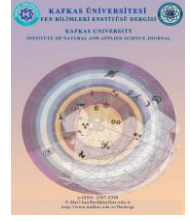




Kafkas Üniversitesi Fen Bilimleri Enstitüsü Dergisi Institute of Natural and Applied Science Journal

Dergi ana sayfası/ Journal home page: <https://dergipark.org.tr/tr/pub/kujs>



E-ISSN: 2587-2389

Sayma Verisi Modelleri Üzerine Bir Karşılaştırma: Konut Sayısına Etki Eden Faktörler Türkiye Örneği

Onur ŞENTÜRK^{1*}, Hülya OLMUŞ¹

¹ Gazi Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik, Ankara, Türkiye

(İlk Gönderim / Received: 10. 04. 2024, Kabul / Accepted: 11. 09. 2024, Online Yayın / Published Online: 20. 11. 2024)

Anahtar Kelimeler:

Sayma Dayalı Regresyon Modelleri,
Sıfır Yığılmalı Regresyon Modelleri,
Sıfır Kesilmiş Regresyon Modelleri.

Özet: Yoksulluk çok boyutlu bir kavramdır. Yoksulluk göstergelerden bir tanesi hanenin sahip olduğu konut sayısıdır. Bu çalışmada, hane halkı bireyinin sahip olduğu konut sayısına etki eden faktörleri belirlemek için sayıma dayalı regresyon modelleri kullanılmıştır. Ayrıca, veriye en iyi uyum sağlayan regresyon modeli araştırılmıştır. Sayıma dayalı regresyon modellerinden en sık kullanılanlar klasik sayıma dayalı regresyon modelleri ve sıfır yığılmalı sayıma dayalı regresyon modelleridir. Ancak literatürde önerilmiş diğer bir regresyon modeli sıfır kesilmiş sayıma dayalı regresyon modelleridir. Bu modeller tüm veriyi analiz etmenin yarattığı zaman ve maliyet kaybının önüne geçmektedir. Bu nedenle, bu modeller sayıma dayalı verilerin olduğu durumlarda modellemede kullanılmak için iyi bir seçenektir. Çeşitli sayıma dayalı regresyon modelleri uygulamasını TÜİK'in yaptığı Gelir ve Yaşam Koşulları Araştırması veri setine uygulanmıştır. Çalışmada ele alınan modellerin performans değerlendirilmesi yapılmıştır. Bu değerlendirmeler için Akaike Bilgi Kriteri ve Log olabilirlik değeri kullanılmıştır. Sonuç olarak, sıfır kesilmiş negatif binom regresyon modeli gerçek veri setine en iyi uyum gösteren modeldir.

A Comparison on Count Data Models: Factors Affecting the Number of Houses Example of Türkiye

Keywords:

Count Data Regression Models,
Zero Inflated Regression Models,
Zero Truncated Regression
Models.

Abstract: Poverty is a multidimensional concept. One of the indicators of poverty is the number of houses owned by the household. In this study, counting regression models were used to determine the factors affecting the number of houses owned by household members. Moreover, it was investigated which regression model would best fit the data. The most commonly used count regression models are classical count regression models (such as Poisson and negative binomial) and zero truncated regression models. However, another count regression model proposed in the literature is zero-truncated count regression models. These models prevent the loss of time and cost caused by analyzing all the data when there is a desired range in the data. Therefore, these models are a good option to use in modeling situations where count data is available. Various count regression models were applied to the Income and Living Conditions Survey data set by TURKSTAT. The performance evaluation of the models considered in the study was made. Akaike Information Criterion and Log Likelihood value were used to compare the suitability of the models. As a result, the zero-truncated negative binomial regression model is the model that best fits the real data set.

*İlgiliyazar: senturkonur@yandex.com
DOI: [10.58688/kujs.1467396](https://doi.org/10.58688/kujs.1467396)

1. GİRİŞ

Regresyon analizi istatistiksel analizler sırasında en sık kullanılan yöntemlerdendir. Çok sayıda regresyon türü mevcuttur. Doğru regresyon modeli seçiminde yanıt değişkenin durumu önem arz etmektedir. Çünkü doğru regresyon modeli ile sağlıklı parametre tahminleri elde edilebilir. Diğer veri türlerinde olduğu gibi herhangi bir sayıya dayalı olarak elde edilen değişkenlerin bulunduğu verilerde de bu duruma uygun regresyon modellerinin seçilmesi önemlidir. Sayıya dayalı verilerde göz ardı edilmemesi gereken bir diğer nokta ise veri setinde var olan sıfırların yoğunluğudur. Veride fazla sayıda sıfır olması durumunda sıfır yığılmalı modellerin tercih edilmesi gerekmektedir.

Yapılan bu çalışmada sayıya dayalı verilerin analizinde kullanılan bazı yöntemler ele alınmıştır. Bu veri setlerinde, sıfır değerinde yığılma söz konusudur. Bu tarz verilerin analizinde klasik doğrusal regresyon analizi varsayımları sağlanamamaktadır. Doğrusal regresyon yerine sayıya dayalı veriyi modellemek için literatürde farklı regresyon modelleri vardır. Bu modeller; Poisson Regresyon, Negatif Binom Regresyon, Sıfır Yığılmalı Poisson Regresyon, Sıfır Yığılmalı Negatif Binom Regresyon, Sıfır Kesilmiş Poisson Regresyon ve Sıfır Kesilmiş Negatif Binom Regresyon modelleridir.

Son yıllarda literatürde konu ile ilgili olarak yapılan çalışmalar incelenmiştir. Bu çalışmalardan bazılarında özet bilgiler sunulmuştur. Yeşilova, Kaydan ve Kaya (2010), sıfır yığılmalı veri setlerinde aşırı yayılım problemini ele alarak sıfır yığılmalı negatif binom regresyon modelinin, sıfır yığılmalı Poisson modeline kıyasla daha iyi performans gösterdiğini bulmuşlardır. Bu model, bitki koruma çalışmalarında böcek yumurtası verilerinde sıklıkla karşılaşılan fazla sıfır gözlemlerinin analizinde etkili bir yaklaşım sunmaktadır. Türkiye'deki gençlerin günlük sigara tüketimlerini etkileyen faktörleri belirlemek için Tüzen ve Erbaş (2017), sayıya dayalı farklı regresyon modelleri kullanmıştır. Bu çalışmada, sıfır yığılmalı negatif iki terimli ve negatif Binom Hurdle regresyon modellerinin veriye en iyi uyan modeller olduğuna karar verilmiştir. Altun (2018), aşırı yayılım ve sıfır yığılmalı veri kümelerinin modellenmesi için sıfır yığılmalı Poisson-Lindley regresyon modelini önermiştir. Yapılan uygulamalar, bu modelin sıfır yığılmalı ve aşırı yayılım gösteren veri setlerinde, sıfır yığılmalı Poisson regresyon modeline göre daha iyi uyum sağladığını göstermektedir. Kim ve arkadaşları (2019), sıfır yığılmalı regresyon modellerini aşırı sıcaktan ölen bireylere ilişkin ölüm sayısını tahmin etmek için kullanmışlardır. Çalışma sonucunda sıfır yığılmalı Poisson regresyonun, sıfır değeri alan gözlemler durumunda iyi bir istatistiksel yaklaşım olduğunu göstermişlerdir. Alwani ve ark. (2021), sayıya dayalı regresyon modellerini incelemişler ve sıfır kesilmiş modelleri kullanarak Malezya'da hava kirliliği ve iklim değişkenleri gibi faktörler arasındaki ilişkiyi incelemişlerdir. Çalışma sonucunda, sıfır kesilmiş negatif binom regresyon modelinin en iyi model olduğunu vurgulamışlardır. Worku ve ark. (2022), 2019 yılında Etiyopya'da anne başına doğan çocuk sayısını tahmin etmek için sıfır kesilmiş regresyon modellerini kullanmışlardır. Bu veri için sıfır kesilmiş negatif binom modelin sıfır kesilmiş Poisson modelden daha iyi

olduğunu ortaya koymuşlardır. Ayrıca doğum sayısına etki eden faktörleri de incelemişler ve eğitim düzeyinin en etkili değişken olduğunu ortaya koymuşlardır. Lawal (2022), NHIS verilerine en uygun sıfır kesilmiş regresyon modelini araştırmıştır. Bu amaçla sıfır kesilmiş genelleştirilmiş Poisson, sıfır kesilmiş Poisson ve sıfır kesilmiş negatif binom regresyon modellerinin veriye uyumunu incelemiştir. Sıfır kesilmiş Poisson modelin daha iyi uyum sağladığı sonucuna ulaşmıştır. Gevrekçi ve ark. (2022), Holstein sığırlarında ölü doğum verilerini modellemek için Poisson, negatif binom, sıfır-yığılmalı Poisson, sıfır-yığılmalı negatif binom, Poisson-logit hurdle ve negatif binom-logit hurdle regresyon modellerini karşılaştırmıştır. Modellerin performansı değerlendirilmiş ve negatif binom-logit hurdle modelinin en iyi uyumu sağladığı belirlenmiştir. Cinsiyet, doğurganlık durumu ve sürü-yıl-mevsim faktörlerinin ölü doğum üzerindeki etkileri önemli bulunmuş, erkeklerde ölü doğum oranının daha yüksek olduğu ve doğurganlık arttıkça azaldığı gözlenmiştir. Durmuş ve İşçi Güneri (2020), aşırı yayılım içeren veri setlerinde Poisson regresyon modelinin yetersiz kaldığını ve bu durumlarda genelleştirilmiş Poisson regresyon modelinin daha uygun olduğunu göstermişlerdir. 1984-2017 yılları arasında Türkiye'deki grev sayıları üzerine yapılan çalışmada, genelleştirilmiş Poisson modelinin daha iyi uyum sağladığı belirlenmiştir. İşçi Güneri ve Durmuş (2021) aşırı yayılım içeren sayısal verilerin analizinde Poisson regresyon modelinin yetersiz kaldığını ve bu durumlarda genelleştirilmiş Poisson, genelleştirilmiş negatif binom ve sıfır-yığılmalı yayılım modellerinin daha uygun olduğunu belirtmektedir. Çalışmalarında, çeşitli genelleştirilmiş regresyon modellerinin aşırı sıfır içeren veri setleri üzerindeki performanslarını karşılaştırmışlardır.

Yapılan bu çalışmada Poisson Regresyon, Negatif Binom Regresyon, Sıfır Yığılmalı Poisson Regresyon, Sıfır Yığılmalı Negatif Binom Regresyon, Sıfır Kesilmiş Poisson Regresyon, Sıfır Kesilmiş Negatif Binom Regresyon modelleri ele alınmıştır. 2018 yılına ait Türkiye İstatistik Kurumu (TÜİK) Gelir ve Yaşam Koşulları Araştırması (GYKA) gerçek verisi kullanılarak analizler gerçekleştirilmiştir. GYKA, Avrupa Birliği İstatistik Ofisi (Eurostat) tarafından koordine edilen ve TÜİK tarafından yürütülen bir araştırmadır. GYKA, 2006 yılından itibaren her yıl düzenli olarak yapılmaktadır. Araştırma, hanhalklarının gelir düzeyleri, yaşam koşulları ve sosyal içerme durumları hakkında bilgi toplamak amacıyla yapılmaktadır. Araştırmanın ana hedefi, hanhalklarının ekonomik ve sosyal durumunu değerlendirerek yoksulluk ve sosyal dışlanma ile ilgili politikaların geliştirilmesine katkıda bulunmaktır. Söz konusu bu modellerin veriye uygunluğunu göstermek için Akaike Bilgi Kriteri ve Log Olabilirlik değerleri kullanılmıştır. En uygun bulunan model için parametre tahminleri ve bu tahminlere ilişkin yorumlar yapılmıştır.

2. MATERYAL VE YÖNTEM

Sayıya dayalı veri, negatif olmayan tamsayı değeri alan, sayıdan kaynaklanan verilerdir. Bu verilere başta sağlık olmak üzere aktüerya, eğitim, dış hekimliği ve çevre bilimi gibi farklı alanlarda sıklıkla rastlamak mümkündür.

Bu veriler ile gerçek dünya problemlerinde sıklıkla karşılaşıldığından modelleme çalışmaları sırasında da sayıya

dayalı regresyon modellerinin kullanımı yaygındır. En çok tercih edilen modellerin başında Poisson, negatif binom regresyon ile bunların sıfır yığılmalı halleri ve sıfır kesilmiş halleri gelmektedir. Bu bölümde kısaca bu yöntemler tanıtılacaktır.

2.1. Poisson Regrasyon (PR)

PR modeli, sayıma dayalı verinin modellenmesinde sıklıkla kullanılmaktadır ve bu alanda ortaya konulan ilk modellerin başında gelmektedir. Bu nedenle, farklı sayıma dayalı modellerin temelini oluşturmaktadır.

PR, belli bir zaman aralığında gerçekleşen olaylar ile açıklayıcı değişkenler arasında bir bağlantı kurmak amacıyla kullanılır. Bu bağlantı, log bağlantı fonksiyonu kullanılarak yapıldığı için Loglineer model olarak adlandırılmaktadır (Agresti, 2002).

Yanıt değişkeni Poisson dağılıma sahip olduğunda, veriyi analiz etmek için PR modeli kullanılabilir. Modelde Y_i , i . gözlem için yanıt değişkeni olsun. Poisson dağılımının λ ortalamasına sahip olduğu varsayalım. Poisson dağılımına ilişkin olasılık fonksiyonu Eş. 1. de verilmiştir.

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

Beklenen değer ve varyansı birbirine eşittir.

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

Y_i 'nin beklenen değerinin negatif değerler almamasını sağlamak için, beklenen değer ile açıklayıcı değişkenler arasındaki ilişkiyi gösteren bağlantı fonksiyonu (Cameron ve Trivedi, 2013):

$$\log(\lambda_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m \text{ veya}$$

$$\lambda_i = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m) = e^{x_i' \beta} \text{ şeklinde gösterilir.}$$

Burada $\beta_0, \beta_1, \dots, \beta_m$ bilinmeyen parametreleri temsil eder.

2.2. Negatif Binom Regrasyon (NBR) Modeli

NBR modeli, aşırı yayılım gösteren verilerin modellenmesinde sıklıkla kullanılır. Aşırı yayılım var olması halinde PR modelinin tercih edilmesi durumunda yanlış parametre tahminlerine ve hatalı sonuçlara neden olacaktır (Hilbe, 2014).

λ ortalama ve α yayılım parametresini göstermek üzere bu modelin olasılık fonksiyonu Eş. 2 de verilmiştir.

$$P(Y_i = y_i) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1) \Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i} \right)^{\alpha^{-1}} \left(\frac{\lambda_i}{\alpha^{-1} + \lambda_i} \right)^{y_i}, \alpha > 0 \quad (2)$$

(2) nolu denklemde; α , aşırı yayılım parametresini gösterir. Aşırı yayılım parametresi sıfır olursa model, Poisson modele yakınsar. Ayrıca, λ ve α 'nın sıfırdan büyük olması aşırı yayılımın bir göstergesidir. Bu model için beklenen değeri ve varyansı aşağıdaki gibidir.

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

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

NBR'de, PR için yazılan eşitliklerde verilen log bağlantı fonksiyonu kullanılarak beklenen değer ile açıklayıcı değişkenler arasındaki ilişki ifade edilebilir.

2.3. Sıfır Yığılmalı Poisson Regrasyon (ZIP)

ZIP modelinde yanıt değişkeni iki farklı veri grubundan oluşur. İlk grup, yalnızca sıfır değerlerini alan Poisson dağılımlı verilerden, ikinci grup ise daima sıfır değerini içermektedir. Bu modelde, yanıt değişkeninin hangi veri grubuna dahil olduğunu belirlemek için lojit fonksiyonu kullanılır. Poisson dağılımına uygun bulunan grup PR ile modellenir (Min ve Agresti, 2005). ZIP modeline ilişkin olasılık fonksiyonu Eş. 3 de verilmiştir.

$$P(Y = y_i) = f(x) = \begin{cases} w_i + (1 - w_i) \exp(-\lambda_i), & y_i = 0 \\ (1 - w_i) \frac{\lambda_i^{y_i}}{y_i!} \exp(-\lambda_i), & y_i > 0 \end{cases} \quad (3)$$

Burada $0 \leq w_i \leq 1$ dir. w_i , verideki sıfır yığılma oranını ifade eder ve $w_i = 0$ ise ZIP model PR'ye döner. $w_i > 0$ olması sıfırda var olan yığılmanın göstergesidir.

Bu modelde, log kısmı için $\log(\lambda) = B\beta$ denklemi, lojit kısmı için ise denklemler aşağıdaki gibidir. Bu denklemlerde B ve G ortak değişken matrisleridir.

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

$$w = \frac{e^{G_\gamma}}{1 + e^{G_\gamma}}$$

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

2.4. Sıfır Yığılmalı Negatif Binom Regrasyon (ZINB) Modeli

ZIP modeline alternatif olarak geliştirilen ZINB modeli, aşırı yayılım yanında sıfır değerlerinde de aşırı yığılmanın olduğu durumlarda verileri modellemek için önerilmiştir (Greene, 1994). ZINB modeli için olasılık fonksiyonu Eş. 4 de verilmiştir:

$$P(Y = y_i) = \begin{cases} p_i + (1 - p_i) \left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i} \right)^{\alpha^{-1}}, & y_i = 0 \\ (1 - p_i) \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1) \Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i} \right)^{\alpha^{-1}} \\ \left(\frac{\lambda_i}{\alpha^{-1} + \lambda_i} \right)^{y_i}, & y_i > 0 \end{cases} \quad (4)$$

2.5. Sıfır Kesilmiş Poisson Regrasyon (ZTPR) Modeli

ZTPR modelinin, klasik Poisson modelinden farkı yanıt değişkeninin sıfır değerini almamasıdır. ZTPR modelinin olasılık fonksiyonu Eş. 5 de verilmiştir.

$$P(Y_i = j | Y_i > 0) = \frac{\exp(-\lambda_i) \lambda_i^j}{j!} [1 - F_p(0)]^{-1} = \frac{\lambda_i^j}{(\exp(\lambda_i) - 1) j!} \quad (5)$$

Bu model denkleminde, j yalnızca 0'dan büyük pozitif tamsayı değerlerini alır (Grogger ve Carson, 1991). $F_p(0)$, Poisson dağılımının dağılım fonksiyonunda 0 değerini almasını ifade eder.

2.6. Sıfır Kesilmiş Negatif Binom Regresyon (ZTNB) Modeli

ZTNB modeli, kesilmiş sayıma dayalı verilerde pozitif değerler alan yanıt değişkeninin aşırı yayılım göstermesi durumunda yansız tahminler elde etmek için kullanılır. Bu modele ilişkin olasılık fonksiyonu Eş. 6 da verilmiştir.

$$P(Y_i = j | Y_i > 0) = \frac{\Gamma(j + \frac{1}{\alpha})}{\Gamma(j+1)\Gamma(\frac{1}{\alpha})} (\alpha\lambda_i)^j [1 + \alpha\lambda_i]^{-(j+1/\alpha)} [1 - F_{NB}(0)]^{-1} \quad (6)$$

Bu model denkleminde, j sıfırdan büyük bir tamsayıyı göstermek üzere, Γ gama fonksiyonunu, α aşırı yayılım parametresini, λ ortalamayı ve $F_{NB}(0)$ ise negatif binom dağılımının dağılım fonksiyonunda 0 değerini almasını ifade eder.

2.7. Model Seçimi

Bu çalışmada en uygun olan modeli belirlemek için, çalışmada kullanılan sayıma dayalı regresyon modellerine ait Akaike Bilgi Kriteri (AIC) ve log olabirlik değerleri kullanılmıştır.

$$AIC = -2L + 2k = -2(L - k) \quad (7)$$

(7) nolu denklemde,

L : log olabirliği

k : tahmin edici veya parametre sayısını

n : modeldeki gözlem sayısını (Akaike, 1973; Hilbe, 2014) göstermektedir. En küçük AIC değerine sahip modelin tercih edilmesi gerekmektedir.

Log olabirlik (LL) değeri için ise durum AIC de olanın tam tersidir. Yani burada en büyük LL değeri alan modelin en iyi model olduğu sonucuna ulaşılır.

3. BULGULAR VE TARTIŞMA

Bu çalışmada Türkiye İstatistik Kurumu'nun (TÜİK) Gelir ve Yaşam Koşullar Araştırması (GYKA) 2018 yılı mikro veri seti kullanılmıştır. GYKA' da gelir ve yoksulluk ile diğer yaşam koşullarına ilişkin çeşitli göstergeler hesaplanmaktadır. Bu göstergeler ile konut, gayrimenkul sahipliği, eğitim, iş gücü durumu ve gelir durumu gibi çeşitli kategorilerde bilgiler derlenmektedir.

Bu çalışmada, hanehalkının oturduğu konut dışında sahip olduğu konut adedi yanıt değişkeni olarak ele alınmıştır. Amaç, hanehalkının oturduğu konut dışında sahip olduğu toplam konut sayısına etki eden diğer faktörleri araştırmaktır. Bunun yanında, açıklayıcı değişken olarak hanehalkı sorumlusuna ait yaş, eğitim durumu, cinsiyet, hanenin yıllık geliri, geçinme durumu ve çalışma durumu alınmıştır.

Yapılan analizler R Studio programı ile yapılmış ve Political Science Computational Laboratory (pscl) ve Modern Applied Statistics with S (mass) paketleri modelleme aşamasında kullanılmıştır. Bu çalışmada verilerin aşırı yayılım gösterdiği analiz edilerek gösterilmiştir. Poisson regresyon modeli için Pearson Ki-kare istatistiği hesaplanmıştır. Eğer Pearson Ki-kare istatistiğinin serbestlik derecesine bölünmesiyle elde edilen değer 1'den büyükse, aşırı yayılım olasılığı yüksektir. $1.47 > 1$ olduğu için veri setinde aşırı yayılım olasılığı yüksektir. Negatif Binom regresyon modelinin AIC değerlerinin de daha düşük olduğu ve aşırı yayılımın varlığına işaret ettiği görülmektedir.

Sayıma dayalı yanıt değişken olan "konut sayısı" na ilişkin hanelere göre dağılımı Tablo 3.1'de verilmiştir.

Tablo 3.1. Konut sayısı dağılım tablosu.

	Konut sayısı		
	Sayı	Yüzde (%)	
	0	19100	79.40
	1	3636	15.10
	2	840	3.50
Aldığı	3	261	1.10
Değer	4	115	0.50
	5 ve üstü	101	0.40
	Toplam	24053	100
	değer		

Tablo 3.1'e göre oturan konut haricinde başka bir konuta sahip olmayan hanelerin oranı %79.40 iken, oturduğu konut dışında yalnızca 1 konutu mevcut olan hanelerin oranı da %15.10'a karşılık gelmektedir.

Tablo 3.2'de sayıma dayalı regresyon modellerinden veriye uygun olan en iyi modeli belirlemek için AIC ve LL değerleri elde edilmiştir. Bu çizelgeye göre veri setini en iyi açıklayan en küçük AIC ve en büyük LL değerine sahip olan Sıfır Kesilmiş Negatif Binom Regresyon modelidir.

Tablo 3.2. AIC ve LL değerleri.

	AIC	LL
PR	31717.010	-15837.510
NB	30181.670	-15068.840
ZIP	30740.600	-15328.300
ZINB	30101.600	-15007.800
ZTNB	8463.342	-4209.671

Tablo 3.3 incelendiğinde, çalışma durumu değişkeninin istihdam kategorisi, cinsiyet değişkeninin kadın kategorisi, zor geçinme durumu ve gelir durumunun tüm kategorileri %95 güven düzeyinde istatistiksel olarak anlamlı bulunmuştur. Bu çizelgeye göre aşağıdaki önemli yorumlamalar yapılabilir:

Hanehalkı sorumlusunun cinsiyeti erkek olan hanelerin, hanehalkı sorumlusunun cinsiyeti kadın olan hanelere göre oturduğu konut dışında sahip olduğu konut sayısının 1.68 kat daha fazla olduğu gözlenmiştir.

Tablo 3.3. ZTNB modeline ait parametre tahminleri.

Değişkenler	Tahmin Standart Hata z değeri p değeri $e^{\hat{\beta}_i}$					
	Sabit Terim	-10.804	47.318	-0.228	0.819	0.00002
Yaş	25-34	-0.428	0.607	-0.706	0.480	0.651
	35-44	0.0832	0.586	0.142	0.887	1.086
	45-54	0.287	0.584	0.491	0.623	1.332
	55-64	0.630	0.584	1.078	0.280	1.878
	65 >	1.101	0.587	1.875	0.060	3.008
Çalışma Durumu	İstihdam	-0.296	0.085	-3.473	0.000515 *	0.743
	Cinsiyet	Kadın	-0.516	0.108	-4.764	1.90e-06 *
Geçinme Durumu	Zor	0.868	0.241	3.593	0.000327 *	2.382
	Biraz zor	0.308	0.212	1.454	0.145	1.361
	Biraz kolay	0.263	0.167	1.575	0.115	1.301
	Kolay	0.055	0.124	0.445	0.656	1.056
	Çok kolay	-0.139	0.089	-1.570	0.116	0.869
Gelir Durumu	25 000-50 000 TL	0.448	0.114	3.914	9.07e-05 *	1.565
	50 000-75 000 TL	0.595	0.131	4.517	6.28e-06 *	1.814
	75 000 TL >	1.103	0.142	7.731	1.07e-14 *	3.015
Eğitim Durumu	okuryazar değil	0.230	0.239	0.963	0.335	1.259
	lise altı	0.306	0.170	1.800	0.071	1.359
	lise	-0.009	0.191	-0.050	0.959	0.990
	yükseköğrenim	0.061	0.187	0.328	0.743	1.063

Geçinme durumu zor olan hanelerin, geçinme durumu çok zor olan hanelere göre oturduğu konut dışında sahip olduğu konut sayısının 2.38 kat daha fazla olduğu ifade edilir.

Yıllık geliri 75000 TL üstünde olanların, yıllık geliri 25000 altında olanlara göre hane halkının oturduğu konut dışında sahip olduğu konut sayısının 3.02 kat daha fazla olduğu, yıllık geliri 50000 TL – 75000 TL ve yıllık geliri 25000 TL – 50000 TL arasında olan hanelerin, yıllık geliri 25000 altında olanlara göre oturduğu konut dışında sahip olduğu konut sayısının sırasıyla 1,81 kat ve 1,56 kat daha fazla olduğu görülmektedir.

4. SONUÇ

Bu makalede, sıklıkla tercih edilen sayıma dayalı regresyon modelleri ele alınmıştır. Gerçek bir veri seti kullanılarak hangi modelin veriye daha uygun olduğu belirlenmeye çalışılmıştır. Veri setine en uygun model için parametre tahmin sonuçları verilmiştir. Ayrıca kullanılan tüm modellere ait model seçim kriterlerinin sonuçlarına yer verilmiştir.

2018 yılı verilerine göre, Türkiye nüfusunun %59'unun kendilerine ait bir konutu olduğu elde edilmiştir (TÜİK, 2018). Bu nedenle, sayıma dayalı yanıt değişkeni konut sayısı olarak alınmıştır. Bu değişken sağa çarpık bir dağılıma sahip olup, fazla sayıda sıfır içermektedir. Yanıt değişkeni olan konut sayısının yaş, cinsiyet, eğitim durumu, çalışma durumu, hanehalkı geliri ve geçinme durumu ile ilişkisi incelenmiştir. Yapılan bu çalışmada amaç, konut sayısına etki eden faktörlerin belirlenmesidir. Yapılan analizler sonucunda, sıfır kesilmiş negatif binom regresyon modelinin veriyi temsil eden en iyi model olduğu belirlenmiştir.

Literatürde sayıma dayalı regresyon modelleri konusu geniş yer bulmaktadır. Ancak kesilmiş (truncated) regresyon modelleri üzerine yapılan çalışma sayısı diğer regresyon

modellerine göre daha azdır. Bu modele ilişkin çalışmalar genişletilebilir.

5. KAYNAKLAR

- Agresti A. (2002). *Categorical data analysis* (Second Edition), New Jersey: Wiley & Sons Incorporation.
- 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.
- Alwani, Z. Z., Ibrahim, A. I. N., Yunus, R. M., & Yusof, F. (2021). Application of zero-truncated count data regression models to air-pollution disease. In *Journal of Physics: Conference Series* (Vol. 1988, No. 1, p. 012096). IOP Publishing.
- Cameron, A. C., and Trivedi, P. K. (2013). *Regression analysis of count data* (Second edition). New York: Cambridge university press, 128-132.
- Durmuş, B., & Güneri, Ö. İ. (2020). An application of the generalized poisson model for over dispersion data on the number of strikes between 1984 and 2017. *Alphanumeric Journal*, 8(2), 249-260.
- Gevrekci, Y., Guneri, O. I., Takma, C., & Yesilova, A. (2022). Comparison of different count models for investigation of some environmental factors affecting

stillbirth in holsteins. *Indian Journal of Animal Research*, 56(9), 1158-1163.

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.

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-Arastirmasi2018-30755>. Son erişim tarihi: 25.09.2020.

İşçi Güneri, Ö., & Durmuş, B. (2021). Models for Overdispersion Count Data with Generalized Distribution: An Application to Parasites Intensity. *Journal of New Theory*(35), 48-61. <https://doi.org/10.53570/jnt.902066>

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.

Lawal, B. H. (2022). Zero-Truncated Models applied to the Nigerian National Health Insurance Data. *BENIN JOURNAL OF STATISTICS* , Vol. 5, pp. 1– 20.

Min, Y., and Agresti, A. (2005). Random effect models for repeated measures of zeroinflated count data. *Statistical modelling*, 5(1), 1-19.

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.

Worku, G., Tadesse, G., Arega, A., & Tesfaw, D. (2022). Determinants of the number of children born in Ethiopia, evidenced from 2019 miniEDHS: Using zero-truncated count regression models.

Yesilova, A., Kaydan, M. B., & Kaya, Y. (2010). MODELING INSECT-EGG DATA WITH EXCESS ZEROS USING ZERO-INFLATED REGRESSION MODELS. *Hacettepe Journal of Mathematics and Statistics*, 39(2), 273-282.