A Comparison of the Performance of Four Confidence Intervals for Population Size Based on a Capture –Recapture Design

Jianjun Hua[1], Paul I. Nelson[2]
[1]Research Computing, Computing Services, Dartmouth College, Hanover, New Hampshire 03755
[2]Department of Statistics, Kansas State University, Manhattan, Kansas 66506

Abstract

We use simulation to study and compare the performance, in terms of coverage rate and length, of four methods of constructing confidence intervals for population size based on a two-stage Capture-Recapture (*CR*) experiment. Two methods are based on the asymptotic normality of point estimators and two are obtained from inverting chisquare and likelihiood ratio tests. In the scope of the settings we studied, we found that the method based on inverting a chisquare test is best and that none of the methods performs well if sampling fractions are small. As a practical matter, our conclusion is that that *CR* designs are most useful for relatively small populations, such as endangered species, where there may be a rough prior estimate of poulation size to guide sample size selections or in a populations where large samples are easy to obtain.

Keywords: survey sampling; hypergeometric distribution.

1. Introduction

How many endangered or threatened animals, such as Giant Pandas, Bald Eagles, whales, polar bears, etc., are still living in their natural habitats? How many fish are there in the Connecticut River? How many people have type II diabetes? How many faults are there in newly-developed software? These are but few instances of the important problem of estimating a population size N when a census cannot be taken. As described in Lohr(2010) and Buckland et al (2000) and many other references, capture–recapture (CR) experiments, described below, have long been used as multi-stage sampling designs to obtain estimators, denoted $\hat{N}$, of population size. The behavior of $\hat{N}$ depends heavily on N and the sampling fractions, all unknown quantities. In particular, the distribution of $\hat{N}$ in finite samples can be highly skewed and far from normality so that reporting its standard error may not convey an accurate assessment of the uncertainty that should be attached to $\hat{N}$ as an estimator of N. Instead, a confidence interval for N is often reported to augment the point estimator $\hat{N}$. For example, Boyce and Andel (2012) used a two stage *CR* experiment, as described in the example below, to estimate the number of adult sockeye salmon in the Taku river in calendar year 2009 and obtained an estimate $\hat{N}$ = 85528 and a nominal 95% confidence interval [77395, 93361]. Here, we use simulation to study the behavior of several methods of constructing nominal 95% confidence intervals for N in terms of actual coverage rate and mean length. Although there is an

extensive literature on the *CR* method, we have not found any references to comparative studies similar to ours.

2. Capture-Recapture Experiments and Estimation of Population Size

2.1 Basic Idea

The basic idea of the Capture-Recapture (*CR*) Method is to use the "overlap information" contained in different samples from a population to estimate some characteristics of a population, such as the population size, survival rate, etc. In its simplest, two-stage form, a Capture-Recapture experiment randomly selects n_1 units from the population and marks them. A captured fish, for example, could be marked by placing a tag on its tail. The tagged units are then returned to the population and assumed to be in their original conditions. A second random sample of size n_2 is then selected, resulting in x marked units. We use simulation to explore the performance of four methods of constructing confidence intervals for size N based on observing (n_1, n_2, x). These assumptions imply that the number of marks may be modeled as the number of common elements obtained from independent simple random samples without replacement from the same population. Taking, as we do, the sample sizes n_1 and n_2 to be fixed by design, x, the number of marks in the second sample, is the observed value of a random variable X having the hypergeometric distribution with mass function given in by

$$f(x|n_1, n_2, N) = \frac{\binom{n_1}{x}\binom{N-n_1}{n_2-x}}{\binom{N}{n_2}}, \quad (1)$$

where $\max\{0, n_1 + n_2 - N\} \leq x \leq \min(n_1, n_2)$.

There are more complex versions, not considered here, of *CR* designs that allow, for example, stratifed sampling, more than two stages, time dependent and varying capture rates among individuals, immigration and emigaration. See Darroch (1961), Buckland et al (2000), Cormak (1992), Chao (1989) and Jolly (1965).

2.2 Estimation from CR Experiments

Since $E(X / n_2) = n_1 / N$, the proportion of marked items in the population, the method of moments point estimate of N is given by

$$\hat{N} = n_1 n_2 / x \ , \quad (2)$$

if $x > 0$ and undefined otherwise. For $x > 0$, an approximate variance of $\hat{N}$ is given in Lohr (2010) by

$$\hat{V}(\hat{N}) = (n_1 n_2 / x)^2 (n_2 - x) / (x(n_2 - 1)) \quad (3)$$

$$\approx n_1^2 n_2 (n_2 - x) / x^3 \quad .$$

Note that the uncertainty in $\hat{N}$ increases rapidly as x decreases. In this simple form, capture-recapture estimation is a special case of ratio estimation of a population total and up to approriate round off to an integer, $\hat{N}$ is also the maximum likelihood estimator. If the sample sizes n_1 and n_2 are not fixed by design, $\hat{V}(\hat{N})$ may be interpreted as being conditional on them. Properties of $\hat{N}$ were studied by Chapman (1951) , who showed that although $\hat{N}$ is a best asymptotically normal estimate of N as $N \to \infty$, it is biased, and its bias can be large for small samples. However, when $n_1 + n_2 \geq \mathrm{N}$, his modified, unbiased estimate, which may even be used when $x = 0$, is given by

$$\tilde{N} = (n_1 + 1)(n_2 + 1) / (x + 1) - 1 \,. \tag{4}$$

An approximate variance of Ñ when x > 0 was given in Seber (1970) and Wittes (1972) by

$$\hat{V}(\tilde{N}) = (n_1 + 1)(n_2 + 1)(n_1 - x)(n_2 - x) / [(x + 1)^2 (x + 2)] \tag{5}$$

Although $n_1 + n_2 \geq \mathrm{N}$ is a very restrictive and unrealistic condition, the use of (4) and (5) has some flexibility. Researchers sometimes apply them to their *CR* studies even if $n_1 + n_2 < \mathrm{N}$. Specifically, in our simulation study, the Chapman estimation method was used as a substitute for two of the methods we investigate which are not defined whenever no marked units are recaptured in the second sample.

2.3 Some Methods for Constructing Confidence Sets for N

We consider several commonly used methods for constructing approximate 1-α confidence sets for the unknown population size *N*.

(i) *Asymptotic Normality*: $\hat{N} \pm z_{\alpha/2}\sqrt{\hat{V}}$, where $\hat{N}$ is the estimator given in (2) when $x > 0$ and or $\sqrt{\hat{V}(\tilde{N})}$, where Ñ is the Chapman's estimator given in (4) and z_γ is the 100(1-γ) percentage point of a standard normal distribution .

(ii) *Inverting a Test for N*: We used two versions of this approach described in Lohr (2010). The first is based on a chisquare test for independence in the only partially observed two-way contingency table given below, where $x_{22} = N - x_{11} - x_{12} - x_{21}$ is the unknown number of units observed in neither sample. Let $p(x_{22}^*)$ be the p-value obtained by carrying out the chisquare test with $x_{22} = x_{22}^*$. The set $\{N; p(x_{22}) > 1 - \alpha\}$, constructed using an iterative search is then an approximate $1 - \alpha$ confidence set for *N*.

Table 1: Two-way Contingency Table

		In Sample 2? Yes	No	
In Sample 1?	Yes	$x_{11}(=x)$	x_{12}	$x_{1+}(=n_1)$
	No	x_{21}	x_{22}	x_{2+}
		$x_{+1}(=n_2)$	x_{+2}	$x_{++}(=N)$

Similarly, we also inverted a likelihood ratio test of H_0: $N=N_0$ vs. $H_1 : N \neq N_0$ with critical region $\{x; -2\text{Log}\left[\frac{L(N_0)}{L(\hat{N})}\right] > \chi^2_{1-\alpha}(1)\}$, the likelihood L($N$) from is given by $L(N) = \binom{N-n_1}{n_2-x} / \binom{N}{n_1}$. For $1-\alpha$ = 0.95, the confidence level we used through out, the confidence set CI is expicitly given by

$$CI = \{N; L(\hat{N}) / L(N) \leq e^{1.92}\},$$

where $\hat{N}$ is the estimate given in (2) for $x > 0$. Chapman estimator was used when $x = 0$ Note that $\hat{N} \in CI$ and that CI contains values less than and greater than $\hat{N}$. To iteratively construct CI, suppose first that $N^{\#} \in CI$. Then, to find values of N greater than $N^{\#}$ that may be in CI, check if

$$L(\hat{N}) / L(N^{\#}+1) = [L(\hat{N}) / L(N^{\#})] L(N^{\#}) / L(N^{\#}+1) \leq e^{1.92}.$$

Note that many terms cancel in $L(\hat{N}) / L(N^{\#}+1)$ and the process is started with $N^{\#} = \hat{N}$. Proceed similarly using $L(\hat{N}) / L(N^{\#}-1)$ to find values of N < $\hat{N}$ that are in CI. This method is not defined when x=0.

2.4 Example

The design used in the example cited above on the estimation of the number of Salmon in the Taku River actually used a stratifeid sample consisting of two strata. However, since the second startum was small and following Darroch (1961), we may use our simpler setup as an approximation with n_1 = 3135, n_2 = 11217 and x = 405 observed marks in the second sample. Hence $\hat{N}$ = 3135*11217/405 ≈ 86828, which is close to the value of 85528 reported by Boyce and Andel (2012), along with their approximate 95% interval given as [77395, 93661]. Our approximate 95% intervals are: [78526, 95130] using the asymptotic normality of $\hat{N}$; [78937, 94359] using the asymptotic normality of $\tilde{N}$; [79480, 94996] based on inverting the chisquare test; [79562, 95104] based on inverting the likelihood ratio test. Note the similarity among the intervals and how wide they are.

3. Simulation Study

3.1 Basic Procedure

Simulation studies were implemented out and visualized using R and SAS to evaluate the performance, in terms of actual coverage rates , and mean width of the four *CR* methods of constructing confidence intervals for the population size *N*: *Asymptotic Normality of* $\hat{N}$, *Asymptotic Normality of* $\tilde{N}$ (Chapman estimator), *Inverting the* χ^2 *Test* and *Inverting a Likelihood Ratio Test*, denoted respectively *NormAN*, *NormCM*, *InvCQ* and *InvLR*, based on the data (n_1, n_2, x). Data were generated using the assumpions given in Section 2 and the resulting hypergeometric distribution of *X* given in (1). Specifically, two independent, simple random samples of sizes (n_1 and n_2) without replacement were selected from a predetermined population of size *N*. The units found in both samples, namely "overlap", were used as the recapture value *x*. The sample sizes were represented as proportions (p_1 and p_2) of the population size *N*, namely $n_1 = Np_1$ and $n_2 = Np_2$, where p_1 and p_2 are looped from 0.05 to 0.45 in steps of 0.10, resulting in twenty five pairs, and N set equal to 1000 and 5000. For each pair (p_1, p_2) 1000 independent data intervals were generated resulting in 1000 values of the number of marked items in the second sample. Confidence sets were constructed as described above. Estimated coverage rates, and mean half widths were recorded in order to compare the four methods' performance. Note that all four methods are used on each data set. The observed coverage rates, denoted $\hat{C}$, and estimated mean half width of the 1000 intervals, denoted *MHW*, were obtained for each of the twenty five settings and displayed in Tables 2 and 3.

Table 2: Comparisons (N=1000): NormAN vs NormCM vs InvCQ vs InvLR

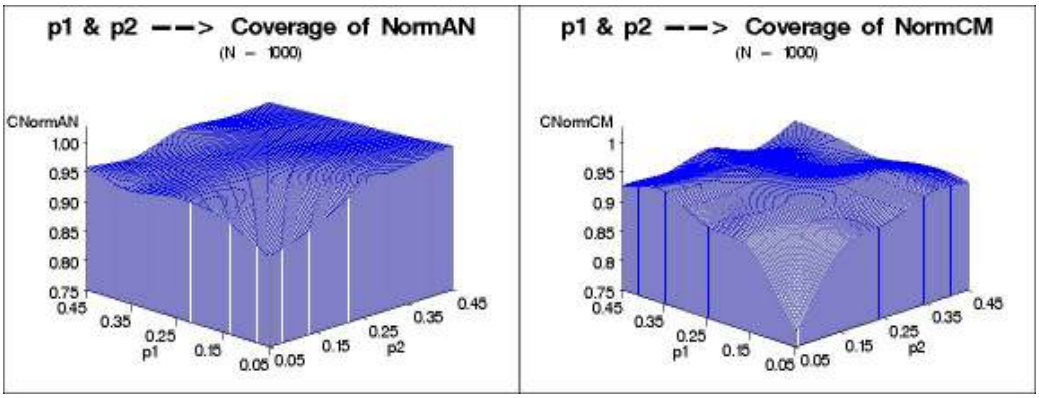
ID	p1	p2	CNormAN	CNormCM	CInvCQ	CInvLR	MHWNormAN	MHWNormCM	MHWInvCQ	MHWInvLR
1	0.05	0.05	0.896	0.774	0.969	0.894	2022	959	940	5657
2	0.05	0.15	0.925	0.885	0.959	0.921	941	621	623	1349
3	0.05	0.25	0.968	0.903	0.953	0.953	602	448	501	603
4	0.05	0.35	0.982	0.933	0.947	0.960	485	355	406	431
5	0.05	0.45	0.993	0.932	0.966	0.966	420	291	327	336
6	0.15	0.05	0.937	0.886	0.970	0.927	871	619	629	1165
7	0.15	0.15	0.951	0.925	0.948	0.948	405	344	381	400
8	0.15	0.25	0.974	0.940	0.958	0.947	310	256	274	279
9	0.15	0.35	0.986	0.955	0.968	0.955	257	201	211	213
10	0.15	0.45	0.991	0.942	0.960	0.960	222	161	167	167
11	0.25	0.05	0.948	0.902	0.963	0.963	532	441	499	593
12	0.25	0.15	0.964	0.936	0.957	0.945	289	254	273	277
13	0.25	0.25	0.976	0.946	0.955	0.948	218	184	191	192
14	0.25	0.35	0.982	0.933	0.943	0.943	184	146	150	150
15	0.25	0.45	0.992	0.933	0.941	0.941	161	118	121	120
16	0.35	0.05	0.939	0.939	0.935	0.959	413	361	413	440
17	0.35	0.15	0.960	0.947	0.957	0.938	225	201	211	213
18	0.35	0.25	0.975	0.949	0.958	0.958	172	147	151	150
19	0.35	0.35	0.985	0.939	0.952	0.952	143	114	116	116
20	0.35	0.45	0.992	0.949	0.953	0.947	127	93	94	94
21	0.45	0.05	0.957	0.926	0.962	0.962	325	292	328	338
22	0.45	0.15	0.954	0.928	0.945	0.945	178	160	166	166
23	0.45	0.25	0.983	0.950	0.954	0.954	139	119	122	121
24	0.45	0.35	0.984	0.928	0.946	0.932	116	93	94	93
25	0.45	0.45	0.993	0.960	0.959	0.959	103	76	76	75

Table 3: Comparisons (N=5000): NormAN vs NormCM vs InvCQ vs InvLR

ID	p1	p2	CNormAN	CNormCM	CInvCQ	CInvLR	MHWNormAN	MHWNormCM	MHWInvCQ	MHWInvLR
1	0.05	0.05	0.966	0.889	0.977	0.961	3135	2567	2679	3517
2	0.05	0.15	0.961	0.947	0.951	0.943	1631	1430	1526	1567
3	0.05	0.25	0.979	0.942	0.949	0.949	1245	1050	1094	1106
4	0.05	0.35	0.989	0.942	0.945	0.958	1034	819	843	848
5	0.05	0.45	0.996	0.954	0.963	0.963	911	667	683	685
6	0.15	0.05	0.956	0.945	0.950	0.937	1553	1437	1537	1575
7	0.15	0.15	0.972	0.935	0.955	0.949	862	782	799	804
8	0.15	0.25	0.978	0.951	0.960	0.966	663	569	577	577
9	0.15	0.35	0.982	0.929	0.951	0.946	558	447	451	451
10	0.15	0.45	0.984	0.937	0.937	0.934	493	364	367	366
11	0.25	0.05	0.968	0.958	0.960	0.960	1106	1046	1090	1102
12	0.25	0.15	0.969	0.947	0.935	0.942	624	570	577	578
13	0.25	0.25	0.985	0.964	0.964	0.964	483	416	419	419
14	0.25	0.35	0.981	0.944	0.949	0.945	406	326	328	327
15	0.25	0.45	0.991	0.952	0.957	0.957	359	265	266	265
16	0.35	0.05	0.961	0.948	0.949	0.953	861	822	846	851
17	0.35	0.15	0.970	0.949	0.956	0.950	488	447	451	451
18	0.35	0.25	0.976	0.958	0.959	0.958	379	327	328	328
19	0.35	0.35	0.989	0.953	0.954	0.945	319	257	258	257
20	0.35	0.45	0.993	0.935	0.934	0.934	282	209	209	208
21	0.45	0.05	0.965	0.953	0.954	0.954	694	666	681	683
22	0.45	0.15	0.964	0.952	0.955	0.952	397	364	367	366
23	0.45	0.25	0.977	0.945	0.947	0.947	306	265	265	265
24	0.45	0.35	0.980	0.948	0.954	0.954	259	208	208	207
25	0.45	0.45	0.994	0.953	0.943	0.943	229	170	170	169

3.2 Coverage Rates:

Cases in Tables 2 and 3 where the true coverage rate $C = E(\hat{C})$ is statistically significantly different from the target value of .95 are marked in red. Note that true rates are (i) mostly high for *NormAN*; (ii) often off target for *NormCM*, especially for N = 1000; (iii) similar and mostly *acceptable* for InvCQ and *InvLR*; (iv) better for N = 5000 than for N = 1000; (iv) unstable in the first row, corresponding to the smallest sampling prportions. Overall, *InvLR* and *InvCQ* appear to be the methods of choice in terms of attaining the desired coverage rate. These conclusions are supported by the surface plots of estimated rates in Figure 1 for N = 1000. Similar plots for N = 5000 are omitted.



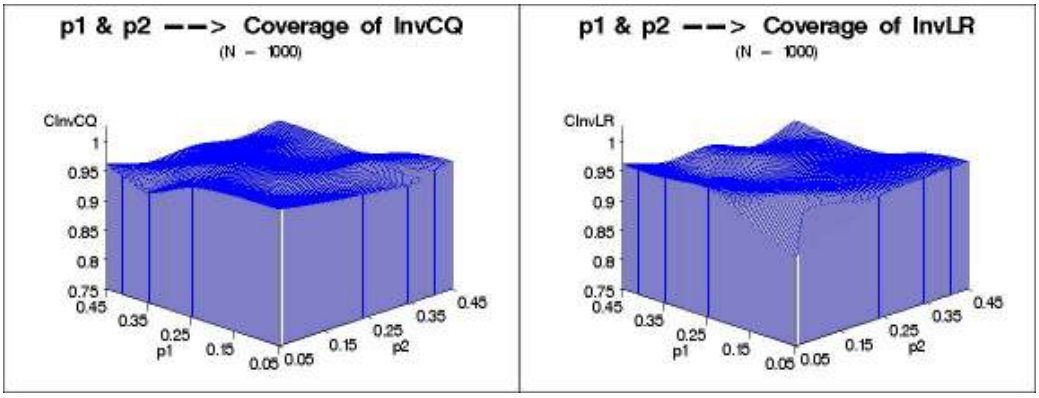


Figure 1: Surface Plots of Estimated Coverage Rates N = 1000

3.3 Analyses of Mean Half Width

We only investigate *InvCQ* and *InvLR* here since the other two methods were considerably less reliable in holding their target coverage rates, especially for N = 1000. Also, to facilitate comparing interval widths across population sizes, we analyzed Relative Mean Half Width (*RMHW*) = *MHW*/*N*. As can be seen from Tables 2 and 3, *RMHW* is so large for $p_1 = p_2 = .05$ that including it would have distorted the plots. Accordingly, this case was deleted. Side by side boxplots of the remaining twenty four cases of *RMHW*'s presented in Figures 2 and 3 below indicate that the intervals are narrower for N = 5000 than for N = 1000 and similar for the two methods. Specifically, the first and third quartiles of *RMHW* 's are: (i) For *InvLR* : 0.13 and .43 (N= 1000) and 0.06, 0.17 (N = 5000); For *InvCQ* : 0.13, 0.40 (N = 5000); 0.06, .17 (N= 5000). Furthermore, in all 24 cases *RMHW* for N = 1000 is greater than for N = 5000, for both methods. The decrease in RMHW for fixed sampling proportions in going from N = 1000 to N = 5000 reflects the corresponding increases in sample sizes. As can be seen in Tables 2 and Table 3, confidence interval widths can be very wide unless sampling proportions, which are rarely known, are large .

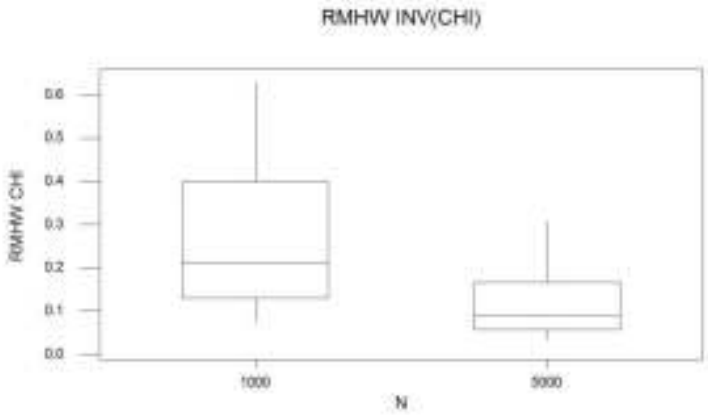


Figure 2: Boxplots of Relative Mean Confidence Intervals *InvCQ*

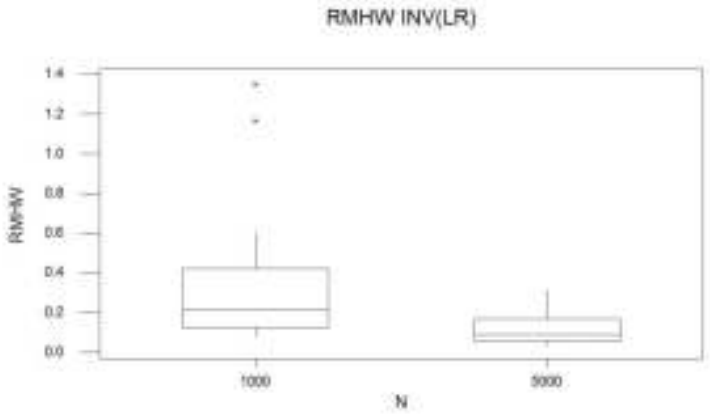


Figure 3: Boxplots of Relative Mean Half Widths of Confidence Intervals *InvLR*

We also used least squares regression analysis to quantitatively explore the relationship between RMHW and sampling fractions $(p_{1,}p_2)$ by fitting the model $RMHW = \beta_0 + \beta_1 p_1 + \beta_2 p_2 + \beta_3 p_1 p_2 + \varepsilon$ and obtined the fitted surfaces

for *InvLR*

$RMHW = 1.44 - 2.82\, p_1 \text{ -2.79p}_2 + 5.37 p_1 p_2$, $N = 1000,\ R^2 = .74,$

$RMHW = 0.38 - 0.60\, p_1 \text{ -0.60p}_2 + .80 p_1 p_2$, $N = 5000,\ R^2 = .84,$ and

for *InvCQ:*

$RMHW = 0.76 - 1.07\, p_1 \text{ -1.07p}_2 + .97 p_1 p_2$, $N = 1000,\ R^2 = .83,$

$RHMH = 0.37 - 0.57\, p_1 \text{ -0.57p}_2 + .74 p_1 p_2$, $N = 5000,\ R^2 = .82.$

Note that for both methods, *RMHW* , as expected , decreases as each sampling fraction inreases, the other being held fixed and that this rate of change decreases as the other sampling fraction increases, within the range of sampling fractions in our study. To the extent that these fitted surfaces refelct the true surfaces, decreases in length achieved by increasing one sampling fraction while keeping the other fraction fixed are similar for each fitted model. The greater similarity of the fitted surfaces for *InvLR* and *InvCQ* for N= 5000 than for N = 1000 appears to reflect the instability and large inflation in *RMHW* for *InvLR* as (p_1, p_2) aproach zero for the smaller population size.

Consequently, although the two methods are similar in terms of coverage rate and relative mean half width in most cases, we prefer *InvCQ* because of its greater stability for the smaller sampling fractions.

4. Conclusion

In the scope of the settings we studied, we found that the method based on inverting a chisquare test is best and that none of the methods performs well if sampling fractions are small. As a practical matter, our conclusion is that that *CR* designs are most useful for relatively small populations, such as endangered species, where there may be a rough prior estimate of poulation size to guide sample size selections or in a populations where large samples are easy to obtain.

Future studies of this type should include the bootstrap method given in Bucklandand Garthwaite (1991) and more complex CR designs that can better handle the zero recapture problem, especially those including more than two stages.

Acknowledgement

The authors would like to express gratitude to Dr. Leigh Murray, Dr. Haiyan Wang and Dr. Juan Du for their valuable input during this research.

References

Boyce, I. M**.** and Andel, J. E. (2012). 'Mark-Recapture Studies of Taku River Adult Sockeye Salmon Stocks in 2009'. *Pacific Salmon Commission Technical Report No. 29.*

Buckland, S. T and Garthwaite, P.H. (1991). 'Quantifying Precision of Mark-Recapture Estimates Using the Bootstrap and Related Methods'. *Biometrics*, *47*, pp 255-268.

Buckland, S. T., Goudie, B. J. and Borchers, D. L. (2000). 'Wildlife Population Assessment and Future Directions'. *Biometrics,* 56, pp 1-12.

Chapman, D. G. (1951). 'Some Poperties of the Hypergeometric Geometric Distribution With Applications to Zoological Census'. *Vol. 1, Univ. Calif. Pub. Statist.*

Cormack, R. M. (1992). 'Interval Estimation for Mark-Recapture of Closed Populations'. *Biometrics, 48*, pp567-576.

Darroch, J. N. (1961). 'The Two-Sample Capture –Recapture Census when Tagging and Sampling are Stratified'. *Biometrika, 48*, pp 241-260.

Jolly, G. M. (1965). 'Explicit Estimates from Capture-Recapture Data with Both Death and Immigration –Stochastic Model'. *Biometrika, 52,* pp 225-247.

Lohr, S. L. (2010). 'Sampling Design and Analysis'*, Second Edition*, Brooks/Cole, Boston.

Seber, G. A. F. (1970). 'The effects of trap response on tag-recapture estimates', *Biometrika* 26, 13–22.

Wittes, J. T. (1972). 'On the bias and estimated variance of chapman's two-sample capture-recapture population estimate', *Biometrics* 28, 592–597.