Hacettepe Journal of Mathematics and Statistics
Volume 39 (2) (2010), 273 – 282

MODELING INSECT-EGG DATA WITH EXCESS ZEROS USING ZERO-INFLATED REGRESSION MODELS

Abdullah Yeşilova*†, M. Bora Kaydan‡ and Yılmaz Kaya§

Received 11 : 02 : 2009 : Accepted 04 : 01 : 2010

Abstract

As zero-inflated observations occur very often in studies on plant protection, models taking into account zero-inflated observations are frequently required. Especially, zero-inflated observations occur in large numbers for insects whose post-oviposition period lasts long, or that generally lay their eggs during the first days of the oviposition period. For the data used in this study, 1114 (43.84%) of the 2541 observations were zero. In the selection of an appropriate regression model, zero-inflated negative binomial regression was chosen as the best model. In all regression models, the day of laying and the three different hosts were seen to have a significant effect on daily egg numbers ($p < 0.01$).

Keywords: Zero-inflated count data, Overdispersion, Zero-inflated models, Hurdle models.

2000 AMS Classification: 91 G 70.

1. Introduction

As is well known, the mean and variance are equal to one another in a Poisson distribution. However, in applications, it is not always possible to realize this equality [1,2,3,5,9,23]. If the variance is higher than the mean, it is described as overdispersion, while if the variance is lower than the mean, it is described as underdispersion [2,7,14,21,24]. In data sets, generally overdispersion, but rarely underdispersion is seen. In such cases, applying Poisson regression (PR) causes biased parameter estimations [7].

*Yuzuncu Yil University, Faculty of Agriculture, Department of Animal Science, Biometry & Genetic Unit, Van, Turkey. E-mail: yesilova@yyu.edu.tr

†Corresponding Author.

‡Yuzuncu Yil University, Faculty of Agriculture, Department of Plant Production, Van, Turkey. E-mail: bkaydan@yyu.edu.tr

§Yuzuncu Yil University, Vocational High School, Department of Computer Programming, Van, Turkey. E-mail: ykaya@yyu.edu.tr

When there is overdispersion in a data set, it is better to use a negative binomial (NB) regression model [11,14,16,18,19]. In an NB regression model, parameter estimations are obtained by considering the effect that stems from overdispersion.

In studies of insect eggs for plant protection it is seen that 70–80 % of total egg laying is realized during the first week, and beginning from the 12^{th} day of oviposition, egg numbers decrease rapidly. Since the number of zero inflated observations is quite high after the 12^{th} day, the use of zero-inflated methods is required. In such a case, using a zero-inflated Poisson (ZIP) regression model is a suitable approach for analyzing a dependent variable having many zero observations [2,4,6,15,17,19].

ZIP assumes that the population consists of two different types of observation. The first is based on zero counts, while the second has a Poisson distribution [4,19,25]. In addition, overdispersion is also likely in data sets having excess zero observations. In such cases, a zero-inflated negative binominal (ZINB) regression model is an alternative method that can be used [2,10, 11,14,18,19]. As in ZIP regression, in ZINB regression the observations with zero data and those without zero data are modeled in different ways. Moreover, the hurdle model is used to analyze data sets having an excess of zero observations. The Hurdle model is composed of two stages. In the first stage we have binary responses indicating positive counts (2.1) against zero counts (0); the second one is the period when only positive counts take place. Binary responses are modeled using a binary model [8,20]. On the other hand, positive counts are modeled by using zero-truncated models. The binary part uses logit, probit or complimentary loglog, whereas the count part uses Poisson, geometric, and negative binomial [10]. Generally, for the positive counts section, Poisson hurdle (PH) or negative binomial hurdle (NBH) is used.

In models of PR, NB, ZINB, PH and NBH, parameter estimations are obtained with the maximum likelihood (ML) method by using the expectation and maximization (EM) algorithm [18]. In the selection of a suitable model, the Akaiki information criteria (AIC) and Bayesian information criteria (BIC) can be used. The model with the lowest selection criteria is accepted as the best model [8,24].

The aim of this study is to apply different regression methods in the analysis of daily egg numbers left on three different hosts by *Phenacoccus aceris* (Signoret) in 2002 and 2003.

Motivating Example Mapple scale *Phenacoccus aceris* (Signoret) (Hemiptera: Pseudococcidae), reared on different host plants (*Acer negundo*, *Acer pseudoplatanus* and *Fraxinus excelsior*) was used as the main material in this study. Daily egg laying numbers of P. aceris on the three different host plant were determined in 2002 and 2003. The studies to determine the number of eggs laid daily by P. aceris on different host plants were conducted on *Acer negundo* L., *Acer pseudoplatanus* L. and *Fraxinus excelsior* L., which were grown in the Ankara University Plant Protection area. For this aim, the host plants were infested by mealybug. When the mealybug reached adulthood, they were put onto separate cells on the leaves and observed daily until all the specimens had died and all the eggs had been counted each day. In this way, the total and daily number of eggs laid by one female, and the duration of the oviposition period, were determined.

The Experiments were replicated at least 15 times for each host plant.

Data used in this study were obtained from Kaydan and Kılınçer [12].

2. Methods

Poisson Regression

Suppose that the dependent variable y_i is distributed according to the Poisson distribution. The logarithm of μ, which is the mean of the Poisson distribution, is assumed to be a linear function of the independent variables (x_i). The PR model is,

$$\text{Pr}(y_i/\mu_i, x_i) = \exp\big(-\mu_i\big)\mu_i^{y_i}\mu_i^{y_i}y_i!,\ y_i = 0, 1, \dots. \tag{2.1}$$

When the independent variables are given, the likelihood function for the PR model is,

$$L(\beta, y_i, x_i) = \sum_{i=1}^{n}\Big[y_i x_i'\beta - \exp\big(x_i'\beta\big) - \ln y_i!\Big] \tag{2.2}$$

In equation (2.2), β is an unknown parameter. The first derivatives of the log-likelihood with respect to β are [10,13],

$$\frac{\partial(L(\beta, y_i, x_i))}{\partial\beta} = \sum\big(y_i - \exp\big(x_i\beta\big)\big)x_i.$$

The ML estimation for β is obtained by setting these equal to zero.

Negative Binomial Regression

The negative binomial regression model uses a log link function between the dependent variable and independent variables. The NB regression model is [14],

$$P(y; \mu, \alpha) = \frac{\Gamma(y + \alpha^{-1})}{y!\Gamma(\alpha^{-1})}\Big(\frac{\alpha\mu}{1 + \alpha\mu}\Big)^{y}\Big(\frac{1}{1 + \alpha\mu}\Big)^{\alpha^{-1}}\ \alpha > 0. \tag{2.3}$$

In equation (2.3), α is an arbitrary parameter showing the overdispersion level. The Log-likelihood function for the NB regression model is [10,16],

$$\ln L(\beta, \alpha, y) = \sum_{i=1}^{n}\left[\sum_{t=0}^{y_i^{-1}}\ln(1 + \alpha t) + y_i \ln\mu_i - (y_i + \alpha^{-1})\ln(1 + \alpha\mu_i)\right].$$

By equating the first derivatives of the log-likelihood with respect to β and α to zero, the ML estimations can be written, respectively, as

$$\frac{\partial \ln L(\beta, \alpha, y)}{\partial\beta_j} = \sum_{i=1}^{n}\frac{y_i - \mu_i}{\mu_i(1 + \alpha\mu_i)}\Big(\frac{\partial\mu_i}{\partial\beta_i}\Big) = 0,\ j = 1, \dots, p,$$

$$\frac{\partial \ln L(\beta, \alpha, y)}{\partial\alpha} = \sum_{i=1}^{n}\left\{\sum_{t=0}^{y_i^{-1}}\Big(\frac{t}{1 + \alpha t}\Big) + \alpha^{-2}\ln(1 + \alpha\mu_i) - \frac{(y_i + \alpha^{-1})\mu_i}{1 + \alpha\mu_i}\right\} = 0$$

Zero inflated Poisson Regression

In order to explain the extra zeros in the variable y_i, the ZIP regression is [15],

$$\text{Pr}(y_i/x_i) = \begin{cases}\pi_i + (1 - \pi_i)\exp(-\mu_i), & y_i = 0,\\ (1 - \pi_i)\exp(-\mu_i)\mu_i^{y_i}/y_i!, & y_i > 0.\end{cases} \tag{2.4}$$

In equation (2.4), π_i represents the probability of the existence of extra zeros. The log likelihood function for the independent variable y_i can be written as [25],

$$\begin{aligned} L &= \sum_{i=1}^{n}\left(I_{y_i=0}\log\big(\pi_i + (1 - \pi_i)e^{-\mu_i}\big) + I_{y_i>0}\log\left(\big(1 - \pi_i\big)\frac{\mu_i^{y_i}e^{\mu_i}}{y_i!}\right)\right)\\ &= \sum_{i=1}^{N} I_{y_i=0}\log\big(\pi_i + \big(1 - \pi_i\big)e^{-\mu_i}\big)\\ &\qquad + I_{y_i>0}\big(\log\big(1 - \pi_i\big) + y_i\log\mu_i - \mu_i - \log y_i!\big)\big). \end{aligned} \tag{2.5}$$

The expression $I.$ occurring in equation (2.5) is the indicator function for the specified event. After this, the parameters μ_i and π_i can be obtained by using the link functions,

$$\log(\mu) = B\beta, \tag{2.6}$$

and

$$\log\left(\frac{\pi}{1-\pi}\right) = G\gamma. \tag{2.7}$$

In equations (2.6) and (2.7), B and G are covariate matrices, β and γ unknown parameter vectors. The ML estimations for β and γ can be obtained by using the EM algorithm [26].

Zero inflated Negative Binomial Regression

In the modeling of a dependent variable y_i with many zero values, an alternative regression method is ZINB. The ZINB regression model [19] is:

$$\Pr(y_i/x_i) = \begin{cases} \pi_i + (1-\pi_i)(1+\alpha\mu_i)^{-\alpha^{-1}}, & y_i = 0, \\ (1-\pi_i)\frac{\Gamma(y_i+\frac{1}{\alpha})}{y_i!\Gamma(\frac{1}{\alpha})}\frac{(\alpha\mu_i)^{y_i}}{(1+\alpha\mu_i)^{y_i+\frac{1}{\alpha}}}, & y_i > 0. \end{cases} \tag{2.8}$$

In equation (2.8), the parameters π_i and μ_i depend on the covariates, and ($\alpha \geq 0$) is an overdispersion parameter. The ZINB log likelihood function [26] for y_i is,

$$\begin{aligned} L(\mu,\alpha,\pi;y) &= \sum_i \Bigg(I_{y_i=0} \log(\pi_i + (1-\pi_i)(1+\alpha\mu_i)^{-\alpha^{-1}} \\ &\qquad + I_{y_i>0} \log\left((1-\pi_i)\frac{\Gamma(y_i+\frac{1}{\alpha})}{y_i!\Gamma(\frac{1}{\alpha})}\frac{(\alpha\mu_i)^{y_i}}{(1+\alpha\mu_i)^{y_i+\frac{1}{\alpha}}}\right)\Bigg) \\ &= \sum_i \Bigg(I_{y_i=0} \log\left(\pi_i + (1-\pi_i)(1+\alpha\mu_i)^{-\alpha^{-1}}\right) \\ &\qquad + I_{y_i>0}\Big(\log(1-\pi_i) - \frac{1}{\alpha}\log(1+\alpha\mu_i) - y_i \log\left(1+\frac{1}{\alpha\mu_i}\right) \\ &\qquad + \log\Gamma\Big(y_i + \frac{1}{k}\Big) - \log\Gamma\Big(\frac{1}{\alpha}\Big) - \log y_i!\Big)\Bigg). \end{aligned} \tag{2.9}$$

The expression $I_\cdot$ occurring in equation (2.9) is the indicator function for the specified event. The model description suggested by Lambert [15] can be given as,

$$\log(\mu) = X\beta \text{ and } \log\left(\frac{\pi}{1-\pi}\right) = G\gamma.$$

Here, X and G are covariate matrices, β and γ are unknown parameter vectors of dimension $(p+1)\times 1$ and $(q+1)\times 1$, respectively. The ML estimations for β, α and γ can be obtained by using the EM algorithm [26].

Poisson Hurdle Model

Positive observations based on truncated count data ($y_i > 0$) are called the *Poisson hurdle model* when they are modeled using the Poisson distribution. We suppose the y_i are independent observations based on count data, and consider that the probability of $y_i = 0$ is $1 - p(x)$, and of $y_i \sim$ truncated Poisson$(\lambda(z))$ is $p(x)$. Here, x and z are covariate matrices. The PH model [8] is:

$$\begin{aligned} P(y_i = 0/x) &= 1 - p(x), \\ P(y_i = q/x, z) &= \frac{p(x)\exp(-\lambda(z))\lambda(z)^q}{q!(1-\exp(-\lambda(z)))},\ q = 1, 2, \ldots. \end{aligned} \tag{2.10}$$

In equation (2.10), $p(x)$ and $\lambda(z)$ are modeled using the logit and log-linear functions, respectively. That is, $\lambda(z)$ and p_i are,

$$\log(\lambda(z)) = x_i'\beta, \tag{2.11}$$

$$\operatorname{logit}(p_i) = z_i'\alpha. \tag{2.12}$$

The β and α given in equation (2.11) and equation (2.12), respectively, are unknown parameter vectors. The Log likelihood for PH can be written as follows:

$$
\begin{aligned}
L &= \sum_{y_i>0} x_i\beta - \sum_{i=1}^{n} \log(1+\exp(x_i\beta)) \\
&\qquad + \sum_{y_i>0} [y_i z_i \alpha - \exp(z_i\alpha) - \log(1-\exp(-\exp(z_i\alpha))) - \log(y_i!)] \\
&= L(\beta) + L(\alpha)
\end{aligned} \tag{2.13}
$$

Estimations of the unknown parameters β and α are obtained by ML using equation (13).

Negative Binomial Hurdle Model

In the negative binomial hurdle, the binomial probability model determining the zero or non-zero results of the basic count dependent variable and the truncated count model, based on the positive count truncated count model, are conjoined in [10] using the following log-likelihood:

$$L = \ln(f(0)) + \{\ln[1-f(0)] + \ln P(t)\}. \tag{2.14}$$

In equation (2.14), $f(0)$ represents the probability of the binary part and $p(j)$ the probability of a positive count. In the case where the logit model is used, the probability of zero is,

$$f(0) = P(y=0; x) = 1/(1+\exp(xb1)).$$

and $1-f(0)$ is $\exp(xb1)/(1+\exp(xb1))$.

The log likelihood function for both parts of the NBH model can be written as:

$$
\begin{aligned}
L = \text{cond}\,\{&y == 0, \ln(1/1-\exp(xb1), \ln(\exp(xb1)/1+\exp(xb1)) \\
&+ y * \ln(\exp(xb)/1+\exp(xb)) - \ln(1+\exp(xb)/\alpha) \\
&+ \ln\Gamma(y+1/\alpha) - \ln\Gamma(1/\alpha) - \ln(1-(1+\exp(xb))(-1/\alpha))\}
\end{aligned}
$$

3. Model selection

The Akaiki Information Criteria (AIC) and Bayesian Information Criteria (BIC) are goodness of criteria used for model selection. Many Monte-Carlo simulations indicate that the BIC and AIC selection criteria need to be used together [8,24]. Generally, they are described as follows:

$$AIC = -2L + 2p \tag{3.1}$$

and

$$BIC = -2L + p\ln(n). \tag{3.2}$$

In equations (3.1) and (3.2), L indicates the log likelihood value, p indicates the parameter number and n indicates sample size.

4. Results

In this study, the necessary analyses were done by using the Stata 10 and R statistical software programs. Daily egg numbers laid on the three hosts by *Phenacoccus aceris* (Signoret) were included in the model as dependent variable, while years, days, and hosts were integrated as independent variables in the model. The distribution of the numbers of eggs laid on various hosts in 2002 is given in Figure 1. When the numbers of daily eggs is examined, it is seen that maximum egg laying takes place on different days for each of the three hosts.

Figure 1. The numbers of eggs laid daily on various hosts by Phenacoccus aceris (Signoret) in 2002

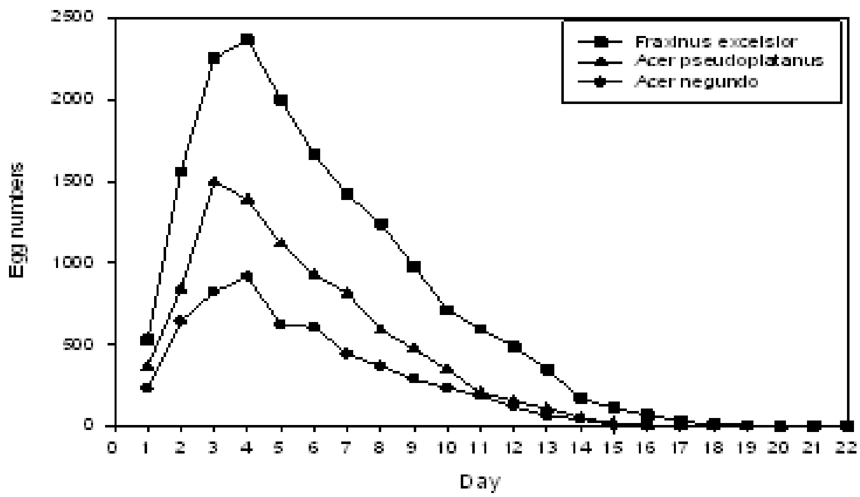
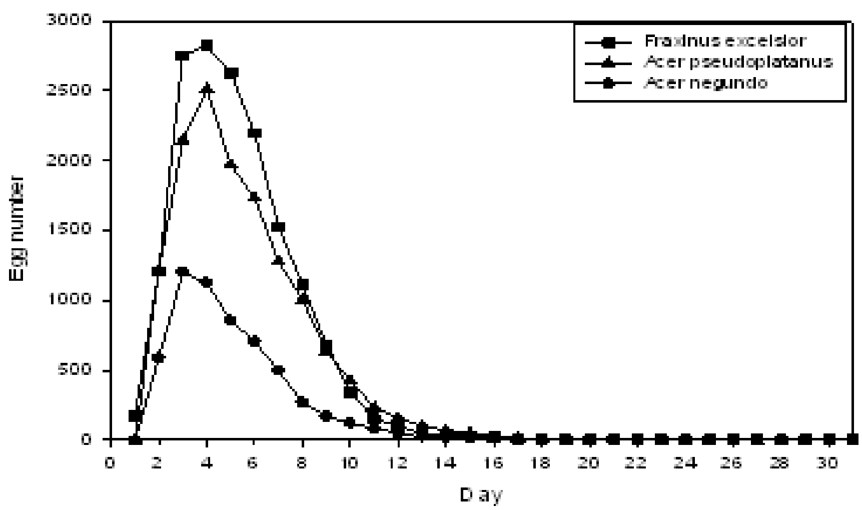


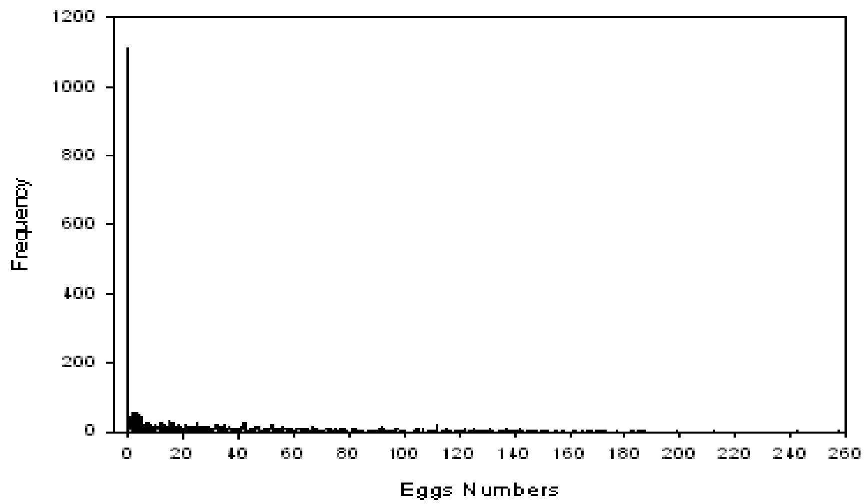
Figure 2. The numbers of eggs laid daily on various hosts by Phenacoccus aceris (Signoret) in 2003



In Figure 2, the daily egg numbers of *P. aceris* nourished on different hosts are shown for 2003. As in 2002, for the three different hosts most of the eggs were laid during the first seven days (605 eggs in F. excelsior, 494 eggs in A. pseudoplatanus, 383 eggs in A. negundo), while almost all of them were left within 15 days.

Of the 2541 observed values used in this study, 1114 (43.84%) had value zero. Daily egg numbers for *P. aceris* are given in Figure 3. The distribution of the data is skewed to the right because of excess zeros.

Figure 3. Daily frequency distribution of eggs



Goodness of statistics determining whether regression methods such as Poisson and logistic are applicable are very essential [21]. In PR analyses, deviance and Pearson Chi-square goodness of statistics indicating overdispersion were obtained as 50.233 and 58.579, respectively. The fact that the goodness of statistics were higher than (1) shows that there was overdispersion in the data set. The AIC BIC model selection criteria for the models PR, NB, ZIP, ZINB, PH, and NBH are given in Table 1. The model selection criteria given in Table 1 produce widely differing results. The model with the smallest AIC and BIC was ZINB regression. Therefore, the ZINB model shown in Table 1 with bold letters was chosen as the best model.

Table 1. Model selection criteria for PR, NB, ZIP, ZINB, PH and NBH

Model	Log-likelihood	AIC	BIC
PR	-30713.550	61433.10	61450.620
ZIP	-17970.360	35946.72	35964.240
NB	-7624.960	15255.92	15273.440
ZINB	**-6278.950**	**12563.90**	**12581.420**
PH	-18890.830	37787.660	37805.181
NBH	-7181.140	14368.280	14385.801

The ML parameter estimations and standard errors obtained for the PR, NB, ZIP, ZINB and NBH regression models are given in Table 2. According to Table 2, while the day laid and the host plant had a significant effect on the number of eggs in PR and NBH ($p < 0.01$), the year was not significant ($p > 0.05$). In the NB, ZIP and ZINB regression models, the year, day laid, and host plant had a significant effect on the number of eggs ($p < 0.01$).

Table 2. Parameter estimations and standard errors for the PR, NB, ZIP, ZINB, PH and NBH models[1] ($p < 0.01$)

Estimation (standard error)				
Method	Intercept	Year	Day	Host plant
PR	5.139[1] (0.158)	0.012 (0.008)	-0.188[1] (0.001)	-0.022[1] (0.005)
NB	7.067[1] (0.128)	-0.267[1] (0.056)	-0.353[1] (0.006)	-0.311[1] (0.036)
ZIP	4.943[1] (0.016)	0.241[1] (0.008)	-0.152[1] (0.001)	-0.258[1] (0.005)
ZINB	5.651[1] (0.085)	-0.080[1] (0.037)	0.216[1] (0.005)	-0.287[1] (0.024)
PH	4.943[1] (0.016)	0.240[1] (0.008)	-0.152[1] (0.001)	-0.258[1] (0.005)
NBH	5.711[1] (0.088)	0.073 (0.039)	-0.222[1] (0.005)	-0.292[1] (0.025)

The Voung statistic has been computed to compare the PR model with the ZIP, and the NB model with the ZINB. The Voung statistics of ZIP against PR was obtained as 11.51 and determined as being significant ($p < 0.01$). The Voung statistics of ZINB against NB was obtained as 21.27 and also determined as being significant ($p < 0.01$)). For a large sample size and under the null hypothesis, the Voung statistic has an assymptotic normal distribution. The likelihood ratio test calculated in order to compare the models of ZIP and ZINB turned out to be very important ($p < 0.01$). That is to say, the ZINB model is better than ZIP. In terms of the results obtained, the goodness of criteria, the Voung statistics, and ratio tests were in parallel with each other.

5. Discussion

In this study, as overdispersion had a great effect, the PR goodness of criteria given in Table 1 was seen to be higher than the other regression models. Some reasons for overdispersion can be explained as the use of a wrong link function, differences between observations, the lack of important terms that need to be in the model, and small sample size [24,27].

The fact that the highest egg numbers for the three hosts were in the third and fourth days of oviposition, with the decrease in egg numbers accelerating during the following days, indicates that the day the eggs are laid is very significant. Similarly, it has been found that the effect of the host is significant, together with the fact that total egg numbers left on the hosts are different from each other. In 2002, most of the eggs left on *A. Pseudoplatanus* were confirmed to have been laid on the third day of oviposition. Moreover, most of the eggs on *F. excelsior* and *A. negundo* appeared on the fourth day of oviposition. In 2003, on the other hand, most of the eggs on A. negundo were seen on the third day of oviposition. Most of the eggs on *F. excelsior* and *A. pseudoplatanus* were seen on the fourth day of oviposition. Most of the eggs were seen in the first half (first 15 days) of the egg laying period. In this period, the egg numbers laid decreases, one or two eggs are seen for a few days till the end of the mature life of the insect, and are generally seen as zero eggs [12].

Senapati and Ghose [22] report that *Planococcoides bengalensis*, Ghose and Ghose (Hemiptera: Pseudococcidae), laid most of their eggs on the second day of the egg laying period, 70-80 % of total egg laying taking place during the first week, and that egg numbers decrease considerably beginning from the 12^{th} day of oviposition. Especially, for insects having quite a long post oviposition period, or for insects laying their eggs during the first days of the oviposition period, the abundance of zero inflated observations makes it necessary to apply zero-inflated methods.

References

[1] Agresti, A. *Categorical Data Analysis* (John and Wiley & Sons, Incorporation, New Jersey, Canada, 1997).

[2] Banik, S. and Kibria, B. M. G. *On some discrete models and their comparisons: An empirical comparative study*, Proceedings of The 5th Sino-International Symposium on Probability, Statistics, and Quantitative Management KU/FGU/JUFE Taipei, Taiwan, ROC May 17, 41–56, 2008 (ICAQM/CDMS, 2008).

[3] Böhning, D. *Zero-inflated Poisson models and C.A.MAN: A tutorial collection of evidence*, Biometrical Journal **40**(6), 833–843, 1998.

[4] Böhning, D., Dietz, E. and Schlattmann, P. *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, 1999.

[5] Cameron, A. C. and Trivedi, P. K. *Regression Analysis of Count Data* (Cambridge University Press, New York, 1998).

[6] Cheung, Y. B.*Zero-inflated models for regression analysis of count data: A study of growth and development*, Statistics in Medicine **21**, 1461–1469, 2002.

[7] Cox, R. *Some remarks on overdispersion*, Biometrika **70**, 269–274, 1983.

[8] Dalrymple, M. L., Hudson, I. L and Ford, R. P. K. *Finite mixture, zero-inflated Poisson and Hurdle models with application to SIDS*, Computational Statistics & Data Analysis **41**, 491–504, 2003.

[9] Frome, E. D., Kutner, M. H. and Beauchamp, J. J. *Regression analysis of Poisson-distributed data*, Journal of American Statistical Association **68** (344), 935–940, 1973.

[10] Hilbe, J. M. *Negative Binomial Regression* (Cambridge, UK, 2007).

[11] Jansakul, N. *Fitting a zero-inflated negative binomial model via R*, In: Proceedings 20th International Workshop on Statistical Modelling. Sidney, Australia, 277–284, 2005.

[12] Kaydan, M. B. and Kılınçer, N. *Investigation on egg laying character of Phenacoccus aceris (Signoret) (Coccoidea:Hemiptera: Pseudococcidae) on different host plant species*, Ankara University Journal of Agriculture **13** (3), 224–230, 2007.

[13] Khoshgoftaar, T. M., Gao, K. and Szabo, R. M. *Comparing software fault predictions of pure and zero-inflated Poisson regression models*, International Journal of Systems Science **36**(9), 707–715, 2005.

[14] Kibria, B. M. G. *Applications of some discrete regression models for count data*, Pakistan Journal of Statistics and Operation Research, **2** (1), 1–16, 2006.

[15] Lambert, D. *Zero-inflated Poisson regression, with an application to defects in manufacturing*. Technometrics **34**(1), 1–13, 1992.

[16] Lawles, J. F. *Negative binomial and mixed Poisson regression*, The Canadian Journal of Statistcs **15** (3), 209–225, 1987.

[17] Lee, A. H., Wang, K. and Yau, K. K. W. *Analysis of zero-inflated Poisson data incorporating extent of exposure*, Biometrical Journal **43**(7), 963–975, 2001.

[18] Long, J. S. and Freese, J. *Regression Models for Categorical Depentent Variable Using Stata* (Stata Press Publication, StataCorp LD Collage Station, Texas, USA , 2006).

[19] Ridout, M., Hinde, J. and Demetrio, C. G. B. *A score test for a zero-inflated Poisson regression model against zero-inflated negative binomial alteratves*, Biometrics **57**, 219–233, 2001.

[20] Rose, C. E., Martin, S. W., Wannemuehler, K. A. and Plikaytis, B. D. *On the zero-inflated and Hurdle models for modelling vaccine adverse event count data*, Journal of Biopharmaceutical Statistics **16**, 463–481, 2006.

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

[22] Senapati, S. K. and Ghose, S. K. *Biology of the mealybug Planococcoides bengalensis Ghosh and Ghose (Homoptera: Pseudococcidae)*, Environment & Ecology **6**, 648–652, 1998.

[23] Stokes, M. E., Davis, C. S. and Koch, G. G. *Categorical Data Analysis Using the SAS System* (John Wiley & Sons Incorporated, USA, 2000).

[24] Wang, P., Puterman, M. L., Cockburn, I. M. and Le, N. *Mixed Poisson regression models with covariate dependent rates*, Biometrics **52**, 381–400, 1996.

[25] Yau, K. K. W. and Lee, A. H. *Zero-inflated Poisson regression with random effects to evaluate an occupational injury prevention programme*, Statistics in Medicine **20**, 2907–2920, 2001.

[26] Yau, Z. *Score Tests for Generalization and Zore-Inflation in Count Data Modeling* (Unpublished Ph.D. Dissertation, University of South Caroline, Columbia, 2002).

[27] Yesilova, A. *The Use of Mixed Poisson Regression Models for Categorical Data in Biology* (Unpublished Ph.D. Dissertation, Yüzüncü Yıl University, Van, Turkey, 2003).