Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/050116

Modernization of the Indian Decennial Census: An Illustration of Vaccination Coverage for Validity of Estimates

Dilip C Nath and Bhushita Patowari*

Department of Statistics, Gauhati University, Guwahati-781014, Assam, India

Received: 14 Sep. 2015, Revised: 7 Feb. 2016, Accepted: 11 Feb. 2016
Published online: 1 Mar. 2016

Abstract: Census is a complete and systematic enumeration of members of a given population in a particular region or geographical area. It has variety of advantages but also suffers from a number of difficulties. In this regard an effort has been made to test whether sample survey provides sufficient information as that of complete enumeration or not. The data has been taken from the survey Comparison of Two Survey Methodologies to Estimates Total Vaccination Coverage" sponsored by Indian Council of Medical Research (ICMR), New Delhi. It has been observed that sample estimate gives fairly good estimate of population parameter.

Keywords: complete enumeration, sample survey, immunization

1 Introduction

The census is older than the Chinese, Egyptian, Greek and Roman civilizations, dating back to the Babylonians in 4000 BC who used a census as an essential guide. From around 2,500 BC the Egyptians used censuses to work out the scale of the labour force they would need to build their pyramids. China too, began to take censuses around this time [1]. The Romans conducted censuses every five years, calling upon every man and his family to return to his place of birth to be counted in order to keep track of the population. The census played a crucial role in the administration of the people of an expanding Roman Empire, and was used to determine taxes. It provided a register of citizens and their property from which their duties and privileges could be listed [1]. Censuses are executed differently around the world. The regularity of the census can vary from every year to every ten years to being completely irregular. The information gathered also changes by country [2].

The decennial census is a count of the entire population that occurs once every ten years. The decennial censuses are planned and conducted with a high degree of professional competence. Indian Census is the single largest source of a variety of statistical information on different characteristics of the people of India. It is the most credible source of information on demography, economic activity, literacy and education, housing and household amenities, urbanisation, fertility and mortality, scheduled castes and scheduled tribes, language, religion, migration, disability and many other socio-cultural and demographic data [3].

The census provides the nation with essential data that go well beyond a simple population count. This information helps the Central and State Governments in planning and formulation of various policies [4]. The obligation for the Central Government to conduct a census or decennial enumeration of the population for the purpose of apportioning the Parliament is part of the Constitution. Besides, the delimitation/ reservation of Constituencies - Parliamentary/ Assembly/ Panchayats and other Local Bodies is also done on the basis of the demographic data thrown up by the Census.

With a history of more than 130 years, this reliable, time tested exercise has been bringing out a veritable wealth of statistics every 10 years, beginning from 1872 when the first census was conducted in India non-synchronously in different parts [3]. Conducting population census in a country like India, with great diversity of physical features, is undisputedly the biggest administrative exercise of peace time. The wealth of information collected through census on houses, amenities available to the households, socio economic and cultural characteristics of the population makes Indian Census the richest

* Corresponding author e-mail: dilipc.nath@gmail.com

and the only source for planners, research scholars, administrators and other data users. The planning and execution of Indian Census is challenging and fascinating [5].

In a study Kish (1979) mentioned that methods of survey sampling have a large literature. The methods are oriented to design for distinct and separate surveys, but those methods have general applicability. Smaller and more concentrated samples may suffice for difficult data, for national statistics and for large domains [6]. Waksberg (1968) discusses some aspects of survey sampling to speed up the processing of census results. In that study it is also mentioned that a number of countries, generally the more industrialized ones, have extended the use of sampling to the basic collection of data, to improve timeliness and reduce costs [7]. Davern et al. (2007) evaluated the historical microdata samples of the IPUMS project from 1850-1930 in order to determine the impact of sample design on standard error estimates and how to apply modern standard error estimation software to historical census samples. They examined the impact of complex sample designs on standard error estimates [8]. Census and survey data had been used to estimate poverty and inequality for Small Areas (Tarozzi and Deaton, 2008) [9]. Various problems and convenience of sample survey and complete enumeration has been discussed (Kish and Verma (1986), Waksberg and Hanson (1965), Ruggles and Menard. (1995), Davern et al. (2009)) [10,11,1,12,13].

2 Utility of Census Data

As part of an integrated programme of data collection, the population census is the primary source of basic national population data required for administrative purposes and for many aspects of economic and social research and planning. The value of the census results is increased if they can be employed together with the results of other investigations, as in the use of the census data as a base of benchmark for current statistics. The usefulness of the census is also enhanced if it can furnish the information needed for conducting other statistical investigations. It can, for example, provide a statistical frame for other censuses and sample surveys. The purpose of a continuing programme of data collection can best be served, therefore, if the relationship between the population census and other statistical investigations is considered when census planning is under way and if provision is made for facilitating the use of the census and its results in connexion with intercensal sample surveys, with continuous population registers, with other types of censuses and with civil registration and vital statistics, and with labour force, educational and similar statistics. The use of consistent concepts and definitions throughout an integrated programme of data collection is essential if the advantages of these relationships are to be fully realized [14].

3 2011 Census

The recently concluded Census 2011 is the 15th National Census of the Country in the unbroken series since 1872 and the seventh after Independence. It is remarkable that the great historical tradition of conducting a Census has been maintained in spite of several adversities like wars, epidemics, natural calamities, political unrest etc. Very few countries in the world can boast of such a glorious tradition [3].

4 Problems with total enumeration

Some striking problems have emerged over the past several decades. The 2011 Indian census had been completed during the year 2012. The Office of the Registrar General of India (ORGI) has started releasing data in phases. It will take another 2-3 years to get many tabulated data and by the time the ORGI will be busy for planning for next (2021) Census enumeration. The cost of Census taking in such a large country as India is Enormous [3]. The cost of Census 2011 has been estimated at Rs 2200 crores. A total of 2.7 million functionaries worked in the conduct of the census. The census schedules were canvassed in 16 languages. A total of 340 million schedules were printed [4]. The costs of taking the census have escalated sharply, even after allowing for inflation and population growth. A major cause of the cost increase was due to costs associated with the long form the content of the census. Also non-sampling error is there in complete enumeration. It is well known fact that larger the size of the sample smaller the sampling error but larger the non-sampling error. Kish (1979) has discussed the relative advantages and disadvantages of samples and censuses together with the variety of ways for using samples to good advantage in connection with censuses [6].

4.1 A few basic questions to ponder

- Is it worth to continue trying to count every last person with traditional census methods of physical enumeration?

- Is it possible to improve the accuracy of the census count with respect to its most important attributes by supplementing a reduced intensity of traditional enumeration with statistical estimates of the number and characteristics of those not directly enumerated?
- Is the use statistical estimation for completing the count, will help to achieve substantial cost savings in the next census, even as accuracy is being improved?
- Is dropping the long form in the next decennial census and substituting a continuous monthly survey to obtain relevant data will be sufficient for planning?

4.2 Task

- First task is to investigate whether and to what extent various types of essential data can best be collected by the decennial census or by other means.
- Second task is to consider and recommend the most cost-effective methods of conducting the census and otherwise collecting census-type data.

4.3 Recommendations

- Identify, other than government, who are census data users?
- To what extent Governments (Central and States) are using the Census data.
- Identify: (i) Items required for Complete enumeration e.g., House Listing, Population Count etc. (ii) Items those can be estimated by Statistical Methodology based on Sample data e.g., Fertility, Mortality, Migration etc.

5 Objective

The main objective of the study is to ascertain whether estimates based on sample survey can provide valid estimate of the population parameter or not. Here in this paper this has been illustrated through an example of vaccination coverage in a metropolitan city Guwahati, Assam, India.

6 Illustrated Example

Let us illustrate with an example whether sample survey can provide reliable estimate of parameter or not. This example is based on child vaccination coverage particularly Hepatitis B (at birth) vaccine. For this study data has been taken from the survey “Comparison of Two Survey Methodologies to Estimates Total Vaccination Coverage” sponsored by Indian Council of Medical Research (ICMR), New Delhi which has carried out in Guwahati City, Assam, India for the period January to October 2011.

7 Methods

The data set for this analysis is primary data and the study area is Guwahati, the capital city of Assam, India. It has been collected by using two stage (30×30) cluster sampling and systematic random sampling. The questionnaire has been developed to collect household information and information of the eligible members regarding vaccination. All children aged between 6 months to 5 years have been considered as the subject of the study population. The vaccination coverage Hepatitis B (at birth) is considered. According to Guwahati Municipal Corporation (GMC) the study area consists of 60 numbers of municipal wards and from that 30 wards are selected randomly. After random selection of the municipal wards of the study area 30 units from each ward are selected. For this purpose first household is selected where an eligible child is found in a particular area of a selected ward and thereafter each consecutive household has been selected where eligible members have found until 30 units (eligible children) are completed. After completion of two stage (30×30) cluster survey sampling systematic sampling is carried out in the same ward. For this purpose first household is selected randomly from the households in a particular lane (where survey is already carried out under two stage (30×30) cluster sampling) and thereafter each household at an interval of 10 households are selected till the required number (30 numbers of eligible children) is not completed. Both These procedures have been carried out in each selected ward. Altogether 1800 numbers of sampling units have been collected (900 in each sampling). Information of immunization is collected from record of vaccination cards.

8 Statistical Analysis

First of all it has been observed how many household have to visit in each ward to cover 30 units of sample under two stage cluster sampling. It is found that to cover 30 units of sample (eligible households) 105 households have to visit in a particular ward (ward number 2) (Table 1). On the basis of that number (i.e. 105) how many household will be eligible household among the recorded (as per GMC) household in that ward has been generated. It is found that 453 household will be eligible household in ward number 2. Similarly for all ward eligible household has been generated together with combined one. This constitutes the study population and from that population numbers of vaccinated children against Hepatitis B (at birth) have been obtained for each ward separately. After that dividing this number of vaccinated children against Hepatitis B (at birth) by respective population of each ward population proportion has been obtained as P. We have generated relevant data for all eligible Households. Also we have generated the population close-to-reality from already drawn sample.

Here Hepatitis B (at birth) vaccine coverage under both sampling methods has been estimated. To compare sample proportion with population proportion the null hypothesis is

$H_0 : P = P_0$

Against the alternative

$H_1 : P \neq P_0$

The test statistic (for single proportion) is given by

$$Z = \frac{p - P_0}{SE(p)} \tag{1}$$

where

$$SE(p) = \sqrt{\frac{N-n}{N-1} . \frac{PQ}{n}} \tag{2}$$

and 95 % Confidence Interval of P is

$$p - 1.96SE(p) \leqslant P \leqslant p + 1.96SE(p) \tag{3}$$

where p is the sample proportion and P_0 is population proportion.

9 Results and discussion

Estimated sample and population proportion of Hepatitis B (at birth) vaccine coverage is given in Table 2. An important result is that sample proportion of combined estimate of hepatitis B (at birth) vaccine coverage for two stage cluster sampling is 0.58 and 0.56 for systematic sampling whereas it is 0.59 for population proportion. They are very near to each other and it can be said that these estimates are almost same. Table 3 gives values of Z-statistic and confidence interval for Hepatitis B (at birth) vaccine coverage. From the Z-statistic it is found that there is no significant differences between the estimates of hepatitis B (at birth) vaccine coverage for the two stage cluster sample proportion and population proportions. Similarly for systematic sample proportion and estimated population proportion it is seen that Z-statistics are insignificant except for the ward number 43 and we may accept the null hypothesis that sample proportions and population proportions are equal. In case of two stage cluster sampling minimum and maximum differences of confidence intervals are 0.22 and 0.35 respectively whereas for systematic sampling these differences are 0.21 and 0.35 respectively. It is seen that the all the differences of confidence intervals (CI) are almost same for the two sampling methods.

Ericksen (1973) presented a new method of postcensal estimation in which the symptomatic information is combined with sample data by means of a regression format [15]. Kish and Verma (1986) studied uses of censuses for improving samples and of samples for improving censuses, and propose a method for cumulating data from rolling (or rotating) periodic (weekly, monthly or quarterly) samples specifically designed to cover the population in detail over designed spans (annual and quinquennial). They also compared sample survey to censuses [10]. This current work may open up possibilities of new areas for researchers. It may also help in different regional level, state level or country level to implement different policy or scheme for which total enumeration is required. Instead of total enumeration sample estimate will reduce time, cost, manpower and related difficulties as survey sampling can be designed to obtain broad areas of complex data which is deep in content.

10 Conclusion

From the study it can be concluded that sample estimates also give us reliable estimates of population parameters and in this study it is population proportion of Hepatitis B (at birth) vaccination coverage.

Appendix

Table 1: Household covered in each ward under two stage (30 $\times$ 30) cluster sampling

Sl. No.	Ward No.	Number of household visited to complete 30 units in two stage (30 $\times$ 30) cluster sample	Total household (as per GMC record)
1	2	105	1585
2	4	215	570
3	5	133	3094
4	11	127	2768
5	12	105	5243
6	15	163	3450
7	17	216	1990
8	18	122	1971
9	24	130	2062
10	25	149	4185
11	26	123	2655
12	33	139	1329
13	35	108	2950
14	36	126	2721
15	37	84	2511
16	38	123	1723
17	40	137	1509
18	42	181	4027
19	43	168	3891
20	46	166	1330
21	47	168	1125
22	48	112	2397
23	50	147	2084
24	51	188	5140
25	53	118	2073
26	54	177	2835
27	55	125	2747
28	57	167	3855
29	59	235	4664
30	60	192	4674
Total Household visited		4449	83158

Table 2: Estimated sample and population proportions of Hepatitis B (at birth) vaccine coverage

Sl. No.	Ward No.	No.of children vaccinated (30 × 30 cluster)	Total no.of children (30 × 30 cluster)	Estimated sample proportion (30 × 30 cluster) P_1	No.of children vaccinated (systematic sample)	Total no.of children (systematic sample)	Estimated sample proportion (systematic sample) P_2	Generated no.of population	Generated no.of children vaccinated (population)	Population proportion P
1	2	5	30	0.17	4	30	0.13	453	49	0.11
2	4	21	30	0.70	20	30	0.67	80	55	0.69
3	5	7	30	0.25	14	30	0.48	654	228	0.35
4	11	28	30	0.93	25	30	0.83	654	538	0.82
5	12	16	30	0.53	17	30	0.57	1498	825	0.55
6	15	18	30	0.60	18	30	0.60	635	403	0.63
7	17	14	30	0.47	13	30	0.43	276	115	0.42
8	18	21	30	0.70	21	30	0.70	485	315	0.65
9	24	5	30	0.17	8	30	0.27	476	91	0.19
10	25	20	30	0.67	14	30	0.47	843	497	0.59
11	26	19	30	0.63	20	30	0.67	648	482	0.74
12	33	22	30	0.73	24	30	0.80	287	236	0.82
13	35	12	30	0.40	16	30	0.53	819	394	0.48
14	36	19	30	0.63	17	30	0.57	648	467	0.72
15	37	19	30	0.63	11	30	0.37	897	431	0.48
16	38	13	30	0.43	15	30	0.50	420	207	0.49
17	40	16	30	0.53	11	30	0.37	330	162	0.49
18	42	22	30	0.73	19	30	0.63	667	459	0.69
19	43	20	30	0.67	15	30	0.50	695	473	0.68
20	46	16	30	0.53	16	30	0.53	240	134	0.56
21	47	17	30	0.57	16	30	0.53	201	121	0.60
22	48	18	30	0.60	16	30	0.53	642	371	0.58
23	50	24	30	0.80	22	30	0.73	425	284	0.67
24	51	19	30	0.63	20	30	0.67	820	571	0.70
25	53	20	30	0.67	19	30	0.66	526	413	0.79
26	54	19	30	0.63	15	30	0.50	481	294	0.61
27	55	11	30	0.37	16	30	0.53	659	275	0.42
28	57	21	30	0.70	20	30	0.67	693	448	0.65
29	59	17	30	0.57	19	30	0.63	595	359	0.60
30	60	25	30	0.83	20	30	0.67	730	543	0.74
31	combined	524	900	0.58	501	900	0.56	17477	10240	0.59

Table 3: Z-values and confidence interval (CI) for Hepatitis B (at birth) vaccine coverage

Sl. No.	Ward No.	Z-value (for two stage cluster sample)	CI	Z value (for systematic sample)	CI
1	2	1.13	(0.06,0.28)	0.40	(0.02,0.24)
2	4	0.19	(0.57, 0.83)	0.26	(0.54,0.80)
3	5	1.16	(0.08, 0.42)	1.54	(0.31,0.65)
4	11	1.57	(0.80, 1.06)	0.11	(0.70,0.96)
5	12	0.23	(0.35, 0.71)	0.21	(0.39,0.75)
6	15	0.40	(0.43, 0.77)	0.40	(0.43,0.77)
7	17	0.63	(0.30, 0.64)	0.16	(0.26,0.60)
8	18	0.60	(0.53, 0.87)	0.60	(0.53,0.87)
9	24	0.30	(0.03, 0.31)	1.13	(0.13,0.41)
10	25	0.91	(0.50, 0.84)	1.35	(0.30,0.64)
11	26	1.46	(0.48, 0.78)	0.95	(0.52,0.82)
12	33	1.40	(0.60, 0.86)	0.34	(0.67,0.93)
13	35	0.90	(0.22, 0.58)	0.55	(0.35,0.71)
14	36	1.13	(0.47, 0.79)	1.88	(0.41,0.73)
15	37	1.67	(0.45, 0.81)	1.23	(0.19,0.55)
16	38	0.71	(0.26, 0.60)	0.08	(0.33,0.67)
17	40	0.45	(0.36, 0.70)	1.39	(0.20,0.54)
18	42	0.51	(0.57, 0.89)	0.70	(0.47,0.79)
19	43	0.13	(0.51, 0.83)	2.17*	(0.34,0.66)
20	46	0.33	(0.36, 0.70)	0.33	(0.36,0.70)
21	47	0.39	(0.41, 0.73)	0.87	(0.37,0.69)
22	48	0.25	(0.43, 0.77)	0.54	(0.36,0.70)
23	50	1.59	(0.64, 0.96)	0.74	(0.57,0.89)
24	51	0.80	(0.47, 0.79)	0.32	(0.51,0.83)
25	53	1.58	(0.53, 0.81)	1.72	(0.52,0.80)
26	54	0.22	(0.46, 0.80)	1.29	(0.33,0.67)
27	55	0.54	(0.20, 0.54)	1.28	(0.36,0.70)
28	57	0.63	(0.53, 0.87)	0.28	(0.50,0.84)
29	59	0.38	(0.40, 0.74)	0.31	(0.46,0.80)
30	60	1.10	(0.68, 0.98)	0.95	(0.52,0.82)
31	combined	0.07	(0.40, 0.76)	0.29	(0.38,0.74)

*significant at 5% probability level

References

[1] Census history. Source: http://www.ons.gov.uk/ons/guide-method/census/2011/how-our-census-works/about-censuses/census-history/census-taking-in-the-ancient-world/index.html.
[2] Census-New World Encyclopedia. Source:http://www.newworldencyclopedia.org/entry/Census.
[3] Census of India 2011–A Story of Innovations. Source:pib.nic.in/newsite/efeatures.aspx?relid=74556.
[4] Census of India 2011-Archieves-spotlite-national portal of India. Source:http://www.archive.india.gov.in/spotlight/spotlight_archive.php?id=80.
[5] History of Census in India. Source:http://www.censusindia.gov.in/Ad_Campaign/drop_in_articles/05-History_of_Census_in_India.pdf.
[6] L. Kish. Samples and Censuses. International Statistical Review **47(2)**, 99-109 (1979).
[7] J. Waksberg. The role of sampling in population census: Its effect on timeliness and accuracy. Demography **5(1)**, 362-373 (1968).
[8] M. Davern, S. Ruggles, T. Swenson and J.M. Oakes (2007). Drawing Statistical Inferences from Historical Census Data, Working Paper No. 2007-02. Source:http://www.pop.umn.edu/sites/www.pop.umn.edu/files/Working%20Paper%202007-02.pdf.
[9] A. Tarozzi and A. Deaton (2008). Using Census and Survey Data to Estimate Poverty and Inequality for Small Areas. Source:http://public.econ.duke.edu/
[10] L. Kish and V. Verma. Complete census and samples. Journal of Official Statistics **2(4)**, 381-395 (1986).
[11] J. Waksberg and R.H. Hanson. Sampling applications in censuses of population and housing. Washington: US Census Bureau, Technical paper no 13 (1965).
[12] S. Ruggles and R.R. Menard (1995). Public Use Microdata Sample of the 1880 United States Census of Population: User's Guide and Technical Documentation. Source:https://usa.ipums.org/usa/resources/codebooks/1880_PUMS_codebook.pdf.
[13] M. Davern, S. Ruggles, T. Swenson, J.T. Alexander and J.M. Oakes. Drawing Statistical Inferences from Historical Census Data, 1850?1950. Demography **46(3)**, :589–603 (2009).
[14] Census of India-Objects, essential features and utility of census. Source:http://censusindia.gov.in/Data_Products/Library/Indian_perceptive_link/Census_Objectives_link/censusobjectives.htm.
[15] E.P. Ericksen. A method for combining sample survey data and symptomatic indicators to obtain population estimates for local areas. Demography **10(2)**, 137-160 (1973).

Dilip C. Nath is currently Professor of Statistics in Gauhati University. Dr. Nath had been trained at different national and international Universities/Institutes viz., Gauhati University, Delhi University, Banaras Hindu University, Duke University, USA, University of Washington, USA and Max Plank Institute for Demographic Research, Germany. At Gauhati University he teaches several courses in Statistics, Demography and Biostatistics in the department of Statistics. He has guided successfully 20 Ph.D scholars so far to get their degree in Statistics on different topics like: fertility, reproductive health, aging and health of the elderly population, disease burden. So far Dr. Nath has collaborated with 50+ national and international scholars. Dr. Nath had been awarded the Rockefeller Foundation Fellowship (1991-93) and the Andrew Mellon Fellowship (1997-98).

Bhushita Patowari is currently pursuing Ph.D in the Department of Statistics, Gauhati University, Guwahati. Her main research interests are survey sampling and epidemiology.