THE COORDINATE - FREE ESTIMATION IN FINITE POPULATION SAMPLING

by

J. Rodrigues and S.N. Elian

THE COORDINATE - FREE ESTIMATION IN FINITE POPULATION SAMPLING

by

Josemar Rodrigues and Silvia Nagib Elian

Instituto de Matematica e Estatistica - USP
Rua do Matao ,1010, Sao Paulo , Brasil

This article presents the coordinate - free approach to extend a
result of Bellhouse (1987) for the estimation of a finite
population total under model - based inference to include models
with correlated error structure. The result is applied to derive the
optimal linear- model unbiased estimator of the finite population
total when the observations are correlated .

Key Words : Optimal estimation ; correlated error structure ;
Coordinate - free models ; Efficiency comparisons .

1- INTRODUCTION

Royall (1970a) , Fuller (1970) and Bellhouse (1987) gave a result
for estimation of a finite population total under the
superpopulation model. This result allows the estimator obtained by
classical parametric procedure to be used to obtain an estimator of
the finite population total. In Bellhouse 's paper (1987) , it was
assumed that the N measurement y of the finite populations units
were independent and had unequal model means. In this paper this
result is extended to include models with correlated error
structure. As in Bellhouse (1987) , it is shown that the result can
be used to obtain optimal model - unbiased estimators of the finite
population total $T = 1'y$, where $1'$ is an N - dimensional vector of
ones and

$$y' = (y_1, \ldots, y_N) ,$$

using the Gauss - Markov Theorem . Bellhouse 's result (1987) is as
follows: Suppose that the statistician has some knowledge of the
population expressed by the following coordinate - free model :

$$(i) \qquad E[y] = b ,$$

where $b' = (b_1, \ldots, b_N)$ (unknown vector) and V is a known diagonal covariance matrix. To gain information about the linear function $l'y$, the statistician, using a sampling plan ,selects a sample S . As in Pereira and Rodrigues (1983) , $i_k = 1$ if the kth unit ($k = 1, \ldots, N$) belongs to S and $i_k = 0$ if the kth unit does not belong to S. The distribution of the random vector $(i_1, \ldots, i_N)$ is called sampling design and will be represented by P . The quantity of interest , $l'y$, may be partitioned as

$$(2) \qquad T = l'y = l'I_S y + l'(I - I_S)y$$

where I_S is a diagonal matrix of order N with kth diagonal element being i_k ($k = 1, \ldots, N$) and I is the identity matrix of order N. To estimate T , the model mean

$$(3) \qquad E[\, l'(I - I_S)y \,] = l'(I - I_S)b = b_{nS}$$

is estimated. Suppose two estimators of b_{nS} , say $b_S^{(1)}$ and $b_S^{(2)}$, are available so that two estimtors of T using the prediction approach are

$$(4) \qquad T_t = l'I_S y + b_S^{(t)} \qquad , t = 1 , 2 \ldots$$

The Bellhouse 's result (1987) is : for any sampling plan P of size n , for every sample S , $E[\, b_S^{(1)} - b_{nS} \,]^2 \leq E[\, b_S^{(2)} - b_{nS} \,]^2$ then

$$(5) \qquad E_P E[\, T_1 - T \,]^2 \leq E_P E[\, T_2 - T \,]^2$$

The operators E and E_P denote expectation with respect to the coordinate - free model and the sampling design P , respectively. Bellhouse restricted attention to diagonal covariance matrix. In Section 2, we extend this result to non - diagonal covariance matrix by using a coordinate - free version of Tam 's result (1986).

2 - RESULTS WITH NON - DIAGONAL COVARIANCE MATRIX

For simplicity, the following notation is used in the sequel:

$$y_S = I_S y \ , \ y_{nS} = (I - I_S)y \ , \ V_1 = I_S V \cdot I_S \ ,$$

$$V_2 = (I - I_S)V(I - I_S) \ , \ V_{12} = V'_{21} = I_S V (I - I_S) \ ,$$

$$\bar{V}_1 = \text{the generalized inverse of } V_1 .$$

Let us assume that the covariance matrix in model (1) is not diagonal and

$$b_S^{(t)} = L'^{(t)} \ I_S y \quad , \ t = 1,2 \quad \text{two linear estimators}$$

of $l'(I - I_S)b$, where $L'^{(t)}$ is an N - dimensional vector.

The following Lemma is the coordinate - free version of Tam's result (1986) and will be useful to abtain our main result.

LEMMA. Under the coordinate - free model defined in (1) with non-diagonal covariance matrix , for all linear estimator $h'I_S y$ of T , we have

$$E [h'I_S y - T]^2 = E [(h'I_S y - l'\bar{V}_1 I_S V)(I_S y - I_S b)]^2$$

$$+ l'V_1 - l'\bar{V}_1 I_S V I_S V_1 + [h'I_S b - l'b]^2.$$

The proof of this Lemma follows standard computation and hence omitted. The following Theorem is an extention of Bellhouse's result to include models with non - diagonal covariance matrix.

THEOREM. For any sampling plan P and under the coordinate – free model (1) with non – diagonal covariance matrix

$$E_p E [T_1 - T]^2 < E_p E [T_2 - T]^2,$$ if and only if, for all S such that $P(S) > 0$.

(9)
$$E \{ [\overset{(1)}{b_S} - 1'V_{21} V_1 I_S y] - [1' (I - I_S) - 1'V_{21} V_1 I_S] \overline{b}_S \}^2$$

$$\leq E \{ [\overset{(2)}{b_S} - 1'V_{21} V_1 I_S y] - [1' (I - I_S) - 1'V_{21} V_1 I_S] \overline{b}_S \}^2 .$$

Proof. Since the expectation operators E and E_p are independent , the result easily follows from the above Lemma . If V_{21} is the null matrix then the Bellhouse' s result follows.

3. APPLICATIONS OF THE RESULT

The theorem in Section 2 is useful to construct optimal model – based unbiased estimators for T. Suppose there is a N*p – covariate matrix X such that $b = Xu$, where u is an unknown p – dimensional vector. The least – square of u is

(10)
$$\overline{u} = (I_S X V_1 X'I_S)X'I_S V_1 y .$$

Then , for any given sample , by the Gauss – Markov Theorem the variance of

(11)
$$b_S - 1'V_{21} V_1 I_S y$$
is minimum among the

variances of estimators of $[1'(I - I_S) - 1' V_{21} V_1 I_S] Xu$ if

(12)
$$b_S = 1'V_{21} V_1 I_S y + [1' (I - I_S) - 1' V_{21} V_1 I_S]Xu$$

$$= 1'(I - I_S)[Xu + V_{S1} I_S V_1 (y - I_S Xu)] .$$

On applying the result of Secton 2 , $T = l'I_s\, y_s + b_s$ is the optimal estimator of T among all linear model − unbiased estimators of T (see Criterion 3 in Sarndal , 1978). Theorem 2.1 of Royall (1976) trivially follows from (12) if in (9) we take $P(S) = 1$ for some sample S. Similar arguments as in Bellhouse (1987) can be applied to find ridge estimators of the finite population total under models with non − diagonal covariance matrix .

REFERENCES.

Bellhouse , D.R. (1987), " Model − Based Estimation in Finite Population Sampling ", American Statistician Association , 41,4,260 − 262.

Fuller , W.A. (1970) , " Simple Estimators for the Mean of Skewed Populations", technical report , Iowa State University ,Dept. of Statistics.

Pereira, C.A.B. and Rodrigues , J. (1983), " Robust Linear Prediction in Finite Populations ", Int. Statist. Rev., 51, 293 −300.

Royall, R.M. (1970), " On Finite Population Sampling Theory Under Certain Linear Regression Models", Biometrica , 57. 377 − 387.

−−−−−−−−−−−−−−(1976), " The Linear Least − Squares Prediction Approach to Two − Stage Sampling ", Journal of the American Statistical Association, 71, 657 − 664.

Sarndal, C.E., (1978) ," Designed − Based Model Inference in Survey Sampling ", Scand. J. Statist. 5, 27 − 52.

Tam , S.M. (1986), " Characterization of Best Model − Based Predictors in Survey Sampling ", Biometrica, 73,1,232 − 235.

RELATÓRIO TÉCNICO
DO
DEPARTAMENTO DE ESTATÍSTICA

TÍTULOS PUBLICADOS EM 1987

8701 - ACHCAR, J.A. & BOLFARINE, H.; Constant Hazard Against a Change-Point Alternative: A Bayesian Approach with Censored Data, São Paulo, IME-USP, 1987, 20p.

8702 - RODRIGUES, J.; Some Results on Restricted Bayes Least Squares Predictors for Finite Populations, São Paulo, IME-USP, 1987, 16p.

8703 - LEITE, J.G., BOLFARINE, H. & RODRIGUES, J.; Exact Expression for the Posterior Mode of a Finite Population Size: Capture-Recapture Sequential Sampling, São Paulo, IME-USP, 1987, 14p.

8704 - RODRIGUES, J., BOLFARINE, H. & LEITE, J.G.; A Bayesian Analysis in Closed Animal Populations from Capture Recapture Experiments with Trap Response, São Paulo, IME-USP, 1987, 21p.

8705 - PAULINO, C.D.M.; Analysis of Categorical Data with Full and Partial Classification: A Survey of the Conditional Maximum Likelihood and Weighted Least Squares Approaches, São Paulo, IME-USP, 1987, 52p.

8706 - CORDEIRO, G.M. & BOLFARINE, H.; Prediction in a Finite Population under a Generalized Linear Model, São Paulo, IME-USP, 1987, 21p.

8707 - RODRIGUES, J. & BOLFARINE, H.; Nonlinear Bayesian Least-Squares Theory and the Inverse Linear Regression, São Paulo, IME-USP, 1987, 15p.

8708 - RODRIGUES, J. & BOLFARINE, H.; A Note on Bayesian Least-Squares Estimators of Time-Varying Regression Coefficients, São Paulo, IME-USP, 1987, 11p.

8709 - ACHCAR, J.A., BOLFARINE, H. & RODRIGUES, J.; Inverse Gaussian Distribution: A Bayesian Approach, São Paulo, IME-USP, 1987, 20p.

8710 - CORDEIRO, G.M. & PAULA, G.A.; Improved Likelihood Ratio Statistics for Exponential Family Nonlinear Models, São Paulo, IME-USP, 1987, 26p.

8711 - SINGER, J.M.; PERES, C.A.& HARLE, C.E.; On the Hardy-Weinberg Equilibrium in Generalized ABO Systems, São Paulo, IME-USP, 1987, 16p.

8712 - BOLFARINE, H. & RODRIGUES, J.; A Review and Some Extensions on Distribution Free Bayesian Approaches for Estimation and Prediction, São Paulo, IME-USP, 1987, 19p.

8713 - RODRIGUES, J.; BOLFARINE, H. & LEITE, J.G.; A Simple Nonparametric Bayes Solution to the Estimation of the Size of a Closed Animal Population, São Paulo, IME-USP, 1987, 11p.

8714 - BUENO, V.C.; Generalizing Importance of Components for Multistate Monotone Systems, São Paulo, IME-USP, 1987, 12p.

8801 - PEREIRA, C.A.B. & WECHSLER, S.; On the Concept of P-value, São Paulo, IME-USP, 1988, 22p.

8802 - ZACKS, S., PEREIRA, C.A.B. & LEITE, J.G.; Bayes Sequential Estimation of the Size of a Finite Population, São Paulo, IME-USP, 1988, 23p.

8803 - BOLFARINE, H.; Finite Population Prediction Under Dynamic Generalized Linear Models, São Paulo, IME-USP, 1988, 21p.

8804 - BOLFARINE, H.; Minimax Prediction in Finite Populations, São Paulo, IME-USP, 1988, 18p.

8805 - SINGER, J.M. & ANDRADE, D.F.; On the Choice of Appropriate Error Terms for Testing the General Linear Hypothesis in Profile Analysis, São Paulo, IME-USP, 1988, 23p.

8806 - DACHS, J.N.W. & PAULA, G.A.; Testing for Ordered Rate Rations in Follow-up Studies with Incidence Density Data, São Paulo, IME-USP, 1988, 18p.

8807 - CORDEIRO, G.M. & PAULA, G.A.; Estimation, Significance Tests and Diagnostic Methods for the Non-Exponential Family Nonlinear Models, São Paulo, IME-USP, 1988, 29p.