Analysis of Over-Dispersed Count Data: Application to Obligate Parasite *Pasteuria penetrans*

IOANNIS VAGELAS
Department of Agriculture Crop Production and Rural Environment,
University of Thessaly, Volos,
GREECE

Abstract: - In this article we present with STATA regression models suitable for analyzing over-dispersed count outcomes. Specifically, the Negative Binomial regression can be an appropriate choice for modeling count variables, usually for over-dispersed count outcome variables. The common problem with count data with zeroes is that the empirical data often show more zeroes than would be expected under either Poisson or the Negative Binomial model. We concluded, this publications showcases that Zero-inflated models can be used to model count data that has excessive zero counts.

Keywords: - Biological control; *P. penetrans*; over-dispersion; Excess zero-count data; Zero Inflation.

1 Introduction

Plant-parasitic nematodes such as the *Meloidogyne* species are recognized as major agricultural pathogens worldwide. *Meloidogyne javanica* is one of the most damaging crop parasites often causing heavy losses. Nowadays, the most reliable practices to control the pathogen are preventive e.g. crop rotation including the choice of plant varieties or the use of biological control agents such as the obligate hyperparasitic bacterium Pasteuria penetrans [1]. The literature review shows that the most widely studied bacterial pathogen of *Meloidogyne* species (root-knot nematodes) is in the genus *Pasteuria*. *Pasteuria penetrans* is a mycelial, endospore-forming, bacterial parasite that has shown remarkable potential as a biological control agent of second-stage juvenile (J2) of root-knot nematodes. The biological control potential of *Pasteuria* spp. has been demonstrated on many crops and has been reported to develop endospores only in females of *Meloidogyne* spp. [1]. Based on previous research [2], attachment count data were observed to be over-dispersed concerning high numbers of spores attaching on each J2 at 6 and 9 h after spore application. It was concluded that the negative binomial distribution was found to be the most acceptable model to fit the observed data sets considering that *P. penetrans* spores are clumped. This issue of over-dispersion with zeros exists in a dataset [2] we recently analyzed. Based on this class of distributions, we tested two approaches to adjust the over-dispersed count data with zeros [3-5]. The first approach was to scale the variance of the Poisson distribution by submitting a dispersion parameter and multiplying it by the variance. The second approach was to test another probability distribution to handle the count data dispersion, such as the Negative binomial the Zero-inflated Poisson (ZIP) or the Zero-inflated negative binomial (Zinb) model.

Overall, in this paper, we employed and compare these different models with a particular focus, on the over-dispersed count data with zeros.

Moreover, this paper attempts to encourage researchers dealing with biological data not to ignore the over-dispersion which statistically influence the conclusions by underestimating the variability of the data.

2 Materials and Methods

2.1 *Meloidogyne* spp. Culture

A culture of *M. javanica* was maintained on tomato plants (cherry tomato variety Tiny Tim) in the glasshouse. Eggs were collected by dissolving the gelatinous matrix into a solution of 0.5% sodium hypochlorite (NaOCl) (10% commercial bleach), passing the solution through a 200-mesh (75 μm) sieve, nested over a 500-mesh (26 μm) sieve and rinsing the eggs under slow running tap water to remove residual NaOCl [6]. Second stage juveniles (J2) were then hatched using standard laboratory practices [7].

2.2 *P. penetrans* Spore Attachment Data

Attachment tests on freshly hatched J2 were conducted in 2.5-cm Petri dishes using standard techniques [8]. Spore suspension of *P. penetrans* (Nematech Co. Ltd Japan) were prepared in tap water [9], and fresh J2s of root-knot nematodes were exposed to 5000 spores per Petri dish [10]. All dishes were placed in a 28°C incubator. Nematodes were observed under an inverted microscope at ×200 magnification [10] and numbers of *P. penetrans* spores attached per nematode were recorded (Table 1). For the data-set (Table 1), a total of 36 random nematodes were examined for *P. penetrans* spore attachment (att) after incubation of the Petri dishes at 28°C for 1, 3, 6 and 9 h.

2.3 Statistical Analysis

In this research, we use a dataset Table 1 based on the attachment counts of the bacterium *P. penetrans* spores to root-knot nematodes cuticle. This dataset which is presented in the results as Table 1, contains four independent variables (time 1, 3, 6, and 9hrs) and is analyzed with STATA 9.1 for Windows Statistical Software [11], to demonstrate the application of Poisson, negative binomial, over-dispersed Poisson, and Zero-inflated Poisson approach to modeling over-dispersion in count data and explain with those models the excess zeroes [12-14].

Moreover, to estimate these calculates the commands used to produce the STATA output were *sum ()* for descriptive statistics, *glm (), family(poisson) nolog* for poisson model, *nbreg (), nolog* for Negative binomial regression and *zip (), inflate () nolog* for zero-inflated Poisson model. All the above commands are on Stata's manual, e.g., stata.com/help.cgi?poisson for the Poisson regression, stata.com/help.cgi?nbreg for the Negative binomial regression, stata.com/help.cgi?zip for the Zero-inflated Poisson and stata.com/help.cgi?zinb for the Zero-inflated negative binomial.

3 Results

3.1 *P. penetrans* spore Attachment Data

The Table 1 shows the number *P. penetrans* spores attached (att) to nematode cuticle. These data characterized by excess zeros were used to model *P. penetrans'*s attachment.

Table 1. Table showing the observed values of *P. penetrans* spores attachment

Variable (h)	observed values attachment (att)
1	0; 1; 3; 4; 3; 4; 1; 0; 2; 1; 1; 3; 3; 2; 1; 2; 2; 1; 1; 2; 0; 3; 2; 2; 1; 1; 1; 2; 0; 1; 3; 0; 3; 1; 1; 1
3	2; 0; 3; 4; 1; 7; 3; 1; 7; 6; 2; 4; 1; 7; 4; 1; 8; 5; 0; 8; 3; 6; 4; 7; 6; 2; 0; 4; 6; 4; 3; 8; 1; 0; 1; 5
6	1; 6; 6; 8; 0; 5; 15; 4; 0; 4; 5; 9; 6; 7; 7; 14; 7; 1; 9; 9; 0; 6; 8; 3; 5; 5; 4; 8; 10; 0; 0; 3; 7; 12; 4; 2
9	7; 2; 3; 9; 7; 3; 5; 6; 14; 12; 6; 10; 12; 2; 11; 3; 8; 19; 11; 9; 7; 6; 12; 8; 0; 0; 4; 13; 16; 3; 1; 8; 12; 3; 0; 10

3.2 Statistical Analysis of *P. penetrans*'s Attachment

Descriptive statistics for counts variables (var1-4) time 1, 3, 6 and 9hrs (h), were presented in Table 2. The Table 2 shows that for var. 3h the mean is twice the variance, and for var. 6h, and 9h the mean is thrice the variance.

Table 2. Descriptive statistics for var1, var2, var3, var4

Variable (h)	Obs.	Mean	Variance	Std. Dev.	Min	Max
1	36	1.638	1.265	1.125	0	4
3	36	3.722	6.663	2.581	0	8
6	36	5.555	14.768	3.842	0	15
9	36	7.278	22.778	4.773	0	19

This suggests over-dispersion means the assumptions of the Poisson model are not met and make the Negative Binomial distribution a useful over-dispersed alternative to the Poisson distribution.

Based on the above observations we presume that the variance is proportional rather than equal to the mean, therefore I divide Pearson's chi-squared by its d.f. in order to estimate the scale parameter φ. Specifically, $var(Y) = E(Y) = \varphi\mu$ applying that if φ = 1, the variance equals the mean and the Poisson

model is obtained. If the φ value is more than one (φ > 1), the data are over-dispersed when compared to Poisson.

The expected value of this statistic is n- p if the model is correct. Equating the statistic to its anticipation and solving for φ gives the estimate 1.

$\hat{\varphi} = \chi^2_p / n - p$ estimate 1 (equation 1)

In our model Table 3 Generalized linear models, $\chi^2_p = 292.272$ for a model with p = 3 parameters on n = 144 observations, which leads to $\hat{\varphi}$ = 292.272 /140 = 2.087. From Table 3 we have that the Poisson estimates the standard errors by $\sqrt{2.087}$ = 1.4448, which inflates them by 49.4%. The high value of Pearson's χ^2 and the p χ2 indicates a lack of fit, which is not related to specification problems but rather to over-dispersion. From Table 3, the Pearson's χ^2 statistic divided by its degree of freedom (df) leads to 2.087 indicating over-dispersion and however the Deviance statistic divided by its df leads to 2.495 recommended over-dispersion.

As shown above another method for modeling over-dispersion in count data is to start with a Poisson regression model and add a multiplicative random effect θ to represent unobserved heterogeneity,

This directs to the Negative Binomial regression model (nbreg) as presented in Table 4. The Negative Binomial model (nbreg), Table 4, gives estimates that are very akin to the Poisson model, and retain the same interpretation with average unobserved characteristics as the count variables contain zeroes Table 1.

The conditional probability distribution of the outcome Y (equation 1), specified an unobserved θ variable of Poisson with mean and variance $\theta\mu$ (equation 1), proposed that the data of Table1 would be Poisson if only we could observe θ. In our results, we do the hypothesis that θ captures unobserved factors that increase (if $\theta > 1$) or decrease (if $\theta < 1$). In our results (Table 5), the output uses alpha which is equal to 0.32245 to label the variance of the unobservable, recommending that the data would not be Poisson.

As was noted above in the STATA log the Negative Binomial model provides estimates that are very similar to the Poisson model Table 4, and have the same interpretation. The standard errors resemble the over-dispersed Poisson standard errors, and both are more considerable than the reference Poisson errors.

Maybe our distributional assumption is similar but the nbreg STATA test of Table 5, leads to χ^2_{LR} = 69.04 which is highly significant (p=0.000)

suggested that the Negative Binomial is better than the Poisson model. Furthermore, it is obvious that $AIC_{Poisson}$, $BIC_{Poisson}$ Table 3, and $AIC_{Neg.Binomial}$, $BIC_{Neg.Binomial}$ Table 6, has very different values suggested the Negative Binomial model fits considerably better than the Poisson, but even has deviance suggested that empirical data probably show fewer zeroes than would be expected under either model. To solve that, we used a Zero-Inflated (ZIP) Model for frequent zero-valued observations and a zero-inflated negative binomial (ZINB) model for modeling over-dispersion and excess zeros.

In an analysis of ZIP Table 7, the Poisson model predicts only 8% of the 11,8% of the nematodes without *P. penetrans* spores. The Poisson model prognosticates 8%, so it underestimates the zeroes by five percentage points.

The ZIP model proposes that there are two latent categories of nematodes, the "always zero" and another the "not always zero", where the count has a Poisson distribution with mean and variance > 0. To solve it we were used in STATA the zero-inflated Poisson command in the inflate() option Table 8. The inflate equation of the outcome values Table 8, shows that the related to the ZIP model where the probability of "always zero" exists in 3, 6 and 9 hrs of incubation. Table 8 shows that the incubation period (3, 6 and 9hrs) is a significant predictor of being in the "always zero" class. Furthermore, we can observe from Table 9 that as alpha = 0 is significantly different from zero signifying that our data are over-dispersed and that a zero-inflated negative binomial model is more proper than a zero-inflated Poisson model. Moreover, the Zinb model in the option zfitnb Table 10, solve the problem with the excess zeroes, predicting that 10.5% of the tested nematodes were unencumbered with *P. penetrans* spores a very close to the observed (zobs) value of 11.8 Table 7.

Table 3. Poisson regression and a GLMs Poisson to accommodate the excess variation.

STATA $_{command}$ poisson att i.h, vce(robust

Poisson regression			Number of obs =**144**		
			Wald chi2(**3**) =**101.43**		
			Prob > chi2 =**0.0000**		
Log pseudolikelihood = **-380.23145**			Pseudo R2 =**0.1667**		
att	Coef.	Robust Std. Err.	z	P>|z| [95% Conf.	Interval]
h					
3	.820	.160	5.10	0.000 .504	1.135
6	1.22	.160	7.60	0.000 .905	1.535
9	1.49	.156	9.52	0.000 1.183	1.797
_cons	.494	.113	4.36	0.000 .2721	.715

STATA command glm att i.h, family(poisson) nolog

Generalized linear models		No. of obs = **144**				
Optimization	: **ML**	Residual df = **140**				
Deviance	= **349.353**	(1/df) Deviance = **2.495**				
Pearson	= **292.272**	(1/df) Pearson = **2.087**				
Variance function: **V(u) = u**		[**Poisson**]				
Link function : **g(u) = ln(u)**		[**Log**]				
Log likelihood	= **-380.231**	AIC = **5.336**				
		BIC =**-346.41**				

att	Coef.	OIM Std.Err.	z	P>|z|	[95% Conf.	Interval]
h						
3	.820	.156	5.25	0.000	.514	1.126
6	1.22	.148	8.24	0.000	.930	1.511
9	1.49	.144	10.35	0.000	1.20	1.773
_cons	.494	.130	3.79	0.000	.238	.749

Table 4. Comparing estimates and standard errors based on the Negative Binomial model and on the Poisson regression model.

Variable	poisson	overdisp	nbreg
att			
h			
3	.82030	.82030	.82030
	.15624	.22575	.20586
6	1.2207	1.2207	1.2207
	.14815	.21406	.19979
9	1.4908	1.4908	1.4908
	.14410	.20821	.1968
	.49401	.49401	.49401
_cons	.13018	.1881	.16104
lnalpha			
_cons			-1.1286952
			.22315033

Table 5. Negative Binomial regression model.
STATA command nbreg att i.h

Negative binomial regression			Number of obs = **144**		
			LR chi2(**3**) = **53.91**		
Dispersion		= **mean**	Prob > chi2 = **0.0000**		
Log likelihood = **-345.70896**			Pseudo R2 = **0.0723**		

att	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval
h						
3	.8203	.2058	3.98	0.000	.41680	1.2237
6	1.2207	.1997	6.11	0.000	.82918	1.6123
9	1.4908	.1968	7.57	0.000	1.1050	1.8765
_cons	.49401	.1610	3.07	0.002	.17838	.80965
/lnalpha	-1.128	.2231			-1.566	-.69132
alpha	.32345	.0721			.20886	.50091

Likelihood-ratio test of alpha=0: **chibar2(01) = 69.04 Prob>=chibar2 =0.000**

Table 6. A Negative Binomial model to accommodate the excess variation.
STATA command glm att i.h, family(nb `v') nolog

Generalized linear models	No. of obs = **144**
Optimization : **ML**	Residual df = **140**
	Scale parameter = 1
Deviance = **164.3046481**	1/df) Deviance = **1.173605**
Pearson = **112.0212999**	(1/df) Pearson = **.8001521**
Variance function: **V(u) = u+(.3235)u^2**	[**Neg. Binomial**] Link [**Log**]
function : **g(u) = ln(u)**	
	AIC = **4.857069**
Log likelihood =**-345.7089557**	BIC = **-531.4692**

Table 7. Zero-Inflated Poisson (ZIP) model.
STATA command sum zobs zfitp

Variable	Obs	Mean	Std.Dev.	Min	Max
zobs	**144**	**0.11855**	**0.3238**	**0**	**1**
zfitp	**144**	**0.08057**	**0.0807**	**0.00069**	**0.19419**

Table 8. Zero-Inflated Poisson (ZIP) model.
STATA command zip att i.h, inflate(h) vce(robust)

Zero-inflated Poisson regression	Number of obs = **144**
	Nonzero obs = **127**
	Zero obs = **17**
Inflation model = **logit**	Wald chi2(**3**)= **128.01**
Log pseudolikelihood = **-350.0632**	Prob > chi2 = **0.0000**

att	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
att						
h						
3	.87726	.1518	5.78	0.000	.57966	1.1748
6	1.3274	.14411	9.21	0.000	1.045	1.6099
9	1.5371	.14561	10.56	0.000	1.2517	1.8226
_cons	.53439	.1096	4.88	0.000	.31956	.74921
inflate						
h	.07184	.09275	0.77	0.439	-.10994	.25363
_cons	-2.6985	.6298	-4.28	0.000	-3.9330	-1.4641

Table 9. Zero-Inflated Negative Binomial (Zinb) model.

STATA $_{command}$ zinb att i.h, inflate(h) vuong zip

Zero-inflated negative binomial regression					Number of obs = **144**	
					Nonzero obs = **127**	
					Zero obs = **17**	
Inflation model = **logit**					LR chi2(**3**)= **68.18**	
Log pseudolikelihood = **-338.4195**					Prob > chi2 = **0.0000**	
att	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
att						
h						
3	**.8618**	**.18628**	**4.63**	**0.000**	**.49678**	**1.227016**
6	**1.3262**	**.17982**	**7.38**	**0.000**	**.97380**	**1.678686**
9	**1.5489**	**.17491**	**8.86**	**0.000**	**1.2060**	**1.891729**
_cons	**.51906**	**.14880**	**3.49**	**0.000**	**.22741**	**.810722**
inflate						
h	**.11522**	**.13260**	**0.87**	**0.385**	**-.14468**	**.3751219**
_cons	**-3.143**	**.94447**	**-3.33**	**0.001**	**-4.9946**	**-1.292392**
/lnalpha	**-1.8996**	**.33253**	**-5.71**	**0.000**	**-2.5514**	**-1.247904**
alpha	**.14961**	**.04975**			**.07797**	**.2871059**

Likelihood-ratio test of alpha=0: chibar2(01) = 23.29 **Pr>=chibar2 = 0.0000**

Vuong test of zinb vs. standard negative binomial: **z = 1.74** **Pr>z = 0.0410**

Table 10. Zero-Inflated Negative Binomial (Zinb) model with zfitnb option.

Variable	Obs	Mean	Std.Dev.	Min	Max
zfitnb	**144**	**0.1051**	**0.9739**	**0.237**	**0.268**

Finally, to choose between the Negative Binomial and Zero-inflated models the Vuong test, Table 9, suggests that the Zero-inflated negative binomial model is a significant better Pr>z = 0.0410, over a standard negative binomial model.

4 Discussion

According to our data (Table 1), the procedure discussed above confirmed that the Zero-inflated negative binomial model is the more appropriate model to estimate *P. penetrans* spore attachment count data. The Negative binomial model is also the preferred model as time of exposure increased (e.g. 6 or 9 h), confirming the original conclusion reported by Vagelas [2, 15] that the data are over-dispersed within a specific time period (e.g. 6). As the variance is greater than the mean [16-21], examination of the variability using the Negative binomial was an acceptable model to describe *P. penetrans* over-dispersion as an aggregating organism, leading to the conclusion that the bacteria are clumped and clustered under natural conditions.

Moreover, in this paper, we used the Poisson, the Negative binomial and the Zero-inflated Poisson model [22], to deal with count data especially, when the data are over-dispersed and contain excessive zero counts [23]. In addition to the Poisson model, we used the Negative binomial and Zero-inflated models to the count data of Table 1. The first step should be a summary of Statistics (e.g. the observed data's sample mean and variance) to measure if the data are over-dispersed. Second, we illustrated the estimate of the dispersion parameter [24], as deviance or Pearson's χ^2 statistic divided by the degrees of freedom, which is often used to indicate over-dispersion for Poisson models. Third, as the Poisson model is a special case of the Negative Binomial when $\sigma^2 = 0$, we apply a likelihood ratio test to compare the two models. Finally, we suggested that the Zero-inflated models should be applied to estimate the dispersion parameters of the Negative Binomial [25] and offer an explanation for the excessive zero counts [26-27].

5 Conclusion

In biological research including biocontrol, count data with a large proportion of zeros are often recorded [28-30]. For those count data Poisson regression is often used when the conditional variance is equal to the mean (equi-dispersion), or alternative since the variance exceeds the mean (over-dispersion), the negative binomial model was used to fit the count data. Since the over-dispersion has occurred, the probability of a structural zero, true zeros, and excess zero needs to test further. Further, apart from the negative binomial model, the Zero-inflated models can be an alternative to overcome over-dispersed data and can be used as a model to offer an explanation for the excess zeros condition in the counts.

References:

[1] Chen Z X, Dickson D, Review of *Pasteuria penetrans*: Biology, ecology, and biological control potential, *Journal of Nematology*, Vol.30,No.3, 1998, p. 313.

[2] Vagelas I. Dennett M D, Pembroke B, Gowen S. R, Fitting the negative binomial distribution to *Pasteuria penetrans* spore attachment on root-knot nematodes and predicting probability of spore attachments using a Markov chain model, *Biocontrol Science and Technology*, Vol.23, No.11, 2013, pp.1296-1306.

[3] Hougaard P, Lee M L T, Whitmore G A, Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes, *Biometrics*, 1197, pp. 1225-1238.

[4] Lee J H, Han G, Fulp W J, Giuliano A R, Analysis of overdispersed count data: application to the Human Papillomavirus Infection in Men (HIM) Study, *Epidemiology & Infection*, Vol.140, No.6, 2012, pp.1087-1094.

[5] Queiroz F F, Lemonte A J, A broad class of zero- or- one inflated regression models for rates and proportions, *Canadian Journal of Statistics*,Vol.49, No.2, 2021, pp.566-590.

[6] Hooper D J, Extraction of nematodes from plant materials. In J. F. Southey (Ed.), Laboratory methods for working with plant and soil nematodes (6th ed., pp. 51–58). 1986, London: Her Majesty's Stationary Office.

[7] Whitehead A G, Hemming J R, A comparison of some quantitative methods of extracting small vermiform nematodes from soil, *Annals of Applied Biology*, Vol.55, 1965, pp. 25–38.

[8] Davies KG, Kerry BR, Flynn C A, Observations on the Pathogenicity of Pasteuria penetrans, a Parasite of Root-knot Nematodes, Annals of Applied Biology, Vol.112, 1988, pp. 491–501.

[9] Vagelas I, Pembroke B, Gowen S R, Techniques for image analysis of movement of juveniles of root-knot nematodes encumbered with Pasteuria penetrans spores. *Biocontrol science and technology*, Vol.21, No.2, 2011, pp. 239-250.

[10] Vagelas I K, Dennett M D, Pembroke B, Gowen S R, Adhering *Pasteuria penetrans* endospores affect movements of root-knot nematode juveniles. *Phytopathologia Mediterranea*, 2012, pp. 618-624.

[11] Stata .do-Files and Data Sets in Stata Format, Statistics Using Stata, 2020, pp. 686–687. doi:10.1017/9781108770163.022

[12] Chipeta M G, Ngwira B M, Simoonga C, Kazembe L N, Zero adjusted models with applications to analysing helminths count data, *BMC research notes,* Vol.7, No.1, 2014, pp. 1-11.

[13] White G C, Bennetts R E, Analysis of frequency count data using the negative binomial distribution. *Ecology*, Vol.77, No.8, 1996, pp. 2549-2557.

[14] Cheung Y B, Zero- inflated models for regression analysis of count data: a study of growth and development, *Statistics in medicine*, Vol.21, No.10, 2002, pp. 1461-1469.

[15] Vagelas I, Data analysis and modeling of *Pasteuria penetrans* spore attachment. International *Journal of Agriculture & Environmental Science*, Vol.7, No.5, 2020, pp. 108–113. doi:10.14445/23942568/ijaes-v7i5p116

[16] Bliss C I, Fisher R A, Fitting the negative binomial distribution to biological data, *Biometrics*, Vol.9, 1953, pp. 176–200. doi:10.2307/3001850

[17] Ross G J S, Preece D A, The negative binomial distribution, *The Statistician*, Vol.34, 1985, pp. 323–336. doi:10.2307/2987659

[18] Gschlobl S, Czado C, Modelling count data with overdispersion and spatial effects *Statistical Papers*, Vol.49, 2008, pp. 531–522. doi:10.1007/s00362-006-0031-6

[19] Morel G J, Nagaraj K J, A finite mixture distribution for modelling multinomial extra variation, *Biometrika*, Vol.80, 1993, pp. 363–371. doi:10.1093/biomet/80.2.363

[20] Richards S A, Dealing with overdispersed count data in applied ecology, J*ournal of Applied Ecology*, Vol.45, 2008, pp. 218–227. doi:10.1111/j.1365-2664.2007.01377.x

[21] Xie M, He B, Goh T N, Zero-inflated Poisson model in statistical process control, C*omputational statistics & data analysis*, Vol.38, No.2, 2001, pp. 191-201.

[22] Gilthorpe M S, Frydenberg M, Cheng Y, Baelum V, Modelling count data with excessive zeros: The need for class prediction in zero- inflated models and the issue of data generation in choosing between zero- inflated and generic mixture models for dental caries data, *Statistics in medicine*, Vol.28, No.28, 2009, pp. 3539-3553.

[23] Piegorsch W W, Maximum likelihood estimation for the negative binomial dispersion parameter, *Biometrics*, 1990, pp. 863-867.

[24] Ridout M, Hinde J, Demétrio C G, A score test for testing a zero inflated Poisson regression model against zero inflated negative binomial alternatives, *Biometrics*, Vol.57, No.1, 2001, pp. 219-223.

[25] Güneri Ö İ, Durmuş B, İncekirik A, Comparison of some count models in case of excessive zeros: An application, *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, Vol.20, No.40, 2021, pp. 247-268.

[26] Feng C X, A comparison of zero-inflated and hurdle models for modeling zero-inflated count data, *Journal of Statistical Distributions and Applications*, Vol.8, No.1, 2021, pp. 1-19.

[27] Ridout M, Demétrio C G, Hinde J, Models for count data with many zeros. In *Proceedings of the XIXth international biometric conference* Cape Town, South Africa: *International Biometric Society*, Vol. 19, 1998, pp. 179-192

[28] Jiang S, Xiao G, Koh A Y, Kim J, Li Q, Zhan X, A Bayesian zero-inflated negative binomial regression model for the integrative analysis of microbiome data. *Biostatistics*, Vol.22, No.3, 2021, pp. 522-540.

[29] Xu T, Demmer R T, Li G, Zero- inflated Poisson factor model with application to microbiome read counts. *Biometrics*, Vol.77, No.1, 2021, pp. 91-101.

[30] Mod H K, Buri A, Yashiro E, Guex N, Malard L, Pinto-Figueroa E, Guisan A, Predicting spatial patterns of soil bacteria under current and future environmental conditions. *The ISME journal*, Vol.15, No.9, 2021, pp. 2547-2560.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)
Ioannis Vagelas, defined the methodology, gathered the data, analyzed and writing the paper.