Robust Linear Prediction in Finite Populations

Carlos Alberto de Bragança Pereira; Josemar Rodrigues

International Statistical Review / Revue Internationale de Statistique, Vol. 51, No. 3. (Dec., 1983), pp. 293-300.

Stable URL:
http://links.jstor.org/sici?sici=0306-7734%28198312%2951%3A3%3C293%3ARLPIFP%3E2.0.CO%3B2-6

International Statistical Review / Revue Internationale de Statistique is currently published by International Statistical Institute (ISI).

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/isi.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org
Thu Jan 24 20:11:06 2008

International Statistical Review, **51** (1983), pp. 293–300. Longman Group Limited/Printed in Great Britain
© International Statistical Institute

Robust Linear Prediction in Finite Populations

Carlos Alberto de Bragança Pereira and Josemar Rodrigues

Instituto de Matemática e Estatística, Universidade de São Paulo, Caixa Postal 20.570 (Ag. Iguatemi), São Paulo, Brasil

Summary

This paper shows how a result of Zyskind (1967) on characterization of best linear estimators for general linear models may be applied to generalize and unify the results of the theory of linear prediction in survey sampling. The robust linear predictor is characterized in a general situation. By using indicators of sample elements it is shown that the natural order of the population quantities may be maintained throughout the analysis.

Key words: Balanced sample; Extended-balanced sample; Extended-overbalanced sample; Linear prediction; Overbalanced sample; Robustness; Superparameter; Superpopulation model; ξ-(xi-)optimality; ξ-(xi-)model; ξ-(xi-)unbiasedness.

1 Introduction

A great deal of attention has been dedicated lately to the problem of robustness of linear predictions in survey sampling under the superpopulation approach. See, for example, Royall & Herson (1973a,b), Scott, Brewer & Ho (1978), Tallis (1978), and Rodrigues (1983). However, some of the properties studied in those papers are restricted to particular cases or are not completely explored. Tallis (1978) proves a necessary and sufficient condition for robustness of the expansion predictor. Royall & Herson (1973a) introduce a sufficient condition for robustness when the alternative is a polynomial model. Scott, Brewer & Ho (1978) extend Royall & Herson's results (1973a) by using a general variance function in the primary model and prove a sufficient condition for robustness.

In the present paper it is shown how the works of Zyskind (1967) and Kruskal (1968) for the general theory of linear models may be applied to the theory of linear prediction in survey sampling. First it is noticed that by using indicator functions of the sample elements one does not need, as usual, to restrict oneself to a particular sample and re-order the vector of population quantities. Then, some known results are restated with the new notation and the sufficient condition given by Scott, Brewer & Ho (1978) is shown to be necessary even in the general case.

Section 2 introduces some notation and emphasizes the point of view of the authors. Section 3 presents a small survey on linear predictors. Finally, § 4 is dedicated to the main contribution of this work. Some known examples are used to illustrate the notation and the results discussed.

2 Preliminaries

In this section we introduce the notation and review some of the concepts used in the sequel.

The finite population of identifiable units is represented by $\mathcal{P} = \{1, 2, \ldots, N\}$, where N, the population size, is known. Associated with each unit (or label) k of $\mathcal{P}$, there is an *unknown* quantity y_k. The populational vector of these unknown quantities is represented by $\mathbf{y}' = (y_1, \ldots, y_N)$. Here, $\mathbf{y}$ is a column vector and $\mathbf{y}'$ is its transpose. In addition to the quantity y_k of the unit k, it is supposed that a vector $X_k' = (x_{k1}, \ldots, x_{kM})$ of M well-defined *known* quantities is associated with k $(k = 1, \ldots, N)$. That is, for every unit k of $\mathcal{P}$, there are associated $M+1$ quantities, $y_k, x_{k1}, \ldots, x_{kM}$, where all but y_k are known. We represent by $\mathbf{X}$ the matrix of order $(N \times M)$ whose row k $(k = 1, \ldots, N)$ is the row vector X_k'.

In order to gain information about a linear function $\mathbf{l}'\mathbf{y}$, where $\mathbf{l}' = (l_1, \ldots, l_N)$ is known, the statistician selects a sample S $(S \subset \mathcal{P})$ of n units from $\mathcal{P}$ and obtains the data $D = \{(k; y_k); k \in S\}$. To represent the elements of S we denote by i_k $(k = 1, \ldots, N)$ the indicator function that indicates whether k belongs to S. That is, $i_k = 1$ if $k \in S$ and $i_k = 0$ if $k \in \mathcal{P} - S$. It is worth emphasizing that S is a subset of n $(< N)$ elements of $\mathcal{P}$ and D is the set of elements of S together with their associated quantities y_k's. In addition we notice that the quantity of interest may be partitioned as $\mathbf{l}'\mathbf{y} = \mathbf{l}'I_S\mathbf{y} + \mathbf{l}'(I - I_S)\mathbf{y}$, where I_S is a diagonal matrix of order N with its kth diagonal element being i_k $(k = 1, \ldots, N)$ and I is the identity matrix of order N. It is clear that after D has been observed, $\mathbf{l}'I_S\mathbf{y}$ is known and the part of $\mathbf{l}'\mathbf{y}$ that remains unknown is $\mathbf{l}'(I - I_S)\mathbf{y}$.

We use in this paper the superpopulation approach to survey sampling; see, for example, Cassel, Sarndal & Wretman (1977). Under this perspective, $\mathbf{y}'$ is considered to be a realization of a random vector $\mathbf{Y}' = (Y_1, \ldots, Y_N)$ whose joint distribution depends on a 'superparameter' λ which is unknown. Usually, the introduction of λ in the context (note that the quantity of interest is $\mathbf{l}'\mathbf{y}$ not λ) is to pin-point the 'information' that is contained in all Y_k's. That is, when λ is fixed, $Y_1, Y_2, \ldots, Y_N$ are mutually independent. However, in this paper the meaning of λ may be considered in more general terms; for example, when λ is fixed the correlation among any two different components of $\mathbf{Y}$ is zero.

After looking at $\mathbf{y}$ as a realization of a random vector, the objective of the study is to predict the value of $\mathbf{l}'(I - I_S)\mathbf{y}$, the unobserved factor of $\mathbf{l}'\mathbf{y}$, using all the auxiliary quantities that form the matrix $\mathbf{X}$. For this, it is desirable to consider models that relate $\mathbf{y}$ to $\mathbf{X}$. The relation among $\mathbf{y}$ and $\mathbf{X}$ is considered when some characteristics of the distribution of $\mathbf{Y}$ have been specified. In the next section the models studied are completely described.

We end this section with some relevant remarks.

Remark 1. The indicators i_k $(k = 1, \ldots, N)$ emphasize the role of the sample design in this kind of approach. Since D is a minimal sufficient statistic (Basu, 1969) we may define as sample design the joint distribution of $i_1, i_2, \ldots, i_N$. With this notation, we do not have, as other authors do, to re-order the elements of $\mathbf{Y}$ in such a way that the elements of S have their correspondent y_k's in the first n positions of $\mathbf{Y}$; see, for example Tallis (1978).

Remark 2. Since the sample design is the same for every value of λ, the vector $(i_1, \ldots, i_N)$ is an ancillary statistic and by the conditionality principle (Basu, 1975) we consider the matrix I_S as known at the inference time.

Remark 3. The quantities y_k that are associated with the units in $\mathcal{P} - S$ are not observed and hence they are in a sense indistinguishable of their respective random quantities Y_k. Consequently, we shall use lower case letters for both random variables and their realizations.

3 Linear predictors

The general background model is described in this section. Some well-known results are stated in terms of our notation, where we use the indicators $i_1, i_2, \ldots, i_N$ and the matrices I_S.

The model that relates $\mathbf{X}$ to $\mathbf{y}$ is described by the following constraints.

(i) Linear: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, where $\mathbf{e}' = (e_1, e_2, \ldots, e_N)$ is a random vector and $\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_M)$ is an unknown vector of constants.

(ii) First moment: $E[\mathbf{e}] = \mathbf{0}$, where $E[\mathbf{e}]$ is the mean vector of $\mathbf{e}$ and $\mathbf{0}' = (0, \ldots, 0)$ is the null vector of order N.

(iii) Covariance structure: $\text{var}[\mathbf{e}] = \sigma^2\mathbf{V}$, where $\text{var}[\mathbf{e}]$ represents the covariance matrix of $\mathbf{e}$, σ^2 is an unknown positive constant, and $\mathbf{V}$ is a diagonal matrix of order N whose kth $(k = 1, \ldots, N)$ diagonal element is a known nonnegative constant that depends on X_k' the kth row of $\mathbf{X}$. This is equivalent to considering a nonnegative real-valued function f such that

$$\mathbf{V} = \text{diag}\,[f(X_1) \ldots f(X_N)].$$

Scott *et al.* (1978) describe some real-life situations where special cases of f intuitively appear.

We notice now that, since any symmetric nonnegative-definite matrix may be diagonalized, there is no loss of generality in considering $\mathbf{V}$ as diagonal. In fact, when $\mathbf{V}$ is not diagonal, the diagonalization suggested here is to be done on the matrix $I_S\mathbf{V}I_S$, the covariance matrix of the sample data. On the other hand, if an element of its diagonal is zero, say $f(X_k) = 0$, then y_k, the corresponding element in $\mathbf{y}$, is considered to be completely known *a priori* and so it could be put aside for the inference. In that way, again without loss of generality, $\mathbf{V}$ may be considered as a positive-definite diagonal matrix.

We must point out that the superparameter λ, described in § 2, corresponds here to the vector $(\beta_1, \beta_2, \ldots, \beta_M, \sigma^2)$ of $M+1$ unknown quantities. On the other hand, the data D is here represented by $I_S\mathbf{y}$. This permits us to define a linear function of the data as a linear combination of elements of $I_S\mathbf{y}$, say $\mathbf{h}'I_S\mathbf{y}$, where $\mathbf{h}' = (h_1, \ldots, h_N)$. This function $\mathbf{h}'I_S\mathbf{y}$ is said to be a linear predictor if it is used to predict any unknown quantity that is a function of $\mathbf{y}$. The reader must note that, although the components h_i of $\mathbf{h}$ are considered for the whole population, the product $\mathbf{h}'I_S$ actually depends only on the components which correspond to the sample units. Consequently, our concern must be addressed only to these sample units of $\mathbf{h}$ and the remaining may be taken arbitrarily. Still relevant is the fact that the rules to choose $\mathbf{h}'I_S$ discussed in this paper have a strong sample dependence. For example, by using these rules, one may conclude that h_1 must take different values for the samples S and S' even when unit 1 belongs to both samples.

For simplicity, in the sequel the model described by constraints (i), (ii) and (iii) is called the ξ-model. In that way, when we say ξ-property we mean a property of this ξ-model.

As described before, our statistical problem is to predict $\mathbf{l}'(I - I_S)\mathbf{y}$ using the data $I_S\mathbf{y}$. This is equivalent to predicting $\mathbf{l}'\mathbf{y}$ since $\mathbf{l}'I_S\mathbf{y}$ is completely known after the data have been observed. The reason for the choice of $\mathbf{l}'\mathbf{y}$, in the place of $\mathbf{l}'(I - I_S)\mathbf{y}$, as the objective of our inference is because we want to relate this work with previous ones.

The concept of unbiasedness in prediction theory is as follows.

Definition 1. A linear predictor $\mathbf{h}'I_S\mathbf{y}$ of $\mathbf{l}'\mathbf{y}$ is said to be ξ-unbiased if $E_\xi[\mathbf{h}'I_S\mathbf{y} - \mathbf{l}'\mathbf{y}] = 0$, where $E_\xi[.]$ is the expectation operator under the ξ-model.

We end this section with the following version of a well-known result in linear prediction (Royall, 1976). As usual, if A is a square matrix, A^- denotes its generalized inverse and A^{-1} its ordinary inverse when A is nonsingular. The class of all linear ξ-unbiased predictors of $\mathbf{l}'\mathbf{y}$ is denoted by $\mathcal{U}_l$. Finally we write

$$\mathbf{B} = [\mathbf{X}' I_S \mathbf{V}^{-1} I_S \mathbf{X}]^- \mathbf{X}' I_S \mathbf{V}^{-1}.$$

THEOREM 1. *Among all elements of* $\mathcal{U}_l$, $\mathbf{h}' I_S \mathbf{y}$ *has the property that* $E_\xi[\mathbf{h}' I_S \mathbf{y} - \mathbf{l}'\mathbf{y}]^2$ *is minimized by choosing h according to*

$$\mathbf{h}' I_S \mathbf{y} = \mathbf{l}' I_S \mathbf{y} + \mathbf{l}'(I - I_S)\mathbf{X}\hat{\beta}, \tag{1}$$

where $\hat{\beta} = \mathbf{B} I_S \mathbf{y} = \mathbf{B}\mathbf{y}$.

This result may be proved by recalling that $I_S\mathbf{y} = (I_S\mathbf{X})\beta + I_S\mathbf{e}$, by taking $(I_S\mathbf{V})^- = I_S\mathbf{V}^{-1}$, and by using the results of the general theory of linear models; see, for example, Rao (1971, p. 301).

Note that the predictor $\mathbf{h}' I_S \mathbf{y}$ satisfying equation (1) relates itself to the ξ-model only through the matrix $\mathbf{X}$ and the function f. Accordingly, we shall write $T(\mathbf{X}, f)$ for the linear function $\mathbf{h}' I_S \mathbf{y}$ that satisfies the conditions of Theorem 1; that is, $T(\mathbf{X}, f)$ is a linear ξ-unbiased predictor of minimum ξ-mean squared error.

4 Robust linear prediction

In this section general results of the theory of linear models are used to obtain a characterization of robustness in linear prediction. We start by restating the robustness problem.

Let us denote by $\mathbf{X}^*$ a matrix of order $N \times (M+J)$, where the first M columns are the columns of $\mathbf{X}$ and the last J columns are $Z_1, \ldots, Z_J$ which form a matrix $\mathbf{Z}$ of order $(N \times J)$. As usual we write $\mathbf{X}^* = [\mathbf{X}, \mathbf{Z}]$ to indicate this construction. Analogously we write

$$\beta^* = \begin{bmatrix} \beta \\ \gamma \end{bmatrix}$$

to indicate that the first M components of β^* are those of β and the last J are the components of another column vector γ. In order to state the robustness problem we consider an alternative ξ^*-model which is defined as the ξ-model having $\mathbf{X}^*$ (known) replacing $\mathbf{X}$, β^* (unknown) replacing β, and a function f^* (known) replacing f. It is clear: (a) that f^* is a function of $M+J$ arguments, and note that it is evaluated in each row $(X_k^*)'$ of X^*; (b) that $\mathbf{V}$ must be replaced by $\mathbf{V}^*$, a diagonal matrix whose kth $(k = 1, \ldots, N)$ diagonal element is $f^*(X_k^*)$; and (c) that $\mathbf{e}$ must be replaced by $\mathbf{e}^*$, another column random vector of N elements. By recalling the 'ξ-best' linear predictor $T(\mathbf{X}, f)$ defined at the end of § 3, and using the above notation, we introduce the following definition which describes the problem of robustness in our context.

Definition 2. The ξ-best predictor $T(\mathbf{X}, f)$ of $\mathbf{l}'\mathbf{y}$ is said to be robust for the ξ^*-model if it is ξ^*-best; that is, $T(\mathbf{X}, f) = T(\mathbf{X}^*, f^*)$.

It is clear that if $T(\mathbf{X}, f)$ satisfies Definition 2, then besides being ξ-unbiased it is ξ^*-unbiased. The following result introduces a necessary and sufficient condition to have both kinds of unbiasedness for $T(\mathbf{X}, f)$.

THEOREM 2. *The linear predictor* $T(\mathbf{X}, f)$ *is* ξ^**-unbiased if and only if the following*

equation holds:

$$\mathbf{l}'(I-I_S)\mathbf{XBZ}=\mathbf{l}'(I-I_S)\mathbf{Z}. \tag{2}$$

Proof. By the definition of $T(\mathbf{X}, f)$ we have

$$T(\mathbf{X}, f)-\mathbf{l}'\mathbf{y}=\mathbf{l}'(I-I_S)\mathbf{X}\hat{\beta}-\mathbf{l}'(I-I_S)\mathbf{y},$$

where $\hat{\beta}$ is defined in Theorem 1. That is,

$$T(\mathbf{X}, f)-\mathbf{l}'\mathbf{y}=\mathbf{l}'(I-I_S)[\mathbf{XB}-I]\mathbf{y}.$$

Since $T(\mathbf{X}, f)$ is ξ-unbiased,

$$\mathbf{l}'(I-I_S)[\mathbf{XB}-I]\mathbf{X}=0.$$

It is clear now that $T(\mathbf{X}, f)$ is ξ^*-unbiased if and only if

$$\mathbf{l}'(I-I_S)[\mathbf{XB}-I]\mathbf{X}^*=0.$$

The proof is completed by using $\mathbf{X}^*=[\mathbf{X}, \mathbf{Z}]$ in this last equation.

The following example shows how condition (2) may be considered as a generalization of the concept of balanced sample.

Example 1. With the above notation, consider $\mathbf{X}$ as being 1_N, the column vector of N elements equal to unity. Suppose that $\mathbf{V}=I$ and that $\mathbf{Z}$ is a column vector of order N whose elements are $x_1, \ldots, x_N$; that is,

$$\mathbf{X}^*=\begin{bmatrix}1 & x_1\\ \vdots & \vdots\\ 1 & x_N\end{bmatrix}.$$

If condition (2) holds, then

$$\mathbf{l}'(I-I_S)1_N[1_N'I_S1_N]^{-}1_N'I_S\mathbf{Z}=\mathbf{l}'(I-I_S)1_N\left(\frac{1}{n}\sum_{k=1}^{N} i_k x_k\right)=\mathbf{l}'(I-I_S)\mathbf{Z}.$$

If $\mathbf{l}=1_N$, then we finally have

$$\frac{(N-n)}{n}\sum_{k=1}^{N} i_k x_k=\sum_{k=1}^{N}(1-i_k)x_k,$$

which is the balanced-sample (of first-order) property; see Royall & Herson (1973a).

The following Lemma is Theorem 3 of Zyskind (1967) for which Kruskal (1968) states an equivalent version. It is strongly used to prove the next theorem.

LEMMA. *A linear function $\omega'\mathbf{y}$ is a best linear unbiased estimator of its expectation if and only if $\mathbf{V}\omega$ belongs to the space generated by the columns of $\mathbf{X}$.*

The result below is the main contribution of this paper. It introduces a necessary and sufficient condition for linear robustness.

THEOREM 3. *The ξ-best linear predictor of $\mathbf{l}'\mathbf{y}$, $T(\mathbf{X}, f)$, is robust for the ξ^*-model, that is $T(\mathbf{X}, f)=T(\mathbf{X}^*, f^*)$, if and only if T is ξ^*-unbiased and there exists a vector $C'=(c_0, c_1, \ldots, c_{M+J-1})$ of scalars such that*

$$I_S\mathbf{V}^*\mathbf{B}'\mathbf{X}'(I-I_S)\mathbf{l}=I_S\mathbf{X}^*C.$$

Proof. Note first that to say that $T(\mathbf{X}, f)$ is a ξ^*-best predictor of $\mathbf{l}'\mathbf{y}$ is equivalent to saying that $T(\mathbf{X}, f) - \mathbf{l}'I_S\mathbf{y}$ is a best linear unbiased estimator for its expectation under the ξ^*-model. By the above Lemma this holds if and only if $I_S\mathbf{V}^*I_S\mathbf{B}'\mathbf{X}'(I - I_S)\mathbf{l}$ belongs to the subspace generated by the columns of $I_S\mathbf{X}^*$. By the definition of $\mathbf{B}$, $I_S\mathbf{B}'$ simplifies to $\mathbf{B}'$ and then $I_S\mathbf{V}^*I_S\mathbf{B}' = I_S\mathbf{V}^*\mathbf{B}'$ which completes the necessity part of the proof. To obtain the sufficiency we use Theorem 2 to conclude that $T(\mathbf{X}, f)$ is ξ^*-unbiased and again the above Lemma to guarantee that $T(\mathbf{X}, f) - \mathbf{l}'I_S\mathbf{y}$ is a best linear unbiased estimator of its expectation under the ξ^*-model.

In order to illustrate the above results we present the following examples.

Example 2 (Tallis, 1978). Here, the elements of the models are

$$\mathbf{l} = 1_N, \quad \mathbf{X} = 1_N, \quad \mathbf{V} = I, \quad \mathbf{Z} = \begin{bmatrix} x_1 & x_1^2 & \dots & x_1^J \\ \vdots & \vdots & & \vdots \\ x_N & x_N^2 & \dots & x_N^J \end{bmatrix}.$$

Note that $\mathbf{B} = (1/n)1_N'I_S$ and if equation (2) holds then

$$\frac{1}{n}\sum_{k=1}^{N} i_k x_k^j = \frac{1}{N-n}\sum_{k=1}^{N}(1-i_k)x_k^j$$

for every $j = 1, \dots, J$. This system of equations is the balanced-sample property (Royall & Herson, 1973a). It follows then that

$$T(\mathbf{X}, f) = \frac{N}{n}\sum_{k=1}^{N} i_k y_k,$$

which is known as the expansion predictor, is the ξ-best predictor of $T = \sum y_k$, where the sum is over $k = 1, \dots, N$, and with a balanced sample it becomes ξ^*-unbiased. By applying Theorem 3, it follows that $Nn^{-1}\sum i_k y_k$, where the sum is over $k = 1, \dots, N$, is the ξ^*-best predictor of T if and only if the sample is balanced and

$$f^*(x_k) = \sum_{j=0}^{J} c_j x_k^j$$

for a vector $(c_0, c_1, \dots, c_J)$ and for every k such that $i_k = 1$ (sample element).

Example 3 (Rodrigues, 1983). Suppose that the elements of the model are now

$$\mathbf{X} = \mathbf{l} = 1_N, \quad \mathbf{V} = \operatorname{diag}(f(x_1) \dots f(x_N)), \quad \mathbf{Z} = \begin{bmatrix} x_1 & x_1^2 & \dots & x_1^J \\ \vdots & \vdots & & \vdots \\ x_N & x_N^2 & \dots & x_N^J \end{bmatrix}.$$

This implies that

$$\mathbf{B} = \left(\sum_{k=1}^{N} \frac{i_k}{f(x_k)}\right)^{-1}\left(\frac{i_1}{f(x_1)}, \dots, \frac{i_N}{f(x_N)}\right)$$

and if equation (2) holds we have, in the place of the balanced sample, that

$$\left(\sum \frac{i_k}{f(x_k)}\right)^{-1}\sum \frac{i_k x_k^j}{f(x_k)} = \frac{\sum (1-i_k)x_k^j}{N-n} \quad (j = 0, 1, \dots, J),$$

where the sums are over $k = 1, \dots, N$. This is named, here, the extended-balanced-sample property since when $f(x) = 1$ it reduces to the balanced-sample case. By Theorem 2, this

extended-balanced-sample property is necessary and sufficient to make the ξ-best predictor

$$T(\mathbf{X}, f) = \sum_{k=1}^{N} i_k y_k + (N-n)\left(\sum_{k=1}^{N} \frac{i_k}{f(x_k)}\right)^{-1}\left(\sum_{k=1}^{N} \frac{i_k y_k}{f(x_k)}\right)$$

a ξ^*-unbiased predictor. By Theorem 3 it becomes ξ^*-best for predicting T if and only if the sample is extended balanced and

$$f^*(x_k) = f(x_k) \sum_{j=0}^{J} c_j x_k^j$$

for a vector $(c_0, c_1, \ldots, c_J)$ and for every k such that $i_k = 1$.

Example 4 (Scott, Brewer & Ho, 1978). In Example 3 consider the first column of $\mathbf{Z}$ as 1_N and

$$\mathbf{X} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}$$

in the place of 1_N.

Now we have

$$\mathbf{B} = \left[\sum_{k=1}^{N} \frac{i_k x_k^2}{f(x_k)}\right]^{-1}\left(\frac{x_1 i_1}{f(x_1)}, \ldots, \frac{x_N i_N}{f(x_N)}\right)$$

and if equation (2) holds, then

$$\left[\sum_{k=1}^{N} \frac{i_k x_k^2}{f(x_k)}\right]^{-1}\left[\sum_{k=1}^{N} \frac{i_k x_k^{j+1}}{f(x_k)}\right] = \left[\sum_{k=1}^{N} (1-i_k)x_k\right]^{-1}\left[\sum_{k=1}^{N} (1-i_k)x_k^j\right]$$

for $j = 0, 1, \ldots, J$. This system of equations defines a generalization of the overbalanced-sample property $[f(x) = x^2]$ of Scott, Brewer & Ho (1978). Here this general property is named extended overbalanced. Note that if f is the identity function, then this property reduces to the balanced-sample case. With the extended-overbalanced property, the linear predictor

$$T(\mathbf{X}, f) = \sum_{k=1}^{N} i_k y_k + \left[\sum_{k=1}^{N} \frac{i_k x_k^2}{f(x_k)}\right]^{-1}\left[\sum_{k=1}^{N} (1-i_k)x_k\right]\left[\sum_{k=1}^{N} \frac{i_k x_k}{f(x_k)} y_k\right]$$

besides being ξ-unbiased is ξ^*-unbiased in predicting T. By Theorem 1, it is ξ-best for T and, by applying Theorem 3, it becomes ξ^*-best if and only if the sample is extended overbalanced and

$$f^*(x_k) = f(x_k) \sum_{j=0}^{J} c_j x_k^{j-1}$$

for a vector $(c_0, c_1, \ldots, c_J)$ and for every k such that $i_k = 1$.

As a final remark we would like to emphasize that the existence of $C = (c_0, c_1, \ldots, c_J)$ that appears in the above examples is needed for the sample only, not for the whole population. We did not use a notation showing this because, as we have suggested in § 1, the sample may be considered as fixed at the time of the inference. On the other hand, the properties of balanced sample, overbalanced sample, extended-balanced sample and extended-overbalanced sample introduce a way of qualifying a sample and suggest that intentional sampling may be the procedure to be followed when auxiliary information (here the matrices $\mathbf{X}$ and $\mathbf{X}^*$) has been used.

Acknowledgements

The authors appreciate very much the encouragements they have been receiving from Professor Richard Royall. We are also thankful to Professor José Ferreira de Carvalho and the referees for their helpful suggestions.

References

Basu, D. (1969). Role of the sufficiency and likelihood principles in sample survey theory. *Sankhyā* A **31,** 441–454.

Basu, D. (1975). Statistical information and likelihood. *Sankhyā* A **37,** 1–71.

Cassel, C.M., Sarndal, C.E. & Wretman, J.H. (1977). *Foundations of Inference in Survey Sampling.* New York: Wiley.

Kruskal, W. (1968). When are Gauss–Markov and least squares estimators identical? A coordinate-free approach. *Ann. Math. Statist.* **39,** 70–75.

Rao, C.R. (1971). *Linear Statistical Inference and its Applications.* New York: Wiley.

Rodrigues, J. (1983). Robust estimation and finite population. *Prob. Math. Statist.* **4**. To appear.

Royall, R.M. (1976). The linear least squares prediction approach to two-stage sampling. *J. Am. Statist. Assoc.* **71,** 657–664.

Royall, R.M. & Herson, J. (1973a). Robust estimation in finite population I. *J. Am. Statist. Assoc.* **68,** 880–889.

Royall, R.M. & Herson, J. (1973b). Robust estimation in finite population II. *J. Am. Statist. Assoc.* **68,** 890–893.

Scott, A.J., Brewer, K.W. & Ho, E.W. (1978). Finite population sampling and robust estimation. *J. Am. Statist. Assoc.* **73,** 359–361.

Tallis, G.M. (1978). Note on robust estimation in finite populations. *Sankhyā* C **40,** 136–138.

Zyskind, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *Ann. Math. Statist.* **38,** 1092–1109.

Résumé

Cette étude montre comment un résultat obtenu par Zyskind (1967) pour la caractérisation des meilleures éstimateurs linéaires de modèles linéaires généraux peut être utilisé dans la généralisation et unification des résultats de la théorie de la prédiction linéaire pour des populations finies. Le Prédicteur Linéaire Robuste est caractérisé dans une situation générale. A l'aide d'indicateurs des éléments des échantillons, nous montrons que l'ordre naturel des caractéristiques numériques de la population peut être conservé tout au long de l'analyse.

[*Paper received June* 1982, *revised May* 1983]

http://www.jstor.org

LINKED CITATIONS

- Page 1 of 2 -

You have printed the following article:

Robust Linear Prediction in Finite Populations

Carlos Alberto de Bragança Pereira; Josemar Rodrigues

International Statistical Review / Revue Internationale de Statistique, Vol. 51, No. 3. (Dec., 1983), pp. 293-300.

Stable URL:

http://links.jstor.org/sici?sici=0306-7734%28198312%2951%3A3%3C293%3ARLPIFP%3E2.0.CO%3B2-6

This article references the following linked citations. If you are trying to access articles from an off-campus location, you may be required to first logon via your library web site to access JSTOR. Please visit your library's website or contact a librarian to learn about options for remote access to JSTOR.

References

When are Gauss-Markov and Least Squares Estimators Identical? A Coordinate-Free Approach

William Kruskal

The Annals of Mathematical Statistics, Vol. 39, No. 1. (Feb., 1968), pp. 70-75.

Stable URL:

http://links.jstor.org/sici?sici=0003-4851%28196802%2939%3A1%3C70%3AWAGALS%3E2.0.CO%3B2-E

The Linear Least-Squares Prediction Approach to Two-Stage Sampling

Richard M. Royall

Journal of the American Statistical Association, Vol. 71, No. 355. (Sep., 1976), pp. 657-664.

Stable URL:

http://links.jstor.org/sici?sici=0162-1459%28197609%2971%3A355%3C657%3ATLLPAT%3E2.0.CO%3B2-Q

Robust Estimation in Finite Populations I

Richard M. Royall; Jay Herson

Journal of the American Statistical Association, Vol. 68, No. 344. (Dec., 1973), pp. 880-889.

Stable URL:

http://links.jstor.org/sici?sici=0162-1459%28197312%2968%3A344%3C880%3AREIFPI%3E2.0.CO%3B2-E

http://www.jstor.org

LINKED CITATIONS

Robust Estimation in Finite Populations II: Stratification on a Size Variable

Richard M. Royall; Jay Herson

Journal of the American Statistical Association, Vol. 68, No. 344. (Dec., 1973), pp. 890-893.

Stable URL:

http://links.jstor.org/sici?sici=0162-1459%28197312%2968%3A344%3C890%3AREIFPI%3E2.0.CO%3B2-B

Finite Population Sampling and Robust Estimation

A. J. Scott; K. R. W. Brewer; E. W. H. Ho

Journal of the American Statistical Association, Vol. 73, No. 362. (Jun., 1978), pp. 359-361.

Stable URL:

http://links.jstor.org/sici?sici=0162-1459%28197806%2973%3A362%3C359%3AFPSARE%3E2.0.CO%3B2-M

On Canonical Forms, Non-Negative Covariance Matrices and Best and Simple Least Squares Linear Estimators in Linear Models

George Zyskind

The Annals of Mathematical Statistics, Vol. 38, No. 4. (Aug., 1967), pp. 1092-1109.

Stable URL:

http://links.jstor.org/sici?sici=0003-4851%28196708%2938%3A4%3C1092%3AOCFNCM%3E2.0.CO%3B2-K