

PAPER DETAILS

TITLE: Sifir Deger Agirlikli Sayima Dayali Olarak Elde Edilen Bagimli Degiskenin Modellenmesinde
Kullanilan Regresyon Yöntemleri

AUTHORS: Abdullah YESILOVA,Baris KAKI,Ismail KASAP

PAGES: 1-9

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/2587690>

SIFIR DEĞER AĞIRLIKLIL SAYIMA DAYALI OLARAK ELDE EDİLEN BAĞIMLI DEĞİŞKENİN MODELLENMESİNDE KULLANILAN REGRESYON YÖNTEMLERİ

Abdullah YEŞİLOVA* Barış KAKİ*
İsmail KASAP**

ÖZET

Çalışmada, sıfır değer ağırlıklı sayıma dayalı olarak elde edilen bağımlı değişkenin analizi için Poisson regresyonu, negatif Binom regresyonu, sıfır ağırlıklı Poisson regresyonu ve sıfır ağırlıklı negatif Binom regresyonu incelenmiştir. Veri kümesinde sıfır değerlerinin çok olması ve gözlemler arasındaki büyük farklılıktan dolayı aşırı yayılımın önemli bir etkiye sahip olduğu saptanmıştır. Uygun model seçiminde Akaike ve Bayesçi bilgi ölçütleri kullanılmıştır. Bunun sonucunda, sıfır ağırlıklı negatif Binom regresyon modeli en uygun model olarak seçilmiştir. Uyum ölçütleri sonucunda, sıfır ağırlıklı Poisson regresyonun, Poisson regresyonuna ve sıfır ağırlıklı negatif Binom regresyonunda, negatif Binom regresyona tercih edilebileceği saptanmıştır. Sıfır ağırlıklı negatif Binom regresyonunda, modele alınan avcı akarın (Zetzellia mali), sıcaklığın ve ilaçlamanın zararlı akar Panonychus ulmi Koch'un tüm dönemleri toplamı üzerine etkileri önemli bulunmuştur ($p<0.01$).

Anahtar kelimeler: Aşırı Yayılım, Negatif Binom Regresyonu, Poisson Regresyonu, Sıfır Ağırlıklı Negatif Binom Regresyonu, Sıfır Ağırlıklı Poisson Regresyonu.

1. GİRİŞ

Poisson regresyonu sayıma dayalı olarak elde edilen verilerin analizinde yoğun olarak uygulanmaktadır (Frome vd, 1973; Böhning, 1994; Cameron ve Trivedi, 1998; Stokes vd, 2000). Poisson Regresyon (PR) analizi, bağımsız değişkenler ile sayıma dayalı olarak elde edilen bağımlı değişken arasındaki ilişkiyi açıklamaktadır. PR'de bağımsız değişkenlerin doğrusal yapısını, bağımlı değişkenin beklenen değerine bağlayan bağlantı fonksiyonu, logaritmik dönüşüm ile verilmektedir (McCullagh ve Nelder, 1989; Breslow, 1990). Bilindiği gibi, Poisson dağılımında ortalama ile varyans birbirine eşittir. Ancak uygulamada bu eşitliği sağlamak her zaman mümkün değildir. Varyansın ortalamaдан büyük çıkması aşırı yayılım (overdispersion) ve küçük çıkması da az yayılım (underdispersion) olarak tanımlanmaktadır (Breslow, 1990; Böhning, 1994; Cameron ve Trivedi, 1998; Stokes vd, 2000; SAS, 2007). Veri kümelerinde genellikle aşırı yayılım, nadiren de az yayılım ile karşılaşmaktadır. Böyle durumlarda PR'yi uygulamak, yanlış parametre tahminlerinin elde edilmesine neden olur (Cox, 1983; Cameron ve Trivedi, 1998). Veri setinde aşırı yayılım söz konusu olduğunda Negatif Binom Regresyon (NBR)'nin kullanılması daha uygun olmaktadır (SAS, 2007). NBR modeli, PR ile aynı bağlantı fonksiyonunu kullanmaktadır. Özellikle veri kümesinde aşırı yayılım söz konusu olduğunda NBR modeli, PR modeline tercih edilmektedir.

* Yrd. Doç. Dr., Yüzüncü Yıl Üniversitesi, Ziraat Fakültesi, Zootehni Bölümü, Biyometri- Genetik ABD, 65080 VAN, e-mail: yesilova@yyu.edu.tr

** Doç. Dr., Yüzüncü Yıl Üniversitesi, Ziraat Fakültesi, Bitki Koruma Bölümü. Entomoloji ABD, 65080 VAN, e-mail: ikasap@yyu.edu.tr

Uygulamada, Poisson dağılımı gösteren sayıma dayalı olarak elde edilen değerler beklenenden daha fazla sayıda sıfır değerine sahip olabilir. Böyle bir durumda fazla sayıda sıfır değerine sahip bağımlı değişkenin modellenmesinde, sıfır ağırlıklı Poisson regresyonunun (Zero-Inflated Poisson Regression=ZIP) kullanılması uygun bir yaklaşımdır (Lambert, 1992; Böhning, 1998; Böhning vd, 1999; Yau ve Lee, 2001; Dalrymple vd, 2003). ZIP, kitlenin iki farklı tip veriden oluştuğunu varsaymaktadır. Bunlardan birincisi, sıfır değerlerine sahip olabilen Poisson dağılımlı sayıma dayalı veriler, buna karşın ikinci tip ise daima sıfır değerleri alan veriler olmaktadır. ZIP dağılımı, Poisson ortalaması μ ve bireylerin oranı π olmak üzere iki parametreye sahiptir (Böhning vd., 1999; Ridout vd., 2001; Lee ve Wang, 2001). Bununla birlikte, yukarıda bahsedilen aşırı yayılım, sıfır değerlerinin çok olduğu veri setlerinde de söz konusudur. Böyle durumlarda, sıfır ağırlıklı negatif Binom regresyonu (Zero-Inflated Negatif Binomial Regression=ZINB) kullanılan alternatif bir modeldir (Hall, 2000; Ridout vd, 2001; Yau ve Lee, 2001). Standart Poisson regresyonunda, aşırı yayılım parametre tahminleri üzerinde küçük bir etkiye sahiptir, ancak standart hataların yanlış tahmin edilmesine neden olmaktadır (Cox, 1983). Bu yanlış tahminleme yarı olabilirlik (Quasi-likelihood) yönteminin kullanılması ile giderilebilir. PR, NBR, ZIP ve ZINB için parametre tahminleri en çok olabilirlik (maximum likelihood) yöntemi ile elde edilmektedir. Uygun model seçiminde Akaike bilgi ölçütü ile Bayesçi bilgi ölçütü kullanılmaktadır.

2. YÖNTEM

2.1 Materyal

Çalışmanın materyalini, Van merkez ilçesinden seçilen bir bahçede, Mayıs-Ekim ayı sonuna kadar starking elma ağaçlarından, haftalık olarak alınan yaprak örnekleri üzerindeki, zararlı akar *Panonychus ulmi* Koch (Acarina: Tetranychidae) ile bu akarın avcısı olan *Zetzellia mali* (Ewing) (Acarina: Stigmaeidae)'ye ait sayımlar oluşturmuştur. Yaprak örnekleri, toplam 10 ağaçtan ve ağaç başına 10 yaprak olarak toplanmıştır. Toplanan yapraklar laboratuvara getirilerek, stereobinoküler mikroskop altında incelenerek, yaprağın her iki yüzeyi üzerindeki akarlar sayılıp, kaydedilmiştir. Denemenin yürütüldüğü bahçe 02.07.2002 ve 27.07.2002 tarihlerinde iki defa Fluvalinate etkili bir tarımsal savaş ilacı ile ilaçlanmış ve bu ilaçlamanın akarlar üzerine etkisi araştırılmıştır. Bunun dışında, Van merkez ilçesinin aylık ortalama sıcaklık değerleri, Devlet Meteoroloji İstasyonundan alınmış ve akarların popülasyonuna etkisi değerlendirilmiştir.

2.2 Yöntem

Bu bölümde PR, NBR, ZIP ve ZINB modellerine ilişkin teorik bilgiler verilecektir. Dört model için parametre tahminleri en çok olabilirlik yöntemi kullanılarak, elde edilmektedir.

2.2.1 Poisson Regresyonu

PR’de ilgilenilen olayın gözlenen sayısı olan y_i bağımlı değişkenin Poisson dağılışına sahip olduğu varsayılmaktadır. Poisson ortalaması olan μ ’nün logaritmasının, bağımsız değişkenlerin bir doğrusal fonksiyonu olduğu varsayılmaktadır (Nelder ve Wedderburn, 1972; SAS, 2007). Log bağlantı fonksiyonlu Poisson regresyon modeli,

$$\Pr(y_i / x_i) = \exp(-\mu_i) \mu_i^{y_i} / y_i! \quad y_i=0,1,\dots \quad (1)$$

biçiminde verilmektedir. Eşitlik (1)’de

$$\mu_i = \exp(x_i \beta)$$

olup, x_i kovaryet (ortak değişken) vektörü, β ise tahmin edilecek olan bilinmeyen parametre vektörüdür. PR’de α negatif olmayan yayılım parametresi olmak üzere,

$$V(\mu) = \alpha \mu \quad (2)$$

biçiminde yazılabilir. Eşitlik (2)’de, $\alpha > 1$ ise aşırı yayılım, $\alpha < 1$ ise az yayılım ve $\alpha = 1$ ise yayılım olmadığını göstermektedir (SAS, 2007).

2.2.2 Negatif Binom Regresyonu

NBR bağımlı değişken ile bağımsız değişkenler vektörü arasında, log bağlantı fonksiyonunu kullanmaktadır. NBR modeli,

$$\Pr(Y = y_i / x_i) = \frac{\Gamma(y_i + \alpha^{-1})}{y_i! \Gamma(\alpha^{-1})} (\mu_i \alpha / (1 + \alpha \mu_i))^{y_i} (1 / (1 + \alpha \mu_i))^{\alpha^{-1}} \quad \alpha > 0 \quad (3)$$

biçiminde verilmektedir. Eşitlik (3)’te, α aşırı yayılımın derecesini gösteren yardımcı parametre değeridir. α değeri sıfıra yaklaştığında, negatif Binom model PR modeline dönüşür. NBR modelde ortalama ve varyans,

$$E(y_i | x_i) = \mu_i$$

ve

$$Var(y_i | x_i) = \mu_i(x) + \alpha \mu_i^2$$

olarak verilebilir (Lawless, 1987; Cheung, 2002).

2.2.3 Sıfır Ağırlıklı Poisson Regresyonu

y_i extra sıfırların sayısını açıklamak için, ZIP modeli,

$$\Pr(y_i/x_i) = \begin{cases} \pi_i + (1 - \pi_i) \exp(-\mu_i), & y_i = 0 \\ (1 - \pi_i) \exp(-\mu_i) \mu_i^{y_i} / y_i!, & y_i > 0 \end{cases} \quad (4)$$

biçiminde yazılabilir (Ridout vd, 2001; Cheung, 2002). Eşitlik (4)'te, π_i ekstra sıfırların olma olasılığını göstermektedir. Bundan dolayı $y_i = 0$ olan bireyler, iki gruptan oluşmuş şekilde tanımlanır. Bu gruplardan biri, deneğin Poisson süreci göstermediği, diğeri ise deneklerin

$$\exp(-\mu_i) \mu_i^0 / 0! = \exp(-\mu_i)$$

olmasından dolayı, sıfır değerleri alan μ ortalamalı Poisson dağılımına aittir. Dağılımın ortalama ve varyansı aşağıdaki biçimde yazılabilir (Ridout vd, 2001; Dalrymple vd, 2003).

$$E(y_i) = (1 - \pi_i) \mu_i \text{ ve } \text{Var}(y_i) = (1 - \pi_i) \mu_i (1 + \pi_i \mu_i)$$

2.2.4 Sıfır Ağırlıklı Negatif Binom Regresyonu

Sıfır değerlerinin çok fazla olduğu y_i bağımlı değişkeninin modellenmesinde alternatif regresyon yöntemi, ZINB'dir. ZINB regresyon modeli,

$$\Pr(y_i/x_i) = \begin{cases} \pi_i + (1 - \pi_i) (1 + \alpha \mu_i^c)^{-\mu_i^{1-c}/\alpha}, & y_i = 0 \\ (1 - \pi_i) \frac{\Gamma(y_i + \mu_i^{1-c}/\alpha)}{y_i! \Gamma(\mu_i^{1-c}/\alpha)} (1 + \alpha \mu_i^c)^{-\mu_i^{1-c}/\alpha} (1 + \mu_i^{-c}/c)^{-y_i}, & y_i > 0 \end{cases} \quad (5)$$

biçiminde yazılabilir (Ridout vd, 2001). Eşitlik (5)'de, $(\alpha \geq 0)$ ortak değişkenlere bağlı olmadığı varsayılan bir yayılım parametresidir. ZINB dağılımının ortalama ve varyansı aşağıdaki biçimde yazılabilir.

$$E(Y_i) = (1 - \pi_i) \mu_i \text{ ve } \text{Var}(Y_i) = (1 - \pi_i) \mu_i (1 + \pi_i \mu_i + \alpha \mu_i^c)$$

Bu dağılım limit $\alpha \rightarrow 0$ gittiğinde, ZIP dağılımına dönüşür. Negatif Binom dağılımın ortalaması μ_i 'dir. c endeksi negatif Binom dağılımın formunu tanımlamaktadır. $c=0$ için negatif Binom dağılımın varyansı

$$(1 + \alpha) \mu_i$$

olup, $c=1$ için ise

$$\mu_i + \alpha \mu_i^2$$

olur (Ridout vd, 2001).

2.2.5 Uyum Ölçütleri

Akaiki bilgi ölçütü (Akaiki Information Criteria=AIC) ve Bayesçi bilgi ölçütü (Bayesian Information Criteria=BIC) model uyumu için kullanılan uyum ölçütleridir. Birçok Monte-Carlo simülasyonu BIC, AIC uyum kriterinin birlikte kullanılması gerektiğini göstermektedir (Muthen ve Muthen, 2002; Dalrymple vd, 2003). En küçük uyum ölçütlerine sahip model, en iyi model olarak kabul edilir. Genel olarak;

$$AIC = -2 \log L + 2r \quad (6)$$

ve

$$BIC = -2 \log L + r \ln(n) \quad (7)$$

biçiminde tanımlanır. Eşitlik (6) ve (7)'de, LogL karışımli Poisson regresyon modelinde iterasyon bittikten sonra elde edilen log-olabilirlik değerini, r parametre sayısını ve n örnek büyüklüğünü göstermektedir.

3. BULGULAR

Çalışmada, gerekli analizler SAS ve MPlus istatistik yazılım programları kullanılarak yapılmıştır. Tüm dönemlere ait toplam akar sayısı bağımlı değişken, ilaçlama, sıcaklık ve *Zetzellia mali* (avcı) bağımsız değişkenler olarak, modele dahil edilmiştir. PR modelinde aşırı yayılım değeri 128.0860 olarak bulunmuştur. Söz konusu yayılım değerinin (1) değerinden çok büyük çıkması, veri kümesinde aşırı bir yayılım olduğunu gösterir. Dört farklı regresyon modeli için uyum ölçütleri Tablo 1'de verilmiştir. Çalışmada kullanılan 1292 gözlem değerinin, 888'i sıfır değerlidir. Veri kümesinde sıfır değerlerinin çok olması ve gözlemler arasındaki büyük farklılıklar aşırı yayılıma neden olmuştur. Bundan dolayı Tablo 1'de verilen uyum ölçütleri birbirlerinden oldukça farklı bulunmuştur. Bunun sonucu olarak, ZIP uyum ölçütleri PR'ye ve ZINB uyum ölçütleri de NBR'ye göre daha küçük çıkmıştır. Tablo 1'de koyu harflerle gösterilen, ZINB modeli en iyi model olarak seçilmiştir.

Tablo 1. Poisson ve negatif Binom regresyonu ile sıfır ağırlıklı Poisson ve negatif Binom regresyonu için uyum ölçütleri

Model	Log-Olabilirlik	AIC	BIC
PR	-148760.41	297522.82	297527.93
ZIP	-63541.62	127085.24	127090.75
NBR	-2812.49	5626.98	5632.09
ZINB	-2103.29	4208.59	4213.70

PR, NBR, ZIP ve ZINB için modele alınan değişkenlere ilişkin en çok olabilirlik yöntemi ile elde edilen parametre tahmin değerleri ve standart hataları Tablo 2'de verilmiştir. Tablo 2'ye göre, PR'de toplam akar sayısı üzerine avcı akarın etkisi önemli bulunmuşken ($p < 0.01$), ilaçlamanın ve sıcaklığın etkisi önemsiz bulunmuştur. NBR'de

toplam akar sayısı üzerine avcı akarın etkisi önemli bulunmuşken ($p < 0.01$), ilaçlamanın ve sıcaklığın etkisi önemsiz bulunmuştur. Elde edilen parametre tahminleri bakımından PR ile NBR benzer sonuçlar vermiştir. ZIP ve ZINB’de, modele alınan tüm bağımsız değişkenlerin, toplam akar sayısı üzerine etkileri önemli bulunmuştur ($p < 0.01$). Sıcaklığın etkisi hem PR’de, hemde NBR’de önemsiz çıkmıştır. Sıcaklık, akarın gelişmesi ve çoğalması üzerine önemli etkenlerden birisi iken, PR ve NBR’de önemsiz bulunması, veri kümesinde büyük bir aşırı yayılım olduğunu göstermektedir. Tablo 2’ye göre, sıfır ağırlıklı bağımlı değişkenin (toplam akar sayısı) tahmin değeri ZIP ve ZINB modellerinde sırasıyla, -48.705 ve -49.462 olarak elde edilmiş ve önemli etkiye sahip oldukları saptanmıştır ($p < 0.01$). Böylece bağımlı değişkendeki sıfır değerlerinin çokluğu ZIP ve ZINB regresyon yöntemlerinde önemli bir etkiye sahip olmuştur. Bu bağlamda, sıfır değerlerinin önemli etkiye sahip olmasından dolayı, sıfır değer ağırlıklı regresyon modellerinin PR ve NBR’ye tercih edilmesi gerekmektedir.

Tablo 2. PR, NBR, ZIP ve ZINB modelleri için parametre tahminleri ve standart hatalar

Parametreler	PR Tahmin (standart hata)	NBR Tahmin (standart hata)	ZIP Tahmin (standart hata)	ZINB Tahmin (standart hata)
Intercept	4.543 ¹ (0.022)	5.767 ¹ (1.089)	7.981 ¹ (0.023)	10.536 ¹ (0.738)
<i>Zetzellia mali</i> (avcı)	-0.059 ¹ (0.001)	-0.099 ¹ (0.015)	-0.176 ¹ (0.002)	-0.103 ¹ (0.013)
İlaçlama	-0.010 (0.015)	0.521 (0.499)	-0.064 ¹ (0.015)	0.923 ¹ (0.283)
Sıcaklık	-0.024 (0.001)	-0.116 (0.073)	-0.120 ¹ (0.001)	-0.328 ¹ (0.454)
İntercept (sıfır ağırlıklı)	-	-	25.654 ¹ (12539.09)	26.179 ¹ (17314.07)
Toplam akar (sıfır ağırlıklı)	-	-	-48.705 ¹ (18273.72)	-49.462 ¹ (22860.9)

¹ $p < 0.01$

4. TARTIŞMA VE SONUÇ

Çalışmada, aşırı yayılım büyük bir etkiye sahip olduğundan dolayı, dört farklı regresyon modelinde uyum ölçütleri ve parametre tahmin değerleri birbirinden oldukça farklı bulunmuştur. Genel olarak en küçük uyum ölçütlerine sahip model, en iyi model olarak tanımlanmaktadır (Ridout vd, 2001; Dalrymple vd, 2003). Uyum ölçütlerine bakıldığında en uygun modelin, ZINB modeli olduğu saptanmıştır. Sayıma dayalı olarak elde edilen bağımlı değişkende sıfır değerlerinin çokluğu ve aşırı yayılımın olması, ZINB modelinin uygun model olarak seçilmesinde etkili olmuştur. Sayıma dayalı olarak elde edilen bağımlı değişkende sıfırların sayısının az olduğu durumlarda, PR ve NBR modellerinin kullanılması daha uygun olmaktadır. Ancak böyle bir veri kümesinde aşırı yayılım söz konusu ise, NBR regresyon modeli PR modeline tercih edilir (Agresti, 1997; Cameron ve Trivedi, 1998; Stokes vd., 2000; Muthen ve Muthen, 2002; SAS, 2007).

Aşırı yayılımın büyük bir etkiye sahip olmasından dolayı ZINB modelinden elde edilen sonuçların kullanılması gerekmektedir. Tablo 2'ye göre toplam akar (sıfır ağırlıklı) etkisinin tahmin değerinin önemli çıkması, sıfır değer ağırlıklı regresyon modellerinin kullanılması gerektiğini göstermektedir. Sıcaklık, akarın gelişmesi ve çoğalması üzerine önemli etkenlerden birisi iken, (Kasap, 2004), PR ve NBR'de sıcaklığın etkisi önemsiz çıkmıştır. ZIP ve ZINB'de ise sıcaklığın etkisinin toplam akar sayısı üzerine etkileri önemli bulunmuştur. Sıcaklık akarın gelişmesi ve çoğalması üzerinde önemli bir etkiye sahip olmasına rağmen, PR ve NBR'de önemsiz çıkması, bu modellerinin böyle yapıdaki veri kümelerine uygulanmaması gerektiği sonucunu doğurmaktadır. Benzer durum ilaçlama için de geçerli olmuştur. PR ve NBR'de ilaçlamanın etkisi önemsiz bulunmuşken, ZIP ve ZINB'de önemli bulunmuştur. Sıfır değer ağırlıklı bağımlı değişkenin (toplam akar sayısı) tahmin değeri ZIP ve ZINB modellerinde sırasıyla, -48.705 ve -49.462 önemli bulunması, uygulanan ZIP ve ZINB modellerinin PR ve NBR'ye göre daha uygun olduğunu göstermektedir. Başka bir ifadeyle bağımlı değişkendeki sıfır sayılarının çok olması, sıfır değerlerini dikkate alan sıfır değer ağırlıklı regresyon modellerinin kullanılmasını gerektirmektedir. Bağımlı değişkende yer alan sıfır değerlerine ilişkin tahmin değerinin önemsiz çıkması, ZIP, ZINB ve NBR sonuçları arasında pek farklılık olmadığına göstergesi olabilir. Bu bağlamda elde edilen bütün bulgular ile uyum ölçütleri birbirini desteklemektedir. ZIP ve ZINB modellerinin gerek uyum ölçütleri bakımından, gerekse parametre tahminleri bakımından PR ve NBR modellerine göre daha iyi olduğu saptanmıştır. Sonuç olarak, sıfır değer ağırlıklı sayıma dayalı olarak elde edilen bağımlı değişkenin modellenmesinde, sıfır ağırlıklı yöntemlerin (ZIP ve ZINB) kullanılması gerekmektedir.

5. KAYNAKLAR

- Agresti, A., 1997. *Categorical Data Analysis*. New Jersey, Canada; John and Wiley & Sons, Incorporation.
- Böhning, D., 1994. A Note on a Test for Poisson Overdispersion. *Biometrika*, 81, 418-419.
- Böhning, D., 1998. Zero- Inflated Poisson Models and C.A.MAN: A Tutorial Collection of Evidence. *Biometrical Journal*, 40(7), 833-843.
- Böhning, D., Dietz, E ve Schlattmann, P., 1999. The Zero-Inflated Poisson Model and the Decayed, Missing and Filled Teeth Index in Dental Epidemiology. *Journal of Royal Statistical Society, A*, 162, 195–209.
- Breslow, N., 1990. Tests of Hypotheses in Overdispersed Poisson Regression and Other Quasi-Likelihood Models. *Journal of American Statistical Association*, 85(410), 565-571.
- Cameron, A.C ve Trivedi, P.K., 1998. *Regression Analysis of Count Data*. New York: Cambridge University Pres.
- Cheung, Y.B., 2002. Zero-Inflated Models for Regression Analysis of Count Data: A Study of Growth and Development. *Statistics in Medicine*, 21, 1461-1469.
- Cox, R., 1983. Some Remarks on Overdispersion. *Biometrika*, 70, 269-274.

Dalrymple, M.L., Hudson, I.L ve Ford, R.P.K., 2003. Finite Mixture, Zero-Inflated Poisson and Hurdle Models with Application to SIDS. *Computational Statistics & Data Analysis* 41, 491-504.

Frome, E.D., Kutner, M.H ve Beauchamp, J.J., 1973. Regression Analysis of Poisson-Distributed Data. *Journal of American Statistical Association*, 68(344), 935-940.

Hall, D.A., 2000. Zero-Inflated Poisson and Negative Binomial Regression with Random Effects: A Case Study. *Biometrics*, 56, 1030-1039.

Kasap, İ., 2004. Effect of Different Apple Cultivars and of Temperatures on Biology and Life Table Parameters of TwoSpotted Spider Mite, *Tetranychus Urticae* Koch (Acarina: Tetranychidae). *Phytoparasitica*, 32(1): 73-82.

Lambert, D., 1992. Zero-Inflated Poisson Regression, with an Application to Defects in Mnaufacturin. *Technometrics*, 34(1), 1-13.

Lawles, J.F., 1987. Negative Binomial and Mixed Poisson Regression. *The Canadian Journal of Statistics*, 15(3), 209-225.

Lee, A.H ve Wang, K., 2001. Analysis of Zero-Inflated Poisson Data Incorporating Extent of Exposure. *Biometrical Journal*, 43(8), 963-975.

McCullagh, P ve Nelder, J.A., 1989. *Generalized Linear Models*. Second Edition, London, UK, Chapman and Hall.

Muthen, L.K ve Muthen, B., 2002. *MPlus: User's Guide*. Los Angeles, CA: Muthén & Muthén.

Nelder, J.A ve Wedderburn, R.W.M., 1972. Generalized Linear Models. *Journal of Royal Statistical Society A*, 135(3), 370- 384.

Ridout, M., Hinde, J ve Demetrio, C.G.B., 2001. A Score Test for a Zero-Inflated Poisson Regression Model Against Zero-Inflated Negative Binomial Alteratves. *Biometrics*, 57, 219-233.

SAS., 2007. *SAS/Stat. Software*. Hangen and Enhanced, USA: SAS, Institute. Incorporation.

Stokes, M.E., Davis, C.S ve Koch, G.G., 2000. *Categorical Data Analysis Using the SAS System*. USA; John and Wiley & Sons, Incorporation.

Yau, K.K.W ve Lee, A.H., 2001. Zero-Inflated Poisson Regression with Random Effects to Evaluate an Occupational Injury Prevention Programme. *Statistics in Medicine*, 20, 2907-2920.

REGRESSION METHODS USED IN MODELLING OF DEPENDENT VARIABLE OBTAINED BASED ON ZERO-INFLATED COUNT DATA

ABSTRACT

In this study, Poisson regression, negative binomial regression, zero-inflated Poisson regression, and zero-inflated negative binomial regression were investigated to analyze dependent variable obtained based on zero-inflated counting. It was determined that overdispersion had a significant effect because there were many zero values in data set and there were great difference among the observations. Akaike and Bayesian information criteria were used to choose the most appropriate model. In conclusion, zero-inflated negative binomial regression was chosen as the most appropriate model. It was determined that zero-inflated Poisson regression could be preferred to Poisson regression, and zero-inflated negative binomial regression could be preferred to negative binomial regression. In zero-inflated negative binomial regression, it was determined that predator acar (Zetzellia mali), temperature, and spraying in the model had significant effects ($p < 0.01$) on all stages of the harmful pest acar (Panonychus ulmi Koch).

Key Words: Negative Binomial Regression, Overdispersion, Poisson Regression, Zero-Inflated Negative Binomial regression, Zero-Inflated Poisson Regression.