ON THE PRESENTATION OF THE RESULTS OF SAMPLE SURVEYS AS LEGAL EVIDENCE

By

W. EDWARDS DEMING

Graduate School of Business Administration
New York University

Reprinted from
The Journal of the American Statistical Association,
Vol. 49, December 1954: pp. 814-825

ON THE PRESENTATION OF THE RESULTS OF SAMPLE SURVEYS AS LEGAL EVIDENCE

BY

W. EDWARDS DEMING

GRADUATE SCHOOL OF BUSINESS ADMINISTRATION
NEW YORK UNIVERSITY

Reprinted from

THE JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION,
Vol. 49, December 1954: pp. 814-825

ON THE PRESENTATION OF THE RESULTS OF SAMPLE SURVEYS AS LEGAL EVIDENCE*

W. EDWARDS DEMING
New York University

PURPOSE OF THIS PAPER

THE purpose here is to view some of the problems that confront the statistician when he presents the results of a sample survey as legal evidence. One particular point is that the statistician, if he is to make his work useful, must distinguish between (a) what he as a statistician may say about the *precision* of the results of his survey, and (b) what an expert in the substantive field may conclude about the *usefulness* of the results. The statistician can testify only to the former, and possibly also about the variance between investigators, and between different methods, if he measured these differences. As a secondary purpose, we shall enquire into the meaning of a standard error, and its relation to a complete count and to the usefulness of the results—a point that is often overlooked, not only in testimony but in statistical reports.

I have no magic nor all the answers to all the questions and difficulties that the statistician will encounter when he presents results as evidence. It is possible, however, to share some experiences with colleagues in this increasingly important role of statistical surveys; to acquaint them with some of the kinds of problems that may arise; and to suggest some general principles that will help the statistician to make his work more useful than it would be otherwise.

At the outset I may explain that this paper will deal only with probability samples. The defence of any other kind of sample is hardly a problem for a statistician anyhow, but rather for the substantive expert who may have enough knowledge of the material and of its variability to feel that he can testify one way or another with respect to the interpretation of the results of a judgment sample.

A statistical survey, formulated and carried out by the dictates of the theory of probability, is to the statistician an exciting and remarkable achievement. It produces man's best empirical knowledge, and it provides an objective measure of the amount of knowledge in the survey. The precision desired can be aimed at and hit pretty accurately by planning *in advance* with the aid of the theory of probability and with bits of knowledge with respect to certain proportions, means, correlations, variances, and other statistical measures of the sampling units in the frame. Then,

* Presented at a meeting of the American Statistical Association in Chicago, 29 December 1952.

after the survey is completed, the precision that was actually reached is calculable and expressible in an international standard of measure (the standard error) from the results of the survey itself. This final measure of the precision is objective, and is not a matter of opinion. It is not biased by incorrect assumptions that went into the planning.

The statistician when he presents his results as legal evidence finds himself nevertheless at an uncomfortable disadvantage. He is usually talking to scholars, but not to fellow statisticians, nor indeed to other scientists, nor relating the results of his research to a trusting client or sponsor. He is teaching, but the techniques of the class room will not necessarily be the best ones for the presentation of evidence.

Scholars in other disciplines are not all acquainted with the achievements of probability sampling, yet the statistician must somehow explain his methods to them. Some of the people that he must deal with in legal evidence know sampling only as a failure to predict an election; they know not the distinction between (a) the standard error of sampling, (b) the errors common to complete counts and to samples, and (c) the error of a prediction. Other people think of sampling as a selection by judgment, carried out by someone who has established a reputation by a run of successes in the past. To still others, sampling is a desperate risk, a hazardous aimless random drawing of areas or of other elements to which anything may happen, and concerning which nothing can really ever be known except by comparison with a complete count, for which a sample is only a substitute to save time and money.

In my own experience, a man questioned the existence of the theory for estimating the variance of a mean, originated by Gauss 120 years ago, and now used all over the world. I was once accused of "pyramiding" my results (whatever that is), because I took the average of the averages of my 10 subsamples for an estimate of the whole, wherefore my standard error "must be viewed with some doubt."

DIRECT TESTIMONY: CROSS-EXAMINATION

There is first of all direct testimony, wherein the statistician presents his results, after careful preparation in advance. He will usually simply read his direct testimony into the record from typed copy. Direct testimony may take the form of questions and answers, the questions being read by the lawyer who has engaged the statistician. The questions should be framed so that they display the results of the survey in the form of valid statistical inferences. The questions must not sound as if they were digging for particular answers, even though both sides in the case know full well that the statistician is reading from prepared copy, and that the answers are exactly what the statistician believes to be essential to his methods and to his results, regardless of the questions.

In the preparation of testimony, the lawyer who engaged the statistician will not try to influence the content of the statistician's statements. He will try to help the statistician to state his procedures and his inferences so that they will be clear. The inferences must be only what the statistician can support, as a scientist seeking truth. To bring out the truth in a scientific inference, one must not only state what he believes to be true, but he must say it so that his listeners will understand what he means, and not think that he has said something that he did not mean. A good lawyer can help immeasurably in achieving this aim.

In giving evidence, the statistician is not fighting a case for either side. He is an expert witness, and he should appear as a professional man, with the sole aim of presenting the truth. This means that he must tell to the best of his ability what the figures mean. He should describe in full any difficulties that he encountered, and their possible limitations on the interpretations of his data.

In some courts one may not read prepared testimony, in which case one can only prepare to present his testimony without the aid of his typed copy. He will of course still be able to present tables and charts, called exhibits.

Usually during or immediately following direct testimony the opposing side asks only questions that will clear up simple failure to recognize technical terms, or to clarify some events with respect to their sequence in time. Questions that may bring out flaws in the testimony, they will usually reserve for further study, following which they will call the statistician to the stand for cross-examination. Here the questions are often well-prepared in advance, but the statistician must answer *ex tempore.* Here the statistician may find himself very uncomfortable to find that statements and interpretations that he thought were clear and objective are now misunderstood and misinterpreted, and his statistical principles challenged.

When cross-examination comes, no matter what question comes, relevant or irrelevant, do the best that you can with it. Be cautious to stay within the field of competence that you have testified to in your qualifications (vide infra). Groundwork in your direct testimony, in an attempt to give clear explanations of your procedures, of the statistical interpretation of your results and of their standard errors, will help to keep the cross-examination on the track and to bring out the inherent scientific truth contained in your survey.

To present the results of a survey in a case where millions of dollars are involved, to ears unfamiliar with the power of modern statistical practice, is an experience that purifies the statistician's thinking. Sometimes the listeners are glad to accept the results of a good survey, and to learn something about modern survey-methods. At other times, they will declare that the statistician's methods are new and untried, that his results are therefore not acceptable evidence; that his sample was too small; or finally, foresooth, that he has not explained the entire theory of sampling so that everyone can understand exactly what he did and why, and that there is therefore no basis by which to judge whether his results have any meaning.

ESSENTIAL INGREDIENTS OF THE DIRECT TESTIMONY

The statistician's statement of his qualifications, which usually comes in the first part of his direct testimony, is important. It is evidence by which the examiner or judge may decide, if the question arises, whether the statistician is qualified. It should therefore contain a full account of the statistician's education and relevant experience.

He may then present the purpose of the survey (an example of an assignment will occur later), what he endeavored to do, the methods that he prescribed, the basis for these methods, the system and the observations by which he satisfied himself that the procedures that he prescribed were understood and followed rigidly and faithfully; finally, the results and their standard errors and their interpretation; also the possible effects of any biases inherent in the procedure, and the possible effects of any difficulties encountered. All these points will go into the direct testimony.

He should tell in simple words what the procedures actually were. He should limit theory to a few simple and well-established principles that illustrate the sampling procedures and the interpretation of the results. The truth and the whole truth means clarity, so that anyone may judge whether your results and your interpretation of the standard error are what you say they are. You can not hope to give a whole course in the theory of sampling, but you can make your procedures and their validity clear without doing so. The most convincing argument concerning your procedures and of your interpretations is that they conform to established international standards, and that they are used in a wide variety of experience. In this connection the document written by the United Nations Sub-Commission on Statistical Sampling entitled, "The presentation of sampling survey results" (UN Series C, No. 1, 1950) is of assistance; likewise the "Manual on the Quality Control of Materials" (1951) and other recommended and standard practices of the American Society for Testing Materials, many of which have been adopted as standards in other parts of the world.

A formula can cause trouble unless you explain pretty expertly how you used it. If in direct testimony you say that you used a formula to calculate in advance the size of sample required, when the fact is that you made a rough mental calculation and tempered it with judgment, or that you made the calculation years ago for similar work, and really did not make a fresh detailed calculation for this job, or if you did make one and then modified the answer to allow for some possible additional variance not fully represented in the formula, or to allow for some possible heavy additional cost of inspection or of interviewing because of probable bad weather in February, and if now upon cross-examination when people start asking questions you can not get exactly the same sample-size out of your formula as you actually used, you may find yourself very uncomfortable. The trouble is that people not accustomed to formulas will not understand how one uses theory.

If you say that a certain constant in your formula for the required sample-size represents your advance estimate of the variability of the material that you sampled, someone may accuse you of prejudging the answer. The fact is, however, that this advance estimate does not invalidate in the slightest the standard error calculated from the results, nor cause any bias in the procedure. You must make this clear in your direct testimony.

In practice sample-sizes are based on both theory and experience, even though you do not make a fresh calculation for every sample-design. Theory is part of your experience. Without theory, experience has no meaning. Theory and experience together produce scientific advances. All this can be made clear, I believe.

I proceed now to describe some of the other problems of exposition that have arisen, and to offer some suggestions toward meeting them.

IMPLICIT FAITH IN THE COMPLETE COVERAGE, AND IN THE 10 PER CENT SAMPLE

A complete coverage, no matter how carried out, and even though it is incomplete (as complete counts too often are), has weight in evidence. A sample, unless it is a 10 per cent sample, has two strikes against it to start with. People who are not statisticians assume that the sheer size of a complete coverage will somehow cover up its incompleteness and the flaws in the method of measurement or in the interviewing. They believe that a judgment sample, if it is big enough, will do the same; and that it will in addition overcome biases of the unknown probabilities of selection.

A 10 per cent sample has almost equal standing with a complete count —maybe even better than a 15 per cent sample. Why, or what 10 per cent, is hardly ever questioned, even by experts in quantitative subject-matter.

The statistician, in the explanation of his sampling procedure, faces such preconceived ideas. The precision of a small sample, selected and estimated by an efficient probability procedure, will require justification. 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 over-all physical condition of the entire aerial plant which might be worth $200,000,000. But without very careful preparation to dispel preconceived ideas about complete counts and 10 per cent 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.

The troubles that people have in understanding the power of a small sample are often tied up with failure to understand that it is the absolute size (n) of the sample, and not its proportion (n/N) to the whole, which determines the standard error 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, because the big city is bigger.

With careful preparation, you can dispel such misunderstandings in an entertaining way, and in simple language. You can explain with black and white beans the statistical principles used, and why it is that the standard error of a sample is in practice hardly influenced at all by the size of the lot that it was drawn from. You can portray vividly how a pint jar of dried beans scooped up from a larger mixture of black and white beans will provide an estimate of the proportion black in the mixture; and that a sample of less than a pint would probably be sufficient. You may then observe, and your listeners will agree, that the mixture could as well be a carload of beans as a bushel of beans: the sample provides as good an estimate of the proportion black for the carload as it does for the bushel, *provided* that in both cases the mixture is thoroughly mixed (an illustration borrowed from testimony presented by Professor John W. Tukey). In practice we accomplish thorough mixing with the use of a table of random numbers—a tool indispensable today in science.

The nigh total failure of the size of a lot to have any influence on the standard error of a random sample drawn thereform is illustrated by charts in Eugene L. Grant's book, *Statistical Quality Control* (McGraw-Hill, 1946), page 345. Incidentally, such citations will often help the statistician's listeners to appreciate the fact that his methods are in universal use. One may usefully refer to the ever-expanding dependence of all kinds of scientific, industrial, agricultural, and medical research on statistical theory; the use of statistical methods to attain extreme precision in industrial production; the necessity for proper statistical design in the comparison of two industrial processes, machines, or medical treatments, the growing reliance, in many parts of the world, on probability samples in social and economic studies that are to guide important decisions.

If you succeed in making your explanation clear, you will help your listeners to appreciate the contribution of modern statistical principles and techniques to scientific truth. They may be grateful, in the long run.

Complex terms, flourished too freely, may alienate your listeners. Rely on patience, truth, and simple language. You can not afford to lose the attention of the examiner or judge; he is in position to protect truth and accuracy of statement. In cross-examination keep him on your side by your fairness and willingness to try to clear up any questions concerned with your sample.

PRECISION, ACCURACY, AND STANDARD ERROR

Two concepts that are important to make clear in any presentation are precision and accuracy. Most statisticians probably think that they know what these words mean. I must confess that experience under the fire of cross-examination taught me some new angles to their meaning, and taught me the importance of explaining in advance the limitations of a standard error.

Precision is expressible by an international standard, viz., the standard error. It measures the average of the differences between a complete coverage and a long series of estimates formed from samples drawn from this complete coverage by a particular procedure of drawing, and processed by a particular estimating formula.

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

The so-called "expected value" of a sampling procedure (which of course includes the formula for the estimate) is the same as the result of an attempted complete coverage of the same frame that the samples are to be drawn from (except for a possible bias in the formula, for which an upper and innocuous limit will be known). Both the complete coverage and the sample are subject to the same uncertainties and errors, such as inadequate supervision, nonresponse, wrong information, missing information, failure of workers to cover their whole assignments, and to find all the people or all the items. The only difference is that the sample has sampling error, which is the one error that we are best able to govern and to measure. The statistician measures the uncertainty introduced by sampling. The substantive expert judges whether the same operations would give accurate and useful information if applied to the entire frame.

The statistician will have drawn up the statistical procedures for the survey (the design of the sample, the instructions for drawing it, the instructions for tabulating the results and for computing the estimates and their standard errors). During the progress of the work, he should be on hand as often and as long as necessary to know that the company that retained him is following his instructions meticulously. He is then in a position to defend the validity of the standard error. If at any time he is not satisfied with the performance of the workers, it is better for him to terminate at once his relationship with the client. He should be sure that this responsibility is clear beforehand.

A statistician will occasionally be called upon to give his opinion in regard to procedures that another statistician has drawn up and testified to, or to give his interpretation of the results, including the standard error. After he has a chance to examine the procedures, he may testify, if he agrees, that they are one of many possible probability designs, and that IF they were followed meticulously, the results and the standard errors have certain interpretations, which he may give if called upon to do so. He may require, before he testifies, that certain calculations be carried out, to help him to examine the magnitudes of any biases that he may suspect. He may require calculations of skewness, if he suspects that the estimate of the standard error is not sufficiently firm. The results of these investigations will guide his conclusions and his testimony concerning the precision of the results of the survey. He must not be satisfied to testify to what he knows; he must explain how certain aspects of the survey that he had no opportunity to examine could possibly affect the results.

Even with familiarity with the job, and no matter how satisfied the statistician may be with the execution thereof, he can still not testify to the inherent usefulness of the result. Unfortunately, he has no standard error of the usefulness of a result. Testimony on the usefulness of the results will be left to the substantive expert—the engineer, the chemist, the physician, the population expert, the agricultural expert. The usefulness of a result is not a problem of sampling; it deals rather with the method of measurement and with reasons why the method used will produce data that will satisfy a particular need. The method would be the same whether the survey were a complete coverage or a sample.

In cross-examination the opposition may tempt the statistician beyond the sphere of his competence. The statistician must try to answer all questions politely and simply, yet he must stay within the limitations of his own ability and of the standard error. He certainly has a right to say he does not know the answer to a question that is beyond his competence and beyond his direct knowledge.

Although he can not testify to the inherent usefulness of the result, the statistician can certainly make it clear that he would not have associated himself with the study had he not been sure in advance that it would be executed rigidly in conformance with his specifications, and that the methods of inspection, interviewing, and questioning, although beyond his qualifications, would be satisfactory and produce useful data. He may do this without professing to be an expert in the subject-matter, as he may declare that he has confidence in Mr. So and So (expert in the subject-matter), who has testified, or will, concerning these things.

This division of responsibility between the statistician and the expert in the subject-matter should not be difficult to explain, but it is easy to forget to do it; and still easier later on in cross-examination to be lured across the border into the subject-matter and into trouble.

The following excerpt represents a statistician's attempt in direct testimony to state what his job was; to put a limitation on his assignment, and hence on what he could testify to in cross-examination. 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.

Q.[1] 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 to prescribe statistical procedures by which to select samples of items of plant for inspection,

[1] 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.

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).

Q. Does Company Exhibit No. 112 (Sampling Procedures for Drawing the Items of Property for Field Inspection) contain the procedures that you prescribed?

A. Yes sir, it does.

Later on came the following explanation of the standard error of the result:

Q. What is your interpretation of the standard error of this study?

A. The sampling precision of this study is expressed by the over-all standard error, which turned out to be .19 per cent. This standard error is not a matter of opinion nor of expert judgment, but is objective, as it is calculated by the laws of probability from the results themselves. The interpretation of this standard error is simple: I may say with a high degree of assurance that the maximum uncertainty that one may attach to the over-all per cent condition because of the introduction of sampling, can not be, at the outside, more than three times the standard error. In other words, any uncertainty in the figure 74.5% (the final result) which can be attributed to the fact that the company used samples instead of a complete and total inspection of every item, with the same care as was exercised on the samples, were such a thing possible, can not exceed .57 per cent.

THE PERMANENCE OF THE STANDARD ERROR CONTRASTED WITH THE TEMPORAL CHARACTER OF ACCURACY AND USEFULNESS

In a probability sample (the only type of survey to be considered here) the precision is calculable from the results, as I mentioned in the opening paragraphs. In practice, the size of the sample will be sufficient to provide a firm estimate of the standard error. This was the case in the excerpt above. One may say with a high degree of assurance that in a long series of repetitions of this sampling procedure, only about 2.3 per cent of the results would fall 2 standard errors above the result of the complete coverage, and about 2.3 per cent of the results would fall 2 standard errors below. Practically none of the long series would fall beyond 3 standard errors either way.

It is another thing, however, to say whether the complete coverage, were it possible, would produce useful information. The inherent accuracy of the method of measurement (the interviewing, the questionnaire, the method of inspection), and the usefulness of the information, whether obtained by a complete coverage or by a sample, is a matter for the substantive expert to testify to, as explained earlier.

The main difference between a sample and a complete count is that the sample possesses an error of sampling. The statistician testifies in regard to this. A sufficient degree of precision is necessary for the usefulness of sample results, but it does not guarantee their usefulness.

The inherent accuracy and usefulness of the procedures of measurement will change from time to time as the substantive experts develop new concepts of the kind of information that they require to solve new and changing problems. Anyone who has followed the changing concepts of the characteristics of the labor force, or the changing concepts of a farm, or of family-budget studies, or the changing concepts of the desirable characteristics of fibres and of textiles, will know that no definitions or methods of measurements stay fixed.

In contrast, the standard error of a procedure of sampling remains fixed with time; likewise the interpretation of the standard error. The validity of the standard error does not depend on the economy or cleverness of the design of the sample. It depends only on careful execution and on the rigid use of probability methods in accordance with some prescribed statistical plan. For this reason, the standard error of a sampling procedure and its interpretation, remain valid, even though new advances in theory point the way to more economical sampling procedures by which to obtain the same standard error.

REFERENCES

There is apparently no previous literature that deals with the presentation of modern statistical procedures and their results in legal evidence. Fortunately, however, sampling has received attention from the legal standpoint in a paper by Frank R. Kennedy, who supplied copious remarks and references to cases in which samples of one kind or another were offered in evidence.[2]

In conclusion, it is a pleasure to acknowledge aid from a number of friends, chiefly from Mr. Melvin F. Wingersky, Attorney at Law, and a member of this Association. This acknowledgment is of special interest, because he cross-examined me on several occasions with great vigor. Valuable help came also from Mr. Harlow A. Coxe, General Staff Engineer of the Illinois Bell Telephone Company; also from Mr. Howard L. Jones, statistician, and from Mr. Gordon Winks, General attorney, all with the same company. Finally, I have had the benefit of a long association and many conversations on this subject with Professors John W. Tukey and Frederick F. Stephan of Princeton.

[2] Frank R. Kennedy, "Some legal aspects of sampling," Industrial Quality Control, Vol. vii (January and March 1951).