VERİ BİLİMİ DERGİSİ
www.dergipark.gov.tr/veri

Kırpılmış Poisson Regresyon Analizi ve Bir Uygulama

Necati ERİLLİ[1]*, Seçil KARTAL[2]

[1]Sivas Cumhuriyet Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, Sivas
[2] Sivas Cumhuriyet Üniversitesi, Sosyal Bilimleri Enstitüsü, Ekonometri A.B.D., Sivas

Makale Bilgisi

Başvuru:
26/10/2020
Kabul:
20/12/2020

Özet

Regresyon analizi aralarında neden sonuç ilişkisi olan iki veya daha fazla değişkenin aralarındaki ilişkiyi gözlemek ve konuya ait öngörüde bulunabilmek için elde edilen matematiksel bir modelle belirtilen istatistiksel bir yöntemdir. Poisson regresyonu sayıma dayalı olarak elde edilen bağımlı değişkenin modellenmesinde kullanılır. Bununla birlikte, bağımlı değişkenin sayıma dayalı olarak elde edildiğinde, bağımsız değişken kategorileri için relatif risk değerini de hesaplamaktadır. Kırpılmış Poisson Regresyon Analizi, Poisson Regresyon Analizinde uç değerlerin analizlerde hesaplama veya yorum zorlukları çıkarması sebebiyle tanıtılan bir yöntemdir. Bu çalışmada kırpılmış poisson regresyonu, 3 farklı veri üzerinde uygulanmış ve sonuçlar klasik poisson regresyon model ile karşılaştırılmıştır. Çalışmadan elde edilen sonuçlara göre aşırı değerler içeren bağımlı değişken yapılarında klasik poisson regresyonu yerine kırpılmış poisson regresyon analizin kullanılması önerilmektedir. Böylece klasik poisson regresyon analizinde bağımlı değişkenin aşırı uç değerlerden veya belirlenen limitlerde yanlı sonuçlar vermesinin de önüne geçilebilir.
Anahtar Kelimeler: *Kırpılmış Poisson Regresyon Analizi, Poisson Regresyon Analizi, Log-likelihood*

Truncated Poisson Regression Analysis and an Application

Abstract

Regression analysis is a statistical method specified with a mathematical model obtained in order to observe the relationship between two or more variables that have a cause and effect relationship and to make predictions about the subject. Poisson regression is used to model the dependent variable based on count. However, when the dependent variable is obtained based on count, it also calculates the relative risk value for the independent variable categories. Truncated Poisson Regression Analysis Method is introduced because extreme values cause calculation or interpretation difficulties in Poisson Regression Analysis. In this study, the Truncated Poisson regression was applied on 3 different data and the results were compared with the classical Poisson regression model. According to the results obtained from the study, it is recommended to use Truncated Poisson regression analysis instead of classical Poisson regression for dependent variable structures with extreme values. Thus, in classical Poisson regression analysis, it can be prevented that the dependent variable gives biased results from extreme values or at specified limits.
Keywords: *Truncated Poisson Regression Analysis, Poisson Regression Analysis, Log-likelihood*

* İletişim e-posta: aerilli@cumhuriyet.edu.tr

1 Giriş

Regresyon analizi aralarında neden sonuç ilişkisi olan iki veya daha fazla değişkenin aralarındaki ilişkiyi gözlemek ve konuya ait öngörüde bulunabilmek için elde edilen matematiksel bir modelle belirtilen istatistiksel bir yöntemdir. Basit doğrusal regresyon modeli bir bağımlı bir bağımsız değişken olmak üzere aralarındaki fonksiyonel ilişkiyi gözlemlerken, çoklu doğrusal regresyon modeli bir bağımlı ve birden fazla bağımsız değişken arasındaki fonksiyonel ilişkiyi incelemektedir. Regresyon analizinin amacı yanıt değişken ile açıklayıcı değişken arasındaki ilişkiyi matematiksel modellerle belirterek, geleceğe yönelik öngörüde bulunmaktır [1]. Bu konuda geliştirilmiş birçok istatistiksel yöntemden bahsetmek mümkündür. Bu yöntemlerin uygulanacağı veri setlerinde yanıt değişkenin normal dağılım içerikli olması, açıklayıcı değişkenlerin normal dağılım gösteren değişken ya da değişkenlerden meydana gelmesi ve hata terimlerinin varyansının normal dağılımlı olması gerekmektedir [2].

Doğrusal regresyon ve varyans analizi gibi doğrusal modellerin temel varsayımlarından biri, hataların normal bir dağılım göstermesidir. Sürekli bir bağımlı değişken çarpık hale geldiğinde bu varsayımı karşılamak için, bağımlı değişkenin dönüşümü ile elde edilen yeni bir model yapısı ile yaklaşık normal olan hatalar üretebilir. Bununla birlikte, çoğu zaman ilgilendiğimiz yanıt değişkeni sürekli değil, kategorik ya da kesikli olarak karşımıza çıkmaktadır. Bu durumda, basit bir dönüşüm bile normal dağılmış hataları üretemeyecektir.

Yaygın bir örnek, yanıt değişkeninin bir olayın sayılan oluşum sayısı olmasıdır. Sayımların dağılımı kesiklidir, sürekli değildir ve negatif olmayan değerler ile sınırlıdır. Bu verilere sıradan bir doğrusal regresyon modeli uygulamanın iki sorunu vardır. İlk olarak, sayım verilerinin pek çok dağılımı, veri kümesinde 0 değerine sahip birçok gözlemle pozitif olarak çarpıtılmıştır. Veri kümesindeki 0'ların yüksek olması, çarpık bir dağılımın normal bir dağılıma dönüşmesini engeller. İkinci olarak, regresyon modelinin teorik olarak imkânsız olan negatif tahmini değerler üretmesi oldukça muhtemeldir.

Sayım verileri günlük olaylarda daha çok karşımıza çıkar. Bu tür verilerin daha iyi anlaşılması ve verilerle ilgili önemli bilgilerin çıkarılması, bazı istatistiksel analiz veya modellemeler gerektirir. Sayma verileri, herhangi bir olayın belirli bir süre içerisinde kaç kez meydana geldiğini gösterir. Sayma modellerindeki ilk gelişmeler aktüerya, biyoistatistik ve demografide gözlemlenmiştir. Bu modeller son zamanlarda ekonomi, siyaset bilimleri ve sosyolojide de geniş bir biçimde kullanılmaktadır [3].

Sayma verisi için regresyon modellerinin geliştirilmesindeki dönüm noktası, Poisson Regresyonun özel bir durumu olduğu "Genelleştirilmiş Doğrusal Modellerin" ortaya çıkmasıdır. İlk olarak Nelder ve Wedderburn tarafından tanımlanmış, McCullagh ve Nelder tarafından detaylandırılmıştır [4, 5]. Sayma veri regresyon modelleri, kesikli regresyonun özel bir türüdür. Negatif olmayan kesikli değerlerden elde edilir. Bu verilerde doğrusal regresyon modelinin uygulanması öngörülen katsayıların yanlı olmasına neden olur. Sayım verisini elde etmek için yararlanılan en yaygın regresyon modeli "Poisson Regresyon", varsayımların sağlanamadığı durumda ise "Negatif Binom Regresyon" modelidir [6]. Farklı sayım verileri farklı özelliklere sahip olabilir ve bu nedenle belirli sayım veri modelleriyle kullanılamaz. Poisson regresyon modeli, sayım verilerinin analizi için bir temel sağlar. Birçok uygulayıcı, bu modelin tüm varsayımlarının karşılandığından emin olmadan bile sayım verilerini içeren veri analizi ile karşı karşıya kaldıklarında Poisson modelini kullanmayı seçmektedir.

Poisson modeli, hataların ortalama ve varyansının eşit olduğunu varsayar. Ancak pratikte genellikle hataların varyansı ortalamadan daha büyüktür (ancak daha küçük de olabilir). Varyans ortalamadan daha büyük olduğunda, Poisson modelinin iyi çalışan iki uzantısı vardır. Aşırı dağınık Poisson modelinde, varyansın ortalamadan ne kadar büyük olduğunu tahmin eden ekstra bir parametre dahil edilir. Bu parametre tahmini daha sonra daha büyük varyansın p değerleri üzerindeki etkilerini düzeltmek için kullanılır. Bir alternatif, negatif bir iki terimli modeldir. Negatif binom dağılımı, dağılımın parametresinin kendisinin bir rastgele değişken olarak kabul edildiği Poisson dağılımının bir şeklidir.

Bu çalışmada poisson regresyon analizinin özel bir hali olan kırpılmış poisson regresyonu incelenmiştir. Yabancı literatürde az da olsa farklı yönleri ile incelenen bu konu hakkında yerli literatür yok denecek kadar azdır. Çalışmanın yerli literatüre katkısının –özellikle kanser çalışmalarında- olması beklenmektedir. Çalışmanın ikinci bölümünde poisson regresyonu ve kırpılmış poisson regresyonu ve kısa literatür incelemesi yer almaktadır. Uygulama bölümünde daha önce farklı çalışmalarda yer almış 3 gerçek veri üzerinde poisson-kırpılmış poisson regresyonu karşılaştırılması yapılmış ve yorumlanmıştır.

2 Poisson Regresyon Analizi

Poisson Regresyon analizi, bağımsız değişkenler ile sayımla ulaşılan bağımlı değişken arasındaki bağlantıyı gösteren bir yöntemdir. Poisson regresyon analizindeki temel alınan yapı, Y_i yanıt değişkeninin kesikli bağımsız Poisson rastlantı değişkeni olmasıdır. Kesiklilik sebebiyle normallik varsayımı sonucu elde edilememesiyle klasik doğrusal regresyon analizine ek olarak gösterilen metotlardan biridir [7, 8]. Poisson Regresyon modeli, Poisson dağılımının ortalamasına göre belirlenir ve model Eşitlik.1'de ifade edilmiştir [9].

$$P(y_i / x_i) = \begin{cases} \dfrac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, y = 0,1,2,\ldots \end{cases} \quad (1)$$

Formülde kullanılan λ_i parametresi $\lambda_i = E\left(y_i / x_i\right) = \mu\left(x_i\right) = c_i f\left(x_i, \beta\right); i = 1,\ldots,n$ şeklinde tanımlanır. Burada y_i; genellikle bir denemedeki başarısızlık sayısını, μ_i; denemedeki olayın ortalama oluş sayısını, c_i; ilgilenilen olay için riskli toplam kişi sayısı ya da kitle genişliğini ve $f\left(x_i, \beta\right)$ regresyon hız fonksiyonunu göstermektedir. Poisson Regresyon analizinde en çok kullanılan regresyon fonksiyonu logdoğrusal model olmakla birlikte doğrusal ve doğrusal olmayan modellerde kullanılabilmektedir. Log-doğrusal modeli kullanarak Poisson regresyon modelinin ortalama parametresi; $\mu_i = E\left(y_i / x_i\right) = e^{x_i' \beta}$ olarak tanımlanmıştır.

Poisson regresyon modelinde parametrelerin tahminleri genel olarak en çok olabilirlik yöntemi ile gerçekleştirilmektedir. En çok olabilirlik tahmini için koşullu ortalamada bağımlılık şartı sağlanmalıdır. Bunun yanı sıra bağımlı değişken Y'nin Poisson dağılımından gelmesi gerekmektedir. İkinci bir şart olarak ise en çok olabilirlik standart hataları ve t istatistikleri kullanarak hesaplanan istatistiksel sonuçlarda koşullu varyans ve ortalamanın eşit olması şartı da sağlanmalıdır [10].

2.1 *Kırpılmış Poisson Regresyon Analizi*

Kırpılmış Poisson Regresyon Analizi, Poisson Regresyon Analizinde uç değerlerin analizlerde hesaplama veya yorum zorlukları çıkarması sebebiyle tanıtılan bir yöntemdir. Kırpılmış ortalama hesabına dayanan kırpılmış Poisson Regresyon analizi, temelde üç ana grupta incelenebilir: Soldan Kırpılmış, Sağdan Kırpılmış ve çift yönlü kırpılmış Poisson Regresyon Analizi.

Sayımla elde edilen verilerde $k = c_L$ noktası için $y < c_L$ değerleri gözlenmemişse bu tarz verilere soldan kırpılmış veriler, $k = c_R$ noktası için $y > c_R$ değerleri gözlenmemişse bu tarz verilere ise sağdan kırpılmış veriler adı verilmektedir [11].

Soldan kırpılmış poisson regresyonu genellikle tıbbi çalışmalarda görülür. Mesela, hastaların hastalık başladıktan sonra hayatta kalma süresi üzerinde çalışmak isteyen birileri için uygun yöntem olarak karşımıza çıkmaktadır. Benzer şekilde aktüeryal araştırmalarda, astronomik, demografik, epidemiyolojik, güvenilirlik testlerinde ve diğer çalışmalarda da yaygındır. Sağdan kırpılma genellikle, her bir bireyi ayırt etmedeki yetersizlik nedeniyle yüksek sayma zorluğu olduğunda veya yüksek değerli sayımlar için yardımcı malzemeler sorun çıkardığında ortaya çıkar. İki yönlü kırpılmış Poisson regresyon analizi ise, soldan ve sağdan kırpılmış Poisson verisinin birleşimi ile yapılan modellerdir. Bu tür bir kırpılma bazen, yalnızca içinde en az bir birey bulunan ekim alanlarının korunduğu ve yüksek yoğunlukların sayılmasının zor olduğu botanik çalışmalarda ortaya çıkar [12].

Poisson olasılık dağılımı yardımıyla, soldan kırpılmış model Eşitlik.2'de verildiği gibi tanımlanabilir:

$$P(Y_i = y_i | Y_i > c_L) = \frac{\lambda_i^{y_i}}{\left[e^{\lambda_i} - \sum_{k=0}^{c_L} \frac{\lambda_i^k}{k!}\right] y_i!} \quad (2)$$

Burada $\lambda_i = e^{x_i^T \beta}$ ve $(i = 1,\ldots,n)$ olarak tanımlanmıştır.

Sağdan kırpılmış model de soldan kırpılmış modele benzer şekilde Eşitlik.3'de olduğu gibi verilir:

$$P(Y_i = y_i | Y_i \le c_R) = \frac{\lambda_i^{y_i}}{\left[\sum_{k=0}^{c_R} \frac{\lambda_i^k}{k!}\right] y_i!} \quad (3)$$

Sağdan kırpılma, her bir bireyi ayırt edememe nedeniyle yüksek sayıları saymada zorluk olduğunda veya sayma aparatı yüksek sayım için sorun çıkardığında ortaya çıkar [13].

İki yönlü kırpılmış poisson regresyon modelleri sağdan ve soldan kırpılmış modelleridir birleşimi olarak tanımlanabilir ve model Eşitlik.4'de verildiği gibidir:

$$P(Y_i = y_i | x_i; c_L \leq y_i \leq c_R) = \frac{\lambda_i^{y_i}}{\left[\sum_{k=c_L}^{c_R} \frac{\lambda_i^k}{k!}\right] y_i!} \quad (4)$$

Kırpılmış sayım verileri genetik, epidemiyoloji, astronomik, demografik, güvenilirlik testi ve diğer birçok çalışmada kullanımı mevcuttur. Bu nedenle, kırpılmış verilerin dağılımının yanlış tanımlanması, ilk koşullu kırpılma anının da yanlış belirleneceği anlamına gelir. Bu yanlış tanımlamanın tutarsız β tahmin ediciler olmasına sebep olacağından, tanımlamanın kırpılmış regresyon aşamasında önemli bir yer tuttuğuna dikkat edilmelidir [14].

Literatürde kırpılmış poisson regresyon konusunda yapılmış çalışmalar sınırlıdır. Grogger ve Carson, kırpılmış sayma verileri modelleri üzerinde çalışmalar yapmışlardır [15]. Çalışmada kırpılmış poisson ve kırpılmış negatif binom dağılımları ile elde edilmiş regresyon modelleri, simülasyon ve gerçek veri yardımıyla karşılaştırılmış ve yorumlanmıştır. Xie ve Aickin, çalışmalarında kırpılmış Poisson parametrelerini log-lineer regresyon modeli ile birleştirip, maksimum olabilirlik yöntemi ile regresyon parametrelerini tahmin etmişlerdir [11]. Küçük hücreli kanser verisi üzerine yaptıkları uygulamada, kolonoskopi sonuçlarından yararlanarak model sonuçlarını yorumlamışlardır. Fokianos, zaman serileri için kırpılmış poisson regresyon analizi modelleri için kısmi olabilirlik fonksiyonlarını incelemiştir [16]. Van der Heijden vd. poisson dağılımı göstermiş bir anakitle için kırpılmış poisson regresyon modelleri yardımıyla nokta ve aralık tahminlerinde bulunmuşlardır [17]. Bunun için Horvitz-Thompson modeli yardımıyla sıfır kırpılmış poisson regresyon modeli kullanmışlardır. Gurmu ve Triverdi ise aşırı dağılım gösteren veri yapılarında sağdan ve soldan kırpılmış regresyon analizi üzerine Monte-Carlo denemeleri yapmışlardır [18]. Yeh vd. sansürlü verilerde sıfır kırpılmış poisson regresyonu üzerinde çalışmalar yapmışlardır [19]. Meme kanseri verisi üzerinde önerilen yöntem uygulanmış ve yorumlanmıştır. Puza vd., bayesçi kırpılmış veriler üzerinde çalışmışlardır [20]. Çalışmalarında özellikle sıfır kırpılmış poisson regresyon analizi üzerinde yoğunlaşmışlar, önerilen yöntemin sonuçlarını Monte Carlo simülasyonu ile göstermişlerdir. Finney ve Varley, kırpılmış poisson parametreleri hakkında yapılmış önceki çalışmaları derlemiş, homojenlik test sonuçlarını göstermişlerdir [21]. Van der Heijden vd. çalışmalarında, örneklem hacmi azaldığında, kırpılmış poisson regresyon model tahmin sonuçlarının kovaryans değerlerinin de azaldıklarını göstermişlerdir [22]. Adesina vd., sıfır kırpılmış sayma modelleri için bayesci modelleri tanıtmışlardır [23]. Monte Carlo simülasyon verileri yarımıyla kırpılmış poisson, Quasi poisson, negatif binom regresyon modelleri karşılaştırılmış ve model farklılıkları yorumlanmıştır. Suaiee, yaptığı tez çalışmasında soldan, sağdan ve iki yönlü kırpılmış poisson regresyon modellerini incelemiş ve rastgele etkilere sahip Çift Kesilmiş Poisson regresyon modeline dayalı bir regresyon modeli sınıfı formüle etmiştir [ ]. Çalışmada sonuç olarak iki yönlü ve soldan kırpılmış modellerin, kümelenmiş iki yönlü kırpılmış verilerde daha başarılı sonuçlar verdiği belirlenmiştir.

3 Uygulama

Uygulamada 3 farklı veri seti kullanılmıştır. 1. veri seti 1493 gözlemden oluşmaktadır. Bağımlı değişken hastaların hastanede yatma süreleri (gün), bağımsız değişkenler yaş, sağlık sigortası olup olmaması ve hastanın ölüm parametreleridir [24]. 2. veri seti 120 gözlemden oluşmaktadır. Bağımlı değişken trafikte meydana gelen kaza sayılarını, bağımsız değişkenler ise kaza yapan kadın sürücü olup olmadığı, kaza anındaki buzlanma şartları, havanın yağışlı olup olmaması ve kazaya karşına sürücünün ehliyet değişkenlerden oluşmaktadır [25]. 3. veri seti 435 gözlemden oluşmaktadır. Bağımlı değişken akciğer kanseri (durum) bağımsız değişken takip süresi (gün), teşhis ve yaş parametreleridir [26]. Analizler STATA.15 paket programı ile yapılmıştır. Her veri setine Poisson ve Kırpılmış poisson analizleri uygulanmış ve parametre tahminleri elde edilmiştir. Ayrıca uygun görülen gözlem değerleri için sağdan veya soldan kırpılmış poisson analizleri de yapılarak model karşılaştırmaları yapılmıştır. İstatistiksel testler için önem seviyesi 0,05 olarak alınmış ve model karşılaştırması log-lilkelihood model seçim kriterine göre yapılmıştır. Tablo.1'de ilk veri seti için kırpılmış poisson regresyon katsayı tahmin sonuçları verilmiştir.

Tablo 1. Birinci veri seti için kırpılmış poisson regresyonu parametre tahmin sonuçları

Bağımlı Değişken: Hastanede yatma süresi (Gün)			
	Katsayı	*Std. Hata*	*p*
Yaş	-0,014430	0,00503	0,004
Sağlık Sigortası	-0,135903	0,02374	0,000
Ölüm	-0,203770	0,01837	0,000
Sabit	2,435808	0,02733	0,000

Tablo.1'deki sonuçlara göre tüm katsayıların anlamlı olduğu, hastaların yatma süresilerine yaş, sağlık sigortası ve ölüm değişkenlerinin ters yönde etki yaptıkları görülmektedir. Tablo.2'de 75 yaş ve üzerindekilerin (Y=75) kırpıldığı, sağdan kırpılmış poisson regresyonu parametre tahmin sonuçları verilmiştir.

Tablo 2. Birinci veri setinde Y=75 için sağdan kırpılmış poisson regresyonu parametre sonuçları

Bağımlı Değişken: Hastanede yatma süresi (Gün)			
	Katsayı	*Std. Hata*	*p*
Yaş	-0,014430	0,005032	0,004
Sağlık Sigortası	-0,135745	0,023722	0,000
Ölüm	-0,203572	0,018359	0,000
Sabit	2,435763	0,027322	0,000

Tablo.2'deki sonuçlara göre tüm katsayıların anlamlı olduğu, katsayıların kırpılmış poisson regresyon analizi sonuçları gibi bağımlı değişkene ters yönde etki yaptıkları görülmektedir. Benzer şekilde veriye Y=80 ve Y=85 için de sağdan kırpılmış poisson regresyonu analizi uygulanmış, Y=75 için elde edilen sonuçların aynısı elde edilmiştir. Böylece hastanede kalma üst limiti 75 ve daha fazlası olduğunda sonuçların değişmediği görülmüştür. Tablo.3'de ise poisson regresyon analizi katsayı tahmin sonuçları verilmiştir.

Tablo.3 Birinci veri seti için poisson regresyonu parametre tahmin sonuçları

Bağımlı Değişken: Hastanede yatma süresi (Gün)			
	Katsayı	*Std. Hata*	*p*
Yaş	-0,01443	0,005032	0,004
Sağlık Sigortası	-0,13574	0,023722	0,000
Ölüm	-0,20357	0,018359	0,000
Sabit	2,435763	0,027322	0,000

Tablo.3'de verilen Poisson Regresyonu Sonuçları, Y=75 ile sağdan kırpılmış poisson regresyonu ile elde edilen sonuçlarla aynı çıktığı görülmektedir. Tablo.4'de ise yukarıda hesaplamaları yapılan analiz sonuçlarını karşılaştırmak için her modelden elde edilen log-likelihood değerleri verilmiştir. Buna göre tüm modellerin seçim kriterlerinin aynı olduğu, tahmin sonuçları bakımından aralarında fark olmadıkları söylenebilir.

Tablo.4 Birinci veri için Model karşılaştırma değerleri

Yöntem	*Log Likelihood*
Poisson Regresyon	-6908,9611
Kırpılmış Poisson Regresyon	-6908,9611
Y=75 Sağdan Kırp.Poisson Reg.	-6908,9611

2.Veri setinde trafik kaza sayılarının belirleyicileri araştırılmıştır. Bağımsız değişken olarak kaza yapan kadın sürücü sayısı, karayolundaki buzlanma durumu, havanın yağış miktarı ve sürücünün ehliyetli olarak araba kullanma yılı alınmıştır. Veri seti 120 gözlemden oluşmaktadır. Tablo.5'de 1.veri seti için kırpılmış poisson regresyon sonuçları verilmiştir.

Tablo 5. İkinci veri seti için kırpılmış poisson regresyonu parametre tahmin sonuçları

Bağımlı Değişken: Kaza Sayısı			
	Katsayı	*Std. Hata*	*p*
Kadın Sürücü	0,2541357	0,0268678	0,000
Buzlanma	0,3525811	0,1549728	0,023
Yağış	0,0812222	0,0490753	0,098
Ehliyet	0,0010086	0,0078509	0,898
Sabit	-0,363825	0,2845958	0,201

Tablo.5'deki analiz sonuçlarına göre Kaza yapan kadın sürücü sayısı ve buzlanma katsayıları anlamlı bulunmuştur ($p<0,05$). Tablo.6'da ise kaza sayısı üst limit değeri 12 (Y=12) olmak üzere sağdan kırpılmış poisson regresyon analizi sonuçları verilmiştir.

Tablo 6. Y=12 için kırpılmış poisson regresyonu parametre tahmin sonuçları

Bağımlı Değişken: Kaza Sayısı			
	Katsayı	*Std. Hata*	*p*
Kadın Sürücü	0,3388595	0,0346714	0,000
Buzlanma	0,2648092	0,1353393	0,050
Yağış	0,138236	0,0516932	0,011
Ehliyet	-0,0031997	0,0077323	0,679
Sabit	-0,3840296	0,2618801	0,143

Tablo 4.10'daki analiz sonuçlarına göre Kaza yapan kadın sürücüler, buzlanma ve yağış katsayıları anlamlı bulunmuştur ($p<0,05$). Ehliyet katsayısı ise Tablo.5'deki sonuçlara göre bağımlı değişkene

pozitif yönde etki ederken, Y=12 için kırpılmış regresyon sonucuna göre ters yönde etki ettiği görülmektedir. Tablo.7'de ise kaza sayısı üst limit değeri 13 (Y=13) olmak üzere sağdan kırpılmış poisson regresyon analizi sonuçları verilmiştir.

Tablo 7. Y=13 için Kırpılmış Poisson Regresyonu parametre tahmin sonuçları

Bağımlı Değişken: Kaza Sayısı			
	Katsayı	*Std. Hata*	*p*
Kadın Sürücü	0,3051441	0,0320429	0,000
Buzlanma	0,2587887	0,134121	0,054
Yağış	0,1130618	0,0494358	0,022
Ehliyet	-0,0022542	0,0075217	0,764
Sabit	-0,2692183	0,2538042	0,289

Tablo.7'deki analiz sonuçlarına göre kaza yapan kadın sürücü sayısı ve yağış katsayıları anlamlı bulunmuştur ($p<0,05$). Buzlanma değişkeni bir önceki sonuca göre istatistiksel olarak anlamsız çıkmıştır. Tablo.8'de aynı veri seti için poisson regresyon sonuçları verilmiştir.

Tablo 8. İkinci veri seti için poisson regresyonu parametre tahmin sonuçları

Bağımlı Değişken: Kaza Sayısı			
	Katsayı	*Std. Hata*	*p*
Kadın Sürücü	0,235541	0,0249888	0,000
Buzlanma	0,2696443	0,1331659	0,043
Yağış	0,0730315	0,0461197	0,113
Ehliyet	-0,0002633	0,0071511	0,971
Sabit	-0,0516263	0,2412063	0,831

Tablo.8'deki analiz sonuçlarına göre kaza yapan kadın sürücüler ve buzlanma katsayıları anlamlı bulunmuştur ($p<0,05$). 2.veri seti için yapılan tahmin sonuçları Tablo.9'da verilen log-likelihood değerleri ile karşılaştırılmıştır.

Tablo 9. İkinci veri seti için model karşılaştırma değerleri

Yöntem	Log Likelihood
Poisson Regresyon Analizi	-207,71766
Kırpılmış Poisson Regresyon Analizi	-197,74791
Y=12 için Sağdan Kırp. Poisson Reg.	-191,49098
Y=13 için Sağdan Kırp. Poisson Reg.	-197,49715

Tablo.9'daki sonuçlara göre mutlak değerce minimum olan Y=12 için Kırpılmış Poisson Regresyon, trafik kazası verisine göre en iyi sonuçları vermiştir. Kırpılmış Poisson Regresyon sonuçlarının Poisson Regresyonuna göre daha iyi olduğu görülmektedir.

3.Veri seti; Akciğer kanseri teşhisi konmuş 435 hastaya aittir. Bağımlı değişken olarak hastanın durumu (yaşıyor, vefat etmiş, haber alınamıyor), takip süreci (ay), teşhis konma süresi (yıl) ve hastanın yaşı değişkenleri kullanılmıştır. Tablo.10'da 3. veri seti için Kırpılmış Poisson Regresyon Analizi sonuçları verilmiştir.

Tablo 10. Üçüncü veri seti için kırpılmış poisson regresyonu parametre tahmin sonuçları

Bağımlı Değişken: Durum (Gün)			
	Katsayı	*Std. Hata*	*p*
Takip	0,1077154	0,0078407	0,000
Teşhis	-1,288745	0,0966472	0,000
Yaş	-0,0113409	0,0050292	0,024
Sabit	4,5006	0,4097263	0,000

Tablo.10'daki analiz sonuçlarına göre tüm katsayılar anlamlı bulunmuştur ($p<0,05$). Teşhis konma süresi ve hastanın yaşı değişkenleri, hastanın durumuna beklenildiği gibi ters yönde etki ettikleri görülmektedir. Tablo.11'de akciğer kanserinde üst limit değeri 4 (Y=4) olmak üzere sağdan Kırpılmış Poisson Regresyon Analizi sonuçları verilmiştir.

Tablo 11. Y=4 için Kırpılmış Poisson Regresyonu parametre tahmin sonuçları

Bağımlı Değişken: Durum (Gün)			
	Katsayı	*Std. Hata*	*p*
Takip	0,0505754	0,0045003	0,000
Teşhis	-0,5035377	0,041578	0,000
Yaş	-0,0111734	0,0047581	0,019
Sabit	3,213129	0,3635542	0,000

Tablo.11'deki analiz sonuçlarına göre de tüm katsayılar anlamlı bulunmuştur ($p<0,05$). Son olarak Tablo.12'de kanser verisi için Poisson Regresyonu Sonuçları verilmiştir.

Tablo 12. Üçüncü veri seti için poisson regresyonu parametre tahmin sonuçları

Bağımlı Değişken: Durum (Gün)			
	Katsayı	*Std. Hata*	*p*
Takip	0,0279125	0,002844	0,000
Teşhis	-0,304516	0,0310039	0,000
Yaş	-0,005449	0,0033517	0,104
Sabit	1,851292	0,2441593	0,000

Tablo.12'deki sonuçlarına bakıldığında yaş değişkenin istatistiksel olarak anlamlı olmadığı diğer değişkenlerin ise anlamlı oldukları görülmektedir. 3.veri seti için yapılan tahmin sonuçları Tablo.13'de verilen log-likelihood değerleri ile karşılaştırılmıştır.

Tablo 13. Üçüncü veri için model karşılaştırma değerleri

Yöntem	Log Likelihood
Poisson Regresyon Analizi	-684,53859
Kırpılmış Poisson Reg. Analizi	-401,44897
Y=4 için Sağdan Kırp. Poisson Reg.	-575,92031

Tablo.13'deki sonuçlara göre mutlak değerce minimum olan Y=4 için Kırpılmış Poisson Regresyon, kanser verisine göre en iyi sonuçları vermiştir. Tüm kırpılmış Poisson Regresyon sonuçlarının Poisson Regresyonuna göre daha iyi olduğu görülmektedir.

4 Sonuç

Regresyon analizi, istatistiksel tahmin yöntemlerinin en önemlisi ve belki de en çok tercih edilen yöntemidir. Amaç; eldeki verileri temsil eden ve anlamlı katsayılara sahip bir model kurmak, verinin gelecek değerlerini başarılı bir şekilde tahmin etmektir. Regresyon analizi, kullanım amacına, verinin yapısına ve uygulama türüne göre çeşitlilik göstermektedir. Poisson regresyon analizi, sayımla elde edilen bağımlı değişken ile bağımsız değişkenler arasındaki bağlantıyı çözümleyen bir analiz metodudur. Poisson Regresyon modeli, ölümle sonuçlanan kaza analizlerinde, evli çiftlerin çeşitli sebeplerden dolayı boşanma oranları ile ilgili analizlerde ve akademik alanda erkek ve kadın akademisyenlerin unvan açısında yükselmeleri dahil olmak üzere insanların hayatlarına dair sosyal, ekonomik, sağlık ve eğitim gibi pek çok alanda kullanılmaktadır.

Poisson Regresyon Analizinin farklı veri yapılarında kullanılmak üzere farklı çeşitleri vardır. Bunlardan biri de aşırı değere sahip verilerde başarılı sonuçlar verdiği düşünülen Kırpılmış Poisson Regresyon Analizidir. Burada bahsedilen aşırı değer kavramının, klasik regresyon için kullanılan kavramdan biraz farklı olduğunu belirtmek gerekir. Sağlık verileri ile yapılan çalışmalarda (özellikle kanser çalışmalarında) 1 yaş bile aşırı değer olarak sayılabilir. Bir kanser türünde hastaların 73 yaş ile 74 yaş arasındaki hastalığa karşı gösterilen semptomların, hastalığın 3. evresinden 4. evresine geçiş kadar belirgin olabilir. Bu yüzden bu çalışmada belirlenen bazı sınırlar ile yapılan analizlerde aşırı değer kavramı kullanılmıştır.

Bu çalışmada Kırpılmış Poisson Regresyon Analizi tanıtılmış ve uygulama bölümünde klasik Poisson Regresyon Analizi ile karşılaştırılmıştır. Gerçek veri tipinde üç farklı veri seti üzerinde yapılan uygulama sonuçlarına göre Kırpılmış Poisson Regresyon yöntemiyle elde edilen sonuçların klasik Poisson Regresyonuna göre daha başarılı sonuçlar verdiğini söyleyebiliriz. Aşırı değerler içeren bağımlı değişken yapılarında veya bağımlı değişkenin bazı değerlerinin altı veya üstünün değerlerinde sonuçların değişip değişmediğinin arandığı durumlarda kırpılmış poisson regresyon analizinin kullanılması önerilmektedir. Böylece klasik poisson regresyon analizinde bağımlı değişkenin aşırı uç değerlerden veya belirlenen limitlerde yanlı sonuçlar vermesinin de önüne geçilebilir.

Kaynaklar

[1] Gujarati DN. *Temel Ekonometri*, Çev. Ümit Şenesen ve Gülay Günlük Şenesen, Literatür Yayıncılık, İstanbul, 1999.

[2] Çokluk Ö. "Lojistik Regresyon Analizi: Kavram ve Uygulama". *Educational Scienses: Theory & Practice* 10 (3) Summer, s: 1357-1407, 2010.

[3] Dinarcan GN. Sayma Verisi için Regresyon Modelleri ve Bir Uygulama. Yayınlanmamış Yüksek Lisans Tezi, İstatistik Ana Bilim Dalı, Fen Bilimleri Enstitüsü, Hacettepe Üniversitesi, s.9, Ankara, 2018.

[4] Nelder JA, Wedderburn RWM. "Generalized Linear Models". *Journal of the Royal Statistical Society Series A*, 135, 3, 370-384, 1972.

[5] McCullagh P, Nelder JA. *Generalized Linear Models*, Chapman and Hall, London, 511, UK, 1989.

[6] Cameron AC, Trivedi P. "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators," *Journal of Applied Econometrics*, 1, 29–53, 1986.

[7] Frome EL, Kutner MH, Beauchamp JJ. "Regression Analysis of Poisson Distributed Data", *Journal of American Statistical Association*, 68, 935- 940, 1973.

[8] Frome EL. "The Analysis of Rates using Poisson Regression Models". *Biometrics*, 39, 665-674, 1983.

[9] Cameron AC, Trivedi P. *Regression Analysis of Count Data*, Cambridge University Press, West Nyack, NY, USA, 1998.

[10] Açıkyürek G. Poisson Regresyon ve bir uygulama. Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Ankara, 2016.

[11] Xie T, Aickin M. "A truncated poisson regression model with applications to occurrence of adenomatous polyps". *Statistics in medicine*, vol. 16, 1845-1857, 1997.

[12] Suaiee AMA. Double Truncated Poisson Regression Model with Random Effects. Dissertations, paper 260, University of Northern Colorado, USA, 2013.

[13] Moore PG. "A Note on Truncated Poisson Distributions". Biometrics, 10, p.402-406, 1954.

[14] Cameron AC, Triverdi PK. *Regression Analysis of Count Data*. Cambridge. UK. Cambridge University Press.

[15] Grogger JT, Carson RT. "Models for Truncated Counts". *Journal of Applied Econometrics*, v.6, no:3, p. 225-238, 1991.

[16] Fokianos K. "Truncated Poisson Regression for Time Series of Counts". *Scand. J. Statist.*, 28, 2001.

[17] van der HeijdenPGM, Bustami R, Cruyff MJLF, Engbersen G, van Houwelingen HC. "Point and Interval Estimation of the Population Size Using the Truncated Poisson Regression Model". Statistical Modelling, 3, p. 305-322, 2003.

[18] Gurmu S, Triverdi PK. "Overdispersion Tests for Truncated Poisson Regression Models". *Journal of Econometrics,* 54, p 347-370, 1992.

[19] Yeh HW, Gajewski B, Mukhopadhyay P, Behbod F. "The Zero-Truncated Poisson With Right Censoring: An Application to Translational Breast Cancer Research". *Statistics in Biopharmaceutical Research,* 4:3, 252-263, 2012.

[20] Puza BD, Johnson HL, O'neill TJ, Barry SC. "Bayesian Truncated Poisson Regression with Application to Dutch Illegal Immigrant Data". *Communications in Statistics - Simulation and Computation*, 37: 1565–1577, 2008.

[21] Finney DJ, Varley GC. "An Example of the Truncated Poisson Distribution". *Biometrics*, Vol. 11, No. 3, 1955.

[22] Van der Heijden PGM, Cruyff M, Van Houwelingen HC. "Estimating the Size of a Criminal Population from Police Records Using the Truncated Poisson Regression Model". *Statistica Neerlandica*, Vol. 57, nr. 3, pp. 289–304, 2003.

[23] Adesina OS, Agunbiade DA, Oguntunde PE, Adesina TF. "Bayesian Models for Zero Truncated Count Data". *Asian Journal of Probability and Statistics*, 4(1), p. 1-12, 2019.

[24] SAS. *SAS/Stat, Software*, Hangen and Enhanced, SAS, Institute, Incorporation, USA, 2007.

[25] Tamar, M. Poisson Regresyonu, Fırat Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Yayınlanmamış Yüksek Lisans Tezi, s.19., 2013.

[26] Zorlutuna Ş, Erilli NA, Yücel B. "Tobit Regresyon Analizi ile Akciğer Kanseri Çalışması: Sivas İli Örneği". *Eurasian Econometrics, Statistics and Emprical Economics Journal*, s.13-22, 2016.