Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v8n1p112

Two stage sampling design for estimation of total fertility rate: with an illustration for slum dweller married woman

By Dilip C. Nath, Aditi Das Deka

Published: 26 April 2015

Electronic Journal of Applied Statistical Analysis
Vol. 08, Issue 01, April 2015, 112-121
DOI: 10.1285/i20705948v8n1p112

Two stage sampling design for estimation of total fertility rate: with an illustration for slum dweller married woman

Dilip C. Nath, Aditi Das Deka*

Department of Statistics, Gauhati University, Assam,India

Published: 26 April 2015

Two Stage Sampling procedure has been found to be much more acceptable for demographic surveys owing to the lower cost involved and also due to the fact that it gives rise to a good representative sample. As such, it has been found to be much acceptable for estimating total fertility rate (TFR), an important measure of fertility. The objective of this paper is to estimate the TFR in the slum pockets of Guwahati city in Assam under two stage sampling technique and to bring out a comparative analysis of the TFR calculated under the proposed sampling design and under the Simple Random Sampling (SRS) method.

The study reveals that the overall fertility scenario in the slum areas of Guwahati, as far as TFR is concerned, when calculated under two stage sampling technique is above the replacement level. However, the TFR calculated by SRS method is found to be almost equal to the replacement level, implying under estimation of TFR when calculated by this method. Moreover, it has been seen that the adjusted TFRs derived after adjusting the biases marginally increases in case of SRS technique and decreases in case of two stage sampling technique. The MSE s are found to be higher in case of conventional technique, implying that the sampling error involved in the conventional method is more as compared to two stage sampling technique. We may thus consider two stage sampling method as a better technique for estimating TFR.

*Corresponding author: aditidasdeka@gmail.com

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

keywords: Simple random sampling, cluster sampling, two stage sampling, age-specific fertility rate, mean square error (MSE).

1 Introduction

The Total Fertility Rate (TFR) is perhaps the most commonly used standardized fertility measure because it is ideal for comparative purposes and is a comprehensive summary measure readily understood, at least as a general concept. Associated with total fertility rate is the concept of replacement rate. The replacement rate is the number of children each woman needs to have to maintain current population levels or what is known as zero population growth for her and her partner. Ultimately, this is the only thing that matters in determining long-term population growth. In most developing countries, demographers are faced with the problem of estimation of demographic rates either due to limited data or due to varying data quality. Methods for estimating fertility rates in developing countries over time, based on different data sources, and assessing its uncertainty are lacking. Most of the literature on quality of fertility data has been focused on the development of indirect estimation techniques[Brass et al. (1964); Brass (1968, 1996); Feeney (2008); Trussell (1975); Hill et al. (1983)]. These techniques deal with biases that are caused by recall lapse errors in retrospective estimates of fertility rates [Becker and Mahmud (1984); Potter (1977); Pullum and Stokes (1997); Som (1973)]. The indirect estimation methods are correct for reporting biases by reconciling information from recent fertility (in the last year or years) with lifetime fertility. They are typically based on one data source only, and underlying assumptions can give problems with respect to the accuracy of the indirect estimates [Moultrie and Dorrington (2008)]. The methods deal with bias only, and not the difference in the error variance of the observations.

2 Rationale of the Study and its objectives

Direct selection of sampling units from the population under study as in Simple Random Sampling (SRS) is not recommended for demographic surveys owing to huge expenses on one hand and non acceptability of results on the other. Even cluster sampling is not a favourable alternative in such situations as it may not give rise to representative sample. Moreover, for calculation of demographic estimates like TFR, etc., it is advisable to apply mathematical formula based on the sampling techniques used instead of relying on the conventional formulae, which is based on SRS technique. The present study is based on Two Stage Sampling procedure and it has been found to be much acceptable for estimating total fertility rate (TFR), an important measure of fertility [Rust and Rao (1996)[.The objective of this paper is to estimate the TFR under Two Stage Sampling Technique with recourse to Probability Proportional to Size (PPS) sampling for selecting the number of subunits from each first stage units (fsu). Through this study an attempt has been made to bring out a comparative analysis of the TFR calculated under the proposed sampling method and under SRS method.

3 Study Population and Sampling Technique

There are 25 slum pockets in the Greater Guwahati area, the total number of households in these pockets being recorded as 24,603 with a total population of 15,6906 out of which 7,1995 are females [6]. From these 25 slum pockets, 5 slum pockets have been selected for carrying out the survey. The female respondents who were the residents of the said slums and who were married and were in the age group 15-49 years were interviewed personally and necessary information were collected. For our study, we have resorted to two stage sampling procedure, so as to obtain a better estimate of fertility for the population under consideration. In our study, the first stage units (fsu) i.e., the slums are selected by SRS technique and the second stage units (ssu) i.e., the households or rather the individuals (married women in the age group 15- 49) have been selected using Linear Systematic Sampling Technique. As regards the number of subunits to be selected from each fsu, recourse has been taken to Probability Proportional to Size. In the first stage, with application of SRS technique 5 slum pockets have been selected for carrying out the survey.
The selected slums are:
1. Islampur
2. Rajabari
3. Athgaon
4. Hatigaon (Sijubari)
5. Santipur (East)

In the second stage, for selection of households or rather the individuals (i.e.,ssu) from each of the selected slum pockets we have made use of Linear Systematic Sampling Technique. This technique requires that the ratio of the number of ssu's in the ith fsu (M_i) to the selected number of ssu (m_i), is an integer and not a fractional number i.e. $M_i/m_i = k$. Although, in our case k is a fractional number, yet we have resorted to Linear Systematic Sampling owing to the fact that we have approximated the fractional numbers to its nearest integer

Table 1: fsu size and number of ssu in the sample

S.no	Name of Slum Pocket	fsu size(M_i)	No. of selected ssu (m_i)	Ratio (M_i/m_i)
1	Islampur	1896	315	6.01
2	Rajabari	520	95	6.11
3	Athgaon	1275	210	6.07
4	Hatigaon (Sijubari)	3492	576	6.06
5	Santipur(East)	995	154	6.06
	Total	8178	1350	

Since the integer k is approximately equal to 6 in all the cases, the possible systematic samples were drawn in such a manner that every 6th household is selected until the desired numbers of individuals (ssu) are selected.

4 Estimation of Age- Specific Fertility Rate (ASFR) and Total Fertility Rate

4.1 Simple random sampling (SRS) technique

Conventionally, TFR is calculated using the SRS technique and is obtained by cumulating the ASFRs of female residents in the reproductive age group (15- 49), of a specified geographic area (nation, state, county, etc.) during a specified time period (usually a calendar year) multiplied by h, the age interval usually taken as 5 years. Mathematically, $TFR = ()xh$, where ASFRis each age-specific birth rate defined as $ASFR =_h a_x =_h b_x/_h f_x$, where $_hb_x$ is the number of live births to mothers in the age group $(x, x + h)$ and $_hf_x$ is the number of resident women in that age group.

4.2 Two Stage Sampling technique

We shall consider the two stage sampling technique with SRSWR at first stage and Linear Systematic Sampling at second stage. We define:
= number of fsu in the population.
M_i = number of households in the i^{th} fsu.
= number of fsu drawn out of N under SRSWR.
m_i = number of households drawn from M_i = households in the i^{th} selected fsu, which are selected using systematic sampling and whose sizes are determined by PPS. $_hF_x$ = total number of women in the reproductive age range i.e.,$15 - 49$ years.
$_hB_x$ = = total number of births to these women in the reference 2 year period preceeding the survey.
$_hf_x$ = total number of women in the age group $(x, x + h)$ in the sample.
$_hb_x$ = total number of births by the women in the age group (x, x+h) in the sample.
$_hf_{x_{ij}}$ = total number of women in the age group $(x, x + h)$ in the j^{th} household of the i^{th} fsu.
$_hb_{x_{ij}}$ =total number of births by the women in the age group (x, x+h) in the jth household of the i^{th} fsu.
P_i = inclusion probability of the i^{th} fsu in the sample.

The Age Specific Fertility Rate (ASFR) for the study population of age group $(x, x+h)$ is defined as:

$$_hA_x =_h B_x/_hF_x \tag{1}$$

Now, our main objective is to estimate the Total Fertility Rate (TFR) of the study population. Thus:

$$TFR = h \sum {_hA_x} \tag{2}$$

Then under the proposed sampling design we have:

$$ {}_hb_x = \sum_{i=1}^{n} \{M_{i.h}\bar{b}_{xi}\}/P_i \tag{3} $$

where ${}_h\bar{b}_x$ is an unbiased estimate of ${}_h\bar{B}_x$ and where:

$$ {}_hB_x = \sum_{i=1}^{N} \{M_{i.h}\bar{B}_{xi}\}/P_i \tag{4} $$

$$ {}_h\bar{B}_{xi} = \{\frac{1}{M_i}\} \sum_{i=1}^{M_i} {}_h\bar{b}_{x_{ij}} \tag{5} $$

$$ {}_h\bar{b}_{xi} = \{\frac{1}{m_i}\} \sum_{i=1}^{m_i} {}_hb_{x_{ij}} \tag{6} $$

and,

$$ {}_hf_x = \sum_{i=1}^{n} \{M_{i.h}\bar{f}_{xi}\}/P_i \tag{7} $$

where ${}_hf_x$ is an unbiased estimate of ${}_hF_x$ and where:

$$ {}_hF_x = \sum_{i=1}^{N} \{M_{i.h}\bar{F}_{xi}\}/P_i \tag{8} $$

$$ {}_h\bar{F}_{xi} = \{\frac{1}{M_i}\} \sum_{i=1}^{M_i} {}_hf_{x_{ij}} \tag{9} $$

$$ {}_h\bar{f}_{xi} = \{\frac{1}{m_i}\} \sum_{i=1}^{m_i} {}_h\bar{f}_{x_{ij}} \tag{10} $$

Then the sample estimates of ASFR and TFR are defined as:

$$ {}_ha_x =_h b_x/{}_hf_x \tag{11} $$

$$ TFR_{(est)} = h\sum {}_ha_x \tag{12} $$

5 Estimation of variance of TFR

5.1 Simple random sampling (SRS) technique

We know that under SRS sample mean is an unbiased estimate of population mean. Thus, we have:

$$E({}_hf_x) =_h \bar{F}_x \tag{13}$$

$$E({}_hb_x) =_h \bar{B}_x \tag{14}$$

Where ${}_hf_x$ and ${}_hb_x$ are the sample estimates of ${}_h\bar{F}_x$ and ${}_h\bar{B}_x$ respectively. Using Δ method of expansion for ratio estimator [Murthy et al. (1967)],$O(n^{-2})$ approximations to the bias and the mean square of ${}_ha_x$, we have:

$${}_hB_x =_h A_x[\{V({}_h\bar{f}_x)/{}_h\bar{F}_x^2\} - Cov({}_h\bar{f}_{x.h}\bar{b}_x)/{}_h\bar{F}_{x.h}\bar{B}_x] \tag{15}$$

$${}_hB_x =_h A_x[\{CV({}_h\bar{f}_x)\}^2 - \rho_hf_{x,h}b_xCV({}_h\bar{f}_x)CV({}_h\bar{b}_x)] \tag{16}$$

and,

$$MSE({}_ha_x) =_h A_x^2[\{CV({}_h\bar{f}_x)^2 + \{CV({}_h\bar{b}_x)\}^2 - 2\rho_hf_{x,h}b_xCV({}_h\bar{f}_x)CV({}_h\bar{b}_x)] \tag{17}$$

where $\rho_{{}_hf_x,{}_hb_x}$ is the correlation co-efficient between ${}_hf_x$ and ${}_hb_x$ and $CV({}_h\bar{f}_x)$and $CV({}_h\bar{b}_x)$ are the respective coefficient of variation. Thus, we have:

$$\begin{gathered} E({}_ha_x -_h A_x)({}_ha_y -_h A_y) = \\ {}_hA_{xh}A_y[\rho_{{}_hf_x,{}_hb_x}CV({}_h\bar{f}_x)CV({}_h\bar{f}_y) + \rho_hf_{x,h}b_xCV({}_hf_x)CV({}_hf_y) - \\ \rho_hf_xCV({}_hfx)CV({}_h\bar{b}_y) - \bar{\rho}_{{}_hb_x,{}_hf_y}CV({}_hf_y)] \end{gathered} \tag{18}$$

Thus, the estimated adjusted TFR and its MSE are defined by-

$$TFR(adj) = \sum[{}_ha_[x - est.B({}_ha_x] = TFR_{SRS} - h\sum est.B({}_ha_x) \tag{19}$$

$$\begin{gathered} Est.MSEofTFR(adj) = h^2[\sum_i Est/MSE({}_ha_x) + \\ \sum_i\sum_j est.E({}_ha_x -_h A_x)({}_ha_y -_h A_y) - \sum_i est/B({}_ha_x)]^2 \end{gathered} \tag{20}$$

Estimators $est.B({}_ha_x)$, $est.MSE({}_ha_x)$ and $est.E({}_ha_x -_h A_x)({}_ha_y -_h A_y)$ of $B({}_ha_x)$, $MSE({}_ha_x)$ and $E({}_ha_x -_h A_x)({}_ha_y -_h A_y)$ respectively can be obtained replacing population parameters by the corresponding sample unbiased estimators in their respective equations.

5.2 Two Stage Sampling technique

Let $E_i, i = 1, 2$ denote the expectation over the i^{th} stage sampling and $d_i = 1$, if i^{th} fsu is included in the sample $= 0$, otherwise It is to be noted that

$E_2[{}_hb_{xi}]/m_i = {}_h \bar{B}_{xi}$, as systematic sample mean is an unbiased estimator of population mean. As a consequence, we get-

$$E({}_hb_x) = E_1[\sum M_i E_2\{{}_h\bar{b}_{x_i}\}/m_i]/P_i \tag{21}$$

$$E({}_hb_x) = {}_h B_x \tag{22}$$

showing that ${}_hb_x$is an unbiased estimate of ${}_hB_x$. Further, taking $V_i, i = 1, 2$ as the variance for the i^{th} stage sampling, by definition, we have-

$$V({}_hb_x) = V_1 E_2({}_hb_x) + E_1 V_2({}_hb_x) \tag{23}$$

$$\begin{aligned} V({}_hb_x) = \sum\sum(P_iP_j - P_{ij})[\{M_{i.h}\bar{B}_{xi}\}/P_i - M_{j.h}\bar{B}_{xj}\}/P_j\}]^2 + \\ \sum\{M_i^2/P_i\}S_h^2 b_{x_i}/m_i[1 + (m_i - 1)\rho_{hb_{xi}}] \end{aligned} \tag{24}$$

where $\rho_{{}_hb_{xi}}$ is the intra class correlation co-eff. Of births to the women in the age group $(x, x + h)$ in the i^{th} fsu and $S_h^2 b_{xi}$ is the population variance of the i^{th} fsu. To obtain an unbiased estimator of $V({}_hb_x)$, we note that -

$$\begin{aligned} E[\sum\sum(P_iP_j - P_{ij})\{M_{i,h}\bar{b}_{xi}/P_i - \{M_{j.h}\bar{b}_{xj}/P_j\}\}]^2 = V({}_hh_x) - \\ E[\sum M_i^2/P_i Est.V_2({}_hb_{xi}/m_i)] \end{aligned} \tag{25}$$

Therefore, an unbiased estimator of $V({}_hb_x)$ is of the form -

$$\begin{aligned} V({}_hh_x) = \sum\sum\{(P_iP_j - P_{ij})\}\{M_{i.h}\bar{b}_{xi}/P_i - M_{j.h}\bar{b}_{xj}/P_j\}^2 \\ + \sum M_i^2/P_i[\{1/m_i(m_i - 1)\} \sum_h \bar{b}x_{ij} -_h \bar{b}x_i] \end{aligned} \tag{26}$$

Similar results for women of age group $(x, x + h)$ can be obtained. As in the usual ratio estimator hax is a bias estimator of ${}_hA_x$. Thus, using the Δ - method of expansion Rust and Rao (1996), the first order approximation to the bias and mean square error of ax are obtained as -

$$B({}_ha_x) = {}_h A_x[\{CV({}_hf_x)\}^2 = \rho_{{}_hf_x,{}_hb_xCV({}_hf_xCV({}_hb_x))}] \tag{27}$$

and

$$MSE({}_ha_x) = {}_h A_x^2[\{CV({}_hf_x)\}^2 + \{CV({}_hb_x)\}^2 - \rho_{{}_hf_x,{}_hb_x}CV({}_hf_x)CV({}_hb_x)] \tag{28}$$

where $\rho_{hfx,hbx}$ is the correlation co-efficient between ${}_hf_x$ and ${}_hb_x$ and $CV({}_hf_x)$ and$CV({}_hb_x)$ are the coefficient of variation of ${}_hf_x$ and ${}_hb_x$ respectively. Equivalently,

$$\begin{aligned} E({}_ha_{xh}A_x)({}_ha_{yh}A_y) ={}_h A_{xh}A_y[\rho_{hfx,hbx}CV({}_hf_x)CV({}_hb_x) \\ +\rho_{hbx,hby}CV({}_hb_y) - \rho_h f_{x,h}\, b_yCV({}_hb_y)- \\ \rho_h b_{x,h}\, f_yCV({}_hb_x).CV({}_hf_y)] \end{aligned} \tag{29}$$

Estimators $est.B({}_ha_x)$, $est.MSE({}_ha_x)$ and $est.E({}_ha_x -_h A_x)({}_ha_y -_h A_y)$ of $B({}_ha_x)$, $MSE({}_ha_x)$ and $E({}_ha_x -_h A_x)$ $({}_ha_y -_h A_y)$ respectively can be obtained replacing population parameters by the corresponding sample unbiased estimators in their respective equations. Thus, the estimated adjusted TFR and its MSE are defined by-

$$TFR(adj) = \sum[{}_ha_x - est.B({}_ha_x)] \tag{30}$$

$$Est.MSeofTFR(adj) = \sum_i Est.MSE({}_ha_x) + \sum\sum est.E({}_ha_x -_h A_x)({}_ha_y -_h A_y)-$$

$$\sum est.B({}_ha_x)$$

^{2}(31)

Table 2: Comparison of ASFR and TFR under SRS method and under Two Stage Sampling Design along with Adjusted TFR and MSE of TFR

	Age specific fertility rate	
Age Group	SRS	Two Stage Sampling
15-19	0.0395	0.0224
20-24	0.1168	0.1139
25-29	0.1448	0.1887
30-34	0.0729	0.0985
35-39	0.0261	0.0185
40-44	0.0092	0.0043
45-49	0	0
Total Fertility Rate	2.047	2.2312
Adjusted TFR	2.061	2.1943
MSE of TFR	0.8372	0.6391

6 Conclusion

The various estimates of ASFRs for the different age groups and the TFRs as obtained under the two methods of sampling along with the adjusted TFR and the mean square error of the estimated TFR are shown in table 2. The ASFRs calculated for various age groups by the conventional (SRS) method and under Two Stage Sampling design gives slightly different values. It might be a reflection of the fact that in some of the slums, for some specific age groups, the number of births in the two year reference period is stated to be zero. In such cases, the ASFR obtained under the proposed sampling frame gives slightly lesser values than the ones obtained by the SRS method. Also it has been seen that the ASFRs calculated for the different age groups in the age interval 15-49 years with 5- yearly interval show an increasing trend with age up to 29 years. It is the highest for the age group $25 - 29(0.1887)$ years and then gradually decreases and becomes zero for the age group $45 - 49$ years. Moreover, the TFR calculated by the SRS method is found to be approximately equal to 2.047, which is almost equal to the replacement level. The TFR obtained under Two Stage Sampling design is found to be approximately equal to 2.2312, which is higher than the replacement level, implying under estimation of TFR by the conventional method. It has been seen that the adjusted TFRs derived after adjusting the biases marginally increases in case of SRS technique and decreases in case of Two Stage Sampling technique. The MSE s are found to be 0.8372 and 0.6391 respectively, implying that the sampling error involved in the SRS method is more as compared to Two Stage Sampling technique. Moreover, it is worth mentioning that researchers are often found to estimate various demographic measures using the formulae based on SRS technique, irrespective of the sampling technique used in their studies. This practice often provide with estimates which are not accurate. This fact has been reflected in our results as provided in TABLE 2, where TFR obtained under SRS technique involved more sampling error as compared to Two Stage Sampling technique. This is not because SRS doesnot provide with accurate results but because of the fact that the present study is based on Two Stage Sampling technique and formulae based on this technique have provided with more accurate results. We may thus infer that for estimation of TFR or any other demographic estimates it is advisable to apply formulae based on the sampling technique used in that particular study instead of the ones used in the conventional set up so as to obtain more accurate results.

References

Becker, S. and Mahmud, S. (1984). A validation study of backward and forward pregnancy histories in matlab bangladesh.

Brass, W. (1968). demography of tropical africa.

Brass, W. (1996). Demographic data analysis in less developed countries: 1946–1996. *Population Studies*, 50(3):451–467.

Brass, W., for Africa, U. N. E. C., et al. (1964). *Uses of census or survey data for the estimation of vital rates.* UN.

Feeney, G. (2008). A new interpretation of brass p/f ratio method applicable when fertility is declining.[electronic resource].

Hill, K., Zlotnik, H., Trussell, J., of International Economic, U. N. D., Division, S. A. P., et al. (1983). *Manual X: indirect techniques for demographic estimation.* UN.

Moultrie, T. and Dorrington, R. (2008). Sources of error and bias in methods of fertility estimation contingent on the p/f ratio in a time of declining fertility and rising mortality. *Demographic Research*, 19(46):1635–1662.

Murthy, M. N. et al. (1967). Sampling theory and methods. *Sampling theory and methods.*

Potter, J. E. (1977). Problems in using birth-history analysis to estimate trends in fertility. *Population Studies*, 31(2):335–364.

Pullum, T. W. and Stokes, S. L. (1997). Identifying and adjusting for recall error with application to fertility surveys. *Survey Measurement and Process Quality*, pages 711–732.

Rust, K. F. and Rao, J. (1996). Variance estimation for complex surveys using replication techniques. *Statistical methods in medical research*, 5(3):283–310.

Som, R. K. (1973). *Recall lapse in demographic enquiries.* Asia Pub. House.

Trussell, T. J. (1975). A re-estimation of the multiplying factors for the brass technique for determining childhood survivorship rates. *Population Studies*, 29(1):97–107.