WHAT IT IS

Sampling is the process of selecting part of a population to determine parameters and characteristics of the whole population. The objective of sampling is to gather data on a limited number of observations (people, things, processes, documents, etc.) which represent the larger group about which more descriptive, normative, or cause-and-effect statements need to be made. Since it is rarely feasible to study an entire population (i.e. do a census), sampling must suffice. However, unless the sample represents the population, sampling accomplishes little.

Sampling may be random or purposeful. The major difference between the two is that random sampling is more confirmatory while purposeful sampling is more exploratory. Both types of sampling may be applied to:

- **Attributes** to reach a conclusion about a population in terms of the proportion, percentage, or total number of items which possess some characteristic (attribute) or fall into some defined classification. Attribute sampling is commonly used when testing internal controls or compliance with laws, policies, or procedures. Typically, the classifications defined when sampling for attributes are some variation of "in compliance" or "not in compliance." Examples of attribute sampling includes tests of compliance with voucher processing controls, tests of compliance with controls over fixed asset additions, and surveys which provide demographic information or answers to "yes/no" questions.
- **Variables** to draw conclusions about a population in terms of numbers, such as dollar amounts. It is usually used in substantive testing to determine the reasonableness of recorded amounts. Generally, variable sampling involves calculation of the difference between the actual and recorded values of a sample and projecting this difference across the population. Examples of variable sampling include tests of inventory quantities, tests of total assessed taxes uncollected, and surveys which gather ratings or interval variable data.

DEFINITIONS

Confidence level relates to the percentage of times the sample accurately represents the population. A 95 percent confidence level indicates that if 100 samples were drawn from the population, 95 would be representative (i.e. within a specified margin of error). However, no way exists to know whether any given single sample is representative or not. To increase the confidence level, one must increase the sample size. Confidence levels are usually set arbitrarily at 95 percent or 99 percent. Project cost may determine the confidence level since sample size may be reduced if the risk of an unrepresentative sample is increased.

Desired precision (d) (margin of error) relates to the amount of deviation from the estimate one is willing to accept. In an opinion poll (simple random sampling), figures quoted often have a listed "margin of error" of, for example, $\pm$ 3 percent. This means if 52 percent of respondents say they will vote for a political candidate, the actual percentage in the population may be as low as 49 percent or as high as 55 percent. In attribute sampling, the desired precision relates to the deviation from the expected percent error rate which is acceptable. For example, the expected percent error rate for transactions being tested may be three percent; however the auditor/evaluator may consider up to seven percent acceptable. The desired precision here would be four percent. Desired precision may be determined by project cost, past history, auditor/evaluator judgment, arbitrary decision, or some combination of these factors. The larger the deviation from the estimate considered acceptable, the smaller the sample needed.

Expected percent error rate (p) relates to the percentage of the population being sampled expected to contain some characteristic (such as being in error or out of compliance). This figure is typically based on past experience. However, it may also be determined by taking a preliminary random or purposeful sample. The AICPA and other sources of audit/evaluation literature often use the phrase "expected population deviation rate" to refer to the expected percent error rate.

Normal distribution occurs when data are distributed in a symmetrical bell shape such that the number of data points decreases the farther they are from the mean (average).

Purposeful sampling is the use of auditor/evaluator judgment when determining sample size. It is also known as nonstatistical, nonprobability, or judgmental sampling. Purposeful sampling can gather information which can later be studied in depth. For this reason, it can be quite useful at the beginning of an audit/evaluation since it can help determine the characteristics of a population or sample in which one is most interested. However, the more specific the purpose of sampling becomes, the harder it is to draw conclusions or make predictions about the population based on information derived from the sample. (See the appendix for more information.)

Random sampling is the use of a sample to estimate the parameters of a population to measure and control sampling risk. It is also known as statistical or probability sampling. Random sampling is the preferred technique. Since every member of the population is equally likely to be in the sample, random sampling makes it easier to achieve external validity. That is, the units sampled are more likely to represent the entire population. Note that most of the statistical rules about sampling presume the use of random samples.

Risk of overreliance is a term used in attribute sampling which relates to the confidence level. The risk of overreliance is the risk that the number of instances of non-compliance detected in a sample indicate the current level of reliance on entity internal controls is acceptable when the true population of errors warrants a lesser degree of reliance. Like the confidence level, the risk of overreliance measures the probability that the sample is not representative.

Standard deviation (σ) indicates the dispersion of data which are normally distributed. The actual standard deviation of a population is rarely known since this can only be determined by a census (100 percent sample). Typically, the sample standard deviation (s) is used in calculations once samples have been drawn.

Tolerable rate is a term used in AICPA and other audit/evaluation literature in attribute estimation tables. The tolerable rate indicates the maximum acceptable error rate in the population.

z relates to confidence level. The z-value is a standardized value which indicates how far from the population mean (or proportion) a sample mean (or proportion) can be and still represent the population. The z-value is critical when calculating sample size. The most common z-values for two-tailed tests are:

Confidence Level	z
99 %	2.58
98 %	2.33
95 %	1.96
90 %	1.64

The most common z-values for one-tailed tests are:

Confidence Level	z
99 %	2.33
98 %	2.05
95 %	1.64
90 %	1.29

WHEN TO USE IT

Use **random sampling** when generalizing the results from the sample to the population as a whole. The different types of random sampling are:

- **Simple** random sampling, which is used to select random samples from a homogenous population such that each item in the population has an equal probability of being included in the sample. For example, one could assign each Texas county a number and select numbers at random. The same could be done with social security numbers, client identification numbers, etc.
- **Stratified** random sampling, which is used to break the sample into groups (strata) on the assumption that the characteristic of interest varies more widely between groups than within groups. Groups are determined by the similarity of their members. Stratified sampling can ensure that specific groups of interest are represented in the sample even when the sample size is small. For example, instead of lumping all clients into one large group, one could stratify the sample by age, occupation, or some other criterion. Stratified samples require a bit more work but often yield more information from the same data set. Stratification also helps maintain precision with smaller samples by reducing the variance. However, groups must be mutually exclusive, and sample size in each groups must achieve the desired level of precision.
- **Cluster** (area) samples are simple random samples of groups (clusters) of observations. Since the characteristics of each population (cluster) are expected to be like those of other clusters, a given cluster should represent the population. Clusters help when it is difficult or costly to generate a list of the population, as required by simple random sampling. For example, assume one wanted to determine average annual household income in a large city. A random

sample would require a list of all households. A stratified sample would require both a list of all households and some stratification criteria. But, if one lets each city block be a cluster, one could randomly select a sample of clusters and then interview all households within only those clusters. This would be much simpler than driving all over interviewing randomly selected individuals.

Purposeful sampling is not based on statistical theory; however, it is quite commonly used and is considered appropriate in many situations. Random sampling may be too resource intensive or inefficient. Generalizing results from purposeful sampling to the entire population is often inappropriate. Purposeful sampling can be an effective time saver, however, when auditors/evaluators know from prior experience where problems are most likely to develop. It is also quite useful during scoping. (See the appendix for more information.)

HOW TO PREPARE IT

- Determine the objectives of sampling based on the project objectives.

- Define the population and the sample. Since the informational objectives are known at this point in the process, the identity and characteristics of the population should be fairly clear. The most important thing is to define the population rigorously and thoroughly since this definition will serve as the benchmark against which each potential sample is measured. In defining the population and the sample, consider these questions:
  - What is the larger group (population) about which general statements or predictions are required? What are the characteristics of this group, and which of these characteristics are of greatest interest?
  - Under which circumstances is one most likely to find information about the characteristics of interest? When, where, and why do the members of the population exhibit these characteristics?
  - Which subgroup (sample) of the population is most likely to exhibit the characteristics of interest in a way that allows gathering data in an efficient, effective, and economical way?
  - When and where is one most likely to capture the sample data sought?

  When answering this last question, take stock of information already available, such as statewide data bases, agency records, mailing or telephone directories and lists, client lists, and so on. Other issues to consider include:
  - How much will it cost to design and execute the sample?
  - Has this study been done before?
  - If data are available from multiple sources, which source is best?
  - What kinds of statements, predictions, or decisions might arise from the data?
  - How precise must the results be?
  - Are the resources required to collect and analyze the data at hand?

- Choose the sampling units (attributes or variables) that best meet the informational objectives.

- Choose the sampling strategy based on the audit/evaluation question. Examples of sampling strategies include:
  - — **Classical variable sampling** which includes simple, stratified, and cluster sampling. This strategy may be expensive depending upon the required sample size and level of access to the sampling units. However, it allows generalization of results to the population.
  - — **Discovery sampling** is a form of attribute sampling which identifies critical errors or irregularities in a population when the occurrence rate is assumed to be close to zero, such as for fraud detection. Discovery sampling is considered exploratory. The objective is to draw a sample large enough to provide some level of assurance that at least one irregularity will be detected. Depending on the number of irregularities found in the sample, additional testing may be warranted.
  - — **Interval (systematic) sampling** selects the sample by taking every nth item in the population. This method is not as statistically defensible as random sampling; however, it is often more convenient. Interval sampling should be used with caution since it increases the possibility of sample bias in a significant number of situations. Note that the starting point must be randomly chosen.
  - — **Probability-proportional-to-size (dollar-unit) sampling,** which uses attribute sampling to estimate dollar amounts. This method is commonly used when auditors/evaluators want to direct the sample toward larger balances and transactions. For example, a population of transactions to be sampled for testing contains many transactions under $300 and several over $10,000. This sampling method increases the likelihood that large transactions will be selected.
  - — **Purposeful sampling** is often the most efficient method. However, it is also the least statistically defensible. Typically, the sample will be based on auditor/evaluator judgment and experience. Purposeful sampling is often used when generalizing specific error rates and confidence levels to the population is not an overriding concern.
  - — **Stop-or-go (sequential) sampling** is a type of attribute sampling used when the objective is to minimize sample size. In stop-or-go, rather than taking one sample, multiple samples are taken with criteria for selecting successive samples determined by the results of prior samples. Subsequent samples are not as small as would occur if only one sample is taken.

- Determine the size of the sample. (See examples below.)

- Carry out sampling.

- Analyze sample results.

- Report results to the appropriate parties and add relevant analyses, conclusions, and other documents to the project working papers.

FORMULAS FOR DETERMINING SAMPLE SIZE

The ability to extrapolate sample characteristics to the whole population is a function of sample size. In other words, statistical accuracy depends on the number of items in the sample.

For **variable sampling** using simple random samples, the formula for determining the sample size is:

$$n = \frac{z^2 * \sigma^2}{d^2}$$

where: n = sample size
z = z-value associated with a desired confidence level
σ = standard deviation of the variable in the population
d = desired precision (margin of error)

Note that as the sample size (n) decreases, the margin of error (d) must increase. That is, larger samples yield lower margins of error, by design.

Note, too, that σ is an estimate since it can be known only if the whole population is sampled. How can one estimate σ then, especially if the sample has not been drawn yet? Three sources are initially available:

- previous research in the same area
- standard deviation from a prior or preliminary sample
- range of observations in a preliminary sample, i.e.

$$\frac{\text{Highest Value - Lowest Value}}{4}$$

CAUTION: Once an actual sample has been taken, the sample size should be recalculated based on the known standard deviation (s instead of σ). If post-sample sample size is substantially less (25 percent or so) than pre-sample sample size, problems may exist. It may be necessary to revise the confidence level and/or margin of error. Throwing out observations or readministering the test to a larger sample may also be needed.

Example: The Department of Human Services wants to know the age distribution of children moving to the Rio Grande Valley in a year. The estimate must be accurate with 95 percent confidence. The margin of error has been set at .25 years. The standard deviation σ has been estimated from previous work as two years.

What should the survey sample size be?

$$n = \frac{1.96^2 * 2^2}{.25^2} = 246$$

The sample size (n) can be adjusted by a correction factor when the above formula yields a sample size larger than the total population from which the sample is drawn. The correction factors are as follows:

$$n = \frac{N * (z^2 * \sigma^2)}{(N - 1) * d^2 + (z^2 * \sigma^2)}$$

or

$$n = \frac{[z^2 * \sigma^2]}{[\ d^2\ ]} * [1 - (n/N)]$$

The first is probably easier for calculating sample size, and the second is probably easier for calculating margin of error (if we solve for d^2), though a spreadsheet macro can do either quite easily.

In both cases:

N	=	Population Size
n	=	Sample Size
z	=	Value from z-table Associated with a Given confidence Level (Risk of Over reliance)
σ	=	Estimated Population Deviation Rate (Expected Percent Error Rate)
d	=	Desired Precision (Margin of Error)

For **attribute sampling** using simple random samples, the formula for determining sample size is:

$$n = \frac{z^2 * p * (1 - p)}{d^2}$$

where: n = sample size
z = z-value associated with a desired confidence level (or risk risk of Over reliance)
p = expected percent error rate
d = desired precision (margin of error)

Note that the most stringent way to use this formula is to posit that p = .5, i.e. that the population is equally divided between errors and non-errors. This makes (1 - p) = .5 as well. This approach gives you the largest value for n when looking at attributes.

Example: During an internal control review, an auditor tests checks with amounts over $5,000. Such checks should have two signatures. The desired confidence is 95 percent (i.e. risk of overreliance = 5 percent), the expected percent error rate is 7.5 percent, and the desired precision (margin of error) is 7.5 percent. How many checks should one test?

$$n = \frac{1.96^2 * (.925 * .075)}{.075^2} = 48$$

Note that the sample size obtained by using this formula may differ from that indicated in an attribute estimation table. If the sample size is in the table, use it since it is the more precise calculation of sample size. These tables normally do not go beyond an expected error rate of seven percent. In this case the table provides a sufficient approximation of sample size.

USING AN ATTRIBUTE ESTIMATION TABLE

These tables have been developed to save the auditor time and are commonly used to determine sample size. They should be available in texts on audit sampling (see books listed in the resources section). The tables are based on assumptions about the population being tested.

The first assumption is the desired confidence. The desired confidence is often referred to as the risk of overreliance on internal controls. The second assumption is the expected percent error rate for the population (p). This must be estimated from previous work or from a preliminary sample. Finally, some acceptable upper limit on the error rate (tolerable rate) must exist. This is often set through past experience or audit/evaluation convention, although the cost of the project may also be a factor. Note also that the desired precision percentage (d) used in the above formula for attribute sample size can be calculated as:

$$d = \text{tolerable rate} - p$$

Attribute estimation tables generally provide an expected number of errors or deviations associated with a given set of assumptions. If the number of actual errors detected exceeds this number, a need may exist to extend the test to additional items. Qualitative evaluation may be necessary if it appears that the level of reliance on the entity's internal controls must be reduced.

A NOTE ABOUT SURVEY RESPONSE RATES AND VALIDITY OF DATA

A survey sampling plan can be sabotaged by low response rates. Statisticians suggest a minimum 70 percent percent response to maintain internal validity, i.e. to make sure that the responses represent the opinions of everyone surveyed.

What if a 70 percent response rate is not achieved? This does not mean that all work is for nothing. It just means that one must decrease the level of information from that of a finding, conclusion, or prediction to that of a suggestion, interpretation, or observation. If one is already triangulating research, this outcome may not be disastrous since it may still be useful in confirming what was found under another research strategy.

If a response rate is low, one should also look at the non-respondents to see if patterns exist between them. For example, are many non-respondents from a particular place or in a certain age range? Such findings may help raise the response rate or suggest that one needs to revise one's findings. Optimally, the reasons for which people do not respond should be randomly distributed.