

PAPER DETAILS

TITLE: Evaluation of Overdispersed Data Set by Using Generalized Linear Mixed Model Approach

AUTHORS: Gazel SER,Abdullah YESILOVA

PAGES: 266-273

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/222912>

Araştırma Makalesi/Research Article (Original Paper)

Aşırı Yayılımlı Veri Setinin Genelleştirilmiş Doğrusal Karışık Model Yaklaşımı Kullanılarak Değerlendirilmesi

Gazel SER*, Abdullah YEŞİLOVA

Yüzüncü Yıl Üniversitesi, Ziraat Fakültesi, Zootehni Bölümü, 65080, Van
*e-posta: gazelser@gmail.com; Tel: +90 (432)2251056 /22640

Özet: Bu çalışmada, Poisson dağılımına sahip veri setlerinde sıkça karşılaşılan varyansın ortalamadan büyük olması durumunda ortaya çıkan, aşırı yayılım sorununun incelenmesi amaçlanmıştır. Çalışmada, Türkiye İstatistik Kurumundan (TÜİK) elde edilen 2010-2015 yılları arasında, 18 ildeki 6 aydan küçük oğlak varlığına ilişkin veriler kullanılmıştır. Veri setinde, aşırı yayılım sorununun incelenmesi amacıyla genelleştirilmiş doğrusal karışık model yaklaşımında, üç model algoritması oluşturulmuştur. Çalışma, iki aşamada gerçekleştirilmiştir. İlk aşamada, veri setinde aşırı yayılım durumu oluşturulan model algoritmasında belirlenmiş, ikinci aşamada ise aşırı yayılım sorununun giderilmesi amacıyla iki farklı model algoritması kullanılmıştır. Aşırı yayılım durumundan elde edilen standart hatalar, aşırı yayılımın giderildiği duruma göre daha küçük elde edilmiştir. Bununla beraber, oğlak varlığına ilişkin olarak aşırı yayılım durumunda, yıllar arasındaki farklılık oldukça önemli ($p<0.0001$) bulunurken, aşırı yayılım olmadığı durumda, yılların etkisi önemsiz bulunmuştur. Sonuç olarak, aşırı yayılım sorununun göz ardı edilmesi durumunda elde edilen bulgular yanıltıcı olabilmektedir. Bu nedenle, genelleştirilmiş doğrusal karışık model yaklaşımında Poisson dağılımına alternatif olarak Negatif Binom dağılımının kullanılması ya da oluşturulan model algoritmalarında Poisson dağılımı varsayımı altında şansa bağlı etkilerin ilave edilmesi, aşırı yayılımın giderilmesinde etkili alternatif çözümler olarak belirlenmiştir.

Anahtar kelimeler: Aşırı yayılım, Genelleştirilmiş doğrusal karışık model, Poisson dağılımı

Evaluation of Overdispersed Data Set by Using Generalized Linear Mixed Model Approach

Abstract: The aim of this study is to investigate the problem of overdispersion frequently observed in the data sets having Poisson distribution, in which the variation is larger than the mean. Data taken Turkish Statistical Institute (TUIK) covered between 2010 and 2015 were consist of kids reared in eighteen city of Turkey. Three different model algorithm are generated in the generalized linear mixed model approach to eliminate the problem of overdispersion in the data set. The study was conducted in two steps. In the first step, we specified the model algorithm showing overdispersion case. In the second step, however, we used the two model algorithm to overcome the elimination of this overdispersion problem. The standard errors estimated in the case of overdispersion were smaller than the case where overdispersion was eliminated. Nevertheless, in case of overdispersion in the data set, it was determined that there were statistically significant differences among factors of years for population size of kid ($P<0.0001$), whereas these differences among years for population size of kids were not significant when overdispersion were eliminated statistically. Consequently, the present results showed that overdispersion in data set can led to important misunderstandings, if the case of this overdispersion that data sets is ignored. In generalized linear mixed model approach, as an alternative, the use of negative binomial distribution instead of Poisson distribution or adding the random effects under Poisson distribution assumption in model algorithms occurred presents the effective alternative solutions to overcome the overdispersion case.

Keywords: Overdispersion, Generalized linear mixed model, Poisson distribution

Giriş

Sayımla elde edilen gözlemler Poisson dağılımı göstermektedir. Bağımlı değişkeninin Poisson dağılımına sahip olması durumunda, bu dağılımın en belirgin özelliği ve aynı zamanda kısıtlayıcılığı olan populasyon ortalamasının, varyansına eşit olması varsayımı sağlanmalıdır. Ancak, uygulamada bu varsayım her zaman sağlanamamaktadır. Eğer varyans ortalamadan büyükse veri setinde aşırı yayılım (overdispersion) ya da ekstra değişim adı verilen bir sorunla karşılaşılmaktadır (Sezgin ve Deniz 2004; Yeşilova ve ark. 2006; Gbur ve ark. 2012; Stroup 2013). Sayımla elde edilen bağımlı değişkenin varyansının, sayım aralığının genişlemesine bağlı olarak artma eğilimi göstermesi, ilişkili gözlemlerin varlığı, gözlemler arasındaki heterojenlik, gözlemlerin kümelenmesi, veri setinde çok sayıda sıfır değerinin varlığı, hatalı model kullanılması veya önemli açıklayıcı değişkenlerin modelde yer almaması gibi nedenler veri setinde aşırı yayılıma neden olabilmektedir (Sezgin ve Deniz 2004; Harwood 2013; Morel 2014). Veri setinde var olan aşırı yayılımın göz ardı edilmesi durumunda, standart hataların ve önemlilik değerlerinin beklenenden daha küçük elde edilmesiyle önemsiz olan ilişkilerin önemli olması, oldukça büyük I. tip hata oranlarının ve yetersiz güven aralıklarının elde edilmesine neden olmaktadır (Harwood 2013; Stroup 2013). Dolayısıyla aşırı yayılım sorununun, bazı istatistiksel çözüm yöntemleri kullanılarak giderilmesi gerekmektedir. Bu yöntemlerden, en çok tercih edileni ise Poisson dağılımına alternatif olarak, Negatif Binom dağılımının kullanılmasıdır (Stroup 2013).

Sayımla elde edilen gözlemler arasında var olan çok büyük değişimden dolayı verilerin sahip olduğu dağılımın şekli, sağa doğru aşırı çarpık olmaktadır. Bu tip veri setlerinde, uygulanan çeşitli transformasyon (örneğin, logaritmik ya da karekök transformasyonu) yöntemleri, normallik varsayımı sağlama noktasında yetersiz kalmaktadır (Gbur ve ark. 2012). Bu nedenle, Poisson dağılımına sahip sayım verilerinin analizinde, genelleştirilmiş doğrusal model (generalized linear model, GLM) veya şansa etkilerin modele dahil edilmesi sonucunda geliştirilen genelleştirilmiş doğrusal karışık model (generalized linear mixed model, GLMM) yaklaşımları kullanılmaktadır. Bu yaklaşımlarda, bağımlı değişkenin gösterdiği dağılıma göre özelleştirilmiş bir bağlantı fonksiyonu kullanarak çözümleme yapılmaktadır. Örneğin, Poisson ya da Negatif Binom dağılımına sahip bir veri setinde *log* bağlantı fonksiyonu kullanılmaktadır. Bu modeller, açıklayıcı değişkenler ve bağımlı değişkenin beklenen değeri ya da ortalaması arasında, doğrusal bir bağlantı fonksiyonu kullanarak çözümleme yapmaktadır. Veriler orijinal ölçeğinde kalırken, ortalama model her dağılıma göre özelleştirilen farklı bir ölçekte tanımlanmaktadır. Bununla beraber, GLM ya da GLMM modellerde kullanılan bağlantı fonksiyonu ile bağımlı değişkene uygulanan transformasyon işlemi aynı değildir. Örneğin, sayımla elde edilen bir veri setinde normal dağılım varsayımının sağlanabilmesi için bağımlı değişkene uygulanan logaritmik transformasyon işlemiyle, Poisson dağılımına sahip bir veri setine uygulanacak GLM ya da GLMM modellerde kullanılan *log* bağlantı fonksiyonu aynı değildir (Işık 2011; Gbur ve ark. 2012).

Bu çalışmada, genelleştirilmiş doğrusal karışık modelde aşırı yayılımın belirlenmesi ve aşırı yayılım sorununun çözümüne yönelik olarak bazı alternatif model algoritmaları tanıtılarak, bir uygulama ile bu modellerin değerlendirilmesi amaçlanmıştır.

Materyal ve Metot

Uygulama verisi

Bu çalışmada, Türkiye İstatistik Kurumu (TÜİK) tarafından oluşturulan hayvancılık istatistiklerine ilişkin bilgilerden yararlanılmıştır. Veri kümesi, ülkemizdeki bazı illerin 2010-2015 yılları arasındaki tiftik keçisi oğlaklarının varlığına ilişkin, sayıma dayalı verilerden oluşturulmuştur (TÜİK 2016). Bu zaman aralığında, bazı yıllara ait verileri olmayan iller çalışmaya alınmamış ve toplamda 18 ile ait bilgiler kullanılmıştır.

İstatistik analiz

Genelleştirilmiş doğrusal karışık model analizleri, SAS 9.4 programında GLIMMIX prosedüründe gerçekleştirilmiştir (SAS 2014). Oluşturulan üç model algoritmasında, 6 aydan küçük oğlak (dişi-erkek) varlığına ilişkin sayım verisi bağımlı değişken, il ve yıl etkisi ise bağımsız değişkenler olarak ele alınmıştır. GLIMMIX prosedüründe aşırı yayılımın durumunun belirlenmesi ve aşırı yayılımın giderilmesine yönelik olarak oluşturulan algoritmalar için Gbur ve ark. (2012) ve Stroup (2013)

kaynaklarından yararlanılmıştır. Çalışmada, öncelikle veri setinde aşırı yayılım durumu belirlenmiş (Model 1), sonrasında aşırı yayılımın giderilmesi için iki alternatif model algoritması (Model 2 ve 3) oluşturulmuştur. Buna göre, Poisson dağılımı varsayımı altında aşırı yayılımın belirlenmesi amacıyla kullanılan ilk model algoritması (Model 1) aşağıda verilmiştir.

Model 1

```
ods rtf;  
proc glimmix data=tuik;  
class yıl il;  
model oglak= yıl/ solution dist=Poisson link=log;  
random intercept/subject=il;  
lsmeans yıl/ilink;  
run;  
ods rtf close;
```

Genelleştirilmiş doğrusal karışık modelde aşırı yayılımın belirlenmesinde, GLIMMIX’de “method=” opsiyonu kullanılarak integral yaklaşımları (laplace ve ya quadrature) tanımlandığında uyum ölçütü olarak pearson ki-kare/serbestlik derecesi elde edilmektedir. Ancak “method=” opsiyonu tanımlanmadığında, genelleştirilmiş ki- kare/serbestlik derecesi uyum ölçütü elde edilir. Bu her iki uyum istatistiğinin, 1’den büyük olması durumunda veri setinde aşırı yayılımın olduğu belirlenmektedir (Stroup 2013). “Class” ifadesinde, kategorik değişkenler (yıl ve il) tanımlanmıştır. Model ifadesinde, bağımlı ve bağımsız değişkenler tanımlanmıştır. “Solution” ifadesi, regresyon parametre tahminlerinin elde edilmesini sağlayan bir opsiyondur. “Dist=” opsiyonunda verinin dağılımı, “link=” opsiyonunda da her dağılıma göre özelleştirilen bağlantı fonksiyonu tanımlanmaktadır. “Random” ifadesi, rasgele değişkenler olarak regresyon katsayılarını tanımlar. Bu modelde sadece kesim (intercept) şansa bağlı olarak kabul edildiğinden, bu model algoritması aynı zamanda rasgele kesim noktası modelidir. “Subject” ifadesi ise tekrarlanan (ardışık) ölçümlerin alındığı değişkeni tanımlamak için kullanılmaktadır.

Aşırı yayılım sorununun çözümüne yönelik olarak oluşturulan ilk model algoritması (Model 2) aşağıda verilmiştir. Model 2’de, Model 1 algoritmasında herhangi bir değişiklik yapılmadan verilerin Negatif Binom dağılımına sahip olduğu varsayılmıştır. Negatif Binom dağılımı, Poisson dağılımında türetildiğinden her iki dağılım aynı bağlantı fonksiyonunu (link=log) kullanmaktadır.

Model 2

```
ods rtf;  
proc glimmix data=tuik;  
class yıl il;  
model oglak= yıl/ solution dist=negbin link=log;  
random intercept/subject=il;  
lsmeans yıl/ilink;  
run;  
ods rtf close;
```

Aşırı yayılımın çözümünde kullanılan ikinci model algoritmasında (Model 3) ise Poisson dağılımı varsayımı altında “random” ifadesine, kesimin (intercept) yanı sıra bir şansa bağlı terim (yıl) daha eklenerek rasgele kesim ve eğim modeli oluşturulmuştur.

Model 3

```
ods rtf;  
proc glimmix data=tuik;  
class yıl il;  
model oglak=yıl/solution dist=Poisson link=log;  
random intercept yıl/subject=il;  
lsmeans yıl/ilink;  
run;  
ods rtf close;
```

Genelleştirilmiş Doğrusal Karışık Model

Genelleştirilmiş doğrusal karışık model, bağımlı değişkenin normal dağılım göstermediği durumda kullanılan ve doğrusal karışık model ile genelleştirilmiş doğrusal modelleri birleştiren bir yöntemdir. Yöntem, genelleştirilmiş doğrusal modellerin bir uzantısı olarak geliştirilmiş olup, doğrusal tahmin edicilere sabit etkilerin yanı sıra şansa bağlı etkilerin dahil edilmesiyle oluşturulmuştur. Bilindiği üzere karışık modelde, doğrusal tahminleyicide, en az bir tane sabit etki ve en az bir tanede şansa bağlı etki yer almaktadır. Buna göre doğrusal tahminleyicinin matris formu;

$$\eta = g(E[Y | u]) = X\beta + Zu \quad (1)$$

şeklinde yazılır. Buna göre GLMM'in genel formu;

$$g(E(Y | u)) = X\beta + Zu \text{ veya } E(Y | u) = g^{-1}(X\beta + Zu) = g^{-1}(\eta) = \mu \quad (2)$$

$$E(u) = 0 \text{ ve } Var(u) = G$$

Eşitlik 2'de $g(\cdot)$, bağlantı (link) fonksiyonu ve $g^{-1}(\cdot)$ ters bağlantı fonksiyonu, Y ($n \times 1$) boyutlu cevap vektörü, X ($n \times p$) boyutlu modeldeki sabit etkilere (β) ilişkin desen matrisi, β ($p \times 1$) boyutlu sabit etkilerin katsayılarının vektörü, Z ($n \times q$) boyutlu modeldeki şansa bağlı etkilere (u) ilişkin desen matrisi, u ($q \times 1$) boyutlu şansa bağlı etkilerin katsayılar vektörüdür. Şansa bağlı etkilerin (u), ortalaması 0 ve varyans-kovaryans matrisi G olup, çok değişkenli normal dağılıma sahiptir ($u \sim MVN(0, G)$) (Işık 2011; Gbur ve ark. 2012). Poisson ve Negatif Binom dağılımlarda, ortalama (μ) ve doğrusal tahminleyici ($X\beta$) arasındaki bağlantı (link) fonksiyonu aynıdır. Her iki dağılım $\log$ bağlantı fonksiyonunu kullanmaktadır. Buna göre Poisson ya da Negatif Binom dağılıma sahip doğrusal regresyon modeli, aşağıdaki şekilde ifade edilebilir.

$$\log(\mu_i) = \sum_{i=1}^n \beta_0 + \beta_1 X_{1i} + \dots + \beta_j X_{ji} \quad (3)$$

$\log$ bağlantı fonksiyonunun önemli bir özelliği, tahmin edilen değerlerin daima pozitif olmasıdır ($\mu > 0$) (Hilbe 2014).

Aşırı yayılım durumu

Poisson dağılımının, önemli özelliklerinden biri varyansın ortalamaya eşit olmasıdır. Bu durum eşit yayılım olarak bilinmektedir ($Var(Y) = E(Y) = \mu$). Ancak, bu eşitliğin bozulması durumunda aşırı yayılım ya da az yayılım sorunuyla karşılaşılır. Veri kümesinde varyansın, ortalamadan büyük olması ($Var(Y) > \mu$) durumu veya bir modelde gözlenen varyansın, beklenen varyanstan daha büyük olması durumu aşırı yayılım (overdispersion) olarak nitelendirilir. Bununla beraber, varyansın ortalamadan çok küçük olması durumu ise ($Var(Y) < \mu$) az yayılım olarak bilinmektedir. Bu çalışma kapsamında, sadece aşırı yayılım durumu incelenmiştir. Aşırı yayılım, aynı zamanda heterojenlik ya da ilişkili gözlemler olması durumunda karşılaşılan bir sorundur (Işık 2011; Hilbe 2014). Belirli bir modelde, verideki aşırı yayılımı belirlemek için yayılım parametresi (dispersion parameter, ϕ) hesaplanır ve aşağıdaki şekilde ifade edilir.

Yayılım parametresi (ya da ölçeklendirilmiş sapma)=Sapma (Devians)/serbestlik derecesi

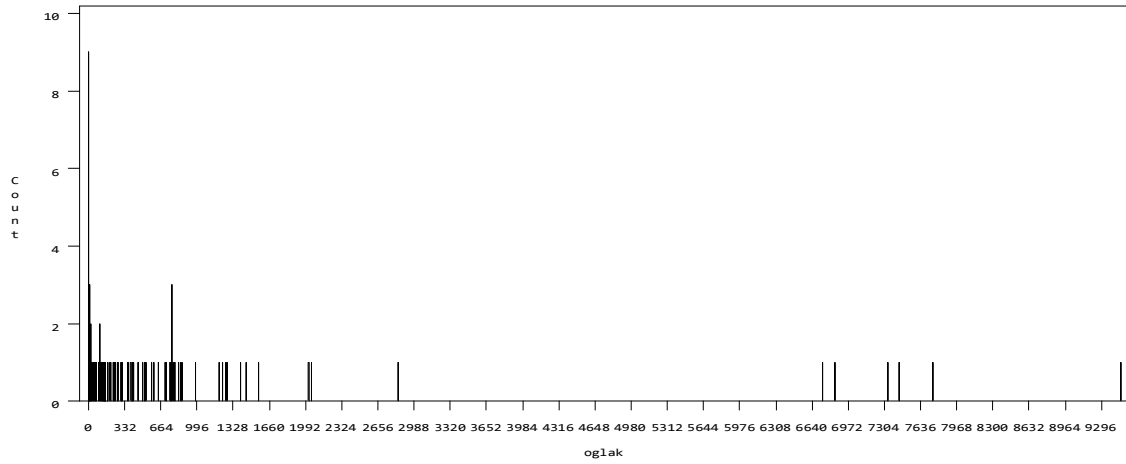
Yayılım parametresinin, 1 değerinden büyük olması aşırı yayılım ($\phi > 1$, overdispersion), küçük çıkması az yayılım ($\phi < 1$, underdispersion) ve 1 değerine eşit çıkması ($\phi = 1$) ise aşırı yayılım olmadığının bir göstergesidir (Yeşilova ve ark. 2006; Işık 2011).

Genelleştirilmiş doğrusal karışık modelde aşırı yayılımın belirlenmesi

GLMM’de aşırı yayılımın belirlenmesinde uyum istatistikleri (örneğin, genelleştirilmiş ki-kare/sd veya Pearson ki-kare/sd) hesaplanmaktadır. Bu uyum istatistiklerinin 1’den büyük olması, veride aşırı yayılımın varlığının bir kanıtıdır (Stroup 2013; Hilbe 2014). GLMM ya da GLM gibi modellerde, aşırı yayılım kontrol edilmelidir. GLM, doğrusal tahminleyici, gözlemlerin dağılımı ve bağlantı fonksiyonu olmak üzere üç bileşenden oluşur. Bunlara ek olarak GLMM’de, şansa bağlı etkiler varyans-kovaryans varsayımlarıyla ilişkilendirilmiştir. Bu modellerde, bağlantı fonksiyonunun hatalı seçimi, önemli açıklayıcı değişkenlerin modelde yer almaması, veride gözlemler arasındaki aşırı değişkenlik ve gözlemlerin dağılımının hatalı belirlenmesi aşırı yayılım sorununu ortaya çıkaran başlıca nedenlerdir (Gbur ve ark. 2012).

Bulgular ve Tartışma

Verilerin dağılımına ilişkin histogram grafiği Şekil 1’de verilmiştir.



Şekil 1. Oğlak verilerinin dağılımına ilişkin grafik

Şekil 1’de gözlem değerleri arasında var olan çok büyük değişimden dolayı, verilerin sahip olduğu dağılımın şekli, sağa doğru aşırı çarpık olduğu görülmektedir. Gbur ve ark. (2012)’na göre, bu tür verilere uygulanan *log* transformasyonunun, çarpıklığın giderilmesi üzerine bir etkisi olmadığı belirtilmiştir.

Genelleştirilmiş doğrusal karışık modelde aşırı yayılımın incelenmesine ilişkin sonuçlar

Çizelge 1’de her üç model algoritmasında aşırı yayılıma ilişkin uyum istatistikleri sonuçları verilmiştir.

Çizelge 1. Aşırı yayılım durumuna ilişkin uyum istatistiği sonuçları

Modeller	Dağılım	Uyum istatistiği	Hesap Değeri
Model 1	Poisson Dağılımı	Genelleştirilmiş Ki-Kare/Sd ¹	199.13
Model 2	Negatif Binomial Dağılımı	Genelleştirilmiş Ki-Kare/Sd	1.00
Model 3	Poisson Dağılımı	Genelleştirilmiş Ki-Kare/Sd	1.08

¹ Sd: Serbestlik derecesi

Çizelge 1’de 6 aydan küçük oğlak (dişi-erkek) varlığına ilişkin veri setinin Poisson dağılımına sahip olduğu varsayımı altında aşırı yayılım ($199.13 > 1$) olduğu belirlenmiştir. Model 2’de ise alternatif dağılım olarak Negatif Binomun tercih edilmesi durumunda uyum istatistiği 1’e eşit olarak elde edilmiş ve yayılım sorununun giderildiği belirlenmiştir. Yapılan bazı çalışmalarda (Sezgin ve Deniz 2004; Yeşilova ve ark. 2006; Stroup 2013), Poisson dağılımına sahip veri setlerinde aşırı yayılımın söz konusu olması durumunda, Poisson dağılımından türetilen ve aynı bağlantı fonksiyonuna sahip olan Negatif Binom dağılımının kullanılması, bu sorunun çözümünde önemli bir yol olduğu belirtilmiştir.

Negatif Binom dağılım, verideki heterojenliği ya da aşırı değişkenliği düzeltmek için kullanılan yayılım parametresine sahiptir. Dolayısıyla, verideki aşırı yayılımın giderilmesine yönelik olarak bir çözümleme yaparak daha doğru tahminlerin elde edilmesine olanak sağlamaktadır. Ancak Negatif Binom dağılım, az yayılım (underdispersion) durumundaki Poisson verilerinde kullanılmamaktadır (Hilbe 2014).

Gbur ve ark. (2012)’na göre şansa bağlı etkilere bağlı olarak değişimin önemli olması durumunda, bu etkilerin ihmal edilmesinin de aşırı yayılıma neden olduğu ifade edilmiştir. Bununla birlikte, oğlak varlığına ilişkin veri setinde yıllar arasındaki değişkenliğinin önemli olduğu varsayılarak, Poisson dağılımı varsayımı altında Model 3’te kesimin (intercept) yanı sıra yıl etkisi şansa bağlı terim olarak ilave edilmiş ve bunun sonucunda uyum istatistiğinin 1’e oldukça yakın çıktığı belirlenmiştir.

Bu çalışmada aynı zamanda, integral yaklaşımları da modellerde denenmiş ancak elde edilen pearson ki-kare/serbestlik derecesi istatistikleri Model 2’de 0.50 ve Model 3’de 0.10 olarak elde edilmiştir. Bu sonuçlar, az yayılım durumunu ortaya çıkardığından, aşırı yayılım durumu için genelleştirilmiş ki-kare/serbestlik derecesi uyum ölçütü tercih edilmiştir.

Çizelge 2’de aşırı yayılım durumunda (Model 1) ve aşırı yayılımın giderildiği durumlarda (Model 2 ve Model 3) elde edilen sabit etkilere ilişkin sonuçlar verilmiştir. Buna göre, aşırı yayılımın olduğu Model 1’den elde edilen standart hatalar, Model 2 ve 3 ile karşılaştırıldığında oldukça küçük elde edilmiştir. Bu sonuçlara paralel olarak, Koç ve ark. (2013) tarafından yapılan çalışmada, aşırı yayılım durumunda elde edilen standart hataların, aşırı yayılımın giderildiği duruma göre daha küçük çıktığı ifade edilmiştir. Çizelge 2’de yılların etkisi Model 1’de önemliyken, Model 2 ve 3’te önemsiz çıktığı görülmektedir. Dolayısıyla, aşırı yayılım sorununun giderilmemesi durumunda, yanıltıcı sonuçların elde edilebileceği belirlenmiştir. Bu sonuçlara göre Model 2 ve 3 aşırı yayımlı verilerin modellenmesinde, Model 1’e göre daha uygun modellerdir.

Çizelge 3’te kurulan her üç modelden elde edilen en küçük kareler ortalamalarına ilişkin sonuçlar verilmiştir. Bu sonuçlar, model kodlarında “lsmeans yıl/ ilink” opsiyonu kullanılarak elde edilmiştir. “lsmeans yıl” kodunda, tüm istatistiki sonuçlar log-bağlantı ölçeğinden (Model ölçeği) elde edilirken, ters bağlantı kodunun eklenmesiyle (ilink) verinin orijinal ölçeğinden ortalama, delta metot kullanılarak ise ortalamanın standart hataları elde edilmektedir (Veri ölçeği) (SAS 2008). Buna göre, Çizelge 3’te aşırı yayılımın belirlendiği Model 1’den elde edilen standart hatalar, Model 2 ve 3’e göre daha küçük elde edilmiştir. Model ölçeğinden elde edilen tahminler kullanılarak, veri ölçeğine ilişkin ortalamalar elde edilmektedir. Örneğin, Model 1 sütunundaki 2010 yılına ait ortalama; $\eta = \log(\mu) = 5.6707$ ve $\mu_{2010} = e^{\eta} = e^{5.6707} = 290.24$ şeklinde elde edilmektedir.

Çizelge 2. Sabit etkilere ilişkin sonuçlar

Model 1			
Yıllar¹	Tahmin	Std. Hata	Pr> t
Sabit	5.8004	0.3165	<.0001
2010	-0.1296	0.0115	<.0001
2011	-0.1025	0.0114	<.0001
2012	-0.1067	0.0115	<.0001
2013	-0.2303	0.0118	<.0001
2014	-0.0512	0.0113	<.0001
Model 2			
Sabit	5.6295	0.3898	<.0001
2010	0.3891	0.3161	0.2217
2011	0.2300	0.3161	0.4689
2012	-0.3654	0.3165	0.2516
2013	-0.2621	0.3164	0.4098
2014	0.2397	0.3161	0.4503
Model 3			
Sabit	4.9415	0.5321	<.0001
2010	0.6293	0.4680	0.1823
2011	0.5329	0.4683	0.2584
2012	-0.2864	0.4728	0.5462
2013	-0.3376	0.4742	0.4785
2014	0.2813	0.4714	0.5524

¹: 2015 yılı, kurulan modellerde referans parametre olduğundan çizelgede yer verilmemiştir.

Çizelge 3. Yıllara göre oğlak sayımlarına ait en küçük kareler ortalamaları

Model 1				
Yıl	Tahmin	Std. Hata	Ortalama	Ortalamanın Standart Hatası
2010	5.6707	0.3165	290.24	91.86
2011	5.6979	0.3165	298.24	94.39
2012	5.6937	0.3165	296.98	93.99
2013	5.5700	0.3165	262.44	83.07
2014	5.7492	0.3165	313.93	99.35
2015	5.8004	0.3165	330.42	104.57
Model (Log) ölçeği			Veri ölçeği	
Model 2				
2010	6.0186	0.3896	410.99	160.13
2011	5.8595	0.3897	350.55	136.60
2012	5.2641	0.3900	193.28	75.38
2013	5.3674	0.3899	214.30	83.56
2014	5.8692	0.3897	353.98	137.93
2015	5.6295	0.3898	278.52	108.56
Model (Log) ölçeği			Veri ölçeği	
Model 3				
2010	5.5708	0.5280	262.66	138.69
2011	5.4744	0.5283	238.51	126.01
2012	4.6551	0.5326	105.12	55.99
2013	4.6039	0.5340	99.88	53.33
2014	5.2228	0.5315	185.45	98.57
2015	4.9415	0.5321	139.98	74.49
Model (Log) ölçeği			Veri ölçeği	

Sonuç olarak, sayımla elde edilen veri setlerinde, Poisson dağılımı varsayımı altında aşırı yayılım durumu incelenmelidir. Çünkü, aşırı yayılım varlığında elde edilen sonuçlar, yapılacak tahminler üzerinde önemli bir etkiye sahip olmaktadır. Bu nedenle, genelleştirilmiş doğrusal karışık model kullanılarak oluşturulacak modellerde, alternatif dağılım olarak Negatif Binom dağılımının tercih edilmesi veya şansa bağlı terim ilavesi, aşırı yayılımın giderilmesinde etkili çözüm yolları olarak önerilebilir.

Kaynaklar

- Gbur EE, Stroup WW, McCarter KS, Durham S, Young LJ, Christman M, West M, Kramer M (2012). Analysis of Generalized Linear Mixed Models in the Agricultural and Natural Resources Sciences. ASA, CSSA and SSSA, Inc., Madison, WI, USA.
- Harwood J (2013). Analysis of Overdispersed Data in SAS. <http://chipts.ucla.edu/downloads/738>. (Erişim tarihi: 09.02.2016).
- Hilbe J M (2014). Modeling Count Data, 1nd ed., Cambridge Univ. Press, New York.
- Işık F (2011). Generalized Linear Mixed Models: An introduction for tree breeders and pathologists. Statistic Session class notes, Fourth International Workshop on the Genetics of Host-Parasite Interactions in Forestry, July 31 – August 5, Eugene, Oregon, USA.
- Koç H, Cengiz MA, Koç T, Dündar E (2013). Aşırı yayımlı veriler için genelleştirilmiş Poisson karma modellerin hava kirliliği üzerine bir uygulaması. IAAOJ Scientific Science 1(2): 3-7.
- Morel JG (2014). Analysis of data with overdispersion using the SAS system. SAS Global Forum, 23-26 Mart, Washington, DC.
- SAS (2008). SAS/STAT® 9.2 User's Guide The GLIMMIX Procedure (Book Excerpt). 1nd electronic book. SAS Institute Inc., Cary, NC, USA.
- SAS (2014). SAS/STAT. Statistical Analysis System for Windows. Release 9.4. SAS Institute Inc., USA.
- Sezgin HF, Deniz E (2004). Poisson regresyon modelinde aşırı yayılım durumu ve Negatif Binomial regresyon analizinin Türkiye grev sayıları üzerine bir uygulaması. Yönetim 15(48): 17-25.
- Stroup WW (2013). Generalized Linear Mixed Models: Modern Concepts, Methods and Applications. CRC Press, New York.
- TÜİK (2016). <https://biruni.tuik.gov.tr/hayvancilikapp/hayvancilik.zul> (Erişim tarihi: 24.02.2016).
- Yeşilova A, Yılmaz A, Kaki B (2006). Norduz erkek kuzularının bazı kesikli üreme davranış özelliklerinin analizinde doğrusal olmayan regresyon modellerinin kullanılması. Yüzüncü Yıl Üniversitesi Tarım Bilimleri Dergisi (YYUJAGRSCI) 16(2): 87-92.