Estimating Under-Reporting of Highway Accidents using Capture-Recapture Method: A Case Study of Bureau of Highway I (Chiang Mai) THAILAND

Suwannee Kannithet [1] and Krisana Lanumteang [2*]

[1] Division of Statistics, Maejo University, ChiangMai, 50290, THAILAND, suwannee300736@gmail.com

[2*] Division of Statistics, Maejo University, ChiangMai, 50290, THAILAND, k.lanumteang@mju.ac.th

(* corresponding author)

Abstract

Incompleteness of registration is one of the key performance indicators showing the quality of management information system. This is an initiative and viable research topic for researchers. We demonstrate the use of capture-recapture techniques in estimating the size of underreporting highway accidents. Capture-recapture methods have been widely used to estimate the size of an elusive target population. These methods are basically required for the frequency counts of identifying unique units. The repeating recorded accidents on each route number and control section in the year 2015 are our variables of interests. The data were collected from the highway accident information management system (HAIMS). In this paper, we only focus on the highway undertaken by the Bureau of Highway I (ChiangMai), THAILAND connecting four provinces (ChiangMai, Lamphun, Lampang and Maehongson). A variety of estimators based on both homogeneous and heterogeneous Poisson models are considered, these include: maximum likelihood, Chao's, Zelterman's and Lanumteang-Böhning's estimator. They yielded reasonable and similar results. The proportion of underreporting routes and control sections were approximately 2.00%–12.50% and 1.22%–17.35%, respectively.

Keywords: capture-recapture, highway accident, population size estimation

1. Introduction

Capture-recapture experiments have been traditionally applied in ecological sciences. These are used to estimate the animal abundance, the size of wildlife animals and the demographic factors that affecting population size, for example, birth, mortality, immigration and emigration rates. The classical capture-recapture model is single-mark experiment whereas the identifying system and the counting of listed cases from multiple sources are major concerned in nowadays. These methods are tended to be widely applied in a variety of other fields such as estimating the size of an elusive human population in the life and medical sciences, and the social sciences [1-2]. For example, the number of illegal activities are investigated such as the number of drug users, the number of violators of a law or the number of illegal immigrants, see [3-4]. In addition, there is a great deal of interest in estimating the number of outbreaks of a disease and determining the completeness of a disease registry in public health science [5].

In this study, we examine the use of capture-recapture methods to estimate the number of underreporting cases in registry system. We addressed the reporting of road accidents as our case study. A recent study by the University of Michigan's Transportation Research Institute confirmed that Thailand ranked number two in the university's study of road fatalities in the world, with 44 road deaths per 100,000 people. Thailand has made headlines on several occasions in recent years due to its appalling road safety record. The government announced in 2011 that it sought to cut the number of deaths from vehicle accidents by half by 2020, a commitment that is part of its decade-long campaign to improve traffic safety [6]. Thus, road safety in Thailand remains a serious issue. Although there are many organizations taking responsible for accident data collection and analysis, it can be sensible to assume that there are some missing in recording or reporting cases. This leads to the incompleteness of identifying cases in the information management system and it remains one of active research.

2. Research Methodology

2.1 Data Set

We used the secondary data from the Highway Accident Information Management System (HAIMS), http://haims.doh.go.th/. HAIMS is an information system that collects accident reports from highway offices and highway districts around Thailand and it is used to generate reports and analyze accident information throughout the country. In this paper, we only focus on the reports of accidents on the highway undertaken by the Bureau of Highway I (ChiangMai), Department of Highway (DOH), in the year 2015. The Bureau of Highway I (ChiangMai) consists of four provinces;

ChiangMai, Lamphun, Lampang and Maehongson. The repeating recorded accidents on each highway number (route number) and each control section are our variables of interests. In this region, there are 99 unique routes, which divided into 151 control sections (part of each highway number). And, there were 538 reporting cases of accidents in 2015. These reported events occurred on only 49 routes and 81 control sections. The frequency count of reporting cases on unique routes and control sections are showed in Table1 and Table2, respectively whereas raw data are gave in Appendix.

Table 1: Count distribution of repeated reporting accident on distinct route number

j	1	2	3	4	5	6	7	8	9	10+	n
f_j	12	13	4	2	1	4	3	0	1	9	49

Note : f_j denote the frequencies of routes reported an accident j times. The third highest repeating were $f_{134} = 1, f_{90} = 1$ and $f_{50} = 2$, respectively.

Table 2: Count distribution of repeated reporting accident on distinct control section

j	1	2	3	4	5	6	7	8	9	10+	n
f_j	18	16	12	4	2	4	4	2	3	16	81

Note : f_j denote the frequencies of control sections reported an accident j times. The third highest repeating were $f_{59} = 1, f_{30} = 1$ and $f_{29} = 2$, respectively.

These repeated indentifying is an essential structure of capture-recapture data. Here, we named the data set in Table1 and Table2 as *"Data Set I"* and *"Data Set II"*, respectively. The review of general background of capture-recapture will be presented in the next subsection.

2.2 Formulating Problem

According to a capture-recapture experiment, we have that $f_1, f_2, f_3, ..., f_m$ denote the frequencies of units observed $1,2,...,m$ times during the periods of study, and $n = \sum_{j=1}^{m} f_j$ is the total number of observed distinct units. Let f_0 be the number of unobserved or hidden cases. Hence, the size of target population (N) can be written as:

$$N = f_0 + f_1 + f_2 + ... + f_m = f_0 + n . \tag{1}$$

If f_0 can be estimated, N can be easily obtained. In addition, suppose that $1 - p_0$ is the probability that the elements are observed as the sample of size n, where p_0 is the probability of the unobserved elements. Therefore, $N = Np_0 + N(1 - p_0)$. As $N(1 - p_0)$ is the expected number of observed cases which can be estimated by n, this leads to the simple equation to estimate the population size N. Consequently, this equation can be solved for estimating N of the form:

$$\hat{N} = \frac{n}{1 - p_0}. \tag{2}$$

However, f_0 and p_0 are typically unknown and the estimator in (1) and (2) would require an estimator of f_0 and p_0 . Due to the fact that capture-recapture methods deal with count data, the Poisson model is sensibly chosen for the probability density function of the model. In this paper, four methodologies of estimating N, f_0 and p_0 based on homogeneous and heterogeneous Poisson models are considered.

2.2.1 Maximum Likelihood Estimator ($\hat{N}_{MLE}$)

The maximum likelihood method is one of the most popular techniques used to derive estimators. For capture-recapture experiments, it is required to look at zero truncation since zeroes (individuals which are not identified) have not been observed in the identifying systems. Let Y_i be the number of times that individual i^{th} was identified, where $i = 1,2,3,...,n$. The zero-truncated Poisson distribution is an elementary model for the probability function of Y. In addition, as $f_1, f_2, f_3, ..., f_m$ denote the frequencies of units observed $1,2,3,...,m$ times and $\sum_{j=1}^{m} f_j = n$, the likelihood function for this zero-truncated Poisson distribution is:

$$L(\lambda) = \prod_{j=1}^{m} \left(\frac{\exp(-\lambda)\lambda^j}{j!(1 - \exp(-\lambda))} \right)^{f_j}, j = 1,2,3,...,m . \tag{3}$$

Differentiating the log-likelihood function of (3) with respect to λ and setting the results equal to zero gives the MLE satisfying the relation:

$$\bar{y} = \frac{\hat{\lambda}_{MLE}}{1 - \exp(-\hat{\lambda}_{MLE})}, \text{ where } \bar{y} = \frac{1}{n}\sum_{j=1}^{m} jf_j . \tag{4}$$

Unfortunately, the estimator in (4) is not the closed form, therefore iterative method is typically required for solving $\hat{\lambda}_{MLE}$ via EM algorithm. Here, the initial value of $\hat{\lambda}_{MLE}$ is simply chosen as a sample mean. As a result of replacing $\hat{\lambda}_{MLE}$ in (2), the estimator of population size from MLE method is readily provided as:

$$\hat{N}_{MLE} = \frac{n}{1 - \exp(-\hat{\lambda}_{MLE})} . \tag{6}$$

A variance of (6) can simply be estimated as:

$$\hat{V}(\hat{N}_{MLE}) = \frac{\hat{N}_{MLE}}{(\exp(\frac{\sum_{j=1}^{m} jf_j}{\hat{N}_{MLE}}) - \frac{\sum_{j=1}^{m} jf_j}{\hat{N}_{MLE}} - 1)}, \tag{7}$$

see [7].

2.2.2 Zelterman's Estimator ($\hat{N}_{Zel}$)

Zelterman [8] proposed a family of robust estimators of the parameter $\psi = \exp(-\lambda)$ under the zero-truncated Poisson capture probability. The estimator of [8] can be simply derived as a consequence of the property of the Poisson distribution. Recall the zero-truncated Poisson probability, $f_+(j;\lambda) = \dfrac{\exp(-\lambda)\lambda^j}{j!(1-\exp(-\lambda))}$, where j is the number of times identifying distinct units, $j = 1,2,\ldots$. Then, we have that $\lambda = \dfrac{(j+1)f_+(j+1;\lambda)}{f_+(j;\lambda)}$. We can estimate $f_+(j;\lambda)$ and $f_+(j+1;\lambda)$ by their associated observed frequency counts f_j and f_{j+1}, respectively. Thus, it is deduced that:

$$\hat{\lambda} = \frac{(j+1)f_{j+1}}{f_j} \qquad (8)$$

As a result of (8), the family of estimators of the parameter $\psi = \exp(-\lambda)$ can be found as $Q_j = \exp(-\dfrac{(j+1)f_{j+1}}{f_j})$, $j = 1,2,3,\ldots$. In practice, [8] argued that the most reliable value of j to be chosen are one or two observed frequencies. These will be more similar to those individuals that were not observed. Therefore, taking $j = 1$, (2) in terms of Zelterman's estimator can be finally found as:

$$\hat{N}_{Zel} = \frac{n}{1-\exp(-\frac{2f_2}{f_1})}. \qquad (9)$$

A simple variance formula for (9) can be obtained as:

$$\hat{V}(\hat{N}_{Zel}) = n\left(\frac{\exp(-\frac{2f_2}{f_1})}{(1-\exp(-\frac{2f_2}{f_1}))^2}\right)\{1 + n\left(\frac{\exp(-\frac{2f_2}{f_1})}{(1-\exp(-\frac{2f_2}{f_1}))^2}\right)\left(\frac{2f_2}{f_1}\right)^2\left(\frac{1}{f_1}+\frac{1}{f_2}\right)\} \qquad (10)$$

, see [9].

2.2.3 Chao's Estimator ($\hat{N}_{Chao}$)

Chao [10] proposed an important lower bound for the population size N under the heterogeneous Poisson population. It is more appropriate to incorporate heterogeneity of the identifying probability because the actual target population may consist of a variety of subgroups. She supposed that the capture probability is $p_j = \int_0^\infty \dfrac{\exp(-\lambda)\lambda^j}{j!} f(\lambda)d\lambda$, where $f(\lambda)$ represents an arbitrary distribution of the model parameter λ in the population, the heterogeneity distribution. This estimator is simply derived in the sense of a nonparametric way by using the Cauchy-Schwartz inequality. For any two random variables, X and Y, we have that:

$$(E(XY))^2 \le E(X^2)E(Y^2)$$

or

$$\left(\int_0^\infty u(\lambda)\upsilon(\lambda)f(\lambda)d\lambda\right)^2 \le \left(\int_0^\infty u(\lambda)^2 f(\lambda)d\lambda\right)\left(\int_0^\infty \upsilon(\lambda)^2 f(\lambda)d\lambda\right) \qquad (11)$$

If we let $u(\lambda) = \sqrt{\exp(-\lambda)\lambda^{j-1}}$ and $\upsilon(\lambda) = \sqrt{\exp(-\lambda)\lambda^{j+1}}$, we have that $u(\lambda)\upsilon(\lambda) = \exp(-\lambda)\lambda^j$. Then, the inequality (11) can be written as:

$$\left(\int_0^\infty \exp(-\lambda)\lambda^j f(\lambda)d\lambda\right)^2 \le \left(\int_0^\infty \exp(-\lambda)\lambda^{j-1} f(\lambda)d\lambda\right)\left(\int_0^\infty \exp(-\lambda)\lambda^{j+1} f(\lambda)d\lambda\right)$$

$$\left(\frac{j!}{j!}\int_0^\infty \exp(-\lambda)\lambda^j f(\lambda)d\lambda\right)^2 \le \left(\frac{(j-1)!}{(j-1)!}\int_0^\infty \exp(-\lambda)\lambda^{j-1} f(\lambda)d\lambda\right)$$

$$\times\left(\frac{(j+1)!}{(j+1)!}\int_0^\infty \exp(-\lambda)\lambda^{j+1} f(\lambda)d\lambda\right)$$

$$\frac{jp_j}{p_{j-1}} \le \frac{(j+1)p_{j+1}}{p_j} \qquad (12)$$

Replacing the probability p_j in (12) by their associated observed frequency for $j = 1$ this achieves the lower bound of estimating the number of unobserved cases,

$$\hat{f}_{0Chao} \ge \frac{f_1^2}{2f_2}, \qquad (13)$$

where this inequality will hold on its expected value asymptotically. Finally, adding $\hat{f}_{0Chao}$ to the number of observed cases n leads to the Chao's lower bound estimator as:

$$\hat{N}_{Chao} = n + \frac{f_1^2}{2f_2}. \qquad (14)$$

Chao also provided an approximate variance formula given in (15) using a standard asymptotic, approach which can be written as:

$$\hat{V}(\hat{N}_{Chao}) = \left(\frac{1}{4}\right)^2 \frac{f_1^4}{f_2^3} + \frac{f_1^3}{f_2^2} + \frac{1}{2}\frac{f_1^2}{f_2}. \qquad (15)$$

Alternatively, Böhning [9] derived another form of a variance estimator of (14) by conditioning. This approximate variance is closely associated with (15) as follows:

$$\hat{V}(\hat{N}_{Chao}) = \frac{1}{4}\frac{f_1^4}{f_2^3} + \frac{f_1^3}{f_2^2} + \frac{1}{2}\frac{f_1^2}{f_2} - \frac{1}{4}\frac{f_1^2}{nf_2^2} - \frac{1}{2}\frac{f_1^4}{f_2(2nf_2+f_1^2)}. \qquad (16)$$

2.2.4 Lanumteang-Böhning Estimator ($\hat{N}_{LB}$)

Lanumteang and Böhning [11] used a linear model for ratios of neighbouring frequency counts of observed individuals based on a Poisson-Gamma mixture to provide an estimator of population size. Let the capture probability be the Poisson-Gamma mixture as well as negative binomial, $p_j = \dfrac{\Gamma(k+j)}{\Gamma(j+1)\Gamma(k)}\theta^k(1-\theta)^j$. Then, we achieve $r_j = jp_j/p_{j-1} = (k+j-1)(1-\theta)$. This clearly implies that there is a linear relationship between r_j and j, $r_j = (k-1)(1-\theta) + (1-\theta)j$. A Taylor expansion of log r_j around $(k-1)$ is

$$\log(r_j) = \log(k+j-1) + \log(1-\theta)$$
$$\approx \underbrace{\log(1-\theta) + \log(k-1)}_{\alpha} + \underbrace{(1/(k-1))}_{\beta} j \,. \quad (17)$$

Using a logarithmic transformation will guarantee that the population size estimate is feasible. Now, for $j=2$ or $j=3$ in (17) we get $\log(r_2) = \log(\frac{2f_2}{f_1}) = \alpha + 2\beta$ and $\log(r_3) = \log(\frac{3f_3}{f_2}) = \alpha + 3\beta$. Solving these equations in α and β can easily be achieved as

$$\hat{\alpha} = 3\log(\frac{2f_2}{f_1}) - 2\log(\frac{3f_3}{f_2}) \quad (18)$$

and $$\hat{\beta} = \log(\frac{3f_3}{f_2}) - \log(\frac{2f_2}{f_1}) \,. \quad (19)$$

Then, plugging $\hat{\alpha}$ and $\hat{\beta}$ into (19) and using $j=1$, (17) provides $\log(r_1) = \log(\frac{f_1}{f_0}) = \alpha + \beta$, or

$$\log(\frac{f_1}{f_0}) = 3\log(\frac{2f_2}{f_1}) - 2\log(\frac{3f_3}{f_2}) + \log(\frac{3f_3}{f_2}) - \log(\frac{2f_2}{f_1})$$

$$= 2\log(\frac{2f_2}{f_1}) - \log(\frac{3f_3}{f_2}) \,. \quad (20)$$

Finally, we achieve that

$$\log(f_0) = \log(f_1) - \log(\frac{4f_2^2}{f_1^2}) + \log(\frac{3f_3}{f_2}) = \log(\frac{3f_1^3 f_3}{4f_2^3}) \,. \quad (21)$$

Hence, the estimator for f_0 and N, respectively, is

$$\hat{f}_{0LB} = \frac{3f_1^3 f_3}{4f_2^3} \quad (22)$$

and $$\hat{N}_{LB} = n + \frac{3f_1^3 f_3}{4f_2^3} \,. \quad (23)$$

It is clearly seen that (22) is closely associated with Chao's estimator, $\hat{N}_{LB} = n + \frac{f_1^2}{2f_2}\gamma$, where $\gamma = \frac{3f_1\, f_3}{2f_2^2}$.

A variance of (23) can be constructed from:

$$\hat{V}(\hat{N}_{LB}) = (\frac{9}{4})^2 \frac{f_1^5 f_3^2}{f_2^6} \{\frac{f_1}{f_2} + 1\} + (\frac{3}{4})^2 \frac{f_1^6 f_3}{f_2^6} \{1 - \frac{f_3}{n}\} + \frac{\frac{3n}{4} f_1^3 f_3}{n f_2^3 + \frac{3}{4} f_1^3 f_3} \quad (24)$$

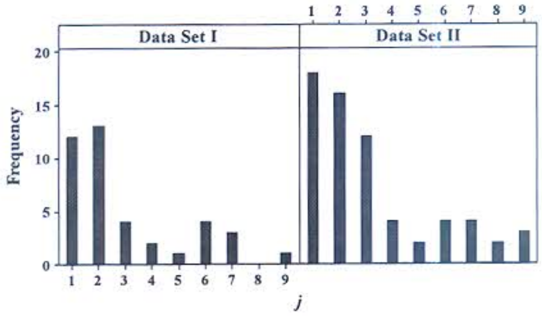
, see [11] for review.

2.2.5 Goodness-of-fit and ratio plot

A method selected as the best from a method set may nonetheless provide a poor description of the data, and this always needs to be checked. As we deal with count data, the chi-square test of goodness of fit was applied to test the distribution of homogeneous Poisson with our data set. On the other hand, the ratio plot was used to detect the heterogeneous Poisson model, see [12] for review.

3. Results

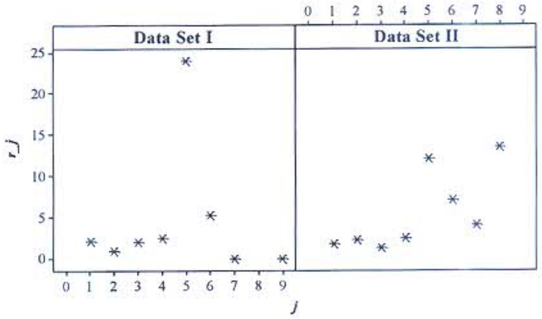
From *Data Set I*, we found that the reporting of accidents were recorded on only 49 distinct routes from 99 routes. This yielded the prevalence rate of an accident on highway undertaken by Bureau of Highway I (Chiangmai) 49.49%. An average of repeated reporting accident on each unique route was 10.98 (*Sd* = 24.02) times of which 12 routes appear only once and 13 routes twice. The highest reporting accident was 134 times, occurring on route number 1 (Phahonyothin Rd). The frequency count of repeated reporting cases is showed in Table 1 and Figure 1. There is evidence to state that Data Set I is not homogeneous Poisson. This data set seems to be over-dispersion, which sample mean and variance are 10.98 and 576.96, respectively. Using chi-square goodness-of-fit test confirms that this data set is not fit homogeneous Poisson, $\chi^2_{df=2} = 168.56$ and p-value < 0.01. Therefore, the Maximum Likelihood and Zelterman's estimator might be not appropriate for this data. On the other hand, the ratio plot $r_j = \frac{(j+1)f_{J+1}}{f_j}$ was applied to detect heterogeneous Poisson. As can be seen from Figure 2, there is a trend of the ratio plot r_j against j. It can be stated that this data might be fitted well heterogeneous Poisson. Thus, Chao's estimator and Lanumteang-Böhning estimator are more sensible for this data set.



Panel variable: Data

(Note: j was truncated above 9)

Figure 1: Frequency count of repeated reporting accidents on distinct units



Panel variable: Data

(Note: j was truncated above 9)

Figure 2: Ratio plot of repeated reporting accidents on distinct units

Table 3 shows summary results of estimating the number of under-reporting cases. Maximum Likelihood yielded the smallest estimator 50 (*Se* = 0.10) routes, only 1 hidden route, whereas Zelterman's estimator gave the highest estimator 56 (*Se* = 6.94). Chao's estimator and Lanumteang-Böhning estimator provided the estimator 55 (*Se* = 4.11) and 52 (*Se* = 3.51) routes. Let $(\hat{f}_0/\hat{N})\times 100\%$ be the proportion of under-reporting accident on highway as well as the proportion of incompleteness reporting. Considered estimators gave this proportion line between 2.00%–12.50%. In contrast, they showed the proportion of completeness registry system 87.50% –98.00%.

Table 3: Estimated number of under-reporting routes

Method	$\hat{N}$	$Se(\hat{N})$	$\hat{f}_0$	$(\hat{f}_0/\hat{N})100\%$
Maximum Likelihood	50	0.10	1	2.00%
Chao	55	4.11	6	10.91%
Lanumteang-Böhning	52	3.51	3	5.77%
Zelterman	56	6.94	7	12.50%

Note: the number of reported accidents on distinct route, $n = 49$

From *Data Set II*, we found that the reporting of accidents were recorded on 81 unique control sections. This gave the prevalence of an accident on distinct control section 53.64%. An average of repeated reporting accident on each control section was 6.64 (*Sd* = 9.06) times. The frequency count of repeated reporting cases is showed in Table 2 and Figure 1. Likely Data set I, there is a linear trend between r_j and *j*, see Figure 2. This is an evidence of the presence of population heterogeneity in repeated recording accident. Therefore, Chao'estimator and Lanumteang-Böhning estimator are fairly to be chosen. From Table 4, Chao's estimator and Lanumteang-Böhning estimator gave similar results which showed the estimator of under-reporting accidents on 92 (*Se* = 6.08) and 94 (*Se* = 9.69) control sections, respectively. These leads to the estimated proportion of under-reporting cases 11.96%–13.83% as well as the proportion of completeness 86.17%–88.04%

Table 4: Estimated number of under-reporting control sections

Method	$\hat{N}$	$Se(\hat{N})$	$\hat{f}_0$	$(\hat{f}_0/\hat{N})100\%$
Maximum Likelihood	82	0.35	1	1.22%
Chao	92	6.08	11	11.96%
Lanumteang-Böhning	94	9.69	13	13.83%
Zelterman	98	12.68	17	17.35%

Note: the number of reported accidents on distinct control section, $n = 81$

4. Conclusion/Discussion

We examined the use of capture-recapture techniques to estimate the under-reporting accidents on highway undertaken by The Bureau of Highway I (ChiangMai) in 2015. The data were collected from Highway Accident Information Management System (HAIMS). The route number and control section were applied to identify repeated reporting an accident cases as well as used to construct the capture-recapture data. We found that the prevalence rate of reporting accidents on distinct route and control section were 49.49% and 53.64%, respectively. In order to estimate the under-reporting cases, Maximum likelihood, Zelterman's, Chao's and Lanumteang-Böhning estimator were considered. They gave similar results. The estimated proportion of incompleteness reporting accident on distinct routes was 2.00%–12.50% whereas the estimated proportion of incompleteness reporting accident on distinct control section was 1.22%–17.35%. If we think of using two items to identify the unique case and then crossed check the consistent results. In this study, both classify yielded sensible results in estimating the under-reporting cases. Using the ratio plot found that our considered data seemed to be heterogeneous Poisson model rather than homogenous once. Therefore, Chao's estimator and Lanumteang-Böhning estimator might be more suitable for this data set. Both methods provided the estimating of 55 and 52 under-reporting routes. This leads to the exact estimated prevalence rate of accident on distinct route 55.56% and 52.53%, respectively. On the other hand, they gave the estimating of 92 and 94 under-reporting control sections. Thus, the exact estimated prevalence rate of accident on unique control sections are 60.93% and 62.25%, respectively.

In this study, we used only data from one source to identify the repeated reporting cases on distinct routes and control sections. Further works, it might be sensible to cooperate with using multiple sources capture-recapture data such as merging data from reporting of police station, hospital and other useful system.

5. Acknowledgements

Both authors are grateful to the Bureau of Highway I (ChiangMai), Department of Highway for providing public access to the data on road accidents. We would like to thank the reviewers for the very helpful comments which lead to considerable improvement of the paper.

6. References

[1] Pollock KH. Capture-recapture models. Journal of the American Statistical Association. 2000; 95: 293-296.

[2] Böhning D, van der Heijden PGM. Recent developments in life and social science applications of capture-recapture methods. AStA Advances in Statistical Analysis. 2009; 93: 1-3.

[3] Hay G, Smit F. Estimating the number of drug injectors from needle exchange data. Addiction Research and Theory. 2003; 11: 235-243.

[4] van der Heijden PGM, Cruy MJLF, van Houwelingen HC. Estimating the size of a criminal population from police records using the truncated Poisson regression model. Statistica Neerlandica. 2003; 57: 289-304.

[5] vab Hest NAH, Smit F, Verhave JP. Underreporting of malaria incidence in the Netherlands:Results from a capture-repcature study. Epidemiology and Infection. 2002; 129: 371-377.

[6] Asiancorrespondent. Thailand's roads 2nd most dangerous in the world [document on the Internet] ;2014 [update 2014 February 25; cited 2016 January 19] Available from: https://asiancorrespondent.com.

[7] Chao A, Lee SM. Estimating the Number of Classes via Sample Coverage. Journal of the American Statistical Association. 1992; 87: 210-217.

[8] Zelterman D. Robust Estimation in truncated discrete distributions with application to capture-recapture experiments. Journal of Statistical Planning and Inference. 1988; 18: 225-237.

[9] Böhning D. A simple variance formula for population size estimators by conditioning. Statistical Methodology. 2008; 5: 410-423.

[10] Chao A. Estimating the population size for capture-recapture data with unequal catchability. Biometrics. 1987; 43: 783-791.

[11] Lanumteang. K, Böhning D. An extension of Chao's estimator of population size based on the first three capture frequency count. Computational Statistics and Data Analysis. 2011; 55: 2302-2311.

[12] Böhning D, Baksh F, Lersuwansri R, Gallagher J. Use of the ratio plot in capture-recapture estimation. Journal of Computational and Graphical Statistics. 2013; 22(1): 135-155.

7. Appendix

Table A1 : The number of reported an accident on each route number

Route	The number of reported cases	Route	The number of reported cases	Route	The number of reported cases
1	134	1010	2	1141	2
11	50	1012	2	1147	2
103	3	1013	1	1157	7
105	1	1014	1	1178	1
106	15	1015	2	1249	3
107	50	1030	1	1252	1
108	90	1033	13	1260	1
109	6	1035	27	1263	4
114	3	1036	2	1287	1
116	22	1037	2	1317	2
118	9	1039	7	1322	2
120	7	1048	1	1348	4
121	5	1088	6	1349	1
127	2	1095	20	1359	2
128	2	1099	1	1361	2
1001	6	1103	6	$n = 49$	*538 Cases in Total*
1009	3	1124	1		