SAKARYA ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ DERGİSİ
SAKARYA UNIVERSITY JOURNAL OF SCIENCE

e-ISSN: 2147-835X
Dergi sayfası: http://dergipark.gov.tr/saufenbilder

Geliş/Received
02.11.2016
Kabul/Accepted
16.01.2017

Doi
10.16984/saufenbilder.298981

Accounting for Zero Inflation of Mussel Parasite Counts Using Discrete Regression Models

Emel Çankaya[1*], Olcay Alpay[1], Ahmet Özer[2]

ABSTRACT

In many ecological applications, the absences of species are inevitable due to either detection faults in samples or uninhabitable conditions for their existence, resulting in high number of zero counts or abundance. Usual practice for modelling such data is regression modelling of log(abundance+1) and it is well know that resulting model is inadequate for prediction purposes. New discrete models accounting for zero abundances, namely zero-inflated regression (ZIP and ZINB), Hurdle-Poisson (HP) and Hurdle-Negative Binomial (HNB) amongst others are widely preferred to the classical regression models. Due to the fact that mussels are one of the economically most important aquatic products of Turkey, the purpose of this study is therefore to examine the performances of these four models in determination of the significant biotic and abiotic factors on the occurrences of Nematopsis legeri parasite harming the existence of Mediterranean mussels (Mytilus galloprovincialis L.). The data collected from the three coastal regions of Sinop city in Turkey showed more than 50% of parasite counts on the average are zero-valued and model comparisons were based on information criterion. The results showed that the probability of the occurrence of this parasite is here best formulated by ZINB or HNB models and influential factors of models were found to be correspondent with ecological differences of the regions.

Keywords: Mediterranean mussel parasites, Poisson regression, Negative binomial regression, Zero-inflated regression, Hurdle model

Kesikli Regresyon Modellerinin Kullanımıyla Midye Parazit Sayılarındaki Sıfır Yoğunluğunun Açıklanması

ÖZ

Ekolojik çalışmaların çoğunda, ya örneklemlerdeki belirleme hataları ya da türlerin varlıkları için elverişsiz yaşam koşulları sebebiyle, türlerin yoklukları kaçınılmazdır ve bunun sonucu olarak çok fazla miktarda sıfır sayısı ya da bolluğu ortaya çıkar. Bu tipteki veriler için genellikle log(bolluk+1) regresyon modellemesi yapılır ve oluşturulan modelinin kestirim için yetersiz olduğu bilinen bir gerçektir. Sıfır sayılı bollukları içeren, aralarında sıfır ağırlıklı regresyon (ZIP ve ZINB), engelli-poisson (HP) ve engelli negatif binom (HNB) bulunan yeni kesikli modeller klasik regresyon modellerine göre daha çok tercih edilmektedir. Türkiye için ekonomik açıdan en önemli deniz ürünlerinden biri olması nedeniyle, bu çalışmada Akdeniz midyelerinin (Mytilus galloprovincialis L.) varoluşlarının tehditi Nematopsis legeri paraziti için önemli biyotik ve abiyotik faktörlerin belirlenmesinde bu dört modelin performanslarının kıyaslanması amaçlanmıştır. Türkiye'nin Sinop şehrinin üç kıyı bölgesinden toplanan parazitlerin sayılarının ortalamada %50'den fazlasının sıfır değerli olduğu görülmüş ve oluşturulan modellerin kıyaslaması bilgi

[*] Sinop University, Faculty of Arts and Sciences, Sinop, Turkey – ecankaya@sinop.edu.tr
1 Sinop University, Faculty of Arts and Sciences, Sinop, Turkey – olcayb@sinop.edu.tr
2 Sinop University, Faculty of Fisheries and Aquatic Sciences, Sinop, Turkey – aozer@sinop.edu.tr

kriterleri ile yapılmıştır. Sonuçlar bu parazitlerin oluşma olasılıklarını en iyi ZINB ve HNB modelleriyle ifade edildiğini göstermiştir ve modellerde etkili faktörlerin bölgelerin çevresel farklılıklarıyla ilişkili olduğu bulunmuştur.

Anahtar Kelimeler: Akdeniz midyesi parazitleri, Poisson regresyon, Negatif binom regresyon, Sıfır ağırlıklı regresyon, Engelli model

1. INTRODUCTION

Communities living within the ecosystems are subject to the integrated, simultaneous influences of multiple processes [1]. The most of the ecological studies have thus been focused on understanding the controlling mechanisms of species occurrence which is generally described by the distribution or abundance of individual species. It is therefore vital to develop a satisfactory theory that is able to make quantitative prediction about the probability of occurrence of such organisms based on environmental and biological factors as precisely as possible.

It is well known that modelling counts of species occurrences or abundances are mostly inappropriate by the classical regression methods because the underlying assumptions such as normality, homoscedasticity etc. are not satisfied, even after the transformation of the data [2]. Although the most natural choice for analyzing species frequencies is the Poisson regression, the distribution of such data often displays over-dispersion, larger variability than expected, which violates the basic equi-dispersion assumption of the Poisson distribution [3]. This restrictive nature has spawned the development of Negative Binomial regression methods that account for over-dispersion with a flexible parameter structure. It is generally most appropriate to describe the distribution of the counts if the over-dispersion is only due to unobserved heterogeneity e.g. ([4], [5]).

In modelling abundance data, it is often necessary to account for excess zeros resulting from the absence of specie from many observational units. Absence may be observed for two reasons; either it is not detected when it is in fact present (accidental/sampling zeros), or its presence at that location or time point is impossible (inevitable/structural zeros). Whatever its cause, such high percentage of zeros contributes to the problem of over-dispersion and in that case, NB may not be appropriate to model the present variability as it underestimates the probability of zeros [6].

Zero inflated models (ZI), also known as conditional models, are typically used when the excess zeros are a mixture of two types of zeros- structural zeros and sampling zeros e.g. ([7] and [8]). Alternative models, namely Hurdle models, are preferred to overcome the problem of over-dispersion due to excess of structural zeros. With this model, all zeros can be described using the binary link, whereas positive counts can be represented via a zero truncated Poisson or NB model [9].

Native to the Mediterranean coastline, Mytilus galloprovincialis L. mussel species is one of the important aquatic products distributing over the littoral fauna of Black Sea. Due to its economic importance to humans, its distribution and impacts (ecological and socio-economic) have been well studied in its other habitats ranging from Mediterranean Sea to Adriatic Sea. However, parasitic pathogens along with several other stress factors can lower the resistance of the mussels and their ability to adapt to changing environmental conditions [10]. Although there have been studies on the impacts of parasitological diseases on both natural and cultivated mussel populations e.g. ([11], [12], [13], [14], [15], [16]), we found no records for the modelling the parasite counts of mussels utilizing above mentioned advanced regression methods.

Therefore, the focus of this study is twofold: we first fit the various models to explore the ecological determinants of the infections by a mussel parasite, namely Nematopsis legeri. Even though different hosts are exposed to the same environmental conditions, they may experience the infections at a different rate due to the uncontrolled covariate variation among subjects or events. Our second primary goal is therefore to assess the abilities of the considered models for the prediction of this parasite counts differing on a wide range, 50% of which includes zero values. Although the application of these models to the analysis of ecological data are not new, to the best of our knowledge, the parasitological analysis of Mediterranean Mussels via these models will be the first record for the area of our sampling.

2. MATERIAL AND METHOD

2.1. Data

Mussels were collected from three locations of Sinop coastline in the Black Sea during the period of August 2012 and July 2013. All locations differ from each other with respect to their ecologically properties; İskele region is a polluted area due to human activities at the inner harbor, Adabaşı region is the most natural and unpolluted area and the last region Sarıada is a site under the effect of both human activities and seasonal fresh water flowing which lowers the salinity level of the water. Details about the collection of mussels and laboratorial processes of parasite identification can be found in [17]. Temperature, Salt, Phosphate-PO4 and Nitrate-NO3 measurements were included to the models as biotic factors; and Sex, Condition Index-CI (i.e. percentage of wet meat weight in total having excluded the shell weight) were also examined as abiotic factors. Statistical summaries of all measurable factors are listed in Table 1. Due to the excessive ecological disparities amongst the three regions, we here preferred to model them separately.

As the size of mussels differs in some amount, we preferred to model parasite counts per unit length instead of counts per mussel that is achieved by including length measurements to the all models as the offset variable (i.e. same scale of exposure).

Table 1. Descriptive statistics of investigated factors

	İskele Region (n=284)		**Adabaşı Region (n=327)**		**Sarıada Region (n=321)**	
	Mean	**StDev**	**Mean**	**StDev**	**Mean**	**StDev**
N. legeri	17.34	49.46	36.20	91.51	275.60	510.30
Temperature	14.42	5.41	15.27	4.81	15.60	5.11
Salt	17.17	0.76	17.15	0.70	17.15	0.67
Nitrate	6.51	0.45	5.86	1.74	5.53	1.34
Phosphate	0.53	0.08	0.55	0.08	0.58	0.11
CI	32.21	9.28	31.80	11.58	31.02	12.88

2.2. Statistical Analyses

The type of regression models we consider here are defined within the generalized linear modelling framework. Models that quantify the relationships between species abundances and environmental characteristics are here typically specified through a 'link' function. For the count response variable, the relationship is generally expressed using a 'log link' [18] as:

$$\log(\lambda) = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k \qquad (1)$$

where μ is the mean rate of the simplest and most commonly specified Poisson model at which an event occurs. β_i are the coefficients that reflect how effective the i^{th} variable is on this relationship. Poisson model for count data, y_i, is therefore:

$$P(Y = y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \quad y_i = 0,1,.. \qquad (2)$$

A major assumption in this model is that mean and variance of observed counts are equal. However, real data generally exhibits more variability than a Poisson model can explain, which often gives misleading relationships between the count outcomes and the predictors.

Violation of the equi-dispersion assumption for Poisson model affects standard errors of the parameter estimates and model fit [5]. Therefore, adaptations are applied to the Poisson model to deal with the overdispersed or underdispersed data. Negative Binomial (NB) is generally used to model overdispersed Poisson data.

$$P(Y = y_i) = \frac{\Gamma(y_i + \theta^{-1})}{\Gamma(\theta^{-1})\Gamma(y_i + 1)} \left(\frac{\theta\lambda_i}{1+\theta\lambda_i}\right)^{y_i} \left(\frac{1}{1+\theta\lambda_i}\right)^{\theta^{-1}} \quad y_i = 0,1,... \qquad (3)$$

where random variable Y has a negative binomial distribution with parameters $\theta \geq 0$ When $\mu \geq 0$ the variance of NB distribution generally exceeds its mean [5].

Zero-Inflated (ZI) models based on the mixture of two-components first of which has a degenerate distribution at zero and the other is a count distribution. Let π be the probability of observing structural zero, a basic ZI model is defined as follows:

$$P(Y = y_i) = \begin{cases} \pi_i + (1-\pi_i)P(S_i = 0) & y_i = 0 \\ (1-\pi_i)P(S_i = y_i) & y_i > 0 \end{cases} \qquad (4)$$

where the probability of the random variable S can be produced from any discrete distribution. When the distribution is selected as Poisson or Negative Binomial, the model is called zero-inflated Poisson (ZIP) or Zero-Inflated Negative Binomial (ZINB) respectively.

Similarly to the ZI models, the hurdle models consist of two parts: the first is a binary outcome model, and the second is a zero-truncated count distribution for the positive observations [19]. With the Poisson and Negative Binomial alternative distributions, Hurdle models abbreviated as HP and HNB are defined respectively as follows:

$$P(Y = y_i) = \begin{cases} w_i & y_i = 0 \\ \frac{(1-w_i)}{(1-e^{-\mu_i})} \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} & y_i > 0 \end{cases} \qquad (5)$$

and

$$P(Y = y_i) = \begin{cases} w_i & y_i = 0 \\ (1-w_i) \frac{\Gamma(y_i + \theta^{-1})}{y_i!\,\Gamma(\theta^{-1})} \frac{(1+\theta^{-1}\mu_i)^{-(y_i+\theta^{-1})}\theta^{-y_i}\mu_i^{y_i}}{1-(1+\theta^{-1}\mu_i)^{\theta^{-1}}} & y_i > 0 \end{cases} \qquad (6)$$

where $w_i = P(Y_i = 0)$.

Expressions for the mean and variance of positive counts under all regression models were listed in Table 2, which shows how much each model incorporates the existing overdispersion. All necessary analyses have been performed using the programs; R v.2.14 and SAS University Edition.

Table 2. Mean and variance of positive counts for all models

Model	Mean	Variance
Poisson Regression	μ	μ
Negative Binomial	μ	$\mu(1+\theta\mu)$
Zero-Inflated Poisson	$(1-\pi)\mu$	$E(Y)[1+\pi\mu]$
Zero-Inflated NB	$(1-\pi)\mu$	$E(Y)[1+\theta\mu+\pi\mu]$
Hurdle Poisson	$\frac{(1-w)\mu}{1-e^{-\mu}}$	$E(Y)[1+\mu-E(Y)]$
Hurdle NB	$\frac{(1-w)\mu}{1-(1+\mu\theta)^{-1/\theta}}$	$E(Y)[1+\mu+\mu\theta-E(Y)]$

2.3. Model Comparison

Traditional approach for the comparison of nested models is to utilize likelihood ratio test which is appropriate for the selection between, for example, the Poisson and the NB, or the ZIP and the ZINB as the first is nested in the second. For non-nested models, however, an information theory based criteria is suggested to be used [20]. As they are commonly used metrics, we here applied Akaike information criteria (AIC), Bayes information criteria (BIC) and corrected Akaike information criteria (AICc) for the comparison of fitted models [21]. The AIC is defined as:

$$AIC = -2LogL + 2k \tag{7}$$

where L is the maximized value of the likelihood function for the estimated model, penalized by the number of parameters k in the model [5].

BIC introduced by Schwarz [22] serves as an asymptotic approximation to a transformation of the Bayesian posterior probability of a candidate model [23]. With the sample size n addition to the penalty term, BIC is specified as:

$$BIC = -2LogL + \ln(n)k \tag{8}$$

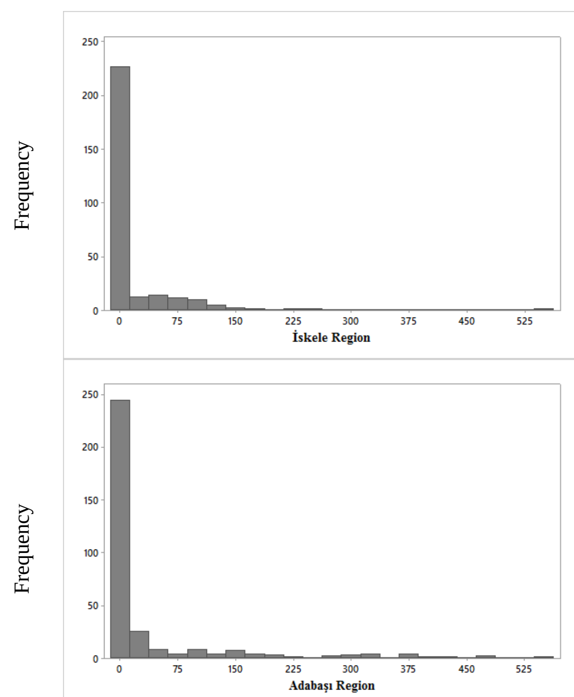
AICc is defined by replacing the penalty term of AIC *2k* with the exact expression for the bias adjustment [23]:

$$AICc = AIC + \frac{2k(k+1)}{(n-k-1)} \tag{9}$$

The model with the lowest value of these information criterions would be taken as the best model in the comparison procedure.

3. RESULTS

Some statistical summaries of both abundances of Nematopsis legeri and factors (biotic and abiotic) investigated in Table 1 reflect the major differences in the intensity of this type of parasitism, which can also be visualized in Figure 1. Sarıada region has the highest abundance and İskele region is the least. Observed inequality of the means and variances violates the assumption of Poisson regression mentioned before. The apparent overdispersion is here mainly caused by high percentages of zeros (79% for İskele, 71% for Adabaşı and 22% for Sarıada) as well as heterogeneities in the variables. We first therefore fit an NB model to parasite counts of three regions. Results in Table 3 suggest that Season has the main influence on the Nematopsis legeri counts for all regions. Temperature and Nitrate are the common influential factors for Adabaşı and Sarıada regions. Amongst the others, the count data of Sarıada seem to be prone to the changes of more factors. Adabaşı region still suggest overdispersion by the value of Pearson statistic.



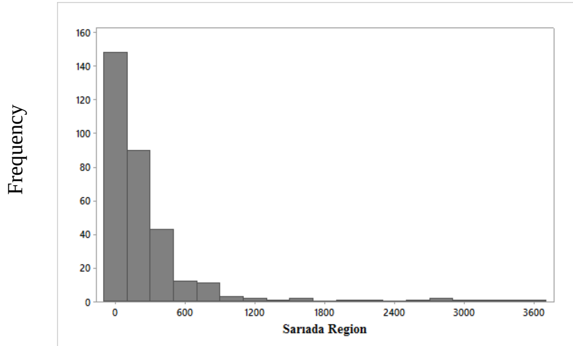


Figure 1. Distribution of Nematopsis legeri for three regions

Determination of which factors to be used in both components of the mixture models, ZI and H, was made by considering the only ones having the significance below 0.25 from the NB models of Table 3. Then, models accounting for excess zeros, ZIP, ZINB, HP and HNB, were constructed and the parameter estimates were obtained as in Table 5-7.

For all regions, performances of these fitted models based on the difference between the actual and the predicted number of zeros were evaluated using the AIC, AICc and BIC. The results in Table 4 suggested that NB model did not perform well as the number of zeros appeared to be over/underestimated for all the regions. The rest of the models fitted for İskele region seem to be equivalently suitable if the prediction performance is of the only concern. Based on the information criterion for the comparison of these models however presented that, moving from Poisson to NB made great improvement in the modelling scope of view. HNB appeared to be the most appropriate model for İskele region. Amongst the considered models for Adabaşı region, the lowest value of calculated information criterion suggested HNB as the best model with a very close prediction percentage of observed zeros. Sarıada region differs ecologically from the others with its high infection intensities which bring about the lowest percentage of zeros in the region. This is also reflected in the model performance as ZINB, differently from the others, is the best model this time to represent the overall pattern in this region.

Counts of Nematopsis legeri parasite could be driven by the biotic factor of “Nitrate” that is the only significant factor in HNB model of İskele region (Table 5). Season, Temperature and Sex are the meaningful factors selected by again HNB model for Adabaşı region (Table 6). Sarıada is the most distinctive region influenced by higher number of influential environmental factors; Season, Salt, Nitrate and Phosphate, as seen in ZINB model of Table 7.

Table 3. Parameter estimates (p-values) for negative binomial regression

Negative Binomial	**İskele Region**	**Adabaşı Region**	**Sarıada Region**
Intercept	-103.62144 **(0.0007*)**	19.73552 **(0.0421*)**	28.57003 **(0.0000*)**
Season	1.84285 **(0.0003*)**	-2.98374 **(0.0000*)**	-1.31227 **(0.0000*)**
Temprature	0.23551 (0.2415)	0.19657 **(0.0000*)**	-0.06318 **(0.0000*)**
Salt	5.72192 **(0.0004*)**	-0.75629 (0.1380)	-1.17469 **(0.0000*)**
Nitrate	-1.32959 (0.1863)	-0.54996 **(0.0000*)**	-0.40810 **(0.0000*)**
Phosphate	10.70273 (0.3336)	1.52818 (0.5616)	2.99367 **(0.0000*)**
CI	0.01159 (0.5970)	-0.02055 (0.2743)	0.00045 (0.9294)

Table 4. Model comparisons based on information criterion

İskele Region (#zero=224 &	**NB**	**ZIP**	**ZINB**	**HP**	**HNB**
AIC	1025.1	3096.6	943.6	2748.5	643.2
AICc	1025.6	3097.0	944.2	2750.1	645.3
BIC	1054.3	3122.0	972.8	2761.2	658.0
-2Log Likelihood	1009.1	3082.6	927.6	2736.5	629.2
Predicted number of	230	223	223	223	223
Zero percentage	80.9%	78.5%	78.5%	78.5%	78.5%
Adabaşı Region (#zero=232 &	**NB**	**ZIP**	**ZINB**	**HP**	**HNB**
AIC	1475.1	5591.9	1400.0	5218.3	1021.9
AICc	1475.6	5592.2	1400.5	5219.2	1023.1
BIC	1505.4	5618.4	1430.3	5233.7	1039.8
-2Log Likelihood	1459.1	5577.9	1384.0	5206.3	1007.9
Predicted number of	194	231	166	231	231
Zero percentage	59.3%	70.6%	50.8%	70.6%	70.6%
Sarıada Region (#zero=71 &	**NB**	**ZIP**	**ZINB**	**HP**	**HNB**

AIC	7046.4	74026.2	3618.2	78788.0	3698.6
AICc	7046.9	74026.7	3618.8	78789.4	3699.2
BIC	7076.6	74056.4	3652.1	78840.8	3726.8
-2Log Likelihood	7030.4	74010.2	3600.2	78760.0	3682.6
Predicted number of	116	71	71	70	70
Zero percentage	36.1%	22.1%	22.1%	21.8%	21.8%

Table 5. Parameter estimates (p-values) for zero-inflated models - İskele region

Count Part	ZIP	ZINB	HP	HNB
Intercept	19.2519 **(0.0000*)**	16.0859 (0.1324)	14.1484 **(0.0000*)**	8.4413 (0.4166)
Season	-0.1310 **(0.0007*)**	-0.0936 (0.6854)	-0.1394 **(0.0004*)**	-0.0375 (0.8668)
Temperature	0.0158 (0.0052)	0.0360 (0.3551)	0.0379 **(0.0000*)**	0.0647 (0.0929)
Salt	-0.6628 **(0.0000*)**	-0.5031 (0.3262)	-0.4548 **(0.0000*)**	-0.1908 (0.7022)
Nitrate	-0.4986 **(0.0000*)**	-0.4953 **(0.0354*)**	-0.5584 **(0.0000*)**	-0.4799 **(0.0355*)**
Sex	0.06964 **(0.0205*)**	-0.0047 (0.9802)	-0.0486 (0.1057)	-0.1431 (0.4389)

Table 6. Parameter estimates (p-values) for zero-inflated models - Adabaşı region

Count Part	ZIP	ZINB	HP	HNB
Intercept	-3.1322 (0.0864)	14.5342 **(0.0287*)**	-3.2332 (0.0721)	2.7583 (0.7058)
Season	-1.0901 **(0.0000*)**	-2.3651 **(0.0000*)**	-1.2199 **(0.0000*)**	-1.3231 **(0.0000*)**
Temperature	0.0239 **(0.0000*)**	0.1654 **(0.0000*)**	0.0132 **(0.0090*)**	0.0884 **(0.0146*)**
Salt	0.4225 **(0.0000*)**	-0.4419 (0.2002)	0.3603 **(0.0003*)**	0.0071 (0.9856)
Nitrate	0.2768 **(0.0000*)**	-0.3055 **(0.0037*)**	0.2486 **(0.0000*)**	0.0839 (0.4094)
Sex	-0.2475 **(0.0000*)**	-0.4693 **(0.0471*)**	-0.3235 **(0.0000*)**	-0.3210 **(0.0322*)**

Table 7. Parameter estimates (p-values) for zero-inflated models - Sarıada region

Count Part	ZIP	ZINB	HP	HNB
Intercept	23.1829 **(0.0000*)**	26.2509 **(0.0000*)**	20.4580 **(0.0000*)**	24.3050 (0.0794)
Season	-0.9678 **(0.0000*)**	-1.1650 **(0.0000*)**	-0.8470 **(0.0000*)**	-1.3118 **(0.0072*)**
Temperature	0.0064 **(0.0000*)**	-0.0059 (0.6854)	0.0210 **(0.0212*)**	-0.0274 (0.6055)
Salt	-0.8447 **(0.0000*)**	-0.9810 **(0.0000*)**	-0.7704 **(0.0000*)**	-1.1446 (0.1120)
Nitrate	-0.4492 **(0.0000*)**	-0.4386 **(0.0000*)**	-0.4989 **(0.0000*)**	-0.5082 **(0.0119*)**
Phosphate	2.7505 **(0.0000*)**	2.5012 **(0.0000*)**	3.7867 **(0.0000*)**	2.5739 (0.1392)
Sex	-0.1167 **(0.0000*)**	-0.1389 (0.3161)	-0.3684 **(0.0016*)**	-0.1902 (0.7085)

4. DISCUSSION

The count nature of an investigated variable implies the use of count regression models, most frequently Poisson or Negative Binomial models. Despite the developments of their new variants to overcome the problem of overdispersion, counts of organisms in ecology these models are still largely preferred because they are easily interpretable. However, it is likely that the observed overdispersion in such studies is caused by large number of zero counts which implies the usage of dual regime count models. Our results for the problem in hand also suggested that the process generating the parasite count is governed by a two-step structure. If the degree of zero excessiveness based on our data is around seventy percent (İskele and Adabaşı regions), Hurdle Negative Binomial model appeared to be substantially superior to the Zero-Inflated models. On the other hand, lower percentages like around twenty (Sarıada region) suggested Zero-Inflated Negative Binomial model as a better model than others. Noting that there are two, structural or sampling, types of zeros, it has been suggested hurdle models are well-suited to represent the sampling zeros while zero-inflated models should be preferred when the investigated problem includes mixture of two types [24]. This fact affects the considered model's ability to better "account for" zero excessiveness in the data.

In the present study, İskele region has rare organism population because this area is exposed to large human activity. Adabaşı region is the most natural ecosystem but intense flows around it make the environment unsuitable for organisms to live. Therefore observed zeros in these areas are more structural. On the other hand, Sarıada region provides more convenient environment for mussels and crabs to populate, which results in high levels of infections by parasites. The observed zeros here could be caused by unsuitable seasonal conditions, and laboratorial or sampling procedures. This might explain why zero-inflated model appeared more suitable.

The general understanding these models from this study can help ecologists to evaluate their data generating processes and make a choice between these models.

Acknowledgements

The authors are grateful to the Scientific and Technological Research Council of Turkey (TÜBİTAK) for sponsoring the project 112O337.

REFERENCES

[1] H. Greene, "Organisms in nature as a central focus for biology," Trends in Ecology and Evolution,

vol. 20, pp. 23-27, 2005.

[2] R. O’Hara and J. Kotze, "Do not log-transform count data," Methods in Ecology and Evolution, vol. 1, pp. 118-122, 2010.

[3] A. Agresti, Categorical Data Analysis, New York: John Wiley& Sons Inc., 2002.

[4] J. Hilbe, Negative Binomial Regression 2nd Edition, UK: Cambridge University Press, 2001.

[5] M. Chipeta, B. Ngwir, C. Simoonga and L. Kazembe, "Zero adjusted models with applications to analysing helminths count data," November 2014. [Online]. Available: http://bmcresnotes.biomedcentral.com/articles/10.1186/1756-0500-7-856.

[6] W. Greene, "Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models," NYU Working Paper No. EC-94-10, 1994.

[7] D. Lambert, "Zero-inflated Poisson regression with an application to defects in manufacturing," Technometrics,, vol. 34, pp. 1-14, 1992.

[8] J. Prasad, Zero-Inflated Censored Regression Models: An Application with Episode of Care Data, USA: M.S. thesis, Dept. Statistics, Brigham Young University, 2009.

[9] J. Cragg, "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods," Econometrica, vol. 39, pp. 829-844, 1971.

[10] H. Williams and A. Jones, Parasitic worms of fish, London: Taylor & Francis, 1994.

[11] G. Murina and A. Solonchenko, "Commensals of Mytilus galloprovincialis in the Black Sea: Urastoma cyprinae (Turbellaria) and Polydora ciliata (Polychaeta)," Hydrobiol, vol. 227, pp. 385-387, 1991.

[12] J. Robledo, J. Caceres-Martinez, R. Suluys and A. Figueras, "The parasitic turbellarian Urastoma cyprinae (Platyhelminthes: Urastomidae) from blue mussel Mytilus galloprovincialis in Spain: occurrence and pathology," Dis Aquat Org, vol. 18, pp. 203-210, 1994.

[13] A. Rayyan, G. Photis and C. Chintiroglou, "Metazoan parasite species in cultured mussel Mytilus galloprovincialis in the Thermaikos Gulf (North Aegean Sea Greece)," Dis Aquat Org, vol. 58, pp. 55-62, 2004.

[14] A. Gaevskaya, "Parasites Diseases and Pests of Mussels (Mytilus Mytilidae)," in I. Protozoa, Sevastopol, 2006.

[15] F. Kayhan, N. Balkıs ve A. Aksu, «İstanbul Balık Halinden Alınan Akdeniz Midyelerinde (Mytilus galloprovincialis) Arsenik Düzeyleri,» Ekoloji, cilt 61, pp. 1-5, 2006.

[16] S. Bhaby, O. Belhsen, A. Errhif and N. Tojo, "Seasonal Dynamics of Parasites on Mediterranean Mussels (Mytilus Galloprovincialis) and Ecological Determinants of the Infections in Southern Alboran Area Morocco," International Journal of Parasitology Research, vol. 5, pp. 116-121, 2013.

[17] A. Ozer and S. Guneydag, "First report of some parasites from Mediterranean mussel Mytilus galloprovincialis Lamarck 1819 collected from the Black Sea coast at Sinop," Turkish Journal of Zoology, vol. 38, pp. 486-490, 2014.

[18] P. McCullagh and J. Nelder, Generalized Linear Models, 2nd Edition, New York: Chapman & Hall, 1989.

[19] J. Potts and J. Elith, "Comparing species abundance models," Ecological Modelling, vol. 199, pp. 153-163, 2006.

[20] G. Sileshi, "Selecting the right statistical model for analysis of insect count data by using information theoretic measures," Bull. Entomol. Res., vol. 96, pp. 479-488, 2006.

[21] S. Richards, "Dealing with overdispersed count data in applied ecology," Journal of Applied Ecology, vol. 45, pp. 218-227, 2008.

[22] G. Schwarz, "Estimating the Dimension of a Model," Annals of Statistics, vol. 6, pp. 461-464, 1978.

[23] A. Neath and J. Cavanaugh, "The Bayesian information criterion: background derivation and applications," WIREs Computational Statistics, vol. 4, pp. 199-203, 2012.

[24] P. Grootendorst, "A comparison of alternative models of prescription drug utilization," Health Economics, vol. 4, pp. 183-198, 1995.