Capture-Recapture Sampling Techniques: Artificial and Real Population Data Analysis

Neelam Bukhari ¹, Sadia Bukhari ², Muhammad Ahmed Shehzad ^{*3}

^{1,3} Department of Statistics, Bahauddin Zakariya, University, Multan

² Higher Education Commission of Pakistan

* Corresponding author: ahmad.shehzad@bzu.edu.pk

Capture-Recapture sampling methods are being used to estimate the population size of an unknown population. These methods are widely used for determining the population size of animals and birds. Literature shows that these methods were purely used by ecologist in the past. But now-a-days, these methods are also used to determine the chronic diseases, for example, cancer patients, HIV Aid patients, etc. Not only this, many studies had conducted to determine the road accidents ratios on different points. This study covered two examples; one was the total number of female drivers at Bahauddin Zakariya University, Multan and the second example was about the total number of male smokers at Bahauddin Zakariya University, Multan. On these two examples, the basic capture-recapture sampling methods were used which involved Lincoln Peterson Index, Chapman Estimate and Schnable Estimate. Then an artificial data with known population was created and same methods were used for analysis. In the end, the results of artificial data and two examples were compared with each other through analysis on excel sheet.

Keywords: Capture- Recapture sampling, Artificial P{opulation, Real Population

ARTICLE INFORMATION

Received: 02 June 2019
Revised: 22 July 2019
Accepted: 30 September 2019

DOI: 10.31580/jpvai.v2i3.930

© Readers Insight Publication

INTRODUCTION

Literature show diverse methods of capture-Recapture sampling methods. Scientists are taking keen interest to investigate the new ways on CRC methods like Manly-Parr method, Thompson and Horvitz Estimate, Bootstap etc. But present study is restricted to the basic methods which are Peterson Index, Chapman Estimate and Schnable Estimate method. The initial method of capture-recapture studies is Lincoln-Peterson Index which is considered as the standard technique (Jibasen, 2011). For the very first time, this method was used for the marine fishes study and waterfowl population studies. But rapidly, the method was being used on human as well (Smallwood, 2013). To estimate the size of unknown population, the organisms are marked as "M", the captured organisms are marked as "C" and the recaptured organism during the procedure are known as "R". Hence the derived formula for estimation through Peterson Index is:

$$N = \frac{C \times M}{R} \quad (i)$$

Later, Lincoln Peterson Index was modified by Chapman in 1951. Chapman estimator is called as the modified Peterson Estimate. These modifications were made to cover up the violation of Lincoln Index assumptions (Pollock, 2010). The formula for population estimate through Chapman is:

$$N = \left\{ \frac{(C+1)(M+1)}{(R+1)} \right\} - 1 \quad (ii)$$

It is evident from the literature that Peterson Index and Chapman estimate are for the closed population (Mineau & Whiteside, 2013). Another procedure of estimation of population is Schnable Estimate where researchers mark the organisms in first capture and then a

second sample is taken over a short period of time. The unmarked organisms in second sample are marked and this process is repeated till the final event. Schnable estimate is determined through following formula:

$$N = \frac{\sum(M_t C_t)}{\sum(R_t)} \quad (iii)$$

Where M_t is the marked individual, C_t is the caught individuals and R_t is recaptured individuals at time t .

Methods

The present study is based on two real population data examples. Example 1 estimates the total number of female drivers at Bahauddin Zakariya University, Multan. For collecting the data, the researcher selected the main gate of university and noted the car registration number of female drivers. In next step, the procedure was repeated and the recaptured female drivers were observed. The population estimate was calculated by using the Peterson Index, Chapman estimate and Schnable estimate. The researcher also computed the mean and variance of given values on excel sheet.

In Example 2, the data of male smoker at Usman Hall (a boys' hostel) at Bahauddin Zakariya University, Multan was collected through identity card numbers and roll numbers. The procedure of estimating the population size was done through basic formulas which involve Lincoln Peterson, Chapman and Schnable estimate.

After real population data an artificial data with known population $N=100$ was generated and same methods were applied to test the results. The researcher created an artificial data set with known population. The collected data was assigned a serial number. This study also covers the comparison of results obtained through artificial data and two real population data.

Results

The results of first real population example (female drivers at Bahauddin Zakariya University, Multan) showed that highest value of population (N=30) and lowest value of population was (N=22) when the procedure of Peterson Estimate was repeated for the six times. The mean was calculated as (mean=21.74) and highest variance was observed as (var=13). Chapman estimate results for first real population example showed that the highest value of estimated population was (N=29) and the lowest value was (N=21). The mean was observed as (mean=21.41) and highest value of variance was recorded as (var=14). Again, the procedure of estimating the population size by Chapman Estimate was run for the six times. The results of Schnable estimate indicated that estimated population for example 1 was (N=25) with a variance (var=5)

Real population Example 2 (male smokers at Usman Hall of Bahauddin Zakariya University, Multan) was analyzed by using the Lincoln-Index, Chapman and Schnable Estimate. The analysis was run on excel sheet. For Peterson-Index, the highest value of population estimate was (N=40) and the lowest value was (N=23). The average estimated population was (mean=26) with a highest variance (var=15). To estimate the population size and average value, the procedure was repeated for the six times. The results of Chapman estimate indicated that the highest value of estimated population was (N=39) and lowest estimated population was (N=23) and mean was calculated as (mean=25) with a highest variance (var=16). The estimation of population through Schnable estimate was found (N=35) with a variance (var=3).

The real population data analysis was done to estimate the population size. Then researcher generated the artificial data with a known population size (N=100). The Peterson, Chapman and Schnable procedure were applied on this artificial data and it was analyzed that how much deviation in results exist for the already known population. The highest value of population estimate was found (N=106) and the lowest value was (N=97) for Peterson Index. The average value was found (mean=100) with a highest variance (var=19). The estimated population for Chapman was recorded as (N=101), the highest value and (N=92) as lowest value. The mean was (mean=97) with a highest variance (var=5). The Schnable Estimate results indicated the estimated population as (N=106) with a variance (var=2).

After that, the researcher compared the results of real population data with artificial data through following table:

Table 1: Comparison Table

Analysis	Artificial Data			Example 1			Example 2		
	Peterson Estimate	Chapman Estimate	Schnable Estimate	Peterson Estimate	Chapman Estimate	Schnable Estimate	Peterson Estimate	Chapman Estimate	Schnable Estimate
Mean/Estimate	100.02	97.89	106.96	21.74	21.41	25.45	26.12	25.85	35.85
Variance	19.73	5.57	2.64	13.64	14.40	5.91	15.00	16.53	3.36

Table 1 represents the comparison of both real population examples with artificial data. The analysis of artificial data indicated that the estimation of population size was very close to the actual size of the population. Hence, in example 1, it is concluded that total number of female drivers at Bahauddin Zakariya University are 25 to 30. Similarly, in example 2, it is concluded that total number of male smokers at Usman Hall of Bahauddin Zakariya University, Multan are 35 to 40.

Discussion

The findings of the current study depicts that the result of all methods are more or less similar. Later, the artificial data analysis with known population also confirms the accuracy of results. It is evident from literature that these basic methods are widely used in wild life and by ecologist. However, this study is unique in nature as none of the study has been with artificial data analysis with known population by using the Peterson Index, Chapman estimate and Schnable estimate. This study also has limitations as it is conducted in closed population and in Bahauddin Zakariya University while if the

study would conducted at city level then result would definitely be different. Further, this study is done by using the simplest methods of capture-recapture techniques. There are lots of other methods for such techniques. They can also be used on these real population examples and later comparison can be made with artificial generated data.

References:

- Chapman, D. G. (1951). Some Properties of the Hypergeometric Distribution with Applications to Zoological Sample Censuses, University of California Publ. Stat. 1(7):13 1-160.
- Jibasen, D. (2011). Capture-recapture Type Models for Estimating the Size of an Elusive Population. A PhD Thesis submitted to the University of Ilorin, Ilorin, Nigeria.
- Pollock KH (2010). Modeling capture, Recapture and Removal Statistics for Estimation of Demographic Parameters for Fish and Wildlife Populations: past, present and future. J. Amer Stat Assoc.
- Mineau, P. & Whiteside, M. (2013). Pesticide Acute Toxicity Is a Better Correlate of US Grassland Bird Declines than Agricultural Intensification. Plos One, 8.
- Smallwood, K.S. (2013) Comparing bird and bat fatality-rate estimates among North American wind-energy projects. Wildlife Society Bulletin, 37, 19-33