ESTIMATORS AND THE ESTIMATION PROCESS

by Charles R. Perry, Jr.

The objective of any survey is to provide information on the characteristics of the population by examining a subset of the population. By analyzing the data from that subset - called a sample - we try to estimate population parameters such as means, totals, ratios, regression coefficients, etc. The goal of estimation is to wring from the sample all possible clues about the characteristics of the population while recognizing that the sample has limits to what it can tell us.

What factors limit the information that the sample can provide? There are two primary limiting factors. The first is sampling error. Sampling error is controlled mostly through the sample design, especially the sample size. The second limiting factor is the nonsampling error, or bias, that affects the data. Nonsampling error actually causes a sample to misrepresent the true characteristics of the population. The sum of the sampling error and the nonsampling error defines the total error associated with a particular sample. This total error limits how much useful information can be obtained from the sample about the population.

The estimation process has two components. The estimator, which is usually a formula or set of formulas, dictates how to calculate the estimate from the sample data. In other words, the formula(s) is the estimator and the actual number produced from the sample data is the estimate. NASS has created a variation on this terminology by using indication to refer to the number produced from the data and letting estimate refer only to the Agricultural Statistics Board estimate - the official number that is set after reviewing all the indications.

There are two types of indications in general use - point indications and interval indications. A point indication is simply a single number calculated from the sample data. An interval indication has two numbers to provide upper and lower bounds on the population parameter. A confidence interval is an example of an interval indication. Although interval indications are a more sophisticated tool, NASS uses point indications for most of its work. However, whenever someone starts discussing CV's, that person is edging toward the interval concept rather than the point concept.

It is possible to have several estimators even with the same sample design. For example, with our area sample we can have a tract estimator, an entire farm estimator, and a weighted estimator. How does one determine the best estimator? Statistical theory has developed a large amount of literature concerning the qualities that make a" good" estimator. Statistical properties such as variance, bias, mean square error, efficiency, sufficiency, and consistency as well as practical considerations such as cost have received a great deal of attention. It is difficult to discuss those issues without using a great many formulas. However, I should point out that different estimators may be "best" in different situations. For example, a

tract estimator of total hogs may perform fairly well, but a tract estimator of pig deaths may perform poorly. Thus, to choose one "best" estimator for all situations becomes very difficult.

The Board process is a tool used by NASS to evaluate multiple indications along with any other available information in order to produce an official estimate. Therefore, in the Board process, even if nowhere else, the relative strengths and weaknesses of each indication must be taken into account.

NASS research has devoted considerable effort toward the investigation of analytic procedures that integrate or partially integrate survey estimates, control data, and other information. The aim is not to have analytic techniques replace the Board process but to provide the Board with more reliable inputs.

A joint effort between the Research and Development Division and the Livestock Branch investigated the use of composite estimators for the June hog series. The aim of this work was to develop analytic procedures for combining farm, tract, weighted, and multiple frame indications in a single indication for use by the Board. Procedures were developed and tested using historical data for eight of the major hog producing States. These composite hog indications were used operationally with satisfactory results in June 1988. NASS has also implemented the use of composite indications for some crop variables.

Estimation is the final step of the survey process - the final fruit of a carefully cultivated tree. During estimation NASS must decide what information has been learned about the agricultural situation. With precise, accurate, and relevant information, we can begin to interpret and understand the ever changing world of agriculture.