ON THE CONTRIBUTIONS OF STANDARDS OF SAMPLING TO LEGAL EVIDENCE AND ACCOUNTING

By

W. Edwards Deming

Graduate School of Business Administration

New York University

Reprinted from CURRENT BUSINESS STUDIES

No. 19, Oct. 1954, pp. 14-32

Published by the

Society of Business Advisory Professions, Inc.

6 Washington Square North

New York 3

ON THE CONTRIBUTIONS OF STANDARDS OF SAMPLING TO LEGAL EVIDENCE AND ACCOUNTING

By W. Edwards Deming

Some advantages to accrue from standards of sampling. Standards are a means of communication and of expedition. Words are standards. The day of the month is a standard. So is the time of day. When two of us agree to meet on Wednesday the 7th at 1 o'clock, we understand each other. The opening of the diaphram on a camera is a standard. I may use a meter constructed in Germany to set the diaphram of my Japanese camera at f:3, and the timing at 1/100th second, for the correct exposure of a film which was made in Rochester. A freight train rolls by and in it are cars from 20 railways in this country and in Canada. Any of them may roll from Halifax to Vancouver via Chicago, San Antonio, Mexico City, El Paso, San Francisco, because they all have the same gauge and require about the same air-pressure to release their brakes, Pullman cars will carry you coast to coast, over several railways. What if the railways chose their own gauges independently? Pilots land aeroplanes everywhere by standard signals and procedures. What if every airport had its own signals? You buy a lamp marked 110 volts, 40 watts, and you may screw it into the socket in practically any city in this country, or in Canada or in Mexico: certainly you may use it wherever the voltage is 110. What if the volt had various meanings in various cities? or what if the voltages were 110 in some cities, 60 in others, 220 in others, 150 in still others? The cost of a lamp would be doubled, because the manufacturer could not make large quantities in any one voltage, and his costs of distribution would be much heavier than they are, as each lot of his product would not only have to be made separately but would have to be beamed toward a particular city. Differences—even small ones—in codes and specifications in various small countries of Europe are more effective than tariffs at raising costs of production and at stifling international commerce.

The existence of commercial standards for materials and for testing them are a very important factor in our high production per man-hour in this country, as these standards cover markets that serve 160 million people, plus our neighbors that use the same standards. A second contributing factor is the use of the statistical method to increase production, and to put scientific meaning into the testing of materials to determine whether they conform to the standard.

Why do I mention some of these obvious economic benefits of standards and of statistical methods? Only to place before you in contrast the ridiculous spectacle of waste and frustration that takes place every day in our own commissions and courts through failure of lawyers and accountants to recognize existing standards of statistical practice. Instead, they argue interminably about exhibits of data that come from surveys or tests scrupulously conducted, while they admit, agree, and pass judgment on statistical evidence that any statistician would disown and repudiate, apparently unaware that standards of statistical performance do exist.

I submit that one should solve statistical problems by statistical methods. When a lawyer acts as his own statistician, the performance is about what you would expect from a statistician who acts as his own lawyer. Standards of statistical performance will help to establish the fact that the precision of a statistical survey need not be in doubt.

I may cite a case of delay and frustration in my own experience. The Illinois Bell Telephone Company used in 1949 modern sampling procedures to select items of property for physical inspection, in order to arrive at a rate-base in terms of reproduction cost. The Illinois Commerce Commission's order ran in part as follows (not verbatim. I have not the order before me):[1]

> The reproduction cost was determined by a tiny sample and is entitled to no consideration.

In 1951, sampling met a slightly kinder fate before the same Commission, but was still discredited with words that connoted complete failure to understand the objective nature of statistical calculations, and failure to appreciate the fact that a statistician, not a lawyer, should be the one to give an opinion on the sampling aspects of a study. The quotation is this:[2]

> Deming's interpretation of his standard error must be viewed with some doubt. His standard error was computed from deviations of averages. These averages themselves, however, were in turn averages of still other averages. . . . Obviously, any such pyramiding of averages would serve to minimize the variations in the raw data.

---

[1] Illinois Commerce Commission: Illinois Bell Telephone Company, in the matter of the proposed advance in rates . . ., Docket No. 36883, 1949.

[2] Illinois Commerce Commission, Docket 39126, 10 Dec. 1951.

I have heard of pyramiding holdings of stock, but not of pyramiding averages. What the writer had in mind is anybody's guess. Anyhow, in 1954, in another case exactly similar, the same Commission hired their own sampling expert, Dr. Leslie Kish of the University of Michigan, who testified in part[3] as follows concerning the sampling procedures that the Company used in 1954, and which were in fact identical with those that the Company had used in 1949:

> The soundness of the statistical theory which underlies the computations of this average, and of its standard error, as well as the meaning of the confidence intervals formed from them, are universally accepted and incontrovertible. By this I mean that there is no controversy; no personal judgment enters into their meaning; they are well-established mathematical consequences. Every trained statistician will make the same interpretation of the results of those computations. . . .
>
> I have made a detailed, independent, and critical examination of the sampling plan, both in its theoretical and practical aspects. I took it to be my duty to find any fault that might make the sampling plan invalid. It is my professional opinion that the sampling plan represents an excellent example of combining theoretical knowledge with sound understanding of the practical requirements to produce a valid, and reliable statistical sampling plan. . . .
>
> In my opinion, the very small standard error of 0.15%, amounting to one part in 500, appears to give sufficient precision.
>
> The examination of the theoretical aspects of a sampling plan and of the formulas connected with it, is a relatively brief assignment for a trained sampler. However, the practical procedures necessary to fit the plan to the intricacies of an actual situation require more detailed, lengthy, and painstaking examination. I have examined the procedures as outlined in company testimony. It is my firm judgment that the sampling procedures, as given, constitute a sound sampling plan which, if followed through, will yield the statistic mentioned above. Secondly, it is my firm opinion that the procedures outlined constitute an eminently practical vehicle for putting the plan into practice. They are clear; they are very complete and comprehensive; and they are tailored to the situation. The procedures take good advantage of the system of numbering items which the company maintains.
>
> Furthermore, the clarity and completeness of the instructions make it a relatively easy and safe task to check the completeness with which the instructions were actually followed through. It would be relatively easy to retrace the steps to most of the units actually selected into the sample. I understand that Dr. Deming actually made some of these checks.

---

[3] Illinois Commerce Commission, Docket 41606, 17 August 1954. The lines that appear here may differ by a word or two from the reporter's record: I have before me only my own notes.

These masterful words of Dr. Kish will, I hope, help to set matters straight in courts and commissions throughout the country. The Commission could have had this opinion from Dr. Kish or from other statisticians 5 years ago, had they taken the trouble to ask for it. In the meanwhile, some of the finest legal and engineering talent on both sides was dissipated in needless argument over methods that are accepted and have been accepted, and used all over the world wherever precision is desired, but apparently not known to the people who carried on the argument.

Existing standards of statistical performance. Scholars of law, accounting, engineering, and economics have frequent need of statistical studies and of sound interpretations thereof. Statistical studies, properly conducted, will become increasingly important to such people, in view of the fact that:

(a) There are standards of performance in statistical surveys of people, opinions, accounts, inventories or of anything that man measures; and also in the methods of statistical inference by which one draws conclusions from such surveys;

1. Recommendations Concerning the Preparation of Reports on Sampling Surveys, produced by the United Nations' Sub-Commission on Statistical Sampling, 1948.

2. Military Standard Sampling Procedures and Tables—Mil-Std 105A, 11 Sept. 1950 (Munitions Board, Washington). Incidentally, these tables are used all over the world.

3. Circular A-46, Standards for Statistical Surveys Carried out by or for Government Agencies. (Bureau of the Budget, 1951. Largely the work of the present author.)

4. Recommended practice in the sampling of physical materials, issued by the American Society for Testing Materials, 1954.

(b) Any two competent statisticians will agree that certain statistical practices will give valid results, and that these results possess valid standard errors. They will agree that other procedures will not do so. Moreover, they will agree on the interpretation of the standard errors, and on the amount of knowledge conveyed by these standard errors.

(c) Inferences concerning causes (as of disease, treatment, defects arising from a system of production, or from a system of accounting, methods of correction, methods in agricultural or in industrial science) MUST be interpreted as samples, whether the statistical study which forms the basis for the inference came from a sample or from a complete investigation of every item. There are standard methods for such inferences.[4]

---

[4] W. Edwards Deming, *Some Theory of Sampling* (John Wiley & Son, 1950), Ch. 7. Also, *Journal of the American Statistical Association,* vol. 48, 1953: pp. 244-255.

(d) Standard statistical plans, used world-wide, offer the only possible way to prescribe meaningful statistical tests in contracts for the delivery of industrial product. These methods are in wide use today. Thousands of dollars worth of product is tested and accepted every hour on the basis of standard statistical plans (such as Mil-Std 105A, mentioned above).

It should be of interest to lawmakers that many laws and regulations, particularly those that deal with commerce, have, as written, no operationally definable meaning. Standard statistical methods offer the means of clarification. For example, there is no way to specify the quality or dimensions desired in 1000 or a million accounting transactions or in 1000 or a million pieces of industrial product except in terms of the sampling plan by which the transactions or the pieces will be accepted or rejected.

What is sampling? Sampling, in the modern sense, is a procedure that conforms to the theory of probability for the selection, examination, and interpretation of the results obtained from a portion of a lot of items (accounts, cases, records, people, families, farms) in a study that is carried on in order to measure some characteristic of the entire lot, or to be able to regulate or to take some action rationally thereon, or on the cause-system that will produce future items.

The theory of probability, and modern statistical methods based on the theory of probability, are man's most dependable tools for making quantitative studies, and will become one of the chief tools of the lawyer and of the accountant, as they have already become in scientific investigations.

The design of a statistical survey, in the sense of today's statistical practice, means THE MOST EXPEDITIOUS WAY TO COLLECT THE DESIRED INFORMATION AND TO INTERPRET IT, and to know the quality of the data and the limitations to the conclusions.

For any given problems, there are usually several possible sampling plans, all valid, of varying degrees of difficulty, some more suitable and some speedier than others; of differing costs. It is universally understood that the statistician will choose, from the numerous plans that are possible (one of which is a 100% investigation) the one that he believes will deliver the required information and the prescribed precision at the lowest cost and with the best despatch compatible with the various administrative and physical limitations that are always present.

Samples less than 100 per cent will often produce results speedier and cheaper than a complete coverage; and in most cases, the sampling-results are more reliable.

A sample carried out by modern procedures is planned in advance, in detail. It would be possible for another statistician to retrace every step of a survey, even to the reading of the random numbers, and to identify each unit selected (person, family, farm, case, card), except when the units are in a continuous flux, as in an active card file. Moreover, he could retrace the tabulations and the formation of the estimates and of their standard errors (cf. the quotations above from Dr. Kish's testimony; also the two sections at the end).

Whatever be the procedure that he prescribes, the precision thereof will never be in doubt, because for any estimate made from the results of a statistical sample, one may calculate limits for the uncertainty that could have arisen from the use of a sample instead of a complete investigation of the same material.

It is also the business of the statistician to interpret the results of the survey with the aid of the theory of probability, particularly in respect to the limitations of the conclusions that arise from the sampling, or (in the case of the analysis of causes) in respect to the limitations that arise from the fact that even a complete coverage of every item must also be interpreted as a sample (cf. the last footnote).

In any case, the chief characteristic of modern sampling is that, after the job is completed, there is never any doubt about the margin of error that the sampling has introduced. This margin of error is as important as any of the results of the sample-survey. It is calculated by the theory of probability, and is above any man's judgment or opinion.

Two statisticians, working independently, may and usually will choose slightly different procedures for any particular job. However, both of them will have valid standard errors, and the difference between the results obtained by the two procedures is predictable in the probability sense.

What sampling is not. Sampling is NOT the substantive expert's selection of "representative" or of "typical" cases, areas, or farms, or of weeks or months from the year. (Do not ask me the meaning of these adjectives: I know not.) It is strange that judgment and experience in the subject-matter, so indispensable in the definition of the problem and in framing the questionnaire, the method of test, or the method of conducting the interview, and in assisting the statistician to choose economical and workable strata and sampling units *(vide infra)* is so useless and hazardous for the final act in which one selects the items for investigation. The reasons are: (1) that there is no way to calculate objectively the margin of error of a sample selected by judgment; (2) such selections have too often in the past led to wrong answers and to wrong decisions.

Expert knowledge, judgment, sincerity, and honesty, are all necessary ingredients of any science, but they are not sufficient to make a sample. There is no substitute for knowledge of statistical theory.

The statistician makes his final selections by means of a standard tool known as a table of random numbers. The statistician's random methods assign a known probability of selection to every item, and they remove the biases that arise from judgment-selections and inferences.

When the selections are made by random numbers, and when additionally the estimates are formed according to rules prescribed by theory, the results will permit inferences to be drawn objectively by the theory of probability. When the selections are made by judgment, inferences may be made only by judgment, but NOT by the theory of probability. True, one sometimes sees unjustifiable calculations of standard errors and unjustifiable inferences in papers on medicine, agricultural science, in accounting, and in marketing and opinion research, when the samples are not random. Such calculations have no meaning; and if they lead to a correct answer, it is so only by luck.

Contrast a statistical plan with a judgment-selection. The statistical plan is fully accountable, and is subject to verification. The judgment-selection is not accountable, and the reasoning that leads to the selections and to the conclusions could never be retraced.

A committee of top executives from a number of companies in a certain industry, well educated and able in their fields, chose recently to study the records in their respective companies for the month of July 1953 to provide legal evidence in a case that threatened their industry. Was this a sample? Of what? Of 1953? Not in my definition. Why? Because I know of no way to place limits on the degree of uncertainty of any useful inferences that might be made from the study of July.

Even four months of the year, January, April, July, and October (a 33% sample) would not constitute a sample that is defensible by probability methods. It is still a sample of size 4, which is a pretty small sample, and it could have a lot of errors and biases.

A much better plan, and only a little more costly than a study of a single month, was easy to design in this particular case. It consisted of a stratified random sample of the records for the full year 1953. Each sampling unit was 10 successive transactions, and the job was laid out in such manner that the uncertainty attributable to sampling can be calculated in a few minutes.

First of all, on a job of this kind, one sets aside for separate treatment all the important and rare transactions, so as to have full information on them (in this case for the year 1953). This is an obligation that the expert in the subject-matter cannot delegate. The statistician will assist in this responsibility by enquiring whether there are rare and important transactions, and how much accuracy the accountant requires in them, and why; and then to set aside for separate (possibly complete) study all of them that can not be left to the main sample. In this case, it was the extra large consignments of ore ($25,000 and over) that had to be separated out. It was a simple matter, as they were already isolated in the records.

In another instance an accountant wished to examine 100,000 transactions. He followed a text-book and drew a sample, and examined the transactions therein. He was disappointed because the sample contained no carloads of alumina, yet there were a few such transactions amongst the 100,000. Was this omission an error of sampling? Yes, it was; but the real fault lay in the planning of the job. No competent statistician would engage himself in sampling the 100,000 transactions without first enquiring from the accountant what kinds and sizes of transactions are important, and what are their relative frequencies of occurrence? He would set transactions of huge money-value, and anything else that is important and rare, into a category for separate treatment, probably 100%. If the cost of this separation is great, the cost is chargeable to the system of filing the records, and not to the sampling thereof.

As a further illustration I may relate here an incident that happens frequently in the life of every statistician. A well-known psychologist came to me in Germany. He was making studies in sociology and psychology. There could be no doubt of the importance of his studies, nor of his competence—as a psychologist. Would I assist his project by drawing a sample of 4 "representative" cities in West Germany in which to conduct these studies? Within these cities, their own expert investigators would select families for the study. They would later, on the basis of these figures, generalize for all of Germany.

The answer was no, with an entreaty to improve their sampling procedures. The right way, as he was interested in accuracy and economy, was to select the families by statistical methods, to assure by stratification and by random numbers not only each family, but also each group, type, and area in Western Germany, a known chance of selection, in a definite manner.

It was fairly easy to show that a proper statistical method would conserve the funds allocated for the study, and that the professor could have much more information for the same money.

I listened to a learned lawyer and economist as he explained how difficult it is to formulate some of the problems of the Federal Trade Commission in a form so that they become statistical—i.e., so that a statistical study (of people, records, sales, revenue) will provide information that will be helpful as evidence. I can well appreciate the difficulties that he described, but I wish that he would not call them problems of sampling. They are difficulties with the subject-matter. The problems of sampling, once the expert formulates the problem, are in most cases relatively straightforward for the statistician.

Newspapers and magazines were full in early 1954 with controversy on the results obtained by the sample used by the Bureau of the Census for the monthly figures on employment and unemployment (incidentally, one of the finest examples of modern sampling procedures and one of the most important monthly figures). Was unemployment going up or down? Most of the writers spoke of errors and possible errors of sampling, when in reality they described the difficulties of defining unemployment. Some of these writers were able economists and should have known better. A man thinks that his company is going to let a few men off, so his wife suddenly looks for work and thinks that she is unemployed. She has not worked for some years, but now, in the face of possible loss of her husband's job, she starts off one fine morning to look for work; comes home without anything definitely lined up. Random numbers had selected this family for the Census sample: the interviewer comes that evening and the woman reports herself to the interviewer as unemployed. Several other women did the same, and the figures for the month showed an alarming increase in females unemployed.

Was this an error of sampling? No, it was simply one of the difficulties of trying to formulate and to understand unemployment in quantitative terms. This is a problem for the experts in the subject-matter. The same difficulties would have existed whether the figures were collected by a complete census (at a cost of $50,000,000) or by the regular monthly sample (at a cost of $75,000—my own figure, not the Census's).

Is it important to make this distinction? It certainly is, if we hope to understand and to use the results of surveys, and to make progress in getting better figures for smoother administration in government and in business.

A sample is NOT a last resort, to be used when a complete investigation is impossible. Rather, it is the first resort: it is the answer to the question, "What is the best way to do the job? What kind of a sample will do it, and how big must it be to deliver the information that is required?" Viewed in this way, sampling can be of inestimable help to the lawyer and to the accountant.

Some common misconceptions about sampling. I collect here some points of importance, commonly misunderstood. Some of these I have or shall elaborate elsewhere in this paper.

1. A sample is NOT a selection by an expert in the subject-matter. A sampling expert is not a man who has enjoyed a run of successful selections by judgment.

2. Judgment and knowledge of the subject-matter are almost as hazardous for judging the margin of error of a sample as they are for making the selections.

3. Judgment and knowledge of the subject-matter are essential to the design of an economical and useful sample. The sampling expert makes use of judgment and knowledge in those stages of the sample-design where they can really be effective. He does not use them for making the selections, nor in making statistical inferences, where judgment and experience have too often led to biased results.

4. The size of sample is no criterion of its precision. The procedure of stratification, the choice of sampling unit, the formulas prescribed for the estimations, are more important. Once these features are fixed, then as we increase the size we gain precision (though the point of diminishing returns comes rapidly).

5. One can not offset bad practice by increasing the size of the sample.

6. The size of the universe is of no practical importance in the size of sample for a particular job. The size of sample prescribed for a universe of 100,000 accounts might be identical with the size for a universe of 1,000,000 similar accounts. I often prescribe the size of sample and its procedure before the company has furnished me with the figures for the universe whence the sample will come.

7. The quality of a statistical sample is built in. It is not a matter of luck. Its margin of error is predictable pretty accurately in advance, and is calculable definitely and objectively afterward from the results of the sample.

8. The statistical part of a job is only the expedition of the job, including the inferences to be drawn therefrom, so far as they are purely statistical. The substantive-expert can not delegate to the statistician his responsibility for specifying the kind of information that he needs, the questionnaire, the method of eliciting the information, the segregation of rare and important items, nor the forms for the tables. It is the substantive-expert, not the statistician, who must testify to the usefulness and meaning of the results. The statistician only testifies to their range of validity, and to the proper method of inference.

9. Many sampling practices supposed by professional workers in substantive fields to be correct statistical practices are indefensible because they are hazardous or too costly, or both.

10. Statistical practice requires scholarly training in theory, experience under competent leadership, and mature judgment in practice.

11. The uncertainty that arises from the use of modern sampling procedures in place of a complete examination of every item is an objective measure, and is not a matter of opinion or of expert judgment.

12. Most of the errors in statistical studies are NOT errors of sampling, but arise from failure of the expert in the subject-matter (accountant, labor expert, engineer, economist) to formulate questions suitable to the solution of the problem, or to specify ways in which to procure the information, or to exhibit it, or to provide a suitable frame (list); and finally, from his failure to interpret the results in accordance with the best knowledge of the subject. Elaboration and examples occur in Chapter 2 in my book.

13. The theory of sampling and the standard error of a result apply as well to studies that involve subjective measurements and opinions, as they do to studies where the measurements are quantitative. (cf. the two paragraphs at the end of the section, "How does one judge a sample?")

Misplaced faith in a complete investigation and in a 10% sample. Another common misconception is so prevalent, so wrong, and so costly, that I give it a separate section.

A complete coverage, no matter how badly it is carried out, and even though it is incomplete (as complete counts too often are), has achieved, through time, unjustifiable weight in evidence. People who are not statisticians assume too often that the sheer size of a complete coverage will somehow cover up its incompleteness and the flaws in the method of eliciting the information.

A 10% sample has almost equal standing with a complete count. Why, or what 10%, is hardly ever questioned, even by the experts in quantitative subject-matter.

The statistician, in the explanation of his sampling procedure in legal evidence, faces such preconceived ideas. It is a fact that the aerial plant in a sample of 1000 to 1500 telephone poles will provide all the precision that one can use for the estimation of the average overall physical condition of the entire aerial plant which might be worth $200,000,000. But without careful preparation to dispel preconceived ideas about complete counts and 10% samples, the statistician must be prepared to face an objection on the ground that a sample of only 1 part in 1000 is not admissible as evidence. The man who objects may, without knowing it, own stock in a woolen mill that purchases a million pounds of wool and pays duty on it on the basis of a sample that weighs from 60 to 100 ounces—1 part in 100,000.

It is the absolute size of the sample, and not its proportion to the whole, which determines the precision of the result. The statistician must be prepared to meet the man who thinks that to reach a prescribed precision in an estimated average rent, for example, a sample of dwelling units from a big city must be bigger than the sample from a small city, simply because the big city is bigger.

Wider use of standards of statistical performances and standards designed especially for surveys in accounting and in legal evidence, would eliminate some of this needless waste and delay, and they would also help to eliminate some of the wrong judicial decisions that are made (a) on the basis of bad statistical evidence, and (b) on failure to recognize good statistical evidence.

How does one judge a sample? One must bear in mind that the judgment of the usefulness of a sample must concern itself with two questions:

For the expert in the subject-matter: Would the results be useful if the sample had been 100% (a complete coverage of all the universe—people, families, farms, loans, records)? On this question one concerns himself with:

a. The information to be obtained: The questionnaire.

b. The method of eliciting the information, such as the examination of the records, or the field-work (selection of workers, their methods, training, supervision).

c. The coding.

d. The forms for the tables.

The above headings are for the specialist in the subject-matter: They would be the same whether the survey was a small sample, or a big one, a good one or a bad one, or a complete coverage.

For the statistician: What is the uncertainty introduced by sampling? (What is the precision, in other words?) This question is purely statistical, and can be answered only by statistical techniques, and only for a sample that has been carried out and processed according to theory. The answer lies in the so-called standard error of an estimate, which is calculated from the returns of the sample itself. It is not a matter of opinion, nor of the judgment of experts in the subject-matter.

I may cite an instance where a regulatory commission hired a well-known engineer to give his opinion concerning a sample that I had drawn from the property of a telephone company, to be used for the determination of the reproduction-cost. The man gave his opinion, after some consideration, based on his long experience as an engineer. The opinion was favorable, but by accident, I believe, because his engineering judgment might have led him to think otherwise: a sample can be judged only by knowledge of sampling, and any two men who possess knowledge requisite for sampling will give the same opinion.

Where his engineering knowledge should have been put to use was in the methods of inspection and test, and in the determination of the weights of the various classes of the property, and in those other aspects of the problem that are the same whether the study is to be carried out by a complete investigation, or by a sample, big, little, good, or bad. On such questions his opinion could have been respected.

On another occasion, an engineer gave it as his considered opinion that a certain statistical sample would be entirely valid if the measurements on the items had been made by calipers and expressed in inches or millimeters, but that as the test of each item was subjective, visual, based on appearance, the sampling plan was not a suitable one for the purpose.

Suppose that the sample had been 100 per cent. Would his argument apply? No. Neither did it apply to a sample less than 100 per cent.

He was confusing sampling error with errors of test or of acquiring the desired information concerning each item. He may have been entirely justified in his doubts about the tests: as a statistician I have no opinion on that matter.

One should regard a sample as a *sample of the labels* that might be tacked on to the item in a complete 100 per cent examination. If the method of test or of acquiring the information or of passing judgment on a single item is not satisfactory, NO SAMPLE, NOT EVEN A COMPLETE COUNT, will give useful results.

Division of responsibility between the statistician and the expert in the subject-matter: An example: The statistician must satisfy himself that the job that he is called upon to do is really a statistical assignment, and not one in the subject-matter. The substantive expert must decide whether the study is worth doing, and must testify to the usefulness of the results.

The following excerpt represents a statistician's attempt in direct testimony to state what his job was; to put a limitation on his assignment; to distinguish between his job and that of the expert in the subject-matter. The case involved the use of samples of items of telephone plant, the aim being to obtain a figure for the average over-all per cent physical condition of the entire property.[5]

> Q. Doctor, for what purpose were you engaged by the Illinois Bell Telephone Company?
>
> A. I was told that this company proposed to make a survey to determine the physical condition of its plant. They asked me

[5] The Illinois Commerce Commission, Docket No. 39126, 1951, and Docket No. 41606, 1954: The Illinois Bell Telephone Company in the matter of the proposed advance in rates. The passage printed here is testimony prepared in advance, and is not necessarily the same word for word in the record.

to prescribe statistical procedures by which to select samples of items of plant for inspection, such as poles, wire, cable, telephones, relays, central office equipment. The samples must determine within narrow limits of precision what result would be obtained for the average over-all per cent condition by a complete 100% inspection of all the items in all the classes of plant that were to be inspected, with the same inspectors, and with the same care as was exercised on the samples, were such a thing possible.

This assignment carried with it the responsibility for prescribing the procedures for summarizing the results for each class of plant, once the code-values assigned by the inspectors were translated into percentages, and for combining the per cent conditions of the several classes of plant into the over-all average per cent conditions of all the classes that were to be inspected. A necessary part of the assignment was to provide procedures by which to calculate the standard error of the precision of the result obtained for the over-all average per cent condition.

My assignment did not include the responsibility for the procedures for inspecting any item, nor for the numerical values that translated the inspectors' codes into percentages. Neither had I any responsibility for determining the weights of the various classes of property. These problems are the same whether one uses sampling or not. These phases of the work have been described by Mr. Coxe (General Staff Engineer).

Some proposed ingredients of standards of sampling procedures for legal evidence. A better understanding of statistical standards will enhance greatly the usefulness of statistical studies in legal evidence and in other work of the lawyer and of the accountant.

A report concerning a statistical survey should convey full information on the following technical matters, so that another statistician could render an opinion on the procedure and on the validity of the statistical conclusions.

1. A statement of the aim of the survey, including the precisions desired (standard errors) for some of the important results.

2. A description of a sampling unit, and of the method of covering a sampling unit once it is drawn into the sample.

3. The source of the frame (lists, maps, rules for the formation of sampling units).

4. A description of the deficiencies of the frame: its failure to cover all the universe, particularly certain segments thereof: the pos-

sible effects of these failures of coverage on the conclusions and on the usefulness of the figures to be derived from the survey. These deficiencies must be brought to the attention of the substantive expert in advance, as the matter is of vital importance. Without approval of the substantive expert to proceed, the study should be abandoned for being incapable of delivering results sufficiently useful.

5. A statement from the substantive expert that lists the rare and important items (if any) that will require segregation and special treatment (perhaps 100%).

6. The rules for drawing the sampling units. This step will include a description of the stratification of the frame or of the sample.

7. The estimates desired. The formulas for the formation of the estimates.

8. The formulas for the standard errors of the estimates of chief importance.

9. The results, calculated from the sample. The standard errors of some of the estimates, as actually calculated from the results.

10. The interpretation of these standard errors as margins of uncertainty introduced by sampling, and as margins of uncertainty in the identification of causes in a system.

11. Tests of the sample against known figures. (Such tests do *not* take the place of the standard errors; and they do *not* prove that the precision or the accuracy of any result is high or low. Some of them are in fact made as supervisory probes before the statistician permits the main collection of information to proceed. These supervisory probes are aimed at the discovery of misunderstandings of the sampling procedure, gaps in the frame, and unsuitable definitions.)

12. Statements of: (a) the statistician's verifications of the prescribed procedure, and of the steps that he took to insure compliance; (b) the limitations of the data because of errors of reporting, because of non-response, and from other difficulties encountered.

13. Relationships between the statistician and his client (next section).

14. Definitions of the frame, of precision, accuracy, standard error, and other simple terms used in statistical parlance.

Precision, accuracy, bias, and standard error. Standards of statistical performance will necessarily contain a few technical terms. Some indispensible terms are precision, accuracy, bias, standard error, efficiency, and validity. None of these is difficult, technically.

Precision is expressible by an international standard, viz., the standard error. These standard errors measure the difference between a complete coverage and a long series of estimates formed from samples drawn from this complete coverage by a particular sampling procedure of selection and estimation.

Great precision or a small standard error attached to an estimate made from a sample does not mean that this estimate is necessarily accurate or useful. It does imply that the results of a complete coverage (a 100 per cent sample) would have been the same as the sample within a very narrow margin of difference, had the complete coverage been carried out with the same definitions, with the same investigators, sharing the load proportionately, and with the same care as they expended on the sample.

The definition of any characteristic, whether it be age, employment status, income from interest, change in liquid assets, goods returned, yield per acre, quality of being defective, or anything else, must be given in terms of an operation or procedure for the measurement of this characteristic. For some characteristics it will be agreed by the experts that there is a preferred procedure, even if in practice some preferred procedures are never or hardly ever used. A preferred procedure is distinguished by the fact that it supposedly gives or would give results nearest to what are needed for a particular end; and also by the fact that it is expensive, or time-consuming, or even impossible to carry out full scale.

A preferred definition for a person's age might be specified as the result of subtracting the date of his birth from the present date. In turn, date of birth might be defined as the figure obtained by examining the person's birth certificate; or if there is no birth certificate, his horoscope, family album, or sworn evidence, or by tying his past with some important event, preference being in the order given. This definition for age would of course be unworkable on a full scale survey. The Division of Passports can of course enforce it, but not you nor I.

In practice, to expedite the survey, we substitute an *un*preferred procedure—for example, merely ask the man his age at his last birthday, or his date of birth, and write down the answer. If the man himself is not at home, enquire of any responsible member of the household.

Besides preferred and unpreferred definitions and ways of asking questions, there are preferred and unpreferred ways of hiring, training, and supervising the field- and office-workers. Four recalls are certainly preferable to three, three to two.

The results of two surveys, one using a preferred definition, the other an unpreferred definition, will be different—often greatly different. The result of a preferred procedure is sometimes called a true value.

An unpreferred procedure is sometimes called a biased procedure.

The bias of an unpreferred or biased procedure is the difference between two complete investigations, one with the preferred procedure, the other with the unpreferred procedure.

It is the expert in the subject-matter, not the statistician, who must decide what is the preferred procedure, and whether a particular proposed unpreferred procedure will be acceptable.

A preferred procedure of today is always subject to modification and new and more useful definitions tomorrow. Hence we are forced to conclude that NEITHER THE ACCURACY NOR THE BIAS OF A PROCEDURE CAN BE KNOWN IN A LOGICAL SENSE, because the preferred procedure is subject to change.

In contrast, THE PRECISION OF A RESULT, ONCE CALCULATED, REMAINS UNCHANGED THROUGHOUT TIME.

Efficiency and inefficiency contrasted with validity. A sample may be inefficient, yet valid. A sample design is efficient if it will deliver a prescribed precision at low cost.

Validity, to the statistician, means the rightful application of probability in the sampling procedure, including the inferences to be drawn from the sample. It has no connexion with efficiency. For example, the validity of an inference drawn from a sample taken and processed today is not impaired by tomorrow's discovery of a more efficient method of sampling.

An efficient design is possible where there is statistical information and experience on which to plan the sample, and to use in the formation of the final estimates. For example, to estimate the total inventory of shoes in all the stores in the country that sell shoes, a rough indication of the size of each store on the list would have cut the cost of the study to perhaps half. In the absence of such statistical information, the sample can not be so efficient, and precision must be achieved through size, at increased cost.

In the end, however, efficient or inefficient, the precision achieved is nevertheless always known.

Working relationships between the statistician and his client. Working relationships are basic to statistical studies. For example, the statistician, before he begins to draw up the procedures for a study, must satisfy himself: (a) that the entire procedure, including the drawing of the sample, the field-work or the testing or other processing of items in the sample, the coding and the tabulating, will be satisfactory; (b) that all people involved in the work agree that the statistician has not only the right but the duty to verify and to retrace and to scrutinize any of the work at any time; (c) that no results or interpretations thereof, nor the methods employed, will be reported or published without his approval.

Both client and statistician must understand these relationships if the statistician's work is to be most effective.

CHAIRMAN DEAN: Our next speaker was at one time an instructor and later an assistant professor at Yale University. He has written one of the standard textbooks on public utility regulation and was formerly consulting economist for the Antitrust Division of the Department of Justice. He has been Director of the Economics Bureau of the Civil Aeronautics Board. At present, he is Assistant Chief Economist of the Federal Trade Commission and Acting Chief of the Division of Economic Evidence and Reports.

We are very fortunate to have with us today, Dr. Irston R. Barnes!

93

ON THE CONTRIBUTIONS OF STANDARDS OF SAMPLING TO LEGAL EVIDENCE AND ACCOUNTING

By

W. Edwards Deming

Graduate School of Business Administration

New York University

Reprinted from CURRENT BUSINESS STUDIES

No. 19, Oct. 1954, pp. 14-32

Published by the

Society of Business Advisory Professions, Inc.

6 Washington Square North

New York 3