Designation: C 970 – 87 (Reapproved 1997)

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

This standard is issued under the fixed designation C 970; 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:*

E 300 Practice for Sampling Industrial Chemicals[2]

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.

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

[1] This practice is under the jurisdiction of ASTM Committee C-26 on Nuclear Fuel Cycle and is the direct responsibility of Subcommittee C26.06 on Statistical Applications.

Current edition approved Jan. 30, 1987. Published March 1987. Originally published as C 970 – 82. Last previous edition C 970 – 82.

[2] *Annual Book of ASTM Standards*, Vol 15.05.

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

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} = measured concentration of the constituent in the k th analysis on the jth sample from the i 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:

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

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

[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).

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

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 the true variances but rather on estimates, the *t*-distribution[4] must be used instead of the Normal. Given the allowable half-width Δ, we have

$$\Delta = s_{\bar{p}} t_{1-\alpha/2}(\nu) \quad (13)$$

where:

$t_{1-\alpha/2}(\nu)$ = value having probability α/2 of being exceeded by a "Student's *t*" variable with degrees of freedom *ν*, and *s*

$s_{\bar{p}}$ = estimated standard deviation of the mean.

Unfortunately, ν depends upon n, m, and r (and if prior data are to be combined with present data in computing $s_{\bar{p}}$ it depends also upon the degrees of freedom appropriate to those data). We therefore proceed iteratively. We guess n, calculate ν (as described in 6.3.2), obtain t from standard tables, and calculate $s_{\bar{p}}$ from Eq 13. We then use this target value and our estimates of the basic variances to obtain an optimum value for n as in 6.2.3. If this optimum value is as large as, but not too much larger than, the guessed value, it should be used. Otherwise, use it in place of the initial guess and repeat the procedure.

6.3.2 The uncertainty in the final $\bar{p}$ will be expressed in terms of an estimated variance $s_{\bar{p}}^2$. The t-factor used with $s_{\bar{p}}$ in Eq 13 has been shown to be approximately correct, provided the degrees of freedom parameter, ν, is properly chosen. Satterthwaite's formula is applicable, whether or not the data from a prior experiment are to be used. In the simple case where only the $n \times m \times r$ data values under consideration are used, the formula[5] is

$$\nu = s_{\bar{p}}^4 \left[\left(\frac{N-n}{nmrN} \right)^2 \frac{MS_b{}^2}{n-1} + \left(\frac{1}{mrN} \right)^2 \frac{MS_s{}^2}{n(m-1)} \right]^{-1} \quad (14)$$

When prior data are combined with these data, the formula is more complicated.

6.3.3 When n and m are both greater than one, the approach given here leads to an unbiased estimate of $\sigma_{\bar{p}}^2$. If n or m, or both, are chosen to be one, then the corresponding mean square(s) (Appendix X1) are undefined. If $n = 1$, no estimate of $\sigma_{\bar{p}}^2$ is available. If $n > 1$ and $m = 1$, then only an overestimate of $\sigma_{\bar{p}}{}^2$ is available: $(1/nmr)\ MS_b$ has expected value $(\sigma_b{}^2/n)\ (N/(N-1)) + (\sigma_s{}^2/nm) + (\sigma_a{}^2/nmr)$, in which the first term is too big by the factor $(N/(N-1))$. Therefore, in order to avoid this problem, it is desirable to choose n greater than one; and unless N is large, also choose m greater than one.

7. Compositing Samples

7.1 In the example of Appendix X3 at least seven samples and seven laboratory analyses (measurements) were needed to reduce the variance of the sample mean to the specified value. Laboratory measurements are usually costly and time consuming. Sampling operations, on the other hand, are relatively inexpensive from the viewpoint of required instrumentation and operator time. Furthermore, in many SNM accountability situations the variance components due to between- and within-container variabilities are not known with the same degree of confidence as the laboratory variance. To reduce the effort in the laboratory and to minimize the random error, it could be desirable to blend samples to form a composite.

7.2 When each container in a lot $(n = N)$ is sampled m times with r analyses per sample, the finite population correction factor in Eq 3 becomes zero and Eq 3 becomes:

$$\sigma_{\bar{p}}^2 = \frac{1}{N} \left(\frac{\sigma_s^2}{m} + \frac{\sigma_a^2}{mr} \right) = \frac{1}{Nm} \left(\sigma_s^2 + \frac{\sigma_a^2}{r} \right) \quad (15)$$

If the m samples from each individual container are composited and thoroughly mixed (Note 4) and each of the N composites is analyzed r times, Eq 15 is replaced by:

$$\sigma_{\bar{p}}^2 = \frac{1}{N} \left(\frac{\sigma_s^2}{m} + \frac{\sigma_a^2}{r} \right) \quad (16)$$

The laboratory effort is still rather large, since even for $r = 1$ a total of $Nr = N$ measurements must be made.

NOTE 4—Thorough mixing is very important to give effective homogenizing of the composite samples, thereby reducing the error from subsampling to a negligibly small value.

7.3 To further reduce the laboratory effort, the m samples from each of the N containers in the lot may be composited into a lot master sample and thoroughly mixed. The contributions to the master sample from each of the N containers should be proportional to the net weights in the corresponding containers. A sub-sample (Note 5) of the composite is then analyzed r times. The variance of the sample mean is given by

$$\sigma_{\bar{p}}{}^2 = \frac{\sigma_s^2}{Nm} + \frac{\sigma_a{}^2}{r} \quad (17)$$

NOTE 5—Dissolution of the material is a step in the laboratory analysis; therefore, the sub-sample must contain an amount of material sufficient for further subdivision into r portions.

7.4 For this latter case, it is shown in Appendix X4 that the values of m and r that minimize "cost" for a given variance bound k are

$$m = \frac{\sigma_s}{N} \left(\frac{\sqrt{c_a}\,\sigma_a + \sqrt{c_s}\,\sigma_s}{k\sqrt{c_s}} \right), \text{ and} \quad (18)$$

$$r = \sigma_a \left(\frac{\sqrt{c_a}\,\sigma_a + \sqrt{c_s}\,\sigma_s}{k\sqrt{c_a}} \right) \quad (19)$$

Finally, the minimum cost is given by

$$\text{minimum cost} = \frac{1}{k} \left(\sqrt{c_a}\,\sigma_a + \sqrt{c_s}\,\sigma_s \right)^2 \quad (20)$$

(Note 6)

Note that, while Eq 18 and Eq 19 give m and r, the values will not generally be integers. If the values are rounded to integers, then Eq 20 is not appropriate for calculating the actual cost corresponding to the chosen m and r. Instead, the cost would be calculated as $c_s Nm + c_a r$.

NOTE 6—If it is desired to minimize the variance for given cost C, the same technique leads to

$$\frac{mN\sqrt{c_s}}{\sigma_s} = \frac{r\sqrt{c_a}}{\sigma_a} = \frac{C}{\sqrt{c_s}\,\sigma_s + \sqrt{c_a}\,\sigma_a}, \text{ and} \quad (21)$$

the minimum variance is given by

$$\text{minimum variance} = \frac{1}{C} \left(\sqrt{c_a}\,\sigma_a + \sqrt{c_s}\,\sigma_s \right)^2 \quad (22)$$

7.5 An example with further discussion is given in Appendix X5.

7.6 Compositing in this way has a major drawback, in that it is impossible to estimate $\sigma_s{}^2$, the within-container variance, on a continuing basis. Quite possibly $\sigma_s{}^2 2$ may change, especially if there has been a change in process conditions or supplier. Periodically, and especially at those times when a change in $\sigma_s{}^2$ might be expected, a number of samples may be drawn from each container and analyzed separately in replicate to establish current estimates of $\sigma_s{}^2$ and $\sigma_a{}^2$.

[5] Mendenhall (*op cit*), p. 352; also Jaech (*op cit*), pp. 157–161.

8. Mechanical and Physical Aspects of Sampling

8.1 The common types of nuclear material encountered are liquid and solids (powder and pellets). In whatever form encountered, the principal task in sampling is to remove a sample that is typical of the bulk material, at least as far as the parameters of interest are concerned. The selection of a procedure and equipment for sampling must be made based on factors such as the following:

8.1.1 Type and form of the material,

8.1.2 Degree of homogeneity,

8.1.3 Stability of the material,

8.1.4 Location of the material,

8.1.5 Purpose and requirements for analyzing the material, and

8.1.6 Accessibility for sampling all units or containers involved.

8.1.7 Practice E 300 and NUREG/CR-0087 present principles and guidelines for sampling materials. The mechanical and physical aspects of sampling are discussed. NUREG/CR-0087 addresses sampling nuclear materials to determine their chemical and isotopic contents.

8.2 *Some Sources of Error*:

8.2.1 There are various sources of error in the sampling process such as nonhomogeneity, contamination of the sample after removal from the bulk material, failure of the equipment, failure of the operator to follow the procedure, bias, and chemical and physical changes in the material during sampling. The latter two sources of error are discussed briefly in 8.2.2 and 8.2.3.

8.2.2 Bias occurs when, in addition to the random errors, all measured values are shifted consistently from the true value in the same direction. Likely sources of bias are improper sampling procedures, faulty sampling devices, and improperly calibrated instruments. The problem is to detect the existence of such biases and to account for them in the results. The solution to this problem usually requires designing an appropriate study.[6] This may not always be possible. For example, if the sample were contaminated to an unknown degree and new samples are not available, it may be impossible to estimate the bias.

8.2.3 An example of errors due to chemical or physical changes occurs with plutonium dioxide powder.[7] This material, which is usually handled as a fine powder with a large surface area, readily picks up or loses water if exposed to a change in humidity. Plutonium dioxide powder can gain over 1 % in weight within a few hours if exposed to an increase in humidity. Therefore, very careful control over conditions must be established and maintained when sampling this material, particularly if it is relocated and then sampled.

[6] Stephens, F. B., et al, *Methods for the Accountability of Uranium Dioxide*, NUREG-75/010, pp. 1–17, U. S. Nuclear Regulatory Commission, National Technical Information Service, Springfield, VA, 1975.

[7] Gutmacher, R. G., et al, *Methods for the Accountability of Plutonium Dioxide*, USAEC Report WASH-1335, 1974.

APPENDIXES

(Nonmandatory Information)

X1. ESTIMATION OF VARIANCES IN A NESTED ANALYSIS OF VARIANCE DESIGN

X1.1 Let X_{ijk} be the kth measurement on the jth sample from the ith container, $k = 1, \ldots, r; j = 1, \ldots, m; i = 1, \ldots, n$. Let

$$X_{ij.} = \sum_{k=1}^{r} X_{ijk}, \bar{X}_{ij.} = \frac{1}{r} X_{ij.}, \tag{X1.1}$$

$$X_{i..} = \sum_{j=1}^{m}\sum_{k=1}^{r} X_{ijk}, \overline{X_{i..}} = \frac{1}{mr} X_{i..}, \text{ and} \tag{X1.2}$$

$$X_{...} = \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{r} X_{ijk}, \bar{X}_{...} = \frac{1}{nmr} X_{...}. \tag{X1.3}$$

X1.1.1 Then the mean squares may appear in a nested analysis of variance (ANOVA) table as follows:

Source	Mean Square	Expected Value of Mean Square
Containers	$MS_b = [mr/(n-1)] \cdot (\Sigma^{n}_{i=1}.(\bar{X}_{i..} - \bar{X}_{...})^2$	$[mrN/(N-1)]\sigma_b + r\sigma_s^2 + \sigma_a^2$
Samples	$MS_s = [r/n(m-1)] \cdot \Sigma^{n}_{i=1}\Sigma^{m}_{j=1}(\bar{X}_{ij.} - \bar{X}_{i..})^2$	$r\sigma_s^2 + \sigma_a^2$
Analyses	$MS_a = [1/nm(r-1)] \cdot \Sigma^{n}_{i=1}\Sigma^{m}_{j=1}\Sigma^{r}_{k=1}(X_{ijk} - \bar{X}_{ij.})^2$	σ_a^2

Note that factor $[(N/(N-1)]$ is due to the finite number of containers. From the preceding table, it is seen that estimates of the variances are as follows:

$$s_a^2 = MS_a, \tag{X1.4}$$

$$s_s^2 = \frac{1}{r}(MS_s - MS_a), \text{ and} \tag{X1.5}$$

$$s_b^2 = \frac{N-1}{Nmr}(MS_b - MS_s). \tag{X1.6}$$

X1.1.2 In practice the latter two estimates could be negative which would require modification of this estimation procedure.[4]

X2. FINDING THE OPTIMAL VALUES OF r AND m FOR MINIMIZING "COST" SUBJECT TO THE CONSTRAINT THAT $\sigma_{\bar{p}}^2 = K$ (see 6.2)

X2.1 The total cost[8] of sampling and analysis is not linear (in *n, m,* and *r*) over the whole range of these variables. However, in the neighborhood of the optimum, a linear approximation is likely to be reasonable. Write the variable part of the cost as

$$c = c_b n + c_s m' + c_a r' \tag{X2.1}$$

where:

m' = mn,
r' = nmr and, as in 6.2.2,
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.

Then applying the Lagrange multiplier technique,[9] we consider the expression

$$L = C + \lambda(\sigma_{\bar{p}}^2 - K), \tag{X2.2}$$

where (see Eq 3, S 5.2):

$$\sigma_{\bar{p}}^2 = \frac{\sigma_b^2}{n} \cdot \frac{N-n}{N-1} + \frac{\sigma_s^2}{m'} + \frac{\sigma_a^2}{r'} \tag{X2.3}$$

Taking partial derivatives with respect to r', m', n, and λ, setting them equal to zero, and solving for λ gives

$$\lambda = \frac{c_a}{\sigma_a^2 r'^2} = \frac{c_s}{\sigma_s^2 m'^2} = \frac{c_b N - 1}{\sigma_b^2 \quad N n^2} \tag{X2.4}$$

From this it follows that the optimum r and m are given by

$$r = \frac{r'}{m'} = \frac{\sigma_a}{\sigma_s}\left(\frac{c_s}{c_a}\right)^{1/2}, \text{ and} \tag{X2.5}$$

$$m = \frac{m'}{n} = \frac{\sigma_s}{\sigma_b}\left(\frac{c_b N - 1}{c_s \quad N}\right)^{1/2} \tag{X2.6}$$

[8] Cost need not be monetary.

[9] Mendenhall, *op cit*, p. 355.

X3. EXAMPLE

X3.1 Find values of *n, m,* and *r* that meet a variance constraint and minimize "cost" (6.2.4). Let $N = 20$; $\sigma_b = 0.3$; $\sigma_s = 0.1$; and $\sigma_a = 0.04$. For $\alpha = 0.05$, $Z_{1-\alpha/2} = 1.96$ and if $\Delta = 0.2$, the target value for $\sigma_{\bar{p}}^2$ is given by:

$$K = \Delta^2 / Z_{1-\alpha/2}^2 = \frac{0.2^2}{1.96^2} = 0.0104 \tag{X3.1}$$

From Eq 3 in X2, $r = \dfrac{0.04}{0.1}\dfrac{c_s^{1/2}}{c_a}$,

which is close to 1 whenever $5 < c_s/c_a < 10$. Since the cost of sampling is unlikely to be more than ten times the cost of an analysis and since $r \geq 1$, r will usually be taken equal to one. Likewise, $m = \dfrac{0.1}{0.3}\left(c_b/c_s \times 19/20\right)1/2$, which is close to 1 whenever $8 < c_b/c_s < 15$. Since $m \geq 1$, m will usually be taken to be one. With $r = m = 1$,

$$\sigma_{\bar{p}}^2 = \frac{0.09}{n}\frac{20-n}{19} + \frac{0.01}{n} + \frac{0.0016}{n} \tag{X3.2}$$

$$= \frac{0.10633}{n} - 0.00473; \tag{X3.3}$$

setting $\sigma_{\bar{p}}^2 = 0.0104$, we obtain $n = 7.03$.

Thus the required variance is (approximately) achieved by taking $n = 7$, $m = r = 1$.

X3.2 The previous paragraph shows that at least 7 containers out of the 20 must be selected, sampled once, and each sample analyzed once to obtain the target value of 0.0104 for $\sigma_{\bar{p}}^2$ and therefore meet the specified confidence interval width (2Δ) of 0.4. The seven containers must be selected at random, that is, each of the 20 containers is assigned one of a sequence of numbers and a random number table is used to select the seven containers.

X3.3 Note that in this case, the variance term involving *only* n, namely $\sigma_b^2/n \cdot N - n/N - 1$, is greater by itself than the variance bound of 0.0104, unless $n \geq 7$; and the other terms contribute so little that for $n = 7$, the total variance is down to the required value even for $m = r = 1$. Therefore, (*a*) we need $n \geq 7$ and (*b*) once $n = 7$, m and r do not need to be any larger than their minimum value, so these optimum values are really independent of the costs. This will not always be so, of course.

X3.4 Significant improvements in the variance $\sigma_{\bar{p}}^2$ can sometimes be achieved with small additional cost by judiciously choosing values for n, m,and r. This is apparent by comparing the sets: $n = 7$, $m = 2$, $r = 1$ and $n = 14$, $m = 1$, $r = 1$. In both situations 14 sampling operations and 14 analyses are required, that is, the total effort is about equal. However, for the second set ($n = 14$), the variance is 0.0029, which is lower by a factor of three than for the first set ($n = 7$), which has a variance of 0.0096. Therefore, the set with $n = 14$ is preferred, unless the cost of choosing additional containers is quite large.

X4. FINDING THE OPTIMAL VALUES FOR r AND m—COMPOSITE SAMPLE CASE (7.4)

X4.1 Selection of m and r to minimize cost for a given variance bound K is achieved by the Lagrange multiplier technique that was used in Appendix X2. The function to be considered is

$$L = c_s Nm + c_a r + \lambda \left(\frac{\sigma_s^2}{Nm} + \frac{\sigma_a^2}{r} - K \right) \quad \text{(X4.1)}$$

Setting the partial derivatives with respect to m and r equal to zero, and solving for λ yields

$$\lambda = \frac{c_s N^2 m^2}{\sigma_s^2} = \frac{c_a r^2}{\sigma_a^2}$$

Then, using $\sigma_{s^2}/Nm + \sigma_{a^2}/r = K$, the values for m and r can be given in symmetric form as

$$\frac{mN\sqrt{c_s}}{\sigma_s} = \frac{r\sqrt{c_a}}{\sigma_a} = \frac{\sqrt{c_a}\,\sigma_a + \sqrt{c_s}\sigma_s}{K} \quad \text{(X4.2)}$$

X5. EXAMPLE

X5.1 Find values of m and r that meet a variance constraint and minimize "cost"—composite sample case (7.5).

X5.1.1 A lot consists of $N = 20$ containers.
Let

$\sigma_s^2 = 0.01$
$\sigma_a^2 = 0.0025$
$c_s = 1$ unit
$c_a = 16$ units

X5.1.1.1 Suppose the variance of $\bar{p}$ is not to exceed $k = 0.001$. Then, from X4,

$$\frac{mN}{0.1} = \frac{4r}{0.05} = \frac{0.2 + 0.1}{0.001} = 300, \quad \text{(X5.1)}$$

so that

$$m = \frac{300 \times 0.1}{20} = 1.5 \quad \text{(X5.2)}$$

$$r = \frac{300 \times 0.05}{4} = 3.75. \quad \text{(X5.3)}$$

X5.1.1.2 The minimum cost is found to be 90 units. If m and r are rounded up to 2 and 4, the actual cost is, of course, $1 \times 20 \times 2 + 16 \cdot 4 = 104$ units, and the actual variance is 0.000875. If instead of compositing, one sample were taken from each container and analyzed, the cost would be 340 units, and the variance would be 0.000625.