı hurdle Poisson regresyonu için rootogram grafiği.....	28
Şekil 4.12. Sıfır yayılımlı hurdle negatif binomial regresyonu için rootogram grafiği..	29
Şekil 4.13. Sıfır yayılımlı Poisson regresyonu için rootogram grafiği.....	29
Şekil 4.14. Sıfır yayılımlı negatif binomial regresyonu için rootogram grafiği.....	29
Şekil 4.15. Poisson regresyonu için rootogram grafiği	30
Şekil 4.16. Negatif binomial regresyonu için rootogram grafiği.....	32
Şekil 4.17. Sıfır yayılımlı Poisson regresyonu için rootogram grafiği.....	32
Şekil 4.18. Sıfır yayılımlı negatif binomial regresyonu için rootogram grafiği.....	32
Şekil 4.19. Sıfır yayılımlı Poisson regresyonu için rootogram grafiği.....	33
Şekil 4.20. Sıfır yayılımlı negatif binomial regresyonu için rootogram grafiği.....	33
Şekil 4.21. Poisson regresyonu için rootogram grafiği	33

Şekil**Sayfa**

Şekil 4.22. Sıfır yayımlı Poisson regresyonu için rootogram grafiği.....	35
Şekil 4.23. Sıfır yayımlı negatif binomial regresyonu için rootogram grafiği.....	36
Şekil 4.24. Negatif binomial regresyonu için rootogram grafiği.....	36
Şekil 4.25. Poisson binomial regresyonu için rootogram grafiği	36
Şekil 4.26. Sıfır yayımlı Poisson regresyonu için rootogram grafiği.....	37
Şekil 4.27. Sıfır yayımlı negatif binomial regresyon için rootogram grafiği.....	37
Şekil 4.28. Poisson regresyonu için rootogram grafiği	38
Şekil 4.29. Negatif binomial regresyonu için rootogram grafiği.....	39

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış bazı simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler

Açıklama

AIC

Akaike bilgi ölçütü

BIC

Bayesian bilgi ölçütü

H-NB

Hurdle negative binomial

H-P

Hurdle poisson

Log

Logaritmik

ML

En çok olabilirlik yöntemi

NB

Negatif binomial

PD

Poisson dağılımı

REHNB

Rasgele etkiler hurdle negatif binom

Std

Standart

ZINB

Sıfır yayımlı negatif binomial

ZIP

Sıfır yayımlı poisson

ZTNB

Sıfır değer kesilmiş negatif binomial

ZTP

Sıfır değer kesilmiş poisson



1. GİRİŞ

Sayıma dayalı olarak elde edilen bağımlı değişken modellenirken poisson regresyonu yaygın olarak kullanılmaktadır. Bununla birlikte, sayıma dayalı olarak elde edilen değişkenlerde, çalışmanın doğası gereği çok fazla sayıda sıfır değerli gözlem olabilir. Örneğin, ağaç yapraklarındaki akar sayımları, sıcaklığın düşük olduğu aylarda sıfır değerini içermektedir. Bu tip gözlemler, veri setinin bir parçası olduğundan dolayı modele dahil edilmesi gerekmektedir. Başka bir deyişle, sıfır değer ağırlıklı gözlemlerin, yok sayılmaması gerekir. Bunun önüne geçmek için sıfır değer ağırlıklı regresyon yöntemleri kullanılmaktadır. Bununla birlikte, sayım ile elde edilmiş bağımlı değişken sıfır değeri içermeyebilir, ya da daha az sayıda sıfır değeri içerebilir. Bu gibi durumlarda ise sıfır kesilmiş regresyon yöntemleri kullanılmaktadır.

Sayım ile elde edilen bağımlı değişkendeki sıfır gözlemlerin sayısına bağlı olarak, hangi doğrusal olmayan regresyonun kullanılması gerektiği, önemli bir sorun teşkil etmektedir. Sıfır yayılımı modellemek için hem Poisson hem de negatif binomial regresyonu esas alan birçok yöntem mevcuttur. Bağımlı değişkendeki sıfırların belli oranlarda silinmesi ile hangi modelin etkili olacağı belirlenebilmektedir. Bu nedenle bağımlı değişkende farkı oranlarda sıfır gözlemlerin silinmesi sonucunda, bahsedilen yöntemlerinin etkinlikleri daha iyi ortaya konulabilir. Böylece, sayım ile elde edilen bağımlı değişkendeki sıfır gözlemlerin oranlarına göre, hangi doğrusal regresyon yönteminin kullanılması gerektiği belirlenebilir.

Bu tez çalışmasında, sayıma dayalı olarak elde edilen bağımlı değişkendeki sıfır değerlerinin % 25, %50, %75 ve %100 oranlarında silinmesi sonucunda, model uyum ölçütlerine bağlı olarak, hangi doğrusal olmayan regresyon modelinin uygun olacağının belirlenmesi amaçlanmıştır. Ayrıca her biri farklı oranlarda kesilmiş sıfır gözlemler için oluşturulan regresyon modellerinin birbirleriyle olan etkinlikleri ele alınacaktır.

2. KAYNAK BİLDİRİŞLERİ

Sayıma dayalı olarak elde edilen değişkenler değeri en küçük sıfır (0) ve üst sınırı belli olmayan değerler içermektedir. Bu tür veri kümelerinde negatif değerler yoktur. Sayıma dayalı olarak elde edilen verilerde küçük değerlerin yoğunluğu büyük değerlerden çok daha fazla olduğundan verilerin dağılımı sağa doğru çarpıktır (Tin, 2008; Vergne ve ark., 2012). Bu tip veriler belirli bir zaman aralığında, birbirinden bağımsız ve düşük olasılık değerleri ile gerçekleşmektedir. Bilindiği gibi, PD'da ortalama ile varyans birbirine eşittir ve eş yayılım (equidispersion) olarak adlandırılmaktadır. Ancak uygulamalarda bu eşitliği sağlamak her zaman mümkün değildir. Varyansın ortalamadan büyük olması aşırı yayılım (overdispersion ya da extra Poisson variation) ve küçük olması da az yayılım (underdispersion) olarak tanımlanmaktadır (Yeşilova ve ark., 2007). Veri kümelerinde genellikle aşırı yayılım, nadiren de az yayılım ile karşılaşılmaktadır.

Poisson dağılımda beklenen sıfır sayısı, bağımlı değişkeninin ortalamasına ve gözlem sayısına bağlıdır. Veri kümesindeki sıfır değerleri sistematik sıfırlar (true zeros, structural, systematic) ve şansa bağlı sıfırlar (false zeros, samples, random) şeklinde iki farklı sınıfa ayrılmaktadır. Sistematik sıfırlar, bağımlı değişkeninin doğrudan sıfır (0) olarak görüldüğü sıfırlardır. Gözlemlerin büyük bir kısmının sıfır olduğu veri kümelerinde sıfır değerlerin analiz dışı tutulması doğru olmayan sonuçların elde edilmesine yol açmaktadır. Ayrıca beklenenden fazla sıfır değeri içeren veri kümesinin yaklaşık normal dağıldığını varsayarak standart regresyon yöntemlerin uygulanması da yanlış parametre tahminlerin elde edilmesine neden olabilir (Mghimbeigi ve ark., 2009; Iddi, 2013; Shiyuka, 2018). Sıfır değerli sayıma dayalı olarak elde edilen verilerin dağılımı genellikle sağa doğru çarpıktır. Bu tür verilerin normal dağılımını sağlamak için değişkenin logaritmik değeri alınmaktadır. Ancak bağımlı değişkenindeki sıfır değerlerin doğal logaritması alınamamaktadır. Sıfır değer logaritmasının tanımsız olmasından dolayı sıfır değerlerine genellikle çok küçük bir değer (0.001) eklendikten sonra logaritması alınır (King, 1989). Bu durumda veri kümesindeki sıfır değerlerine eklenen sabit değer logaritmik yayılımı söz konusu olur. Bu durum yanlış parametre

tahminlerine neden olabilir (McCullagh ve Nelder, 1989; Agresti, 1997; Cameron ve Trivedi, 1998).

Kontrollü klinik çalışmalarında, bağımlı değişkenler, belirli bir zaman periyodu boyunca (örneğin, ilaç kullanım süresi, aylık riskli cinsel ilişki periyodu) semptomların sayısı veya risk davranışlarının sayısı gibi tamsayı veya sayımlar şeklini alır. Bu tip değişkenler genellikle normal dağılım göstermezler. T-testi, varyans analizi ve kovaryans analizi gibi parametrik testler; en küçük kareler yöntemini kullanmakta ve değişkenlerin normal olarak dağıldığını varsaymaktadır. Bu varsayımın sağlanmaması, sapmalı bir tahminlerin elde edilmesine neden olabilir. Bazı durumlarda, varsayımların sağlanmaması durumunda uygulanan dönüşümlerde yetersiz kalmaktadır. Bu nedenle; son yirmi otuz yılda Poisson, negatif binomial, sıfır değer ağırlıklı yöntemler ve hurdle modelleri de içeren normal dağılım göstermeyen veriler için standart parametrik olmayan istatistiksel modellerin artan şekilde tercih edildiği görülmüştür. Hastaneye gelen hasta sayıları, bir kavşakta meydana gelen kaza sayısı gibi olayların modellenmesinde, Poisson, negatif binomial, sıfır değer ağırlıklı ve hurdle modelleri yaygın olarak kullanılmaktadır (Safari ve ark., 2012; Sarul ve Şahin, 2015).

HIV risk davranışı (CTN-0019) vakalarını azaltmak için Madde Bağımlılığı Klinik Araştırmaları Ağı Ulusal Enstitüsü tarafından yürütülen davranışsal müdahalelerin çok alanlı klinik çalışmasından, madde bağımlılığını etkileyen değişkenlerin modellenmesinde Poisson, negatif binomial, sıfır değer ağırlıklı yöntemler ve hurdle modelleri kullanılmıştır(Hu ve ark., 2011).

Poisson regresyon modeli genellikle bağımlı değişkenin sayım ile elde edildiği durumlarda kullanılmaktadır. Birçok durumda bağımlı değişkende sıfır değerinin çok olmasından dolayı; ortalama ile bağımlı değişkenin varyansı birbirine eşit değildir. Bu nedenle, Poisson regresyon modeli bu tip verilerin analizi için uygun değildir. Dolayısıyla, aşırı yayılım problemini gidermek için negatif binomial regresyon, sıfır değer ağırlık yöntemler ile hurdle modelini kullanılmaktadır. Ayrıca, bazı durumlarda, bağımlı değişken bazı değerler için sansürlenmiş olabilir. Sansürlenmiş sayıma dayalı olarak elde edilen bağımlı değişkenin analizinde, hurdle negatif binomiyal regresyonu kullanılmıştır. ML yöntemi kullanılarak regresyon parametrelerinin hesaplanması üzerinde durulmuş ve regresyon modeline uyum iyiliği incelenmiştir (Sakthivel ve Rajitha, 2017).

Sigortacılıkta adil bir fiyatlandırma sağlamak önemli bir konudur. Poisson süreci varsayımları sağlandığında, Poisson dağılımı, sayım verileri için temel modeldir. Sigortacılar, tenzili muafiyet ve hasarsızlık indirimi almak için, adli tazminat taleplerini belirtme eğiliminde olmadıklarından dolayı sigorta verilerinde, beklenenden daha fazla sıfır değerleri vardır. Ancak, sıfır değerleri hesaba katılmadığında, portföyde daha iyi fiyatlandırma elde etmek için bilgi eksikliğine sebep olur. Hasar talep sıklığı verilerinin analizi için yan Poisson ile sıfır değer ağırlıklı regresyon yöntemleri kullanılmıştır. Elde AIC değerlerine göre hurdle modellerinin, diğer modellere nazaran daha iyi sonuç verdiği görülmüştür (Hofstetter ve ark., 2016).

Hayat-dışı sigortacılıkta, hasar talep sıklığının modellenmesi hayati önem taşımaktadır. Hayat-dışı sigortadaki hasar talep sayımlarında sıfır değerlerinin çokluğundan dolayı Poisson veya negatif binomial regresyonlarının kullanılması uygun değildir. Bu sıfır değerlerinin analiz dışı tutulması, hayat-dışı sigorta portföyünü oluşturacak olan kesin oranı elde etmek için bilgi eksikliğine yol açabilir. Hayat-dışı sigortacılık verilerinin analizinde sıfır değer ağırlıklı Poisson regresyon modeli ile hurdle modeli kullanılmıştır. Negatif binomial hurdle modelin diğer modellere nazaran daha iyi sonuç verdiği saptanmıştır (Mouatassim, 2012).

Diş epidemiyolojisinde; çürük (Ç), eksik (E) ve dolgu (D) dişler veya yüzeyler indeksi (ÇED indeksi) sık kullanılan bir ölçüttür. ÇED indeksi, çürük diş olmayan bireyler için sıfır gözlemlerinin çok olduğu ve dağılımının kuvvetli bir pozitif çarpık olmaktadır. Böyle durumlarda standart doğrusal modellerin kullanılması uygun değildir. Burada, çürük diş sayısı ve onu etkileyen değişkenlerin modellenmesinde hurdle modelleri kullanılmıştır. Uyum ölçütlerine göre hurdle modellerin diğer sıfır değer ağırlıklı regresyon yöntemlerine göre daha iyi sonuç verdiği belirlenmiştir (Zulkifli ve ark., 2011).

Sıfır değer ağırlıklı regresyon modellerinin performansını test etmek için Monte Carlo yöntemi kullanılmak bağımlı değişken üzerinde farklı olanlarda sıfır ve sıfır olmayan gözlemler türetilmiştir. Her bir farklı oranda silinen gözlemler için sıfır değer ağırlıklı regresyon yöntemleri uygulanmıştır. Böyle AIC istatistiği kullanılarak bu regresyon modelleri karşılaştırılmıştır. Karşılaştırmalar neticesinde sıfır değer ağırlıklı negatif binomial regresyon daha iyi sonuç verdiği saptanmıştır (Miller ve Miller, 2008).

Klinisyenlerin, hastalığın şiddetini ve prognozunu saptamak ve müteakip tedaviyi planlamak için meme kanseri hastalarında, ilgili nodların sayısını tahmin etmeleri gerekmektedir. İlgili nodların dağılımı genellikle aşırı yayılma - beklenenden daha büyük bir değişkenlik göstermektedir. Şimdiye kadar, aşırı yayılmanın sadece gözlemlenmemiş heterojenliğe bağlı olduğu varsayılarak, bu dağılımı tanımlamak için negatif binomial regresyon modeli kullanılmıştır. İlgili nodüllerin dağılımı, aşırı yayılmaya yol açabilen çok fazla sayıda sıfırın büyük bir bölümünü (negatif nodüller) içermektedir. Çalışmada, 1993 Ocak- 2005 Ocak tarihleri arasında Hindistan'daki bir üniversite hastanesinde aksiller diseksiyon yapılan 1152 hastanın verileri incelenmiştir. İlgili nodüllerin sayısını tahmin etmek amacıyla sıfır değer ağırlıklı regresyon modelleri uygulanmış ve karşılaştırılmıştır. Aynı zamanda, tüm fazla sıfırların, beklenen sonuç için biraz risk altında olan kişilerden geldiği bu tür popülasyonlarda sıfır değer ağırlıklı modellerin kullanılması daha uygundur. Negatif binom regresyon modeli, verilere Poisson, sıfır hurdle/değer ağırlıklı Poisson regresyon modellerinden daha iyi uyum sağlamıştır. Bununla birlikte, elde edilen uyum ölçütlerine göre; sıfır hurdle/değer ağırlıklı negatif binomial regresyon modelleri, ilgili nodül sayısını negatif binom modelinden çok daha doğru tahmin etmiştir. Bu, ilgili nodüllerin sayısının, sadece gözlemlenmemiş heterojenlik nedeniyle değil, aynı zamanda veri setindeki fazla sayıdaki negatif nodüller nedeniyle de aşırı değişkenlik gösterdiğini düşündürmektedir. Bu analizde sadece cilt değişiklikleri ve primer bölgeler, negatif nodüllerle ilişkili olmuştur. Bununla birlikte; doğum sayısını deri değişiklikleri, primer bölge ve tümörün boyutu daha fazla sayıda nodülle etkilemiştir. Neredeyse eşit performanslar söz konusu olduğunda, nodal tutulumunda “yüksek riskli” olan negatif nodüllerin bir tahminini sağladığı için nodal frekans tanımlanırken hurdle modeli yerine sıfır değer ağırlıklı negatif binom modeli tercih edilmelidir (Dwivedi ve ark., 2010).

Orman yangınları doğal kaynaklarda yıkıcı hasara yol açabilir. Bununla birlikte, ciddi ekonomik ve sosyal etkileri de beraberinde getirebilir. Bir orman yangını için uygun koşulların oluşmasında meteorolojik faktörler kritik bir rol oynamaktadır. Orman yangını olaylarının etkin tahmini, kayıpları önleyebilir veya en aza indirebilir. Çalışmada, dağınık olması muhtemel olan ve sıfır aşırılığı (yangın oluşmaması) sayımı içeren yangın oluşum verilerini analiz etmek için sıfır değer ağırlıklı regresyon modelleri kullanılmıştır. Yani; bu veriler genellikle Poisson modeli, negatif binomial

model, sıfır değer ağırlıklı modeller ve hurdle modelleri gibi sayma verileri modelleri kullanılarak analiz edilmiştir. Çalışmanın verileri Çin'deki Guizhou eyaletinin Qiannan özerk ilçesinden toplanmıştır. 1996-1997 yılları arasında Ocak-Nisan arası yangın oluşumu verilerini (bahar yangını mevsimi) kullanarak, sayma verileri modellerine rastgele etkiler eklendi. Bu çalışmanın sonuçları, sıfır değer ağırlıklı negatif binomial modeli ile elde edilen tahminin gerçek orman yangını oluşumuna uyumu açısından daha etkili ve güvenilir bir temel sağlamıştır. Ayrıca karışık etki modelinin, orman yangın tahmininde ilgili sabit etkili modelinden daha iyi performans gösterdiğini göstermiştir. Tüm meteorolojik faktörler arasında, bağıl nem ve rüzgar hızının yangın oluşumu ile yüksek derecede ilişkili olduğu saptanmıştır (Xiao ve ark., 2015).

Ekolojik çalışmaların birçoğunda, örneklerdeki tespit hatalarından ya da türlerin varlıklarını sürdürebilmeleri için yaşama elverişsiz koşullardan dolayı, türlerin yokluğu kaçınılmazdır ve bu da çok fazla sıfır sayısına neden olur. Bu tür verilerin modellenmesinde $\log(\text{bağımlı değişken}+1)$ dönüşümü yapılarak doğrusal modeller kullanılmaktadır. Bu da, sapmalı parametre tahminlerinin elde edilmesine yol açmaktadır. Bu tip veriler kesikli olduklarından ve fazla sayıda sıfır değerini içerdiklerinden dolayı sıfır değer ağırlıklı Poisson ve negatif binomial modeller ile hurdle modellerinin kullanılması daha uygundur. Midyelerin, ekonomik açıdan Türkiye'nin en önemli deniz ürünlerinden biri olması nedeniyle, çalışmada, Akdeniz midyelerinin (*Mytilus galloprovincialis* L.) varlığı için bir tehdit oluşturan Nematopsis legeri parazitin oluşmasında önemli biyotik ve abiyotik faktörlerin belirlenmesinde bu dört modelin performanslarının incelenmesidir. Türkiye'nin Sinop şehrinin üç kıyı bölgesinden toplanan veriler, parazit sayılarının ortalama % 50'sinden fazlasının sıfır değerli olduğunu göstermiş ve model karşılaştırmaları bilgi ölçütlerine dayanarak yapılmıştır. Elde edilen sonuçlara göre, bu parazitin oluşma olasılığının en iyi sıfır değer ağırlıklı negatif binomial regresyonu ile negatif binomial hurdle modeli ile ifade edildiğini ve bu modellerde etkili faktörlerinin, bölgelerin ekolojik farklılıklarıyla uyumunu göstermiştir (Çankaya ve ark., 2017).

Mekânsal ağırlık matrisinin belirlenmesi meselesi, şu anda mekânsal ekonometri literatüründe büyük ölçüde tartışılırken, uzaklık tanımı şimdiye kadar daha az dikkat çekmiştir. Uzaklık ölçüsünün seçimi, genellikle geçirilmiştir, ama en sonunda öklid uzaklığı kullanılmıştır. Paris bölgesinde yapılan bir çalışmada; yüksek trafik sıklığı,

hız sınırları ya da fiziksel olarak geçilmesi mümkün olmayan bariyerler, komşu bölgeler arasındaki bağlantıyı azaltabilir veya tamamen ortadan kaldırabilir; bu da mekânsal etkileri belirtmede öklid uzaklığı seçiminin sorgulanmasına neden olabilir. Çeşitli mesafe ölçülerini karşılaştırmak için bazı faaliyet sektörlerine yönelik hurdle-Poisson modellerinin olasılık karışımı geliştirilmiştir. Her bir model sınıfı, yayılma etkilerini kaydetmek için farklı bir mesafe tanımını kullanmaktadır. Öklid mesafesi, iki yol arasındaki mesafe (tikanıklık durumunda ve tikanıklık olmadan), toplu taşıma mesafesi ve ilgili seyahat süreleri kullanılmıştır. Genel olarak, elde edilen sonuçlar, kuruluşların konumlarının ana belirleyicileri ile ilgili olarak literatürle uyumludur. Bununla birlikte, inşaat gibi bazı faaliyet sektörlerinde, hususi araçlar için en yoğun zamanda seyahat süresinin, mekansal yayılmaları doğru bir şekilde kaydetme olasılığının en yüksek olduğu, oysa gayrimenkul gibi diğer sektörler için, öklid mesafesinin biraz daha baskın olduğu görülmüştür. Bu, mekansal yayılmaların, muhtemel faaliyet sektörüne bağlı olarak farklı yollarla yönlendirildiğini göstermektedir. Ayrıca, birkaç gizli sınıf kullanan hurdle-Poisson modellerinin önerilen karışımının, tek bir mesafe ölçüsüne dayanan "saf" hurdle-Poisson modellerinden önemli ölçüde daha iyi performans sergilediği ortaya çıkarılmıştır (Buczowska ve ark., 2016).

Doğum öncesi bakım, gebeliğin düzenli takibi yoluyla maternal fetal sonuçların en iyi şekilde kullanılmasını amaçlayan önleyici bir obstetrik sağlık programıdır. Dünya Sağlık Örgütü, normal gebelik için en az dört kez doğum öncesi bakımı tavsiye etse bile, Etiyopya gibi gelişmekte olan ülkelerden elde edilen mevcut kanıtlar, farklı nedenlerden dolayı bundan faydalanan çok az kadının olduğunu belirtmektedir. Etiyopya'nın kırsal bölgelerinde hamile kadınlar arasında doğum öncesi bakım hizmeti ziyaretlerinin sayısındaki engellerin belirleyici etkenlerini analiz etmek için 2011 yılı ENSA (Etiyopya Nüfus ve Sağlık Araştırması)'dan 1127 hamile kadın kullanılmıştır. Verilerde çok fazla sıfır sayısı (%51.5) olduğu; varyansın, 7.196, ortalamasından, 1.85, çok daha yüksek olduğu saptanmıştır. Doğum öncesi bakım sayıları ve onu etkileyen değişkenleri modellemek için Poisson, negatif binomial, sıfır değer ağırlıklı poisson, sıfır değer ağırlıklı negatif binomial, Poisson hurdle ve negatif binomial hurdle regresyonları uygulanmıştır. Bu modellerin her birisi, olabilirlik oranı testi, young testi ve AIC kullanılarak karşılaştırılmıştır. Analiz sonucunda, Etiyopya'nın kırsal bölgelerinde hamilelik sırasında modern ziyaretlerine erişimin, ulusal değere göre düşük

olduğu bulunmuştur. Farkındalık eksikliği, eğitim eksikliği, ağır iş yükü, yoksulluk ve sağlık bilgilendirme eksikliği, doğum öncesi bakım ziyaretlerine katılmama ile önemli ölçüde ilişkili olmuştur. Hurdle Poisson regresyon modelinin, fazla sayıda sıfır sayısı ve sıfır bulunmayan sonuçlardaki yüksek değişkenlikle karakterize edilen verilerle daha iyi uyum sağladığı saptanmıştır (Sekata, 2015).

Rasgele etkiler panel veri hurdle modelleri, Colorado'daki I-70 otoyolunun dağlık kesimlerindeki günlük kaza (çarpma) sıklığını araştırmak için oluşturulmuştur. Yol Hava Bilgi Sisteminde (RWIS) gerçek zamanlı trafik ve hava durumu ve yol yüzey koşulları, yol özelliklerini içeren modellerle birleştirilmektedir. Rasgele etkiler hurdle negatif binom (REHNB) modeli, diğer üç rakip modelle birlikte günlük kaza sıklığını incelemek için geliştirilmiştir. Önerilen model, gözlemlerin seri korelasyonunu, dengesiz panel-veri yapısını ve baskın sıfırları dikkate almaktadır. Bazı istatistiksel testlere dayanarak REHNB modeli, tipik bir dağlık otoyol için dört aday modelden en uygun olanı olarak tanımlanmaktadır. Sonuçlar, şu şekildedir: (1) kısa süreli kaza sıklığı verilerindeki aşırı dağılım, kaza verilerindeki hem çok fazla sayıdaki sıfırlardan hem de gözlemlenmemiş heterojenlikten kaynaklanmaktadır; ve (2) REHNB modeli, bu tip veriler için uygundur. Ayrıca, hava koşulları, yol yüzey koşulları ve trafik koşullarını içeren zamanla değişen değişkenlerin, kaza sıklığında önemli rol oynadığı bulunmuştur. Metodolojik ilerlemelerin yanı sıra, önerilen teknoloji, dünya çapında daha erişilebilir hale gelen RWIS gibi alan izleme verilerinin ayrıntılı verilerini kullanarak mühendislik uygulamalarının kısa süreli kaza frekansı modelleri geliştirmesi için büyük bir potansiyel taşımaktadır (Chen ve ark., 2016).

Poisson regresyon modeli ve negatif binom regresyon modeli gibi sayım modelleri, yazılım hata tahminlerini bulmak için kullanılabilir. Bu tür tahminlerin yardımıyla geliştirme ekibi, operasyonel yazılımın kalitesini artırabilir. Belirli bir yazılım sistemi için, hatalı modül sayısı çok az olduğunda sıfır değer ağırlıklı ve hurdle sayım modelleri daha uygun olabilir. İlgili literatür, yazılım kalitesinin tahmininde sayım modellerin uygulanması ile ilgili nicel yönlendirmeden yoksundur. Bu çalışma, yazılım hata tahmini bağlamında sekiz sayım modelinin kapsamlı bir deneysel araştırmasını sunmaktadır. Farklı kriterlere göre sayım modelleri için karşılaştırmalı hipotez testini, model seçimini ve performans değerlendirmesini kapsamaktadır. Sunulan vaka çalışması, tam ölçekli bir endüstriyel yazılım sistemine aittir. Hipotez

testlerinden ve model seçim tekniklerinin elde edilen bilgilerin, sayım modellerinin yordayıcı performansları ile tutarlı olmadığı gözlenmiştir. Ayrıca, bir kritere dayalı yapılan karşılaştırmalı analiz, başka bir kriterin analizi ile eşleşmemiştir. Bununla birlikte, belirli bir kriterle ilgili olarak, bir sayım modelinin performansı hem uyum hem de test veri setleri açısından tutarlıdır. Bu, kullanılan bir modelin, belirli bir kritere göre iyi kabul edilmesi durumunda, modelin aynı kritere dayanarak iyi bir tahmin vermesini sağlamaktadır. Sekiz modelin göreceli performansları, tek yönlü ANOVA modeli ve Tukey çoklu karşılaştırma tekniği ile değerlendirilmiştir. Bu karşılaştırmalı çalışma, belirli bir yazılım sisteminin kalitesini tahmin etmek için en iyi sayım modelinin seçilmesinde faydalıdır (Gao ve Khoshgoftaar, 2007).

Bir ikili (binari) içsel olarak belirlendiği sayma verileri regresyon modelleri için bir tahminleyici önerilmektedir. Bu tahmin edici, sayımların koşullu olasılık fonksiyonu için esnek bir form kullanması bakımından önceki yaklaşımlardan ayrılır. Bir Monte Carlo deneyi kullanarak, tahmincimizin uyumu geliştirdiğini ve ortalama doğru-üstel olması yönünde kısıtlayan alternatiflerle karşılaştırıldığında, tahmin edicinin sayım üzerindeki etkisine dair daha güvenilir bir tahmin sağladığını gösteriyoruz. ABD'de ailelerin yaptığı seyahat sayısına dair bir uygulamada, elde edilen işlem etkisinin tahmininin, doğru-üstel bir ortalama altında elde edilen sonuçtan oldukça farklı olduğunu görmekteyiz (Romeu ve Vera-Hernandez, 2005).

Bireyler ve kuruluşlar için Facebook ve benzeri sosyal medya platformlarını kullanmak giderek yaygınlaşmaktadır. Bunlar, sağlıkla ilgili bilgilerin yayılması, tartışılması ve araştırılması dâhil olmak üzere birçok çeşitli amaçla kullanılmaktadır. ABD Federal sağlık kurumları, sosyal medya kullanıcılarının, sağlıkla ilgili tartışmaları okumak, yaymak, tanıtmak ve teşvik etmek için bu platformlardan yararlanmalarını sağlamaktadır. Ancak, farklı kurumlar ve iletişimleri farklı düzeylerde katılım almaktadır. Bu çalışmada, etkileşimle ilişkili faktörleri tanımlamak için istatistiksel modeller kullanılmıştır. 24 sağlık kuruluşuna ait 72 Facebook hesabından 45.000'in üzerinde Facebook gönderi analiz edilmiştir. Hesap kullanımı, kullanıcı aktivitesi, bu mesajların duygusu ve içeriği incelenmiştir. Katılım süreleri ile ilişkili faktörleri belirlemek için Cox oransal hazard modeli ve katılım düzeyiyle ilgili faktörleri tanımlamak için de hurdle regresyon modeli kullanılmıştır.

Yapılan analizlerin sonucunda, kurumların ve hesapların, sosyal medya kullanımları ve gerçekleştirdikleri etkinlikler bakımından büyük ölçüde değişiklik gösterdiği ortaya koyulmuştur. Örneğin, istatistiksel analizler, fotoğraf veya video gibi daha fazla görsel ipucu içeren ya da olumlu duygular ifade eden Facebook gönderilerinin daha fazla katılım sağladığını göstermiştir. Ayrıca, meslek ya da kurum gibi belirli konulardaki gönderilerin, katılım süresini olumsuz etkilediği görülmüştür. İlk olarak, Facebook'ta ABD Federal sağlık kuruluşlarına gösterilen katılımın kapsamlı bir analizi sunulmuştur. Ayrıca, yöntemlerimizin sağlamlığını göstermek için bu çalışmadan elde edilen bulgular, benzer bir konuya odaklanan ancak Twitter üzerine yapılan daha önceki bir çalışmamızla karşılaştırılmıştır (Bhattacharya ve ark., 2017).





3. MATERYAL VE YÖNTEM

3.1. Materyal

Çalışmanın materyali Bitki koruma alanından elde edilmiş olan gerçek bir veri setinden oluşmuştur. Van ilinde yaygın olarak bulunan Starking Delicious ve Golden Delicious elma çeşitleri (*Malus communis* L. Rosaceae) üzerinde bulunan *Aculus schlechtendali* (Nal.) (Acari: Eriophyidae) üzerinde populasyon yoğunluk çalışmaları 2016 yılında yapılmıştır. *A. schlechtendali*'nin *M. communis* L. üzerinde farklı zamanlarda (Haziran-Eylül ayları arasında) protogyn ve deutogyn evreleri, yumurtaları ve nimf pupa dönemleri belirlenmiştir. Değişik zamanlardaki akar sayımları bağımlı değişkeni oluştururken; sıcaklık, nem, çeşit ve ay gibi değişkenler ise bağımsız değişkenler olarak modele dahil edilmiştir.

3.2. Yöntem

3.2.1. Poisson regresyonu

Poisson regresyonu regresyon analizinin genelleştirilmiş doğrusal modellerinden biridir. Bu regresyon yöntemi, sayım ile elde edilmiş verilerin analizinde ve ihtimal tablolarının analizinde kullanılır. Özellikle ihtimal tabloları modellenirken log-doğrusal regresyon adıyla da bilinir.

Modelde bağımlı değişken Poisson dağılımı gösterirken, log bağlantı fonksiyonu ile bu bağımlı değişken bağımsız değişkenlerle ilişkilendirilir ve doğrusal bir eşitlik elde edilir (Lambert, 1992; Yesilova ve ark., 2016)

Bağımsız değişkenler $X_1, X_2, \dots, X_m$ iken;

$$\log(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m \quad (3.1)$$

şeklinde log bağlantı fonksiyonu kullanılmaktadır. Burada μ bağımsız değişkenlerin üssel bir fonksiyonudur. Log olabilirlik fonksiyonu ise

$$L(\beta/y_i, x_i) = \sum_{i=1}^n [y_i x_i' \beta - \exp(x_i' \beta) - \ln y_i!] \quad (3.2)$$

şeklindedir (Khoshgoftaar ve ark., 2005)

3.2.2. Negatif binomial regresyonu

Negatif binomial (NB) regresyonu, bağımsız değişkende aşırı yayılım görüldüğünde Poisson regresyonu yerine kullanılan metottur. Log bağlantı fonksiyonu bağımlı değişken ile bağımsız değişkenleri ilişkilendirir. Olasılık fonksiyonu;

$$P_r(Y) = \frac{\Gamma(y_i + \frac{1}{\alpha})(\alpha\mu_i)^{y_i}}{y_i! \Gamma(\frac{1}{\alpha})(1 + \alpha\mu_i)^{y_i + \frac{1}{\alpha}}} , \quad \alpha > 0 \quad (3.3)$$

biçiminde verilmektedir. Eşitlikte μ_i dağılımın ortalaması, α NB yayılım parametresini belirtmektedir (Achim ark., 2007; Hilbe, 2007). NB regresyon modelinin log olabilirlik fonksiyonu,

$$L(\beta, \alpha, y) = \sum_{i=1}^n \left[\frac{1}{\alpha} \log(1 + \alpha\mu_i) - y_i \log\left(1 + \frac{1}{\alpha\mu_i}\right) + \log\Gamma\left(y_i + \frac{1}{\alpha}\right) - \log\Gamma\left(\frac{1}{\alpha}\right) - \log y_i! \right] \quad (3.4)$$

şeklinde yazılabilir.

3.2.3. Kesilmiş modeller

Araştırmacının amacı doğrultusunda veya modellenecek veri setine uygun olarak bazen bir değer üzerinde ya da bir değer altında olan veriler kesilip atılır. Sağ kesilmiş modeller belirli bir değer üzerinde olan tüm verileri sıfır kabul eder, bu değerler

modelden atılmış olur. Sol kesilmiş modellerde ise belirli bir değer altındaki değerler modelden atılır. Bazen de bağımlı değişkende belirli bir değer altında ya da üstünde hiç değer gözlemlenmez. Bu durumlarda da kesilmiş modeller tercih edilir. Değeri kesilmiş veri setleri modellenirken buna uygun olarak geliştirilen kesilmiş regresyon modelleri kullanılır (Hilbe, 2007; Rose ve ark., 2006). Bu çalışmada özel olarak işlenen ve en çok kullanılan sıfır değer kesilmiş regresyon modelleri incelenecektir.

3.2.4. Sıfır değer kesilmiş Poisson regresyonu

Sıfır değer kesilmiş Poisson (ZTP), pozitif Poisson olarak da bilinir. Bu dağılımda Poisson dağılımından farklı olarak tüm değerler pozitif ve sıfır değeri (0) içermez.

Sıfır değer kesikli Poisson modeli ;

$$P(y_i | y_i > 0, \lambda) = \frac{P(y_i | \lambda)}{P(y_i > 0, | \lambda)} = \frac{\exp(-\lambda) \lambda^{y_i}}{y_i! (1 - \exp(-\lambda))} \quad (3.5)$$

$$P(y_i > 0, | \lambda) = (1 - \exp(-\lambda)), \quad y_i = 1, 2, \dots, \quad i = 1, 2, \dots, N, \quad (3.6)$$

olarak verilir. i-inci gözlem için log-likelihood ;

$$l(y_i; \mu_i | y_i > 0) = y_i [X_i^T \beta] - \exp(X_i^T \beta) - \ln(y_i!) - \ln[1 - \exp\{-\exp(X_i^T \beta)\}] \quad (3.7)$$

ZTP modeli için log-likelihood;

$$\begin{aligned} l(y_i; \mu_i | y_i > 0) \\ = \sum y_i [X_i^T \beta] - \exp(X_i^T \beta) - \ln(y_i!) - \ln [1 \\ - \exp\{-\exp(X_i^T \beta)\}] \end{aligned} \quad (3.8)$$

şeklinde verilir (Hilbe, 2014).

3.2.5. Sıfır değer kesilmiş negatif binomial regresyonu

Sıfır değer kesilmiş negatif binomial regresyonu pozitif değerlere sahip bağımlı değişken ($y_i > 0$) aynı zamanda aşırı yayılım gösteriyorsa bu tip bağımlı değişkenlerin modellenmesinde kullanılan regresyon yöntemidir. ZTNB için Log-likelihood ;

$$L(\mu; y|y > 0) = \sum_{i=1}^n \{L_{NB2} - \ln[1 - \{1 + \alpha \exp(x_i' \beta)\}^{-\frac{1}{\alpha}}]\} \quad (3.9)$$

biçiminde yazılabilir. Eşitlikte yer alan L_{NB2} ise;

$$L(\beta; y, \alpha) = \sum_{i=1}^n \left\{ y_i \log \left(\frac{\alpha \exp(x_i' \beta)}{1 + \alpha \exp(x_i' \beta)} \right) - \left(\frac{1}{\alpha} \right) \log(1 + \alpha \exp(x_i' \beta)) \right. \\ \left. + \log \Gamma\left(y_i + \frac{1}{\alpha}\right) - \log \Gamma(y_i + 1) - \log \Gamma\left(\frac{1}{\alpha}\right) \right\} \quad (3.10)$$

3.2.6. Poisson hurdle regresyon modeli

Poisson Hurdle Regresyon Modeli; Sınırlandırılmış sayıma dayalı olarak elde edilen pozitif gözlem değerleri ($y_i > 0$) Poisson dağılımı kullanılarak modellendiğinde, Poisson hurdle model olarak adlandırılmaktadır. y_i , $i=1,2,\dots,n$ birbirinden bağımsız sayıma dayalı olarak elde edilen gözlem değerleri olsun. $y_i=0$ olma olasılığı $1-p(x)$ ve $y_i \sim \text{truncated Poisson}(\lambda(z))$ olma olasılığı $p(x)$ olsun. Burada x ve z kovariet matrisleridir. Poisson hurdle model (Dalrymple ve ark., 2003),

$$P(y_i = 0/x) = 1 - p(x)$$

$$P(y_i = q/x, z) = \frac{p(x) \exp(-\lambda(z)) \lambda(z)^q}{q! (1 - \exp(-\lambda(z)))}, q=1,2 \quad (3.11)$$

$p(x)$ ve $\lambda(z)$ sırasıyla logit ve log doğrusal fonksiyonları kullanılarak,

$$\log(\lambda(z)) = x_i' \beta \quad (3.12)$$

$$\log it(p_i) = z_i' \alpha \quad (3.13)$$

biçiminde modellenmektedirler (Lambert, 1992). Eşitlik 3.12 ve eşitlik 3.13'de verilen β ve α sırasıyla bilinmeyen parametre vektörleridir.

Bilinmeyen β, α parametrelerinin tahmin edilmesinde en çok olabilirlik (maximum likelihood) yöntemi kullanılmaktadır. Poisson hurdle için log-olabilirlik aşağıdaki gibi yazılabilir.

$$\begin{aligned} LL &= \sum_{y_i > 0} x_i \beta - \sum_{i=1}^n \log(1 + \exp(x_i \beta)) \\ &+ \sum_{y_i > 0} \left[y_i z_i \alpha - \exp(z_i \alpha) - \log(1 - \exp(-\exp(z_i \alpha))) - \log(y_i!) \right] \\ &= LL(\beta) + LL(\alpha) \end{aligned} \quad (3.14)$$

3.2.7. Negatif binomial hurdle regresyon modeli

Negatif binomial hurdle modelde, sayıma dayalı olarak elde edilen bağımlı değişkenin sıfır ya da sıfır değerli olmama sonuçlarını belirleyen binomial olasılık modeli ile pozitif sonuçları tanımlayan sınırlandırılmış sayıma dayalı modeli için verilen log-olabilirlik fonksiyonu aşağıdaki gibi yazılabilir (Long ve Freese, 2006; Hilbe, 2007),

$$LL = \ln(f(0)) + \{\ln[1 - f(0)] + \ln P(j)\} \quad (3.15)$$

Eşitlikte 3.15'te, $f(0)$ modelin binary kısmının olasılığını, $p(j)$ pozitif sayımın olasılığını göstermektedir. Logit model kullanıldığı durumda, sıfır değerli gözlemin olasılığı,

$$f(0) = P(y = 0; x) = 1 / (1 + \exp(xb_1))$$

ve

1- $f(0)$ ise (sıfırdan büyük gözlemler),

$\exp(xb_1) / (1 + \exp(xb_1))$ biçiminde yazılabilir.

Böylece negatif binomial hurdle modelin her iki kısmı için log-olabilirlik fonksiyonu aşağıdaki gibi yazılabilir.

$$\begin{aligned} \text{cond} \{ y = 0, \ln(1 / (1 - \exp(xb_1))) \}, \\ \ln(\exp(xb_1) / (1 + \exp(xb_1))) + y * \ln(\exp(xb) / (1 + \exp(xb))) \\ - \ln(1 + \exp(xb)) / \alpha + \ln \Gamma(y + 1 / \alpha) - \ln \Gamma(y + 1) - \ln \Gamma(1 / \alpha) \\ - \ln(1 - (1 + \exp(xb))(-1 / \alpha)) \} \end{aligned} \quad (3.16)$$

Eşitlik 3.6'da verilen log-olabilirlik fonksiyonun maksimize edilmesiyle, bilinmeyen parametrelerin maksimum olabilirlik tahminleri elde edilmektedir.

3.2.8. Model seçimi

Akaike bilgi ölçütü (AIC) model uyumu için kullanılan uyum ölçütleridir. üç ölçüt genel bilgi ölçütü (information criterion=IC)'den türetilmiştir. IC,

$$IC = -2\log L + Cr \quad (3.17)$$

biçiminde yazılabilir. Burada logL modelde iterasyon bittikten sonra elde edilen log olabilirlik değerini, r parametre sayısını ve n örnek büyüklüğünü göstermektedir. Eğer

C=0 ise IC klasik LR istatistiğine eşittir. C=2 olduğunda IC=AIC ve C=LogN olduğunda IC=BIC olmaktadır. Yapılan çalışmalar BIC, AIC uyum ölçütlerinin birlikte kullanılması gerektiğini göstermektedir (Dalrymple ve ark., 2003). En küçük uyum ölçütlerine sahip model, en iyi model olarak kabul edilir. AIC bilgi ölçütü;

$$AIC = -2\log L + 2r \quad (3.18)$$

biçiminde hesaplanmaktadır (Liu ve Cela, 2008).

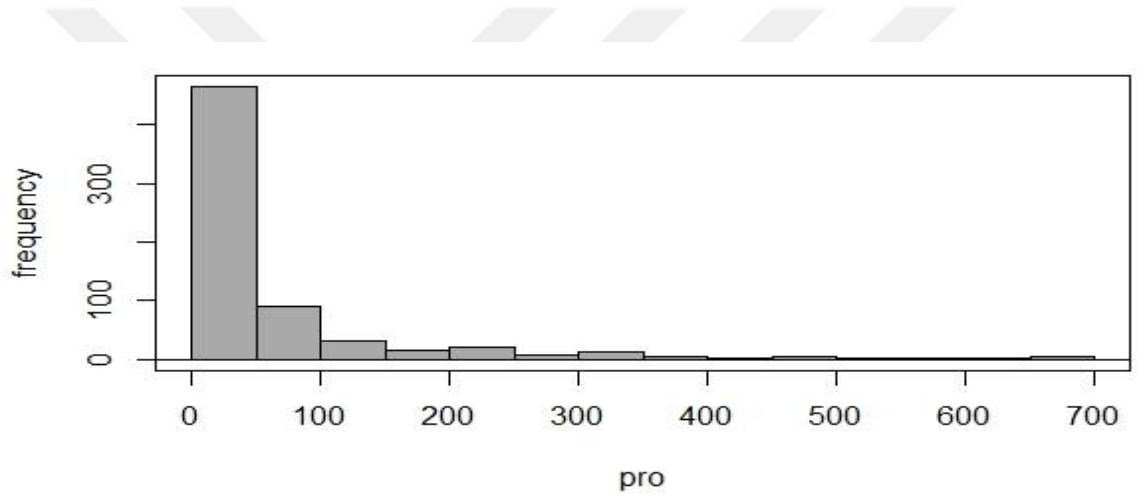




4. BULGULAR

4.1. Tanıtıcı İstatistikler

Çalışmada, bağımsız değişken elma yaprakları üzerindeki protogyn akar sayımları bağımlı değişken; aylar, çeşit, nem ve sıcaklık ise bağımsız değişken olarak modele dahil edilmiştir. Bağımlı değişkendeki toplam 660 gözlemin %26.2'si sıfır değerden oluşmaktadır. Bağımlı değişkenin frekans dağılımı Şekil 4.1'de verilmiştir. Sıfır değerlerinin çokluğundan dolayı, dağılımın şekli sağa çarpık olarak elde edilmiştir.



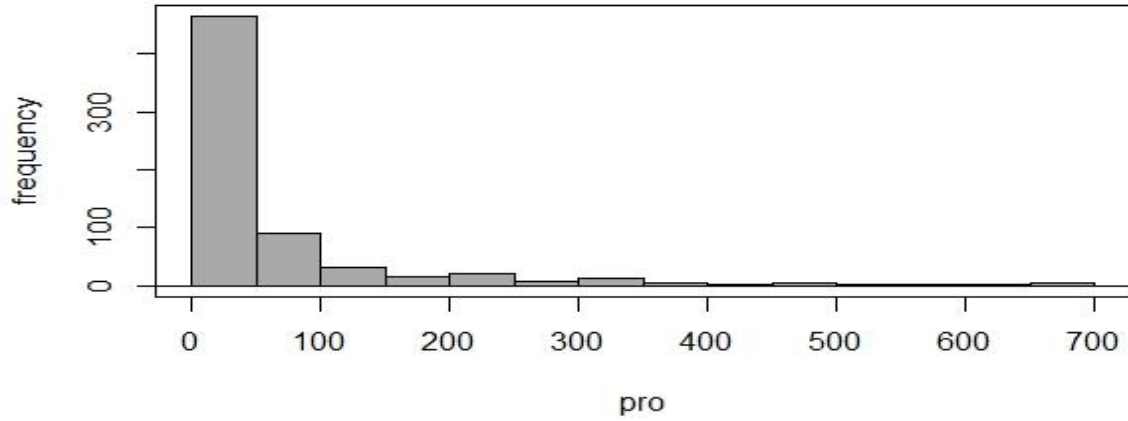
Şekil 4.1. Bağımlı değişkenin dağılımı.

4.2. Parametre Tahminleri

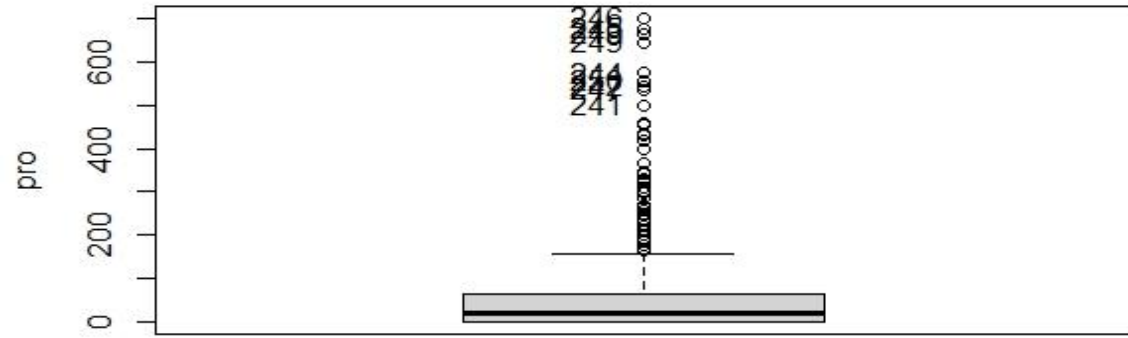
Bu bölümde veri setinin sırasıyla; ham veri seti durumu, sıfır gözlemlerin rasgele %25'nin, sıfır gözlemlerin rasgele %50'sinin, sıfır gözlemlerin rasgele %75'nin ve sıfır gözlemlerin tamamının silinmesi durumunda, hangi doğrusal olmayan regresyonun, veri setini açıklayan en iyi model olduğu, uyum ölçütlerine bakılarak karar verilmiştir.

4.2.1. Ham veriler için elde edilen parametre tahminleri

Bağımlı değişkenin dağılım grafiği ve box-plot grafiği sırasıyla Şekil 4.2 ve Şekil 4.3’de verilmiştir. Sıfır gözlemlerin çok sayıda olmasından dolayı dağılımın şekli oldukça sağa çarpık olmuştur. Hem sıfır gözlemlerin çok sayıda olması hem de aykırı değerlerin olması box-plot grafiğine yansımıştır.



Şekil 4.2. Bağımlı değişkenin dağılımı.



Şekil 4.3. Bağımlı değişkenin box-plot dağılımı.

Tüm sıfır gözlemlerin bağımlı değişkende olması durumunda, veri setini en iyi açıklayan doğrusal olmayan modeli belirlemek için elde edilen AIC ve BIC uyum ölçütleri ile Vounge istatistiği Çizelge 4.1’de verilmiştir. AIC ve BIC model uyum ölçütlerine bakıldığında, en küçük değerler sıfır yayılımlı hurdle negatif binomial regresyonunda elde edilmiştir. Vounge istatistiği sonucunda da en iyi modelin sıfır yayılımlı negatif binomial hurdle regresyon olduğu saptanmıştır.

Çizelge 4.1.Tüm sıfır gözlemler modelde olduğunda model uyum ölçütleri

MODEL	AIC	BIC	VUONG
Poisson(P)	25519	25559	NB>P
Negatif Binomial(NB)	5448	5474	-
Zero Inflated Poisson(ZIP)	7367	7371	H-NB >ZIP
Zero Inflated Negative Binomial (ZINB)	5085	5126	H-NB >ZINB
Hurdle Poisson(H-P)	7056	7068	H-NB >H-P
Hurdle Negative Binomial(H-NB)	5062	5111	-

Tüm sıfır gözlemler modeldeyken, veri setini en iyi açıklayan model sıfır yayımlı negatif binomial hurdle regresyonuna ilişkin parametre tahminleri Çizelge 4.2’de verilmiştir. Ay ve çeşit bağımsız değişkenlerinden sırasıyla nisan ayı ve Starking elma çeşidi referans kategorisi olarak alınmıştır. Nem ve sıcaklık numerik bağımsız değişkenlerin bağımsız değişken üzerindeki etkileri istatistiksel olarak önemli bulunmuştur ($p<0.001$).

Çizelge 4.2. Sıfır yayımlı negatif binomial hurdle regresyonu için parametre tahminleri

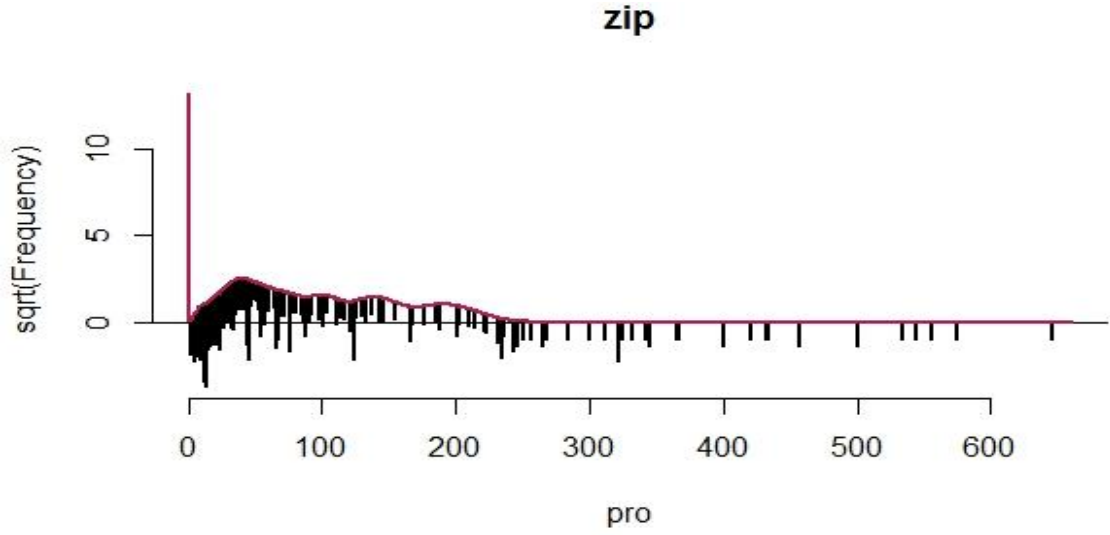
Sayım modeli için parametre tahminleri (truncated negbin with log link):				
Değişken	Std. Hata	z değeri	Pr(> z)	Önem düzeyi
(Intercept)	5.21054	1.38818	3.754	0.000174 ***
Çeşit	-0.24933	0.08941	-2.789	0.005293 **
Aylar	-0.65302	0.05315	-12.286	< 2e-16 ***
Sıcaklık	0.36301	0.02942	12.337	< 2e-16 ***
Nem	-0.06199	0.01342	-4.619	3.86e-06 ***
Log(theta)	0.05978	0.06845	0.873	0.382518

Sıfır gözlemler için Hurdle negatif binomial için parametre tahminleri (binomial with logit link):

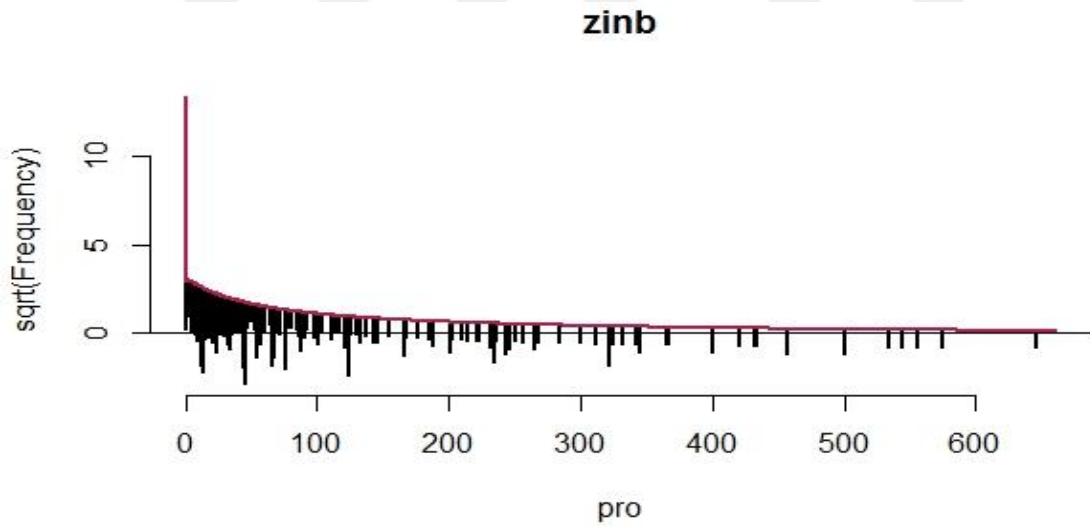
Değişken	Std. Hata	z değeri	Pr(> z)	Önem düzeyi
(Intercept)	1019.6742	89089.3907	0.011	0.991
Sıcaklık	28.8259	5669.6347	0.005	0.996
Nem	-19.0606	77.3536	-0.246	0.805
Aylar	-19.5936	1888.1023	-0.010	0.992
Çeşit	0.3335	0.8099	0.412	0.681

Rootogram grafiklerinde barlar tahmin edilen ve gözlenen akar sayılarının kareköklerinin farkını göstermektedir. Çizgiler ise tahmin edilen akar sayısının karekökü olarak ifade edilmektedir. Barlar x eksenine ne kadar yakınsa model o kadar iyidir şeklinde yorum yapılabilir. Poisson modeline ait grafikte diğer modellere göre

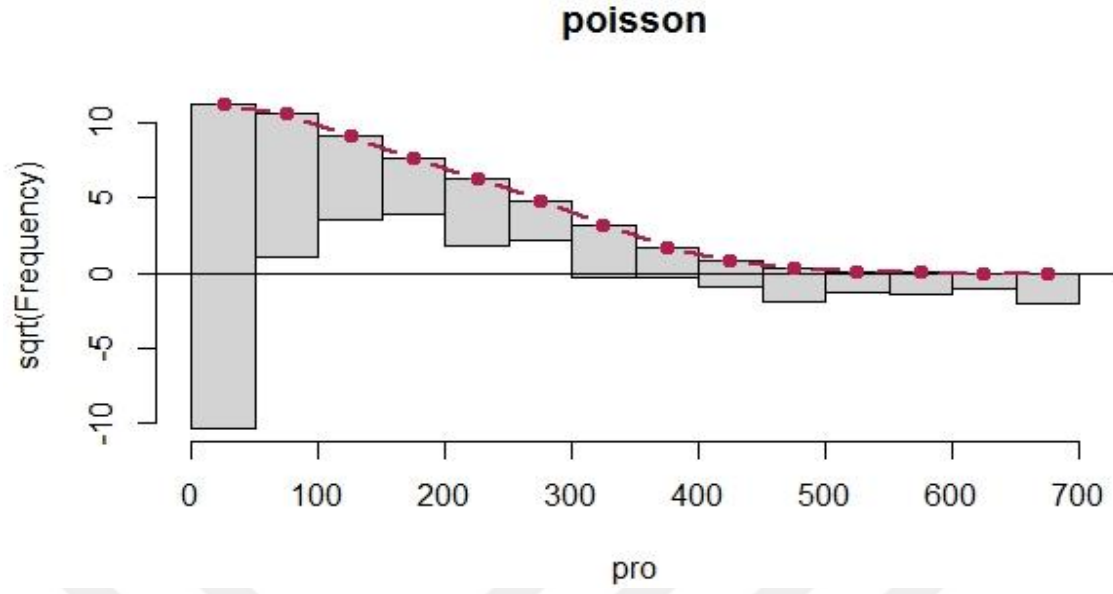
barların sıfıra daha uzakta olduğu ve sıfır yayımlı negatif binomial hurdle regresyonu modelinde ise en yakın olduğu görülmektedir (Şekil 4.3, Şekil 4.8). Bu çalışmada kullanılan modellere ait rootogram grafikleri uyum ölçütleri ile ulaşılan karşılaştırma sonuçlarını desteklemektedir.



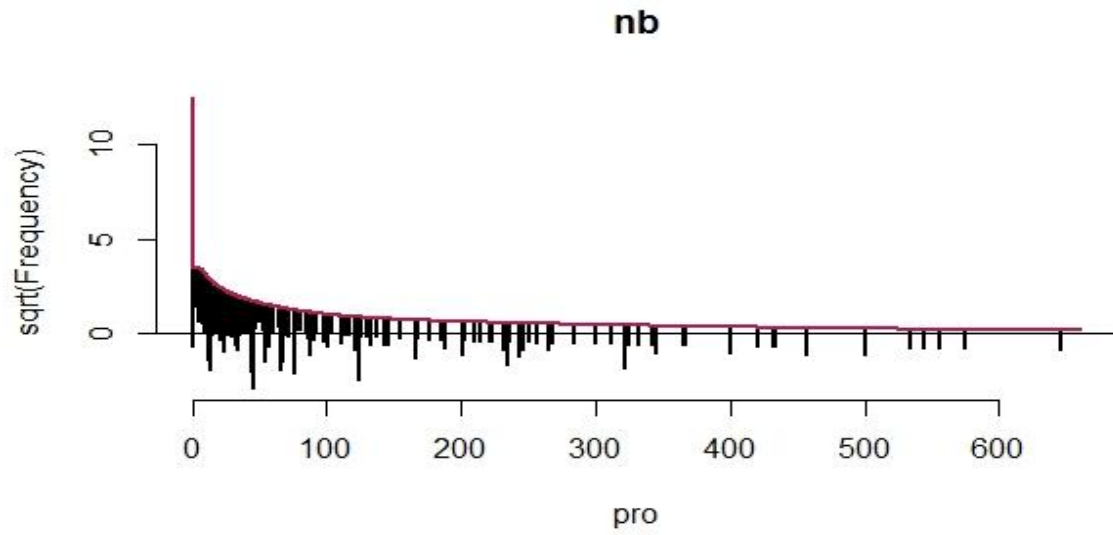
Şekil 4.4. Sıfır yayımlı Poisson regresyonu için rootogram grafiği.



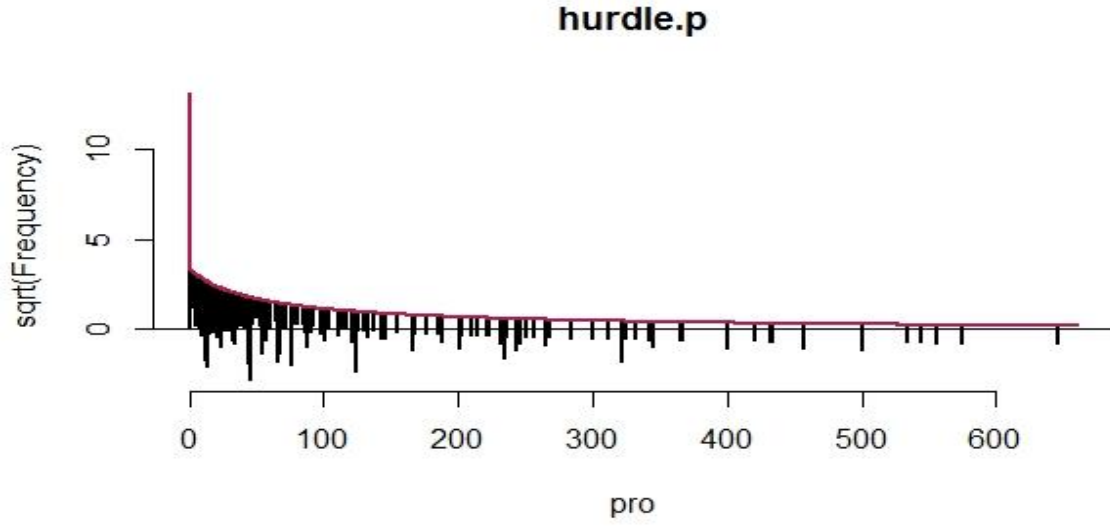
Şekil 4.5. Sıfır yayımlı negatif binomial regresyonu için rootogram grafiği.



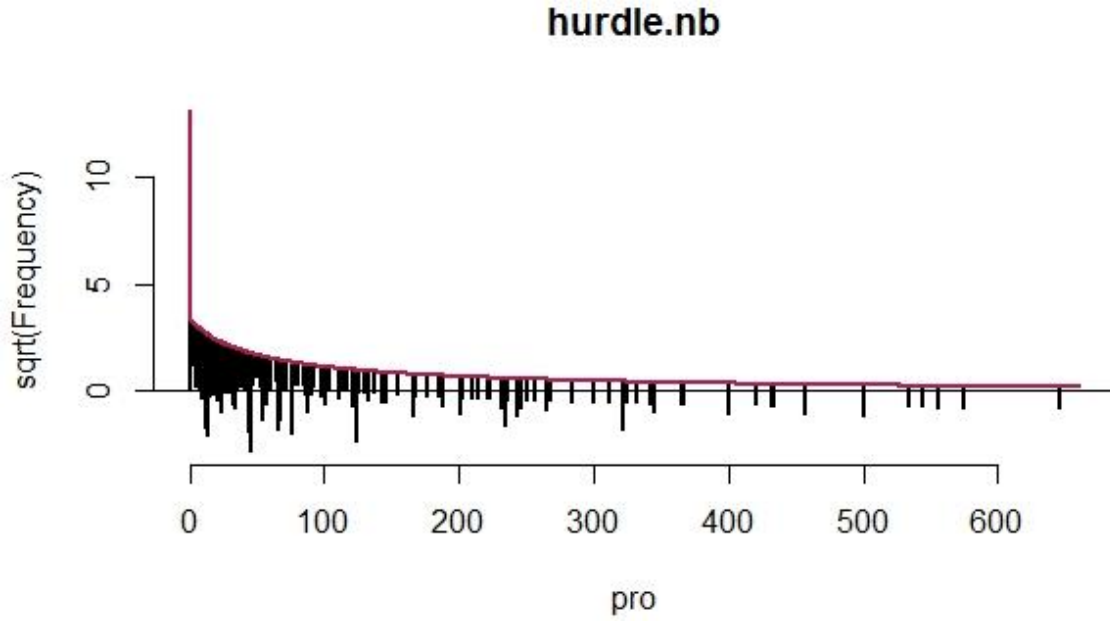
Şekil 4.6. Poisson regresyonu için rootogram grafiği.



Şekil 4.7. Negatif binomial regresyonu için rootogram grafiği.



Şekil 4.8. Sıfır yayımlı hurdle Poisson regresyonu için rootogram grafiği.



Şekil 4.9. Sıfır yayımlı hurdle negatif binomial regresyonu için rootogram grafiği.

4.2.2. Veri集中的 sıfır gözlemlerinin %25'i silindiğinde elde edilen parametre tahminleri

Bağımlı değişkende sıfır gözlemlerinin %25'i rasgele silindiğinde, veri setini en iyi açıklayan doğrusal olmayan modeli belirtmek için elde edilen AIC ve BIC uyum ölçütleri ile Vounğ istatistiği Çizelge 4.3'de verilmiştir. AIC ve BIC model uyum ölçütlerine bakıldığında, en küçük değerlerin sıfır yayımlı negatif binomial hurdle

regresyonunda elde edilmiştir. Vounng istatistiği sonucunda da en iyi modelin sıfır yayılımlı negatif binomial hurdle regresyon olduğu saptanmıştır.

Çizelge 4.3. Gözlemlerin %25'i silindiğinde model uyum ölçütleri

MODEL	AIC	BIC	VUONG
Poisson(P)	25308	25347	NB>P
Negatif Binomial(NB)	5394	5420	-
Zero Inflated Poisson(ZIP)	7325	7349	H-NB>ZIP
Zero Inflated Negative Binomial (ZINB)	5078	5118	H-NB>ZINB
Hurdle Poisson(H-P)	7038	7051	H-NB>H-P
Hurdle Negative Binomial(H-NB)	5058	5107	-

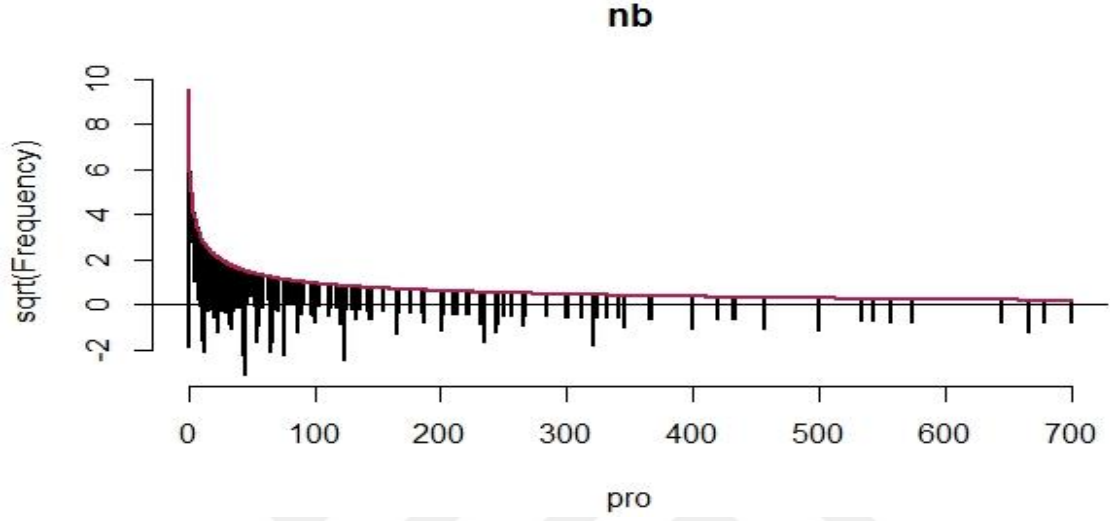
Tüm sıfır gözlemler modeldeyken, veri setini en iyi açıklayan model sıfır yayılımlı negatif binomial hurdle regresyonuna ilişkin parametre tahminleri Çizelge 4.4'de verilmiştir. Ay, çeşit, nem ve sıcaklık bağımsız değişkenlerin bağımlı değişken üzerindeki etkileri istatistiksel olarak önemli bulunmuştur ($p<0.001$).

Çizelge 4.4. Sıfır yayılımlı negatif binomial hurdle regresyonu için parametre tahminleri

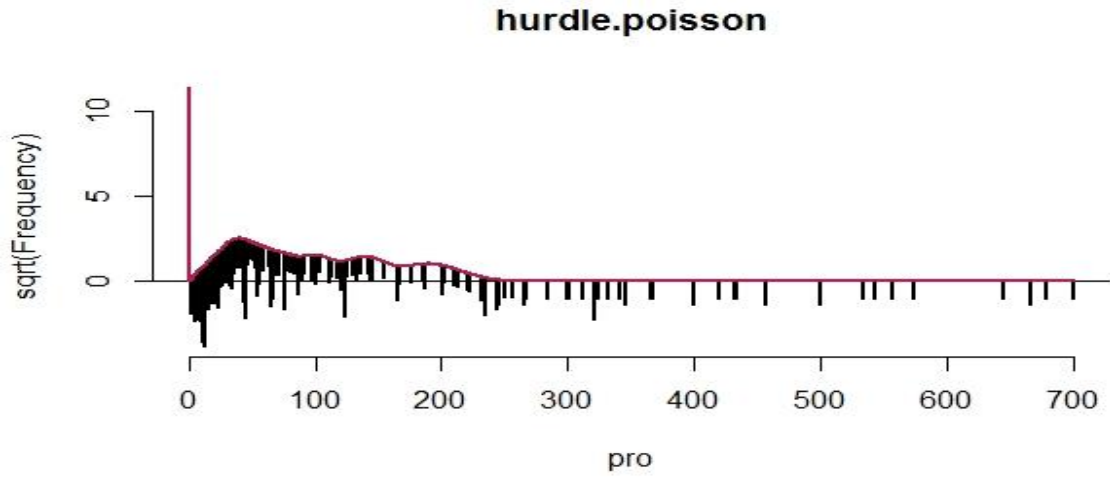
Sayım modeli için parametre tahminleri (truncated negbin with log link):				
Değişken	Std. Hata	z değeri	Pr(> z)	Önem düzeyi
(Intercept)	5.21055	1.38818	3.754	0.000174 ***
Çeşit	-0.24933	0.08941	-2.789	0.005293 **
Aylar	-0.65302	0.05315	-12.286	< 2e-16 ***
Nem	-0.06199	0.01342	-4.619	3.86e-06 ***
Sıcaklık	0.36301	0.02942	12.337	< 2e-16 ***
Log(theta)	0.05978	0.06845	0.873	0.382518
Sıfır gözlemler için Hurdle negatif binomial için parametre tahminleri (binomial with logit link):				
Değişken	Std. Hata	z değeri	Pr(> z)	Önem düzeyi
(Intercept)	1041.7241	86958.4443	0.012	0.990
Çeşit	0.1883	0.8369	0.225	0.822
Aylar	-19.7913	1837.1545	-0.011	0.991
Nem	-19.3510	67.2267	-0.288	0.773
Sıcaklık	28.9193	5521.5758	0.005	0.996

Çizgiler ise tahmin edilen akar sayısının karekökü olarak ifade edilmektedir. Barlar x eksenine ne kadar yakınsa model o kadar iyidir şeklinde yorum yapılabilir.

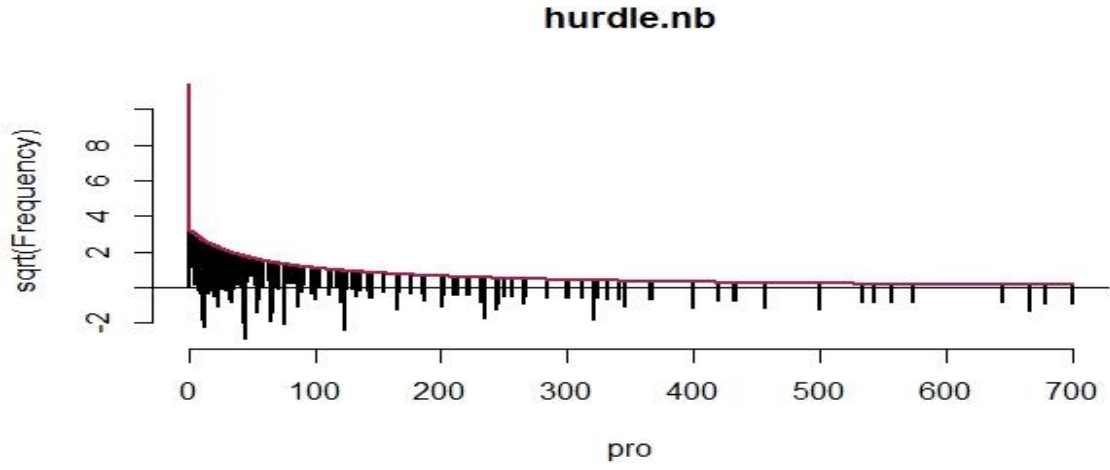
Poisson modeline ait grafikte diğer modellere göre barların sıfıra daha uzakta olduğu ve sıfır yayılımlı negatif binomial hurdle regresyonu modelinde ise en yakın olduğu görülmektedir (Şekil 4.9, Şekil 4.14). Bu çalışmada kullanılan modellere ait rootogram grafikleri uyum ölçütleri ile ulaşılan karşılaştırma sonuçlarını desteklemektedir.



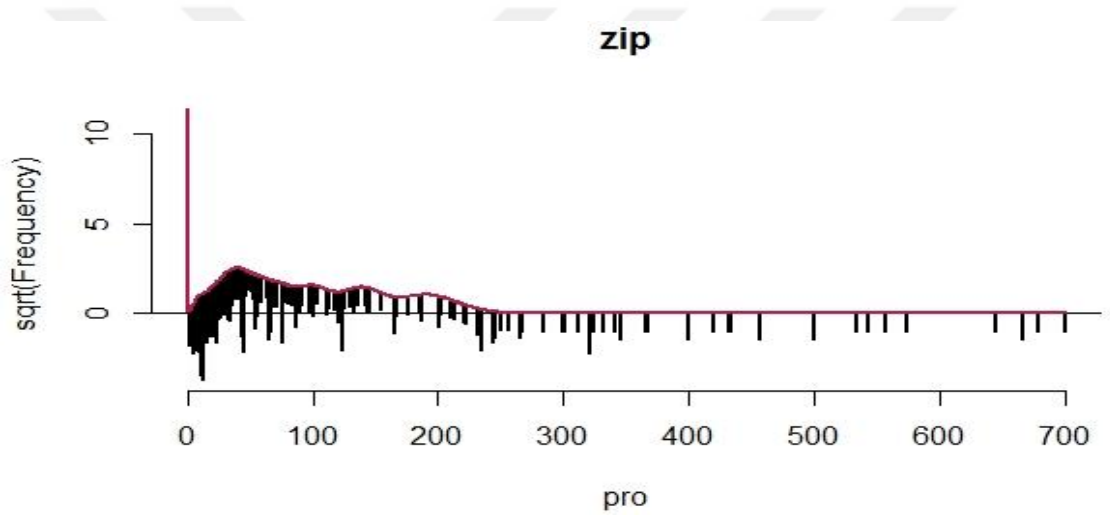
Şekil 4.10. Negatif binomial regresyonu için rootogram grafiği.



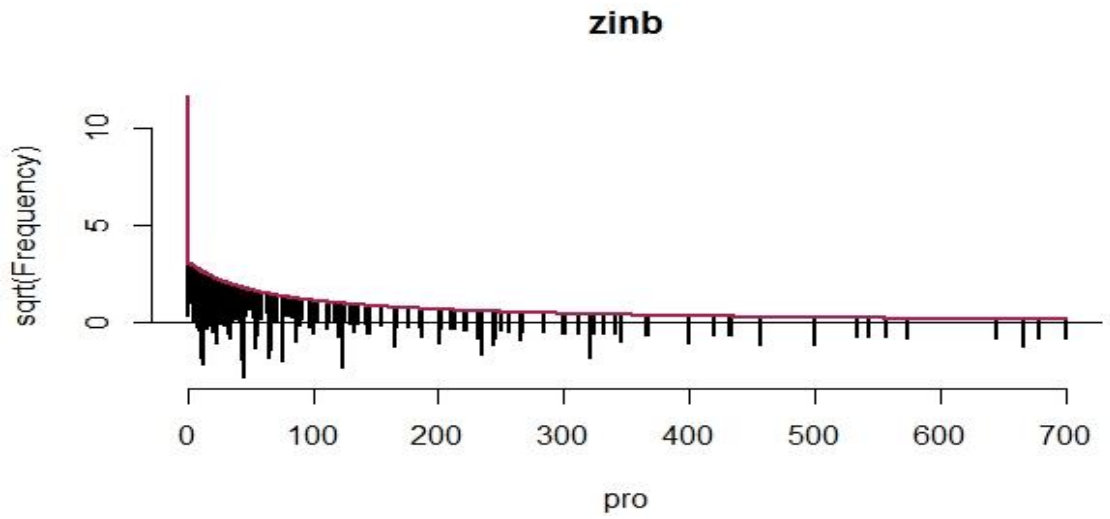
Şekil 4.11. Sıfır yayılımlı hurdle Poisson regresyonu için rootogram grafiği.



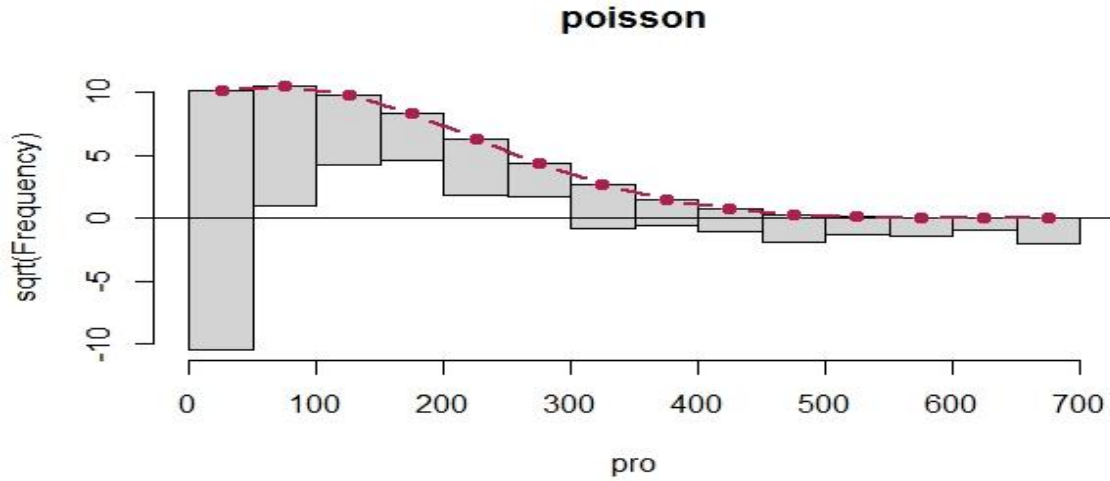
Şekil 4.12. Sıfır yayımlı hurdle negatif binomial regresyonu için rootogram grafiği.



Şekil 4.13. Sıfır yayımlı Poisson regresyonu için rootogram grafiği.



Şekil 4.14. Sıfır yayımlı negatif binomial regresyonu için rootogram grafiği.



Şekil 4.15. Poisson regresyonu için rootogram grafiği.

4.2.3. Veri集中的 sıfır gözlemlerinin %50'si silindiğinde elde edilen parametre tahminleri

Bağımlı değişkende sıfır gözlemlerin %50'si rasgele silindiğinde, veri setini en iyi açıklayan doğrusal olmayan modeli belirlemek için elde edilen AIC ve BIC uyum ölçütleri ile Vounğ istatistiği Çizelge 4.5'de verilmiştir. AIC ve BIC model uyum ölçütlerine bakıldığında, en küçük değerler sıfır yayılımlı negatif binomial hurdle regresyonundan elde edilmiştir. Vounğ istatistiği sonucunda da en iyi modelin sıfır yayılımlı negatif binomial hurdle regresyon olduğu saptanmıştır.

Çizelge 4.5. Gözlemlerin %50'si silindiğinde model uyum ölçütleri

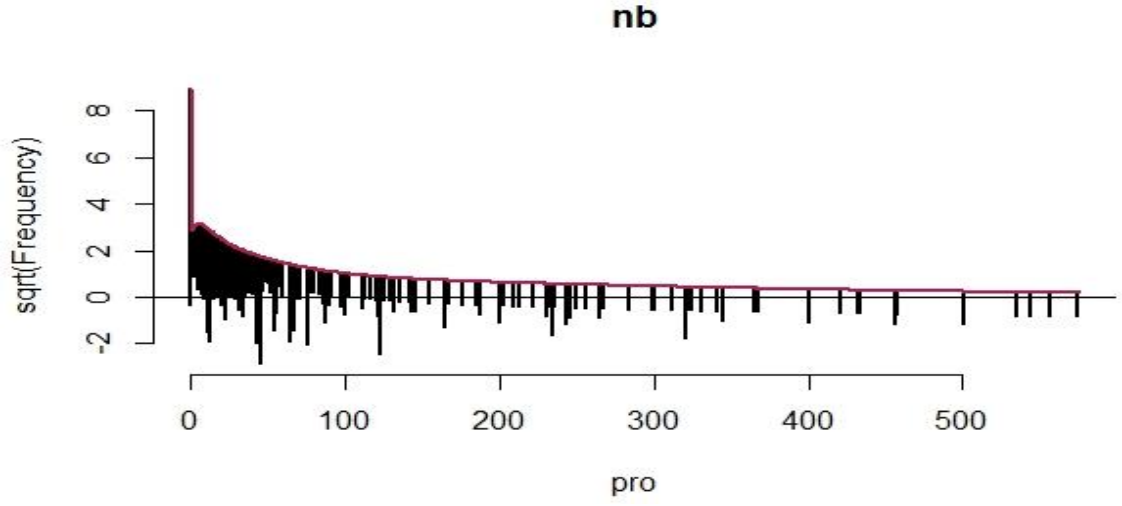
MODEL	AIC	BIC	VUONG
Poisson(P)	25184	25223	NB>P
Negatif Binomial(NB)	5302	5328	-
Zero Inflated Poisson(ZIP)	7317	7324	H-NB>ZIP
Zero Inflated Negative Binomial (ZINB)	5072	5111	H-NB>ZINB
Hurdle Poisson(H-P)	7020	7038	H-NB>H-P
Hurdle Negative Binomial(H-NB)	5026	5097	-

Tüm sıfır gözlemler modeldeyken, veri setini en iyi açıklayan model sıfır yayılımlı negatif binomial hurdle regresyonuna ilişkin parametre tahminleri Çizelge 4.6'da verilmiştir. Ay, çeşit, nem ve sıcaklık bağımsız değişkenlerin bağımlı değişken üzerindeki etkileri istatistiksel olarak önemli bulunmuştur ($p < 0.001$).

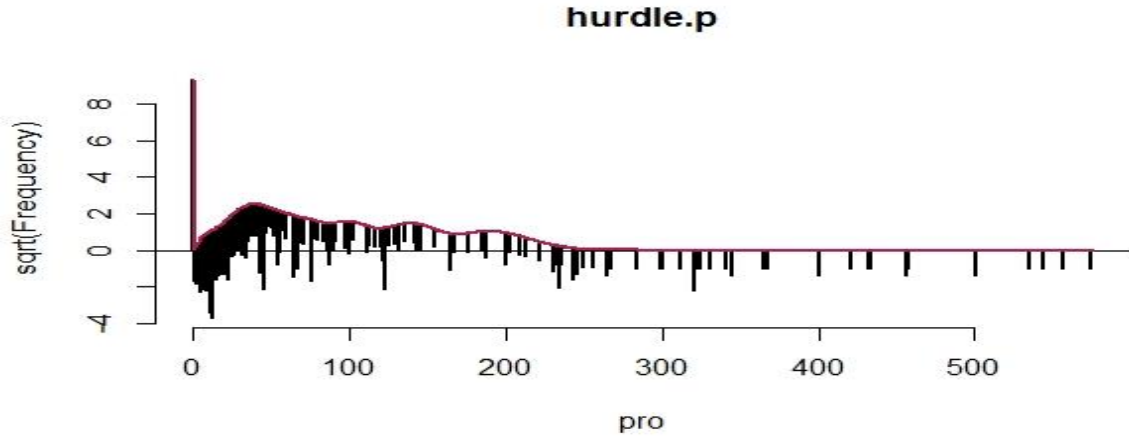
Çizelge 4.6. Sıfır yayımlı negatif binomial hurdle regresyonu için parametre tahminleri

Sayım modeli için parametre tahminleri (truncated negbin with log link):				
Değişken	Std. Hata	z değeri	Pr(> z)	Önem düzeyi
(Intercept)	5.21055	1.38818	3.754	0.000174 ***
Çeşit	-0.24933	0.08941	-2.789	0.005293 **
Aylar	-0.65302	0.05315	-12.286	< 2e-16 ***
Sıcaklık	0.36301	0.02942	12.337	< 2e-16 ***
Nem	-0.06199	0.01342	-4.619	3.86e-06 ***
Log(theta)	0.05978	0.06845	0.873	0.382518
Sıfır gözlemler için Hurdle negatif binomial için parametre tahminleri (binomial with logit link):				
Değişken	Std. Hata	z değeri	Pr(> z)	Önem düzeyi
(Intercept)	-106.262	32.420	-3.278	0.00105 **
Sıcaklık	3.694	0.775	4.767	1.87e-06 ***
Nem	0.445	0.262	1.699	0.08939

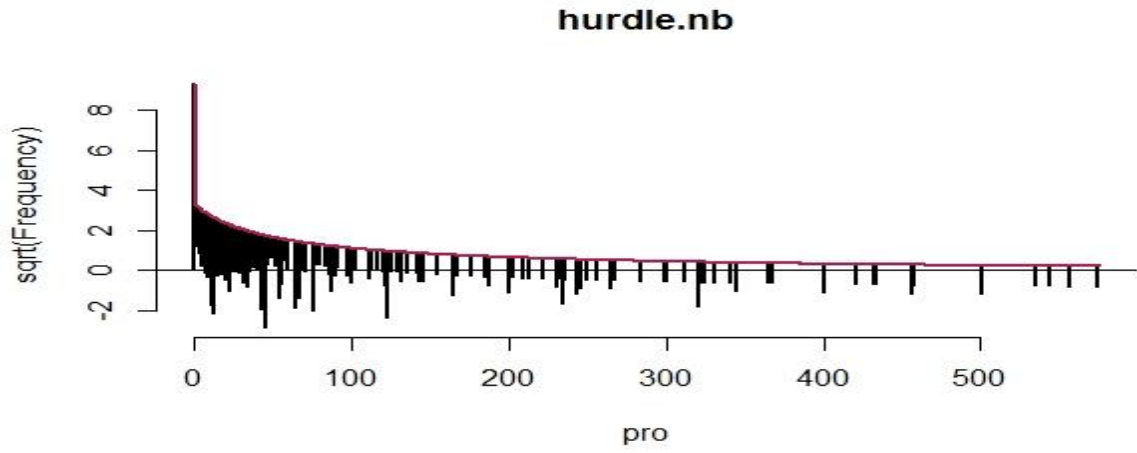
Rootogram grafiklerinde barlar tahmin edilen ve gözlenen akar sayılarının kareköklerinin farkını göstermektedir. Çizgiler ise tahmin edilen akar sayısının karekökü olarak ifade edilmektedir. Barlar x eksenine ne kadar yakınsa model o kadar iyidir şeklinde yorum yapılabilir. Poisson modeline ait grafikte diğer modellere göre barların sıfıra daha uzakta olduğu ve sıfır yayımlı hurdle negatif binomial regresyonu modelinde ise en yakın olduğu görülmektedir (Şekil 4.15, Şekil 4.20). Bu çalışmada kullanılan modellere ait rootogram grafikleri uyum ölçütleri ile ulaşılan karşılaştırma sonuçlarını desteklemektedir.



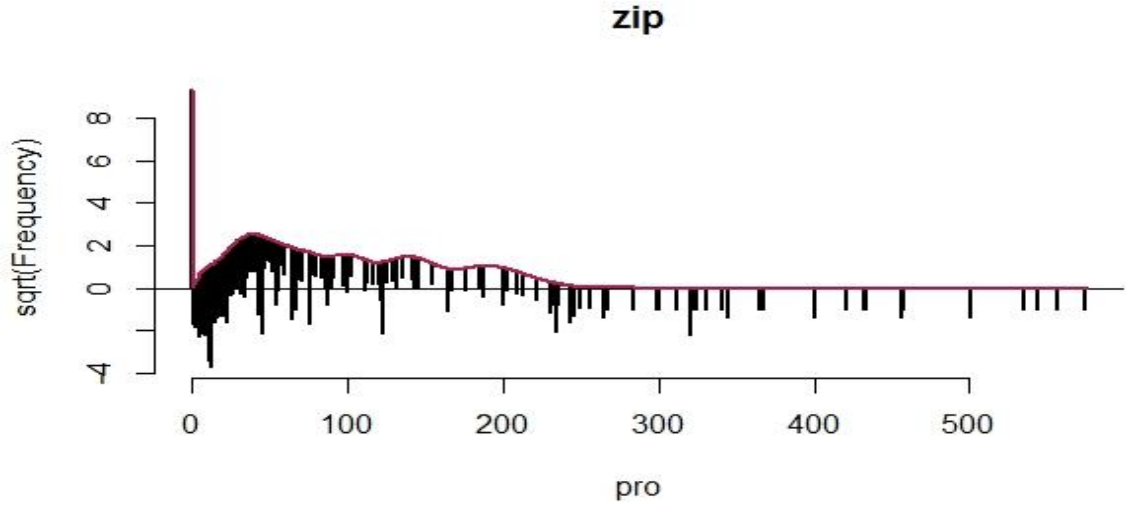
Şekil 4.16. Negatif binomial regresyonu için rootogram grafiği.



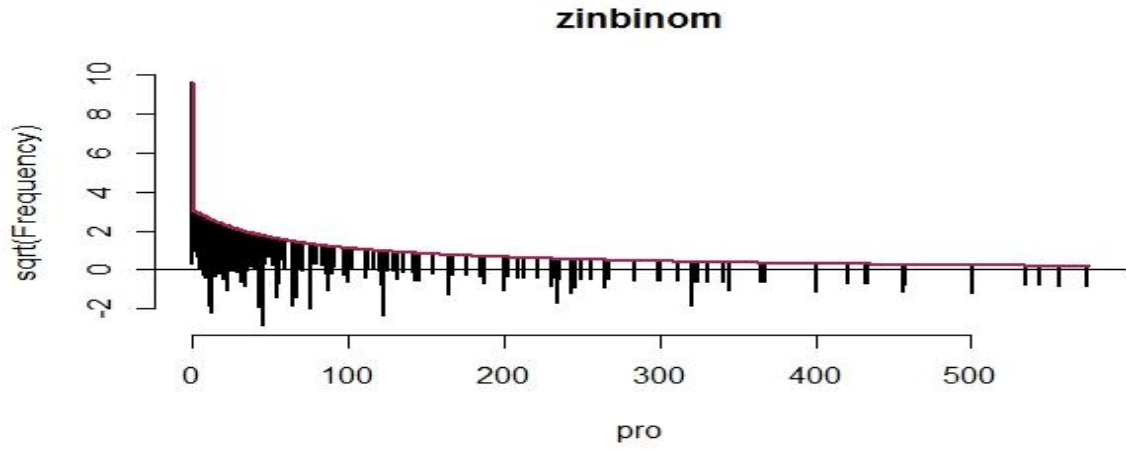
Şekil 4.17. Sıfır yayımlı Poisson regresyonu için rootogram grafiği.



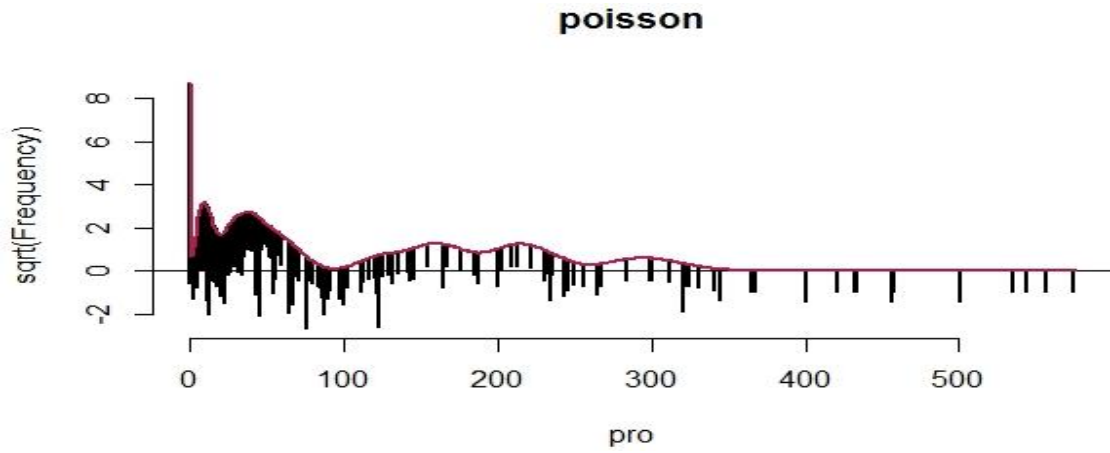
Şekil 4.18. Sıfır yayımlı negatif binomial regresyonu için rootogram grafiği.



Şekil 4.19. Sıfır yayımlı Poisson regresyonu için rootogram grafiği.



Şekil 4.20. Sıfır yayımlı negatif binomial regresyonu için rootogram grafiği.



Şekil 4.21. Poisson regresyonu için rootogram grafiği.

4.2.4. Veri setindeki sıfır gözlemlerinin %75'i silindiğinde elde edilen parametre tahminleri

Bağımlı değişkende sıfır gözlemlerin %75'i rasgele silindiğinde, veri setini en iyi açıklayan doğrusal olmayan modeli belirlemek için elde edilen AIC ve BIC uyum ölçütleri ile Vounç istatistiği Çizelge 4.7'de verilmiştir. AIC ve BIC model uyum ölçütlerine bakıldığında, en küçük değerlerin sıfır yayımlı negatif binomial hurdle regresyonunda elde edilmiştir. Vounç istatistiği sonucunda da en iyi modelin sıfır yayımlı negatif binomial hurdle regresyon olduğu saptanmıştır.

Çizelge 4.7. Gözlemlerin %75'i silindiğinde model uyum ölçütleri

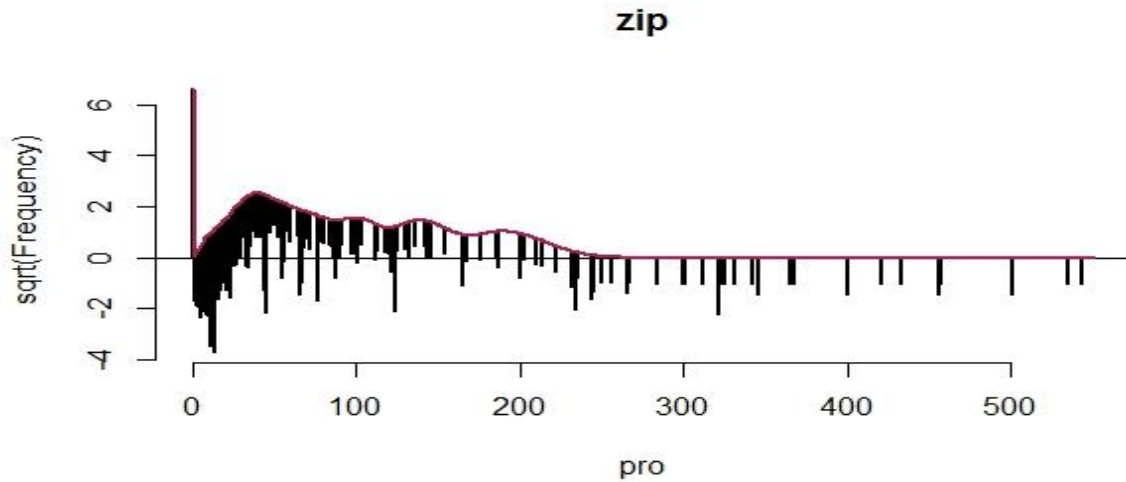
MODEL	AIC	BIC	VUONG
Poisson(P)	25061	25099	NB>P
Negatif Binomial(NB)	5058	5126	-
Zero Inflated Poisson(ZIP)	7285	7317	ZINB >ZIP
Zero Inflated Negative Binomial (ZINB)	4997	4839	-
Hurdle Poisson(H-P)	6823	6935	ZINB >H-P
Hurdle Negative Binomial(H-NB)	5046	5078	ZINB>H-NB

Tüm sıfır gözlemler modeldeyken, veri setini en iyi açıklayan model sıfır yayımlı negatif binomial hurdle regresyonuna ilişkin parametre tahminleri Çizelge 4.6'da verilmiştir. Ay, çeşit, nem ve sıcaklık bağımsız değişkenlerin bağımlı değişken üzerindeki etkileri istatistiksel olarak önemli bulunmuştur ($p<0.001$).

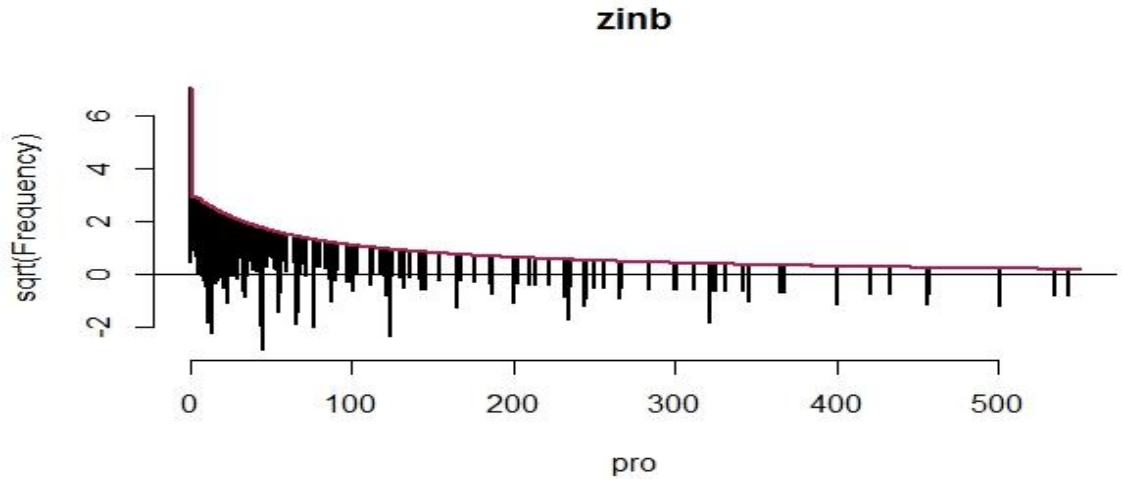
Çizelge 4.8. Sıfır yayımlı negatif binomial regresyonu için parametre tahminleri

Sayım modeli için parametre tahminleri (truncated negbin with log link):			
Bağımsız değişken	Parametre Tahmini	Standart Hata	z değeri
Sabit	5.21055	1.38818	3.754 ***
Çeşit	-0.24933	0.08941	-2.789 **
Aylar	-0.65302	0.05315	-12.286 ***
Sıcaklık	0.36301	0.02942	12.337 ***
Nem	-0.06199	0.01342	-4.619 ***
Log(theta)	0.05978	0.06845	0.873
Sıfır gözlemler için Hurdle negatif binomial için parametre tahminleri (binomial with logit link):			
Bağımsız değişken	Parametre Tahmini	Standart Hata	z değeri
Sabit	-97.3654	35.7635	-2.722 **
Sıcaklık	3.3927	0.8003	4.240 ***
Nem	0.4144	0.2993	1.385

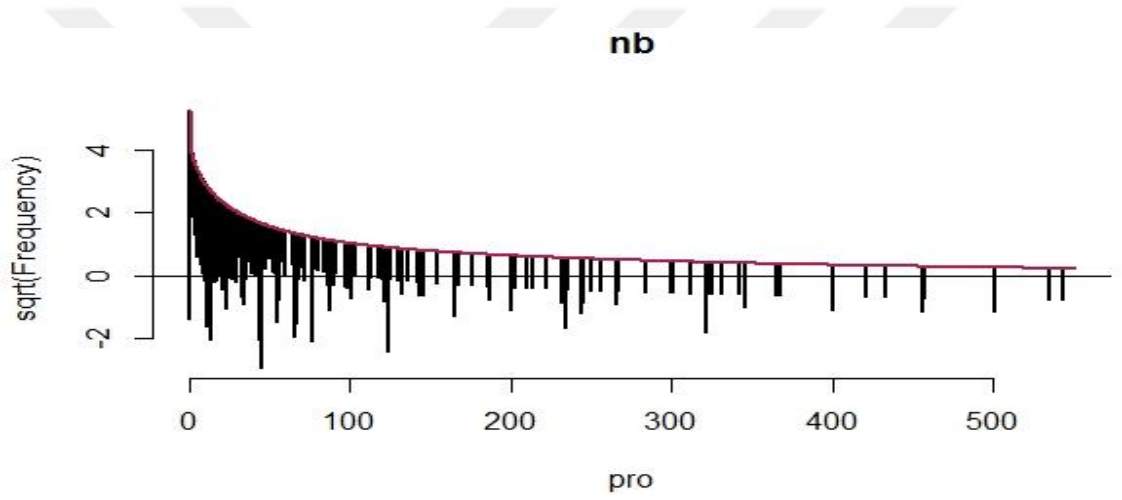
Rootogram grafiklerinde barlar tahmin edilen ve gözlenen akar sayılarının kareköklerinin farkını göstermektedir. Çizgiler ise tahmin edilen akar sayısının karekökü olarak ifade edilmektedir. Barlar x eksenine ne kadar yakınsa model o kadar iyidir şeklinde yorum yapılabilir. Poisson modeline ait grafikte diğer modellere göre barların sıfıra daha uzakta olduğu ve sıfır yayımlı hurdle negatif binomial regresyonu modelinde ise en yakın olduğu görülmektedir (Şekil 4.21, Şekil 4.26). Bu çalışmada kullanılan modellere ait rootogram grafikleri uyum ölçütleri ile ulaşılan karşılaştırma sonuçlarını desteklemektedir.



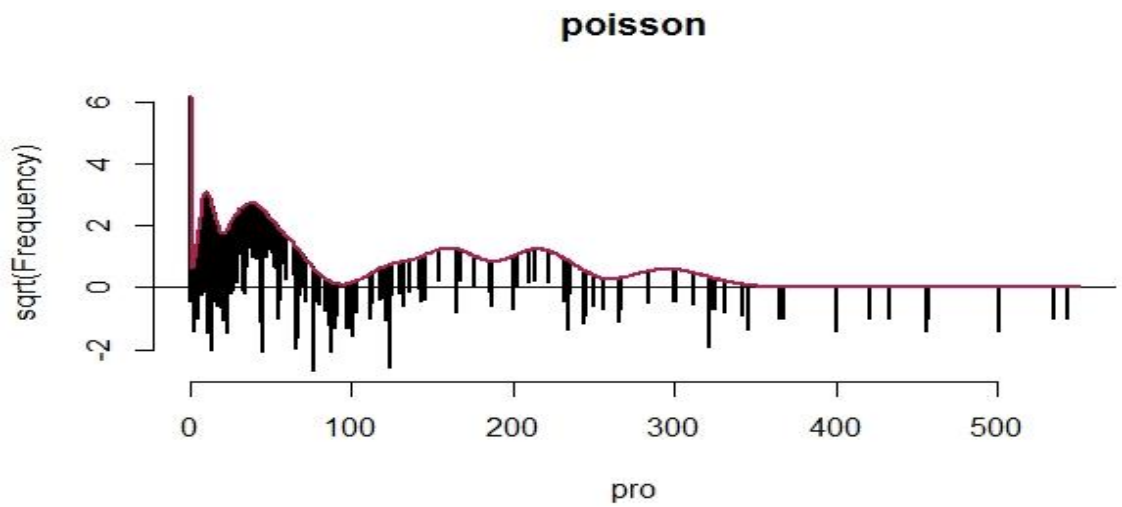
Şekil 4.22. Sıfır yayımlı Poisson regresyonu için rootogram grafiği.



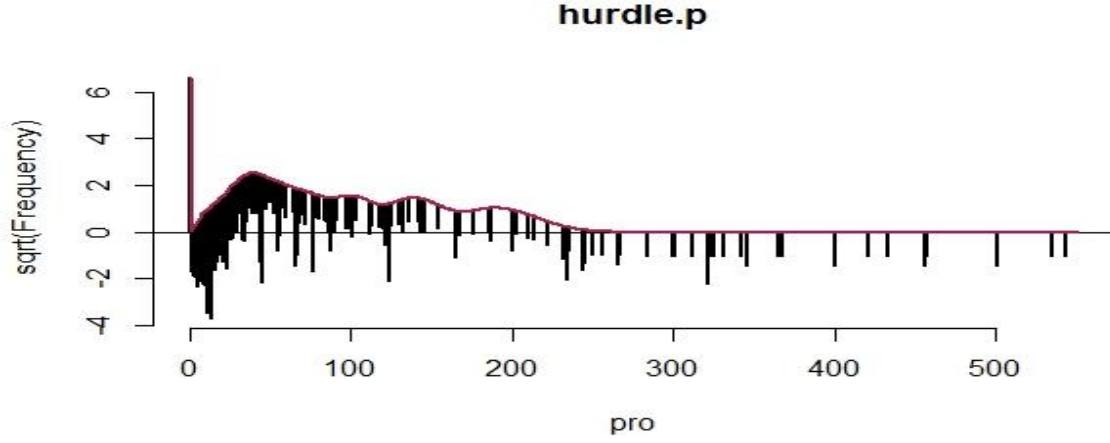
Şekil 4.23. Sıfır yayımlı negatif binomial regresyonu için rootogram grafiği.



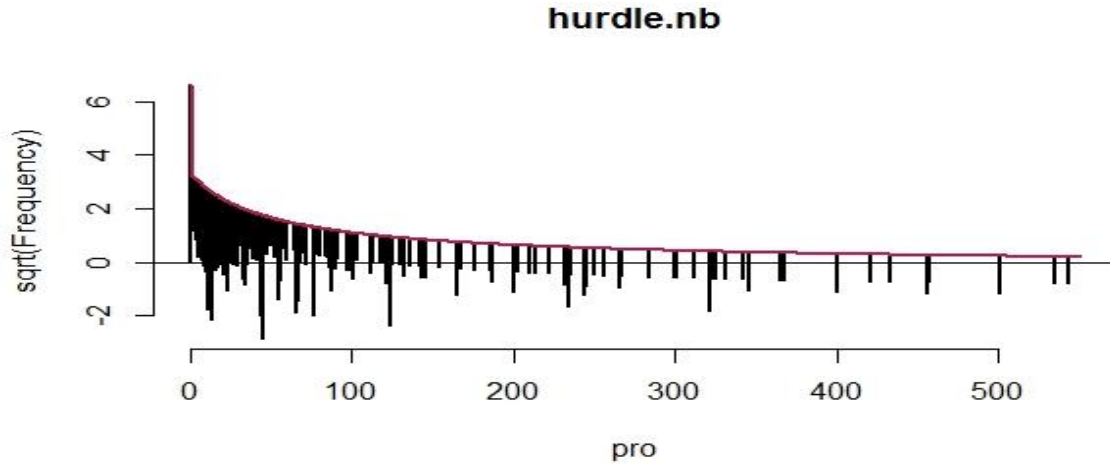
Şekil 4.24. Negatif binomial regresyonu için rootogram grafiği.



Şekil 4.25. Poisson binomial regresyonu için rootogram grafiği.



Şekil 4.26. Sıfır yayımlı Poisson regresyonu için rootogram grafiği.



Şekil 4.27. Sıfır yayımlı negatif binomial regresyon için rootogram grafiği.

4.2.5. Veri集中的 sıfır gözlemlerinin tamamı silindiğinde elde edilen parametre tahminleri

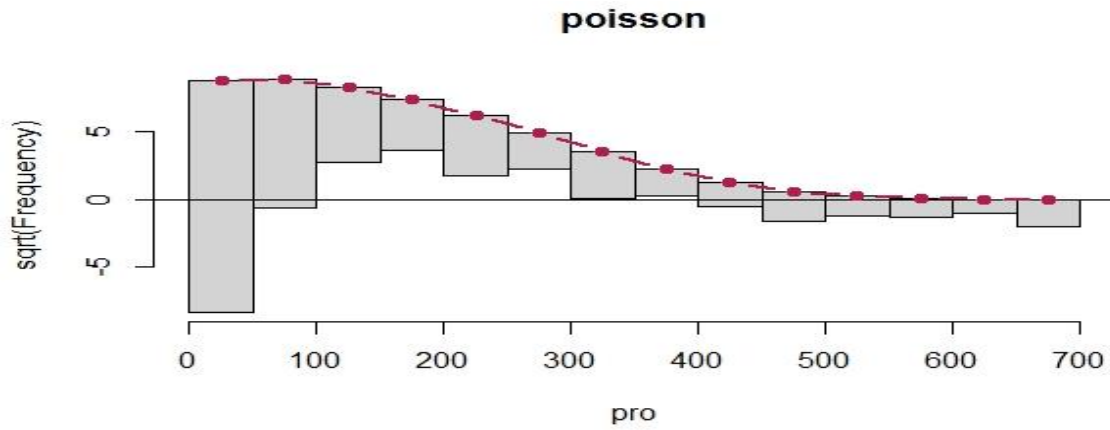
Çizelge 4.9. Sıfır gözlemlerin tamamı silindiğinde model uyum ölçütleri

MODEL	AIC	BIC
Poisson(P)	5787	5825
Negatif Binomial(NB)	5028	5053
Zero Truncated Poisson(ZTP)	24858	24892
Zero Truncated Negative Binomial (ZTNB)	4733	4770

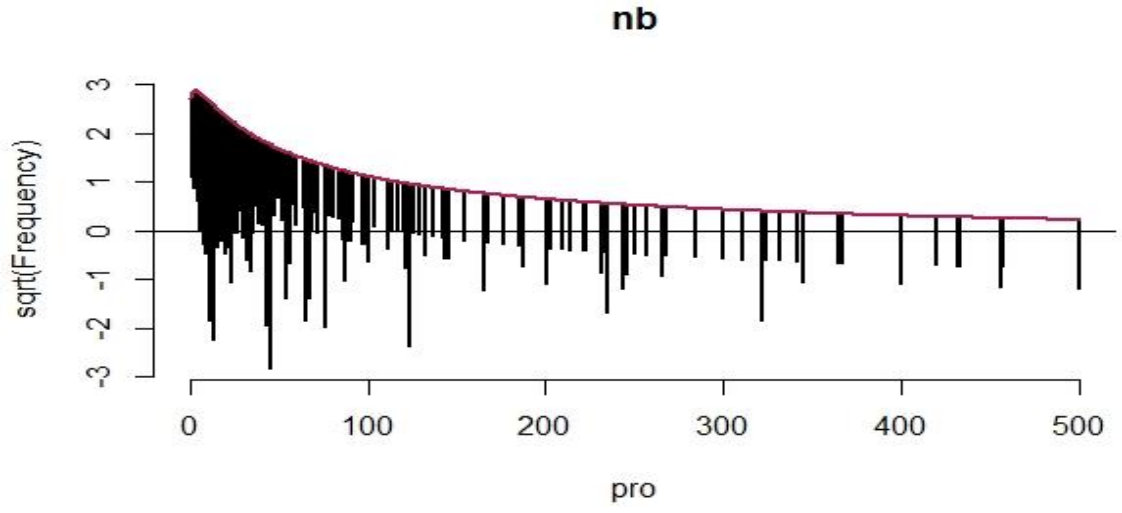
Çizelge 4.10. Sıfır kesilmiş negatif binomial regresyonu için parametre tahminleri

Bağımsız değişken	Parametre Tahmini	Standart Hata	z-değeri
Sabit1	-5.75661	18.62149	-0.309
Sabit2	0.63768	0.06766	9.424 ***
Çeşit	-0.199881	0.06798	-2.939 **
Ay1	9.50639	18.57829	0.512
Ay2	11.08883	18.57890	0.597
Ay3	9.80338	18.57897	0.528
Ay4	8.38652	18.57856	0.451
Sıcaklık	0.04086	0.02887	1.415
Nem	-0.01397	0.01351	-1.034

Rootogram grafiklerinde barlar tahmin edilen ve gözlenen akar sayılarının kareköklerinin farkını göstermektedir. Çizgiler ise tahmin edilen akar sayısının karekökü olarak ifade edilmektedir. Barlar x eksenine ne kadar yakınsa model o kadar iyidir şeklinde yorum yapılabilir. Poisson modeline ait grafikte diğer modellere göre barların sıfıra daha uzakta olduğu ve sıfır yayılımı negatif binomial hurdle regresyonu modelinde ise en yakın olduğu görülmektedir (Şekil 4.28, Şekil 4.29). Bu çalışmada kullanılan modellere ait rootogram grafikleri uyum ölçütleri ile ulaşılan karşılaştırma sonuçlarını desteklemektedir.



Şekil 4.28. Poisson regresyonu için rootogram grafiği.



Şekil 4.29. Negatif binomial regresyonu için rootogram grafiği.



5. TARTIŞMA VE SONUÇ

Bu çalışmada, Poisson dağılımı gösteren bağımlı değişkendeki sıfır dağılımına göre Poisson ve türevleri olan doğrusal regresyon modellerinin uyum ölçütü bazında karşılaştırılması yapılmıştır. Böylece farklı miktarda rasgele sıfır gözlemleri silindiğinde, hangi doğrusal olmayan regresyonun ön plana çıktığı belirlenmiştir. Başlangıçta, bağımlı değişkendeki sıfır gözlemlerin oranı %26.2 (173 gözlem) olarak verilmiştir. Daha sonrasında sırasıyla; %25, %50, %75 ve %100 oranlarında rasgele sıfır gözlemler silinmiştir. Bu bağlamda, elde edilen sonuçlar Çizelge 5.11’de özetlenmiştir.

Çizelge 5.11’e bakıldığında, sıfır gözlemlerin sayısı azaldıkça AIC ve BIC uyum istatistikleri de küçülmektedir. Buna benzer yapılan birçok çalışmada bu azalmaya rastlanmaktadır (Böhning ve ark., 1999). Bağımlı değişkendeki gözlemlerin %26.2’si sıfır gözlemden oluşmasından dolayı sıfır yayılım, uyum ölçütlerine de yansımıştır. Özellikle sıfır gözlemler %50’ye kadar silindiğinde, sıfır yayımlı negatif binomial hurdle regresyonu ön plana çıkmıştır. H-NBR modeli hem AIC ve BIC uyum istatistikleri hem de Vuong istatistiği bakımından en iyi model olarak belirlenmiştir. Yapılan çalışmalarda bağımlı değişkende sıfır gözlemlerin çok olması durumunda genellikle H-NBR modeli en iyi model olarak belirlenmiştir (Xie ve ark., 2001; Xia ve ark., 2012). Özellikle sıfır gözlemlerin daha çok olduğu durumlarda H-NBR modeli daha ön plana çıkmaktadır.

Bu çalışmada bağımlı değişkendeki aşırı yayılım 59.127 olarak elde edilmiştir. Poisson ortalaması ile varyansı birbirine eşit olduğunda aşırı yayılım değeri 1’e eşit olmaktadır. Bu çalışmada bağımlı değişkendeki aşırı yayılım bu bahsedilen değerden çok büyük olarak elde edilmiştir. Bu bağlamda, bağımlı değişkendeki sıfır yayılımın yanında bir de gözlemler arasında da çok farklılıklar olduğunu göstermiştir. Çizelge 5.11’e bakıldığında aşırı yayılım, regresyon modellerinin uyum ölçütlerine de yansımıştır. Özellikle Poisson modelde uyum ölçütleri diğerlerine nazaran çok yüksek bulunmuştur. NBR modeli aşırı yayılımı modele dahil ettiğinden dolayı AIC ve BIC uyum ölçütleri sırasıyla 25519’dan 5448’e ve 25559’dan 5474 düşmüştür. Zaten negatif binomial esaslı modellere bakıldığında hem aşırı yayılım hem de sıfır yayılım

bakımından en etkili modeller olarak elde edilmişlerdir (Martin ve ark., 2005; Czado ve Min, 2006).

Sıfır gözlemlerin çok azalması durumunda (%75'i silindiğinde), NBR modeli H-NBR modelinden daha iyi sonuç vermiştir. Çünkü sıfırlarında sayısı çok azaldığından dolayı, geriye aşırı yayılım kalmaktadır. NBR modeli de aşırı yayılımı modellediğinden dolayı en iyi model seçilmiştir. Sıfır gözlemlerin tamamı silindiği durumda ise sıfır değer kesilmiş negatif binomial regresyon en iyi model seçilmiştir. Kesilmiş modellerin esası, sıfır gözlemlerin tamamını işlem dışı bırakmaktır (Belasco ve Ghosh, 2008; Yu ve ark., 2013). Bu nedenle bağımlı değişkende sıfır gözlem olmadığı durumlarda, kesilmiş modeller en iyi model olmaktadır.

Çizelge 5.11. Model uyum ölçütlerinin regresyon modellerine göre dağılımı

Sıfır gözlemlerin oranı	Regresyon modeli	Uyum ölçütleri		En iyi model
		AIC	BIC	
Ham veri	PR	25519	25559	H-NBR
	NBR	5448	5474	
	ZIP	7367	7371	
	ZINB	5085	5126	
	H-PR	7056	7068	
	H-NBR	5062	5111	
%25'i silindiğinde	PR	25308	25347	H-NBR
	NBR	5394	5420	
	ZIP	7325	7349	
	ZINB	5078	5118	
	H-PR	7038	7051	
	H-NBR	5058	5107	
%50'si silindiğinde	PR	25184	25223	H-NBR
	NBR	5302	5328	
	ZIP	7317	7324	
	ZINB	5072	5111	
	H-PR	7020	7038	
	H-NBR	5026	5097	
%75'i silindiğinde	PR	25061	25099	ZINB
	NBR	5058	5126	
	ZIP	7285	7317	
	ZINB	4997	4839	
	H-PR	6823	6935	
	H-NBR	5046	5078	
%100'ü silindiğinde	PR	5787	5825	ZTNB
	NBR	5028	5053	
	ZTP	24858	24892	
	ZTNB	4733	4770	

6. KAYNAKLAR

- Achim, Z., Christian, K., Simon, J., 2007. Regression Models for Count Data in R. Hilbe, J., M., 2007. **Negative Binomial Regression**. Cambridge, UK.
- Agresti, A., 1997. **Categorical Data Analysis**. New Jersey, Canada; John and Wiley & Sons, Incorporation
- Belasco, E. J., Ghosh, S. K., 2008. Modeling censored data using mixture regression models with an application to cattle production yields. **Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting**, Orlando, Florida, July 27 - 29, 2008.
- Bhattacharya, S., Srinivasan, P., Polgreen, P., 2017. Social media engagement analysis of US Federal health agencies on Facebook. **BMC Medical Informatics and Decision Making**, **17** (1): 1-12.
- Böhning, D., Dietz, E., Schlattmann, P., 1999. The zero-inflated Poisson model and The decayed, missing and filled teeth index in dental epidemiology. **Journal of Royal Statistical Society A**, **162**: 195–209.
- Buczkowska, S., Coulombel, N., De Lapparent, M., 2016. Euclidean versus network distance in business location: A probabilistic mixture of hurdle-Poisson models. **Hal open science**, hal-0137757v2f.
- Cameron, A.C., Trivedi, P.K., 1998. **Regression Analysis of Count Data**. New York: Cambridge University Press
- Chen, F., Ma, X., Chen, S., Yang, L., 2016. Crash frequency analysis using hurdle models with random effects considering short-term panel data. **International Journal of Environmental Research and Public Health**, **13** (11): 1043.
- Czado, C., Min, A., 2006. Testing for zero-modification in count regression models. **Statistical Modelling**: 474.
- Çankaya, E., Alpay, O., Özer, A., 2017. Accounting for Zero Inflation of Mussel Parasite Counts Using Discrete Regression Models. **Sakarya University Journal of Science**, **21** (3): 378-384.
- Dwivedi, A. K., Dwivedi, S. N., Deo, S., Shukla, R., Kopras, E., 2010. Statistical models for predicting number of involved nodes in breast cancer patients. **Health**, **2** (7): 641.
- Gao, K., Khoshgoftaar, T. M., 2007. A comprehensive empirical study of count models for software fault prediction. **IEEE Transactions on Reliability**, **56** (2): 223-236.
- Hofstetter, H., Dusseldorp, E., Zeileis, A., Schuller, A. A., 2016. Modeling caries experience: advantages of the use of the hurdle model. **Caries Research**, **50** (6): 517-526.
- Hu, M. C., Pavlicova, M., Nunes, E. V., 2011. Zero-inflated and hurdle models of count data with extra zeros: examples from an HIV-risk reduction intervention trial. **The American Journal of Drug and Alcohol Abuse**, **37** (5): 367-375.
- Iddi, S., Molenberghs, G., 2013. A marginalized model for zero-inflated, overdispersed and correlated count data. **Electronic Journal of Applied Statistical Analysis**, **6** (2): 149-165.
- Lambert, D., 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. **Technometrics**, **34** (1): 1-13.
- Liu, W., Cela, J., 2008. Count data models in SAS. **SAS Global Forum**. Paper 371.

- Martin, T. G., Wintle, B. A., Rhodes, J. R., Kuhnert, P. M., Field, S. A., Low-Choy, S. J., Possingham, H. P., 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters*, **8** (11): 1235-1246.
- McCullagh, P., Nelder, J.A., 1989. *Generalized Linear Models*. Second Edition, London, UK, Chapman and Hall.
- Miller, M. D., Miller, J., 2008. No Zero Left Behind: Comparing the fit for zero-inflation models as a function of skew and proportion of zeros. *Interstat Statistics on the Internet*.
- Moghimbeigi, A., Eshraghian, M. R., Mohammad, K., McArdle, B., 2009. A score test for zero-inflation in multilevel count data. *Computational Statistics & Data Analysis*, **53** (4): 1239-1248.
- Mouatassim, Y., 2012. Zero-modified discrete distributions for operational risk modelling. *The Journal of Risk Finance*, **13**: 476 – 490.
- Romeu, A., Vera-Hernández, M., 2005. Counts with an endogenous binary regressor: A series expansion approach. *The Econometrics Journal*, **8**(1), 1-22.
- Rose, C. E., Martin, S. W., Wannemuehler, K. A., Plikaytis, B. D., 2006. On the use of zero-inflated and hurdle models for modeling vaccine adverse event count. *J Biopharm Stat.*, **16** (4): 463-481.
- Saffari, S. E., Adnan, R., Greene, W., 2012. Hurdle negative binomial regression model with right censored count data. *Sort-Statistics and Operations Research Transactions*, 181-194.
- Sakthivel, K. M., Rajitha, C. S., 2017. A comparative study of zero-inflated, hurdle models with artificial neural network in claim count modeling. *International Journal of Statistics and Systems*, **12** (2): 265-276.
- Sarul, L. S., Sahin, S., 2015. An application of claim frequency data using zero inflated and hurdle models in general insurance. *Journal of Business Economics and Finance*, **4** (4).
- Sekata D., 2015. Modeling the number of antenatal care service visits among pregnant women in rural ethiopia: zero inflated and hurdle model specifications. *International Journal of Healthcare Sciences*, **3** (1): 332-355.
- Shiyuka, N., 2018. *Hurdle negative binomial model for motor vehicle crash injuries in Namibia (masters degrees thesis, unpublished)*. University of Namibia.
- Tin, A., 2008. Modeling zero-inflated count data with underdispersion and overdispersion. *SAS Global Forum*. Paper 372.
- Vergne, T., Calavas, D., Cazeau, G., Durand, B., Dufour, B., Grosbois, V., 2012. A Bayesian zero-truncated approach for analysing capture–recapture count data from classical scrapie surveillance in France. *Preventive veterinary medicine*, **105** (1-2): 127-135.
- Xia, Y., Morrison-Beedy, D., Ma, J., Feng, C., Cross, W., Tu, X., 2012. Modeling count outcomes from HIV risk reduction interventions: A comparison of competing statistical models for count responses. *AIDS Research and Treatment*, 593569.
- Xiao, Y., Zhang, X., Ji, P., 2015. Modeling forest fire occurrences using count-data mixed models in qiannan autonomous prefecture of Guizhou Province in China. *PloS One*, **10** (3): e0120621.
- Xie, M., He, B., Goh, T. N., 2001. Zero-inflated Poisson model in statistical process control. *Computational Statistics & Data Analysis*, **38**: 191-201.

- Yesilova, A., Kaki, B., Kasap, 2007. Sıfır deger ağırlıklı sayıma dayalı olarak elde Edilen bağımlı degiskenin modellenmesinde kullanılan regresyon yöntemler. ***İstatistik Arastırma Dergisi***, 5 (1): 1-9.
- Yeşilova, A., Denizhan, E., 2016. Modeling mite counts using poisson and negative binomial regressions. ***Fresenius Environmental Bulletin***, 25 (11): 5062-5066
- Yu, Q., Chen, R., Tang, W., He, H., Gallop, R., Crits-Christoph, P., Tu, X. M., 2013. Distribution-free models for longitudinal count responses with overdispersion and structural zeros. ***Statistics in Medicine***, 32 (14): 2390-2405.
- Zulkifli, M., Ismail, N., Razali, A. M., 2011. Zero-inflated Poisson versus zero-inflated negative binomial: Application to theft insurance data. ***The 7th IMT-GT International Conference on Mathematics, Statistics and its Applications***. Bangkok, Thailand.





ÖZ GEÇMİŞ

İlk ve orta öğrenimini Cumhuriyet İlköğretim Okulu' nda, lise öğrenimini Elmadağ Lisesi' nde (2003-2006) Ankara' nın Elmadağ ilçesinde tamamladı. Lisans eğitimini Kırıkkale Üniversitesi' nde Fen Edebiyat Fakültesi İstatistik Bölümünde 2008-2012 yılları arasında tamamladı. 2019 yılında Van Yüzüncü Yıl Üniversitesi Ziraat Fakültesinde Zootečni Anabilim Dalı' nda yüksek lisansa başladı. 2012-2017 yılları arasında Türkiye İstatistik Kurumu Van Bölge Müdürlüğü'nde İstatistikçi olarak görev yaptı. 4 Aralık 2017 yılından itibaren Türkiye Cumhuriyeti Yargıtay Başkanlığı'nda İstatistikçi olarak görev yapmaktadır. Evli ve bir çocuk annesidir.



**VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
LİSANSÜSTÜ TEZ ORJİNALLİK RAPORU**

Tarih: ../.. /... /...

Tez Başlığı / Konusu:

Yukarıda başlığı/konusu belirlenen tez çalışmamın Kapak sayfası, Giriş, Ana bölümler ve Sonuç bölümlerinden oluşan toplam ... sayfalık kısmına ilişkin, ../.../.... tarihinde şahsım/tez danışmanım tarafından Turnitin intihal tespit programından aşağıda belirtilen filtreleme uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % (yüzde ...) dir.

Uygulanan filtreler aşağıda verilmiştir:

- Materyal ve yöntem hariç,
- Kaynaklar hariç,
- Tezden çıkan yayınlar hariç,
- 7 kelimedenden daha az örtüşme içeren metin kısımları hariç (Limit inatch size to 7 words)

Van Yüzüncü Yıl Üniversitesi Lisansüstü Tez Orijinallik Raporu Alınması ve Kullanılmasına İlişkin Yönergeyi inceledim ve bu yönergede belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini bilgilerinize arz ederim.

Tarih ve İmza

Adı Soyadı:

Öğrenci No:

Anabilim Dalı:

Programı:

Statüsü: ☐ Yüksek Lisans

☐ Doktora

DANIŞMAN ONAYI
UYGUNDUR

ENSTİTÜ ONAY
UYGUNDUR