NOTICE: This standard has either been superceded and replaced by a new version or discontinued.
Contact ASTM International (www.astm.org) for the latest information.

Designation: D 6842 – 02

Standard Guide for Designing Cost-Effective Sampling and Measurement Plans by Use of Estimated Uncertainty and Its Components in Waste Management Decision-Making[1]

This standard is issued under the fixed designation D 6842; 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 Waste management decisions generally involve uncertainty because of the fact that decisions are based on the use of sample data. When uncertainty can be reduced or controlled, a better decision can be achieved. One way to reduce or control uncertainty is through the estimation and control of the components contributing to the overall uncertainty (or variance). Control of the sizes of these variance components is an optimization process. The optimizations results can be used to either improve an existing sampling and analysis plan (if it should be found to be inadequate for decision-making purposes) or to optimize a new plan by directing resources to where the overall variance can be reduced the most.

1.2 Estimation of the variance components from the total variance starts with the sampling and measurement process. The process involves two different kinds of uncertainties: random and systematic. The former is associated with imprecision of the data, while the latter is associated with bias of the data. This guide will discuss only sources of uncertainty of a random nature.

1.3 There may be many sources of uncertainty in waste management decisions. However, this guide does not intend to address the issue of how these sources are identified. It is the responsibility of the stakeholders and their technical staff to analyze the sampling and measurement processes in order to identify the potentially significant sources of uncertainty. After identifying these sources, this guide will provide guidance on how to collect and analyze data to obtain an estimate of the total uncertainty and its components.

2. Terminology

2.1 *analysis of variance (ANOVA)*, *n*—a statistical method of decomposing (or breaking down) the total variance and estimating or testing its contributing component variances for statistical significance.

2.2 *balanced design*, *n*—a statistical study where replication in each of the levels of ANOVA is identical.

2.3 *measurement process*, *n*—the method and procedure of obtaining and measuring samples or their subsamples to produce sample data.

2.4 *sampling process*, *n*—the method and procedure of collecting physical samples from a defined population.

2.5 *unbalanced design*, *n*—a statistical study where replication in some or all of the levels of ANOVA is not identical.

3. Significance and Use

3.1 This guide will evaluate sample data that contain a high level of uncertainty for decision-making purposes and, when it is feasible, design a statistical study to estimate and reduce the sources of uncertainty. Oftentimes, historical data may be available and adequate for this purpose and no new study is needed.

3.1.1 This approach will help the stakeholders better understand where the greatest sources of uncertainty are in the sampling and analysis process. Resources can be directed to where it can most reduce the overall uncertainty.

3.1.2 Sampling and analysis design under this approach can often be cost-efficient because (*a*) the reduction in uncertainty can be done by statistical means alone and (*b*) the reduction can be translated into a lower number of analyses.

3.2 This guide is limited to the situation where a decision is based on the mean of a population. It will only include discussions of a balanced design for the collection and analysis of sample data in order to estimate the sources of uncertainty. References to unbalanced designs are provided where appropriate.

4. Uncertainty in Decision-Making

4.1 *Decision-Making Based on Data*:

4.1.1 When waste management decision-making is based on data and when the data come from a subset of a population, the data can be used to calculate quantities such as mean, median, or percentage for the purpose of estimating the true value of these quantities in the population. These estimates can be used to make conclusions or decisions about the population on

[1] This guide is under the jurisdiction of ASTM Committee D34 on Waste Management and is the direct responsibility of Subcommittee D34.01 on Sampling and Monitoring.

Current edition approved Dec. 10, 2002. Published February 2003.

Copyright © ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959, United States.

NOTICE: This standard has either been superceded and replaced by a new version or discontinued. Contact ASTM International (www.astm.org) for the latest information.

issues such as: (*1*) Is the average concentration of a contaminant at a certain site higher or lower than a regulatory standard? (*2*) Has the cleanup standard been met?

4.1.2 However, these estimates involve uncertainty because of uncertainties in the sampling and measurement processes. The total uncertainty associated with an estimate can be derived from the sample data and it is usually expressed as the variance or standard deviation of the estimate. The estimate and its variance can be used to define the level of confidence in decision-making. For example, they can be used to calculate the upper and lower confidence limits, where the width of the confidence limits is a measure of uncertainty in decision-making.

4.1.3 An example of high data uncertainty and low confidence in decision-making can occur when the sample mean concentration of a site is substantially below a regulatory limit while its upper confidence limit is higher than the regulatory limit. In this case, a reduction in uncertainty will lead to better decision-making. That is, there is a higher probability that the correct decision about the true concentration can be reached and the appropriate action taken.

4.2 *Sampling and Measurement Process*:

4.2.1 When the confidence level is not at the level desired by the decision-makers, the data from the sampling and measurement processes can be analyzed to identify significant sources of contributors to the total variance. This guide will permit project managers to focus on the large sources of uncertainty and allocate resources for their reduction. That, in turn, will improve the sampling and measurement processes and achieve a higher level of confidence in project decisions.

4.2.2 This guide is limited to the situation when a decision needs to be made regarding the mean of a population.

4.2.3 This guide is also limited to the discussions of a balanced design for the collection and analysis of sample data in order to estimate the sources of uncertainty. An example of a balanced design is given in Table 1. In Table 1, the letter "*m*" indicates the number of subsamples taken from a field sample and the letter "*k*" indicates the number of replicate analyses performed on each subsample. Note that there is an equal number of subsamples for each of the field samples and an equal number of replicate analyses for each of the subsamples in Table 1. It is this equality in replication at the subsampling level and at the replicate analysis level that constitutes a balanced design. When there is inequality at any of the levels, it is called an unbalanced design. References to unbalanced designs will be provided where appropriate.

TABLE 1 Study Design for the Example Sampling and Measurement Process Described in Section 5[A]

Field Sample No.	Subsample No.	Replicate No.	Value
1	1	1	X_{111}
		n	X_{11n}
	m	1	X_{1m1}
		n	X_{1mn}
f	1	1	X_{f11}
		n	X_{f1n}
	m	1	X_{fm1}
		n	X_{fmn}

[A] $f, m, n \geq 2$.

TABLE 2 Example Data of TPH (ppm) for a 3-Stage Sampling and Measurement Process

Field Sample	Sub-sample	TPH in Replicate			Subsample Total	Field Sample Total	Grand Total
		1	2	3			
1	1	10	11	11	32		
	2	8	7	8	23	55	
2	1	5	6	5	16		
	2	4	4	6	14	30	85

4.2.4 A typical sampling and measurement process goes through three stages:

4.2.4.1 The collection of field samples,

4.2.4.2 Taking of subsamples from the field samples in the laboratory, and

4.2.4.3 Duplicate analysis of the subsamples.

4.2.5 The variances associated with each of these stages are known as the sampling variance, subsampling variance, and analytical variance, respectively. The sum of these variances constitutes the total variance in decision-making. The total variance and its contributing components can be estimated from the data when the sampling and measurement process is designed for such purposes. For this guide, the 3-stage sampling and measurement process above will be used as a model for discussion purposes. When other processes are appropriate, consult a statistician.

5. Estimation of Total Variance and Its Components

5.1 *Study Design and Example Data*:

5.1.1 Under any sampling and measurement process, the total variance and its components can be estimated only when the data are collected according to a design. In particular, for the 3-stage process described in 4.2, the variances can be estimated only when there are multiple field samples, where multiple subsamples are taken from each of the field samples and when each of the subsamples is in turn analyzed in multiple replicates (duplicate, triplicate, etc.). The word "multiple" here implies two or more, with two being the minimum requirement. The optimal numbers of field samples, subsamples and replicates will depend on the sizes of their respective variance components and the costs associated with the collection or analysis regarding these components. When the costs are negligible, then they will depend solely on the relative sizes of the variance components alone.

5.1.2 An example of such a study design may appear as noted in Table 1. Example data of TPH concentrations collected from a hypothetical site may appear as shown in Table 2, with the addition of the last 3 columns for the statistical method Analysis of Variance (ANOVA). Note that the data in Table 2 is a balanced design in that the number of subsamples per field sample is equal at 2 and the number of replicate analyses per subsample is equal at 3.

5.1.3 An unbalanced design occurs when the number of subsamples is not equal among the field samples or when the number of replicates is not equal among the subsamples. In this case, the estimation of the variance components becomes more complicated. In this situation, consult a statistician. Some

NOTICE: This standard has either been superceded and replaced by a new version or discontinued. Contact ASTM International (www.astm.org) for the latest information.

statistical software programs such as Statgraphics Plus (1993)[2] allow for the estimation of variance components when the design is unbalanced. Because the use of different algorithms in the estimation procedure may produce different results, these programs need to be used with care.

5.2 *Estimation of Total Uncertainty and Its Components*:

5.2.1 This section will discuss data uncertainty using the example data in Table 2. The data in Table 2 represent a two-way random effects model, the two random effect variables being the "field samples" and "subsamples." It is also called a nested design in that the replicates are "nested" within each subsample and the subsamples are "nested" within a field sample. This method of analysis can be found in most statistical textbooks (for example, Snedecor and Cochran, 1967).[3] In order to carry out this analysis, let:

X_{ijk} = TPH value for the *k*th replicate of the *j*th subsample from the *i*th field sample, where $i = 1, \ldots, f$, $j = 1, \ldots, m$, $k = 1, \ldots, n$
X_{ij} = sum of replicate TPH values for subsample *j* from field sample *i*
X_i = sum of all TPH values for field sample *i*
X = grand total
f = number of field samples (= 2 in the example)
m = number of subsamples per field sample (= 2 in the example)
n = number of replicate analyses per subsample (= 3 in the example)

5.2.2 Calculate:

$C = (X\ldots)^2/(fmn) = (85)^2/[(2)(2)(3)] = 602.08$

SS(total) = total sum of squares
$= \Sigma X_{ijk}^2 - C$
$= 10^2 + 11^2 + 11^2 + 8^2 + \ldots\ldots + 4^2 + 6^2$
$= 673.00 - 602.08$
$= 70.92$

SS(subsamples) = sum of squares due to subsamples
$= \Sigma X_{ij.}^2/n - C$
$= (32^2 + 23^2 + 16^2 + 14^2) / 3 - 602.08$
$= 66.25$

SS(field samples) $= \Sigma X_{i..}^2/(mn) - C$
$= (55^2 + 30^2) / (2\times3) - 602.08$
$= 52.08$

SS(subsamples in field samples) = SS(subsamples) − SS(field samples)
$= 66.25 - 52.08$
$= 14.17$

SS(replicates) = sum of squares due to replicates
= SS(total) − SS(field samples) − SS(subsamples in field samples)
$= 70.92 - 52.08 - 14.17$

5.2.3 An ANOVA table can be constructed using the above quantities:

5.2.4 Note that the "expected mean squares" in Table 3 is a function of the variance components in the sampling and measurement process, where σ_k^2 = variance component due to replicate analyses, σ_j^2 = variance component due to subsampling within a field sample, and σ_k^2 = variance component due to field sampling.

TABLE 3 ANOVA Table for TPH (Nested Design)[A]

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares (MS)	Expected MS
Field samples	$f - 1 = 1$	52.08	52.08	$\sigma_k^2 + n\sigma_j^2 + mn\sigma_i^2$
Subsamples in field samples	$f(m - 1) = 2$	14.17	7.08	$\sigma_k^2 + n\sigma_j^2$
Replicate analyses	$fm(n - 1) = 8$	4.67	0.58	σ_k^2
Total	$fmn - 1 = 11$	70.92	6.45	

[A](mean squares) = (sum of squares) / (degrees of freedom).

5.2.5 Thus, the variance components can be obtained by subtracting one row from the other and then divided by the appropriate divisor as follows:

5.2.5.1 From row 3 of Table 3, we obtain the variance component due to replicate analyses:

$$\sigma_k^2 = 0.58$$

5.2.5.2 From rows 2 and 3, we obtain the variance component due to subsampling:

$$\sigma_j^2 = (7.08 - 0.58) / 3 = 2.17$$

5.2.5.3 From rows 1 and 2, we obtain the variance component due to field sampling:

$$\sigma_j^2 = (52.08 - 7.08) / [(2)(3)] = 7.50$$

5.2.6 Given these estimated variance components, the estimated total variance of one single analysis from one subsample taken from one field sample is:

$$\sigma_T^2 = \sigma_i^2 + \sigma_j^2 + \sigma_k^2 = 10.25 \quad (1)$$

5.2.7 The estimated variance components are summarized in Table 4:

TABLE 4 Variance Components from Analysis of Variance of TPH

Source of Variation	Variance Component	Percentage
Field samples (σ_i^2)	7.50	73.2
Subsampling (σ_j^2)	2.17	21.1
Analytical error (σ_k^2)	0.58	5.7
Total	10.25	100.0

5.2.8 The last column of Table 4 shows that the greatest contributor to the total variance is field sampling, accounting for 73.2 % of the total variance. Second to field sampling is subsampling, accounting for 21.1 %, while analytical error is only 5.7 %.

5.2.9 The results in Tables 3 and 4 can be obtained using software programs such as Statgraphics Plus (1993) or SAS (1993).[2,4]

5.2.10 These results imply that we can reduce the total uncertainty or variance by first focusing on field sampling variance (σ_i^2), and then laboratory subsampling variance (σ_j^2). This is discussed in the next section.

5.3 *Improving Existing Design or Optimizing a New Design*:

[2] Statgraphics Plus, "User's Manual—Nested Design," Version 7, Manugistics, Inc., 215 E. Jefferson St., Rockville, MD, 1993, pp. N1-N5.

[3] Snedecor, George W., and Cochran, William G., "Statistical Methods," 6th ed., The Iowa State University Press, Ames, IA, 1967, Section 10.16, pp. 285-288.

[4] "SAS/STAT User's Guide: The VARCOMP Procedure," Version 6, 4th ed., Vol 2, SAS Institute Inc., Cary, NC, 1993, pp. 1661-1673.

NOTICE: This standard has either been superceded and replaced by a new version or discontinued.
Contact ASTM International (www.astm.org) for the latest information.

5.3.1 Uncertainty about inference on the population mean is measured by the variance of the sample mean. In the 3-stage sampling and measurement process, the sample mean is the average of "*f*" field samples, with "*m*" subsamples taken from each field sample and each subsample analyzed "*n*" times (data from Table 2). Thus, the variance of the sample mean ($X\ldots$) is:

$$Var(X\ldots) = \sigma_i^2/f + \sigma_j^2/(fm) + \sigma_k^2/(fmn) = 7.50/2 + 2.17/4 + 0.58/12 = 4.341 \quad (2)$$

5.3.2 Eq 2 provides information on how to reduce uncertainty in the inference about the population mean.

5.3.2.1 All the denominators of the three terms on the right-hand side contain the term "*f*" for the number of field samples. Thus, an increase in "*f*" can effectively reduce the variance of the sample mean. Next in effectiveness is an increase in "*m*" as it appears on two terms containing the largest variance components (σ_i^2 and σ_j^2). And the last is an increase in "*n*" as it appears on only the term containing the smallest variance component (σ_k^2).

5.3.2.2 In the numerators of the three terms on the right hand-side, the variance component for field sampling (σ_i^2) is the largest in size. Thus, an increase in "*f*," its denominator, can most effectively reduce the variance of the mean. Next in effectiveness is an increase in "*m*".

5.3.2.3 Note that the variance of the sample mean, *Var* ($X\ldots$), has degrees of freedom of $f(m - 1) = 2$ (see row 2 of Table 3). These degrees of freedom can be used to obtain the tabled *t*-value when calculating confidence limits for the mean. The tabled *t*-value with this 2 degrees of freedom is larger than other *t*-values with larger degrees of freedom. This large *t*-value will lead to wider confidence limits and therefore is a less precise inference about the population mean. If more precise inference is needed, an increase in the number of field samples "*f*" will produce narrower confidence limits (or higher confidence) much faster than an increase in "*m*," as a result of larger degrees of freedom for the *t*-value.

NOTE 1—All the factors in the preceeding sections need to be considered jointly to find the desired solution.

5.3.3 Eq 2 can also be used to allocate resources to achieve a desired level of precision (the variance of the sample mean). Alternatively, given a desired level of precision, the optimal combination of "*f*," "*m*," and "*n*" can be found.

5.3.4 The following will discuss three different applications of these principles. The first application presents the way to determine the lowest number of samples to achieve a given level of precision. The second illustrates how to achieve the highest level of precision within a fixed budget. And finally, the third approach presents a means of maximizing precision while minimizing cost. The decision of which approach to choose will depend on the overall project objectives. The third approach represents an opportunity to balance between cost and precision and achieve an optimal solution.

5.3.5 For the example data in 5.1, the variance and standard deviation of the sample mean can be simulated for various values of *f*, *m*, and *n*. Table 5 gives some limited simulation results for illustrative purposes. In real applications, more extensive simulations may be required.

TABLE 5 Examples of Resource Allocation and Sample Variance and Standard Deviation

No. of Field Samples (*f*)	No. of Subsamples (*m*)	No. of Replicates (*n*)	Total Number of Analysis	Sample Variance	Sample Standard Deviation
1	1	1	1	10.25	3.20
1	1	2	2	9.96	3.16
1	1	3	3	9.86	3.14
1	1	4	4	9.82	3.13
1	1	5	5	9.79	3.13
1	2	1	2	8.88	2.98
1	2	2	4	8.73	2.95
1	2	3	6	8.68	2.95
1	2	4	8	8.66	2.94
1	2	5	10	8.64	2.94
1	3	1	3	8.42	2.90
1	3	2	6	8.32	2.88
1	3	3	9	8.29	2.88
1	3	4	12	8.27	2.88
1	3	5	15	8.26	2.87
2	1	1	2	5.13	2.26
2	1	2	4	4.98	2.23
2	1	3	6	4.93	2.22
2	1	4	8	4.91	2.22
2	1	5	10	4.89	2.21
2	2	1	4	4.44	2.11
2	2	2	8	4.37	2.09
2	2	3	12	4.34	2.08
2	2	4	16	4.33	2.08
2	2	5	20	4.32	2.08
2	3	1	6	4.21	2.05
2	3	2	12	4.16	2.04
2	3	3	18	4.14	2.04
2	3	4	24	4.14	2.03
2	3	5	30	4.13	2.03
3	1	1	3	3.42	1.85
3	1	2	6	3.32	1.82
3	1	3	9	3.29	1.81
3	1	4	12	3.27	1.81
3	1	5	15	3.26	1.81
3	2	1	6	2.96	1.72
3	2	2	12	2.91	1.71
3	2	3	18	2.89	1.70
3	2	4	24	2.89	1.70
3	2	5	30	2.88	1.70
3	3	1	9	2.81	1.67
3	3	2	18	2.77	1.67
3	3	3	27	2.76	1.66
3	3	4	36	2.76	1.66
3	3	5	45	2.75	1.66
4	1	1	4	2.56	1.60
4	1	2	8	2.49	1.58
4	1	3	12	2.47	1.57
4	1	4	16	2.45	1.57
4	1	5	20	2.45	1.56
4	2	1	8	2.22	1.49
4	2	2	16	2.18	1.48
4	2	3	24	2.17	1.47
4	2	4	32	2.16	1.47
4	2	5	40	2.16	1.47
4	3	1	12	2.10	1.45
4	3	2	24	2.08	1.44
4	3	3	36	2.07	1.44
4	3	4	48	2.07	1.44
4	3	5	60	2.07	1.44

5.3.5.1 Given a desired level of precision, find the minimum cost (or an optimal combination of "*f*," "*m*," and "*n*").

(1) Any combination of (*f*, *m*, *n*) in Table 5 represents a cost for sampling and analysis.

NOTICE: This standard has either been superceded and replaced by a new version or discontinued.
Contact ASTM International (www.astm.org) for the latest information.

(2) If sampling and subsampling costs are assumed to be negligible, the total analytical cost for any (*f*, *m*, *n*) combination is:

$$\text{Total cost} = (fmn)C_a \quad (3)$$

where:

(fmn) = the total number of analyses required, and
C_a = cost of an analysis.

(3) Oftentimes sampling cost is not negligible. A detailed analysis of the sampling cost is then required. Assuming there is a fixed cost (*F*) to move the sampling equipment to the field and the cost of taking a physical sample is C_f (and assuming that subsampling cost is negligible), then the total cost for any (*F*, *m*, *n*) combination is:

$$\text{Total cost} = F + f\,C_f + (fmn)C_a = F + f(C_f + mnC_a) \quad (4)$$

where:

$f\,C_f$ = cost of taking *f* field samples, exclusive of *F*, and
$(fmn)C_a$ = cost of analyzing (*fmn*) subsamples.

(4) Depending on the actual situation, either Eq 3 or Eq 4 can be calculated and included in Table 5. These results will allow the stakeholders to identify where the lowest cost is for a given precision (as represented by either the sample standard deviation or variance in the table).

5.3.5.2 Given a budget, find the highest level of precision.

(1) The variance of the sample mean, *Var* (*X*…), can be calculated for various combinations of "*f*," "*m*," and "*n*". The combination that produces the smallest value for *Var* (*X*…) and meets the total resource or cost requirements is the one to adopt. This is an effective way of determining the number of field samples to take (determination for "*f*"), the number of subsamples to take from each field sample (determination for "*m*"), and the number of replicate analyses for each subsample (determination for "*n*").

(2) Given a budget for a fixed number of analyses, Table 5 can be used to search for the smallest sample variance for that fixed number of analyses.

(3) For example, the objectives may be: (*a*) to augment the data in Table 2 to achieve a reduced overall sample variance, (*b*) to maintain the balanced design given in Table 2, and (*c*) to meet a budget of no mare than 10 new analyses. Table 5 indicates that a combination of 3 field samples, 2 subsamples per field sample and 3 analyses per subsample will give a sample variance of 2.89. This combination represents (*a*) an increase, from the data in Table 2, of 1 new field sample to be subsampled twice, which in turn is analyzed in 3 replicates (for a total of 6 new analyses) and (*b*) the new sample variance is 2.89, a substantial reduction from the original variance of 4.341. This reduction of 33 % in the sample variance will improve the statistical confidence in decision-making.

(4) If the objective is to use the results in Table 5 to optimally design a new sampling and measurement plan, then these objectives need to be specified in detailvf. For example, if the only objective is to perform no more than a total of 4 analyses, Table 5 indicates that the combination of ($f = 4$, $m = 1$, $n = 1$), for a total of 4 analyses, the sample variance is only 2.56, smaller than any other feasible combinations in Table 5. Since Table 5 is limited in simulation results, more extensive simulation may be needed for more complex applications.

5.3.5.3 Combination of increased precision and reduced cost.

(1) Approaches 5.3.5.1 and 5.3.5.2 often can be used in combination to simultaneously achieve an increase in precision and a reduction in cost.

(2) For example, the sample variance for the example data in Table 2 is 4.341 (from Eq 2), requiring a total of 12 analyses. Table 5 indicates that many combinations of (*f*, *m*, *n*) equal to or smaller than 12 analyses have a smaller sample variance. For example, for a total of 3 analyses (3 field samples, 1 subsample and 1 single analysis), a sample variance as small as 3.42 can be obtained. This represents not only a reduction in cost (number of analyses), but also an increase in precision (3.42 versus 4.341), assuming that sampling cost is negligible. Other combinations may be considered depending on project objectives. When sampling cost is not negligible, additional calculations need to be made.

6. Keywords

6.1 analysis of variance; cost-efficient; decision-making; optimization; precision; sampling and measurement process; sampling plan; sources of uncertainty; variance; variance components

ASTM International takes no position respecting the validity of any patent rights asserted in connection with any item mentioned in this standard. Users of this standard are expressly advised that determination of the validity of any such patent rights, and the risk of infringement of such rights, are entirely their own responsibility.

This standard is subject to revision at any time by the responsible technical committee and must be reviewed every five years and if not revised, either reapproved or withdrawn. Your comments are invited either for revision of this standard or for additional standards and should be addressed to ASTM International Headquarters. Your comments will receive careful consideration at a meeting of the responsible technical committee, which you may attend. If you feel that your comments have not received a fair hearing you should make your views known to the ASTM Committee on Standards, at the address shown below.

This standard is copyrighted by ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959, United States. Individual reprints (single or multiple copies) of this standard may be obtained by contacting ASTM at the above address or at 610-832-9585 (phone), 610-832-9555 (fax), or service@astm.org (e-mail); or through the ASTM website (www.astm.org).