NONSAMPLING ERRORS IN AN AGRICULTURE SURVEY

David Faulkenberry, Oregon State University and Robert Tortora, U.S. Department of Agriculture

One of the main concepts in statistics, and the basis for much of our inference theory, is the notion of an internal measure of variability calculated from a particular data set as the basis for assessing the variability one would expect if the study were repeated. Yet statisticians realize that whether we are dealing with a field trials experiment or a public opinion survey, we often do not have a realistic assessment of the "replication effect" in the sense of another replication of the study. That is, we know that it is not terribly unusual for two different studies that were supposed to be estimating the same population characteristics to give somewhat conflicting reports. We also realize that in any particular statistical study these nonsampling errors, even though they exist in the simplest of measurement processes (Faulkenberry and Tortora (1979)), may be difficult if not impossible to identify and measure. Even though this is the case it is still important that we experiment with ways of demonstrating, measuring, and modeling nonsampling error so that we have a more realistic notion of their effort on statistical inferences and are aware of the limitations of the typical measure of error.

This talk is about a study we were involved in with the U.S. Department of Agriculture where, for some of the variables measured in the survey, we had an opportunity to get an alternative measure to compare with the survey responses. What I would like to do today is to make some general comments about nonsampling errors, then discuss some of the concepts, methods, and problems of the particular nonsampling errors study we were involved with. I will not say much about particular numberical results, and I don't think these are as important as the study itself, but for those of you interested there should be a USDA report out this coming fall or winter.

General Remarks on Nonsampling Errors

Hansen, Hurwitz, and Pritzker give three factors they consider relevant to evaluating the outcome (statistics) of a statistical survey:

(1) ideal goals - defined or implied by the purposes to be served by the survey
(2) defined goals - more operationally feasible
(3) expected values - expectation of the operations actually carried out.

In terms of observations or measurements we might rephrase these as (1) the ideal measurement, (2) the agreed upon measurement, and (3) the actual measurement. Anderson et. al. give the following diagram of possible error sources in a survey. This illustrates the overall complexity. Realistically, we should place nonsampling errors in the discrepancy between (2) and (3). This is something that has to be assessed and accepted. (See Figure 1)

The Study

The purpose of the study reported on here was to measure and analyze the effect of some of the nonsampling errors in the Farm Production and Expenditures Survey which is an annual survey conducted by the U. S. Department of Agriculture. The study was made possible by using data from the Farm Management System at Kansas State University. When the FPES was conducted in Kansas in February and March of 1979, a sample of farm operators that use the Kansas State Farm Management System was placed in the regular FPES sample. These farm operators were used in the survey as regular FPES sample units and the questionnaires went through the same processing. The enumerator did not know that these units were from the KSU Management System where accurate records were available and that these questionnaires would be analyzed separately. After completion of the survey and processing of the data, the KSU Management System operator's questionnaires were taken out. Survey data obtained from these farm operators was then compared with their data from the KSU Farm Management System as a basis for the study. The farm operators belonging to the KSFMS are not expected to be typical of the population of farm operators in Kansas, so it is not intended that the inference apply to the whole population.

Now the type measurements made in the FPES are more precise than, say, those in a survey of public opinion but less precise than those of a survey of age. That is, there is a distinction between (1) and (2), i.e., the ideal or conceptual vs the operational measure. What we are saying then is that the KSFMS data, since it is the basis for financial statements for many of these operations, is a reasonable concept of an operational measure for many of the variables measured in the survey. Taking the KSFMS data then as a reasonable defined goal, we used the difference as a measure of nonsampling errors.

Sample Design

In order to provide a sample with a wide dispersion of type of farm operator the list of farm operators in the KSFMS was stratified by value of gross farm income and by the geographical region. A stratified random sample was taken and these operators were included as part of the FPES. Sample size and response rate information is given in the following table:

Total Sample	Completed Interviews	Refusals	Inaccessable	Dropouts from KSFMS
150	75*	28	29	18

*Five completed did not have KSU records so this left 70 for analysis. One other questionnaire was later dropped as a result of the outlier analysis.

Initial considerations of the FPES questionnaire and the KSFMS data resulted in the selection of 38 variables that were thought to be comparable. Ten of these were later dropped from the study either because further analysis showed that for these variables the KSFMS values were not really comparable to the FPES values or because there was an insufficient number of observations on the variable.

Data Analysis

Analysis of the data from the study consisted of the following steps:

1. Description
   a. Descriptive statistics of the FPES values, the KSFMS values, and the difference.
   b. Histograms of the above three variables.
   c. Scatter plots of FPES vs KSFMS

2. Outlier analysis
   It became clear that some of the differences in FPES values and KSFMS values were unusual and required special attention. A straightforward outlier test was applied to screen the data and identify extreme values. The questionnaires and KSFMS values associated with these extreme values were rechecked to determine if there was some error or if the value seemed legitimate.
3. After editing the data with the outlier analysis, estimates of bias and hypothesis tests were made.
4. Estimates of population totals for the population consisting of KSFMS operators were made to determine the effect of bias on the usual inference statements.

To give you an idea of the type data we are talking about, I have chosen two variables to give some results for:

(1) Total cattle and calves
(2) Total fuel and petroleum

We would expect the survey information on number of cattle and calves to be more accurate than fuel expense. Summary statistics for the data (unweighted) are given in the following tables and figures.

Total Cattle and Calves

	Mean	Standard Deviation	Median	Range
Survey	183.3	243.2	95.0	1170.0
Management Information Sys.	172.7	232.5	92.0	1020
Difference	10.4	99.4	0.0	714.0

Total Fuel and Petroleum

	Mean	Standard Deviation	Median	Range
Survey	7188	8208	4518	46,516
Management Information Sys.	5099	4546	3640	21,862
Difference	1295	4085	502	30,569

(See Figures 2 and 3)

As can be seen from a preliminary look at this data, while there is fairly good agreement, there are some large differences, there is an estimated positive bias (Survey - Information System), and there is an additional variance component contribution. The data of this type for all variables was carefully screened (which is quite tedious and time consuming) and weighted estimates of bias and variance components were calculated.

Conclusions - Questions - Shortcomings

It is unlikely that this particular study will lead to adjustments in the survey estimates, and we would not recommend it because

(1) the study was primarily exploratory
(2) there is some question as to how comparable the KSFMS and FPES data should be, and
(3) the number of usable cases is large enough to test some hypotheses but not large enough to get precise estimates of the various components in the total error.

What then is the value, if any, of studies such as this one? I think there are several. It provides additional empirical information in an area of inquiry where there is little and where there is much needed. It focuses attention on the differences among

(1) the ideal measurement
(2) the agreed to measurement, and
(3) the actual measurement

It may lead to questions that would result in an improved questionnaire, result in improved editing procedures, or a decision to measure different variables, or it may encourage the combination of theoretical and empirical work necessary to develop a more realistic error model and designs to help fill in the components in this model.

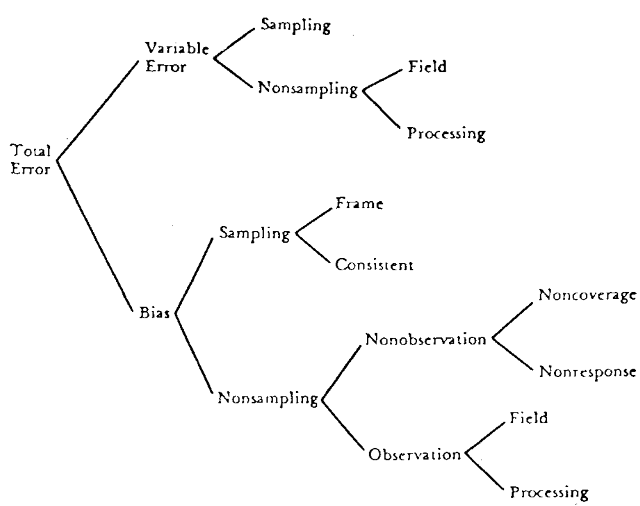
In any event, this is a most important area of inquiry for statisticians to concern themselves with. This is especially true in large surveys where we wish precise estimates and realistic assessments of this precision. The problems are very difficult but each study of this type provides a little more insight and awareness.

References

Andersen, R., J. Kasper, M. Frankel, and Associates. Total Survey Error. Jossey-Bass Publishers, 1979.

Faulkenberry, G. D. and R. D. Tortora. A Study of Measurement Error Suitable for a Classroom Example. The American Statistician, February, 1979.

Hansen, M., W. N. Hurwitz, and L. Pritzker. Standardization of Procedures for the Evaluation of Data: Measurement Errors and Statistical Standards in the Bureau of the Census. Report on the American Statistical Assn. Conference on The Transfer of Methodology Between Academic and Government Statisticians. March, 1978.



Classification of Sources of Survey Error

Figure 1

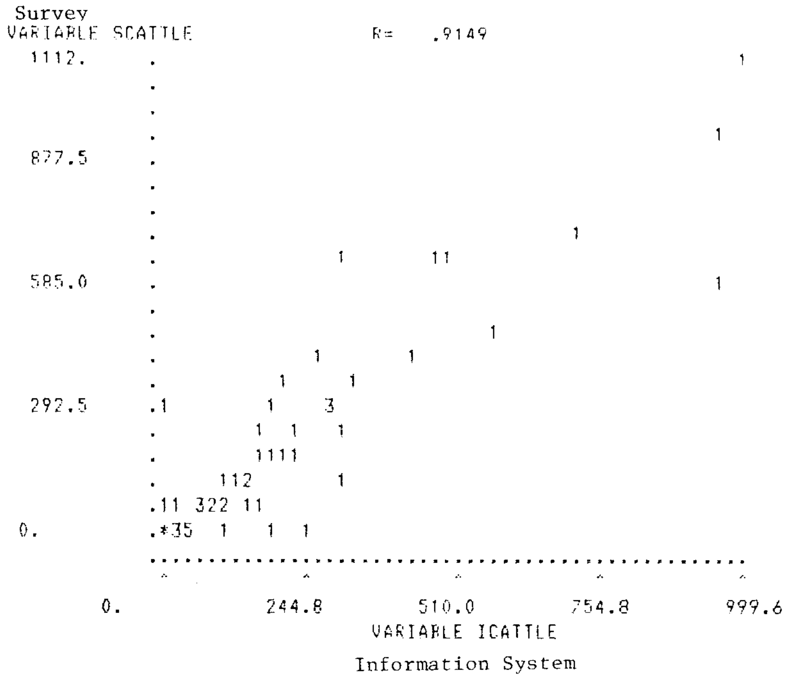


Figure 2

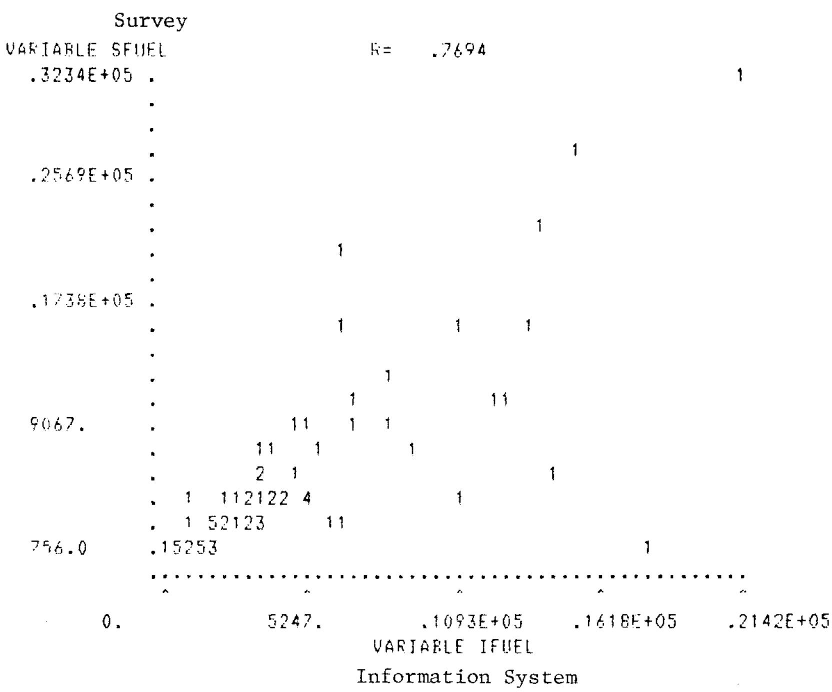


Figure 3