

**SIFIR DEĞER AĞIRLIKLIL VERİLERİN ANALİZİNDE SIFIR DEĞER
AĞIRLIKLIL REGRESYON MODELLERİN İNCELENMESİ**

*INVESTIGATING OF ZERO-INFLATED REGRESSION MODELS IN ANALYSIS
ZERO-INFLATED DATA*

Aslı Gizem KARACA* Hülya OLMUŞ**

*Geliş Tarihi: 02.08.2017
(Received)*

*Kabul Tarihi: 25.09.2018
(Accepted)*

ÖZ: Aşırı sıfır ve / veya aşırı yayılıma sahip sayma verilerinin modellenmesinde Poisson Regresyon, Negatif Binom Regresyon, Sıfır Değer Ağırlıklı Poisson Regresyon ve Sıfır Değer Ağırlıklı Negatif Binom Regresyon yöntemleri kullanılmaktadır. Bu çalışmada, herhangi bir hizmet sektöründe çalışan personelin müşterilerden şikâyet alma sayısını cinsiyet, yaş, eğitim ve tecrübe değişkenlerinin etkilediği düşünülmektedir. Sıfır değer ağırlıklı yöntemleri değerlendirmek için elde edilen sayma verisi analiz edilmiştir. Ayrıca, Akaike Bilgi Kriteri regresyon modellerini değerlendirmek için kullanılmıştır. Uygulamada 2016 yılının her ayı için hangi modelin uygun olduğu tespit edilmiş ve bu modellere ait parametre tahminlerine ilişkin yorumlamalar yapılmıştır.¹

Anahtar Kelimeler: sayma verisi, aşırı sıfır, sıfır değer ağırlıklı veri, sıfır değer ağırlıklı modeller

ABSTRACT: Poisson Regression, Negative Binomial Regression, Zero-Inflated Poisson Regression and Zero-Inflated Negative Binomial Regression methods are user to modelling count data with excess zeros and / or over dispersion. In this study, it is considered that the number of complaints received from customers in any service sector is influenced by gender, age, education and experience variables. These data are the analyzed to assess zero-inflated procedures. Also, Akaike Information Criterion is used to evaluate the regression models. In practice, it is determined which model is suitable for each month of the year 2016 and interpretations are made about parameter estimations of appropriate models.

Keywords: count data, excess zeros, zero-inflated data, zero-inflated regression models

1. GİRİŞ

Sayma dayalı olarak elde edilen, gerçekleşme sayısının negatif olmayan tamsayı değerlerinden oluştuğu veriler sayma verisi olarak adlandırılır. Sayma verileri biyoistatistik, iktisat, demografi, eğitim bilimleri, sosyoloji ve aktüeryal bilimlerde sıkça kullanılmaktadır.

* Gazi Üniversitesi, asligizem@gmail.com

** Doç. Dr, Gazi Üniversitesi, hulya@gazi.edu.tr

¹Bu çalışma ilk yazarın, ikinci yazar danışmanlığında hazırladığı yüksek lisans tezinden üretilmiştir.

Sayma verisi kullanılarak elde edilen değişkenlerin bulunduğu veri setlerinde kimi zaman sıfır değerleri ağırlık göstermektedir. Ortaya çıkan sıfır değerler elde edilmiş durumlarına göre yapısal sıfır ve örnekleme sıfırı 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 değerlere de *örnekleme sıfırı* denir. Örneğin; araç sahibi olmayan bir kişinin belirlenen süre içinde belli bir akaryakıt istasyonundan aldığı akaryakıt litre miktarı yapısal sıfır; araç sahibi olan kişinin belirlenen süre içinde belli bir akaryakıt istasyonundan akaryakıt almaması durumu örnekleme sıfırına örnektir.

Sıfır değerlerin ağırlık gösterdiği durumlar ortaya çıktığında veri setinin dağılımı sağa çarpıklık göstermekte ve bu durum doğrusal regresyon yöntemi için gerekli olan normal dağılım varsayımının sağlanamamasına yol açmaktadır. Bu tür durumlarda elde edilen sıfır değerleri için dönüşüm yöntemleri uygulamak veya sıfır değerleri göz ardı etmek hatalı, yanlış sonuçlar ortaya çıkmasına neden olmaktadır. Ortaya çıkan sıfır değerlerini göz önünde bulundurarak, hatalı sonuçlar ortaya çıkmasını engellemek ve doğru değerlendirmeler yapmak için alternatif analiz yöntemleri uygulanmaktadır.

Sayma verisi kullanılarak elde edilen sıfır ağırlıklı verilerin analizinde doğrusal regresyon yöntemi yerine daha duyarlı analizler ortaya çıkartmak için sıfır değerleri dikkate alan Poisson Regresyon (PR), Negatif Binom Regresyon (NBR), Sıfır Değer Ağırlıklı Poisson Regresyon (ZIP) ve Sıfır Değer Ağırlıklı Negatif Binom Regresyon (ZINB) modelleri kullanılmaktadır.

ZIP model Lambert tarafından 1992 yılında imalat kusurları ile ilgili çalışmasında geliştirilmiştir. Greene tarafından 1994 yılında aşırı yayılım gösteren veriler için NB modelinin genişletilmiş bir formu olan ZINB model tanımlamıştır. Özmen ve Famoye (2007) yapısal sıfırlar içeren zoolojik veriler üzerine yaptıkları çalışmada sıfır değer ağırlıklı modelleri karşılaştırmış ve uygun model seçimi yapmışlardır. Berk ve MacDonald (2008), kriminolojik araştırmalarda aşırı yayılım ve PR modeli üzerinde çalışma sunmuşlardır. Yang, Hardin ve Addy (2009), ZIP model için aşırı yayılımları test etmede kullanılan modeller üzerine çalışma sunmuşlardır. Hu, Pavlicova ve Nunes (2011), HIV risk taşıma durumu ile sayma verilerinden elde edilen aşırı sıfırların analizi üzerine çalışmışlardır. Kaya ve Yeşilova (2012) tarafından yapılan çalışmada üniversite içindeki e-posta trafiği incelenmiş ve veri seti üzerinden en uygun modelin seçimi belirlenmiştir. Peng (2013) tarafından sakatlanma verileri üzerinden çalışma yapılmış ve çeşitli değişkenler üzerinden uygun model seçimi yapılmıştır. Kong, Xu, Levy ve Datta 2014 yılında biyoistatistik çalışmalarına bir örnek olarak, diş çürüğü ile florür

uygulaması arasındaki ilişkiyi ZINB model ile modellenmesi üzerine uygulama çalışması yapmışlardır.

Bu çalışmada herhangi bir hizmet sektöründen alınan veriler kullanılarak, personellerin şikâyet alma sayılarını etkileyen faktörlerin açıklanmasında PR, NB, ZIP ve ZINB modellerinden hangisinin en uygun model olduğunun belirlenmesi üzerine çalışılmıştır.

1.1 Poisson Regresyon Modeli

Poisson regresyon (PR) modeli sayıma dayalı verilerinin analizinde tutarlı bir değerlendirme yapılmasına olanak sağlayan, sıfır değer ağırlıklı verilerin yorumlanmasında kullanılan temel yöntemdir (Ridout, Demetrio, Hinde ,1998; Peng,2013).

PR modelinde ortalama ile varyansın birbirine eşitliği söz konusudur. Ancak çoğu zaman bu eşitlik sağlanamamaktadır. Böyle durumlar ya aşırı yayılım ya da az yayılım olarak ifade edilmektedir.

Kullanılan verilerin kesikli yapıya sahip olduğu durumlarda doğrusal regresyon modelleri kullanılarak yapılacak analizler etkisiz, tutarsız ve güvenilmez sonuçlar verebilir. Özellikle, sayma ölçeğinde elde edilen veriler için kullanılabilir en etkin modellerden biri PR modelidir.

PR model denklemi aşağıdaki gibidir.

$$Pr(Y = y_i) = \exp(-\mu_i) \frac{\mu_i^{y_i}}{y_i!}, \quad y_i = 0,1,2, \dots$$

Poisson dağılımında ortalama ve varyans birbirine eşittir ve $E(Y) = V(Y) = \mu$ şeklindedir.

1.2 Negatif Binom Regresyon Modeli

Negatif Binom (NB) regresyonu, Poisson regresyonun özel bir durumu olmakla beraber yapılan uygulamalarda sıfır değerlerin aşırı yayılım gösterdiği sonuçlar ortaya çıktığı durumlarda alternatif yöntem olarak kullanılmaktadır. Sıfır değerlerin aşırı yayılım göstermesi halinde koşullu varyans değeri, koşullu ortalama değerini aşmaktadır (Beaujean, Morgan, 2016). Bu tür durumlarda, alternatif yöntem olarak NB regresyon modelleri kullanılması yapılan çalışmanın doğru değerlendirilmesi bakımından önem taşımaktadır.

NB regresyon model denklemi aşağıdaki gibidir.

$$Pr(Y = y_i) = \frac{\Gamma(y_i + a^{-1})}{y_i! \Gamma(a^{-1})} \left(\frac{a^{-1}}{a^{-1} + \mu_i} \right)^{a^{-1}} \left(\frac{\mu_i}{a^{-1} + \mu_i} \right)^{y_i}$$

NB regresyon modelinin beklenen değeri $E(Y_i) = \mu_i$ ve varyansı ise $Var(Y_i) = \mu_i(1 + a\mu_i)$ eşitliği ile ifade edilmektedir.

1.3 Sıfır Değer Ağırlıklı Poisson Regresyon Modeli

Sıfır değer ağırlıklı modellerin temel şartı elde edilen veri setlerinde sıfır değerlerin ağırlığı ve aşırı yayılım göstermesi durumudur (Akinpelu, Yusuf, Akpa, Gbolahan, 2016). Aşırı yayılım gösteren verileri analiz etmede kullanılan alternatif yöntemlerden birisi sıfır değer ağırlıklı Poisson regresyon (ZIP) modelidir.

ZIP modeli ilk olarak Diane Lambert (1992) tarafından üretim kusurları uygulaması ile ilgili çalışmasında tanımlanmıştır.

ZIP modelinde elde edilen sıfırların iki şekilde elde edildiğini varsayılır. Bunlar; daima sıfır grubundan gelen yani yapısal sıfırlar; diğeri ise her zaman sıfır olmayan gruptan gelen yani örnekleme sıfırı olarak tanımlanan gruptan gelen değerlerdir (Peng, 2013).

ZIP model denklemi aşağıdaki gibidir.

$$Pr(Y = y_i) = \begin{cases} p_i + (1 - p_i) \exp(-\mu_i) & y_i = 0 \\ (1 - p_i) \frac{\mu_i^{y_i}}{y_i!} \exp(-\mu_i) & y_i > 0 \end{cases}$$

Bu model denkleminde $0 \leq p_i \leq 1$ ve $\mu_i > 0$ 'dür.

ZIP modelinin ortalaması $E(Y_i) = (1 - p_i) \mu_i$ ve varyansı $Var(Y_i) = (1 - p_i) \mu_i (1 + p_i \mu_i)$ şeklinde gösterilmektedir.

$p_i = 0$ ise ZIP modeli PR dönüşür. $p_i > 0$ olması durumu aşırı yayılım göstergesidir.

ZIP modeli iki parçalı bir modeldir. Bu parçalardan Log fonksiyonu hem yapısal sıfırdan gelen hem de örnekleme sıfırından gelen sıfırlar ile Poisson ve Negatif Binom dağılımlarından gelen pozitif sayıları modellemede kullanılır. Diğer parça ise Logit fonksiyonudur. Bu kısım veri setindeki sıfırları modellemek için kullanılır (Peng,2013).

1.4 Sıfır Değer Ağırlıklı Negatif Binom Regresyon Modeli

Sıfır değer ağırlıklı negatif binom regresyon (ZINB) modeli veri setlerinde sıfır ağırlığı bulunması ve aşırı yayılım olması durumlarında alternatif bir yöntem olarak kullanılır. Bu model Greene (1994) tarafından NB modelinin gelişmiş bir şekli olarak tanımlanmıştır.

ZINB model denklemi aşağıdaki gibidir.

$$Pr(Y = y_i) = \begin{cases} \omega_i + (1 - \omega_i) \left(\frac{a^{-1}}{a^{-1} + \mu_i} \right)^{a^{-1}}, & y_i = 0 \\ (1 - \omega_i) \frac{\Gamma(y_i + a^{-1})}{y_i! \Gamma(a^{-1})} \left(\frac{a^{-1}}{a^{-1} + \mu_i} \right)^{a^{-1}} \left(\frac{\mu_i}{a^{-1} + \mu_i} \right)^{y_i}, & y_i > 0 \end{cases}$$

ZINB modelinin beklenen değeri $E(Y_i) = (1 - \omega_i)\mu_i$ ve varyansı $Var(Y_i) = E(Y_i) (1 + a\mu_i + \omega_i\mu_i)$ şeklinde tanımlanmaktadır.

ZINB model $\omega_i = 0$ olması durumunda NB regresyona; limit $a \rightarrow 0$ olması durumunda ise ZIP regresyona dönüşmektedir.

$a > 0$ ve $\omega_i > 0$ olması durumlarında ZINB modelinde aşırı yayılım söz konusudur (İsmail, Zamani, 2013).

ZINB fonksiyonu ZIP model gibi Log ve Logit parçalarında oluşmaktadır.

2. MODEL SEÇİMİ

Sıfır değer ağırlıklı verilerin analizi için kullanılan modeller arasından hangi modelin daha uygun olduğuna karar vermek için bazı testler kullanılmaktadır. Bu çalışmada Akaike Bilgi Kriteri (AIC) testi kullanılarak uygun model seçimi yapılmıştır. AIC model denklemi;

$$AIC = -2 \ln L + 2p$$

şeklinde dir. Burada, $\ln L$ uygun logaritmik olabilirlik değerini; p ise parametrelerin sayısını belirtir. AIC değeri küçük olan model diğer modellere göre veri seti için daha uygundur yorumu yapılır. Ayrıca, Logaritmik Olabilirlik (Log-Olabilirlik) değeri üzerinden model seçimi yapılırken ise, en büyük log-olabilirlik değerine sahip model uygun model olarak belirlenir.

3. YÖNTEM

Bu çalışmada herhangi bir hizmet sektöründen alınan veriler kullanılarak, personellerin şikâyet alma sayılarını etkileyen faktörlerin açıklanmasında PR, NB, ZIP ve ZINB modellerinden hangisinin en uygun model olduğunun belirlenmesi üzerine çalışılmıştır.

Hizmet, insanların mecburi ve mecburi olmayan ihtiyaçlarının tamamına hitap edecek şekilde hem yaşam standartlarının yükseltilebilmesi adına hem de temel ihtiyaçların karşılanması adına ortaya konan eylemlerin bütünüdür.

Hizmet sektörü, insan hayatının bütününe kapsayan ihtiyaçlara karşılık veren hizmet parçalarının bütünüdür. Herhangi bir ihtiyaç durumunda hizmet sağlayıcısının olmaması insanların mağduriyetine sebep olmakta bazı durumlarda ise insan hayatını tehlikeye sokabilmektedir.

Hizmet veren işletmelerin imajı, müşterilere sunulan hizmet kalitesi ile doğru orantılı olarak değerlendirilmektedir. Bu nedenle çalışan personellerin müşterilere karşı tutumu işletmeler için büyük önem arz etmektedir. Hizmet kalitesini arttırmak için çalışanların almış oldukları şikâyetleri değerlendirmek ve bu konuda önlem almak oldukça önemlidir. Müşterilerin aldıkları hizmetlerle ilgili yapmış oldukları geri bildirimler işletmeler tarafından titizlikle değerlendirilmeli ve şikâyet alma oranını etkileyen faktörler belirlenmelidir.

Günümüzde hizmet alan kişilerin şikâyet veya memnuniyetini dile getirebileceği birçok kanal bulunmaktadır. Müşteriler yazılı ve sözlü olarak işletme içinde, e-posta, çağrı merkezi, işletmenin internet sitesi üzerinden online olarak vs. şikâyetlerini gerekli yerlere ulaştırabilmektedirler. Hizmet sektörü verileri şikâyet alma oranları sıfır ağırlığı gösterdiğinden bu çalışmada kullanılan modeller için uygun veri setlerini oluşturmaktadır.

Bu amaçla elde edilen veriler üzerinden personellerin şikâyet alma durumlarını cinsiyet, yaş, eğitim durumu ve tecrübe değişkenleri üzerinden analiz ederek en uygun modelin belirlenmesi ve bu model üzerinden anlamlı değişkenlerin yorumlanması üzerine uygulama çalışması yapılmıştır.

4. BULGULAR

Elde edilen veri seti, 12 aylık bölümler halinde değerlendirmeye alınmıştır. R – studio programı kullanılarak tüm aylara ilişkin 4 model üzerinden parametre tahminleri elde edilmiş, en uygun model seçimi analizi yapılmış ve ilgili modellere ait anlamlı bulunan değişkenler için yorumlamalar yapılmıştır.

Kullanılan veri setinde ele alınan değişkenler ve düzeyleri aşağıdaki gibidir.

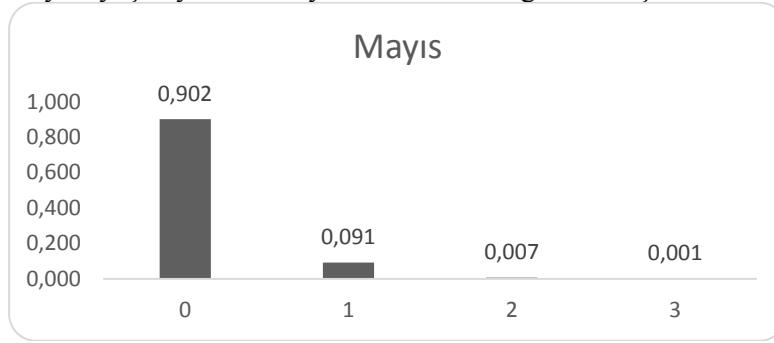
Tablo 1: Çalışmada kullanılan değişken ve değişken düzeyleri

Değişken	Gruplar
Cinsiyet	Kadın
	Erkek
Yaş	< 20
	21-30
	31-40
	>41
Eğitim Durumu	Lise
	Ön Lisans
	Lisans
	Yüksek Lisans
Tecrübe(Ay)	<12
	13-24
	25-48
	>49

Bu kısımda ilk olarak PR, NB, ZIP ve ZINB modellerinin en uygun model olarak belirlendiği 4 aya ait detaylı analiz sonuçları verilmiştir. Sonrasında ilgili 4 ay dışında kalan aylar için genel bilgiler verilmiş ve yorumlamaları yapılmıştır.

İlk olarak Mayıs ayı analiz sonuçları gösterilmiştir. Mayıs ayında değerlendirmeye alınan çalışanlardan kadın personellerin oranı 0.536 iken erkek operatörlerin oranı 0.464'tür. Çalışanların 0.706'sı 21 -30 yaş aralığında yer alırken, 0.044 ile 20 yaş altı ve 0.021 ile 41 yaş üstü en az orana sahip yaş aralıklarıdır. Çalışan personelin eğitim seviyesi durumlarına baktığımızda Lise mezunu olanlar 0.575 ile en çok orana sahiptir. Bu durumu 0.266 ile ön lisans mezunu olanlar takip etmektedir. Lisans mezunu olan personel sayısı 0.158'lik paya sahiptir. Yüksek lisans mezunu çalışanların oranı 0.001'dir.

Mayıs ayı şikâyet alma sayıları Grafik 1'de gösterilmiştir.



Grafik 1: Mayıs ayı şikâyet oranları

Mayıs ayı içerisinde hiç şikâyet almayan çalışanlar 0.902'lik paya sahiptir. Bir kere şikâyet alan personel oranı ise 0.091'dir. Grafikten de görüldüğü gibi çalışılan veride sıfır ağırlığı gözlemlenmektedir.

Tablo 2'de Mayıs ayı verileri için en uygun modeli belirleme amacıyla elde edilen AIC ve Log-olabilirlik değerleri verilmiştir.

Tablo 2: Mayıs ayı için model seçimi

	PR	NB	ZIP	ZINB
AIC	2017.269	2015.548	2012.301	2012.214
Log- Olabilirlik	-1003.634	-1001.774	-996.150	-995.107

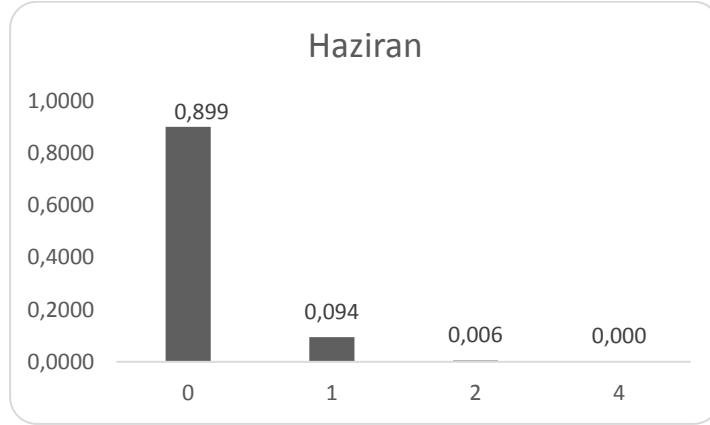
Tablo 2'ye göre en küçük AIC değerine sahip modelin ZINB regresyon modeli olduğu görülmektedir. Log-olabilirlik değerlerine bakıldığında da en büyük değere sahip model ZINB modeli olmuştur. Bu modele ait parametre tahmin değerleri Tablo 3'de verilmiştir.

Tablo 3: ZINB model parametre tahmini

Log	Tahmin	Standart hata	z-değeri	p-değeri
Sabit Terim	-2.311	0.369	-6.247	4.19e-10 *
Cinsiyet	0.057	0.132	0.438	0.661
Yaş	0.118	0.132	0.892	0.372
Eğitim Dur.	0.084	0.105	0.797	0.425
Tecrübe	-0.156	0.080	-1.941	0.052
Logit	Tahmin	Standart hata	z-değeri	p-değeri
Sabit Terim	-8.333	5.479	-1.521	0.120
Cinsiyet	-0.109	1.058	0.103	0.918
Yaş	2.596	1.222	2.125	0.033*
Eğitim Dur.	2.125	1.026	2.071	0.038*
Tecrübe	-3.174	1.701	-1.865	0.062

ZINB modeli için log tablosu kullanılarak elde edilen parametre değerlerine bakıldığında istatistiksel olarak anlamlı çıkan değişken bulunmamaktadır. Logit tablosu kullanılarak elde edilen parametre değerlerine bakıldığında ise, yaş ve eğitim durumu değişkenlerinin $\alpha=0.05$ anlamlılık düzeyinde istatistiksel olarak anlamlı bulunduğu görülmektedir. Şikâyet alma oranı üzerinde yaş ve eğitim durumu değişkenleri pozitif yönde istatistiksel olarak anlamlı bir etki göstermiştir. Çalışanların yaşları arttıkça ve eğitim durumları yükseldikçe şikâyet alma oranları artmıştır. Genel olarak daha yüksek eğitim durumuna sahip çalışanların diğer çalışanlara göre çok daha az şikâyet aldığı gözlemlenmektedir. Mayıs ayında bu duruma ters bir sonuç çıkma nedeni dönemsel olarak çalışan personellerin eğitim durumlarında meydana gelen değişkenlik gösterilebilir. Mayıs ayında diğer aylara göre eğitim seviyesi yüksek personel sayısında artış meydana gelmiş olabilir. Ayrıca elde edilen sonuçlara göre önemli bulgulardan biri, çalışan personelin yaş aralığında meydana gelen artış şikâyet alma oranını yaklaşık olarak 9 kat arttırmıştır. Aynı zamanda, personellerin eğitim durumlarındaki yükseliş ise şikâyet alma oranını yaklaşık olarak yine 9 kat olarak arttırmıştır.

İkinci olarak Haziran ayına ait değerlendirme sonuçları paylaşılmıştır. Haziran ayı şikâyet almama oranı 0.899 olarak gerçekleşmiştir. Yani yaklaşık olarak %90'lık sıfır ağırlığı gözlemlenmektedir. Şikâyet alma sayılarına ilişkin oranlar Grafik 2'de gösterilmiştir.



Grafik 2: Haziran ayı şikâyet oranları

Haziran ayı model seçimi değerlendirmesine ait veriler Tablo 4'de verilmiştir.

Tablo 4: Haziran ayı için model seçimi

	PR	NB	ZIP	ZINB
AIC	1668.582	1670.121	1665.620	1667.620
Log- Olabilirlik	-829.291	-829.060	-822.808	-822.809

Tablo değerlerine bakıldığında en düşük AIC değerine sahip model ZIP modeli çıkmıştır. Log-olabilirlik değerlerine bakıldığında da en büyük değere sahip model ZIP modeli olmuştur. Bu modele ait parametre tahmin değerleri Tablo 5'de verilmiştir.

Tablo 5: ZIP model parametre tahmini

Log	Tahmin	Standart hata	z-değeri	p-değeri
Sabit Terim	-1.239	0.626	-1.979	0.048*
Cinsiyet	0.054	0.185	0.294	0.769
Yaş	-0.045	0.185	-0.241	0.809
Eğitim Dur.	-0.100	0.132	-0.762	0.446
Tecrübe	-0.256	0.120	-2.125	0.033*
Logit	Tahmin	Standart hata	z-değeri	p-değeri
Sabit Terim	0.341	2.032	0.168	0.867
Cinsiyet	-0.365	0.610	-0.599	0.549
Yaş	1.507	0.713	2.113	0.035 *
Eğitim Dur.	-0.738	0.497	-1.484	0.138
Tecrübe	-2.023	0.941	-2.150	0.032 *

ZIP modeli log tablosu kullanılarak elde edilen parametre değerlerine bakıldığında, tecrübe değişkeninin $\alpha=0.05$ anlamlılık düzeyinde anlamlı bulunduğu görülmektedir.

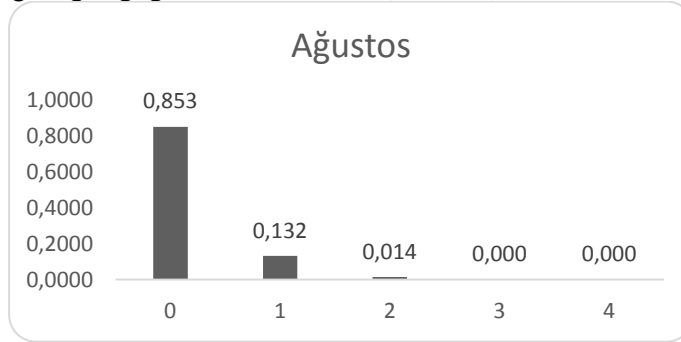
Şikâyet alma oranı üzerinde tecrübe değişkeni negatif yönde bir etki göstermektedir. Personellerin tecrübelerinde artış oldukça şikâyet alma oranları azalmıştır ki bu beklenen bir durumdur. Çalışan personelin tecrübelerindeki artış şikâyet alma oranını yaklaşık olarak %13 azaltmıştır.

ZIP modeli logit tablosu kullanılarak elde edilen parametre değerlerine bakıldığında, şikâyet alma oranı üzerinde yaş değişkeni pozitif yönde, tecrübe değişkeni ise negatif yönde istatistiksel olarak anlamlı bir etki göstermiştir. Elde edilen önemli bulgulardan biri, çalışan personellerin yaşlarında meydana gelen artış şikâyet alma oranını yaklaşık olarak 4.5 kat arttırmıştır. Yaş aralığında artış meydana geldikçe şikâyet alma oranı yükselmiştir.

Personellerin tecrübe durumlarındaki yükseliş ise şikâyet alma oranını yaklaşık olarak %87 oranında azaltmıştır. Personellerin tecrübeleri arttıkça şikâyet alma oranlarında düşüş meydana gelmiştir.

Burada yaş ve tecrübe değişkenlerinin zıt yönde etki göstermesi çelişki olarak değerlendirilmemelidir. Her zaman yaş aralığı yüksek olan personel tecrübelidir şeklinde bir durum oluşmamaktadır. Çıkan sonuçların yorumunda bu bilgi dikkate alınmalıdır.

Üçüncü olarak Ağustos ayına ait veriler değerlendirilmiştir. Ağustos ayında şikâyet almama oranı 0.852 olarak gerçekleşmiştir. Yani yaklaşık olarak %85'lik sıfır değer ağırlığı gözlemlenmektedir (Grafik 3).



Grafik 3: Ağustos ayı şikâyet oranları

Ağustos ayında hiç şikâyet almayan çalışanların oranının yaklaşık %85 olduğu grafikten görülmektedir. 1 kez şikâyet alan çalışan oranı yaklaşık olarak %13'tür. Ağustos ayı model seçimi değerlendirmesine ait veriler Tablo 6'da verilmiştir.

Tablo 6: Ağustos ayı için model seçimi

	PR	NB	ZIP	ZINB
AIC	2145.790	2145.157	2150.923	2152.749
Log- Olabilirlik	-1067.895	-1064.578	-1065.462	-1065.374

Tablo değerlerine bakıldığında Ağustos ayı için en düşük AIC değerine sahip modelin NB modeli olduğu belirlenmiştir. Log-olabilirlik değerlerine bakıldığında da en büyük değere sahip model NB modeli olmuştur. Bu modele ait parametre tahmin değerleri Tablo 7’de verilmiştir.

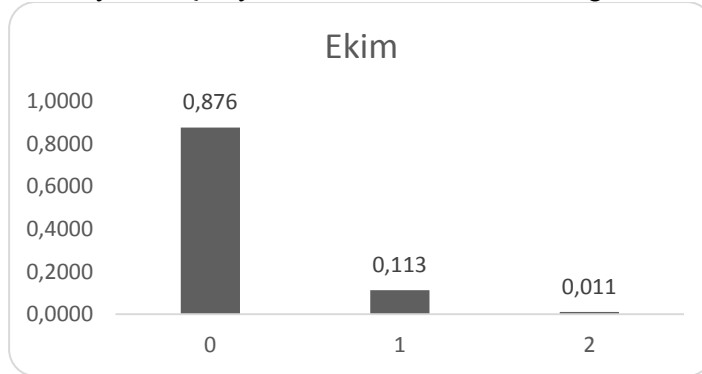
Tablo 7: NB modeli için parametre tahmini

	Tahmin	Standart hata	z-değeri	p-değeri
Sabit Terim	-1.509	0.314	-4.801	1.58e-06 *
Cinsiyet	0.150	0.107	1.397	0.162
Yaş	-0.231	0.101	-2.273	0.023 *
Eğitim Dur.	-0.052	0.072	-0.723	0.470
Tecrübe	0.026	0.060	0.429	0.668

Tablo 7 değerlerine bakıldığında çalışan personellerin yaş durumu, $\alpha=0.05$ anlamlılık düzeyinde şikâyet alma oranı üzerinde negatif yönde istatistiksel olarak anlamlı bir etki göstermiştir. Çalışan personelin yaşlarında meydana gelen artış şikâyet alma oranını yaklaşık olarak %6 azaltmıştır. Ağustos ayı verilerine göre çalışan personelin yaş aralığı arttıkça şikâyet alma oranında azalış gözlemlenmiştir.

Son olarak ise Ekim ayına ait değerlendirme sonuçları gösterilmiştir. Ekim ayında çalışan personelin şikâyet almama oranı yaklaşık olarak %87 olmuştur. 1 kez şikâyet alan personelin oranı yaklaşık olarak %11’dir.

Ekim ayına ait şikâyet alma oranları Grafik 4 ‘te gösterilmiştir.



Grafik 4: Ekim ayı şikâyet oranları

Ekim ayı model seçimi değerlendirmesine ait veriler Tablo 8’de verilmiştir.

Tablo 8: Ekim ayı model seçimi

	PR	NB	ZIP	ZINB
AIC	1918.525	1919.774	1922.736	1924.737
Log- Olabilirlik	-951.262	-953.887	-951.368	-951.368

Ekim ayına ait model seçimi tablo değerlerine bakıldığında en düşük AIC değerine sahip modelin PR modeli olduğu belirlenmiştir. Log-olabilirlik değerlerine bakıldığında da en büyük değere sahip model PR modeli olmuştur.

Bu modeline ait parametre tahmini değerleri Tablo 9’da verilmiştir.

Tablo 9: PR modeli için parametre tahmini

	Tahmin	Standart hata	z-değeri	p-değeri
Sabit Terim	-1.345	0.333	-4.034	5.48e-05 *
Cinsiyet	-0.038	0.114	-0.330	0.742
Yaş	-0.131	0.105	-1.245	0.213
Eğitim Dur.	-0.241	0.081	-2.978	0.003 *
Tecrübe	0.023	0.064	0.354	0.723

Tablo 9 değerlerine bakıldığında eğitim durumu değişkeninin $\alpha=0.05$ anlamlılık düzeyinde istatistiksel anlamlı olarak bulunduğu görülmektedir.

Şikâyet alma olasılığı üzerinde eğitim durumu değişkeni negatif yönde istatistiksel olarak anlamlı bir etki göstermiştir. Çalışan personelin eğitim durumundaki değişiklik şikâyet alma oranını yaklaşık olarak %22 azaltmıştır.

Yani Ekim ayı verilerine göre çalışan personelin eğitim durumu yükseldikçe şikâyet alma olasılığında azalış gözlemlenmiştir.

Tüm aylara ilişkin AIC değerlerine bakılarak elde edilen en uygun modeller ve veri setlerindeki sıfır değer ağırlıkları Tablo 10’da gösterilmiştir.

Tablo 10: Tüm Aylara İlişkin Uygun Modeller ve Sıfır Değer Ağırlıkları

Aylar	En Düşük AIC Değeri	Sıfır Değer Ağırlığı
Ocak	ZIP	0.893
Şubat	NB	0.895
Mart	ZIP	0.863
Nisan	NB	0.889
Mayıs	ZINB	0.902
Haziran	ZIP	0.899
Temmuz	PR	0.891
Ağustos	NB	0.853
Eylül	PR	0.878
Ekim	PR	0.876
Kasım	NB	0.873
Aralık	PR	0.847

Ocak ayında en uygun model olarak ZIP modeli belirlenmiştir. ZIP modeline ait parametre tahminlerine bakıldığında eğitim durumu değişkeninin negatif yönde

anamlı bir etki gösterdiği görülmüş; personellerin eğitim durumlarında meydana gelen artış şikâyet alma oranını azaltmıştır.

Şubat, Mart, Nisan, Temmuz, Eylül, Kasım ve Aralık aylarına ilişkin parametre tahminleri incelendiğinde istatistiksel olarak anlamlı değişken veya değişkenler olmamıştır.

5. SONUÇ

Günümüzde alınan hizmetlerle ilgili olarak müşterilerin memnun kalmadıkları ya da memnun oldukları durumları dile getirebileceği birçok alan bulunmaktadır. İşletmelerin kullanmakta oldukları internet siteleri, e-posta, çağrı merkezi ve direkt olarak işletme sorumluları bu bildirim alanlarına örnektir. Hizmet sektörü verilerinde, çalışan personellerin şikâyet alma oranları sıfır değer ağırlığı gösterdiğinden, bu çalışma için uygun veri setlerini oluşturmuştur.

Bu amaçla kullanılan veri üzerinden işletmelerde çalışan personellerin şikâyet alma durumları cinsiyet, yaş, eğitim durumu ve tecrübe değişkenleri üzerinden analiz edilmiş, aylık olarak değerlendirilen veriler için en uygun modelin hangisi olduğu gösterilmiş ve belirlenen model üzerinden anlamlı bulunan değişkenler yorumlanmıştır.

Kullanılan veri setlerinde beklenenden fazla sayıda sıfır ağırlığı gözlenmesi halinde Poisson Regresyon, Negatif Binom Regresyon, Sıfır Değer Ağırlıklı Poisson Regresyon ve Sıfır Değer Ağırlıklı Negatif Binom Regresyon modelleri kullanılarak analizler yapmak daha uygun olmaktadır.

Bu çalışmada kullanılan veri setleri de sıfır değer ağırlığı göstermektedir. 12 ay için ayrı ayrı yapılan uygulama neticesinde veriyi açıklamada her ay için farklı modellerin uygunluk gösterdiği belirlenmiştir.

Elde edilen uygun modellere göre yapılan parametre tahminlerinde; Ocak ayında personellerin eğitim durumlarında meydana gelen artışın şikâyet alma oranını azalttığı; Mayıs ayında personellerin yaş ve eğitim durumlarında meydana gelen artışın şikâyet alma oranını arttırdığı;

Haziran ayında personellerin tecrübelerindeki artışın şikâyet alma oranını azalttığı, yaş aralığındaki artışın ise şikâyet alma oranını arttırdığı;

Ağustos ayında personellerin yaşlarındaki artışın şikâyet alma oranını azalttığı; Ekim ayında ise personellerin eğitim durumlarındaki artışın şikâyet alma oranını azalttığı yönünde istatistiksel olarak anlamlı bulgular elde edilmiştir.

Yapılan analiz sonuçlarına göre, sıfır değerlerinin ağırlıkta olduğu aylara ilişkin değerlendirmelerde genel olarak sıfır değer ağırlıklı modellerin veri setinin açıklamada daha uygun modeller olduğu gözlenmiştir. Sıfır değer ağırlığının azaldığı durumlarda ise Poisson Regresyon ve Negatif Binom Regresyon modelleri en uygun modeller olarak elde edilmiştir.

KAYNAKÇA

- Akinpelu, K.P., Yusuf, O.B., Akpa, O.M. and Gbolahan, A.O. Zero “Inflated Regression Models with Application to Malaria Surveillance Data”, *International Journal of Statistics and Applications*, 6(4), 223-224.
- Beaujean A.A. and Morgan G.B. (2016). “Tutorial on Using Regression Models with Count Outcomes Using R”. *Practical Assessment Research & Evaluation*, 21(2), 1531-7714.
- Berk, R. And MacDonald, J.M. (2008). “Overdispersion and Poisson Regression”, *Journal of Quantitative Criminology*, 24(3), 269-284.
- Greene, W.H. (1994). Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models, *New York University*, New York.
- 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”. *Am J Drug Alcohol Abuse*, 37(5), 367-375.
- Ismail, N. ve Zamani H. (2013). “Estimation of Claim Count Data Using Negative Binomial, Generalized Poisson, Zero-Inflated Negative Binomial and Zero-Inflated Generalized Poisson Regression Models”. *Casualty Actuarial Society E-Forum*
- Kaya, Y. ve Yeşilova, A. (2012). “E-Posta Trafikinin Sıfır Değer Ağırlıklı Regresyon Yöntemleri Kullanılarak İncelenmesi”. *Anadolu Üniversitesi Bilim ve Teknik Dergisi*, 13(1), 51-63.
- 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.
- Özmen, İ. and Famoye, F. (2007). “Count Regression Models with an Application to Zoological Data Containing Structural Zeros”. *Journal of Data Science*, 5(2017), 491-502.
- Peng, J. (2013). Count Data Models for Injury Data from the National Health Interview Survey, (M. Sc. Thesis), The Ohio State University Graduate Program in Public Health, Columbus.
- Ridout, M., Demetrio, C.G.B. and Hinde, J. (1998). “Models for Count Data with Many Zeros”. International Biometric Conference, Cape Town.
- Yang, Z., Hardin, J.W. and Addy, C. (2009). “Testing Overdispersion in the Zero-Inflated Poisson Model”. *Journal of Statistics Planning and Inference*, 139, 3340-3353.