Designation: C970 – 87 (Reapproved 2006)

Standard Practice for Sampling Special Nuclear Materials in Multi-Container Lots[1]

This standard is issued under the fixed designation C970; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This practice provides an aid in designing a sampling and analysis plan for the purpose of minimizing random error in the measurement of the amount of nuclear material in a lot consisting of several containers. The problem addressed is the selection of the number of containers to be sampled, the number of samples to be taken from each sampled container, and the number of aliquot analyses to be performed on each sample.

1.2 This practice provides examples for application as well as the necessary development for understanding the statistics involved. The uniqueness of most situations does not allow presentation of step-by-step procedures for designing sampling plans. It is recommended that a statistician experienced in materials sampling be consulted when developing such plans.

1.3 The values stated in SI units are to be regarded as the standard.

1.4 *This standard does not purport to address all of the safety problems, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

E300 Practice for Sampling Industrial Chemicals

2.2 *Other Standard:*

NUREG/CR-0087, Considerations for Sampling Nuclear Materials for SNM Accounting Measurements[3]

3. Terminology Definitions

3.1 *analysis of variance*—the body of statistical theory, methods, and practice in which the variation in a set of measurements, as measured by the sum of squares of the measurements, is partitioned into several component sums of squares, each attributable to some meaningful cause (source of variation).

3.2 *confidence interval*—(*a*) an *interval estimator* used to bound the value of a population parameter and to which a measure of confidence can be associated, and (*b*) the *interval estimate*, based on a realization of a sample drawn from the population of interest, that bounds the value of a population parameter [with at least a stated confidence].

3.3 *Estimation, Estimator, Estimate*:

3.3.1 *Estimation*, in statistics, has a specific meaning, considerably different from the common interpretation of guessing, playing a hunch, or grabbing out of the air. Instead, estimation is the process of following certain statistical principles to derive an approximation (estimate) to the unknown value of a population parameter. This estimate is based on the information available in a sample drawn from the population.

3.4 *estimator*—a function of a sample (X_1, X_2, ... , X_n) used to estimate a population parameter.

NOTE 1—An estimator is a random variable; therefore, not every realization (x_1, x_2, ... , x_n) of the sample (X_1, X_2, ... , X_n) will lead to the same value (realization) of the estimator. An estimator can be a function that, when evaluated, results in a single value or results in an interval or region of values. In the former case the estimator is called a *point* estimator, and in the latter case it is referred to as an *interval* estimator.

3.5 *estimate, (a*: *n)*—a particular value or values realized by applying an estimator to a particular realization of a sample, that is, to a particular set of sample values (x_1, x_2, ... , x_n). (*b*: *v*)—to use an estimator.

3.6 *nested design*— one of a particular class of experimental designs, characterized by "nesting" of the sources of variation: for *each* sampled value of a variable *A*, a given number of values of a second variable *B* is sampled; for each of these, a given number of values of the next variable *C* is sampled, etc. The result is that each line of the "Expected Value of Mean Square" column in an analysis of variance table contains all but one of the terms of the preceding line.

3.7 *random variable*— a variable that takes on any one of the values in its range according to a [fixed] probability distribution. (Synonyms: chance variable, stochastic variable, variate.)

3.8 *standard deviation (s.d.)*—the positive square root of the variance.

[1] This practice is under the jurisdiction of ASTM Committee C26 on Nuclear Fuel Cycle and is the direct responsibility of Subcommittee C26.08 on Quality Assurance, Statistical Applications, and Reference Materials.

Current edition approved Jan. 1, 2006. Published February 2006. Originally approved in 1982. Last previous edition approved in 1997 as C970 – 87 (1997). DOI: 10.1520/C0970-87R06.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

[3] Available from National Technical Information Service, Springfield, VA 22161.

3.9 *variance*—(*a*: *population*) the expected value of the square of the difference between a random variable and its own expected value; that is, the second moment about the mean. (*b*: *sample*) The sum of squared deviations from the sample mean divided by one less than the number of values involved.

4. Significance and Use

4.1 Plans for sampling and analysis of nuclear material are designed with two purposes in mind: the first is related to material accountability and the second to material specifications.

4.2 For the accounting of special nuclear material, sampling and analysis plans should be established to determine the quantity of special nuclear material held in inventory, shipped between buyers and sellers, or discarded. Likewise, material specification requires the determination of the quantity of nuclear material present. Inevitably there is uncertainty associated with such measurements. This practice presents a tool for developing sampling plans that control the random error component of this uncertainty.

4.3 Precision and accuracy statements are highly desirable, if not required, to qualify measurement methods. This practice relates to "precision" that is generally a statement on the random error component of uncertainty.

5. Designing the Sampling Plan—Measuring Random Error

5.1 The random error component of measurement uncertainty is due to the various random errors involved in each operation such as weighing, sampling, and analysis. The quantification of the random error is usually given in terms of the variance of the mean of the measurements. When analyzing a lot of nuclear material to estimate the true concentration, p, of a constituent such as uranium, the sample mean, $\bar{p}$, is the calculated estimator. The variance of $\bar{p}$, $\sigma_{\bar{p}}^2$, is a measure of the random error associated with the measurement process. This practice deals primarily with random error; measurement process systematic error will be discussed briefly in 8.2.

5.2 To estimate the true concentration, p, in a lot consisting of N containers using a completely balanced nested design, randomly select n of the N containers; from each of the n containers, randomly select m samples; perform r laboratory analyses on each of the nm samples. (It is assumed that the amount of material withdrawn for samples is only a small fraction of the total quantity of material.) Let

$$X_{ijk} = \text{measured concentration of the constituent in the } k \text{ th analysis on the } j\text{th sample from the } i \text{ th container, or} = p + b_i + s_{ij} + a_{ijk}. \quad (1)$$

where:

p = true concentration,
b_i = effect due to container i,
s_{ij} = effect due to the j^{th} sample from container i, and
a_{ijk} = effect due to the k^{th} analysis on the j^{th} sample from container i.

Then, if each container holds the same amount of material, (Note 2), the sample mean

$$\bar{p} = \bar{X} = \frac{1}{nmr}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{r} X_{ijk} \quad (2)$$

is an estimator of the true value p. The *true* variance of $\bar{p}$ is then

$$\sigma_{\bar{p}}^2 = \frac{\sigma_b^2(N-n)}{n\ N-1} + \frac{\sigma_s^2}{nm} + \frac{\sigma_a^2}{nmr} \quad (3)$$

where:

σ_b^2 = true variance among the N containers in the given lot, defined as $N^{-1}\Sigma p_i^2 - N^{-2}(\Sigma p_i)^2$;
σ_s^2 = true variance among samples taken from a single container,
σ_a^2 = true variance of the laboratory analysis on a homogeneous sample, and
$\frac{N-n}{N-1}$ = finite population correction factor.

NOTE 2—If the ith container has g_i grams of material, then the true average concentration is $\Sigma_1^N w_i p_i$, where $w_i = g_i/\Sigma_1^N g_i$. However, the *variance* of the corresponding estimate can still be calculated as shown in this guideline; the true variance will be only slightly larger if the g_i values do not differ too much. For example, if the s.d. of the g_i were 20 % of the average g_i, it can be shown that the s.d. of p would be underestimated by about 2 % of the true standard deviation; for g_i's having s.d.'s of 10 % or 30 % of their average, the underestimation is 0.5 % or 4.5 % respectively. Note that a set of 25 weights g_i, uniformly spread from 3.3 to 6.7 kg, has a s.d. equal to 20 % of the average (5 kg). (It is assumed that errors in the estimation of net weights are insignificant compared to differences between containers, sampling variability, and analytical uncertainty, or both.)

5.3 Since the true variances σ_b^2, σ_s^2, and σ_a^2 are generally unknown, they may be estimated using appropriate data. Those data can be historical data obtained from analyzing production samples, as long as there have been no changes in the process with time. If such data are not available, as for example during the start-up of a facility or after a change in process conditions, a designed experiment is required to obtain estimates of the variances.[4]

5.4 An estimate $s_{\bar{p}}^2$ of the variance of the sample mean can be obtained from Eq 3, by inserting estimates of the variances appearing there. If a designed experiment is performed, the estimates can be obtained from the mean squares.

It is shown in Appendix X1 that estimates of the variances are as follows:

$$s_a^2 = MS_a, \quad (4)$$

$$s_s^2 = \frac{1}{r}(MS_s - MS_a), \quad (5)$$

$$s_b^2 = \frac{N-1}{Nmr}(MS_b - MS_s), \quad (6)$$

where:

MS_a, MS_b, and MS_s are the "mean squares" for analyses, containers and samples. The estimated variance of $\bar{p}$ is obtained by replacing the true variances in Eq 3 by their estimates:

[4] This topic can be found in many standard statistical texts, for example, Brownlee, K. A., *Statistical Theory and Methodology in Science and Engineering*, 2nd ed., John Wiley and Sons, New York, 1965; Bennett, C. A., and Franklin, N. L., *Statistical Analysis in Chemistry and the Chemical Industry*, John Wiley and Sons, New York, 1954; Mendenhall, William, *Introduction to Linear Models and the Design and Analysis of Experiments*, Duxbury Press, Belmont, CA, 1968; and in Jaech, J. L., "Statistical Methods in Nuclear Material Control," (TID-26298, USAEC, 1973).

$$s_{\bar{p}}^2 = \frac{1}{n}\frac{N-n}{N-1}s_b^2 + \frac{1}{nm}s_s^2 + \frac{1}{nmr}s_a^2 \quad (7)$$

Finally, expressed in terms of the mean squares, this becomes

$$s_{\bar{p}}^2 = \frac{1}{nmr}\frac{N-n}{N}MS_b + \frac{1}{Nmr}MS_s. \quad (8)$$

5.5 The variance of the sample mean, $\sigma_{\bar{p}}^2$, or its estimate, $s_{\bar{p}}^2$, is used to calculate confidence limits for the quantity and concentration of nuclear materials. Therefore, it is desirable to reduce this variance and, in this way, reduce the random error. Obviously, this can be done by using large values of *n, m,* and *r* (large number of samples and laboratory analyses). The cost and time required by that approach could be prohibitive. Another approach is to improve the overall process such that the basic variances σ_b^2, σ_s^2, σ_a^2 are reduced.

5.6 Eq 8 gives an estimate of the variance$\sigma_{\bar{p}}^2$ for any given *n, m,* and *r* and therefore can be used for comparing different sampling plans. An example of two sampling plans involving the same number of analyses but having different random errors is given in Appendix X3.

5.7 When one has fixed resources within which the sampling plan must function, the question arises as how to allocate these resources to obtain the "best" sampling plan. Sections 6 and 7 discuss this problem when "cost" is considered. "Cost" is used generically here—it need not be a monetary quantity; it could be time or something else.

6. Determining Sample Sizes

6.1 There are two common situations in which sampling plans must be developed for use in nuclear material measurement when there are constraints on resources. In the first situation a constraint is imposed upon the "cost" of sampling and analysis. In this case, the problem is to find a plan that minimizes the variance of the sample mean (minimizes random error) subject to the cost constraint. In the second situation, a constraint is imposed upon the variance of the sample mean (upon the random error) and the problem is to find a plan which minimizes cost subject to this constraint. Since this latter problem is the most frequently encountered, methods for its solution will be given. The former problem, for which the solution technique closely parallels the one given, will be covered in footnotes.

6.2 *Component Variances Are Known*:

6.2.1 If the variance constraint is expressed as a maximum value for the width, 2Δ, of a confidence interval for *p*, it can be transformed immediately to a maximum value for $\sigma_{\bar{p}}$, by using the relationship

$$\Delta = (Z_{1-\alpha/2})\sigma_{\bar{p}} \quad (9)$$

where:

$Z_{1-\alpha/2}$ = value having a probability α/2 of being exceeded by a standard normal variate.

Therefore, if Δ is limited to Δ_o, say, then $\sigma_{\bar{p}}$ is limited to $\Delta_o / Z_{1-\alpha/2}$. Since the minimum cost is achieved when the constraint is barely satisfied, we need to minimize cost subject to the constraint

$$\sigma_{\bar{p}}^2 = K \quad (10)$$

where *K* is a constant, either specified directly or computed from Δ_o and α.

6.2.2 When the underlying variances are known from previous history, the problem of achieving a minimum cost within a stated confidence interval width reduces to finding a suitable set of values for *n, m,* and *r*. In Appendix X2 it is shown that the optimum *r* and *m* are given by

$$r = \frac{\sigma_a}{\sigma_s}\left(\frac{c_s}{c_a}\right)^{1/2} \quad (11)$$

$$m = \frac{\sigma_s}{\sigma_b}\left(\frac{c_b}{c_s}\frac{N-1}{N}\right)^{1/2} \quad (12)$$

where:

c_b = marginal cost of choosing one additional container and preparing it for sampling,

c_s = marginal cost of drawing an additional sample from a container and preparing it for analysis, and

c_a = marginal cost of an additional laboratory analysis.

Therefore, the optimum values for *r* and *m* do not depend on *n*, and in fact can be calculated immediately from the variances, the "costs," and *N*.

6.2.3 Once *m* and *r* are determined and inserted into Eq 3, $\sigma_{\bar{p}}^2$ is seen to be a monotonic decreasing function of *n*, so that one need only make *n* large enough to achieve the required bound on $\sigma_{\bar{p}}^2$(Note 3). Letting $c_s = c_a = c_b = 1.0$ provides the optimum values of *r, m,* and *n* when costs are considered equal. In practice, the optimum values for *m* and *r* obtained this way are unlikely to be integers. Unless these values are very close to integers, it is prudent to consider both bracketing values, that is, if the optimum value for *r* is 1.4, try both *r = 1* and *r = 2*. The reason is that the final value of *n* will generally be different and it is not clear beforehand which set of values of *r, m,* and *n* will achieve the required variance at minimum cost. It is also possible to use different values of *m* (or *r*, or both) for different containers or samples, or both, to obtain a non-integer "effective" value of *m* (or *r*, or both). In this case, $\bar{p}$ should be replaced by a weighted average; $\sigma_{\bar{p}}^2$ becomes more complicated; and the expected values of the mean squares also become more complicated, as does the estimate of $\sigma_{\bar{p}}^2$. The advice of a statistician is strongly suggested if this approach is being considered.

NOTE 3—The *same* values of *m* and *r* provide minimum variance for given cost. When these are inserted into the cost function, it is seen to be proportional to *n*, so that *n* should be chosen as large as the cost constraint will allow.

6.2.4 An example with further discussion is given in Appendix X3.

6.3 *Component Variances Are Not Known*:

6.3.1 The approach to finding values for *n, m,* and *r* described in Appendix X2 is also valid when the basic variances are not known, provided some estimates of these variances are available. As in 6.2, values for *m* and *r* can be obtained from estimates of the variances and cost factors. There is a complication in the calculation of an optimum value of *n*, however: since the final uncertainty will be based not on