Black Sea Journal of Engineering and Science
Open Access Journal
e-ISSN: 2619-8991

BSPublishers

Araştırma Makalesi (Research Article)
Cilt 1 - Sayı 4: 140-146 / Ekim 2018
(Volume 1 - Issue 4: 140-146 / October 2018)

POİSSON REGRESYON TAHMİN YÖNTEMLERİNİN KARŞILAŞTIRILMASI

Müslüme MEMİŞ[1], Hasan ÖNDER[2*]

[1]Doğubeyazıt İlçe Tarım Müdürlüğü, Doğubeyazıt, Ağrı, Türkiye

[2] Ondokuz Mayıs Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, 55139 Samsun, Türkiye

Gönderi: *06 Temmuz 2018;* ***Kabul:*** *13 Eylül 2018;* ***Yayınlanma:*** *01 Ekim 2018*
*(**Received:** July 06, 2018; **Accepted:** September 13, 2018; **Published:** October 01, 2018)*

Özet

Çoğu bilimsel çalışmanın amacı bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi matematiksek modellerle açıklayarak, bu modellerin kullanılması ile geleceğe yönelik tahminler elde etmektir. Sayıma dayalı olarak elde edilen verilerin analizinde Poisson regresyon modeli pek çok alanda yaygın olarak kullanılmaktadır. Bu çalışma, Poisson regresyon analizinde tahmin yöntemlerinden; Poisson en çok olabilirlik tahmin yöntemi ve genelleştirilmiş doğrusal modeller tahmin yöntemlerinin karşılaştırılarak hangi yöntemin daha uygun olduğu konusunda araştırmacılara yol göstermek amacıyla yapılmıştır. Yöntemlerin karşılaştırılması için 100, 500 ve 1000 örnek büyüklüklerinde yapay veri kullanılmıştır. Yapılan bu çalışmada sonuç olarak kullanılan parametre tahmin yöntemleri arasında uyum iyiliği bakımından farklılık olmadığı tespit edilmiştir. Ancak, en çok olabilirlik tahmin edicisinin ürettiği standart hata değerlerinin daha yüksek olmasından dolayı genelleştirilmiş doğrusal modeller yöntemin daha güvenilir olduğu ve küçük örnek büyüklüklerinde de daha güvenilir tahmin yapabildiği bulunmuştur. Sonuç olarak, Poisson regresyon analizinde genelleştirilmiş doğrusal modeller yönteminin kullanılması önerilmiştir.

Anahtar sözcükler: Poisson regresyon, Genelleştirilmiş doğrusal modeller, En çok olabilirlik

Comparison of Poisson regression estimation methods

Abstract: The aim of many scientific studies is to explain relationships between response variable and explanatory variables with mathematical models and to acquire prudential predictions with these models. Poisson regression models are commonly used for analyzing the data based on counting processes. This study aimed to guide the researchers for determining appropriate Poisson regression estimation method (Poisson Maximum Likelihood and Generalized Linear Model). In comparison of methods, artificial data were used with sample size of 100, 500 and 1000. It was concluded that there were no differences among parameter estimation methods in terms of goodness of fit. However, it was detected that generalized linear models method was more reliable than maximum likelihood method because maximum likelihood estimator produced high standard error for the parameters. In addition, generalized linear models were more reliable for small sample sizes because of estimated lover standard errors. As a result, it was suggested that generalized linear models should be used in Poisson regression analysis.

Keywords: Poisson regression, Generalized linear models, Maximum likelihood

***Corresponding author:** Ondokuz Mayıs Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, 55139 Samsun, Türkiye
Email: hasanonder@gmail.com (H. ÖNDER)

1. Giriş

Biyoloji, tıp, ekonomi, fizik, kimya ve sosyal bilimler gibi birçok alanda yaygın olarak kullanılmakta olan regresyon analizi, aralarında sebep – sonuç ilişkisi bulunan iki veya daha fazla değişken arasındaki ilişkiyi inceleyen ve bu ilişkiyi modellemek için kullanılan istatistiksel analiz yöntemidir (Vural, 2007). Regresyon analizinde incelenen değişkenler sürekli ya da kesikli yapıda olabilmektedir, veri yapısına bağlı olarak farklı regresyon modellerinin kullanılması gerekmektedir (Özarıcı, 1996).Regresyon analizinin asıl amacı yanıt değişkeni ile açıklayıcı değişkenler arasındaki ilişkiyi matematiksek modellerle açıklayarak, bu modellerin kullanılması ile geleceğe yönelik tahminler elde etmektir. Elde edilen modelin doğru bir şekilde oluşturulması, analiz ve yorumların geçerliliğinde etkin bir rol oynamaktadır. Bu konuda geliştirilmiş birçok istatistiksel yöntem söz konusudur (Karadavut ve ark., 2005) ve kullanılacak yöntemin seçiminde temel ölçüt değişkenlerin sahip olduğu veri yapısıdır.Sayıma dayalı olarak elde edilen verilerin analizinde Poisson regresyon modeli pek çok alanda yaygın olarak kullanılmaktadır. Poisson regresyon analizi, açıklayıcı değişkenler ile sayıma dayalı olarak elde edilen yanıt değişkeni arasındaki ilişkiyi açıklamaktadır. Poisson regresyonunda açıklayıcı değişkenlerin doğrusal yapısını yanıt değişkeninin beklenen değerine bağlayan bağlantı fonksiyonu logaritmik dönüşüm ile ifade edilmektedir (Frome, 1983).Araştırmadan elde edilen verilerin ölçeğinin sürekli yapıda olmadığı, başka bir ifadeyle kategorik veriye sahip olunduğunda doğrusal regresyon modelleri kullanmak yapılacak analizleri etkisiz, tutarsız ve güvenilmez sonuçlar verebilir. (Arı ve Önder, 2013).Bazı çalışmalarda verilerin kesikli olması durumunda doğrusal regresyon kullanılarak yapılacak analizler iki açıdan sorun oluşturmaktadır. Birincisi kuramsal olarak mümkün olmayan negatif parametre tahmininin elde edilmesi, ikincisi ise çoğu değerlerin sıfır olmasından dolayı dağılışın sağa çarpık olmasıdır (Frome ve ark., 1973; Cox, 1983; SAS, 2005).

Bu çalışma, Poisson regresyon analizinde tahmin yöntemlerinden; Poisson en çok olabilirlik tahmin yöntemi (ML) ve genelleştirilmiş doğrusal modeller tahmin yöntemlerinin (GLM) karşılaştırılarak hangi yöntemin daha uygun olduğu konusunda araştırmacılara yol göstermek amacıyla yapılmıştır.

2. Materyal ve Metot

2.2. Materyal

Çalışmada kullanılan veriler MINITAB 12.0 yazımı kullanılarak 100, 500 ve 1000 örnek büyüklüğünde yapay olarak üretilmiştir. Elde edilen veri Kolmogorow-Smirnow tek örnek testi ile tekrar analiz edilmiş ve Poisson dağılışına uygunluğu test edilmiştir. Analiz sonuçlarına göre elde edilen I. tip hata değerleri örnek büyüklükleri için sırasıyla 0.997, 0.965 ve 0.975 olarak elde edilmiş olup, üretilen verinin Poisson dağılışı gösterdiği anlaşılmıştır.

2.2. Metot

2.2.1. Poisson regresyon

Poisson regresyon, incelenmek istenen zaman periyodunda meydana gelmiş olan olaylarda belirlenen açıklayıcı değişkenlerdeki problemlerde uygulanmaktadır. Model; süreksiz ve negatif olmayan, sayılabilir verileri içermelerinden dolayı, beklenen sayıların logaritmasının açıklayıcı değişkenlerin doğrusal bir fonksiyonu olduğu varsayımından yola çıkmaktadır. Özel açıklayıcı değişkenlerin regresyon katsayılarını, diğer tüm açıklayıcı değişkenlerin sabit tutulduğu durumda açıklayıcı değişkendeki bir birimin artışından önce ve sonra beklenen sayıların oranının logaritması olarak yorumlanabilir (Köleoğlu, 2006).Yanıt değişkenin 0, 1, 2,…,n gibi kesikli değer aldığı fakat kategorik olmadığı durumlar vardır. Kesikli ve kategorik olmayan, nadir olaylarla ilişkili yanıt değişkenli model, bazı varsayımlar altında Poisson regresyon modeli olarak adlandırılır. Bu model daha çok sayma verilerini analiz etmek için kullanılır (Akın, 2002).Model üstel bir model olması nedeniyle katsayı yorumlamalarında zorluk ve karmaşıklık yaratması dezavantaj olmasına rağmen, yanıt değişkenin sayma verilerinden oluştuğu durumlarda doğrusal regresyon analizine alternatif olabilen bir modeldir (Deniz, 2005). Poisson regresyon modeli; sayılabilir verilerin ortalamasını içeren çarpımsal modellerde, açıklayıcı değişkenler arasında şartlı veya marjinal bir bağımlılık olduğunda esnek bir model olmasından dolayı araştırmacılara kolaylık sağlayan bir model özelliği taşımaktadır (Lloyd, 1999). Poisson regresyon, lojistik regresyondan sonra en genel olan ikinci genelleştirilmiş modeldir. Bu modelinin en belirgin özelliği, ortalama ile varyansın birbirine eşit olmasıdır. Ancak çoğu uygulamalarda, bu eşitliği sağlamak mümkün olmamaktadır. Poisson dağılışında; varyansın ortalamadan büyük olması aşırı yayılım (overdispersion) ve varyansın ortalamadan küçük olması hali ise az yayılım (underdispersion) olarak bilinir. Poisson regresyonunda ilgilenen olayın sayısı olan Y bağımlı değişkenin; $x_1, x_2, ..., x_n$ bağımsız değişkenleri verilmişken Poisson dağılışına sahip olduğu varsayılmaktadır. Bu durumda Poisson ortalaması μ'nün logaritmasının bağımsız değişkenlerinin bir doğrusal fonksiyonu olduğu varsayılmaktadır (SAS,2005; Yeşilov ve ark, 2006). Söz konusu fonksiyona göre Poisson regresyon modeli;

$$\log(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X2 + \cdots + \beta_m X_m \quad (1)$$

biçiminde verilmektedir. Eşitlikte μ bağımsız değişkenlerin üstel bir fonksiyon olmaktadır. μ'yü,

$$\mu = \exp\left(\beta_0 + \beta_1 X_1 + \beta_2 X2 + \cdots + \beta_m X_m\right) \quad (2)$$

biçiminde yazabiliriz.

Poisson dağılımı belirli bir zaman, belirli bir alan veya hacimde gerçekleşebilecek olay sayısının tanımlayan bir dağılımdır. Dağılım 18. yüzyılda yaşamış olan Fransız matematikçi Poisson tarafından tanıtılmıştır. Sonraki yıllarda farklı bilim adamları tarafından üzerinde çalışılan ve genişletilerek bugünkü halini alan dağılım ilk temellerini atan ünlü matematikçinin adını taşımaktadır (Gürsakal, 1997). Poisson dağılımını aşağıda formülde verildiği gibidir;

$$P(y;\mu) = \frac{\mu^y e^{-\mu}}{y!}, \; y = 0, 1, 2, \ldots \quad (3)$$

bu eşitlikte y_i, istenen olayların meydana gelme sayısıdır ve dağılımın parametresi μ_i 'dir. Poisson dağılımının şartlı beklenen değeri $E(y_i) = \mu_i$ 'dir. μ_i'nün açıklayıcı değişkenlere bağlı olarak tanımlanması durumunda Poisson regresyon modeli elde edilmektedir. µ_i genellikle $\mu_i = e^{x'\beta}$ şeklinde tanımlanır. Burada *x* açıklayıcı değişken vektörü ve β ise tahmin edilecek parametre vektörünü göstermektedir. Poisson dağılımın ortalaması;

$$\mu_i = E(y_i / x_i) = \exp(x_i'\beta) \quad (4)$$

şeklinde gösterilir. İstatistik literatür de bu model log-doğrusal model olarak ifade edilmektedir. Poisson dağılımında ortalama-varyans birbirine eşittir.

$$\mu_i = E(y_i / x_i) = V(y_i / x_i) \quad (5)$$

Ortalama ve varyansın eşitliğine eşit yayılım denir. Ancak uygulamada sayma değişkenler genellikle ortalamadan daha büyük varyansa sahip olduklarından aşırı yayılım gösterirler. Verinin aşırı yayılım göstermesine gözlemlenen sıfır değerlerin sayısının, Poisson modeli ile ortaya konulan sıfır değerlerini aşması ve gözlenmemiş heterojenlik gibi durumlar neden olmaktadır (Kibar, 2008). Modeldeki aşırı yayılım katsayı tahminini etkilemez, fakat tahminin standart hatasını etkisi altında olmaya sebep verir, böylece modelin güvenirliğini yükseltir (AL- Ghirbal ve AL- Ghamdi,2006).

2.2.2. Genelleştirilmiş doğrusal modeller

GLM, şans değişkenleri için olabilirlik tabanlı yaklaşımları kullanarak bir regresyon modelinde parametre tahmini yapmaktadır (Breslow ve Clayton, 1993). GLM bunu yaparken ilk önce şans değişkenlerinin sahip olduğu dağılışı üssel dağılış ailesinde tanımlar. Daha sonra değişkenlerin kendisinin yerine değişkenlere ait beklenen değerlerinin fonksiyonunu kullanır. GLM'de olasılık fonksiyonu;

$$f(y;\theta,\phi) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y;\phi)\right] \quad (6)$$

şeklinde olmakradır . Burada ϕ yayılım ve θ doğal parametresi olmaktadır. *α, b* ve *c* dağılışın ne olduğunu belirlemektedir. Poisson dağılışına ait herhangi bir gözlem değerinin olasılık fonksiyonu;

$$P(y;\mu) = \frac{\mu^y e^{-\mu}}{y!}, \; y = 0, 1, 2, \ldots \quad (7)$$

biçiminde verilmektedir. Bu olasılık fonksiyonunun log olabilirlik fonksiyonu, GLM kullanılarak

$$l(\mu; y) = y\log(\mu) - \mu - \log(y!) \quad (8)$$

biçiminde verilir. Burada, terimler karşılaştırıldığında $\log(\mu) = \theta$, $\mu = b(\theta)$, $\log(y!) = c(y,\phi)$ ve $a(\phi) = 1$ olmaktadır (Donson, 1990; Littell ve ark., 1996; Okut ve ark., 1999; Yeşilova ve ark., 2006).

2.2.3. Poisson regresyon analizinde katsayıların tahmini

Poisson regresyon analizinde yanıt değişkeni y_i' nin dağılışına göre, $\hat{\beta}$ tahmin edicilerini hesaplama yöntemleri değişiklik göstermektedir. En çok olabilirlik yöntemi ve genelleştirilmiş doğrusal modeller bu yöntemlerden en sık kullanılan ve en çok bilinenleridir.

Poisson En Çok Olabilirlik tahmin yöntemi

Poisson modeli, temelde doğrusal olmayan bir regresyondur. Fakat maksimum olabilirlik teknikleri ile parametre tahmini elde etmek oldukça kolaydır. Poisson regresyon modelinin standart tahmin edicisi Maksimum Olabilirlik tahmin edicisidir. Bağımsız gözlemler için, maksimum olabilirlik fonksiyonu;

$$L(y,\beta) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} \frac{e^{-\mu_i}\mu_i^{y_i}}{(y_i)!} = \frac{\left(\prod_{i=1}^{n} \mu_i^{y_i}\right) e^{-\sum_{i=1}^{n}\mu_i}}{\prod_{i=1}^{n}(y_i)!}$$

(9)

şeklindedir. Log-olabilirlik fonksiyonu;

$$\ell(\beta) = \sum_{i=1}^{n} y_i x_i'\beta - \sum_{i=1}^{n} e^{x_i'\beta} - \sum_{i=1}^{n} \ln(y_i!) \qquad (10)$$

eşitlikteki gibi olur (Arıcan,2010). Buna bağlı olarak Poisson MLE $\hat{\beta}_p$ değeri;

$$\sum_{i=1}^{n} x_i (y_i - \exp(x_i'\beta)) = 0 \qquad (11)$$

ifadesinden hesaplanır. $\hat{\beta}$ değeri birinci derecen türev alınarak hesaplanır.

$\hat{\beta}_p$ değerini hesaplamak için analitik çözüm yoktur. Çözüm için genellikle iterasyon yöntemi kullanılır. Bu yönteme, iteratif olarak yeniden ağırlıklandırılmış en küçük kareler (IRWLS) denilmektedir. Uygulamada genellikle 10 ya da daha az iterasyon yapmak yeterlidir. Verilen bilgiler ile uygulanan modeller doğrultusunda varyans değeri için;

$$V_{ML}\left[\hat{\beta}_P\right] = \left(\sum_{i=1}^{n} \mu_i x_i x_i'\right)^{-1} \qquad (12)$$

sonucuna ulaşılır (Deniz , 2005).

Genelleştirilmiş Doğrusal Modeller tahmin yöntemi

GLM yönteminde verilerin orijinal dağılışını üslü (exponential) formda yazılarak, parametre tahminleri en çok olabilirlik (ML) yöntemleriyle elde edilmektedir. Bazı durumlarda gözlem değerleri yi normal dağılışlı olmayabilir. GLM, standart doğrusal modellerle verilerin orijinal dağılışını esas alarak ML yöntemi ile parametre tahmini yapar (Yeşilova ve ark, 2006). GLM' de gözlem değerlerinin olasılık yoğunluk fonksiyonu;

$$f(y_i, \theta_i, \phi) = \exp\left\{\left[\frac{y_i\theta_i - b(\theta_i)}{\alpha(\phi)}\right] + c(y_i, \phi)\right\} \qquad (13)$$

şeklinde olmakta *α, b* ve c dağılışın ne olduğunu belirlemektedir. Burada θ doğal ve ϕ ölçek parametresi olmaktadır (Yeşilova ve ark., 2006).Genelleştirilmiş doğrusal modeller yardımıyla hesaplanan Poisson tahmin edicisi $\hat{\beta}_{GLM}$,

$$\sum_{i=1}^{n} \frac{1}{\phi}(y_i - \exp(x_i'\beta))x_i = 0 \qquad (14)$$

ifadesinden hesaplanmaktadır (Cameron ve Trivedi, 1998), varyans ise

$$V_{GML}\left[\hat{\beta}_P\right] = \phi\left(\sum_{i=1}^{n} \mu_i x_i x_i'\right)^{-1} \qquad (15)$$

eşitliğinden hesaplanmaktadır.

2.2.4. Modelin uyum iyiliğinin sınanması

Lineer regresyon modellerinde regresyon doğrusunun verilere ne ölçüde uyumlu olduğunun ölçütü, bir veri kümesine uyarlanmış regresyon doğrusunun uyum iyiliği olarak isimlendirilebilir (Gujaratti, 1999). Parametreler tahmin edildikten sonra gözlemlerin modelin şekli etrafındaki dağılımlarının ölçülmesi gerekmektedir. Çünkü gözlemler tahmin edilen modelin çizilen şekline ne kadar yakınsa modelin uyum iyiliği de o kadar yüksek olacaktır. Bir başka deyişle, *Y*' deki değişimim açıklayıcı değişkenlerdeki değişmelerle açıklanması o kadar iyi olacaktır (Koutsoyiannis, 1989). Poisson regresyon modelinin uyum iyiliğinin sınanmasında; Pearson χ^2, sapmalar istatistiği, yapay R^2 ölçümü, Akaike Bilgi Ölçütü (AIC) ve Bayes Bilgi Ölçütü (BIC) yaygın olarak kullanılan ölçütlerdir.

Pearson istatistiği

μ_i ortalamalı ve ω_i varyanslı yanıt değişkeni y_i' ye ait herhangi bir model için standart uyum iyiliği ölçüm yöntemi Pearson istatistiğidir ve

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu})^2}{\hat{\omega}_i} \qquad (16)$$

olarak ifade edilir. Bu değer serinin yayılımının aşırı olup olmadığını belirlemede kullanılır. Burada $\hat{\mu}_i$ ve $\hat{\omega}_i$ değerleri μ_i ve ω_i'nin tahmin değerleridir. Hesaplanan χ^2 değeri, $\hat{\mu}$ için belirlenmiş serbestlik derecesi (*n-k*) ile karşılaştırılır. Bu formül Poisson regresyon için uygulandığında, $\omega_i = \mu_i$ olacaktır ve

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu})^2}{\hat{\mu}_i} \qquad (17)$$

şeklini alacaktır. Hesaplanan χ^2 değerinin serbestlik derecesine oranının 1'den büyük değer alması verilerin modele uygun olmadığını ve aşırı yayılım durumunun varlığını ifade etmektedir (Deniz , 2005).

Sapma istatistiği

Uyum iyiliğinin ölçülmesinde kullanılan diğer bir teknik de sapma istatistiğidir. Bu istatistik değerine aynı zamanda "G2 istatistiği" de denilmektedir ve

$$G^2 = 2\sum_{i=1}^{n} y_i \ln\left(\frac{y_i}{\mu_i}\right) \qquad (18)$$

şeklinde ifade edilmektedir. Bu istatistik değeri 0'a yakınsıyor ise model uyumu artıyor denilebilir. Eğer bu istatistik değeri tam 0'a eşit ise model uyumunun mükemmel olduğu söylenebilir (Deniz , 2005).

Akaike bilgi ölçütü

Akaike tarafından önerilen ve farklı modellerin karşılaştırılmasında yaygın olarak kullanılan ölçüt, Akaike bilgi ölçütü (Akaike Information Criterion: AIC) olarak tanımlanır. Akaike bilgi ölçütü;

$$AIC = -2\log L + 2q \tag{19}$$

şeklinde ifade edilmektedir. Bu eşitlikte L; log olabilirlik fonksiyonunun maksimum değerini, q; açıklayıcı değişken sayısını göstermektedir. En küçük AIC değerine sahip model en iyi model olarak kabul edilmektedir (Ercanlı ve ark, 2012).

Bayes bilgi ölçütü

Bayes bilgi ölçütü Akaike bilgi ölçütü gibi veriler ve model arasında uygunluğu ölçen yöntemlerden biridir. Bayes bilgi ölçütü;

$$BIC = -2\ln(L) + q\ln(N) \tag{20}$$

şeklinde gibi ifade edilmektedir. Eşitlikte L; log olabilirlik fonksiyonunun maksimum değerini, q; açıklayıcı değişken sayısını N; örnek büyüklüğünü göstermektedir (Ercanlı ve ark, 2012).

3. Bulgular ve Tartışma

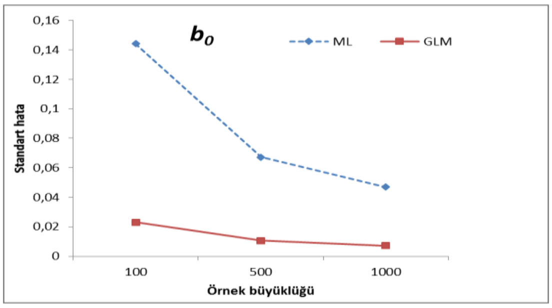
Yapay veri kullanılarak 100, 500 ve 1000 örnek büyüklükleri için elde edilen analiz sonuçlarına göre en çok olabilirlik (ML) ve genelleştirilmiş doğrusal modeller (GLM) yöntemlerinden üretmiş olduğu regresyon sabiti, regresyon katsayısı ve bunlara ait standart hata değerleri ve ilgili katsayılara ait önem testi istatistikleri Tablo 1'de verilmiştir.

Tablo 1. Farklı örnek büyüklükleri için Poisson regresyon yöntemlerinden elde edilen analiz sonuçları

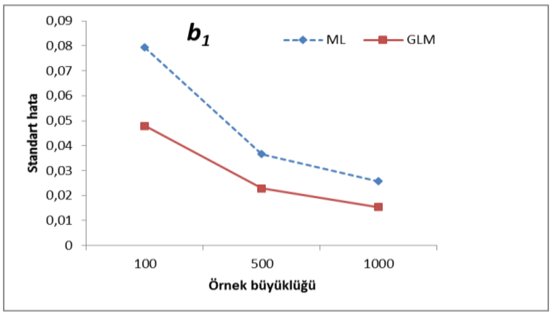
Yöntemler	Örnek büyüklüğü		Katsayılar	Standart hata	χ^2	*P*
ML	N=100	b_0	1.1953	0.1441	68.77	<0.0001
		b_1	0.5820	0.0793	53.90	<0.001
	N=500	b_0	1.2061	0.0673	321.46	<0.0001
		b_1	0.5730	0.0366	244.88	<0.0001
	N=1000	b_0	1.1756	0.0470	625.71	<0.0001
		b_1	0.5901	0.0257	528.80	<0.0001
GLM	N=100	b_0	2.3590	0.0231	10472.9680	<0.0001
		b_1	0.5820	0.0479	147.403000	<0.0001
	N=500	b_0	2.3520	0.0105	50488.4750	<0.0001
		b_1	0.5730	0.0229	623.720000	<0.0001
	N=1000	b_0	2.3560	0.0072	108309.602	<0.0001
		b_1	0.5900	0.0154	1462.52700	<0.0001

Tablo 1 incelendiğinde, Poisson en çok olabilirlik ve Poisson genelleştirilmiş doğrusal modeller yöntemlerine ait regresyon sabiti olan b0 değerlerinin sayısal farklılık gösterdiği anacak regresyon katsayısı olan b1 değerleri bakımından aynı sonuçları ürettikleri görülmektedir. Ancak parametre önem testi olan değerleri incelendiğinde GLM yönteminden elde edilen değerlerin ML yönteminden elde edilenlere göre çok daha yüksek olduğu gözlemlenmiş olup Demidenko (2007) tarafından yapılan çalışmanın sonuçlarına uyumluluk göstermektedir. Bu durum Russo ve ark. (2012) tarafından da belirtildiği şekilde GLM yönteminden elde edilen sonuçlarda I. tip hata olasılığının ML yöntemine göre çok daha düşük olduğunu göstermektedir.

Katsayılara ait standart hata değerlerinin daha iyi yorumlanabilmesi amacıyla Demidenko (2007)' nin önerdiği şekilde b_0 ve b_1 için sırasıyla standart hata değerleri Şekil 1 ve Şekil 2'de grafik olarak verilmiştir. Şekil 1 ve Şekil 2 incelendiğinde her iki yöntemde de örnek büyüklüğü arttıkça b_0 ve b_1 için standart hata değerlerinin beklendiği şekilde azaldığı gözlemlenmektedir. Her iki katsayı için de GLM yönteminden elde edilen standart hata değerlerinin tüm örnek büyüklüklerinde ML yöntemine göre daha düşük elde edildiği görülmektedir. Faria ve Soromenho (2012) ML yönteminden elde edilen standart hata değerlerinin daha yüksek bulunduğunu belirtmiş olup mevcut sonuçları destekler niteliktedir. Ayrıca Şekil 1 incelendiğinde b0 için elde edilen standart hata değerinin ML yönteminde örnek büyüklüğüne bağlı olarak hızlı bir azalış gösterdiği ancak GLM yönteminden elde edilen standart hata değerlerinin örnek büyüklğne bağlı olarak göstermiş olduğu azalış eğiminin daha düşük olduğu anlaşılmaktadır. Şekil 2. incelendiğinde b_1 için elde edilen standart hata değerlerinin her iki yöntemde de örnek büyüklüğüne bağlı değişiminin benzer olduğu söylenebilir ki bu durum GLM yönteminin küçük örnek büyüklüklerinde de daha güvenilir tahminler yapabildiğini göstermektedir. Bunun sebebi, Oral (2011)'in yapmış olduğu çalışmada belirttiği şekilde GLM yönteminin kanonik bağlantı fonksiyonlarını kullanmasından dolayı daha güvenilir olmasından kaynaklanıyor olabilir ki GLM yöntemi temelde ML yöntemini esas almakta ve iterasyon ile sonuca ulaşmaktadır (Bolker ve ark., 2009). Büyük örnek durumunda ML ve GLM yönteminden elde edilen Poisson regresyon parametrelerinin standart hatalarının yakınlaşması Min (2005) tarafından yöntemlerin asimptotik özelliklerine atfedilmiştir. Russo ve ark. (2012) her iki yöntemin de büyük örnek durumunda aynı sonuçları verdiği ancak GLM yönteminin daha güvenilir olduğunu belirtmiş olup, mevcut araştırma bulgularını desteklemektedir.



Şekil 1. Farklı örnek büyüklüklerinde b_0 için ML ve GLM yöntemlerinden elde edilen standart hata değerleri.



Şekil 2. Farklı örnek büyüklüklerinde b_1 için ML ve GLM yöntemlerinden elde edilen standart hata değerleri.

ML ve GLM yöntemlerinin karşılaştırılabilmesi için kullanılan model uyum iyiliği istatistikleri Tablo 2'de verilmiştir. Regresyon modelinin uyum iyiliğinin sınanması için sapma istatistiği, Pearson χ^2, AIC ve BIC ölçütlerin sonuçları değerlendirilmiştir. Her iki yöntemde de elde edilen yayılım parametlerinin 0.3 civarlarında değerlere sahip olduğu ve bu durumda Poisson regresyon analizinin uygun bir yöntem olduğunun söylenebileceği anlaşılmaktadır. AIC ve BIC ölçütleri incelendiğinde farklı örnek büyüklüklerinde iki yöntem arasında uyum iyiliği bakımından bir farklılığın olmadığı gözlemlenmektedir. Her iki yöntemin ürettiği standart hata değerleri yorumlandığında Wang ve Fameo (1997)' nin de belirttiği şekilde GLM yönteminin önerilebileceği söylenebilir.

3. Sonuç ve Öneriler

Yürütülen bu çalışmanın sonuçları incelendiğinde tüm örnek büyüklüklerinde b1 katsayıların tahmininde en çok olabilirlik ve genelleştirilmiş doğrusal modeller yöntemlerinden elde edilen değerlerin aynı olduğu görülmektedir. AIC ve BIC değerlerine göre modeller karşılaştırıldığında iki yöntemin uyum iyiliği bakımından benzer olduğu söylenebilir. Ancak GLM yönteminden elde edilen katsayılara ait standart hata değerlerinin daha düşük olması ve bu yöntemde küçük örnek durumunda da daha düşük standart hataya sahip tahminler yapabiliyor olması GLM yönteminin sayılarak elde edilen yanıt değişkeni varlığında Poisson regresyon analizi için önerilebileceğini göstermektedir.

İleride bu konu ile yapılacak çalışmalarda ölçüm hatalarının etkileri ve aşırı ya da az yayılım durumlarının değerlendirilmesinin yararlı olacağı söylenebilir.

Tablo 2. Farklı örnek büyüklükleri için model uyum iyiliği sonuçları

Yöntem	Örnek Büyüklüğü	Sapma Yayılım Parametresi(ϕ)	Pearson χ^2 Yayılım Parametresi(ϕ)	AIC	BIC
ML	N=100	0.3685	0.3657	439.30680	444.5172
	N=500	0.3856	0.3926	2198.8691	2207.2983
	N=1000	0.3587	0.3616	4360.0526	4369.0646
GLM	N=100	0.368	0.366	439.307	444.517
	N=500	0.386	0.393	2199.3	2207.3
	N=1000	0.359	0.362	4360.3	4370.3

References

Akın F. 2002. Kalitatif tercih modelleri analizi, Bursa, Ekin Kitabevi.

Al-Ghirbal AS, Al-Ghamdi AS. 2006. Predecting severe accidents rates at roundabouts using Poisson distribution. TRB Annual Meeting, TRB Paper 06-1684.

Arı A, Önder H. 2013. Farklı Veri Yapılarında Kullanılabilecek Regresyon Yöntemleri, OMÜ Anad Tar Bilim Derg, 28(3):168-174.

Arıcan E. 2010. Nitel yanıt değişkene sahip regresyon

modellerinde tahmin yöntemleri. ÇÜ Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi.
Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White JS. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol, 24(3): 127 – 135.
Breslow NE, Clayton DG. 1993. Aproximate inference generalized linear mixed models. JASA, 88 (421): 9-25.
Cameron AC, Trivedi PK. 1998. Regression analysis of count data. Cambridge University Press. s. 411, UK.
Cox R. 1983. Some remarks on over dispersion. Biometrika, 70: 269-274.
Demidenko E. 2007. Poisson Regression for Clustered Data. Inter Stat Rev, 75(1): 96–113.
Deniz Ö. 2005. Poisson Regreyon Analizi, İstanbul Ticaret Üniv Fen Bilim Derg, 4(7): 59-72.
Dobson J. 1990. An Introduction to Generalized Models. New York: Chapman and Hall.
Ercanlı İ, Kahriman A, Yavuz H. 2012. Trabzon Orman Bölge Müdürlüğü doğu ladini-sarıçam karışık meşcereleri için karışık etkili doğrusal olmayan regresyon denklemleri ile doğu ladini çap-boy modellerinin geliştirilmesi. SDÜ Orman Fak Derg, 13: 75-84.
Faria S, Soromenho G. 2012. Comparison ofEMand SEM Algorithms in Poisson Regression Models: A Simulation Study. Commun Stat Simul Comput, 41: 497–509.
Frome ED, Kutner MH, Beauchamp JJ. 1973. Regression Analysis of Poisson-Distributed Data. J Am Stat Assoc, 68(344): 935-940.
Frome EL. 1983. The Analysis of Rates Using Poisson RegressionModels, Biometrics, 39: 665-674.
Gujaratti DF. 1999. Temel Ekonometri, Çev. Ümit Şenesen ve Gülay Günlük Şenesen, İstanbul: Literatür Yayıncılık.
Gürsakal N. 1997. Bilgisayar Uygulamalı 1, Bursa: Marmara Kitabevi.
Karadavut U, Genç A, Tozluca A, Kınacı İ, Aksoyak Ş, Palta Ç, Pekgör A. 2005. Nohut (Cicer Arietinum L.) bitkisinde verime etki eden bazı karakterlerin alternatif regresyon yöntemleriyle karşılaştırılması. Tarım Bilim Derg, 11(3): 328-333.
Kibar FT. 2008. Trafik kazaları ve Trabzon bölünmüş sahil yolu örneğinde kaza tahmin modelinin oluşturulması. Karadeniz Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi.
Koutsoyiannis A. 1989. Ekonometri Kuramı, Çev: Ümit Şenesen ve Gülay Günlük Şenesen, Ankara: Verso yayıncılık.
Köleoğlu N. 2006. Olay zamanı analizinde tesadüfi etkiler poisson regresyon modeli ile gözlemlenemeyen heterojenliğin incelenmesi. Marmara Üniversitesi, Sosyal Bilimler Enstitüsü, Doktora Tezi.
Littell CR, Milliken AG, Stroup WW, Wolfinger DR. 1996. SAS System for Mixed Models, SAS Institute Inc., Cary, NC.
Lloyd CJ. 1999. Statistical Analysis of Categorical Data, New York.
Min C. 2005. Consistency and asymptotic normality of the maximum likelihood estimator in a zero-inflated generalized Poisson regression. Sonderforschungsbereich 386, 423: 1 – 28.
Okut H, Gökdere SA, Yeşilova A. 1999. Aplication Generalized Linear Mixed Models. III. National Conference of the Italian Biometric Society, Roma.
Oral E. 2011, Parameter Estimation in Generalized Linear Models through Modified Maximum Likelihood, Bulletin of the International Statistical Institute.
Özarıcı Ö. 1996. Farklı not sistemlerinde öğrencinin başarılı olma olasılığının probit regresyon analiziyle değerlendirilmesi. Osmangazi Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi.
Russo S, Flender D. da Silva GF. 2012. Poisson Regression Models for Count Data: Use in the Number of Deaths in the Santo Angelo (Brazil). J Basic Appl Sci, 8: 266 – 269.
SAS, 2005. SAS/STAT Software: Hangen and Enhanced. SAS, Inst. Inc., USA.
Yeşilova A, Yılmaz A, Kaki B. 2006. Norduz erkek kuzularının bazı kesikli üreme davranış özelliklerinin analizinde doğrusal olmayan regresyon modellerin kullanılması, Yüzüncü Yıl Üniv, Ziraat Fak, Tarım Bilim Derg, 16(2): 87-92.
Vural A. 2007. Aykırı değerlerin regresyon modellerine etkileri ve sağlam kestiriciler. Marmara Üniversitesi, Sosyal Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul.
Wang W, Famoye F. 1997. Modeling Household Fertility Decisions with Generalized Poisson Regression. J Popul Econ, 10: 273-283.