

## Proposal to obtain the optimal sample size of pests with an excess of zeros

---

Luis Gabriel Otero-Prevost<sup>1</sup>  
Juan A. Villanueva-Jiménez<sup>1,5</sup>  
Gustavo Ramírez-Valverde<sup>2</sup>  
Mónica C. Vargas-Mendoza<sup>1</sup>  
Carlos M. Becerril-Pérez<sup>1,2</sup>  
Lauro Soto-Rojas<sup>2</sup>

1 Colegio de Postgraduados-Campus Veracruz. Carretera Xalapa-Veracruz km 88.5, Manlio F. Altamirano, Veracruz, México. CP. 91963.

2 Colegio de Postgraduados-Campus Montecillo. Carretera México-Texcoco km 36.5, Montecillo, Texcoco, México. CP. 56230.

**Autor para correspondencia:** [javj@colpos.mx](mailto:javj@colpos.mx).

---

### Abstract

In sampling of pests with low densities, it is common to obtain a large number of zeros, which is difficult to manage since the Poisson and negative binomial probability distributions are not suitable for modeling and equations to estimate the optimal sample size are not available. In this study model the excess of zeros by estimating parameters through the methods of moments and maximum likelihood of the zero-inflated Poisson and zero-inflated negative binomial distributions, and to derive equations to calculate the optimal sample size. Systematic sampling was used to select 100 trees per grove of Río Red grapefruit (*Citrus paradisi* Macfad) at Finca Sayula, Veracruz, Mexico (latitude 19.20722, longitude -96.35194), from June to July 2021 and January 2022. The number of leafminers (*Phyllocnistis citrella* Stainton) and aphids (*Toxoptera citricida* Kirkaldy) present in three leaves per shoot per tree, considered as a sample unit, was counted. Simulations were performed in RStudio with different proportions of zero (0.1, 0.4, and 0.6) to compare the parameters obtained in the field using the methods of moments and maximum likelihood. Equations were derived to estimate the optimal sample size in studies of pests with low densities, based on the zero-inflated Poisson and zero-inflated negative binomial probability distributions. The method of moments yields optimal sample sizes smaller than those obtained by maximum likelihood, because they distinguish the origin from zero, so its use is recommended.

### Keywords:

sampling, zero-inflated negative binomial, zero-inflated Poisson.

---



License (open-access): Este es un artículo publicado en acceso abierto bajo una licencia **Creative Commons**

## Introduction

In the population dynamics of pest organisms, count data reflect the presence and abundance of species in a fixed period of time (Hashim *et al.*, 2021). It is common for samples of pest populations to present values of zero in excess due to the complex interactions between biotic and abiotic components, to the inherent characteristics of pest species, to spatial-temporal dependencies, to unexplained environmental heterogeneity (Zou *et al.*, 2021) and agroecological control techniques (Villanueva-Jimenez *et al.*, 2017; García-González *et al.*, 2018).

The study and monitoring of the periods in which pest organisms have excess zeros can be very useful since they allow carrying out preventive management of their populations and recognizing early stages of pest invasion for the application of preventive management methods, such as those offered by precision agriculture (Jankielsohn, 2017; Clay *et al.*, 2018), as well as the use of combat tactics before pests cause damage to crops, which would prevent the abusive use of organic-synthetic pesticides, thus also reducing damage to the environment (Shannon *et al.*, 2018; Talaviya *et al.*, 2020).

The excess of zeros is a theoretical and practical problem that arises when the high frequency of zeros alters the probabilities expected by the discrete variable distributions of Poisson and negative binomial (Yesilova *et al.*, 2010; Hashim *et al.*, 2021; Haslett *et al.*, 2022) and no attention has been paid to the mechanisms that explain the origin of zero despite its impact on the estimation of population parameters in species of pest organisms (Haslett *et al.*, 2022).

For the study of pest populations in agroecosystems, it is proposed to analyze the excess of zeros from the proposals of (Mullahy, 1986; Lambert 1992); that is, recognize two possible origins of zero, distinguishing between structural zero (plants without susceptible shoots for the establishment of a pest) and non-structural zero (plants with susceptible shoots free of the pest and susceptible shoots plagued), model zero by its origin with binomial distributions (Lambert, 1992; Zou *et al.*, 2021; Haslett *et al.*, 2022) and depending on the observed value of counts greater than zero, study the effect of overdispersion (Hall, 2000; Cheung, 2002; Doyle, 2009).

In pest counts, the optimal sample size equations for the Poisson or negative binomial distribution are used on a recurring basis, but due to the excess of zeros, the estimated optimal sample sizes are so large as to be impractical (Southwood and Henderson, 2000); however, in integrated pest management, there are no equations that estimate the optimal sample size of zero-inflated distributions, nor proposals that consider the origin of zero.

Equations estimating the optimal sample size are proposed here (Karandinos, 1976), which are adjusted to zero-inflated distributions. The objectives of the present research were: model the excess of zeros, estimate the parameters using the methods of moments and maximum likelihood of the zero-inflated Poisson and zero-inflated negative binomial distributions, and derive equations to calculate the optimal sample size.

## Materials and methods

For the estimation of the optimal sample size, the excess of zeros was modeled; the parameters were determined by the methods of moments and maximum likelihood of the zero-inflated Poisson and zero-inflated negative binomial distributions and the equations for calculating the sample size were derived.

### Modeling excess zeros

To model the excess of zeros, the following stages were performed: i) the absence of plant tissue that allows the pest to be housed was included as a cause of extra-zeros. In this way, there were two origins: the 'structural zero', when there is no susceptible tissue in the plant that can be occupied by the pest and the 'non-structural' zero, when there is adequate tissue in the plant, but it is not inhabited by a pest.

With this definition, the frequency of structural zero was modeled using a binomial distribution (Mullahy, 1986). Where: X is the number of structural zeros present in a sample size n, therefore:

$$X \sim B(n, p_e)$$

. Where:  $p_e$  is the proportion of structural zeros and  $q_e = 1 - p_e$  is the proportion of susceptible plant tissue free from the presence of the pest (non-structural zero), plus the plant tissue inhabited by the target species (positive integer values).

Thus, the probability function of the random variable X or the number of structural zeros in the sample of size n is given by:

$$P[X = x] = \binom{n}{x} p_e^x (1 - p_e)^{n-x}$$

1). If  $p_e$  is very large, it means that the hosts have little tissue susceptible to damage. ii) the probability of presence-absence of the pest was estimated as a conditioned variable of a binomial distribution. If Y is the number of non-structural zeros (susceptible tissue without the presence of pest) in a sample of size n, then:

$$Y|X \sim B(n-x, p_{ne})$$

$$P[Y = y|X = x] = \binom{n-x}{y} p_{ne}^y (1 - p_{ne})^{n-x-y}$$

2). Where:  $p_{ne}$  is the probability of occurrence of a non-structural zero; then, in a sample of size n,  $X = x$  is the number of structural zeros in the sample,  $Y = y$  is the number of non-structural zeros, while  $n - x - y$  is the number of units of plant tissue with the presence of a pest; in this way,  $q_{ne} = 1 - p_{ne}$  represents the proportion of the population of susceptible tissue, inhabited by the organism of interest; iii) to model the abundance of a pest that excludes structural zeros, Poisson count distribution was used when the mean is equal to the variance (equidispersion) and the negative binomial distribution when the variance is greater than the mean (Hilbe, 2011).

The Poisson distribution is used on a sample  $n - x$  when Y is the number of insects in a sample unit that is not a structural zero, so it is possible to use:

$$P(Y = y | \text{it is non-structural zero}) = \frac{e^{-\lambda} \lambda^y}{y!} \text{ for } y = 0, 1, 2, n$$

3). Where:  $\lambda$  is the mean of the number of insects in the population, excluding structural zeros (ie., sample units without susceptible tissue are not considered).

With overdispersion, the negative binomial is used, where Y is the number of insects in a unit that is not a structural zero:

$$P(Y = y | \text{it is non-structural zero}) = \frac{\Gamma(y + \frac{1}{k}) (k\lambda)^y}{\Gamma(y + 1) \Gamma(\frac{1}{k}) (1 + k\lambda)^{y + \frac{1}{k}}} \text{ for } y = 0, 1, 2, 3, n$$

4). Where:  $\lambda$  is the mean of the number of insects in the population, excluding structural zeros;  $k$  is an overdispersion parameter and  $\Gamma(y)$  is the gamma mathematical function. In this way, estimates are not affected by excess zeros (structural zeros).

It can be noted that, under this scheme, the probability of a non-structural zero is given by:

$$e^{-\lambda}$$

if it is Poisson and  $(1+k\lambda)^{-\frac{1}{k}}$  if it is negative binomial. The probability of a structural zero in both cases is  $p_e$ ; iv) to model the abundance of the pest considering the mixture of structural and non-structural zeros (the two origins of zero), there are two cases. If the mean and variance were equal (equidispersion), the population was modeled with the zero-inflated Poisson distribution (Lambert, 1992; Zou *et al.*, 2021) as follows:

$$P(Y=y) = \begin{cases} p_e + (1-p_e)e^{-\lambda} & \text{if } y=0 \\ (1-p_e)e^{-\lambda}\lambda^y/y! & \text{if } y>0 \end{cases}$$

5). The mean of this distribution is  $(1-p_e)\lambda$ ; in addition, the variance is  $(1-p_e)\lambda(1+\lambda p_e)$ . In the second case, when overdispersion was found, the zero-inflated negative binomial (ZINB) distribution was used (Fang *et al.*, 2016). Where:

$$P(Y=y) = \begin{cases} p_e + (1-p_e)(1+k\lambda)^{-\frac{1}{k}} & \text{if } y=0 \\ (1-p_e) \frac{\Gamma(y+\frac{1}{k})(k\lambda)^x}{\Gamma(y+1)\Gamma(\frac{1}{k})(1+k\lambda)^{y+\frac{1}{k}}} & \text{if } y>0 \end{cases}$$

6). The mean of this distribution is  $(1-p_e)\lambda$ ; in addition, the variance is  $(1-p_e)\lambda(1+\lambda(p_e+k))$

### Parameter estimation

To obtain the parameters of the distributions i) zero-inflated Poisson; and ii) zero-inflated negative binomial, the methods of moments and maximum likelihood were used. a) For the zero-inflated Poisson distribution, the moment estimators for  $p_e$  and  $\lambda$ , given respectively by (Banik and Kibria, 2009) are used:

$$\hat{\lambda}_{lm} = \frac{\bar{y}}{1-\bar{y}^2+s^2\bar{y}}; \hat{p}_e = \frac{s^2-\bar{y}}{\bar{y}+s^2-\bar{y}}, \text{ with } s^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

7). With  $\hat{\lambda}_m$  the estimator of moments of the mean,  $\bar{y}$  the sample mean,  $s^2$  the sample variance and  $\hat{p}_e$  the estimator of moments of occurrence of structural zero.

The maximum likelihood estimators for  $p_e$  and  $\lambda$  are obtained by maximizing the log-likelihood function given by:

$$\log L(k, \lambda) = \sum_{i=1}^n I_{(y_i=0)} \log(p_e + (1-p_e)e^{-\lambda}) + \sum_{i=1}^n I_{(y_i>0)} \log((1-p_e)e^{-\lambda}\lambda^{y_i}/y_i!)$$

8); b) for the zero-inflated negative binomial distribution, there are no moment estimators for  $p_e$ ,  $k$  and  $\lambda$  (Banik and Kibria, 2009; Hilbe, 2011). Since the excess zeros are structural (without

susceptible tissue), with  $X = x$  structural zeros in a sample of size  $n$  and since  $X \sim B(n, p_e)$ , then the moment estimator of  $p_e$  is given:

$$\hat{p}_e = \frac{X}{n}$$

9). If structural zeros are excluded, the  $n-x$  elements of the sample have a negative binomial distribution, with estimators of moments of  $k$  and  $\lambda$  given by:

$$\hat{\lambda}_m = \bar{y}_{n-x}, \hat{k} = \frac{s_{n-x}^2 - \bar{y}_{n-x}}{\bar{y}_{n-x}^2} - 1, \text{ with: } \bar{y}_{n-x} = \frac{1}{n-x} \sum_{i=1}^{n-x} y_i, y; s_{n-x}^2 = \frac{1}{n-x} \sum_{i=1}^{n-x} (y_i - \bar{y}_{n-x})^2$$

10). Where:  $\lambda_k$  is the parameter of the mean estimated by the method of moments,  $\bar{y}_{n-x}$  the sample mean,  $s_{n-x}^2$  the sample variance; and  $\hat{k}$  the estimator of moments of the dispersion parameter.

The maximum likelihood estimator for  $p_e$ ,  $k$  and  $\lambda$  are obtained by maximizing the log-likelihood function given by:

$$\log L(k, \lambda) = \sum_{i=1}^n I_{(y_i=0)} \log \left( p_e + (1-p_e) \left( 1 + k\lambda \right)^{-\frac{1}{k}} \right) + \sum_{i=1}^n I_{(y_i>0)} \log \left( (1-p_e) \frac{\Gamma\left(x + \frac{1}{k}\right) (k\lambda)^x}{\Gamma(x+1) \Gamma\left(\frac{1}{k}\right) (1+k\lambda)^{x+\frac{1}{k}}} \right)$$

11). Based on the above, it is proposed to use the moment estimators of the negative binomial distribution (Banik and Kibria, 2009), but excluding structural zeros from the equation, as an approximation to the moments of the zero-inflated negative binomial distribution.

## Derivation of equations

To derive the equations of optimal sample size, the parameters obtained from the models iii and iv were substituted in the equations of Karandinos (1976), related to the coefficient of variation (CV), the fixed proportion of the mean ( $D \cdot x$ ) and half of a confidence interval ( $h$ ) (Table 1). The values of CV,  $D$ , and  $h$  are arbitrary, so the value used in each case depends on the precision defined in each research (Ramírez *et al.*, 2013; Taherdoost, 2016). The coefficient of variation used was 25% (0.25), proposed by Southwood and Henderson (2000), a level suitable for ecological studies.

**Table 1. Equations proposed for estimating the optimal sample size of pests with low densities.**

Distribution	Optimal sample size*, based on:		
	Coefficient of variation	Proportion of the mean $D$	Confidence Interval $h$
General	$n = \left( \frac{\sigma}{\mu C} \right)^2$	$n = \left( \frac{Z_{\alpha/2}}{D} \right)^2 \frac{\sigma^2}{\mu^2}$	$n = \left( \frac{Z_{\alpha/2}}{h} \right)^2 \sigma^2$
Poisson	$n = \frac{1}{\lambda C^2}$	$n = \left( \frac{Z_{\alpha/2}}{D} \right)^2 \frac{1}{\lambda}$	$n = \left( \frac{Z_{\alpha/2}}{h} \right)^2 \lambda$

Negative binomial	$n = \frac{\frac{1}{\lambda} + \frac{1}{k}}{CV^2}$	$n = \frac{(Z_{\alpha/2})^2 \left( \frac{1}{\lambda} + \frac{1}{k} \right)}{D^2}$	$n = \left( \frac{Z_{\alpha/2}}{h} \right)^2 \left( \frac{k\bar{x} + \bar{x}^2}{k} \right)$
Zero-inflated Poisson	$n = \frac{(1+p_e)\lambda}{(1-p_e)CV^2}$	$n = \frac{(Z_{\alpha/2})^2 (1+p_e)\lambda}{(1-p_e)\lambda D^2}$	$n = \left( \frac{Z_{\alpha/2}}{h} \right)^2 (1-p_e) (1+p_e)$
Zero-inflated negative binomial	$n = \frac{(1+\lambda(p_e+k))}{(1-p_e)CV^2}$	$n = \frac{(Z_{\alpha/2})^2 (1+\lambda(p_e+k))}{(1-p_e)\lambda D^2}$	$n = \left( \frac{Z_{\alpha/2}}{h} \right)^2 (1-p_e) (1+(p_e+k))$

\*= to obtain the optimal sample size, the values of  $\lambda$ ,  $p_e$  and  $k$  are replaced by their estimators.

### Field samplings vs simulations

Six systematic samplings ( $n = 100$ ) were carried out in three Río Red grapefruit (*Citrus paradisi* Macfad) groves at Finca Sayula, SPR de RL de CV, Veracruz, Mexico (latitude 19.20722, longitude -96.35194). Sampling data were direct counts in small units (three leaves per shoot per tree), conducted during the months of June and July 2021 and January 2022.

Three of the samplings were carried out to detect the presence of the citrus leafminer *Phyllocnistis citrella* Stainton and three more to detect the presence of the citrus tristeza virus vector aphid *Toxoptera citricida* Kirkaldy. In addition, three samplings were simulated with zero-inflated Poisson and three samplings with zero-inflated negative binomial; both with  $n = 100$ , randomly generated numbers. The simulations were performed with RStudio using the programs rbinom (100, size = 1, prob = 0.1, 0.4, 0.6), rpois (100-x, 1.5), rnbinom (100, 1.5) and zeroinfl (x#1 | 1, dist = 'poisson', 'negbin') of the vgam and pscl libraries.

For the six field samplings, three of *P. citrella* (Table 2) and three of *T. citricida* (Table 3), and for the six simulations (Table 4), the simulated and observed proportion of structural zeros, the non-structural zeros, the overdispersion parameter  $k$ , the probability of structural zero and the optimal sample size were estimated using the coefficient of variation equations, proportion of mean and half confidence interval (Table 1).

**Table 2. Optimal sample size estimates using the method of moments and log-likelihood, with zero-inflated Poisson and zero-inflated negative binomial in populations of *Phyllocnistis citrella* Stainton with excess zeros in the State of Veracruz.**

Sampling	Method	Probability distribution	$Pr_{sz} / Pr_{nsz}$	$k$	$p_e$	CV	$D \bar{x}$	$h$
1	log-lik mom	ZIP ZINB	0.33/0.43	$1.4e^{-5}$ 1.29	0.67 0.67	81 69 70 75	81 69 70 75	51 51 - 351
		ZIP ZINB			0.629 0.33			
2	log-lik mom	ZIP ZINB	0.27/0.45	$1.9e^{-5}$ 2.69	0.537 0.537	53 55 43 102	53 55 43 102	41 41 - 472
		ZIP ZINB			0.465 0.27			
3	log-lik mom	ZIP ZINB	0.13/0.46	$8.1e^{-6}$ 1.35	0.543 0.543	54 34 47 42	54 34 47 42	148 148
		ZIP ZINB			0.499 0.13			50 1151

log-lik= log-likelihood; mom= moments;  $Pr_{sz}$ = proportion of structural zeros;  $Pr_{nsz}$ = proportion of non-structural zeros;  $k$ = overdispersion parameter;  $p_e$ = estimated probability of structural zero; optimal sample size by CV= coefficient of variation;  $D \bar{x}$  = proportion of the mean;  $h$ = half the confidence interval amplitude.

**Table 3. Optimal sample size estimates using the method of moments and log-likelihood, with zero-inflated Poisson and zero-inflated negative binomial in populations of *Toxoptera citricida* Kirkaldy with excess zeros in the State of Veracruz.**

Sampling	Method	Probability distribution	$Pr_{sz}/Pr_{nsz}$	k	$p_e$	CV	$D \bar{x}$	h
1	log-lik mom	ZIP ZINB	0.33/0.64	181.8 0.02	0.97 0.97	1061 2994	1061 2994	- 24686
		ZIP ZINB			0.987 0.33	2447 18	2447 18	- 1207
2	log-lik mom	ZIP ZINB	0.27/0.68	0.426 0.056	0.95 0.949	623 450	623 450	2266 3945
		ZIP ZINB			0.96 0.27	801 17	801 17	- 983
3	log-lik mom	ZIP ZINB	0.13/0.84	0.474 0.025	0.97 0.969	1050 779	1050 779	5738 8486
		ZIP ZINB			0.978 0.13	1475 12.55	1475 12.55	- 854

log-lik= log-likelihood; mom= moments;  $Pr_{sz}$ = proportion of structural zeros;  $Pr_{nsz}$ = proportion of non-structural zeros; k= overdispersion parameter;  $p_e$ = estimated probability of structural zero; optimal sample size by: CV= coefficient of variation;  $D \bar{x}$  = proportion of the mean; h= half the confidence interval amplitude.

**Table 4. Optimal sample size estimates in zero-inflated Poisson and zero-inflated negative binomial simulations generated in RStudio.**

Sampling	Method	Probability distribution	$Pr_{sz}$	k	$p_e$	CV	$D \bar{x}$	h
ZIPS1	log-lik	ZIP	0.1	$4.8e^{-5}$	0.089	19	19	29
ZIPS2	log-lik	ZIP	0.4	0.107	0.479	45	45	31
ZIPS3	log-lik	ZIP	0.6	$1e^{-5}$	0.476	45	45	22
ZINBS1	log-lik	ZINB	0.1	2.221	0.005	39	39	664
ZINBS2	log-lik	ZINB	0.4	0.623	0.429	32	32	1268
ZINBS3	log-lik	ZINB	0.6	0.656	0.651	62	62	1935

ZIPS= zero-inflated Poisson simulations (1-3); ZINBS= zero-inflated negative binomial simulations (1-3); log-lik= log-likelihood;  $Pr_{sz}$ = proportion of structural zeros; k= overdispersion parameter;  $p_e$ = estimated probability of structural zero; optimal sample size by: CV= coefficient of variation;  $D \bar{x}$  = proportion of the mean; h= half the confidence interval amplitude.

## Results and discussion

### Equations proposed for estimating the optimal sample size of pests with excess zeros

The equations proposed to estimate the optimal sample size of pests with excess zeros are detailed in the methodology (Table 1).

### Optimal sample size

It was found that the optimal sample size calculated by the proportion of the mean ( $D \bar{x} = 0.5$ ) is equivalent to the coefficient of variation (CV) proposed by Southwood and Henderson (2000). For the estimation of optimal sample size by half the confidence interval (h), no system that allowed equivalence with the coefficient of variation or proportion of the mean was found.

The optimal sample size of half the confidence interval (h) increased as the overdispersion parameter (k) increased, resulting in very large or difficult-to-estimate optimal sample sizes when pest populations have excess zeros (Tables 2, 3 and 4).

The estimation of the optimal sample size by log-likelihood of the parameter  $k$  of the samples of *P. citrella* (Table 2) indicated that the samples have zero-inflated Poisson distribution. The  $k$  estimated by the moment method of the zero-inflated negative binomial distribution, by excluding structural zeros, showed that non-structural zeros and positive integer values had overdispersion.

This result is consistent with that reported by Banik and Kibria (2009), who indicated that, by conditioning or eliminating the structural zeros of a population modeled with a zero-inflated Poisson distribution, it can also be modeled with a negative binomial distribution, provided that the data of the non-structural component present overdispersion.

The values of  $p_e$  for the methods of moments and log-likelihood for zero-inflated Poisson were similar, therefore, both methods are efficient for the estimation of the parameters. The estimated sample sizes for *P. citrella* are smaller when estimated by moments than by log-likelihood, even when the number of structural zeros ( $Pr_{sz}$ ) is greater; however, the difference between the two estimates is not very large ( $< 20$  units).

The effect of overdispersion significantly affected the sample size estimated by  $h$ ; for *P. citrella*, the results indicate that estimation by CV or by  $D\bar{x}$  is preferable since, although the interval ranges from 47 to 70, the sample size is smaller than that obtained by Poisson and negative binomial, because the methods proposed here consider the number of structural and non-structural zeros.

In the samplings of *T. citricida* (Table 3), an insect with a high tendency to aggregation, the  $k$  values estimated by log-likelihood indicate populations with zero-inflated negative binomial distribution. The value of  $k$  by the method of the moments resulted in low values, which indicates that, when excluding the structural component, the few sample units found with pest presented low variation.

The result is interesting since populations with zero-inflated negative binomial distribution present random distribution at the farm level, but the few occupied trees had a high number of individuals, indicating aggregation, in accordance with the biology of the insect. The exclusion of structural zero, the frequency of non-structural zeros, and the reduction of variation in counts with positive integer values resulted in sample sizes very small for CV and  $D\bar{x}$  estimated with the moment method.

The optimal sample size of the zero-inflated negative binomial distribution, calculated by moments, is smaller because it distinguishes the different origins of zero. By considering only the non-structural zeros and the positive integer values for the estimation of the sample size, a difference was established with the parameters estimated by log-likelihood that does not distinguish the origin of zero. Therefore, the method of moments for zero-inflated Poisson and zero-inflated negative binomial allows estimating optimal sample sizes similar to or smaller than those estimated by maximum likelihood.

In the simulations (Table 4), it was observed that, as the number of structural zeros increased, the sample size increased in both distributions since, as the sample size was only estimated by the log-likelihood method, when simulating, the origin of zero is not distinguished. In addition, the estimated value of the overdispersion parameter  $k$  is consistent with the values obtained in the field.

For zero-inflated Poisson, very small  $k$  values were obtained due to the proximity of the mean and variance values, while for the simulations of the zero-inflated negative binomial, the overdispersion parameter was greater than zero, indicating overdispersion, similar to that reported by Zou *et al.* (2021); Haslett *et al.* (2022).

## Conclusions

The zero-inflated Poisson and zero-inflated negative binomial probability distributions allow modeling populations of pest organisms with low densities and excess zeros. The parameters obtained by the moment method distinguish the origin of zero and estimate optimal sample sizes equivalent to or less than those estimated by log-likelihood, which does not distinguish the origin of zero. A zero-inflated Poisson population can also be modeled with a negative binomial distribution, provided that the non-structural component is overdispersed.

The estimation of the optimal sample size in pest populations with excess zeros can be performed equivalently with the coefficient of variation (CV) equation and the mean proportion ( $\bar{D}_x$ ) equation. On the other hand, the estimation of the optimal sample size with the equation of the half the confidence interval (h) depends on the value of the overdispersion parameter (k), since it does not have a fixed value that allows establishing an equivalence.

## Bibliography

- 1 Banik, S. and Kibria, B. M. G. 2009. On some discrete distributions and their applications with real life data. USA. *JMASM*. 8(2):423-447. <https://doi.org/10.22237/jmasm/1257034020> .
- 2 Cheung, Y. B. 2002. Zero inflated models for regression analysis of count data: a study of growth and development. USA. *Statist. Med.* 21(10):1461-1469. <https://doi.org/10.1002/sim.1088>.
- 3 Clay, S. A.; French, B. W. and Mathew, F. M. 2018. Pest measurement and management. *In: precision agriculture basics*. Shanon, D. K.; Clay, D.E. and Kitchen N. R. (eds.). Ed. ASA, CSSA, and SSSA Books. USA. 93-102 pp. <https://doi.org/10.2134/precisionagbasics.2016.0090> .
- 4 Doyle, S. R. 2009. Examples of computing power for zero-inflated and over dispersed count data. USA. *JMASM*. 8(2):360-376. <https://doi.org/10.22237/jmasm/1257033720> .
- 5 Fang, R.; Wagner, B. D.; Harris, J. K. and Fillon, S. A. 2016. Zero inflated negative binomial mixed models: and important application to two microbial organisms important in oesophagitis. UK. *Epidemiol. Infect.* 144(1):2447-2455. <http://doi.org/10.1017/S0950268816000662>.
- 6 García-González, J. C.; López-Collado, J.; García-García, C. G.; Villanueva-Jiménez, J. A. y Nava-Tablada, M. E. 2018. Factores bióticos, abióticos y agronómicos que afectan las poblaciones de adultos de mosca pinta (Hemiptera: Cercopidae) en cultivos de caña de azúcar en Veracruz, México. México. *Acta Zool. Mex.* 33(3):508-517. <https://doi.org/10.21829/azm.2017.3331152>.
- 7 Hall, D. B. 2000. Zero inflated Poisson and binomial regression with random effects: a case study. USA. *Biometrics*. 56(1):1030-1039. <https://doi.org/10.1111/j.0006-341x.2000.01030.x>.
- 8 Hashim, L. H.; Hashim, K. H. and Shiker, M. A. K. 2021. An application comparison of two Poisson models on zero count data. UK. *journal of physics: conference series*, 1818(012165):1-12. <http://doi:10.1088/1742-6596/1818/1/012165>.
- 9 Haslett, J.; Parnel, A. C.; Hinde, J. and de Andrade, M. R., 2022. Modelling excess of zeros in count data: a new perspective on modelling approaches. USA. *International statistical review*. 90(2):216-236. <https://doi.org/10.1111/insr.12479>.
- 10 Hilbe, J. M. 2011. *Negative binomial regression*. Cambridge University Press. 2<sup>a</sup> Ed. UK. 346-399 pp.
- 11 Jankielsohn, A. 2017. The redesign of suitable agricultural crop ecosystems by increasing natural ecosystem services provided by insects. Hong Kong SAR China. *Advances in ecological and environmental research*. 1(1):365-381. <http://www.ss-pub.org/wp-content/uploads/2017/09/AEER2017040501-1.pdf>.
- 12 Karandinos, M. G. 1976. Optimum sample size and comments on one published formula. USA. *Bull. Entomol. Soc. Amer.* 22(4):417-421. <https://doi.org/10.1093/besa/22.4.417> .
- 13 Lambert, D. 1992. Zero inflated Poisson regression, with an application to defects manufacturing. USA. *Technometrics*. 34(1):1-14. <https://doi.org/10.2307/1269547>.
- 14 Mullahy, J. 1986. Specification and testing of some modified count data models. Netherlands. *J. Econ.* 33(1):341-365. [https://doi.org/10.1016/0304-4076\(86\)90002-3](https://doi.org/10.1016/0304-4076(86)90002-3) .

- 15 Ramírez, I. C.; Barrera, C. J. y Correa, J. C. 2013. Efecto del tamaño de muestra y el número de réplicas bootstrap. Colombia. *Inycompe*. 15(1):93-101. <https://www.redalyc.org/articulo.oa?id=291329165008>.
- 16 Shannon, D. K.; Clay, D. E. and Sudduth, K. A. 2018. And introduction to precision agriculture. *In: precision agriculture basics* . Shanon, D. K.; Clay, D.E. and Kitchen N. R. (eds.). Ed. ASA, CSSA, and SSSA Books. USA. 1-12 pp. <https://doi.org/10.2134/precisionagbasics.2016.0084>.
- 17 Southwood, T. R. E. and Henderson, P. A. 2000. *Ecological methods*. Blackwell science. 3<sup>rd</sup> Ed. Oxford, UK. 7-66 pp. <https://www.researchgate.net/publication/260051655-Ecological-Methods-3rd-edition>.
- 18 Taherdoost, H. 2016. Sampling methods in research methodology, how to choose a sampling technique for research. Brazil. *IJARM*. 5(2):18-27. <http://dx.doi.org/10.2139/ssrn.3205035> .
- 19 Talaviya, T.; Shah, D.; Patel, N.; Yagnik, H. and Shah, M. 2020. Implementation of artificial intelligence in agriculture for optimization of irrigation and application of pesticides and herbicides. China. *Artificial Intelligence in Agric*. 4(1):58-73. <https://doi.org/10.1016/j.iiia.2020.04.002>.
- 20 Villanueva-Jiménez, J. A.; Reyes-Pérez, N. y Abato-Zárate, M. 2017. Manejo integrado de plagas y sostenibilidad. *In: agricultura sostenible como base para los agronegocios*. Jarquín, G. R. y Huerta, P. A. (coords.). 1a Ed. Universidad Autónoma de San Luis Potosí. México. 32-42 pp. <https://www.researchgate.net/publication/320779257-Manejo-Integrado-de-Plagas-y-Sostenibilidad> .
- 21 Yesilova, A.; Kaydan, M. B. and Kaya, Y. 2010. Modeling insect-egg data with excess zero using zero-inflated regression models. *Hacettepe J. Math. Stat*. 39(2):273-282. [http://www.hjms.hacettepe.edu.tr/uploads/c879f14e-8c0d-4f30-8bfa-e28658\\_a8fe0b.pdf](http://www.hjms.hacettepe.edu.tr/uploads/c879f14e-8c0d-4f30-8bfa-e28658_a8fe0b.pdf).
- 22 Zou, Y.; Hanning, J. and Young, D. S. 2021. Generalized fiducial inference on the mean of zero inflated Poisson and Poisson hurdle models. Germany. *J Statistical Distributions and Applications*. 8(5):1-15. <https://doi.org/10.1186/s40488-021-00117-0>.





## Proposal to obtain the optimal sample size of pests with an excess of zeros

Journal Information
Journal ID (publisher-id): remexca
Title: Revista mexicana de ciencias agrícolas
Abbreviated Title: Rev. Mex. Cienc. Agríc
ISSN (print): 2007-0934
Publisher: Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias

Article/Issue Information
Date received: 01 November 2023
Date accepted: 01 January 2024
Publication date: 07 February 2024
Publication date: January 2024
Volume: 15
Issue: 1
Electronic Location Identifier: e3618
DOI: 10.29312/remexca.v15i1.3618

**Categories**

Subject: Articles

**Keywords:**

**Keywords:**

sampling  
zero-inflated negative binomial  
zero-inflated Poisson

**Counts**

Figures: 0  
Tables: 4  
Equations: 29  
References: 22  
Pages: 0