



**SAYMA VERİSİ İÇİN REGRESYON MODELLERİ: TÜRKİYE GELİR VE
YAŞAM KOŞULLARI ARAŞTIRMASI ÜZERİNE BİR UYGULAMA**

Onur ŞENTÜRK

**YÜKSEK LİSANS TEZİ
İSTATİSTİK ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

KASIM 2020

Onur ŞENTÜRK tarafından hazırlanan “SAYMA VERİSİ İÇİN REGRESYON MODELLERİ: TÜRKİYE GELİR VE YAŞAM KOŞULLARI ARAŞTIRMASI ÜZERİNE BİR UYGULAMA” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi İstatistik Ana Bilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Doç. Dr. Hülya OLMUŞ

İstatistik Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....

Başkan: Prof. Dr. Serpil AKTAŞ ALTUNAY

İstatistik Ana Bilim Dalı, Hacettepe Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....

Üye: Prof. Dr. Bülent ÇELİK

İstatistik Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....

Tez Savunma Tarihi: 23/11/2020

Jüri tarafından kabul edilen bu çalışmanın Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....

Prof. Dr. Cevriye GENCER

Fen Bilimleri Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Onur ŞENTÜRK

23/11/2020

SAYMA VERİSİ İÇİN REGRESYON MODELLERİ: TÜRKİYE GELİR VE YAŞAM
KOŞULLARI ARAŞTIRMASI ÜZERİNE BİR UYGULAMA

(Yüksek Lisans Tezi)

Onur ŞENTÜRK

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Kasım 2020

ÖZET

Bu çalışmada sayma verisinin analiz edilmesinde kullanılan bazı yöntemler incelenmiştir. Sayma verisinin olduğu veri setlerinde sıfır değerinin yığılma göstermesi sıklıkla rastlanan bir durumdur. Böyle verilerin analizinde klasik regresyon varsayımları sağlanamadığından hatalı veya yanlış sonuçlar elde etmemek için klasik regresyon modelleri kullanılmamalıdır. Bunun yerine bu verileri modellemek için Poisson Regresyon, Negatif Binom Regresyon, Sıfır Yığılmalı Poisson Regresyon, Sıfır Yığılmalı Negatif Binom Regresyon, Sıfır Kesilmiş Poisson Regresyon, Sıfır Kesilmiş Negatif Binom Regresyon, Poisson Hurdle Regresyon ve Negatif Binom Hurdle Regresyon kullanılabilir. Çalışmada sayılan bu yöntemler incelenmiştir. 2018 Türkiye Gelir ve Yaşam Koşulları Araştırması kullanılarak gerçek veri seti ile analiz yapılmıştır. Modellerin uygunluğunu karşılaştırmak amacıyla Akaike Bilgi Kriteri, Log Olabilirlik değeri ve Vuong istatistiği kullanılmıştır. Ayrıca modeller için parametre tahminleri ve yorumlar yapılmıştır. Çalışma sonucunda Sıfır Kesilmiş Negatif Binom Regresyonun veriye uygunluk açısından diğer modellerden daha iyi olduğu görülmüştür.

Bilim Kodu : 20513

Anahtar Kelimeler : Sayma verisi, Poisson Regresyon, Negatif Binom Regresyon, Sıfır Yığılmalı Regresyon Modelleri, Sıfır Kesilmiş Regresyon Modelleri, Hurdle Modeller

Sayfa Adedi : 63

Danışman : Doç. Dr. Hülya OLMUŞ

REGRESSION MODELS FOR COUNT DATA: AN APPLICATION ON INCOME AND
LIVING CONDITIONS SURVEYS IN TURKEY

(M. Sc. Thesis)

Onur ŞENTÜRK

GAZİ UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

November 2020

ABSTRACT

In this study, some methods used in analyzing the data obtained based on counting were examined. It is common for the zero value to show an accumulation in data sets with counting data. In the analysis of such data, classical regression models should not be used in order to obtain erroneous or biased results in the analysis of such data, since classical regression assumptions can not be provided. Instead, Poisson Regression, Negative Binomial Regression, Zero Inflated Poisson Regression, Zero Inflated Negative Binomial Regression, Zero-Truncated Poisson Regression, Zero-Truncated Negative Binomial Regression, Poisson Hurdle Regression and Negative Binomial Hurdle Regression can be used. These methods listed in the study were examined. Analysis was made with real data set using Turkey Income and Living Conditions Survey 2018. Akaike Information Criteria, Log Likelihood value and Vuong statistics were used to compare the fit of the models. In addition, parameter estimates and comments were made for models. As a result of the study, it was seen that Zero-Truncated Negative Binomial Regression is better than other models in terms of data compatibility.

Science Code : 20513

Key Words : Count data, Poisson Regression, Negative Binomial Regression, Zero Inflated Regression models, Zero Truncated Regression models, Hurdle models

Page Number : 63

Supervisor : Assoc. Prof. Dr. Hülya OLMUŞ

TEŞEKKÜR

Yoğun çalışma temposuna rağmen benden yardımlarını, desteğini, sabrını ve bilgisini esirgemeyen değerli hocam Doç. Dr. Hülya OLMUŞ' a ve her zaman yanımda olup destekleri ile beni cesaretlendiren ve bugünlere getiren aileme sonsuz teşekkür ederim.

Tezimi aramızdan ayrılan sevgili babam Ali Ekber ŞENTÜRK' e gururla ithaf etmek isterim.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	ix
SİMGELER VE KISALTMALAR.....	xi
1. GİRİŞ.....	1
2. SAYMA VERİLERİ İÇİN REGRESYON MODELLERİ.....	7
2.1. Sayma Veri Modelleri	7
2.1.1. Poisson regresyon modeli	8
2.1.2. Negatif binom regresyon modeli.....	10
2.2. Sıfır Yığılmalı Sayma Veri Modelleri.....	12
2.2.1. Sıfır yığılmalı poisson regresyon modeli	13
2.2.2. Sıfır yığılmalı negatif binom regresyon modeli	14
2.3. Kesilmiş Sayma Modelleri	15
2.4. Sıfır Kesilmiş Sayma Modelleri	16
2.4.1. Sıfır kesilmiş poisson regresyon	17
2.4.2. Sıfır kesilmiş negatif binom regresyon	18
2.5. Hurdle Modeller	19
2.5.1. Poisson hurdle regresyon modeli	20
2.5.2. Negatif binom hurdle regresyon modeli (NBH)	20
2.6. Model Seçimi	21
3. UYGULAMA.....	25
3.1. Veri Seti	25

	Sayfa
3.2. Betimleyici İstatistikler	26
3.3. Uygulama Sonuçları.....	35
4. SONUÇ VE ÖNERİLER	55
KAYNAKLAR	57
ÖZGEÇMİŞ.....	63

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 3.1. Kullanılan değişkenlere ilişkin temel bilgiler.....	26
Çizelge 3.2. Konut sayısına ilişkin frekans tabloları	27
Çizelge 3.3. Yaş değişkeni düzeyine ilişkin frekans dağılımı	27
Çizelge 3.4. Konut sayısı ve yaş değişkeni düzeyine ilişkin frekans dağılımı	28
Çizelge 3.5. Gelir değişkenine ilişkin frekans dağılımı.....	29
Çizelge 3.6. Gelir değişkeni ve konut sayısına ilişkin frekans dağılımı.....	29
Çizelge 3.7. Eğitim durumu değişkenine ilişkin frekans dağılımı.....	30
Çizelge 3.8. Konut sayısı ve eğitim durumu değişkenine ilişkin frekans dağılımı.....	31
Çizelge 3.9. Çalışma durumu değişkenine ilişkin frekans dağılımı.....	32
Çizelge 3.10. Konut sayısı ve çalışma durumu düzeyine ilişkin frekans dağılımı	32
Çizelge 3.11. Geçinme durumu düzeyine ilişkin sayı ve oranlar	33
Çizelge 3.12. Konut sayısı ve geçinme durumu düzeyine ilişkin frekans dağılımı.....	33
Çizelge 3.13. Cinsiyet değişkenine ilişkin frekans dağılımı.....	34
Çizelge 3.14. Konut sayısı ve cinsiyet düzeyine ilişkin frekans dağılımı	35
Çizelge 3.15. PR modeli parametre tahminleri.....	36
Çizelge 3.16. NB regresyon modeli parametre tahminleri	38
Çizelge 3.17. ZIP regresyon modeli parametre tahminleri (log kısım)	39
Çizelge 3.18. ZIP regresyon modeli parametre tahminleri (lojit kısım).....	41
Çizelge 3.19. ZINB regresyon modeli parametre tahminleri (log kısım)	42
Çizelge 3.20. ZINB regresyon modeli parametre tahminleri (lojit kısım).....	44
Çizelge 3.21. ZTP regresyon modeli parametre tahminleri.....	45
Çizelge 3.22. ZTNB regresyon modeli parametre tahminleri.....	47
Çizelge 3.23. PH regresyon modeli parametre tahminleri (log kısım)	48
Çizelge 3.24. PH regresyon modeli parametre tahminleri (lojit kısım).....	49

Çizelge	Sayfa
Çizelge 3.25. NBH regresyon modeli parametre tahminleri (log kısım).....	51
Çizelge 3.26. NBH regresyon modeli parametre tahminleri (lojit kısım).....	52
Çizelge 3.27. Regresyon modellere ilişkin AIC ve LL değerleri	53
Çizelge 3.28. Vuong test sonuçları	54

SİMGELER VE KISALTMALAR

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

Simgeler

Açıklamalar

λ

Ortalama

α

Aşırı Yayılım Parametresi

β

Regresyon Katsayıları

Kısaltmalar

Açıklamalar

AIC

Akaike Bilgi Kriteri

BIC

Bayesian Bilgi Kriteri

GYKA

Gelir ve Yaşam Koşulları Araştırması

LL

Log Olabilirlik

PR

Poisson Regresyon

PH

Poisson Hurdle Regresyon

NB

Negatif Binom Regresyon

NBH

Negatif Binom Hurdle Regresyon

TÜİK

Türkiye İstatistik Kurumu

ZIP

Sıfır Yığılmalı Poisson Regresyon

ZINB

Sıfır Yığılmalı Negatif Binom Regresyon

ZTP

Sıfır Kesilmiş Poisson Regresyon

ZTNB

Sıfır Kesilmiş Negatif Binom Regresyon

1. GİRİŞ

İstatistik alanında veriyi analiz etmede kullanılan farklı yöntemler vardır. Regresyon analizi bu yöntemlerin başında gelmektedir. Regresyon analizinde kullanılan bağımlı değişkene bağlı olarak farklı modeller oluşturulabilmektedir. Burada temel ayrımlardan bir tanesi bağımlı değişkenin sürekli ya da kesikli dağılımdan gelmesidir. Daha sonra klasik regresyon varsayımlarına uyup uymadığı kontrol edilir. Kesikli dağılımdan gelen bir bağımlı değişken var ise bu değişken sayılabilir bir değer almaktadır. Böyle sayılabilir değer alan modellerde sayma verisi analizinde kullanılan yöntemleri tercih etmek daha uygun olmaktadır. Sayma verisi negatif değerler almazlar. Böyle verilere ekonometri, eğitim, psikoloji, aktüerya, demografi gibi farklı disiplinlerde rastlamak mümkündür.

Sayma verisinin analiz edilmesinde Poisson regresyon, negatif binom regresyon, sıfır yığılmalı regresyon, sıfır kesilmiş regresyon ve hurdle regresyon modelleri uygulanabilir. Eğer bağımlı değişken Poisson dağılımından geliyorsa Poisson regresyon seçilebilir. Ancak burada önemli olan Poisson regresyonda ortalama ve varyansın birbirine eşit olması şartının sağlanmasıdır. Gerçek hayatta bu şartın sağlanması pek mümkün olmamaktadır. Ortalamanın varyanstan büyük olması aşırı yayılım (overdispersion) veya ortalamanın varyanstan küçük olması az yayılım (underdispersion) durumları görülmektedir. Eğer bağımlı değişken negatif binom dağılımından geliyorsa ve aşırı yayılımda gösteriyorsa bu durumda negatif binom regresyon modeli tercih etmek daha doğru olacaktır. Çünkü negatif binom dağılımında aşırı yayılımı gösteren ekstra bir parametre vardır (Sileshi, 2008).

Sayma dayalı olarak elde edilen verilerde sıklıkla gözlenen diğer bir durum verinin çok sayıda sıfır içermesidir. Böyle sıfır yığılmanın olduğu verilerde klasik Poisson ve klasik Negatif Binom regresyon modelleri uygulamak yanlış parametre tahminlerine ve hatalı sonuçlara yol açacaktır. Sıfır yığılma olduğu durumda veri, sıfırların olduğu grup ayrı diğer sayma verisinin olduğu grup ayrı bir grup olacakmış gibi iki şekilde analiz edildiği düşünülür. Sayma verisinden oluşan kısım eğer Poisson dağılımından geliyorsa Sıfır Yığılmalı Poisson Regresyon (Zero Inflated Poisson Regression-ZIP) kullanmak daha uygundur. Eğer veride aşırı yayılım durumu varsa ve bağımlı değişken negatif binoma uygun ise Sıfır Yığılmalı Negatif Binom Regresyon (Zero Inflated Negative Binomial Regression-ZINB) kullanılır (Hall,2000).

Sıfır kesilmiş (Zero Truncated) regresyon modellerinde de sıfır yığılmalı modellere benzer bir yaklaşım vardır. Bağımlı değişkenin geldiği dağılıma göre isimler alır. Burada sıfırdan kesilme, bağımlı değişkenin 0,1,2,... gibi değerler alırken 0 değerlerinin kesilerek verinin kalan pozitif değerler ile modellenmesini ifade etmektedir. Kalan pozitif değerler Poisson dağılımına uygunluk gösteriyorsa Sıfır Kesilmiş Poisson Regresyon (Zero Truncated Poisson Regression-ZTP) modeli tercih edilirken, veride sıfır yığılma varsa ve bağımlı değişken negatif binom dağılımından geliyorsa Sıfır Kesilmiş Negatif Binom Regresyon (Zero Truncated Negative Binomial Regression-ZTNB) modeli kullanılır (Breen,1996).

Sayma verisinin analizinde kullanılan diğer bir yöntem Hurdle regresyon modelleridir. Bu modeller iki aşamalı modellerdir. İlk aşama geçiş aşaması (transition stage) olarak kabul edilir ve bu aşamada 0 değerlerine karşı diğer pozitif değerleri gösteren bir sistem vardır. Sıfırlar 0 diğer pozitif değerler 1 şeklinde bir ikili (binary) gösterim vardır. Son aşama ise olay aşaması (event stage) yani sadece pozitif değerlerin olduğu aşamadır. Sıfır yığılmalı modellere benzer şekilde olay aşamasındaki değerler Poisson dağılımına uyuyorsa Poisson Hurdle (PH) regresyon modeli, negatif binom dağılımına uyuyorsa ve aşırı yayılım varsa ise Negatif Binom Hurdle (NBH) regresyon modeli olarak adlandırılır (Hilbe, 2014). Hurdle modeller bakıldığında sıfır yığılmalı modellere benzemektedir. Ancak hurdle modeller sıfır yığılma olmadığı durumlarda da tercih edilebilir.

Bahsedilen sayma regresyon modelleri dışında sansürlü regresyon modelleri, sıfır yığılmalı gama regresyon modelleri, sıfır yığılmalı Poisson- Lindley regresyon modelleri gibi farklı sayma modelleri de vardır. Literatürde sayma verisi kullanılarak yapılan çalışma sayısı her geçen gün artmaktadır. İlk olarak ZIP model Lambert (1992) tarafından, bir uygulama çalışması olan baskı devre kartlarının üretiminde gözlenen lehimleme hata sayılarının bir model olarak ifade edilmesinde kullanılmıştır. Lambert bu çalışmasında ZIP modelin kullandığı veriyi daha iyi açıkladığını görmüştür. ZINB modeli ise Greene (1994) tarafından önerilmiştir. Greene aşırı yayılım durumu ve sıfır yığılma olduğu durumlarda kullanılacak yeni bir yöntem olarak ZINB modelini sunmuştur. Bu modeli tüketici kredisine ilişkin bir veri setinde uygulamıştır. Carrivick ve arkadaşları (2003), 600 yataklı bir hastanede temizlik işçileri arasında mesleki yaralanma riskini azaltmada işyeri risk değerlendirme ekibinin etkinliğini değerlendirmek için ZIP modeli kullanmışlardır. ZIP modelin verilere uyum sağladığı ve risk değerlendirme ekiplerinin çalışmaları ile temizlik işçilerinin yaralanma oranlarını azalttıklarını görmüşlerdir. Sezgin ve Deniz (2004), 1964-2000 yılları arasında

Türkiye’de görülen grev sayılarını etkileyen faktörleri sıfır yığılmalı modeller kullanarak incelemiştir. Martin ve arkadaşları (2005), belirli bir kuş türünün sayısının modellenmesinde sıfır yığılmalı modeller ve Hurdle modeller kullanmışlardır. Bu çalışmalarında sıfır yığılmayı göz ardı ederlerse parametre tahminlerinde önemli ölçüde farklılıklar olacağını göstermişlerdir. Simons ve arkadaşları (2006), 292 üniversite öğrencisi arasında esrar ilgili sorunlar ile sosyal normlar, dürtüsellik ve kullanım programı arasındaki ilişkiyi incelemiştir. Bu ilişkiyi modellemede de sıfır yığılmalı modellerden ZINB modelini kullanmışlardır. Elde ettikleri sonuçların genel olarak sosyal ve çevresel problemlere ilişkin teorilere uygun olduğunu görmüşlerdir. Famoye ve Singh (2006), aile içi şiddet içeren bir veri için ZIP ve genelleştirilmiş ZIP kullanmışlardır. En çok olabilirlik yöntemini kullanarak model parametrelerinin tahminini sağlamışlardır. Özmen ve Famoye (2007), güneşe maruz kalmaktan ölen caretta yavrularının sayısını yaptıkları bu çalışmalarında sıfır yığılmalı modellerle modellemiştir. Poisson, NB, genelleştirilmiş Poisson, ZIP ve sıfır yığılmalı genelleştirilmiş Poisson modellerini veri setine uygulamışlardır. Genelleştirilmiş Poisson regresyon modelinin daha iyi performans gösterdiğini görmüşlerdir. Puza ve arkadaşları (2008), kesilmiş verilerin regresyon analizine Bayesci bir yaklaşım sunmuşlardır. Hollanda’da yasadışı göçmen verilerine kesilmiş Poisson regresyon yöntemi uygulamıştır. Flynn ve Francis (2009), veri setlerinde sayma verisi olduğu durumda fazla sayıda sıfır olmasının ve doğrusallık varsayımının sağlanmaması sorunlarını ele almışlardır. Sigortacılık alanında hasar sayısının modellenmesinde sıfır yığılmalı modelleri kullanmışlardır. Yang ve arkadaşları (2009), bazı aykırı değer belirleme yöntemleri ve bu yöntemlerin uygulamaları ile sonuçlarını öne sürmüşlerdir. Aykırı değerlerin etkisini ortadan kaldırmak için kesilmiş veya Winsorized ortalamaya dayalı parametre tahminleri önermişlerdir. Sıfır yığılmalı Poisson modelde simülasyon çalışması yaparak sonuçları sunmuşlardır. Hu ve arkadaşları (2011), Poisson, NB, ZIP, PH, ZINB, NBH modellerinin özelliklerini incelemiştir. Altı model, toplum temelli kötü madde kullanımı tedavisine katılan kadın hastalar (N=515) arasında daha önce tamamlanmış bir klinik araştırmadan elde edilen veriler kullanılarak, HIV risk azaltma müdahalesinin korunmasız cinsel olayların sayısı üzerindeki etkisini modellemek için kullanmışlardır. Her modelden elde edilen uyum iyiliği ve tedavi etkisinin tahminleri karşılaştırmışlar ve ZINB modelinin en uygun model olduğunu görmüşlerdir. Nasiri (2011), karma poisson modeli sağlık alanında bir uygulama çalışmasında kullanmıştır. Sağlık hizmetleri ve hastane kabulleri verilerini ele almış ve veride çok sayıda sıfır olduğunu gözlemlemiştir ve bu sorunun üstesinden gelmek için kesilmiş dağılımlar kullanmıştır. Kaya ve Yeşilova (2012), yaptıkları çalışmalarında Yüzüncü Yıl Üniversitesi

(YYÜ) e-posta sunucusundan personelin 2009 bahar eğitim öğretim döneminde yaptıkları e-posta trafiğini incelemişlerdir. Veride sıfır yığılma olmasından dolayı veriye ZIP, ZINB, PH ve NBH regresyon yöntemleri uygulamışlardır. ZINB ve NBH regresyon modellerinin veri için daha uygun olduğu görmüşlerdir. Asrul ve Naingb (2012), çalışmalarında AIDS hastalarının ölüm oranlarını yaş üzerinden modellemede cinsiyet, milliyet, ırk, medeni durum, meslek ve taşıyıcılık şekli değişkenlerinin etkisini sıfır yığılmalı regresyon modelleri kullanarak ortaya koymuşlardır. Bu çalışmalarında, sapma ve Pearson Ki-Kare uyum iyiliği istatistiğini kullanarak veride aşırı yayılım olmadığını göstermişlerdir. Uygun regresyon modelinin seçiminde Akaike Information Criteria (AIC) ve Bayesian Information Criteria (BIC) kullanmışlardır. En uygun modelin ZINB modeli olduğunu göstermişlerdir. Sharma ve Landge (2013), çalışmalarında Hindistan'da yer alan bir karayolu üzerindeki kaza sayılarını sıfır yığılmalı modeller kullanarak incelemişlerdir. Sıfır yığılmalı modellerin, çarpışma sıklığını tahmin etmek için güçlü bir araç olduğu banket genişliği, şerit genişliği, erişim yoğunluğu, nokta hızı ve yıllık ortalama günlük trafik değişkenlerini ele alarak göstermişlerdir. Model uyum iyiliğini ölçmek için ise AIC değerini kullanmışlardır. Kong ve arkadaşları (2014), Sıfır yığılma olan sayma verisinde genelleştirilmiş tahmin denklemlerini (GEE) kullanılabileceğini öne sürmüşler ve sıfır yığılmalı negatif binom (ZINB) modelini kullanmışlardır. Iowa Florür Çalışmasından çıkarılan bir veri seti kullanılarak diş çürükleri ve florüre maruz kalma arasındaki ilişkiyi bahsedilen yöntem ile incelemişlerdir. Wang ve arkadaşları (2015), sağlık hizmeti araştırmalarında genellikle çok sayıda sıfır olduğunu öne sürmüşler ve bu makale için Almanya'da sağlık hizmetleri çalışmalarını sıfır yığılmalı negatif binom model kullanarak incelemişlerdir. Model parametrelerini tahmin etmek için EM algoritmasını önermişlerdir. Beaujean ve Morgan (2016), eğitim alanından kullandıkları veri ile sıfır değer ağırlıklı veri seti için uygun model seçimi hakkında çalışma sunmuşlardır. Chowdhury ve Islam (2016), sıfır kesilmiş iki değişkenli Poisson modeli önermişlerdir. Bunun için marjinal koşullu bir yaklaşım kullanmışlardır. Karayolu verileri üzerinde çalışmışlardır ve iki değişkeni olarak kazaya karışan araç sayısı ile kazazedelerin sayısını almışlardır. Tüzen ve Erbas (2017), Türkiye'deki gençler arasında günlük sigara tüketimiyle ilişkili faktörleri incelemek için Poisson (P), negatif binom (NB), sıfır yığılmalı Poisson (ZIP), sıfır yığılmalı negatif iki terimli (ZINB), Poisson Hurdle (PH) ve negatif binom hurdle (NBH) regresyonlarını sigara tüketim sayım verilerine uygulamışlardır. Buna göre ZINB ve NBH modellerinin tercih edilmesi gerektiğini göstermişlerdir ve Türkiye'de 15-24 yaş grubundaki sigara içenler ve içmeyenler için cinsiyet, istihdam ve evde tütün kullanımının daha etkili faktörler olduğunu

öne sürmüşlerdir. Altun (2018), yapmış olduğu çalışmada sıfır yığılmalı modellere yeni bir yaklaşım olarak sıfır yığılmalı Poisson- Lindley modelini önermiştir. Poisson dağılımının aksine, Poisson-Lindley dağılımı aşırı yayılıma izin vermesi nedeniyle, bu modelin aşırı yayılımlı ve sıfır yığılmalı verileri modellemek adına iyi bir seçenek olduğunu öne sürmüştür. Önerdiği modelin gerçek veri üzerinde uygulama çalışmasını yapmıştır ve Poisson ve Zero-Inflated Poisson regresyon modelleriyle karşılaştırmıştır. Sıfır yığılmalı Poisson-Lindley regresyon modelinin, sıfır yığılmalı ve aşırı yayılım gösteren veri seti için, sıfır yığılmalı Poisson regresyon modelinden daha iyi uyum sağladığını göstermiştir. Olmuş ve Karaca (2018), herhangi bir hizmet sektöründe çalışan personelin müşterilerden şikâyet alma sayısını cinsiyet, yaş, eğitim ve tecrübe değişkenlerinin etkilediği düşünmüşlerdir ve sıfır yığılmalı yöntemleri kullanarak veriyi analiz etmişlerdir. Ayrıca, Akaike Bilgi Kriteri değerlendirmede kullanmışlardır parametre tahminlerine ilişkin yorumlamalar yapmışlardır. Kim ve arkadaşları (2019), yapmış oldukları çalışmalarında Güney Kore’de iklim değişikliği sonucu aşırı ısıya maruz kalan kişilerde yaşanan ölüm sayısını tahmin etmek için sıfır yığılmalı regresyon modelleri kullanmışlardır. Sonuç olarak sıfır yığılmalı Poisson regresyon modelinin, haftalık ısı ölçümü verilerinde sıfır değerli gözlemlerin sık meydana gelmesi sorununu ele aldığından en uygun istatistiksel yaklaşım olduğunu göstermişlerdir. Özen (2020), yapmış olduğu çalışmasında Mersin kent merkezinde sinyalizasyon ile yönetilen dört kollu 39 kavşakta, 2014-2016 yılları arasındaki sadece taşıtların karıştığı ölümlü/yaralanmalı trafik kazaları ile trafik hacmi, kavşak geometrisi ve trafik kontrol özellikleri arasındaki ilişkiyi incelemek için Poisson ve Negatif Binom (NB) regresyon modelleri kullanmıştır.

Bu tez çalışmasının ikinci bölümünde yukarıda geçen regresyon modelleri tanıtılmıştır ve bu modellere ilişkin en uygun modeli seçmekte kullanılacak model seçim kriterleri anlatılmıştır. Tezin üçüncü bölümünde Türkiye İstatistik Kurumu (TÜİK) tarafından yapılan 2018 yılı Gelir ve Yaşam Koşulları Araştırması verileri kullanılarak bu sekiz modele ilişkin parametre tahminleri ve yorumları ile model karşılaştırmaları verilmiştir. Dördüncü ve son bölümde ise uygulama çalışmasından elde edilen sonuçlar incelenmiş ve bu sonuçlara ilişkin yorumlar verilmiştir.

2. SAYMA VERİLERİ İÇİN REGRESYON MODELLERİ

Tez çalışmasının bu bölümünde sayma verisinden oluşan verilerin analizinde kullanmak amacıyla Poisson Regresyon (PR), Negatif Binom Regresyon (NB), Sıfır Yığılmalı Poisson Regresyon (ZIP), Sıfır Yığılmalı Negatif Binom Regresyon (ZINB), Sıfır Kesilmiş Poisson Regresyon (ZTP), Sıfır Kesilmiş Negatif Binom Regresyon (ZTNB), Poisson Hurdle (PH) ve Negatif Binom Hurdle (NBH) modelleri anlatılmıştır.

2.1. Sayma Veri Modelleri

Sayma verisi, herhangi bir olayın bir süre aralığında kaç kez gerçekleştiğini gösteren verilerdir. Bu veriler adından da anlaşılacağı gibi sayımdan kaynaklanan verilerdir. Bu veriler negatif olmayan tamsayı değeri alan rastgele değişkenleri içerir. Yani yanıt değerleri kesikli tam sayılarından oluşur. Sayma verisi eğitim, bankacılık, sağlık, psikoloji ve çevre bilimi gibi farklı disiplinlerde yaygın olarak kullanılmaktadır. Bu verilere örnek olarak kalp krizi sayısı, hastanede yatılan gün sayısı, eğitim çalışmalarında belirli bir süre boyunca bulunmayan öğrenci sayısı ve aile içi şiddete maruz kalan çocuk sayısı verilebilir (Klakattawi ve ark., 2018).

Sayma verisi, normal dağılım özelliği göstermeyen ve daima pozitif değerler alan verilerdir (Khoshgoftaar ve ark., 2005). Sayma verisi olduğu durumda dağılım genellikle sağa çarpıktır. Normallik varsayımı sağlanamadığından klasik regresyon yöntemleri kullanılmaz (Pittmann ve ark., 2020). Eğer klasik yöntemler kullanılırsa yanlış parametre tahminleri elde edilir. Sayma verisinde aşırı yayılım sıklıkla görülür. Aşırı yayılım; varyansın ortalamanın çok üstünde olması durumudur.

Sayma verisini analiz etmek için Poisson regresyon, negatif binom regresyon, sıfır yığılma olduğu durumlarda ise sıfır yığılmalı regresyon modelleri veya veride kesilme olduğu durumlarda kesilmiş regresyon modelleri kullanılabilir.

2.1.1. Poisson regresyon modeli

Poisson dağılımı ilk kez Siméon-Denis Poisson (1781–1840) tarafından 1838'de yayınlanan “Ceza hukuku ve medeni hukuk alanlarındaki hükümlerin olasılığı üzerinde araştırmalar” (Recherches sur la probabilité des jugements en matières criminelles et matière civile) isimli eserde ortaya çıkmıştır.

Poisson dağılımı genellikle belli bir zaman diliminde meydana gelen olayları inceler ve bu olaylar birbirinden bağımsızdır. Dağılımın λ olmak üzere tek bir parametresi vardır. Sağa çarpık bir dağılımdır.

Gözlem sayısı olan n büyük iken değişken sayısı olan p de çok küçük ise binom dağılımı poisson dağılımına yaklaşır. Genel olarak $np \leq 5$ olduğu zaman binom dağılımı yerine poisson dağılımı kullanılabilir. Ayrıca deneme sayısının 20'den büyük olması koşulu vardır. Ortalama büyüdükçe poisson dağılımı normal dağılıma yaklaşır.

Poisson regresyon, sayma verisinin modellenmesinde sıklıkla kullanılmaktadır. Sayma verisini modellemek için kullanılan ilk modellerden biridir. Analizciler için mevcut olan birçok sayma modelinin temelindedir.

Poisson regresyon modeli, bağımsız değişken ile λ parametresi arasında bulunan ilişki durumunu ifade eden Poisson dağılımından çıkarılır. Bu model sayma verisinin analizinin yapılmasına imkân sağlamaktadır (Karaca, 2018). Diğer tüm sayma modelleri Poisson modelin bir varyasyonu olarak ortaya çıkmıştır. Bir Poisson modelinin yanıtı veya bağımlı değişkeni bir sayı olmalıdır. Sürekli bir değişken modellemek istiyorsak, başka bir model seçilmelidir (Hilbe, 2014).

Bu modele ilişkin bazı varsayımlar şunlardır:

- Dağılım, genellikle λ veya μ olarak sembolize edilen tek bir parametre ile modellenir ve kesikli bir dağılımdır. Ortalama ayrıca bir oran parametresi olarak anlaşılır. Zaman, alan veya hacim birimi başına bir ögenin veya olayın gerçekleşmesi için beklenen sayıdır.
- Yanıt terimleri veya y değerleri, negatif olmayan tamsayılardır; yani dağılım $y \geq 0$ olduğu durumlarda kullanılır.

- Gözlemler birbirinden bağımsızdır.
- Saymaya dayalı gözlemlenen hiçbir hücre, ampirik dağılımın ortalamasına bağlı olarak beklenenden daha fazla veya daha az değildir. Örneğin, belirli bir ortalamaya sahip bir Poisson dağılımında veriler beklenenden daha fazla sayıda sıfır içermemelidir. λ değeri büyüdükçe sıfır (0) sayım olasılığı azalır.
- Modelin ortalaması ve varyansı aynıdır. Örneğin daha yüksek ortalama değerlere sahip Poisson dağılımları buna bağlı olarak daha büyük değişkenliğe sahiptir.
- Pearson ki kare istatistiği 1'e yakın bir değere sahiptir. Yanıtın gözlemlenen ve tahmin edilen varyansları aynı olduğunda 1 değeri oluşur.

Yanıt değişkeni bir Poisson dağılımı olduğunda, verileri analiz etmek için Poisson regresyon modeli kullanılabilir. Poisson dağılımının model denklemi Eş. 2.1 de verildiği gibidir.

$$P(y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \quad y_i = 0, 1, \dots \quad (2.1)$$

Beklenen değer ve varyansı birbirine eşittir ve Eş.2.2 de verildiği gibi ifade edilir.

$$E(Y_i) = Var(Y_i) = \lambda_i \quad (2.2)$$

Ortalamanın logaritması doğrusal bir regresyon olarak modellendiğinden, Poisson regresyonunun log-lineer modeller sınıfına ait olduğu söylenmektedir. Bir Poisson modelinin tahmini regresyon katsayıları şu yorumu verir: Bir tahmin edici olan x sayısal olduğunda, buna karşılık gelen üslü katsayı olan Exp (β), tahmin edicideki bir birim artışa karşılık gelen tahmini bir oran olarak yorumlanır (Nava, 2014).

Poisson regresyon belirli bir zaman aralığında meydana gelmiş olan olaylar ile bu olaylar için belirlenen bağımsız değişkenler arasında bir bağlantı kurulmak istendiği durumlarda kullanılmaktadır (Yıldırım, 2019).

Poisson regresyon modelinin kullanıldığı çok sayıda alan vardır. Bu alanlara örnek olarak, kaza analizi çalışmaları, akademik alanda yapılan çalışmalar, aktüerya, boşanma istatistikleri, biyoistatistik ve ekoloji verilebilir.

Poisson regresyon modeli sahip olduğu özelliklerinden dolayı modelde bulunan parametrelerin tahmini için en küçük kareler yöntemi kullanılması mümkün değildir. En küçük kareler yöntemi yerine yaygın olarak en çok olabilirlik yöntemi kullanılır (Deniz, 2005).

Modele ilişkin en çok olabilirlik fonksiyonu;

Log- olabilirlik fonksiyonu,

$$LL = \sum_{i=1}^n (-\lambda_i + y_i \log(\lambda_i) - \log(y_i!)) \quad (2.3)$$

olarak ifade edilir.

Poisson regresyonda uyum iyiliğini test etmek için sıklıkla Pearson Ki- Kare belirleme katsayısı, R^2 değeri, sapma istatistiği, Akaike bilgi ölçütü ve Bayesyen bilgi ölçütü kullanılır.

2.1.2. Negatif binom regresyon modeli

Negatif binom regresyon modeli, istatistikçiler ve analizciler tarafından aşırı yayılım gösteren Poisson verilerini modellemek için yaygın olarak kullanılır. Veride aşırı yayılım durumu olduğunda Poisson regresyon modelini kullanmak yanlış parametre tahminlerine ve geçersiz sonuçlara sebep olmaktadır. Çoğu zaman aşırı yayılımın nedeni bilinmemektedir. Verilerde aşırı sayıda sıfır olması, verilerin birbirileri ile ilişkilendirilmesi veya verilerde sıfır olmaması gibi sebepleri olabilir (Hilbe, 2014).

Negatif binom dağılımı, Poisson ve Gamma dağılımlarının birbirileri ile birleşmelerinden oluşturulabilmektedir (Cameron ve Trivedi, 2013).

Negatif binom, ortalama (λ) ve yayılım (α) olmak üzere iki parametrelili bir modeldir ve bu model denklemi Eş.2.4 de verildiği gibi ifade edilir.

$$P(Y_i = y_i) = h(y_i|\lambda_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i}\right)^{\alpha^{-1}} \left(\frac{\lambda_i}{\alpha^{-1} + \lambda_i}\right)^{y_i}, \alpha > 0 \quad (2.4)$$

Burada α , aşırı yayılım parametresidir ve bu parametre 0 olduğunda bu model Poisson model ile aynı olmaktadır (Sarul ve Şahin, 2015). $\lambda > 0$ ve $\alpha > 0$ olması durumu aşırı yayılım göstergesidir. Bu modele ilişkin beklenen değer ve varyans,

$$E(Y_i|\lambda_i, \alpha) = \lambda_i \quad (2.5)$$

$$Var(Y_i|\lambda_i, \alpha) = \lambda_i(1 + \alpha\lambda_i) \quad (2.6)$$

şeklinde gösterilirken log olabilirlik fonksiyonu,

$$LL = \sum_{i=1}^n \left(\log\left(\frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})}\right) - (y_i + \alpha^{-1})\log(1 + \alpha\lambda_i) + y_i \log(\alpha\lambda_i) \right) \quad (2.7)$$

şeklindedir.

Ortalama ve varyanstan anlaşılacağı gibi yayılım parametresinin eklenmesiyle aşırı yayılımın modele dahil edilmesi sağlanmıştır. α sıfıra yaklaştığında negatif binom regresyon modelinden Poisson regresyon modeli elde edilmektedir. Bu da bu iki modelin iç içe (nested) olduğunu göstermektedir (Xie ve diğerleri, 2013).

Negatif binom, Poisson'dan çok daha geniş bir değişkenlik aralığını modellememizi sağlar.

Klasik negatif binom modeli, sayıların dağılımına Poisson varsayımları altında izin verilenden daha geniş bir şekil sağlayan ikinci bir parametreye (yayılım parametresi) sahip olması haricinde, Poisson dağılımı ile aynı dağılım varsayımlarına sahiptir. Poisson eşitlik varsayımı, ortalama ve varyans değerlerinin aynı olduğu anlamına gelir. Negatif binom için, iki parametre ortalamanın varyansını etkiler. Bu parametreler; yayılım parametresi (α) ve ortalamanın (λ^2) karesidir.

Negatif binom modeli kullanmak için varsayımlar şu şekilde sıralanabilir:

- Bağımlı değişken olan y , negatif olmayan tamsayılardan oluşur.
- λ değeri arttıkça, bağımlı değişkenin sıfır değerini alma olasılığı azalır.

- y , 0 olasılığını alabilmelidir.
- Modele uygunluğu sağlanan y da tahmin edilen değişken, y dağılımının beklenen ortalamasıdır.
- Negatif binom regresyonunun en önemli amacı varyans değerinin ortalamayı aştığı veya gözlemlenen varyansın beklenen varyansı aştığı verileri modellemektir.
- Uygun hale getirilmiş bir negatif binom modeli, 1'e yakın bir dağılım istatistiğine ve alternatif sayma modellerinden daha az bir AIC, BIC ve log olabilirlik istatistiğine sahiptir.
- Tahmini sayma verisinin sayısı, y 'nin dağılımı boyunca gözlemlenen sayma verisinin sayısıyla yaklaşık olarak aynıdır.

2.2. Sıfır Yığılmalı Sayma Veri Modelleri

Sayıma dayalı olarak elde edilen veri kümeleri beklenenden fazla sıfır değeri içerebilir. Veri kümesindeki bu beklenenden fazla sıfır yoğunluğu, sıfır yığılma olarak ifade edilmektedir (Sileshi, 2008). Sıfır yığılmalı verilerin analizinde sıfır değerlerinden kaynaklanan etkiyi dikkate alan sıfır değer yığılmalı regresyon modellerinin uygulanması önerilmektedir (Khoshgoftaar ve ark., 2005). Sıfır yığılmalı verilerin analizinde uygun yöntemlerin kullanılmaması, yanlış parametre tahminlerin elde edilmesi, standart hataların küçülmesi ve tutarsız sonuçların elde edilmesine neden olabilir (Miller, 2007). Sıfır yığılma durumunda ortaya çıkan sıfır değerler elde edilmiş durumlarına göre yapısal sıfır (structural zeros) ve örnekleme sıfırı (sampling zeros) olarak iki şekilde adlandırılmaktadır. Değerlendirilen veri kümesinde elde edilme ihtimali olmadığından yani gözlem yapmanın imkân dâhilinde olmadığı durumlarda hücrenin sıfır değerini almasına yapısal sıfır; elde edilmesi mümkün olduğu halde veri setinde gözlenmeyen, bu nedenle sıfır değer alan gözlemlere de örnekleme sıfırı denir (Karaca, 2018). Örneğin; araç sahibi olmayan bir kişinin belirlenen süre içinde belli bir akaryakıt istasyonundan aldığı akaryakıt litre miktarı yapısal sıfırdır. Araç sahibi olan kişinin belirlenen süre içinde belli bir akaryakıt istasyonundan akaryakıt almaması durumu örnekleme sıfırına örnektir.

Sıfır yığılmalı modellerin başlıcaları sıfır yığılmalı poisson ve sıfır yığılmalı negatif binomdur.

2.2.1. Sıfır yığılmalı poisson regresyon modeli

Sıfır yığılmalı Poisson regresyon modeli (ZIP), ilk olarak Lambert (1992) tarafından geliştirilmiş ve birçok çalışmada kullanılmıştır. Bu model, bağımlı değişkeninin iki farklı veri grubundan oluştuğunu varsaymaktadır. Bunlardan birincisi, sıfır değerlerini alabilecek olan Poisson dağılımlı veri grubu iken, ikinci grup ise daima sıfır değerlerini içermektedir. Bu regresyon modelinde, lojit fonksiyonu bağımlı değişkenin hangi veri grubuna dahil olduğunu belirlemek için kullanılır. Poisson dağılımı gösteren grup ise Poisson regresyon ile modellenir (Min ve Agresti, 2005).

ZIP modeline ilişkin model denklemi aşağıda verildiği gibidir.

$$P(Y = y_i) = f(y_i|w_i, \lambda_i) = f(x) = \begin{cases} w_i + (1 - w_i)\exp(-\lambda_i), & y_i = 0 \\ (1 - w_i) \frac{\lambda_i^{y_i}}{y_i!} \exp(-\lambda_i), & y_i > 0 \end{cases} \quad (2.8)$$

Bu model denkleminde $0 \leq w_i \leq 1$ dir. w_i , verideki sıfır yığılma oranıdır. $w_i = 0$ ise ZIP model Poisson Regresyona dönüşür. $w_i > 0$ olması durumu ise sıfır yığılma göstergesidir.

Modelin beklenen değer ve varyansı ise şu şekilde gösterilir:

$$E(Y_i) = (1 - w_i)\lambda_i \quad (2.9)$$

$$Var(Y_i) = (1 - w_i)\lambda_i(1 + w_i\lambda_i) \quad (2.10)$$

ZIP modelinde bahsedilen log parçaya ait denklem $\log(\lambda) = B\beta$ şeklinde iken, lojit parçasına ait denklemler aşağıda verildiği gibidir.

$$\text{logit}(w) = \log\left(\frac{w}{1-w}\right) = G_\gamma \quad (2.11)$$

$$w = \frac{e^{G_\gamma}}{1 + e^{G_\gamma}} \quad (2.12)$$

$$(1 - w) = \frac{1}{1 + e^{G_\gamma}} \quad (2.13)$$

Burada B ve G ortak değişken (covariate) matrislerdir. Modele ilişkin log olabilirlik fonksiyonu ise şu şekilde ifade edilir:

$$LL(\gamma, \beta, y) = \sum_{y_i=0} \log(e^{G_i\gamma} + \exp(-e^{B_i\beta})) + \sum_{y_i>0} (y_i B_i \beta - e^{B_i\beta}) - \sum_{i=1}^n \log(1 + e^{G_i\gamma}) - \sum_{y_i>0} \log(y_i!) \quad (2.14)$$

Burada G_i ve B_i ; G ve B matrislerinin i'inci satırıdır.

2.2.2. Sıfır yığılmalı negatif binom regresyon modeli

Sıfır yığılmalı Negatif Binom regresyon (ZINB) sıfır değerlerinin çok fazla olduğu bağımlı değişkeninin modellenmesinde kullanılan alternatif regresyon yöntemidir. Bu regresyon modeli, aşırı yayılım sorununun yanı sıra sıfır değerinde aşırı yığılmanın gözlemlendiği verilerin modellenmesi amacıyla ZIP modeline alternatif olarak 1994 yılında Greene tarafından geliştirilmiştir. ZINB, ZIP'tan daha iyi yayılım özelliklerine sahiptir. ZINB, bağımlı değişkenin çok fazla sıfır içermesi durumunda kullanılır ve aşırı yayılım sonuçları elde edilir.

ZINB için model denklemi aşağıda verilmiştir:

$$P(Y = y_i) = f(y_i | \lambda_i) = \begin{cases} p_i + (1 - p_i) \left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i} \right)^{\alpha^{-1}} & , y_i = 0 \\ (1 - p_i) \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1) \Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i} \right)^{\alpha^{-1}} \left(\frac{\lambda_i}{\alpha^{-1} + \lambda_i} \right)^{y_i} & , y_i > 0 \end{cases} \quad (2.15)$$

ZINB modelinin beklenen değer ve varyansı aşağıda verildiği gibidir.

$$E(Y_i) = (1 - p_i) \lambda_i \quad (2.16)$$

$$Var(Y_i) = E(Y_i)(1 + \alpha \lambda_i + p_i \lambda_i) \quad (2.17)$$

ZINB için log-olabilirlik fonksiyonu;

$$LL = (\lambda, \alpha, p_i; y) = \sum_{i=1} [I_{y_i=0} \log[(1 - p_i)(1 + \alpha \lambda_i)^{-\alpha^{-1}} + p_i] + I_{y_i>0} \log \left((1 - p_i) \frac{\Gamma(y_i + \frac{1}{\alpha})(\alpha \lambda_i)^{y_i}}{\frac{1}{\alpha} \Gamma(\frac{1}{\alpha})(1 + \alpha \lambda_i)^{y_i + \frac{1}{\alpha}}} \right) \quad (2.18)$$

$$= \sum_i (I_{y_i=0} \log[(1 - p_i)(1 + \alpha \lambda_i)^{-\alpha^{-1}} + p_i] + I_{y_i>0} (\log(1 - p_i) - \frac{1}{\alpha} \log(1 + \alpha \lambda_i) - y_i \log(1 + \frac{1}{\alpha \lambda_i}) + \log \Gamma(y_i + \frac{1}{\alpha}) - \log \Gamma(\frac{1}{\alpha}) - \log(y_i!))$$

verildiği gibidir. Burada I_i rasgele gösterge fonksiyonudur.

2.3. Kesilmiş Sayma Modelleri

Sayma verisi genellikle negatif olmayan tamsayılar kümesinden oluşur. Bununla birlikte, birçok durumda, sayımların tümü dağılımda gözlenemez. Özellikle, sıfırdan kesme olarak adlandırılan sıfır sayımların kaydedilmediği, yapısal olarak dışlandıkları veya gözlemlenmedikleri durumlara sıklıkla rastlanır (Rodriguez ve ark., 2019). Örnek olarak, hastanede kalış süresi, bir şarkının ilk on listesinde yer aldığı hafta sayısı veya bir arabadaki yolcu sayısı verilebilir.

Kesme işlemi, belirli bir aralıkta bazı gözlem değerlerinin veri setinden atılması şeklinde ifade edilebilir. Özellikle sayım verilerini içeren bazı çalışmalarda bu işlem oldukça etkindir (Cameron ve Trivedi, 2013). Sayıma dayalı verilerde verinin dağılımı, negatif olmayan tamsayıların tümünde değil, sadece bir alt kümesinde gözlemlenebiliyor ise bu verinin dağılımının kesildiği söylenir (Winkelmann, 2000).

Y tane olayın olduğu bir durumda r pozitif bir tam sayı olmak üzere bu r tane pozitif tam sayıdan küçük olanların veri setinden çıkarılması işlemine soldan kesme veya alttan kesme denir. Soldan kesilmiş sayım dağılımı şu şekilde ifade edilebilir:

$$f(y|\theta, y \geq r) = \frac{h(y|\theta)}{1-H(r-1|\theta)}, y = r, r+1, \dots \quad (2.19)$$

Y tane olayın olduğu bir durumda c pozitif bir tam sayı olmak üzere bu c tane pozitif tam sayıdan büyük olanların veri setinden çıkarılması işlemine sağdan kesme veya üstten kesme denir. Sağdan kesilmiş sayım dağılımı şu şekilde ifade edilebilir:

$$f(y|\theta, y \leq c) = \frac{h(y|\theta)}{H(c|\theta)}, y = 0, 1, \dots, c \quad (2.20)$$

Veride belirli bir değerin altında ya da üstünde gözlenen bir değer yoksa kesilmiş modeller kullanılabilir.

2.4. Sıfır Kesilmiş Sayma Modelleri

Sayım modellerinde en sık kullanılan kesilme şekli sıfırda kesilmedir. Bu durumda $r = 1$ 'dir (Gurmu, 1991). Yani, veri seti bir olayın meydana gelmesi ile başlar.

Sıfır kesilmiş regresyon yöntemleri bağımlı değişkenin yapısına göre farklı isimler alabilir. Örneğin dağılımdan sıfır değerleri atılmışsa ve Poisson dağılıyorsa sıfır kesilmiş Poisson (Zero Truncated Poisson=ZTP) regresyonu kullanılmaktadır. Eğer bağımlı değişkende aşırı yayılım durumu söz konusu ise bu durumda sıfır kesilmiş negatif binom (Zero Truncated Negatif Binomial=ZTNB) regresyonu kullanılır (Breen, 1996).

Hardin ve Hilbe (2015), kesilmiş verilerin kesilmemiş dağılıma dayalı regresyon modelleri ile modellendiğinde, tahminlerin yanlı olarak elde edilebileceğini belirtmektedir.

Sıfır kesilmiş sayma verisi çalışmalarına örnek olarak hastanede kalış süresi (Prebensen ve ark., 2015), bir Milli Park'a yapılan ziyaretlerinin sayısı (Martínez-Espiñeira ve Amoako-Tuffour 2008) veya raydan çıkan tren sayısı verilebilir. (Liu ve ark. 2013).

Veri analiz edilmeye başlamadan önce sayıma dayalı olarak elde edilen verilerin dağılımının ortalamasına bakılmalıdır. Eğer düşük bir ortalama söz konusu değilse sıfırdan kesme yöntemlerini kullanmak gerekli değildir. Eğer bağımlı değişkenin ortalaması 5'ten büyük ise bu veri setinde çok sayıda sıfır olması beklenen bir durum değildir. Bu nedenle sıfır kesilmiş regresyon yöntemleri kullanılmadan önce ortalamaya, örneklemin büyüklüğüne, beklenen sıfır sayısına, gerçekleşen sıfır sayısına ve sıfırların örneklem içerisindeki yüzdesine bakılmalıdır (Hilbe, 2014).

Modelde çok sayıda sıfır gözleendiğinde bu sıfırlara ilişkin olasılığı modelden çıkarmak için sıfır kesilmiş dağılımlar kullanılır (Zuur ve diğerleri, 2009). Sıfırdan kesilme bağımlı değişkenin sıfır değerini alamayacağı anlamına gelir. Bu duruma örnek olarak hastaların hastanede kalış süreleri verilebilir. Sıfır kesilmiş veriler bir sorun olarak görülmemelidir. Ortalama küçükse ve bağımlı değişken sıfır değerini içermiyorsa, Genelleştirilmiş lineer modeller tarafından elde edilen tahmini parametreler ve standart hatalar yanlı olabilir. Bu soruna bir çözüm olarak sıfır kesilmiş Poisson ve sıfır kesilmiş negatif binom modelleri önerilmektedir. Bağımlı değişkenin ortalaması görece büyükse, kesme problemini göz ardı

ederek, Poisson \ negatif binom veya genelleştirilmiş doğrusal model uygulanması durumunda, bir sorun olma olasılığı düşüktür.

Sıfırdan kesme modelleri iki şekilde olur:

- Özel olarak sıfır olmayacak şekilde düzenlenen modellerde ve
- Soldan kesilmiş sayma modeli oluşturma şeklindedir.

Sıfır kesilmiş modeller, modelde yapısal sıfırlar hariç tutulduğunda ve Hurdle modellerin uygulanmasında kullanılır (Hilbe, 2014).

2.4.1. Sıfır kesilmiş poisson regresyon

Bu model literatürde pozitif Poisson regresyon modeli olarak da geçmektedir. Klasik Poisson'dan farkı bağımlı değişkenin sıfır değerini almamasıdır. Modelde Y_i , i. gözlem için bağımlı değişken olsun. Poisson ortalaması λ ile dağıldığı varsayalım. Bu şartlar altında sıfır kesilmiş Poisson modeli şu şekilde ifade edilir:

$$P(Y_i = j | Y_i > 0) = \frac{\exp(-\lambda_i) \lambda_i^j}{j!} [1 - F_p(0)]^{-1} = \frac{\lambda_i^j}{(\exp(\lambda_i) - 1)j!} \quad (2.21)$$

burada j yalnızca 0'dan büyük pozitif tamsayı değerlerini almaktadır (Grogger ve Carson, 1991). $F_p(0)$ ifadesi ise Poisson dağılımının dağılım fonksiyonunda 0 değerini almasını ifade eder.

Kesilmiş olasılık fonksiyonu, standart olasılık fonksiyonunun $[1 - F_p(0)]^{-1}$ ile çarpılması sonucu, gözlemlenmeyen sıfırları hesaba katarak onların sayısını (şişirir) artırır.

Bu model için Log olabilirlik fonksiyonu şu şekilde gösterilir:

ZTP modeli için ise:

$$LL = I((y_i; \lambda_i | y_i > 0)) = \sum_{i=1}^m \sum_{y_i} [X^T \beta_i] - \exp(X^T \beta_i) - \ln(y_i!) - \ln[1 - \exp\{-\exp(X^T \beta_i)\}] \quad (2.22)$$

Burada β , regresyon katsayılarını ve m kesilmiş verideki gözlem sayısını ifade etmektedir. Y_i nin koşullu ortalaması ve varyansı:

$$E(Y_i|X_i, Y_i > 0) = \lambda_i[1 - F_p(0)]^{-1} \quad (2.23)$$

$$Var(Y_i|X_i, Y_i > 0) = E(Y_i|X_i, Y_i > 0)[1 - F_p(0)E(Y_i|X_i > 0)] \quad (2.24)$$

Burada $F_p(0)$ ifadesi ise Poisson dağılımının olasılık yoğunluk fonksiyonunda 0 değerini almasını ifade eder.

2.4.2. Sıfır kesilmiş negatif binom regresyon

Kesilmiş sayma verisi kapsamında pozitif değerler alan bağımlı değişkende aşırı yayılımda gözleniyorsa doğru ve yansız tahmin yapabilmek için sıfır kesilmiş negatif binom modeli kullanılır. Bu model Eş.2.25 de verildiği şekilde ifade edilir.

$$P(Y_i = j|Y_i > 0) = \frac{\Gamma(j+\frac{1}{\alpha})}{\Gamma(j+1)\Gamma(\frac{1}{\alpha})} (\alpha\lambda_i)^j [1 + \alpha\lambda_i]^{-(j+1/\alpha)} [1 - F_{NB}(0)]^{-1} \quad (2.25)$$

Burada j sıfırdan büyük bir sayı, Γ gama değeri, α aşırı yayılım parametresi, λ ortalama ve $F_{NB}(0)$ ise negatif binom dağılımının dağılım fonksiyonunda 0 değerini almasını ifade etmektedir.

Modele ilişkin log olabilirlik fonksiyonu ise Eş. 2.26 da verildiği gibidir.

$$LL = \sum_{i=1}^m \log(\Gamma(j + \frac{1}{\alpha})) - \log(\Gamma(\frac{1}{\alpha})) + j \log(\alpha) + jX_i\beta - (j + \frac{1}{\alpha}) \log[1 + \alpha\lambda_i] - \log[1 - (1 + \alpha\lambda_i)^{1/\alpha}] \quad (2.26)$$

Y_i nin koşullu ortalaması ve varyansı:

$$E(Y_i|X_i, Y_i > 0) = \lambda_i[1 - F_{NB}(0)]^{-1} \quad (2.27)$$

$$Var((Y_i|X_i, Y_i > 0) = \frac{E(Y_i|X_i, Y_i > 0)}{F_{NB}(0)^\alpha} [1 - [F_{NB}(0)]^{(1+\alpha)} E(Y_i|X_i, Y_i > 0)] \quad (2.28)$$

şeklindedir (Grogger ve Carson,1991).

2.5. Hurdle Modeller

Hurdle model, sıfır yığılmalı veri setlerini analiz etmede kullanılan alternatif bir metottur. İlk defa Cragg (1971) sansürlü verilerin analizinde önermiştir. Ardından Mullahy (1986) sayma veri modellerine uygulamalarını ve modelin genel yapısını ortaya atmıştır. Hurdle modeller bağımlı değişkenin aldığı değere göre iki aşamadan meydana gelmektedir. Yani bağımlı değişkene ikili (binary) yanıtlı muamelesi yapılır. İlk aşama, verideki sıfırlara karşı pozitif sayımları gösteren geçiş aşamasıdır (transation stage). Diğer aşama ise sadece pozitif değerlerin yer aldığı olay aşamasıdır (events stage). İkili cevapların modellenmesinde lojit fonksiyonu kullanılmaktadır. Pozitif değerlerin modellenmesinde ise sıfır kesilmiş (zero-value truncated) yöntemler kullanılmaktadır. Olay aşamasındaki pozitif değerlerin Poisson dağılımından gelmesi durumunda Poisson Hurdle (PH), NB dağılımdan gelmesi durumunda ise Negatif Binom Hurdle (NBH) regresyon modelleri kullanılmaktadır (Tüzen, 2018). Bu modelde ilk aşama sıfır veya sıfır olmayan sonuçların gelip gelmediğini gösteren binom modelidir. Pozitif değerlerin yer aldığı kesilmiş veriler ise ikinci aşama olarak modellenmektedir. İlk aşama lojit, probit veya complementary loglog kullanılarak modellenmektedir. Pozitif değerlerin olduğu ikinci kısım ise Poisson, geometrik veya negatif binom kullanılarak modellenmektedir (Long ve Freese, 2006).

Hurdle modeller aşırı yayılım durumunda da kullanılabilir (Saffari ve ark., 2012). Sıfır yığılmalı modeller ile Hurdle modeller benzer gibi görünse de aralarında birtakım farklar vardır. İki modelde de süreç iki aşamalıdır. Ancak sıfır yığılmalı modellerde sıfırdan farklı değerlerde tahmin için model oluşturulurken yanıtın hala sıfır değeri alma durumu söz konusudur. Hurdle modellerde ise sıfırdan farklı yanıt değerlerinin modellenmesi yapılırken yanıtın pozitif olma koşulu vardır. Bahsedilen bu iki modelde sıfırların tahmin modelini oluştururken bağımlı değişkene ikili değişken muamelesi yapar. Sıfırdan farklı değerlerde ise sıfır yığılmalı modeller kesikli dağılımları kullanırken, Hurdle modeller kesilmiş (truncated) dağılımları kullanır (Yıldırım,2019).

2.5.1. Poisson hurdle regresyon modeli

Hurdle modellerde olay aşamasında yani ikinci aşamada pozitif değerlerin Poisson dağılımından gelmesi durumunda Poisson Hurdle (PH) regresyon modeli tercih edilir. Poisson hurdle regresyon modeli için eşitlik:

$$P(Y = y) = \begin{cases} w & , y=0 \\ (1-w) \frac{e^{-\lambda} \lambda^y}{(1-e^{-\lambda})\Gamma(\gamma+1)} & , y \geq 1 \end{cases} \quad (2.29)$$

Burada w , bağımlı değişkenin sıfır değerini alma olasılığını göstermektedir.

y_i ($i=1,2,\dots$) ler birbirinden bağımsız ve sayma verisinden oluşan gözlem değerleri olsun. $y_i=0$ olma olasılığı $1 - p(x)$ olsun. y_i kesilmiş Poisson ($\lambda(z)$) olma olasılığı ise $p(x)$ olsun. Burada x ve z ise ortak değişken matrisleridir (Yeşilova, 2009).

$$P(y_i = 0|x) = 1 - p(x) \quad (2.30)$$

$$P(y_i = q|x, z) = \frac{p(x)\exp(-\lambda(z))\lambda(z)^q}{q!(1-\exp(-\lambda(z)))}, q = 1, 2, \dots \quad (2.31)$$

$p(x)$ ve $\lambda(z)$, $\log(\lambda(z)) = x'_i\beta$ ve $\text{logit}(p_i) = z'_i\alpha$ şeklinde de ifade edilebilir. Bu iki eşitlikte yer alan α ve β bilinmeyen parametrelerini tahmin etmek için kullanılan Poisson Hurdle için log olabilirlik fonksiyonu aşağıdaki gibidir:

$$LL = \sum_{y_i > 0} x_i \beta - \sum_{i=1}^n \log(1 + \exp(x_i \beta)) + \sum_{y_i > 0} [y_i z_i \alpha - \exp(z_i \alpha) - \log(1 - \exp(-\exp(z_i \alpha))) - \log(y_i!)] \quad (2.32)$$

$$LL = L(\beta) + L(\alpha) \quad (2.33)$$

2.5.2. Negatif binom hurdle regresyon modeli (NBH)

Poisson hurdle modeline oldukça benzer bir modeldir. Klasik negatif binom ve sıfır yığılmalı negatif binomda olduğu gibi burada da aşırı yayılım olması durumunda tercih edilir. Poisson hurdle modelinden farklı olarak ikinci aşamada pozitif değerler negatif binom dağılımından gelmektedir. Bu modele ilişkin denklem:

$$P(Y = y) = \begin{cases} w & , y=0 \\ (1-w) \frac{\Gamma(y + \frac{1}{\alpha})^w}{[1 - (\frac{1}{1+\alpha\lambda})^{\frac{1}{\alpha}}] \Gamma(y+1) \Gamma(\frac{1}{\alpha})} (\frac{1}{1+\alpha\lambda})^{\frac{1}{\alpha}} (\frac{\alpha\lambda}{1+\alpha\lambda})^y & , y \geq 1 \end{cases} \quad (2.34)$$

şeklindedir. Burada α aşırı yayılım parametresidir.

Log olabilirlik fonksiyonu:

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

Burada $f(0)$, modelin 0 ve 1 değeri alan ikili(binary) kısmının olasılığı iken $p(j)$ pozitif değerlerin olasılığıdır.

2.6. Model Seçimi

Bu tez çalışmasında kullanılan modeller arasında hangi modelin kullanılan veri setini daha iyi temsil ettiğine karar vermek için çeşitli testler kullanılmıştır. Bu testler Akaike Bilgi Kriteri (AIC), Log olabilirlik ve Vuong testidir.

AIC, Japon istatistikçi Hirotugu Akaike tarafından ortaya konulmuştur (Akaike, 1973). Günümüzde en sık kullanılan istatistiklerden biridir. AIC istatistiği genellikle iki biçimde bulunur. Bu biçimlerden geleneksel kullanım şu şekildedir:

$$AIC = -2L + 2k = -2(L - k)$$

Diğer gösterim şekli ise,

$AIC = \frac{-2L+2k}{n} = -2(L - k)/n$ şeklindedir. Burada L değeri, log olabilirliği, k tahmin edici (ya da parametre) sayısını ve n modeldeki gözlem sayısını ifade etmektedir (Hilbe, 2014).

AIC istatistiğini hesaplamada genellikle ilk verilen eşitlik kullanılmaktadır. Bazı analistler farklı örneklem büyüklüğündeki modelleri karşılaştırmak için de bu istatistiği

kullanabilmektedir.

AIC, model için kullanılan açıklayıcı değişkenlerin doğru ve yeterli sayıda belirlenerek modele hiçbir katkıda bulunmayan veya modelin veriye uyumunu olumsuz yönde etkileyen açıklayıcı değişkenlerin modelden çıkartılması amacıyla da kullanılmaktadır (Tüzen, 2018).

AIC değerine ilişkin önemli bir özellik model karşılaştırmalarında her zaman en küçük AIC değerini veren modelin tercih edilmesidir.

Log olabilirlik (LL) değeri üzerinden model seçimi yapılırken, en büyük LL değerine sahip model uygun model olarak belirlenir (Karaca, 2018).

Bu tez çalışmasında kullanılan diğer bir test ise sıfır yığılmalı verilerin analiz edilmesinde sıklıkla kullanılan Vuong testidir. Vuong testi, 1989 yılında Vuong tarafından geliştirilen ve iç içe geçmeyen (non-nested) modellerin karşılaştırılmasında kullanılan hipotez testlerinden biridir (Tüzel, 2011). İç içe model karşılaştırmalarının dışında olası ikili modeller de Vuong testiyle karşılaştırılabilmektedir. Bu sayede aşırı sıfır değerinin tahmin edilmesinde hangi modellerin uygun olabileceği belirlenebilmektedir. Vuong testinde pozitif test istatistiği birinci modelin ikinci modele göre daha tercih edilebilir olduğunu, negatif olması durumunda ise ikinci modelin tercih edilebileceği yorumu yapılır (Yıldırım, 2019).

$p_1(.)$ ve $p_2(.)$ olasılık yoğunluk fonksiyonları olmak üzere Vuong testi için denklemler şu şekilde ifade edilir:

$$V = \frac{\bar{m}\sqrt{n}}{sd(m)}$$

Burada $\bar{m}$, m_i 'nin ortalamasını ifade etmektedir. m_i ise $m_i = \ln\left(\frac{p_{1i}(y_i)}{p_{2i}(y_i)}\right)$ şeklinde gösterilir.

Burada $sd(m)$; standart sapmayı, m iki olasılık yoğunluk fonksiyonu arasındaki oranın logaritmasını ve n ise örnek çapını temsil etmektedir.

Vuong test istatistiği standart normal dağılımlıdır. Vuong test değerinin yorumlanmasına örnek olarak, 0,05 anlamlılık düzeyi için, V değeri 1,96'dan büyükse, ilk model gerçek

modele “daha yakındır” şeklinde bir yorum yapılabilirken eğer V değeri -1,96'dan küçükse ikinci model gerçek modele “daha yakındır” yorumu yapılabilir. Hesaplanan değer bu değerler (-1,96;1,96) arasında ise modeller gerçek modele “daha yakın” değildir ya da birinci veya ikinci modeli kullanma arasında fark yoktur yorumu yapılır (İsmail ve Zamani, 2013).

3. UYGULAMA

Bu bölümde Türkiye İstatistik Kurumu'ndan alınan veriler kullanılarak, hanehalkı sorumlusunun sahip olduğu konut sayısını etkileyen faktörlerin belirlenmesinde PR, NB, ZIP, ZINB, ZTP, ZTNB, PH ve NBH modellerinden hangisinin veri setini daha iyi temsil ettiği araştırılmıştır. Bu araştırma sırasında kullanılan analizler R Studio 3.6.1 paket programı kullanılarak yapılmıştır ve R Studio içinde var olan “MASS ve pscl” kütüphanelerinden yararlanılmıştır (Cran.r, 2020).

3.1. Veri Seti

Bu çalışmada Türkiye İstatistik Kurumu'nun (TÜİK) 2018 yılında yaptığı Gelir ve Yaşam Koşullar Araştırması (GYKA) mikro veri seti kullanılmıştır. 2006 yılından itibaren her yıl düzenli olarak yapılan, Avrupa Birliğine uyum çerçevesinde, amacı gelir dağılımı yanında, yaşam koşulları, sosyal dışlanma ve göreceli gelir yoksulluğu gibi konularda veri üretmek olan, “panel anket” yönteminin kullanıldığı GYKA uygulamaya başlamıştır (TÜİK,2019).

Araştırmada, gelir, yoksulluk, sosyal dışlanma ve diğer yaşam koşullarına ilişkin göstergeleri hesaplamak üzere konut, ekonomik durum, sosyal dışlanma, gayrimenkul sahipliği, eğitim, demografi, sağlık durumu, iş gücü durumu, gelir durumu gibi kategorilerde bilgi derlenmektedir:

Uygulama için haneler filtrelenerek analizler yapılmıştır. Hanehalkı sorumlusunu; hanenin yönetiminde söz sahibi olan, hanehalkının hukuki, sosyal ve ekonomik planlama ve karar sürecinde en etkin rol oynayan belirli bir yaştan üzerindeki hanehalkı ferdi şeklinde tanımlayabiliriz. Araştırmaya katılan hanehalkının oturduğu konut dışında sahip olduğu konut adedi bağımlı değişken olarak ele alınmıştır. Bu değişkenin bağımlı değişken olarak seçilmesindeki kriter çok sayıda 0 (sıfır) içermesidir. Yapılan çalışma, hanehalkının oturduğu konut dışında sahip olduğu toplam konut sayısını etkileyen faktörleri ortaya koymak amacıyla hazırlanmıştır. Ayrıca ferde (hanehalkı sorumlusuna) özgü özelliklerden yaş, cinsiyet, eğitim durumu ile hanenin yıllık geliri, faal durumu ve geçinme durumu açıklayıcı değişken olarak alınmıştır.

Uygulamada kullanılan bağımsız değişkenler, açıklamaları ve düzeyleri Çizelge 3.1’de gösterildiği gibidir.

Çizelge 3.1. Kullanılan değişkenlere ilişkin temel bilgiler

Değişkenler	Açıklama	Düzey
Oturulan konut hariç sahip olunan konut sayısı (Bağımlı Değişken)	2017 yılı içerisinde sahip olunan konut sayısı (oturulan konut hariç)	
Yaş	Hanehalkı sorumlusu yaş bilgisi	1. 25 Yaş altı 2. 25 – 34 Yaş 3. 35 – 44 Yaş 4. 45 – 54 Yaş 5. 55 – 64 Yaş 6. 65 Yaş ve üstü
Cinsiyet	Hanehalkı sorumlusu cinsiyet bilgisi	1. Erkek 2. Kadın
Eğitim Durumu	Hanehalkı sorumlusunun en son bitirdiği eğitim seviyesi	1. Bir okul bitirmeyen 2. Okur yazar olmayan 3. Lise altı 4. Lise 5. Yükseköğretim
Faal Durum	Hanehalkı sorumlusunun şu anki istihdam durumu	1. İş gücüne dahil olmayan 2. İşsiz 3. İstihdam
Hanehalkı Geliri	Hanehalkının 2017 yılında elde ettiği toplam yıllık geliri	1. 25 000 TL’den az 2. 25 000 – 50 000 TL 3. 50 000 – 75 000 TL 4. 75 000 TL üstü
Geçinme Durumu	Hanenin gerekli harcamaları yapabilme durumu	1. Çok zor 2. Zor 3. Biraz zor 4. Biraz kolay 5. Kolay 6. Çok kolay

3.2. Betimleyici İstatistikler

TÜİK, GYKA anketine katılım gösteren 24 053 bireyin konut sayısına göre dağılımları ve araştırmaya katılım gösteren bireylerin demografik özelliklerine göre dağılımları betimleyici istatistikler başlığı altında yer alan tablo ve grafiklerle gösterilmiştir.

Çizelge 3.2. Konut sayısına ilişkin frekans tabloları

KONUT SAYISI			
		Sayı	Yüzde
Aldığı Değer	0	19 100	79,40
	1	3 636	15,10
	2	840	3,50
	3	261	1,10
	4	115	,50
	5	43	,20
	6	21	,10
	7	15	,10
	8	9	,00
	9	1	,00
	10	5	,00
	14	2	,00
	15	2	,00
	17	1	,00
	25	1	,00
	42	1	,00
	Toplam	24 053	100,00

Çizelge 3.2 incelendiğinde oturduğu konut dışında başka bir konuta sahip olmayan bireylerin sayısının 19 100 olduğu görülmüştür. Bu da ankete katılan bireylerin yaklaşık %79 u dur. Oturduğu konut hariç 1 tane konutu olan bireylerin sayısı 3 636 dır. Bu da yaklaşık olarak %15 dir.

Çizelge 3.3. Yaş değişkeni düzeyine ilişkin frekans dağılımı

Yaş Değişkeni			
		Sayı	Yüzde
Düzey	1. 25 Yaş altı	512	2,10
	2. 25-34	3 439	14,30
	3. 35-44	5 577	23,20
	4. 45-54	5 301	22,00
	5. 55-64	4 647	19,30
	6. 65 yaş ve üzeri	4 577	19,00
	Toplam	24 053	100,00

Çizelge 3.4. Konut sayısı ve yaş değişkeni düzeyine ilişkin frekans dağılımı

		Yaş Değişkeni						Toplam
		25 yaş altı	25-34	35-44	45-54	55-64	65 yaş ve üzeri	
Konut Sayısı	0	488	3 168	4 783	4 142	3 211	3 308	19 100
	1	20	243	641	899	1 020	813	3 636
	2	3	19	118	173	259	268	840
	3	1	6	16	48	96	94	261
	4	0	2	10	19	34	50	115
	5	0	0	4	7	13	19	43
	6	0	0	1	5	4	11	21
	7	0	0	1	5	4	5	15
	8	0	0	2	2	2	3	9
	9	0	0	0	0	0	1	1
	10	0	0	0	1	2	2	5
	14	0	1	0	0	1	0	2
	15	0	0	1	0	0	1	2
	17	0	0	0	0	0	1	1
	25	0	0	0	0	0	1	1
	42	0	0	0	0	1	0	1
Toplam		512	3 439	5 577	5 301	4 647	4 577	24 053

Çizelge 3.3 e göre; 25 yaşından küçük olanların oranı %2 (512 kişi), 25-34 yaş aralığında olanların oranı %14 (3 439 kişi), 35-44 yaş aralığında olanların oranı %23 (5 577 kişi), 45-54 yaş aralığında olanların oranı %22 (5 301 kişi), 55-64 yaş aralığında olanların oranı %19 (4 647 kişi) ve son olarak 65 yaş ve üzerinde olanların oranı %19 (4 577 kişi) dur. Dolayısıyla örneklemin çoğunluğunun 35-44 yaş aralığında dağıldığı yorumu yapılabilir.

Konut sayısı ve yaş değişkenine ilişkin frekans dağılımı incelendiğinde; 25 yaşından küçük kişilerin olduğu aralıkta 24 gözlemin, 25-34 yaş aralığında yer alan kişilerin olduğu aralıkta 271 gözlemin, 35-44 yaş aralığında yer alan kişilerin olduğu aralıkta 794 gözlemin, 45-54 yaş aralığında yer alan kişilerin olduğu aralıkta 1 159 gözlemin, 55-64 yaş aralığında yer alan kişilerin olduğu aralıkta 1 416 gözlemin, 65 yaş ve üstü düzeyde yer alan kişilerin olduğu aralıkta 1 269 gözlemin sahip olduğu konut sayısının sıfırdan farklı bir değer aldığı görülmektedir. 25 yaş altı kişilerin sahip olduğu konut sayısı en fazla üçtür. Çizelge 3.4 incelendiğinde; yaş ilerledikçe kişilerin sahip olduğu konut sayısında artış olduğu görülmektedir. Bu durumun çeşitli sebepleri olabilir.

Çizelge 3.5. Gelir değişkenine ilişkin frekans dağılımı

Gelir Değişkeni			
		Sayı	Yüzde
Düzey	1. 25 Bin TL altı	7 116	29,60
	2. 25 Bin TL-50 Bin TL	9 584	39,80
	3. 50 Bin TL-75 Bin TL	4 119	17,10
	4. 75 Bin TL ve üzeri	3 234	13,40
	Toplam	24 053	100,00

Çizelge 3.6. Gelir değişkeni ve konut sayısına ilişkin frekans dağılımı

		Gelir Değişkeni				Toplam
		25 Bin TL Altı	25 Bin TL-50 Bin TL	50 Bin TL-75 Bin TL	75 Bin TL Ve Üzeri	
Konut Sayısı	0	6 392	7 772	2 983	1 953	19 100
	1	581	1 363	842	850	3 636
	2	98	305	199	238	840
	3	28	88	54	91	261
	4	12	33	26	44	115
	5	2	9	7	25	43
	6	3	7	5	6	21
	7	0	5	1	9	15
	8	0	1	1	7	9
	9	0	0	0	1	1
	10	0	0	1	4	5
	14	0	0	0	2	2
	15	0	1	0	1	2
	17	0	0	0	1	1
	25	0	0	0	1	1
	42	0	0	0	1	1
Toplam		7 116	9 584	4 119	3 234	24 053

Çizelge 3.5 e göre ankete katılanların gelir düzeyleri incelendiğinde; yıllık geliri 25 Bin TL altında olanların oranı %29 (7 116 kişi), 25 Bin TL- 50 Bin TL aralığında olanların oranı %39 (9 584 kişi), 50 Bin TL- 75 Bin TL aralığında olanların oranı %17 (4 119 kişi) ve son olarak 75 Bin TL ve üzerinde olanların oranı %13 (3 234 kişi) tür. Dolayısıyla örneklemin çoğunluğunun gelirinin 25 Bin TL- 50 Bin TL aralığında dağıldığı yorumu yapılabilir.

Konut sayısı ve gelir düzeyi değişkenine ilişkin frekans dağılımı incelendiğinde; geliri 25 Bin TL den az olan hanelerin olduğu aralıkta 724 gözlemin, geliri 25 Bin TL- 50 Bin TL aralığında yer alan hanelerin olduğu aralıkta 1 812 gözlemin, geliri 50 Bin TL-75 Bin TL aralığında yer alan hanelerin olduğu aralıkta 1 136 gözlemin, geliri 75 Bin TL ve üstü düzeyde yer alan hanelerin olduğu aralıkta 1 281 gözlemin sahip olduğu konut sayısının sıfırdan farklı bir değer aldığı görülmektedir. 25 Bin TL altı gelire sahip hanelerin sahip olduğu konut sayısı en fazla altıdır. Tablo incelendiğinde; 25 Bin TL- 50 Bin TL gelir aralığında yer alan hanelerin daha fazla konutu olduğu görülmektedir. Bu durum çeşitli sebepleri olabilir. Gelir arttıkça kişiler konut alımından çok başka alanlarda gelirini değerlendirmeye gitmiş olabilir.

Çizelge 3.7. Eğitim durumu değişkenine ilişkin frekans dağılımı

Eğitim Durumu Değişkeni			
		Sayı	Yüzde
Düzye	1. Bir okul bitirmedi	1 400	5,80
	2. Okuryazar değil	1 856	7,70
	3. Lise altı	12 706	52,80
	4. Lise	4 001	16,60
	5. Yükseköğretim	4 090	17,00
	Toplam	24 053	100,00

Çizelge 3.8. Konut sayısı ve eğitim durumu değişkenine ilişkin frekans dağılımı

		Eğitim Durumu Değişkeni					Toplam
		Bir Okul Bitirmedi	Okur Yazar Değil	Lise Altı	Lise	Yükseköğretim	
Konut Sayısı	0	1 173	1 675	10 057	3 220	2 975	19 100
	1	164	133	1 913	601	825	3 636
	2	44	33	451	121	191	840
	3	14	9	153	33	52	261
	4	3	5	67	15	25	115
	5	2	0	27	5	9	43
	6	0	1	15	2	3	21
	7	0	0	8	1	6	15
	8	0	0	8	1	0	9
	9	0	0	1	0	0	1
	10	0	0	3	0	2	5
	14	0	0	0	1	1	2
	15	0	0	1	1	0	2
	17	0	0	1	0	0	1
	25	0	0	1	0	0	1
	42	0	0	0	0	1	1
Toplam		1 400	1 856	12 706	4 001	4 090	24 053

Çizelge 3.7 ye göre ankete katılan hanehalkı sorumlularının eğitim düzeyleri incelendiğinde; herhangi bir okul bitirmeyen kişilerin oranı %5 (1 400 kişi), okuryazar olmayan kişilerin oranı %7 (1 856 kişi), öğrenim düzeyi lise altı olan kişilerin oranı %52 (12 706 kişi), lise düzeyinde eğitim gören kişilerin oranı %16 (4 001 kişi) ve son olarak yükseköğrenim düzeyinde eğitim görenlerin oranı %17 (4 090 kişi) dir. Dolayısıyla örneklemin çoğunluğunun eğitim düzeyinin lise altı olduğu yorumu yapılabilir.

Çizelge 3.8 incelendiğinde; herhangi bir okul bitirmeyen kişilerin olduğu aralıkta 227 gözlemin, eğitim düzeyi okuryazar olmayan kişilerin olduğu aralıkta 181 gözlemin, eğitim düzeyi lise altı olan kişilerin olduğu aralıkta 2 649 gözlemin, eğitim düzeyi lise olan kişilerin olduğu aralıkta 781 gözlemin, eğitim düzeyi yükseköğrenim olan kişilerin olduğu aralıkta 1 115 gözlemin sahip olduğu konut sayısının sıfırdan farklı bir değer aldığı görülmektedir. Bir okul bitirmeyen kişilerin sahip olduğu konut sayısı en fazla beştir. Tablo incelendiğinde; lise altı eğitim düzeyinde yer alan kişilerin daha fazla konutu olduğu görülmektedir.

Çizelge 3.9. Çalışma durumu değişkenine ilişkin frekans dağılımı

Çalışma Durumu Değişkeni			
		Sayı	Yüzde
Düzey	1. İş gücüne dahil olmayan	8 417	35,00
	2. İşsiz	834	3,50
	3. İstihdam	14 802	61,50
	Toplam	24 053	100,00

Çizelge 3.10. Konut sayısı ve çalışma durumu düzeyine ilişkin frekans dağılımı

		Çalışma Durumu Düzeyleri			Toplam
		İş Gücüne Dahil Olmayan	İşsiz	İstihdam	
Konut Sayısı	0	6 256	764	12 080	19 100
	1	1 516	57	2 063	3 636
	2	387	9	444	840
	3	148	3	110	261
	4	57	1	57	115
	5	21	0	22	43
	6	13	0	8	21
	7	6	0	9	15
	8	5	0	4	9
	9	0	0	1	1
	10	3	0	2	5
	14	1	0	1	2
	15	1	0	1	2
	17	1	0	0	1
	25	1	0	0	1
	42	1	0	0	1
Toplam		8 417	834	14 802	24 053

Çizelge 3.9 a göre ankete katılanların çalışma düzeyleri incelendiğinde; iş gücüne dahil olmayan kişilerin oranı %35 (8 417 kişi), işsiz kişilerin oranı %3 (835 kişi) ve son olarak istihdam edilenlerin oranı %61 (14 802 kişi) dir. Dolayısıyla örneklemin çoğunluğunun istihdam edilmiş kişiler olduğu yorumu yapılabilir.

Konut sayısı ve çalışma düzeyine ilişkin frekans tablosu incelendiğinde; iş gücüne dahil olmayan kişilerin olduğu aralıkta 2 161 gözlemin, işsiz kişilerin olduğu aralıkta 70 gözlemin,

istihdam edilen kişilerin olduğu aralıkta 2 722 gözlemin sahip olduğu konut sayısının sıfırdan farklı bir değer aldığı görülmektedir. İşsiz kişilerin konut sayısı en fazla dördttür. Çizelge 3.10 incelendiğinde; istihdam edilmiş kişilerin daha fazla konutu olduğu görülmektedir.

Çizelge 3.11. Geçinme durumu düzeyine ilişkin sayı ve oranlar

Geçinme Durumu Değişkeni			
		Sayı	Yüzde
Düzyey	1. Çok zor	832	3,50
	2. Zor	2 556	10,60
	3. Biraz zor	4 223	17,60
	4. Biraz kolay	10 287	42,80
	5. Kolay	5 023	20,90
	6. Çok kolay	1 132	4,70
	Toplam	24 053	100,00

Çizelge 3.12. Konut sayısı ve geçinme durumu düzeyine ilişkin frekans dağılımı

		Geçinme Durumu Düzeyleri						Toplam
		Çok Zor	Zor	Biraz Zor	Biraz Kolay	Kolay	Çok Kolay	
Konut Sayısı	0	768	2 272	3 573	8 237	3 584	666	19 100
	1	55	227	500	1 573	1 004	277	3 636
	2	7	40	99	328	274	92	840
	3	1	14	30	94	84	38	261
	4	1	2	14	35	39	24	115
	5	0	0	3	11	16	13	43
	6	0	1	3	4	10	3	21
	7	0	0	1	4	4	6	15
	8	0	0	0	1	3	5	9
	9	0	0	0	0	1	0	1
	10	0	0	0	0	2	3	5
	14	0	0	0	0	0	2	2
	15	0	0	0	0	1	1	2
	17	0	0	0	0	0	1	1
	25	0	0	0	0	0	1	1
	42	0	0	0	0	1	0	1
Toplam		832	2 556	4 223	10 287	5 023	1 132	24 053

Çizelge 3.11 e göre ankete katılanların geçinme düzeyleri incelendiğinde; çok zor geçinen kişilerin oranı %3 (832 kişi), zor geçinen kişilerin oranı %10 (2 256 kişi), biraz zor geçinen kişilerin oranı %17 (4 223 kişi), biraz kolay geçinen kişilerin oranı %42 (10 287), kolay geçinen kişilerin oranı %20 (5 023) ve son olarak çok kolay geçinen kişilerin oranı %4 (1 132 kişi) dir. Dolayısıyla örneklemin çoğunluğunun biraz kolay geçinebilen kişilerden oluştuğu yorumu yapılabilir.

Konut sayısı ve geçinme düzeyine ilişkin frekans dağılımı incelendiğinde; çok zor geçinen kişilerin olduğu aralıkta 64 gözlemin, zor geçinen kişilerin olduğu aralıkta 284 gözlemin, biraz kolay geçinen kişilerin olduğu aralıkta 650 gözlemin, kolay geçinebilen kişilerin olduğu aralıkta 1 439 gözlemin ve son olarak çok kolay geçinebilen kişilerin olduğu aralıkta 466 gözlemin konut sayısının sıfırdan farklı bir değer aldığı görülmektedir. Çok zor geçinebilen kişilerin konut sayısı en fazla dördür. Tablo incelendiğinde; biraz kolay geçinebilen kişilerin daha fazla konutu olduğu görülmektedir.

Çizelge 3.13. Cinsiyet değişkenine ilişkin frekans dağılımı

Cinsiyet			
		Sayı	Yüzde
Düzey	1. Erkek	19 536	81,20
	2. Kadın	4 517	18,80
	Toplam	24 053	100,00

Çizelge 3.14. Konut sayısı ve cinsiyet düzeyine ilişkin frekans dağılımı

		CİNSİYET		Toplam
		Erkek	Kadın	
Konut Sayısı	0	15 294	3 806	19 100
	1	3 070	566	3 636
	2	744	96	840
	3	233	28	261
	4	102	13	115
	5	38	5	43
	6	21	0	21
	7	13	2	15
	8	9	0	9
	9	1	0	1
	10	5	0	5
	14	2	0	2
	15	2	0	2
	17	1	0	1
	25	0	1	1
	42	1	0	1
Toplam		19 536	4 517	24 053

Çizelge 3.13 e göre hanehalkı sorumlusu erkek oranının %81 (19 536 kişi) kadın oranının ise %18 (4 517 kişi) olduğu görülmektedir. Dolayısıyla örneklemdaki erkek sayısının kadın sayısına göre daha fazla olduğu yorumu yapılabilmektedir.

Cinsiyet konut sayısı tablosu incelendiğinde erkeklerden oluşan 15 294 gözlemin başka konutu olmadığı; 1 gözlemin ise 42 adet konutu olduğu görülmektedir. Kadınlardan oluşan 3 806 gözlemin başka bir konuta sahip olmadığı; 1 gözlemin ise 25 adet konutunun olduğu görülmektedir. Sonuç olarak çizelgeden anlaşılabileceği üzere; erkeklerin sahip olduğu konut sayısı kadınlara oranla daha fazladır.

3.3. Uygulama Sonuçları

Kullanılan veri için veriyi en iyi temsil eden modelin belirlenmesi amacıyla PR, NB, ZIP, ZINB, ZTP, ZTNB, PH ve NBH regresyon modellerine ilişkin parametre tahminleri verilmiştir. Yorumlama yapılırken referans kategorileri değişkenlerin ilk düzeyi alınmıştır. Örneğin yaş değişkeni için referans ifade 25 yaş altı gruptur.

Çizelge 3.15. PR modeli parametre tahminleri

	Tahmin	Standart Hata	z değeri	p değeri	$e^{\hat{\beta}_i}$
Sabit Terim	-3,54951	0,19852	-17,880	< 2e-16 *	0,02874
Yaş (25-34)	0,29610	0,19444	1,523	0,1278	1,34460
Yaş (35-44)	0,94099	0,18888	4,982	6,29e-07 *	2,56252
Yaş (45-54)	1,35527	0,18807	7,206	5,75e-13 *	3,87781
Yaş (55-64)	1,76980	0,18783	9,422	< 2e-16 *	5,86968
Yaş (65 ve üzeri)	2,07082	0,18847	10,987	< 2e-16 *	7,93132
Çalışma Durumu (İşsiz)	-0,57950	0,11057	-5,241	1,60e-07 *	0,56018
Çalışma Durumu (İstihdam)	-0,25676	0,03079	-8,339	< 2e-16 *	0,77355
Cinsiyet (Kadın)	-0,28513	0,03754	-7,596	3,05e-14 *	0,75192
Geçinme Durumu(Zor)	0,79747	0,07829	10,186	< 2e-16 *	2,21992
Geçinme Durumu(Biraz zor)	0,02092	0,06835	0,306	0,7595	1,02114
Geçinme Durumu(Biraz kolay)	0,10693	0,05469	1,955	0,0506	1,11286
Geçinme Durumu(Kolay)	0,01466	0,04186	0,350	0,7262	1,01477
Geçinme Durumu(Çok kolay)	-0,03865	0,03110	-1,243	0,2139	0,96209
Gelir Durumu (25 000-50 000tl)	0,61350	0,04012	15,293	< 2e-16 *	1,84688
Gelir Durumu (50 000-75 000tl)	0,92892	0,04531	20,503	< 2e-16 *	2,53177
Gelir Durumu(75 000 TL üstü)	1,35856	0,04867	27,912	< 2e-16 *	3,89059
Eğitim Durumu(okuryazar değil)	-0,43579	0,08540	-5,103	3,34e-07 *	0,64675
Eğitim Durumu(lise altı)	0,37463	0,05982	6,263	3,78e-10 *	1,45445
Eğitim Durumu(lise)	0,31941	0,06661	4,795	1,62e-06 *	1,37632
Eğitim Durumu(yükseköğrenim)	0,45163	0,06537	6,909	4,89e-12 *	1,57087

Çizelge 3.15 de belirtilen p-değerleri incelendiğinde; 25-34 yaş hariç diğer yaş değişken düzeyleri, çalışma durumu değişkeni, cinsiyet değişkeni, gelir durumu değişkeni, eğitim durumu değişkeni ve geçinme durumu zor değişkenlerinin %95 güven düzeyinde istatistiksel olarak anlamlı olduğu görülmektedir.

Hanehalkı sorumlusu 25 yaş altı olan hanelere göre hanehalkı sorumlusu 35 ile 44 yaş arasında olan bir hanehalkının oturduğu konut hariç sahip olduğu konut sayısının 2,56 kat daha fazla olduğu söylenebilir. Yine hanehalkı sorumlusu 45 yaş ile 54 yaş arasında olan haneler 3,88 kat daha fazla, hanehalkı sorumlusu 55 yaş ile 64 yaş arasında olan haneler 5,87 kat daha fazla ve hanehalkı sorumlusu 65 yaş ve üstünde olan haneler 7,93 kat daha fazla konuta sahiptir. Yani yaş artıkça konut sayısına etki artmıştır.

Hanehalkı sorumlusu işsiz olan hanelere göre hanehalkı sorumlusu iş gücüne dahil olmayan hanelerin oturduğu konut hariç sahip olduğu konut sayısı 1,79 (1/0,56) kat daha fazladır. Yine hanehalkı sorumlusu istihdamda olan hanelere göre hanehalkı sorumlusu iş gücüne dahil olmayan hanelerin sahip olduğu konut sayısı 1,30 (1/0,77) kat daha fazladır.

Hanehalkı sorumlusu erkekler, kadınlara göre sahip olduğu konut sayısı 1,33 (1/0,75) kat daha fazladır.

Geçinme durumu zor geçinen haneler, çok zor geçinenlere göre 2,21 kat daha fazla konuta sahiptir.

Gelir durumu değişkeninde ise durum şu şekildedir: Yıllık geliri 25 000 ile 50 000 arası olan haneler, yıllık geliri 25 000 altında olan hanelere göre 1,84 kat daha fazla konuta sahipken, yıllık geliri 50 000 ile 75 000 arası olanlar yıllık geliri 25 000 altında olanlara göre 2,53 kat daha fazla ve son olarak yıllık geliri 75 000 üstünde yer alan haneler, yıllık geliri 25 000 altında olanlara göre 3,89 kat daha fazla konuta sahiptir.

Hanehalkı sorumlusu bir okul bitirmeyen haneler, hanehalkı sorumlusu okuryazar olmayan hanelere göre 1,56 (1/0,64) kat daha fazla konuta sahip iken, hanehalkı sorumlusu eğitim durumu lise altı eğitim düzeyine sahip haneler, hanehalkı sorumlusu bir okul bitirmeyen hanelere göre 1,45 kat daha fazla, hanehalkı sorumlusu lise eğitim düzeyine sahip haneler, hanehalkı sorumlusu bir okul bitirmeyen hanelere göre 1,38 kat daha fazla etki etmekte ve son olarak hanehalkı sorumlusu eğitim durumu yükseköğretim düzeyine sahip haneler, hanehalkı sorumlusu bir okul bitirmeyen hanelere göre 1,57 kat daha fazla etki etmektedir.

Çizelge 3.16. NB regresyon modeli parametre tahminleri

	Tahmin	Standart Hata	z değeri	p değeri	e^{β_i}
Sabit Terim	-3,56622	0,21135	-16,874	< 2e-16 *	0,02826
Yaş (25-34)	0,28548	0,20433	1,397	0,1624	1,33040
Yaş (35-44)	0,92715	0,19851	4,671	3,00e-06 *	2,52730
Yaş (45-54)	1,35123	0,19775	6,833	8,31e-12 *	3,86217
Yaş (55-64)	1,78652	0,19765	9,039	< 2e-16 *	5,96865
Yaş (65 ve üzeri)	2,09111	0,19879	10,519	< 2e-16 *	8,09389
Çalışma Durumu (İşsiz)	-0,59495	0,12308	-4,834	1,34e-06 *	0,55159
Çalışma Durumu (İstihdam)	-0,28469	0,03901	-7,297	2,94e-13 *	0,75225
Cinsiyet (Kadın)	-0,30074	0,04505	-6,675	2,47e-11 *	0,74027
Geçinme Durumu(Zor)	0,78165	0,08744	8,939	< 2e-16 *	2,18507
Geçinme Durumu(Biraz zor)	0,02925	0,07614	0,384	0,7009	1,02968
Geçinme Durumu(Biraz kolay)	0,11254	0,06123	1,838	0,0661	1,11912
Geçinme Durumu(Kolay)	0,02492	0,04744	0,525	0,5994	1,02523
Geçinme Durumu(Çok kolay)	-0,03989	0,03583	-1,113	0,2655	0,96090
Gelir Durumu (25 000-50 000tl)	0,62620	0,04548	13,768	< 2e-16 *	1,87049
Gelir Durumu (50 000-75 000tl)	0,97236	0,05286	18,395	< 2e-16 *	2,64418
Gelir Durumu(75 000 TL üstü)	1,41932	0,05879	24,141	< 2e-16 *	4,13431
Eğitim Durumu(okuryazar değil)	-0,42859	0,09669	-4,433	9,31e-06 *	0,65143
Eğitim Durumu(lise altı)	0,37976	0,07078	5,365	8,09e-08 *	1,46193
Eğitim Durumu(lise)	0,33112	0,07947	4,167	3,09e-05 *	1,39253
Eğitim Durumu(yükseköğrenim)	0,46907	0,07889	5,946	2,75e-09 *	1,59851

Çizelge 3.16 da belirtilen p-değerleri incelendiğinde; 25-34 yaş hariç diğer yaş değişken düzeyleri, çalışma durumu değişkeni, cinsiyet değişkeni, gelir durumu değişkeni, eğitim durumu değişkeni ve biraz kolay geçinme durumu değişkenlerinin %95 güven düzeyinde istatistiksel olarak anlamlı olduğu görülmektedir.

Hanehalkı sorumlusu 25 yaş altında olanlara göre hanehalkı sorumlusu 35 ile 44 yaş arasında olan hanehalkının oturduğu konut hariç sahip olduğu konut sayısı 2,53 kat daha fazla olduğu söylenebilir. Hanehalkı sorumlusu 25 yaş altında olanlara göre hanehalkı sorumlusu 45 yaş ile 54 yaş arasında olanlar 3,86 kat daha fazla, hanehalkı sorumlusu 55 yaş ile 64 yaş arasında olanlar 5,97 kat daha fazla ve hanehalkı sorumlusu 65 yaş ve üstünde olan haneler ise 8,09 kat daha fazla konuta sahiptir. Yani yaş artıkcı oturduğu konut hariç sahip olunan konut sayısı artmıştır.

Hanehalkı sorumlusu işsiz hanelere göre hanehalkı sorumlusu iş gücüne dahil olmayan hanelerin oturduğu konut hariç sahip olduğu konut sayısı 1,81 (1/0,55) kat daha fazladır. Yine Hanehalkı sorumlusu istihdamda olan hanelere göre hanehalkı sorumlusu iş gücüne dahil olmayan haneler 1,33 (1/0,75) kat daha fazla konuta sahiptir.

Hanehalkı sorumlusu erkek olan haneler, kadınlara göre 1,35 (1/0,74) kat daha fazla konuta sahiptir.

Geçinme durumu zor olan haneler, çok zor geçinenlere göre 2,19 kat daha fazla konuta sahiptir.

Gelir durumu değişkeninde ise durum şu şekildedir: Yıllık geliri 25 000 altında olan hanelere göre, yıllık geliri 25 000 ile 50 000 arasında olan haneler 1,87 kat daha fazla, yıllık geliri 50 000 ile 75 000 arası olan haneler 2,64 kat daha fazla ve son olarak yıllık geliri 75 000 üstünde yer alan haneler 4,13 kat daha fazla konuta sahiptir.

Hanehalkı sorumlusu bir okul bitirmeyen haneler, hanehalkı sorumlusu okur yazar olmayan hanelere göre oturduğu konut hariç sahip olduğu konut sayısı 1,54 (1/0,65) kat daha fazla iken, hanehalkı sorumlusu lise altı eğitim düzeyine sahip haneler, hanehalkı sorumlusu bir okul bitirmeyen hanelere göre 1.46 kat daha fazla, hanehalkı sorumlusu lise eğitim düzeyine sahip haneler, hanehalkı sorumlusu bir okul bitirmeyen hanelere göre 1,39 kat daha fazla ve son olarak hanehalkı sorumlusu eğitim durumu yükseköğretim düzeyine sahip haneler, hanehalkı sorumlusu bir okul bitirmeyen hanelere göre 1,60 kat daha fazla konuta sahiptir.

Çizelge 3.17. ZIP regresyon modeli parametre tahminleri (log kısım)

(Log)	Tahmin	Standart Hata	z değeri	p değeri	$e^{\hat{\beta}_i}$
Sabit Terim	-1,44642	0,43929	-3,293	0,000993 *	0,23541
Yaş (25-34)	-0,25576	0,43825	-0,584	0,559490	0,77433
Yaş (35-44)	0,23761	0,42335	0,561	0,574620	1,26821
Yaş (45-54)	0,39569	0,42190	0,938	0,348302	1,48541
Yaş (55-64)	0,62480	0,42177	1,481	0,138511	1,86787
Yaş (65 ve üzeri)	1,01660	0,42307	2,403	0,016264*	2,76378
Çalışma Durumu (İşsiz)	-0,19221	0,21142	-0,909	0,363281	0,82513
Çalışma Durumu (İstihdam)	-0,23506	0,04742	-4,957	7,17e-07 *	0,79052
Cinsiyet (Kadın)	-0,31727	0,06214	-5,106	3,29e-07 *	0,72813
Geçinme Durumu(Zor)	0,74280	0,17710	4,194	2,74e-05 *	2,10181
Geçinme Durumu(Biraz zor)	0,20004	0,15716	1,273	0,203086	1,22145
Geçinme Durumu(Biraz kolay)	0,19624	0,12055	1,628	0,103561	1,21682
Geçinme Durumu(Kolay)	0,01262	0,08503	0,148	0,882047	1,01270
Geçinme Durumu(Çok kolay)	-0,10864	0,05793	-1,875	0,060733	0,89705
Gelir Durumu (25 000-50 000tl)	0,36423	0,07431	4,902	9,50e-07 *	1,43941
Gelir Durumu (50 000-75 000tl)	0,47919	0,08204	5,841	5,19e-09 *	1,61477
Gelir Durumu(75 000 TL üstü)	0,86194	0,08622	9,997	< 2e-16 *	2,36775
Eğitim Durumu(okuryazar değil)	0,17861	0,14525	1,230	0,218828	1,19555
Eğitim Durumu(lise altı)	0,30785	0,09975	3,086	0,002027 *	1,36050
Eğitim Durumu(lise)	0,03259	0,10961	0,297	0,766215	1,03313
Eğitim Durumu(yükseköğrenim)	0,10900	0,10660	1,022	0,306566	1,11516

Sıfır yığılmalı poisson regresyonun log kısmına göre (sayma verisi olan kısım) 65 yaş ve üstü yaş değişkeni, istihdam çalışma düzeyi değişkeni, cinsiyet, geçinme durumunun 2.düzeyi zor, gelir durumu ve lise altı eğitim durumu istatistiksel olarak anlamlı bulunmuştur ($p<0,05$).

Hanehalkı sorumlusu 25 yaş altı olanlara göre hanehalkı sorumlusu 65 yaş ve üzerinde olan hanehalkının oturduğu konut hariç sahip olduğu konut sayısı 2,76 kat daha fazladır.

Hanehalkı sorumlusu istihdamda olanlara göre hanehalkı sorumlusu iş gücüne dahil olmayanların oturduğu konut hariç sahip olduğu konut sayısı 1,26 (1/0,79) kat daha fazladır.

Hanehalkı sorumlusu erkek olan bir hanehalkının oturduğu konut hariç sahip olduğu konut sayısı, hanehalkı sorumlusu kadınlara göre 1,37 (1/0,73) kat daha fazladır.

Hanehalkının geçinme durumu düzeyi çok zor olan hanelere göre geçinme durumu zor düzeyinde olanların oturduğu konut hariç sahip olduğu konut sayısı 2,10 kat daha fazladır.

Gelir durumu değişkeninde ise durum şu şekildedir: Yıllık geliri 25 000 ile 50 000 arası olan hanelerin oturduğu konut hariç sahip olduğu konut sayısı, yıllık geliri 25 000 altında olan hanelere göre 1,44 kat daha fazla iken, yıllık geliri 50 000 ile 75 000 olanlar yıllık geliri 25 000 altında olanlara göre 1,61 kat daha fazla ve son olarak yıllık geliri 75 000 üstünde yer alan haneler, yıllık geliri 25 000 altında olanlara göre 2,37 kat daha fazla konuta sahiptir.

Hanehalkı sorumlusu lise altı eğitim düzeyine sahip haneler, hanehalkı sorumlusu bir okul bitirmeyenlere göre oturduğu konut hariç sahip olduğu konut sayısı 1,36 kat daha fazladır.

Çizelge 3.18. ZIP regresyon modeli parametre tahminleri (lojit kısım)

(Lojit)	Tahmin	Standart Hata	z değeri	p değeri	$e^{\hat{\beta}_i}$
Sabit Terim	2,547819	0,572472	4,451	8,56e-06 *	12,779202
Yaş (25-34)	-0,777447	0,568426	-1,368	0,171400	0,459578
Yaş (35-44)	-1,076657	0,537240	-2,004	0,045064 *	0,340733
Yaş (45-54)	-1,642868	0,536351	-3,063	0,002191 *	0,193425
Yaş (55-64)	-2,152734	0,538207	-4,000	6,34e-05 *	0,116166
Yaş (65 ve üzeri)	-1,894465	0,538905	-3,515	0,000439 *	0,150399
Çalışma Durumu (İşsiz)	0,654563	0,310937	2,105	0,035280 *	1,924301
Çalışma Durumu (İstihdam)	0,099921	0,103729	0,963	0,335402	1,105084
Cinsiyet (Kadın)	-0,056576	0,122972	-0,460	0,645466	0,944995
Geçinme Durumu(Zor)	0,051069	0,279244	0,183	0,854890	1,052396
Geçinme Durumu(Biraz zor)	0,372215	0,244220	1,524	0,127485	1,450945
Geçinme Durumu(Biraz kolay)	0,236590	0,189869	1,246	0,212738	1,266922
Geçinme Durumu(Kolay)	0,002151	0,138775	0,016	0,987631	1,002153
Geçinme Durumu(Çok kolay)	-0,136635	0,100785	-1,356	0,175192	0,872289
Gelir Durumu (25 000-50 000tl)	-0,459543	0,118374	-3,882	0,000104 *	0,631572
Gelir Durumu (50 000-75 000tl)	-0,992823	0,151022	-6,574	4,90e-11 *	0,370529
Gelir Durumu(75 000 TL üstü)	-1,173760	0,166648	-7,043	1,88e-12 *	0,309202
Eğitim Durumu(okuryazar değil)	0,963325	0,217767	4,424	9,70e-06 *	2,620395
Eğitim Durumu(lise altı)	-0,176360	0,176518	-0,999	0,317743	0,838316
Eğitim Durumu(lise)	-0,713535	0,212053	-3,365	0,000766 *	0,489909
Eğitim Durumu(yükseköğrenim)	-0,898538	0,209925	-4,280	1,87e-05 *	0,407164

Sıfır yığılmalı poisson regresyonun lojit kısmına göre (sıfır yığılmalı olan kısım) yaş değişkeninde 25-34 yaş, çalışma durumunda işsiz, gelir durumu ve eğitim durumu lise altı hariç istatistiksel olarak anlamlı bulunmuştur ($p < 0,05$).

Hanehalkı sorumlusu 25 yaş altında olan haneler, hanehalkı sorumlusu 35 ile 44 yaş arasında yer alan bir haneye göre oturduğu konut hariç sahip olduğu konut sayısı 2,94 (1/0,34) kat daha fazla, hanehalkı sorumlusu 45 ile 54 yaş arasında yer alan hanelere göre 5,26 (1/0,19) kat daha fazla, hanehalkı sorumlusu 55 ile 64 yaş arasında yer alan hanelere göre 9,09 (1/0,12) kat daha fazla ve son olarak hanehalkı sorumlusu 65 yaş ve üzerinde olan hanelere göre ise 6,67 (1/0,15) kat daha fazla konuta sahiptir.

Hanehalkı sorumlusu iş gücüne dahil olmayanların olduğu hanelere göre hanehalkı sorumlusu işsiz olanların oturduğu konut hariç sahip olduğu konut sayısı 1,92 kat daha fazladır.

Gelir durumu değişkeninde ise durum şu şekildedir: Yıllık geliri 25 000 altında olan haneler, yıllık geliri 25 000 ile 50 000 arası olan haneler göre 1,59 (1/0,63) kat daha fazla, yıllık geliri 50 000 ile 75 000 arası olan hanelere göre 2,70 (1/0,37) kat daha fazla ve son olarak yıllık geliri 75 000 üstünde yer alan hanelere göre 3,23 (1/0,31) kat daha fazla konuta sahiptir.

Hanehalkı sorumlusu bir okul bitirmeyen hanelere göre hanehalkı sorumlusu okuryazar olmayan hanelerin oturduğu konut hariç sahip olduğu konut sayısı 2,62 kat daha fazla iken, hanehalkı sorumlusu eğitim durumu lise düzeyi olan hanelere göre, hanehalkı sorumlusu bir okul bitirmeyen haneler 2,08 (1/0,49) kat daha fazla ve son olarak hanehalkı sorumlusu eğitim durumu yükseköğretim düzeyi olanlara göre hane sorumlusu bir okul bitirmeyenlerden oluşan haneler 2,44 (1/0,41) kat daha fazla konuta sahiptir.

Çizelge 3.19. ZINB regresyon modeli parametre tahminleri (log kısım)

(Log)	Tahmin	Standart Hata	z değeri	p değeri	$e^{\hat{\beta}_i}$
Sabit Terim	-2.361326	0.517310	-4.565	5.00e-06 *	0.094295
Yaş (25-34)	-0.228752	0.482436	-0.474	0.63539	0.795526
Yaş (35-44)	0.389572	0.480328	0.811	0.41733	1.476349
Yaş (45-54)	0.616388	0.489004	1.260	0.20749	1.852226
Yaş (55-64)	0.929180	0.496126	1.873	0.06109	2.532432
Yaş (65 ve üzeri)	1.261941	0.493429	2.557	0.01054 *	3.532271
Çalışma Durumu (İşsiz)	-0.432076	0.185210	-2.333	0.01965 *	0.649160
Çalışma Durumu (İstihdam)	-0.246530	0.045636	-5.402	6.59e-08 *	0.781508
Cinsiyet (Kadın)	-2.361326	0.517310	-4.565	5.00e-06 *	0.094295
Geçinme Durumu(Zor)	0.917878	0.129362	7.095	1.29e-12 *	2.503971
Geçinme Durumu(Biraz zor)	-0.02768	0.114482	-0.242	0.80889	0.972700
Geçinme Durumu(Biraz kolay)	0.211603	0.090190	2.346	0.01897 *	1.235657
Geçinme Durumu(Kolay)	-0.002474	0.068098	-0.036	0.97101	0.997529
Geçinme Durumu(Çok kolay)	-0.027523	0.049504	-0.556	0.57823	0.972852
Gelir Durumu (25 000-50 000tl)	0.458059	0.072445	6.323	2.57e-10 *	1.581002
Gelir Durumu (50 000-75 000tl)	0.637236	0.085240	7.476	7.67e-14 *	1.891246
Gelir Durumu(75 000 TL üstü)	1.019535	0.089419	11.402	< 2e-16 *	2.771906
Eğitim Durumu(okuryazar değil)	0.005480	0.173487	0.032	0.97480	1.005495
Eğitim Durumu(lise altı)	0.264673	0.101819	2.599	0.00934*	1.303005
Eğitim Durumu(lise)	0.183111	0.109773	1.668	0.09530	1.200948
Eğitim Durumu(yükseköğrenim)	0.308726	0.108676	2.841	0.00450*	1.361689

Sıfır yığılmalı nb regresyonun log kısmına göre (sayma verisi olan kısım) yaş değişkeninde 65 yaş ve üzeri, çalışma durumu, cinsiyet, geçinme durumunda zor ve biraz kolay, gelir durumu ve eğitim durumunda lise altı ve yükseköğrenim istatistiksel olarak anlamlı bulunmuştur ($p < 0,05$).

Hanehalkı sorumlusu 25 yaş altı grubuna göre 65 yaş ve üzerinde olan hanehalkının oturduğu konut hariç sahip olduğu konut sayısı 3,53 kat daha fazladır.

Hanehalkı sorumlusu istihdamda olanların olduğu gruba göre hanehalkı sorumlusu iş gücüne dahil olmayanların oturduğu konut hariç sahip olduğu konut sayısı 1,28 (1/0,65) kat daha fazla iken hanehalkı sorumlusu işsiz olanlara göre ise iş gücüne dahil olmayanların konut

sayısı 1,54 (1/0,78) kat daha fazladır.

Hanehalkı sorumlusu kadınlara göre, erkeklerin oturduğu konut hariç sahip olduğu konut sayısı 11,11 (1/0,09) kat daha fazladır.

Hanehalkının geçinme durumu düzeyi çok zor düzeyinde olanlara göre geçinme durumu zor düzeyinde olanların oturduğu konut hariç sahip olduğu konut sayısı 2,50 kat daha fazla iken, biraz kolay düzeyinde olan haneler 1,24 kat daha fazladır.

Gelir durumu değişkeninde ise durum şu şekildedir: Yıllık geliri 25 000 ile 50 000 arası geliri olan hanelerin oturduğu konut hariç sahip olduğu konut sayısı yıllık geliri 25 000 altında olan hanelere göre 1,58 kat daha fazla iken, yıllık geliri 50 000 ile 75 000 arası geliri olan haneler, yıllık geliri 25 000 altında olanlara göre 1,89 kat daha ve son olarak yıllık geliri 75 000 üstünde yer alan haneler, yıllık geliri 25 000 altında olanlara göre 2,77 kat daha fazla konuta sahiptir.

Hanehalkı sorumlusu lise altı eğitim düzeyine sahip haneler, hanehalkı sorumlusu bir okul bitirmeyenlere göre 1,30 kat daha fazla konuta sahip iken, hanehalkı sorumlusu eğitim durumu yükseköğretim olanlar, bir okul bitirmeyenlere göre 1,36 kat daha fazla konuta sahiptir.

Çizelge 3.20. ZINB regresyon modeli parametre tahminleri (lojit kısım)

(Lojit)	Tahmin	Standart Hata	z değeri	p değeri	$e^{\hat{\beta}_i}$
Sabit Terim	2,49605	1,04509	2,388	0,016924 *	12,13447
Yaş (25-34)	-1,21389	0,92840	-1,308	0,191039	0,29704
Yaş (35-44)	-1,30234	0,91357	-1,426	0,153997	0,27189
Yaş (45-54)	-2,49270	0,93894	-2,655	0,007935 *	0,08269
Yaş (55-64)	-4,12570	1,04341	-3,954	7,68e-05 *	0,01615
Yaş (65 ve üzeri)	-3,27745	0,96237	-3,406	0,000660 *	0,03772
Çalışma Durumu (İşsiz)	0,61424	0,61424	1,000	0,317310	1,84825
Çalışma Durumu (İstihdam)	0,25497	0,28838	0,884	0,376606	1,29042
Cinsiyet (Kadın)	0,04109	0,27894	0,147	0,882893	1,04195
Geçinme Durumu(Zor)	1,32081	0,70104	1,884	0,059554	3,74645
Geçinme Durumu(Biraz zor)	0,09304	0,65105	0,143	0,886356	1,09751
Geçinme Durumu(Biraz kolay)	0,90187	0,46451	1,942	0,052191	2,46421
Geçinme Durumu(Kolay)	0,08542	0,30124	0,284	0,776756	1,08917
Geçinme Durumu(Çok kolay)	0,14323	0,20790	0,689	0,490885	1,15400
Gelir Durumu (25 000-50 000tl)	-0,77308	0,23111	-3,345	0,000823 *	0,46159
Gelir Durumu (50 000-75 000tl)	-2,32219	0,56391	-4,118	3,82e-05 *	0,09806
Gelir Durumu(75 000 TL üstü)	-4,01158	1,18872	-3,375	0,000739 *	0,01810
Eğitim Durumu(okuryazar değil)	1,37459	0,46967	2,927	0,003426 *	3,95346
Eğitim Durumu(lise altı)	-0,87434	0,42214	-2,071	0,038340 *	0,41714
Eğitim Durumu(lise)	-1,20934	0,46102	-2,623	0,008711 *	0,29839
Eğitim Durumu(yükseköğrenim)	-1,16979	0,48375	-2,418	0,015598 *	0,31043

Sıfır yığılmalı NB regresyonun lojit kısmına göre (sıfır yığılmalı verileri olan kısım) yaş değişkeninde 45-54, 55-64 ve 65 yaş ve üzeri, gelir durumu ve eğitim durumu istatistiksel olarak anlamlı bulunmuştur ($p < 0,05$).

Hanehalkı sorumlusu 45 ile 54 yaş düzeyinde yer alan hanelere göre hanehalkı sorumlusu 25 yaş altı grubunda yer alan haneler 12,50 (1/0,08) kat daha fazla konuta sahip, hanehalkı sorumlusu 55 ile 64 yaş düzeyinde olan haneler ve hanehalkı sorumlusu 65 yaş ve üzerinde olan hanelere göre hanehalkı sorumlusu 25 yaş altı grubunda yer alan haneler sırasıyla 100 (1/0,02) ve 33,33 (1/0,04) kat daha fazla konuta sahiptir.

Gelir durumu değişkeninde ise durum şu şekildedir: Yıllık geliri 25 000 altı olan haneler, yıllık geliri 25 000 ile 50 000 arası olan hanelere göre 2,17 (1/0,46) kat daha fazla konuta sahipken, yıllık geliri 50 000 ile 75 000 arası olan hanelere ve yıllık geliri 75 000 üstünde yer alan hanelere göre sırasıyla 10 (1/0,1) ve 50 (1/0,02) kat daha fazla konuta sahiptir.

Hanehalkı sorumlusu bir okul bitirmeyen düzeye sahip hanelere göre hanehalkı sorumlusu okuryazar olmayan haneler 3,95 kat daha fazla konuta sahipken, hanehalkı sorumlusu lise altı düzeyinde olan hanelere göre hane sorumlusu bir okul bitirmeyen haneler 2,38 (1/0,42)

kat daha fazla konuta sahiptir. Hanehalkı sorumlusu bir okul bitirmeyen haneler, hanehalkı sorumlusu lise düzeyine sahip hanelere göre 3,33 (1/0,30) kat daha fazla ve son olarak hanehalkı sorumlusu eğitim durumu yükseköğretim düzeyinde olan hanelere göre de 3,23 (1/0,31) kat daha fazla konuta sahiptir.

Çizelge 3.21. ZTP regresyon modeli parametre tahminleri

	Tahmin	Standart Hata	z değeri	p değeri	$e^{\hat{\beta}_i}$
Sabit Terim	-1,407148	0,453308	-3,104	0,00191 *	0,244841
Yaş (25-34)	-0,217694	0,454342	-0,479	0,63184	0,804372
Yaş (35-44)	0,184943	0,437621	0,423	0,67258	1,203150
Yaş (45-54)	0,330924	0,435633	0,760	0,44747	1,392254
Yaş (55-64)	0,598724	0,435044	1,376	0,16875	1,819795
Yaş (65 ve üzeri)	0,970855	0,435814	2,228	0,02590 *	2,640201
Çalışma Durumu (İşsiz)	-0,273129	0,232022	-1,177	0,23913	0,760995
Çalışma Durumu (İstihdam)	-0,227764	0,049811	-4,573	4,82e-06 *	0,796312
Cinsiyet (Kadın)	-0,343462	0,068202	-5,036	4,75e-07 *	0,709310
Geçinme Durumu(Zor)	0,717448	0,179217	4,003	6,25e-05 *	2,049197
Geçinme Durumu(Biraz zor)	0,174119	0,159194	1,094	0,27407	1,190197
Geçinme Durumu(Biraz kolay)	0,205528	0,122903	1,672	0,09447	1,228173
Geçinme Durumu(Kolay)	0,006799	0,087804	0,077	0,93828	1,006822
Geçinme Durumu(Çok kolay)	-0,135889	0,060176	-2,258	0,02393 *	0,872940
Gelir Durumu (25 000-50 000tl)	0,347506	0,077199	4,501	6,75e-06 *	1,415533
Gelir Durumu (50 000-75 000tl)	0,440717	0,085126	5,177	2,25e-07 *	1,553821
Gelir Durumu(75 000 TL üstü)	0,883843	0,087553	10,095	< 2e-16 *	2,420183
Eğitim Durumu(okuryazar değil)	0,189770	0,152188	1,247	0,21242	1,208972
Eğitim Durumu(lise altı)	0,324804	0,105388	3,082	0,00206 *	1,383759
Eğitim Durumu(lise)	0,038918	0,117378	0,332	0,74022	1,039685
Eğitim Durumu(yükseköğrenim)	0,084442	0,113306	0,745	0,45612	1,088110

Çizelge de belirtilen p-değerleri incelendiğinde; 65 yaş ve üzeri yaş değişkeni, çalışma durumu (istihdam) değişkeni, cinsiyet değişkeni, geçinme durumu (zor ve çok kolay) değişkeni, gelir durumu değişkeni ve lise altı eğitim durumu değişkenlerinin %95 güven düzeyinde istatistiksel olarak anlamlı olduğu görülmektedir.

Hanehalkı sorumlusu 65 yaş ve üzeri olan hanehalkının oturduğu konut hariç sahip olduğu konut sayısı, hanehalkı sorumlusu 25 yaş altında olanlara göre 2,64 kat daha fazla olduğu görülmektedir.

Hanehalkı sorumlusu istihdamda olanların olduğu gruba göre iş gücüne dahil olmayanların oturduğu konut hariç sahip olduğu konut sayısı 1,25 (1/0,80) kat daha fazla olduğu görülmektedir.

Hanehalkı sorumlusu kadın olan hanelere göre, hanehalkı sorumlusu erkeklerden oluşan hanelerin oturduğu konut hariç sahip olduğu konut sayısı 1,41 (1/0,71) kat daha fazladır.

Hanehalkının geçinme durumu düzeyi çok zor olanlara göre geçinme durumu zor düzeyinde olanların oturduğu konut hariç sahip olduğu konut sayısı 2,05 kat daha fazla iken, geçinme durumu düzeyi çok kolay olanlara göre geçinme durumu çok zor olan haneler 1,15 (1/0,87) kat daha konut sayısına sahiptir.

Gelir durumu değişkeninde ise durum şu şekildedir: Yıllık geliri 25 000 ile 50 000 arasında olan hanehalkının oturduğu konut hariç sahip olduğu konut sayısı yıllık geliri 25 000 den az olanlara göre 1,42 kat daha fazla, yıllık geliri 50 000 ile 75 000 arasında olan hanehalkının oturduğu konut hariç sahip olduğu konut sayısı yıllık geliri 25 000 den az olanlara göre 1,55 kat daha fazla ve son olarak yıllık geliri 75 000 den fazla olan hanehalkının oturduğu konut hariç sahip olduğu konut sayısı yıllık geliri 25 000 den az olanlara göre 2,42 kat daha fazla olduğu görülmektedir.

Hanehalkı sorumlusu lise altı eğitim düzeyinde olan hanelerin oturduğu konut hariç sahip olduğu konut sayısı hanehalkı sorumlusu bir okul bitirmeyenlere göre 1,38 kat daha fazladır.

Çizelge 3.22. ZTNB regresyon modeli parametre tahminleri

	Tahmin	Standart Hata	z değeri	p değeri	$e^{\hat{\beta}_i}$
Sabit Terim	-10,803985	47,318178	-0,228	0,819393	0,000020
Yaş (25-34)	-0,428827	0,607301	-0,706	0,480113	0,651273
Yaş (35-44)	0,083302	0,586734	0,142	0,887100	1,086870
Yaş (45-54)	0,287228	0,584694	0,491	0,623254	1,332728
Yaş (55-64)	0,630416	0,584638	1,078	0,280899	1,878392
Yaş (65 ve üzeri)	1,101372	0,587337	1,875	0,060766	3,008291
Çalışma Durumu (İşsiz)	-0,310421	0,328639	-0,945	0,344881	0,733138
Çalışma Durumu (İstihdam)	-0,296606	0,085407	-3,473	0,000515 *	0,743337
Cinsiyet (Kadın)	-0,516002	0,108322	-4,764	1,90e-06 *	0,596902
Geçinme Durumu(Zor)	0,868239	0,241652	3,593	0,000327 *	2,382711
Geçinme Durumu(Biraz zor)	0,308832	0,212384	1,454	0,145912	1,361834
Geçinme Durumu(Biraz kolay)	0,263859	0,167525	1,575	0,115247	1,301945
Geçinme Durumu(Kolay)	0,055208	0,124179	0,445	0,656623	1,056760
Geçinme Durumu(Çok kolay)	-0,139968	0,089140	-1,570	0,116365	0,869386
Gelir Durumu (25 000-50 000tl)	0,448239	0,114518	3,914	9,07e-05 *	1,565553
Gelir Durumu (50 000-75 000tl)	0,595779	0,131908	4,517	6,28e-06 *	1,814444
Gelir Durumu(75 000 TL üstü)	1,103745	0,142768	7,731	1,07e-14 *	3,015438
Eğitim Durumu(okuryazar değil)	0,230731	0,239516	0,963	0,335387	1,259520
Eğitim Durumu(lise altı)	0,306759	0,170442	1,800	0,071894	1,359013
Eğitim Durumu(lise)	-0,009637	0,191841	-0,050	0,959935	0,990409
Eğitim Durumu(yükseköğrenim)	0,061362	0,187338	0,328	0,743253	1,063284

Çizelgede belirtilen p-değerleri incelendiğinde; çalışma durumu (istihdam) değişkeni, cinsiyet değişkeni, zor geçinme durumu değişkeni ve gelir durumu değişkenleri %95 güven düzeyinde istatistiksel olarak anlamlı olduğu görülmektedir.

Hanehalkı sorumlusu istihdamda olanlara göre hane sorumlusu iş gücüne dahil olmayanların oturduğu konut hariç sahip olduğu konut sayısı 1,35 (1/0,74) kat daha fazladır.

Hanehalkı sorumlusu erkek olan bir hanehalkının oturduğu konut hariç sahip olduğu konut sayısı hanehalkı sorumlusu kadınlara göre 1,68 (1/0,60) kat daha fazladır.

Gelir durumu değişkeninde ise durum şu şekildedir: Yıllık geliri 25 000 ile 50 000 arasında olan hanehalkının oturduğu konut hariç sahip olduğu konut sayısı yıllık geliri 25 000 den az olan hanelere göre 1,57 kat daha fazla, yıllık geliri 50 000 ile 75 000 arasında olan hanehalkının oturduğu konut hariç sahip olduğu konut sayısı yıllık geliri 25 000 den az olanlara göre göre 1,81 kat daha fazla ve son olarak yıllık geliri 75 000 üstünde olan hanehalkının oturduğu konut hariç sahip olduğu konut sayısı yıllık geliri 25 000 altında olanlara göre 3,02 kat daha fazla olduğu görülmektedir.

Çizelge 3.23. PH regresyon modeli parametre tahminleri (log kısım)

(Log)	Tahmin	Standart Hata	z değeri	p değeri	$e^{\hat{\beta}_i}$
Sabit Terim	-1,407143	0,453309	-3,104	0,00191 *	0,244842
Yaş (25-34)	-0,217688	0,454343	-0,479	0,63185	0,804376
Yaş (35-44)	0,184949	0,437623	0,423	0,67257	1,203157
Yaş (45-54)	0,330928	0,435634	0,760	0,44747	1,392260
Yaş (55-64)	0,598726	0,435045	1,376	0,16875	1,819799
Yaş (65 ve üzeri)	0,970856	0,435815	2,228	0,02590 *	2,640204
Çalışma Durumu (İşsiz)	-0,273124	0,232021	-1,177	0,23913	0,760998
Çalışma Durumu (İstihdam)	-0,227769	0,049811	-4,573	4,82e-06 *	0,796308
Cinsiyet (Kadın)	-0,343466	0,068202	-5,036	4,75e-07 *	0,709308
Geçinme Durumu(Zor)	0,717471	0,179220	4,003	6,25e-05 *	2,049244
Geçinme Durumu(Biraz zor)	0,174099	0,159197	1,094	0,27413	1,190173
Geçinme Durumu(Biraz kolay)	0,205544	0,122904	1,672	0,09445	1,228193
Geçinme Durumu(Kolay)	0,006788	0,087804	0,077	0,93838	1,006811
Geçinme Durumu(Çok kolay)	-0,135884	0,060176	-2,258	0,02394 *	0,872944
Gelir Durumu (25 000-50 000tl)	0,347500	0,077199	4,501	6,75e-06 *	1,415524
Gelir Durumu (50 000-75 000tl)	0,440710	0,085126	5,177	2,25e-07 *	1,553810
Gelir Durumu(75 000 TL üstü)	0,883839	0,087553	10,095	< 2e-16 *	2,420173
Eğitim Durumu(okuryazar değil)	0,189766	0,152187	1,247	0,21242	1,208967
Eğitim Durumu(lise altı)	0,324799	0,105388	3,082	0,00206 *	1,383752
Eğitim Durumu(lise)	0,038913	0,117377	0,332	0,74025	1,039680
Eğitim Durumu(yükseköğrenim)	0,084437	0,113306	0,745	0,45614	1,088104

Poisson hurdle regresyonun log kısmına göre 65 yaş ve üzeri yaş değişkeni, istihdam çalışma durumu, cinsiyet, geçinme durumu zor ve çok kolay, gelir durumu ve eğitim durumu lise altı istatistiksel olarak anlamlı bulunmuştur ($p < 0,05$).

Hanehalkı sorumlusu 25 yaş altı grubuna göre 65 yaş ve üzerinde olan hanehalkının oturduğu konut hariç sahip olduğu konut sayısı 2,64 kat daha fazladır.

Hanehalkı sorumlusu istihdamda olan hanelere göre hanehalkı sorumlusu iş gücüne dahil olmayan hanelerin oturduğu konut hariç sahip olduğu konut sayısı 1,25 (1/0,80) kat daha fazladır.

Hanehalkı sorumlusu kadınlara göre, erkeklerin oturduğu konut hariç sahip olduğu konut sayısı 1,41 (1/0,71) kat daha fazladır.

Hanehalkının geçinme durumu düzeyi çok zor düzeyinde olanlara göre geçinme durumu zor düzeyinde olanların oturduğu konut hariç sahip olduğu konut sayısı 2,05 kat daha fazla iken, çok kolay düzeyinde olan hanelere göre geçinme durumu çok zor olanların konutları 50 (1/0,87) kat daha fazladır.

Gelir durumu değişkeninde ise durum şu şekildedir: Yıllık geliri 25 000 ile 50 000 arası geliri olan hanelerin oturduğu konut hariç sahip olduğu konut sayısı yıllık geliri 25 000 altında olan hanelere göre 1,42 kat daha fazla iken, yıllık geliri 50 000 ile 75 000 arası geliri olanlar 25 000 geliri olanlara göre 1,55 kat daha fazla ve son olarak yıllık geliri 75 000 üstünde yer alan haneler, yıllık geliri 25 000 altında olanlara göre 2,42 kat daha fazla konuta sahiptir.

Hanehalkı sorumlusu lise altı eğitim düzeyine sahip haneler, hanehalkı sorumlusu bir okul bitirmeyenlere göre 1,38 kat daha fazla konuta sahiptir.

Çizelge 3.24. PH regresyon modeli parametre tahminleri (lojit kısım)

(Lojit)	Tahmin	Standart Hata	z değeri	p değeri	$e^{\hat{\beta}_i}$
Sabit Terim	-3,859197	0,232218	-16,619	< 2e-16*	0,021085
Yaş (25-34)	0,340948	0,221995	1,536	0,125	1,406280
Yaş (35-44)	1,021872	0,216060	4,730	2,25e-06*	2,778391
Yaş (45-54)	1,526716	0,215369	7,089	1,35e-12*	4,603036
Yaş (55-64)	2,027878	0,215550	9,408	< 2e-16 *	7,597946
Yaş (65 ve üzeri)	2,259596	0,217369	10,395	< 2e-16 *	9,579218
Çalışma Durumu (İşsiz)	-0,629927	0,136886	-4,602	4,19e-06 *	0,532631
Çalışma Durumu (İstihdam)	-0,295513	0,047060	-6,280	3,40e-10*	0,744150
Cinsiyet (Kadın)	-0,255363	0,051920	-4,918	8,73e-07 *	0,774635
Geçinme Durumu(Zor)	0,704758	0,098535	7,152	8,53e-13 *	2,023357
Geçinme Durumu(Biraz zor)	-0,045447	0,085901	-0,529	0,597	0,955570
Geçinme Durumu(Biraz kolay)	0,031370	0,069113	0,454	0,650	1,031867
Geçinme Durumu(Kolay)	0,003195	0,053735	0,059	0,953	1,003200
Geçinme Durumu(Çok kolay)	-0,010396	0,040889	-0,254	0,799	0,989658
Gelir Durumu (25 000-50 000tl)	0,657026	0,051313	12,804	2e-16 *	1,929047
Gelir Durumu (50 000-75 000tl)	1,081382	0,060636	17,834	2e-16 *	2,948752
Gelir Durumu(75 000 TL üstü)	1,532173	0,069661	21,995	< 2e-16 *	4,628223
Eğitim Durumu(okuryazar değil)	-0,585340	0,111344	-5,257	1,46e-07*	0,556916
Eğitim Durumu(lise altı)	0,392894	0,081862	4,799	1,59e-06*	1,481261
Eğitim Durumu(lise)	0,448167	0,092003	4,871	1,11e-06*	1,565440
Eğitim Durumu(yükseköğrenim)	0,648942	0,092294	7,031	2,05e-12*	1,913515

Poisson hurdle regresyonun lojit kısmına göre yaş değişkeni (2. Düzey hariç), çalışma durumu, cinsiyet, geçinme durumu (2. Düzey), gelir durumu ve eğitim durumu istatistiksel olarak anlamlı bulunmuştur ($p < 0,05$).

Hanehalkı sorumlusu 25 yaş altında olanlara göre hanehalkı sorumlusu 35 ile 44 yaş arasında yer alan bir hanenin oturduğu konut hariç sahip olduğu konut sayısı 2,78 kat daha fazla iken, hanehalkı sorumlusu 45 ile 54 yaş arasında yer alan haneler 4,60 kat daha fazla, hanehalkı sorumlusu 55 ile 64 yaş arasında yer alan haneler 7,60 kat daha fazla ve son olarak hanehalkı sorumlusu 65 yaş ve üzerinde olan haneler 9,58 kat daha fazla konuta sahiptir.

Hanehalkı sorumlusu işsiz olanlara göre hanehalkı sorumlusu iş gücüne dahil olmayanların oturduğu konut hariç sahip olduğu konut sayısı 1,89 (1/0,53) kat daha fazla iken, hane sorumlusu istihdamda olanlara göre hane sorumluları iş gücüne dahil olmayanların olduğu konut sayısı 1,35 (1/0,74) kat daha fazla konuta sahiptir.

Hanehalkı sorumlusu kadın olan hanelere göre erkek olan hanelerin konut sayısı 1,30 (1/0,77) kat daha fazladır.

Geçinme durumu zor düzeyde olanların konut sayısı çok zor olan hanelere göre 2,02 kat daha fazladır.

Gelir durumu değişkeninde ise durum şu şekildedir: Yıllık geliri 25 000 ile 50 000 arası olan haneler yıllık geliri 25 000 altı olan hanelere göre 1,93 kat daha fazla, yıllık geliri 50 000 ile 75 000 arası olan haneler, yıllık geliri 25 000 altında olanlara göre 2,95 kat daha fazla ve son olarak yıllık geliri 75 000 üstünde yer alan haneler, yıllık geliri 25 000 altında olanlara göre 4,63 kat daha fazla konuta sahiptir.

Hanehalkı sorumlusu okuryazar olmayan hanelere göre hanehalkı sorumlusu bir okul bitirmeyen hanelerin oturduğu konut hariç sahip olduğu konut sayısı 1,79 (1/0,56) kat daha fazla iken, hanehalkı sorumlusu lise altı eğitim düzeyine sahip haneler, hanehalkı sorumlusu bir okul bitirmeyen hanelere göre 1,48 kat daha fazla, hanehalkı sorumlusu lise eğitim düzeyine sahip haneler, hanehalkı sorumlusu bir okul bitirmeyen hanelere göre 1,57 kat daha fazla ve son olarak hanehalkı sorumlusu eğitim durumu yükseköğretim düzeyine sahip haneler, hanehalkı sorumlusu bir okul bitirmeyen hanelere göre 1,91 kat daha fazla konuta sahiptir.

Çizelge 3.25. NBH regresyon modeli parametre tahminleri (log kısım)

(Log)	Tahmin	Standart Hata	z değeri	p değeri	$e^{\hat{\beta}_i}$
Sabit Terim	-12,657842	119,900438	-0,106	0,915924	0,00000
Yaş (25-34)	-0,430136	0,607234	-0,708	0,478726	0,65042
Yaş (35-44)	0,082046	0,586664	0,140	0,888776	1,08551
Yaş (45-54)	0,285982	0,584624	0,489	0,624719	1,33107
Yaş (55-64)	0,629169	0,584568	1,076	0,281794	1,87605
Yaş (65 ve üzeri)	1,100147	0,587268	1,873	0,061023	3,00461
Çalışma Durumu (İşsiz)	-0,310462	0,328644	-0,945	0,344825	0,73311
Çalışma Durumu (İstihdam)	-0,296610	0,085409	-3,473	0,000515 *	0,74333
Cinsiyet (Kadın)	-0,515999	0,108323	-4,764	1,90e-06 *	0,59690
Geçinme Durumu(Zor)	0,868053	0,241646	3,592	0,000328 *	2,38227
Geçinme Durumu(Biraz zor)	0,309002	0,212377	1,455	0,145677	1,36207
Geçinme Durumu(Biraz kolay)	0,263729	0,167522	1,574	0,115419	1,30178
Geçinme Durumu(Kolay)	0,055271	0,124179	0,445	0,656254	1,05683
Geçinme Durumu(Çok kolay)	-0,140008	0,089141	-1,571	0,116268	0,86935
Gelir Durumu (25 000-50 000tl)	0,448238	0,114520	3,914	9,08e-05 *	1,56555
Gelir Durumu (50 000-75 000tl)	0,595785	0,131910	4,517	6,28e-06 *	1,81445
Gelir Durumu(75 000 TL üstü)	1,103762	0,142770	7,731	1,07e-14 *	3,01549
Eğitim Durumu(okuryazar değil)	0,230781	0,239521	0,964	0,335292	1,25958
Eğitim Durumu(lise altı)	0,306833	0,170445	1,800	0,071832	1,35911
Eğitim Durumu(lise)	-0,009577	0,191844	-0,050	0,960184	0,99047
Eğitim Durumu(yükseköğrenim)	0,061457	0,187342	0,328	0,742875	1,06338

Negatif binom hurdle regresyonun log kısmına göre çalışma durumu (istihdam), cinsiyet, geçinme durumu zor ve gelir durumu istatistiksel olarak anlamlı çıkmıştır ($p < 0,05$).

Hanehalkı sorumlusu istihdamda olan hanelere göre hanehalkı sorumlusu iş gücüne dahil olmayanların oturduğu konut hariç sahip olduğu konut sayısı 1,35 (1/0,74) kat daha fazladır.

Hanehalkı sorumlusu kadın olan hanelere göre erkek olan hanelerin konut sayısı 1,67 (1/0,60) kat daha fazladır.

Geçinme durumu zor düzeyde olanların konut sayısı çok zor olan hanelere göre 2,38 kat daha fazladır.

Gelir durumu değişkeninde ise durum şu şekildedir: Yıllık geliri 25 000 ile 50 000 arası olan haneler yıllık geliri 25 000 altı olan hanelere göre 1,57 kat daha fazla, yıllık geliri 50 000 ile 75 000 arası olan haneler, yıllık geliri 25 000 altında olanlara göre 1,81 kat daha fazla ve son olarak yıllık geliri 75 000 üstünde yer alan haneler, yıllık geliri 25 000 altında olanlara göre 3,02 kat daha fazla konuta sahiptir.

Çizelge 3.26. NBH regresyon modeli parametre tahminleri (lojit kısım)

(Lojit)	Tahmin	Standart Hata	z değeri	p değeri	$e^{\hat{\beta}_i}$
Sabit Terim	-3,859197	0,232218	-16,619	< 2e-16 *	0,021085
Yaş (25-34)	0,340948	0,221995	1,536	0,125	1,406280
Yaş (35-44)	1,021872	0,216060	4,730	2,25e-06 *	2,778391
Yaş (45-54)	1,526716	0,215369	7,089	1,35e-12 *	4,603036
Yaş (55-64)	2,027878	0,215550	9,408	< 2e-16 *	7,597946
Yaş (65 ve üzeri)	2,259596	0,217369	10,395	< 2e-16 *	9,579218
Çalışma Durumu (İşsiz)	-0,629927	0,136886	-4,602	4,19e-06 *	0,532631
Çalışma Durumu (İstihdam)	-0,295513	0,047060	-6,280	3,40e-10 *	0,744150
Cinsiyet (Kadın)	-0,255363	0,051920	-4,918	8,73e-07 *	0,774635
Geçinme Durumu(Zor)	0,704758	0,098535	7,152	8,53e-13 *	2,023357
Geçinme Durumu(Biraz zor)	-0,045447	0,085901	-0,529	0,59747	0,955570
Geçinme Durumu(Biraz kolay)	0,031370	0,069113	0,454	0,65070	1,031867
Geçinme Durumu(Kolay)	0,003195	0,053735	0,059	0,95335	1,003200
Geçinme Durumu(Çok kolay)	-0,010396	0,040889	-0,254	0,79989	0,989658
Gelir Durumu (25 000-50 000tl)	0,657026	0,051313	12,804	< 2e-16 *	1,929047
Gelir Durumu (50 000-75 000tl)	1,081382	0,060636	17,834	< 2e-16 *	2,948752
Gelir Durumu(75 000 TL üstü)	1,532173	0,069661	21,995	< 2e-16 *	4,628223
Eğitim Durumu(okuryazar değil)	-0,58534	0,111344	-5,257	1,46e-07 *	0,556916
Eğitim Durumu(lise altı)	0,392894	0,081862	4,799	1,59e-06 *	1,481261
Eğitim Durumu(lise)	0,448167	0,092003	4,871	1,11e-06 *	1,565440
Eğitim Durumu(yükseköğrenim)	0,64894	0,092294	7,031	2,05e-12 *	1,913515

Negatif binom hurdle regresyonun lojit kısmına göre 25-34 yaş değişkeni, çalışma durumu, cinsiyet, geçinme durumu zor, gelir durumu ve eğitim durumu değişkeni istatistiksel olarak anlamlı çıkmıştır ($p < 0,05$).

Hanehalkı sorumlusu 25 yaş altında olanlara göre hanehalkı sorumlusu 35 ile 44 yaş arasında yer alan bir hanenin oturduğu konut hariç sahip olduğu konut sayısı 2,78 kat daha fazla iken, hanehalkı sorumlusu 45 ile 54 yaş arasında yer alan haneler 4,60 kat daha fazla, hanehalkı sorumlusu 55 ile 64 yaş arasında yer alan haneler 7,60 kat daha fazla ve son olarak hanehalkı sorumlusu 65 yaş ve üzerinde olan haneler 9,58 kat daha fazla konuta sahiptir.

Hanehalkı sorumlusu işsiz olan hanelere göre hanehalkı sorumlusu iş gücüne dahil olmayanların oturduğu konut hariç sahip olduğu konut sayısı 1,89 (1/0,53) kat daha fazladır. Hanehalkı sorumlusu iş gücüne dahil olmayanların konut sayısı, hane sorumlusu istihdamda olan hanelerden 1,35 (1/0,74) kat daha fazla konuta sahiptir.

Hanehalkı sorumlusu kadın olan hanelere göre erkek olan hanelerin konut sayısı 1,30 (1/0,77) kat daha fazladır.

Geçinme durumu zor düzeyde olanların konut sayısı çok zor olan hanelere göre 2,02 kat daha fazladır.

Gelir durumu değişkeninde ise durum şu şekildedir: Yıllık geliri 25 000 ile 50 000 arası olan haneler yıllık geliri 25 000 altı olan hanelere göre 1,93 kat daha fazla, yıllık geliri 50 000 ile 75 000 arası olan haneler, yıllık geliri 25 000 altında olanlara göre 2,95 kat daha fazla ve son olarak yıllık geliri 75 000 üstünde yer alan haneler, yıllık geliri 25 000 altında olanlara göre 4,63 kat daha fazla konuta sahiptir.

Hanehalkı sorumlusu bir okul bitirmeyen haneler, hanehalkı sorumlusu okur yazar olmayanlara göre oturduğu konut hariç sahip olduğu konut sayısı 1,79 (1/0,56) kat daha fazla iken, hanehalkı sorumlusu lise altı eğitim düzeyine sahip haneler, hanehalkı sorumlusu bir okul bitirmeyen hanelere göre 1,48 kat daha fazla, hanehalkı sorumlusu lise eğitim düzeyine sahip haneler, hanehalkı sorumlusu bir okul bitirmeyen hanelere göre 1,57 kat daha fazla ve son olarak hanehalkı sorumlusu eğitim durumu yükseköğretim düzeyine sahip haneler, hanehalkı sorumlusu bir okul bitirmeyen hanelere göre 1,91 kat daha fazla konuta sahiptir.

Yukarıda verilen regresyon modellerinden hangi modelin veriyi daha iyi temsil ettiğine AIC ve LL değerlerine bakılarak karar verilmiştir.

Çizelge 3.27. Regresyon modellere ilişkin AIC ve LL değerleri

	AIC	LL
PR	31717,01	-15837,51
NB	30181,67	-15068,84
ZIP	30740,6	-15328,3
ZINB	30101,6	-15007,8
ZTP	9249,659	-4603,829
ZTNB	8463,342	-4209,671
PH	30747,41	-15331,71
NBH	29961,1	-14937,55

Çizelge 3.27 ye göre veri setini en iyi açıklayan en küçük AIC ve en büyük LL değerine sahip olan Sıfır Kesilmiş Negatif Binom'dur.

Modellerin ikili olarak karşılaştırılmasında ise Vuong istatistiği değeri kullanılmıştır.

Çizelge 3.28. Vuong test sonuçları

	V İSTATİSTİĞİ	P DEĞERİ
ZIP- PR	10,358282	2,22e-16
ZINB- NB	2,952497	0,00157607
PR- PH	-10,813974	2,22e-16
NB-NBH	-7,042317	9,4535e-13
ZIP-PH	0,7510646	0,22631
ZINB-NBH	-5,363848	4,0734e-08

Vuong istatistiklerine bakıldığında ZIP ile PR modelleri karşılaştırıldığında ZIP modelin PR modelden daha uygun (10,36, $p<0,05$) olduğu, ZINB ile NB modelleri karşılaştırıldığında ise ZINB modelin NB modelden daha uygun (2,95, $p<0,05$) olduğu saptanmıştır.

PR ile PH modelleri karşılaştırıldığında ise PH modelin PR modelden daha uygun (-10,81, $p<0,05$) olduğu, NB ile NBH modelleri karşılaştırıldığında ise NBH modelin NB modelden daha uygun (-7,04, $p<0,05$) olduğu saptanmıştır. ZIP ile PH karşılaştırdığımızda v istatistiği 0,75 (-1,96;1,96) olduğundan karşılaştırılamamaktadır. ZINB ile NBH karşılaştırdığımızda ise NBH modelin ZINB modelden daha uygun (-5,36, $p<0,05$) olduğu ortaya çıkmıştır.

4. SONUÇ VE ÖNERİLER

Regresyon istatistiki analizlerde en çok kullanılan yöntemlerin başında gelmektedir. Birçok farklı regresyon çeşidi vardır. Bağımlı değişkenin durumuna göre en doğru regresyon modelini seçmek sağlıklı parametre tahminleri açısından önemlidir. Sayma yoluyla elde edilen değişkenlerin bulunduğu verilerde buna uygun regresyon modelleri seçilmelidir. Böyle verilerde dikkat edilmesi gereken önemli bir nokta ise veri setindeki sıfırların yoğunluğudur. Eğer çok sayıda sıfır var ise sıfır yığılmalı modeller seçilmelidir.

Bu çalışmada literatürde en çok kullanılan sayma regresyon modellerinden sekiz tanesi seçilerek gerçek bir veri seti üzerinde hangi modelin veriye daha uygun olduğu tespit edilmeye çalışılmıştır. Tezde bahsedilen sekiz model için ayrı ayrı parametre tahminleri ve model seçim kriterleri sonuçları verilmiştir. Model seçim kriterlerinden AIC, log olabilirlik ve Vuong istatistiği kullanılmıştır.

TÜİK tarafından 2018 yılında yapılmış olan GYKA veri seti bölüm ikide bahsedilen regresyon modellerine uygulanmıştır. Değişken seçiminde literatürde yapılmış olan çalışmalar dikkate alınmıştır. Bağımlı değişken olarak konut sayısı ele alınmıştır. TÜİK verilerine göre nüfusun %59 unun kendilerine ait bir konutu olduğu görülmüştür (TÜİK, 2018). Regresyon analizi yapılmadan önce konut sayısı değişkeni incelenmiş olup bu değişkenin sağa çarpık bir yapıda olduğu, yüksek oranda sıfır içerdiği ve varyansın ortalamadan büyük olduğu görülmüştür. Konut sayısının seçilen bağımsız değişkenler (yaş, cinsiyet, eğitim durumu, faal durumu, hanehalkı geliri ve geçinme durumu) ile olan ilişkisi betimleyici istatistikler bölümünde verilmiştir. Ardından da veri setine sekiz adet regresyon modeli uygulanmıştır. Burada amaç, konut sayısına etkisi olan faktörlerin belirlenmesidir. Bu sekiz model sonuçlarına bakıldığında gelir durumu değişkeninin bütün modellerde anlamlı bulunduğu görülmüştür. Diğer bütün modeller için parametrelere ait yorumlar tezin uygulama bölümünde anlatılmıştır. AIC ve LL değerlerine göre sıfır kesilmiş negatif binom regresyon modelinin diğer modellere göre daha iyi olduğu görülmüştür. Bu durum ikili karşılaştırmaların yapıldığı Vuong testi sonuçlarında da görülmektedir.

Sonuç olarak bu tez çalışmasında konut sayısına sahip olmada önemli olan faktörler sayma regresyon modelleri kullanılarak yorumlanmış olup veriye en uygun regresyon modelinin ne

olduđu araştırılmıştır. Yapılan araştırma sonucunda seçilen model seçim kriterlerine göre bu veri seti için sıfır kesilmiş regresyon modellerinin diğ er modellere kıyasla daha iyi olduđu görülmüştür. Literatürde sayma regresyon modelleri sıklıkla kullanılmaktadır. Ancak sayma regresyon modelleri kapsamında yer alan kesilmiş (truncated) regresyon modellerine ait çalışma sayısı oldukça azdır. Veride istenilen bir aralık olduğ unda tüm veriyi analiz etmek hem zaman kaybına hem de maliyet kaybına yol açacaktır. Ayrıca yanlış parametre tahminleri oluşacak ve elde edilecek sonuçlar sağlıklı olacaktır. Bu nedenle kesilmiş regresyon analizi uygulamak oldukça önemlidir. Bu çalışmanın sayma regresyon modelleri alanında çalışma yapacak kişilere sıfır kesilmiş modellerin diğ er modellerle karşılaştırılması ve parametre tahminlerinin yorumlanması açısından fayda sağlayacağı düşünülmektedir. Analizlerden elde edilen sonuçlar hanelerin konut sayısına hangi değişkenlerin etki ettiğini göstermektedir. Konut sayısının yoksullukla olan ilişkisi düşünülerek yoksulluđu azaltmak için konut sayısı bir kriter olarak seçilmek istenirse bu çalışmada anlamlı bulunan değişkenler göz önüne alınarak uygun çalışmalar yapılabilir. Bu sayede yoksulluk oranı düşürülebilir.

KAYNAKLAR

- Akaike, H. (1973). *Information theory and extension of the maximum likelihood principle*, Second International Symposium on Information Theory, Budapest: Akademiai Kiado, 267–281.
- Altun, E. (2018). A new zero-inflated regression model with application. *İstatistikçiler Dergisi: İstatistik ve Aktüerya*, 11(2), 73-80.
- Asrul A. A. M. and Naingb, N. N. (2012). Analysis death rate of age model with excess zeros using zero inflated negative binomial and negative binomial death rate: mortality aids co-infection patients, kelantan malaysia. *Procedia Economics and Finance*, 2, 275-283.
- Beaujean A. A. and Morgan G. B. (2016). Tutorial on using regression models with count outcomes using R. *Practical Assessment Research & Evaluation*, 21(2), 1-19.
- Breen, R. (1996). The Tobit model for censored data. *Regression Models*, 12-33.
- Cameron, A. C., and Trivedi, P. K. (2013). *Regression analysis of count data* (Second edition). New York: Cambridge university press, 128-132.
- Carrivick, P. J. W., Lee, A.H. and Yau, K. K. W. (2003). Zero-inflated Poisson Modeling to Evaluate Occupational Safety Interventions. *Safety Science*, 41(1), 53-63.
- Chowdhury, R. I. and Islam, M. A. (2016). Zero truncated bivariate Poisson model: marginal-conditional modeling approach with an application to traffic accident data. *Applied Mathematics*, 7, 1589-1598.
- Deniz, Ö. (2005). Poisson regresyon analizi. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, (7), 59-72.
- Famoye, F. and Singh, K. P. (2006). Zero-inflated generalized poisson regression model with an application to domestic violence data. *Journal of Data Science*, 4, 117-130.
- Flynn, M. and Francis, L. A. (2009). More flexible GLMs zero-inflated models and hybrid models. *Casualty Actuarial Society*, 2009, 148-224.
- Nava, M. (2014). *Generalized Linear Regression Models For Count Data*, Master Thesis, California State University Department of Mathematics and Statistics, Long Beach, 3-5.
- Greene, W. H., (1994). Accounting for excess zeros and sample selection in poisson and negative binomial regression models. *New York University Working Paper*, 94(10), 1-37.
- Grogger, J. T., and Carson, R. T. (1991). Models for truncated counts. *Journal of applied econometrics*, 6(3), 225-238.

- Gurmu, S. (1991). Tests for detecting overdispersion in the positive poisson regression model. *Journal of Business and Economic Statistics*, 9, 215–222.
- Hall, D. B. (2000). Zero-inflated Poisson and negative binomial regression with random effects: A case study. *Biometrics*, 56(4), 1030-1039.
- Hardin, J. W. and Hilbe, J. M. (2015). Regression models for count data from truncated distributions. *The Stata Journal*. 15, 226– 246.
- Hilbe, J. M. (2014). *Modelling Count Data*. New York: Cambridge University Press, 20-170.
- İnternet: <https://data.tuik.gov.tr/Bulten/Index?p=Gelir-ve-Yasam-Kosullari-Arastirmasi-2018-30755>. Son erişim tarihi: 25.09.2020.
- İnternet: Cran.r-project: <https://cran.r-project.org/web/packages/pscl/index.html>. Son erişim: 28/11/2020.
- Hu, M. C., Pavlicova, M. and 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.
- Ismail, N. and Zamani, H. (2013). Estimation of claim count data using negative binomial, generalized Poisson, zero-inflated negative binomial and zero-inflated generalized Poisson regression models. *In Casualty Actuarial Society E-Forum*. 41(20), 1-28.
- Karaca, A. G. (2018). *Sayma Verileri İçin Regresyon Modellerinin Karşılaştırılması Üzerine Bir Uygulama*, Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 5-11.
- Karaca, A., G. ve Olmuş, H. (2018). Sıfır değer ağırlıklı verilerin analizinde sıfır değer ağırlıklı regresyon modellerin incelenmesi. *Trakya Üniversitesi Sosyal Bilimler Dergisi*, 20(2), 105-118.
- Kaya, Y., ve Yeşilova, A. (2012). E-Posta trafiğinin sıfır değer ağırlıklı regresyon yöntemleri kullanılarak incelenmesi. *Anadolu University of Sciences & Technology-A: Applied Sciences & Engineering*, 13(1), 51-63.
- Khoshgoftaar, T. M., Gao, K. and Szabo, R. M. (2005). Comparing software fault predictions of pure and zero- inflated Poisson regression models. *International Journal of Systems Science*. 36(11), 707-715.
- Kim, D. W., Deo, R. C., Park, S. J., Lee, J. S., and Lee, W. S. (2019). Weekly heat wave death prediction model using zero-inflated regression approach. *Theoretical and Applied Climatology*, 137(1-2), 823-838.
- Klakattawi, H., Vinciotti, V., and Yu, K. (2018). A simple and adaptive dispersion regression model for count data. *Entropy*, 20(2), 142.

- Kong, M., Xu, S., Levy, S. M. and Datta, S. (2014). GEE Type inference for clustered zero-inflated negative binomial regression with application to dental caries. *Computational Statistics and Data Analysis*, 85, 54-66.
- Lambert, D. (1992). Zero-inflated Poisson regression: with an application to defects in manufacturing. *Technometrics*, 34(1), 1-14.
- Liu, X., Saat, M. R., Qin, X. and Barkan, C. P. (2013). Analysis of US freight-train derailment severity using zero-truncated negative binomial regression and quantile regression. *Accident Analysis & Prevention*, 59, 87-93.
- Martin, T. G., Wintle, B. A., Rhodes, J. R., Kuhnert, P. M., Field, S. A., Low-Choy, S. J., Tyre, A. J. and Possingham, H. P. (2005). Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters*, 8 (11), 1235-1246.
- Martinez-Espineira, R., and Amoako-Tuffour, J. (2008). Recreation demand analysis under truncation, overdispersion, and endogenous stratification: An application to Gros Morne National Park. *Journal of environmental management*, 88(4), 1320-1332.
- Martínez-Rodríguez, A. M., Conde-Sánchez, A., & Olmo-Jiménez, M. J. (2019). A new approach to truncated regression for count data. *ASTA Advances in Statistical Analysis*, 103(4), 503-526.
- Miller, J., M. (2007). *Comparing Poisson, Hurdle and Zip Model Fit Under Varying Degrees Of Skew And Zero-Inflation*, Doctoral Thesis, University of Florida, Florida, 21-23.
- Min, Y., and Agresti, A. (2005). Random effect models for repeated measures of zero-inflated count data. *Statistical modelling*, 5(1), 1-19.
- Murat, Ö. (2020). Dört kollu sinyalize kentsel kavşaklarda trafik kazalarının sıklığını etkileyen faktörlerin incelenmesi. *Teknik Dergi*, 31(3), 10033-10053.
- Nasiri, P. (2011). Estimation parameter of zero truncated mixed Poisson models. *International Journal of Mathematical Analysis*, 5(10), 465-470.
- Özmen, İ. and Famoye, F. (2007). Count regression models with an application to zoological data containing structural zeros. *Journal of Data Science*, 5, 491- 502.
- Pittman, B., Buta, E., Krishnan-Sarin, S., O'Malley, S. S., Liss, T., and Gueorguieva, R. (2020). Models for analyzing zero-inflated and overdispersed count data: an application to cigarette and marijuana use. *Nicotine and Tobacco Research*, 22(8), 1390-1398.
- Prebensen, N. K., Altin, M., and Uysal, M. (2015). Length of stay: a case of Northern Norway. *Scandinavian Journal of Hospitality and Tourism*, 15(1), 28-47.

- Puza, B. D., Johnson, H. L., O'Neill, T. J., and Barry, S. C. (2008). Bayesian truncated Poisson regression with application to Dutch illegal immigrant data. *Communications in Statistics—Simulation and Computation*, 37(8), 1565-1577.
- Rose, C. E., Martin, S. W., Wannemuehler, K. A., and Plikaytis, B. D. (2006). On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. *Journal of biopharmaceutical statistics*, 16(4), 463-481.
- Sarul, L. S., and 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), 732-743.
- Sezgin F. H. ve Deniz E. (2004). Poisson regresyon modelinde aşırı yayılım durumu ve negatif binomial regresyon analizinin Türkiye grev sayıları üzerine bir uygulaması. *Yönetim*, 15(48), 17-25.
- Sharma, A. K., and Landge, V., S. (2013). Zero inflated negative binomial for modeling heavy vehicle crash rate on indian rural highway. *International Journal of Advances in Engineering and Technology*, 5(2), 292.
- Sileshi, G. (2008). The excess-zero problem in soil animal count data and choice of appropriate models for statistical inference. *Pedobiologia*, 52, 1-17.
- Simons, J. S., Neal, D. J. and Gaher, R. M. (2006). Risk for marijuana-related problems among college students: an application of zero-inflated negative binomial regression. *The American Journal of Drug and Alcohol Abuse*, 32 (1), 41–53.
- Tüzel, S. (2011). *Hasar Sıklıkları İçin Sıfır Yığılmalı Kesikli Modeller*, Yüksek Lisans Tezi, Hacettepe Üniversitesi, Fen bilimleri Enstitüsü, Ankara, 39.
- Tüzen, M., F. and Erbaş, S. (2017). A comparison of count data models with an application to daily cigarette consumption of young persons. *Communications In Statistics Theory And Methods*, 47(23), 5825-5844.
- Ucal, M., Ş. (2006). Ekonometrik model seçim kriterleri üzerine kısa bir inceleme. *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 7(2), 41-57.
- Wang, Z., Ma, S., and Wang, C., Y. (2015). Variable selection for zero-inflated and overdispersed data with application to health care demand in Germany. *Biometrical Journal*, 57(5), 867-884.
- Winkelmann, R. (2000). *Econometric Analysis of Count Data* (Third Edition). New York: Springer, 133-135.
- Xie, H., Tao, J., McHugo, G. J. and Drake, R. E. (2013). Comparing statistical methods for analyzing skewed longitudinal count data with many zeros: An example of smoking cessation. *Journal of Substance Abuse Treatment*, 45 (1), 99–108.

- Yang, J., Xie, M. and Goh, T. N. (2009). Outlier identification and robust parameter estimation in a zero-inflated Poisson model. *Journal of Applied Statistics*, 38(2), 421-430.
- Yıldırım, G. (2019). *Poisson Ve Negatif Binom Regresyon Modelleri*, Yüksek Lisans Tezi, Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Adana, 31-69.
- Zuur, A., F., Ieno, I., N., Walker, N., J., Saveliev, A., A., and Smith, G., M. (2009). *Mixed Effects Models and Extensions in Ecology with R*. New York: Springer, 206-209.

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : ŞENTÜRK, Onur
 Uyruğu : T.C.
 Doğum tarihi ve yeri : 09.09.1989, Ankara
 Medeni hali : Bekar
 e-posta : senturkonur@yandex.com



Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Yüksek lisans	Gazi Üniversitesi / İstatistik	Devam ediyor
Lisans	Ankara Üniversitesi / İstatistik	2012
Lise	İncirli Yabancı Dil Ağırlıklı Lise	2007

İş Deneyimi

Yıl	Yer	Görev
2015-Halen	Türkiye İstatistik Kurumu	Uzman
2014-2015	Türkiye İstatistik Kurumu	İstatistikçi

Yabancı Dil

İngilizce

Yayınlar

Şentürk, O., ve Olmuş, H. (2020) *Sıfır Yığılmalı Regresyon Modelleri Üzerine Bir Uygulama*. 3. Uygulamalı Bilimler Kongresi Özet Bildiri Kitabı. Ordu, 47.

Hobiler

Spor, seyahat, kitap



GAZİ GELECEKTİR..