The Survey As A Measurement Instrument

By Earl E. Houseman

Cost-benefit ratios for surveys are related to the congruence between objectives and survey measurement capabilities. A high degree of congruence requires careful matching of objectives and survey design. Thus, emphasis is placed on survey planning and improvement of relationships between survey designs and objectives. The degree of congruence has a significant effect on the efficiency of research and the interpretations and value of results.

Key words: Survey techniques; survey planning; probability sampling; statistical inference; evaluating survey data; congruence.

There should be better communication between agricultural economists as users and survey statisticians as producers of statistical information. Owing to differences in training, duties, and philosophic principles, they have different perspectives. Both could benefit from freer exchange of views in the interest of improving surveys as measurement instruments for various purposes. Admittedly, many of the remarks in this article are impressions and hypotheses emerging from years of experience as a sampling statistician and consultant. I hope some of the items or points of view that are presented will be helpful to social scientists and will suggest some avenues for exploration. I also hope that this article will contribute to better understanding and cooperation between social scientists and survey statisticians.

Introduction

The survey is used to measure numerous kinds of characteristics or quantities pertaining to a wide variety of populations. We have much commonsense understanding of the capability of various physical measurement devices and we appreciate the need to choose an instrument that is suitable for the purpose. For example, it might be important to find out whether the difference in diameters of two automobile pistons, machined to the same specifications, is less than some specified amount. Would you use an ordinary yardstick? The incongruence between the objective and the measurement tool is obvious. When surveys are involved, congruity between objectives and errors of measurement is often obscure depending upon intelligence about the capability of the survey. A major purpose of the science of survey technology is enabling better judgment in the development of survey plans in relation to objectives.

With reference to planning research, Cox[1] in 1951 stated: "The statistician who expects that his contribution to the planning will involve statistical theory, finds repeatedly that he makes his most valuable contribution simply by getting the investigator to explain why he is doing the experiment, to justify the experimental treatments and to defend his claim that the experiment, when completed, will enable its objectives to be realized." "Survey" may be substituted for experiment. Cox's statement contains a message for social scientists who are planning surveys and for professional survey statisticians. The writer's general experience is in accord with Cox's observation.

The experienced applied mathematical statistician is a student of variability (measurement error). He is a statistical engineer. A major part of his business is to understand variability and to know how to cope with it effectively. With experience he acquires an unusual insight of patterns and magnitudes of variability that commonly exist in the world we live in. He develops good judgment of the efficiency of alternative survey designs in relation to objectives. In most cases, he can provide good indications of what the precision (and to some extent the accuracy) of the results will be. In survey planning, this kind of expertise is one of the essential inputs to achieving a high degree of congruence (or a good match) between objectives and the measurement instrument.

With reference to measurement problems, it is possible to draw many analogies between the physical and social sciences. The quantities to be measured need

[1] Gertrude M. Cox, The Value and Usefulness of Statistics in Research. A lecture presented in the Department of Agriculture Auditorium, January 11, 1951.

to be defined or specified. The physical measurement instrument and the measurement "instrument" for a survey (namely the sample including the questionnaire, etc.) have similar roles. Skills and procedures involved in using measurement instruments have a bearing on the accuracy of the results. Errors of measurement, both constant (bias) and variable (random), always exist to some degree in either case. We have heard about calibration of physical measurement instruments. Surveys also need "calibration." However, the main point under consideration is the matter of achieving good congruity between objectives and the measurement instruments.

The writer is unable to state an exact definition of congruity; but it is clear that the cost-benefit ratio for a survey is related to the degree of congruence between objectives and the measurement capability of the survey. Degree of congruence is a major factor affecting efficiency of research, interpretation and value of results, and rate of progress. Hence, a good understanding of what is involved in good congruity and how to achieve it is needed.

Planners of surveys agree that a survey should make a contribution to knowledge, but they have widely varying views as to the kind of a survey that is appropriate for a particular purpose. A major responsibility of the statistician is to be as expert as possible regarding the capability of a survey as a measurement instrument and the best way to fashion it to assure success with reference to a particular set of objectives. In practice, cooperative team effort is usually needed to obtain congruity of objectives, design, and resources in the interest of achieving the most favorable cost-benefit ratio.

Perhaps some readers of this article are unfamiliar with the technical meaning of "precision" and "accuracy" as words describing properties of an estimate, so let's review the concepts briefly:

Precision of an estimate is measured in terms of its standard error, which is a measure of random variation of an estimate from its "expected" value. The expected value of an estimate might differ from the "target" value. The differences between the expected and target values is bias, or a constant component of error that is not measured by the standard error.

Accuracy is a measure of the total error of an estimate. It pertains to the possible deviations of an estimate from the target value and is a combination of the standard error and bias.

The use of the words "precision" and "accuracy" in statistics is analogous to their use in the physical sciences. One may speak of a precision instrument in a laboratory, meaning that the instrument is capable of making very exact measurements or detecting very small differences. But the readings will not necessarily be accurate unless the instrument is correctly calibrated and is functioning properly. An accurate instrument must be relatively free of bias as well as precise. A clock, for example, that is always exactly 10 minutes fast is precise but inaccurate. A sample may provide an estimate that is precise (low standard error), but high precision alone does not mean that the estimate is accurate. Note that, although we commonly speak of the "accuracy of an estimate," statements of accuracy or precision reflect attributes of the entire system that generates an estimate. System refers to the whole survey process; that is, sample design, editing specifications, method of estimation, the way questions are asked, etc.

Some Examples Involving Congruity

Can we agree that an estimate is of undetermined value and hence of no value when we have absolutely no knowledge whatever (not even subjective experience) about its accuracy? That is, the utility of an estimate is a function of the nature and amount of information we have about its accuracy, which depends upon the measurement instrument and the care with which it is used. Here are some examples that illustrate a few aspects of the congruity problem:

1. A survey was proposed to get a measure of an average cost per unit. The sample was rather small. Discussion revealed that the average cost was already known within about 10 or 15 percent. The survey was proposed because a much more accurate answer was needed, but the proposed survey design and sample size were such that the prospects were not favorable for getting better information. If the survey had gone forward on this basis, an answer that "looked" all right, or a "satisfying" result, might have been obtained. But how does one judge the accuracy of an estimate? This is a key point in achieving congruity. It is a question involving many facets of survey technology and hence a subject outside of the scope of this article.

Fortunately, with regard to the case just mentioned, it was possible to revamp data specification and analysis plans so there would be a direct tie with data that already existed for each unit in the population. This meant that the objective of acquiring an accurate estimate of the average unit cost could be fulfilled by adopting a more efficient design without increasing the sample size. Thus, when knowledge of measurement capability of the proposal was brought into consideration, a different view of the job was formulated and a measurement device was fashioned which would more closely meet the objectives. Frequently, in a case of this type, objectives must be curtailed or resources increased to achieve congruity.

Usually, the more that is known about the subject under investigation, the more complex and stringent the survey requirements become. So the task of planning (that is, adjusting objectives, design, and resources) can be a major undertaking—but an undertaking that is necessary to avoid fruitless effort or inefficient expenditure of resources. There have been cases where an investigator repeated essentially the same survey several times trying to get better answers to an important problem. Typically, the first survey or two contributed an appreciable amount of information. But, from there on, the additional surveys added very little because the efforts were the equivalent of trying to get a better measurement of the same quantity with the same yardstick. There are many variations of this type of situation—the general picture being that as information about a subject is acquired, the survey as a measurement instrument remains essentially unchanged and the returns from the efforts diminish. In this type of situation, congruity diminishes because improved measurement instruments are not employed to meet the more exact requirements that develop. In the allocation of funds for specific projects too much weight is often given to the importance of the problem and not enough to the prospects for a good return from the investment.

2. Let us look briefly at another type of case—one where the congruity was very good but the capability of the measurement instrument was not adequately understood. An investigator had worked very closely with a statistician on the design of an experiment to determine which of two methods of transportation had the lesser detrimental impact on the quality of a product. The design of the experiment was excellent and it was conducted with unusual care. The observed difference between the two methods was not statistically significant—a disappointing result because the investigator firmly believed that a difference between the two methods existed. The investigator returned to the statistician for advice on the number of additional replications that might be needed to achieve statistical significance. He had overlooked one important point; the experimental error was less than one-half of 1 percent. Thorough checking showed that the analysis and arithmetic were without error. Discussion between the investigator and the statistician led to a conclusion that the experiment was accurate enough to detect any difference between the two methods that was large enough to be of practical importance. Hence, at this decision point, the question of additional replication shifted to a question of whether to continue the work under another set of experimental conditions. Note how knowledge of the error associated with the measurement instrument influenced the interpretation of results, and changed the research objectives.

3. Knowledge of variance components can have a bearing on the effectiveness of research. For example, a research group was trying to develop better physical devices for sampling and estimating the quality (grade) of an agricultural product stored in bulk. It happened, owing to the group's lack of knowledge about various components of variance, that the research efforts were *not* oriented to reducing the one component of variance that was much greater than any other. After some fruitless effort, an analysis of variance pointed the direction for more productive research.

4. It is common for an investigator to propose a project wherein there is doubt about whether some of the objectives can be met. But let's consider those cases where even the key objectives are in such conflict with the capabilities of the proposed measurement instrument that the prospects for success are nil. When an investigator receives comments to that effect, one of two kinds of reaction is likely to be elicited. Some investigators are anxious to correct the situation and to reach common understanding about congruity. This type of investigator will usually have resolved matters of congruity between objectives and survey plans before arriving at a final proposal.

The other kind of reaction tends to be defensive. The investigator will explain the great importance of the survey, and the extent of the backing, as though the urgency were so overwhelming as to justify any survey tool. In this case, the investigator's own statements are sometimes incongruous. If the problem is of the great importance portrayed, then planning a survey that will do a satisfactory measurement job is also important. The time-worn argument that "a little information is better than none" is often used. It appears that some investigators haven't recognized that an estimate with low accuracy might have a negative value, or add little or nothing to what is already known. Please ponder the question: "What constitutes new or additional information?" As a minimum, one should try to plan a survey so there is a reasonable chance that it will make a positive contribution.

5. Sometimes the objectives of a particular project are limited more than necessary. To illustrate, let's assume a national survey of 20,000 households. A typical survey plan might call for allocating the sample

measurement capability of the proposal was brought into consideration, a different view of the job was formulated and a measurement device was fashioned which would more closely meet the objectives. Frequently, in a case of this type, objectives must be curtailed or resources increased to achieve congruity.

Usually, the more that is known about the subject under investigation, the more complex and stringent the survey requirements become. So the task of planning (that is, adjusting objectives, design, and resources) can be a major undertaking—but an undertaking that is necessary to avoid fruitless effort or inefficient expenditure of resources. There have been cases where an investigator repeated essentially the same survey several times trying to get better answers to an important problem. Typically, the first survey or two contributed an appreciable amount of information. But, from there on, the additional surveys added very little because the efforts were the equivalent of trying to get a better measurement of the same quantity with the same yardstick. There are many variations of this type of situation—the general picture being that as information about a subject is acquired, the survey as a measurement instrument remains essentially unchanged and the returns from the efforts diminish. In this type of situation, congruity diminishes because improved measurement instruments are not employed to meet the more exact requirements that develop. In the allocation of funds for specific projects too much weight is often given to the importance of the problem and not enough to the prospects for a good return from the investment.

2. Let us look briefly at another type of case—one where the congruity was very good but the capability of the measurement instrument was not adequately understood. An investigator had worked very closely with a statistician on the design of an experiment to determine which of two methods of transportation had the lesser detrimental impact on the quality of a product. The design of the experiment was excellent and it was conducted with unusual care. The observed difference between the two methods was not statistically significant—a disappointing result because the investigator firmly believed that a difference between the two methods existed. The investigator returned to the statistician for advice on the number of additional replications that might be needed to achieve statistical significance. He had overlooked one important point; the experimental error was less than one-half of 1 percent. Thorough checking showed that the analysis and arithmetic were without error. Discussion between the investigator and the statistician led to a conclusion that the experiment was accurate enough to detect any difference between the two methods that was large enough to be of practical importance. Hence, at this decision point, the question of additional replication shifted to a question of whether to continue the work under another set of experimental conditions. Note how knowledge of the error associated with the measurement instrument influenced the interpretation of results, and changed the research objectives.

3. Knowledge of variance components can have a bearing on the effectiveness of research. For example, a research group was trying to develop better physical devices for sampling and estimating the quality (grade) of an agricultural product stored in bulk. It happened, owing to the group's lack of knowledge about various components of variance, that the research efforts were *not* oriented to reducing the one component of variance that was much greater than any other. After some fruitless effort, an analysis of variance pointed the direction for more productive research.

4. It is common for an investigator to propose a project wherein there is doubt about whether some of the objectives can be met. But let's consider those cases where even the key objectives are in such conflict with the capabilities of the proposed measurement instrument that the prospects for success are nil. When an investigator receives comments to that effect, one of two kinds of reaction is likely to be elicited. Some investigators are anxious to correct the situation and to reach common understanding about congruity. This type of investigator will usually have resolved matters of congruity between objectives and survey plans before arriving at a final proposal.

The other kind of reaction tends to be defensive. The investigator will explain the great importance of the survey, and the extent of the backing, as though the urgency were so overwhelming as to justify any survey tool. In this case, the investigator's own statements are sometimes incongruous. If the problem is of the great importance portrayed, then planning a survey that will do a satisfactory measurement job is also important. The time-worn argument that "a little information is better than none" is often used. It appears that some investigators haven't recognized that an estimate with low accuracy might have a negative value, or add little or nothing to what is already known. Please ponder the question: "What constitutes new or additional information?" As a minimum, one should try to plan a survey so there is a reasonable chance that it will make a positive contribution.

5. Sometimes the objectives of a particular project are limited more than necessary. To illustrate, let's assume a national survey of 20,000 households. A typical survey plan might call for allocating the sample

about equally to four regions to accommodate regional as well as national tabulations. Is this the best use of available resources? For example, it might be feasible to divide the total sample of 20,000 in the time dimension to provide temporal as well as spatial comparisons. Much depends upon the kind of items involved and the nature of the problem. For some purposes, it is possible that regional comparisons involving an average of four points in time might be more useful than comparisons for one point. Secondly, a sample of 5,000 for each time period could be designed so that regional comparisons for all periods combined would have essentially the same sampling errors as a single sample of 20,000 at one time.

Thus, changing plans to provide temporal comparisons doesn't necessarily mean sacrificing accuracy of regional comparisons. Even though each time period is represented by an independent sample of 5,000, the sampling errors for time comparisons (depending on the kind of items measured) could be about the same as, or substantially less than, the sampling errors for regional comparisons. The cost of a 20,000-household sample survey conducted at one time would be less than the cost of surveying 5,000 at four points in time. But the example illustrates that, in planning, interplay between objectives and survey plans is needed. The importance of insight regarding possible alternatives in relation to objectives and cost is obvious.

Numerous additional illustrations could be cited to emphasize the value of knowing as much as possible about components of error, the capability of a survey, and the best way to design a survey for a complex set of objectives. Knowing how to approach a problem is of major importance. It helps to have a good perspective of various aspects of planning and to resolve a problem into a logical framework. For example, data specification problems are often construed as sampling problems. In that case, it might help to think in terms of the data and analytical specifications that would be appropriate if all units in the population were to be included in the survey. This can help resolve definitions of population parameters to be estimated, as well as data specifications pertaining to individual units of observation. When the planning has reached this point, one is in a good position to resolve matters of the sample design and the questionnaire, but throughout the planning process the idea of achieving congruity should stand out. After the objectives and the survey plans are in congruity, the survey technicians are free to follow through with details of sample design and selection as well as questionnaire structure and content.

Assurance that congruity is good requires a clear common understanding of analytical plans which, under the time pressures involved, are too o[illegible] overlooked. Incidentally, in the writer's view, objectives have not been fully specified until tabulation plans, at least in broad outline, have been formulated. There have been many cases where important omissions in a questionnaire have occurred because analysis plans had not received adequate attention before the questionnaire was put in final form.

Usually, out of a broad or general set of objectives a few key objectives can be identified which *must* be achieved with some recognizable degree of accuracy. These key objectives receive most of the weight in setting sample design and size specifications. Lesser objectives are then accommodated to the extent feasible. In other words, usually one does not anticipate complete fulfillment of objectives, so the main strategy is to seek assurance that key objectives are satisfied.

Unfortunately, the amount of lead time for planning surveys usually leaves much to be desired. Be aware that the pressures of time often call for very good planning rather than going forward with something that is "half-baked." It is possible that a little extra time spent on careful planning could reduce the lapse of time between inception and completion. In other words, planning that clarifies procedures, c[illegible] processes to essentials, and foresees and eliminates potential delaying snags might speed up the whole job as well as provide a higher quality product. Although the pressures of time might reduce the opportunity for thorough planning, time pressure should not become an excuse for poor planning.

Because the approach to applying probability sampling stresses definitions of parameters and level of accuracy needed, much progress has been made toward better congruity since probability sampling became generally accepted. We are beginning to acquire a broad basis for evaluating and improving the returns from investments in statistical programs. We should give more attention to the whole problem of inference, using a cooperative approach among the disciplines involved.

Inference—A Bridge With Two Spans

The writer has a general concept of the problem of making inferences which he has found helpful in approaching many problems and in responding to questions about inference. The inference bridge may [illegible] thought of as consisting of a statistical span and a nonstatistical span. Statistical inference refers to an inference that is founded in probability theory; for example, an inference made from a probability sample about the population from which it was drawn. No attempt is being made to be rigorous. For present purposes, the key point of distinction between the two inference spans is that (1) the nonstatistical span refers to inference that extends beyond the specific population parameters that estimates pertain to, and (2) the statistical span refers to inference from sample data to the population which the sample represents.

Survey objectives and designs involve both spans to some degree. Sometimes there is a tradeoff between (1) making the statistical span short and strong and the nonstatistical span long and weak, or (2) making the statistical span longer (and perhaps weaker) so the nonstatistical span will be shorter. Which inference bridge is best? The whole inference bridge should be kept in mind when planning surveys and stating conclusions from survey results.

Incidentally, from several points of view including inference, censuses and samples are alike. A census may be viewed simply as a large sample. In either, the problems of what to measure, how to measure, and what to infer are essentially the same. Statisticians measure the accuracy of an estimate with reference to a sampling distribution. In a very real sense, results from a census are also estimates and conceptually have an error distribution equivalent to a sampling distribution. Hence, "survey" may include censuses as well as samples.

A nonstatistical inference span always exists. In many instances, there is no statistical span, depending upon how the statistical span is defined. Although one might be able to limit interpretations of data strictly to the population involved, the nonstatistical span exists because action inference decisions, beyond the population to which data relate, are inescapable. There are numerous reasons for this, but one ever-present reasons is the simple fact that the passing of time never ceases; so, to some degree, actions always relate to something that differs from what the data represent. Also, owing to the incomplete nature of information, action decisions generally involve considerations other than information contained in estimates. Hence, there is uncertainty associated with a decision even if all estimates bearing on the decision are without error.

Users of data are generally concerned about the accuracy of data with which they are working. One should be equally concerned, perhaps even more concerned, about the relevance of the data to decision problems. In the writer's view, relevance of an estimate refers to its potential contribution to a decision when there is no error in the estimate. An investment that reduces the sampling error of an estimate provides no return unless the estimate is relevant. We need to know more about how the level of accuracy that is worth purchasing is related to degree of relevance and to what is at stake in the decisions or actions.

Users of data want to be able to make good decisions. They want "reliable" estimates. Assuming the word reliable refers to something that can be counted upon to do whatever is required or expected, a reliable estimate must be accurate and relevant. Hence, degree of reliability might be considered a function of accuracy and relevance. Reliability is sometimes unwittingly equated only with accuracy. We need to give much more attention to evaluating the tradeoff between accuracy of estimates and filling data gaps or doing other things that will strengthen the inference bridge, such as developing improved analytical or decision models.

As the accuracy of a relevant estimate (or set of estimates) is improved, the statistical span of the inference bridge is improved. But it seems intuitively clear that a point is reached where the nonstatistical span becomes comparatively weak and where additional accuracy of estimates will contribute practically nothing to improving the inference bridge. This point is especially important in light of the fact that the marginal cost of additional increments of accuracy increases at a rapid rate. Also, it seems clear, intuitively, that the marginal value of an estimate diminishes as the error becomes very small. Hence, as investments for higher and higher levels of accuracy are considered, a question that becomes increasingly critical is, "How much is any given level of accuracy in an estimate worth?"

It is important to recognize accuracy as a function of two major components—random error (standard error) and bias. The relationship of bias and of standard error to the value of an estimate might be very different; in addition, survey methods and costs differ with regard to reducing standard error and to reducing bias. (Incidentally, the reader should have in mind the quite general situation that as sample size is increased, the standard error of an estimate might become small relative to bias in the estimate.) To illustrate the point about bias versus standard error, consider a time series. The value of the estimates in the series might be increased by a large amount if the standard error were reduced from, for example, 3 percent to 1 percent. But suppose there is an unknown bias of 2 percent in all estimates in the series. If this bias were discovered and reduced, how much, if any, would the value of the series be improved? That type of question has been, and will probably continue to be, debated for decades. However, the writer is simply trying to point out that in his view the value of an estimate, as a minimum and to the extent possible, should be treated as a function of relevance, standard error, and bias. This suggests that some econometric model builders should attempt to separate the contribution of an estimate (a variable in the model) with regard to relevance, standard error, and bias. This type of information could provide very valuable guidance in building and directing statistical programs.

Figure 1 depicts relationships of value and cost of an estimate to standard error of the estimate. Bias is assumed to be negligible at least in terms of effect on the value of the estimate. Much is known about the relationship between cost and sampling error, whereas the relationship between value and sampling error is a matter of conjecture. In figure 1, the maximum point, A, of the benefit curve corresponds to "relevance" of the estimate as discussed earlier. It is conceivable that the maximum for some variables could be negative, because a completely irrelevant variable in a model or decision situation could make a negative contribution even though it contained no error whatever.

With reference to figure 1, proposed surveys need careful examination for possible improvement of the benefit-cost ratio when the sampling standard errors for key items are less than B or greater than C, where B and C are arbitrary amounts. If the sampling errors are greater than C, usually additional resources are called for to increase sample size although adjustments in objectives or survey plans can sometimes be made to improve the cost-benefit ratio without additional cost. If the proposed plans indicate sampling errors that are lower than B, some appropriate actions that might be taken are:

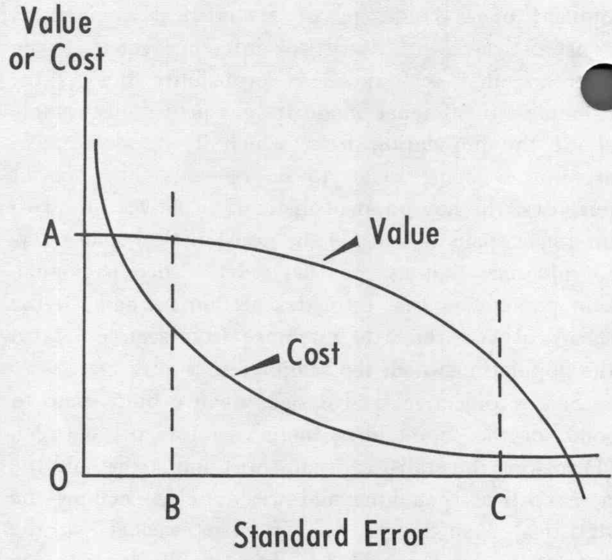

Figure 1.—Value and cost of estimate as functions of standard error.

(1) Broaden the objectives;

(2) Reduce sample size and invest the savings in another project; or

(3) Reduce sample size and invest savings in efforts to reduce bias if appreciable biases are likely to exist.

Perhaps, as knowledge about the value curve is acquired, a criterion for determining the optimal level of standard error will emerge—namely, finding the point where the vertical distance between the value curve and the cost curve is maximized.