Jour. Ind. Soc. Ag. Statistics
50(1), 1997 : 18–28

Estimation of Parametric Functions in Repeat Surveys

Jagbir Singh and O.P. Kathuria[1]
Indian Agricultural Statistics Research Institute, New Delhi-110012
(Received : October, 1994)

SUMMARY

In sampling on h-occasions, Minimum Variance Linear Unbiased Estimator (MVLUE) of any linear parametric function has been developed using Hilbert space method. The problem of estimation of parametric functions has been dealt with under any sampling design for univariate or multivariate populations.

Key words: Minimum variance linear unbiased estimation, Elementary estimate, Hilbert space, Best weights, Dispersion matrix, Univariate/Multivariate population, Hadamard product of matrices.

1. Introduction

On the basis of study of rotating samples in repeat surveys over a number of occasions or periods Minimum Variance Linear Unbiased Estimator (MVLUE) of parameter for a level, change or average over time can be obtained. The work done by Jessen [4], Yates [12], Patterson [8], Tikkiwal [11], Eckler [1], Rao and Graham [9], Narain *et al.* [7] Krishnaiah and Rao [5], Holt and Skinner [3] and Nandram [6] may be mentioned in this connection. Gurney and Daly [2] have defined an elementary estimate as one which does not make use of the survey data for any other time period except that period to which the estimates refer. And with this elementary estimate as a conceptual base, using the Hilbert space approach, they have derived two matrices viz. a matrix C (say) of the best weights to be used on all the elementary estimators in order to yield the MVLUE's of level parameters and a dispersion matrix D (say) of these MVLUE's. Often in a repeat survey it is also desired to obtain MVLUE's of parametric functions such as change between any two occasions, an average or total over some or all the occasions or in general, any linear combination of population parameters. Data on income in all surveys are collected in nominal forms. But in all planning exercises it is worked out in terms of constant prices i.e. after taking into account the inflation.

1 Institute of Statistics & Applied Economics, Makerere University, Kampala, Uganda

Mathematically

$$\overline{Y} = \frac{1}{K} \sum_{t=1}^{K} \frac{\overline{Y}_t}{C_t}$$

where,

$\overline{Y}_t$ = Average income per capita for time period t

C_t = Inflation over base period for time period t for working out $\overline{Y}$ on constant prices

In the Health Sector the number (B) of births, averted by using various methods of Family Planning i.e. sterlization or spacing / temporary methods (e.g. Cu. T, Tablets etc.) are estimated as a linear function of beneficiaries adopting various methods.

Mathematically

$$B = \sum_{i}^{M} k_i \, Y_i$$

where k_i = 1, for permanent method and less than one for spacing methods.

This problem has been dealt with in this paper for an univariate or multivariate population.

2. *Estimation of Parametric Functions*

In general, two types of sampling patterns are frequently noticed in the literature on successive sampling viz. (i) a sampling pattern in which a subsample is retained on all the seasons of years but the other sub-sample remains afresh on all the seasons of the years and (ii) a sampling pattern in which a sub-sample is retained in all the seasons of years but the other sub-sample remains afresh only on all the years not on the seasons. In order to obtain MVLUE's of parametric functions only the latter sampling pattern has been adopted in this paper leaving the former as subject matter of further research.

2.1 Sampling Pattern of the Problem

Suppose elementary estimators x_{mi} and x_{ui} along with their variances $V(x_{mi})$ and $V(x_{ui})$ correspond to two independent samples mi and ui selected at the i-th occasion of a repeat survey for multivariate or univariate population such that sample mi is matched over all the h-occasions while the samples ui's are matched over only some of the h-occasions as shown in the following diagram :

Season/ Occasion	Samples								
1	m1	u1							
2	m1	u1							
3	m1	u1							
4	m1		u4						
5	m1		u4						
6	m1		u4						
...	...			...					
...	...				...				
i − 2	m1					ui − 2			
i − 1	m1					ui − 2			
i	m1					ui − 2			
...	...						...		
...	...							...	
h − 2	m1								uh-2
h − 1	m1								uh-2
h	m1								uh-2

Here we have considered a situation where a year comprises of three seasons/occasions.

The sampling design followed in the selection of all the samples mi's and ui's might be the same or it might be varied from sample to sample for all the occasions. We shall assume here that the expected value of the elementary estimators corresponding to the two independent samples selected on a given occasion for a characteristic is the same.

2.2 Notations

Let

$$x_m = (x_{m1}\ x_{m2} \cdots x_{mi} \cdots x_{mh})$$

$$x_u = (x_{u1}\ x_{u2} \cdots x_{ui} \cdots x_{uh})$$

$$x = (x_m\ x_u)$$

$$s_{mi}, s_{ui} = \sqrt{V(x_{mi})},\ \sqrt{V(x_{ui})}$$

$$s_m = (s_{m1}\ s_{m2} \cdots s_{mi} \cdots s_{mh})$$

$$\mathbf{s}_u = (s_{u1}\ s_{u2} \ldots s_{ui} \ldots s_{uh})$$

$$\mathbf{s} = (\mathbf{s}_m\ \mathbf{s}_u)'$$

θ_i = Population parameter on i–th occasion, $\forall\ i = 1, 2, \ldots, h$

z_i = MVLUE for θ_i

$$\theta = (\theta_1\ \theta_2 \ldots \theta_i \ldots \theta_h\ \theta_1\ \theta_2 \ldots \theta_i \ldots \theta_h)$$

$$= (\Sigma\, \theta_i\, u_i)$$

$$\mathbf{U} = [u_1'\ u_2' \ldots u_i' \ldots u_h']'$$

$$= [I_h\ I_h]$$

= Design matrix or a matrix of the sampling pattern

$$E(\mathbf{x}) = \theta$$

$$E(\mathbf{z}) = \theta$$

$$\mathbf{R}_{mh} = [\rho_{ij}]$$

= Correlation matrix of $\mathbf{x}_m'$ under any correlation model

$$\mathbf{R}_{uh} = [\rho_{ij}']$$

= Correlation matrix of $\mathbf{x}_u'$ under any correlation model

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{mh} & \mathbf{O}_h \\ \mathbf{O}_h & \mathbf{R}_{uh} \end{bmatrix}$$

= Correlation matrix of $\mathbf{x}'$

$\mathbf{K}$ = Dispersion matrix of $\mathbf{x}'$

= Hadamard product (.) of $\mathbf{R}$ and $\mathbf{s}'\,\mathbf{s}$

$$= \begin{bmatrix} \mathbf{R}_{mh} \cdot \mathbf{s}_m'\,\mathbf{s}_m & \mathbf{O}_h \\ \mathbf{O}_h & \mathbf{R}_{uh} \cdot \mathbf{s}_u'\,\mathbf{s}_u \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{K}_{mh} & \mathbf{O}_h \\ \mathbf{O}_h & \mathbf{K}_{uh} \end{bmatrix}, \quad \text{(say)}$$

2.3 Results

The matrix **C** of the best weights and the dispersion matrix **D** of MVLUE's **z**′ of θ′ are obtained as follows:

$$\mathbf{C} = \mathbf{U}'(\mathbf{U}\mathbf{K}^{-1}\mathbf{U}')^{-1}\mathbf{U}\mathbf{K}^{-1}$$

$$= [\mathbf{K}_{mh}^{-1} + \mathbf{K}_{uh}^{-1}]^{-1} \begin{bmatrix} \mathbf{I}_h & \mathbf{I}_h \\ \mathbf{I}_h & \mathbf{I}_h \end{bmatrix} \begin{bmatrix} \mathbf{K}_{mh}^{-1} & \mathbf{0}_h \\ \mathbf{0}_h & \mathbf{K}_{uh}^{-1} \end{bmatrix}$$

$$\mathbf{D} = \mathbf{U}'(\mathbf{U}\mathbf{K}^{-1}\mathbf{U}')^{-1}\mathbf{U}$$

$$= [\mathbf{K}_{mh}^{-1} + \mathbf{K}_{uh}^{-1}]^{-1} \begin{bmatrix} \mathbf{I}_h & \mathbf{I}_h \\ \mathbf{I}_h & \mathbf{I}_h \end{bmatrix}$$

For detailed steps for derivation of **C** and **D** matrices readers are referred to Gurney and Daly [2].

Let us further define a matrix **B** (say) of the coefficients of θ′ in its various parametric functions as follows:

$$\mathbf{B} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \vdots \\ \beta_i \\ \vdots \\ \vdots \\ \beta_h \\ \beta_{h+1} \\ \vdots \\ \vdots \\ \beta_{h+h'} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ -1 & 1 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ 1/h & 1/h & \cdots & 1/h & \cdots & 1/h & 0 & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{I}_h & \mathbf{0}_h \\ \mathbf{B}_{1_{h' \times h}} & \mathbf{0}_{h' \times h} \end{bmatrix}$$

where

$$\beta_j = (\beta_1\, \beta_2 \ldots\ldots \beta_j \ldots\ldots \beta_h\, 0\, 0 \ldots\ldots 0 \ldots\ldots 0)$$
$$\forall\; j = 1, 2, \ldots, i, \ldots, h, h+1, \ldots, h+h'$$
= Row vector with the coefficients of z_j's or θ_j's in their any linear combination as the components

Then following lemmas (Appendices A and B) due to Singh [10], the MVLUE's for the parametric functions $\mathbf{B}\,\theta'$ (i.e. including level parameters) are given by

$$(\hat{\mathbf{B}\,\theta'}) = \mathbf{B\,C\,x'}$$

$$= \begin{bmatrix} (K_{mh}^{-1} + K_{uh}^{-1})^{-1} & (K_{mh}^{-1}x_m' + K_{uh}^{-1}x_u') \\ \mathbf{B}_{1_{h'\times h}} (K_{mh}^{-1} + K_{uh}^{-1})^{-1} & (K_{mh}^{-1}x_m' + K_{uh}^{-1}x_u') \end{bmatrix} \tag{1}$$

and the corresponding dispersion matrix by

$$V(\hat{\mathbf{B}\,\theta'}) = \mathbf{B}V(\hat{\theta}')\mathbf{B}'$$

$$= \mathbf{B\,D\,B'}$$

$$= \begin{bmatrix} (K_{mh}^{-1} + K_{uh}^{-1})^{-1} & (K_{mh}^{-1} + K_{uh}^{-1})^{-1}\,\mathbf{B}_{1_{h\times h'}}' \\ \mathbf{B}_{1_{h'\times h}} (K_{mh}^{-1} + K_{uh}^{-1})^{-1} & \mathbf{B}_{1_{h'\times h}} (K_{mh}^{-1} + K_{uh}^{-1})^{-1}\,\mathbf{B}_{1_{h\times h'}}' \end{bmatrix} \tag{2}$$

The structure of (1) and (2) immediately prompts us to prove the following theorem :

2.4 Theorem

In a h-occasion repeat survey MVLUE's for the level parameters and for their parametric functions are separable.

Proof. In Section 2.3 evidently it is the structure of the matrix

$$\mathbf{B} = \begin{bmatrix} I_h & 0_h \\ \mathbf{B}_{1_{h'\times h}} & 0_{h'\times h} \end{bmatrix}$$

which helps us to obtain the MVLUE's for both the level parameters and for their parametric functions. Therefore, to prove the result let us consider the following two cases:

Case 1:

When $\mathbf{B} = [I_h\;\; 0_h]$

Then MVLUE's for the level parameters are given by

$$
\begin{aligned}
(\hat{B\theta'}) &= \mathbf{B\,C\,x'} \\
&= [I_h\; 0_h]Cx' \\
&= (K_{mh}^{-1} + K_{uh}^{-1})^{-1}\,(K_{mh}^{-1}x_m' + K_{uh}^{-1}x_u')
\end{aligned}
\tag{3}
$$

and their corresponding variances by the dispersion matrix

$$
\begin{aligned}
V(\hat{B\theta'}) &= [I_h\; 0_h]\,\mathbf{D}\begin{bmatrix} I_h \\ 0_h \end{bmatrix} \\
&= (K_{mh}^{-1} + K_{uh}^{-1})^{-1}
\end{aligned}
\tag{4}
$$

Case 2:

When $\mathbf{B} = [\mathbf{B}_{1_{h'\times h}}\; 0_{h'\times h}]$

Similarly as in Case 1, here also the MVLUE's for the parametric functions of the level parameters are given by

$$(\hat{B\theta'}) = \mathbf{B}_{1_{h'\times h}}(K_{mh}^{-1} + K_{uh}^{-1})^{-1}\,(K_{mh}^{-1}x_m' + K_{uh}^{-1}x_u') \tag{5}$$

and their corresponding variances by the dispersion matrix

$$V(\hat{B\theta'}) = \mathbf{B}_{1_{h'\times h}}(K_{mh}^{-1} + K_{uh}^{-1})^{-1}\,\mathbf{B}_{1_{h\times h'}} \tag{6}$$

Therefore, as shown by the structures of the matrix B in Cases 1 and 2, the MVLUE's for the level parameters and for their parametric functions are given by the relations (3) and (5) and their corresponding dispersion matrices by the relations (4) and (6) respectively.

Obviously the relations (3) and (5) and (4) and (6) are components of the relations (1) and (2) respectively. Hence the theorem.

Appendix A

Lemma - 1. The best weights for any linear combination of the MVLUE's are generated from the same linear combination of the individual best weights of the corresponding MVLUE's.

Proof. $\theta = (\theta_1\, \theta_2 \ldots \theta_j \ldots \theta_h)$

= Population parameter–value row vector

$$\mathbf{z} = (z_1\, z_2 \ldots z_j \ldots z_h)$$
$$= \text{MVLU estimators' row vector}$$
$$\mathbf{x} = (x_1\, x_2 \ldots x_j \ldots x_h)$$
$$= \text{Elementary estimators row vector}$$
$$E(\mathbf{x}') = \theta'$$
$$\mathbf{C} = [c_1'\, c_2' \ldots c_i' \ldots c_h']'$$
$$= [c_{ij}] \qquad \forall\ i, j = 1, 2, \ldots, h$$
$$= \text{Matrix of the best weights to be used on the elementary estimators for providing } z_j\text{'s}$$
$$\beta = (\beta_1\, \beta_2 \ldots \beta_j \ldots \beta_h)$$
$$= \text{Row vector with the coefficients of } z_j\text{'s or } \theta_j\text{'s in their any linear combination as the components}$$

Then a linear combination of the parameters and their MVLUE's is defined by $\beta\theta'$ and $\beta\mathbf{z}'$ respectively, such that

$$\beta\mathbf{z}' = (\widehat{\beta\, \theta'})$$

Obviously $\mathbf{z}' = \mathbf{C}\mathbf{x}'$

Hence we get

$$z_j = c_i\mathbf{x}' \qquad \forall\ i = j = 1, 2, \ldots, h$$

Further we know that

$$(\widehat{\beta\theta'}) = \beta\mathbf{z}'$$
$$= \beta\mathbf{C}\mathbf{x}'$$
$$= (\Sigma\beta_j c_i)\mathbf{x}' \qquad \forall\ i = j = 1, 2, \ldots, h \quad (1)$$
$$= \left(\sum_{i=j} \beta_j c_{i1}\ \sum_{i=j} \beta_j c_{12} \cdots \sum_{i=j} \beta_j c_{ij'} \cdots \sum_{i=j} \beta_j c_{ih}\right)\mathbf{x}'$$
$$\forall\ j' = 1, 2, \ldots h \qquad (2)$$

From (2) we conclude that the best weight component

$$\sum_{i=j} \beta_j c_{ij}' = \beta_1 c_{1j}' + \beta_2 c_{2j}' + \ldots + \beta_j c_{ij}' + \ldots + \beta_h c_{hj}'$$

assigned to x_j is the same linear combination of the best weights assigned to x_j in z' as is the linear combination $\beta z'$, $\forall\ j = j' = 1, 2, \ldots h$.

Appendix B

Lemma - 2. The MVLU estimator $\hat{G}$ for level, change, average, total or any other linear combination of the population parameters is given by

$$\hat{G} = \alpha x'$$

and its variance by

$$V(\hat{G}) = V(\alpha x') = \sum_{i=j} \beta_j^2 d_{ii} + 2 \sum_{(i=j)<j'} \beta_j \beta_j' d_{ij}'$$

where

α = row vector of the best weights to be used on x'

β_j, β_j' = j–th and j′–th components of the row vector β

d_{ii} = i–th diagonal element of dispersion matrix D $(= [d_{ij}])$ of z_j's

$= V(z_j) \qquad \forall\ i = j = 1, 2, \ldots, h$

and

d_{ij}' = (i, j′)–th element of D

$= \mathrm{Cov}(z_j, z_{j'}) \qquad \forall (i = j) < j' = 1, 2, \ldots, h$

Proof. Let

$D = [d_{ij}]$ = Dispersion matrix of the MVLUE's z_j's of the population parameters θ_j's for the levels, $\forall\ i, j = 1, 2, \ldots, h$ and

$\alpha = (\alpha_1\ \alpha_2 \ldots \alpha_j \ldots \alpha_h)$

= Row vector of the best weights depending upon β_j's and C

$\forall\ i, j, j' = 1, 2, \ldots, h$

Denote "$\sum_{i=j} \beta_j c_i$" in lemma-1 by 'α'

Also from (1) of lemma-1, we get

$$\hat{(\beta\theta')} = \beta z'$$
$$= \beta C x'$$
$$= \alpha x'$$

Hence,

$$\beta z' = \alpha x' \tag{3}$$

After defining $\beta z'$ by $\hat{G}$ in (3), we get the desired result

$$\hat{G} = \alpha x'$$
$$= (\sum_{i=j} \beta_j c_{i1}, \sum_{i=j} \beta_j c_{i2} \cdots \sum_{i=j} \beta_j c_{ij}' \cdots \sum_{i=j} \beta_j c_{ih})x' \tag{4}$$

where a specific estimator $\hat{G}$ is ascertained by a suitable choice of an arbitrary vector β.

To obtain variance of $\hat{G}$, let us consider now the dispersion matrix of z' given by

$$D = [d_{ij}]$$

Obviously

$$V(\hat{G}) = V(\beta z')$$
$$= \beta V(z')\beta'$$
$$= \beta D \beta'$$
$$= \sum_{i=j,j'} \beta_j \beta_j' d_{ij}'$$
$$= \sum_{i=j} \beta_j^2 d_{ii} + 2 \sum_{(i=j)<j'} \beta_j \beta_j' d_{ij}'$$

Hence the lemma.

ACKNOWLEDGEMENTS

Authors are thankful to the referees for their valuable comments/suggestions. The authors are also thankful to Dr. Padam Singh, Director, I.R.M.S., New Delhi for rendering his valuable suggestions for the improvement of text of the paper.

REFERENCES

[1] Eckler, A.R., 1955. Rotation sampling. *Ann. Math. Statist.*, **26**, 664-685.

[2] Gurney, M. and Daly, J.F., 1965. A multivariate approach to estimation in periodic sample surveys. *Proc. Soc. Stat. Sec.2, J. Amer. Statist. Assoc.*, 242-257.

[3] Holt, D. and Skinner, C.J., 1989. Components of change in surveys. *Instt. Stat. Rev.*, **57**(1), 1-18.

[4] Jessen, R.J., 1942. Statistical investigation of a sample survey for obtaining farm facts. *Iowa Agri. Expt. Station Res.*, Bull No. 304.

[5] Krishnaiah, P.R. and Rao, C.R., 1988. *Sampling in Time eds.* Handbook of Statistics, **6**, 187-211.

[6] Nandram, Balgobin, 1994. Bayesian predictive inference for multivariate sample surveys. *J. Off. Statist.*, **10**(2), 167-179.

[7] Narain, P., Kathuria, O.P. and Srivastava, A.K., 1987. Measuring change in agricultural surveys. *Invited Paper, 46th session of the International Statistical Institute*, Tokyo, 1-15.

[8] Patterson, H.D., 1950. Sampling on successive occasions with partial replacement of units, *J. Roy. Statist. Soc.*, **B 12**, 241-255.

[9] Rao, J.N.K. and Graham, J.E., 1964. Rotation designs for sampling on repeated occasions. *J. Amer. Statist. Assoc.*, **59**, 492-509.

[10] Singh, Jagbir, 1989. A study on estimation of change in sampling on successive occasions. Unpublished Ph.D. thesis, I.A.R.I., New Delhi.

[11] Tikkiwal, B.D., 1951. Theory of successive sampling. Unpublished thesis of Diploma, I.C.A.R., New Delhi.

[12] Yates, F. 1949. *Sampling Methods for Census and Surveys*, Charles Griffin and Company Ltd., London.