Athens Institute for Education and Research
ATINER

ATINER's Conference Paper Series
MAT2013-0682

Interactive Linear Models in Survey Sampling

Pulakesh Maiti
Associate Professor
Indian Statistical Institute
India

Athens Institute for Education and Research
8 Valaoritou Street, Kolonaki, 10671 Athens, Greece
Tel: + 30 210 3634210 Fax: + 30 210 3634209
Email: info@atiner.gr URL: www.atiner.gr
URL Conference Papers Series: www.atiner.gr/papers.htm

Printed in Athens, Greece by the Athens Institute for Education and Research.

ISSN **2241-2891**
4/11/2013

An Introduction to ATINER's Conference Paper Series

ATINER started to publish this conference papers series in 2012. It includes only the papers submitted for publication after they were presented at one of the conferences organized by our Institute every year. The papers published in the series have not been refereed and are published as they were submitted by the author. The series serves two purposes. First, we want to disseminate the information as fast as possible. Second, by doing so, the authors can receive comments useful to revise their papers before they are considered for publication in one of ATINER's books, following our standard procedures of a blind review.

This paper should be cited as follows:

Maiti, P. (2013) **"Interactive Linear Models in Survey Sampling"** Athens: ATINER'S Conference Paper Series, No: MAT2013-0682.

Interactive Linear Models in Survey Sampling

Pulakesh Maiti
Associate Professor
Indian Statistical Institute
India

Abstract

Considered is a linear 'interactive' model in the context of survey sampling. This situation arises when investigator and/or supervisor interventions are contemplated in the responses. Blinded situation has been discussed herein.

Considered is the set-up of simple i.e., direct response on a quantitative response variable Y in the context of a finite labeled population of size N.

It so happens that in actual surveys, we need investigators and often some supervisors as well. We depict a situation wherein there are possibilities of investigators' and / or supervisors' intervention effects on the response profile finally received by the data collection agency. Of course, these effects may be assumed to be random, having mean zero, non-interactive within and between the two sets of 'people'.

The problem is to unbiasedly estimate the finite population total of the response variable Y by incorporating a fixed size (n) sampling design and by administering the survey design in situations where in the above two types of random effects are likely to be present.

Keywords:

Corresponding Author:

Introduction

Denote by 'i' a Responding Unit [RU] in the sample of size n and by S[i] the number of schedule-based observations collected on this particular unit. Naturally, S[i] is based on the 'survey design' used for this unit in combination with the investigators and the supervisors.

We may write S[i] = ΣΣ I[i; (j; k)]
where I[i; (j; k)] = 1 if (j; k)-combination of the investigator and the supervisor have both worked on a schedule assigned to the ith responding unit.

Naturally, for any triplet [i; (j; k)]; I[i; (j; k)] >= 0 while S[i] > 0 for each responding unit. Whenever I[i; (j; k)] = 1, we will denote by Y[i;(j;k)] the underlying response on the study variable.

Model for Intervention Effects

Consider a finite population of N units and let us adopt an SRSWOR(N, n) sample of size 'n'. Denote by Y[i] the response on the ith responding unit; i=1, 2, …, n i.e., the 'data' accrued from the field.

Without any intervention effect on the part of the investigators/supervisors, we would have regarded the above data as 'error-free' and so usual estimation techniques could be routinely used. Thus, for example, sample mean would be the usual unbiased estimator for the population mean.

However, we want to examine the possibility of intervention by one or the other group or possibly by both and so we postulate a linear model of the following form, as applied to Y[i;(j;k)]:

Y[i;(j;k)] = TR[i] + IR[j] + S[k] + e[i;(j;k)]
where
TR[i] is the true response from Respondent labelled 'i';
IR[j] is the intervention effect of Investigator labeled 'j' and
S[k] is that of the Supervisor labeled 'k'.
The last term is the so-called error term.

As usual, we assume that the errors and the intervention effects are all randomly distributed with means 0's, variances σ_e^2; σ_{IR}^2 and σ_S^2 respectively while all pairwise effects / interventions are uncorrelated.

Model & Data Perspective

At this stage, we need to differentiate between two distinct scenarios:
(i) Blinded Submission;
(ii) Unblinded Submission.

In case the submission is blinded, each supervisor treats each response profile as a separate document and treats it as an isolated document - without the knowledge of identification of the interviewer/investigator.

In the other case, the supervisor also receives information about the identity of the interviewer/investigator along with response profiles.

I will only discuss the first scenario.

Illustrative Example

To fix ideas, we consider a simple example of N=700 Respondents, clustered in M=70 Large Units of 10 each. We treat the Clusters as 'Responding Unit [RU]' for our study and draw SRSWOR(M=70,m=7) Clusters [so that effectively we have n = 70 ultimate units]. We consider 7 Investigators and 2 Supervisors and follow the network of RU versus Investigator versus Supervisor as exhibited in the following Table
Intervention Network:

(I) : (j = 1; k = 1); (j = 5; k = 2); (j = 7; k = 2);
(II) : (j = 1; k = 1); (j = 2; k = 1); (j = 6; k = 2);
(III) : (j = 2; k = 1); (j = 3; k = 1); (j = 7; k = 2);
(IV) : (j = 1; k = 1); (j = 3; k = 1); (j = 4; k = 1);(j = 4; k = 2);
(V) : (j = 2; k = 1); (j = 4; k = 1); (j = 4; k = 2); (j = 5; k = 2);
(V I) : (j = 3; k = 1); (j = 5; k = 2); (j = 6; k = 2);
(V II) : (j = 4; k = 1); (j = 4; k = 2); (j = 6; k = 2);(j = 7; k = 2)

Note : Derived from the BIBD(7, 7, 3, 3, 1)

Interpretation

There are 3 data points for the first Respondent-Set I - as collected independently by the investigators 1; 5; 7. Both the supervisors are involved for further processing of the 3 responses derived by the 3 Investigators. While Supervisor # 1 deals with data collected by Investigator # 1, the other two responses are handled by the Supervisor # 2. For Blinded Submission, we straightaway take the average of the three responses and use this as the representative figure for the first responding set/unit. This we do for all other responding sets as well.

Processing of Data

Note that there are altogether 24 data points and the respondent unit-wise frequency distributions are given by 3; 3; 3; 4; 4; 3; 4 respectively. We denote by Y the vector of 24 observations and by A the 7 X 24 incidence matrix of the population units versus the observations as per the Survey Design.

We derive below Model Expectation and Model Variance of sample means for each respondent unit.

Performance of Sample Means

Model Assumptions IMPLY:
Model Expectation = True Value
Computations of the model-based variances and covariances are quite involved.
For example,
Σ11=dispersion matrix of Y[I;(1;1)]; Y[I;(5;2)]; Y[I;(7;2)]
= dispersion matrix of (IR[1] + S[1] + e[I:(1;1)]; IR[5] + S[2] + e[I;(5;2)]; IR[7] + S[2] + e[I;(7;2)])

Computation of Σ11

$$\begin{bmatrix} \sigma_e^2+\sigma_{IR}^2+\sigma_S^2 & 0 & 0 \\ 0 & \sigma_e^2+\sigma_{IR}^2+\sigma_S^2 & \sigma_S^2 \\ 0 & \sigma_S^2 & \sigma_e^2+\sigma_{IR}^2+\sigma_S^2 \end{bmatrix}$$

Therefore, $V_M\ Y(I..)$ is given by

$$\left[3\sigma_e^2+3\sigma_{IR}^2+5\sigma_S^2\right]/9\,.$$

Covariance Computations

Likewise, all variance terms can be computed.

Next we need to compute all model-based covariances of sample means for the 7 responding units.
For example,
Σ12 = Cov_M ([Y[I;(1;1)]; Y[I;(5;2)]; Y[I;(7;2)]]; [Y[II;(1;1)]; Y[II;(2;1)]; Y[II;(6;2)]]) =

$$\begin{bmatrix} \sigma_{IR}^2+\sigma_S^2 & \sigma_S^2 & 0 \\ 0 & 0 & \sigma_S^2 \\ 0 & \sigma_S^2 & \sigma_S^2 \end{bmatrix}$$

Therefore, Cov_M (Y[I]..;Y[II]..) = 1′(Σ12)1/9 = $\left[\sigma_{IR}^2+4\sigma_S^2\right]/9$.

Similarly, the rest can be computed.

Data Analysis Under Blinded Submission

As usual, estimate of Finite Population Total T(TR) is provided by

$\hat{T}(TR) = \Sigma$ Y[i]../∏[i];

$V(\hat{T}(TR))$ has 2 components :

$V(\hat{T}(TR)) = V_1 E_2 + E_1 V_2$

E_2 & V_2 refer to Model Exp. & Model Var.

E_1 & V_1 : Design-based Exp. & Var. require standard computations;

E_2 provides TR-values qnd

$V_2 = \Sigma V_M$ [Y[i]..]/$\Pi^2[i]$

+ Σ Σ Cov_M [Y[i].., Y[j]..]/ ∏[i] ∏[j]

All components have been evaluated.

V_1 E_2 needs a careful handling since $TR^2_{[i]}$ are involved.

It is the difference between two expressions given by
First Expression:

M^2(1/m-1/M)[Σ $(Y_{[i]}.. - Y_{[i]}..)^2$ /m(m-1)];

Second Expression : M^2 $(1/m - 1/M)$ times

[(m-1) Σ σ_{ii} – Σ Σ σ_{ij} /m(m-1)]

Under the assumed model, σ_{ii} & σ_{ij} have been computed.

$\Sigma\ \sigma_{ii}$ = [25/12] $\sigma_e^2 + [59/24]\ \sigma_{IR}^2 + [143/36]\sigma_S^2$;

$\Sigma\Sigma\sigma_{ij} = [22/9]\ \sigma_{IR}^2 + [191/36)]\sigma_S^2$.

Final Results
Estimate of Finite Popl. Total = M (Σ Y[i]../m)
Variance Estimate is in TWO PARTS :
First Part : Usual Contribution from the data given by

M^2 (1/m-1/M) ΣΣ $(\text{Y[i]..-Y[j]..})^2$/m(m-1);

Second Part : contribution from variance components given by

[M/m] $\Sigma\sigma_{ii} + [M(M-1)/m\ (m-1)\Sigma\Sigma\sigma_{ij}]$

which simplifies to [M/m] [(25/12) $\sigma_e^2 + (59/24)\sigma_{IR}^2 + (143/36)\sigma_S^2$]

+ [2 M (M-1)/m(m-1)] [(22/9) $\sigma_{IR}^2 + (191/36)\ \sigma_S^2$].

References

Dalenius Tore(1974). The Ends and Means of Total Survey Design. Stockholm: The University of Stockholm.

Hansen, Morris H. William N. Hurwitz, Etes. Marks, And W. Parker Mauldin(1951). Response Errors in Survey, JASA, 46, 147-190.

Hartley, H.O. and Rao. J.N.K.(1978). The Estimation of Non Sampling Variance Components in Sample Surveys, N. Krishnam Namboodiri, ed. Survey Sampling and Measurement, New York; Academic.