MJFAS
Malaysian Journal of Fundamental and Applied Sciences

KERANA TUHAN UNTUK MANUSIA

1972-2022

RESEARCH ARTICLE

On Variance Estimation for the Population Size Estimator under One-Inflated Positive Poisson Distribution

Razik Ridzuan Mohd Tajuddin*, Noriszura Ismail, Kamarulzaman Ibrahim

Department of Mathematical Sciences, Universiti Kebangsaan Malaysia 43600 UKM Bangi, Selangor, Malaysia

Abstract Let $\hat{N}_{OIPP}$ be the Horvitz-Thompson estimator for the population size with one-inflated positive Poisson distribution as the underlying distribution. We estimate the variance of this estimator using conditional expectation technique and provide some descriptions on the variance and its associated confidence interval based on simulation study and real data applications.

Keywords: Horvitz-Thompson estimator, capture-recapture, count inflation, zero-truncated Poisson

Introduction

One-inflated positive Poisson distribution [1], denoted as OIPP, has been recently introduced to cater for positive count data with large number of ones. The probability mass function for a random variable Y which follows the OIPP distribution is given as

$$\Pr(Y=y|\omega,\lambda)=\begin{cases}\omega+(1-\omega)\dfrac{\lambda}{\exp(\lambda)-1}; y=1\\ (1-\omega)\dfrac{\lambda^y}{y!\,[\exp(\lambda)-1]}; y\geq 2,\end{cases}$$

where $0<\omega<1$ and $\lambda>0$. Parameter ω refers to the one-inflation parameter whereas parameter λ refers to the rate parameter.

Godwin and Böhning [1] justify the use of the OIPP distribution because in a capture-recapture setting, if subjects gain sufficient information from the initial capture that gives them the desire and capability to avoid being recaptured, the resulting capture-recapture data will have excess of ones. The author proceeds to develop an estimator for the population size which is in the form of Horvitz-Thompson estimator [2], by considering the OIPP distribution as the underlying distribution of the population.

In a capture-recapture framework, population size, N can be written as the sum of observed members, n and unobserved members, n_0 of the population. The unobserved members, n_0 of the population can also be written as Np_0, which is the product of the population size and the proportion of unobserved members, p_0. The population size, N can be estimated using $\hat{N}=n/(1-p_0)$ with variance $Var(\hat{N})=Np_0(1-p_0)/(1-p_0)^2$ [3]. However, since $E(n)=N(1-p_0)$, the variance can be estimated using $\widehat{Var}(\hat{N})=np_0/(1-p_0)^2$ [3]. Böhning [3] highlighted that p_0 are often not available and must be estimated from modelling, which further adds to the variation in the estimation of the population size. Hence, using $\widehat{Var}(\hat{N})$ given above is insufficient.

Therefore, when dealing with unknown p_0, the unknown p_0 needs to be estimated first in order to estimate N. When considering the OIPP distribution as the distribution of the population, the unknown p_0 can be

***For correspondence:**
razikridzuan@siswa.ukm.edu.my

Received: 20 August 2021
Accepted: 19 April 2022

estimated using Poisson distribution with parameter $\hat{\lambda}$, which is estimated from fitting the OIPP distribution to the data using maximum likelihood estimation, resulting in $\hat{p}_0 = \exp(-\hat{\lambda})$. Assuming that the population follows the OIPP distribution, the population size estimator can be written as $\hat{N}_{OIPP} = n/[1-\exp(-\hat{\lambda})]$ [1]. The $\hat{\lambda}$ is the maximum likelihood estimator for the parameter λ of the OIPP distribution, which can be obtained by solving [4]

$$(\hat{\lambda} - A)\exp(\hat{\lambda}) + (A-1)\hat{\lambda} + A = 0, \quad (1)$$

where $A = (nm_1 - n_1)/(n - n_1)$, $n = \sum_{x=1}^{\infty} n_x$, n_x is the frequency of x-valued data, m_1 is the sample mean. Tajuddin et al. [4] has found out that the $\hat{\lambda}$ is asymptotically unbiased, consistent and efficient compared to other estimators obtained from the ordinary least square approach, the method of moment and the ratio of probability. Similarly, the maximum likelihood estimator of ω can be written as [4]

$$\hat{\omega} = \frac{n_1[\exp(\hat{\lambda}) - 1] - n\hat{\lambda}}{n[\exp(\hat{\lambda}) - 1 - \hat{\lambda}]}.$$

Note that, if the population size follows a zero-truncated Poisson distribution, the same population size estimator can be used, however the $\hat{\lambda}$ will be the maximum likelihood estimator of λ using the zero-truncated Poisson distribution, which can be obtained by solving [5]

$$\hat{\lambda}[\exp(\hat{\lambda}) - \bar{x}] + \bar{x} = 0,$$

where $\bar{x}$ is the sample mean.

1. Estimation on the variance of Horvitz-Thompson estimator under the OIPP distribution

A simple and general formula in obtaining the variance of the population size estimator has been proposed by Böhning [3], which makes use of the idea that the nonzero counts, n and the parameter of the distribution, $\hat{\lambda}$ are two random variables, where n follows a binomial distribution with parameter N and $g(\lambda)$, and $g(\lambda) = 1 - f(0|\lambda)$. This idea has been applied for different population size estimator [6-9]. Generally, the variance can be written as

$$Var(\hat{N}) = Var_{\hat{\lambda},n}\left[\frac{n}{g(\hat{\lambda})}\right] = Var_n\left\{E_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right]\right\} + E_n\left\{Var_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right]\right\}. \quad (2)$$

From equation (2), it is clear that the variation in the estimated population size come from two sources [3]. The first term emerges as the binomial random variation involved in sampling the n data from the population with probability $g(\lambda)$ and size N. The second term arises as the variation in estimating λ based on n data. Consider the first term on the RHS of equation (2) and by using delta method,

$$E_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right] \approx \frac{n}{g(\lambda)}.$$

The δ-method is used by approximating the expected value of the transformed variable with the transformation of the expected value [10]. Since $n \sim Binomial(N, g(\lambda))$, we obtain

$$Var_n\left\{E_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right]\right\} \approx Var_n\left[\frac{n}{g(\lambda)}\right] = \frac{Ng(\lambda)[1-g(\lambda)]}{g(\lambda)^2}.$$

This term can be estimated by substituting λ with $\hat{\lambda}$ and $Ng(\lambda)$ with n, resulting in

$$\widehat{Var}_n\left\{E_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right]\right\} = \frac{n[1-g(\hat{\lambda})]}{g(\hat{\lambda})^2}.$$

Note that for Poisson distribution, $g(\lambda) = 1 - f(0|\lambda) = 1 - \exp(-\lambda)$. Let

$$H(\lambda) = \frac{1-g(\lambda)}{g(\lambda)^2} = \frac{g'(\lambda)}{g(\lambda)^2} = \frac{\exp(-\lambda)}{[1-\exp(-\lambda)]^2}.$$

We obtain,

$$\widehat{Var}_n\left\{E_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right]\right\} = nH(\hat{\lambda}). \tag{3}$$

Consider the second term in equation (2) and assume that

$$E_n\left\{Var_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right]\right\} \approx Var_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right].$$

Note that by using delta method,

$$Var_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right] = n^2 Var_{\hat{\lambda}|n}\left[\frac{1}{g(\hat{\lambda})}\right] \approx n^2\left[\frac{g'(\lambda)}{g(\lambda)^2}\right]^2 Var_{\hat{\lambda}|n}(\hat{\lambda}) = [nH(\lambda)]^2 Var_{\hat{\lambda}|n}(\hat{\lambda}),$$

where $Var_{\hat{\lambda}|n}(\hat{\lambda}) = \lambda/n$ is the variance of the maximum likelihood estimator for the parameter λ of Poisson distribution. With substitution and some algebraic manipulation, the second term in equation (3) can be written as

$$E_n\left\{Var_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right]\right\} \approx n\lambda[H(\lambda)]^2,$$

and by substituting λ with $\hat{\lambda}$, the expected value of this term can be estimated as

$$\hat{E}_n\left\{Var_{\hat{\lambda}|n}\left[\frac{n}{g(\hat{\lambda})}\right]\right\} = n\hat{\lambda}\left[H(\hat{\lambda})\right]^2. \tag{4}$$

Therefore, the estimated variance of the population size estimator after some simplification and by combining equations (3) and (4), can be written as

$$\widehat{Var}(\hat{N}_{OIPP}) = nH(\hat{\lambda})\left[1 + \hat{\lambda}H(\hat{\lambda})\right]. \tag{5}$$

Under symmetric normal distribution assumption, the confidence interval for $\hat{N}_{OIPP}$ can be approximated using

$$\hat{N}_{OIPP} \pm z_{0.025} SE(\hat{N}_{OIPP}), \tag{6}$$

where $z_{0.025} = 1.96$ and $SE(\hat{N}_{OIPP}) = \sqrt{V\hat{a}r(\hat{N}_{OIPP})} = \sqrt{nH(\hat{\lambda})\{1 + \hat{\lambda}H(\hat{\lambda})\}}$. The variance in equation (5) and the confidence interval in equation (5) are estimated by substituting $\hat{\lambda}$, obtained from fitting data to the OIPP distribution. Note that the inflation parameter ω is missing entirely from the formulae in equations (5-6). Also, note that the variance and confidence interval of the population size estimated using a zero-truncated Poisson distribution have similar formulae as in equations (5-6). However, the $\hat{\lambda}$ value is obtained by fitting the zero- truncated Poisson distribution to the data.

2. Some results on the variance and the confidence interval

A simulation study investigated by Godwin and Böhning [1] showed that the estimator, $\hat{N}_{OIPP}$ produces smaller percentage of bias values and the percentage of root mean squared error values when λ is large ($\lambda = 1$ vs $\lambda = 2$). However, the increment in the value of ω does not change the percentage of bias values and the percentage of root mean squared error values significantly (see Table 1 [1]). Since, $V\hat{a}r(\hat{N}_{OIPP})$ in equation (5) and its associated 95% confidence interval in equation (6) only depend on the parameter $\hat{\lambda}$, the effect of the parameter ω to is almost close to negligible. Therefore, when the population data are generated from the OIPP distribution, we hypothesize that the estimated variance for the population size estimator is smaller and the 95% confidence interval for the population size estimator is narrower as λ increases with no significant influence from parameter ω.

It is clear that for a given value of λ, as n increases, the variance estimation in equation (5) increases. However, it is unclear the effect of λ on the variance. The question on the indirect effect of the parameter ω on the variance still lingers even though the parameter ω is absent from the variance estimation formula. Therefore, a simulation study is conducted to investigate the effects of both parameters towards the variance estimation and subsequently, its effect on the 95% confidence interval. The hypothesis for the simulation study is that when the population data are generated from the OIPP distribution, the estimated variance (confidence interval) for the population size estimator is smaller (narrower) as λ increases with no significant influence from ω. Using the same simulated data, the accuracy of the standard error and the confidence interval given in (6) can be investigated. Note that, any changes in the standard error directly reflects on the variance itself. Therefore, for simplification, the standard error is considered in investigating the hypothesis. The algorithm for the simulation study is given below.

- Step 1: Simulate $N = 5000$ data from the OIPP distribution with parameters $\lambda = 1.0$ and $\omega = 0.2$.
- Step 2: Estimate $\hat{\lambda}$ using maximum likelihood estimation given in equation (1).
- Step 3: Obtain the $\hat{N}_{OIPP}$ and $SE(\hat{N}_{OIPP})$ from the data.
- Step 3: Repeat steps 1 and 2 for 5000 times and obtain 5000 values for $\hat{N}_{OIPP}$ and $SE(\hat{N}_{OIPP})$.
- Step 4: Obtain the average values for $\hat{N}_{OIPP}$ and $SE(\hat{N}_{OIPP})$ and the standard deviation, $SD(\hat{N}_{OIPP})$ for 5000 values of $\hat{N}_{OIPP}$. The $SD(\hat{N}_{OIPP})$ will provide the standard error based on the simulated data.
- Step 5: Tabulate the results and repeat the algorithms by varying parameters values ($\lambda = 2.0, 3.0, 4.0,\ 5.0$ and $\omega = 0.4, 0.6$).
- Step 6: Compare the confidence interval of the estimated population size using $SE(\hat{N}_{OIPP})$ and $SD(\hat{N}_{OIPP})$ to investigate the accuracy of the formula given in (6).

The results of the simulation study are given in Table 1 and Table 2. Table 1 shows the percentage of relative absolute errors in calculating the variability in 5000 estimates of the population size. The formula for the percentage of relative absolute errors is given as

$$e = \frac{\left|E\left[SE(\hat{N}_{OIPP})\right] - SD(\hat{N}_{OIPP})\right|}{SD(\hat{N}_{OIPP})} \times 100,$$

where $E\left[SE(\hat{N}_{OIPP})\right]$ is the average standard error values obtain from equation (6) and $SD(\hat{N}_{OIPP})$ is the standard deviation based on 5000 estimates of the population size.

From Table 1, for any given values of λ and ω, the $E(\hat{N}_{OIPP})$ from the 5000 estimates of the population size is close to the true population size, supporting the findings of Godwin and Böhning [1]. For a given ω, as λ increases, both the values of both $E\left[SE(\hat{N}_{OIPP})\right]$ and $SD(\hat{N}_{OIPP})$ become close which can also be seen from the decreasing percentage values of e. This shows that the proposed standard error formula and consequently the proposed variance, are accurate in estimating the variability in the simulated population size estimates. For a given λ, as ω increases, the $SD(\hat{N}_{OIPP})$ increases and this is bound to happen when dealing with simulated data, where some variability cannot be avoided. However, for a larger value of λ ($\lambda \geq 3.0$),the $SD(\hat{N}_{OIPP})$ values are close. It is expected that the $E\left[SE(\hat{N}_{OIPP})\right]$ has a constant value because the ω is absent in the $SE(\hat{N}_{OIPP})$ formula in (6). Furthermore, the $E(\hat{N}_{OIPP})$ values do not change significantly when ω increases.

Table 2 shows the 95% confidence interval based on the $SD(\hat{N}_{OIPP})$ and the $E\left[SE(\hat{N}_{OIPP})\right]$, From Table 2, as λ increases, the 95% confidence interval using both standard deviation and standard error formulae become narrower. However, for any value of λ, as ω increases, the confidence interval using standard error in equation (6) does not change significantly. The changes can be noticed for the confidence interval using standard deviation of the simulated data. Therefore, from Table 1 and Table 2, the hypothesis has been addressed and it can be concluded that when the population data are generated from the OIPP distribution, the estimated variance for the population size estimator is smaller as λ increases. Consequently, the 95% confidence interval for the population size estimator becomes narrower as λ increases. However, the parameter ω plays no significant role in determining the variance and the 95% confidence interval for the population size.

3. Some applications to real data

To illustrate the variance estimation and the confidence interval for population size estimator, we use classic examples of number of eggs cells and gall cells observed on flower-heads, initially studied by Finney and Varley [11]. The resulting estimated population size for these datasets explain the observed eggs and gall cells by Finney and Varley [11] and the unobserved eggs and gall cells on flower-heads.

Table 3 provides four datasets studied. The datasets refer to the number of eggs and gall cells on flower-heads in 1935 and 1936.

Table 1 Results of the simulation study on the estimated population size and percentage of relative absolute errors in calculating the variability in 5000 estimates of the population size from the simulated data and the conditional expectation technique in equation (5) when both parameters λ and ω vary with true population size, $N = 5000$.

ω	λ	$E(\hat{N}_{OIPP})$	$SD(\hat{N}_{OIPP})$	$E[SE(\hat{N}_{OIPP})]$	e (%)
0.2	1.0	5006.17	144.76	75.00	48.2
	2.0	5001.00	41.24	32.69	20.7
	3.0	5000.11	18.76	17.48	6.8
	4.0	5000.18	10.12	10.02	1.0
	5.0	4999.99	6.02	5.92	1.7
0.4	1.0	5009.12	164.43	75.10	54.3
	2.0	5000.06	44.11	32.67	25.9
	3.0	4999.81	19.90	17.47	12.2
	4.0	5000.23	10.47	10.02	4.3
	5.0	4999.98	6.03	5.92	1.8
0.6	1.0	5013.97	201.89	75.26	62.7
	2.0	5001.61	50.48	32.69	35.2
	3.0	5000.57	21.09	17.48	17.1
	4.0	5000.24	10.47	10.03	4.2
	5.0	4999.95	6.05	5.93	2.0

Table 2 Results of the simulation study on the estimated population size and the 95% confidence interval for the estimated population size from the simulated data and the conditional expectation technique in equation (6) when both parameters λ and ω vary with true population size, $N = 5000$.

ω	λ	$E(\hat{N}_{OIPP})$	95% Confidence interval based on $SD(\hat{N}_{OIPP})$	95% Confidence interval based on $E[SE(\hat{N}_{OIPP})]$
0.2	1.0	5006.17	(4722.44, 5289.90)	(4859.17, 5153.17)
	2.0	5001.00	(4920.17, 5081.83)	(4936.93, 5065.07)
	3.0	5000.11	(4963.34, 5036.88)	(4965.85, 5034.37)
	4.0	5000.18	(4980.34, 5020.02)	(4980.54, 5019.82)
	5.0	4999.99	(4988.19, 5011.78)	(4988.39, 5011.59)
0.4	1.0	5009.12	(4686.84, 5331.40)	(4861.92, 5156.32)
	2.0	5000.06	(4913.60, 5086.52)	(4936.03, 5064.09)
	3.0	4999.81	(4960.81, 5038.81)	(4965.57, 5034.05)
	4.0	5000.23	(4979.71, 5020.75)	(4980.59, 5019.87)
	5.0	4999.98	(4988.16, 5011.80)	(4988.38, 5011.58)
0.6	1.0	5013.97	(4618.27, 5409.67)	(4866.46, 5161.48)
	2.0	5001.61	(4902.67, 5100.55)	(4937.54, 5065.68)
	3.0	5000.57	(4959.23, 5041.91)	(4966.31, 5034.83)
	4.0	5000.24	(4979.72, 5020.76)	(4980.58, 5019.90)

5.0	4999.95	(4988.09, 5011.81)	(4988.33, 5011.57)

Table 3 Datasets on the number of eggs and gall cells on flower-heads in 1935 and 1936.

Data	Year	Counts									Total
		1	2	3	4	5	6	7	8	9+	
Number of eggs	1935	29	38	36	23	8	5	5	2	2	148
	1936	22	18	18	11	9	6	3	0	1	88
Number of gall cells	1935	287	272	196	79	29	20	2	0	1	886
	1936	90	96	57	26	10	4	5	0	1	289

To show the usability of the variance estimation and the confidence interval for population size estimator for human population data, the Dutch illegal immigrants, which was initially studied by Van der Heijden et al. [12] and re-investigated in the context of one-inflation by Godwin and Böhning [1], is reconsidered in this study. The data on Dutch illegal immigrants refers to the frequency of apprehensions of illegal immigrants, who were unable to be effectively expelled, is given in Table 4.

Table 4 Dataset on the Dutch illegal immigrants

Counts						Total
1	2	3	4	5	6	
1645	183	388	13	1	1	1881

Using datasets in Table 3 and Table 4, the population size of the eggs and the gall cells as well as the Dutch illegal immigrants with its associated confidence interval are estimated. The chi-squared goodness-of-fit test is used to show that the OIPP distribution adequately fits the data. Table 5 summarizes the model fittings of the five datasets, the estimated population size with its 95% confidence interval.

Table 5 Model fittings of the five datasets to the OIPP distribution, variance estimation and 95% confidence interval (95% CI) for population size estimator, $\hat{N}_{OIPP}$.

Data	Year	$\hat{N}_{OIPP}$ (SE)	95% CI	$\hat{\lambda}$	$\hat{\omega}$	χ^2 (p-value)
Number of eggs [11]	1935	156 (3.21)	(150,163)	2.934	0.037	4.139 (0.247)
	1936	92 (2.10)	(88,96)	3.212	0.133	1.202 (0.753)
Number of gall cells [11]	1935	1028 (15.04)	(999,1058)	1.977	0.009	6.903 (0.075)
	1936	332 (8.14)	(316,348)	2.048	0.012	7.312 (0.063)
Dutch illegal immigrants [12]	-	2501 (79.03)	(2346,2656)	0.793	0.636	3.310 (0.069)

Based on Table 5, it can be observed that the OIPP distribution adequately fits all datasets. For the number of eggs and galls cells, the estimated proportion of excess ones ranges from 0.9% to 13.3% for the number of eggs and galls cells. For an example, the total population size for the eggs in 1935 is 156, which means that a total of 8 eggs are unobserved, whereas the number of unobserved eggs in 1936 is 4. For gall cells, the number of unobserved gall cells in 1935 is 142 whereas in 1935, the unobserved gall cells are 43. Finney and Varley [9] did mention that the unobserved eggs and gall cells are not reported because the counting process is destructive. Moreover, some eggs fail to become gall cells [11]. Using the $\hat{\lambda}$ reported by Finney and Varley [11], for these four datasets, we can obtain the estimated population size, the standard error and the 95% confidence interval for the data. Taking the number of eggs in 1936 as an example, the value of $\hat{\lambda}$ reported by Finney and Varley [11] is 2.860 and this value is significantly different from the one reported in Table 5. Using the formula of the population size estimator and equation 6, the estimated population size and the 95% confidence interval due to Finney and Varley [11] are 93 and (88,98) with a slightly bigger standard error of 2.59. This shows that the proposed variance estimate based on the Horvitz-Thompson estimator under the OIPP distribution is useful in obtaining the confidence interval for the estimated population size.

For the Dutch illegal immigrants, the estimated proportion of excess ones is 63.6% with the estimated population size of 2501, meaning that there are 856 illegal immigrants who did not get caught and effectively expelled yet. This often happens because either the apprehended illegal immigrants refuse to mention their nationality or their home country refuses to accept them [12]. Based on the null model which is the truncated Poisson distribution, used by Van der Heijden et al. [12], the estimated population size is 7080 with the 95% confidence interval of (6363, 7797), which is way higher than our estimate of 2501 and its associated 95% confidence interval of (2346, 2656). However, it is trivial to see that the standard error based on the confidence interval due to Van der Heijden et al. [12] is about 366, whereas the standard error from the proposed formula is only about 79, which further shows how reliable and useful the proposed variance estimation is.

Concluding remarks

Assuming that the population follows the one-inflated positive Poisson distribution (OIPP), the variance estimation for the population size estimator in the form of Horvitz-Thompson estimator is developed. Using the variance estimation, the 95% confidence interval is given under the symmetric normal approximation. The changes in the variance and the confidence interval relies heavily on the parameter λ.

The error in the difference in the simulated standard deviation and the theoretical standard error approach to zero as λ increases. The increment in of ω on the other hand does affect the standard deviation of the simulated data but this is due to the unavoidable variation happening during simulating the data. However, any changes in ω does not affect the standard error as seen from equation (6) and further supported by the simulation results in Table 1. Similarly, as λ increases, the standard error decreases, which subsequently makes the confidence interval narrower. Similar observation can be made when the simulated data is used to obtain the standard deviation and subsequently, the confidence interval. However, for a given value of λ, any changes in ω only affects the confidence interval due to the standard deviation of the simulated data and not due to the standard error in equation (6). Applications to real datasets also show the usability of the proposed variance for population size estimator based on the OIPP distribution.

Therefore, the simulation study supports the hypothesis stated that when the population data are generated from the OIPP distribution, the estimated variance (confidence interval) for the population size estimator is smaller (narrower) as λ increases with no significant influence from ω. It is found that the proposed variance and the standard error can accurately estimate the variability in the estimates of population size, and subsequently provide an accurate 95% confidence interval for a given value of λ and for any value of ω.

Since the formulae on the estimated variance and the 95% confidence interval rely on the $\hat{\lambda}$, the only limitation to the formulae is the $\hat{\lambda}$ itself. The better the estimator of λ, the better the estimated variance and the 95% confidence interval will be. The formulae on the estimated variance and the 95% confidence interval can be further investigated in the context of not upweighting the excess ones in the data. By not upweighting the excess ones, a new estimator [13] can be obtained and both variances from this study and [13] can be compared.

Conflict of Interest Statements

The authors have no conflict of interests.

References

[1] R. Godwin and D. Böhning, "Estimation of the population size by using the one-inflated positive Poisson model," Journal of the Royal Statistical Society, Series C (Applied Statistics), vol. 66, no. 2, pp. 425-448, 2017.
[2] D. G. Horvitz and D. J. Thompson, "A generalization of sampling without replacement from a finite universe," Journal of the American Statistical Association, vol. 47, no. 260, pp. 663-685, 1952.
[3] D. Böhning, "A simple variance formula for population size estimators by conditioning," Statistical Methodology, vol. 5, no. 5, pp. 410-423, 2008.
[4] R. R. M. Tajuddin, N. Ismail and K. Ibrahim, "Comparison of estimation methods for one-inflated positive Poisson distribution," Journal of Taibah University for Science, vol. 15, no. 1, pp. 869-881, 2021.
[5] F. N. David and N. L. Johnson, "The truncated Poisson," Biometrics, vol. 8, no. 4, pp. 275-285, 1952.
[6] O. Anan, D. Böhning and A. Maruotti, "Population size estimation and heterogeneity in capture-recapture data: a linear regression estimator based on the Conway-Maxwell-Poisson distribution," Statistical Methods & Applications, vol. 26, no. 1, pp. 49-79, 2016.
[7] O. Anan, D. Böhning and A. Maruotti, "On the Turing estimator in capture-recapture count data under the geometric distribution," Metrika, vol. 82, no. 2, pp. 149-172, 2019.
[8] D. Böhning, P. Kaskasamkul and P. G. M. Van der Heijden, "A modification of Chao's lower bound estimator in the case of one-inflation," Metrika, vol. 82, no. 3, pp. 361-384, 2019.
[9] R. R. M. Tajuddin, N. Ismail and K. Ibrahim, "Estimating population size of criminals: a new Horvitz–Thompson estimator under one-inflated positive Poisson-Lindley model," Crime and Delinquency, 2021.
[10] G. Casella and R. L. Berger, "Statistical inference (2nd edition)", Duxbury Press, Belmont, 2002.
[11] D. J. Finney and G. C. Varley, "An example of the truncated Poisson distribution," Biometrics, vol. 11, no. 3, pp. 387-394, 1955.
[12] P. G. M. Van der Heijden, R. Bustami, M. J. Cruyff, G. Engbersen and H. C. Van Houwelingen, "Point and interval estimation of the population size using the truncated Poisson regression model," Statistical Modelling, vol. 3, no. 4, 305-322, 2003.
[13] D. Böhning and P. G. M. Van der Heijden, "The identity of the zero-truncated, one-inflated likelihood and the zero-one-truncated likelihood for general count densities with an application to drink-driving in Britain," The Annals of Applied Statistics, vol. 13, no. 2, 1198-1211, 2019.