We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Estimation of Measles Immunization Coverage in Guwahati by Ranked Set Sampling

Vivek Verma and Dilip C. Nath

Abstract

In order to study the efficacy of the ranked set sampling (RSS), as an alternative procedure, for estimation of the proportion of children aged 12–23 immunized against measles vaccine, a study is conducted in slum and non-slum regions of Guwahati, the capital of Assam, India. The RSS-based approach in the cases of both perfect and imperfect rankings is compared with its counterpart simple random sampling (SRS). The results revealed that estimates based on RSS with set size (4) are very close to Census report for Assam (2012) and has less variability than the SRS estimator. RSS-based estimates for different choices of probability of ranking error (ρ) are not only more accurate but are more precise and efficient than the SRS procedure, and also suggest that a the procedure of RSS better than the classical SRS.

Keywords: relative precision, relative saving, order statistics, maximum likelihood estimator

1. Introduction

In public health related studies, the virus of measles is considered as highly epidemic and responsible for serious diseases. According to the medical dictionary, measles virus infects the lungs in children, which can cause pneumonia in older children, can cause inflammation of the brain, called encephalitis, which can cause seizures and brain damage [1]. As a preventive measure, adequate vaccination is introduced in early childhood to acquire immunity against measles virus. According to the Integrated Child Development Services Program (ICDS) in India, a child should have received the basic vaccines (BCG, polio, DPT and measles) at 12–23 months of age.

In this regard, the World Health Organization (WHO) launched the Expanded Program on Immunization (EPI) in May 1974 to immunize children around the world. Since then, it has been widely used to access coverage. India launched the National Vaccination Program entitled Expanded Programme of Immunization in 1978 with the introduction of the BCG, OPV, DPT and anti-typhoid-paratyphoid vaccine. The EPI was renamed and launched, with a major change in list of vaccinations, as the Universal Vaccination Program on 19 November 1985. The measles vaccine [2] has been added to latest schedule.

To study the vaccination coverage a few among various sampling procedures and are the most popular *viz.,* lot quality assurance sampling (LQAS) [3, 4],

systematic sampling, cluster sampling [5] and stratified sampling [6], are based on simple random sampling (SRS) either at one stage or subsequent stage. In most of the practical situations to obtain actual measurement of an observation is not relatively easy and economical, but ranking a small subset of auxiliary information about a sample is relatively easy, economical and reliable. McIntyre in 1952 [7] have introduced the procedure of ranked set sampling (RSS), as an alternative, which is highly beneficial and much superior to the standard SRS, for estimating population parameters. In RSS procedure a set of units is randomly drawn from the population and the selected units are ranked by using judgment method or by other means which does not require actual measurement. Only the unit possessing lowest rank is measured for this set of units. Next, a second set of units of the same size is drawn as earlier and ranked; then the unit at the second lowest position is measured. This procedure of ranking and measuring is continued until we get as many observations as the size of the set. This entire procedure is regarded as a cycle. The process of cycle formation will be continued until the desired sample size is obtained for analysis.

Based on real-life primary data, in the present study, the procedure of RSS is investigated against SRS for estimating the proportion of children aged 12–23 months of slum and non-slum households of Guwahati, the capital city of state Assam, India, who are not immunized with the vaccine against measles. The information on a total of 500 (260 slum and 240 non-slum) households, were obtained after being ranked using mother's age 15–49 years (in months), as auxiliary variable is used as auxiliary variable. The data of the same size is obtained by following the SRS procedure, for the evaluation of the performance and effectiveness of the RSS estimator compared to SRS estimator.

2. Sampling design

The study population is a representative cross-sectional sample of children aged 12–23 months to mothers aged 15–49 of Guwahati City, India. Five hundred households, following both SRS and RSS procedures, having the children of age from 6 months to 5 years were identified for the present study. Following the SRS technique, from both slum and non-slum regions, a sample 250 households were obtained. Among the selected 500 households under RSS, 260 were residents of slum region whereas remaining 240 are of non-slum parts of the Guwahati City. The variable considered in the study is the proportion of children in Guwahati City, India who are not immunized with the vaccine against measles. It is assumed that the case of receiving the vaccination of children usually depends on the awareness of the child's mother on immunization. Lower the age of a mother who has children aged 12–23 months, lower would be the consciousness as expected. Therefore, the age of the mother (in months) was used as a ranking variable in RSS for classification of the sample. Information on children was obtained by a face to face interview with the mothers. Mothers were selected through RSS conducted in Guwahati City.

The observation were divided into m sets of size four ($s = 4$) each. The observations under RSS procedure are obtained through the following steps.

1. A simple random sample of units s^2 is selected from the target population and is distributed randomly in sets s, each with s units.

2. In each set s, the units are ranked according to the age of the mother. Since the samples in different sets are ranked according to the age of the mother (1, 2, 3, 4).

Obviously, there is a high chance of having ties then in this situation, the observations are ordered systematically in sequence, as explained by Terpstra and Nelson (2005).

3. From the first set, the unit corresponds to the mother with the lowest age (in months) is selected. From the second set, the unit corresponds to the mother with the second lowest age is selected and so on. Finally, from *sth* set, the unit corresponding to the mother with the highest age is selected. The other $s(s-1)$ sampled units are discarded from the data set.

4. Steps 1–3, called a cycle, are repeated m times to obtain a ranked set sample of size ms.

Corresponding to each selected mother, information regarding whether her child is administrated with measles vaccination or not is collected.

Suppose X is the binary response that takes value "1" if the child is vaccinated with measles vaccine and "0" otherwise. Let $\{X_{[i]j}, i = 1, 2, \cdots, s, j = 1, 2, \cdots, m\}$ represent a ranked set sample of size ms, where $X_{[i]j}$ takes the values "1" or "0" according as the jth child in the ith ranking class is vaccinated or not. By virtue of ranked set sampling all $X_{[i]j}'s$ are independently distributed. Here for any i from 1 to s, $X_{[i]1}, X_{[i]2}, \cdots, X_{[i]m}$ are independently and identically (i.i.d.) distributed. Here $X_{[i]1}$ can be regarded as the ith order statistic corresponding to a simple random sample of s observations, say, $(X_1, X_2, \cdots, X_s)$ on X. Obviously, X_is have the common probability mass function (p.m.f.) given by

$$f(x|p) = p^x(1-p)^{1-x}, \quad x = 0, 1, \quad 0 < p < 1, \tag{1}$$

where p is the probability that a child is vaccinated with measles vaccine in the population. Now we have, for any i, $1 \le i \le s$

$$P(X_{[i]1} = 1) = P(\text{among } X_i's, \text{ at least } s - i + 1 \text{ X} - \text{values are } 1)$$

$$= \sum_{r=s-i+1}^{s} \binom{s}{i} p^r (1-p)^{s-r} = p_{[i]}, \text{ say,}$$

$$P(X_{[i]} = 0) = \left(1 - p_{[i]}\right).$$

Obviously, $p_{[i]}$ is the proportion, in ith class, of children who received the vaccination and p is the overall proportion of children receiving the vaccine in entire target population. Here it can be easily shown that

$$\sum_{r=1}^{s} p_{[r]} = sp. \tag{2}$$

3. Estimation of parameter p

For a dichotomous population, estimation of population proportion p, based on ranked set samples have already been investigated [8–14]. A method for estimation [15] of p using RSS for the situations where the binary variable is obtained from a

continuous variable. Let $X_{RSS} = \left(\mathbf{X}_{[1]}, \cdots, \mathbf{X}_{[i]}, \cdots, \mathbf{X}_{[s]}\right)$, $\mathbf{X}_{[i]} = \left(X_{[i]1}, \cdots, X_{[i]m}\right)^T$ and $\mathbf{p} = \left(p_{[1]} \cdots, p_{[i]}, \cdots, p_{[s]}\right)$. Here $X_{[i]1}, \cdots, X_{[i]m}$ are i.i.d. Binomial $\left(1, p_{[i]}\right)$, $i = 1, 2, \cdots, s$. So the joint p.m.f. of $X_{[i]j}, j = 1, 2, \cdots, m$ is given by

$$f\left(\mathbf{x}_{[i]} | p_{[i]}\right) = \prod_{j=1}^{m} p_{[i]}^{x_{[i]j}} \left(1 - p_{[i]}\right)^{1-x_{[i]j}} = p_{[i]}^{z_i} \left(1 - p_{[i]}\right)^{m-z_i} \qquad (3)$$

where $z_i = \sum_{i=1}^{m} x_{[i]j}$, the number of children getting vaccinated observed in ith ranking class of the given ranked set sample. Obviously, $Z_i = \sum_{i=1}^{m} X_{[i]j} \sim$ Binomial $\left(m, p_{[i]}\right)$, independently for all $i = 1, 2, \cdots, s$. Then the joint p.m.f. of the whole sample $\mathbf{X}_{RSS}$ is of the form

$$f(\mathbf{x}_{RSS} | \mathbf{p}) = \prod_{i=1}^{s} f\left(\mathbf{x}_{[i]} | p_{[i]}\right) = \prod_{i=1}^{s} p_{[i]}^{z_i} \left(1 - p_{[i]}\right)^{m-z_i}. \qquad (4)$$

Applying standard maximum likelihood (ML) principle the ML estimate of p under ranked set sampling, is given by

$$\hat{p}_{RSS} = arg \ \max_{p} \ f(\mathbf{x}_{RSS} | p)$$

Given the RSS data, the form of likelihood function of p is complicated and hence the MLE of p is difficult to obtain directly. Alternatively, for $i = 1, 2, \cdots, s$, one can separately derive ML estimate, say, $\hat{p}_{[i]}$ of based on the likelihood function (3) and then MLE of p can be formulated by using the relation (2) as

$$\hat{p}_{RSS} = \frac{1}{s} \sum_{i=1}^{s} \hat{p}_{[i]}.$$

Here, it can be shown that, for each $i = 1, 2, \cdots, s$, $\frac{Z_i}{m}$ would be the maximum likelihood estimator of $p_{[i]}$ and hence we get

$$\hat{p}_{RSS} = \frac{1}{s} \sum_{i=1}^{s} \frac{Z_i}{m} = \overline{X}_{RSS}, \ \text{say,} \qquad (5)$$

where $\overline{X}_{RSS}$ is the overall mean of the ranked set sample.

Let $Y_1, Y_2, \cdots, Y_n$ are the observations drawn according to SRS design and $\hat{p}_{SRS}$ be an unbiased estimator of p, then the corresponding unbiased estimator of p can be obtained, as

$$\hat{p}_{SRS} = \sum_{i=1}^{n} \frac{y_i}{n} \qquad (6)$$

4. Effect of imperfect ranking

The discussion done in the previous sections are based on the assumption that the ranking procedure produces the correct order statistics. But, a perfect ranking

mechanism is very rare in practice and hence some error in judgment ranking is inevitable. So it is necessary to study how much robust the proposed procedure is against when the rankings are not perfect. Estimation of p under perfect and imperfect unbalanced RSS [13, 16, 17] are discussed.

Let $X_{[i]}$ and $X_{(i)}$ denote, respectively, the ith judgment order and i^{th} true order statistic while a set of s units are ordered. In the presence of ranking error, $X_{[i]}$ is not necessarily equal to $X_{(i)}$. Let π_{ij} denote the probability that the ith judgment order statistic actually have the true rank j, for $i = 1(1)s, j = 1(1)s$. Assume that π_{ij}s satisfy the conditions

$$0 \leq \pi_{ij} \leq 1, \quad \pi_{ij} = \pi_{ji}, \forall (i,j), \quad \sum_{j=1}^{s} \pi_{ij} = 1, \forall i, \quad \sum_{i=1}^{s} \pi_{ij} = 1, \forall j.$$

That is, $\boldsymbol{\pi} = (\pi_{ij})$ is a doubly stochastic symmetric matrix of order $s \times s$. Under this assumption the distribution of $X_{[i]}$ would be changed to Binomial $(1, p_{[i]}^*)$, for each $i = 1(1)s$, where

$$p_{[i]}^* = \sum_{j=1}^{s} \pi_{ij} p_{[j]}.$$

Equivalently,

$$\boldsymbol{p}^* = \boldsymbol{\pi} \boldsymbol{p}$$

where $\boldsymbol{p} = \left(p_{[1]}, p_{[2]}, \cdots, p_{[s]}\right)'$ and $\boldsymbol{p}^* = \left(p_{[1]}^*, p_{[2]}^*, \cdots, p_{[s]}^*\right)'$. For the present purpose we take the particular form of $\boldsymbol{\pi}$ as in [18, 19],

$$\boldsymbol{\pi} = \begin{pmatrix} \rho & \delta & \delta & \dots & \delta \\ & \rho & \delta & \dots & \delta \\ & & \rho & \dots & \delta \\ & & & \ddots & \vdots \\ & & & & \rho \end{pmatrix}, \quad 0 < \rho < 1, \quad \delta = \frac{1-\rho}{s-1}.$$

Here "$\rho = 1$" corresponds to the case of perfect ranking. Under the above probability model for imperfect ranking some consequential facts are justified below.

1. $\sum_{i=1}^{s} p_{[i]}^* = sp$, that is, the present imperfect ranking mechanism is consistent

Justification:

$$\sum_{i=1}^{s} p_{[i]}^* = \sum_{i=1}^{s} \sum_{j=1}^{s} \pi_{ij} p_{[j]} = \sum_{j=1}^{s} \left(\sum_{i=1}^{s} \pi_{ij} \right) p_{[j]} = \sum_{j=1}^{s} p_{[j]} = sp$$

and hence

$$\sum_{i=1}^{s} \left(1 - p_{[i]}^*\right) = s(1-p),$$

so that

$$\frac{1}{s}\sum_{i=1}^{s} P\big(X_{[i]} = x\big) = P(X = x), \ for \ x = 0, 1.$$

2. For equal sample size, i.e., $n = ms$

$$Variance\big(\hat{p}_{RSS}\big) \leq Variance\big(\hat{p}_{SRS}\big)$$

Justification: For all $i = 1, 2, \cdots, s$, setting

$$a_i = \sum_{j=1}^{s} \pi_{ij} p_{[j]}, \quad b_i = \sum_{j=1}^{s} \pi_{ij}\big(1 - p_{[j]}\big),$$

one can get under ranked set sampling with the presence of ranking error

$$\begin{aligned}
\text{Variance } (\hat{p}) &= \frac{1}{ms^2}\sum_{i=1}^{s} p_{[i]}^*\big(1 - p_{[i]}^*\big) \\
&= \frac{1}{ms^2}\sum_{i=1}^{s} \left(\sum_{j=1}^{s} \pi_{ij} p_{[j]}\right)\left(\sum_{j=1}^{s} \pi_{ij}\big(1 - p_{[j]}\big)\right) \\
&= \frac{1}{ns}\sum_{i=1}^{s} a_i b_i, \quad \text{say,}
\end{aligned}$$

Also, for the above choice of $\boldsymbol{\pi}$-matrix, it is verified that

$$a_i = \delta sp + (\rho - \delta)p_{[i]} \ \text{ and } \ b_i = \delta s(1 - p) + (\rho - \delta)\big(1 - p_{[i]}\big). \qquad (7)$$

Now, from the definition of $p_{[i]}$s, we argue that $\big\{p_{[i]}\big\}_{i=1(1)s}$ is a non-decreasing sequence and subsequently, among two sequences $\{a_i\}_{i=1(1)s}$ and $\{b_i\}_{i=1(1)s}$, one is non-decreasing and the other is non-increasing. So, from Chebyshev's inequality for a_is and b_is we have

$$\frac{1}{s}\sum_{i=1}^{s} a_i b_i \leq \left\{\frac{1}{s}\sum_{i=1}^{s} a_i\right\}\left\{\frac{1}{s}\sum_{i=1}^{s} b_i\right\}.$$

As we know that the variance of $\hat{p}$ in SRS is $\frac{p(1-p)}{n}$, the required justification follows from the fact that

$$\sum_{i=1}^{s} a_i = sp, \quad \sum_{i=1}^{s} b_i = s(1 - p).$$

The justification in case of perfect ranking follows automatically by taking $\rho = 1$ in the above proof.

5. Comparison between $\hat{p}_{RSS}$ and $\hat{p}_{SRS}$

It is easy to argue that the ML estimates $\hat{p}_{[1]}, \hat{p}_{[2]}, \cdots, \hat{p}_{[s]}$ are statistically independent as the variables $Z_i's$ are independently distributed. Again, substituting the

value $\frac{Z_i}{m}$ of $\hat{p}_{[i]}$ in Eq. (5), the estimate $\hat{p}_{RSS}$ can be shown to be identical with the overall mean of the given ranked set sample. It is also readily verified that $\hat{p}_{SRS}$ is an unbiased estimator of p.

The comparison of effectiveness and efficiency of the estimators based on simple random samples and ranked set samples, is obtained on the basis of criteria *viz.,* relative precision (RP) and relative saving (RS). The expressions for RP, RS and MSE of the estimators are described below as

$$RP = \frac{E(\hat{p}_{SRS} - p)^2}{E(\hat{p}_{RSS} - p)^2} = \frac{V(\hat{p}_{SRS})}{V(\hat{p}_{RSS})} \tag{8}$$

$$RS = \frac{V(\hat{p}_{SRS}) - V(\hat{p}_{RSS})}{V(\hat{p}_{SRS})}. \tag{9}$$

6. Result and discussion

Table 1 shows the estimates of proportion of measles immunized children in Assam, under SRS and RSS are very different but RSS based estimates 0.80 and 0.92 are very close to Census report for Assam (2012) [20] true value, which is 0.84 (rural) and 0.90 (urban), and has less variability than the SRS estimator and are very distinct from the true values. The estimate based on RSS is found to be 58% and 142%, for slum and non-slum region, respectively, more precision than that of SRS. Here, smaller the value of ρ represents higher will be the ranking error in RSS. The performance of estimates even in imperfect situation as compare to SRS, for different choices of the ranking error probability $\rho = 0.2, 0.6, 0.9$, shows the estimates based on RSS is 48%, 50% and 56% for slum region, and 139%, 140% and 141% for non-slum, respectively, more precision as compare to SRS. RSS also shows a saving of 37%(59%) under perfect and a minimum of 32%(58%) under imperfect in slum (non-slum) as compare to SRS.

Region	Criteria	n	SRS	s	m	RSS perfect	RSS imperfect		
							$\rho = 0.2$	$\rho = 0.6$	$\rho = 0.9$
Slum	Estimate	250	0.64	4	65	0.80	0.80	0.80	0.80
	Variance		0.92			0.58	0.62	0.62	0.59
	RP					1.58	1.48	1.50	1.56
	RS					36.83	32.29	33.27	35.70
Non-slum	Estimate	250	0.74	4	60	0.92	0.92	0.92	0.92
	Variance		0.76			0.32	0.32	0.32	0.32
	RP					2.42	2.39	2.40	2.41
	RS					58.60	58.22	58.30	58.51

Table 1.
Estimate of proportion of immunized children (p), variance (in 10^{-3}), relative precision and relative saving (in %), in different regions of Guwahati under SRS, RSS perfect and imperfect procedures.

7. Conclusion

The present study revealed that RSS based estimates in both of perfect and imperfect situations, performs better than SRS based estimates. It should also be

emphasized in context of estimation of proportion of measles immunization coverage in slum and non-slum region of Assam, RSS based estimates for different choices of accuracy (ρ) are not only more accurate but are more precise and efficient than the SRS procedure, and also suggest that the procedure of RSS is better than the classical SRS. Therefore, based on the obtained results one can recommend to adopt RSS procedure in epidemiological application and in other health related studies so that it will help in planning to build a healthy and disease free environment.

Acknowledgements

Author details

Vivek Verma[1]* and Dilip C. Nath[2]

1 Department of Neurology, All India Institute of Medical Sciences (AIIMS), New Delhi, India

2 Administration, Assam University, Silchar, Assam, India

*Address all correspondence to: viv_verma456@yahoo.com

IntechOpen

References

[1] Orenstein WA, Perry RT, Halsey NA. The clinical significance of measles: A review. The Journal of Infectious Diseases. 2004;**189**(Suppl 1):S4-S16

[2] Lahariya C. A brief history of vaccines & vaccination in India. Indian Journal of Medical Research. 2014;**139**: 491-511

[3] Hoshaw-Woodard S. Description and Comparison of the Methods of Cluster Sampling and Lot Quality Assurance Sampling to Assess Immunization Coverage. Geneva: Department of Vaccines and Biologicals, World Health Organization; 2001

[4] Datta A, Mog C, Das S, Datta S. A cross-sectional study to assess the immunization coverage and vaccine dropout rates among 12 to 23 months old children in a rural area of Tripura. International Journal. 2017;**6**(2):394

[5] Singh J, Jain DC, Sharma RS, Verghese T. Evaluation of immunization coverage by lot quality assurance sampling compared with 30-cluster sampling in a primary health Centre in India. Bulletin of the World Health Organization. 1996;**74**(3):269

[6] Murhekar MV, Kamaraj P, Kanagasabai K, Elavarasu G, Rajasekar TD, Boopathi K, et al. Coverage of childhood vaccination among children aged 12–23 months, Tamil Nadu, 2015, India. Indian Journal of Medical Research. 2017;**145**(3):377

[7] McIntyre GA. A method for unbiased selective sampling, using ranked sets. Australian Journal of Agricultural Research. 1952;**3**(4):385-390

[8] Terpstra JT, Miller ZA. Exact inference for a population proportion based on a ranked set sample. Communications in Statistics-

Simulation and Computation. 2006; **35**(1):19-26

[9] Kvam PH. Ranked set sampling based on binary water quality data with covariates. Journal of Agricultural, Biological, and Environmental Statistics. 2003;**8**(3):271-279

[10] Chen H. Alternative ranked set sample estimators for the variance of a sample proportion. Applied Statistics Research Progress. 2008;**35**

[11] Chen H, Stasny EA, Wolfe DA. Ranked set sampling for efficient estimation of a population proportion. Statistics in Medicine. 2005;**24**(21): 3319-3329

[12] Chen H, Stasny EA, Wolfe DA. Improved procedures for estimation of disease prevalence using ranked set sampling. Biometrical Journal. 2007; **49**(4):530-538

[13] Chen H, Stasny AE, Wolfe AD, MacEachern NS. Unbalanced ranked set sampling for estimating a population proportion under imperfect rankings. Communications in Statistics: Theory and Methods. 2009;**38**(12):2116-2125

[14] Verma V, Nath CD, Das R. Bayesian bounds for population proportion under ranked set sampling. Communications in Statistics-Simulation and Computation. 2019;**48**(2):478-493

[15] Verma V, Das R, Nath CD. Bayesian Cramer-Rao lower bound of variances under ranked set sampling. Materials Today Proceedings. 2018;**5**(1): 1747-1758

[16] Lacayo H, Neerchal NK, Sinha BK. Ranked set sampling from a dichotomous population. Journal of Applied Statistical Science. 2002;**11**(1): 83-90

[17] Chen H, Stasny EA, Wolfe DA. Unbalanced ranked set sampling for estimating a population proportion. Biometrics. 2006;**62**(1):150-158

[18] Terpstra JT, Nelson EJ. Optimal rank set sampling estimates for a population proportion. Journal of Statistical Planning and Inference. 2005;**127**(1):309-321

[19] Hatefi A, Jozani MJ. Fisher information in different types of perfect and imperfect ranked set samples from finite mixture models. Journal of Multivariate Analysis. 2013;**119**:16-31

[20] Registrar General of India. Annual Health Survey 2011-12 FACT SHEET Assam. New Delhi: Ministry of Health and Family Welfare, Department of Family Welfare; 2012. Available at: http://www.censusindia.gov.in/vital_sta tistics/AHSBulletins/AHS_Factsheets_ 2011_12/Assam _Factsheet_2011–12.pdf