Statistical Sampling Revisited

By Neal B. Hitzig

Auditing standards are undergoing revision in the wake of recent, massive audit failures. Legislative and regulatory bodies are focusing more critically on auditors than ever before. Yet, contemplated revisions to auditing standards leave untouched ambiguities and unresolved issues that have reduced the effectiveness of the authoritative literature for decades. One of the longest-standing issues concerns the role and appropriateness of statistical sampling as a substantive audit testing procedure.

Background

Throughout the 1960s and '70s, the largest accounting firms devoted extensive resources to the development and implementation of statistical sampling procedures. The firms wrote new policies and guidance, developed time-sharing and batch computer programs, and trained specialized staff. Monetary unit sampling was developed and became a widespread audit tool. The AICPA issued Statement on Auditing Procedure (SAP) 54 and published *Statistical Auditing*, by Donald M. Roberts.

Then, in 1980, the Auditing Standards Board (ASB) issued SAS 39, *Audit Sampling* (AU 350). Members of the Statistical Sampling Subcommittee that wrote SAS 39, which included this author, expected that the imposition of risk, materiality, and selection requirements would further establish statistical sampling as a principal audit testing procedure. In fact, the opposite has occurred, largely because the ASB gave nonstatistical sampling equal evidentiary weight.

Substantive Tests

Substantive tests are intended to detect and estimate misstatement in accounts and classes of transactions. The authoritative literature recognizes two types of substantive tests: tests of details and analytical procedures. Except in those cases where complete enumeration of an accounting population is feasible (as in certain computer-assisted auditing techniques), the audit sample is a principal approach to performing the test of details.

Many auditors apply sampling to test controls, despite concerns that such applications may not reveal the information that an auditor seeks. For example, the initialing of documents does not mean that the documents are correct (if that is what initialing purports to signify); it means only that the documents were initialed. Similarly, the fact than an invoice is correctly priced does not mean that a price-checking control functioned properly, because the invoice may have been properly priced in the first place. These examples demonstrate why testing preventive controls with tests of details may not inform the auditor that the subject controls are functioning as intended.

Nonstatistical Sampling

AU 350 does not provide a definition of nonstatistical sampling. It states only that “[t]here are two approaches to audit sampling: nonstatistical and statistical” (AU 350.03). The AICPA’s Audit Guide, *Audit Sampling*, provides the following definition:

Any sampling procedure that does not measure the risk is a nonstatistical sampling procedure. Even though the auditor rigorously selects a random sample, the sampling procedure is a nonstatistical sampling application if the auditor does not make a statistical evaluation of the sample results. (AAG-SAM 2.18)

This statement establishes that an auditor may label a sampling technique “nonstatistical” without regard to the manner of sample selection. Thus, even though the Audit Guide acknowledges the well-known ability of statistical sampling to measure sampling risk, it nevertheless sanctions an auditor’s decision to ignore available statistical theory and rely instead on judgment.
or intuition in interpreting the results of a sampling procedure. In short, the guide gives guesswork equal status with measurability. Such a view is potentially hazardous, because the auditor is permitted to ignore facts that are readily discernable to any practitioner, or legal adversary, who is knowledgeable in the application of statistical methodology.

Why would an auditor prefer nonstatistical sampling, knowing of the availability of objective statistical procedures? Various reasons, restated in the 2001 edition of the Audit Guide, have been cited as the impediments: the cost of training, the cost of sample selection, the cost of sample evaluation. With the passage of time, these reasons have become progressively weaker. Mandatory continuing professional education is now a reality, so there should be little reason for auditors not to advance their skills in sampling techniques. As to the implementation costs associated with the selection and evaluation of random samples, the ready availability of computers and off-the-shelf software has greatly mitigated, if not eliminated, these factors as relevant considerations.

In short, a nonstatistical sample is selected by the exercise of judgment, and not by chance. Haphazard, judgmental, and purposive sampling are some of the terms that describe a nonstatistical sample.

Statistical Sampling

AU 350 and the Audit Guide approach statistical sampling in a roundabout way. The Audit Guide states:

**Statistical sampling helps the auditor (1) design an efficient sample, (2) measure the sufficiency of the evidential matter obtained, and (3) quantitatively evaluate the sample results.**

Statistical sampling uses the laws of probability to measure sampling risk. (AAG-SAM 2.17)

Although the foregoing statements are correct, they do not define statistical sampling per se.

**Statistical sampling is probability sampling.** In probability sampling, every item in the population under audit has a known chance of selection. The decision as to which items in the population are to be selected is left to the laws of chance, not to judgment. The most common probability sampling methods in auditing are equal probability (such as simple random and systematic sampling) and sampling with probability proportional to size (such as monetary unit sampling).

The prominent feature of statistical sampling is its ability to measure risk. The measurement instrument is the confidence interval, which gives a calculated range of values for the estimated amount of misstatement in a population. The measurability of statistical sampling distinguishes it from so-called judgment sampling, where the decision as to the items selected for examination is left to the judgment of the auditor. **Statistical sampling is a measurement tool.** When applied in a substantive test of details, it measures misstatement in an account or class of transactions. Its ability to measure arises from the selection method used, which is probability sampling. Lawyers, judges, and statisticians have explicitly recognized these features of statistical sampling. The Special Committee on Empirical Data in Decision Making, Recommendation on Pretrial Proceeding in Cases with Voluminous Data, made the following statement (see Appendix F, in Fienberg, S.E., ed., *The Evolving Role of Statistical Assessments as Evidence in the Courts*, 1989):

> When a survey is based on probability sampling, the probabilities or risks of sampling misstatements of various sizes can be calculated. This requires the application of *appropriate* statistical formulas. Assessments of sampling misstatement are very often expressed in terms of a standard misstatement. This is a universally accepted measure of the margin of error in a survey result that is attributable to sampling.

This illuminating report should serve to alert auditors to the growing use of statistically based evidence in litigation and, by implication, to the risks they face should they ignore the information contained in samples.

The implication is clear: Ignore the formulas applicable to the results of a probability sample and rely instead on intuition at your own risk.

Some auditors believe that they must calculate a sample size beforehand for an audit sample to be statistical. This is incorrect. Any probability sample can be subjected to evaluation by application of the laws of probability, however arbitrary the choice of sample size. Failure to calculate beforehand usually results in samples that are either too large or too small for the auditor’s objectives. They are, nevertheless, statistical.

Statistical and nonstatistical sampling methods are defined in terms of the method by which a sample is selected, not in terms of a decision by the auditor not to apply statistical methods, even to a random sample.

**When Is Statistical Sampling Appropriate?**
Statistical sampling is appropriate whenever an auditor wishes to draw a conclusion about a population without performing an examination of all the items composing that population. Moreover, statistical sampling is appropriate when the auditor has no prior knowledge as to which specific items in a population are misstated.

An important concern that affects the sampling decision is the practicability of selecting a probability sample. If files are computerized and 100% verification cannot be performed by computer-assisted audit techniques, then probability sampling is most likely to be the practical approach. If files are not computerized and the population is large (as a rough rule of thumb, a large population has more than 500 items), then probability sampling may still be practicable. If a population of manual records is maintained in numerical order, a computer application may be used to select random numbers that identify the items to be selected, even items at multiple locations. The items are then located by hand. If the population is not maintained in numerical order, then systematic selection (select every kth item after a random start) may be performed. Systematic selection is one of the easiest procedures to apply, although proper application requires counting through the population. Although many caution that systematic selection is subject to bias because a key characteristic of the population under examination may coincide with the selection interval, in more than 30 years of practice, the author has never observed this to be even a remote practical concern.

Sampling Risk

AU 350 states “[s]ampling risk arises from the possibility that, when a test … is restricted to a sample, the auditor’s conclusions may be different from the conclusions he would reach if the test were applied in the same way to all items in the population.” (AU 350.10) AU 350 also identified two aspects of sampling risk:

The risk of incorrect acceptance is the risk that the sample supports the conclusion that the recorded account balance is not materially misstated when it is materially misstated. (AU 350.12)

The risk of incorrect rejection is the risk that the sample supports the conclusion that the recorded balance is materially misstated when it is not materially misstated. (AU 350.12)

In practice, it is convenient to think of the foregoing in terms of detection risk and estimation risk, respectively.

Detection risk is the chance that a sample will fail to detect misstatement that actually exceeds the auditor’s specified maximum tolerable amount. “Detection” refers to the decision rule that an auditor applies to decide whether a misstatement is tolerable under the circumstances. A commonly employed rule is the comparison of the calculated upper confidence limit of misstatement with the specified maximum tolerable amount. In SAS 39 terms, the upper confidence limit is the projected misstatement plus the allowance for sampling risk. If the calculated limit is greater than the maximum tolerable amount, the auditor decides that misstatement may exceed the tolerable amount. Otherwise, the auditor decides that misstatement, if it exists, is tolerable. If a properly designed sample discloses no misstatements, the auditor may then decide that misstatement in the population under audit does not exceed the maximum tolerable amount.

Detection risk is principally a planning concept. The auditor specifies it beforehand and uses it as one of the factors that determines the appropriate extent of testing reflected in the sample size. If misstatements are detected, on the other hand, the estimation risk becomes the key risk under consideration. Estimation risk is the chance that the actual amount of misstatement will not be within the calculated confidence interval. SAS 39 is dismissive of this risk, which it labels the risk of incorrect rejection, as being merely an efficiency issue. AU 350.12 states:

[If] the auditor’s evaluation leads him to the initial erroneous conclusion that a balance is materially misstated when it is not, the application of additional audit procedures and consideration of other audit evidence would ordinarily lead the auditor to the correct conclusion.

This is misleading. An auditor does not know that his conclusion is incorrect; only that the evidence suggests that the population may be materially misstated. Frequently, this is sufficient for action, and no further audit evidence is needed, even if it were practicable to extend testing or to apply alternate procedures. More seriously, AU 350.12 invites the auditor to disregard the results of an unfavorable sample outcome and subordinate it to other, contradictory evidence whose reliability may be less than that of the sample.
Moreover, if the results of an audit sample are sufficiently precise, they may provide the basis for the proposal of an adjusting journal entry by the auditor. In such a case, the appropriate risk consideration is that the adjustment is materially correct. The calculated confidence interval provides the basis for that assessment. Estimation risk is the complement of the confidence level.

**Statistical Sampling and Audit Decisions**

The auditor uses a sample to decide whether misstatement exists and whether it may exceed the tolerable misstatement. This is the essence of the detection objective of a substantive test of details. While it is possible to design a sample to control for both the detection and estimation risk, audit samples often are designed only with the detection objective in mind. Nonetheless, if a properly selected random sample has disclosed misstatement, that sample can always be used to obtain a confidence interval on the amount of misstatement, regardless of the planning decisions and the consequent sample size.

For convenience, interval estimates may be classified into six basic categories, each of which is informative in its own way as to the extent of misstatement in the population. The possibilities are discussed below in terms of tolerable misstatement (TM), which is $600,000 in the examples, the lower confidence limit (LCL) on the estimated misstatement, and the upper confidence limit (UCL) on the estimated misstatement. The projected misstatement (that is, point estimate) is not needed, as the following examples will show. More importantly, the projected misstatement could be misleading. A projection (or point estimate) is merely one outcome in a sample space. Its principal function is to be locator for the confidence interval. It provides no information as to its margin of error. For example, 10 missstatements of $100 each will yield the same point estimate as one $1,000 misstatement, but the latter’s margin of error is greater.

**Example 1.** If neither confidence limit exceeds the tolerable misstatement and $0 is included within the confidence interval, then the auditor would decide that misstatement, if present, is no greater than tolerable misstatement. This case suggests that the amount of misstatement might also be trivial. (See the Exhibit, Figure 1.)

This is the most favorable outcome. This outcome can arise even if misstatements are detected. For example, many misstatements of very small magnitude might yield such a confidence interval. The auditor would conclude that net misstatement, if it exists, does not exceed $200,000 of understatement or $400,000 of overstatement. Because neither amount exceeds $600,000, the auditor may conclude that net misstatement is tolerable. Because $0 is within the confidence interval, it is possible that net misstatement may be $0.

Except for situations where the sample discloses no misstatement, this case does not apply when the auditor is performing tests of overstatement, such as for the existence or the lower of cost or market.

**Example 2.** If neither confidence limit exceeds the tolerable misstatement and $0 is outside the confidence interval, then the auditor would decide that the population is misstated, but the amount of misstatement is no greater than the tolerable misstatement. (See the Exhibit, Figure 2.)

This is similar to Example 1, except that the sample evidence indicates some misstatement. That is, the auditor may be confident that the population is overstated by at least $150,000, but not by more than $400,000.

**Example 3.** This case is the same as above, except that one of the confidence limits exceeds the tolerable misstatement. The auditor would conclude that the population is misstated and that the total misstatement may be greater than the tolerable misstatement, but it also may be less. The auditor cannot accept the population as being fairly stated on the sample evidence provided. (See the Exhibit, Figure 3.)

This situation arises when the disclosed misstatements exceed the auditor’s expectation. This can occur in a sample even though the actual population misstatement is as expected. In fact, if the actual population misstatement is equal to the amount expected by the auditor and used to determine sample size, then there is roughly a 50% chance that the sample’s projected misstatement will be greater than the expected misstatement. In the context of AU 350’s approach to interpretation of results, this outcome would imply that the risk of intolerable misstatement is greater than the level specified by the auditor as the risk of incorrect acceptance.

**Example 4.** This is a common outcome of audit samples. It is the outcome to be expected if the difference between the actual (but unknown) misstatement and tolerable misstatement is less than the precision of the sample estimate.

Extending the audit sample in such a circumstance often only confirms the initial finding, albeit more precisely, because the range of the confidence interval decreases as the sample size increases. In this case, an adjusting journal entry might be proposed. Whether a possible adjustment would be passed over is a question that would await the completion of the audit.

Example 4. In this case, just one of the confidence limits exceeds the tolerable misstatement, but the lower limit is negative and the upper limit is positive. The results indicate that the population may be overstated by as much as $800,000 (greater than the tolerable misstatement) or it may be understated by as much as $300,000 (less than the tolerable misstatement). The net misstatement could also be $0. Nevertheless, because one of the limits exceeds tolerable misstatement, the auditor may not conclude that the population is fairly stated. (See the Exhibit, Figure 4.)

This outcome can be the result of either the projected misstatement exceeding expectation or the variability of the misstatements in the sample being larger than planned. This situation is common to inventory valuation tests, such as price tests, where large, offsetting misstatements are disclosed. The result strongly suggests significant weakness in controls.

Example 5. In this case, the confidence limits are positive and negative and both exceed the tolerable misstatement. The interval ranges from $800,000 of understatement to $800,000 of overstatement. The misstatement may exceed the tolerable amount or it may be trivial. In this case, the sample results are too imprecise for an audit decision at the specified confidence level. (See the Exhibit, Figure 5.)

As in Example 4, of which Example 5 is a more extreme example, this result is not uncommon to tests of inventory valuation, where misstatements are more numerous than anticipated and vary greatly as to magnitude and can be both under- and overstated. While the results are not sufficiently precise for an audit adjustment (in fact, no adjustment may be needed), results such as these demonstrate that accounting controls, if they exist, are ineffective. In addition, the result questions whether sufficient evidence has been obtained.

Example 6. In this case, both confidence limits are positive (or both negative) and both exceed the tolerable misstatement, then the auditor would decide that misstatement indeed exceeds the tolerable amount. In this case, where the overstatement may range from $800,000 to $1,600,000, an adjusting journal entry would be likely. (See the Exhibit, Figure 6.)

Statistical Sampling and Audit Actions

The auditor has three courses of action when a misstatement is discovered:

- Waive the misstatement
- Do more work
- Propose an adjusting journal entry.

The question of whether the sample evidence is sufficient for an audit conclusion about the population depends upon the size of the confidence interval and the amount of tolerable misstatement. If the length of the interval (from LCL to UCL) is less than twice the tolerable misstatement, then there is some materially correct value within the interval. The auditor’s objective is not to estimate the amount of misstatement with pinpoint precision. If an adjustment is to be made, the auditor should be able to propose an amount that will reduce any remaining misstatement to an amount that is no greater than the tolerable misstatement.

Given the risk level specified by the auditor when evaluating the sample, an adjusting journal entry (AJE) can be proposed that reduces the misstatement in the population to an amount that is no greater than the tolerable misstatement. Suppose that a 90% confidence interval yields a lower limit of $800,000 and an upper limit of $1,600,000, and that the tolerable misstatement is $600,000. The range of the interval ($800,000) is less than two times the tolerable misstatement. Exhibit Figure 7 shows that a materially correct AJE can be booked within a range of values from $1 million to $1,400,000. In other words, any value within the confidence interval would be a tolerably correct AJE if both confidence limits are within the tolerable misstatement of the proposed adjustment. The risk would be no greater than the specified estimation risk.

Examination of Figure 7 should make it evident why two-sided interval estimation is important in cases where adjusting journal entries are being considered. Auditing literature has, in recent years, focused exclusively on the upper confidence limit of misstatement (that is, the confidence limit further from zero). Such a focus does not provide adequate basis for proposing sufficiently correct adjustments. By looking at only the upper limit, the auditor could inadvertently propose too large an adjustment, turning a case that was intolerably overstated into one that is intolerably understated. Only by reference to the lower confidence limit can the auditor avoid such an outcome. The Audit Guide is not clear regarding the foregoing, providing only a one-sentence approach to audit adjustments (AAG-SAM 7.36).

Does Statistical Sampling Undermine Auditor Judgment?

Many auditors continue to resist applying statistical sampling. In addition to objections to the cost of training, the cost of sample selection, and the cost of sample evaluation, some auditors have expressed concern that statistical sampling impedes...
Auditor judgment. This assertion is no truer than the assertion that laboratory biopsy is an impediment to a physician’s exercise of judgment. Auditor judgment is essential in several key respects: in deciding tolerable misstatement, in choosing the method for selecting the sample, in analyzing and assessing the population’s characteristics (such as the expected misstatement and variability of misstatement amounts), in deciding the appropriate risk level, and in deciding the method of estimation. If the auditor suspects that some population categories are more likely to contain misstatement, a sampling plan to accommodate such judgments can be devised.

Judgment is not applied in the random selection process, which is left to the operation of the laws of chance, and in the construction of the confidence interval after the sample results are available.

The ASB and the Public Company Accounting Oversight Board should provide explicit recognition of the superiority of statistical sampling in situations where the auditor has no specific knowledge as to the location and amounts of individual misstatements in an accounting population. The recently published Audit Guide, which “includes increased coverage of nonstatistical audit sampling,” is a step in the wrong direction. It is time for the profession to acknowledge that audit sampling is a decision tool that calls for the application of objective, defensible techniques, not guesswork.

Neal B. Hitzig, PhD, CPA, is professor of accounting and information systems at Queens College (CUNY). He is a member of the Auditing Standards and Procedures Committee of the NYSSCPA and a retired partner of Ernst & Young.