United States
Environmental Protection
Agency

Policy, Planning
and Evaluation
(2163)

EPA 230-R-94-004
December 1992

Statistical Methods For Evaluating The Attainment Of Cleanup Standards

Volume 3: Reference-Based Standards For Soils And Solid Media

Statistical Methods For Evaluating The Attainment Of Cleanup Standards

Volume 3: Reference-Based Standards For Soils And Solid Media

Environmental Statistics and Information Division (2163)
Office of Policy, Planning, and Evaluation
U.S. Environmental Protection Agency
401 M Street, SW
Washington, D.C. 20460

June, 1994

DISCLAIMER

This work was prepared for the U.S. Environmental Protection Agency under an interagency agreement with the Department of Energy. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government. The United States Government makes no warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government.

Available to the public from the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161.

ACKNOWLEDGEMENTS

This report was prepared for the U.S. Environmental Protection Agency by Richard O. Gilbert and J. C. Simpson of Pacific Northwest Laboratory. Many individuals have contributed to this document. Rick Bates, Pacific Northwest Laboratory, provided peer review of draft chapters and insightful comments onpractical aspects of statistical procedures. Technical guidance and review were provided by the members of the Statistical Policy Branch of USEPA. Sharon McLees provided editorial support. Sharon Popp and Darlene Winter typed the multiple drafts. The authors thank all these individuals for their support and fine work.

EXECUTIVE SUMMARY

This document is the third volume in a series of volumes sponsored by the U.S. Environmental Protection Agency (EPA), Statistical Policy Branch, that provide statistical methods for evaluating the attainment of cleanup standards at Superfund sites. Volume 1 (USEPA 1989a) provides sampling designs and tests for evaluating attainment of risk-based standards for soils and solid media. Volume 2 (USEPA 1992) provides designs and tests for evaluating attainment of risk-based standards for groundwater.

The purpose of this third volume is to provide statistical procedures for designing sampling programs and conducting statistical tests to determine whether pollution parameters in remediated soils and solid media at Superfund sites attain site-specific reference-based standards. This document is written for individuals who may not have extensive training or experience with statistical methods. The intended audience includes EPA regional remedial project managers, Superfund-site potentially responsible parties, state environmental protection agencies, and contractors for these groups.

This document recommends dividing a remediated Superfund site, when necessary, into "cleanup units" and using statistical tests to compare each cleanup unit with an appropriately chosen, site-specific reference area. For each cleanup unit, samples are collected on a random-start equilateral triangular grid except when the remedial-action method may leave contamination in a pattern that could be missed by a triangular grid. In the latter case, unaligned grid sampling is recommended. The measurements for a given pollution parameter in the cleanup unit are compared with measurements obtained using triangular-grid or unaligned grid sampling in the reference area.

The comparison of measurements in the reference area and cleanup unit is made using two nonparametric statistical tests: the Wilcoxon Rank Sum (WRS) test (also called the Mann-Whitney test), the Quantile test, and a simple "hot measurement" comparison. The WRS test has more power than the Quantile test to detect uniform failure of remedial action throughout the cleanup unit. The Quantile test has more power than the WRS test to detect when remedial action has failed in only a few areas within the cleanup unit. The hot-measurement comparison consists of determining if any measurements in the remediated cleanup unit exceed a specified upper limit value, H_m. If so, then additional remedial action is required, at least locally, regardless of the outcome of the WRS and Quantile tests. This document recommends that all three tests should be conducted for each cleanup unit because the tests detect different types of residual contamination patterns in the cleanup units.

Chapter 1 discusses the purpose of this document, the intended audience and use of the document, and the steps that must be taken to evaluate whether a Superfund site has attained a reference-based standard.

Chapter 2 discusses 1) the hypotheses that are being tested by the WRS and Quantile tests and how they differ from the hypotheses used in Volumes 1 and 2, 2) Type I and Type II decision errors and why they should be specified

before collecting samples and conducting tests, and 3) the assumptions used in this volume.

Chapter 3 discusses statistical data analysis issues associated with environmental pollution measurements and how these issues are handled by the statistical procedures discussed in this document. The issues discussed are: non-normally distributed data, large variability in reference data sets, composite samples, pooling data, the reduced power to detect non-attainment of reference-based cleanup standards when multiple tests are conducted, measurements that are less than the limit of detection, outliers, the effect of residual contamination patterns on test performance, multivariate tests, and missing or unusable data.

Chapter 4 discusses the steps needed to define "attainment objectives" and "design specifications," which are crucial parts of the testing process. Definitions are given of "cleanup units," "reference region," and "reference areas." Some criteria for selecting reference areas are provided, and the cleanup standards associated with the WRS and Quantile tests are discussed. We also discuss the hot-measurement comparison and how it complements the WRS and Quantile tests to improve the probability of detecting non-attainment of reference-based cleanup standards.

Chapter 5 gives specific directions and examples for how to select sampling locations in the reference areas and the cleanup units. In this document, sampling on an equilaterial triangular grid is recommended because it provides a uniform coverage of the area being sampled and, in general, provides a higher probability of hitting hot spots than other sampling designs. However, unaligned grid sampling is recommended if the residual contamination in the remediated cleanup unit is in a systematic pattern that might not be detected by samples collected on a triangular grid pattern.

Chapters 6 and 7 explain how to use the WRS test and the Quantile test, respectively, and how to determine the number of samples to collect in the reference area and the cleanup units. Several examples illustrate the procedures. Chapter 6 also has a short discussion of when the familiar t test for two data sets may be used in place of the WRS test. In Chapter 7, we also compare the power of the WRS and Quantile tests to provide guidance on which test is most likely to detect non-attainment of the reference-based standard in various situations.

Finally, statistical tables and a glossary of terms are provided in Appendices A and B, respectively.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF BOXES

CHAPTER 1. INTRODUCTION

This is the third in a series of documents funded by the U.S. Environmental Protection Agency (EPA), Statistical Policy Branch, that describe and illustrate statistical procedures to test whether Superfund cleanup standards have been attained. These documents were prepared because neither the Superfund legislation in the Superfund Amendments and Reauthorization Act of 1986 (SARA) nor EPA regulations or guidance for Superfund sites specify how to verify that the cleanup standards have been attained.

Volume I (USEPA 1989a) in this series describes procedures for testing whether concentrations in remediated soil and solid media are statistically below a specified generic or site-specific risk-based cleanup standard or an applicable or relevant and appropriate requirement (ARAR). The statistical procedures in Volume I are appropriate when the risk-based standard is a fixed (constant) value.

The statistical procedures in Volume II (USEPA 1992) may be used to evaluate whether concentrations in groundwater at Superfund sites are statistically below a site-specific risk-based fixed-value (constant) standard.

1.1 Purpose of This Document

This document, Volume III, offers statistical procedures for designing a sampling program and conducting statistical tests to determine whether pollution parameter concentrations in remediated soils and solid media attain a site-specific reference-based cleanup standard. The objective is to detect when the distribution of measurements for the remediated cleanup unit is "shifted" in part or in whole to the right (to higher values) of the reference distribution.

Figure 1.1 shows the steps in evaluating whether remedial action at a Superfund site has resulted in attainment of the site-specific reference-based cleanup standard. Each of the steps are discussed in this document in sections identified in Figure 1.1.

1.2 Intended Audience and Use

Volume III is written primarily for individuals who may not have extensive training or experience with statistical methods for environmental data. The intended audience includes EPA regional remedial project managers, potentially responsible parties for Superfund sites, state environmental protection agencies, and contractors for these groups.

Volume III may be used in a variety of Superfund program activities:

- Emergency or Routine Removal Action: Verifying that contamination concentration levels in soil that remain after emergency or routine removal of contamination attain the reference-based cleanup standard.

- Evaluating Remediation Technologies: Evaluating whether a remediation technology is capable of attaining the reference-based cleanup standard.

- Final Status Survey: Conducting a final status survey to determine whether completed remedial action has resulted in the attainment of the reference-based cleanup standard.

- Superfund Enforcement: Providing an enhanced technical basis for negotiations between the EPA and owners/operators, consent decree stipulations, responsible party oversight, and presentations of results.

This document is not a EPA regulation. There is no EPA requirement that the statistical procedures discussed here must be used. This document should not be used as a cookbook or as a replacement for scientific and engineering judgement. It is essential to maintain a continuing dialogue among all members of the remedial-action assessment team, including soil scientists, engineers, geologists, hydrologists, geochemists, analytical chemists, and statisticians.

This document discusses only the statistical aspects of assessing the effectiveness of remedial actions. It does not address issues that pertain to other areas of expertise needed for assessing effectiveness of remedial actions such as soil remediation techniques and chemical analysis methods. Table 1.1, which is an updated version of Table 1.1 in USEPA (1989a), lists EPA guidance documents that give methods for collecting and evaluating soils data.

In this volume, the reader is advised to consult a statistician for additional guidance when the discussion and examples in this report are not adequate for the situation. Data used in the examples in this document are for data collected at actual Superfund sites.

1.3 Summary

This document gives statistical procedures for evaluating whether pollution parameter concentrations in remediated soil and solid media at Superfund sites are statistically above site-specific reference-based cleanup standards. The variability in the reference-area and cleanup-unit measurements is taken into account by the testing procedures.

The intended audience for this document includes EPA regional managers, Superfund site responsible parties, state environmental protection agencies, and contractors for these groups. This document can be applied to implement and evaluate emergency or routine removal actions, remedial response activities, final status surveys, and Superfund enforcement.

Due to the importance of technical aspects other than statistics to Superfund assessment, it is essential that all members of the assessment team interact on a continuing basis to develop the best technical approach to assessing the effectiveness of remedial action.

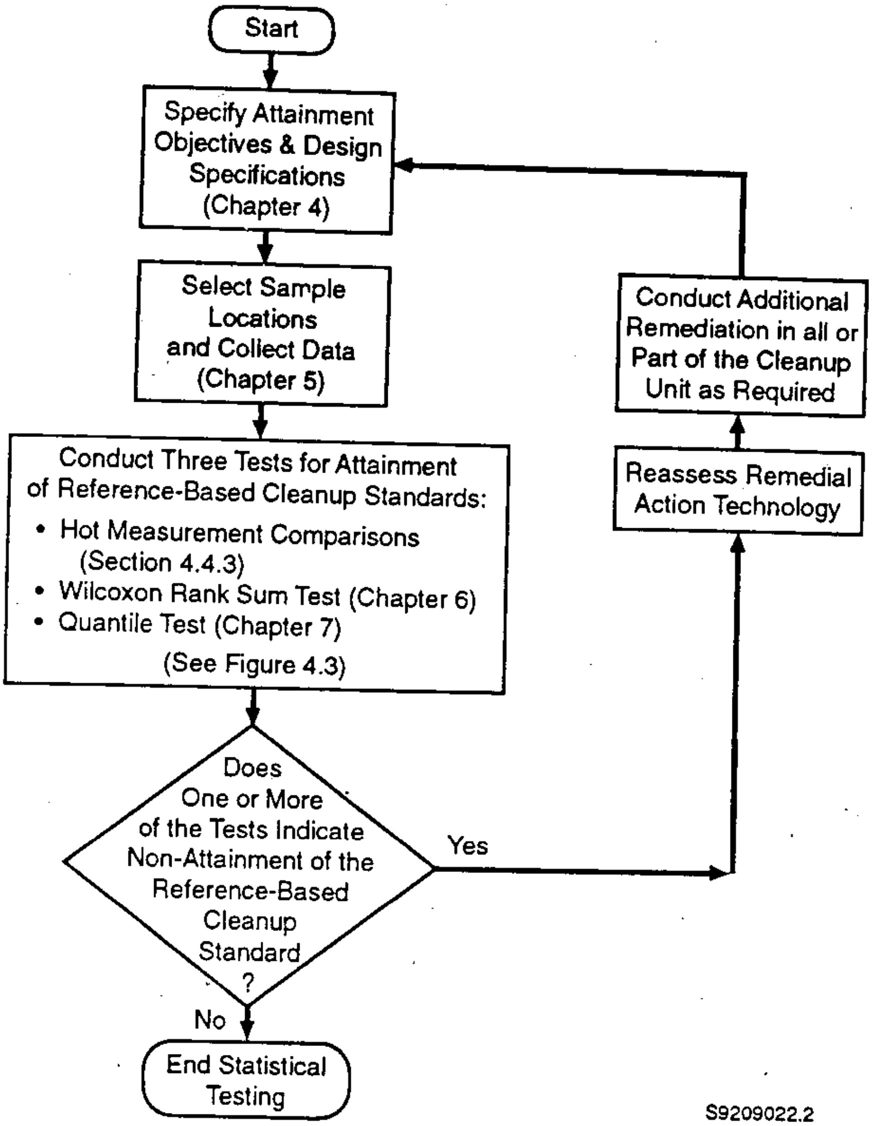


FIGURE 1.1. Steps in Evaluating Whether a Site Has Attained the Reference-Based Cleanup Standard

TABLE 1.1. Guidance Documents that Present Methodologies for Collecting and Evaluating Soils Data

Title	Sponsoring Office	Date	ID Number
Preparation of Soil Sampling Protocol: Techniques and Strategies	EMSL-LV ORD	August 1983	EPA 600/4-83-020
Verification of PCB Spill Cleanup by Sampling and Analysis	OTS OPTS	August 1985	EPA 560/5-85-026
Guidance Document for Cleanup of Surface Impoundment Sites	OERR OSWER	June 1986	OSWER Directive 9380.0-6
Test Methods for Evaluating Solid Waste	OSW OSWER	November 1987	SW-846
Draft Surface Impoundment Clean Closure Guidance Manual	OSW OSWER	March 1987	OSWER Directive 9476.0-8.C
Data Quality Objectives for Remedial Response Activities: Development Process	OERR OSWER	March 1987	EPA-540/G-87/003
Data Quality Objectives for Remedial Response Activities: Example Scenario RI/FS Activities at a Site with Contaminated Soils and Ground Water	OERR OSWER	March 1987	EPA 540/G-87/004
Soil Sampling Quality Assurance User's Guide, 2nd Edition	EMSL-LV ORD	March 1989	EPA 600/4-89-043

CHAPTER 2.0 MAKING DECISIONS USING STATISTICAL TESTS

This chapter discusses concepts that are needed for a better understanding of the tests described in this volume. We begin by discussing why statistical tests are useful for evaluating the attainment of cleanup standards. Then, the following statistical concepts and their application in this document are presented: null and alternative hypotheses, Type I and Type II decision errors, and test assumptions.

2.1 Why Statistical Tests are Used

In Chapter 2 of Volume I (USEPA 1989a) the following question was considered:

> "Why should I use statistical methods and complicate the remedial verification process?"

The answer given in Volume 1, which is also appropriate here, was essentially that statistical methods allow for specifying (controlling) the probabilities of making decision errors and for extrapolating from a set of measurements to the entire site in a scientifically valid fashion. However, it should be recognized that statistical tests cannot prove with 100% assurance that the cleanup standard has been achieved, even when the data have been collected using protocols and statistical designs of high quality. Furthermore, if the data have not been collected using good protocols and design, the statistical test will be of little or no value. Appropriate data must be obtained for a statistical test to be valid.

2.2 Hypothesis Formulation

Before a statistical test is performed it is necessary to clearly state the null hypothesis (H_o) and the alternative hypothesis (H_a). The H_o is assumed to be true unless the statistical test indicates that it should be rejected in favor of the H_a.

The hypotheses used in this document are:

H_o: Reference-Based Cleanup Standard Achieved (2.1)

H_a: Reference-Based Cleanup Standard Not Achieved

The hypotheses used in Volumes I and II (USEPA 1989a, 1992) are the reverse of those in Equation 2.1:

H_o: Risk-Based Cleanup Standard Not Achieved

H_a: Risk-Based Cleanup Standard Achieved

(2.2)

The hypotheses in Equation 2.2 are not used here for reference-based cleanup standards because they would require that most site measurements be less than the reference measurements before accepting H_a (Equation 2.2) that the cleanup standard has been attained. The authors of this report consider that requirement to be unreasonable. The hypotheses used in this document (Equation 2.1) are also used in USEPA (1989b, p. 4-8) to test for differences between contaminant concentrations in a reference area and a site of interest.

It should be understood that the use of the hypotheses in Equation 2.1 will, in general, allow some site measurements to be larger than some reference-area measurements without rejecting the null hypotheses that the reference-based cleanup standard has been achieved. The real question addressed by the statistical tests in this document (Chapters 6 and 7) is whether the site measurements are sufficiently larger to be considered significantly (statistically) different from reference-area measurements.

2.3 Decision Errors

Two types of decision errors can be made when a statistical test is performed:

1. Type I Error: Rejecting H_o when it is true.

   The maximum allowed probability of a Type I Error is denoted by α. For the hypotheses used in this document (Equation 2.1), a Type I Error occurs when the test incorrectly indicates that the cleanup standard has not been achieved. This decision error may lead to unnecessary additional remedial action.

2. Type II Error: Accepting H_o when it is false.

   The specified allowed probability of a Type II Error is denoted by β. For the hypotheses used in this document (Equation 2.1), a Type II Error occurs when the test incorrectly indicates that the standard has been achieved. This decision error may lead to not performing needed additional remedial action.

Acceptable values of α and β must be specified as part of the procedure for determining the number of samples to collect for conducting a statistical test. The number of samples collected in the reference area and in a remediated cleanup unit must be sufficient to assure that β does not exceed its specified level. Methods for determining the number of samples are given in Chapters 6 and 7.

Type I and Type II decision errors are illustrated in Figure 2.1. The "power" or ability of a test to detect when a remedial cleanup unit does not meet the standard is 1 - β. Clearly, a test should have high power, but α should also be small so that unnecessary additional remedial action seldom occurs. Unfortunately, smaller specified values of α and β require a larger number of measurements. Specifying small values of α and β may result in more samples than can be accomodated by the budget.

DECISION BASED ON SAMPLE DATA	TRUE CONDITION	
	STANDARD ACHIEVED	STANDARD NOT ACHIEVED
STANDARD ACHIEVED	Correct Decision (Probability = 1 - α)	Type II Error (Probability = β)
STANDARD NOT ACHIEVED	Type I Error (Probability = α)	Correct Decision (Power = 1 - β)

FIGURE 2.1. Type I (α) and Type II (β) Decision Errors

Regarding the choice of α, if there are many cleanup units and each unit requires a separate decision, then for approximately 100α% of those units the H_o will be incorrectly rejected and hence incorrectly declared to not meet the standard. Hence, if a larger value of α is used, the number of cleanup units for which H_o is incorrectly rejected will also be larger. This situation could lead to unnecessary resampling of cleanup units that actually met the standard. On the other hand, if larger values of α are used, the number of samples required from each cleanup unit will be smaller, thereby reducing cost.

Regarding power (1 - β), it should be understood that power is a function whose value in practice depends on the magnitude of the size of the actual non-zero (and positive) difference between reference-area and cleanup-unit measurements. As shown in Chapters 6 and 7, the number of samples depends not only on α and β, but also on the size of the positive difference that must be detected by the statistical test with specified power 1 - β.

2.4 Assumptions

The following assumptions are used in this document.

1. A suitable reference area has been selected (see Section 4.2.2).

2. The reference area contains no contamination from the cleanup unit being evaluated.

3. Contaminant concentrations in the reference area do not present a significant risk to man or the environment.

4. There is no requirement that the cleanup unit be remediated to levels less than those in the reference area even when the contaminant occurs naturally in the reference area or has been deposited in the reference area from anthropogenic (human-made, non-site) sources of pollution such as from industry or automobiles.

5. Contaminant concentrations in the reference area and in cleanup units do not change after samples are collected in these areas.

6. Contaminant concentrations in the reference area and at the remediated site do not cycle or have short-term variability during the sampling period. If such cycles are expected to occur, the reference area and the cleanup unit must be sampled during the same time period to eliminate or reduce temporal effects.

7. Measurements in the reference area and the remediated site are not spatially correlated. See Section 3.8 for discussion.

2.5 Summary

Statistical methods should be used to test for attainment of cleanup standards because they allow for specifying and controlling the probabilities of making decision errors and for extrapolating from a set of measurements to the entire cleanup unit in a scientifically valid fashion.

In this document the null hypothesis being tested is

H_o: Reference-Based Cleanup Standard Achieved.

The alternative hypothesis that is accepted if H_o is rejected is

H_a: Reference-Based Cleanup Standard Not Achieved.

The use of this H_o and H_a implies that the cleanup unit will be accepted as not needing further remediation if the measurements from the cleanup unit are not demonstrably larger, in a distribution sense, than the site-specific reference-area measurements. This H_o and H_a, which are the reverse of those used in Volumes 1 and 2 (USEPA 1989a, USEPA 1992), are used here because the authors believe it is unreasonable to require cleanup units to be remediated to achieve residual concentrations less than what are present in the reference area.

Two types of decisions errors can be made when using a statistical test: A Type I error (rejecting the null hypothesis when it is true) and a Type II error (accepting the null hypothesis when it is false). Acceptable probabilities that these two errors occur must be specified as part of the procedure for determining the number of samples to collect in the reference area and remediated cleanup units. See Chapters 4, 6 and 7 for further details.

CHAPTER 3.0 STATISTICAL DATA ANALYSIS ISSUES

There are several data analysis issues that must be considered when selecting sampling plans and statistical tests to assess attainment of cleanup standards. In this chapter we discuss these issues and the approaches used in this document to address them.

3.1 Non-Normally Distributed Data

Many statistical tests were developed assuming the measurements have a normal (Gaussian) distribution. However, experience has shown that measurements of contaminant concentrations in soil and solid media are seldom normally distributed.

In this document we recommend and discuss non-parametric statistical tests, i.e., tests that do not require that the measurements be normally distributed. If the measurements should happen to be normally distributed, these nonparametric tests will have slightly less power than their parametric counterparts that were developed specifically for normally distributed data. However, the nonparametric tests may have greater power than their parametric counterparts when the data are not normally distributed.

3.2 Large Variability in Reference Data

Measurements of chemical concentrations in a reference area may be highly variable and have distributions that are asymmetric with a long tail to the right (i.e., there are a few measurements that appear to be unusually large). The reference area distribution could also be multimodal. For a given number of samples, large variability tends to reduce the power, 1 - ß, of statistical tests (Section 2.3) to detect non-attainment of standards. It is important to use the most powerful tests possible and to collect enough samples to achieve the required power. This document illustrates procedures to determine the number of samples needed to achieve adequate power (Chapters 6 and 7).

3.3 Composite Samples

A composite sample is a sample formed by collecting several samples and combining them (or selected portions of them) into a new sample, which is then thoroughly mixed before being analysed (in part or as a whole) for contaminant concentrations. Composite samples may be used to estimate the average concentration for the cleanup unit with less laboratory analysis cost. Also, compositing may increase the power of statistical tests to detect non-attainment of reference-based standards. This increased power could occur because compositing may decrease the variability among the measurements obtained from composite samples. However, compositing methods must not be adopted without carefully evaluating their variability and the representativeness of the area being sampled. This important topic is discussed further in Section 4.3.1.

3.4 Pooling Data

If several data sets have been collected in the reference area at different times or in difference portions of the area, consideration should be given to whether the data should be combined (pooled) before a test for attainment of reference-area standards is made. Such pooling of data, when appropriate, will tend to increase the power to detect when the reference-area standard has not been attained.

Pooling of data sets should only be done when all the data were selected using the same sample collection, handling, and preparation procedures. For example, all samples should be collected from the same soil horizon, and the same soil compositing technique should be used. Also, if the data sets were collected at different times, pooling should not be done if the average or variability of the data change over time. Such time changes will tend to increase the Type I and Type II error rates of tests.

To illustrate the effect of using different sample-collection methods, suppose the depth of surface-soil samples was different for two reference-area data sets. Then it would not be appropriate to combine the data sets if contaminant concentrations change with depth. One data set would tend to have higher concentrations (and perhaps higher variability) than the other set, due entirely to the method used to collect the soil samples. Hence, the variability of the data in the combined data set would be larger than for either data set, which could reduce the power and increase the Type I error rate of the test for attainment of the reference-area standard. However, the increased number of samples may mitigate these effects.

It is not correct to pool data simply to achieve a desired test result. For example, it may be known that soil samples collected previously in a subsection of the reference area have higher concentrations than the data collected more recently on a grid over the entire reference area. Suppose that a statistical test that compares the grid data to data collected in a cleanup unit indicates that the cleanup unit requires additional remediation. It would not be correct to pool the subsection and the grid data in an attempt to reverse the test result. Instead, additional soil samples should be collected in the reference area to determine if the higher concentrations in the subsection can be confirmed. If so, then consideration should be given to whether the subsection should be part of the reference area that is compared with the cleanup unit. The problem becomes one of deciding whether the boundary of the reference area should be changed.

3.5 Multiple Tests

Many statistical tests may be conducted at a Superfund site because many pollutants are present at the site and/or because a separate decision is needed for each cleanup unit. When multiple tests are conducted, the probability that at least one of the tests will incorrectly indicate that the standard has not been attained will be greater than the specified α (probability of a Type I Error for a given test). If each of u independent statistical tests are performed at the α significance level when all cleanup units are in compliance with standards, then the probability all u tests will indicate attainment of compliance is $p = (1 - \alpha)^u$. For example, if $\alpha = 0.05$ and $u = 25$, then $p = (0.95)^{25} = 0.28$, and if $u = 100$, then $p = (0.95)^{100} = 0.0059$. Hence, as the number of tests, u, is increased the probability approaches 0 that all u tests will correctly indicate attainment of the standard.

This problem has led to the development of multiple comparison tests, which are discussed in, e.g., Hochberg and Tamhane (1987) and Miller (1981). Two multiple comparison tests that could potentially be used for testing attainment of reference-based standards are those by Dunnett (1955, 1964) and Steel (1959). In general, for these tests, the α level of each individual test is made small enough to maintain the overall α level (i.e., the α level for all tests taken as a group) at the required level. However, unless there is an appropriate increase in the number of measurements, the multiple-comparison tests may have very low power to detect the failure to reduce contamination to reference levels.

Because of this severe loss of power, we do not recommend using multiple comparison techniques when testing for the attainment of reference-based cleanup standards when the number of tests is large. Also, practical limitations in field remedial-action activities may prevent doing statistical testing until several cleanup units or pollution parameters can be tested simultaneously.

Rather than conduct multiple comparison tests, we recommend conducting each test at the usual α level (say 0.01 or 0.05) so that the power of each test is maintained. The problem of large numbers of false positives (Type I errors) when multiple-comparison tests are not used can be handled by collecting additional representative samples in those cleanup units for which test(s) indicated non-attainment of the reference-based standard.

When there are several contaminants in a cleanup unit that must be tested for attainment of reference standards, an alternative approach to multiple comparison tests is to conduct a multivariate test. Multivariate tests are discussed in Section 3.9.

3.6 Data Less Than the Limit of Detection

Frequently, measurements of pollution parameters in soil and solid media will be reported by the analytical laboratory as being less than the analytical limit of detection. These measurements are often called "less-than data," and data sets containing less-than data are called censored data sets. Aside from the problems of how a chemist determines the detection limit and its exact meaning [see USEPA (1989a; pp. 2-15) and Lambert, et al. (1991)], there is the problem of how to conduct valid statistical tests when less-than data are present. Some papers that discuss statistical aspects of this problem are Gilbert and Kinnison (1981), Gleit (1985), Gilliom and Helsel (1986), Helsel and Gilliom (1986), Gilbert (1987), Millard and Deverel (1988), Helsel and Cohn (1988), Helsel (1990), and Atwood, et al. (1991). The WRS and Quantile tests discussed in this document allow for less-than measurements to be present in the reference area and the cleanup units, as discussed in Chapters 6 and 7.

3.7 Outliers

Outliers are measurements that are unusually large relative to most of the measurements in the data set. Many tests have been proposed to detect outliers from a specified distribution such as the Normal (Gaussian) distribution; see e.g., Beckman and Cook (1983), Hawkins (1980), Barnett and Lewis (1985), and Gilbert (1987). Tests for outliers may be used as part of the data validation process wherein data are screened and examined in various ways before they are placed in a data file and used in statistical tests to evaluate attainment of cleanup standards. However, it is very important that no datum should be discarded solely on the basis of an outlier test. Indeed, there is always a small chance (the specified Type I error probability) that the outlier test incorrectly declares the suspect datum to be an outlier. But more important, outliers may not be mistakes at all, but rather an indication of the presence of hot spots, in which case the Superfund site may require further remediation.

Outlier tests are primarily useful for identifying data that may require further evalution to determine if they are the result of mistakes. If no mistakes are found, the outlier should be accepted as a valid datum and used in the test for attainment of the reference-based standard. We note that the Quantile Test (Chapter 7) can be viewed as a test for multiple outliers in the cleanup-unit data set, where the standard for comparison is the data set for the site-specific reference area.

3.8 Spatial Patterns in Data

The statistical tests described in this document assume that there is no correlation among the samples collected on the equilateral triangular grid spacing for the reference areas and cleanup units. If the data are correlated, then the Type I and Type II error rates will be different than their specified values. Chapter 10 in Volume 1 (USEPA 1989a) discusses geostatistical methods that take into account spatial correlation when assessing compliance with risk-based standards. Cressie (1991) and Isaaks and Srivastava (1989) provide additional information about geostatistical methods.

As discussed in Chapter 5, this document recommends that whenever possible, samples should be collected on an equilateral triangular grid. One advantage of this design is that if spatial correlation is present at the grid spacing used, the data may be suitable for estimating the spatial correlation structure using geostatistical methods.

3.9 Multivariate Tests

In many cases, more than one contaminant will be present in a cleanup unit. Suppose there were K > 1 contaminants present in soil at the site before remedial action. Then one may consider conducting a multivariate statistical test of the null hypothesis that the cleanup standards of all K contaminants have been achieved, versus the alternative hypothesis that the cleanup standard has not been achieved for one or more of the K contaminants. Two such (nonparametric) tests are the multivariate multisample Wilcoxon Rank Sum test and the multivariate multisample median test (Schwertman 1985).

However, a discussion of these tests is beyond the scope of this report. Also, additional studies to evaluate the power of these tests for Superfund applications is needed before they can be recommended for use.

3.10 Missing or Unusable Data

Missing or unusable data can occur with any sampling program. Samples can be mislabeled, lost, held too long before analysis, or they may not meet quality control standards. As discussed in Volume I (USEPA 1989a), the pattern of missing data should be examined to determine if a bias in statistical tests could arise.

Also, to account for the likelihood of missing or unusable data, it is prudent to increase the number of samples that would otherwise be collected. Let n be the number of samples that would be collected if no missing or unusable data are expected. Let R be the expected rate of missing or unusable data based on past experience. Then the total number of samples to collect, n_f, is (from USEPA 1989a, pp. 2-15):

$$\boxed{n_f = n / (1 - R)} \tag{3.1}$$

The use of Equation 3.1 will give some assurance that enough samples will be collected to meet specified Type I and Type II error-rate requirements.

3.11 Summary

This chapter discusses statistical data analysis problems and how they influence the choice of sampling plans and tests. This document emphasizes the use of nonparametric tests because of the possibility that environmental pollution measurements from reference areas and cleanup units will not be normally distributed.

Large data variability tends to reduce the power of statistical tests. This document gives procedures for determining the number of samples required to achieve required power.

When using compositing methods, careful consideration must be given to whether the data from composite samples will be meaningful for assessing attainment of reference-based standards.

Although multiple comparison tests can be used to limit to a specified level the number of cleanup units incorrectly categorized as needing additional remedial action, these tests are not recommended here because they can result in a severe loss of power to detect when a cleanup unit needs additional remedial action. A preferred approach is to take additional samples in cleanup units for which statistical tests indicated additional remedial action may be required.

The nonparametric tests discussed in this document can be conducted when data sets are censored if the number of less-than data is not too large.

Outliers (unusually large measurements) should not be removed from the data set unless they can be shown to be actual mistakes or errors.

The data analysis and testing procedures in this document require that measurements are not spatically correlated at the spacing used for the equilateral triangular grid. However, if measurements are spatially correlated at the grid spacing, then geostatistical methods should be considered for use (USEPA 1989a; Cressie 1991; Isaaks and Srivastava 1989).

When more than one contaminant is present in a cleanup unit, it may be possible to use a multivariate statistical procedure to test whether one or more of the reference standards has not been attained, rather than conduct a series of univariate tests for the individual contaminants. However, the performance of multivariate tests for Superfund applications has not been sufficiently evaluated to permit a recommendation for their use. The reader should consult a statistican for assistance in applying multivariate tests.

Compensation for anticipated missing or unusable data can be made by increasing the number of samples using Equation 3.1.

CHAPTER 4. ATTAINMENT OBJECTIVES AND THE DESIGN SPECIFICATION PROCESS

In this chapter we discuss attainment objectives and the design specification process, which are important parts of the Data Quality Objectives (DQOs) process that should be followed when testing for the attainment of site-specific reference-based cleanup standards. Figure 4.1 gives the sequence of steps needed to define attainment objectives and design specifications. The figure also indicates the sections in this report where each step is discussed. We begin this chapter with a brief discussion of DQOs.

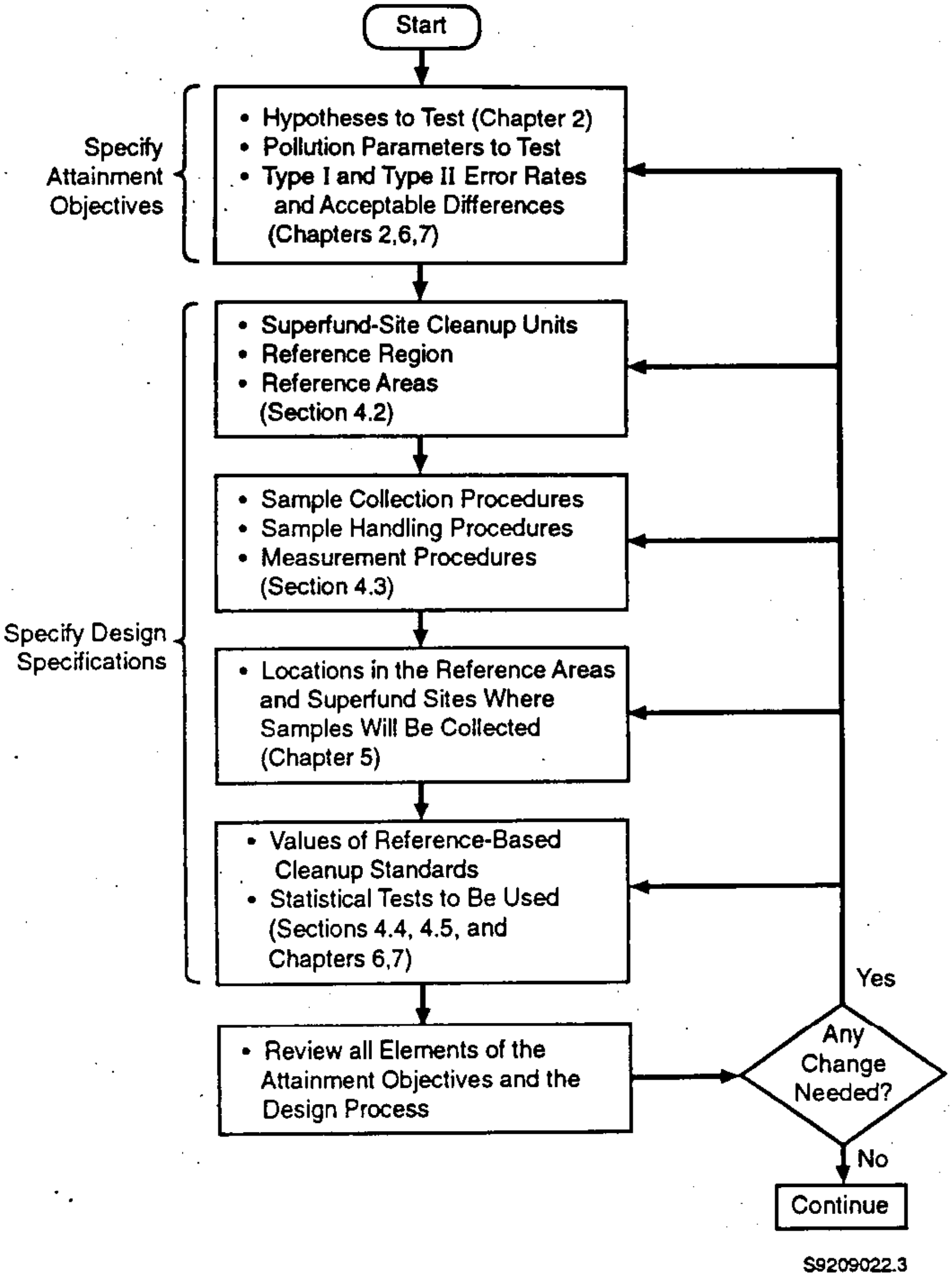
4.1 Data Quality Objectives (DQOs)

Data Quality Objectives (DQOs) are qualitative and quantitative statements that specify the type and quality of data that are required for the specified objective.

As indicated above, the development of attainment objectives and design specifications, which are discussed in this chapter and in Chapter 5, are an important part of the DQO process. The DQO process addresses the following issues (USEPA 1989a, 1987a, and 1987b):

- the objective of the sampling effort
- the decision to be made
- the reasons environmental data are needed and how they will be used
- time and resource constraints on data collection
- detailed description of the data to be collected
- specifications regarding the domain of the decision
- the consequences of an incorrect decision attributable to inadequate environmental data
- the calculations, statistical or otherwise, that will be performed on the data to arrive at the result, including the statistics that will be used to summarize the data and the "action level" (cleanup standard) to which the summary statistic will be compared
- the level of uncertainty that the decision maker is willing to accept in the results derived from the environmental data

All of the above items should be addressed when planning a sampling program to test for the attainment of cleanup standards. Neptune et al. (1990) and Ryti and Neptune (1991) illustrate the development and use of DQOs for Superfund-site remediation projects.



<u>FIGURE 4.1</u>. Steps in Defining Attainment Objectives and the Design Specifications

4.1.1 Attainment Objectives

Attainment Objectives are objectives that must be attained by the sampling program. Attainment objectives are developed by re-expressing the general goal of "testing for attainment of reference-based cleanup standards" in terms of testing specific pollution parameters using specific null and alternative hypotheses, Type I and Type II error rates, and an acceptable "average" difference. Hypotheses and error rates were introduced in Chapter 2. Examples of these concepts are given in Chapters 6 and 7.

It is necessary to specify acceptable Type I and Type II error rates as part of the procedure for determining the number of samples to collect in the reference area and the remediated cleanup units. When the number of samples to be collected is determined in an ad hoc manner without clear-cut numerical Type I and Type II error rates, it is more likely that the Superfund-site owner/operator will be requested or required to collect additional samples at possibly great cost with no clear end point in sight.

4.1.2 Design Specification Process

The Design Specification Process is the process of specifying the field sampling design, cleanup standards, statistical tests, number of samples, and the sample collection, handling, measurement, and quality assurance procedures that are needed to achieve the attainment objectives.

4.2 Specifying the Sampling Design

The first step in the design specification process (Figure 4.1) is to specify the site-specific reference region, the reference area(s) within the reference region, and the cleanup unit(s) within the Superfund site being remediated. These geographical areas, which are illustrated in Figure 4.2, are defined below.

4.2.1 Definitions

Cleanup Units:

Geographical areas of specified size and shape at the remediated Superfund site for which separate decisions will be made regarding the attainment of the applicable reference-based cleanup standard for each designated pollution parameter.

Reference Areas:

Geographical areas from which representative reference samples are selected for comparison with samples collected in cleanup units at the remediated Superfund site.

Reference Region:

The geographical region within which reference areas are selected.

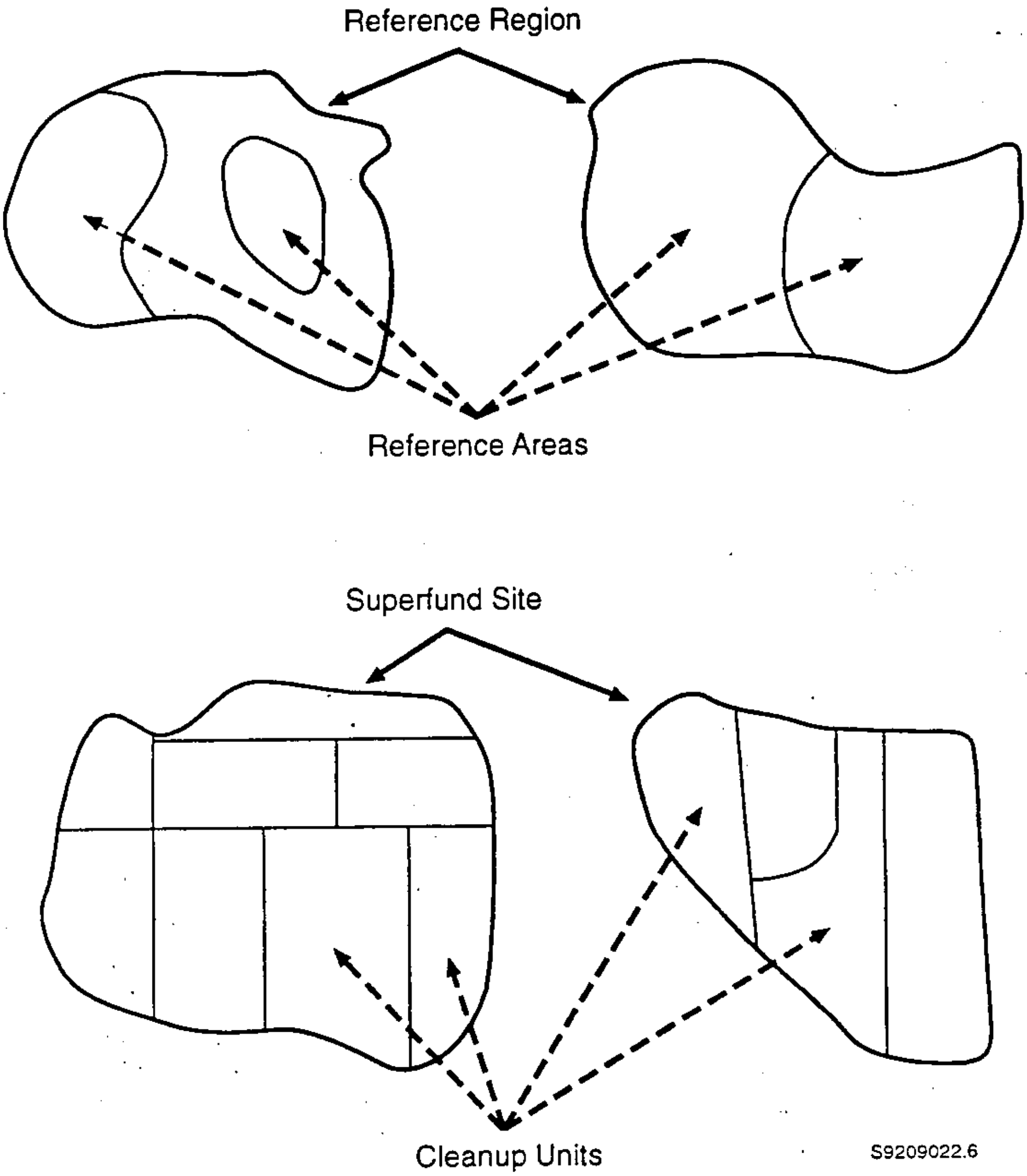


FIGURE 4.2. Geographical Areas at the Superfund Site and the Site-Specific Reference Region

4.2.2 Design Considerations

The remediated Superfund site may have one, a few, or many cleanup units. A separate set of soil samples is collected and measured in each cleanup unit for comparison with the same type of samples and measurements from the applicable reference area. The number, location, size, and shape of cleanup units may differ depending on interrelated factors such as the size and topography of the site, cost and convenience factors, the type of remedial action that was used, the expected patterns of residual contamination that might remain after remedial action, and assessed risks to the public if the reference-area cleanup standard is not attained. Whenever possible all cleanup units should be approximately the same size so that the number of samples and the distances between samples in the field will not be greatly different for the cleanup units. For similar reasons, it is desirable for the reference area to be approximately the same size as the applicable cleanup unit. However the reference area should be large enough to encompass the full range of background conditions.

Neither the reference region nor the Superfund site will necessarily be one contiguous area (Figure 4.2). At some Superfund Sites a single reference area (perhaps the entire reference region) may be appropriate for all cleanup units. At other sites, the physical, chemical, or biological characteristics of different cleanup units may differ enough to warrant matching each cleanup unit with its own unique reference area within the reference region.

In some situations, reference areas that are closest to but unaffected by the cleanup unit may be preferred, assuming spatial proximity implies similarity of reference area concentrations. If concentrations differ systematically within the reference region the reference areas may contain quite different concentration levels. In this case, different cleanup units would have a different cleanup standard, which may not be reasonable. In this situation, consideration may be given to using the entire reference region as the reference area for all cleanup units, as proposed in DOE (1992) for the Hanford Site in Washington State.

In some cases, a buffer zone that surrounds the Superfund Site should be established as a distinct cleanup unit (or units) from which soil samples are collected and evaluated for attainment of reference-based cleanup standards. The buffer zone may consist of the area that could have been contaminated as a result of remedial-action activities and/or environmental transport mechanisms (e.g., wind and water movement, or redistribution by wildlife) during or following remedial action.

Neptune et al. (1990) point out that, in general, dividing the Superfund site into spatially distinct cleanup units for testing purposes may result in missing an unacceptably contaminated area that lies across two or more cleanup units. However, the likelihood of missing a contaminated area should be reduced if the Quantile test (Chapter 7) and the hot-measurement comparison (Section 4.4.3 below) are used.

In some cases information may not be available to do a completely defendable job of matching a cleanup unit with a reference area. In this document we assume that either the required information is available to achieve an acceptable matching or that environmental samples will be collected to provide that information. General criteria for selecting reference areas are given in the next section.

4.2.3 Criteria for Selecting Reference Areas

The following criteria should guide the selection of the reference region and reference areas (Liggett 1984):

1. The reference region and reference area(s) must be free of contamination from the remediated site.

2. The distribution of pollution-parameter concentrations in the applicable reference area should be the same as the distribution of concentrations that would be present in the cleanup unit if that unit had never become contaminated by man's local activities at the site.

   The soil of the reference area(s) is allowed to contain concentrations that are naturally occurring or arise from the activities of man on a regional or worldwide basis. Examples of such anthropogenic sources of pollution parameters include low concentrations of persistent organic compounds that have been used globally and low concentrations of radionuclides that were distributed via worldwide fallout (DOE 1992).

3. A reference area selected for comparison with a given cleanup unit or set of cleanup units should not differ from those cleanup units in physical, chemical, or biological characteristics that might cause measurements in the reference area and the cleanup unit to differ.

Selecting reference areas that satisfy these criterion will require professional judgement supported by historical and/or new measurements of soil samples.

4.3 Procedures for Collecting, Handling, and Measuring Samples

The procedures used to collect, handle, and measure environmental samples from the reference areas and the cleanup units must be developed, documented, and followed with care. Also, to the extent possible, these procedures should be the same for the remediated cleanup units and the applicable reference areas. If these conditions are not met, the resulting measurements may be biased or unnecessarily variable, in which case the statistical test results may be meaningless and/or the test may have little power to detect when the reference-based standard has not been attained. The documents listed in Table 1.1 (Chapter 1) provide information on procedures for soil sample collecting, handling, and measurements.

4.3.1 Subsampling and Composite Sampling

It is important to carefully consider and document:

- the type of composite samples, if any, that will be formed
- whether the entire sample (or composite sample) or only one or more portions (aliquots) from the sample (or composite sample) will be measured.

In general, the variance of measurements of pollution parameters for composite samples collected over time or space will tend to be smaller than the variance of noncomposited samples. One implication of this phenomenon is that if composite samples are used, the same compositing methods must be used in the reference area and the remediated cleanup unit. Otherwise, the measurements in the two areas will not be comparable and the statistical tests will not be valid. Also, the compositing process may average out (mask) small areas that have relatively high concentrations.

Before a decision is made to collect composite samples the following conditions should be met:

- All stakeholders must agree that a measurement obtained from a specific type of composite sample is the appropriate metric for making cleanup decisions.
- The sample collection and handling procedures must be specifically designed to collect and adequately mix composite samples according to a written protocol.
- The same procedures must be used to collect, mix, and analyze composite samples in the reference area and the remediated cleanup unit.

Additional information on statistical aspects of compositing is given by Duncan (1962), Elder et al. (1980), Rohde (1976), Schaeffer et al. (1980), Schaeffer and Janardan (1978), Gilbert (1987), Garner et al. (1988), Bolgiano et al. (1990), and Neptune et al. (1990). The statistician on the remedial-action planning team should be consulted regarding the design of any sampling program that may involve composite sampling.

4.3.2 Quality Assurance and Quality Control

Quality assurance and quality control methods and procedures for collecting and handing samples must be an integral part of the soil sampling program. This topic is discussed in USEPA (1984, 1987a, 1987b), Brown and Black (1983), Taylor and Stanley (1985), Garner (1985), Taylor (1987) and Keith (1991).

4.4 Specification of the Reference-Based Cleanup Standard

Two types of cleanup standards are used in this document. The first type of standard is a specific value of a statistical parameter associated with the statistical tests discussed in Sections 4.4.1 and 4.4.2 below. The second type of standard is a specific upper-limit concentration value, H_m, for the pollution parameter of interest, as discussed in Section 4.4.3.

4.4.1 Wilcoxon Rank Sum Test

When the Wilcoxon Rank Sum (WRS) test (Hollander and Wolfe 1973, Gilbert 1987) is used, the applicable statistical parameter is P_r and the standard is $P_r = 1/2$, where

P_r = probability that a measurement of a sample collected at a random location in the cleanup unit is greater than a measurement of a sample collected at a random location in the reference area.

If $P_r > 1/2$, then the remedial action in that cleanup unit has not been complete. In this document the WRS test (Chapter 6) is used to detect when $P_r > 1/2$.

4.4.2 Quantile Test

When the Quantile test (Johnson et al. 1987) is used, the applicable parameters are ϵ and Δ/σ, and the standard is $\epsilon = 0$ and $\Delta/\sigma = 0$, where

ϵ = proportion of the soil in the remediated cleanup unit that has not been remediated to levels in the reference area, and

Δ/σ = amount (in units of standard deviation) that the distribution of 100ϵ% of the measurements in the remediated cleanup unit is shifted to the right (to higher measurements) of the distribution in the reference area.

If $\epsilon > 0$, then $\Delta/\sigma > 0$ and the remedial action has not been complete. In this document the Quantile test (Chapter 7) is used to detect when $\epsilon > 0$.

4.4.3 Hot-Measurement Comparison

The hot-measurement comparison consists of comparing each measurement from the cleanup unit with a upper-limit concentration value, H_m. The cleanup standard is this specific value of H_m, where

H_m = a concentration value such that any measurement from the remediated cleanup unit that is equal to or greater than H_m indicates an area of relatively high concentrations that must be remediated, regardless of the outcome of the WRS or Quantile tests.

Of course, there must be assurance that the measurement(s) that equals or exceeds H_m is not the result of a mistake or of inappropriate sample collection, handling, or analysis procedures. The selected value of H_m might be based on a site-specific risk assessment or an estimated upper confidence limit (such as the 95th) for an upper quantile (such as the 95th) of the distribution of measurements from the reference area. The value of H_m or the procedure used to determine H_m must be determined by negotiation between the EPA (and/or a comparable state agency) and the Superfund-site owner or operator.

The hot-measurement comparison is used in conjunction with the WRS and Quantile tests because the latter two tests can fail to reject H_0 when only a very few high measurements in the cleanup unit are obtained. The use of H_m may be viewed as insurance that unusually large measurements will receive proper attention regardless of the outcome of the WRS and Quantile tests.

4.5 Selection of the Statistical Test

Two important criteria for the selection of a statistical test are:

- the power of the test to detect non-attainment of the standard
- the sensitivity of the test results to the presence of less-than values.

The WRS Test has more power than the Quantile test to detect when the remediated cleanup unit has concentrations uniformly higher than the reference area. However, the WRS test allows for fewer less-than measurements than does the Quantile Test. As a general rule, the WRS test should be avoided if more than about 40% of the measurements in either the reference area or the cleanup unit are less-than data.

The Quantile Test has more power than the WRS Test to detect when only a small portion of the remediated cleanup unit has not been successfully remediated. Also, the Quantile test can be used even when a fairly large proportion of the cleanup-unit measurements (more than 50%) are below the limit of detection.

As illustrated in Figure 4.3, the WRS and Quantile tests are conducted for each remediated cleanup unit so that both types of unsuccessful remediation (uniform and spotty) can be detected. Also, the hot measurement (H_m) comparison (Section 4.4.3) is conducted in each unit to assure that a single or a very few unusually large measurements receive proper attention.

4.6 Number of Samples: General Strategy

In general, the number of samples required for the WRS test and the Quantile test will differ for specified Type I and Type II error rates. The following procedure is recommended for determining the number of samples to collect:

1. If the remedial-action procedure is likely to leave concentrations in the cleanup unit that are uniform in value over space, then the number of samples should be greater than or equal to the number of samples determined using the procedures given in Section 6.2 for the WRS test.

2. If the remedial action procedure is likely to leave spotty (non-uniform) rather than uniform (over space) concentrations in the cleanup unit, then the number of samples should be greater than or equal to the number

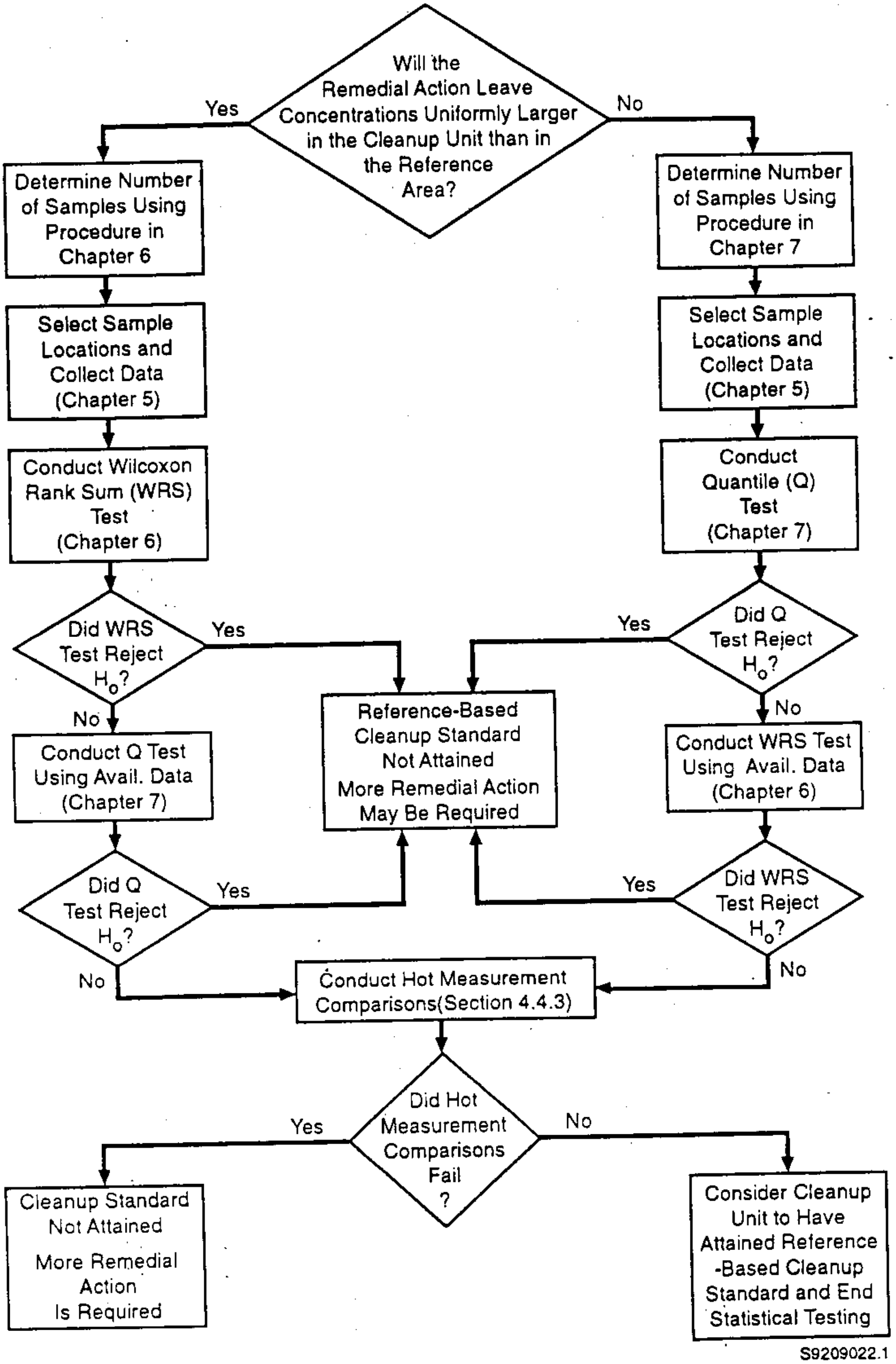


FIGURE 4.3. Sequence of Testing for Attainment of Reference-Based Cleanup Standards

determined using the procedure described in Section 7.2 for the Quantile test.

3. If there is very little difference between the number of samples determined for the two tests, or if there is little or no information available about whether the remedial action procedure is more likely to leave spotty or uniform contamination, then the larger of the number of samples for the WRS and Quantile tests should be used.

4. When determining the required number of samples, we recommend first selecting the overall Type I error level (α) desired for both tests combined. Then divide this overall error level by 2 and use this smaller value to determine the number of samples using the procedures in Sections 6.2 and 7.2. For example, if an overall type I error level of $\alpha = 0.05$ is desired, then determine the number of samples using $\alpha/2 = 0.025$.

5. If it is necessary to detect isolated hot spots of specified size and shape with specified probability, then the number of samples needed to to detect hot spots with specified probability, as described in USEPA (1989a, Chapter 9) or Gilbert (1987), should be used. If the number of samples determined using that approach is larger than the number of samples obtained using the methods in Section 6.2 or 7.2, then more samples than indicated by those latter methods could be collected. This approach would increase the power of the WRS test and the Quantile test to levels greater than the specified minimum power ($1 - \beta$).

4.7 Summary

Attainment objectives and the design specification process must be carefully specified as part of the process of testing for compliance with site-specific reference-based cleanup standards.

Steps in Defining Attainment Objectives:

1. **Specify the Pollution Parameters to be Tested.** These parameters should be listed for each cleanup unit.

2. **Specify the Null and Alternative Hypotheses.** The hypotheses used in this document are given by Equations 2.1, 6.2 and 7.2.

3. **Specify the Type I and Type II Error Rates for the Tests.** The specification of Type I and Type II error rates is part of the process of determing the number of samples that must be collected. This process is illustrated in Chapters 6 and 7 for the WRS and Quantile tests, respectively.

Steps in the Design Specification Process:

1. **Specify the Cleanup Units.** The remediated Superfund site may be divided into two or more geographical cleanup units for which separate decisions will be made concerning attainment of reference standards.

2. **Specify the Reference Region.** The reference region defines the region within which all site-specific reference samples will be collected.

3. **Specify the Reference Area(s).** Reference areas are defined areas within the reference region that are chosen because their physical, chemical and biological characteristics are similar to those characteristics in specified cleanup units. Different cleanup units and/or pollution parameters may require different reference areas.

4. **Specify the Sample Collection, Handling, and Measurement Procedures.** Clearly define and document the type and size of soil or solid-media samples, the sample handling procedures, and the measurement procedures. These procedures should be identical for the reference area and the remediated cleanup units. If it is impossible for the procedures to be identical, then experiments should be conducted to determine the effect of non-identical procedures on the measured values and the conclusions drawn from statistical tests for non-attainment.

5. **Specify Sample Locations in the Reference Area(s) and the Cleanup Unit(s)** Methods for determining sample locations are given in Chapter 5.

6. **Specify the Values of the Cleanup Standard.** Specify the value of H_m (a concentration value) for the hot-measurement comparison. The cleanup standards for the WRS and Quantile tests are $P_r = 1/2$ and $\epsilon = 0$, $\Delta/\sigma = 0$, respectively. These tests are discussed and illustrated in Chapters 6 and 7, respectively.

7. **Determine the Number of Samples to Collect.** The procedure in Sections 4.6, 6.2 and 7.2 are used to determine the number of samples to collect.

8. **Review all Elements of the Attainment Objectives.** Review and revise, if necessary, the attainment objectives and design specifications.

CHAPTER 5. SELECTING SAMPLE LOCATIONS

After the attainment objectives and the design specifications (Chapter 4) have been defined, attention should be directed to specifying how to select locations where samples will be collected, which is the topic of this chapter.

5.1 Selecting Sampling Locations in Reference Areas and Cleanup Units

There are many ways to select sampling locations. USEPA (1989a) shows how to use simple random sampling, stratified random sampling, systematic sampling, or sequential sampling to select sampling locations for assessing if a soils remediation effort at a Superfund site has succeeded in attaining a risk-based standard.

In this document, we recommend collecting samples in reference areas and cleanup units on a random-start equilateral triangular grid except when the remedial-action method may leave contamination in a pattern that could be missed by a triangular grid, in which case unaligned grid sampling is recommended.

The triangular pattern has the following advantages:

- It is relatively easy to use.
- It provides a uniform coverage of the area being sampled, whereas simple random or stratified random sampling can leave subareas that are not sampled.
- Samples collected on a triangular grid are well suited for estimating the spatial correlation structure of the contamination, which is required information if geostatistical procedures (USEPA 1989a; Cressie 1991; Isaaks and Srivastava 1989) are used to evaluate the attainment of cleanup standards.
- The probability of hitting a hot spot of specified elliptical shape one or more times is almost always greater using a triangular grid than using a square grid when the density of sample points is the same for both types of grids for the areas being investigated (Singer 1975).

However, caution is needed when using the triangular (or any regular) grid. The grid points (sampling locations) must not correspond to patterns of high or low concentrations. If such a correspondence exists, the measurements and statistical test results could be very misleading. In that case, simple random sampling within each cleanup unit could be used, but a uniform coverage would not be achieved. Alternatively, the unaligned grid (Gilbert 1987, p. 94; Cochran 1977, p. 228; Berry and Baker 1968), which incorporates an element of randomness in the choice of sampling locations, should do a better job of avoiding biased sampling while retaining the advantage of uniform coverage.

The decision not to recommend stratified random sampling in this document is based on the following considerations. When stratified random sampling is used, the remediated Superfund site is divided into relatively homogeneous subareas (strata) and a simple random sample is collected in each area. This method was applied in USEPA (1989a) to the situation where a test is made to determine whether the entire remediated Superfund site (all cleanup units combined) met a risk-based standard. By dividing the total area into homogeneous strata, a better estimate of the mean concentration in the remediated site can be obtained, which tends to increase the power of the test.

However, in this document, the view is taken that if sufficient information is available to split up the Superfund site into internally homogeneous areas (cleanup units), then a separate test for compliance with the reference standard should be made in each area. With this approach, there is no interest in conducting a test for the entire Superfund site, and hence no need to use stratified random sampling.

5.2 Determining Sampling Points in an Equilateral Triangular Grid Pattern

In this section we show how to set up an equilateral triangular sampling grid in a reference area(s) and in any cleanup unit. If a square grid is used, the reader is directed to USEPA (1989a) for the procedure to determine sample locations. The main steps in the process for the triangular grid are as follows (from USEPA 1989a):

1. Draw a map of the area(s) to be sampled as illustrated in Figure 5.1.
2. Locate a random sampling point using the procedure in Box 5.1.
3. Determine the approximate sampling locations on the triangular grid using the procedure in Box 5.2.
4. Ignore any sampling locations that fall outside the area to be sampled.

Using this procedure, the number of sampling points on the triangular grid within the sampling area may differ from the desired number n depending on the shape of the area. If the number of points is greater than the desired number, use all the points. If the number of points is less than the desired number, select the remaining points at individual random locations within the sampling area using the procedure in Box 5.1 for each additional point.

5.3 Determining Exact Sample Locations

The procedure in Section 5.2 gives the approximate sampling points in the field. As indicated in USEPA (1989a), the points are approximate because "the sampling coordinates were rounded to distances that are easy to measure, the measurement has some inaccuracies, and there is judgment on the part of the field staff in locating the sample point." USEPA (1989a) recommends a procedure to locate the exact sample collection point that avoids subjective bias factors such as "difficulty in collecting a sample, the presence of vegetation, or the color of the soil".

The recommended methods for locating exact sample collecting points in the field are given in Box 5.3 (from USEPA 1989a). Box 5.4 gives an example of setting up a triangular grid and determining exact sample locations.

5.4 Summary

In this chapter, a method for determining sampling locations in reference areas and cleanup units on a random-start equilateral triangular pattern is discussed and illustrated. The random-start equilateral triangular grid pattern is the method of choice because:

- it is easy to implement
- it provides a uniform coverage of the area to be sampled
- the data are well suited for estimating the spatial correlation structure of the contamination
- the probability of hitting an elliptical hot spot one or more times is almost always larger if an equilateral triangular grid rather than a square grid is used.

A triangular or any other systematic grid sampling plan can lead to invalid statistical tests if the grid points happen to be located in patches of only relatively high or low concentrations. If that situation is likely to occur, then the unaligned grid design may be preferred.

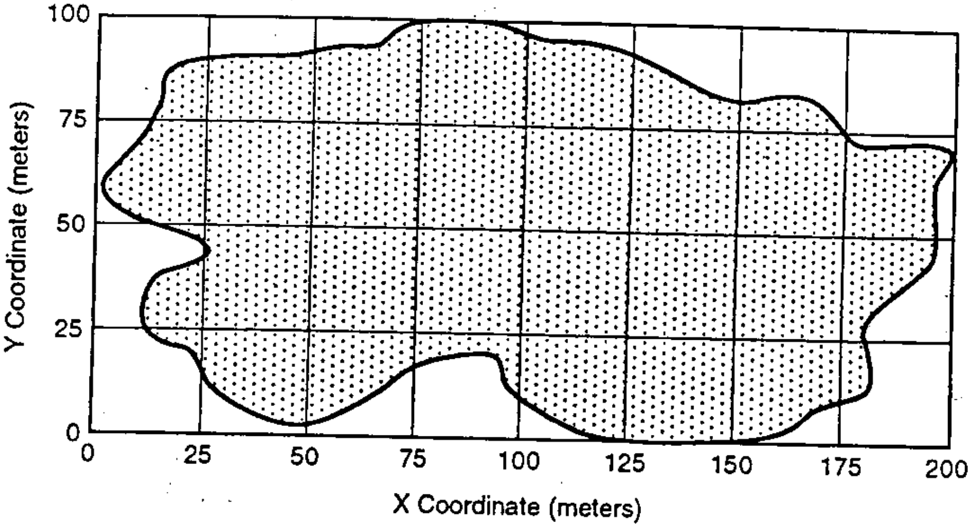
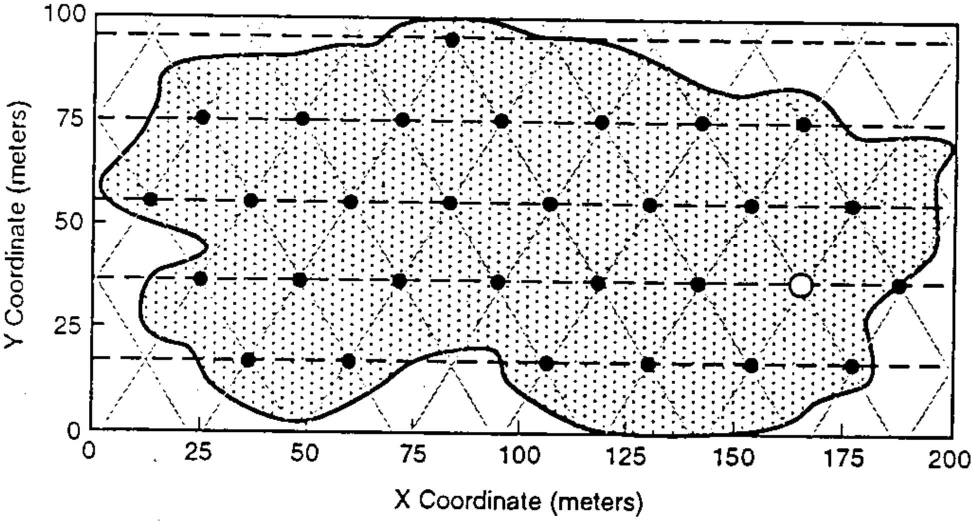


FIGURE 5.1. Map of an Area to be Sampled



S9209022.9

FIGURE 5.2. Map of an Area to be Sampled Showing a Triangular Sampling Grid

BOX 5.1

STEPS FOR DETERMINING A RANDOM POINT WITHIN A DEFINED AREA*

1. Determine the location (X, Y) in the defined area:

   $$X = X_{min} + RND_1 \times (X_{max} - X_{min})$$

   $$Y = Y_{min} + RND_2 \times (Y_{max} - Y_{min})$$

   where RND_1 and RND_2 are random numbers between 0 and 1 obtained using a calculator, computer software or a random number table**. X_{max}, X_{min}, Y_{max} and Y_{min} are the corners of a rectangular area that encloses the area to be sampled. These corners are illustrated in Figure 5.1 for the case $X_{min} = 0$, $X_{max} = 200$, $Y_{min} = 0$, and $Y_{max} = 100$.

2. If the computed (X, Y) from Step 1 is outside the area to be sampled, return to Step 1. Otherwise, go to Step 3.

3. Determine the random location (X_i, Y_i) as follows:

   Round X from Step 1 to the nearest unit, e.g., 1 or 5 meters, that can be easily located in the field. Denote this nearest unit by X_i.

   Round Y from Step 1 to the nearest unit that can be easily located in the field. Denote this nearest unit by Y_i.

   (X_i, Y_i) is the desired random point.

---

* This procedure is similar to the procedure in USEPA (1989a).

** Random number tables are found in many statistics books, e.g., Table A1 in Snedecor and Cochran (1980).

BOX 5.2

PROCEDURE FOR FINDING APPROXIMATE SAMPLING LOCATIONS ON A TRIANGULAR GRID*

1. Determine the surface area, A, of the area to be sampled.

2. Determine the total number of sampling locations, n, required in the area (see Chapters 6 and 7).

3. Compute L as follows:

$$L = \left(\frac{A}{0.866\ n} \right)^{1/2}$$

4. Draw a line parallel to the X axis through the point (X_i, Y_i) that was obtained using the procedure in Box 5.1. Mark off points a distance L apart on this line.

5. To lay out the next row, find the midpoint between the last two points along the line and mark a point at a distance 0.866 L perpendicular to the next line. This is the first point of the next line.

6. Mark off points a distance L apart on this new line.

7. Repeat steps 5 and 6 until the n points throughout the entire area to be sampled have been determined.

*This procedure is from USEPA (1989a). A similar procedure is in Kelso and Cox (1986).

BOX 5.3

STEPS FOR DETERMINING EXACT SAMPLING LOCATIONS STARTING FROM POINTS ON A TRIANGULAR GRID

1. Determine the n points on a triangular grid using the Procedure in Box 5.2.
2. Let M be the accuracy to which distances were measured in the field to determine the triangular grid. For example, M might be 1 meter.
3. At each of the locations on the triangular grid, choose a random* distance (between -M to M) to go in the X direction and then a random distance (from -M to M) to go in the Y direction, to determine the exact sample location.
4. Collect the samples at the exact sample locations determined in Step 3.
5. Record the exact locations where the samples were collected.

* Random numbers can be generated using a calculator in the field. Alternatively, they could be determined prior to going out to the field using a calculator, random number table, or a computer.

BOX 5.4

EXAMPLE OF SETTING UP A TRIANGULAR GRID AND DETERMINING EXACT SAMPLE LOCATIONS IN THE FIELD

This example is illustrated in Figure 5.2.

1. From Figure 5.1 we find $X_{min} = 0$, $Y_{min} = 0$, $X_{max} = 200$, and $Y_{max} = 100$.

2. Suppose a random number generator on a calculator is used to obtain the random numbers 0.037 and 0.457 between 0 and 1.

3. Using Step 1 in Box 5.1:

$$X = 0 + 0.037*(200 - 0) = 7.4 = 7$$
$$Y = 0 + 0.457*(100 - 0) = 45.7 = 46$$

This point, (X, Y) = (7, 46), is outside the sampled area. Therefore, repeating the process we obtain random numbers 0.820 and 0.360, for which

$$X = 0 + 0.820(200 - 0) = 164$$
$$Y = 0 + 0.360(100 - 0) = 36$$

Therefore, (X, Y) = (164, 36) is the random starting point for the triangular grid (Figure 5.2). We assume that measurements can be made to the nearest meter in the field.

4. The surface area of the sample area in Figure 5.1 is A = 14,025 square meters. Suppose the number of locations where samples will be collected is n = 30. (Methods for determining n are given in Chapters 6 and 7.)

5. Use the formula for L in Box 5.2:

$$L = (14{,}025/0.866*30)^{1/2} = 23.23 -23$$

6. Draw a line parallel to the X axis through the point (164, 36). Mark off points 23 meters apart on this line.

7. Find the midpoint between the last two points along the line and mark a point at a distance 0.866*23 = 19.92 -20 meters perpendicular to the line at that midpoint. This point is the first sample location on the next line.

BOX 5.4 (continued)

8. Mark off points at distance L = 23 meters apart on this new line.

9. Repeat steps 7 and 8 until the triangular grid is determined.

10. In this example, the exact number of sample locations (30) is obtained. Hence, no random locations need to be determined.

11. For each of the 30 sample locations, determine the exact sample locations by selecting a random distance between -1 and 1 meter to go in the X direction and a random distance from -1 to 1 meter to go in the Y direction. The distance from -1 to 1 meter is used because in this example the accuracy to which distances were measured in the field to determine the triangular grid was 1 meter. Record the exact sampling location.

CHAPTER 6. WILCOXON RANK SUM (WRS) TEST

In this chapter we show how to use the Wilcoxon Rank Sum (WRS) test to assess whether a cleanup unit at a remediated Superfund site has attained the site-specific reference-based cleanup standard for a pollution parameter. In Chapter 7 we show how to conduct the Quantile test for that purpose. As discussed in Chapter 4, both the WRS test and the Quantile test should be performed for each remediated cleanup unit because the two tests detect different types of non-attainment. The WRS test has more power than the Quantile test to detect when remedial action has resulted in cleanup-unit contamination levels that are still uniformly (over space) larger than in the reference area. The Quantile test has better power than the WRS test to detect when remedial action has failed in only a few areas within the cleanup unit.

Briefly, the WRS test is performed by first listing the combined reference-area and cleanup-unit measurements from smallest to largest and assigning the ranks 1, 2, ... to the ordered values. Then the ranks of the measurements from the cleanup unit are summed and used to compute the statistic Z_{rs}, which is compared to a critical value from the standard normal distribution. If Z_{rs} is greater than or equal to the critical value, then we conclude that the cleanup unit has not attained the reference-area cleanup standard.

In Section 6.1 we begin by discussing the appropriate form of the testing hypotheses for the WRS test. Then we show how to determine the number of samples to collect (Section 6.2) and how to perform the test (Section 6.3). In Section 6.4 we briefly discuss the two-sample t test, a test that may be preferred to the WRS test under special, although usually unrealistic, conditions. The chapter concludes with a summary in Section 6.5.

6.1 Hypotheses and the Reference-Based Cleanup Standard

As stated in Section 2.2, the hypotheses used in this document are:

H_o: Reference-Based Cleanup Standard Achieved

H_a: Reference-Based Cleanup Standard Not Achieved (6.1)

where H_o is assumed to be true unless the test indicates H_o should be rejected in favor of H_a. When H_o is true, the distribution of measurements in the reference area is very similar in shape and central tendency (average) to the distribution of measurements in the remediated cleanup unit.

When using the WRS test, the above hypotheses are restated as follows:

$$\boxed{\begin{array}{ll} H_o: & P_r = 1/2 \\ H_a: & P_r > 1/2 \end{array}} \quad (6.2)$$

where

P_r = probability that a measurement of a sample collected at a random location in the cleanup unit is greater than a measurement of a sample collected at a random location in the reference area.

As stated in Chapter 4 (Section 4.4.1), the cleanup standard for the WRS test is the value of P_r given in the H_o. Hence, from Equation 6.2, the standard is $P_r = 1/2$. Indeed, if the distribution of measurements at the remediated cleanup unit is identical to the distribution of measurements in the applicable reference area, then P_r equals 1/2. However, if P_r is actually larger than 1/2, then some of the distribution of measurements in the remediated cleanup unit lay to the right of the distribution for the reference area.

When determining the number of samples to collect, it is necessary to specify a value of P_r that is greater than 1/2, as well as the required power of the WRS test to reject H_o when P_r equals that specified value. This procedure is discussed and illustrated in the next section.

6.2 Number of Samples

Noether (1987) developed for the WRS test a formula (Equation 6.3) that may be used for computing the approximate total number of samples (N) to collect in the reference area and in the cleanup unit being compared with the reference area. This formula can be used regardless of the shape of the reference-area and cleanup-unit distributions. We note that an approximate formula for computing N for any specified (known) distribution is provided by Lehman (1975, Equation 2.33). He also gives an approximate formula for the special case of a normal (Gaussian) distribution (his Equation 2.34). However, Noether's formula may be used when the distribution is unknown, which is frequently the case.

Noether's formula, when divided by the factor 1 - R to account for expected missing or unusable data (see Equation 3.1 in Chapter 3), is

$$N = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2}{12c(1 - c)(P_r - 0.5)^2(1 - R)} \tag{6.3}$$

= total number of required samples,

where

α = specified Type I error rate (see Chapter 2)

β = specified Type II error rate (see Chapter 2)

$Z_{1-\alpha}$ = the value that cuts off $(100\alpha)\%$ of the upper tail of the standard normal distribution

$Z_{1-\beta}$ = the value that cuts off $(100\beta)\%$ of the upper tail of the standard normal distribution

c = specified proportion of the total number of required samples, N, that will be collected in the reference area (see Section 6.2.1 below)

m = number of samples required in the reference area

P_r = specified probability greater than 1/2 and less than 1.0 that a measurement of a sample collected at a random location in the cleanup unit is greater than a measurement of a sample collected at a random location in the reference area.

R = expected rate of missing or unusable data (Chapter 3, Equation 3.1)

Recall from Section 4.6 that the value of α (first parameter in the above list) should be one half of the overall Type I error rate for the WRS and Quantile tests combined. For example, if an overall Type I error rate of 0.10 is required for the WRS and Quantile tests combined, then the number of samples required for the WRS test should be determined using $\alpha = 0.05$.

Some typical values of $Z_{1-\alpha}$ and $Z_{1-\beta}$ for use in Equation 6.3 are given in Table 6.1. The values in Table 6.1 are from Table A.1 (Appendix A), which is a table of the cumulative standard normal (Gaussian) distribution.

Equation 6.3 gives the total number of samples, i.e., the sum of the number of samples for the reference area and the number of samples for the cleanup unit being compared with that reference area. This total number, N,

TABLE 6.1. Some Values of Z_ϕ that May be Used to Compute N Using Equation 6.3

ϕ	Z_ϕ
0.700	0.524
0.800	0.842
0.900	1.282
0.950	1.645
0.975	1.960
0.990	2.326

---

* These and other values of Z_ϕ were obtained from Table A.1 in Appendix A.

is apportioned to the reference area and the cleanup unit using the specified proportion c defined above:

$$\boxed{\begin{aligned} m &= cN \\ &= \text{number of samples required} \\ &\quad\ \text{in the reference area} \end{aligned}} \qquad (6.4)$$

and

$$\boxed{\begin{aligned} n &= (1 - c)N \\ &= \text{number of samples required} \\ &\quad\ \text{in the cleanup unit} \end{aligned}} \qquad (6.5)$$

where N is computed using Equation 6.3.

If there are several cleanup units that will be compared with a reference area, then n measurements from each cleanup unit would be required.

6.2.1 Determining c, the Proportion of Samples for the Reference Area

The value of c to use in Equations 6.3, 6.4 and 6.5 for a given pollution parameter can be determined by specifying

- the number of cleanup units, h, that will be compared to the reference area, and
- the ratio of standard deviations, $v = \sigma_r/\sigma_c$

where

σ_r = standard deviation of the measurements for the reference area and

σ_c = standard deviation of the measurements for the remediated cleanup units.

We assume that σ_c is the same for all remediated cleanup units.

The number of cleanup units, h, will usually be known, but the ratio v can only be estimated from collected samples and/or other information.

Case 1: v Equal to 1

In some situations it may be reasonable to assume that the standard deviation for the cleanup units, σ_c, will be approximately equal to the standard deviation for the reference area, σ_r. In that case, v will be approximately equal to 1. If it is assumed that v = 1, then c can be determined using the following equation (from Hochberg and Tamhane 1987, p. 202):

$$\boxed{c = \frac{h^{1/2}}{h^{1/2} + 1}} \tag{6.6}$$

When this equation is used, we are in effect assuming that v = 1 and that the measurements of the specified pollution parameter in the reference and remediated cleanup units are normally distributed. Some values of c computed using Equation 6.6 for various values of h are given in Table 6.2.

TABLE 6.2. Values of c for Various Values of the Number of Cleanup Units (h) when $\sigma_r/\sigma_c = 1$.

Number of Cleanup Units (h)	Proportion of Samples to be Collected from Reference Area (c)
1	0.50
2	0.59
4	0.67
6	0.71
10	0.76
15	0.79
20	0.82
50	0.88
100	0.91

Suppose, for example, that h = 4 remediated cleanup units will be compared with an applicable reference area and the standard deviations for all h cleanup units and the reference area are approximately equal. Then we would use c = 0.67 in Equation 6.3 to determine N. Also, Equations 6.4 and 6.5 would be used to determine m and n, respectively, where m is the number of measurements to take in the reference area and n is the number of measurements to take in each of the four cleanup units.

Case 2: v Not Equal to 1

If there is no reason to expect that the standard deviation of measurements for the cleanup units and the reference area will be equal, then c can be computed using

$$\boxed{c = \frac{v^2 h^{1/2}}{v^2 h^{1/2} + 1}} \qquad (6.7)$$

For example, suppose there are h = 2 cleanup units and v = 2 (i.e., the standard deviation for the reference area is twice as large as that for the cleanup units). Then Equation 6.7 gives

$$c = \frac{(2)^2 * 2^{1/2}}{(2)^2 * 2^{1/2} + 1} = 0.85$$

This value of c would be used in Equations 6.3, 6.4, and 6.5 to determine N, m and n as before.

For another example, suppose there are h = 2 cleanup units, but that v = 1/2 (i.e., the standard deviation for the reference area is only half as large as that for the cleanup units). Then Equation 6.7 yields

$$c = \frac{(1/2)^2 * 2^{1/2}}{(1/2)^2 * 2^{1/2} + 1} = 0.26$$

which is used in Equations 6.3, 6.4 and 6.5 to determine N, m and n.

These two examples illustrate that the allocation of measurements, c, between the reference area and the cleanup units can be very different for different values of v.

Examples 6.1 and 6.2 (Boxes 6.1 and 6.2) illustrate how to use Equations 6.3 through 6.6.

BOX 6.1

EXAMPLE 6.1

COMPUTING THE NUMBER OF SAMPLES NEEDED FOR THE WILCOXON RANK SUM TEST WHEN ONLY ONE CLEANUP UNIT WILL BE COMPARED WITH THE REFERENCE AREA

1. State the question:

How many samples are required to test H_o versus H_a (Equation 6.2) using the WRS test when we require a Type I error rate of $\alpha = 0.05$ and power $1-\beta = 0.70$ when $P_r = 0.75$? Suppose we expect about 10% of the data to be missing or unusable and we assume the standard deviations of reference-area and cleanup-unit measurement distributions are equal.

2. Specifications given in the question:

$1 - \beta = 0.70$ $\quad P_r = 0.75$
$\alpha = 0.05$ $\quad R = 0.10$
$c = 0.50$ (from Equation 6.6)

3. Using Equation 6.3 and the appropriate values of Z_ϕ from Table 6.1:

$$N = \frac{(1.645 + 0.524)^2}{12*0.5(1 - 0.5)(0.75 - 0.5)^2(1 - 0.10)}$$

$$= \frac{4.7046}{0.1687}$$

$$= 27.9 \text{ or } 28$$

Using Equations 6.4 and 6.5:

$m = 0.5*28 = 14$
$n = 0.5*28 = 14$

4. Conclusion:

A total of 14 samples is needed in both the reference area and the cleanup unit. As discussed in Chapter 5, this document recommends collecting the samples in each area from a random-start equilateral triangular grid.

BOX 6.2

EXAMPLE 6.2

COMPUTING THE NUMBER OF SAMPLES NEEDED FOR THE WILCOXON RANK SUM TEST WHEN <u>TWO CLEANUP UNITS</u> WILL BE COMPARED WITH THE REFERENCE AREA

1. State the question:

How many samples are required to test H_0 versus H_a using the WRS test when we require a Type I error rate of $\alpha = 0.05$ and power = 0.80 when $P_r = 0.70$? Suppose we expect about 5% of the data to be missing or unusable and that we assume the standard deviations for the reference area and cleanup units are equal.

2. Specifications given in the question:

$1 - \beta = 0.80$ $P_r = 0.70$
$\alpha = 0.05$ $R = 0.05$
$c = 0.59$ (from Equation 6.6)

3. Using Equation 6.3 and the appropriate values of Z_ϕ from Table 6.1:

$$N = \frac{(1.645 + 0.842)^2}{12*0.59(1 - 0.59)(0.70 - 0.5)^2(1 - 0.05)}$$

$$= \frac{6.185}{0.110}$$

$$= 56.07$$

Using Equations 6.4 and 6.5:

$$m = 0.59*56.07 = 33.1 \text{ or } 34$$
$$n_1 = n_2 = 0.41*56.07 = 22.99 \text{ or } 23$$

4. Conclusions:

34 samples need to be collected in the reference area and 23 samples need to be collected in each of the cleanup units. This document recommends collecting samples from a random-start equilateral triangular grid.

6.2.2 Methods for Determining P_r

A value of the probability P_r must be specified when Equation 6.3 is used to determine N. However, it may be difficult to understand what a specific value of P_r really means in terms of the differences in the distributions of measurements in the reference area and the cleanup units. Two ways of alleviating this problem are discussed below.

6.2.2.1 The Odds Ratio, d, Used to Determine a Value of P_r

Rather than specify P_r, it may be easier to understand a value of the odds ratio, d, where

$$\boxed{d = \frac{P_r}{1 - P_r} = \frac{\text{probability a measurement from the cleanup unit is larger than one from the reference area}}{\text{probability a measurement from the cleanup unit is smaller than one from the reference area}}} \qquad (6.8)$$

For example, we might want to have a specified power $1 - \beta$ that the WRS test will indicate the cleanup unit needs additional remedial action when $d = 2$, i.e., when the probability a measurement obtained at random from the cleanup unit is larger than one from the reference area is <u>twice</u> as large as the probability it is smaller than an observation from the reference area. Once a value of d is specified, P_r is easily obtained using the equation

$$\boxed{P_r = \frac{d}{1 + d}} \qquad (6.9)$$

This value of P_r is then used in Equation 6.3 to determine N.

Some values of P_r for selected values of d are given in Table 6.3, as determined using Equation 6.9.

TABLE 6.3. Values of P_r for Selected Values of the Odds Ratio d (Equation 6.9)

d	P_r	d	P_r
1.2	0.55	5	0.83
1.5	0.60	6	0.86
2	0.67	10	0.91
3	0.75	20	0.95
4	0.80	100	0.99

6.2.2.2 The Amount of Relative Shift, Δ/σ, Used to Determine a Value of P_r

Rather than specify P_r directly or by first specifying d, one could think in terms of the amount of relative shift, Δ/σ, in the cleanup-unit distribution to the right (to higher values) of the reference distribution that is important to detect with specified power $1 - \beta$. Then, if the measurements of the pollution parameter in both the reference area and the cleanup units are normally distributed with the same standard deviation, σ, this Δ/σ can be transformed into the equivalent value of P_r using the equation

$$\boxed{P_r = \phi(0.707\Delta/\sigma)} \qquad (6.10)$$

where

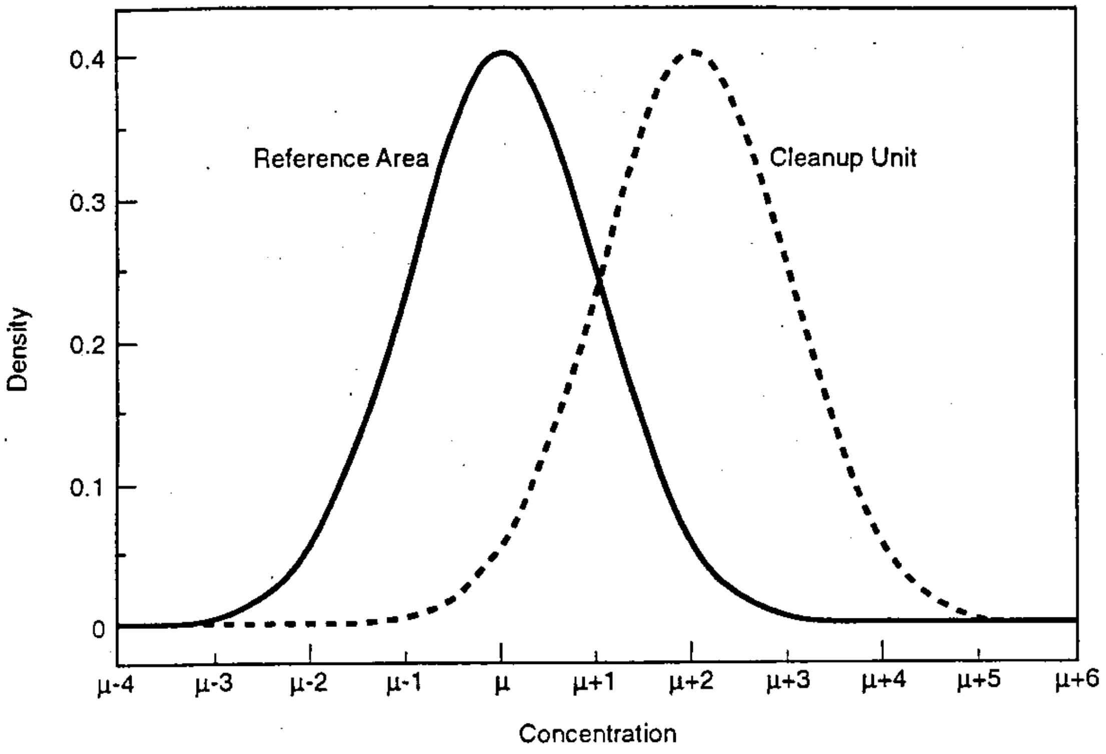
$\phi(0.707\Delta/\sigma)$ = probability that a measurement drawn at random from a normal distribution with mean 0 and standard deviation 1 will be less than $0.707\Delta/\sigma$.

The probability $\phi(0.707\Delta/\sigma)$ is determined from Table A.1 in Appendix A. This value of ϕ, i.e., of P_r, can then be used in Equation 6.3 to determine N.

For example, suppose the measurements of a pollution parameter in the reference area and cleanup unit are both normally distributed with the same standard deviation $\sigma = 1$ ppm. Further, suppose the cleanup-unit distribution is shifted to the right of the reference-area distribution by the amount $\Delta = 2$ ppm. (This example is illustrated in Figure 6.1.) Then $\Delta/\sigma = 2$, Equation 6.10, and Table A.1 give

$$P_r = \phi(0.707*2/1) = \phi(1.414) = 0.921$$

Some values of P_r computed using Equation 6.10 for selected values of Δ/σ are given in Table 6.4.



S9209022.8

FIGURE 6.1. Illustration of When the Distribution of Measurements for a Pollution Parameter in the Remediated Cleanup Unit is Shifted Two Units to the Right of the Reference Area Distribution for that Pollution Parameter.

TABLE 6.4. Values of P_r Computed Using Equation 6.10 when the Reference-Area and Cleanup-Unit Measurements are Normally Distributed with the Same Standard Deviation, σ, and the Cleanup-Unit Distribution is Shifted an Amount Δ/σ to the Right of the Reference Area Distribution

P_r	Δ/σ	P_r	Δ/σ
0.50	0.00	0.80	1.19
0.55	0.18	0.85	1.47
0.60	0.36	0.90	1.81
0.65	0.55	0.95	2.33
0.70	0.74	0.99	3.29
0.75	0.95		

It is also possible to determine N using Figure 6.2 once a value of P_r has been determined. However, Figure 6.2 may be used only for the special case of m = n for when both the reference-area and cleanup-unit measurements are normally distributed with the same σ. If Figure 6.2 is used when c is not equal to 1/2, the value of N obtained from that figure must be multiplied by the factor

$$\boxed{F = \frac{0.25}{c\ (1-c)}}$$

In summary, the procedure for determining P_r and then N when the reference-area and cleanup-unit distributions are both normal with the same standard deviation σ is:

1. Specify the amount of shift in units of standard deviation, Δ/σ, that must be detected with power $1 - \beta$.
2. Use the ratio Δ/σ, Equation 6.10, and Table A.1 to determine P_r.
3. Use P_r in Equation 6.3 or Figure 6.2 to determine N.
4. If Figure 6.2 is used and c is not equal to 1/2, then multiply the N obtained from Figure 6.2 by the factor F (Equation 6.11) to determine the required N.

This procedure is illustrated in Box 6.3 and Box 6.4 when Figure 6.2 is used to determine N.

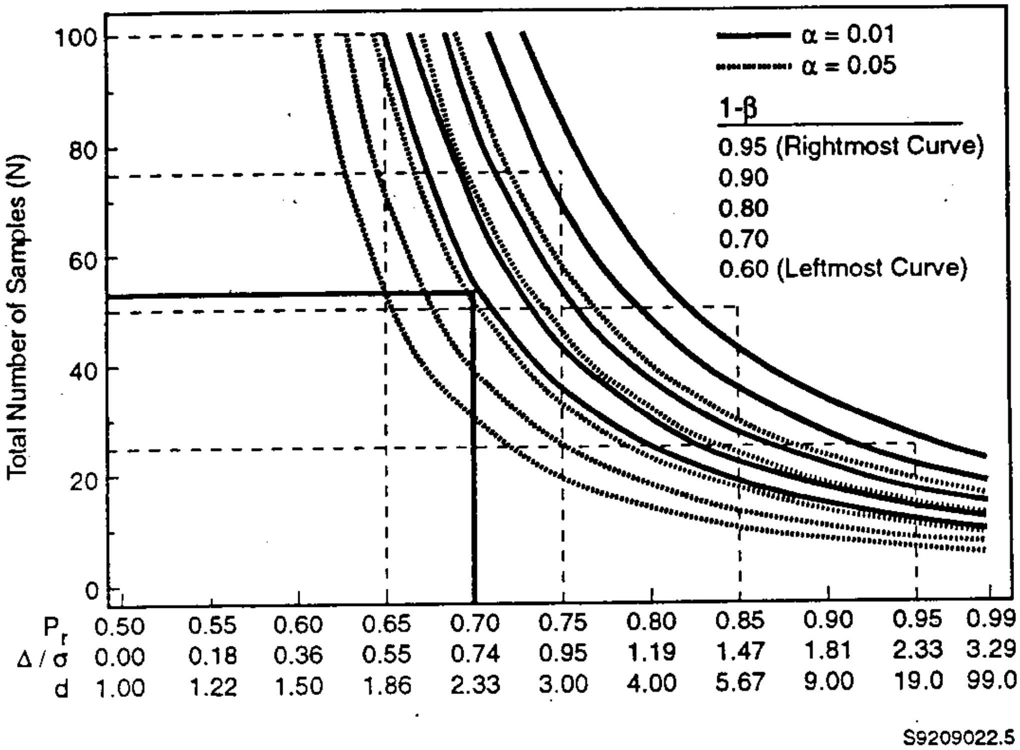


FIGURE 6.2. Power (1 - ß) of the Wilcoxon Rank Sum Test when n = m or the Distribution of Measurements for a Pollution Parameter in the Reference Area and Remediated Cleanup Unit are Both Normally Distributed with the Same Standard Deviation, σ.

6.3 Procedure for Conducting the Wilcoxon Rank Sum Test

For each cleanup unit and pollution parameter, use the following procedure to compute the WRS test statistic and to determine on the basis of that statistic if the cleanup unit being compared with the reference area has attained the reference-area standard. This procedure is illustrated in Box 6.5 and Box 6.6.

1. Collect the m samples in the reference area and the n samples in the cleanup unit (m + n = N).

2. Measure each of the N samples for the pollution parameter of interest.

3. Consider all N data as one data set. Rank the N data from 1 to N; that is, assign the rank 1 to the smallest datum, the rank 2 to the next smallest datum,..., and the rank N to the largest datum.

4. If several data are tied, i.e., have the same value, assign them the midrank, that is, the average of the ranks that would otherwise be assigned to those data.

5. If some of the reference-area and/or cleanup-unit data are less-than data, i.e., data less than the limit of detection, consider these less-than data to be tied at a value less than the smallest measured (detected) value in the combined data set. Assign the midrank for the group of less-than data to each less-than datum. For example, if there were 10 less-than data among the reference and cleanup-unit measurements, they would each receive the rank 5.5, which is the average of the ranks from 1 to 10. The assumption that all less-than measurements are less than the smallest detected measurement should not be made lightly because it may not be true for some pollution parameters, as pointed out by Lambert et al. (1991). However, the development of statistical testing procedures to handle this situation are beyond the scope of this document.

   The above procedure is applicable when all measurements have the same limit of detection. When there are multiple limits of detection, the adjustments given in Millard and Deveral (1988) may be used.

   Do not compute the WRS test if more than 40% of either the reference-area or cleanup unit measurements are less-than values. However, still conduct the Quantile test described in Chapter 7.

6. Sum the ranks of the n samples from the cleanup unit. Denote this sum by W_{rs}.

7. If both m and n are less than or equal to 10 and no ties are present, conduct the test of H_o versus H_a (Equation 6.2) by comparing W_{rs} to the appropriate critical value in Table A.5 in Hollander and Wolfe (1973). Then go to Step 12 below.

8. If both m and n are greater than 10 go to Step 9. If m is less than 10 and n is greater than 10, or if n is less than 10 and m is greater than 10, or if both m and n are less than or equal to 10 and ties are present, then consult a statistician to generate the required tables.

9. If both m and n are greater than 10 and ties are not present, compute Equation 6.12 and go to Step 11.

BOX 6.3

EXAMPLE 6.3

USING FIGURE 6.2 TO COMPUTE THE NUMBER OF SAMPLES NEEDED FOR THE WILCOXON RANK SUM TEST WHEN ONLY ONE CLEANUP UNIT WILL BE COMPARED WITH THE REFERENCE AREA

1. State the question:

   How many samples are required to test H_o versus H_a (Equation 6.2) using the WRS test with power 0.70 when we require a Type I error rate of $\alpha = 0.05$ and when $\Delta/\sigma = 0.95$, i.e., when $P_r = 0.75$ (from Table 6.4)? Assume the reference-area and cleanup-unit distributions are normal with the same σ. Suppose we expect about 10% of the data to be missing or unusable.

2. Specifications given in the question.

   $1 - \beta = 0.70$ $\quad\Delta/\sigma = 0.95$
   $\alpha = 0.05$ $\quad R = 0.10$
   $c = 0.50$ (from Equation 6.6)

3. From Figure 6.2, using the line for $\alpha = 0.05$ and $1 - \beta = 0.70$, which is the second light line from the left, at the point $P_r = 0.75$ gives

   $N = 25$

   which is divided by $1 - R = 0.90$ to obtain the final $N = 27.7$ or 28.

4. Then, $m = n = 0.5*28 = 14$, which are the same results obtained in Box 6.1 using Equation 6.3.

BOX 6.4

EXAMPLE 6.4

USING FIGURE 6.2 TO COMPUTE THE NUMBER OF SAMPLES NEEDED FOR THE WILCOXON RANK SUM TEST WHEN TWO CLEANUP UNITS WILL BE COMPARED WITH THE REFERENCE AREA

1. State the question:

   How many samples are required to test H_0 versus H_a using the WRS test with power 0.80 when we require a Type I error rate of $\alpha = 0.05$, and when $\Delta/\sigma = 0.74$ or $P_r = 0.70$ (from Table 6.4)? We assume the reference-area and the two cleanup-unit distributions are normal with the same σ. Suppose we expect about 5% of the data to be missing or unusable.

2. Specifications given in the question:

   $1 - \beta = 0.80$ $\quad\Delta/\sigma = 0.74$
   $\alpha = 0.05$ $\quad R = 0.05$
   $c = 0.59$ (from Equation 6.6)

3. From Figure 6.2, using the line for $\alpha = 0.05$ and $1 - \beta = 0.80$, which is the third light line from the left, at the point $P_r = 0.70$ gives $N = 53$.

4. Compute the product FN, where F is computed using Equation 6.11.

   $F = 0.25/(0.59*0.41) = 1.033$.

   $FN = 1.033*N = 1.033*53 = 54.75$.

5. Compute $FN/(1-R)$ to obtain the final N.

   $FN/(1-R) = 54.75/0.95 = 57.63$.

6. Compute $m = cN$ and $n = (1-c)N$.

   $m = 0.59*N = 0.59*57.63 = 34.002$ or 35
   $n_1 = n_2 = 0.41*N = 0.41*57.63 = 23.63$ or 24

$$\boxed{Z_{rs} = \frac{W_{rs} - n(N + 1)/2}{[mn(N+1)/12]^{1/2}}} \qquad (6.12)$$

10. If both m and n are greater than 10 and ties are present, compute

$$\boxed{Z_{rs} = \frac{W_{rs} - n(N+1)/2}{\left\{ (nm/12) \left[N+1 - \sum_{j=1}^{g} t_j(t_j^2-1)/N(N-1) \right] \right\}^{1/2}}} \qquad (6.13)$$

where g is the number of tied groups and t_j is the number of tied measurements in the jth group.

11. Reject H_o (cleanup standard attained) and accept H_a (cleanup standard not attained) if Z_{rs} (from Equation 6.12 or 6.13, whichever was used) is greater than or equal to $Z_{1-\alpha}$, where $Z_{1-\alpha}$ (from Table A.1) is the value that cuts off 100α% of the upper tail of the standard normal distribution.

12. If H_o is not rejected, conduct the Quantile test (Chapter 7). Also, compare each measurement from the cleanup unit to the hot measurement value, H_m. If any measurement exceeds H_m, then additional remedial action is needed at least locally (see Section 4.4.3).

In Example 6.5 (Box 6.5), the WRS test indicated the cleanup unit had not attained the cleanup standard of $P_r = 1/2$. This test result occurred because most of the small ranks were for the reference area and most of the large ranks were for the cleanup unit. Hence, W_{rs} was large enough for H_o to be rejected.

In Example 6.6 (Box 6.6), the WRS test indicated that the H_o: $P_r = 1/2$ cannot be rejected even though 14 cleanup-site measurements exceeded the largest reference-area measurement. In this example, the WRS test did not reject H_o because the reference-area measurements fell in the middle of the distribution of the cleanup-unit measurements. Hence, the cleanup unit had small as well as large ranks so that W_{rs} was not large enough to reject H_o. This example illustrates why it is necessary to also conduct the H_m (hot-measurement) comparison (Section 4.4.3) and the Quantile test (Chapter 7). Example 6.6 also illustrates the need to have statistical software to compute the WRS test when the number of measurements is large. Hand calculations become tedious and prone to error.

Examples 6.5 and 6.6 illustrate that the WRS test can be conducted even when less-than data are present. As a general guideline, the WRS test should not be used if more than 40% of either reference-area and cleanup-unit measurements are less-than data. However, the Quantile test (Chapter 7) can still be used in that situation.

6.4 The Two-Sample t Test

If the distribution of measurements for both the reference area and the cleanup unit are normally (Guassian) distributed and if no measurements are below the limit of detection, then the two-sample t test (Snedecor and Cochran 1980, pp. 89-98) could be used in place of the WRS test. However, the WRS test is preferred to the t test because it should have about the same or more power than the t test for most types of distributions. Lehmann (1975, pp. 76-81) compares the power of the WRS test and the two-sample t test when no measurements below the limit of detection are present. Helsel and Hirsch (1987) discuss the power of the WRS test when data less than the limit of detection are present. Further discussion of power is given here in Chapter 7.

6.5 Summary

This chapter describes and illustrates how to use the Wilcoxon Rank Sum (WRS) test to evaluate whether a cleanup unit has attained the reference-based cleanup standard. The WRS test is used to decide whether to reject

H_o: The remediated cleanup unit has attained the reference-based cleanup standard

and accept

H_a: The remediated cleanup unit has not attained the reference-based cleanup standard

The number of samples required for the WRS test may be determined using Equations 6.3, 6.4, and 6.5. The allocation of samples to the reference area and the cleanup unit can be approximated using Equation 6.6 or 6.7. Equation 6.6 is used if the standard deviations of measurements in the reference area and the applicable cleanup unit are equal. Equation 6.7 is used for the unequal case.

The number of samples may also be obtained using the curves in Figure 6.2 for the special case of $m = n$ if the reference-area and cleanup-unit measurements are normally distributed and each distribution has the same standard deviation, σ.

A value for the parameter P_r must be specified in Equation 6.3 to determine the required number of samples. Three ways of specifying this value of P_r are provided:

- direct specification of a value of P_r

- by first specifying the odds ratio, d, and converting d to P_r using Equation 6.9
- by first specifying the amount of relative shift, Δ/σ, in the distribution of cleanup-unit measurements to the right of the reference-area distribution, and then using Equation 6.10 to determine P_r.

The WRS test statistic is computed using Equation 6.12 or 6.13. Equation 6.13 is used when tied measurements are present.

If some of the reference-area and/or cleanup-unit measurements are less-than data, the WRS test can still be computed by considering these less-than data to be tied at a value less than the smallest measured value in the combined data set. The WRS test should not be computed if more than 40% of either the reference-area or cleanup unit measurements are less-than values. However, the Quantile test described in Chapter 7 can still be conducted.

The two-sample t test can be used in place of the WRS test if the data are normally distributed and if no measurements are below the limit of detection.

BOX 6.5

EXAMPLE 6.5

TESTING PROCEDURE FOR THE WILCOXON RANK SUM TEST

1. Suppose that the number of samples was determined using the specification in Example 1 (Box 6.1), namely,

$1 - \beta = 0.70$
$\alpha = 0.05$
$c = 0.50$
$P_r = 0.75$
$R = 0.10$

For these specifications we found that m = n = 14.

2. Rank the reference-area and cleanup-unit measurements from 1 to 28, arranging the data and their ranks as illustrated. Measurements below the limit of detection are denoted by ND and assumed to be less than the smallest value reported for the combined data sets. The data are lead measurements (mg/Kg).

Reference Area		Cleanup Unit	
Data	Rank	Data	Rank
ND	3	ND	3
ND	3		
ND	3		
ND	3		
39	6		
		48	7
49	8		
		51	9
53	10		
59	11		
61	12		
65	13		
67	14		
70	15		
72	16		
75	17		

Continued on next page

BOX 6.5 (Continued)

Reference Area		Cleanup Unit	
Data	Rank	Data	Rank
		80	18
		82	19
		89	20
		100	21
		150	22
		164	23
		193	24
		208	25
		257	26
		265	27
		705	28

$$W_{rs} = 272$$

3. The sum of the ranks of the cleanup unit is

$$W_{rs} = 3 + 7 \ldots + 27 + 28 = 272.$$

4. Compute Z_{rs} using Equation 6.13 because ties are present. There are $t = 5$ tied values for the $g = 1$ group of ties (ND values). We obtained:

$$Z_{rs} = \frac{272 - 14(28 + 1)/2}{\left\{ (14*14/12) \left[28 + 1 - 5(5*5 - 1)/28(28 - 1) \right] \right\}^{1/2}}$$

$$= \frac{69}{21.704} = 3.18$$

5. From the standard normal distribution table (Table A.1) we find that $Z_{1-\alpha} = 1.645$ for $\alpha = 0.05$ ($\alpha = 0.05$, the Type I error rate for the test, was specified in Step 1 above). Since $3.18 > 1.645$, we reject the null hypothesis H_o: $P_r = 1/2$ and accept the alternative hypothesis H_a: $P_r > 1/2$.

6. Conclusion:

The cleanup unit does not attain the cleanup standard of $P_r = 1/2$.

BOX 6.6

EXAMPLE 6.6

TESTING PROCEDURE FOR THE WILCOXON RANK SUM TEST

This example is based on measurements of 1,2,3,4-Tetrachlorobenzene (TcCB) (ppb) taken at a contaminated site and a site-specific reference area. There are $m = 47$ measurements in the reference area and $n = 77$ measurements in the cleanup unit for a total of 124 measurements. Although the samples were not located on a triangular grid, we shall assume here that the data are representative of the two areas. Although m and n were not determined using the procedure described in this document, i.e., by specifying values for α, $1 - \beta$, c, P_r, and R, the data are useful for illustrating computations. We shall set the Type I error rate, α, at 0.05.

1. Rank the reference-area and cleanup-unit measurements from 1 to 124.

Reference Area		Cleanup Unit		
Data	Rank	Data	Rank	t_j
		ND	1	
		0.09	2.5	2
		0.09	2.5	
		0.12	4.5	2
		0.12	4.5	
		0.14	6	
		0.16	7	
		0.17	9	3
		0.17	9	
		0.17	9	
		0.18	11	
		0.19	12	
		0.20	13.5	2
		0.20	13.5	
		0.21	15.5	2
		0.21	15.5	
0.22	18.5	0.22	18.5	4
		0.22	18.5	
		0.22	18.5	
0.23	21.5	0.23	21.5	2

Continued on next page

BOX 6.6 (CONTINUED)

Reference Area		Cleanup Unit		
Data	Rank	Data	Rank	t_j
		0.24	23	
		0.25	25.5	4
		0.25	25.5	
		0.25	25.5	
0.26	28.5	0.26	28.5	2
0.27	30			
0.28	32.5	0.28	32.5	4
0.28	32.5	0.28	32.5	
0.29	35.5	0.29	35.5	2
		0.31	37	
0.33	39.5	0.33	39.5	4
		0.33	39.5	
		0.33	39.5	3
0.34	42.5	0.34	42.5	2
0.35	44			
0.38	46.5	0.37	45	
0.39	49	0.38	46.5	2
0.39	49	0.39	49	3
		0.40	51	
0.42	52.5			2
0.42	52.5			
0.43	55	0.43	55	3
		0.43	55	
0.45	57			
0.46	58			
		0.47	59	
0.48	61	0.48	61	3
		0.48	61	
		0.49	63	
0.50	64.5			2
0.50	64.5			
0.51	67	0.51	67	3
		0.51	67	
0.52	69			
0.54	70.5	0.54	70.5	2
0.56	72.5			2
0.56	72.5			

Continued on next page

BOX 6.6 (CONTINUED)

Reference Area		Cleanup Unit		
Data	Rank	Data	Rank	t_j
0.57	74.5			2
0.57	74.5			
0.60	76.5	0.60	76.5	2
		0.61	78	
0.62	79.5	0.62	79.5	2
0.63	81			
0.67	82			
0.69	83			
0.72	84			
0.74	85			
		0.75	86	
0.76	87			
0.79	88			
0.81	89			
0.82	90.5	0.82	90.5	2
0.84	92			
		0.85	93	
0.89	94			
		0.92	95	
		0.94	96	
		1.05	97	
		1.10	98.5	2
		1.10	98.5	
1.11	100			
1.13	101			
1.14	102.5			
1.14	102.5			
		1.19	104	
1.20	105			
		1.22	106	
1.33	107.5	1.33	107.5	2
		1.39	109.5	2
		1.39	109.5	
		1.52	111	
		1.53	112	
		1.73	113	

Continued on next page

BOX 6.6 (CONTINUED)

Reference Area		Cleanup Unit		
Data	Rank	Data	Rank	t_j
		2.35	114	
		2.46	115	
		2.59	116	
		2.61	117	
		3.06	118	
		3.29	119	
		5.56	120	
		6.61	121	
		18.40	122	
		51.97	123	
		168.64	124	
		W_{rs} =	4585	

2. The sum of the ranks of the cleanup unit is

$$W_{rs} = 1 + 2.5 + 2.5 \ldots + 123 + 124 = 4585.$$

Note: If the ranks assigned to the m samples from the reference area are summed and denoted by W_{rb}, then

$$W_{rb} + W_{rs} = N(N + 1)/2.$$

In this example it is less effort to calculate W_{rb} and compute

$$W_{rs} = N(N + 1)/2 - W_{rb} = 124*125/2 - 3165$$

$$= 4585$$

rather than compute W_{rs} directly as was done above.

3. Compute Z_{rs} using Equation 6.13. There are g = 30 groups of ties: 21 groups with $t_j = 2$; 5 groups with $t_j = 3$; and 4 groups with $t_j = 4$. Therefore,

t_j	Number of Groups	$t_j(t_j^2 - 1)$	Product of Column 2 and Column 3
2	21	6	126
3	5	24	120
4	4	60	240
			Sum = 486

Continued on next page

BOX 6.6 (Continued)

Therefore, $\sum t_j(t_j^2 - 1)/2 = 486$. Therefore,

$$Z_{rs} = \frac{4585 - 77(124 + 1)/2}{\left\{ (77*47/12) \left[124 + 1 - 486/(124(124-1)) \right] \right\}^{1/2}}$$

$$= \frac{-227.5}{194.13}$$

$$= -1.17$$

4. From Table A.1 we find that $Z_{0.95} = 1.645$. Since -1.17 is not greater than 1.645, we cannot reject the null hypothesis H_o: $P_r = 1/2$.

5. Conclusion: There is no statistical evidence that the cleanup unit has not attained the cleanup standard of $P_r = 1/2$.

6. Conduct the Quantile test (conducted in Box 7.5, Chapter 7).

7. Determine if any measurements are greater than H_m. If so, additional remedial action is required at least locally around the sampling locations for those samples.

CHAPTER 7. QUANTILE TEST

In this chapter we show how to use the Quantile test (Johnson et al. 1987) to decide if the cleanup unit has attained the reference-based cleanup standard. As indicated in Chapter 6, we recommend that both the WRS test and the Quantile test, as well as the hot-measurement comparison (Section 4.4.3), be performed for each cleanup unit. If one or more of these tests rejects the null hypothesis (that the cleanup standard is achieved) for a given cleanup unit, then the site-specific reference-based cleanup standard has not been attained for that unit. The Quantile test is more powerful than the WRS test for detecting when only one or a few small portions of the cleanup unit have concentrations larger than those in the reference area. Also, the Quantile test can be used when a large proportion of the data is below the limit of detection.

Briefly, the Quantile test is performed by first listing the combined reference-area and cleanup-unit measurements from smallest to largest as was done for the WRS test (Chapter 6). Then, among the largest r measurements of the combined data sets, a count is made of the number of measurements, k, that are from the cleanup unit. If k is sufficiently large, then we conclude that the cleanup unit has not attained the reference-area cleanup standard.

In Section 7.1, the null and alternative hypotheses that are used with the Quantile test are defined and illustrated. In Section 7.2 we describe and illustrate how to use a table look-up procedure to determine the number of samples and to conduct the test for the case of equal numbers of samples in the reference area and the cleanup unit. A procedure for conducting the Quantile test for an arbitrary number of reference-area and cleanup-unit measurements is given in Section 7.3. In Section 7.4, we compare the power of the WRS and Quantile tests to provide guidance on which test is most likely to detect non-attainment of the cleanup standard in various situations. A summary is provided in Section 7.5.

7.1 Hypotheses and the Cleanup Standard

As stated in Section 2.2, the hypotheses used in this document are:

(7.1)

H_o: Reference-Based Cleanup Standard Achieved

H_a: Reference-Based Cleanup Standard Not Achieved

where H_o is assumed to be true unless the test indicates H_o should be rejected in favor of H_a.

When using the Quantile test, the above hypotheses are restated as:

$$\boxed{\begin{array}{l} H_o: \epsilon = 0,\ \Delta/\sigma = 0 \\ \\ H_a: \epsilon > 0,\ \Delta/\sigma > 0 \end{array}} \tag{7.2}$$

where

ϵ = the proportion of the soil in the cleanup unit that has not been remediated to reference-area levels

Δ/σ = amount (in units of standard deviation, σ) that the distribution of $100\epsilon\%$ of the measurements in the remediated cleanup unit is shifted to the right (to higher measurements) of the distribution in the reference area.

Please note that the relative shift, Δ/σ, is also used for the WRS test (Section 6.2.2.2). However, Δ/σ for the WRS test is applicable to the entire distribution of measurements in the cleanup unit rather than to only a proportion ϵ of the measurements.

The cleanup standard for the Quantile test is the value of ϵ and Δ/σ given in the H_o. Hence, the cleanup standard is $\epsilon = 0$ and $\Delta/\sigma = 0$, i.e., that all the cleanup-unit soil has been remediated such that the distribution of measurements for a given pollution parameter is the same in both the cleanup unit and the applicable reference area. The cleanup unit has not attained the reference-based cleanup standard for a given pollution parameter if any portion of the soil in the cleanup unit has concentrations such that the distribution of measurements for the unit is significantly shifted to the right of the reference-area distribution.

7.1.1 Examples of Distributions

Figures 7.1 and 7.2 illustrate the distribution of measurements for a hypothetical pollution parameter in a remediated cleanup unit and the reference area to which it is being compared. In Figure 7.1, $\epsilon = 0.10$ and $\Delta/\sigma = 4$, i.e., the measurements of the pollution parameter in $100\epsilon\% = 100(0.10)\% = 10\%$ of the cleanup unit have a distribution that is shifted to the right of the distribution of that pollution parameter in the reference area by $\Delta/\sigma = 4$ standard-deviation units. As seen in Figure 7.1, when Δ/σ is this large, the distribution of measurements for the entire cleanup unit has a distinct bimodal appearance. The Quantile test has more power than the WRS test for this situation.

In Figure 7.2, $\epsilon = 0.25$ and $\Delta/\sigma = 1$, i.e., the measurements in $100(0.25)\% = 25\%$ of the cleanup unit have a distribution that is shifted to the right of that of the reference area by $\Delta/\sigma = 1$ standard-deviation unit. Figure 7.2 illustrates that when Δ/σ is small, the distribution of

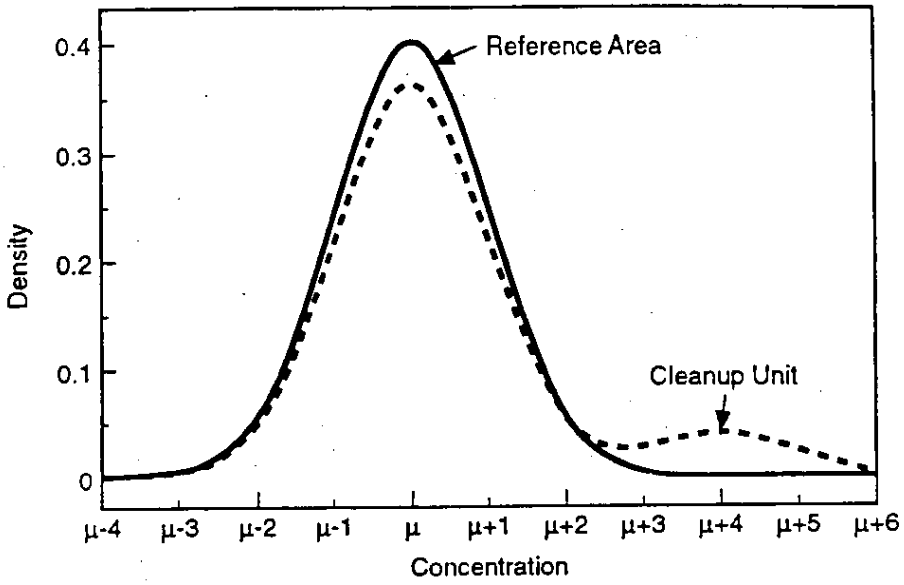


FIGURE 7.1. Hypothetical Distribution of Measurements for a Pollution Parameter in the Reference Area and for a Remediated Cleanup Unit. $\epsilon = 0.10$ and $\Delta/\sigma = 4$ for the Cleanup Unit.

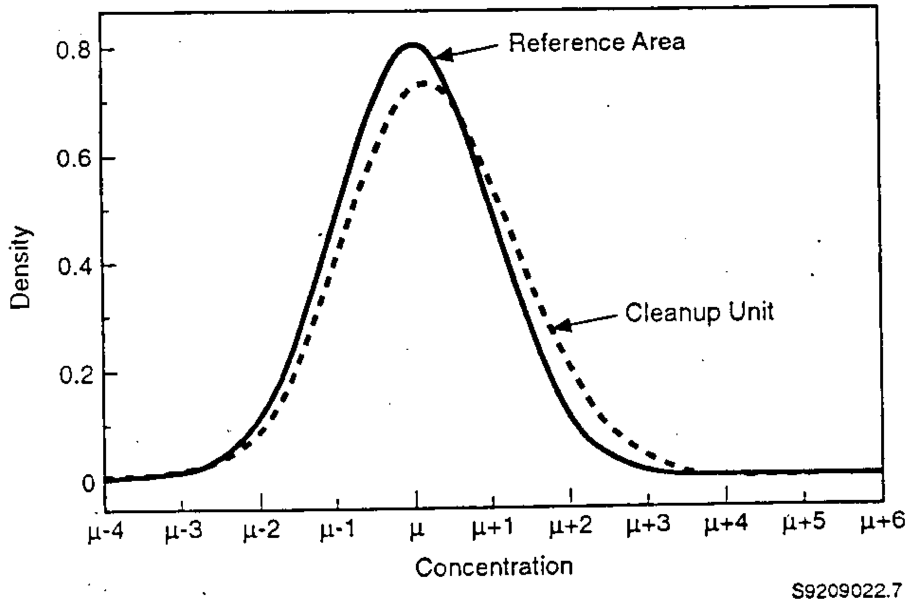


FIGURE 7.2. Hypothetical Distribution of Measurements for a Pollution Parameter in the Reference Area and for a Remediated Cleanup Unit. $\epsilon = 0.25$ and $\Delta/\sigma = 1$ for the Cleanup Unit.

measurements for the entire cleanup unit does not have a bimodal appearance. The WRS test has more power than the Quantile test for this situation.

When $\epsilon = 1$, then the shape of the distribution of measurements in the cleanup unit is the same as that for the distribution in the reference area, but the former distribution is shifted to the right by the amount $\Delta/\sigma > 0$. In that case, and more generally whenever ϵ is close to 1, the WRS test will have more power than the Quantile test.

7.2 Determining the Number of Samples and Conducting the Quantile Test

The procedure for determining the number of samples and conducting the Quantile test for a given pollution parameter is described and illustrated in this section. This procedure uses Tables A.2, A.3, A.4, and A.5 in Appendix A. These tables give the power of the Quantile and WRS tests to reject H_o for different combinations of α, ϵ, Δ/σ, m, and n for the special case of m = n. (See Section 7.3 for unequal m and n.) The power required for the Quantile test is used to determine the number of samples needed for the Quantile test, as discussed below.

Tables A.2 through A.5 were obtained using computer simulations (10,000 iterations) for the case where the residual contamination is distributed at random throughout the cleanup unit. The reference-area and cleanup-unit measurements were assumed to be normally (Gaussian) distributed. In reality, of course, the measurements may not be Gaussian, and residual contamination may exist in local areas, strips, or spatial patterns depending on the particular cleanup method that was used. Hence, the power results in Tables A.2 through A.5 are approximate, as are the number of samples determined using those tables.

The power of the WRS test in Tables A.2 through A.5 is supplemental information that may be compared with the power of the Quantile test to determine which test has the most power for given parameter values (α, ϵ, Δ/σ, and m = n). See Section 7.4 for discussion.

The procedure for using Tables A.2 through A.5 to determine the number of required measurements (m = n) and to conduct the Quantile test for each cleanup unit and pollution parameter is as follows:

1. Specify the Type I error rate, α, required for the test. The available options in this document are α equal to 0.01, 0.025, 0.05 and 0.10.

   Note: Recall from Section 4.6 that the selected value of α for the Quantile test should be one half the Type I Error rate selected for the combined WRS and Quantile tests.

2. Specify the values of ϵ and Δ/σ that are important to detect.

3. Specify the required power of the Quantile test, $1 - \beta$, to detect the specified values of ϵ and Δ/σ.

4. Use Table A.2, A.3, A.4 or A.5 as appropriate to determine m_{rc}, r, and k, where

m_{rc} = number of measurements that are needed from both the reference area and the cleanup unit to yield the required power for the specified ϵ and Δ/σ ($m_{rc} = n = m$)

r = number of largest measurements among the $N = 2m_{rc}$ combined reference-area and cleanup-unit measurements that must be examined

k = number of measurements from the cleanup unit that are among the r largest measurements.

Table A.2 is used if $\alpha = 0.01$ was specified in Step 1. Table A.3, A.4, or A.5 is used if $\alpha = 0.025$, 0.05, or 0.10 was specified in Step 1.

Note: The actual α level for the Quantile test frequently is not equal to the nominal specified level. This discrepancy, which is usually small enough to be ignored in practice, occurs whenever there are no values of r and k for which the actual α level will equal the specified level. For example, suppose the desired (specified) α level is 0.01. Turning to Table A.2 we see that when $m_{rc} = 10$, $r = 5$, and $k = 5$, the actual α level for the Quantile test is 0.015 instead of 0.01, a difference of 0.005. For other combinations of m_{rc}, r, and k in Table A.2, the actual α level for the Quantile test is usually slightly different from the nominal 0.01, but the differences are very small.

5. Compute

$$m_f = \frac{m_{rc}}{1 - R}$$

= number of samples to collect in both the reference area and cleanup unit

where R is the rate of missing or unusable data that is expected to occur. (Recall from Section 3.10 that unusable data are those that are mislabeled, lost, held too long before analysis, or do not meet quality control standards. Note that measurements less than the limit of detection are "usable".)

6. Collect m_f samples in the reference area and m_f samples in the cleanup unit for a total of $N_f = 2m_f$ samples.

7. Measure each of the N_f samples for the required pollution parameter.

8. Order from smallest to largest the combined reference-area and cleanup-unit measurements for the pollution parameter. If measurements less than the limit of detection are present in either the reference-area or cleanup-unit data sets, consider them to have a value less than the rth largest measured value in the combined data set (counting down from the maximum measurement). If this assumption is not realistic, consult a statistician.

   Note: Recall that for the WRS test (Section 6.3), a more restrictive assumption was necessary, i.e., that measurements less than the limit of detection were assumed to be less than the smallest measured value in the combined data set. This assumption for the WRS test can be relaxed for the Quantile test because the latter test only uses the r largest measurements in the combined data set. If fewer than r measurements are greater than the limit of detection, then the Quantile test cannot be performed.

   Note: The actual number of usable measurements (which includes measurements less than the limit of detection) from the reference area and the cleanup-unit area that are ordered in Step 8 may be different from the m or m_f because of missing or unusable measurements. However, the values of r and k determined from Table A.2, A.3, A.4 or A.5 in Step 4 can still be used to conduct the test as long as the final number of usable measurements in each area does not differ from m by more than about 10%. If the deviation is greater than 10% the testing procedure in Section 7.3 may be used.

9. If the rth largest measurement (counting down from the largest measurement) is among a group of tied (equal-in-value) measurements, then increase r to include the entire set of tied measurements. Also increase k by the same amount. For example, suppose from Step 4 we have that r = 10 and k = 7. Suppose the 7th through 12th largest measurements (counting down from the maximum measurement) have the same value. Then we would increase r from 10 to 12 and increase k from 7 to 9.

   By increasing k by the same amount as r we are assured that α remains less than the specified alpha. However, it is possible that a smaller increase in k would result in larger power while still giving an α that was less than the specified alpha. The optimum value of k for a selected r can be determined by computing α using Equation 7.3 (Section 7.3.2) for different values of k. The optimum k is the largest k that still gives a computed (actual) α less than or equal to the specified α.

10. Reject H_o and accept H_a (Equation 7.2) if k or more of the largest r measurements in the combined reference-area and cleanup-unit data sets are from the cleanup unit. As indicated in Step 8 above, the Quantile test uses only the largest r measurements so that only r measurements must be greater than the limit of detection. However, the full set of

N_f samples must be collected and analyzed even though only the largest r are actually used by the Quantile test.

11. If H_o is rejected, the Quantile test has indicated that the remediated cleanup unit does not attain the reference-based cleanup standard ($\epsilon = 0$, $\Delta/\sigma = 0$) and that additional remedial action may be needed.

    If H_o is not rejected, conduct the WRS test and the hot-measurement (H_m) comparison.

Examples of this procedure are given in Box 7.1 and Box 7.2. The example in Box 7.1 is for the case of relatively large ϵ and small Δ/σ, i.e., when a large portion of the remediated cleanup unit is slightly contaminated above the reference-area standard. The example in Box 7.2 is for the case of small ϵ and large Δ/σ, i.e., when a small proportion of the cleanup unit is highly contaminated relative to reference-area concentrations.

Note: The values of r and k used in Tables A.2 through A.5 are not the only values that will achieve the desired α level for the Quantile test. Among all combinations of r and k that will achieve an α level test, the combination with the smallest value of r was selected for use in the tables. This smallest value of r was selected because it gave the highest power for the Quantile test.

7.3 Procedure for Conducting the Quantile Test for an Arbitrary Number of Samples

In this section we describe how to conduct the Quantile test for an arbitrary (not necessarily equal) number of measurements from the reference area and the cleanup unit. A simple but approximate table look-up procedure for conducting the test is described in Section 7.3.1. An exact procedure that requires computations is described in Section 7.3.2.

Recall that in Section 7.2 the required power of the Quantile test was used (in conjunction with specified α, ϵ and Δ/σ) to determine $m = n = m_{rc}$ (as well as r and k). However, in this section it is assumed that the data have already been collected and there is no opportunity or desire to collect additional data. Hence, there is no opportunity to determine m and n on the basis of required power. The reader is cautioned that conducting the Quantile test using whatever data is available may yield a Quantile test that has insufficient power. The main reason for including Section 7.3 in this document is to provide a method for conducting the Quantile test when m is not equal to n. Section 7.3 would not be needed if power tables similar to Tables A.2 through A.5 were available for when m is not equal to n.

7.3.1 Table Look-Up Procedure

A simple table look-up procedure for conducting the Quantile test when m and n are specified a priori is given in this section. It is assumed that m and n representative measurements have been obtained from the reference area and the cleanup unit, respectively. The procedure in this section is

BOX 7.1

EXAMPLE 7.1

NUMBER OF SAMPLES AND CONDUCTING THE QUANTILE TEST

1. State the goal:

   Suppose we want to collect enough samples to be able to test H_o: $\epsilon = 0$, $\Delta/\sigma = 0$ versus H_a: $\epsilon > 0$, $\Delta/\sigma > 0$ using the Quantile test so that the test has an approximate power $(1 - \beta)$ of at least 0.70 of detecting when 40% of the remediated cleanup unit has measurements with a distribution that is shifted to the right of the reference-area distribution by 1.5 standard-deviation units. Suppose we require a Type I error rate of $\alpha = 0.05$ for the test and we expect about 5% of the data to be missing or unusable.

2. Specifications given in the above goal statement:

   $\alpha = 0.05$ $\epsilon = 0.4$
   $1 - \beta = 0.70$ $\Delta/\sigma = 1.5$
   $R = 0.05$

3. Using Table A.4 (since $\alpha = 0.05$ was specified) we find by examining the approximate powers in the body of the table corresponding to $\Delta/\sigma = 1.5$ and $\epsilon = 0.40$ that $m = n = 50$, $r = 10$ and $k = 8$. Hence, 50 usable measurements are needed from the reference area and from the cleanup unit.

   The test consists of rejecting the H_o if $k = 8$ or more of the $r = 10$ largest measurements among the 100 measurements are from the cleanup unit.

4. Divide $m_{rc} = 50$ by $(1 - R) = 0.95$ to obtain $m_f = 52.6$, or 53.

5. Collect 53 samples in both the reference area and the cleanup unit.

6. Order the 106 measurements from smallest to largest. Assume that measurements less than the limit of detection are smaller than the rth largest measured value in the combined data set (counting down from the maximum measurements).

Continued on the next page.

BOX 7.1 (Continued)

7. If the rth largest measurement (counting from the largest measurement) is among a group of tied measurements, increase r and k accordingly as illustrated in Step 9 of Section 7.2.

8. Using these values of r and k, and the value of m and n, compute the actual α level of the Quantile test using Equation (7.3). If the actual α level is too far below the required α level (0.05 in this example), decrease k by one and recompute Equation (7.3). Continue in this way to find the smallest k for which Equation (7.3) does not exceed 0.05.

9. If the number of usable measurements in both the reference area and the cleanup unit is greater than (m - 0.10m) = 50 - 5 = 45, then reject H_o and accept H_a if k or more of the largest 10 of the m + n measurements are from the cleanup unit.

10. If the number of usable measurements in either area is less than 45, then use the testing procedure in Section 7.3.

BOX 7.2

EXAMPLE 7.2

NUMBER OF SAMPLES AND CONDUCTING THE QUANTILE TEST

1. State the Goal:

   Suppose we want to collect enough samples to be able to test H_o: $\epsilon = 0$, $\Delta/\sigma = 0$ versus H_a: $\epsilon > 0$, $\Delta/\sigma > 0$ using the Quantile test so that the test has a power of at least 0.70 of detecting when 10% of the remediated cleanup unit has measurements with a distribution that is shifted to the right of the background distribution by 4 standard-deviation units. Suppose we specify $\alpha = 0.05$ and expect about 5% missing or unusable data.

2. Specifications given in the goal statement:

   $\alpha = 0.05$
   $1 - \beta = 0.70$
   $R = 0.05$

   $\epsilon = 0.1$
   $\Delta/\sigma = 4.0$

3. Using Table A.4 (since $\alpha = 0.05$ was specified) we find by examining the approximate powers in the body of the table corresponding to $\epsilon = 0.10$ and $\Delta/\sigma = 4.0$ that $m = 75$, $r = 10$ and $k = 8$. The testing procedure is to obtain 75 usable measurements in both the reference area and the cleanup unit and to reject the H_o and accept the H_a if $k = 8$ or more of the $r = 10$ largest measurements among the 150 usable measurements are from the cleanup unit.

4. Divide $m_{rc} = 75$ by $1 - R = 0.95$ to obtain $m_f = 78.9$ or 79.

5. Collect $m_f = 79$ samples in both the reference area and the cleanup unit. Suppose 2 reference-area and 3 cleanup-unit samples are lost so that the number of usable measurements is 77 in the reference area and 76 in the cleanup unit.

Continued on the next page.

BOX 7.2 (Continued)

6. Use Equation (7.3) to compute the actual α level when $m = 77$, $n = 76$, $r = 10$, and $k = 8$ to make sure that the actual level is close to the required value, 0.05. If the difference is too large, change k by one and recompute α using Equation (7.3). Repeat this process until the actual α level is sufficiently close to the required level. ("Sufficiently close" is defined by the user.)

7. Order the 153 measurements from smallest to largest. Suppose there are no tied measurements.

8. Since fewer than 10% of the required 75 measurements were lost, reject H_o and accept H_a if k (determined in Step 6 above) or more of the largest $r = 10$ of the 153 measurements are from the cleanup unit.

approximate because the Type I error rate, α, of the test may not be exactly what is required. However, the difference between the actual and required levels will usually be small. Moreover, the exact α level may be computed as explained in Section 7.3.2.

The testing procedure is as follows:

1. Specify the required Type I error rate, α. The available options in this document are α equal to 0.01, 0.025, 0.05 and 0.10.

2. Turn to Table A.6, A.7, A.8, or A.9 in Appendix A if α is 0.01, 0.025, 0.05, or 0.10, respectively.

3. Enter the selected table with m and n (the number of reference-area and cleanup-unit measurements, respectively) to find

   - values of r and k needed for the Quantile test

   - actual α level for the test for these values of r and k (the actual α may differ slightly from the required α level in Step 1)

4. If the table has no values of r and k for the values of m and n, enter the table at the closest tabled values of m and n. In that case, the α level in the table will apply to the tabled values of m and n, not the actual values of m and n. However, the α level for the actual m and n can be computed using Equation (7.3).

5. Order from smallest to largest the combined $m + n = N$ reference-area and cleanup-unit measurements for the pollution parameter. If measurements less than the limit of detection are present in either data set, assume that their value is less than the rth largest measured value in the combined data set of N measurements (counting down from the maximum measurement). If fewer than r measurements are greater than the limit of detection, then the Quantile test cannot be performed.

6. If the rth largest measurement (counting down from the maximum measurement) is among a group of tied (equal-in-value) measurements, then increase r to include that entire set of tied measurements. Also increase k by the same amount. For example, suppose from Step 3 we have $r = 6$ and $k = 6$. Suppose the 5th through 8th largest measurements (counting down from the maximum measurement) have the same value. Then we would increase both r and k from 6 to 8. (See the note in Step 9 of Section 7.2.)

7. Count the number, k, of measurements from the cleanup unit that are among the r largest measurements of the ordered N measurements, where r and k were determined in Step 3 (or Step 6 if the rth largest measurement is among a group of tied measurements).

8. If the observed k (from Step 7) is greater than or equal to the tabled value of k, then reject H_0 and conclude that the cleanup unit has not attained the reference area cleanup standard ($\epsilon = 0$ and $\Delta/\sigma = 0$).

9. If H_o is not rejected, then do the WRS test and compare the hot-measurement standard, H_m, (see Section 4.4.3) with measurements from the remediated cleanup unit. If the WRS test indicates the H_o should be rejected, then additional remedial action may be necessary. If one or more cleanup-unit measurements exceed H_m, then additional remedial action is needed, at least in the local area (see Section 4.4.3).

This procedure is illustrated with an example in Box 7.3.

7.3.2 Computational Method

A method for conducting the Quantile test that provides a way of computing the actual α level that applies to the test is given in this section. This procedure allows one to change r and k so that the actual and required α levels are sufficiently close in value (see Step 4). The first three steps below are the same as in Section 7.3.1.

1. Specify the required Type I error rate, α. The available options in this document are α equal to 0.01, 0.025, 0.05 and 0.10.

2. Turn to Table A.6, A.7, A.8, or A.9 in Appendix A if α is 0.01, 0.025, 0.05, or 0.10, respectively.

3. Enter the selected table with m and n (the number of reference-area and cleanup-unit measurements, respectively) to find

   - values of r and k needed for the Quantile test
   - actual α level for the test for these values of r and k.

4. If the table has no values of r and k for the values of m and n in Step 3, enter the table at the closest tabled values of m and n. The α level given in the table along with r and k applies to the tabled values of m and n rather than to the actual values of m and n. Compute the actual level of α, i.e., that level of α that corresponds to the actual m and n:

Actual Type I Error

$$\alpha = \frac{\sum_{i=k}^{r} \binom{m+n-r}{n-i} \binom{r}{i}}{\binom{m+n}{n}}$$

BOX 7.3

EXAMPLE 7.3

TABLE LOOK-UP TESTING PROCEDURE FOR THE QUANTILE TEST

1. We illustrate the Quantile test using the lead measurements listed in Box 6.5 (Chapter 6). There are 14 lead measurements in both the reference area and the cleanup unit. Suppose we specify $\alpha = 0.05$ for this Quantile test.

2. Turn to Table A.8 (because the table is for $\alpha = 0.05$). We see that there are no entries in that table for $m = n = 14$. Hence, we enter the table with $n = m = 15$, the values closest to 14. For $n = m = 15$ we find $r = 4$ and $k = 4$. Hence, the test consists of rejecting the H_o if all 4 of the 4 largest measurements among the 28 measurements are from the cleanup unit.

3. The $N = 28$ largest measurements are ordered from smallest to largest in Box 6.5.

4. From Box 6.5, we see that all 4 of the $r = 4$ largest measurements are from the cleanup unit. That is, $k = 4$.

5. Conclusion:

   Because $k = 4$, we reject the H_o and conclude that the cleanup unit has not attained the cleanup standard of $\epsilon = 0$ and $\Delta/\sigma = 0$. The Type I error level of this test is approximately 0.05.

   Note: The exact Type I error level, α, for this test is not given in Table A.8 because the table does not provide r, k, and α for $m = n = 14$. However, the exact α level can be computed using Equation (7.3) in Section 7.3.2.

where m and n are the actual number of reference-area and cleanup-unit measurements, r and k are from Step 3 above, and

$$\binom{a}{b} = \frac{a!}{b!(a - b)!}$$

$$a! = a*(a-1)*(a-2)*...*2*1,$$

where a! is called "a factorial".

Note: If Equation (7.3) is calculated using a hand calculator, use the calculation procedure of multiplying fractions illustrated in Examples 7.4 and 7.5 (Boxes 7.4 and 7.5) to guard against calculator overflow. Factorials can be evaluated with the help of tables of the logarithms of factorials found in, e.g., Rohlf and Sokal (1981) and Pearson and Hartley (1962). To avoid tedious and error-prone calculations, it is best to use computer software to compute α, especially if k is substantially less than r. Examples of commercially available statistical software packages are SAS (1990), Minitab (1990) and SYSTAT (1990).

If the computed actual α [Equation (7.3)] is sufficiently close to the required α level, go to Step 5. If not, increase and/or decrease r and/or k by one unit and recompute the actual α [Equation (7.3)] in an attempt to find an actual α that is sufficiently close to the required α. On the basis of these computations, select the values of r and of k that give an actual α level closest to the required α level. Note that since r and k are discrete numbers, it is nearly impossible for the actual α level to exactly equal the required level.

5. Order from smallest to largest the combined m + n = N reference-area and cleanup-unit measurements for the pollution parameter. If measurements less than the limit of detection are present in either the data sets, assume that their value is less than the rth largest measured value in the combined data set of N measurements (counting down from the maximum measurement). If fewer than r measurements (from Step 3 or 4) are greater than the limit of detection, then the Quantile test cannot be performed.

6. If the rth largest measurement (counting down from the maximum measurement) is among a group of tied (equal-in-value) measurements, then increase r to include that entire set of tied measurements. Also increase k by the same amount. For example, suppose from Steps 3 or 4 we have r = 6 and k = 6. Suppose the 5th through 8th largest measurements (counting down from the maximum measurement) have the same value. Then we would increase both r and k from 6 to 8.

7. Count the number, k, of measurements from the cleanup unit that are among the r largest measurements of the ordered N measurements, where r was determined in Steps 3 or 4 (or Step 6 if the rth largest measurement is among a group of tied measurements).

8. If $r \leq 20$, go to Step 9. If $r > 20$, go to Step 10.

Note: Rather than use steps 9 through 13 below to determine whether to reject the H_o, one can use the simpler procedure in steps 7 through 9 in Section 7.3.1. However, Equation (7.4) or Equation (7.5) can be used to compute P (defined below). Reporting this P level provides more information than just a "reject H_o" or "do not reject H_o" statement.

9. Compute the probability, P, of obtaining a value of k as large or larger than the observed k if, in fact, the H_o [Equation 7.2)] is really true, i.e., if all of the soil in the cleanup unit has really been remediated to reference-area levels:

$$\boxed{P = \frac{\sum_{i=k}^{r} \binom{m+n-r}{n-i}\binom{r}{i}}{\binom{m+n}{n}}} \qquad (7.4)$$

where m and n are the actual number of reference-area and cleanup-unit measurements, and r and k are from Step 3, 4, or 6.

Go to Step 11.

10. Use the following procedure to determine the probability, P, of obtaining a value of k as large or larger than the observed k if the null hypothesis, H_o [Equation (7.2)] is really true.

Compute

$$\text{XBAR} = \frac{nr}{m+n}$$

= mean of the hypergeometric distribution

$$\text{SD} = \left[\frac{mnr\ (m+n-r)}{(m+n)^2\ (m+n-1)}\right]^{1/2} \qquad (7.5)$$

= standard deviation of the hypergeometric distribution, and

$$\boxed{Z = \frac{k - 0.5 - XBAR}{SD}}$$

Enter Table A.1 with the computed value of Z to determine P, as illustrated in Box 7.5.

11. Reject H_o and accept H_a if $P \leq$ actual α level. Do not reject H_o if $P >$ actual α level.

12. If H_o is rejected, conclude that the remediated cleanup unit does not attain the reference-area standard ($\epsilon = 0$, $\Delta/\sigma = 0$).

13. If H_o is not rejected, then do the WRS test and compare the hot-measurement standard H_m (see Section 4.4.3) with the measurements in the remediated cleanup unit. If the WRS test is significant, then some type of additional remedial action may be needed. If one or more cleanup-unit measurements exceed H_m, then additional remedial action is needed, at least in the local area (see Section 4.4.3).

The test procedures in this section are illustrated in Boxes 7.4, 7.5, and 7.6.

7.4 Considerations in Choosing Between the Quantile Test and the Wilcoxon Rank Sum Test

This document recommends that both the WRS and Quantile tests be conducted for each cleanup unit. In this section we compare the power of the WRS and Quantile tests to provide guidance on which test is most likely to detect non-attainment of the reference-based standard in various situations. We also discuss the difficulty in practice of choosing which test to use, which is the basis for our recommendation to always conduct both tests.

Figure 7.3 shows the power curves of the Quantile and WRS Tests when $\alpha = 0.05$ and $m = n = 50$. The power curves of the Quantile test are for when $r = 10$ and $k = 8$. As seen in Figure 7.3, the power of each test increases as ϵ or Δ/σ increase. However, the increase in power of the two tests occurs at different rates. For example, as indicated in Table 7.1 (from Figure 7.3), the power of 0.7 can be achieved for several different combinations of Δ/σ and ϵ.

TABLE 7.1 Some Values of Δ/σ and ϵ for Which the Power of the Quantile Test and the WRS Test is 0.70 (from Figure 7.3)

Δ/σ	ϵ	Test
4.0	0.15	Quantile
	0.22	WRS
3.0	0.16	Quantile
	0.26	WRS
2.0	0.24	Quantile
	0.30	WRS
1.5	0.35	WRS
	0.36	Quantile
1.0	0.48	WRS
	0.68	Quantile
0.5	0.89	WRS

The results in Table 7.1 show that when the area in the cleanup unit with residual contamination is small (ϵ small) and the level of contamination is high (Δ/σ high), the Quantile test has more power than the WRS test. However, when the area with residual contamination is large (ϵ large) and the level of contamination is small (Δ/σ small), then the WRS test has more power than the Quantile test. An examination of Tables A.2 through A.5 will further illustrate this effect. It should be noted that when both the area and level of residual contamination is small, neither test will have sufficient power to determine if the cleanup unit is not in compliance unless a very large number of samples (m and n both over 100) are taken. If both the area and level of residual contamination is large, then both the Quantile and WRS tests have sufficient power to detect when the cleanup standard for the cleanup unit has not been attained.

The difficulty in choosing between the Quantile and WRS Tests is in predicting the size (ϵ) of the area in the cleanup unit that has concentrations (Δ/σ) greater than in the reference area. If ϵ and Δ/σ cannot be predicted accurately, then we recommend that both tests be conducted. (Recall that the hot-measurement comparison in Section 4.4.3 is always conducted.) However, it is important to understand that when both tests are conducted on the same set of data, the overall α level for the two tests combined is almost double the α level for each individual test. For example, if both the Quantile and WRS tests are conducted at the $\alpha = 0.05$ level, the combined α level is increased to almost 0.10. This is the reason we recommend

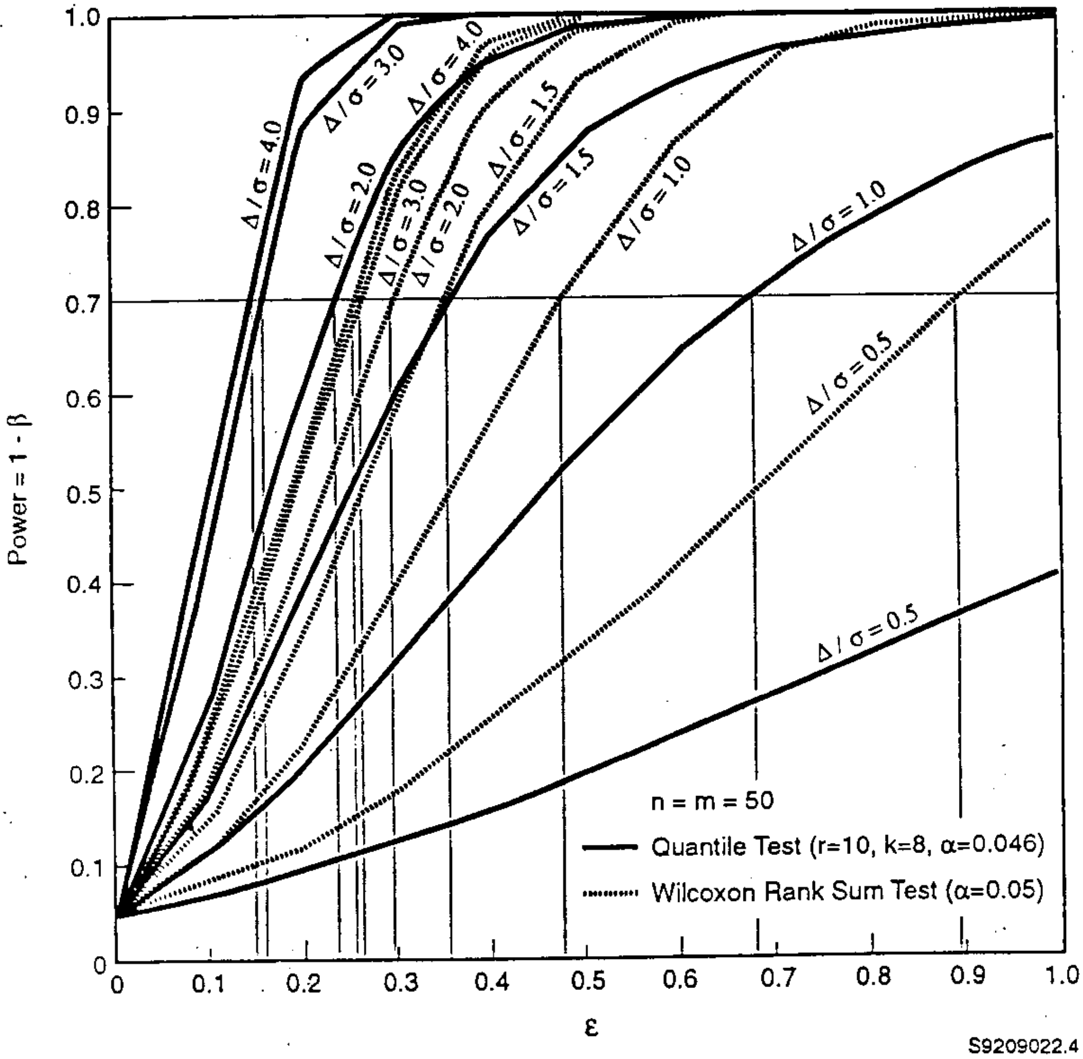


FIGURE 7.3. Power (1 - ß) of the Quantile Test and the Wilcoxon Rank Sum Test for Various Values of ϵ and Δ/σ when $m = n = 50$, $\alpha = 0.05$, $r = 10$, and $k = 8$.

that the overall α level for both tests combined should first be specified. Then both the WRS test and the Quantile test should be conducted at one-half that overall α level rate to achieve the desired overall α level rate.

Rather than computing both tests at the same α level, say $\alpha = 0.05$, which would achieve an overall α level of 0.10, we could use <u>either</u> the WRS test or the Quantile test at the $\alpha = 0.10$ level. The same overall α level of 0.10 would be achieved in both cases. But, is the combined power of both tests computed at the $\alpha = 0.05$ level greater than the power of either test conducted at the $\alpha = 0.10$ level? The answer to this question depends on whether the most powerful of the two tests is selected, which in turn depends on whether enough information about ϵ and Δ/σ is available to select the most powerful test.

As seen in Table 7.2 below, if the correct (most powerful) test is used at the $\alpha = 0.10$ level, then the power of that test is greater than the combined power of both tests conducted at the $\alpha = 0.05$ level. However, if the incorrect (less powerful) test is used at the $\alpha = 0.10$ level, then the power of that test is less than the combined power of both tests when each test is conducted at the $\alpha = 0.05$ level. Hence, conducting both tests guards against using the wrong (less powerful) test. But, when information about ϵ and Δ/σ is available for selecting the most powerful test, the practice of conducting both tests may decrease somewhat the chances of detecting non-attainment of the referance-based cleanup standard.

<u>TABLE 7.2</u> Power of the Quantile Test and the WRS Test and for Both Tests Combined when $n = m = 50$.

Correct Test	Δ/σ	ϵ	Combined Power When Each Test is Conducted at $\alpha = 0.05$	Power of Each Test Conducted at $\alpha = 0.10$	
				Quantile	WRS
WRS	0.5	1.0	0.786	0.486	0.877
Quantile	4.0	0.2	0.931	0.992	0.681

In conclusion:

- conduct both the Quantile and WRS tests to guard against using the wrong (less powerful) test
- if the expected size of ϵ and Δ/σ for the cleanup technology being used is known, then an alternative strategy is to
    - use the Quantile test in preference to the WRS test when it is known that the cleanup technology used at the site will result in a small ϵ and a large Δ/σ

- use the WRS test in preference to the Quantile test when it is known that the cleanup technology used at the site will result in a large ϵ and a small Δ/σ.

We recommend using both tests at least until substantial practical experience has been gained using the selected cleanup technology.

7.5 Summary

This chapter describes and illustrates how to use the Quantile test to evaluate whether a cleanup unit has attained the reference-based cleanup standard. The Quantile test is used to test

H_o: The remediated cleanup unit has attained the reference-based cleanup standard

versus

H_a: The remediated cleanup unit has not attained the reference-based cleanup standard

The number of samples required for the Quantile test can be determined using Tables A.2 through A.5 in Appendix A, which give the power of the Quantile test. These tables are for the case of equal number of samples in the reference area and the cleanup unit, i.e, for $m = n$. Tables A.6 through A.9 in Appendix A can be used to conduct the Quantile test when unequal numbers of samples have been collected and a required power has not been specified.

The Quantile test is more powerful than the WRS test at detecting when small areas (ϵ) in the remediated cleanup unit are contaminated at levels (Δ/σ) greater than in the reference area. Also, the Quantile test can be conducted even when a large proportion of the data set is below the limit of detection. This document recommends using both the Quantile and WRS tests to guard against a loss of power to detect when the reference-based cleanup standard has not been attained.

BOX 7.4

EXAMPLE 7.4

COMPUTING THE ACTUAL α LEVEL FOR THE QUANTILE TEST (CONTINUATION OF EXAMPLE 7.3)

1. In Example 7.3 it was necessary to enter Table A.8 with $m = n = 15$ rather than the actual number of measurements ($m = n = 14$). In Table A.8 for $m = n = 15$ we found $r = 4$, $k = 4$, and $\alpha = 0.05$. But this α level applies to $m = n = 15$, not $m = n = 14$. In accord with Step 4 in Section 7.3 we can use Equation (7.3) to compute the actual Type I error level, α, of the Quantile test conducted in Box 7.3.

2. Using $m = n = 14$ and $r = k = 4$ in Equation (7.3) we obtain

Actual Type I error level (α)

$$= \frac{\binom{28-4}{12-4}\binom{4}{4}}{\binom{28}{14}} = \frac{\binom{24}{10}}{\binom{28}{14}} = \frac{24!14!}{28!10!}$$

$$= \frac{14*13*12*11}{28*27*26*25}$$

$$= \frac{14}{28} * \frac{13}{27} * \frac{12}{26} * \frac{11}{25}$$

$$= 0.049$$

3. We see that the actual α level is 0.049, which is very close to the required α level of 0.05. Therefore, there is no need to change the values of r and k from those determined in Table A.8 using $m = n = 15$. Hence, the Quantile test procedure in Box 7.3 is appropriate.

BOX 7.5

EXAMPLE 7.5

CONDUCTING THE QUANTILE TEST

1. In this example, we illustrate the procedures for the Quantile test discussed in Section 7.3.2. We use the TcCB (ppb) measurements used in Box 6.6 (Chapter 6). There are $m = 47$ measurements from the reference area and $n = 77$ measurements from the cleanup unit, for a total of $N = 124$ measurements. Suppose we require that $\alpha = 0.01$ for the Quantile test, in which case Table A.6 in Appendix A is used for the test.

2. Table A.6 has no tabled values of r, k, and α for $m = 47$ and $n = 77$. Hence, the table is entered with $m = 45$ and $n = 75$, the closest values to m and n that are found in the table. For $m = 45$ and $n = 75$ we find that $r = 9$, $k = 9$, and $\alpha = 0.012$.

3. The α level of 0.012 in Step 2 above applies to $m = 45$, $m = 75$, $r = k = 9$ rather than to $m = 47$, $n = 77$, $r = k = 9$. The α level associated with the Quantile test for the latter set of parameters is computed using Equation (7.3) as follows:

   Actual Type I error level

$$= \frac{\binom{124 - 9}{77 - 9}\binom{9}{9}}{\binom{124}{77}} = \frac{\binom{115}{68}}{\binom{124}{77}} = \frac{115!77!}{68!124!}$$

$$= \frac{77*76*...*69}{124*123*...*116} = \frac{77}{124} * \frac{76}{123} *...* \frac{69}{116} = 0.0117 = 0.012$$

4. Hence, the actual α level for the Quantile test when $m = 47$, $n = 77$, $r = k = 9$ is 0.012, which is very close to the required level of 0.01. Therefore, we shall conduct the Quantile test using $r = k = 9$ even though they were determined by entering Table A.6 with $m = 45$ and $n = 75$.

Continued on the next page.

BOX 7.5 (Continued)

5. The 124 measurements are ordered from smallest to largest in Box 6.6 in Chapter 6. The largest $r = 9$ measurements are all from the cleanup unit. That is $k = 9$. Hence, the observed k and the k from Table A.6 are both equal to 9.

6. Using Steps 7 through 9 in Section 7.3.1 we reject H_o and conclude that the cleanup unit does not attain the reference-based cleanup standard. H_o is rejected because the observed k and the k from Table A.6 are equal in value.

7. The value of P, the probability of obtaining a value of k as large or larger than the observed k if the H_o is really true, is computed using Equation (7.4). We see that the computations for Equation (7.4) are identical to the computations given above in Step 3 for determining the actual α level. Hence, $P = 0.012$. The values of P and the actual α level are equal because the observed k and the k from Table A.6 were both equal to 9.

8. Following Step 11 in Section 7.3.2, we compare P with the actual α level. Since P = actual α level, we reject H_o and conclude that the cleanup unit does not attain the reference-based cleanup standard ($\epsilon = 0$, $\Delta/\sigma = 0$). As expected this conclusion is the same as obtained in Step 6 above.

9. Note that for these same data, the WRS test did <u>not</u> reject H_o (see Box 6.6, Chapter 6). The conclusions from the WRS and Quantile tests differ because the reference-area measurements fall in the middle of the distribution of the cleanup-unit measurements. The WRS test has less power than the Quantile test for this situation.

BOX 7.6

EXAMPLE 7.6

CONDUCTING THE QUANTILE TEST WHEN TIED DATA ARE PRESENT

This example is based on measurements of 2-Chloronaphthalene(CNP) (ppb) taken at a contaminated site and a site-specific reference area.

1. There are m = 77 measurements of CNP in the reference area and n = 58 measurements in the cleanup unit for a total of 135 measurements. We specify $\alpha = 0.05$.

2. Turn to Table A.8 and enter the table with m = 75 and n = 60, the values closests to m = 77 and n = 58. We find that r = 9, k = 7, and $\alpha = 0.05$.

3. Before conducting the Quantile test, we need to look at the data to see if there are tied valeus.

4. The largest 28 measurements in the combined reference-area and cleanup-unit data sets are shown below. The data are ordered from lowest to highest values. The 9th largest measurement (counting down from the maximum) is the 2nd in a group of 5 measurements with the same value (0.012 ppb). Hence, using Step 6 in Section 7.3.2, 23 increase r from 9 to 12, and increase k from 7 to 10.

Reference	
Data	Rank
.	.
.	.
0.10	111.5
0.10	111.5
0.10	111.5
0.10	111.5
0.10	111.5
0.11	119.5
0.11	119.5
0.11	119.5
0.11	119.5
0.12	126
0.12	126

Cleanup Unit	
Data	Rank
.	.
.	.
.	.
.	.
0.10	111.5
0.10	111.5
0.10	111.5
0.11	119.5
0.11	119.5
0.11	119.5
0.11	119.5
0.12	126
0.12	126

Continued on the next page

BOX 7.6 (Continued)

Reference Area		Cleanup Unit	
Data	Rank	Data	Rank
		0.12	126
		0.13	129
		0.14	130.5
		0.14	130.5
0.15	132		
0.16	133		
		0.19	134
		0.32	135

5. Now, calculate the actual α level of the Quantile test for $m = 77$, $n = 58$, $r = 12$ and $k = 10$ to see if that level is sufficiently close to the required 0.05. ("Sufficiently close" is defined by the user.) If not, decrease k by one and recompute the actual α level using Equation (7.3). If necessary, continue in this way until the value of k gives an actual α level that exceeds 0.05. Then increase k by 1. Applying this process yielded the following results:

k	Actual α Level
10	0.00341
9	0.02025
8	0.0759

Therefore, we select $k = 9$. Hence, the Quantile test will consist of rejecting H_o if 9 or more of the largest 12 measurements in the combined data sets are from the cleanup unit. The actual α level test is for this test is $\alpha = 0.020$.

6. The observed k from the above data is seen to be 8, which is less than 9. Therefore, we cannot reject H_o. That is, we cannot reject the hypothesis that the cleanup unit has attained the reference-based cleanup standard.

Continued on next page.

BOX 7.6 (Continued)

7. We may use Equation (7.4) to compute the probability, P, of obtaining a value of k as large or larger than the observed k if, in fact, the H_o is really true. P is computed using Equation (7.4) because $\leq$ 20. Using Equation (7.4) with m = 77, n = 58. r = 12, and k = 8 we compute P = 0.0759, which is greater than the α level, 0.020. From Step 11 in Section 7.3.2, we cannot reject H_o, as indicated in Step 6 above.

8.0 REFERENCES

Atwood, C.L, L.G. Blackwood, G.A. Harris, and C.A. Loehr. 1991. Recommended Methods for Statistical Analysis of Data Containing Less-Than-Detectable Measurements. EGG-SARE-9247, Rev. 1, Idaho National Engineering Laboratory, EG&G Idaho, Inc, Idaho Falls, Idaho.

Barnett, V., and T. Lewis. 1985. Outliers in Statistical Data. 2nd ed. Wiley, New York.

Beckman, R.J., and R.D. Cook. 1983. "Outlier..........." Technometrics 25:119-149.

Berry, B.J.L., and A.M. Baker. 1968. "Geographic Sampling." Spatial Analysis, eds. B.J.L. Berry and D.F. Marble. Prentice-Hall, Englewood Cliffs, New Jersey.

Bolgiano, N.C., G.P. Patil, and C. Taillie. 1990. "Spatial Statistics, Composite Sampling, and Related Issues in Site Characterization with Two Examples." In Proceedings of the Workshop on Superfund Hazardous Waste: Statistical Issues in Characterizing a Site: Protocols, Tools, and Research Needs, eds. H. Lacayo, R.J. Nadeau, G.P. Patil, and L. Zaragoza, pp. 79-117, Pennsylvania State University, Department of Statistics, University Park, Pennsylvania.

Brown, K.W., and S.C. Black. 1983. "Quality Assurance and Quality Control Data Validation Procedures Used for the Love Canal and Dallas Lead Soil Monitoring Programs." Environmental Monitoring and Assessment 3:113-122.

Cochran, W.G. 1977. Sampling Techniques. 3rd ed. Wiley, New York.

Cressie, N.A.C. 1991. Statistics for Spatial Data. Wiley, New York.

DOE. April 1992. Hanford Site Soil Background. DOE/RL-92-24, U.S. Department of Energy, Richland Field Office, Richland, Washington.

Duncan, A.J. 1962. "Bulk Sampling: Problems and Lines of Attack." Technometrics 4(2):319-344.

Dunnett, C.W. 1955. "A Multiple Comparison Procedure for Comparing Several Treatment with a Control." Journal of the American Statistical Association 50:1096-1121.

Dunnett, C.W. 1964. "New Tables for Multiple Comparisons with a Control." Biometrics 20:482-491. .

Elder, R.S., W.O. Thompson, and R.H. Myers. 1980. "Properties of Composite Sampling Procedures." Technometrics 22(2):179-186.

Garner, F.C. 1985. Comprehensive Scheme for Auditing Contract Laboratory Data [interim report]. Lockheed-EMSCO, Las Vegas, Nevada

Garner, F.C., G.L. Robertson, and L.R. Williams. 1988. "Composite Sampling for Environmental Monitoring." Principals of Environmental Sampling, pp. 363-374, American Chemical Society, Washington, D.C.

Gilbert, R.O. 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold, New York.

Gilbert, R.O., and R.R. Kinnison. 1981. "Statistical Methods for Estimating the Mean and Variance from Radionuclide Data Sets Containing Negative, Unreported or Less-Than Values." Health Physics 40:377-390.

Gilliom, R.J., and D.R. Helsel. 1986. "Estimation of Distributional Parameters for Censored Trace Level Water Quality Data, 1. Estimation Techniques." Water Resources Research 22(2):135-146.

Gleit, A. 1985. "Estimation for Small Normal Data Sets with Detection Limits." Environmental Science and Technology 19(12):1201-1206.

Hawkins, D.M. 1980. Identification of Outliers. Chapman and Hall, New York.

Helsel, D.R. 1990. "Less Than Obvious: Statistical Treatment of Data Below the Detection Limit." Environmental Science and Technology 24:1766-1774.

Helsel, D.R., and T.A. Cohn. 1988. "Estimation of Descriptive Statistics for Multiply Censored Water Quality Data." Water Resources Research 24(12):1997-2004.

Helsel, D.R., and R.J. Gilliom. 1986. "Estimation of Distributional Parameters for Censored Trace Level Water Quality Data, 2. Verification and Applications." Water Resources Research 22(2):147-155.

Helsel, D.R. and R. Hirsch. 1987. "Discussion of Applicability of the t test for Detecting Trends in Water Quality Variables." Water Resources Bulletin 24: 201-204.

Hochberg, Y., and A.C. Tamhane. 1987. Multiple Comparison Procedures. Wiley, New York.

Hollander, M., and D.A. Wolfe. 1973. Nonparametric Statistical Methods. Wiley, New York.

Isaaks, E.H., and R.M. Srivastava. 1989. An Introduction to Applied Geostatistics, Oxford University Press, New York.

Johnson, R.A., S. Verrill, and D.H. Moore II. 1987. "Two-Sample Rank Tests for Detecting Changes That Occur in a Small Proportion of the Treated Population." Biometrics 43:641-655.

Keith, L.H. 1991. Environmental Sampling and Analysis: A Practical Guide, Lewis Publishers, Chelsea, Michigan.

Kelso, G.L., and D.C. Cox. May 1986. Field Manual for Grid Sampling of PCB Spill Sites to Verify Cleanup. EPA-560/5-86-0, U.S. Environmental Protection Agency, Washington, D.C.

Lambert, D., B. Peterson, and I. Terpenning. 1991. "Nondetects, Detection Limits, and the Probability of Detection." Journal of the American Statistical Association 86:266-277.

Lehmann, E.L. 1975. NONPARAMETRICS: Statistical Methods Based on Ranks. Holden-Day, Inc., San Francisco, California.

Liggett, W. 1984. "Detecting Elevated Contamination by Comparisons with Background." Environmental Sampling for Hazardous Wastes, eds. G.E. Schweitzer and J.A. Santolucito, pp. 119-128. ACS Symposium Series 267, American Chemical Society, Washington, D.C.

Millard, S.P., and S.J. Deveral. 1988. "Nonparametric Statistical Methods for Comparing Two Sites Based on Data with Multiple Nondetect Limits." Water Resources Research 24(12):2087-2098.

Miller, R.C. 1981. Simultaneous Statistical Inference. 2nd ed. Springer-Verlag, New York.

MINITAB. 1990. MINITAB Statistical Software. Minitab, Inc., State College, Pennsylvania.

Neptune, D., E.P. Brantley, M.J. Messner, and D.I. Michael. 1990. "Quantitative Decision Making in Superfund: A Data Quality Objectives Case Study." Hazardous Materials Control 3(3):18-27.

Noether, G.E. 1987. "Sample size Determination for Some Common Nonparametric Tests." Journal of the American Statistical Association 82:645-647.

Pearson, E.S., and H.O. Hartley. 1962. Biometrika Tables for Statisticians, Volume I. 2nd ed. Cambridge University Press, Cambridge, England.

Rohlf, F.J., and R.R. Sokal. 1981. Statistical Tables. 2nd ed. Freeman and Company, San Francisco, California.

Rohde, C.A. 1976. "Composite Sampling." Biometrics 32:273-282.

Ryti, R.T., and D. Neptune. 1991. "Planning Issues for Superfund Site Remediation." Hazardous Materials Control 4(6):47-53.

SAS. 1990. Statistical Analysis System, Inc., Cary, North Carolina.

Schaeffer, D.J., and D.G. Janardan. 1978. "Theoretical Comparison of Grab and Composite Sampling Programs." Biometrical Journal 20:215-227.

Schaeffer, D.J., H.W. Derster, and D.G. Janardan. 1980. "Grab Versus Composite Sampling: A Primer for Managers and Engineers." Journal of Environmental Management 4:157-163.

Schwertman, N.C. 1985. "Multivariate Median and Rank Sum Tests," Encyclopedia of Statistical Sciences, Vol. 6, eds. S. Kotz, N.L. Johnson, and C.B. Read. Wiley, New York.

Singer, D.A. 1975. "Relative Efficiencies of Square and Triangular Grids in the Search for Elliptically Shaped Resource Targets." Journal of Research of the U.S. Geological Survey 3(2):163-167.

Snedecor, G.W., and W.G. Cochran. 1980. Statistical Methods. 7th ed. Iowa State University Press, Ames, Iowa.

Steel, R.G.D. 1959. "A Multiple Comparison Rank Sum Test: Treatments Versus Control." Biometrics 15:560-572.

SYSTAT. 1990. SYSTAT: The System for Statistics. SYSTAT, Inc., 1800 Sherman Ave., Evanston, Illinois.

Taylor, J.K. 1987. Quality Assurance of Chemical Measurements, Lewis Publishers, Chelsea, Michigan.

Taylor, J.K. and T.W. Stanley (eds). 1985. Quality Assurance for Environmental Measurements, ASTM Special Technical Publication 867, American Society for Testing and Materials, Philadelphia, Pennsylvania.

U.S. Environmental Protection Agency (USEPA). 1984. Soil Sampling Quality Assurance User's Guide. 1st ed. EPA 600/4-84-043, Washington, D.C.

U.S. Environmental Protection Agency (USEPA). 1987a. Data Quality Objectives for Remedial Response Activities: Development Process. EPA 540/G-87/003), Washington, D.C.

U.S. Environmental Protection Agency (USEPA). 1987b. Data Quality Objectives for Remedial Response Activities: Example Scenario RI/FS Activities at a Site with Contaminated Soils and Ground Water. EPA 540/G-87/004, Washington, D.C.

U.S. Environmental Protection Agency (USEPA). 1989a. Methods for Evaluating the Attainment of Cleanup Standards, Volume 1: Soils and Solid Media. EPA 230/02-89-042, Statistical Policy Branch, Washington, D.C.

U.S. Environmental Protection Agency (USEPA). 1992. Statistical Methods for Evaluating the Attainment of Superfund Cleanup Standards, Volume 2: Groundwater. DRAFT, Statistical Policy Branch, Washington, D.C.

U.S. Environmental Protection Agency (USEPA). 1989b. Risk Assessment Guidance for Superfund, Volume I: Human Health Evaluation Manual (Part A), Interium Final. EPA/540/1-89/002. Office of Emergency and Remedial Response, Washington, D.C.

APPENDIX A

STATISTICAL TABLES

APPENDIX A

STATISTICAL TABLES

TABLE A.1. Cumulative Standard Normal Distribution (Values of the Probability ϕ Corresponding to the Value Z_ϕ of a Standard Normal Random Variable)

Z_ϕ	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5674	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Table A.2 Approximate Power and Number of Measurements for the Quantile and Wilcoxon Rank Sum (WRS) Tests for Type I Error Rate α = 0.01 for when m = n. m and n are the Number of Required Measurements from the Reference Area and the Cleanup Unit, respectively.

						Δ/σ							
Test	m=n	r	k	α	ε	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Quantile	10	5	5	0.015	0.1	0.018	0.025	0.029	0.036	0.038	0.045	0.043	0.050
					0.2	0.026	0.040	0.058	0.082	0.102	0.108	0.119	0.122
					0.3	0.032	0.054	0.096	0.146	0.200	0.233	0.264	0.278
					0.4	0.036	0.078	0.149	0.244	0.333	0.418	0.463	0.490
					0.5	0.043	0.100	0.211	0.349	0.495	0.598	0.663	0.697
					0.6	0.050	0.137	0.283	0.469	0.642	0.761	0.821	0.869
					0.7	0.063	0.169	0.359	0.569	0.750	0.875	0.935	0.955
					0.8	0.079	0.207	0.426	0.662	0.848	0.936	0.976	0.992
					0.9	0.080	0.250	0.500	0.745	0.896	0.970	0.993	0.997
					1.0	0.090	0.284	0.564	0.806	0.933	0.982	0.997	1.000
WRS				0.010	0.1	0.014	0.016	0.020	0.019	0.020	0.022	0.025	0.019
					0.2	0.016	0.025	0.030	0.043	0.047	0.050	0.049	0.051
					0.3	0.021	0.037	0.053	0.078	0.093	0.101	0.106	0.107
					0.4	0.026	0.052	0.099	0.132	0.165	0.185	0.197	0.196
					0.5	0.033	0.081	0.152	0.220	0.274	0.316	0.327	0.334
					0.6	0.039	0.118	0.234	0.333	0.438	0.486	0.499	0.514
					0.7	0.052	0.165	0.327	0.505	0.604	0.666	0.691	0.700
					0.8	0.058	0.212	0.458	0.676	0.790	0.835	0.865	0.873
					0.9	0.073	0.280	0.596	0.823	0.926	0.959	0.968	0.973
					1.0	0.089	0.380	0.751	0.946	0.995	1.000	1.000	1.000
Quantile	15	6	6	0.008	0.1	0.011	0.016	0.021	0.027	0.033	0.037	0.039	0.040
					0.2	0.015	0.027	0.047	0.074	0.103	0.129	0.147	0.157
					0.3	0.019	0.043	0.088	0.157	0.237	0.311	0.363	0.393
					0.4	0.024	0.064	0.146	0.272	0.416	0.540	0.623	0.668
					0.5	0.030	0.090	0.216	0.402	0.594	0.740	0.827	0.869
					0.6	0.036	0.121	0.294	0.527	0.737	0.872	0.938	0.964
					0.7	0.043	0.155	0.374	0.635	0.835	0.939	0.980	0.993
					0.8	0.051	0.193	0.450	0.720	0.894	0.969	0.993	0.999
					0.9	0.060	0.232	0.520	0.784	0.929	0.982	0.997	0.999
					1.0	0.070	0.272	0.581	0.831	0.950	0.989	0.998	1.000
WRS				0.010	0.1	0.012	0.017	0.021	0.022	0.029	0.027	0.026	0.027
					0.2	0.016	0.030	0.042	0.056	0.066	0.071	0.072	0.078
					0.3	0.024	0.049	0.089	0.120	0.144	0.158	0.170	0.166
					0.4	0.036	0.080	0.152	0.213	0.274	0.294	0.315	0.321
					0.5	0.042	0.123	0.251	0.356	0.442	0.495	0.514	0.525
					0.6	0.058	0.183	0.374	0.533	0.644	0.703	0.715	0.734
					0.7	0.071	0.258	0.512	0.722	0.825	0.868	0.885	0.900
					0.8	0.091	0.352	0.683	0.878	0.946	0.968	0.975	0.976
					0.9	0.112	0.457	0.821	0.968	0.993	0.998	0.999	1.000
					1.0	0.144	0.574	0.924	0.997	1.000	1.000	1.000	1.000

TABLE A.2 (Continued)

						Δ/σ							
Test	m=n	r	k	α	ε	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Quantile	20	6	6	0.010	0.1	0.014	0.020	0.030	0.042	0.055	0.065	0.071	0.075
					0.2	0.018	0.037	0.070	0.122	0.185	0.246	0.291	0.317
					0.3	0.024	0.059	0.133	0.251	0.392	0.520	0.608	0.658
					0.4	0.031	0.089	0.213	0.402	0.602	0.755	0.845	0.888
					0.5	0.038	0.124	0.302	0.544	0.759	0.891	0.953	0.976
					0.6	0.047	0.163	0.391	0.660	0.856	0.952	0.986	0.996
					0.7	0.056	0.205	0.474	0.746	0.911	0.976	0.995	0.999
					0.8	0.066	0.249	0.547	0.808	0.942	0.987	0.998	1.000
					0.9	0.077	0.292	0.610	0.852	0.960	0.992	0.999	1.000
					1.0	0.089	0.335	0.663	0.883	0.971	0.994	0.999	1.000
WRS				0.010	0.1	0.014	0.017	0.025	0.030	0.032	0.032	0.037	0.037
					0.2	0.018	0.036	0.055	0.076	0.086	0.096	0.105	0.100
					0.3	0.030	0.065	0.119	0.165	0.204	0.228	0.237	0.248
					0.4	0.040	0.109	0.221	0.314	0.377	0.420	0.432	0.449
					0.5	0.055	0.179	0.357	0.499	0.600	0.646	0.672	0.679
					0.6	0.074	0.259	0.511	0.704	0.802	0.838	0.859	0.867
					0.7	0.094	0.368	0.694	0.871	0.932	0.959	0.962	0.967
					0.8	0.123	0.483	0.838	0.958	0.988	0.995	0.996	0.997
					0.9	0.163	0.617	0.937	0.994	1.000	1.000	1.000	1.000
					1.0	0.194	0.741	0.983	1.000	1.000	1.000	1.000	1.000
Quantile	25	6	6	0.008	0.1	0.017	0.025	0.038	0.059	0.079	0.096	0.119	0.120
					0.2	0.024	0.045	0.091	0.170	0.266	0.368	0.445	0.490
					0.3	0.029	0.074	0.176	0.332	0.514	0.683	0.776	0.826
					0.4	0.037	0.107	0.272	0.503	0.723	0.866	0.940	0.970
					0.5	0.044	0.148	0.383	0.647	0.846	0.944	0.983	0.995
					0.6	0.055	0.193	0.453	0.739	0.907	0.978	0.995	0.999
					0.7	0.064	0.240	0.539	0.810	0.942	0.987	0.998	1.000
					0.8	0.082	0.288	0.609	0.857	0.961	0.992	0.998	1.000
					0.9	0.091	0.336	0.674	0.892	0.971	0.995	0.999	1.000
					1.0	0.105	0.380	0.715	0.909	0.978	0.997	0.999	1.000
WRS				0.010	0.1	0.017	0.022	0.028	0.037	0.038	0.037	0.038	0.039
					0.2	0.022	0.046	0.069	0.096	0.113	0.120	0.129	0.123
					0.3	0.033	0.083	0.150	0.218	0.262	0.297	0.313	0.307
					0.4	0.047	0.138	0.277	0.404	0.481	0.538	0.557	0.559
					0.5	0.069	0.229	0.448	0.620	0.722	0.761	0.791	0.796
					0.6	0.088	0.338	0.639	0.820	0.889	0.923	0.937	0.940
					0.7	0.126	0.469	0.804	0.935	0.976	0.989	0.991	0.991
					0.8	0.153	0.616	0.920	0.990	0.997	0.999	0.999	1.000
					0.9	0.207	0.738	0.977	0.999	1.000	1.000	1.000	1.000
					1.0	0.262	0.841	0.996	1.000	1.000	1.000	1.000	1.000

TABLE A.2 (Continued)

						Δ/σ							
Test	m=n	r	k	α	ϵ	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Quantile	30	6	6	0.013	0.1	0.018	0.024	0.052	0.069	0.108	0.136	0.171	0.187
					0.2	0.024	0.055	0.115	0.218	0.357	0.494	0.584	0.644
					0.3	0.028	0.085	0.214	0.410	0.623	0.785	0.881	0.923
					0.4	0.038	0.134	0.316	0.581	0.808	0.928	0.976	0.991
					0.5	0.051	0.169	0.419	0.702	0.895	0.972	0.993	0.998
					0.6	0.060	0.233	0.521	0.790	0.931	0.984	0.998	0.999
					0.7	0.074	0.279	0.592	0.839	0.959	0.994	0.999	1.000
					0.8	0.088	0.324	0.659	0.885	0.974	0.996	0.999	1.000
					0.9	0.102	0.373	0.701	0.906	0.979	0.997	0.999	1.000
					1.0	0.117	0.416	0.755	0.923	0.986	0.998	1.000	1.000
WRS				0.010	0.1	0.016	0.022	0.033	0.038	0.038	0.042	0.049	0.045
					0.2	0.023	0.050	0.075	0.104	0.134	0.143	0.149	0.151
					0.3	0.036	0.097	0.173	0.260	0.320	0.355	0.361	0.362
					0.4	0.054	0.165	0.335	0.476	0.563	0.607	0.637	0.643
					0.5	0.079	0.280	0.527	0.714	0.795	0.836	0.863	0.869
					0.6	0.106	0.401	0.719	0.884	0.948	0.962	0.971	0.971
					0.7	0.145	0.552	0.875	0.973	0.992	0.996	0.998	0.998
					0.8	0.182	0.696	0.962	0.997	0.999	1.000	1.000	1.000
					0.9	0.248	0.822	0.993	1.000	1.000	1.000	1.000	1.000
					1.0	0.310	0.908	1.000	1.000	1.000	1.000	1.000	1.000
Quantile	40	15	12	0.010	0.1	0.016	0.026	0.043	0.062	0.078	0.089	0.094	0.095
					0.2	0.024	0.059	0.128	0.224	0.318	0.384	0.417	0.430
					0.3	0.035	0.113	0.277	0.491	0.669	0.769	0.814	0.830
					0.4	0.049	0.188	0.463	0.744	0.901	0.958	0.975	0.980
					0.5	0.067	0.280	0.641	0.898	0.981	0.996	0.999	0.999
					0.6	0.088	0.382	0.779	0.965	0.997	1.000	1.000	1.000
					0.7	0.112	0.484	0.872	0.989	1.000	1.000	1.000	1.000
					0.8	0.140	0.579	0.928	0.996	1.000	1.000	1.000	1.000
					0.9	0.171	0.664	0.960	0.999	1.000	1.000	1.000	1.000
					1.0	0.205	0.735	0.978	1.000	1.000	1.000	1.000	1.000
WRS				0.010	0.1	0.018	0.024	0.037	0.044	0.052	0.058	0.054	0.057
					0.2	0.029	0.058	0.109	0.147	0.189	0.192	0.210	0.209
					0.3	0.046	0.131	0.255	0.356	0.422	0.474	0.485	0.497
					0.4	0.071	0.240	0.451	0.619	0.718	0.760	0.784	0.787
					0.5	0.101	0.376	0.680	0.853	0.909	0.940	0.950	0.950
					0.6	0.141	0.542	0.858	0.965	0.988	0.994	0.994	0.995
					0.7	0.197	0.693	0.957	0.996	0.999	1.000	1.000	1.000
					0.8	0.262	0.836	0.994	1.000	1.000	1.000	1.000	1.000
					0.9	0.335	0.930	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.423	0.975	1.000	1.000	1.000	1.000	1.000	1.000

TABLE A.2 (Continued)

						δ/σ							
Test	m=n	r	k	α	ϵ	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Quantile	50	15	12	0.011	0.1	0.019	0.033	0.059	0.092	0.125	0.149	0.161	0.166
					0.2	0.029	0.078	0.182	0.335	0.485	0.588	0.641	0.662
					0.3	0.043	0.149	0.376	0.650	0.837	0.920	0.949	0.959
					0.4	0.061	0.243	0.583	0.864	0.971	0.994	0.998	0.999
					0.5	0.083	0.352	0.750	0.957	0.996	1.000	1.000	1.000
					0.6	0.108	0.464	0.861	0.987	1.000	1.000	1.000	1.000
					0.7	0.138	0.568	0.925	0.996	1.000	1.000	1.000	1.000
					0.8	0.171	0.660	0.960	0.999	1.000	1.000	1.000	1.000
					0.9	0.207	0.737	0.979	1.000	1.000	1.000	1.000	1.000
					1.0	0.245	0.798	0.988	1.000	1.000	1.000	1.000	1.000
WRS				0.010	0.1	0.018	0.030	0.043	0.051	0.062	0.065	0.068	0.068
					0.2	0.033	0.073	0.133	0.190	0.229	0.250	0.261	0.261
					0.3	0.053	0.162	0.311	0.440	0.531	0.579	0.595	0.607
					0.4	0.080	0.299	0.566	0.729	0.819	0.861	0.872	0.882
					0.5	0.126	0.458	0.787	0.926	0.963	0.979	0.984	0.985
					0.6	0.180	0.648	0.934	0.988	0.997	0.999	0.999	0.999
					0.7	0.254	0.810	0.986	1.000	1.000	1.000	1.000	1.000
					0.8	0.336	0.920	0.998	1.000	1.000	1.000	1.000	1.000
					0.9	0.429	0.975	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.521	0.993	1.000	1.000	1.000	1.000	1.000	1.000
Quantile	60	10	9	0.008	0.1	0.014	0.028	0.058	0.113	0.189	0.266	0.323	0.354
					0.2	0.022	0.066	0.186	0.401	0.640	0.808	0.890	0.923
					0.3	0.032	0.125	0.365	0.687	0.902	0.978	0.995	0.998
					0.4	0.045	0.201	0.540	0.854	0.976	0.998	1.000	1.000
					0.5	0.060	0.285	0.680	0.932	0.993	1.000	1.000	1.000
					0.6	0.078	0.370	0.779	0.966	0.998	1.000	1.000	1.000
					0.7	0.098	0.451	0.847	0.982	0.999	1.000	1.000	1.000
					0.8	0.121	0.525	0.892	0.990	1.000	1.000	1.000	1.000
					0.9	0.144	0.591	0.923	0.994	1.000	1.000	1.000	1.000
					1.0	0.170	0.648	0.943	0.996	1.000	1.000	1.000	1.000
WRS				0.010	0.1	0.019	0.033	0.048	0.061	0.072	0.074	0.078	0.082
					0.2	0.032	0.095	0.160	0.234	0.280	0.313	0.328	0.332
					0.3	0.058	0.192	0.382	0.538	0.624	0.669	0.698	0.707
					0.4	0.096	0.365	0.652	0.824	0.892	0.924	0.928	0.936
					0.5	0.149	0.560	0.865	0.966	0.986	0.994	0.993	0.996
					0.6	0.218	0.750	0.973	0.997	0.999	1.000	1.000	1.000
					0.7	0.301	0.888	0.995	1.000	1.000	1.000	1.000	1.000
					0.8	0.408	0.960	1.000	1.000	1.000	1.000	1.000	1.000
					0.9	0.515	0.990	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.619	0.998	1.000	1.000	1.000	1.000	1.000	1.000

						Δ/σ							
Test	m=n	r	k	α	ε	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Quantile	75	10	9	0.009	0.1	0.015	0.032	0.074	0.157	0.277	0.401	0.492	0.543
					0.2	0.024	0.080	0.236	0.508	0.771	0.915	0.968	0.984
					0.3	0.036	0.151	0.440	0.780	0.953	0.994	0.999	1.000
					0.4	0.051	0.238	0.618	0.907	0.989	0.999	1.000	1.000
					0.5	0.069	0.330	0.745	0.958	0.997	1.000	1.000	1.000
					0.6	0.089	0.420	0.830	0.980	0.999	1.000	1.000	1.000
					0.7	0.112	0.503	0.884	0.989	0.999	1.000	1.000	1.000
					0.8	0.137	0.576	0.920	0.994	1.000	1.000	1.000	1.000
					0.9	0.163	0.639	0.943	0.996	1.000	1.000	1.000	1.000
					1.0	0.191	0.692	0.958	0.998	1.000	1.000	1.000	1.000
WRS				0.010	0.1	0.020	0.037	0.060	0.076	0.090	0.098	0.100	0.103
					0.2	0.041	0.110	0.204	0.304	0.355	0.394	0.414	0.411
					0.3	0.070	0.248	0.471	0.647	0.743	0.776	0.806	0.806
					0.4	0.123	0.451	0.763	0.909	0.948	0.969	0.977	0.977
					0.5	0.192	0.671	0.937	0.989	0.997	0.998	0.999	0.999
					0.6	0.285	0.846	0.992	0.999	1.000	1.000	1.000	1.000
					0.7	0.385	0.950	1.000	1.000	1.000	1.000	1.000	1.000
					0.8	0.510	0.990	1.000	1.000	1.000	1.000	1.000	1.000
					0.9	0.623	0.998	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.726	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Quantile	100	10	9	0.009	0.1	0.017	0.039	0.100	0.230	0.421	0.607	0.730	0.792
					0.2	0.027	0.100	0.310	0.641	0.888	0.978	0.996	0.999
					0.3	0.041	0.187	0.536	0.866	0.982	0.999	1.000	1.000
					0.4	0.059	0.288	0.704	0.949	0.996	1.000	1.000	1.000
					0.5	0.080	0.389	0.813	0.978	0.999	1.000	1.000	1.000
					0.6	0.[illegible]03	0.483	0.879	0.989	1.000	1.000	1.000	1.000
					0.7	0.[illegible]30	0.565	0.919	0.994	1.000	1.000	1.000	1.000
					0.8	0.[illegible]58	0.635	0.945	0.997	1.000	1.000	1.000	1.000
					0.9	0.[illegible]87	0.693	0.961	0.998	1.000	1.000	1.000	1.000
					1.0	0.[illegible]17	0.742	0.971	0.999	1.000	1.000	1.000	1.000
WRS				0.010	0.1	0.[illegible]25	0.048	0.072	0.101	0.112	0.123	0.130	0.134
					0.2	0.[illegible]55	0.146	0.272	0.392	0.484	0.509	0.539	0.550
					0.3	0.[illegible]93	0.332	0.611	0.787	0.862	0.896	0.909	0.914
					0.4	0.[illegible]68	0.586	0.888	0.971	0.989	0.994	0.997	0.996
					0.5	0.[illegible]62	0.817	0.982	0.999	1.000	1.000	1.000	1.000
					0.6	0.[illegible]77	0.936	0.999	1.000	1.000	1.000	1.000	1.000
					0.7	0.[illegible]21	0.989	1.000	1.000	1.000	1.000	1.000	1.000
					0.8	0.[illegible]48	0.999	1.000	1.000	1.000	1.000	1.000	1.000
					0.9	0.[illegible]69	1.000	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.[illegible]67	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table A.3 Approximate Power and Number of Measurements for the Quantile and Wilcoxon Rank Sum (WRS) Tests for Type I Error Rate α = 0.025 for when m = n. m and n are the Number of Required Measurements from the Reference Area and the Cleanup Unit, respectively.

						Δ/σ							
Test	m=n	r	k	α	ϵ	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Quantile	10	7	6	0.029	0.1	0.034	0.042	0.051	0.055	0.056	0.061	0.062	0.063
					0.2	0.042	0.064	0.083	0.100	0.111	0.117	0.122	0.124
					0.3	0.049	0.084	0.135	0.176	0.202	0.219	0.230	0.237
					0.4	0.065	0.124	0.197	0.281	0.333	0.374	0.396	0.409
					0.5	0.076	0.152	0.272	0.398	0.503	0.554	0.582	0.604
					0.6	0.084	0.198	0.370	0.549	0.670	0.736	0.772	0.785
					0.7	0.102	0.249	0.468	0.678	0.809	0.878	0.903	0.921
					0.8	0.116	0.311	0.565	0.787	0.911	0.962	0.980	0.981
					0.9	0.137	0.370	0.658	0.874	0.965	0.991	0.999	0.999
					1.0	0.150	0.423	0.735	0.927	0.987	0.999	1.000	1.000
WRS				0.025	0.1	0.033	0.039	0.048	0.051	0.054	0.055	0.062	0.061
					0.2	0.043	0.056	0.081	0.095	0.105	0.112	0.115	0.114
					0.3	0.053	0.088	0.124	0.160	0.188	0.198	0.212	0.209
					0.4	0.062	0.125	0.187	0.260	0.300	0.320	0.336	0.352
					0.5	0.075	0.169	0.277	0.379	0.443	0.486	0.499	0.507
					0.6	0.093	0.221	0.388	0.512	0.609	0.656	0.684	0.683
					0.7	0.109	0.292	0.506	0.669	0.772	0.809	0.829	0.844
					0.8	0.132	0.366	0.638	0.819	0.891	0.930	0.934	0.943
					0.9	0.158	0.456	0.770	0.919	0.975	0.989	0.992	0.993
					1.0	0.184	0.559	0.873	0.986	0.999	1.000	1.000	1.000
Quantile	15	5	5	0.021	0.1	0.025	0.036	0.046	0.063	0.086	0.085	0.092	0.096
					0.2	0.034	0.060	0.094	0.151	0.201	0.250	0.291	0.300
					0.3	0.044	0.090	0.162	0.277	0.396	0.489	0.553	0.596
					0.4	0.052	0.123	0.244	0.411	0.584	0.723	0.789	0.829
					0.5	0.066	0.156	0.329	0.556	0.739	0.858	0.923	0.948
					0.6	0.073	0.213	0.421	0.658	0.842	0.931	2.975	0.989
					0.7	0.086	0.250	0.498	0.743	0.903	0.973	0.992	0.998
					0.8	0.097	0.297	0.561	0.812	0.936	0.986	0.997	1.000
					0.9	0.110	0.331	0.632	0.856	0.961	0.990	0.998	1.000
					1.0	0.122	0.372	0.684	0.889	0.969	0.994	0.999	1.000
WRS				0.025	0.1	0.034	0.039	0.050	0.055	0.060	0.065	0.064	0.064
					0.2	0.044	0.070	0.093	0.120	0.142	0.138	0.149	0.154
					0.3	0.055	0.113	0.163	0.215	0.254	0.275	0.288	0.290
					0.4	0.076	0.163	0.262	0.355	0.420	0.467	0.475	0.472
					0.5	0.092	0.221	0.393	0.513	0.616	0.657	0.669	0.682
					0.6	0.112	0.311	0.539	0.700	0.789	0.829	0.848	0.851
					0.7	0.147	0.407	0.702	0.843	0.915	0.938	0.948	0.952
					0.8	0.167	0.504	0.817	0.941	0.979	0.989	0.992	0.991
					0.9	0.212	0.620	0.907	0.990	0.998	0.999	1.000	1.000
					1.0	0.251	0.733	0.969	1.000	1.000	1.000	1.000	1.000

Table A.3 (Continued)

						Δ/σ							
Test	m=n	r	k	α	ε	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Quantile	20	5	5	0.024	0.1	0.031	0.043	0.063	0.084	0.114	0.138	0.143	0.160
					0.2	0.038	0.072	0.127	0.217	0.309	0.402	0.462	0.495
					0.3	0.046	0.110	0.225	0.381	0.555	0.687	0.760	0.813
					0.4	0.059	0.150	0.318	0.538	0.723	0.868	0.925	0.954
					0.5	0.075	0.202	0.414	0.669	0.854	0.941	0.979	0.993
					0.6	0.088	0.251	0.512	0.761	0.907	0.976	0.995	0.998
					0.7	0.105	0.303	0.600	0.827	0.945	0.987	0.998	1.000
					0.8	0.112	0.346	0.645	0.868	0.966	0.991	0.998	1.000
					0.9	0.129	0.394	0.708	0.898	0.977	0.994	1.000	1.000
					1.0	0.150	0.431	0.743	0.923	0.980	0.997	1.000	1.000
WRS				0.025	0.1	0.035	0.047	0.059	0.065	0.065	0.069	0.079	0.074
					0.2	0.049	0.077	0.114	0.145	0.170	0.177	0.184	0.185
					0.3	0.060	0.131	0.205	0.276	0.322	0.353	0.365	0.377
					0.4	0.082	0.199	0.338	0.453	0.534	0.577	0.591	0.612
					0.5	0.104	0.286	0.501	0.644	0.743	0.781	0.798	0.807
					0.6	0.145	0.391	0.666	0.819	0.885	0.922	0.925	0.931
					0.7	0.179	0.519	0.808	0.936	0.972	0.982	0.987	0.989
					0.8	0.221	0.639	0.915	0.985	0.996	0.998	0.999	0.999
					0.9	0.274	0.751	0.972	0.998	1.000	1.000	1.000	1.000
					1.0	0.321	0.850	0.995	1.000	1.000	1.000	1.000	1.000
Quantile	25	5	5	0.025	0.1	0.03	0.053	0.081	0.113	0.157	0.188	0.215	0.234
					0.2	0.051	0.084	0.160	0.275	0.422	0.532	0.616	0.666
					0.3	0.051	0.128	0.273	0.463	0.662	0.804	0.885	0.918
					0.4	0.068	0.187	0.388	0.633	0.821	0.927	0.970	0.987
					0.5	0.083	0.233	0.480	0.746	0.901	0.972	0.993	0.998
					0.6	0.095	0.294	0.576	0.818	0.945	0.987	0.997	1.000
					0.7	0.115	0.346	0.648	0.870	0.964	0.995	0.998	1.000
					0.8	0.128	0.385	0.708	0.898	0.976	0.995	1.000	1.000
					0.9	0.142	0.437	0.744	0.924	0.983	0.997	1.000	1.000
					1.0	0.166	0.468	0.783	0.941	0.988	0.998	1.000	1.000
WRS				0.025	0.1	0.036	0.051	0.060	0.073	0.082	0.082	0.083	0.086
					0.2	0.053	0.089	0.132	0.172	0.202	0.205	0.225	0.225
					0.3	0.072	0.153	0.244	0.341	0.391	0.420	0.449	0.444
					0.4	0.101	0.247	0.412	0.550	0.638	0.666	0.693	0.700
					0.5	0.127	0.354	0.599	0.749	0.825	0.855	0.877	0.885
					0.6	0.162	0.484	0.760	0.898	0.945	0.967	0.973	0.972
					0.7	0.217	0.619	0.893	0.974	0.990	0.995	0.997	0.997
					0.8	0.265	0.755	0.962	0.996	1.000	1.000	1.000	1.000
					0.9	0.335	0.842	0.991	1.000	1.000	1.000	1.000	1.000
					1.0	0.391	0.924	1.000	1.000	1.000	1.000	1.000	1.000

Table A.3 (Continued)

						Δ/σ							
Test	m=n	r	k	α	ϵ	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Quantile	30	5	5	0.026	0.1	0.037	0.048	0.088	0.137	0.194	0.253	0.295	0.316
					0.2	0.043	0.098	0.187	0.332	0.495	0.644	0.734	0.795
					0.3	0.056	0.142	0.306	0.535	0.745	0.880	0.941	0.965
					0.4	0.074	0.197	0.432	0.691	0.874	0.958	0.988	0.998
					0.5	0.089	0.256	0.536	0.792	0.929	0.981	0.996	1.000
					0.6	0.107	0.317	0.620	0.853	0.962	0.992	0.999	1.000
					0.7	0.126	0.368	0.680	0.891	0.975	0.995	0.999	1.000
					0.8	0.146	0.419	0.737	0.919	0.982	0.997	0.999	1.000
					0.9	0.160	0.467	0.769	0.935	0.988	0.998	1.000	1.000
					1.0	0.173	0.497	0.807	0.949	0.989	0.998	1.000	1.000
WRS				0.025	0.1	0.039	0.052	0.073	0.082	0.089	0.089	0.096	0.094
					0.2	0.055	0.098	0.160	0.197	0.234	0.250	0.256	0.262
					0.3	0.081	0.181	0.291	0.401	0.462	0.493	0.517	0.521
					0.4	0.112	0.283	0.475	0.628	0.707	0.755	0.769	0.777
					0.5	0.149	0.422	0.679	0.829	0.894	0.921	0.931	0.931
					0.6	0.200	0.552	0.836	0.944	0.978	0.985	0.988	0.988
					0.7	0.250	0.700	0.939	0.991	0.997	0.999	0.999	0.999
					0.8	0.308	0.820	0.986	0.999	1.000	1.000	1.000	1.000
					0.9	0.387	0.906	0.998	1.000	1.000	1.000	1.000	1.000
					1.0	0.469	0.962	1.000	1.000	1.000	1.000	1.000	1.000
Quantile	40	5	5	0.027	0.1	0.036	0.061	0.110	0.180	0.273	0.371	0.438	0.490
					0.2	0.058	0.114	0.233	0.430	0.645	0.793	0.887	0.924
					0.3	0.068	0.166	0.374	0.641	0.841	0.946	0.984	0.996
					0.4	0.079	0.229	0.507	0.777	0.923	0.984	0.998	1.000
					0.5	0.102	0.295	0.607	0.841	0.961	0.993	0.999	1.000
					0.6	0.116	0.360	0.682	0.891	0.977	0.995	0.999	1.000
					0.7	0.137	0.416	0.735	0.920	0.984	0.998	1.000	1.000
					0.8	0.160	0.469	0.790	0.943	0.988	0.999	1.000	1.000
					0.9	0.187	0.519	0.822	0.952	0.993	0.999	1.000	1.000
					1.0	0.202	0.556	0.847	0.961	0.993	1.000	1.000	1.000
WRS				0.025	0.1	0.039	0.059	0.080	0.092	0.110	0.113	0.115	0.117
					0.2	0.058	0.125	0.199	0.257	0.295	0.322	0.339	0.344
					0.3	0.091	0.232	0.375	0.499	0.579	0.611	0.636	0.641
					0.4	0.142	0.357	0.602	0.757	0.823	0.873	0.881	0.880
					0.5	0.190	0.516	0.800	0.919	0.961	0.972	0.978	0.980
					0.6	0.251	0.690	0.930	0.986	0.995	0.998	0.998	0.999
					0.7	0.317	0.821	0.983	0.999	1.000	1.000	1.000	1.000
					0.8	0.398	0.915	0.998	1.000	1.000	1.000	1.000	1.000
					0.9	0.488	0.970	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.574	0.991	1.000	1.000	1.000	1.000	1.000	1.000

Table A.3 (Continued)

						Δ/σ							
Test	m=n	r	k	α	ϵ	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Quantile	50	11	9	0.026	0.1	0.037	0.064	0.116	0.176	0.251	0.308	0.339	0.358
					0.2	0.052	0.138	0.289	0.496	0.685	0.803	0.854	0.876
					0.3	0.080	0.230	0.512	0.778	0.925	0.975	0.991	0.994
					0.4	0.105	0.342	0.691	0.918	0.989	0.998	1.000	1.000
					0.5	0.134	0.435	0.806	0.972	0.998	1.000	1.000	1.000
					0.6	0.171	0.541	0.894	0.991	1.000	1.000	1.000	1.000
					0.7	0.199	0.627	0.935	0.996	1.000	1.000	1.000	1.000
					0.8	0.243	0.706	0.961	0.999	1.000	1.000	1.000	1.000
					0.9	0.282	0.769	0.978	1.000	1.000	1.000	1.000	1.000
					1.0	0.312	0.818	0.984	1.000	1.000	1.000	1.000	1.000
WRS				0.025	0.1	0.041	0.066	0.091	0.112	0.121	0.122	0.130	0.133
					0.2	0.067	0.144	0.234	0.313	0.356	0.380	0.399	0.404
					0.3	0.102	0.274	0.460	0.594	0.677	0.715	0.740	0.743
					0.4	0.148	0.427	0.703	0.842	0.898	0.929	0.940	0.945
					0.5	0.224	0.617	0.879	0.966	0.984	0.991	0.995	0.994
					0.6	0.292	0.785	0.970	0.996	0.999	1.000	1.000	1.000
					0.7	0.388	0.901	0.995	1.000	1.000	1.000	1.000	1.000
					0.8	0.485	0.966	1.000	1.000	1.000	1.000	1.000	1.000
					0.9	0.589	0.990	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.666	0.998	1.000	1.000	1.000	1.000	1.000	1.000
Quantile	60	11	9	0.027	0.1	0.043	0.076	0.136	0.217	0.329	0.409	0.465	0.480
					0.2	0.064	0.157	0.344	0.591	0.792	0.897	0.942	0.953
					0.3	0.084	0.261	0.563	0.850	0.965	0.994	0.998	0.999
					0.4	0.107	0.374	0.750	0.952	0.995	1.000	1.000	1.000
					0.5	0.141	0.485	0.860	0.986	0.999	1.000	1.000	1.000
					0.6	0.183	0.586	0.917	0.994	1.000	1.000	1.000	1.000
					0.7	0.221	0.676	0.952	0.998	1.000	1.000	1.000	1.000
					0.8	0.258	0.745	0.974	0.999	1.000	1.000	1.000	1.000
					0.9	0.301	0.806	0.982	1.000	1.000	1.000	1.000	1.000
					1.0	0.340	0.848	0.991	1.000	1.000	1.000	1.000	1.000
WRS				0.025	0.1	0.046	0.072	0.096	0.123	0.140	0.145	0.146	0.149
					0.2	0.076	0.163	0.270	0.347	0.414	0.447	0.465	0.475
					0.3	0.117	0.320	0.526	0.671	0.755	0.802	0.807	0.814
					0.4	0.176	0.501	0.779	0.902	0.946	0.963	0.972	0.972
					0.5	0.252	0.705	0.936	0.984	0.995	0.998	0.998	0.998
					0.6	0.344	0.856	0.989	0.999	1.000	1.000	1.000	1.000
					0.7	0.450	0.949	0.998	1.000	1.000	1.000	1.000	1.000
					0.8	0.566	0.982	1.000	1.000	1.000	1.000	1.000	1.000
					0.9	0.653	0.997	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.754	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table A.3 (Continued)

						Δ/σ							
Test	m=n	r	k	α	ε	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Quantile	75	14	11	0.023	0.1	0.036	0.078	0.142	0.242	0.361	0.450	0.507	0.526
					0.2	0.060	0.166	0.391	0.661	0.857	0.934	0.969	0.975
					0.3	0.082	0.293	0.644	0.906	0.987	0.999	1.000	1.000
					0.4	0.124	0.429	0.822	0.981	0.999	1.000	1.000	1.000
					0.5	0.159	0.561	0.918	0.996	1.000	1.000	1.000	1.000
					0.6	0.202	0.671	0.963	0.999	1.000	1.000	1.000	1.000
					0.7	0.243	0.761	0.982	1.000	1.000	1.000	1.000	1.000
					0.8	0.289	0.829	0.991	1.000	1.000	1.000	1.000	1.000
					0.9	0.339	0.878	0.995	1.000	1.000	1.000	1.000	1.000
					1.0	0.385	0.910	0.998	1.000	1.000	1.000	1.000	1.000
WRS				0.025	0.1	0.048	0.075	0.113	0.145	0.166	0.175	0.180	0.176
					0.2	0.086	0.192	0.324	0.439	0.497	0.532	0.556	0.567
					0.3	0.134	0.387	0.621	0.774	0.843	0.877	0.889	0.897
					0.4	0.213	0.603	0.868	0.958	0.981	0.987	0.990	0.991
					0.5	0.313	0.796	0.971	0.997	1.000	1.000	1.000	1.000
					0.6	0.420	0.923	0.997	1.000	1.000	1.000	1.000	1.000
					0.7	0.540	0.977	1.000	1.000	1.000	1.000	1.000	1.000
					0.8	0.654	0.995	1.000	1.000	1.000	1.000	1.000	1.000
					0.9	0.756	1.000	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.838	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Quantile	100	14	11	0.024	0.1	0.042	0.090	0.192	0.352	0.537	0.662	0.726	0.771
					0.2	0.065	0.205	0.497	0.797	0.953	0.991	0.997	0.999
					0.3	0.099	0.363	0.753	0.964	0.997	1.000	1.000	1.000
					0.4	0.138	0.509	0.891	0.993	1.000	1.000	1.000	1.000
					0.5	0.180	0.625	0.953	0.999	1.000	1.000	1.000	1.000
					0.6	0.234	0.745	0.980	1.000	1.000	1.000	1.000	1.000
					0.7	0.274	0.823	0.990	1.000	1.000	1.000	1.000	1.000
					0.8	0.333	0.874	0.995	1.000	1.000	1.000	1.000	1.000
					0.9	0.378	0.911	0.998	1.000	1.000	1.000	1.000	1.000
					1.0	0.440	0.938	0.999	1.000	1.000	1.000	1.000	1.000
WRS				0.025	0.1	0.055	0.093	0.134	0.176	0.203	0.217	0.215	0.231
					0.2	0.097	0.241	0.408	0.541	0.623	0.666	0.675	0.678
					0.3	0.173	0.486	0.752	0.875	0.926	0.948	0.958	0.959
					0.4	0.273	0.726	0.946	0.987	0.996	0.998	0.999	0.999
					0.5	0.392	0.900	0.994	1.000	1.000	1.000	1.000	1.000
					0.6	0.529	0.976	1.000	1.000	1.000	1.000	1.000	1.000
					0.7	0.665	0.996	1.000	1.000	1.000	1.000	1.000	1.000
					0.8	0.777	1.000	1.000	1.000	1.000	1.000	1.000	1.000
					0.9	0.875	1.000	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.933	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table A.4 Approximate Power and Number of Measurements for the Quantile and Wilcoxon Rank Sum (WRS) Tests for Type I Error Rate $\alpha = 0.05$ for when $m = n$. m and n are the Number of Required Measurements from the Reference Area and the Cleanup Unit, respectively.

						Δ/σ							
Test	m=n	r	k	α	ϵ	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Quantile	10	4	4	0.043	0.1	0.052	0.065	0.079	0.094	0.105	0.113	0.117	0.119
					0.2	0.062	0.092	0.132	0.177	0.218	0.250	0.270	0.280
					0.3	0.074	0.125	0.199	0.287	0.372	0.437	0.479	0.500
					0.4	0.086	0.162	0.276	0.411	0.536	0.629	0.686	0.714
					0.5	0.098	0.203	0.358	0.533	0.683	0.786	0.843	0.869
					0.6	0.112	0.247	0.439	0.641	0.797	0.890	0.936	0.955
					0.7	0.127	0.291	0.516	0.729	0.874	0.948	0.978	0.989
					0.8	0.142	0.336	0.584	0.796	0.921	0.975	0.993	0.998
					0.9	0.157	0.379	0.644	0.845	0.948	0.986	0.997	0.999
					1.0	0.173	0.422	0.695	0.880	0.964	0.992	0.998	1.000
WRS				0.050	0.1	0.065	0.076	0.091	0.095	0.101	0.111	0.104	0.101
					0.2	0.080	0.109	0.138	0.158	0.174	0.182	0.199	0.193
					0.3	0.101	0.149	0.211	0.263	0.294	0.302	0.310	0.309
					0.4	0.110	0.197	0.291	0.376	0.435	0.445	0.469	0.476
					0.5	0.136	0.259	0.404	0.506	0.576	0.619	0.632	0.632
					0.6	0.159	0.330	0.522	0.653	0.731	0.768	0.792	0.795
					0.7	0.194	0.413	0.636	0.785	0.862	0.892	0.899	0.907
					0.8	0.216	0.495	0.751	0.895	0.949	0.966	0.971	0.975
					0.9	0.256	0.587	0.855	0.966	0.989	0.994	0.997	0.998
					1.0	0.282	0.677	0.939	0.995	1.000	1.000	1.000	1.000
Quantile	15	4	4	0.050	0.1	0.062	0.081	0.106	0.136	0.164	0.186	0.200	0.207
					0.2	0.075	0.120	0.187	0.273	0.361	0.433	0.481	0.507
					0.3	0.090	0.165	0.284	0.431	0.572	0.680	0.745	0.779
					0.4	0.105	0.215	0.384	0.577	0.740	0.847	0.903	0.928
					0.5	0.122	0.267	0.478	0.694	0.850	0.934	0.970	0.983
					0.6	0.139	0.318	0.562	0.780	0.913	0.971	0.991	0.997
					0.7	0.157	0.369	0.633	0.839	0.947	0.986	0.997	0.999
					0.8	0.175	0.417	0.692	0.881	0.965	0.992	0.999	1.000
					0.9	0.194	0.462	0.739	0.909	0.976	0.995	0.999	1.000
					1.0	0.213	0.504	0.778	0.928	0.983	0.997	0.999	1.000
WRS				0.050	0.1	0.072	0.084	0.105	0.109	0.121	0.120	0.126	0.128
					0.2	0.085	0.132	0.168	0.206	0.229	0.241	0.241	0.245
					0.3	0.110	0.193	0.270	0.338	0.391	0.414	0.415	0.418
					0.4	0.134	0.253	0.385	0.498	0.558	0.593	0.616	0.626
					0.5	0.168	0.347	0.536	0.664	0.738	0.770	0.793	0.791
					0.6	0.200	0.448	0.683	0.804	0.878	0.904	0.916	0.922
					0.7	0.234	0.546	0.802	0.914	0.959	0.972	0.976	0.979
					0.8	0.279	0.654	0.898	0.975	0.992	0.996	0.997	0.998
					0.9	0.330	0.753	0.959	0.997	1.000	1.000	1.000	1.000
					1.0	0.369	0.841	0.988	1.000	1.000	1.000	1.000	1.000

Table A.4 (Continued)

						Δ/σ							
Test	m=n	r	k	α	ϵ	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Quantile	20	4	4	0.053	0.1	0.067	0.091	0.127	0.173	0.220	0.261	0.290	0.306
					0.2	0.083	0.139	0.232	0.354	0.481	0.586	0.655	0.693
					0.3	0.099	0.194	0.347	0.535	0.704	0.821	0.885	0.915
					0.4	0.118	0.252	0.458	0.678	0.842	0.932	0.970	0.984
					0.5	0.136	0.310	0.555	0.779	0.915	0.973	0.992	0.998
					0.6	0.156	0.366	0.634	0.845	0.951	0.988	0.998	1.000
					0.7	0.176	0.419	0.699	0.888	0.969	0.994	0.999	1.000
					0.8	0.197	0.468	0.749	0.916	0.979	0.996	0.999	1.000
					0.9	0.217	0.513	0.789	0.936	0.985	0.997	1.000	1.000
					1.0	0.238	0.554	0.821	0.949	0.989	0.998	1.000	1.000
WRS				0.050	0.1	0.066	0.090	0.108	0.122	0.125	0.134	0.134	0.137
					0.2	0.091	0.145	0.191	0.244	0.262	0.277	0.288	0.291
					0.3	0.122	0.213	0.321	0.406	0.459	0.489	0.489	0.496
					0.4	0.151	0.303	0.461	0.586	0.657	0.699	0.711	0.721
					0.5	0.187	0.407	0.629	0.767	0.836	0.864	0.877	0.883
					0.6	0.232	0.532	0.775	0.893	0.945	0.959	0.965	0.971
					0.7	0.283	0.652	0.896	0.968	0.988	0.994	0.995	0.995
					0.8	0.331	0.758	0.959	0.994	0.999	0.999	1.000	1.000
					0.9	0.386	0.849	0.989	0.999	1.000	1.000	1.000	1.000
					1.0	0.451	0.917	0.998	1.000	1.000	1.000	1.000	1.000
Quantile	25	7	6	0.049	0.1	0.065	0.091	0.127	0.169	0.206	0.233	0.248	0.254
					0.2	0.083	0.149	0.251	0.375	0.491	0.573	0.618	0.639
					0.3	0.104	0.219	0.399	0.599	0.755	0.845	0.887	0.903
					0.4	0.127	0.297	0.544	0.771	0.906	0.962	0.980	0.986
					0.5	0.153	0.377	0.667	0.879	0.968	0.993	0.998	0.999
					0.6	0.179	0.455	0.763	0.937	0.989	0.999	1.000	1.000
					0.7	0.207	0.528	0.832	0.967	0.996	1.000	1.000	1.000
					0.8	0.236	0.594	0.881	0.981	0.998	1.000	1.000	1.000
					0.9	0.265	0.652	0.915	0.989	0.999	1.000	1.000	1.000
					1.0	0.295	0.702	0.938	0.993	1.000	1.000	1.000	1.000
WRS				0.050	0.1	0.072	0.092	0.115	0.137	0.150	0.152	0.151	0.152
					0.2	0.096	0.159	0.229	0.278	0.305	0.333	0.326	0.335
					0.3	0.128	0.243	0.367	0.462	0.536	0.562	0.578	0.587
					0.4	0.169	0.360	0.545	0.685	0.753	0.786	0.802	0.813
					0.5	0.211	0.483	0.727	0.842	0.902	0.928	0.936	0.931
					0.6	0.269	0.614	0.852	0.951	0.973	0.984	0.987	0.987
					0.7	0.325	0.744	0.944	0.990	0.996	0.999	0.999	0.998
					0.8	0.390	0.841	0.983	0.999	1.000	1.000	1.000	1.000
					0.9	0.465	0.913	0.997	1.000	1.000	1.000	1.000	1.000
					1.0	0.530	0.957	1.000	1.000	1.000	1.000	1.000	1.000

Table A.4 (Continued)

						Δ/σ							
Test	m=n	r	k	α	ϵ	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Quantile	30	7	6	0.051	0.1	0.069	0.100	0.146	0.202	0.256	0.297	0.321	0.332
					0.2	0.090	0.167	0.292	0.449	0.592	0.691	0.745	0.769
					0.3	0.113	0.246	0.457	0.681	0.840	0.920	0.951	0.963
					0.4	0.138	0.332	0.607	0.836	0.949	0.986	0.995	0.997
					0.5	0.166	0.417	0.724	0.919	0.985	0.998	1.000	1.000
					0.6	0.195	0.498	0.809	0.959	0.995	1.000	1.000	1.000
					0.7	0.225	0.571	0.868	0.979	0.998	1.000	1.000	1.000
					0.8	0.256	0.635	0.908	0.988	0.999	1.000	1.000	1.000
					0.9	0.288	0.690	0.934	0.993	1.000	1.000	1.000	1.000
					1.0	0.319	0.737	0.952	0.996	1.000	1.000	1.000	1.000
WRS				0.050	0.1	0.073	0.097	0.125	0.136	0.147	0.159	0.170	0.162
					0.2	0.103	0.167	0.241	0.294	0.345	0.364	0.372	0.376
					0.3	0.142	0.265	0.420	0.515	0.581	0.622	0.645	0.646
					0.4	0.178	0.398	0.602	0.743	0.813	0.838	0.856	0.854
					0.5	0.240	0.542	0.787	0.897	0.942	0.952	0.966	0.966
					0.6	0.290	0.679	0.904	0.973	0.991	0.994	0.995	0.996
					0.7	0.353	0.803	0.971	0.996	0.999	1.000	1.000	1.000
					0.8	0.444	0.894	0.994	1.000	1.000	1.000	1.000	1.000
					0.9	0.505	0.950	0.999	1.000	1.000	1.000	1.000	1.000
					1.0	0.596	0.980	1.000	1.000	1.000	1.000	1.000	1.000
Quantile	40	7	6	0.054	0.1	0.075	0.114	0.178	0.264	0.354	0.426	0.471	0.493
					0.2	0.099	0.196	0.363	0.568	0.742	0.848	0.899	0.919
					0.3	0.126	0.290	0.548	0.791	0.929	0.978	0.992	0.996
					0.4	0.155	0.387	0.695	0.907	0.982	0.998	1.000	1.000
					0.5	0.187	0.479	0.798	0.958	0.995	1.000	1.000	1.000
					0.6	0.219	0.561	0.866	0.980	0.998	1.000	1.000	1.000
					0.7	0.253	0.632	0.910	0.989	0.999	1.000	1.000	1.000
					0.8	0.287	0.693	0.938	0.994	1.000	1.000	1.000	1.000
					0.9	0.321	0.743	0.956	0.996	1.000	1.000	1.000	1.000
					1.0	0.354	0.784	0.968	0.998	1.000	1.000	1.000	1.000
WRS				0.050	0.1	0.077	0.109	0.136	0.164	0.178	0.189	0.189	0.202
					0.2	0.113	0.198	0.297	0.365	0.408	0.450	0.450	0.470
					0.3	0.166	0.334	0.509	0.626	0.701	0.741	0.744	0.759
					0.4	0.216	0.489	0.718	0.848	0.899	0.925	0.933	0.937
					0.5	0.279	0.655	0.880	0.959	0.980	0.989	0.990	0.993
					0.6	0.360	0.791	0.962	0.993	0.999	0.999	0.999	0.999
					0.7	0.444	0.897	0.994	0.999	1.000	1.000	1.000	1.000
					0.8	0.519	0.959	0.999	1.000	1.000	1.000	1.000	1.000
					0.9	0.617	0.988	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.699	0.996	1.000	1.000	1.000	1.000	1.000	1.000

Table A.4 (Continued)

						Δ/σ							
Test	m=n	r	k	α	ϵ	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Quantile	50	10	8	0.046	0.1	0.067	0.108	0.176	0.266	0.356	0.423	0.463	0.480
					0.2	0.093	0.201	0.390	0.612	0.783	0.876	0.916	0.931
					0.3	0.123	0.313	0.606	0.850	0.959	0.989	0.996	0.998
					0.4	0.157	0.430	0.767	0.950	0.994	0.999	1.000	1.000
					0.5	0.194	0.540	0.869	0.984	0.999	1.000	1.000	1.000
					0.6	0.234	0.636	0.927	0.995	1.000	1.000	1.000	1.000
					0.7	0.275	0.715	0.959	0.998	1.000	1.000	1.000	1.000
					0.8	0.317	0.778	0.976	0.999	1.000	1.000	1.000	1.000
					0.9	0.359	0.828	0.986	1.000	1.000	1.000	1.000	1.000
					1.0	0.400	0.866	0.991	1.000	1.000	1.000	1.000	1.000
WRS				0.050	0.1	0.083	0.117	0.150	0.183	0.193	0.212	0.213	0.214
					0.2	0.121	0.224	0.338	0.427	0.487	0.513	0.530	0.541
					0.3	0.177	0.394	0.578	0.711	0.779	0.808	0.835	0.829
					0.4	0.246	0.564	0.803	0.904	0.948	0.958	0.968	0.970
					0.5	0.327	0.735	0.936	0.985	0.993	0.997	0.998	0.997
					0.6	0.410	0.865	0.988	0.999	1.000	1.000	1.000	1.000
					0.7	0.506	0.949	0.998	1.000	1.000	1.000	1.000	1.000
					0.8	0.610	0.984	1.000	1.000	1.000	1.000	1.000	1.000
					0.9	0.704	0.995	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.786	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Quantile	60	10	8	0.047	0.1	0.070	0.119	0.203	0.320	0.440	0.532	0.585	0.610
					0.2	0.099	0.224	0.446	0.696	0.865	0.942	0.969	0.977
					0.3	0.132	0.348	0.669	0.901	0.982	0.997	0.999	1.000
					0.4	0.170	0.472	0.818	0.971	0.998	1.000	1.000	1.000
					0.5	0.210	0.584	0.903	0.991	1.000	1.000	1.000	1.000
					0.6	0.253	0.678	0.948	0.997	1.000	1.000	1.000	1.000
					0.7	0.296	0.753	0.971	0.999	1.000	1.000	1.000	1.000
					0.8	0.340	0.811	0.984	1.000	1.000	1.000	1.000	1.000
					0.9	0.384	0.855	0.990	1.000	1.000	1.000	1.000	1.000
					1.0	0.426	0.888	0.994	1.000	1.000	1.000	1.000	1.000
WRS				0.050	0.1	0.084	0.126	0.171	0.204	0.230	0.237	0.240	0.243
					0.2	0.129	0.257	0.390	0.475	0.550	0.578	0.596	0.604
					0.3	0.195	0.435	0.655	0.779	0.841	0.872	0.882	0.893
					0.4	0.282	0.632	0.854	0.947	0.973	0.983	0.985	0.987
					0.5	0.366	0.804	0.966	0.993	0.998	0.999	1.000	1.000
					0.6	0.467	0.920	0.995	1.000	1.000	1.000	1.000	1.000
					0.7	0.583	0.972	0.999	1.000	1.000	1.000	1.000	1.000
					0.8	0.675	0.993	1.000	1.000	1.000	1.000	1.000	1.000
					0.9	0.771	0.999	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.847	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table A.4 (Continued)

						Δ/σ							
Test	m=n	r	k	α	ϵ	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Quantile	75	10	8	0.049	0.1	0.075	0.132	0.240	0.394	0.553	0.672	0.739	0.769
					0.2	0.106	0.254	0.517	0.786	0.934	0.982	0.994	0.996
					0.3	0.143	0.392	0.738	0.944	0.994	1.000	1.000	1.000
					0.4	0.185	0.523	0.867	0.986	0.999	1.000	1.000	1.000
					0.5	0.229	0.635	0.933	0.996	1.000	1.000	1.000	1.000
					0.6	0.275	0.724	0.966	0.999	1.000	1.000	1.000	1.000
					0.7	0.322	0.793	0.981	0.999	1.000	1.000	1.000	1.000
					0.8	0.368	0.844	0.990	1.000	1.000	1.000	1.000	1.000
					0.9	0.413	0.883	0.994	1.000	1.000	1.000	1.000	1.000
					1.0	0.457	0.911	0.996	1.000	1.000	1.000	1.000	1.000
WRS				0.050	0.1	0.090	0.135	0.185	0.221	0.258	0.271	0.278	0.274
					0.2	0.145	0.288	0.443	0.558	0.629	0.661	0.680	0.672
					0.3	0.226	0.509	0.738	0.861	0.906	0.933	0.937	0.942
					0.4	0.314	0.726	0.925	0.977	0.989	0.994	0.995	0.996
					0.5	0.432	0.881	0.989	0.999	1.000	1.000	1.000	1.000
					0.6	0.556	0.956	0.999	1.000	1.000	1.000	1.000	1.000
					0.7	0.664	0.990	1.000	1.000	1.000	1.000	1.000	1.000
					0.8	0.764	0.999	1.000	1.000	1.000	1.000	1.000	1.000
					0.9	0.848	1.000	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.909	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Quantile	100	10	8	0.050	0.1	0.079	0.150	0.293	0.501	0.703	0.833	0.895	0.921
					0.2	0.116	0.294	0.606	0.875	0.978	0.997	1.000	1.000
					0.3	0.157	0.448	0.812	0.975	0.999	1.000	1.000	1.000
					0.4	0.204	0.584	0.914	0.994	1.000	1.000	1.000	1.000
					0.5	0.253	0.693	0.959	0.998	1.000	1.000	1.000	1.000
					0.6	0.303	0.776	0.980	0.999	1.000	1.000	1.000	1.000
					0.7	0.353	0.836	0.989	1.000	1.000	1.000	1.000	1.000
					0.8	0.402	0.879	0.994	1.000	1.000	1.000	1.000	1.000
					0.9	0.449	0.911	0.997	1.000	1.000	1.000	1.000	1.000
					1.0	0.494	0.933	0.998	1.000	1.000	1.000	1.000	1.000
WRS				0.050	0.1	0.101	0.158	0.220	0.271	0.303	0.314	0.332	0.334
					0.2	0.175	0.350	0.542	0.659	0.721	0.772	0.792	0.798
					0.3	0.261	0.604	0.835	0.931	0.961	0.975	0.978	0.982
					0.4	0.385	0.821	0.973	0.993	0.998	0.999	0.999	0.999
					0.5	0.515	0.941	0.998	1.000	1.000	1.000	1.000	1.000
					0.6	0.647	0.987	1.000	1.000	1.000	1.000	1.000	1.000
					0.7	0.770	0.998	1.000	1.000	1.000	1.000	1.000	1.000
					0.8	0.858	1.000	1.000	1.000	1.000	1.000	1.000	1.000
					0.9	0.925	1.000	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.964	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table A.5 Approximate Power and Number of Measurements for the Quantile and Wilcoxon Rank Sum (WRS) Tests for Type I Error Rate $\alpha = 0.10$ for when $m = n$. m and n are the Number of Required Measurements from the Reference Area and the Cleanup Unit, respectively.

						Δ/σ							
Test	m=n	r	k	α	ϵ	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Quantile	10	3	3	0.105	0.1	0.119	0.144	0.174	0.210	0.241	0.249	0.266	0.271
					0.2	0.133	0.197	0.257	0.336	0.410	0.463	0.496	0.512
					0.3	0.1[illegible]6	0.242	0.360	0.486	0.594	0.674	0.715	0.738
					0.4	0.179	0.306	0.457	0.607	0.734	0.822	0.866	0.878
					0.5	0.190	0.351	0.540	0.706	0.836	0.912	0.946	0.960
					0.6	0.227	0.400	0.607	0.789	0.909	0.958	0.983	0.991
					0.7	0.233	0.453	0.683	0.855	0.939	0.983	0.993	0.997
					0.8	0.26[illegible]	0.491	0.735	0.892	0.963	0.991	0.998	1.000
					0.9	0.29[illegible]	0.546	0.773	0.919	0.973	0.995	0.998	1.000
					1.0	0.30[illegible]	0.581	0.803	0.936	0.984	0.998	0.999	1.000
WRS				0.100	0.1	0.131	0.149	0.176	0.173	0.185	0.195	0.202	0.186
					0.2	0.152	0.203	0.235	0.287	0.299	0.315	0.319	0.324
					0.3	0.181	0.263	0.334	0.392	0.428	0.460	0.466	0.473
					0.4	0.205	0.326	0.449	0.520	0.583	0.608	0.630	0.629
					0.5	0.234	0.402	0.564	0.662	0.731	0.762	0.763	0.765
					0.6	0.268	0.487	0.675	0.788	0.846	0.870	0.884	0.886
					0.7	0.302	0.577	0.776	0.891	0.932	0.950	0.952	0.959
					0.8	0.354	0.659	0.871	0.955	0.979	0.988	0.991	0.992
					0.9	0.396	0.732	0.932	0.986	0.997	0.999	0.999	0.999
					1.0	0.435	0.809	0.976	0.999	1.000	1.000	1.000	1.000
Quantile	15	3	3	0.113	0.1	0.131	0.171	0.217	0.262	0.313	0.360	0.386	0.394
					0.2	0.155	0.226	0.327	0.443	0.557	0.644	0.699	0.727
					0.3	0.176	0.285	0.443	0.614	0.749	0.847	0.889	0.912
					0.4	0.208	0.356	0.551	0.741	0.867	0.935	0.967	0.980
					0.5	0.227	0.414	0.644	0.816	0.924	0.975	0.992	0.995
					0.6	0.253	0.472	0.701	0.877	0.961	0.988	0.997	1.000
					0.7	0.271	0.517	0.758	0.909	0.975	0.993	0.999	1.000
					0.8	0.301	0.571	0.794	0.934	0.982	0.996	0.999	1.000
					0.9	0.322	0.603	0.833	0.952	0.988	0.999	1.000	1.000
					1.0	0.347	0.640	0.858	0.956	0.992	0.999	1.000	1.000
WRS				0.100	0.1	0.128	0.157	0.180	0.206	0.215	0.215	0.213	0.215
					0.2	0.163	0.221	0.292	0.342	0.359	0.378	0.375	0.393
					0.3	0.198	0.306	0.418	0.492	0.530	0.560	0.572	0.580
					0.4	0.235	0.407	0.545	0.647	0.704	0.734	0.745	0.757
					0.5	0.282	0.496	0.682	0.802	0.847	0.873	0.889	0.887
					0.6	0.324	0.603	0.814	0.894	0.936	0.954	0.960	0.961
					0.7	0.375	0.696	0.891	0.961	0.983	0.990	0.990	0.992
					0.8	0.425	0.791	0.953	0.991	0.998	0.999	0.999	0.999
					0.9	0.469	0.863	0.984	0.999	1.000	1.000	1.000	1.000
					1.0	0.535	0.923	0.997	1.000	1.000	1.000	1.000	1.000

Table A.5 (Continued)

						Δ/σ							
Test	m=n	r	k	α	ϵ	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Quantile	20	6	5	0.089	0.1	0.115	0.148	0.192	0.230	0.276	0.287	0.308	0.312
					0.2	0.136	0.219	0.325	0.443	0.540	0.605	0.636	0.653
					0.3	0.165	0.290	0.465	0.648	0.771	0.843	0.873	0.885
					0.4	0.190	0.379	0.605	0.793	0.906	0.956	0.972	0.978
					0.5	0.235	0.464	0.714	0.892	0.966	0.992	0.996	0.997
					0.6	0.261	0.522	0.802	0.935	0.988	0.998	1.000	1.000
					0.7	0.281	0.589	0.865	0.969	0.996	1.000	1.000	1.000
					0.8	0.319	0.661	0.902	0.983	0.999	1.000	1.000	1.000
					0.9	0.354	0.711	0.931	0.990	0.999	1.000	1.000	1.000
					1.0	0.380	0.754	0.947	0.994	1.000	1.000	1.000	1.000
WRS				0.100	0.1	0.127	0.156	0.183	0.203	0.212	0.224	0.235	0.233
					0.2	0.164	0.240	0.303	0.358	0.393	0.411	0.424	0.420
					0.3	0.205	0.340	0.454	0.545	0.594	0.624	0.646	0.642
					0.4	0.256	0.440	0.619	0.723	0.781	0.812	0.827	0.823
					0.5	0.292	0.553	0.762	0.868	0.911	0.928	0.935	0.938
					0.6	0.363	0.672	0.872	0.950	0.973	0.979	0.984	0.987
					0.7	0.407	0.772	0.943	0.987	0.995	0.998	0.998	0.998
					0.8	0.470	0.859	0.981	0.998	1.000	1.000	1.000	1.000
					0.9	0.530	0.925	0.997	1.000	1.000	1.000	1.000	1.000
					1.0	0.602	0.959	0.999	1.000	1.000	1.000	1.000	1.000
Quantile	25	6	5	0.093	0.1	0.127	0.167	0.229	0.283	0.333	0.376	0.395	0.403
					0.2	0.150	0.236	0.375	0.529	0.637	0.733	0.769	0.784
					0.3	0.177	0.332	0.532	0.742	0.858	0.922	0.947	0.960
					0.4	0.209	0.420	0.678	0.865	0.955	0.985	0.993	0.996
					0.5	0.238	0.501	0.769	0.934	0.984	0.997	1.000	1.000
					0.6	0.274	0.580	0.848	0.965	0.995	1.000	1.000	1.000
					0.7	0.319	0.651	0.895	0.983	0.998	1.000	1.000	1.000
					0.8	0.350	0.703	0.927	0.992	0.999	1.000	1.000	1.000
					0.9	0.375	0.743	0.949	0.994	1.000	1.000	1.000	1.000
					1.0	0.403	0.786	0.963	0.997	1.000	1.000	1.000	1.000
WRS				0.100	0.1	0.132	0.165	0.193	0.227	0.242	0.234	0.248	0.248
					0.2	0.172	0.254	0.349	0.401	0.445	0.463	0.475	0.480
					0.3	0.215	0.362	0.509	0.607	0.661	0.687	0.711	0.712
					0.4	0.270	0.506	0.685	0.797	0.854	0.873	0.880	0.888
					0.5	0.331	0.623	0.832	0.919	0.952	0.968	0.968	0.967
					0.6	0.392	0.746	0.923	0.977	0.992	0.993	0.995	0.996
					0.7	0.458	0.844	0.972	0.994	0.999	0.999	0.999	1.000
					0.8	0.535	0.915	0.994	1.000	1.000	1.000	1.000	1.000
					0.9	0.595	0.957	0.999	1.000	1.000	1.000	1.000	1.000
					1.0	0.669	0.985	1.000	1.000	1.000	1.000	1.000	1.000

Table A.5 (Continued)

						Δ/σ							
Test	m=n	r	k	α	ε	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Quantile	30	6	5	0.098	0.1	0.124	0.174	0.246	0.318	0.392	0.446	0.482	0.493
					0.2	0.156	0.257	0.418	0.601	0.731	0.821	0.861	0.879
					0.3	0.193	0.357	0.584	0.799	0.912	0.964	0.981	0.984
					0.4	0.221	0.457	0.718	0.906	0.976	0.995	0.999	1.000
					0.5	0.251	0.535	0.812	0.956	0.994	0.999	1.000	1.000
					0.6	0.293	0.612	0.880	0.979	0.998	1.000	1.000	1.000
					0.7	0.325	0.678	0.919	0.987	1.000	1.000	1.000	1.000
					0.8	0.360	0.735	0.943	0.994	1.000	1.000	1.000	1.000
					0.9	0.400	0.777	0.962	0.996	1.000	1.000	1.000	1.000
					1.0	0.430	0.824	0.973	0.999	1.000	1.000	1.000	1.000
WRS				0.100	0.1	0.138	0.179	0.212	0.239	0.256	0.264	0.269	0.265
					0.2	0.177	0.279	0.379	0.448	0.483	0.518	0.521	0.526
					0.3	0.241	0.412	0.563	0.665	0.726	0.755	0.762	0.776
					0.4	0.292	0.542	0.741	0.852	0.895	0.921	0.926	0.922
					0.5	0.358	0.685	0.883	0.950	0.974	0.982	0.987	0.987
					0.6	0.440	0.804	0.953	0.989	0.995	0.998	0.998	0.999
					0.7	0.505	0.893	0.987	0.998	1.000	1.000	1.000	1.000
					0.8	0.587	0.949	0.998	1.000	1.000	1.000	1.000	1.000
					0.9	0.663	0.980	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.730	0.993	1.000	1.000	1.000	1.000	1.000	1.000
Quantile	40	6	5	0.098	0.1	0.134	0.192	0.278	0.393	0.507	0.582	0.624	0.652
					0.2	0.168	0.294	0.492	0.694	0.844	0.924	0.954	0.968
					0.3	0.198	0.403	0.662	0.879	0.966	0.993	0.997	0.999
					0.4	0.239	0.515	0.790	0.946	0.992	0.999	1.000	1.000
					0.5	0.285	0.593	0.874	0.975	0.997	1.000	1.000	1.000
					0.6	0.325	0.665	0.913	0.989	1.000	1.000	1.000	1.000
					0.7	0.360	0.730	0.943	0.995	1.000	1.000	1.000	1.000
					0.8	0.391	0.776	0.962	0.997	1.000	1.000	1.000	1.000
					0.9	0.430	0.811	0.973	0.998	1.000	1.000	1.000	1.000
					1.0	0.465	0.848	0.980	0.999	1.000	1.000	1.000	1.000
				0.100	0.1	0.139	0.189	0.228	0.264	0.281	0.296	0.301	0.303
					0.2	0.197	0.310	0.418	0.501	0.560	0.584	0.601	0.600
					0.3	0.268	0.473	0.647	0.761	0.816	0.839	0.848	0.850
					0.4	0.336	0.635	0.832	0.917	0.951	0.963	0.969	0.969
					0.5	0.423	0.768	0.939	0.983	0.993	0.996	0.996	0.997
					0.6	0.500	0.879	0.986	0.998	0.999	0.999	1.000	1.000
					0.7	0.591	0.947	0.999	1.000	1.000	1.000	1.000	1.000
					0.8	0.672	0.983	1.000	1.000	1.000	1.000	1.000	1.000
					0.9	0.743	0.995	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.818	0.998	1.000	1.000	1.000	1.000	1.000	1.000

Table A.5 (Continued)

						Δ/σ							
Test	m=n	r	k	α	ϵ	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Quantile	50	6	5	0.102	0.1	0.137	0.205	0.310	0.462	0.588	0.694	0.744	0.771
					0.2	0.179	0.326	0.548	0.768	0.913	0.966	0.987	0.992
					0.3	0.215	0.440	0.719	0.914	0.985	0.997	1.000	1.000
					0.4	0.256	0.544	0.834	0.966	0.997	1.000	1.000	1.000
					0.5	0.298	0.631	0.897	0.983	0.999	1.000	1.000	1.000
					0.6	0.340	0.707	0.938	0.994	1.000	1.000	1.000	1.000
					0.7	0.378	0.761	0.957	0.997	1.000	1.000	1.000	1.000
					0.8	0.425	0.804	0.970	0.999	1.000	1.000	1.000	1.000
					0.9	0.456	0.846	0.980	0.999	1.000	1.000	1.000	1.000
					1.0	0.482	0.875	0.986	0.999	1.000	1.000	1.000	1.000
WRS				0.100	0.1	0.145	0.209	0.250	0.289	0.318	0.330	0.340	0.341
					0.2	0.214	0.348	0.480	0.566	0.633	0.668	0.672	0.681
					0.3	0.283	0.536	0.718	0.824	0.871	0.896	0.908	0.904
					0.4	0.379	0.707	0.885	0.957	0.979	0.987	0.985	0.987
					0.5	0.468	0.838	0.971	0.995	0.998	0.999	0.999	0.999
					0.6	0.554	0.931	0.996	0.999	1.000	1.000	1.000	1.000
					0.7	0.652	0.978	1.000	1.000	1.000	1.000	1.000	1.000
					0.8	0.741	0.993	1.000	1.000	1.000	1.000	1.000	1.000
					0.9	0.824	0.999	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.877	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Quantile	60	6	5	0.098	0.1	0.143	0.212	0.331	0.504	0.665	0.790	0.839	0.862
					0.2	0.179	0.345	0.596	0.833	0.945	0.986	0.997	0.998
					0.3	0.219	0.476	0.760	0.941	0.991	1.000	1.000	1.000
					0.4	0.268	0.568	0.861	0.977	0.997	1.000	1.000	1.000
					0.5	0.307	0.668	0.916	0.990	0.999	1.000	1.000	1.000
					0.6	0.356	0.734	0.950	0.996	1.000	1.000	1.000	1.000
					0.7	0.391	0.786	0.968	0.998	1.000	1.000	1.000	1.000
					0.8	0.427	0.826	0.978	0.998	1.000	1.000	1.000	1.000
					0.9	0.476	0.856	0.984	0.999	1.000	1.000	1.000	1.000
					1.0	0.492	0.889	0.989	1.000	1.000	1.000	1.000	1.000
WRS				0.100	0.1	0.161	0.214	0.274	0.312	0.342	0.359	0.366	0.366
					0.2	0.223	0.381	0.528	0.628	0.684	0.719	0.727	0.728
					0.3	0.316	0.571	0.773	0.873	0.915	0.933	0.940	0.945
					0.4	0.410	0.753	0.930	0.978	0.990	0.994	0.994	0.995
					0.5	0.504	0.881	0.986	0.999	1.000	1.000	1.000	1.000
					0.6	0.623	0.959	0.998	1.000	1.000	1.000	1.000	1.000
					0.7	0.718	0.990	1.000	1.000	1.000	1.000	1.000	1.000
					0.8	0.798	0.998	1.000	1.000	1.000	1.000	1.000	1.000
					0.9	0.867	1.000	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.913	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table A.5 (Continued)

						Δ/σ							
Test	m=n	r	k	α	ϵ	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
Quantile	75	6	5	0.102	0.1	0.142	0.226	0.382	0.577	0.748	0.867	0.917	0.942
					0.2	0.188	0.370	0.638	0.868	0.975	0.995	0.999	1.000
					0.3	0.230	0.504	0.807	0.963	0.997	1.000	1.000	1.000
					0.4	0.281	0.608	0.893	0.985	0.999	1.000	1.000	1.000
					0.5	0.316	0.699	0.942	0.995	1.000	1.000	1.000	1.000
					0.6	0.363	0.762	0.963	0.997	1.000	1.000	1.000	1.000
					0.7	0.406	0.816	0.974	0.998	1.000	1.000	1.000	1.000
					0.8	0.445	0.844	0.981	1.000	1.000	1.000	1.000	1.000
					0.9	0.491	0.880	0.989	1.000	1.000	1.000	1.000	1.000
					1.0	0.536	0.905	0.991	1.000	1.000	1.000	1.000	1.000
WRS				0.100	0.1	0.163	0.237	0.295	0.354	0.377	0.391	0.415	0.412
					0.2	0.235	0.417	0.585	0.704	0.757	0.779	0.795	0.798
					0.3	0.341	0.646	0.846	0.923	0.954	0.965	0.973	0.975
					0.4	0.464	0.828	0.964	0.991	0.996	0.998	0.998	0.999
					0.5	0.588	0.937	0.996	1.000	1.000	1.000	1.000	1.000
					0.6	0.686	0.982	0.999	1.000	1.000	1.000	1.000	1.000
					0.7	0.782	0.996	1.000	1.000	1.000	1.000	1.000	1.000
					0.8	0.866	1.000	1.000	1.000	1.000	1.000	1.000	1.000
					0.9	0.917	1.000	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.956	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Quantile	100	6	5	0.104	0.1	0.145	0.248	0.435	0.665	0.847	0.939	0.975	0.986
					0.2	0.192	0.402	0.709	0.922	0.988	0.999	1.000	1.000
					0.3	0.232	0.549	0.851	0.979	0.999	1.000	1.000	1.000
					0.4	0.294	0.658	0.920	0.994	1.000	1.000	1.000	1.000
					0.5	0.342	0.735	0.954	0.996	1.000	1.000	1.000	1.000
					0.6	0.389	0.793	0.975	0.998	1.000	1.000	1.000	1.000
					0.7	0.436	0.845	0.982	0.999	1.000	1.000	1.000	1.000
					0.8	0.468	0.879	0.988	1.000	1.000	1.000	1.000	1.000
					0.9	0.513	0.895	0.992	1.000	1.000	1.000	1.000	1.000
					1.0	0.551	0.919	0.995	1.000	1.000	1.000	1.000	1.000
WRS				0.100	0.1	0.178	0.258	0.345	0.398	0.442	0.464	0.479	0.483
					0.2	0.286	0.494	0.681	0.780	0.837	0.861	0.874	0.875
					0.3	0.396	0.737	0.908	0.970	0.984	0.992	0.992	0.993
					0.4	0.530	0.904	0.986	0.998	1.000	1.000	1.000	1.000
					0.5	0.663	0.975	0.999	1.000	1.000	1.000	1.000	1.000
					0.6	0.780	0.996	1.000	1.000	1.000	1.000	1.000	1.000
					0.7	0.864	1.000	1.000	1.000	1.000	1.000	1.000	1.000
					0.8	0.934	1.000	1.000	1.000	1.000	1.000	1.000	1.000
					0.9	0.964	1.000	1.000	1.000	1.000	1.000	1.000	1.000
					1.0	0.984	1.000	1.000	1.000	1.000	1.000	1.000	1.000

TABLE A.6 Values of r, k, and α for the Quantile Test for Combinations of m and n When α is Approximately Equal to 0.01

Number of Cleanup-Unit Measurements, n

Number of Reference-Area Measurements, m

	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
5			11,11 0.008	13,13 0.015	16,16 0.014	19,19 0.013	22,22 0.013	25,25 0.013	28,28 0.012									r,k α		
10		6,6 0.005	7,7 0.013	9,9 0.012	11,11 0.011	13,13 0.010	14,14 0.014	16,16 0.013	18,18 0.012	19,19 0.015	21,21 0.014	23,23 0.013	25,25 0.012	26,26 0.015	28,28 0.014	30,30 0.013				
15	3,3 0.009	7,6 0.007	6,6 0.008	7,7 0.011	8,8 0.014	10,10 0.009	11,11 0.011	12,12 0.013	13,13 0.014	15,15 0.011	16,16 0.012	17,17 0.013	18,18 0.014	19,19 0.015	21,21 0.012	22,22 0.013	23,23 0.014	24,24 0.015	26,26 0.013	27,27 0.013
20	6,4 0.005	4,4 0.008	5,5 0.009	6,6 0.010	7,7 0.011	8,8 0.011	9,9 0.011	10,10 0.011	11,11 0.011	12,12 0.011	13,13 0.011	14,14 0.012	15,15 0.012	16,16 0.012	17,17 0.012	18,18 0.012	19,19 0.012	19,19 0.015	20,20 0.015	21,21 0.015
25	4,3 0.009	7,5 0.012	4,4 0.015	5,5 0.013	6,6 0.011	7,7 0.010	8,8 0.009	9,9 0.009	9,9 0.014	10,10 0.012	11,11 0.011	12,12 0.011	12,12 0.015	13,13 0.014	14,14 0.013	15,15 0.012	16,16 0.011	16,16 0.014	17,17 0.014	18,18 0.013
30	4,3 0.006	3,3 0.012	4,4 0.009	5,5 0.007	6,6 0.006	6,6 0.012	7,7 0.010	8,8 0.008	8,8 0.013	9,9 0.011	10,10 0.009	10,10 0.013	11,11 0.011	12,11 0.014	12,12 0.013	13,13 0.012	14,14 0.011	14,14 0.014	15,15 0.012	15,15 0.015
35	2,2 0.013	3,3 0.008	4,4 0.006	4,4 0.014	5,5 0.010	6,6 0.007	6,6 0.012	7,7 0.009	7,7 0.014	8,8 0.011	9,9 0.009	9,9 0.013	10,10 0.010	10,10 0.014	11,11 0.011	11,11 0.015	12,12 0.012	13,13 0.011	13,13 0.013	14,14 0.012
40	2,2 0.010	3,3 0.006	7,5 0.013	4,4 0.010	5,5 0.006	5,5 0.012	6,6 0.008	6,6 0.013	7,7 0.009	7,7 0.013	8,8 0.010	8,8 0.014	9,9 0.011	9,9 0.014	10,10 0.011	10,10 0.014	11,11 0.012	11,11 0.014	12,12 0.012	12,12 0.014
45	2,2 0.008	6,4 0.008	3,3 0.013	4,4 0.007	4,4 0.014	5,5 0.008	5,5 0.014	6,6 0.009	6,6 0.013	7,7 0.009	7,7 0.013	8,8 0.009	8,8 0.012	9,9 0.009	9,9 0.012	10,10 0.009	10,10 0.012	10,10 0.015	11,11 0.012	11,11 0.014
50		4,3 0.013	3,3 0.010	4,4 0.005	4,4 0.010	5,5 0.006	5,5 0.010	5,5 0.015	6,6 0.009	6,6 0.013	7,7 0.009	7,7 0.012	8,8 0.009	8,8 0.011	8,8 0.014	9,9 0.011	9,9 0.013	10,10 0.010	10,10 0.012	10,10 0.015
55		4,3 0.010	3,3 0.008	7,5 0.013	4,4 0.008	4,4 0.014	5,5 0.007	5,5 0.011	6,6 0.007	6,6 0.010	6,6 0.014	7,7 0.009	7,7 0.012	8,8 0.008	8,8 0.010	8,8 0.013	9,9 0.009	9,9 0.012	9,9 0.014	10,10 0.011
60		4,3 0.008	3,3 0.007	3,3 0.014	4,4 0.006	4,4 0.011	5,5 0.006	5,5 0.009	5,5 0.013	6,6 0.007	6,6 0.010	6,6 0.014	7,7 0.009	7,7 0.011	7,7 0.014	8,8 0.010	8,8 0.012	8,8 0.015	9,9 0.010	9,9 0.013
65		4,3 0.007	3,3 0.006	3,3 0.012	6,5 0.006	4,4 0.009	4,4 0.013	5,5 0.007	5,5 0.010	5,5 0.014	6,6 0.008	6,6 0.011	6,6 0.014	7,7 0.009	7,7 0.011	7,7 0.014	8,8 0.009	8,8 0.011	8,8 0.014	9,9 0.010
70		2,2 0.014	6,4 0.008	3,3 0.010	7,5 0.013	4,4 0.007	4,4 0.011	5,5 0.005	5,5 0.008	5,5 0.011	5,5 0.015	6,6 0.008	6,6 0.011	6,6 0.014	7,7 0.009	7,7 0.011	7,7 0.013	8,8 0.009	8,8 0.011	8,8 0.013
75		2,2 0.013	4,3 0.014	3,3 0.008	3,3 0.014	4,4 0.006	4,4 0.009	4,4 0.013	5,5 0.006	5,5 0.009	5,5 0.012	6,6 0.007	6,6 0.009	6,6 0.011	6,6 0.014	7,7 0.009	7,7 0.011	7,7 0.013	8,8 0.008	8,8 0.010
80		2,2 0.011	4,3 0.012	3,3 0.007	3,3 0.012	6,5 0.006	4,4 0.008	4,4 0.011	5,5 0.005	5,5 0.007	5,5 0.010	5,5 0.013	6,6 0.007	6,6 0.009	6,6 0.012	6,6 0.014	7,7 0.009	7,7 0.010	7,7 0.013	7,7 0.015
85		2,2 0.010	4,3 0.010	3,3 0.006	3,3 0.011	7,5 0.013	4,4 0.006	4,4 0.009	4,4 0.013	5,5 0.006	5,5 0.008	5,5 0.011	5,5 0.014	6,6 0.008	6,6 0.010	6,6 0.012	6,6 0.014	7,7 0.008	7,7 0.010	7,7 0.012
90			4,3 0.009	3,3 0.005	3,3 0.009	3,3 0.014	4,4 0.005	4,4 0.008	4,4 0.011	5,5 0.005	5,5 0.007	5,5 0.009	5,5 0.012	5,5 0.015	6,6 0.008	6,6 0.010	6,6 0.012	6,6 0.014	7,7 0.008	7,7 0.010
95			4,3 0.008	6,4 0.008	3,3 0.008	3,3 0.013	6,5 0.005	4,4 0.007	4,4 0.010	4,4 0.013	5,5 0.006	5,5 0.008	5,5 0.010	5,5 0.013	6,6 0.007	6,6 0.008	6,6 0.010	6,6 0.012	6,6 0.014	7,7 0.008
100			4,3 0.007	4,3 0.014	3,3 0.007	3,3 0.011	7,5 0.013	4,4 0.006	4,4 0.008	4,4 0.011	4,4 0.015	5,5 0.007	5,5 0.009	5,5 0.011	5,5 0.013	6,6 0.007	6,6 0.008	6,6 0.010	6,6 0.012	6,6 0.014

TABLE A.7 Values of r, k, and α for the Quantile Test for Combinations of m and n When α is Approximately Equal to 0.025

Number of Cleanup-Unit Measurements, n

Number of Reference-Area Measurements, m

	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
5			9,9 0.030	12,12 0.024	15,15 0.021	17,17 0.026	20,20 0.024	22,22 0.028	25,25 0.025									r,k α		
10		7,6 0.029	6,6 0.028	8,8 0.022	9,9 0.029	11,11 0.024	12,12 0.029	14,14 0.025	15,15 0.029	17,17 0.025	18,18 0.029	20,20 0.026	21,21 0.029	23,23 0.026	24,24 0.029	26,26 0.026	27,27 0.029			
15	11,5 0.030	6,5 0.023	5,5 0.021	6,6 0.024	7,7 0.026	8,8 0.027	9,9 0.028	10,10 0.029	11,11 0.030	13,13 0.022	14,14 0.023	15,15 0.023	16,16 0.024	17,17 0.025	18,18 0.025	19,19 0.026	21,21 0.021	21,21 0.027	22,22 0.027	23,23 0.027
20	8,4 0.023	3,3 0.030	4,4 0.026	5,5 0.024	6,6 0.022	7,7 0.020	12,11 0.021	13,12 0.024	9,9 0.028	10,10 0.026	11,11 0.024	12,12 0.023	13,13 0.022	13,13 0.029	14,14 0.027	15,15 0.026	16,16 0.025	17,17 0.024	17,17 0.029	18,18 0.028
25	2,2 0.023	8,5 0.027	6,5 0.021	7,6 0.023	5,5 0.025	6,6 0.020	10,9 0.026	7,7 0.027	8,8 0.023	13,12 0.027	9,9 0.027	10,10 0.024	11,11 0.022	11,11 0.028	12,12 0.025	13,13 0.023	13,13 0.028	14,14 0.025	15,15 0.023	15,15 0.028
30	6,3 0.026	6,4 0.026	9,6 0.026	4,4 0.021	7,6 0.029	5,5 0.026	9,8 0.024	6,6 0.029	7,7 0.023	12,11 0.021	8,8 0.025	9,9 0.021	9,9 0.027	10,10 0.023	10,10 0.029	11,11 0.025	11,11 0.030	12,12 0.026	13,13 0.023	13,13 0.027
35	7,3 0.030	4,3 0.030	3,3 0.023	6,5 0.020	4,4 0.026	10,8 0.022	5,5 0.027	9,8 0.024	6,6 0.027	7,7 0.020	7,7 0.027	8,8 0.021	8,8 0.027	9,9 0.022	9,9 0.027	10,10 0.022	10,10 0.027	11,11 0.022	11,11 0.027	12,12 0.023
40	3,2 0.029	4,3 0.022	8,5 0.028	11,7 0.025	6,5 0.028	4,4 0.030	10,8 0.026	5,5 0.027	9,8 0.023	6,6 0.026	10,9 0.028	7,7 0.024	12,11 0.020	8,8 0.023	8,8 0.029	9,9 0.022	9,9 0.027	10,10 0.021	10,10 0.026	11,11 0.021
45	3,2 0.023	8,4 0.029	6,4 0.030	3,3 0.026	8,6 0.021	4,4 0.023	7,6 0.025	5,5 0.020	5,5 0.028	9,8 0.023	6,6 0.024	10,9 0.026	7,7 0.022	7,7 0.027	8,8 0.020	8,8 0.025	8,8 0.030	9,9 0.023	9,9 0.027	10,10 0.021
50		2,2 0.025	6,4 0.022	3,3 0.021	11,7 0.027	6,5 0.026	4,4 0.026	7,6 0.028	5,5 0.021	5,5 0.028	9,8 0.022	6,6 0.023	6,6 0.029	7,7 0.020	7,7 0.025	12,11 0.020	8,8 0.022	8,8 0.026	13,12 0.027	9,9 0.023
55		2,2 0.022	4,3 0.029	8,5 0.028	3,3 0.028	8,6 0.021	4,4 0.020	4,4 0.029	10,8 0.021	5,5 0.022	5,5 0.028	9,8 0.022	6,6 0.023	6,6 0.028	10,9 0.029	7,7 0.023	7,7 0.027	12,11 0.023	8,8 0.023	8,8 0.027
60		14,5 0.022	4,3 0.024	8,5 0.021	3,3 0.023	11,7 0.029	6,5 0.024	4,4 0.023	7,6 0.023	10,8 0.024	5,5 0.023	5,5 0.029	9,8 0.022	6,6 0.022	6,6 0.027	10,9 0.027	7,7 0.021	7,7 0.025	7,7 0.030	8,8 0.021
65		6,3 0.028	7,4 0.021	6,4 0.025	10,6 0.025	3,3 0.029	8,6 0.021	6,5 0.029	4,4 0.026	7,6 0.026	10,8 0.026	5,5 0.023	5,5 0.029	9,8 0.022	6,6 0.021	6,6 0.026	10,9 0.026	7,7 0.020	7,7 0.024	7,7 0.028
70		6,3 0.024	2,2 0.029	6,4 0.021	8,5 0.028	3,3 0.025	13,8 0.026	6,5 0.023	4,4 0.022	4,4 0.023	7,6 0.028	10,8 0.027	5,5 0.024	5,5 0.029	9,8 0.022	6,6 0.021	6,6 0.025	6,6 0.029	10,9 0.030	7,7 0.022
75		11,4 0.022	2,2 0.026	4,3 0.028	8,5 0.022	3,3 0.022	9,6 0.028	8,6 0.021	6,5 0.027	4,4 0.024	7,6 0.023	7,6 0.030	10,8 0.029	5,5 0.024	5,5 0.029	9,8 0.021	6,6 0.021	6,6 0.024	6,6 0.028	10,9 0.028
80		7,3 0.028	2,2 0.024	4,3 0.024	6,4 0.028	10,6 0.024	3,3 0.027	13,8 0.027	6,5 0.023	4,4 0.020	4,4 0.026	7,6 0.024	10,8 0.023	5,5 0.020	5,5 0.025	5,5 0.029	9,8 0.021	6,6 0.020	6,6 0.024	6,6 0.027
85		3,2 0.029	2,2 0.021	4,3 0.021	6,4 0.023	8,5 0.028	3,3 0.023	9,6 0.030	8,6 0.020	6,5 0.026	4,4 0.022	4,4 0.028	7,6 0.026	10,8 0.024	5,5 0.021	5,5 0.025	5,5 0.029	9,8 0.021	6,6 0.020	6,6 0.023
90			5,3 0.020	11,5 0.027	9,5 0.023	8,5 0.023	3,3 0.021	3,3 0.028	13,8 0.028	6,5 0.022	6,5 0.029	4,4 0.024	4,4 0.029	7,6 0.028	10,8 0.026	5,5 0.022	5,5 0.025	5,5 0.030	9,8 0.021	9,8 0.025
95			10,4 0.029	2,2 0.029	4,3 0.028	6,4 0.029	10,6 0.023	3,3 0.025	11,7 0.026	8,6 0.020	6,5 0.025	4,4 0.021	4,4 0.026	7,6 0.024	7,6 0.029	10,8 0.027	5,5 0.022	5,5 0.026	5,5 0.030	9,8 0.021
100			6,3 0.029	2,2 0.027	4,3 0.025	6,4 0.025	8,5 0.028	3,3 0.022	3,3 0.029	13,8 0.028	6,5 0.022	6,5 0.028	4,4 0.023	4,4 0.027	7,6 0.025	10,8 0.022	10,8 0.028	5,5 0.022	5,5 0.026	5,5 0.030

TABLE A.8 Values of r, k, and α for the Quantile Test for Combinations of m and n When α is Approximately Equal to 0.05

Number of Cleanup-Unit Measurements, n

Number of Reference-Area Measurements, m

	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
5			8,8 0.051	10,10 0.057	13,13 0.043	15,15 0.048	17,17 0.051	19,19 0.054	21,21 0.056									r,k α		
10		4,4 0.043	5,5 0.057	14,12 0.045	8,8 0.046	9,9 0.052	10,10 0.058	12,12 0.046	13,13 0.050	14,14 0.054	15,15 0.057	17,17 0.049	18,18 0.052	19,19 0.055	20,20 0.057	21,21 0.059	23,23 0.053			
15	2,2 0.053	3,3 0.052	4,4 0.050	5,5 0.048	6,6 0.046	7,7 0.045	8,8 0.052	9,9 0.043	9,9 0.060	10,10 0.057	11,11 0.055	12,12 0.054	13,13 0.052	14,14 0.051	15,15 0.050	16,16 0.049	16,16 0.058	17,17 0.057	18,18 0.056	19,19 0.055
20	9,4 0.040	8,5 0.056	6,5 0.040	4,4 0.053	5,5 0.043	9,8 0.052	6,6 0.056	7,7 0.048	8,8 0.043	8,8 0.057	9,9 0.051	10,10 0.046	10,10 0.057	11,11 0.052	12,12 0.048	12,12 0.057	13,13 0.053	14,14 0.049	14,14 0.057	15,15 0.054
25	6,3 0.041	6,4 0.043	3,3 0.046	6,5 0.052	4,4 0.055	5,5 0.041	5,5 0.059	6,6 0.046	11,10 0.042	7,7 0.050	8,8 0.042	8,8 0.053	9,9 0.045	9,9 0.055	10,10 0.048	11,11 0.042	11,11 0.050	11,11 0.058	12,12 0.052	12,12 0.060
30	3,2 0.047	2,2 0.058	10,6 0.052	3,3 0.058	11,8 0.045	4,4 0.056	8,7 0.[illegible]	5,5 0.[illegible]	6,6 0.[illegible]	6,6 0.053	7,7 0.041	7,7 0.052	8,8 0.042	8,8 0.051	9,9 0.042	9,9 0.050	9,9 0.059	10,10 0.049	10,10 0.057	11,11 0.049
35	8,3 0.046	2,2 0.045	6,4 0.058	3,3 0.043	6,5 0.041	4,4 0.040	4,4 0.057	8,7 0.043	5,5 0.051	9,8 0.052	6,6 0.047	6,6 0.057	7,7 0.043	7,7 0.053	8,8 0.041	8,8 0.049	8,8 0.057	9,9 0.046	9,9 0.053	10,10 0.044
40	4,2 0.055	5,3 0.048	4,3 0.057	10,6 0.059	3,3 0.053	6,5 0.048	4,4 0.043	4,4 0.058	8,7 0.042	5,5 0.048	9,8 0.047	6,6 0.042	6,6 0.051	11,10 0.042	7,7 0.045	7,7 0.053	8,8 0.041	8,8 0.048	8,8 0.055	9,9 0.043
45	4,2 0.045	9,4 0.047	2,2 0.059	8,5 0.052	3,3 0.042	8,6 0.041	6,5 0.054	4,4 0.045	4,4 0.058	8,7 0.041	5,5 0.046	5,5 0.057	9,8 0.056	6,6 0.047	6,6 0.055	11,10 0.046	7,7 0.047	7,7 0.054	8,8 0.041	8,8 0.047
50		6,3 0.052	2,2 0.050	6,4 0.051	12,7 0.050	3,3 0.049	8,6 0.049	6,5 0.059	4,4 0.047	4,4 0.059	8,7 0.041	5,5 0.045	5,5 0.054	9,8 0.051	6,6 0.043	6,6 0.050	6,6 0.058	7,7 0.042	7,7 0.048	7,7 0.054
55		3,2 0.059	2,2 0.043	4,3 0.056	8,5 0.058	3,3 0.041	5,4 0.041	6,5 0.046	9,7 0.042	4,4 0.048	4,4 0.059	8,7 0.040	5,5 0.043	5,5 0.052	9,8 0.048	6,6 0.040	6,6 0.047	6,6 0.054	11,10 0.043	7,7 0.043
60		3,2 0.052	5,3 0.052	4,3 0.046	6,4 0.059	3,3 0.035	3,3 0.047	8,6 0.043	6,5 0.051	9,7 0.046	4,4 0.049	4,4 0.059	13,10 0.052	5,5 0.042	5,5 0.050	5,5 0.058	9,8 0.054	6,6 0.044	6,6 0.050	6,6 0.056
65		3,2 0.045	5,3 0.043	2,2 0.053	6,4 0.048	10,6 0.050	3,3 0.040	3,3 0.052	6,5 0.041	6,5 0.055	4,4 0.042	4,4 0.050	4,4 0.060	13,10 0.052	5,5 0.041	5,5 0.048	5,5 0.055	9,8 0.051	6,6 0.041	6,6 0.047
70		8,3 0.057	9,4 0.048	2,2 0.047	4,3 0.055	8,5 0.050	5,4 0.041	3,3 0.046	3,3 0.057	6,5 0.045	6,5 0.058	4,4 0.043	4,4 0.051	4,4 0.060	13,10 0.051	5,5 0.041	5,5 0.047	5,5 0.054	9,8 0.048	9,8 0.057
75		8,3 0.049	6,3 0.056	2,2 0.043	4,3 0.047	6,4 0.054	10,6 0.053	3,3 0.040	3,3 0.051	8,6 0.044	6,5 0.049	9,7 0.041	4,4 0.044	4,4 0.052	5,5 0.060	13,10 0.051	8,7 0.047	5,5 0.046	5,5 0.052	5,5 0.058
80		4,2 0.059	6,3 0.048	5,3 0.053	2,2 0.055	6,4 0.046	8,5 0.055	5,4 0.042	3,3 0.045	3,3 0.055	6,5 0.041	6,5 0.052	9,7 0.043	4,4 0.045	4,4 0.053	7,6 0.058	13,10 0.051	8,7 0.046	5,5 0.045	5,5 0.051
85		4,2 0.054	3,2 0.058	5,3 0.047	2,2 0.050	4,3 0.054	4,3 0.048	10,6 0.056	5,4 0.049	3,3 0.049	3,3 0.059	6,5 0.044	6,5 0.055	9,7 0.046	4,4 0.046	4,4 0.053	7,6 0.059	10,8 0.060	8,7 0.045	5,5 0.044
90			3,2 0.053	5,3 0.041	2,2 0.046	6,4 0.059	6,4 0.051	8,5 0.058	5,4 0.042	3,3 0.044	3,3 0.053	8,6 0.045	6,5 0.047	6,5 0.058	4,4 0.041	4,4 0.047	4,4 0.054	7,6 0.059	10,8 0.060	8,7 0.041
95			3,2 0.048	9,4 0.048	2,2 0.042	2,2 0.056	4,3 0.059	8,5 0.050	10,6 0.058	5,4 0.048	3,3 0.048	3,3 0.056	6,5 0.041	6,5 0.050	9,7 0.040	4,4 0.042	4,4 0.048	4,4 0.054	7,6 0.059	10,8 0.059
100			3,2 0.044	6,3 0.057	5,3 0.054	2,2 0.052	4,3 0.053	6,4 0.056	10,6 0.049	5,4 0.043	3,3 0.043	3,3 0.051	3,3 0.059	6,5 0.044	6,5 0.053	9,7 0.042	4,4 0.043	4,4 0.049	4,4 0.055	7,6 0.059

TABLE A.9 Values of r, k, and α for the Quantile Test for Combinations of m and n When α is Approximately Equal to 0.10

Number of Cleanup-Unit Measurements, n

Number of Reference-Area Measurements, m

	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
5			7,7 0.083	8,8 0.116	10,10 0.109	12,12 0.104	14,14 0.100	15,15 0.117	17,17 0.112									r,k α		
10		3,3 0.105	4,4 0.108	5,5 0.109	6,6 0.109	7,7 0.109	8,8 0.109	9,9 0.109	10,10 0.109	11,11 0.109	12,12 0.109	13,13 0.109	14,14 0.109	15,15 0.109	16,16 0.109	17,17 0.109	18,18 0.109			
15	9,4 0.098	10,6 0.106	3,3 0.112	4,4 0.093	5,5 0.081	5,5 0.117	6,6 0.102	7,7 0.092	7,7 0.118	8,8 0.106	9,9 0.098	9,9 0.118	10,10 0.109	11,11 0.101	11,11 0.118	12,12 0.110	13,13 0.104	13,13 0.118	14,14 0.111	15,15 0.106
20	3,2 0.091	2,2 0.103	5,4 0.093	3,3 0.115	4,4 0.085	4,4 0.119	5,5 0.093	10,9 0.084	6,6 0.099	7,7 0.083	7,7 0.102	8,8 0.088	8,8 0.105	9,9 0.092	9,9 0.107	10,10 0.095	10,10 0.108	11,11 0.098	11,11 0.110	12,12 0.100
25	4,2 0.119	7,4 0.084	8,5 0.112	3,3 0.080	3,3 0.117	4,4 0.080	4,4 0.107	8,7 0.108	5,5 0.101	10,9 0.088	6,6 0.096	6,6 0.114	7,7 0.093	7,7 0.108	8,8 0.091	8,8 0.104	8,8 0.117	9,9 0.100	9,9 0.112	10,10 0.098
30	4,2 0.089	5,3 0.089	2,2 0.106	14,8 0.111	3,3 0.088	3,3 0.119	9,7 0.116	4,4 0.100	8,7 0.093	5,5 0.088	5,5 0.106	6,6 0.080	6,6 0.095	6,6 0.110	7,7 0.087	7,7 0.100	7,7 0.113	8,8 0.092	8,8 0.103	8,8 0.115
35	5,2 0.109	3,2 0.119	2,2 0.086	6,4 0.120	5,4 0.091	3,3 0.093	3,3 0.120	9,7 0.112	4,4 0.094	4,4 0.114	8,7 0.107	5,5 0.094	5,5 0.110	6,6 0.081	6,6 0.094	6,6 0.107	6,6 0.120	7,7 0.094	7,7 0.105	7,7 0.116
40	5,2 0.087	3,2 0.098	5,3 0.119	2,2 0.107	12,7 0.109	5,4 0.102	3,3 0.097	6,5 0.100	9,7 0.109	4,4 0.090	4,4 0.107	8,7 0.097	5,5 0.086	5,5 0.099	5,5 0.112	6,6 0.082	6,6 0.093	6,6 0.104	6,6 0.116	7,7 0.089
45	6,2 0.103	3,2 0.082	5,3 0.094	2,2 0.091	6,4 0.115	7,5 0.086	5,4 0.112	3,3 0.100	6,5 0.101	9,7 0.107	4,4 0.087	4,4 0.102	4,4 0.117	8,7 0.107	5,5 0.091	5,5 0.103	5,5 0.115	6,6 0.083	6,6 0.093	6,6 0.103
50		7,3 0.083	9,4 0.115	7,4 0.097	2,2 0.108	10,6 0.112	5,4 0.090	3,3 0.084	3,3 0.103	6,5 0.102	9,7 0.105	4,4 0.084	4,4 0.098	4,4 0.112	8,7 0.099	5,5 0.084	5,5 0.95	5,5 0.105	5,5 0.116	6,6 0.083
55		4,2 0.109	3,2 0.114	5,3 0.114	2,2 0.095	6,4 0.112	14,8 0.111	5,4 0.098	3,3 0.088	3,3 0.104	6,5 0.103	9,7 0.104	4,4 0.082	4,4 0.095	4,4 0.107	4,4 0.120	8,7 0.107	5,5 0.088	5,5 0.098	5,5 0.108
60		4,2 0.095	3,2 0.100	5,3 0.097	2,2 0.084	2,2 0.109	8,5 0.119	5,4 0.082	5,4 0.105	3,3 0.091	3,3 0.106	6,5 0.103	9,7 0.102	4,4 0.081	4,4 0.092	4,4 0.103	4,4 0.115	8,7 0.100	5,5 0.083	5,5 0.092
65		4,2 0.084	3,2 0.089	5,3 0.082	7,4 0.090	2,2 0.097	6,4 0.110	12,7 0.113	5,4 0.089	5,4 0.111	3,3 0.093	3,3 0.108	6,5 0.104	9,7 0.101	7,6 0.084	4,4 0.090	4,4 0.100	4,4 0.110	8,7 0.094	8,7 0.107
70		5,2 0.115	7,3 0.101	9,4 0.106	5,3 0.112	2,2 0.088	2,2 0.109	8,5 0.114	7,5 0.081	5,4 0.096	3,3 0.083	3,3 0.096	3,3 0.109	6,5 0.104	9,7 0.101	7,6 0.082	4,4 0.088	4,4 0.097	4,4 0.107	4,4 0.117
75		5,2 0.103	7,3 0.088	3,2 0.111	5,3 0.098	7,4 0.101	2,2 0.099	2,2 0.119	10,6 0.117	5,4 0.083	5,4 0.102	3,3 0.085	3,3 0.098	3,3 0.110	6,5 0.105	9,7 0.100	7,6 0.081	4,4 0.086	4,4 0.095	4,4 0.104
80		5,2 0.093	4,2 0.116	3,2 0.101	5,3 0.086	7,4 0.086	2,2 0.091	2,2 0.109	8,5 0.111	14,8 0.110	5,4 0.089	5,4 0.107	3,3 0.088	3,3 0.099	3,3 0.111	6,5 0.105	6,5 0.120	9,7 0.116	4,4 0.084	4,4 0.093
85		5,2 0.084	4,2 0.106	3,2 0.092	9,4 0.117	5,3 0.111	2,2 0.083	2,2 0.101	2,2 0.118	10,6 0.112	7,5 0.084	5,4 0.094	5,4 0.111	3,3 0.090	3,3 0.101	3,3 0.112	6,5 0.105	6,5 0.119	9,7 0.114	4,4 0.083
90			4,2 0.097	3,2 0.085	3,2 0.119	5,3 0.099	7,4 0.095	2,2 0.093	2,2 0.109	8,5 0.108	12,7 0.114	5,4 0.083	5,4 0.099	3,3 0.082	3,3 0.092	3,3 0.102	3,3 0.113	6,5 0.105	6,5 0.119	9,7 0.113
95			4,2 0.089	7,3 0.100	3,2 0.110	5,3 0.089	7,4 0.084	2,2 0.086	2,2 0.102	2,2 0.117	10,6 0.108	14,8 0.117	5,4 0.088	5,4 0.103	3,3 0.084	3,3 0.094	3,3 0.103	3,3 0.113	6,5 0.106	6,5 0.118
100			4,2 0.082	7,3 0.090	3,2 0.102	5,3 0.080	5,3 0.109	2,2 0.080	2,2 0.095	2,2 0.110	6,4 0.118	12,7 0.109	7,5 0.086	5,4 0.093	5,4 0.108	3,3 0.086	3,3 0.095	3,3 0.104	3,3 0.114	6,5 0.106

APPENDIX B

GLOSSARY

APPENDIX B

GLOSSARY

Alpha (α) The specified maximum probability of a Type I Error, i.e., the maximum probability of rejecting the null hypothesis when it is true. In the context of this document, α is the maximum acceptable probability that a statistical test incorrectly indicates that a cleanup unit does not attain the cleanup standard. See Section 2.3.

Alternative Hypothesis See Hypothesis

Attainment Objectives Specifying the design and scope of the sampling study including the chemicals to be tested, the cleanup standards to be attained, the measure or parameter to be compared to the cleanup standard, and the Type I and Type II error rates for the selected statistical tests. See Section 4.1.1 and Chapters 6 and 7.

ARAR Applicable or Relevant and Appropriate Requirement. See Chapter 1.

Beta (β) The probability of a Type II Error, i.e., the probability of accepting the null hypothesis when it is false. In the context of this document, β is the specified, allowable (small) probability that a statistical test incorrectly indicates that the cleanup unit has been successfully remediated. $\beta = 1$ - Power. See Power. See Section 2.3.

c The proportion of the total number of samples in the reference area and cleanup unit that are to be taken in the reference area. c is used with the Wilcoxon Rank Sum (WRS) Test. See Section 6.2.

Cleanup Unit A geographical area of specified size and shape at a remediated Superfund site for which a separate decision will be made whether the unit attains the site-specific reference-based cleanup standard for the designated pollution parameter. See Section 4.2.1.

Cleanup Standard In the context of this document, the cleanup standard for the Wilcoxon Rank Sum (WRS) test and for the Quantile test are specific values of statistical parameters. For the WRS test, the standard is $P_r = 1/2$. For the Quantile test, the standard is $\epsilon = 0$ and $\Delta/\sigma = 0$. See Sections 4.4, 6.1 and 7.1.

Composite Sample A sample formed by collecting several samples and combining them (or selected portions of them) into a new sample which is then thoroughly mixed. See Sections 3.3 and 4.3.1.

DQOs (Data Quality Objectives) Qualitative and quantitative statements that specify the type and quality of data that are required for the specified objective. See Section 4.1.

d Odds ratio: The quantity "probability a measurement from the cleanup unit is larger than one from the reference area" divided by the quantity "probability a measurement from the cleanup unit is smaller than one from the reference area." The odds ratio can be used in place of P_r when determining the number of measurements needed for the Wilcoxon Rank Sum test. See Section 6.2.2.1.

Delta (Δ) The amount that the distribution of measurements for the cleanup unit is shifted to the right of the distribution of measurements of the reference area. In this document, Δ is always divided by σ, the standard deviation of the measurements, so that the shift is always in multiples of standard deviations. See Sections 6.2.2.2 and 7.1.

Design Specification Process The process of determining the sampling and analysis procedures that are needed to demonstrate that the attainment objectives have been achieved. See Sections 4.1.2 and 4.2.

Epsilon (ϵ) The proportion of soil in a cleanup unit that has not been remediated to the reference-based cleanup standard. ϵ is used in the Quantile test. See Section 4.4.2 and Chapter 7.

F A factor used to increase N for the Wilcoxon Rank Sum test to account for unequal m and n. See **N**, **m**, and **n**. See Section 6.2.2.2.

Hot Measurement A measurement of soil for a specified pollution parameter that exceeds the value of H_m established for that pollution parameter. See H_m. See Section 4.4.3

Hypothesis An assumption about a property or characteristic of a population under study. The goal of statistical inference is to decide which of two complementary hypotheses is likely to be true (from USEPA 1989a). In the context of this document, the <u>null hypothesis</u> is that the cleanup unit has been successfully remediated and the <u>alternative hypothesis</u> is that the cleanup unit has not been successfully remediated. See Sections 2.2, 6.1 and 7.1.

H_m A concentration value such that any measurement from the cleanup unit at the remediated site that is larger than H_m indicates an area of relatively high concentration that must be removed. The "H_m test" is used in conjunction with both the Wilcoxon Rank Sum test and the Quantile test. See Section 4.4.3.

h The number of cleanup units that will be compared to a specified reference area. See Section 6.2.1

k	When conducting the Quantile test, k is the number of measurements from the cleanup unit that are among the r largest measurements of the combined set of reference area and cleanup unit measurements. See Quantile test. See P. See Sections 7.2 and 7.3.
Less-Than Data	Measurements that are less than the limit of detection. The tests in this document allow for less-than data to occur. See Sections 3.6, 6.3, 7.2 and 7.3.
m	The number of measurements required from the reference area to conduct a statistical test with specified Type I and Type II error rates. See Sections 6.2 and 7.2.
Missing or Unusable Data	Data (measurements) that are mislabeled, lost, held too long before analysis, or do not meet quality control standards. In this document "less-than" data are not considered to be missing or unusable data. See R. See Sections 3.10, 6.2 and 7.2.
Multiple-Comparison Test	A test constructed so that the Type I error rate for a whole group of individual tests does not exceed a specified α level. In the context of this document, many tests may be needed at a Superfund site because of multiple pollutants, cleanup areas, times, etc. See Section 3.5.
N	$N = m + n =$ the total number of measurements required from the reference area and a cleanup unit being compared with the reference area. See m and n. See Sections 6.2 and 7.2
n	Number of measurements required from the cleanup unit to conduct a statistical test that has specified Type I and Type II error rates. See Sections 6.2 and 7.2.
n_f	The number of samples that should be collected in an area to assure that the required number of measurements from that area for conducting statistical tests is obtained. $n_f = n/(1 - R)$. See R. See Sections 3.10, 6.2, and 7.2.
Nonparametric Test	A test based on relatively few assumptions about the exact form of the underlying probability distributions of the measurements. As a consequence, nonparametric tests are valid for a fairly broad class of distributions. The Wilcoxon Rank Sum test and the Quantile test are nonparametric tests. See Section 3.1 and Chapters 6 and 7.
Normal (Gaussian) Distribution	A family of bell-shaped distributions described by the mean and variance, μ and σ^2. Refer to a statistical text (e.g., Gilbert 1987) for a formal definition. See Standard Normal Distribution. See Sections 3.1, 6.2, and 7.3.
Outlier	Measurements that are unusually large relative to the bulk of the measurements in the data set. See Section 3.7.

P When conducting the Quantile test, P is the probability of obtaining a value of k as large or larger than the observed K if the null hypothesis is true. See k. See Section 7.3.2.

Power $(1 - \beta)$ The probability of rejecting the null hypothesis when it is false. Power = 1 - Type II error rate. In the context of this document, the power of a test is the probability the test will correctly indicate when a cleanup unit has not been successfully remediated. See **Beta (β)**. See Section 2.3 and Chapters 6 and 7.

P_r The probability that a measurement of a sample collected at a random location in the cleanup unit is greater than a measurement of a sample collected at a random location in the reference area. See Section 4.4.1 and Chapter 6.

Quantile Test A nonparametric test, illustrated in Chapter 7, that looks at only the r largest measurements of the N combined reference area and cleanup unit measurements. If a sufficiently large number of these r measurements are from the cleanup unit, then the test indicates the remediated cleanup unit has not attained the reference-based cleanup standard. See Section 4.4.2 and Chapter 7.

R The rate of missing or unusable pollution parameter measurements expected to occur for samples collected in reference areas or cleanup units. See **Missing or Unusable Data**. See n_f.

Reference Areas Geographical areas from which representative reference samples will be selected for comparison with samples collected in specific cleanup units at the remediated Superfund site. See Section 4.2.1.

Reference Region The geographical region from which reference areas will be selected for comparison with cleanup units. See Section 4.2.1.

Representative Measurement A measurement that is selected using a procedure in such a way that it, in combination with other representative measurements, will give an accurate picture of the phenomenon being studied.

Standard Normal Distribution A normal (Gaussian) distribution with $\mu = 0$ and $\sigma^2 = 1$. See **Normal (Gaussian) Distribution**. See Table A.1.

Stratified Random Sampling In the context of this document, stratified random sampling refers to dividing the Superfund Site into nonoverlapping cleanup units and collecting soil samples at randomly selected locations within each cleanup unit. See Section 5.1

Tandem Testing When two or more statistical tests are conducted using the same data set. See Section 4.5 and Chapters 6 and 7.

Tied Measurements Two or more measurements that have the same value. See Sections 6.3 and 7.2.

Triangular Sampling Grid A grid of sampling locations that is arranged in a triangular pattern. See Chapter 5.

Two-Sample t Test A test described in most statistics books that may be used in place of the Wilcoxon Rank Sum test if the reference area and cleanup unit measurements are known to be normally (Gaussian) distributed and there are no less-than measurements in either data set. See Section 6.4.

Wilcoxon Rank Sum (WRS) Test The nonparametric test, illustrated in Chapter 6, to detect when the remedial action has failed more or less uniformly throughout the cleanup unit to achieve the reference-based cleanup standard. See Section 4.4.1 and Chapter 6.

$Z_{1-\phi}$ A value from the standard normal distribution that cuts off $(100\phi)\%$ of the upper tail of the standard normal distribution. See Standard Normal Distribution.