Animal Science Papers and Reports vol. 37 (2019) no. 4, 355-363
Institute of Genetics and Animal Breeding, Jastrzębiec, Poland

A comparison of count regression models describing the number of mummified piglets

José Candelario Segura Correa[1]*, José Alfonso Hinojosa Cuellar[2], Ramiro Avalos-Ramírez[3]

[1] Campus of Biological and Agricutural Sciences, Autonomous University of Yucatan km 15.5 Merida-Xmatkuil road, Merida, Yucatan, Mexico,

[2] Popular University of the Chontalpa, Cardenas- Huimanguillo road, km 2 Paso y Playa. H. Cardenas. Tabasco, Mexico

[3] Autonomous University of Nuevo Leon. Faculty of Veterinary and Animal Science, Campus of Agricultural Sciences, Francisco Villa w/n Avenue, Ex-Hacienda El Canada, ZP 66050. General Escobedo, Nuevo Leon, Mexico

(Accepted October 15, 2019)

The aim of this study was to compare count regression models to identify the best model and evaluate the effect of year, season, parity and litter size of sows at farrowing on the number of mummified fetuses per litter. Data from a commercial production unit in Yucatan, Mexico were used. Management and feeding of the animals were typical of the region. Poisson, negative binomial, zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB) and normal distribution regression models were evaluated and their goodness-of-fit was compared. Among the 1689 litters 60.27% had zero mummies. The mean number of mummies was 0.649±1.028 for litter for data including zeros and 1.634±1.025 for data with at least one mummified fetus. The ratio of the Pearson Chi-square value/degrees of freedom was significantly different from 1 for Poisson, negative binomial and ZIP models, indicating overdispersion (variance> mean) of the trait studied. However, the ZINB model indicated no overdispersion (P=0.0684). The ZINB model was considered the best model based on log likelihood, AIC and BIC. The significance of factors varied according to the model used. Furthermore, Poisson models had minor standard errors and in consequence tended to reject the null hypothesis more easily than negative binomial models. Under the ZINB model, year and parity number had significant effects on the number of mummified fetuses per litter (P<0.05). In conclusion, parity number was an important risk factor associated to the number of mummified piglets. The best model that describes the number of mummified piglets was the zero inflated negative binomial regression procedure.

*Corresponding author:jose.segura@correo.uady.mx

KEY WORDS: binomial / Negative binomial, number of dead fetuses / Poisson / Zero inflated Poisson / Zero inflated negative

The presence of mummified piglets in the litter causes significant economic losses in pig production units. Therefore, it is important to identify risk factors associated to it. However, the use of inappropriate statistical models could influence the results. Count data are commonly used to describe animal production traits, such as litter size, number of stillborn piglets, mummified fetuses, number of oocytes collected, preantral ovarian follicles etc. During statistical analysis of observational or experimental data, different procedures are applied, such as the analysis of variance, non-parametric, logistic and Poisson regression [Borges *et al*. 2005, Espitia and Galindez 2011, Segura and Solorio 2013, Silva-Santos *et al*. 2014]. Count data, however, do not follow normal or binomial distribution and consequently the application of models that use such distributions may produce bias and loss of information [Agresti 1996]. Furthermore, linear regression is not an appropriate analytical method for count data due to skewed distribution and overdispersion (variance greater than the mean). The categorisation of count data into binary type data to be used in chi-square and logistic regression procedures is expected to lead to loss of information and incorrect inferences.

Discrete distributions such as the Poisson, negative binomial, zero inflated Poisson and zero inflated negative binomial distributions are applicable in modelling count data that take positive integer values and where the data are obtained from counting rather than ranking [Cameron and Trivedi 1998]. In view of the inconvenience of treating count data as continuous or binomial data, Poisson regression or negative binomial regression are used to model such data [Tani *et al*. 2016]. However, the probability of zeros based on those distributions may not account for excess zero counts, which could be the case for the number of mummified fetuses per litter or other excess zero traits. In such a case, ignoring excess zeros will bias the estimation of parameters [Cheung 2002]. Zero-inflated regression models consider the evaluation of the data as a two-part analysis, an all-zero subset and another subset following the Poisson or negative binomial distribution. Some studies indicate that zero-inflated models are the best to solve the issue of excess zeros in human and animal health [Denwood *et al.* 2008, Rose *et al*. 2006 Kipnis *et al*. 2009]. However, to the best of the authors' knowledge, only one paper applied zero inflated methodology to describe the number of piglets dead in the litter [Varona and Sorensen 2010].

The objectives of this study were to compare the Poisson, negative binomial, zero-inflated Poisson and zero-inflated negative binomial regression procedures to detect the best model and evaluate the effect of year, season, parity and litter size of sows at farrowing on the number of mummified fetuses per litter.

Material and methods

Data on the number of mummified fetuses per litter from a commercial production unit in Yucatan, Mexico were used. Data were obtained from the 2011 to 2014 records registered in an electronic recording system (PigCHAMP ®). The production unit was a one-site farm with the capacity for 220 sows.

Both artificial insemination and natural mating were carried out for reproduction purposes. Estrus detection was at 6:00 and 18:00 h every day using a boar and sows were inseminated at 12 to 18 h, after estrus had been diagnosed. Sows were fed commercial diets based on the stage of production. Young sows were provided 2.6 kg/day of feed with 3,000 kcal EM/kg, 16% crude protein and 0.8% lysine, whereas adult sows received 3.2 kg/day of feed. Replacements gilts were obtained from the same farm.

Statistical analysis

The information used for the study included sow identification, sow parity number, date of farrowing, litter size (total number of piglets born) and number of mummified fetuses per litter. Potential risk factors in the model to describe the response variable included year of farrowing, season of farrowing, parity number (1 to 7) and litter size (as a continuous covariate).

The Poisson, negative binomial, zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) regression models were used and their goodness-of-fit was compared. In addition, in some papers count data are analysed using models assuming normal distribution, arguing robustness of that procedure, therefore a model with normal distribution was also evaluated. The presence of overdispersion of the data was tested. The regression models evaluated in this study are suitable for count data, but ZINB was expected to provide a better goodness of fit for the information on mummified fetuses. The models that described the number of mummified fetuses (y) were:

The Poisson and negative binomial regression models correspond to the Poisson and negative binomial sections of the ZIP and ZINB models given below.

The ZIP regression model is a two-section model, consisting of a binary and a count model section to account for excessive zeros in the response variable.

$$p(y_i) = \begin{Bmatrix} \pi_i + (1-\pi_i)e^{-\mu} & y_i = 0 & \textit{logit section} \\ (1-\pi_i)\dfrac{e^{-\mu}\mu^{y_i}}{y_i!} & y_i > 0 & \textit{Poisson section} \end{Bmatrix}$$

Similarly, the ZINB model refers to a two-section model including an all-zero section and a section following the negative binomial distribution. The probability density function of the ZINB model is:

$$p(y_i) = \begin{Bmatrix} \pi_i + (1-\pi_i)\left(\dfrac{1}{1+k\mu}\right)^{k^{-1}} & y_i = 0 & \textit{logit section} \\ (1-\pi_i)\dfrac{\Gamma(y_i + k^{-1})(k\mu)^{y_i}}{y_i!\,\Gamma(k^{-1})(1+k\mu)^{y_i + k^{-1}}} & y_i > 0 & \textit{NB section} \end{Bmatrix}$$

where: π_i is the zero inflated probability, μ is the mean of the number of mummified fetuses and k is the dispersion parameter. The Ln and logit link functions were used for parameters μ and π_i.

In the logit part of the zero-inflated models, the explanation of regression coefficients is equal to those in standard logistic regression models. In the Poisson or negative binomial (NB) sections, the explanations are equal to the traditional Poisson or negative binomial regression models (i.e. as rates or means).

In this study the SAS version 9.4 (2012) was used to construct the regression models. The goodness of fit for the models was assessed by the Akaike information criterion (AIC), Bayes Information criterion (BIC) and likelihood ratio tests. In addition, the ratio of Pearson chi-square/degrees of freedom was used to identify overdispersion. AIC and BIC are penalised likelihood criteria used to select best predictor models in regression. In both criteria the model with the lowest values is the best. For the likelihood ratio test the greatest its value, the best is the model.

Results and discussion

The percentage of mummified fetuses per litter is shown in Figure 1. Among the total of 1689 litters, 60.27% had zero mummies (no mummified fetus in the litter) and it did not follow a normal distribution according to the Shapiro-Wilk test. The average number of mummies was 0.649 ± 1.028 for data including zeros and 1.634 ± 1.025 for data with at least one mummified fetus.

Based on the model fit criteria (log likelihood, AIC and BIC), the best model that

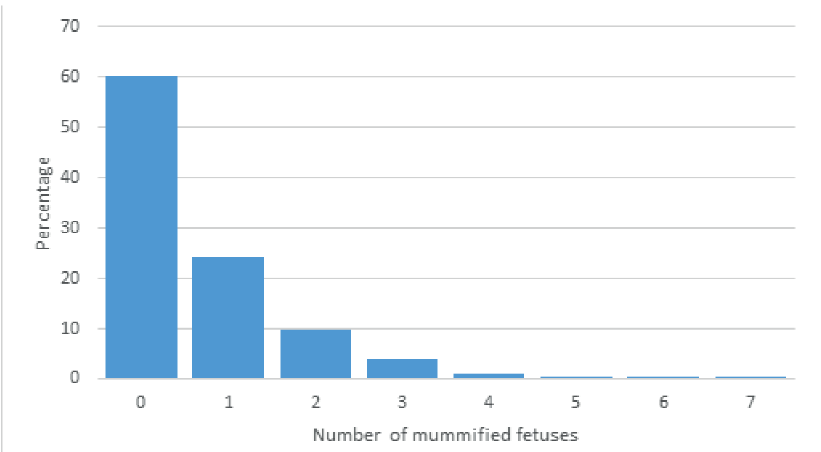

Fig. 1. Percenages of mummified fewtuses per litter in a production unit in the tropics.

described the number of mummies per litter in this study was the ZINB model (Tab. 1). It was followed by the negative binomial model. The worst model fit was found for the one with the normal distribution. The ratio of the Pearson Chi-square value/degrees

Table 1. Fitting goodness statistics of regression models and significance of over-dispersion

Model	Log likelihood	AIC	BIC	Valor/df	P value
Poisson	-1414.9	3634.1	3704.7	1.4156	0
Poisson	-1414.9	3634.1	3704.7	1.4156	0
NB	-1373.3	3552.9	3698.9	1.0613	0.0399
ZIP	-1363.8	3558.0	3699.2	1.0993	0.0027
ZINB	-1747.6	3549.1	3695.8	1.0521	0.0684
Normal	-2348.8	4725.7	4801.7	1.0078	0.4069

AIC – Akaike's information criterion; BIC – Bayesian information criterion; NB – Negative binomial; ZINB – Zero-inflated negative binomial; ZIP – Zero-inflated Poisson.

of freedom was significantly different from 1 for the Poisson, negative binomial and ZIP models indicating overdispersion (variance> mean) in the trait studied. The ZINB model showed no overdispersion (p=0.0684), as it was found for the normal model (0.4069). The ZINB model fitted the data better than the ZIP model, because the ZINB and negative binomial models take account of overdispersion. Some studies, in areas of knowledge other than animal production, report that the ZINB model is better for count data when compared to the Poisson models [Akram *et al*. 2006; Rose *et al*. 2006, Akinpelu *et al*. 2016, Xu *et al*. 2017]. Zero inflated models evaluate the influence of risk factors on the number of mummified fetuses checking the rate of the number of mummies per litter (count section) and the excess of zeros (litters with no mummified fetuses; logit section).

Regression coefficients by risk factors studied for the models evaluated are shown in Tables 2 and 3. An important thing was that the significance of factors varied depending on the model used. For the best model (ZINB, count section) and ZIP model (count section) only litter size had an influence on the rate of mummified fetuses. However, the Poisson model showed also the effect of year and season of farrowing and the negative binomial of season. The Poisson, negative binomial and ZIP models had minor standard errors and in consequence have minor p values, which in turn tended to reject the null hypothesis more easily than the ZINB model. With respect to the zero section of the zero inflated models and logistic model the; ZINB gave similar results to ZIP (year and litter size effects); in contrast, the logistic model indicated, in addition, the effect of season. Based on the recorded results a comparison of the effect of the risk factors studied here with those reported in the literature was made based on the ZINB model results.

The lack of year of farrowing effects on the number of mummies per litter agrees with the results of Rico and Gomez [1982], but is not consistent with those of Espitia and Galíndez [2011], who using normal distributed regression models found a significant effect of year on the rate of mummies per litter (number of mummies per litter x 100 /litter size). Year is a factor commonly found to be significant, but it is not of biological interest. Season, on the other hand, did not influence the number

Table 2. Coefficients±standard errors for the Poison, negative binomial and normal distributed regression models for number of mummified fetuses in pigs

Factor	Poisson	Negative binomial	Normal
Year of farrowing	0.0281	0.0559	0.2040
2011	-0.3136±0.1065	-0.3274±0.1241	-0.1628±0.0763
2012	-0.0526±0.0831	-0.0155±0.0994	-0.0310±0.0660
2013	-0.0726±0.0789	-0.0684±0.0940	-0.00494±0.0626
2014	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000
Season of farrowing	0.0006	0.0038	0.0058
1	0.2522±0.0806	0.2511±0.0949	0.1598±0.0609
2	0.2780±0.0786	0.2918±0.0931	0.1734±0.0591
3	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000
Parity number	0.1024	0.1732	<0.0001
1	-0.1045±0.1414	-0.1116±0.1782	-0.2018±0.01320
2	-0.2540±0.1457	-0.2764±0.1820	-0.2721±0.1328
3	-0.2812±0.1449	-0.3053±0.1822	-0.2775±0.1344
4	-0.3730±0.1488	-0.4214±0.1877	-0.3205±0.1375
5	-0.1977±0.1551	-0.2557±0.1959	-0.2007±0.1455
6	-0.2048±0.1701	-0.2073±0.2122	-0.2293±0.1537
7	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000
Litter size	0.1605±0.0102	0.1622±0.0124	0.0919±0.0778

Table 3. Coefficients±standard errors for the Zero inflated Poisson (ZIP) and Zero inflated negative binomial (ZINB) regression models for the number of mummified fetuses in pigs

Factor	ZIP count	ZIP zero	ZINB count	ZINB zero
Year of farrowing*	0.0747	0.0023	0.3720	0.0429
2011	-0.0288±0.1491	0.6819±0.3221	-0.0287±0.1705	0.9054±0.4614
2012	-0.3004±0.1281	-1.1028±0.5637	-0.2452±0.1467	-1.4258±1.1309
2013	-0.2029±0.1133	-0.4369±0.3220	-0.1805±0.1286	-0.4988±0.4915
2014	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000
Season of farrowing*	0.4423	0.5574	0.4770	0.7269
dry	0.1258±0.1236	-0.3462±0.3312	0.1482±0.1368	-0.3830±0.4900
rainy	0.1669±0.1334	-0.3383±0.3657	0.1801±0.1581	-0.4051±0.6098
windy	0.0000±0.0000	0.0000±0.0000	0.0000±0.000	0.0000±0.0000
Parity number*	0.0717	0.5532	0.1319	0.6571
1	0.0856±0.1923	0.0856±0.6210	0.0046±0.2216	0.5522±0.9188
2	-0.2604±0.2016	-0.2604±0.6569	-0.3015±0.2366	-0.0142±0.9961
3	-0.3041±0.2006	-0.3041±0.6778	-0.3769±0.2295	-0.3951±1.0716
4	-0.2954±0.2039	-0.2954±0.7101	-0.3455±0.2404	0.2434±1.1137
5	-0.1187±0.2007	-0.1187±0.6926	-0.1465±0.2271	0.5107±0.9803
6	-0.1586±0.2592	-0.1586±0.8506	-0.2226±0.2956	0.0477±1.3744
7	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000
Litter size	0.1057±0.0165	-0.1886±0.0450	0.1107±0.0178	-0.2331±0.0585

*Values in this row correspond to *p* values of the factor.

of mummified fetuses, which agrees with the observations reported by Espitia and Galíndez [2011]. However, Segura *et al*. [2013] mentioned that high temperatures during the dry season could be the cause for an increase in the number of dead and mummified fetuses.

In this study parity number had no effect on the number of mummified fetuses. In contrast, Spotter and Distl [2000] indicated higher mortalities in sows with 4 to 6 farrowings due to an increase of litter size, which could have increased the probability of a greater number of dead and mummified fetuses. In addition, Pedrosa *et al*. [2000] and Espitia and Galíndez [2011] found an association of parity number of the sow and percentage of mummies in the litter. The discrepancy between studies may be due to variations between regions and management systems, but also to the inequality regression models using different distributional assumptions. As mentioned before, inappropriate models could provide different significant results for the factors studied.

The effect of litter size, found in all the models used in this study, emphasises the importance of this factor for the number of mummified fetuses. Here, the natural log of the number of mummies was predicted to increase 0.1107 (ZINB model) times per each extra piglet in the litter, while in terms of relative risks it was an 11.7% (exp(0.1107)) increase in the number of mummies per each extra piglet in the litter. Similarly, Espitia and Galindez [2011], using linear regression, reported a significant effect of litter size and found a regression coefficient of 1.09%. They interpreted this result as an increase of 1.09% mummified fetuses per each extra piglet in the litter above the mean. The increase in the number of mummies with an increase in litter size could be explained by the limited space in the uterus and nutrients available per fetus, which diminish as the litter size increases [Johnson *et al*. 1999].

The logit part of the ZINB model showed that year and litter size were significant risk factors for the occurrence of mummies. As the litter size increased, the probablity of the occurrence of at least one mummy in the litter decreased (-0.2091+ 0.0188). In the logistic regression model (Tab. 4) year, season and litter size were associated with increased probabilities of one or more mummies in the Therefore, categorising the count data could lead to different results and as a consequence logistic regression may

Table 4. Logistic regression coefficients for some risk factors on the presence of at least one mummified fetus in the litter

Factor	Beta	EE	OR	95%CI(OR)
Year of farrowing				
2011	0.4851	0.1783	1.624	1.145, 2.304
2012	-0.2535	0.1449	0.776	0.584, 1.031
2013	-0.0962	0.1387	0.908	0.692, 1.192
2014	0	.	1.	.
Season of farrowing				0.0018
dry	-0.3658	0.1367	0.694	0.531, 0.907
rainy	-0.4516	0.1330	0.637	0.491, 0.826
windy	0	.	.1	.
Parity number				0.7004
1	0.4854	0.2905	1.625	0.920, 2.871
2	0.4046	0.2929	1.499	0.844, 2.661
3	0.4249	0.2958	1.529	0.857, 2.731
4	0.5681	0.3028	1.765	0.975, 3.195
5	0.4499	0.3198	1.568	0.838. 2.935
6	0.4171	0.3374	1.518	0.783, 2.940
7	0	.	1.	.
Litter size	-0.2091	0.0188	0.811	0.782. 0.842

not be an appropriate model for count outcome studies. Segura-Correa *et al*. [2013] using logistic regression found a significant effect of farm, year, season, parity and litter size on the presence of mummies in the litter.

According to Xu *et al*. [2017], the limitations to analyse count data are related with the lack of upper limits of data with the Poisson or negative binomial distributions. However, in animal production and animal health datasets the response count variables always have an upper limit. In this study, the variable number of mummies ranged from 0 to 7, which did not show a wide distribution of the data. This might be a reason for the poor goodness of fit for the Poisson and negative binomial models. Secondly, some risk factors could be confounded, such as parity and litter size. However, regression models are expected to adjust for a given factor, keeping the other factors constant.

Conclusion

Based on the test for overdispersion and goodness of fit criteria for the regressions models used the ZINB model described better the number of mummified fetuses per litter in sows under the tropical conditions of this study. Based on the ZINB regression model litter size was the most important factor influencing the number of mummified fetuses.

Acknowledgment. *The authors thank the owner of the production unit for providing the information for this study.*

REFERENCES

1. AGRESTI,A., 1996 – An introduction to categorical data analysis. *Wiley,* New York, USA*:* p. 74-83.
2. AKINPELU K.P., YUSUF O.B., AKPA O.M., GBOLAHAN A.O., 2016 – Zero inflated regression models with application to malaria surveillance data. ***International Journal of Statistics and Applications*** 6(4), 223-234.
3. AKRAM K., PEDERSEN-BJERGAARD U., CARSTENSEN B, BORCH-JOHNSEN K., THORSTEINSSON B., 2006 – Frequency and risk factors of several hypoglycaemia in insulin-treated type 2 diabetes: a cross-sectional survey. ***Diabetic Medicine*** 23, 750-756.
4. BORGES V. F., BERNARDI M.L., BORTOLOZZO F. P., WENTZ I., 2005 – Risk factors for stillbirth and foetal mummification in four Brazilian swine herds. **P*reventive Veterinary Medicine*** 70, 165-176.
5. CAMERON A. C., TRIVEDI P.K., 1998 – Regression analysis of count data. Cambridge University Press, London. pp.1-3.
6. 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*.*
7. DENWOOD M., STEAR M., MAHEUS L., REID S., TOFT N., INNOCENT G., 2008 – The distribution of the pathogenic nematode Nematodirus battus in lambs is zero-inflated. ***Parasitology*** 135*,* 1225-1235.
8. ESPITIA A.,GALÍNDEZ R., 2011 – Estudio de algunos factores genéticos y no genéticos que afectan el número de momias y la mortalidad al nacimiento de cerdos Duroc y Hampshire (Study of some genetic and non genetic factors that affect the number of mummies and mortality at birth in Duroc

and Hampshire pigs). ***Revista Facultad de Agrononomía*** **(UCV)** 37, 38-44. In Spanish, summary in English.

9. JOHNSON R., NIELSEN M.,CASEY D., 1999 – Response in ovulation rate, embryonic survival, and litter traits in swine to 14 generations of selection to increase litter size. ***Journal of Animal Science*** 76, 29-35.
10. KIPNIS V., MIDTHUNE D., BUCKMAN D.W., DODD K.W., GUENTHER P.M., KREBS-SMITH S.M., SUBAR A.F., TOOZE J.A., CARROLL R.J., FREEDMAN L.S., 2009 – Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes*. **Biometrics**.* 65*,* 1003-1010.
11. RICO C., GOMEZ J., 1982 – Factores genéticos y ambientales que influyen en el comportamiento reproductivo de la raza Duroc en Cuba. II. Influencia en la mortalidad (Genetic and environmetal factors that influence the reproductive performance of the Duroc breed in Cuba.). **R*evista Cubana de Ciencia Agrícola*** 16, 35-40. In Spanish, summary in English.
12. SAS, 2012 – User´s Guide. Version 9.4. Cary, North Carolina, USA.
13. SEGURA-CORREA J.C., SOLORIO-RIVERA J.L., 2013 – Risk factors for stillborn pigs and mummified fetuses in two swine farms in southeastern Mexico. ***Livestock Research for Rural Development*** 25(10). http://www.lrrd.org/lrrd25/10/solo25173.htm
14. SILVA- SANTOS K.C., SANTOS G.M.G., MOTOTTI F., MARCANTONIO T. N., SENEDA M. M., 2014 – Comparison of antral and preantral ovarian follicle populations between Bos indicus and Bos indicus-taurus cows with high and low antral follicle counts. ***Reproduction in Domestic Animals*** 49, 48-51.
15. SPOTTER A., DISTL O., 2006. Genetic approaches to the improvement of fertility traits in the pig. ***Veterinary Journal*** 172, 234-247.
16. TANI S., IIDA R., KOKETSU Y., 2016 – Climatic factors, parity and total number of pigs born associated with occurrences and number of stillborn piglets during hot or cold season in breeding herds. ***Veterinary Medicine and Animal Sciences*** 4, 1-8.
17. VARONA L., SORENSEN D. 2010 – A genetic analysis of mortality in pigs. ***Genetics*** 184, 277-284.
18. XU T., ZHU G., HAN S., 2017 – Study of depression influencing factors with zero-inflated regression models in a large-scale population survey. ***BMJ Open 7:e016471***. http://dx.doi.org/10.1136/bmjopen-2017-016471.