Revista Mexicana de Ciencias Agrícolas
REMEXCA
DOI: https://doi.org/10.29312/remexca.v15i1.3618
Artículo

Propuesta para obtener el tamaño de muestra óptimo de plagas con exceso de ceros

Luis Gabriel Otero-Prevost[1]
Juan A. Villanueva-Jiménez[1,§]
Gustavo Ramírez-Valverde[2]
Mónica C. Vargas-Mendoza[1]
Carlos M. Becerril-Pérez[1,2]
Lauro Soto-Rojas[2]

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.

Resumen

En muestreos de plagas con densidades bajas es común obtener gran cantidad de ceros, lo que es difícil de manejar, ya que las distribuciones de probabilidad Poisson y binomial negativa no son adecuadas para su modelación y no se dispone de ecuaciones para estimar el tamaño de muestra óptimo. En este estudio se modelo el exceso de ceros mediante la estimación de parámetros a través de los métodos de momentos y de máxima verosimilitud de las distribuciones Poisson cero inflado y binomial negativa cero inflado, y derivar ecuaciones para calcular el tamaño de muestra óptima. Se utilizó muestreo sistemático para seleccionar 100 árboles por huerto de toronja (*Citrus paradisi* Macfad) Río Red, en la Finca Sayula, Veracruz, México (latitud 19.20722, longitud -96.35194), de junio a julio 2021 y enero 2022. Se contó el número de minadores (*Phyllocnistis citrella* Stainton) y pulgones (*Toxoptera citricida* Kirkaldy) presentes en tres hojas por brote por árbol, consideradas como unidad muestral. Se realizaron simulaciones en RStudio con diferentes proporciones de cero (0.1, 0.4 y 0.6) para comparar los parámetros obtenidos en campo, mediante el método de los momentos y máxima verosimilitud. Se derivaron ecuaciones para estimar el tamaño de muestra óptimo en estudios de plagas con densidades bajas, a partir de las distribuciones de probabilidad Poisson cero inflado y binomial negativa cero inflado. El método de los momentos arroja tamaños de muestra óptimos menores a aquellos obtenidos mediante máxima verosimilitud, debido a que distinguen el origen del cero, por lo que se recomienda su uso.

Palabras clave:

binomial negativa cero inflado, muestreo, Poisson cero inflado

DOI: https://doi.org/10.29312/remexca.v15i1.3618

Introducción

En la dinámica poblacional de organismos plaga, los datos de conteos reflejan la presencia y abundancia de las especies en un periodo fijo de tiempo (Hashim *et al.*, 2021). Debido a las complejas interacciones entre los componentes bióticos y abióticos, a las características inherentes de las especies plaga, a dependencias espacio-temporales, a la heterogeneidad ambiental no explicada (Zou *et al.*, 2021) y a las técnicas agroecológicas de control (Villanueva-Jiménez *et al.*, 2017; García-González *et al.*, 2018) es común que en los muestreos de poblaciones plaga se presenten valores de cero en exceso.

El estudio y el monitoreo de los periodos en que los organismos plaga presentan exceso de ceros puede ser de gran utilidad, ya que permiten realizar el manejo preventivo de sus poblaciones y reconocer etapas tempranas de invasión de plagas para la aplicación de métodos preventivos de manejo, tales como los que oferta la agricultura de precisión (Jankielsohn, 2017; Clay *et al.*, 2018), así como el uso de las tácticas de combate antes de que las plagas causen daño en los cultivos, lo que evitaría el uso abusivo del plaguicidas organosintéticos, con lo que también se reduciría el daño al ambiente (Shannon *et al.*, 2018; Talaviya *et al.*, 2020).

El exceso de ceros es un problema teórico y práctico que se presenta cuando la elevada frecuencia de ceros, altera las probabilidades esperadas por las distribuciones de variables discretas Poisson y binomial negativa (Yesilova *et al.*, 2010; Hashim *et al.*, 2021; Haslett *et al.*, 2022) y no se ha puesto atención en los mecanismos que explican el origen del cero, a pesar de sus repercusiones en la estimación de los parámetros poblacionales en especies de organismos plaga (Haslett *et al.*, 2022).

Para el estudio de poblaciones de plagas en agroecosistemas, se propone analizar el exceso de cero a partir de las propuestas de (Mullahy, 1986; Lambert 1992); es decir, reconocer dos posibles orígenes del cero distinguiendo entre cero estructural (plantas sin brotes susceptibles para el establecimiento de una plaga) y cero no estructural (plantas con brotes susceptibles libres de la plaga y brotes susceptibles plagados), modelar el cero por su origen con distribuciones binomiales (Lambert, 1992; Zou *et al.*, 2021: Haslett *et al.*, 2022) y dependiendo del valor observado de los conteos mayores que cero, estudiar el efecto de la sobredispersión (Hall, 2000; Cheung, 2002; Doyle, 2009).

En conteos de plagas se emplean de manera recurrente las ecuaciones de tamaño óptimo de muestra para la distribución Poisson o binomial negativa, pero debido al exceso de ceros, los tamaños de muestra óptimos estimados son tan grandes que resultan imprácticos (Southwood y Henderson, 2000); sin embargo, en el manejo integrado de plagas no se dispone de ecuaciones que estimen el tamaño de muestra óptimo de distribuciones cero infladas, ni propuestas que consideren el origen del cero.

Aquí se proponen ecuaciones que estiman el tamaño de muestra óptimo (Karandinos, 1976), ajustadas a las distribuciones cero infladas. Los objetivos de la presente investigación fueron: modelar el exceso de ceros, estimar los parámetros mediante los métodos de los momentos y máxima verosimilitud de las distribuciones Poisson cero inflado y binomial negativa cero inflado, y derivar ecuaciones para calcular el tamaño de muestra óptimo.

Revista Mexicana de Ciencias Agrícolas

DOI: https://doi.org/10.29312/remexca.v15i1.3618

Materiales y métodos

Para la estimación del tamaño de muestra óptimo, se modeló el exceso de ceros; se determinaron los parámetros mediante los métodos de momentos y máxima verosimilitud de las distribuciones Poisson cero inflado y binomial negativa cero inflado y se derivaron las ecuaciones de cálculo del tamaño de muestra.

Modelado del exceso de ceros

Para modelar el exceso de ceros, se realizaron las siguientes etapas: i) se incluyó como causa de extra-ceros, la ausencia de tejido vegetal que permite alojar a la plaga. De esta manera, se tuvieron dos orígenes: el 'cero estructural', cuando no existe en la planta tejido susceptible que pueda ser ocupado la plaga y el cero 'no estructural', cuando si existe en la planta tejido adecuado, pero no está habitado por una plaga.

Con esta definición, se modeló la frecuencia del cero estructural mediante una distribución binomial (Mullahy, 1986). Donde: X es el número de ceros estructurales presentes en una muestra tamaño n, por tanto: $X \sim B(n,p_e)$. Donde: P_e es la proporción de ceros estructurales y $q_e = 1- p_e$ es la proporción de tejido vegetal susceptible libre de la presencia de la plaga (cero no estructural), más el tejido vegetal habitado por la especie blanco (valores enteros positivos).

De esta manera, la función de probabilidad de la variable aleatoria X o el número de ceros estructurales en la muestra de tamaño n, está dada por:

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

1). Si p_e es muy grande significa que los hospederos presentan poco tejido susceptible de ser dañado; ii) se estimó la probabilidad de presencia-ausencia de la plaga, como una variable condicionada de una distribución binomial. Si Y es el número de ceros no estructurales (tejido susceptible sin la presencia de plaga) en una muestra de tamaño n, entonces:

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

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

2). Donde: p_{ne} es la probabilidad de ocurrencia de un cero no estructural; entonces, en una muestra de tamaño n, X = x es el número de ceros estructurales en la muestra, Y = y es el número de ceros no estructurales, mientras que n-x-y es el número de unidades de tejido vegetal con la presencia de una plaga; de esta forma, $q_{ne} = 1- p_{ne}$ representa la proporción de la población del tejido susceptible, habitado por el organismo de interés; iii) para modelar la abundancia de una plaga que excluya los ceros estructurales, se emplearon las distribuciones de conteos Poisson cuando la media sea igual que la varianza (equidispersión) y binomial negativa cuando la varianza sea mayor que la media (Hilbe, 2011).

La distribución Poisson se utiliza en una muestra n- x cuando Y es el número de insectos en una unidad muestral que no es un cero estructural, por lo que se puede utilizar:

$$P(Y=y|\text{es cero no estructural})=\frac{e^{-\lambda}\lambda^y}{y!} \text{ para } y=0, 1, 2, n$$

DOI: https://doi.org/10.29312/remexca.v15i1.3618

3). Donde: (es la media del número de insectos en la población excluyendo los ceros estructurales (es decir, no se toman en cuenta las unidades muestrales sin tejido susceptible).

Con sobredispersión se emplea la binomial negativa, donde sea Y el número de insectos en una unidad que no es un cero estructural:

$$P(Y=y|\,\text{es cero no estructural})=\frac{\Gamma\left(y+\frac{1}{k}\right)(k\lambda)^{y}}{\Gamma(y+1)\Gamma\left(\frac{1}{k}\right)(1+k\lambda)^{y+\frac{1}{k}}}\ \text{para } y=0,1,2,3,n$$

4). Donde: (es la media del número de insectos en la población, excluyendo los ceros estructurales; k es un parámetro de sobre dispersión y ((y) es la función matemática gamma. De esta manera, las estimaciones no se ven afectadas por el exceso de ceros (ceros estructurales).

Puede notarse que, bajo este esquema, la probabilidad de un cero no estructural está dado por

$e^{-\lambda}$

si es Poisson y (1+kλ)^(-1/k) si es binomial negativa. La probabilidad de un cero estructural en ambos casos es p_e; iv) para modelar la abundancia de la plaga considerando la mezcla los ceros estructurales y no estructurales (los dos orígenes del cero) se tienen dos casos. Si la media y varianza es igual (equidispersión), se modeló la población con la distribución Poisson cero inflada (Lambert, 1992; Zou *et al.*, 2021) de la siguiente manera:

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

5). La media de esta distribución es (1- p_e)# ; además, la varianza es (1- p_e)λ(1+λp_e) En el segundo caso, cuando se encontró sobredispersión, se utilizó la distribución binomial negativa cero inflada (BNCI) (Fang *et al.*, 2016). Donde:

$$P(Y=y)=\begin{Bmatrix} p_e+(1-p_e)(1+k\lambda)^{-\frac{1}{k}}\text{si } y=0 \\ (1-p_e)\dfrac{\Gamma\left(y+\frac{1}{k}\right)(k\lambda)^{x}}{\Gamma(y+1)\Gamma\left(\frac{1}{k}\right)(1+k\lambda)^{y+\frac{1}{k}}}\text{si } y>0 \end{Bmatrix}$$

6). La media de esta distribución es (1- p_e)"λ" ; además, la varianza es (1- p_e)λ(1+λ(p_e+k)).

Estimación de los parámetros

Para obtener los parámetros de las distribuciones i) Poisson cero inflada; y ii) binomial negativa cero inflada, se emplearon los métodos de los momentos y máxima verosimilitud. a) para la distribución Poisson cero inflada se utilizan los estimadores de momentos para p_e y λ, dados respectivamente por (Banik y Kibria, 2009):

$$\hat{\lambda}_m=\frac{\bar{y}}{1-\bar{y}^2+s^2-\bar{y}}\ ;\ \hat{p_e}=\frac{s^2-\bar{y}}{\bar{y}^2+s^2-\bar{y}},\ \text{con } s^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2$$

7). Con $\hat{\lambda}_m$ el estimador de momentos de la media, $\bar{y}$ la media muestral, s^2 la varianza muestral y $\hat{p}_e$ el estimador de momentos de ocurrencia de cero estructural.

Revista Mexicana de Ciencias Agrícolas
REMEXCA
DOI: https://doi.org/10.29312/remexca.v15i1.3618

El estimador de máxima verosimilitud para p_e y λ se obtienen maximizando la función de log-verosimilitud dada por:

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

8); b) para la distribución binomial negativa cero inflada no existen los estimadores de momentos para p_e, k y λ(Banik y Kibria, 2009; Hilbe, 2011). Dado que el exceso de ceros es estructural (sin tejido susceptible), con $X = x$ ceros estructurales en una muestra de tamaño n y que $X\sim B(n,p_e)$, entonces el estimador de momentos de p_e está dado por $\hat{p}_e=\frac{x}{n}$ 9). Si se excluyen los ceros estructurales, los $n-x$ elementos de la muestra presentan una distribución binomial negativa, con estimadores de momentos de k y (dados por:

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

10). Donde: $\hat{\lambda}_m$ es el parámetro de la media estimado por el método de los momentos, $\bar{y}_{n-x}$ la media muestral, s_{n-x}^2 la varianza muestral; y $\hat{k}$ el estimador de momentos del parámetro de dispersión.

El estimador de máxima verosimilitud para p_e, k y λ se obtienen maximizando la función de log-verosimilitud dada por:

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

11). Con base en lo anterior, se propone usar los estimadores de momentos de la binomial negativa (Banik y Kibria, 2009), pero excluyendo de la ecuación los ceros estructurales, como una aproximación a los momentos de la binomial negativa cero inflada.

Derivación de las ecuaciones

Para derivar las ecuaciones de tamaño de muestra óptimo, se sustituyeron los parámetros obtenidos de los modelos iii y iv en las ecuaciones de Karandinos (1976), relacionadas con el coeficiente de variación (CV), la proporción fija de la media ($D\bar{x}$) y la mitad de un intervalo de confianza (h) (Cuadro 1). Los valores de CV, D y h son arbitrarios, por lo que el valor empleado en cada caso depende de la precisión que se defina en cada investigación (Ramírez *et al.*, 2013; Taherdoost, 2016). Se utilizó el coeficiente de variación de 25% (0.25), propuesto por Southwood y Henderson (2000), nivel adecuado para estudios ecológicos.

Revista Mexicana de Ciencias Agrícolas
REMEXCA
DOI: https://doi.org/10.29312/remexca.v15i1.3618

Cuadro 1. Ecuaciones propuestas para la estimación del tamaño de muestra óptimo de plagas con densidades bajas.

Distribución	Tamaño de muestra óptimo*, con base en:		
	Coeficiente de variación	Proporción de la media D	Intervalo de confianza 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	$n=\frac{1}{\lambda C^2}$	$n=\left(\frac{Z_{\alpha/2}}{D}\right)^2\frac{1}{\lambda}$	$n=\left(\frac{Z_{\alpha/2}}{h}\right)^2\lambda$
Binomial negativa	$n=\frac{\frac{1}{\lambda}+\frac{1}{k}}{CV^2}$	$n=\frac{(Z_{\alpha/2})^2\left(\frac{1}{\lambda}+\frac{1}{k}\right)}{D^2}$	$n=\left(\frac{Z_{\alpha/2}}{h}\right)^2\left(\frac{k\bar{x}+\bar{x}^2}{k}\right)$
Poisson cero inflado	$n=\frac{(1+p_e)\lambda}{(1-p_e)CV^2}$	$n=\frac{(Z_{\alpha/2})^2(1+p_e)\lambda}{(1-p_e)\lambda D^2}$	$n=\left(\frac{Z_{\alpha/2}}{h}\right)^2(1-p_e)\,(1+p_e)$
Binomial negativa cero inflado	$n=\frac{(1+\lambda(p_e+k))}{(1-p_e)CV^2}$	$n=\frac{(Z_{\alpha/2})^2(1+\lambda(p_e+k))}{(1-p_e)\lambda D^2}$	$n=\left(\frac{Z_{\alpha/2}}{h}\right)^2(1-p_e)\,(1+(p_e+k))$

*= para obtener el tamaño de muestra óptimo, se sustituyen los valores de λ, p_e y k por sus estimadores.

Muestreos en campo *vs* simulaciones

Se realizaron seis muestreos sistemáticos (n= 100) en tres huertos de toronja (*Citrus paradisi* Macfad) Río Red en la Finca Sayula, SPR de RL de CV, Veracruz, México (latitud 19.20722, longitud -96.35194). Los datos de los muestreos fueron conteos directos en unidades pequeñas (tres hojas por brote por árbol), realizados durante los meses de junio y julio de 2021 y enero de 2022.

Tres de los muestreos se efectuaron para detectar la presencia del minador de la hoja de los cítricos *Phyllocnistis citrella* Stainton y tres más para detectar la presencia del pulgón vector del virus de la tristeza de los cítricos *Toxoptera citricida* Kirkaldy. Además, se simularon tres muestreos con Poisson cero inflado y tres muestreos con binomial negativa cero inflado; ambos con n= 100, números generados al azar. Las simulaciones se realizaron con RStudio, mediante los programas rbinom (100, size = 1, prob = 0.1, 0.4, 0.6), rpois (100-x, 1.5), rnbinom (100, 1.5) y zeroinfl (x~1 | 1, dist = ‘poisson’, ‘negbin’) de las librerías vgam y pscl.

Para los seis muestreos de campo, tres de *P. citrella* (Cuadro 2) y tres de *T. citricida* (Cuadro 3) y para las seis simulaciones (Cuadro 4), se estimó la proporción simulada y observada de ceros estructurales, los ceros no estructurales, el parámetro *k* de sobredispersión, la probabilidad de cero estructural y el tamaño de muestra óptimo mediante las ecuaciones de coeficiente de variación, proporción de la media y mitad de intervalo de confianza (Cuadro 1).

Revista Mexicana de Ciencias Agrícolas
REMEXCA
DOI: https://doi.org/10.29312/remexca.v15i1.3618

Cuadro 2. Estimaciones de tamaño de muestra óptimo mediante el método de los momentos y log-verosimilitud, con Poisson cero inflado y binomial negativa cero inflada en poblaciones de *Phyllocnistis citrella* Stainton con exceso de ceros en el Estado de Veracruz.

Muestreo	Método	Distribución de probabilidad	Pr_{ce} / Pr_{cne}	k	p_e	CV	D $\bar{x}$	h
1	log-ver mom	PCI BNCI PCI BNCI	0.33/0.43	$1.4e^{-5}$ 1.29	0.67 0.67 0.629 0.33	81 69 70 75	81 69 70 75	51 51 - 351
2	log-ver mom	PCI BNCI PCI BNCI	0.27/0.45	$1.9e^{-5}$ 2.69	0.537 0.537 0.465 0.27	53 55 43 102	53 55 43 102	41 41 - 472
3	log-ver mom	PCI BNCI PCI BNCI	0.13/0.46	$8.1e^{-6}$ 1.35	0.543 0.543 0.499 0.13	54 34 47 42	54 34 47 42	148 148 50 1151

log-ver= log-verosimilitud; mom= momentos; Pr_{ce}= proporción de cero estructural; Pr_{cne}= proporción de cero no estructural; k= parámetro de sobredispersión; p_e= probabilidad estimada de cero estructural; tamaño de muestra óptimo por: CV= coeficiente de variación; D $\bar{x}$ = proporción de la media; h= mitad de amplitud de intervalo de confianza.

Cuadro 3

Estimaciones de tamaño de muestra óptimo mediante el método de los momentos y log-verosimilitud, con Poisson cero inflado y binomial negativa cero inflada en poblaciones de *Toxoptera citricida* Kirkaldy con exceso de ceros en el Estado de Veracruz.

Muestreo	Método	Distribución de probabilidad	Pr_{ce}/ Pr_{cne}	k	p_e	CV	D $\bar{x}$	h
1	log-ver mom	PCI BNCI PCI BNCI	0.33/0.64	181.8 0.02	0.97 0.97 0.987 0.33	1061 2994 2447 18	1061 2994 2447 18	- 24686 - 1207
2	log-ver mom	PCI BNCI PCI BNCI	0.27/0.68	0.426 0.056	0.95 0.949 0.96 0.27	623 450 801 17	623 450 801 17	2266 3945 - 983
3	log-ver mom	PCI BNCI PCI BNCI	0.13/0.84	0.474 0.025	0.97 0.969 0.978 0.13	1050 779 1475 12.55	1050 779 1475 12.55	5738 8486 - 854

log-ver= log-verosimilitud; mom= momentos; Pr_{ce}= proporción de cero estructural; Pr_{cne}= proporción de cero no estructural; k= parámetro de sobredispersión; p_e= probabilidad estimada de cero estructural; tamaño de muestra óptimo por CV= coeficiente de variación; D $\bar{x}$ = proporción de la media; h= mitad de amplitud de intervalo de confianza.

Cuadro 4. Estimaciones de tamaño de muestra óptimo en simulaciones Poisson cero inflada y binomial negativa cero infladas generadas en RStudio.

Muestreo	Método	Distribución de probabilidad	Pr_{ce}	k	p_e	CV	D $\bar{x}$	h
SPCI1	log-ver	PCI	0.1	$4.8e^{-5}$	0.089	19	19	29
SPCI2	log-ver	PCI	0.4	0.107	0.479	45	45	31
SPCI3	log-ver	PCI	0.6	$1e^{-5}$	0.476	45	45	22
SBNCI1	log-ver	BNCI	0.1	2.221	0.005	39	39	664
SBNCI2	log-ver	BNCI	0.4	0.623	0.429	32	32	1268
SBNCI3	log-ver	BNCI	0.6	0.656	0.651	62	62	1935

SPCI= simulaciones Poisson cero inflado (1-3); SBNCI= simulaciones binomiales negativa cero inflado (1-3); log-ver= log-verosimilitud; Pr_{ce}= proporción de cero estructural; k= parámetro de sobredispersión; p_e= probabilidad estimada de cero estructural; tamaño de muestra óptimo por: CV= coeficiente de variación; D $\bar{x}$ = proporción de la media; h= mitad de amplitud de intervalo de confianza.

DOI: https://doi.org/10.29312/remexca.v15i1.3618

Resultados y discusión

Ecuaciones propuestas para estimar el tamaño de muestra óptimo de plagas con exceso de ceros

Las ecuaciones propuestas para estimar el tamaño de muestra óptimo de plagas con exceso de ceros se detallan en la metodología (Cuadro 1).

Tamaño de muestra óptimo

Se encontró que el tamaño de muestra óptimo calculado por la proporción de la media ($D_{\bar{x}}$= 0.5) es equivalente al coeficiente de variación (CV) propuesto por Southwood y Henderson (2000). Para la estimación de tamaño de muestra óptimo por la mitad del intervalo de confianza (h), no se encontró un sistema que permitiera la equivalencia con el coeficiente de variación o proporción de la media.

El tamaño de muestra óptimo de la mitad del intervalo de confianza (h) incrementó conforme aumentó el parámetro de sobredispersión (k), y resultó en tamaños de muestra óptimo muy grandes o difíciles de estimar, cuando las poblaciones de plagas presentan exceso de ceros (Cuadros 2, 3 y 4).

La estimación del tamaño de muestra óptimo por log-verosimilitud del parámetro k de las muestras de *P. citrella* (Cuadro 2) indicó que los muestreos tienen distribución Poisson cero inflado. La k estimada por el método de momentos de la distribución binomial negativa cero inflado, al excluir los ceros estructurales, mostró que los ceros no estructurales y los valores enteros positivos presentaban sobredispersión.

Este resultado concuerda con lo reportado por Banik y Kibria (2009), quienes indicaron que al condicionar o eliminar los ceros estructurales de una población modelada con una distribución Poisson cero inflado, también se puede modelar con una distribución binomial negativa, siempre que los datos del componente no estructural presenten sobredispersión.

Los valores de p_e para los métodos de momentos y de log-verosimilitud para Poisson cero inflado fueron similares, por tanto, ambos métodos son eficientes para la estimación de los parámetros. Los tamaños de muestra estimados para *P. citrella* son menores cuando se estiman por momentos que por log-versomilitud, incluso cuando el número de ceros estructurales (Pr_{ce}) es mayor; sin embargo, la diferencia entre ambas estimaciones no es muy grande (< 20 unidades).

El efecto de la sobredispersión afectó de manera importante el tamaño de muestra estimado por h; para *P. citrella* los resultados indican que es preferible la estimación por CV o por $D_{\bar{x}}$, ya que aunque el intervalo oscila entre 47 y 70, el tamaño de muestra es menor al obtenido por Poisson y binomial negativa, debido a que los métodos aquí propuestos consideran el número de ceros estructurales y no estructurales.

En los muestreos de *T. citricida* (Cuadro 3), insecto con alta tendencia a la agregación, los valores de k estimados por log-verosimilitud indican poblaciones con distribución binomial negativa cero inflado. El valor de k por el método de los momentos resultó en valores bajos, lo que indica que, al excluir el componente estructural, las pocas unidades muestrales encontradas con plaga presentaron baja variación.

DOI: https://doi.org/10.29312/remexca.v15i1.3618

El resultado es interesante, ya que las poblaciones con distribución binomial negativa cero inflado presentan distribución al azar al nivel de predio, pero los pocos árboles ocupados tuvieron un elevado número de individuos, lo que indica agregación, en concordancia con la biología del insecto. La exclusión del cero estructural, la frecuencia de los ceros no estructurales y la reducción de la variación en los conteos con valores enteros positivos dieron por resultado tamaños de muestra muy pequeños para el CV y $D_{\bar{x}}$ estimados con el método de los momentos.

El tamaño de muestra óptimo de la distribución binomial negativa cero inflada, calculada por momentos, es menor porque distingue los diferentes orígenes del cero. Al tomar en cuenta para la estimación del tamaño de muestra solo los ceros no estructurales y los valores enteros positivos, se estableció una diferencia con los parámetros estimados por log-verosimilitud que no distingue el origen del cero. Por tanto, el método de los momentos para Poisson cero inflado y binomial negativa cero inflada permite estimar tamaños de muestra óptimos similares o menores que los estimados por máxima verosimilitud.

En las simulaciones (Cuadro 4) se observó que, conforme incrementaba el número de ceros estructurales, incrementaba el tamaño de muestra en ambas distribuciones, ya que como sólo se estimó el tamaño de muestra por el método de log-verosimilitud, al simular no se distingue el origen del cero. Además, el valor estimado del parámetro de sobredispersión k es congruente con los valores obtenidos en campo.

Para Poisson cero inflado se obtuvieron valores de k muy pequeños, debido a la cercanía de los valores de media y varianza, mientras que para las simulaciones de la binomial negativa cero inflada, el parámetro de sobredispersión fue mayor que cero, lo que indica sobredispersión, similar a lo reportado por Zou *et al*. (2021); Haslett *et al*. (2022).

Conclusiones

Las distribuciones de probabilidad Poisson cero inflado y binomial negativa cero inflada permiten modelar poblaciones de organismos plaga con densidades bajas y exceso de ceros. Los parámetros obtenidos por el método de los momentos distinguen el origen del cero y estiman tamaños de muestra óptimos equivalentes o menores a los estimados por log-verosimilitud, la cual no distingue el origen del cero. Una población Poisson cero inflada puede modelarse también con una distribución binomial negativa, siempre que el componente no estructural presente sobredispersión.

La estimación del tamaño de muestra óptimo en poblaciones de plagas con exceso de ceros se puede realizar de manera equivalente con la ecuación del coeficiente de variación (CV) y la ecuación de la proporción de la media ($D_{\bar{x}}$). Por otro lado, la estimación del tamaño de muestra óptimo con la ecuación de la mitad del intervalo de confianza (h) depende del valor del parámetro de sobredispersión (k), ya que no tiene un valor fijo que permita establecer una equivalencia.

Bibliografía

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.

DOI: https://doi.org/10.29312/remexca.v15i1.3618

(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[a] 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 .

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[rd] 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

DOI: https://doi.org/10.29312/remexca.v15i1.3618

and herbicides. China. Artificial Intelligence in Agric. 4(1):58-73. https://doi.org/10.1016/j.aiia.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.

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.

Propuesta para obtener el tamaño de muestra óptimo de plagas con exceso de ceros

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: Artículo

Palabras claves:

Palabras claves:

binomial negativa cero inflado
muestreo
Poisson cero inflado

Counts

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