2-1951

Minimum Standards for Testing and Sampling Procedures

Herbert G. Bowles

Herbert G. Bowles is a Certified Public Accountant and a resident partner of Lybrand, Ross Bros. & Montgomery, a national firm of certified public accountants. Mr. Bowles is first vice-president of the Los Angeles Chapter of the National Association of Cost Accountants and a state director of the California Society of CPA's.

This excellent paper was presented by Mr. Bowles at the Public Relations Meeting of the Los Angeles Chapter of ASWA. We are indeed fortunate in being able to bring his message to the attention of the readers of "The Woman C.P.A."

MINIMUM STANDARDS FOR TESTING AND SAMPLING PROCEDURES

HERBERT G. BOWLES

Historically, the use of testing and sampling techniques in the conduct of business became necessary at the same time that accounting records themselves became necessary, that is, when the volume of business transactions grew so large that the proprietor could no longer personally remember and check each one.

The sampling of agricultural and other products for indications of quality must have been practiced from the very dawn of civilization. Human nature changes slowly, if at all, and the outstretched hand of the housewife at the Brooklyn fruit stand must have had its counterpart thousands of years ago. Then, as now, the best sampling procedure was to pinch every apple in the bin. Practical considerations largely prevent taking full advantage of this best of all sampling methods.

During World War II, a statistical sampling technique was developed by the National Defense Research Committee for use in testing munitions products. This method has been applied in some degree to professional auditing problems and deserves the careful study of all accountants whose work includes the selection and application of sampling and testing techniques.

Although sampling tests have always had the common objective of satisfying the tester and that, if pursued further, the tests would disclose conditions similar to those already found, there have been no generally recognized testing standards or minimum requirements upon which such a conclusion could be predicated with reasonable assurance. This conclusion has been in the field of professional judgment and may properly always remain there. The new techniques referred to do, however, offer a mathematically and statistically, and, therefore, scientifically sound basis for a conclusion as to the probable character of all the data in the group from which the particular sample was drawn, and this assurance is important to a professional conclusion.

If professional accountants, internal auditors, and others making sampling tests of accounting data should be required to substantiate their work and justify the conclusions which they have reached, it could be very helpful to be able to show that the testing procedure was scientifically designed and that the area of substantiation required was therefore limited to selection of methods, interpretation of results, and to other pertinent considerations. This happy position is not held today, in the usual case, because of the unlimited variation in types of sampling techniques, and because there is no generally accepted auditing standard for minimum requirements.

It will be helpful to review, at this point, the weaknesses, or areas of vulnerability, in sampling techniques commonly utilized, which are not present in the scientifically sound sequential sampling technique which will be described shortly.

A percentage method is used frequently and is sometimes applied loosely. If a 5% test is programmed and if budgeted time runs out when only 4% of the data have been examined, the remaining items may be given a 1/10 of 1% scrutiny or none at all.

The percentage method does not function well with extremely large numbers of items because of the burdensome time requirements. One per cent of 100,000 checks is 1,000 checks and one per cent of 100,000,000 sales rung up on cash registers is 1,000,000 of such sales. The use of inflexible percentages for sampling may require so much auditing time as to be impracticable, or result in such low percentages as to constitute only token testing. The selection of a percentage sample when the percentage method is used is influenced by individual bias and preference. One out of every twenty is 5% but it may be every twentieth item or it may be one-twentieth of the total in a group off the top, off the bottom, or from the middle in the body of detail. There is a tendency for samplers to develop habits, such as selecting April and August for test months, or checking every tenth page footing. These habits become known as time passes and some risk is bound to be incurred that personnel responsible for the paper work will take liberties with those records resonably safe from selection by examiners.

Some auditors co-operatively leave their special colored pencil tick marks in the records year after year, to provide interested clerks with a reliable program of what will probably be examined in the next audit. In addition to these and other weaknesses in conventional sampling procedures, they usually fail in an important mathematical and statistical criterion. Each one of the entire body of data should have an equal chance of being selected in the sample. If every tenth item is drawn, the other nine have no chance whatever of being examined.

Not only should every item have an equal chance of being drawn in the sample, but it must be "replaced" after drawing, and be available, therefore, to be drawn again.

If drawn again, and this is not unlikely, the results of the sampling should list the item twice, with the same effect as if it had been two different items.

In the case of a calendar year corporation for which proofs of cash are prepared for two selected months as a test, how often is the month of February selected? Almost never, because the auditors usually are still working on the December 31 balance sheet during February and feel that February is too early in the year to be a typical month, or has too few days, or is likely to be freer from good test characteristics for other reasons.

Factors which may influence both the manner and extent of sampling when no minimum standard is utilized include the available fee, illness, temperamental attitude or caprice on the part of the individual making the tests and local conditions for conducting the tests. All of these factors can be eliminated if scientific sampling is applied, and the area for application of professional judgment can then be restricted to consideration of other pertinent matters.

Scientific sampling involves the selection of a number of specimens at random. The tabulated number of errors found, in relationship to the number of items in the sample, is then analyzed to determine which of three courses of action should be taken, as follows:

1. Conclude that the data from which the sample was drawn are acceptable.
2. Conclude that the data from which the sample was drawn are unacceptable.
3. Make further sampling tests of the data in order to reach an assured conclusion.

Minimum numbers required for a particular sample are computed by use of a formula. The minimum number is an absolute number and is not affected by the total number of items to be tested. If then, we are to test 100 cleared bank checks, or 1,000, or 100,000, or 100,000,000 the minimum number required to be drawn as a sample is the same. One suggested formula which may become widely used produces the number 77 as the minimum. The sample number is determined by use of a mathematical equation containing four elements, as follows:

1. The percentage of errors which is considered acceptable, that is, the maximum number of errors which may be indicated to be present without requiring a conclusion that the

data are unreliable. As an example, .001 (or less) or 1/10 of 1%, or an average of one error in every 1,000 items.

2. The percentage of errors which, if indicated to be present, justifies a conclusion that the data are unreliable, are not substantially correct, and are unacceptable, from an audit viewpoint. Unless the data are rechecked in detail and corrected throughout, this finding may require a qualification in a report or a disclaimer of any opinion. As an example, .03 (or more) or 3% or an average of 3 errors in every 100 items.
3. The mathematical probability that, on the average, there will, by chance, be samples drawn which are not representative and will indicate a satisfactory body of data when, in fact, the data are not satisfactory. For example, .10 or 10% or once in ten times.
4. The mathematical probability that, on the average, there will, by chance, be samples drawn which are not representative and will indicate an unsatisfactory body of data when, in fact, the entire group are within acceptable limits of error. For example, .05, or 5%, or once in twenty times.

The examples cited, that is, the setting of an acceptable incidence of error at 1/10 of 1%, an unacceptable incidence of error at 3%, and utilizing probability factors for erroneously reaching a favorable conclusion once in ten times and for erroneously reaching an unfavorable conclusion once in twenty times, produce the absolute number 77 as the minimum sample required to be drawn and examined.

One of my colleagues, in discussing the use of an absolute number, such as 77, commented upon the great disparity between such a relatively small number and the number which might be required to be examined in a 1% test, or 1,000,000 cash sales out of a total of 100,000,000. "Do you mean to tell me," he said, "that equal reliance can be placed on a test of 77 out of 100,000,000 as on a test of 1,000,000 out of 100,000,000?" I asked him if he had ever gone to the beach and tested the water with his toe before going in, and what percentage of the Pacific Ocean he thought that he had sampled.

The question is not what portion, or share or percentage of the whole should be examined but how few items need to be examined to reach a conclusion as to the probable character of the others, regardless of their number. In the illustration cited, assuming that no errors were disclosed through examination of the 77 items, the conclusion can be supported that the entire 100,000,000 items are probably substantially correct. If the 1% test had been consistently applied by examining every tenth item, and if this method had become known to persons responsible for producing the data, the examination of 1,000,000 would not warrant as much assurance as to the probable character of the other 99,000,000 as would the scientific test of 77.

It is of course true, that, other things being equal, the larger the sample, the more assured the result in any testing program.

It should be pointed out that scientific sampling deals with numbers of errors indicated, but not with the amounts of such errors. An inventory extension transposition error of $990 would be one error; also an inventory extension error of $.01; or of $100,000. For this reason, data to be sequentially sampled should first be arranged in dollar amount groupings, in so far as this is feasible, and separate tests then applied to each group. In this connection, examiners will continue to be interested in the character of and reason for errors disclosed by any sampling technique and to recognize those resulting from wilful manipulation or other fraudulent practice.

The question is frequently asked: How are formula factors selected and how great is the variation in the resulting absolute number for the minimum sample?

Let us suppose that an internal auditor is testing performance of personnel responsible for the receiving function and is sampling receiving report forms for completeness. Assume that a receiving slip is not dated. If other prenumbered forms before and after the undated one are complete and the approximate date of the incomplete one can be determined, it would not appear to be too serious an omission. For this test the internal auditor might select a factor as high as 5% or 6% of errors as an acceptable showing, and set his unacceptable error factor at 15% or 20%.

But consider an examination of signatures to corporate bank checks to test proper authorization by the board of directors. Normally, no errors would be expected, and the factors might be 1/1000 of 1%, and 1%, respectively, for acceptance and rejection.

Percentage acceptance and rejection factors must be selected for each sampling operation. Probability factors may also be modified to better fit the needs of individual situations. In the selection of the four formula factors, professional judgment must be employed, and in interpretation of the results of sampling tests, professional judgment must again be exercised.

To illustrate the range in absolute numbers for samples, a selected group of combinations has been tabulated utilizing the probability factors already mentioned for erroneously accepting or rejecting the data on the basis of the results of a particular sample, that is, 1 in 10 and 1 in 20, respectively, as follows:

	Range	
	From	*To*
Acceptable error percentages	.001	.03
Unacceptable error percentages	.02	.10
Minimum sample size	27	222
Average sample size if of acceptable error character	37	1420
Average sample size if of unacceptable error character	13	1524

As may be inferred from its name the sequential sampling technique contemplates the drawing off of a sample, the tabulation of the results, and the determination at that point of whether an additional sample should be taken or whether a reasonable conclusion can then be reached as to the probable character of the entire body of data.

The application of the method in practice utilizes a chart or a table for ready reference. The chart is a simple right angle, the horizontal arm being graduated for number of items sampled, and the vertical arm being graduated for number of errors disclosed. As the sampling proceeds, the varying numbers of errors developed with increasing numbers of items sampled can be plotted as points within the right angle and a line drawn connecting the points.

The position of these points and their connecting lines in relationship to two fixed diagonal lines from the lower left to the upper right determines whether the test is sufficient or whether additional tests need to be made. If the plotted points fall below the lower of the two fixed diagonal lines no further tests are required and the data may be accepted; if above the upper of the two fixed diagonal lines, the conclusion can be supported that the data are unacceptable; if between the two fixed diagonal lines, no conclusion is possible and the drawing of additional samples is indicated.

The two fixed diagonal lines are determined by the formula used, that is, by the selected criteria as to maximum acceptable percentage of errors, minimum unacceptable percentage of errors, and probability factors for erroneous conclusions from the particular sample or samples drawn. By varying the elements of the formula a number of different sets of two fixed diagonals each may be constructed and the plotted results of the sampling may be interpreted in the light of different assumptions as to composition of the data.

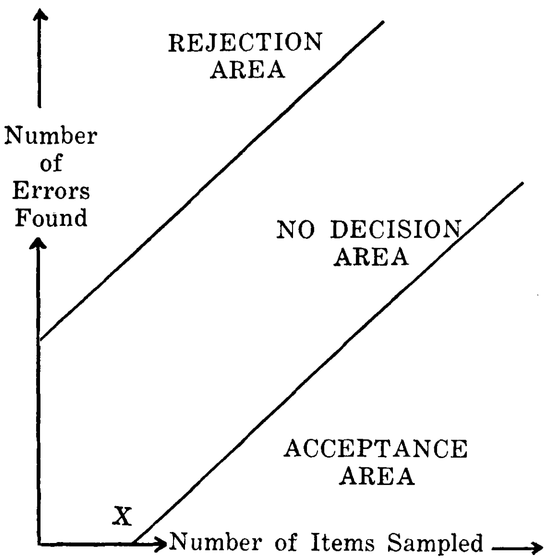
To those disinclined to use pictorial displays in graphic charts a table may be constructed which will indicate at what point in the sampling procedure a conclusion is possible.

Illustrative graphs and tables, together with an explanation of the scientific sequential sampling technique are available in a book recently published by the University of California Press, Berkeley, California, and written by Lawrence L. Vance, Associate Professor at the University of California at Berkeley, entitled *SCIENTIFIC METHOD FOR AUDITING.* This book contains much of the material presented in this paper and has been freely drawn upon for basic principles.

The mathematics of the technique is too involved for discussion at this meeting; and interested persons are referred to the book by Vance or to other publications listed in the excellent bibliography in that book. A word may be said about random numbers and their selection for use in pulling items to be sampled. Because of conscious or unconscious personal bias and preference for certain numbers, it is considered more reliable to utilize mathematically tested random number sequences which may be obtained from published tables. There is no objection, of course, to any method of selection of items to be sampled if the basic principle is observed; that is, if every item has an equal chance to be drawn in the sample each time that an item is drawn. The term "subjective randomization" has been used to describe the technique whereby samples are selected by what the sampler believes to be a truly random technique but without use of mechanical shuffling or sorting devices or reference to published tables of random numbers.

Sequential sampling techniques have had only limited application in professional auditing practice. The formulas described in this paper were developed during World War II and their full adaptability to the needs of the profession will be demonstrated only through general acceptance by professional accountants in the years to come. Such general acceptance may not be expected to occur rapidly. The primary purpose of this paper is to invite the attention of the profession to the techniques; to the possibilities of adapting them generally in professional auditing practice; to the possible desirability of establishing minimum auditing standards for sampling tests; and to emphasize the dynamic, ever-changing, ever-questioning attitude which the profession must foster and nourish if it is to continue to provide acknowledged leadership in those phases of economic development in which accountancy plays such a vital role.

GUIDE CHART
for
PLOTTING RESULTS OF
SEQUENTIAL SAMPLING TESTS



X=Point determining minimum number required to be sampled to justify acceptance.

Adapted from the book,
Scientific Method for Auditing, by Vance.

GUIDE TABLE
for
REFERRING RESULTS OF
SEQUENTIAL SAMPLING TESTS

Risk Factors For:
Erroneously accepting undesirable data .10
Erroneously rejecting desirable data .05

Quality Factors:
Acceptable Error Percentage .005
Unacceptable Error Percentage .03

Number of Items Sampled	Number of Errors Found Which Will: Permit Acceptance	Not Permit Decision	Require Rejection
2		1	2
29		1	2
89	0	1, 2	3
100	0	1, 2	3
160	1	2, 3	4
171	1	2, 3	4
239	2	3, 4	5
243	2	3. 4	5
303	3	4, 5	6
314	3	4, 5	6
374	4	5, 6	7
385	4	5, 6	7

Risk Factors For:
Erroneously accepting undesirable data .10
Erroneously rejecting desirable data .05

Quality Factors:
Acceptable Error Percentage .005
Unacceptable Error Percentage .05

Number of Items Sampled	Number of Errors Found Which Will: Permit Acceptance	Not Permit Decision	Require Rejection
2		1	2
39		1	2
49	0	1, 2	3
89	0	1, 2	3
100	1	2, 3	4
140	1	2, 3	4
151	2	3, 4	5
191	2	3, 4	5
201	3	4, 5	6
241	3	4, 5	6
252	4	5, 6	7
292	4	5, 6	7
303	5	6, 7	8
343	5	6, 7	8

Adapted from the book,
"Scientific Method of Auditing," by Vance.