Sankhyā : The Indian Journal of Statistics
1989, Volume **51**, Series B, Pt. 1, pp. 134-139.

NOTE

SAMPLING BIAS MSE AND CONFIDENCE INTERVALS FOR SOME FERTILITY MEASURES UNDER CLUSTER SAMPLING SCHEME

By A. K. BHATTACHARYYA
Indian Statistical Institute

SUMMARY. The bias of the demographic rate TFR under cluster sampling scheme has been investigated in this paper. Attempt has been made to estimate the sampling bias and provide corrected estimators for the above rate. An estimator for MSE of the estimated TFR has been proposed so that it can be used to give confidence interval for the above estimator. Similar investigation has been made for other demorgraphic rates as well.

0. Introduction

In a previous paper (Bhattacharyya and Mullick, 1987) we have shown that the demographic rates TFR, GRR and NRR are approximately unbiased under simple random sampling scheme. We also gave expression for the MSE of these rates and proposed estimators for MSE for using them in giving confidence interval for the estimators. But in practice most of the demographic surveys do not use simple random sampling for various practical reasons. Instead from cost consideration a cluster samplmg scheme may be preferred for the survey. In this paper we show by considering TFR that the usual fertility rates are biased upto third order of approximation under cluster sampling scheme. Since the sample estimates are biased, we propose estimators for these bias and give corrected or revised estimators. We also attempt to give estimators for the MSE of these corrected estimates which we need for giving confidence intervals for the revised estimates.

The practical aspect of this paper is that in many surveys, as for example fertility survey of a district say, we collect data from villages. Now generally a sample of villages is chosen and further sampling is done within villages. But as far as the household listing is concerned it is done for the complete village and information on the incidence of a birth last year is also collected for the whole village. Thus if a SRSWOR of villages is taken we can use the formulae given in this paper for estimating bias of the fertility rate find corrected estimates and also give confidence interval for these corrected estimates. To save space we will confine ourselves only to TFR. For other rates, we refer to the technical report (Bhattacharyya, 1987).

AMS (1980) *subject classification :* Primary, 62P25.

Key words and phrases : Total Fertility Rate, General Fertility Rate, Net Reproduction Rate.

1. Notations and definitions

Let there be a total of F females in the child bearing age group (w_1, w_2). These F females are divided into N clusters. The size of the i-th cluster is F_i. A simple random sample of n clusters is taken from these N clusters without replacement.

Let F_{ix} = number of females of age x in the i-th cluster.

${}_aF_{ix}$ = number of female in the age group $(x, x+a)$ in the i-th cluster.

B_{ix} = number of single births to the F_{ix} females.

${}_aB_{ix}$ = number of single births to the ${}_aF_{ix}$ females.

Let S denote the sum over the sample and Σ denote the sum over the population.

$$b_x = \underset{i}{S}\, B_{ix}$$

$${}_ab_x = \underset{i}{S}\, {}_aB_{ix}$$

$$f_x = \underset{i}{S}\, F_{ix}$$

$${}_af_x = \underset{i}{S}\, {}_aF_{ix}$$

$$i_x = b_x/f_x$$

$$I_x = \underset{i}{\Sigma}\, B_{ix}/\underset{i}{\Sigma}\, F_{ix} \qquad = B_x/F_x$$

$${}_aI_x = \underset{i}{\Sigma}\, {}_aB_{ix}/\underset{i}{\Sigma}\, {}_aF_{ix} \qquad = {}_aB_x/{}_aF_x$$

Note that i_x and ${}_ai_x$ are the usual estimators of I_x and ${}_aI_x$ are spectively. Also we have the sample estimate of TFR as

$$\widehat{TFR} = \Sigma\, i_x,\ \widehat{TFR}(G) = 5\, \Sigma_5 i_x.$$

Here G is a notation used for rates based on grouped data.

2. Some preliminary results

Here we give expressions for Bias $(i_x) = B(i_x)$, $MSE\ (i_x)$ and $E(i_x - I_x)\ (i_y - I_y)$ which we shall need later. We notice $E(cb_x) = B_x$ and $E(cf_x) = F_x$ where $c = \dfrac{N}{n}$.

Hence by expanding by Δ method for ratio estimators we get the following results by neglecting third and higher order terms

$$B(i_x) = I_x\, c^2 \left\{\frac{1}{F_x^2}\ V(f_x) - \frac{1}{F_x B_x}\ cov\ (b_x, f_x)\right\} \qquad \ldots \quad (2.1)$$

$$MSE(i_x) = I_x^2\, c^2 \left\{ \frac{1}{B_x^2}\ V(b_x) + \frac{1}{F_x^2}\ V(f_x) - \frac{2}{B_x F_x}\ cov\ (b_x, f_x)\right\} \qquad \ldots \quad (2.2)$$

$$E(i_x-I_x)(i_y-I_y) = I_x I_y c^2 \left\{ \frac{cov(b_x, b_y)}{B_x B_y} - \frac{cov(b_x, f_y)}{B_x F_y} - \frac{cov(f_x, b_y)}{F_x B_y} + \frac{cov(f_x, f_y)}{F_x F_y} \right\} \quad \ldots (2.3)$$

Substituting the expressions for variance and covariance terms we can get the expressions for $B(i_x)$, $MSE\ (i_x)$ and $E(i_x-I_x)(i_y-I_y)$. We find that $B(i_x)$ is non-zero. Hence we can give an estimator of $B(i_x)$ by estimating I_x by i_x and putting unbiassed estimators of $V(f_x)$, $cov(b_x f_x)$ and F_x^2 and $F_x B_x$ in (2.1). Similarly we can find estimators for $MSE(i_x)$ and $E(i_x-I_x)(i_y-I_y)$ in the usual manner. Since we shall use the estimator of Bias for getting corrected or revised estimators we explicitly give the expression for $\hat{B}(i_x)$, we also give the estimator of MSE and the estimator of the term of the form $E(i_x-I_x)(i_y-I_y)$.

3. Estimators of BIAS, MSE and $E(i_x-I_x)(i_y-I_y)$

$$\hat{B}(i_x) = i_x c^2 \left\{ \frac{\frac{(N-n)}{N} \underset{i}{S} F_{ix}^2 - \frac{(N-n)}{N(n-1)} \underset{i \neq j}{S} F_{ix} F_{jx}}{\frac{N}{n} \underset{i}{S} F_{ix}^2 + \frac{N(N-1)}{n(n-1)} \underset{i \neq j}{S} F_{ix} F_{jx}} - \frac{\frac{(N-n)}{N} \underset{i}{S} B_{ix} F_{ix} - \frac{(N-n)}{N(n-1)} \underset{i \neq j}{S} B_{ix} F_{jx}}{\frac{N}{n} \underset{i}{S} B_{ix} F_{ix} + \frac{N(N-1)}{n(n-1)} \underset{i \neq j}{S} B_{ix} F_{jx}} \right\} \quad \ldots (3.1)$$

$\hat{B}\ ({}_5i_x)$ can be obtained by replacing i_x by ${}_5i_x$; B_{ix} by ${}_5B_{ix}$

F_{ix} by ${}_5F_{ix}$ in (3.1).

Even if age groups are not of 5 years then also similar results hold.

$$\widehat{MSE}\,(i_x) = i_x^2\, c^2 \left\{ \frac{\frac{(N-n)}{N} \underset{i}{S} B_{ix}^2 - \frac{(N-n)}{N(n-1)} \underset{i \neq j}{S} B_{ix} B_{jx}}{\frac{N}{n} \underset{i}{S} B_{ix}^2 + \frac{N(N-1)}{n(n-1)} \underset{i \neq j}{S} B_{ix} B_{jx}} + \frac{\frac{(N-n)}{N} \underset{i}{S} F_{ix}^2 - \frac{(N-n)}{N(n-1)} \underset{i \neq j}{S} F_{ix} F_{jx}}{\frac{N}{n} \underset{i}{S} F_{ix}^2 + \frac{N(N-1)}{n(n-1)} \underset{i \neq j}{S} F_{ix} F_{jx}} - 2\, \frac{\frac{(N-n)}{N} \underset{i}{S} B_{ix} F_{ix} - \frac{(N-n)}{N(n-1)} \underset{i \neq j}{S} B_{ix} F_{jx}}{\frac{N}{n} \underset{i}{S} B_{ix} F_{ix} + \frac{N(N-1)}{n(n-1)} \underset{i \neq j}{S} B_{ix} F_{jx}} \right\} \quad \ldots (3.2)$$

It is not obvious that $\widehat{\text{MSE}}(i_x)$ is non-negative, hence it is advised that in such cases MSE be taken as zero.

$$\hat{E}(i_x-I_x)(i_y-I_y) = i_x i_y c^2 \left\{ \frac{\frac{(N-n)}{N} \underset{i}{S} B_{ix} B_{iy} - \frac{(N-n)}{N(n-1)} \underset{i\neq j}{S} B_{ix} B_{jy}}{\frac{N}{n} \underset{i}{S} B_{ix} B_{iy} + \frac{N(N-1)}{n(n-1)} \underset{i\neq j}{S} B_{ix} B_{jy}} \right.$$

$$- \frac{\frac{(N-n)}{N} \underset{j}{S} B_{ix} F_{iy} - \frac{N(N-n)}{N(n-1)} \underset{j\neq j}{S} B_{ix} F_{jy}}{\frac{N}{n} \underset{i}{S} B_{ix} F_{iy} + \frac{N(N-1)}{n(n-1)} \underset{i\neq j}{S} B_{ix} F_{jy}}$$

$$- \frac{\frac{(N-n)}{N} \underset{i}{S} F_{ix} B_{iy} - \frac{(N-n)}{N(n-1)} \underset{i\neq j}{S} F_{ix} B_{jy}}{\frac{N}{n} \underset{i}{S} F_{ix} B_{iy} + \frac{N(N-1)}{n(n-1)} \underset{i\neq j}{S} F_{ix} B_{jy}}$$

$$\left. + \frac{\frac{(N-n)}{N} \underset{i}{S} F_{ix} F_{iy} - \frac{(N-1)}{N(n-1)} \underset{i\neq j}{S} F_{ix} F_{jy}}{\frac{N}{n} \underset{i}{S} F_{ix} F_{iy} + \frac{N(N-1)}{n(n-1)} \underset{i\neq j}{S} F_{ix} F_{jy}} \right\} \quad \ldots \quad (3.3)$$

The details of these derivations are omitted here and are to be found in (Bhattacharyya, 1987).

4. Revised estimates of age specific fertility rates

Since $i_x - B(i_x)$ is an unbiased estimator of I_x we propose the estimator

$$\hat{I}_x = i_x - \hat{B}(i_x) \quad \ldots \quad (4.1)$$

as a revised estimator of I_x.

We note that

$$E(i_x - B(i_x) - I_x)^2 = MSE(i_x) - \{B(i_x)\}^2$$

Hence we take

$$\widehat{MSE}\,(\hat{I}_x) = M\hat{S}E(i_x) - \{\hat{B}(i_x)\}^2 \quad \ldots \quad (4.2)$$

5. Revised estimator of TER

We denote the revised estimator by subscript c i.e. Revised TFR is denoted by TFR_c

$$\text{Clearly } B(TFR) = \Sigma\, B(i_x) \quad \ldots \quad (5.1)$$

Hence $$TFR_c = TFR - \hat{B}(TFR) = TFR - \Sigma\, \hat{B}(i_x) \quad \ldots \quad (5.2)$$

For grouped data we get

$$TFR_c\,(G) = TFR(G) - 5.\Sigma\hat{B}({}_5i_x) \qquad \dots \quad (5.3)$$

6. Estimators of the MSE of the revised fertility rates

Since $E(i_x - B(i_x) - I_x)\,(i_y - B(i_y) - I_y) = E(i_x - I_x)\,(i_y - I_y) - B(i_x)\,B(i_y)$

we estimate

$$E(\hat{I}_x - I_x)\,(\hat{I}_y - I_y) \text{ by } \{\hat{E}(i_x - I_x)\,(i_y - I_y) - \hat{B}(i_x)\,\hat{B}(i_y)\}.$$

Now

$$MSE(TFR_c) = E(\sum_x \hat{I}_x - \sum_x I_x)^2$$

$$= \sum_x E(\hat{I}_x - I_x)^2 + 2 \sum_{x>y} E(\hat{I}_x - I_x)\,(\hat{I}_y - I_y)$$

$$= \sum_x MSE(\hat{I}_x) + 2 \sum_{x>y} E(\hat{I}_x - I_x)\,(\hat{I}_y - I_y).$$

Therefore we have the estimators of MSE as follows

$$\widehat{MSE}(TFR_c) = \sum_x \widehat{MSE}(\hat{I}_x) + 2 \sum_{x>y} \hat{E}(\hat{I}_x - I_x)\,(\hat{I}_y - I_y)$$

$$= \sum_x \widehat{MSE}(i_x) + 2 \sum_{x>y} \hat{E}(i_x - I_x)\,(i_y - I_y) - \{\Sigma\,\hat{B}(i_x)\}^2$$

$$= \sum_x \widehat{MSE}(i_x) + 2 \sum_{x>y} \hat{E}(i_x - I_x)\,(i_y - I_y) - \{B(TFR)\}^2. \qquad \dots \quad (6.1)$$

Similarly

$$\widehat{MSE}(TFR_c)\,(G) = 25. \quad \{\sum_x \widehat{MSE}({}_5i_x) + 2 \sum_{x>y} \hat{E}({}_5i_x - {}_5I_x)\,({}_5i_y - {}_5I_y\} - \{\hat{B}(TFR(G)\}^2 \qquad \dots \quad (6.2)$$

Since in most case grouping of age is 5 years we mentioned them. But if age grouping is not 5 years or even if they are not equal for all groups, results are similar and have not been explicitly given here.

Conclusion

In this work we have investigated the bias of fertility measures under culster sampling scheme. We have proposed corrected estimators and have tried to give estimators of their MSEs. We can use the standard probability inequalities of the form $P(\,|TFR_c - TFR|\, < t) > 1 - \dfrac{\widehat{MSE}\ (TFR_c)}{t^2}$ for giving confidence intervals.

Acknowledgement. The author is greatful to Dr. R. Bairagi for some helpful discussion, and to the referee for useful suggestions for revisions.

REFERENCES

BHATTACHZRYYA, A. K. and MULLICK, S. K. (1987) : On the bias and mean square error of some fertility measures. *Demography India*, **16**, No. 2, 288-294.

BHATTACHARYYA, A. K. (1987) : Sampling bias MSE and confidence intervals for some fertility measures under cluster sampling scheme. ***P.S.U. Publication*** No. 1, Indian Statistical Institute, Calcutta-700 035.

Paper received : *July*, 1987.
Revised : *August*, 1988.