Demography India (2018) ISSN: 0970-454X
Vol. 47, Issue: 1, pp: 71-78

Research Article

Estimation of Standard Error of Demographic and Health Parameters for NFHS-4

A. K. Tiwari[1] & Shivam Mishra[2,*]

Abstract

A sample survey involves always sampling errors along with non-sampling errors. Non-sampling errors are the results of mistakes made during execution of data collection and data processing and it is impossible to avoid and tough to evaluate statistically. On the other hand, sampling error is a degree of the variability which is not known exactly, it can be estimated from the survey outcomes. It is usually measured in terms of the standard error for a specific statistic, but it is quite complex. In this paper an attempt has been made to develop a simple procedure to estimate standard error of some demographic and health parameters for NFHS-4 data.

Introduction

Data is the pillar of research and survey is a research tool used to collect the data from people to obtain information and facts on various topics. Survey has a variety of purposes and can be conducted in various ways subject to the methodology preferred and the objectives to be achieved. Basically, a survey comprises asking people for information through questionnaire/observations. There are two technique of collecting survey data about the population- Census and sampling. Census refers to the well-organized procedure of gathering, recording and analyzing information concerning all the members of the population. It is an official and complete enumeration of the population. While sampling is the process of selecting sample from the population, so carefully chosen to represent the characteristics of the larger group or population. This method is used for statistical testing, where it is not possible to consider all members or observations, as the population size is very large. As statistical inferences are based on the observations taken from the sample, the selection of the appropriate representative sample is utmost important. On the basis of the data collected from the representative samples, the conclusion is drawn for the whole population.

A sample survey contains a certain amount of sampling errors along with non-sampling errors. Non-sampling errors are the results of mistakes made during execution of data collection and data processing. Non-sampling errors are impossible to avoid and tough to evaluate statistically. On the other hand, sampling error is a degree of the variability among all possible samples. Although the degree of variability is not known exactly, it can be estimated from the survey outcomes. It is usually measured in terms of the standard error for a specific statistic which is the square root of the variance. The standard error is normally used to calculate confidence intervals of various statistics apart from providing a clearer picture of possible variations in the statistics over possible samples and thus provides an account of the reliability of an estimate from the sample. Obviously, it is highly dependent on the sampling design used and the nature of the statistic (estimate) itself.

About National Family Health Survey (NFHS)

The National Family Health Survey (NFHS) programme, commenced in the early 1990s, has seemed as are lying source of data to monitor and evaluate the progress of family planning,

[1]Associate Professor, Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi-221005
[2]Research Scholar, Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi-221005
*Corresponding Author: Shivam Mishra, Email: shivam.stats@gmail.com

reproductive and child health programmes in India and its states and union territories. The first survey (NFHS-1) was conducted in 1992-93 succeeded by NFHS-2 in 1998-99 and NFHS-3 in 2005-06. The fourth phase of these nationwide surveys - National Family Health Survey (NFHS-4) was conducted in 2015-16. Like previous surveys, NFHS-4 was designed to provide estimates of important demographic parameters like family welfare, reproductive, maternal and child health, infant and child mortality, nutrition and other health related problems. In this survey, a uniform sample design, questionnaires, field processes and procedures for biomarker measurements are used all over the country to ensure the highest probable data quality and to enable comparability across the states.NFHS-4 is visualized to help managers and policymakers to set standards and to examine progress over time in India. MoreoverNFHS-4 data will help to identify the effectiveness of current programmes and the requirement of new programmes in the area of family welfare, health and nutrition and other related areas.

Sample Design of NFHS-4

The National Family Health Survey (NFHS) has used a complicated sampling design for the selection of households (given in appendix). The design of NFHS is a multistage stratified random sampling. The sample for NFHS-4 is a stratified sample selected in two stages from the sampling frame (i.e. all Census Enumeration Blocks for Census 2011). For stratification, each district was split into urban and rural areas. In the first stage of sample selection, Primary Sampling Units (PSUs) were selected. This selection is implemented with probability proportional to the PSU size along with independent selection in each sampling stratum with the sample allocation from the Municipal Corporation Offices (MCOs) slums list and the census sampling frame. For the state module, a subsample of about one-third of the sample PSUs was randomly selected with the sample allocation. A household listing was in each of the selected PSUs before the main survey. In this process, a list was prepared by visiting each of the selected PSUs and registering all residential households found in the PSU. During this operation, the selected PSUs with an estimated number of households greater than 300 were segmented into segments of 100-150 households. Finally, an NFHS-4 cluster is a PSU or part of a PSU. The resultant list of households operated as the sampling frame for the selection of households in the second stage. In the second stage of selection, a fixed number of 22 households per cluster (PSU or part of PSU) was selected with an equal probability systematic selection. No replacements and no changes of the pre-selected households were acceptable in the executing stages in order to avoid bias. Consequently, considering the complicated sampling design, the computation of standard errors of various estimates is cumbersome and the used formula is complex also. To calculate sampling errors of different parameters for NFHS-4, the formula is programmed in SAS software. For estimation of variance of survey estimates (i.e. means and proportions), Taylor linearization method is used in this procedure. In this method, any proportion or mean is treated as a ratio estimate, $r = y/x$, where y represents the total sample value for variable y, and x represents the total number of cases in the group or subgroup under consideration. Standard Error of r is computed by using the formula given below, where the standard error is the square root of the variance:

$$SE^2(r) = \text{var}(r) = \frac{1-f}{x^2}\sum_{h=1}^{H}\left[\frac{m_h}{m_h-1}\left(\sum_{i=1}^{m_h} z_{hi}^2 - \frac{z_h^2}{m_h}\right)\right]$$

in which $z_{hi} = y_{hi} - rx_{hi}$, *and* $z_h = y_h - rx_h$

Where h represents the stratum, which varies from 1 to H,

m_h is the total number of clusters selected in the h_{th} stratum,

y_{hi} is the sum of the weighted values of variable y in the i_{th} cluster in the h_{th} stratum,

x_{hi} is the sum of the weighted number of cases in the i_{th} cluster in the h_{th} stratum, and

f is the overall sampling fraction, which is so small that it is ignored.

In addition to the standard error, the design effect (DEFT) for each estimate is also computed, which is defined as the ratio between the standard error using the given sample design and the standard error that would result if a simple random sample had been used. Obviously, the formula is complex and requires a lot of computational steps. This is mainly because of the complicated sampling design. Tiwari (2002) has estimated the standard error of age specific fertility rate for NFHS-1 by assuming the sampling design as simple random sample. If the sample of respondents had been selected as a simple random sample, it would have been possible to use direct formulas for calculating sampling errors. The standard error of any proportion (proportion of sampling units possessing any specified attribute) is Standard Error $(p) = \sqrt{PQ/n}$

where p is the observed proportion in the sample, P is actual proportion in the population, $Q = 1 - P$ and n is the sample size. An estimate of the standard error can be taken as

$$\widehat{SE}(p) = \sqrt{pq/n} \qquad \text{where} \quad q = 1 - p$$

In this paper, authors have tried to estimate standard error of few demographic and health parameters (currently using any contraceptive method, unmet need for family planning, want no more children, women with any anemia and have heard of HIV/AIDS) identified in NFHS-4 by defining the interrelationship between the actual standard error of parameters computed from the basis of the sampling design used in NFHS and the standard error of the parameters assuming the sampling design to be simple random sampling using regression technique. The standard error of parameters along with base population is published in the NFHS-4 reports for the different states. The regression lines between the standard error on the basis of sampling design of NFHS and standard error corresponding to simple random sampling have been drawn for different parameters treating each state as the unit of observation where standard error corresponding to simple random sample is treated as predictor variable.

Estimation of Standard Error for Demographic and Health Parameters

For the estimation of standard error, authors have selected following demographic and health parameters-

A. Demographic Parameters

1. Currently using any contraceptive method
2. Unmet need for family planning
3. Want no more children

B. Health Parameters

1. Women with any anemia
2. Have heard of HIV/AIDS

A. Standard Error of Demographic Parameters

1. **Currently using any contraceptive method:** To estimate the standard error for parameter 'currently using any contraceptive method', currently married women of age 15-49 years are considered as base population in NFHS-4. Estimate of the standard error for this parameter is computed using given formula

$$\widehat{SE}(p_c) = \sqrt{\frac{p_c q_c}{n_{cm}}}$$

Where, p_c is the proportion of currently married women of age 15-49 years who are using any contraceptive method

q_c is the proportion of currently married women of age 15-49 years who are not using any contraceptive method i.e. $1-p_c$

n_{cm} is the total number of currently married women of age 15-49 years.

The regression equation for the parameter 'currently using any contraceptive method' is found

$$Y = 1.84X - 0.0005$$

Where Y is the standard error of currently using any contraceptive method and X is the computed standard error $\left(\sqrt{p_c q_c / n_{cm}}\right)$ assuming simple random sampling. The value of R^2 is found 0.93 . From the above result it is interesting to note that in the regression line, the value of the intercept (a) is almost zero showing a proportional relationship between Y and X i.e.

$$Y = 1.84X$$

In addition the design effect (DEFT) for currently using any contraceptive method is 1.84, which is defined as the ratio of the standard error using the given sampling design to the standard error that would result if a simple random sample had been used.

2. **Unmet Need for Family Planning:** For estimating the standard error for parameter another demographic variable 'unmet need for family planning', base population is currently married women of age 15-49 years are considered as base population in NFHS-4. Estimate of the standard error for the parameter is computed from given formula

$$\widehat{SE}(p_u) = \sqrt{\frac{p_u q_u}{n_{cm}}}$$

Where, p_u is the proportion of currently married women of age 15-49 years with unmet need for family planning

q_u is $1-p_u$

n_{cm} is the total number of currently married women of age 15-49 years.

The equation of regression line for the parameter 'unmet need for family planning' is found

$$Y = 1.29X + 0.0004$$

Where, Y is the standard error of unmet need for family planningand X is the computed standard error $\left(\sqrt{p_u q_u / n_{cm}}\right)$ assuming simple random sampling. The value of R^2 is found 0.95. Above result shows that in the regression line, the value of the intercept (a) is almost zero which shows a proportional relationship between Y and X i.e.

$$Y = 1.29X$$

From the above equation, it is concluded that the design effect (DEFT) for unmet need for family planning is 1.29

3. **Want no more children:** To estimate the standard error for parameter 'want no more children', currently married women of age 15-49 years are labeled as base population in NFHS-4 report. Estimate of the standard error for this parameter is calculated by using following formula

$$\widehat{SE}(p_w) = \sqrt{\frac{p_w q_w}{n_{cm}}}$$

Where, p_w is the proportion of currently married women of age 15-49 years who want no more children and who are sterilized (either respondent or partner)

q_c is $1 - p_c$

n_{cm} is the total number of currently married women of age 15-49 years.

The equation of regression line for the parameter 'want no more children' is found

$$Y = 1.32X + 0.0004$$

Where Y is the standard error of parameter 'want no more children' and X is the estimated standard error $\left(\sqrt{p_w q_w / n_{cm}}\right)$ assuming simple random sampling. The value of R^2 is found 0.91. From the regression line, it is concluded that the value of the intercept (a) is almost zero and shows a proportional relationship between Y and X i.e.

$$Y = 1.32X$$

In addition the design effect (DEFT) for 'want no more children' is 1.32

B. Standard Error of Health Parameters

1. **Women with any anemia:** In NFHS-4, women of age 15-49 years with an anemia test are treated as base population for estimating standard error of the health parameter 'women with any anemia'. Estimation of the standard error for this parameter is worked out by using given formula

$$\widehat{SE}(p_a) = \sqrt{\frac{p_a q_a}{n_{fa}}}$$

Where, p_a is the proportion of women of age 15-49 years with an anemia test who have recognized anemic,

q_a isthe proportion of women of age 15-49 years with an anemia test who have not recognized anemic i.e. $1 - p_a$,

n_{fa} is the total number of women of age 15-49 years with an anemia test.

The regression equation for the parameter 'women with any anemia' is found

$$Y = 1.24X + 0.0015$$

Where Y is the standard error of health parameter 'women with any anemia' and X is the computed standard error $\left(\sqrt{p_a q_a / n_{fa}}\right)$ assuming simple random sampling. The value of R^2 is found 0.78 and design effect is 1.24

2. **Have heard of HIV/AIDS:** To estimate the standard error for parameter 'have heard of HIV/AIDS', all women of age 15-49 years are considered as base population in NFHS-4. Estimate of the standard error for this parameter is computed using given formula

$$\widehat{SE}(p_h) = \sqrt{\frac{p_h q_h}{n_f}}$$

Where, p_h is the proportion of women of age 15-49 years who have heard about HIV/AIDS,

q_h is the proportion of women of age 15-49 years who have not heard about HIV/AIDS i.e. $1-p_h$,

n_f is the total number of women of age 15-49 years.

The equation of regression line for the parameter 'currently using any contraceptive method' is found

$$Y = 1.64X - 0.0004$$

Where Y is the standard error of parameter 'have heard of HIV/AIDS' and X is the estimated standard error $\left(\sqrt{p_h q_h / n_f}\right)$ assuming sampling design as simple random sampling. The value of R^2 is found0.76 . From the above regression equation, it is found that the value of the intercept (a) is almost zero presenting a proportional relationship between Y and X i.e.

$$Y = 1.64X$$

Here 1.64 may be defined as design effect (DEFT) for the parameter as the ratio of the standard error using the given sampling design to the standard error that would result if sampling design is simple random sample.

Results and Applications

Table 1 shows the observed and estimated standard error of selected demographic parameters for all states (excluding Arunachal Pradesh and Uttarakhand) of country. For the demographic parameter 'unmet need for family planning', result indicates that the estimated standard error is almost equal to observed standard error for 15 states (Assam, Bihar, Gujarat, Jammu & Kashmir, Kerala, Madhya Pradesh, Maharashtra, Manipur, Meghalaya, Nagaland, Odisha, Punjab, Tamil Nadu, Uttar Pradesh and West Bengal) and the little difference was found between estimated and observed standard error for remaining states.

From table 1, for the demographic parameter 'currently using any contraceptive method', it indicates that the estimated standard error is same as observed standard error for many states (Bihar, Jammu & Kashmir, Madhya Pradesh, Manipur, Sikkim, Tamil Nadu, Tripura and Uttar Pradesh) while for other states, the difference between estimated and observed standard error is less. For the demographic parameter 'want no more children', conclusion is drawn that the estimated standard error is equivalent to observed standard error for 14 states (Assam, Bihar, Chhattisgarh, Goa, Gujarat, Haryana, Himachal Pradesh, Kerala, Madhya Pradesh, Manipur, Odisha, Punjab, Tamil Nadu and West Bengal) and for other states, the estimated and observed standard error are quite close to each other.

The estimated and observed standard error for health indicators are given in table 2. From table 2, for the health parameter 'women with any anemia', no difference is found between estimated standard error and observed standard error for 8 states (Bihar, Jammu & Kashmir, Madhya Pradesh, Manipur, Sikkim, Tamil Nadu, Tripura and Uttar Pradesh) while for 14 states, the difference between estimated and observed standard error is 0.001 . The third and fourth column of table 2 shows the observed and estimated standard error for variable 'have heard of HIV/AIDS'.The standard error for other variable may be calculated using the same technique.

In any sample survey, when we report the mean of sample data, it is good to present some indication of the reliability of the data. This can be done using confidence interval. It provides a range that estimates the true population value for a statistic. Confidence interval is defined using standard error. Normally in large scale sample survey, researchers consider complicated sampling design to

represent population. In complicated design, the computation of standard error of various estimates is very difficult. This technique may be used for estimation of the standard error of the different parameters.

References

Tiwari, Abhay Kumar (2002). *A Study of Fertility through Mathematical Models* (Ph.D. thesis). Banaras Hindu University, Varanasi, India.

International Institute for Population Sciences (IIPS) and ICF. 2017. National Family Health Survey (NFHS-4), 2015-16: India. Mumbai: IIPS.

Table 1. Observed and Estimated Standard Error of Demographic Parameters for NFHS-4

S. No.	State	Currently using any Contraceptive Method		Unmet need for Family Planning		Want no more children	
		Observed SE	Estimated SE	Observed SE	Estimated SE	Observed SE	Estimated SE
1	Andhra Pradesh	0.006	0.009	0.003	0.004	0.006	0.007
2	Assam	0.007	0.006	0.004	0.004	0.005	0.005
3	Bihar	0.004	0.004	0.003	0.003	0.004	0.004
4	Chhattisgarh	0.005	0.006	0.003	0.004	0.005	0.005
5	Goa	0.024	0.023	0.015	0.016	0.020	0.020
6	Gujarat	0.008	0.007	0.004	0.004	0.005	0.005
7	Haryana	0.007	0.006	0.004	0.003	0.005	0.005
8	Himachal Pradesh	0.012	0.010	0.007	0.006	0.007	0.007
9	Jammu and Kashmir	0.007	0.007	0.004	0.004	0.005	0.006
10	Jharkhand	0.005	0.006	0.003	0.004	0.004	0.005
11	Karnataka	0.007	0.006	0.004	0.003	0.008	0.005
12	Kerala	0.009	0.010	0.005	0.005	0.008	0.008
13	Madhya Pradesh	0.004	0.004	0.002	0.002	0.003	0.003
14	Maharashtra	0.006	0.005	0.003	0.003	0.005	0.004
15	Manipur	0.008	0.008	0.007	0.007	0.007	0.007
16	Meghalaya	0.011	0.010	0.007	0.007	0.010	0.009
17	Mizoram	0.012	0.010	0.008	0.007	0.011	0.008
18	Nagaland	0.008	0.010	0.007	0.007	0.010	0.009
19	Odisha	0.006	0.005	0.003	0.003	0.004	0.004
20	Punjab	0.005	0.006	0.003	0.003	0.005	0.005
21	Rajasthan	0.004	0.005	0.002	0.003	0.003	0.004
22	Sikkim	0.015	0.015	0.010	0.009	0.010	0.011
23	Tamil Nadu	0.006	0.006	0.003	0.003	0.005	0.005
24	Tripura	0.014	0.014	0.008	0.007	0.009	0.011
25	Uttar Pradesh	0.003	0.003	0.002	0.002	0.002	0.003
26	West Bengal	0.006	0.007	0.003	0.003	0.005	0.005
27	Telangana	0.010	0.012	0.004	0.005	0.008	0.009

Observed Standard Error of Demographic Parameters are taken from NFHS-4 Reports for States

Table 2. Observed and Estimated Standard Error of Health Parameters for NFHS-4

S. No.	State	Women with any Anemia		Have heard of HIV/AIDS	
		Observed SE	Estimated SE	Observed SE	Estimated SE
1	Andhra Pradesh	0.006	0.009	0.007	0.009
2	Assam	0.007	0.006	0.014	0.010
3	Bihar	0.004	0.004	0.010	0.009
4	Chhattisgarh	0.005	0.006	0.011	0.010
5	Goa	0.024	0.023	0.006	0.010
6	Gujarat	0.008	0.007	0.010	0.010
7	Haryana	0.007	0.006	0.009	0.010
8	Himachal Pradesh	0.012	0.010	0.007	0.010
9	Jammu and Kashmir	0.007	0.007	0.006	0.005
10	Jharkhand	0.005	0.006	0.012	0.011
11	Karnataka	0.007	0.006	0.012	0.010
12	Kerala	0.009	0.010	0.003	0.005
13	Madhya Pradesh	0.004	0.004	0.009	0.008
14	Maharashtra	0.006	0.005	0.009	0.008
15	Manipur	0.008	0.008	0.002	0.004
16	Meghalaya	0.011	0.010	0.023	0.018
17	Mizoram	0.012	0.010	0.007	0.006
18	Nagaland	0.008	0.010	0.013	0.012
19	Odisha	0.006	0.005	0.008	0.006
20	Punjab	0.005	0.006	0.004	0.004
21	Rajasthan	0.004	0.005	0.010	0.009
22	Sikkim	0.015	0.015	0.014	0.016
23	Tamil Nadu	0.006	0.006	0.004	0.005
24	Tripura	0.014	0.014	0.016	0.020
25	Uttar Pradesh	0.003	0.003	0.007	0.006
26	West Bengal	0.006	0.007	0.016	0.013
27	Telangana	0.010	0.012	0.011	0.014

Observed Standard Error of Health Parameters are taken from NFHS-4 Reports for States