On Some New Methods and Concepts in Sampling

By

W. EDWARDS DEMING
Consultant in Statistical Surveys
Washington

Professor of Statistics
Graduate School of Business Administration
NEW YORK UNIVERSITY

Reprinted from SCIENCE IN MANAGEMENT,
American Society for Quality Control,
Washington meeting, 5 Mar. 1959: pp. 89-96

ON SOME NEW METHODS AND CONCEPTS IN SAMPLING

By

W. Edwards Deming
Consultant in Statistical Surveys
Washington

The equal complete coverage. The reliability of the information derived from a sample-survey carried out by probability methods is in most ways little different from the reliability that would be inherent in a complete coverage. The only difference is that the results of such samples possess sampling errors. One sometimes hears the statement that samples are more accurate than complete coverages, the argument being that the complete coverages are afflicted to a greater degree with nonsampling errors. A more correct statement is that samples have the possibility of being more accurate, and many of them are, because their smaller size permits better control of the work. The fact is that the quality of either a sample or a complete coverage is only what you build into it. Quality is not an accident in either a sample or a complete coverage.

A number of new concepts are helping statisticians to understand and explain the results of samples and experiments, and to improve the usefulness of sampling. For example, we are all grateful for the word frame as a means of access to the universe, or to as much of the universe as we can afford to cover in the preparation of a survey.* A further new concept is the equal complete coverage, which we define as the coverage of every sampling unit in the same frame that the sample was drawn from, carried out by the same people that worked on the sample, exercising the same care and definitions in the field, and the same rules in the coding and processing.

When we draw a sample, we draw not material from the frame, but samples of the labels that the equal complete coverage has already or would place on every sampling unit. The equal complete coverage is

---

* Frederick F. Stephan, "Practical problems of sampling procedure," Amer. Soc. Rev., vol. 1, 1936: pp. 569-580.

often a conceptual model, as the quality of the performance in the field and in the coding may change when we attempt to increase the size of a sample to 100%.

The equal complete coverage leads to an operational definition of the standard error, which we may now define as the root-mean-square difference between the estimates made by a long series of samples drawn from the same equal complete coverage and processed by a specified sampling procedure.* Failure of the frame to cover all the universe, and deficiencies that are inherent in the questionnaire or in the method of test or interview or in the coding or in the tabulations will affect, first of all, the equal complete coverage, and hence samples drawn therefrom.

The equal complete coverage leads also to a logical definition of the accuracy and the bias of a given survey-technique, which may mean a certain test-method, or a questionnaire, or a technique of interviewing, including the procedures for coding the tests or interviews. Whatever be the survey-technique, a complete coverage of all the sampling units in the frame will produce some result. A new complete coverage conducted with the same survey-technique, and before any changes have taken place, will give results slightly different from the results of the first coverage. This is so because people do not always give the same answer when asked a 2d time. Or, some other member of the household may answer in the 2d complete coverage. Some other survey-technique (different definitions, different questions, different procedures) would give still another figure. Neither figure is right or wrong, and neither of them is a true value. There is no such thing as a true value.

Any result, whatever it be, is the result of applying some set of operations. Although there is no true value, we do have the liberty to define a set of operations and to say that we shall accept the results thereof or that we shall not.

There may be, by agreement of the experts in the subject-matter, for any desired property of the material, a <u>preferred</u> survey-technique.

---

* Morris H. Hansen and W. Edwards Deming, "On an important limitation to the use of data from samples," Bulletin de l'Institut International de Statistique, vol. xxxii, 1950: pp. 214-219.

Unfortunately, it often happens that the preferred technique, usable on a laboratory-scale, is too expensive to apply in a full-scale survey, or it may be objectionable otherwise. The experts must then supply also a working technique. Thus, the preferred technique by which to define a person's age might be to compute the difference in time today and the date on his birth-certificate. But some people don't have birth-certificates at all, and few people have them handy. Moreover, some people would not be happy with an interviewer who asked for birth-certificates. The Passport Division can ask for birth-certificates, but interviewers may only ask the person how old he is, or his date of birth, and record the result. This would be the working technique by which to measure age.

The preferred technique and the working technique will give different results. A working technique is acceptable to the experts if it gives results not too far from the results of the preferred technique.

The difference in the 2 techniques, applied to a complete coverage of the frame, is the <u>bias</u> of the working technique. A working technique is said to be <u>accurate</u> if its bias is small.

<u>Meaning and limitations of the standard error</u>. The standard error as defined above automatically allows for the effect of variable performance in the field-work and in the office, including variable judgment in subjective and visual tests, and for differences between investigators. It allows for the effect of variable mistakes in selection of the sample and in processing the data. This statement follows from the fact that such variations are embedded in the equal complete coverage and hence they appear in samples drawn therefrom.

The margin of sampling error does not detect nor include the effect of persistent errors, such as the selection of convenient units, or of new or old units, instead of units designated by the random numbers; persistent omission of a portion of an area, getting out of bounds, omission of information, persistent overcounts or undercounts, omission of a whole area or of portion of units.

All these errors except persistent wrong selection of sampling units are nonsampling errors--i. e., they are defects in the equal complete coverage, and are not attributable to the use of sampling. Nonsampling errors and blemishes in selection can be detected and their effect measured only by a statistical control or audit, or by outside comparisons. The statistical control or audit is a careful re-test of a small probability sample of the main sample.

The margin of sampling error, coupled with the information in the audit and in any outside comparisons, enables one to evaluate the total loss in reliability that arises from operational errors.

The statistical control is incidentally as necessary for a complete census or for a complete inspection as it is for a sample. It is an important part of our own Census.*

Three standard errors is an international standard for the practical maximum variation between repeated samples, drawn from the same equal complete coverage, and processed in the manner specified. One must of course make proper allowance, when occasion arises, for extreme skewness and for a small number of degrees of freedom.

It follows that if a standard error is small, and if the audit shows few departures from the prescribed sampling procedure, then the result of the sample is the same within narrow limits as the result that would come from the equal complete coverage. Whether the equal complete coverage would be useful or acceptable as evidence in a scientific or legal enquiry is another matter, to be settled by the judgment of experts in the subject-matter and in law. This judgment will be based on the content of the enquiry or on the method of test (whichever is applicable); also on the report of the audit, supplemented by any outside comparisons that are reliable. It will be entirely independent of the margin of sampling error. If the equal complete coverage would not serve the purpose, in the judgment of the experts, then neither would any sample serve the purpose, no matter how small be its standard error.

If the margin of sampling error is too big for the purpose, then the usefulness of the sample is in doubt, regardless of whether the complete coverage would be useful.

Effective division of responsibility. When we focus attention on the equal complete coverage, we see that statistical inference carried

---

* Hansen, Hurwitz, and Madow, *Sample Survey Methods and Theory* (John Wiley, 1953), vol. 1, p. 544, 581.

out by the theory of probability covers only the given frame and the conditions subjected to the sampling procedure. Generalizations to other cities, climates, conditions, and levels, not covered by the frame nor subjected to the sampling procedure, require subjective judgment, to which statistical theory offers little aid. The same remark applies to predictions, except when the samples come from a state of statistical control.

A logical division of responsibility now becomes clear. The content of the questionnaire, and the method of tests or of interviewing, the coding, and the generalization of the results beyond the frame and beyond conditions not subjected to the sampling procedure, are the responsibility of experts in the subject-matter. The design of the sample or experiment, and the objective inferences made with the theory of probability, are the responsibility of the statistician. Both people must of course work together to formulate the problem in statistical terms, and to decide on the precision that will serve the purpose.

<u>A simple example of calculations with a replicated sample</u>. The result of a sample without a standard error is of limited utility. Fortunately, it is possible, by replication of a sample-design, to reduce to simple arithmetic the calculation of the standard error. A simple example will indicate the calculations. A marketing research company conducted a survey in New York in October 1955, and amongst other estimates, the sample produced an estimate of the total number of sampling units.* The frame was the list of blocks by tract in the 1950 Block Statistics, which shows for each tract and for each block the number of occupied dwelling units therein. The number of sampling units in an area was by definition the number of occupied dwelling units shown by the Census of 1950, divided by 10. The sample was replicated in 2 subsamples. There were 16 zones in the frame, each zone being 2000 successive sampling units. The zones were thus small geographic strata. Subsample 1 consisted of 1 random drawing in each zone (stratum). Subsample 2 consisted of a 2d random drawing from each zone. The probability that any dwelling unit or any person attached thereto would be

---

* I am indebted to the firm O'Brien-Sherwood of New York for the figures in the table and for the privilege of working with them on the survey.

drawn into Subsample 1 was thus exactly 1 in 2000, and likewise for Subsample 2, regardless of growth or of loss in the number of dwelling units in an area.

Each sampling unit drawn in the field had an intended size of 2 segments of 5 dwelling units each, on opposite halves of the block. Most of the sampling units did actually contain 2 segments, although a few contained but 1, and a few contained 3. Most segments contained 5 dwelling units, although some contained 3 or 5 dwelling units, and some contained 6 or 7. The interviewers had instructions to vary the sizes of the segments in a block in order to give priority to the boundaries. It is most important that the boundaries be definite and unmistakable. Random numbers in a sealed envelope drew the segments for interview, once the interviewer had created the segments in the block. Use of the half-open interval in the field is a simple way to define sampling units in urbanized areas, with a diagram in case of possible doubt about the shapes or boundaries of a segment.* Variation in the number of segments and in the sizes of segments introduces no change in the probability of selection and causes no bias. The effect on the standard error is automatically included in the calculations (_vide infra_).

As each sampling unit had the same probability as another, the design described draws blocks and other areas with probability in proportion to the number of sampling units designated therefor in the frame. Probabilities in proportion to the square root of the number of dwelling units shown in the block statistics, or in any other proportion, are easy to arrange with the replicated method, simply by increasing or decreasing the number of dwelling units in a sampling unit, and there is no change in the probabilities nor in the method of computation.

The accompanying table shows the counts of dwelling units in the 2 subsamples zone by zone. An unbiased estimate of the total number of dwelling units is

$$X = \frac{1}{2} \times 2000\ (1246 + 1215) = \frac{1}{2} \times 2000\ x = 2{,}461{,}000 \text{ dwelling units}$$

An estimate of the coefficient of variation of X [illegible]

---

* W. Edwards Deming, _Some Theory of Sampling_ (John Wiley and Sons, 1950), p. 82. The replicated method is described by the author in the J. Amer. Stat. Assoc., vol. 51, 1956: pp. 24-53.