&EPA

Methods For Evaluating The Attainment Of Cleanup Standards

Volume 2: Ground Water

United States
Environmental Protection
Agency

Policy, Planning
And Evaluation
(PM-222)

EPA 230-R-92-014
July 1992

Methods for Evaluating the Attainment of Cleanup Standards Volume 2: Ground Water

DISCLAIMER

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF FIGURES

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF BOXES

TABLE OF CONTENTS

xv

TABLE OF CONTENTS

TABLE OF CONTENTS

AUTHORS AND CONTRIBUTORS

This manual represents the combined effort of several organizations and many individuals. The names of the primary contributors, along with the role of each organization, are summarized below.

Westat, Inc., 1650 Research Boulevard, Rockville, Md 20850, Contract No. 68-0l-7359, Task 11 -- research, statistical procedures, draft and final draft report. Key Westat staff included:

John Rogers	Adam Chu
Ralph DiGaetano	Ed Bryant

Contract Coordinator: Robert Clickner

Dynamac Corporation, 11140 Rockville Pike, Rockville, MD 20852 (subcontractor to Westat) -- consultation on the sampling of ground water, treatment alternatives, and chemical analysis. Key Dynamac staff included:

David Lipsky	Richard Dorrler

Wayne Tusa

EPA, **OPPE, Statistical Policy Branch --** Project management, technical input, peer review. Key EPA staff included:

Barnes Johnson	Herbert Lacayo

SRA Technologies, Inc., 4700 King Street, Suite 300, Alexandria, VA 22302, Contract No. 68-01-7379, Delivery Order 16 -- editorial and graphics support, technical review, and preliminary draft preparation. Key SRA staff included:

Marcia Gardner .	Karl Held
Lori Hidinger	Alex Polymenopoulos
Mark Ernstmann	Jocelyn Smith

EXECUTIVE SUMMARY

This document provides regional project managers, on-site coordinators, and their contractors with sampling and analysis methods for evaluating whether ground water remediation has met pm-established cleanup standards for one or more chemical contaminants at a hazardous waste site. The verification of cleanup by evaluating a site relative to a cleanup standard or an applicable or relevant and appropriate requirement (ARAR) is mandated in Section 121 of the Superfund Amendments and Reauthorization Act (SARA). This document, the second in a series, provides sampling and data analysis methods for the purpose of verifying attainment of a cleanup standard in ground water. The fast volume addresses evaluating attainment in soils and solid media.

This document presents statistical methods which can be used to address the uncertainty of whether a site has met a cleanup standard. Sup&fund managers face the uncertainty of having to make a decision about the entire site based only on samples of the ground water at the site, often collected for only a limited time period.

The methods in this document approach cleanup standards as having three components that influence the overall stringency of the standard: first, the magnitude, level, or concentration deemed to be protective of public health and the environment; second, the sampling performed to evaluate whether a site is above or below the standard; and third, the method of comparing sample data to the standard to decide whether the remedial action was successful. All three of these components are important. Failure to address any one these components can result in insufficient levels of cleanup. Managers must look beyond the cleanup level and explore the sampling and analysis methods which will allow confident assessment of the site relative to the cleanup standard

A site manager is likely to confront two major questions in evaluating the attainment of the cleanup standard: (1) is the site really contaminated because a few samples are above the cleanup standard? and (2) is the site really "clean" because the sampling shows the majority of samples to be below the cleanup standard? The statistical methods demonstrated in this guidance document allow for decision making under uncertainty and permit valid extrapolation of information that can be defended and used with confidence to determine whether the site meets the cleanup standard.

The presentation of concepts and solutions to potential problems in assessing ground water attainment begins with **an introduction to the statistical reasoning required to** implement these methods.   Next, the planning activities, requiring input from both statisticians and nonstatisticians, are described. Finally, a series of methodological chapters are presented to address statistical procedures applicable to successive stages in the remediation effort. Each chapter will now be considered in detail.

Chapter 1 provides a brief introduction to the document, including its organization, intended use, and applications for a variety of treatment technologies. A model for the sequence of ground water remediation activities at the site is described. Many areas of expertise must be involved in any remedial action process. This document attempts to address only statistical procedures relevant to evaluating the attainment of cleanup goals.

The cleanup activities at the site will include site investigation, ground water remediation, a post-treatment period allowing the ground water to reach steady state, sampling and analysis to assess attainment, and possible post-cleanup monitoring. Different statistical procedures are applicable at different stages in the cleanup process. The statistical procedures used must account for the changes in the ground water system over time due to natural or man-induced causes. As a result, the discussion makes a distinction between short-term estimates which might be used during remediation and long-term estimates which are used to assess attainment. Also, a slack period of time after treatment and before assessing attainment is strongly recommended to allow any transient effects of treatment to dissipate.

Chapter 2 addresses statistical concepts as they might relate to the evaluation of attainment. The chapter discusses the form of the null and alternate hypothesis, types of errors, statistical power curves, the handling of outliers and values below detection limits, short- versus long-term tests, and assessing wells individually or as a group. Due to the cost of developing new wells, the assessment decision is assumed to be based on established wells. As a result, the statistical conclusions strictly apply only to the water in the sampling wells rather than the ground water in general. The expertise of a hydrogeologist can be useful for making conclusions about the ground water at the site based on the statistical results from the sampled wells.

The procedures in this document favor protection of the environment and human health. If uncertainty is large or the sampling inadequate, these methods conclude that the sample area does not attain the cleanup standard, Therefore, the null hypothesis, in statistical terminology, is that the site does not attain the cleanup standard until sufficient data are acquired to prove otherwise.

Procedures used to combine data from separate wells or contaminants to determine whether the site as a whole attains all relevant cleanup standards are discussed. How the data from separate wells are combined affects the interpretation of the results and the probability of concluding that the overall site attains the cleanup standard. Testing the samples from individual wells or groups of wells is also discussed.

Chapter 3 considers the steps involved in specifying the attainment objectives. Attainment objectives must be specified before the evaluation of whether a site has attained the cleanup standard can be made. Attainment objectives are not specified by statisticians but rather must be provided by a combination of risk assessors, engineers, project managers, and hydrogeologists. Specifying attainment objectives includes specifying the chemicals of concern, the cleanup standards, the wells to be sampled, the statistical criteria for defining attainment, the parameters to be tested, and the precision and confidence level desired.

Chapter 4 discusses the specification of the sampling and analysis plans. The sampling and analysis plans are prerequisites for the statistical methods presented in the following chapters. A discussion of common sampling plan designs and approaches to analysis are presented. The sample designs discussed include simple random sampling, systematic sampling, and sequential sampling.   The analysis plan is developed in conjunction with the sample design.

Chapter 5 provides methods which are appropriate for describing ground water conditions during a specified period of time. These methods are useful for making a quick evaluation of the ground water conditions, such as during remediation. Because the short-term confidence intervals reflect only variation within the sampling period and not long-term trends or shifts between periods, these methods are not appropriate for assessing attainment of the cleanup standards after the planned remediation has been completed. However, these descriptive procedures can be used to estimate means, percentiles,

confidence intervals, tolerance intervals and variability. Equations are also provided to determine the sample size required for each statistical test and to adjust for seasonal variation and serial correlation.

Chapter 6 addresses statistical procedures which are useful during remediation, particularly in deciding when to terminate treatment. Due to the complex dynamics of the ground water flow in response to pumping, other remediation activity, and natural forces, the decision to terminate treatment cannot easily be based on statistical procedures. Deciding when to terminate treatment should be based on a combination of statistical results, expert knowledge, and policy decisions. This chapter provides some basic statistical procedures which can be used to help guide the termination decision, including the use of regression methods for helping to decide when to stop treatment. In particular, procedures are given for estimating the trend in contamination levels and predicting contamination levels at future points in time. General methods for fitting simple linear models and assessing the adequacy of the model ate also discussed.

Chapter 7 discusses general statistical methods for evaluating whether the ground water system has reached steady state and therefore whether sampling to assess attainment can begin. As a result of the treatment used at the site, the ground water system will be disturbed from its natural level of steady state. To reliably **evaluate** whether the ground water can be expected to attain the cleanup standard after remediation, samples must be collected under conditions similar to those which will exist in the future. Thus, the sampling for assessing attainment can only occur when the residual effects of treatment on the ground water are small compared to those of natural forces.

Finding that the ground water has returned to a steady state after terminating remediation efforts is an essential step in establishing of a meaningful test of whether or not the cleanup standards have been attained. There are uncertainties in the process, and to some extent it is judgmental. However, if an adequate amount of data is carefully gathered prior to beginning remediation and after ceasing remediation, reasonable decisions can be made as to whether or not the ground water can be considered to have reached a state of stability. The decision on whether the ground water has reached steady state will be based on a combination of statistical calculations, plots of data, ground water modeling using predictive models, and expert advice from hydrogeologists familiar with the site.

Chapters 8 and 9 present the statistical procedures which can be used to evaluate whether the contaminant concentrations in the sampling wells attain the cleanup standards after the ground water has reached steady state. The suggested methods use either a fixed sample size test (Chapter 8) or a sequential statistical test (Chapter 9). The testing procedures can be applied to either samples from individual wells or wells tested as a group. Chapter 8 presents fixed sample size tests for assessing attainment of the mean: using yearly averages or after adjusting for seasonal variation; using a nonparametric test for proportions; and using a nonparametric confidence interval about the median. Chapter 9 discusses sequential statistical tests for assessing attainment of the mean using yearly averages, assessing attainment of the mean after adjusting for seasonal variation, and assessing attainment using a nonparametric test for proportions. In both fixed sample size tests and sequential tests, the ground water at the site is judged to attain the cleanup standards, if the contaminant levels are below the standard and are not increasing over time. If the ground water at the site attains the cleanup standards, follow-up monitoring is recommended to ensure that the steady state assumption holds.

Although the primary focus of the document is the procedures presented in Chapters 8 and 9 for evaluating attainment, careful consideration of when to terminate treatment and how long to wait for steady state are important in the overall planning. If the treatment is terminated prematurely, excessive time may be spent in evaluating attainment only to have to restart treatment to complete the remediation, followed by a second period of attainment sampling and decision. If the ground water is not at steady state, the possibility of incorrectly determining the attainment status of the site increases.

As an aid to the reader, a glossary of commonly-used terms is provided in Appendix G; calculations and examples are presented in boxes within the text; and worksheets with examples are provided in Appendix B.

1. INTRODUCTION

Congress revised the Superfund legislation in the Superfund Amendments **and Reauthorization**Act of 1986 (SARA). Among other provisions of SARA, section 121 on Cleanup Standards discusses criteria for selecting applicable or relevant and appropriate requirements (ARAR's) for cleanup and includes specific language that requires EPA mandated remedial action to attain the ARAR's.

Neither SARA nor EPA regulations or guidances specify how to determine whether the cleanup standards have been attained. This document offers procedures that can be used to determine whether a site has attained the appropriate cleanup standard after a remedial action.

1.1 General Scope and Features of the Guidance Document

1.1.1 Purpose

This document provides a foundation for decision-making regarding site cleanup by providing methods that statistically **compare**risk standards with field data in a scientifically defensible manner that allows for uncertainty. Statistical procedures can be used for many different purposes in the process of a Superfund site cleanup. The purpose of this document is to provide statistical procedures which can be used to determine if contaminant concentrations measured in selected ground-water wells attain (i.e., are less than) the cleanup standard. This evaluation requires specification of sampling protocols and statistical analysis methods. Figure 1.1 shows the steps involved in the evaluation process to determine whether the cleanup standard has been attained in a selected ground water well.

Figure 1.1   Steps in Evaluating Whether a Ground Water Well Has Attained the cleanup standard
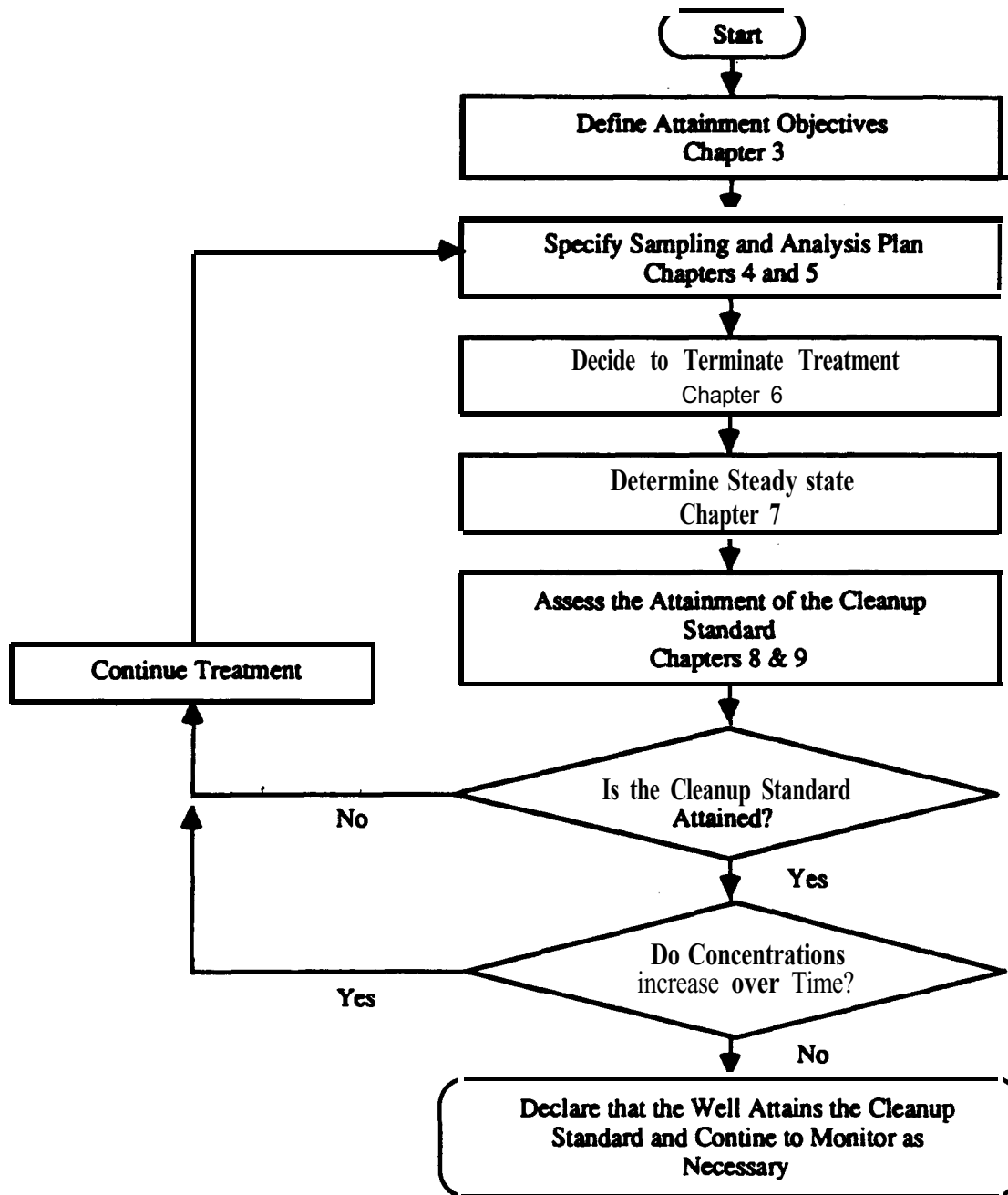
Consider the situation where several samples were taken and the results' indicated that one or two of the samples exceed the cleanup standard. How should this information be used to decide whether the standard has been attained? The mean of the samples might be compared with the standard. The magnitude of the measurements that are larger than the standard might be taken into consideration in making a decision. The location where large measurements occur might provide some insight.

When specifying how attainment is to be defined and deciding how statistical procedures can be used, the following factors are all important:

- The location of the sampling wells and the associated relationship between concentrations in neighboring wells;

- The number of samples to be taken;

- The sampling procedures for selecting and obtaining water samples;

- The data analysis procedures used to test for attainment.

Appendix D lists relevant EPA guidance documents on sampling and evaluating ground water. These documents address both the statistical and technical components of asampling and analysis program. This document is intended to extend the methodologies they provide by addressing statistical issues in the evaluation of the remediation process. This document does not attempt to suggest which standards apply or when they apply (i.e., the "How clean is clean?" issue). Other Superfund guidance documents perform that function.

1.1.2   Intended Audience and Use

This document is intended primarily for Agency personnel (primarily on-site coordinators and regional project managers), responsible parties, and their contractors who are involved with monitoring the progress of ground-water remediation at Superfund sites. Although selected introductory statistical concepts arc reviewed, this document is directed toward readers that have had some prior training or experience applying quantitative methods.

It must be emphasized that this document is intended to provide general direction and assistance to individuals involved in the evaluation of the attainment of cleanup standards. It is not a regulation nor is it formal guidance from the Superfund Office. This manual should not be viewed as a "cookbook" or a replacement for good engineering or statistical judgment

1.1.3 Bibliography, Glossary, Boxes, Worksheets, Examples, and References to "Consult a Statistician"

This document includes a bibliography which provides a point of departure for the more sophisticated or interested user. There are references to primary textbooks, pertinent journal articles, and related guidances.

The glossary (Appendix F) is included to provide short, practical definitions of terminology used in this guidance. Words and phrases appearing in **bold** within the text are listed in the glossary. The glossary does not use theoretical explanations or formulas and, therefore, may not be as precise as the text or alternative sources of information.

Boxes are used throughout the document to separate and highlight equations and example applications of the methods presented. For a quick reference, a listing of all boxes and their page numbers is provided in the index.

A series of worksheets is included (Appendices B and C) to help order and structure the calculations. References to the pertinent sections of the document are located at the top of each worksheet. Example data and calculations are presented in the boxes and the worksheets in Appendix B. The data and sites are hypothetical, but elements of the examples correspond closely to several existing sites.

Finally, the document often directs the reader to "consult a statistician" when more difficult and complicated situations are encountered. A directory of Agency statisticians is available from the Environmental Statistics and Information Division (PM-222) at EPA Headquarters (FTS 260-2680, 202-260-2680).

1.2   Use of this Guidance in Ground-Water Remediation Activities

Standards that apply to Superfund activities normally fall into the category of risk-based standards which arc developed using risk assessment methodologies. Chemical-specific ARARs adopt from other programs often include at least a generalized component of risk. However, risk standards may be specific to a site, developed using a local endangerment evaluation.

Risk-based standards are expressed as a concentration value and, as applied in the Superfund program axe not associated with a standard method of interpretation. Although statistical methods arc used to develop elements of risk-based standards, the estimated uncertainties are not carried through the analysis or used to qualify the standards for use in a field sampling program. Even though risk standards are not accompanied by measures of uncertainty, decisions based on field data collected for the purpose of representing the entire site and validating cleanup will be subject to uncertainty. This document allows decision-making regarding site cleanup by providing methods that statistically compare risk standards with field data in a scientifically defensible manner that allows for **uncertainty.**

Superfund activities where risk-based standards might apply are highly varied. The following discussion provides suggestions for the use of procedures &scribed in this document when implementing or evaluating Superfund activities.

1.2.1   Pump-and-Treat Technology

Ground water is often treated by pumping contaminated ground water out of the ground, treating the water, and discharging the water into local surface waters or municipal treatment plants. The contaminated ground water is gradually replaced by uncontaminated water from the surrounding aquifer or from surface recharge. Pump and treat systems may use a few or many wells. The progress of the remediation depends on where the wells arc placed and the schedule for pumping. Pumping is often planned to extend over many years.

Statistical methods presented in this manual can be used for monitoring the contaminants in both the effluent from the treatment system and the ground water in order to monitor the progress of the remediation.

Project managers must decide when to terminate treatment based on available data, advice from hydrogeologists, and the results of ground-water monitoring and modeling. This manual provides guidance on statistical procedures to help decide when to terminate treatment.

The remediation may temporarily alter ground water levels and flows, which in turn will affect the contaminant concentration levels. After termination of treatment and after the transient effects of the remediation have dissipated, the statistical procedures presented in this manual can be used to assess if the ground-water contaminant concentrations remain at levels which will attain and continue to attain the cleanup standard.

1.2.2 Barrier Methods to Protect Ground Water

If the contamination is relatively immobile and cannot effectively be removed from the ground water using extraction, it is sometimes handled by containment. In such cases, establishing barriers at the surface or around the contamination source may reduce contaminant input to the aquifer, resulting in the reduction of ground-water concentrations to a level which attains the cleanup standard. The barriers include soil caps to prevent surface infiltration, and slurry walls and **other** structures to force ground water to flow away from contamination sources.

The procedures in this manual can be used to establish whether the contamination levels attain the relevant standards after the ground water has established its new levels as a result of changes in ground-water flows.

1.2.3 Biological Treatment

In many situations natural bacteria will adapt to the contamination in the soil and ground water and consume the contaminants, releasing metabolic products. These bacteria will be most effective in consuming the contaminant if the underground environ-

ment can be controlled, including controlling the dissolved oxygen and nutrient levels. Biological treatment of ground water usually involves pumping ground water from down-gradient @cations and injecting enriched ground water at upgradient locations. The changes in the water table levels produce an underground flow carrying the nutrients to and throughout the contaminated soil and aquifer. Progress of the treatment can be monitored by sampling the water being pumped from the ground and measuring contaminant and nutrient concentrations. Biological treatment can also be accomplished above ground using a bioreactor as a component of a pump-and-treat system

Monitoring wells are placed in various patterns throughout, and possibly beyond, the area of contamination. These wells can be used to sample ground water both during treatment to monitor progress and after treatment to assess remediation success using the statistical methods discussed in this document.

1.3   Organization of this Document

The topics covered in each chapter of this document are outlined below.

Chapter 2. Introduction to Statistical Concepts and Decisions: introduces terminology and concepts useful for understanding statistical tests presented in later chapters.

Chapter 3. Specification of Attainment Objectives: discusses specification of the attainment objectives in a way which allows selection of the statistical procedures to be used.

Chapter 4. Design of the Sampling and Analysis Plan: discusses common sampling plan designs and approaches to the analysis.

Chapter 5. Descriptive Statistics: provides basic statistical procedures which are useful in all stages of the remedial effort. The procedures form a basis for the statistical procedures used for assessing attainment.

Chapter 6. Deciding to Terminate Treatment Using Regression Analysis: discusses statistical procedures which can aid the decision-makers who must decide when to terminate treatment.

Chapter 7. Approaching a Steady State After Terminating Remediation: discusses statistical and nonstatistical criteria for determining whether the ground water system is at steady state and/or if additional remediation might be required.

Chapter 8. Assessing Attainment Using Fixed Sample Size Tests: discusses statistical procedures based on fixed sample sizes for deciding whether the concentrations in the ground water attain the relevant cleanup standards

Chapter 9. Assessing Attainment Using Sequential Tests: discusses sequential statistical procedures for deciding whether the concentrations in ground water attain the relevant cleanup standards.

Worksheets: Provided for both practical use at Superfund sites and as examples of the procedures which in being recommended.

1.4   Summary

This document provides a foundation for decision-making regarding site cleanup by providing methods that statistically compare risk standa& with field data in a scientifically defensible manner that allows for uncertainty. In particular, the document provides statistical procedures for assessing whether the Superfund Cleanup Standards for ground water have been attained. The document is written primarily for agency personnel, responsible parties and contractors. Many areas of expertise must be involved in any remedial action process. This document attempts to address only the statistical input required for the attainment decision.

The statistical procedures presented in this document provide methods for comparing risk based standards with field data in a manner that allows for assessing uncertainty. The procedures allow flexibility to accommodate site-specific environmental factors.

To aid the reader, statistical calculations and examples arc provided in boxes separated from the text, and appendices contain a glossary of commonly-used terms; statistical tables and detailed statistical information; worksheets for implementing procedures and calculations explained in the text.

2. INTRODUCTION TO STATISTICAL CONCEPTS AND DECISIONS

This document provides **statistical** procedures to help answer an important question that will arise at Superfund sites undergoing ground water remediation:

**"Do the contaminants in the ground water in designated
wells at the site attain the cleanup standards?"**

The cleanup standard is attained if, as a result of the remedial effort, the previously unacceptably high contaminant concentrations are reduced to a level which is acceptable and can be expected to remain acceptable when judged relative to the cleanup standard.

In order to answer the question above, the following more specific questions must be answered:

- What contaminant(s) must attain the designated cleanup standards?

- How is attainment of the cleanup standards to be defined?

- What is the designated cleanup standard for the contaminant(s) being assessed? and

- Where and when should samples of the ground water be collected?

This chapter discusses each of these topics briefly, followed by an introduction to statistical procedures for assessing the attainment of cleanup standards in ground water at Superfund sites. Also discussed are terminology and statistical concepts which are useful for understanding the statistical tests presented in later chapters. Basic statistical principles and topics which have particular applicability to ground water at Superfund sites are also considered.

Later chapters discuss in detail the specification of attainment objectives and the implementation of statistical procedures required to determine if those objectives have been met at the Superfund site.

2.1   A Note on Terminology

This guidance document assumes that the reader is familiar with statistical procedures and terminology, particularly the concepts of random sampling and hypothesis testing, and the calculation of descriptive statistics such as means, standard deviations, and proportions. An introduction to these statistical procedures can be found in statistical textbooks such as Sokal and Rohlf (1981), and Neter, Wasserman, and Whitmore (1982). The glossary provides a description of the terms and procedures used in this document.

In this document we will use the word clean as a short hand for "attains the cleanup standard" and contaminated for "does not attain the cleanup standard."

The term sample can be used in two different ways. One refers to a physical water sample collected for laboratory analysis while the other refers to a collection of data called a statistical sample. To avoid confusion, the physical water sample will be called a physical sample or water sample. Otherwise, the word sample will refer to a statistical sample i.e. a collection of randomly selected physical samples obtained for assessing attainment of the cleanup standard.

2.2   Background for the Attainment Decision

In general, over time, a Superfund site will go through the following phases:

- ·   Contamination;
- ·   Realization that a problem exists;
- ·   Investigation to determine the extent of the problem;
- ·   Selection of a remediation plan to alleviate the problem;
- ·   Cleanup (which may occur in several steps);
- ·   Termination of cleanup;
- ·   Final determination that the cleanup has achieved the required goals; and
- ·   Termination of the remediation effort.

CHAPTER 2: INTRODUCTION TO STATISTICAL CONCEPTS AND DECISIONS:

This document focuses on the post-cleanup phase and particularly on the sampling and statistical procedures for determining if the site has attained the required cleanup standards.

2.2.1   A Generic Model of Ground-Water Cleanup Progress

During the planning and execution of remedial action and the sampling and analysis for assessing attainment, numerous activities must take place as indicated in the following scenario and illustrated in Figure 2.1. This figure will be used throughout the document to indicate to the reader at which step in the remedial process the procedures being discussed in a chapter ate applicable. A discussion of each step follows Figure 2.1.

Figure 2.1   Example scenario for contaminant measurements in one well during successful remediation action
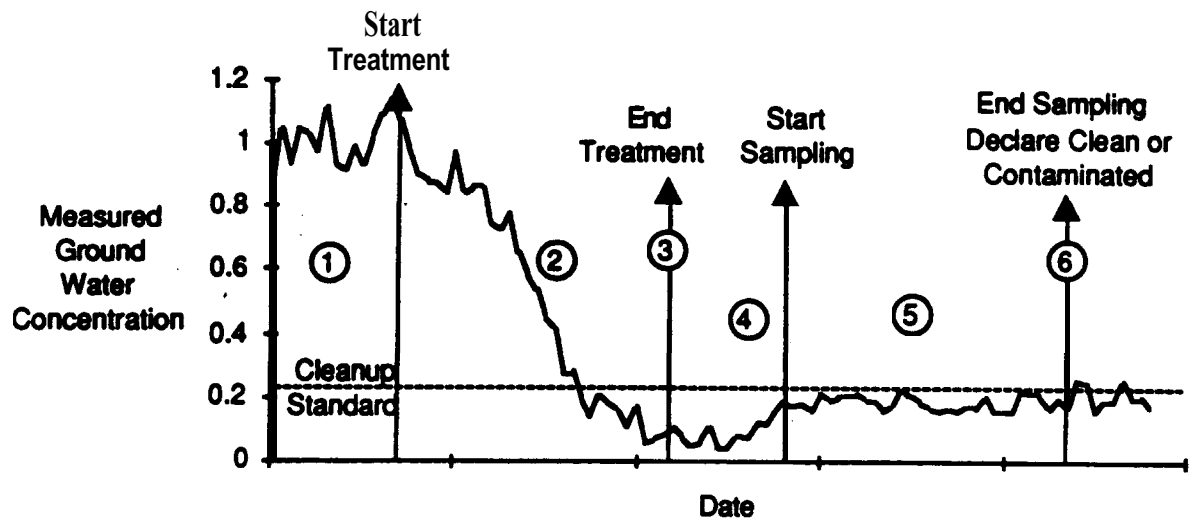
| (1) | Evaluate the site; determine the remedial action to be used | Although **evaluation** of the site and selection of the cleanup technology may require the use of several statistical procedures, this document does not address this aspect of the remedial effort |

(2) **Perform remedial** cleanup

During a successful remedial cleanup, the concentrations of contaminants can be expected to have a decreasing trend. Due to seasonal changes, natural fluctuations, changes in pumping schedules, lab measurement error, etc., the measured concentrations will fluctuate around the trend. Some statistical procedures that could be used to analyze data during treatment are discussed in Chapter 5.

(3) Decide when to terminate remedial **treatment**

Based on both expert knowledge of the ground-water system and data collected during treatment, it must be decided when to terminate treatment and prepare for the sampling and analysis far assessing attainment. Statistical procedures relevant to the termination decision are discussed in Chapter 6. Analysis of data collected during treatment may indicate that the cleanup standards will not be achieved by the chosen cleanup methods, in which case the cleanup technology and goals must be reassessed.

(4) Assess when the ground water concentrations reach steady state

The ground-water system will be disturbed from its natural level and flow by the treatment process, including perhaps pumping or reinjection of ground water. After treatment is terminated, the transient effects will dissipate and the ground-water levels and flows will gradually reach their natural levels. In this process, the contaminant concentrations may change in unpredictable ways. Before the assessment is initiated, the ground water must be able to return to its natural level and flow pattern, called steady state, so that the data collected are relevant to assess conditions in the future. Sampling and analysis during the return to natural conditions are discussed in Chapter 7. The ground water at a particular site will be considered to have achieved steady state if the assumption of steady state is consistent with both statistical tests and the advice of a hydrogeologist familiar with the site. The attainment sampling can begin once it is determined that the site is at steady state.

(5) Sample to assess attainment

After the water levels and flows have reached steady state, sampling to assess attainment of the cleanup standards can begin. Statistical procedures for assessing attainment are presented in Chapters 8 and 9. The statistical tests used may be either fixed sample size tests or sequential tests. At many sites sequential tests will probably be preferred. During the assessment phase, measured concentrations are expected to either fluctuate around a constant or gradually decreasing concentration. If the measurements consistently increase, then either the ground-water system is not at steady state or there is reason to believe that the sources of contamination have not been adequately cleaned up. In this situation, a reassessment of the data is required to determine if more time must pass until the site is at steady state or if additional remedial activity is required.

(6)   Based on statistical tests, determine if the cleanup standard has been attained or not.

If the cleanup standard has been attained, implementation, of periodic sampling to monitor for unanticipated problems", is recommended. The attainment decision is based on several assumptions. From a statistical perspective, the purpose of periodic monitoring after attainment is to check the validity of the assumptions. If the attainment objectives have not been met, the cleanup technology and goals must be reassessed.

Different statistical procedures are needed at different steps in this process. The statistical procedures which are helpful in dedetermining whether to terminate treatment arc different from those used in the attainment decision. In all aspects of the site investigation and remediation, statistical procedures may be required that are not addressed in this document. In this case, consultation with a statistician familiar with ground-water data is **recommended.**

This document takes the approach that:

·   A decision that the ground water in the wells attains the cleanup standard requires the assumption that the ground water can be expected to continue to attain the cleanup standards beyond the termination of sampling, and

·   Data collected while the ground-water system is disturbed by treatment cannot reliably predict concentrations after steady state has been achieved. Therefore, it is recommended that the ground-water system return to steady state before the sampling for assessing attainment commences. The data gathered prior to reaching steady state can be used for guidance in selecting the statistical procedure to employ for assessing attainment

2.2.2   The Contaminants to be Tested

In general, multiple contaminants will be identified at the site prior to remedial action. The mixture of contaminants which are present at any one time or place will depend on many factors.

The discussion in this document assumes that relevant regulatory agencies have specified the contaminants which arc to be used to assess attainment. Conclusions based on the statistical procedures introduced in this document apply only to the compounds actually sampled and the corresponding data analyzed in the statistical tests.

2.2.3   The Ground-Water System to be Tested

Contamination in ground water is measured from water samples collected from wells at specified locations and times. The location of the wells, the times and frequency of the sampling, and the assumptions behind the analyses will affect the interpretation of the statistical results.

This document assumes that the attainment decision will be based on samples from established wells, This document does not make recommendations on where to locate wells for sampling. However, decisions must be made on which wells arc to be used for the assessing attainment. Because wells arc not randomly located throughout an aquifer, the statistical conclusions strictly apply only to the water obtained from the selected wells and not to the aquifer in general. Conclusions about the aquifer must be based on a combination of statistical results for the sampled wells and expert knowledge or beliefs about the ground-water system and not on statistical inference.

Because of the high cost of installing a new well and the possibility of using information from previous investigation stages, this document assumes that the location of wells has been specified by experts in ground-water hydrology and approved by regulatory agencies who arc familiar with the contamination data at the site.

Interpretation of the results of the statistical analysis will depend on a judgment as to whether the wells are in the correct place. If it is necessary to test the assumptions used to select wells, additional wells will have to be established and sampled. In this case, consultation with a statistician is recommended.

2.2.4   The Cleanup Standard

The cleanup standard is the criterion set by EPA against which the measured concentrations are compared to determine if the ground water at the Superfund site is acceptable or nor. If the ground water meets the cleanup standard, then the remediation efforts are judged to be complete. The specification of the cleanup standard by EPA or another regulatory agency may be different for different sites and for different chemicals or mixtures of chemicals. With a mixture of contaminants, the cleanup standard may apply to

an aggregate measure, or, in complex mixtures, the ground water may be required to meet the cleanup standard for every contaminant present. For more information, see *Guidance on Remedial Actions for Contaminated Ground Water at Superfund* Sites (EPA, 1988).

2.2.5   The Definition of Attainment

In order to determine if the contaminant concentrations at the site attain the cleanup standard one must carefully define what concentration is to be compared to the cleanup standard and what criteria are to be used to make the comparison for assessing attainment. This document assumes that either the average concentration or a selected percentile of the concentrations is to be compared to the cleanup standard. The examples in the text usually use the average concentration. The ground water in a well attains the cleanup standard if, based on statistical tests, it is unlikely that the average concentration (or the percentile) is greater than the cleanup Standard.

The statistical procedures for assesing the attainment of the cleanup standard use a basic statistical technique called hypothesis testing. To show that the ground water in the selected wells is actually below the cleanup standard (i.e., attains the cleanup standard), we assume that the water in the wells does not attain the cleanup standard. This assumption is called the null hypothesis. Then data arc collected. If the data arc sufficiently inconsistent with the null hypothesis, the null hypothesis is rejected and we conclude that the water in the well attains the cleanup standard.

The steps involved in hypothesis testing are:

(1)   Establish the null hypothesis, "The contaminant concentrations in the select+ wells do not attain the applicable cleanup standard";

(2)   *collect data; and*

(3)   Based on the data, decide if the ground water attains the cleanup standard:

(a)   If the data are inconsistent with the null hypothesis, conclude that there is sufficient evidence to reject the null hypothesis. Accept the alternate hypothesis that the contaminant concerntrations attain the applicable cleanup standard, i.e., conclude that the ground water is clean.

(b)   Otherwise, conclude that there is insufficient evidence to reject the null hypothesis and that the contaminant concentra-

tions do not attain the cleanup standards, i.e., conclude that the ground water is contaminated.

- To be technically correct, the results of the hypothesis **test** indicate whether the null hypothesis can be rejected with **a** specified level of confidence. In practice, we would conclude that the concentrations do or do not attain the cleanup standards and act as if that conclusion were known as fact rather than subject to error. Therefore to avoid the verbose but technically correct wording above, the results of the hypothesis tests will be worded as concluding that the concentrations either attain **or** do not attain the cleanup standard.

When specifying simplified Superfund site cleanup objectives in consent decrees, records of decision, or work plans, it is extremely important to say that the site shall be cleaned up until the sampling program indicates with reasonable confidence that the concentrations of the contaminants at the entire site are <u>less than the cleanup standard</u>. However, attainment is often wrongly described by saying that concentrations at the site <u>shall not exceed the cleanup stan</u>dard.

2.3 Introduction to Statistical Issues For Assessing Attainment

This section provides a discussion of some basic statistical issues with an emphasis on those with specific application to assessing attainment in ground water. This discussion provides a general background for the specification of attainment objectives in Chapter 3 and the statistical procedures presented in Chapters 4 through 9.

2.3.1 Specification of the Parameter to be Compared to the Cleanup Standard

In order to define a statistical test to determine whether the ground water attains the cleanup standard, the characteristics of the chemical concentrations to be compared to the cleanup standard must be specified. Such characteristics are called parameters. The choice of the parameter to use when assessing attainment at Superfund sites may depend on site specific characteristics and decisions and has not, in general, been specified by EPA.

The parameters discussed in this document are the mean or average concentration and a selected percentile of the concentrations. For example, the rule for deciding if the ground water attains the cleanup standard might be: the ground water is considered clean (orremediated) if the mean concentration is below the cleanup standard based on a statistical test. The following sections define parameters for distributions of data and the statistical properties of these parameters. An understanding of these properties is necessary for determining the appropriate parameter to test

The Distribution of Data Values

This section discusses the characteristics of concentration distributions which might be expected at Superfund sites and how the distribution of concentrations in the ground water can be described using parameters. These topics are discussed in more detail-in Volume I (Sections 2.8 and 3.5).

Consider the set of concentration measurements which would be obtained if all possible ground-water samples from a particular monitoring well over a specified period of time could be collected and analyzed. This set of measurements is called **the population** of ground-water sample measurements. The set of ground-water samples comprising the population may cover a fixed period of time, such as one year, or an unlimited time, such as all future measurements. The set of ground-water measurements can be described mathematically and graphically by the "population distribution function" referred to as the "distribution of the data". Figure 2.2 shows a plot of the population distribution for data from three hypothetical distributions. The vertical axis shows the relative proportion of the population measurements at each concentration value on the horizontal axis. In the plots, the areas under the curve between any two points on the concentration axis represents the percentage of the ground-water measurements that have concentration values within the specified range.

Two distributions, the normal and lognormal distributions, will be used as examples in the following discussion. Both the normal and lognormal distributions are useful in statistical work and can be used to approximate the concentration distributions from wells at Superfund sites. Figure 2.2 shows an example of a normal and a lognormal distribution.

Figure 2.2   **Measures of location:** Mean, **median, 25th percentile, 75th percentile, and 95th percentile for three hypothetical distributions**
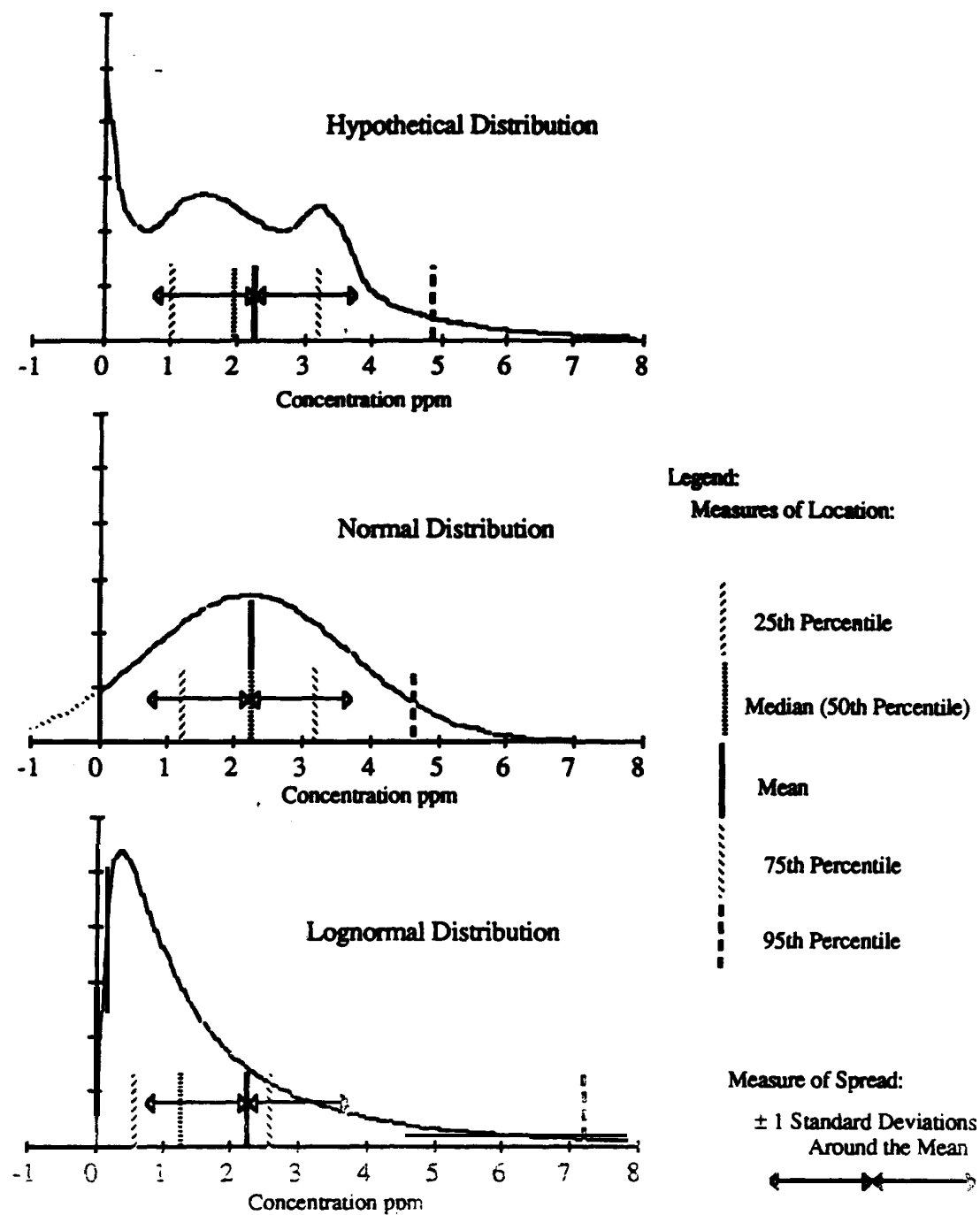
Summary **measures** describing characteristics of **the** population **distribution** are referred to as parameters or population parameters. Three important characteristics of the data described by these parameters:

- **The location of the data;**

- **The spread (or dispersion) of the data; and**

- **The general shape or "skewness" of the data distribution.**

Measures 'of Location

Measures of location (or central tendency) are often used to describe where most of the data lie along the concentration axis of the distribution plot. Examples of such measures of location are:

- "The mean (or average) concentration of all ground-water samples is 17.2 ppm" (i.e., 17.2 is the mean concentration);

- "Half the ground-water samples have concentrations greater than 13 ppm and half less than 13 ppm" (13 is the median concentration); or

- "Concentrations of 5 ppm (rounded to the nearest unit) occur more often than any other concentration value" (the mode is 5 ppm).

Another measure of location is the percentile. The Qth percentile is the concentration which separates the lower Q percent of the ground-water measurements from the upper 100-Q percent of the ground-water measurements. The median is a special percentile, the 50th percentile. The 25th percentile is the concentration which is greater than the lowest 25 percent of the ground-water measurements and less than the remaining 75 percent of the ground-water measurements. Figure 2.2 shows the mean, median, 25th percentile, 75th percentile, and 95th percentile for three distributions introduced previously.

Throughout this document, the Greek letter, μ, (spelled "mu" and pronounced "mew") will be used to denote the population mean. The median will be denoted by X_{50}, and the Qth percentile will be denoted by X_Q.

Measures of Spread

Measures of spread provide information about the variability or dispersion of a set of measurements. Examples of different measures of spread are:

- **The standard deviation or the variance** (the square of the standard deviation). The population standard deviation is denoted by the Greek letter, σ, (pronounced "sigma") throughout this document. If data are normally distributed, two-thirds of the data are within one standard deviation of the mean ;

- **The coefficient of variation** is the ratio of the standard deviation to the mean, $\frac{\sigma}{\mu}$; and

- **The interquartile range** is the difference between the 75th and 25th percentiles of the distribution.

For each distribution in Figure 2.2, the mean and the range of plus and minus one standard deviation around the mean are shown on the plots.

Measure of Skewness

Skewness is a measure of the extent to which a distribution is symmetric or asymmetric. A distribution is symmetric if the shape of the two halves are mirror images of each other about a center line. One common symmetric distribution is the normal distribution, which is often described as having a "bell-shape." Many statistical tests assume that the sample measurements are normally distributed (i.e., have a normal distribution).

The distribution of concentrations is not likely to be symmetric. It may be skewed to the right. That is, the highest measurements (those to the right on the plot of the distribution function) are farther from the mean concentration than are the lowest concentrations. Ground-water measurements often have a skewed distribution which can be approximated by a lognormal distribution (see Gilbert 1987, for additional discussion of the normal and lognormal distributions). Note that for right skewed distributions (e.g., the lognormal distribution in Figure 2.2) the mean is greater than the median.

The three distributions shown in Figure 2.2 have the same mean and standard deviation. Note, however, that the occurrence of particularly high or low concentra-

tions differs for the three distributions. In general, the more skewed the distribution, the more likely are these extreme observations.

Selecting the Parameter to Compare to the Cleanup Standard

In order to determine if the contaminant concentrations attain the cleanup standard the measure of location which is to be compared to the cleanup standard must be specified. Even though the true distribution is unknown, the specified measure of location, or parameter of interest, can be selected based on:

· Information about the distribution from preliminary data;

· Information about the behavior of each parameter for different distributions;

· The effects of various concentrations of the contaminant on human health and the environment; and

· Relevant criteria far protecting human health and the environment.

Chapter 3 discusses in more detail the selection of the mean or a percentile to be compared to the cleanup standard.

2.3.2   Short-term Versus Long-term Tests

Due to fluctuating concentrations over time, the average contaminant concentration over a short period of time may be very different from the average over a long period of time. Figure 2.3 shows a hypothetical series of weekly ground-water concentration measurements collected over a period of 70 weeks (about 16 months). The figure shows the weekly concentration measurements, the average concentration for **weeks** 21 through 46 (6 months), and the long-term average concentration which is obtained from data collected over 50 years (only a portion of which is shown here). From the figure, it can be seen that the short-term average concentration can be very different from the long-term average.

Figure 2.3   Illustration of the difference between a short- and long-term mean concentration
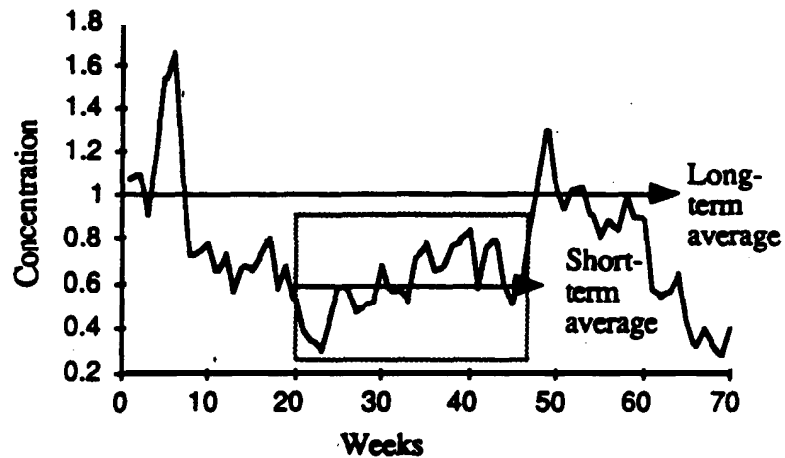
The short-term average is estimated using data collected during the period of interest, in this example during weeks 21 through 46. Similarly the longer term average can be estimated based on data collected over the longer period of interest, perhaps 50 years. Fortunately, by using information on the correlation of the measurements across time, it is usually possible to estimate the long-term average concentration from data collected over a limited period of time. In order to estimate the average concentration for a period which is longer than the data collection period, assumptions must be made which relate the unmeasured future concentrations to the concentrations which are actually measured. These assumptions are stated in terms of a model for the data.

Statistical decisions and estimates that only apply to the sampling period arc referred to here as "short-term" estimates and are presented in Chapter 4. Decisions and estimates that apply to the foreseeable future are called "long-term" estimates. The long-term estimates are made based on the assumption that the ground-water concentrations will behave in a predictable manner. The assumptions take into account the expected natural fluctuations in ground-water flows and contaminant concentrations.

In this document the ground water is said to attain the cleanup standard only if the concentrations attain the cleanup standard for the foreseeable (or at least predictable) future. Thus, long-term estimates and procedures are used to assess attainment. Short-term estimates can be used to make interim managemcnt decisions.

2.3.3   The Role of Statistical Sampling and Inference in Assessing Attainment

When assessing attainment, it is desirable to compare the population mean (or population percentile or other parameter) of the concentrations to the cleanup standard. However, the data for assessing attainment arc derived from a sample, a small proportion of the population. Statistical inference is used to make conclusions about the population parameter from the sample measurements.   For **illustration,** the following discussion assumes that the population mean must be less than the cleanup standard if we arc to conclude that the water in the well attains the cleanup standard.

The mean concentration calculated from the sample data provides an estimate of the population mean. Estimates of concentration levels computed from a statistical sample are subject to "error" in part because they arc based on only a small subset of the population. The use of the term "error" in this context in no way implies that then are mistakes in the data. Rather, "error" is a short hand way of saying that there is variability in the sample estimates from different samples. There are two components to this error sampling error and lab, or measurement, error.

· 　Different samples will yield different estimates of the parameter of interest due to **sampling error.**

• 　Unknown factors in the handling and lab analysis procedures result in errors or variation in the lab measurements, i.e., two lab analyses of the same ground-water sample will usually give slightly different concentration values. This difference is attributed to **lab error** or **measurement error.**

Because the sample mean is subject to error, it cannot be directly compared to the cleanup standard to decide if the population mean is less than the cleanup standard. For example, just because the mean for a particular sample happens to be below the cleanup standard does not mean that the standard has been attained. To make meaningful inferences, it is necessary to obtain a measure of the error (or expressed another way, the precision) associated with the sample mean[1]. An estimate of the error in the sample mean can be calculated from the sample and is referred to as the standard error of the mean. It is a

---

[1] The possible bias in the measurements is assumed to be zero. The quality assurance plan should address the problems of possible bias.

basic measure of the absolute variability of the calculated sample mean from one sample to another.

The standard error of the mean can be used to construct confidence intervals around a sample mean **using** equation (2.1) in Box 2.1. Under general conditions, the interval constructed using equation (2.1) will include the population mean in approximately 95 percent of all samples collected and is called a "95 percent two-sided confidence interval." This useful fact follows from **the Central Limit Theorem** which states that, under fairly general conditions, the distribution of the sample mean is "close" to a normal distribution even though we may not know the distribution of the original data. Note also that the validity of the confidence interval given in Box 2.1 depends on the data being independent in a statistical sense. Independent ground water measurements are obtained when the sample collection times are randomly selected within the sampling period.

When assessing attainment, a two-sided test would be used for pH because both high and low values represent pollution. For most other pollutants, use one-sided confidence intervals because only high values indicate pollution. A 95 percent one-sided confidence interval can be obtained from equation (22) in Box 2.1. The interval from zero (the lowest possible measurement) to this upper endpoint will also include the population mean in approximately 95 percent of all samples collected.

Box 2.1
Construction of Confidence Intervals Under Assumptions of Normality

To construct a 95 percent two-sided confidence interval around a sample mean:

lower endpoint = sample mean - 1.96 * standard error and

upper endpoint = sample mean + 1.96 * standard error.   (2.1)

To construct a 95 percent one-sided confidence interval:

upper endpoint = sample mean + 1.65 * standard error.   (2.2)

Using confidence intervals, the following procedure can be used to make conclusions about the population mean based on a sample of data:

(1)   Calculate the sample mean;

(2)   calculate the standard error of the sample mean;

(3)   Calculate the upper endpoint of the one-sided confidence interval;

(4)   If the upper endpoint of the confidence interval is below the cleanup standard, then conclude that the ground water attains the cleanup standard, otherwise conclude that the ground water does not attain the cleanup standard.

A 95 percent confidence interval will not cover the population parameter in 5 percent of the samples. When using the confidence interval to assess attainment, one will incorrectly concluded that the ground water attains the cleanup standard in up to 5 percent of all samples. Thus, this procedure is said to have a false positive rate of 5 percent. This false positive rate is discussed in detail in the next section.

2.3.4   Specification of Precision and Confidence Levels for Protection Against Adverse Health and Environmental Risks

The validity of the decision that a site meets the cleanup standard depends on how well the samples represent the ground water during the period of sampling, how accurately the samples are analyzed, and the criteria used to define attainment. The true but unknown condition is that the ground water is either clean or contaminated. Similarly, the decisions made using the statistical procedures will result in an attainment or non-attainment decision. The relationship between these two conditions is shown in Table 2.1.

Table 2.1   False positive and negative decisions

Decision based on a statistical sample	True condition in the well:	
	Clean (Attains the cleanup standard)	**Contaminated (Does** not attain the cleanup standard)
Clean	Correct decision	False positive decision
Contaminated	False negative decision	correct decision

As a result of the sampling and measurement uncertainty, one may decide that the site is clean when it is not. In the context of this document, this mistaken conclusion is referred to as a **false positive** finding (statisticians refer to a false positive as a 'Type I error"). There are several points to make regarding false positives:

· Reducing the chance of a false positive decision helps to protect human health and the environment;

· A low false positive rate does not come without cost. The additional cost of lowering false positive rates comes from taking additional samples and using more precise analysis methods;

· The definition of a false positive in this document is exactly the opposite of the more familiar definition of a false positive under RCRA detection and compliance monitoring.

In order to design a statistical test for assessing attainment, those specifying the sampling and analysis objectives must select the maximum acceptable false positive rate (the maximum probability of a false positive decision is denoted by the Greek letter alpha, **a)**. It is usually set at, levels such as 0.10, 0.05, or 0.01 (that is 10%. 5%, or 1%), depending on the potential consequences of declaring that the ground water is clean when in fact it is not. While different false positive rates can be used for each chemical, it is recommended that the same rate be used for all chemicals being investigated. For a further discussion of false positive rates, see Sokal and Rohlf (198 1).

The converse of a false positive decision is a **false negative**decision (or Type II error), the mistake of concluding the ground water requires additional treatment when, in fact, it attains the cleanup standard This error results in the waste of resources in unnecessary treatment. It would be desirable to minimize the probability of false negative **decisions as well as false positive decisions. The Greek letter beta (β) is used to represent** the probability of a false negative decision.

If both α and β can be reduced, the percentage of time that the correct deci-sion will be made will be increased. Unfortunately, simultaneous reduction usually can only be achieved by increasing sample size (the number of samples collected and analyzed), which may be expensive.

The probability of declaring the ground water to be clean will depend on the true mean concentration of the ground water. If the population mean is above the cleanup standard, the ground water will rarely be declared clean (this will only happen if the particular sample chosen has a large associated sampling and/or measurement error). If the population mean is much smaller than the cleanup standard, the ground water will almost always be judged to be clean. This relationship can be plotted for various values of the population mean as in Figure 2.4. The plot shows the probability of declaring the ground water to be clean as a function of a hypothetical population mean, and is referred to as a **power curve.** For practical purposes, in this volume the probability of declaring the site clean is the "power of the test." The following assumptions were made when plotting the example power curve in Figure 2.4: the false positive rate is 5%, the false negative rate **when the true mean, μ_1, is 0.6 is 20%, and the cleanup standard is 1.0.**

If the population mean concentration is equal to or just above the cleanup standard (i.e., does not attain the cleanup standard), the probability of declaring the ground **water to be clean is α; this is the maximum false positive rate.**

For the specification of the attainment objectives (discussed in Chapter 3), the acceptable probabilities of a false positive and false negative decision must be specified. Based on these values and the selected statistical procedures, the required sample size can be calculated.

Figure 2.4   Hypothetical power curve
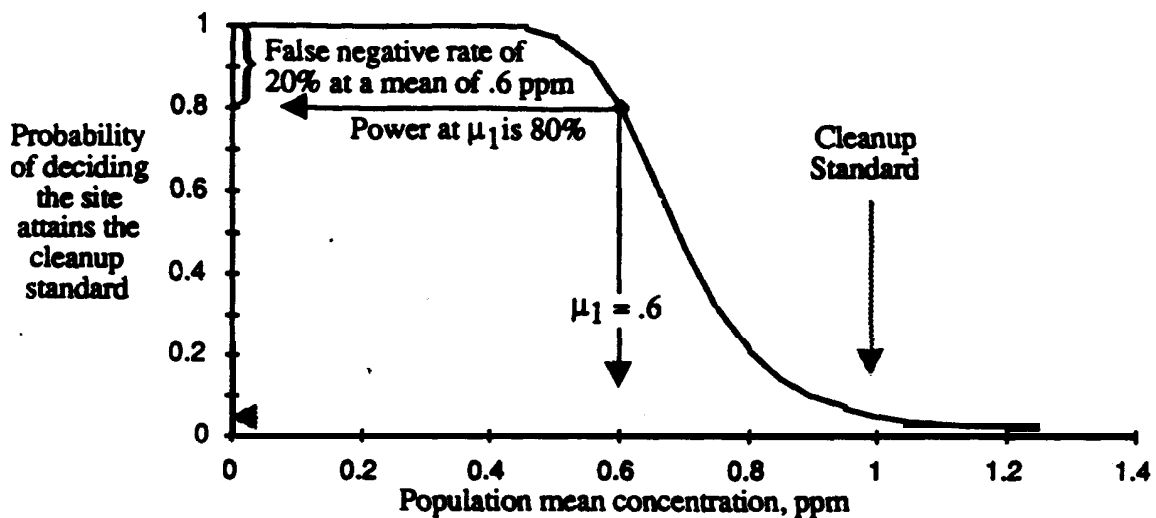
2.3.5   Attainment Decisions Based on Multiple Wells

The ground water will be judged to attain the cleanup standard if the contaminant concentrations in the selected wells are sufficiently low compared to the cleanup standard. Below are two possible ways in which the attainment decision can be based on water samples from multiple wells:

- Assess each well **individually:** make a separate attainment decision for each well; conclude that the ground water at the site attains the cleanup standard if the ground water in each tested well attains the cleanup standard.

- **Associate** selected wells into groups: collect samples in all wells in a group at the same time, combine the results from all wells in the same group into one summary statistic for that time period; conclude that the ground water represented by each group attains the cleanup standard if the summary statistic attains the cleanup standard. Conclude that the ground water at the site attains the cleanup standard if the summary statistics from all groups attain the standard.

The choice of assessing wells individually or as a group has implications for the interpretation of the statistical results and the false positive and false negative probabilities for deciding that the site, as opposed to the well, attains the cleanup standard. These issues are discussed in more detail in the following three sections.

Assessing Multiple Wells Individually

When assessing each well individually, slightly different criteria can be used for each attainment decision. For example, different sample collection schedules can be used for each well. Assessing each well individually may require substantially fewer samples than assessing the wells as a group, depending on the concentrations in the wells.

The attainment decisions for each individual well must be combined to make an attainment decision for the entire site. The only procedure discussed in this document for combining the results from assessments on individual wells is to conclude that the ground water at the site attains the cleanup standard only if the ground water in each well attains the cleanup standard

If many wells are tested the site will not attain the cleanup standard if any one of the wells does not attain the standard. Even if all wells actually attain the cleanup standard, the more wells used to assess attainment, the greater the likelihood of a false negative decision in one well, resulting in an overall non-attainment decision. On the other hand, assessing all wells individually can result in significant protection for human health and the environment because all concentrations must attain the cleanup standard in spite of false negative decisions. Implicit in the above discussion is the conflict of protecting the public health versus the cost of possible overcleaning are over attainment.

Testing Multiple Wells as a Group

When multiple wells are tested as a group, samples must be collected in each well at the same. time and thus the same number of samples will be collected in all wells within a group. At each sample time, the measurements from each well are combined into a summary statistic. The ground water in the group of wells would be declared to attain the cleanup standard if the summary statistic was significantly less than the cleanup standard. Several methods can be used to combine the measurements from all tested wells at each sample time into one summary statistic. Two methods arc:

- Average of measurements from all wells within a group; and
- Take the maximum concentration across all wells within a group.

If the average across all wells must be less than the cleanup standard, then the site may be declared clean if the concentrations in some wells are substantially greater than the cleanup standard as long as concentrations in other wells arc much less than the cleanup stand&d. These differences among wells in a groups can sometimes be minimized by grouping wells with similar concentration levels. On the other hand, requiring that the maximum concentration across all wells attain the cleanup standard assures that each well individually will attain the standard.

If the average concentration across all wells is to be compared to the cleanup standard, a decrease in lab costs may be achieved by compositing the water samples across wells (and possibly across time) and analyzing the contaminant concentrations in the composite samples. Since the recommended number of samples to be composited and the length of the sample period will depend on the serial correlation of the data and several cost and variance estimates, consultation with a statistician is recommended if compositing is considered.

Multiple Statistical Tests

When assessing attainment in multiple wells (or groups of wells) and when assessing attainment far multiple chemicals, two probabilities are of interest: the probability of deciding that one compound in one well (or group of wells) is clean and the probability of deciding that all compounds in all wells (or groups of wells) are clean. The following discussion will be phrased in terms of testing individual wells. However, it also applies to testing groups of wells.

For an individual statistical decision on one compound or well, the maximum probability of a false positive decision is denoted by the Greek letter alpha, a. This may also be called the comparison-wise alpha. When multiple chemicals or wells are being assessed, the overall alpha or experiment-wise alpha is the maximum probability of incorrectly declaring that the all compounds in all ground water wells at the site attain the cleanup standard.*   In this document it is assumed that the site will be declared to have

---

[1]Note that the procedures discussed here for assessing the attainment of the site from the results of multiple statistical tests are different from the typical presentations on "multiple comparison tests" or "experiment-wise versus comparison-wise tests" presented in many introductory statistics textbooks which use a different null hypothesis. Here all tests, rather than any single test, must have a significant result.

attained the cleanup standard only if all contaminants tested attain their specified cleanup
Standard

The probability of deciding that all compounds in all wells attain the cleanup standard, i.e., the overall a, depends on the number of statistical tests performed. If wells are assessed individually, more statistical tests will be performed than when assessing wells as a group. Thus, the decision on whether to group wells is related to the selection of the probabilities of a false positive or false negative decision.

The overall probability of declaring that a site has attained the cleanup standard depends on the:

- Number of contaminants and wells being assessed

- Concentrations of the contaminants being assessed;

- Statistical tests being used for the individual contaminants;

- Correlation between the concentration measurements of different contaminants in the same wells and contaminants in different wells; and

- Decision rules for combining the statistical results from each contaminant and well to decide if the overall site attains the cleanup Standard,

Although the calculation of the overall probability of declaring the site to attain the cleanup standard can be difficult, the following general conclusions can be stated when using the rule that all contaminants (or wells) must attain the cleanup standard:

- The probability of incorrectly deciding that the site attains the cleanup standard, the overall alpha, is always less than or equal to the maximum probability of mistakenly deciding that any one contaminant (or well) attains its cleanup standard (comparison-wise alpha).

- As the number of contaminants being assessed increases, the probability of deciding that the site is clean decreases, regardless of the true status of the site.

Choice of a strategy for combining the results from many statistical tests involves both policy and statistical questions. As a result no general recommendations can

be made in this document. When many contaminants or wells arc being assessed, consultation with a statistician is recommended.

2.3.6 Statistical Versus Predictive Modeling

A model is a mathematical description of the process or phenomenon from which the data are collected. A model provides a framework for extrapolating from the measurements obtained during the data collection period to other periods of time and for describing the important characteristics of the data. Perhaps most importantly, a model serves as a formal description, of the assumptions which are being made about the data. The choice of statistical method used to analyze the data depends on the nature of these assumptions. (See Appendix D for a discussion on modeling the data.)

Mathematical (deterministic) models can be used to predict or simulate the contaminant concentrations, the effect of treatment on the contaminants, the time required far remediation, and the remaining concentrations after remedial action. These models are referred to here as predictive models. To predict future concentrations these models typically use (1) mathematical formulae describing the flow of ground water and contaminants through porous or fractured media, (2) boundary conditions to specify the conditions at the start of the simulation (often based on assumptions), and (3) assumptions about the aquifer conditions. Predictive models are powerful tools, providing predictions in a relatively short time with minimal cost compared to the corresponding field sampling. They allow comparison of the expected results of different treatment alternatives. However, it is difficult to determine the probability of correctly or incorrectly deciding if the ground water attains the cleanup standard using predictive models, in part, due to the many assumptions on which the models are based.

On the other hand, the statistical models and procedures discussed in this document arc based on very few assumptions and can be used whether or not predictive models have been applied at the site. The statistical procedures can also be used as a check on the predictive models. Unlike the predictive models, the statistical models presented in this document for assessing attainment only use measurements from the period after remedial action has been terminated.

While this document makes the assumption that the attainment decision will be based on statistical models and procedures, predictive models and data collected prior to the sampling for the attain-t decision provide a guide as to which wells are to be used for assessing attainment, when to initiate an evaluation, and what criteria are to be used to define attainment of the cleanup standard. If predictive models are used in other ways for the attainment decision, consultation with a statistician is recommended. Due to the complexity of both site conditions and predictive modeling, other procedures which might be used to combine the results of predictive and statistical models are beyond the scope of this document

2.3.7 Practical Problems with the Data Collection and Their Resolution

With any collection of data there are possible problems which must be addressed by the statistical procedures. The problems discussed below are: measurements below the detection limit, missing data and very unusual observations, often called "outliers."

Measurements Below the Detection Limit

The detection limit for a laboratory measurementprocedure is the lowest concentration level which can be determined to be different from a blank. Measurements which arc below the detection limit may be reported in one of several different ways (Gilbert 1987). For example:

· A concentration value, with the notation that the reported concentration is below the detection limit;

· Less than a specified detection limit; or

· Coded as "below the detection limit" with no concentration or detection limit specified.

Special procedures arc required to use the below-detection-limit mesuremets in a statistical analysis. If, due to poor selection of the laboratory analysis method or unanticipated problems with the analysis, the cleanup standard is below the detection limit, the possible statistical procedures which might be used to compare the concentrations to the

cleanup standard are very limited and required many assumptions which are difficult to justify. As a result, this document only addresses the situation where the cleanup standard is greater than the detection limit.

For all of the procedures described in this manual, the following procedures for handling belowdetection-limit measurements are recommended:

·   Whenever the measured concentration for a given water sample is reported by the laboratory, use this concentration in the analysis even though it is below the detection limit;

·   When the concentration is reported as less than a specified detection limit, use the value at the detection limit as the measured concentration in the analysis; and

·   When the laboratory reports that the chemical concentration is "below the detection limit" with no specified detection limit, contact the analytical laboratory to determine the minimum detectable value, and use this value in the analysis. Do not treat below-detection-level measurements as missing.

Using the detection limit for values below the detection limit is conservative; i.e., errs in favor of minimizing health and environmental risks. Other methods of handling below-detection-limit problems can be used, but are more difficult to implement and have the potential of erring in the opposite direction. Selection of a method can be dependent upon the proportion of non-detects. Alternative procedures should be investigated and assessed as to how data are affected Some of these alternative procedures are discussed in the following references on detection limit problems: Bishop, 1985; Clayton et al., 1986; Gilbert, 1981; Gilliom and Helsel, 1986; Helsel and Gilliom, 1986; and Gleit, 1985.

Missing Values

Missing concentration values are different from below-detection measurements in that no information about the missing concentration (either above or below the detection level) is known. Missing values may be due to many factors, including either (1) non-collection of the scheduled sample. (2) loss of the sample before it is analyzed due to shipping or lab problems, or (3) loss of the lab results due to improper recording of results or loss of the data records.

In general, this problem can be minimized with appropriate planning and backup procedures and by using a proper **chain of custody procedures**, careful packaging and handling, clear labeling, and keeping copies of important records.

If the sample is lost shortly after collection, it is recommended that another sample be collected immediately to replace the lost sample as long as the time between the lost and replacement sample is less than half the time between successive samples specified in the sample design. Any deviations to the sampling design, including lost and replacement samples should be reported with the data and analysis. The replacement or substitution of missing data by numerical values is never recommended.

Outliers

In many statistical texts, measurements that are (1) very large or small relative to the rest of the data, or (2) suspected of being unrepresentative of the true concentration at the sample location are often called "outliers." Observations which appear to be unusual may correctly represent unusual concentrations in the field, or may result from unrecognized handling problems, such as contamination, lab measurement, or data recording errors. If a particular observation is suspected to be in error, the error should be identified and corrected, and the corrected value used in the analysis. If no such verification is possible, a statistician should be consulted to provide modifications to the statistical analysis that account for the suspected "outlier." For more background on statistical methods to handle outliers, see Barnett and Lewis (1984).

The handling of outliers is a controversial topic. In this document, all data <u>not known to be in error are considered to be valid</u> because:

- The expected distribution of concentration values may be skewed (i.e., non-symmetric) so that large concentrations which look like "outlicrs" to some analysts may be legitimate;

- The procedures recommended in this document are less sensitive to extremely low concentrations than to extremely high concentrations; and

- High concentrations arc of particular concern for their potential health and environmental impact.

2.4   Limitations and Assumptions of the Procedures Addressed in this Document

Because a single document cannot adequately address the wide variety of situations found at all Superfund sites, this document will only discuss those statistical procedures that are applicable to most sites and can be implemented without **a** detailed knowledge of statistical methods. Although the procedures recommended here will be generally applicable, specific objectives or situations at some sites may require the use of other statistical procedures. Where possible problems are anticipated, the text will recommend consultation with a statistician.

Due to the complex nature of conditions at Superfund sites, this document cannot address all statistical issues applicable either to Superfund sites or to assessing the attainment of cleanup standards. The discussion in this document is based on certain assumptions about what statistical tests will be requited and what the situations at the site will be. For completeness, the major assumptions are reviewed below.

- The contaminants are known;

- The ground water does not attain the cleanup standard until this assumption (that is the null hypothesis) is rejected using a statistical test;

- At the time of sampling for assessing attainment, there are no reasons to believe the ground-water concentrations might increase over time;

- Location of the monitoring and pumping (or treatment) wells arc fixed and arc not to be specified as part of the statistical methods. As a result, the attainment decision strictly applies only to the water in the wells, not to the ground water in general. To draw general conclusions about the ground water, additional assumptions must be made or additional wells must be established; and

    The cleanup standard is greater than the detection limit for all chemicals to be tested.

2.5   Summary

This guidance considers the variety and complexity of ground water conditions at Superfund sites and provides procedures which can be used at most sites and under most conditions. This chapter outlines some of the conditions found at Superfund sites and

some of the assumptions which have been made as a guide to the selection of statistical procedures presented in later chapters.

Errors are possible in evaluating whether a site attains the cleanup standards, resulting in false positive and false negative decisions. Statistical methods provide approaches for balancing these two decision errors and allow extrapolation in a scientifically-valid fashion.

This chapter reviews briefly the statistical concepts that farm a basis for the procedures described in this guidance. These include:

- false positive decision -- a site is thought to be clean when it is not;

- false negative decision -- a site is thought to be contaminated when it is not;

- mean -- the value that corresponds to the "center" of the concentration distribution;

- Qth proportion or percentile -- a value that separates the lower Q percent of the measurements from the upper 100-Q percent of the measurements;

- confidence intervals -- a sample-based estimate of a mean or percentile which is expressed as a range or interval of values which will include the true parameter value with a known probability or confidence;

- null hypothesis -- the prior assumption that the contaminant concentrations in the ground water at the site do not attain the cleanup Standard;

- hypothesis tests -- a statistical procedure far assessing attainment of the ground water by accepting or rejecting the null hypothesis on the basis of data; and

- power curve -- for a specified statistical test and sample size, the probability of concluding that the ground water attains the cleanup standard versus true concentration.

Unlike statistical tests in other circumstances, assessment of ground water requires consideration of the correlation between measurements across time and space. As a result of correlation across time, estimating the short-term and long-term concentrations requires different procedures. The-ground water is defined as attaining the cleanup stan-

dard if the statistical test indicates the long-term mean concentration or concentration percentile at the site attains the cleanup standard

When many wells or contaminants are assessed, careful consideration must be given to the decision procedures which arc used to combine data from separate wells or contaminants in order to determine if the site as **a** whole attains all relevant cleanup standards. How the data from separate wells are combined affects the interpretation of the results and the probability of concluding that the overall site attains the cleanup standard. **A complete discussion of how to assess attainment using multiple wells is beyond the scope of this volume.**

3. SPECIFICATION OF ATTAINMENT OBJECTIVES

This chapter discusses the specification of the attainment objectives, including the specific procedures to be used to assess attainment. The sampling and analysis plans, discussed in the next chapter, outline procedures to be used to assess attainment consistent with the attainment objectives. The specification of **objectives** must be completed by personnel familiar with the following:

- The characteristics of the ground water and contamination present at the waste site;

- The health and environmental risks of the chemicals involved; and

- The costs of sampling, analysis and remediation.

The flow chart in Figure 3.1 summarizes the steps required to specify the sampling and analysis objectives and shows where each step is discussed. In general, specification of **the attainment objectives** for the site under investigation involves specifying the following items:

- The wells to be sampled;

- **The sample collection and handling procedures;**

- The chemicals to be tested and the laboratory test methods to be used;

- The relevant cleanup standard for the chemicals under investigation;

- The parameter (e.g., the mean or a percentile) of the chemical concentration distribution which is to be compared to the cleanup standard

- The "false positive rate" for the statistical test (the confidence level for protection against adverse health and environmental risk);

- The precision to be achieved; and

- Any other secondary **objectives** for which the data are to be used which may affect the choice of statistical procedure.

Figure 3.1   Steps in defining the attainment objectives
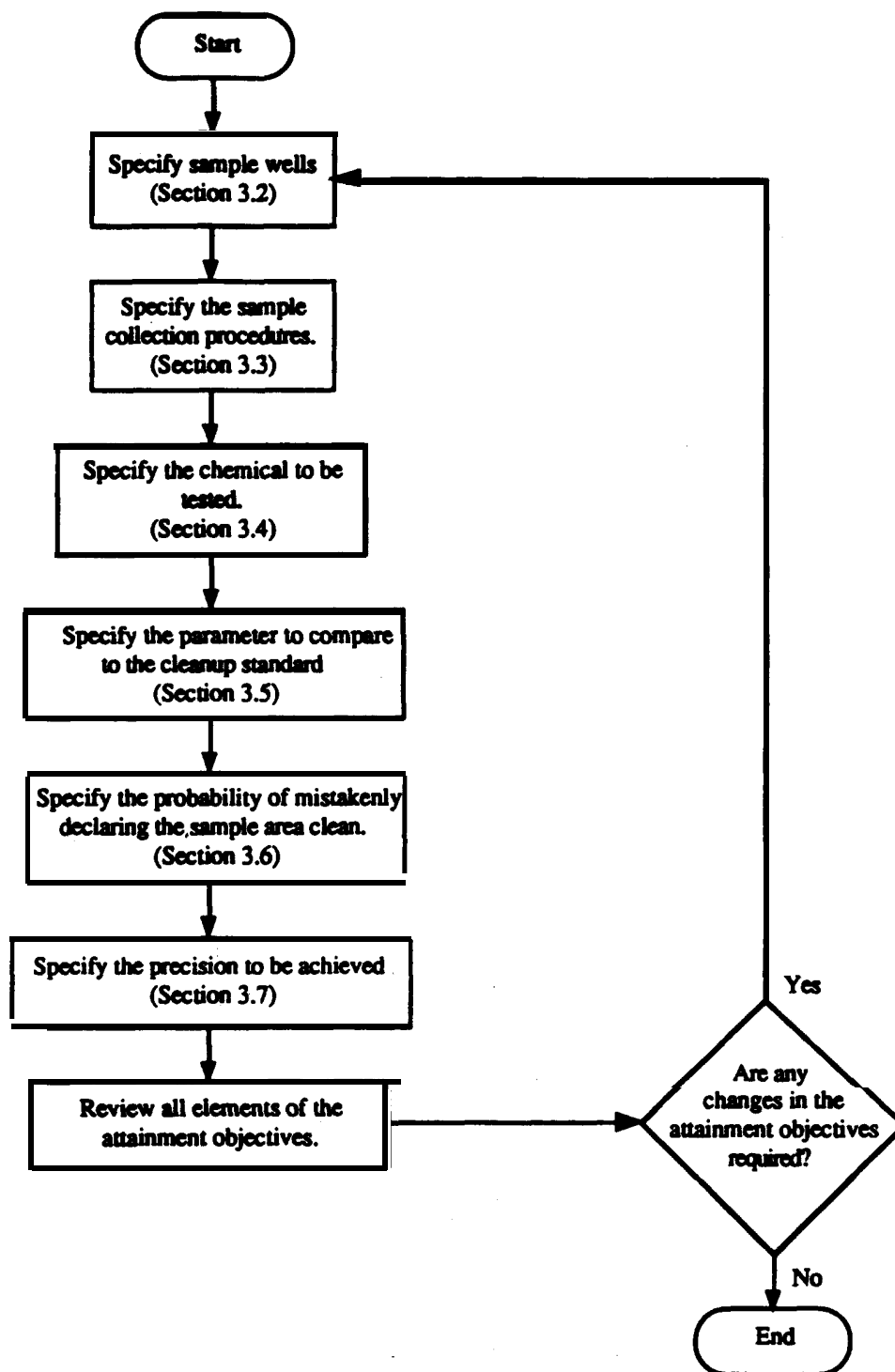
CHAPTER 3: SPECIFICATION OF ATTAINMENT OBJECTIVES

The items which make up the attainment objectives are discussed in detail in the following sections.

3.1 Data Quality Objectives

The Quality Assurance Management staff within EPA has developed requirements and procedures for the development of Data Quality Objectives (DQOs) when environmental data arc collected to support regulatory and programmatic decisions. Although the DQOs are an important part of the attainment objectives, they are discussed in detail elsewhere and will not be addressed here. For more information, readers should refer to U.S. EPA (1987a) and U.S. EPA (1987b).

3 . 2 Specification of the Wells to be Sampled

Wells within the site will be monitored and evaluated with respect to the applicable cleanup standards. Extending inferences from the sampled wells to the ground water in general must be made on the basis of both available data and expert knowledge about the ground-water system and not on the basis of statistical sampling theory. Careful selection of the ground-water wells to be used for assessment is required to ensure that attainment of the cleanup standard in the sampled wells implies to all parties concerned that the ground-water quality has been adequately protected.

Sections 2.2.3 and 2.3.5 provide more discussion on the implications of the decision on which wells must attain the cleanup standard.

3.3 Specification of Sample Collection and Handling Procedures

The results of any statistical analysis are only as good as the data on which it is based. Therefore, an important objective for sampling and analysis plan is to carefully define all aspects of data collection and measurement procedures, including:

- How the ground-water sample is to be collected;
- What equipment and procedures are to be used;

·   How the sample is to be handled between collection and measurements

·   How the laboratory measurements are to be made; and

·   What precision is to be achieved

One reference for guidance on these topics is The Handbook for Sampling and Sample Preservation of Water and Wastewater (U.S. EPA, 1982).

3.4   Specification of the Chemicals to be 'rested and Applicable Cleanup Standards

The chemicals to be tested should be listed. When multiple chemicals are tested, this document assumes that all chemicals must attain the relevant cleanup standard in order for the ground water from the well(s) to be declared clean.

The term "cleanup standard" is a generic term for the value to which the sample measurements must be compared. Throughout this document, the cleanup standard will be denoted by Cs. The cleanup standard for each chemical of concern must be stated at the outset of the study. Cleanup standards are determined by EPA in the process of evaluating site-specific cleanup alternatives. Final selection of the cleanup standard depends on many factors. These factors are discussed in Guidance on Remedial Actions for Contaminated Ground Water at Superfund Sites [Interim Final] (U.S. EPA, 1988).

3.5   Specification of the Parameters to Test

In order to define a statistical test to determine if the contaminant concentrations in ground water well(s) attain the cleanup standard, the characteristic of the concentrations which is to be compared to the cleanup standard must be specified. Such characteristics are called parameters. The two parameters discussed in this document for testing individual wells are the mean concentration and a specified percentile of the concentrations such as the median or the 90th percentile of the ground-water concentrations. The following sections discuss the criteria for selecting the parameters to test. These parameters have been defined previously in Section 2.3.1.

3.5.1 **Selecting the Parameters to Investigate**

- Criteria for selecting the parameter to use in the statistical attainment decision are:

- The criteria used to develop the risk-based standards, if known;

- Whether the effects of the contaminant being measured are acute or chronic;

- The relative sample sizes required;

- The likelihood of finding concentration measurements below the Cleanup standard; and

- The relative spread of the data.

For example, if the cleanup standard is a risk-based standard developed for the mean concentration over a specified period of time, it is logical that the cleanup standard be compared to the mean concentration. Alternatively, if the cleanup standard is a risk-based standard developed for extreme concentrations which should rarely be exceeded. it is logical to test an upper percentile of the concentration distribution.

Many considerations may go into the selection of the parameter to test. Table 3.1 presents criteria and conditions that support or contradict the use of each parameter.

Some general rules for selecting the parameter to test are:

(1) If the chemical contaminant of concern has short-term or acute effects on human health or the environment, testing of upper percentiles is recommended, with higher percentiles being chosen for testing when the distribution of contamination has a higher coefficient of variation.

(2) If the chemical contaminant of concern has long-term or chronic effects on human health or the environment, Table 3.2 shows the recommended parameter based on the coefficient of variation of the data and the likelihood of measurements below the detection level.

CHAPTER 3: SPECIFICATION OF ATTAINMENT OBJECTIVES

Table 3.1   Points to consider when trying to choose among the mean, upper proportion/percentile, a median

Parameter	Points to consider
Mean	1) Easy to calculate and estimate a confidence interval. 2) Useful when the cleanup standard has been based on consideration of carcinogenic a chronic health effects a long-term average exposure. 3) Useful when the data have little variation from sample to sample or season to season. 4) If the data have a large coefficient of variation (greater than about 1.5) testing the mean can require more samples than for testing an upper percentile in order to provide the same protection to human health and the environment 5) Can have high false positive rates with small sample sizes and highly skewed data, i.e. when the contamination levels are generally low with only occasional short periods of high contamination. 6) Not as powerful for testing attainment when there is a large proportion of less-thandetection-limit values. 7) Is adversely affected by outliers or errors in a few data values.
Upper Proportion Percentile	1) Requiring that an upper percentile be less than the cleanup standard can limit the occurrence of samples with high concentrations, depending on the selected percentile. 2) Unaffected by less-thandetection-limit values, as long as the detection limit is less than the cleanup standard. 3) If the health effects of the contaminant axe acute, extreme concentrations are of concern and are best tested by ensuring that a large proportion of the measurements are below a cleanup standard. 4) The proportion of the samples that must be below the cleanup standard must be chosen. 5) For highly variable or skewed data, can provide similar protection of human health and the environment with a smaller sample size than when testing the mean. 6) Is relatively unaffected by a small number of outliers.

Table 3.1   Points to consider when trying to choose among the mean, upper proportion/percentile, or median (continued)

Parameter	Points to Consider
Median	1) Has benefits over the mean because it is not as heavily influenced by outliers and highly variable data, and can be used with a large number of less-than-detection-limit values. 2) Has many of the positive features of the mean, in particular its usefulness for evaluating cleanup standards based on carcinogenic or chronic health effects and long-term average exposure. 3) For positively skewed data, the median is lower than the mean and therefore testing the median provides less protection for human health and the environment than testing the mean. 4) Retains some negative features of the mean in that testing the median will not limit the occurrence of extreme values.

Table 3.2   Recommended parameters to test when comparing the cleanup standard to the concentration of a chemical with chronic effects'

	Proportion of the data with concentrations below the detection limit:	
	(Perhaps 30%)	High (Perhaps > 30%)
Large Coefficient of Variation (Perhaps cv > 1.5)	MeanOr Upper Percentile (Upper percentile requires fewer samples)	Upper Percentile
Intermediate Coefficient of Variation (Perhaps 1.5 > cv > .5)	Mean or Upper Percentile	Upper Percentile
Small Coefficient of Variation (Perhaps cv < .5)	Mean or Median	Median

---

[1]Based on Westat simulations and analysis summarized in an internal Westat memo.

3.5.2   Multiple Attainment Criteria

In some situations two or more parameters might be chosen. For example, both the mean and an upper percentile can be tested using the rule that the ground water attains the cleanup standard if <u>both</u> parameters are below the cleanup standard.

Other more complicated criteria may be used to assess the attainment to the cleanup criteria. Examples of multiple criteria are:

·   It is desirable that most of the ground-water samples have concentrations below the cleanup standard and that the concentrations which are above the cleanup standard are not too large. This may be accomplished by testing if the 75th percentile is below the cleanup standard and the mean of those concentrations which are above the cleanup standard is less than twice the cleanup standard. This combination of tests can be performed with modifications of the methods presented in this document.

·   It is desirable that the mean concentration be less than the cleanup standard and that the standard deviation of the data be small. This may be accomplished by testing if the mean is below the cleanup standard and the standard deviation is below a specified value. This document does not address testing the standard deviation, variance, or coefficient of variation against a standard.

For testing of multiple criteria not discussed in the guidance document, consultation with a statistician is recommended.

3.6   Specification of Confidence Levels for Protection Against Adverse Health and Environmental Risks

In order to design a statistical test for deciding if the ground water attains the cleanup standard, those specifying the sampling and analysis objectives must select the false positive rate. This rate is the maximum probability that the test results will show the ground water to be clean when it is actually contaminated. It is usually set at levels such as 0.10, 0.05, or 0.01 (that < 10%, 5%, or 1%), depending on the potential consequences of deciding that the ground water is clean when, in fact, it is not clean. While different false positive rates can be used for each chemical, it is recommended that the same rate be used

for all chemicals being investigated.[1] For a further discussion of false positive rates see Section 2.3.4 or Sokal and Rohlf (1981).

3.7   Specification of the Precision to be Achieved

Recision generally refers to the degree to which repeated measurements are similar to one another. In this context it refers to the degree to which estimates from different samples are similar to one another. Decisions based on precise estimates will usually be the same from sample to sample. The desired precision of the statistical test is specified by the desired confidence in the statistical decisions resulting from the statistical test.

Specification of the precision to be achieved is required to completely define the statistical test to use. The precision which is to be achieved can be defined by specifying the-parameter value for which the probability of a false negative decision is to be controlled. For a definition of "false negative" see Section 2.3.4.

To completely define the precision when testing the mean, the following items must be specified:

- α, the false positive rate;

- Cs, the cleanup standard;

- μ_1, the mean concentration at which the false negative rate is to be specified; and

- β, the false negative rate at μ_1.

To completely define the precision when testing percentiles, the following items must be specified:

- α, the false positive rate;

- Cs, the cleanup standard;

---

[1]When testing multiple chemicals from the same ground water samples. the overall false positive rate will be approximately the same as that for individual chemical tests if the concentrations of different chemicals are highly correlated. In situations when the concentrations are not highly correlated, the overall false positive rate for the entire site will be smaller than that specified for the individual chemicals.

P_O, the largest acceptable proportion of ground-water samples with concentrations above the cleanup standard;

- **P_1, the value of the proportion for which the false negative rate is to be specified (comparable to μ_1, when testing means);**

- **β, the false negative rate at P_1.**

The specification of these items is discussed in &tail Chapter 2 of this document and in Chapter 6 and 7 of Volume L The reader should refer to Volume I for detailed instructions on how these items arc to be specified.

3.8   Secondary Objectives

The sampling and analysis data may be used for purposes other than assessing the attainment of the cleanup standards. For example, they may be used to determine the relationship between concentrations of different contaminants, to determine the seasonal patterns in the measurements, or to get measurements on a contaminant not being assessed. **These secondary objectives may** determine what procedure is used to collect the samples or how often the samples arc collected.

3.9   Summary

This chapter discussed the specification of the various items which make up the attainment objectives. The objectives will be specified by EPA, regulatory agencies, and others familiar with the site, the environmental and health risks, and the sampling and remediation costs. As part of the objectives, careful consideration must be given to defining the wells to be tested, the ground-water sampling and analysis procedures, the statistical parameter to be compared to the cleanup standard, and the precision and confidence level desired. The attainment objectives provide the background for developing the sampling and analysis plans discussed in Chapter 4.

4. DESIGN OF THE SAMPLING AND ANALYSIS PLAN

Once the attainment objectives are specified by program and subject matter personnel, statisticians and hydrogeologists can be useful in designing important components of sampling and analysis plans. The sampling plan specifies how the water samples are to be collected, stored, and analyzed, and how many samples to collect. The analysis plan specifies which of the statistical procedures presented in the following chapters are to be used. The sampling and analysis plans are interrelated and must be prepared together. The decision regarding attainment of the cleanup standard can be made only if the field and laboratory procedures (in the sampling plan) provide data that are representative of the ground water and can provide the parameter estimates (from the analysis plan) specified in the attainment objectives.

The specification of the sampling and analysis plans will depend on the characteristics of the waste site and the evidence needed to evaluate attainment. The statistical methods must be consistent with the sample design and attainment objectives. If there appears to be any reason to use different sample designs or analysis plans than those discussed in this guidance, or if there is any reason to change either the sample design or the analysis plan after field data collection has started, it is recommended that a statistician be consulted.

4.1 The Sample Design

The sample design, or sampling plan; outlines the procedure for collecting the data, including the timing, location, and filed procedures for obtaining each physical water sample. The discussion here focuses on the timing of the sample collection activities. Common types of sample design are random sampling and systematic sampling. Either of these sample collection procedures can require a fixed number of samples or use sequential sampling in which the number of samples to be collected is not specified before the sampling period.

4.1.1 Random Sampling

In a random sample design, samples arc collected at random times throughout the sampling period. For example, using simple random sampling 48 sample collection times might be randomly selected within a four year sampling period. Using simple random sampling, some years may have more samples than other years. One alternative to 'simple random sampling is stratified random sample in which 12 samples arc collected in each of four years, with the sample times within each year being randomly selected. In either case, with a simple random sample the time interval between the collection of the water samples will vary. Some samples may be collected within days of each other while at other times there may be many months between samples.

Although random sampling has some advantages when calculating the statistical results for short term tests (Chapter 5), systematic sampling is generally recommended far assessing attainment.

4.1.2 Systematic Sampling

Using a systematic sample with a random start, ground water samples arc collected at regular time intervals, (such as every week, month, three months, year, etc.) starting from the fast sample collection time, which is randomly determined. In this document, the systematic sample with a random start will be referred to as simply a systematic sample.

When sampling ground water, a systematic sample is usually preferred over a simple random sample because:

- Extrapolating from the sample period to future periods is easier with a systematic sample than a simple random sample;

- Seasonal cycles can be easily identified and accounted for in the data analysis;

- A systematic sample will be easier to administer because of the fixed schedule for sampling times; and

- Most ground water samples have been traditionally collected using a systematic sample.

The procedures described in the following chapters assume that either a systematic or random sample is used when collecting data for a short term test and that a systematic sample is collected when assessing attainment. If other sample designs arc considered, consultation with a statistician is recommended. It should be noted that when implementing a systematic sample, care must be taken to capture any periodic seasonal **variations in the data. The seasonal patterns in the data will repeat themselves (after adjust-**ing for measurement errors) following a regular pattern. For example, if ground water measurements at a site exhibit seasonal fluctuations, following the four seasons of the year, collecting data every six months may miss some important aspects of the data, such as high or low measurements, and could present a misleading picture of the status of the site. Because many seasonal patterns will have a yearly cycle (due to yearly patterns in surface water recharge) the text will often refer to the number of samples per year instead of the number of samples per seasonal cycle.

One variation of the standard systematic sample uses a different random start for each years data. For example, if one water sample is collected each month, in the first year samples might be collected on the 17th of each month and in the second year on the 25th of each month, etc. This variation is preferred when there arc large seasonal fluctuations in the data.

Follow the steps below to specify the systematic sample design:

(1)   Determine the period of any seasonal fluctuation (i.e., time period between repeating patterns in the data). This period will usually be a year. If no period is discernible from the data, the use of a one-year period is recommended

(2)   Determine the number of ground water samples, n, to collect in each year (seasonal cycle) and the corresponding sampling period between samples. A minimum of four sample collections per year is recommended.

(3)   Specify the beginning of the attainment sampling period.

(4)   Randomly select a sampling time during the first sampling period.

(5)   Subsequent sampling should be at equal intervals of the sampling period after the first sample is collected.

In practice, the samples need not be collected precisely at the time called fur by the sampling interval. However, the difference between the scheduled sampling time and the actual time of sampling should be small compared to the time between successive

samples. The sample collection of subsequent samples should not be changed if one sample is collected early or later than scheduled. An example of the procedure is presented in Box 4.1.

Box 4.1
Example of Procedure for Specifying a Systematic Sample Design

(1) The seasonal cycle in the measurements is assumed to have a period of one year.

(2) Based on the methods in Chapter 8, it is decided to collect 6 samples per year, one every two **months.**

(3) The attainment sampling period is to start on April 1, 1992

(4) The first sampling time during the first two-month sampling period is randomly selected using successive flips of a coin. Each flip divides the portion of the sampling period being considered into two. Heads chooses the earlier half, tails the later half. After 5 flips, the chosen day for the first sample is April 15.

(5) Samples are scheduled to be collected the 15th of every other month. If one sample is collected on the 20th of a month, the subsequent sample should still be targeted for the 15th of the appropriate month.

4.1.3   Fixed versus Sequential Sampling

For most statistical tests or procedures, the statistical analysis is performed after the entire set of water samples has been collected and the laboratory results arc complete. This procedure uses a fixed sample size test because the number of samples to be collected is established and fixed before the sample collection begins. In sequential testing, the water samples are analyzed in the lab and the statistical analysis is performed as the sample collection proceeds. A statistical analysis of the data collected at any point in time is used to determine whether another sample is to be collected or if the sampling terminates. Sequential statistical tests for data collected using sequential sampling of ground water are discussed in detail in Chapter 9.

CHAPTER 4: DESIGN OF THE SAMPLING AND ANALYSIS PLAN

4.2   The Analysis Plan

Similar to sampling plan, planning an approach to analysis begins before the first physical sample is collected. The first step is to define the attainment objectives, discussed in Chapter 3. If the mean is to be compared to cleanup standards, the statistical methods will be different than if a specified proportion of the samples must have concentrations below the cleanup standard. Second, the analysis plan must be developed in conjunction with the sampling plan discussed earlier in this chapter.

Third, determine the appropriate sample size (i.e. the number of physical samples to be collected) for the selected sample and analysis plan. Whether using a fixed sample size or sequential design, calculate the sample size for the fixed sample size test. Use this sample size for comparing alternate plans. In some cases, the number of samples is determined by economics and budget rather than an evaluation of the required accuracy. Nevertheless, it is important to evaluate the accuracy associated with a prespecified number of samples.

Fourth, the analysis plan will describe the statistical evaluation of the data.

In many cases, specification of the sampling and analysis plan will involve consideration of several alternatives. It may also be an iterative process as the plans are refined. In cases where the costs of meeting the attainment objectives are not acceptable, it may be necessary to reconsider those objectives. When trying to balance cost and precision, decreasing the precision can decrease the sampling and lab costs while increasing the costs of additional remediation due to incorrectly concluding that the ground water does not attain the cleanup standard. In this situation, consultation with a statistician, and possibly an economist, is recommended.

Chapters 8 and 9 offer various statistical methods, depending on attainment objectives and the sampling plan. Table 4.1 presents the locations in this document where various combinations of analysis and sampling plans are discuss&

CHAPTER 4: DESIGN OF THE SAMPLING AND ANALYSIS PLAN

Table 4.1   Locations in this document of discussions of sample designs and analysis for ground water sampling

Type of Evaluation	Analysis Method	Sample Design	
		Fixed Sample Site	Sequential
Continuous Data	Test of the Mean	**Sections 8.3** and 8.4	Sections 9.3 and 9.4
Discrete Data	Test of Proportions	**Section 8.5**	**section 9.5**

4.3   Other Considerations for Ground Water Sampling and Analysis Plans   .

At a minimum, all ground water sampling and analysis plans should' **specify:**

- ·   sampling objective;
- ·   sampling preliminaries;
- ·   Sample collection;
- ·   In-situ field analysis;
- ·   Sample preservation and analysis;
- ·   Chain of custody control;
- ·   Analytical procedures and quantitation limits;
- ·   Field and laboratory QA/QC plans;
- ·   Analysis procedures far any QC data;
- ·   statistical analysis procedures; and
- ·   Interim and final statistics to be provided to project personnel.

For more information on other considerations in ground water sampling and analysis, see RCRA Ground Water Monitoring Technical Enforcement Guidance Document (EPA, 1986b).

4.4   **Summary**

Design of the sampling and analysis plan requires specification of attainment objectives-by program and subject matter personnel. The sampling and analysis objectives can be refined with the assistance of statistical expertise. The sample design and analysis plans go together, therefore, the methods of analysis must be con&tent with the sample design and both must be consistent with the characteristics of the data and the attainment objectives.

Types of sample design include simple random sampling or systematic sampling, and fixed sample size or sequential sampling. This guidance assumes the data will be collected using a systematic sample when assessing attainment.

Steps required to plan an approach to analysis are:

·   Specify the attainment objectives;

·   Develop the analysis plan in conjunction with the sampling plan,

·   Determine the appropriate sample size; and

·   Describe how the resulting data will be evaluated.

5. DESCRIPTIVE STATISTICS AND HYPOTHESIS TESTING

This chapter introduces the reader to some basic statistical procedures that can be used to both describe (or characterize) a set of data, and to test hypotheses and make inferences from the data. The procedures use the mean or a selected percentile from a sample of ground water measurements along with its associated confidence interval. The confidence interval indicates how well the population (a actual) mean on percentile can be estimated from the sample mean or percentile. These parameter estimates and their confidence intervals can be useful in communicating the current status of a clean up effort. Methods of assessing whether the concentrations meet target levels are useful for evaluating progress of the remediation. The statistical procedures given in this chapter arc called "parametric" procedures. These methods usually assume that the underlying distribution of the data is known. Fortunately, the procedures perform well even when these assumptions arc not strictly true; thus they are applicable in many different field conditions (see Conover, 1980). The text notes situations in which the statistical procedures are sensitive to violations of these assumptions. In these cases, consultation with a statistician is **recommended.**

Calculations of means, proportions, percentiles, and their corresponding standard errors and their associated confidence intervals (measures of how precise these estimated means, proportions, or percentiles are) will be described. The statistics and inferential procedures presented in this chapter are appropriate <u>only for es</u>timating short-**term characteristics of contaminant** <u>levels</u>. By "short-term characteristics" we mean characteristics such as the mean or percentile of contaminant concentrations during the fixed period of time during which sampling occurs. For example, data collected over a one year period can be used to characterize the mean contaminant concentrations during the year. Procedures for estimating the long-term mean and for assessing attainment arc discussed in Chapters 8 and 9. The distinction between the methods of this chapter and those given in Chapters 8 and 9 is that inferences based on short-term methods apply only to the specified period of sampling and not to future points of time. The procedures discussed in this chapter can be used in any phase of the remedial effort; however, they will be most 'useful during treatment, as indicated in Figure 5.1. For a further discussion of short- versus long-term tests. see Section 2.3.2.

Figure 5.1   Example scenario for contaminant measurements during successful remedial action
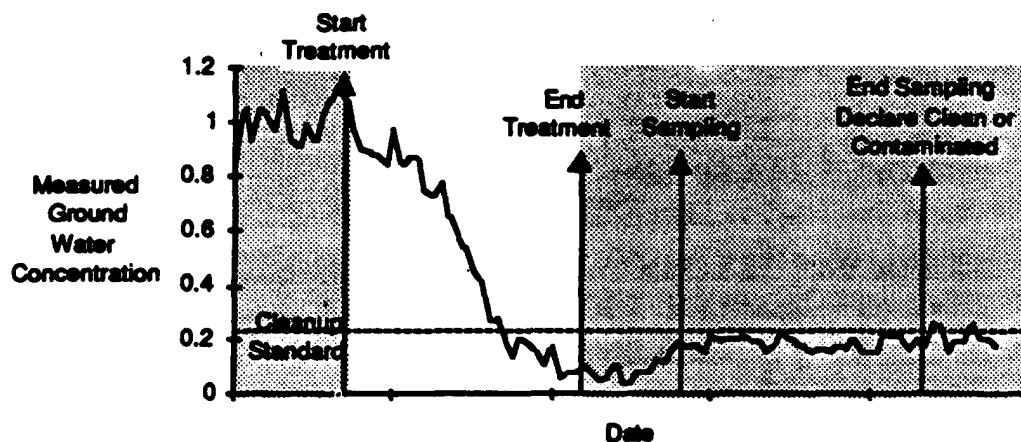
Much of the material on means, percentiles, standard errors and confidence intends has been previously presented in Volume I of this series of guidance documents. To avoid duplication, the discussion of these topics in this chapter is limited to the main points. The reader should refer to Volume I (Section 6.3 and 7.3) for additional details.

Some Notations and Definitions

Unless stated otherwise, the symbols x_1, x_2, ..., x_i, ..., x_N will be used in -this manual to denote the contaminant concentration measurements for N ground-water samples taken at regular intervals during a specified period of time. The subscript on the x's indicates the time order in which the sample was drawn; e.g., x_1 is the first (or oldest) measurement while x_N is the Nth (or latest) measurement. Collectively, the set of x's is referred to as a data set, and, in general, x_i will be used to denote the ith measurement in the data set.

The data set has properties which can be summarized by individual numerical quantities such as the sample mean, standard deviation or percentile (including the median). In general, these numerical quantities are called sample statistics. The sample mean or median provides a measure of the central tendency of the data or the concentration around which the measurements cluster. The sample standard deviation provides a measure of the spread or dispersion of the data, indicating whether the

sample data are relatively close in value or somewhat spread out about the mean. The sample variance is the square of the standard deviation. The computational formulas for these quantities arc given in subsequent sections.

As one of many possible sets of samples which could have been obtained from a ground water well, the mean, standard deviation, or median of the observed sample of measurements, $x_1, x_2, \ldots, x_N$, represent just one of the many possible values that could have been obtained. Different samples will obviously lead to different values of the sample mean, standard deviation or median. This sample-to-sample variability is referred to as sampling error or sampling variability and is used to characterize the precision of sample-based estimates.

The precision of a sample-based estimate is measured by a quantity known as the standard error. For example, an estimate of the standard error of the mean will provide information on the extent to which the sample mean can be expected to vary among different sets of samples, each set collected during the same sample collection period. The standard error can be used to construct confidence intervals. A confidence interval provides a range of values within which we would expect the true parameter value to lie with a specified level of confidence. Statistical applications requiring the use of standard errors and confidence intervals are described in detail in the sections which follow. The standard error differs from the standard deviation in that the standard deviation measures the variability of the individual observations about their mean while the standard error measures the variability of the sample mean among independent samples.

Throughout the remainder of this document, certain mathematical symbols will be used. For reference, some of the frequently-used symbols are summarized in Table 5.1.

Finally, note that the equations that follow assume that there are no missing **observations.** If there arc relatively few missing observations (i.e., five percent or less of the data set have missing data for the chemical measurement under consideration), the ground-water samples with missing data should be deleted from the data set. In this case, all statistics should be calculated with the available data, where the "sample size" now corresponds to the number of samples which have non-missing concentration values. However, if more than five percent of the data arc missing, a statistician should be

CHAPTER 5: DESCRIPTIVE STATISTICS

consulted. Additional comments regarding the treatment of the missing values will be given in the sections where specific statistical procedures are being discussed.

Table 5.1   Summary of notation used in Chapters 5 through 9

Symbol	Definition
x_i	Contaminant measurement for the ith ground water sample. For measurements reported as below detection, x_i = the detection limit.
	In the discussion of regression, the dependent variable, often the sample collection time, sometimes the sample collection time after a transformation.
m	The number of years for which data were collected (usually the analysis will be performed with data obtained over full year periods)
n	The number of sample measurements per year (for monthly data, n = 12; for quarterly data, n = 4). This is also referred to as the number of "seasons" per year
N	The total number of sample measurements (for data obtained over full year periods with no missing values, N = nm)
x_{jk}	An alternative way of denoting a contaminant measurement, where k = 1, 2, ..., m denotes the year; and j = 1, 2, . . . , n denotes the sampling period (season) within the year. If there are no missing values, the subscript for x_{jk} is related to the subscript for x_i in the following manners: i = (k-l)n + j.
$\bar{x}$	The mean (or average) of the N ground water measurements.
s^2	The variance of the N ground water measurements.
s	The standard deviation of the N ground water measurements.
$s_{\bar{x}}$	The standard error of the mean (this is calculated differently for long and short term tests).
Df	The degrees of freedom associated with the standard error of an estimate.
Cs	The cleanup standard relevant to the ground water and the contaminan being tested.

Table 5.1   summary of notation used in Chapters 5 through 9 (continued)

Symbol	Definition
P	The "true" but unknown proportion of the ground water with contaminant concentrations greater than the cleanup Standard.
P_O	The criterion for defining whether the sample area is clean or contaminated using proportions. According to the attainment objectives, the ground water attains the cleanup standard if the proportion of the ground water samples with contaminant concentrations greater than the cleanup standard is less than P_o i.e., the ground water is clean if $P < p_O$.
P_1	The value of P under the alternative hypothesis for which a specified false negative rate is to be controlled.
α	The desired false positive rate for the statistical test. The false positive rate for the statistical procedure is the probability that the ground water will be declared to be clean when it is actually Contaminated.
β	The false negative rate for the statistical procedure is the probability that the ground water will be declared to be contaminated when it is actually clean (see Section 2.3.4 and Table 2.1 for further discussion).
y_i	In calculating proportions, the coded value of x_i. If the concentration in sample i is less than the cleanup standard ($x_i < Cs$), then $y_i = 0$. If the concentration in the sample is greater than or equal to the cleanup standard ($x_i \geq Cs$), then $y_i = 1$. In the discussion of regression, the independent variable, often the contaminant measurement for the ith ground water sample, sometimes the measurement after transformation.
μ	The "true" but unknown mean concentration across the sample area, the population mean.
μ_1	The value of μ under the alternative hypothesis for which a specified false negative rate is to be controlled ($\mu_1 < \mu$).

5.1   Calculating the Mean, Variance, and Standard Deviation of the Data

The basic equation presented in Box 5.1 for calculating the mean and variance (or standard deviation) for a sample of data can be found in any introductory statistics text (e.g., Sokal and Rohlf, 1981 or Neter, Wasserman, and Whitmore, 1982).

Box 5.1
calculating Sample Mean, variance and standard Deviation

Designate the individual data values from a sample of N observations as x_1, x_2, ..., x_N. The sample mean (or arithmetic average) of these observations, indicated by $\bar{x}$, is given by

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} \tag{5.1}$$

The equation for the sample variance, s^2, is

$$s^2 = \frac{\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}}{N-1} = \frac{\sum_{i=1}^{N} x_i^2 - N\bar{x}^2}{N-1} \tag{5.2}$$

The corresponding equation for the standard deviation of the data is

$$s = \sqrt{\frac{\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}}{N-1}} \tag{5.3}$$

Both the variance and standard deviation have N-1 degrees of freedom.

The mean and standard deviation are descriptive statistics that provide information about certain properties of the data set. The mean is a measure of the

concentration around which the individual measurements cluster (the location central tendency). The standard deviation (or equivalently, the variance) provides a measure of the extent to which sample data vary about their mean.

Note that samples with missing data should be excluded from these calculations, in which case N equals the number of samples with non-missing observations. If more than five percent of the data have missing values, consult a statistician.

The term, "Degrees of Freedom," denoted by Df, can be thought of as a measure of the amount of information used to estimate the variance (or standard deviation) and thus reflects the precision of the estimate. For example, the variance and standard deviation calculated from formulas (5.2) and (5.3). respectively, are based on "N-l degrees of freedom."   For other estimates of variance (e.g., see Section 5.2.2 or 5.2.4). the associated degrees of freedom may be different. The degrees of freedom is used in calculating confidence intervals and performing hypothesis tests.

5.2   Calculating the Standard Error of the Mean

The standard error of the mean (denoted by $s_{\bar{x}}$) provides a measure of the precision of the mean concentration obtained from ground-water samples that have been collected over a period of time. The standard error of a statistic (e.g., a mean) reflects the degree to which that statistic will vary from one randomly selected set of samples to another (each of the same size). Small values of $s_{\bar{x}}$ indicate that the mean is relatively precise, whereas large values indicate that the mean is relatively imprecise.

A number of different formulas are available for calculating the standard error of the mean. The appropriate formula to use depends on the behavior of contaminant measurements over time and the sampling design used for sample collection. Four methods of calculating the standard error and the conditions under which they are applicable are discussed below. Care should be taken in each case to insure that an appropriate estimation formula for the standard error is chosen. Appropriate formulas should be decided on a site-by-site basis.

CHAPTER 5: DESCRIPTIVE STATISTICS

General rules for the selection of the formula for calculating the standard error of the mean include:

- If the ground water samples are collected **using** a <u>random</u> sample, use the formulas in section 5.2.1 and Box 5.2.

- If the ground water samples are collected using a <u>systematic</u> sample:

  - Use the formulas in Section 5.2.4 and Box 5.6 unless there are no obvious seasonal patterns or the serial correlations in the data are not significant.

  - Use the formulas in Section 5.2.2 and Box 5.3 if there are obviously no seasonal patterns in the data however the data might be correlated.

  - Use the formulas in Section 5.2.3 and Box 5.4 if there are seasonal patterns in the data and serial correlations in the residuals are not significant.

  - Use the formulas in Section 5.2.1 and Box 5.2 if there are obviously no seasonal patterns in the data and serial correlations in the data are not significant.

  - If there are trends in the data consider using regression methods (Chapter 6). If regression methods are not used and the trends are small relative to the variation of the data, the methods using differences (Sections 5.2.2 and 5.2.4) are preferred over the other methods.

Sections 5.3 and 5.6 discusses procedures for estimating the serial correlation and statistical tests for determining if it is significant.

5.2.1   Treating the Systematic Observations as a Random Sample

The simplest method of estimating the standard error is to treat the systematic sample as a simple random sample (see Section 4.1). In this case, the standard error of the mean (denoted by $s_{\bar{x}}$) is given by the equations in Box 5.2. Formula (5.4) will provide a reasonably good estimate of the standard error if the contamination is distributed randomly with respect to time. The formula may overstate the standard error if there are trends in contamination over time, seasonal patterns or if the data are serially correlated.

Box 5.2
Calculating the Standard Error Treating the Sample
as a Simple Random Sample

$$s_{\bar{x}} = \frac{s}{\sqrt{N}} \qquad (5.4)$$

where s is the standard **deviation** of the data as computed from equation (5.3) and N is the number of non-missing observations. Equation (5.4) is equivalent to

$$s_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}}{N(N-1)}} \qquad (5.5)$$

The degrees of freedom for this estimate of the standard error is N-l.

5.2.2 Estimates From Differences Between Adjacent Observations

Another method in common use is based on overlapping pairs of consecutive observations. That is, observation 1 is paired with observation 2.2 with 3, 3 with 4, and so on. This method often gives a more accurate estimate of the standard error if the serial correlation between successive observations is high. The computational formula for this estimate of the standard error is given in Box 5.3 (e.g., see Kish, 1965, page 119 or Wolter, 1985, page 251).

If the data are independent, that is if the samples are collected using a random sample or if the data have no seasonal patterns or serial correlations, the standard error calculated using equation (5.6) will be less precise than that using equation (5.4). Since most statistics text books assume that the data are independent, these text books present only equation (5.4) for estimating the standard error of the mean. However, when using a systematic sample, the data are rarely independent. When the data are not independent, equation (5.4) may over estimate the standard error of the short term mean. On the other hand, equation (5.6) is preferred because it provides a less biased estimate of the standard error of the short-term mean. Calculation of the standard error using the differences between adjacent observations, equation (5.6), is not appropriate for estimating

the standard error of a long-term mean. Because systematic samples and short term means (i.e., the mean of the limited population being sampled) are often of interest in survey sampling, equation (5.6) is more commonly used in the analysis of sample surveys.

Box 5.3
Calculating the Standard Error Using Estimates Between Adjacent

$$s_{\bar{x}} = \sqrt{\frac{\sum_{i=2}^{N} (x_i - x_{i-1})^2}{2N(N-1)}} \qquad (5.6)$$

The number of degrees of freedom for the standard error given by (5.6) is approximately $\frac{2N}{3}$, as suggested by DuMouchel, Govindarajulu and Rothman (1973). When using this formula, round the approximate degress of freedom down to the next smallest integer.

We suggest that this method of successive differences using overlapping pairs be used to estimate the standard error of the mean unless there are obvious seasonal patterns in the data, or seasonal patterns are expected. If there are seasonal patterns or trends in the data, equation (5.6) will tend to overestimate the standard error. If the sample data reflect seasonal variation, the method for computing the standard error discussed in the next section should be employed.

5.2.3   Calculating the Standard Error After Correcting for Seasonal

The formulas given in the preceding sections for calculating the standard error are are appropriate for data exhibiting seasonal variability. Seasonal variability is generally indicated by a regular pattern that is repeated every year. For example, Figure 5.2 shows 16 chemical observations taken at quarterly intervals. Notice that beginning with the first observation, there is a fairly obvious seasonal pattern in the data. That is, within each year, the first quarter observation tends to have the largest value, while

the third quarter observation tends to have the smallest value. Over the year, the general pattern is for the concentration to start at a high value, decrease in the second quarter, decrease again in the third quarter, and then in the fourth quarter..

Figure 5.2   Example of data from a monitoring well exhibiting as a seasonal pattern
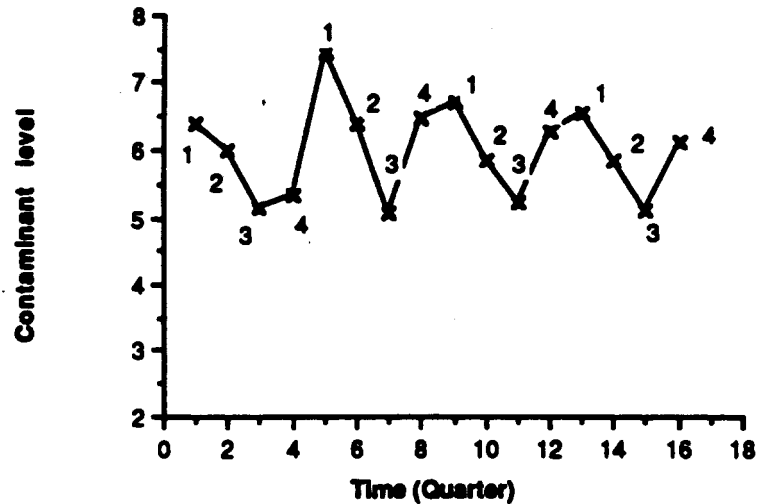
When the data exhibit regular seasonal patterns, the seasonal means should **be calculated separately and then used to "adjust" the sample data. Specifically, let x_{ik}** denote the observed concentration for the ground water sample taken from the jth time point in year k. Let n be the number of "seasons" in a seasonal cycle. Note that if data arc collected every month, then we have n = 12 and j = 1, 2, . . . , 12.   However, if data arc collected quarterly, then we have n = 4 and j = 1, 2, 3, 4.   In general, let j = 1, 2, . . . , n; and k = 1, 2, . . . , m_j, where m_j is the number of non-missing observations that arc available for season j. Note that m_j will equal m (the number of years) far all j (i.e., for all seasons) unless some data arc missing. Even if the seasonal effects arc relatively small, it is recommended that the seasonal means be subtracted from the sample data.   The presence of "significant" seasonal patterns can be formally tested by means of analysis of variance (ANOVA) techniques. A statistician should be consulted for more information about these tests.

CHAPTER 5: DESCRIPTIVE STATISTICS

The equations for the jth seasonal average, the average of the m_j **(non-missing)** sample observations for season j, and the sample _residual_ after correcting for the seasonal means are given in Box 5.4. Additional discussion of methods for adjusting for Seasonality can be found in _Statistical_ _Analysis of Ground-Water Monitoring Data at_ RCRA _Facilities_ (EPA, 1989b).

Box 5.4
Calculating Seasonal Averages and Sample Residuals

The jth seasonal average is:

$$\bar{x}_j = \frac{1}{m_j} \sum_{k=1}^{m_j} x_{jk} \qquad (5.7)$$

where m_j is the number of non-missing observations available for season j.

The sample _residual_ after correcting for the seasonal means is defined by

$$e_{jk} = x_{jk} - \bar{x}_j \qquad (5.8)$$

By subtracting the estimated seasonal means from the measurements, the resulting values, E_{jk} (or residuals), will all have an expected mean of zero and the variation of the e_{jk} about the value zero reflects the general variation of the observations. Using the residuals calculated from formula (5.8). the standard error of the mean can be calculated from the equations in Box 5.5 (e.g., see Neter, Wasserman, and Kutner, 1985, pages 573 and 539). **The term s_e^2 is referred to as the mean square error and is standard output in** many statistical computer packages (e.g.. see Appendix E for details on using SAS to calculate the relevant statistics).

Box 5.5
Calculating the Standard Error After Removing Seasonal Averages

The standard error based on the residuals resulting from removing the seasonal averages is:

$$s_{\bar{x}} = \sqrt{\frac{\sum_{j=1}^{n} \sum_{k=1}^{m_j} e_{jk}^2}{N(N-n)}} \qquad (5.9)$$

where

$$N = \sum_{j=1}^{n} m_j. \qquad (5.10)$$

The degrees of freedom associated with the standard error is Df = N-n.

Note that equation (5.9) can also be written as:

$$s_{\bar{x}} = \sqrt{\frac{s_e^2}{N}} \qquad (5.11)$$

where

$$s_e^2 = \frac{\sum_{j=1}^{n} \sum_{k=1}^{m_j} e_{jk}^2}{N-n} \qquad (5.12)$$

The estimate of s_e^2 above is the same as the mean square error when using one-way analysis of variance.

5.2.4 Calculating the Standard Error After Correcting for Serial Correlation

If the serial correlation of the seasonally adjusted residuals is significant (see Section 5.6), the following formula in Box 5.6 should be used to compute the standard error of the mean, $s_{\bar{x}}$.

```
┌─────────────────────────────────────────────────────────────────┐
│                                                                   │
│                          Box 5.6                                  │
│        Calculating the Standard Error After Removing Seasonal Averages │
│                                                                   │
│    The-standard error based on the residuals resulting from removing the │
│    seasonal averages is:                                          │
│                                                                   │
│                                                                   │
│                            √‾‾‾‾‾‾‾‾‾‾‾‾‾‾                         │
│                                                                   │
│                                                                   │
└─────────────────────────────────────────────────────────────────┘
```

$$s_x = \sqrt{\dfrac{\sum\limits_{i=2}^{N} (e_i - e_{i-1})^2}{2N(N-1)}} \qquad (5.13)$$

The degrees of freedom associated with the standard error given by formula (5.13) is approximately $Df = \dfrac{2(N-n)}{3}$. When using this formula, round the approximate degress of freedom down to the next smallest integer. This equation results from applying equation 5.6 to the residuals from equation 5.8.

5.3   Calculating Lag 1 Serial Correlation

The serial correlation (or autocorrelation) measures the correlation of observations separated in time. Consider the situation where the ground water concentrations are distributed around an average concentration, with no long-term trend or seasonal patterns. The ground water measurements will fluctuate around the mean due to historic fluctuations in the contamination events and the ground water flows and levels. Even though the measurements fluctuate around the mean in what may appear to be a random pattern, the measurements in ground water samples taken close in time (such as on successive days) will typically be more similar than measurements taken far apart in time (such as a year apart). Therefore measurements taken close together in time arc more highly correlated than measurements taken far apart in time. The extent to which successive measurements arc correlated if measured by the serial correlation. The presence of significant serial correlation affects the standard error of the mean.

If serial correlation is present in the data, statistical methods must be selected which will provide correct results when applied to correlated data Some of the statistical procedures described in Chapters 5, through 9 require the calculation of the serial correlation. In general, serial correlations need not be based on observations which immediately follow one another in time sequence ("lag 1" serial correlations). Serial correlations may be defined that are 2 time periods, 3 time periods, etc., apart. These are

referred to as "lag 2", "lag 3", or in general, "lag k" serial correlations. Serial correlations are discussed more fully in Gilbert (1987). page 38 or Box and Jenkins (1976). page 26. only "lag l" serial correlations will be considered in this document,

To calculate the serial correlation, first compute the seasonally adjusted residuals, e_{jk}; using the procedure described in Section 5.2.3. Order the e_{jk},'s chronologically and denote the ith time-ordered **residual**by e_i The serial correlation between the residuals can then be computed as shown in Box 5.7 (see Neter, Wasserman, and Kutner, 1985, page 456).

Box 5.7
Calculating the Correlation from the Residuals After Removing
Seasonal Averages

The sample estimate of the serial correlation of the residuals is:

$$\hat{\phi}_{obs} = \frac{\sum_{i=2}^{N} e_i e_{i-1}}{\sum_{i=1}^{N} e_i^2} \tag{5.14}$$

Where e_i, i = 1, 2, ...,N are the residuals after removing seasonal averages, in the time order in which the samples were collected.

The serial correlation between successive observations, computed from formula (5.14), depends on the time interval between collection of ground-water samples. For example, for quarterly data, $\hat{\phi}_{obs}$ represents the correlation between measurements that are taken three months apart, while, for monthly data, $\hat{\phi}_{obs}$ represents the correlation between measurements that are taken one month apart. Correlations between observations taken at different intervals will generally be different. For estimating sample sizes (Section 5.10) it will be convenient to work with the monthly serial correlation, i.e., the correlation between observations that are one month apart. If the data are not collected at monthly intervals, the formula in Box 5.8 can be used to convert $\hat{\phi}_{obs}$ to a monthly serial correlation $\hat{\phi}$ (see Box and Jenkins, 1970, for more details). Equation (5.15) estimates the monthly correlation from a correlation based on observations separated by t months. For example, for a sample correlation calculated from quarterly data, t = 3. Equation (5.15) is based on

assumptions about the factors which affect the correlations in the measurements. These assumptions become more important as the frequency at which the observations are collected differs from monthly (see Box and Jenkins, 1970, page 57 and Appendix D).

Box 5.8
Estimating the Serial Correlation Between Monthly Observations

The estimated serial correlation between monthly observations based on a sample estimate of the serial correlation between observations separated by t months is:

$$\hat{\phi} = \left(\hat{\phi}_{obs}\right)^{\frac{1}{t}} \qquad (5.15)$$

With data from multiple wells, the estimates of serial correlations can be combined across wells to provide a better estimate when the following conditions are met:

- The contaminant concentration levels in the wells are similar;

- The wells are sampled at the same frequency;

- The wells are sampled for roughly the same period of time; and

- The wells are geographically close.

Under these conditions, the combined estimate of serial correlation is calculated by averaging the estimates calculated for each well.

5.4 Statistical Inferences: What can be Concluded from Sample Data

The first two sections of this chapter dealt with the computation of several types of measures that can be used to characterize the sample data, means, standard errors, and serial correlation coefficients. In addition to characterizing or describing one's data with summary statistics, it is often desirable to draw conclusions from the data, such as an answer to the question: Is the mean concentration less than the cleanup standard?

A general approach to drawing conclusions from the data, also referred to as making inferences from the data, uses a standard structure and process for making such' decisions referred to as "hypothesis testing" in statistical literature. It can be outlined as follows,

1.   Make an assumption about the concentrations which you would like to disprove (e.g.. the average population measure of a contaminant is greater than the cleanup standard of 2.0 ppm). This cleanup standard represents your initial or null hypothesis about the current situation.

2.   Collect a set of data, representing a random sample from the population of interest.

3.   Construct a statistic from the sample data. Assuming that the null hypothesis is true, calculate the expected distribution of the statistic.

4.   If the value of the statistics is consistent with the null hypothesis, conclude that the null hypothesis provides an acceptable description of the present situation.

5.   If the value of the statistic is highly unlikely given the assumed null hypothesis, conclude that the null hypothesis is incorrect.

Of course, sample data may occasionally provide an estimate that is somewhat different from the true value of the population parameters being estimated. For example, the average value of the sample data could be, by chance, much higher than that of the full population. If the sample you happened to collect was substantially different from the population, you might draw the wrong conclusion. Specifically, you might conclude that the value assumed in the null hypothesis had changed when it really had not. This false conclusion would have been arrived at simply by chance, by the luck of randomly selecting a particular set of observations or data values. The probability of incorrectly rejecting the null hypothesis by chance can be controled in the hypothesis test.

If the chance of obtaining a value of a test statistic beyond a specified limit is, say, 5% if the null hypothesis is true, then if the sample value is beyond this limit you have substantial evidence that the null hypothesis is not true. Of course, 5% of the time when the null hypothesis is true a test statistic value will be beyond that specified limit. This probability of incorrectly rejecting the null hypothesis is generally denoted by the symbol a (alpha) in statistical literature. The person(s) making the decision specify the risk of making this type of error (often referred to as a Type I error in statistical literature) prior to analyzing the data. **If one wishes to be conservative, one might choose $\alpha=.01$, allowing**

up to a one percent chance of incorrectly rejecting the null hypothesis. With less concern **about this type of error, one might choose α=.1. A common choice is α=.05.**

Many of the test procedures presented below use confidence intervals. A confidence interval shows the range of values for the parameter of interest for which the test statistics discussed above would not result in the rejection of the null hypothesis.

5.5 The Construction and Interpretation of Confidence Intervals about Means

A confidence intends a range of values which will include the population parameter, such as the population mean, with a known probability or confidence. The confidence interval indicates how closely the mean of a sample drawn from a population approximates the true mean of the population. Any level of confidence can be specified for a confidence interval. For example, a 95 percent confidence interval constructed from sample data will cover the true mean 95 percent of the time. In general, a l00(l-a) percent confidence interval will cover the true mean 100(1-a) percent of the time. As indicated **above the value of α, the probability of a Type I error, must be decided upon and is usually** chosen to be small; e.g., 0.10, 0.05, or 0.01. The general form of a confidence interval for the mean is shown in Box 5.9.

Box 5.9
General Construction of Two-sided Confidence Intervals

A two-side confidence interval for a mean is generally of the form:

$$\bar{x} - t^* s_{\bar{x}} \text{ to } \bar{x} + t^* s_{\bar{x}} \tag{5.16}$$

In equation (5.16) the product t*s$_{\bar{x}}$ represents the distance (in terms of sample standard errors) on either side of the sample average that is likely to include the true population mean. One determines t from a table of the t-distribution giving the probability that the ratio of (a) the difference between the true mean and the sample mean to (b) the sample standard error of the mean exceeds a certain value. To determine t, you actually

need to determine two parameters: α, the probability of a Type I error, and Df, the number of degrees of freedom associated with the standard error. Thus, t is usually expressed as $t_{1-\alpha,Df}$ and the appropriate value of $t_{1-\alpha,Df}$ can be found from a table of the critical values of the t distribution using the row and column associated with the values of $1-\alpha$ and Df (see Appendix A).

Given below are the formulas for *one-* and *two-sided* confidence limits for a population mean (Boxes 5.10 and 5.11). Here, the population (or "true") mean is the conceptual average contamination over all possible ground-water samples taken during the specified time period. The one-sided confidence interval (establishing an acceptable limit on the range of possible values for the population mean on only one side of the sample mean) can be used to test whether the ground water in the well for the (short-term) period of is significantly less than the cleanup standard The two-sided version of the confidence interval can be used to characterize the ground-water contamination levels during the period of sampling.

Box 5.10
General Construction of One-sided Confidence Intervals

The upper one-sided confidence limit for the mean is given by:

$$\mu_{U\alpha} = \bar{x} + t_{1-\alpha,Df}\, s_{\bar{x}} \qquad (5.17)$$

Box 5.11
Construction of Two-sided Confidence Intervals

The corresponding two-sided confidence limits are given by:

$$\mu_{U\alpha/2} = \bar{x} + t_{1-\alpha/2,Df}\, s_{\bar{x}} \qquad (5.18)$$

and

$$\mu_{L\alpha/2} = \bar{x} - t_{1-\alpha/2,Df}\, s_{\bar{x}} \qquad (5.19)$$

In equations (5.17) to (5.19), $1-\alpha$ is the confidence level associated with the interval, $\bar{x}$ is the computed mean level of contamination; $s_{\bar{x}}$ is the corresponding standard

error computed from the appropriate formula in Section 5.2, and Df is the number of degrees of freedom associated with $s_{\bar{x}}$. The degrees of freedom (Df) associated with the standard error depend on the particular formula used. Table 5.2 summarizes the various standard error formulas, their corresponding degrees of freedom, and the conditions under which they should be used. The appropriate value of $t_{1-\alpha,Df}$ can be obtained from Appendix Table A.1. Note that for two-sided intervals, the t-value used is $t_{1-\alpha/2,Df}$ rather than $t_{1-\alpha,Df}$. This reflects a willingness to take the risk of making a Type 1 error for values at both extremes of the distribution instead of just one (using equation 5.17, one cannot make a Type I error at the lower extreme because one's decision about the status quo never changes for extreme low values). The range of values $\mu_{L\alpha/2}$ to $\mu_{U\alpha/2}$ determines 100(1-α) percent lower and upper confidence limits for the true (short-term) average concentration levels during the sampling period.

Table 5.2   Alternative formulas for the standard error of the mean

Formula	Df	When formula should be used
1. $s_{\bar{x}} = \dfrac{s}{\sqrt{N}}$	N-1	Data exhibit no seasonal patterns and no serial correlation (Section 5.2.1)
2. $s_{\bar{x}} = \sqrt{\dfrac{\sum_{i=2}^{N}(x_i - x_{i-1})^2}{2N(N-1)}}$	$\sim\dfrac{2N}{3}$	Data exhibit no seasonal patterns, but may be serially correlated (Section 5.2.2)
3. $s_{\bar{x}} = \sqrt{\dfrac{s_e^2}{N}}$	N-n	Data exhibit a seasonal pattern, but no serial correlation (Section 5.2.3)
4. $s_{\bar{x}} = \sqrt{\dfrac{\sum_{i=2}^{N}(e_i - e_{i-1})^2}{2N(N-1)}}$	$\sim\dfrac{2(N-n)}{3}$	Seasonally-adjusted residuals exhibit serial correlation (Section 5.2.4)

The upper one-sided confidence limit $\mu_{U\alpha}$, defined in equation (5.17) can be used to test whether the average contaminant levels for ground-water samples collected

over a specified period of time is less than the cleanup standard, Cs (see Box 5.12).
Although the rules indicated below can be used to monitor cleanup progress, they should
not be used to assess attainment of the cleanup standard. Procedures for assessing
attainment are given in Chapters 8 and 9.

Box 5.12
Comparing the Short Term Mean to the Cleanup Standard Using

For short-term means, the decision rule to be used to decide whether or not
the ground water is less than the cleanup standard is the following:

If $\mu_{U\alpha}$ < Cs, conclude that the short-term mean ground-water contaminant
concentration is less than the cleanup standard (i.e., μ < Cs).

If $\mu_{U\alpha}$ ≥ Cs, conclude that the short-term mean ground-water contaminant
concentration exceeds the cleanup standard (i.e., μ ≥ Cs)

5.6 Procedures for Testing for Significant Serial Correlation

Different statistical methods may be required if the data have significant
serial correlations. The serial correlation can be estimated using the procedures in Box 5.7.
The Durbin-Watson test and the approximate large sample test in sections 5.6.1 and 5.6.2
can be used to test if the observed serial correlation, $\hat{\phi}_{obs}$, is significantly different from
zero.

5.6.1 Durbin-Watson Test

The discussion here on determining the existence of serial correlation in the
data assumes the knowledge of confidence intervals and hypothesis testing. Sections 5.4
and 2.3.4 provide a discussion of these concepts, if the reader would like to review them.

If there is no serial correlation between observations, the expected value of
$\hat{\phi}_{obs}$ will be close to zero. However, the calculated value of $\hat{\phi}_{obs}$ is unlikely to be zero even
if the actual serial correlation is zero. The Durbin-Watson statistic can be used to test
whether the observed value of $\hat{\phi}_{obs}$ is significantly different from zero. To perform the test

(e.g., see Neter, Wasserman, and Kutner, 1985, page 450). compute the statistic D shown in Box 5.14.

Box 5.13
Example: Calculation of Confidence Intervals

Suppose that 47 monthly ground-water samples were collected over a period of slightly less than 4 years. The measurements for three of the samples were below the detection limit and were replaced in the analysis by the detection limit. Based on these data, the overall mean is .33. Since the data did not exhibit any seasonal patterns but was thought to be serially correlated, equation (5.6) was used to compute the standard error of the mean; i.e., $s_{\bar{x}} = .1025$. The degrees of freedom associated with the standard error is $2N/3 = 2(47)/3 = 31$. Hence, for a two-sided 99 percent confidence interval, $\alpha = 0.01$ and $t_{.005,31} = 2.75$ from Appendix Table A.1.

The required confidence interval for the mean goes from $\bar{x} - t_{1-\alpha/2,Df}\, s_{\bar{x}}$ to $\bar{x} + t_{1-\alpha/2,Df}\, s_{\bar{x}}$ i.e., from $[.33 - 2.75(.1025)]$ to $[.33 + 2.75(.1025)]$ or from .048 to .612 ppm.

For a one-sided 99 percent confidence interval, $\alpha = 0.01$ and $t_{.01,31} = 2.457$ from Appendix Table A.1. The corresponding one-sided confidence interval goes from zero to

$$\mu_{U\alpha} = \bar{x} + t_{1-\alpha,Df}\frac{s}{\sqrt{n}} = .33 + 2.457(.1025) = .58 \text{ ppm.}$$

Since the cleanup standard is $Cs = 0.5$ ppm it is concluded that for the period of observation, there is insufficient evidence to conclude with confidence that the true mean ground-water concentration is less than the cleanup standard. This is the case even though the sample mean happens to be less than the cleanup standard. There is enough variability in the data that a true mean greater than 0.5 ppm cannot be ruled out.

Box 5.14
Calculation of the Durbin-Watson Statistic

$$D = \frac{\sum_{i=2}^{N}(e_i - e_{i-1})^2}{\sum_{i=1}^{N} e_i^2}. \qquad (5.20)$$

If D < d_U, where d_U is the upper "critical" value for the test given in Appendix Table A-6 of the book by Neter, Wasserman, and Kutner, 1985 (pages 1086-1087), conclude that there is a significant serial correlation. If D ≥ d_U, conclude that there is no serial correlation[1]. The Durbin-Watson statistic D is standard output in many regression packages.

5.6.2   An Approximate Large-Sample Test

If N > 50, the following approximate test can be used in place of the Durbin-Watson test (e.g., see Abraham and Ledolter, 1983, page 63).

Box 5.15
Large Sample Confidence Interval for the Serial Correlation

Compute the lower and upper limits, ϕ_L and ϕ_U, defined by

$$\phi_L = \hat{\phi}_{obs} - 2\sqrt{\frac{1}{n}} \qquad (5.21)$$

and

$$\phi_U = \hat{\phi}_{obs} + 2\sqrt{\frac{1}{n}} \qquad (5.22)$$

If the interval from ϕ_L to ϕ_U does not contain the value 0, conclude that the serial correlation is significant. Otherwise, conclude that the serial correlation is not significant.

5.7   Procedures for Testing the Assumption of Normality

Many of the, procedures discussed in this manual assume that the sampling and measurement error follow a normal distribution. In particular, the assumption of normality is critical for the method of tolerance intervals described later in Section 5.8.

---

[1] The decision rule used here is somewhat different from the usual Durbin-Watson test described in most text books. For the applications given in this manual, the recommended decision rule results in deciding that autocorrelation exists unless there is strong evidence to the contrary. Also, the particular value of d_U to use depends on N and "p-1", where p is the number of parameters in the fitted model. See section 4.2.5 for an example

Thus, it will be important to ascertain whether the assumption of normality holds. Some methods for checking the normality assumption are discussed below.

5.7.1   Formal Tests for Normality

The statistical tests used for evaluating whether or not the data follow a specified distribution are called "goodness-of-fit tests."* The computational procedures necessary for performing the goodness-of-fit tests that work best with the normal distribution are beyond the scope of this guidance document. Instead, the user of this document should use one of the statistical packages that implements a goodness-of-fit test. SAS (the Statistical Analysis System) is one such statistical package. A good reference for these tests is the book on nonparametric statistics by Conover (1980). Chapter 6. There axe many different tests for evaluating normality (e.g. D'Agostino, 1970; Filliben, 1975; Mage, 1982; and Shapiro and Wilk, 1985). If a choice is available, the Shapiro-Wilk or the Kolmogorov-Smirnov test with the Lilliefors critical values is recommended.

5.7.2   Normal Probability Plots

A relatively simple way of checking the normality of the data or residuals (such as those obtained from Box 5.4) is to plot the data or residuals ordered by size against their expected values under normality. Their i^{th} expected value will be called EV_i. Such a plot is referred to as a "normal probability plot."

If there are no seasonal effects, the residual e_i, is simply defined to be the difference between the observed value and the sample mean, i.e.,

$$e_i = x_i - \bar{x}. \qquad (5.23)$$

If seasonal variability is present, the residuals should be calculated from formula (5.8). In either case, the ith ordered residual, $e_{(i)}$, for $i = 1, 2, ..., N$, is defined to be the ith smallest value of the e_i's (that is, $e_{(1)} \le e_{(2)} \le ... \le e_{(i)} \le ... \le e_{(N)}$), and its expected value is given approximately by (SAS 1985):

---

[2] These should not be confused with tests for assessing the fit of a regression model which are discussed later in Chapter 5.

$$EV_i = s_{res} \cdot z(\frac{i-.375}{N+.25}), \qquad (5.24)$$

where s_{res} is the **standard deviation of the residuals** and z(.) is given by formula (5.25) below. If formula (5.23) applies --i.e., no seasonal effects are in evidence-- and is used to compute the residuals, then $s_{res} = s$, where s is given by formula (5.3). If formula (5.8) applies--requiring an adjustment for seasonal effects--and is used to compute the residuals, then $s_{res} = s_e$, where s_e^2 is given by formula (5.12). The function z(a) is defined to be the upper 100a percentage point of the standard normal distribution and is approximated by (Joiner and Rosenblatt 1975):

$$z(a) = 4.91[a^{0.14} - (1-a)^{0.14}] \qquad (5.25)$$

Under normality, the plot of the ordered residuals, $e_{(i)}$, against EV_i should fall approximately along a straight line. An example of the use of normal probability plots is given in Section 6.X For more rigorous statistical procedures for testing normality, use the "goodness-of-fit" tests mentioned in Figure 6.17.

5.8 Procedures for Testing Percentiles Using Tolerance Intervals

This section describes a statistical technique for estimating and evaluating percentiles of **a** concentration distribution. The technique is based on tolerance intervals and is not recommended if there are seasonal or other systematic patterns in the data. Moreover, this procedure is relatively sensitive to the assumption that the data (or transformed data) follow a normal distribution. If it is suspected that a normal distribution does not adequately approximate the distribution of the data (even after transformation), tolerance intervals should not be used. Instead, the procedure described later in Section 5.9 should be used.

5.8.1 Calculating a Tolerance Interval

The Qth percentile of a distribution of concentration measurements is that concentration value, say X_Q, for which Q percent of the concentration measurements are less than X_Q and (100-Q) percent of the measurements arc greater than X_Q. For example,

if the value 3.2 represents the 25th percentile for a give population of data, 25% of the data fall below the value 3.2 and 75% are above it. Since the data represent a sample (rather than the population) of concentration values, it is not possible to determine the exact value of X_Q from the sample data. However, with normally distributed data, a $100(1-\alpha)$ percent confidence interval around the desired percentile can be easily computed.

Let $x_1, x_2, ..., x_N$ denote N concentration measurements collected during a specified period of time. As explained in Section 2.3.7, values that are recorded as below the detection limit should be assigned the minimum detectable value (DL). The sample mean, $\bar{x}$, and the sample standard deviation, s, should initially be computed using the basic formulas given in Section 5.1.

Given Q and α, the upper $100(1-\alpha)$ one-sided confidence limit for the true percentile, X_Q, is given by:

$$\hat{x}_Q = \bar{x} + ks \qquad (5.26)$$

where k is a constant that depends on n, a, and $P_0 = (100-Q)/100$. The appropriate values of k can be obtained from Appendix Table A.3. For values not shown in the table, see Guttman (1970).

5.8.2 Inference: Deciding if the True Percentile is Less than the Cleanup Standard

The upper confidence interval as computed from equation (5.26) can be used to test whether the true (unknown) Q^{th} percentile, X_Q, for a specified sampling period is less than a value, Cs. The decision rule to be used to test whether the true percentile is below Cs is:

If $\hat{x}_Q < Cs$, conclude that the Qth percentile of ground-water contaminant concentrations is less than the Cs (i.e., $X_Q < Cs$).

If $\hat{x}_Q \geq Cs$, conclude that the Qth percentile of ground water contaminant concentrations is not less than Cs and may be much greater than Cs.

Box 5.16
Tolerance Intervals: Testing for the 95th Percentile with Lognormal Data

Data for 20 ground-water samples were obtained to determine if the 95th percentile of the contaminant concentrations observed for a two-year period was below the cleanup standard of 100 ppm. A false positive rate of one percent ($\alpha = 0.01$) was specified for the test. The data appeared to follow a lognormal distribution. Therefore, the logarithms of the data (the transformed data) were assumed to have a normal distribution and were analyzed. In the following discussion, x refers to the original data and y refers to the transformed data. Because the log of the data was used, the upper confidence interval on the 95th percentile of the data was compared to the log of the cleanup standard [$\ln(100) = 4.605$].

For the transformed data, the sample mean (the average of the logarithms) is:

$$\bar{y} = \frac{72.372}{20} = 3.619$$

The standard deviation of the transformed observations, s, as calculated from equation (5.3) is 0.715.

For N = 20, $\alpha = .01$ and $P_0 = 5\%$, k = 2.808 (from Appendix Table A.3). Finally, $\hat{x}_{95}$ can be calculated using equation (5.26):

$$\hat{y}_{95} = 3.619 + 2.808(.715) = 5.627$$

Since 5.627 is greater than 4.605 (the cleanup standard in log units), it is concluded that the 95th true percentile may be greater than Cs.

5.9 Procedures for Testing Proportions

An alternative statistical procedure for testing percentiles is based on the proportion of water samples that have contaminant levels exceeding a specified value. As was the case for the method of Section 5.8, this method is not recommended if there are seasonal patterns in the data. If seasonal variability is present, consult a statistician. The equations presented in this section apply if the acceptable proportion of contaminated samples is less than 0.5 and large sample sizes are used.

To apply this test, each sample ground-water measurement should be coded as either equal to or above the cleanup standard, Cs, (coded as " 1") or below Cs (coded as "0"). The statistical analysis is based on the resulting coded data set of O's and 1's. This

test can be applied to any concentration distribution (unlike the method of tolerance intervals which applies only to normally distributed data) and requires only that the cleanup Standard be greater than the detection limit.

Let $x_1, x_2, ..., x_N$ denote N concentration measurements collected during a specified period of time. Corresponding to each measurement x_i, define a coded value $y_i = 1$ if x_i is greater than the cleanup standard and $y_i = 0$, otherwise. The proportion of samples, p, above the cleanup standard can be calculated using the following equations:

$$r = \sum_{i=1}^{N} y_i \qquad (5.27)$$

$$p = \frac{\sum_{i=1}^{N} y_i}{N} = \frac{r}{N} \qquad (5.28)$$

Assuming that the observations are independent, the standard error of the proportion, s_p, is given by:

$$s_p = \sqrt{\frac{p(1-p)}{N}} \qquad (5.29)$$

Formula (5.29) will tend to over estimate the variance if the data have a significant serial correlation. If the data have significant serial correlations, we can use formula (5.6) with the x's replaced by the y's. Note that formulas (5.29) and (5.6) should only be used if N is large; i.e., if $N \geq 10/p$ and $N \geq 10/(1-p)$.

5.9.1 Calculating Confidence Intervals for Proportions

For sufficiently large sample sizes (i.e., $N \geq 10/p$ and $N \geq 10/(1-p)$, i.e. at least 10 samples with measurements above the cleanup standard and 10 with measurements below the cleanup standard), an approximate confidence interval may be constructed using the normal approximation. If there is concern about the sample size N being too small relative to p, a statistician should be consulted.

For large sample sizes, the one-sided $100(1-\alpha)$ percent upper confidence limit is given by:

$$P_{U\alpha} = p + z_{1-\alpha} s_p \qquad (5.30)$$

where p is the proportion of ground-water samples that have concentrations exceeding Cs, and $z_{1-\alpha}$ is the appropriate critical value obtained from the normal distribution (see Appendix Table A.2).

The corresponding two-sided $100(1-\alpha)$ percent confidence limits are given by:

$$P_{U\alpha/2} = p + z_{1-\alpha/2} s_p \qquad (5.31)$$

and

$$P_{L\alpha/2} = p - z_{1-\alpha/2} s_p \qquad (5.32)$$

where $z_{1-\alpha/2}$ is the appropriate critical value obtained from the normal distribution (see Table A.2). The range of values from $P_{L\alpha/2}$ to $P_{U\alpha/2}$ represents a $100(1-\alpha)$ percent confidence interval for the corresponding population proportion.

5.9.2   Inference: Deciding Whether the Observed Proportion Meets the Cleanup Standard

The upper confidence limit as computed from equation (5.30) can be used to test whether the true (unknown) proportion, P, is less than a specified standard, P_0. The decision rule to be used to test whether the true proportion is below P_O is:

If P_{ua}, $< P_O$, conclude that the proportion of ground-water samples with contaminant concentrations exceeding Cs is less than P_O.

If $P_{U\alpha} \geq P_0$, conclude that the proportion of ground-water samples with contaminant concentrations exceeding Cs may be greater than or equal to P_0.

Box 5.17
Calculation of Confidence Intervals

For 184 ground-water samples collected during an 8-year period, 11 samples had concentrations greater than or equal to the cleanup standard.

The proportion of contamination samples is (equation 5.27):

$$p = \frac{r}{N} = \frac{11}{184} = .0598$$

A one-sided confidence interval has an upper limit of (from equations 5.30):

$$p + z_{1-\alpha} \, s_{p'}$$

Assuming $\alpha = 0.05$ (i.e., 95 percent confidence), $z_{1-\alpha} = 1.645$. The standard error of p determined from formula (5.29) is $s_p = 0.0175$.

The confidence interval is thus .0000 to .0598 + .0288 or .0000 to .0886.

5.9.3 Nonparametric Confidence Intervals Around a Median

An alternate approach to testing proportions is to test percentiles. For example, the following two approaches are equivalent: (a) testing to see if less than 50% of the samples have contamination greater than the cleanup standard and (b) testing to see if the median concentration is less than the cleanup standard. The method presented in this section for testing the median can be extended to testing other percentiles, however, the calculations can be cumbersome. If you wish to test percentiles rather than proportions, or to test the median using other confidence intervals than are presented here, consultation with a statistician is recommended

If the data do not adequately follow the normal distribution even after transformation, a nonparametric confidence interval around the median can be constructed. The median concentration equals the mean if the distribution is symmetric (see Section 2.5). The nonparametric confidence interval for the median is generally wider and requires more data than the corresponding confidence interval for the mean based on the normal distribution. Therefore, the normal or log-normal distribution interval should be used whenever it is appropriate.

CHAPTER 5: DESCRIPTIVE STATISTICS

The nonparametric confidence interval for the median requires a minimum of seven (7) observations in order to construct a 98 percent two-sided confidence interval, or a 99 percent one-sided confidence interval. Consequently, it is applicable only for the pooled concentration of compliance wells at a single point in time a for sampling to produce a minimum of seven observations at a single well during the sampling period.

The procedures below for construction of a nonparametric confidence interval for the median concentration follow (U.S. EPA, 1989b). An example is presented in Box 5.19.

(1) **Within each well or group of wells, order the N data from least to greatest, denoting the ordered data by x_1, x_2,...x_N, where x_i is the ith value in the ordered data. Ties do not affect the procedure. If there are ties, order the observations as before, including all of the tied values as separate observations. That is, each of the observations with a common value is included in the ordered list (e.g., 1, 2, 2, 2, 3, 4, etc.). For ties, use the average of the tied ranks.**

(2) Determine the critical values of the order statistics as follows. If the minimum seven observations is used, the critical values arc 1 and 7. Otherwise, find the smallest integer, M, such that the cumulative binomial. distribution with parameters N (the sample size) and p = 0.5 is at least 0.99. Table 5.3 gives the values of M and N+l-M together with the exact confidence coefficient far sample sizes from 4 to 11. For larger samples, use the equation in Box 5.18.

(3) Once M has been determined, find N+l-M and take as the confidence limits the order statistics X_M and $x_{n+1-M.}$ (With the minimum seven observations, use x_1 and x_7.)

(4) Inference: Deciding whether the site meets the cleanup standards.

After calculating the upper one-sided nonparametric confidence limit X_M from (3). use the following rule to decide whether the ground water attains the cleanup standard:

If $x_M < Cs$, conclude the median ground water concentration in the wells during the sampling period is less than the cleanup standard.

If $x_M \geq Cs$, conclude the median ground water concentration in the wells during the sampling period is not less than the cleanup standard.

Table 5.3   **Values** of M and N+l-M and confidence coefficients for small samples

N	M	N+l-M	Two-sided confidence
4	4	1	87.5%
5	5	1	93.8%
6	6	1	96.9%
7	7	1	98.4%
8	8	1	99.2%
9	9	1	99.6%
10	9	2	97.9%
11	10	2	98.8%

Box 5.18
Calculation of M

$$M = \frac{N}{2} + 1 + z_{0.99}\sqrt{\frac{N}{4}} \qquad (5.33)$$

where $z_{0.99}$ is the 99th percentile from the normal distribution and equals 2.33. (From Table A.2 in Appendix A)

Concentration (ppm)	Concentration (ppm)

Table 5.4   Example contamination data used in Box 5.19 to generate nonparametric confidence interval

Sampling Date	Well 1 Concentration (ppm)	Rank	Well 2 Concentration (ppm)	Rank
Jan. 1	3.17	(2)	3.52	(6)
	2.32	(1)	12.32	(15)
	7.37	(11)	2.28	(4)
	4.44	(6)	5.30	(7)
April 1	9.50	(13)	8.12	(11)
	21.36	(16)	3.36	(5)
	5.15	(7)	11.02	(14)
	15.70	(15)	35.05	(16)
July 1	5.58	(8)	2.20	(3)
	3.39	(3)	0.00	(1.5)
	8.44	(12)	9.30	(12)
	10.25	(14)	10.30	(13)
Oct. 1	3.65	(4)	5.93	(8)
	6.15	(9)	6.39	(9)
	6.94	(10)	0.00	(1.5)
	3.74	(5)	6.53	(10)

5.10   Determining Sample Size for Short-Term Analysis and Other Data Collection Issues

The discussion in Chapter 4 assumes that the number of ground-water samples to be analyzed has been previously specified. In general, determination of the number of samples to be collected for analysis must be done before collection of the samples. The appropriate sample size for a particular application will depend upon the desired level of precision, as well as on assumptions about the underlying distribution of the measurements. Given below arc some guidelines for determining sample size for estimating means, percentiles and proportions for short term analyses. When assessing whether remediation has indeed been successful, use the procedures discussed in chapters 8 and 9 to determine the required sample size. Some discussion of various data collection issues is also offered hue.

Box 5.19
Example of Constructing Nonparametric Confidence Intervals

Table 5.4 contains concentrations of a contaminant in parts per million from two hypothetical wells. The data are are assumed to consist of 4 samples taken each quarter for a year, so that 16 observations are available from each well. The data are not normally distributed, neither as raw data nor when log-transformed Thus, the nonparametric confidence interval is used. The Cs is 25 ppm

(1)   The l6 measurements are ordered from the least to greatest within each well separately. The numbers in parentheses beside each concentration in Table 5.4 arc the ranks or order of the observation. For example, in Well 1, the **smallest** observation is 2.32, which has rank 1. The second smallest is 3.17, which has rank 2, and so forth, with the largest observation of 21.36 having rank 16.

(2)   The sample size is large enough so that the approximation (equation 5.33) is used to find M:

$$M = \frac{16}{2} + 1 + 2.33 \sqrt{\frac{16}{4}} = 13.7 \cong 14$$

(3)   The approximate 95 percent confidence limits are given by the N + 1 - M observation (16 + 1 - 14 = 3rd) and the Mth largest observation (14th). For Well 1, the 3rd observation is 3.39 and the 14th observation is 10.25. Thus the confidence limits for Well 1 are (3.99, 10.25). Similarly for Well 2, the 3rd observation and the 14th observation arc found to give the confidence interval (2.20, 11.02). Note that for Well 2 there were two values below the detection. These were assigned a value equal to the detection limit and received the two smallest ranks. Had there been three or more values below the detection, the lower limit of the confidence interval would have been the limit of detection because these values would have been the smallest values and so would have included the third order statistic.

(4)   Neither of the two confidence intervals' upper limit exceeds the cleanup standard of 25 ppm. Therefore, the short-term median ground water concentrations arc less than the cleanup standard.

5.10.1   Sample Sizes for Estimating a Mean

In order to determine the sample size for estimating a mean, some information about the standard deviation, σ, (or equivalently, the variance σ^2) of the

measurements of each contaminant is required. This parameter represents the underlying variability of the conceptual population of contaminant measurements. The symbol "^" is used to denote that $\hat{\sigma}$ is an estimate of σ. In practice, $\hat{\sigma}$ is either obtained from prior data or by conducting a small preliminary investigation. Cochran (1977), pages 78-81, discusses various approaches to determining a preliminary value for $\hat{\sigma}$. Some procedures that are useful in ground-water studies are outlined below.

Use of Data from a Comparable Period

The value $\hat{\sigma}$ may be calculated from existing data which is comparable to the data expected from the sampling effort. Comparable data will have a similar level of contamination and be collected under similar conditions. For calculating the sample size required for assessing attainment, one may be able to use data on contamination levels for the wells under investigation from ground-water samples collected during the period in which steady state is being established. Using the comparable data, the value $\hat{\sigma}$ may be calculated using formula (5.3).

Use of Data Collected Prior to Remedial Action

If data from samples collected prior to remediation are available. the variability of these sample measurements can be used to obtain a rough estimate of σ using the coefficient of variation. The coefficient of variation is defined to be the standard deviation divided by the mean. Remediation will usually result in a lowering of both the mean and the standard deviation of contamination levels. In this case, it might be reasonable to expect the coefficient of variation to remain approximately constant. In this case, estimates of the coefficient of variation from the available data can be used to obtain $\hat{\sigma}$ as follows.

Using this data, let ($\bar{x}$) and s represent the sample mean and sample standard deviation for data collected prior to remedial action, perhaps from a previous study. Calculate ($\bar{x}$) and (s) using the equations in Section 5.1. An estimate $\hat{\sigma}$ of the standard deviation when clean up standards are attained can be computed using the cleanup standard, Cs, where

$$\hat{\sigma} = \frac{Cs \cdot s}{\bar{x}}, \qquad\qquad (5.35)$$

Conducting a Preliminary Study After Remedial Action

The following approach can be used if there are no existing data on contamination levels from which to estimate σ and if there is time to collect preliminary data before sampling begins.

(1)   After achieving steady state conditions (see Chapter 7), collect a preliminary sample of at least $n_1 = 8$ ground-water samples over a minimum period of 2 years. Determine the contamination levels for these samples. The larger the sample size and the longer the period of time over which the samples are collected, the more reliable the estimate of σ. A minimum of four samples per year is recommended so that seasonal variation will be reflected in the estimate.

(2)   From this preliminary sample, compute the estimated standard deviation, s, of the contaminant levels. Use this standard deviation as an estimate of σ.

Box 5.20

Estimating σ from Data Collected Prior to Remedial Action

Suppose that the number of ground-water samples to be taken from a monitoring well prior to remedial action was limited to 10. The concentrations of total PAH'S from the samples axe:

0.24, 2.93, 3.09, 0.14, 0.60, 4.20, 3.81, 2.31, 1.11, and 0.07

Using equations (5.1) and (5.3), the mean concentration is $\bar{x} = 1.85$ ppm and the standard deviation of the measurements is s = 1.60 ppm.

With a cleanup standard of .5 ppm, the value of $\hat{\sigma}$ to use for determining sample size can be obtained from:

$$\hat{\sigma} = \frac{Cs \cdot s}{\bar{x}} = \frac{.5 \cdot 1.60}{1.85} = .43 \text{ ppm}$$

CHAPTER 5: DESCRIPTIVE STATISTICS

A Rough Approximation of the Standard Deviation

If there are no existing data to estimate σ and a preliminary study is not feasible, a very rough approximation for $\hat{\sigma}$ can still be obtained. The approximation is rough because it is based on speculation and judgments concerning the range within which the ground-water measurements are likely to fall. Because the approximation is based on very little data, it is possible that the sample sizes computed from these approximations will be too small to achieve the specified level of precision. Consequently, this method should only be used if no other alternative is available.

The approximation is based on the fact that the range of possible ground-water measurements (i.e., the largest such value minus the smallest such value) provides a measure of the underlying variability of the data. Moreover, if the frequency distribution of the ground-water measurementsof interest is approximately bell-shaped, then virtually all of the measurements can be expected to lie within three standard deviations of the mean. In this case, if R represents the expected range of the data, an estimate of σ is given by

$$\hat{\sigma} = \frac{R}{6}. \tag{5.36}$$

If the data are not bell-shaped, the alternative (conservative) estimate $\hat{\sigma} = $ R/5 should be used.

Formula for Determining Sample Size for Estimating a Mean

The equations for determining sample size require the specification of the following quantities: Cs, μ_1, α, β, $\hat{\sigma}$. Given these quantities, the required sample size can be computed from the following formula (e.g., see Neter, Wasserman, and Whitmore, 1982, page 264 and Appendix F):

$$n = \hat{\sigma}^2 \left\{ \frac{z_{1-\beta} + z_{1-\alpha}}{Cs - \mu_1} \right\}^2 + 2 \tag{5.37}$$

where $z_{1-\beta}$ and $z_{1-\alpha}$ are the critical values for the normal distribution with probabilities of $1-\alpha$ and $1-\beta$ (Table A.2) and the factor of 2 is empirically derived in Appendix F.

5-37

Strictly speaking, formula (5.37) applies to simple random sampling. However, the standard error of a mean based on a systematic sample will usually be less than or equal to the standard error of a mean based on a simple random sample of the same size. Therefore, using the sample size formula given above may provide greater precision than is required.

Box 5.21
Example of Sample Size Calculations

Following the example in Box 5.20, suppose that it is desired to be able to detect a difference of .2 ppm from the cleanup standard of .5 ppm ($Cs = .5$, $\mu_1 = .3$) with a power of .80 (i.e., $\beta = .20$). Also suppose that $\hat{\sigma} = .43$ and $\alpha = .01$.

From tables of the cumulative normal distribution (Appendix Table A.2), we find that $z_{1-\alpha} = 2.326$ and $z_{1-\beta} = 0.842$. Then using formula (5.37)

$$n = \frac{(.43)^2 \, (2.236 + .842)^2}{(.5 - .3)^2} + 2 = 45.8$$

Rounding up, the sample size is 46.

5.10.2   Sample Sizes for Estimating a Percentile Using Tolerance Intervals

To determine the required sample size for tests based on the procedure described in Section 5.8, the following terms need to be defined: P_0, P_1, α, β (e.g., see Volume 1, Section 7.6). Once these terms have been established, the following quantities should be obtained   from Appendix Table A.2:

$z_{1-\beta}$,   the upper β-percentage point of a normal distribution;

$z_{1-\alpha}$,   the upper α-percentage point of a normal distribution;

z_{1-P_0},   the upper P_0-percentage point of a normal distribution; and

z_{1-P_1},   the upper P_1-percentage point of a normal distribution.

The sample size necessary to meet the stated objectives is then (see Guttman, 1970):

$$n = \left\{ \frac{z_{1-\beta} + z_{1-\alpha}}{z_{1-P_0} - z_{1-P_1}} \right\}^2 \tag{5.38}$$

Box 5.22
Calculating Sample Size for Tolerance Intervals

PCB's have contaminated the ground water near a former industrial site. The site managers have decided to use the procedures of Section 5.8 to help decide if the treatment can be terminated. Specifically, after discussion with ground-water experts, they decide to conclude that the treatment can be terminated if the 99th percentile of the PCB concentrations is less than Cs. That is, in the notation of Section 5.8, $P_0 = 1-.99 = .01$. They have also decided to set the false positive rate of the test to $\alpha = .05$. Moreover, they have required the false negative rate to be no more than 20 percent ($\beta = 0.20$) when the actual proportion of contaminated samples is 0.5 percent ($P_1 = .005$).

From Appendix Table A.2, $z_{1-P_0} = z_{.99} = 2.326$; $z_{1-P_1} = z_{.995} = 2.576$; $z_{1-\alpha} = z_{.95} = 1.645$; and $z_{1-\beta} = z_{.80} = 0.842$. Using formula (5.38), the required sample size for each well is:

$$n = \left\{ \frac{.842 + 1.645}{2.326 - 2.576} \right\}^2 = \left\{ \frac{2.487}{-.250} \right\}^2 = 98.96 \approx 99$$

where $z_{1-\beta}$ and $z_{1-\alpha}$ are critical values from the normal distribution associated with probabilities of $1-\alpha$ and $1-\beta$ (Appendix Table A.2).

5.10.3 Sample Sizes for Estimating Proportions

The sample size required for estimating a proportion using the procedures of Section 5.9 depend on the following quantities: P_0, P_1, α, and β. Given these quantities, the sample size can be computed from the following formula (e.g., see Neter, Wasserman, and Whitmore, 1982, page 304):

$$n = \left\{ \frac{z_{1-\beta}\sqrt{P_1(1-P_1)} + z_{1-\alpha}\sqrt{P_0(1-P_0)}}{P_0 - P_1} \right\}^2 \qquad (5.39)$$

Box 5.23
Sample size Determination Estimating Proportions

At a site with corrosive residues in the top soil, much of the contaminated top soil has been removed. However, it is known that the contaminants have leached into the ground water. Wanting to minimize the possibility of future health effects, the site manager would like to know if, in the short term, she can be 95 percent confident ($\alpha = .05$) that less than 10 percent ($P_0 = .10$) of the ground-water samples have concentrations exceeding the cleanup standard. The expected proportion of contaminated ground-water samples is very low, less than 5 percent. The manager wants to be 80 percent confident ($\beta = 1-.80 = .20$) that the ground water will be declared clean if the proportion of contaminated ground water samples is less than 5 percent ($P_1 = .05$).

Using formula (5.39).

$$n = \left\{ \frac{z_{1-\beta}\sqrt{P_1(1-P_1)} + z_{1-\alpha}\sqrt{P_0(1-P_0)}}{P_0 - P_1} \right\}^2$$

$$= \left\{ \frac{.842\sqrt{.05(.95)} + 1.645\sqrt{.10(.90)}}{.10-.05} \right\}^2$$

$$= 183.3$$

Rounding up gives a final sample size of 184.

5.10.4   Collecting the Data

After the sample size and sampling frequency have been specified, collection of the ground-water samples can begin. In collecting the samples, it is important to maintain strict quality control standard and to fully document the sampling procedures. Occasionally, a sample will be lost in the field or the lab. If this happens, it is best to try to collect another sample to replace the missing observation before teaching the next sampling period. Any changes in the sampling protocol should be fully documented.

Data resulting from a sampling program can only be evaluated and interpreted with confidence when adequate quality assurance methods and procedures have been incorporated into the program design. An adequate quality assurance program requires awareness of the sources of error or variation associated with each step of the sampling effort.

If a timely and representative sample of proper size and content is not delivered to the analytical lab, the analysis cannot be expected to give meaningful results. Failing to build in a quality assurance program often results in considerable money spent on sampling and analysis only to find that the samples were not collected in a manner that allows valid conclusions to be drawn from the resulting data. Seen in its broadest sense, the QA program should address the sample design selected, the quality of the ground-water samples, and the care and skill spent on the preparation and testing of the samples.

The samples should reflect what is actually present in the ground water. Improper or careless collection of the samples can likely influence the magnitude of the sample collection error. Sample preparation also introduces quality control issues.

While a full discussion of these topics is beyond the scope of this document, the implementation of an adequate QA program is important

5.10.5 Making Adjustments for Values Below the Detection Limit

Sometimes the reported concentration for a ground-water sample will be below the detection limit (DL) for the sampling and analytical procedure used. The rules outlined in Section 2.3.7 should be used to handle such measurements in the statistical analysis.

5.11 Summary

This chapter introduces the reader to some basic statistical procedures that can be used to both describe (or characterize) a set of data, and to test hypotheses and make inferences from the data. The chapter discusses the calculation of means and proportions. Hypothesis tests-and confidence intervals are discussed for making inferences from the data The statistics and inferential procedures presented in this chapter are appropriate &

for estimating short-term characteristics of contaminant levels. By "short-term characteristics" we mean characteristics such as the mean or percentile of contaminant concentrations during the <u>fixed</u> of time during which sampling occurs. Procedures far estimating the long-term mean and far assessing attainment are discussed in Chapters 8 and 9. The procedures discussed in this chapter can be used in any phase of the remedial effort; however, they will be useful during treatment.

This chapter provided procedures for estimating the sample sizes required for assessing the status of the cleanup effort <u>prior to a final assessment</u> of whether the remediation effort has been successful. It also discussed briefly issues involved in data collection.

6. DECIDING TO TERMINATE TREATMENT USING REGRESSION ANALYSIS

The decision to stop treatment is based on many sources of information including (1) expert knowledge of the ground water system at the site; (2) mathematical modeling of how treatment affects ground water flows and contamination levels; and (3) statistical results from the monitoring wells from which levels of contamination can be model and extrapolated. This chapter is concerned with the third source of information. In particular, it describes how one statistical technique, known as regression analysis, can be used in conjunction with other sources of information to decide when to terminate treatment. The methods given here are applicable to analyzing data from the treatment period indicated by the unshaded portion of Figure 6.1. Methods other than regression analysis, such as time series analysis (Box and Jenkins, 1970) can also be used. However, these methods are usually computer intensive and require the assistance of a statistician familiar with these methods,

Figure 6.1   Example Scenario for Contaminant Measurements During Successful Remedial Action
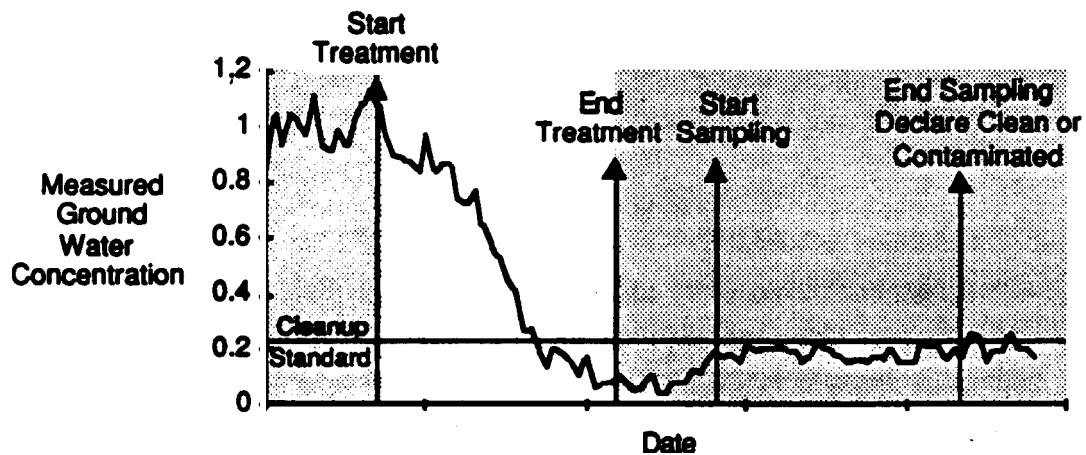
Section 6.1 provides a brief overview of regression analysis and serves as a review of the basic concepts for those readers who have had some previous exposure to the subject. Section 6.2, the major focus of the chapter, provides a discussion of the steps required to implement a regression analysis of ground water remediation data Section 6.3

briefly outlines important considerations in combining statistical and nonstatistical information.

6 . 1   Introduction to Regression Analysis

Regression analysis is a statistical technique far fitting a theoretical curve to a set of sample data For example, as a result of site clean-up, it is expected that contamination levels will decrease over time. Regression analysis provides a method for modeling (i.e., describing) the rate of this decrease. In ground-water monitoring studies, regression techniques can be used to (1) detect trends in contaminant concentration levels over time, (2) determine variables that influence concentration levels, and (3) predict chemical concentrations at future points in time. An example of a situation where a regression analysis might be useful is given in Figure 6.2 which shows a plot of chemical concentrations for 15 monthly samples taken from a hypothetical monitoring well during the period of treatment. As seen from the plot, there is a distinct downward trend in the observed chemical concentrations as a function of time. Moreover, aside from some "random" fluctuation, it appears that the functional relationship between contaminant levels and time can be reasonably approximated by a straight line for the time interval shown. This mathematical relationship is referred to as the regression "curve" or regression model. The goal of a regression analysis is to estimate the underlying functional relationship (i.e., the model), assess the fit of the model, and, if appropriate, use the model to make predictions about future .-observations.

In general, the underlying regression model need not be linear. However, to fix ideas, it is useful to introduce regression methods in the context of the **simple linear regression** model of which the linear relationship in Figure 6.2 is an example. Underlying assumptions, required notation, and the basic framework for simple linear regression analysis are provided in Section 6.1.1. Section 6.1.2 gives the formulas required to fit the regression model. Section 6.1.3 discusses how to evaluate the fit of the regression model using the residuals. Section 6.1.4 discusses how some important regression statistics can be used for inferential purposes (i.e., forming statistically defensible conclusions form the data).

Figure 6.2   Example of a Linear Relationship Between Chemical Concentration Measurements and Time
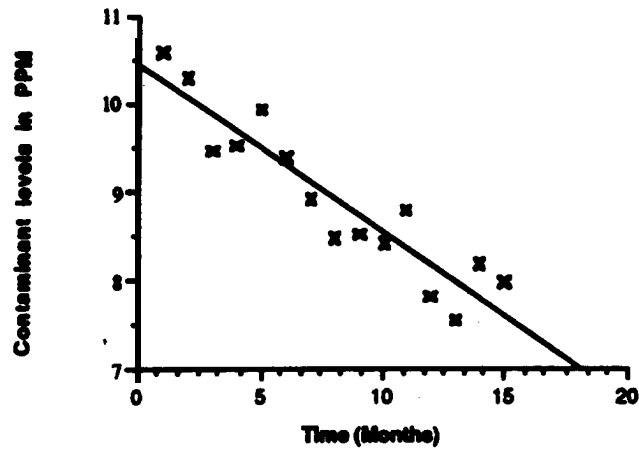
6.1.1   Definitions, Notation, and Assumptions

Assume that a total of N ground water samples have been taken from a monitoring well over a period of time for chemical measurement. Denote the sample collection time for i^{th} sample as t_i and the chemical concentration measurement in the ith sample as c_i where $i = 1, 2, \ldots, N$. Let y_i denote some function of the ith observed **concentration, for example, the identity function, $y_i = c_i$, the square root, $y_i = \sqrt{c_i}$, or the log transformation; $y_i = 1n(C_i)$. Let x_i denote time or** a function **of the time, for example, if** the "time" variable is the original collection time, $x_i = t_i$, if the time variable is the reciprocal of the collection time then $x_i = 1/t_i$, etc. If the samples are collected at regular time intervals, then the time index, i, can be used to measure time in place of the actual collection time, i.e., $x_i = i$ or $x_i = 1/i$ in the examples above.   Note that the notation used in this section is different from that introduced in Chapter 5.

The simple linear regression model relating the concentration measurements to time is defined by equation (6.1) in Box 6.1.

Box 6.1
Simple Linear Regression Model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, ..., N \qquad (6.1)$$

In equation (6.1), β_0 and β_1 are constants referred to as the **regression coefficients**, or alternatively as the **parameters of the model**, and ε_i is a random error. The term "y_i" is often referred to as the dependent, response, or outcome variable. In this document, the outcome variables of interest are contamination levels or related measures. The term "x_i" is also referred to as an independent or explanatory variable. The independent variable (for example the collection time) is generally under the control of the experimenter. The term N represents the number of observations or measurements on which the regression model is based.

The regression coefficients are unknown but can be estimated from the observed data under the assumption that the underlying model is correct. The non random part of the regression model is the formula for a straight line with **y-intercept** equal to β_0 and **slope** equal to β_1. In most regression applications, primary interest centers on the slope parameter. For example, if $x_i = i$ and the slope is negative, then the model states that the chemical concentrations decrease linearly with time, and the value of β_1 gives the rate at which the chemical concentrations decrease.

The random error, ε_i, represents "random" fluctuations of the observed chemical measurements around the hypothesized regression line, $y_i = \beta_0 + \beta_1 x_i$. It reflects the sources of variability not accounted for by the model, e.g., sources of variability due to unassignable or unmeasurable causes. Regression analysis imposes the following assumptions on the errors:

(i) The ε_i's are independent;

(ii) The ε_i's have mean 0 for all values of x_i;

(iii) The ε_i's have constant variance, σ^2, for all values of x_i; and

(iv) The ε_i's are normally distributed.

These assumptions are critical for the Validity Of the statistical tests used in a regression analysis. If they do not hold, steps must be taken to accommodate any departures from the underlying assumptions. Section 6.2.3 describes some simple graphical procedures which can be used to study the aptness of the Underlying assumptions and also indicates some corrective measures when the above assumptions do not hold

Interested readers should refer to Draper and Smith (1966) or Neter, Wasserman, and Kutner (1985) for more details on the theoretical aspects of regression analysis.

6.1.2   Computational Formulas for Simple Linear Regression

The computational formulas for most of the important quantities needed in a simple linear regression analysis are summarized below. These formulas are given primarily for completeness, but have been written in sufficient detail so that they can be used by persons wishing to carry out a simple regression analysis without the aid of a computer, spreadsheet, or scientific calculator. Readers who do not need to know the computational details in a regression analysis should skip this section and go directly to Sections 6.1.3 and 6.1.4, where specific procedures for assessing the fit of the model and making inferences based on regression model arc discussed

Estimates of the slope, β_1, and intercept, β_0, of the regression line are given by the values b_1 and b_0 in equations (6.2) and (6.3) in Box 6.2. 'The statistics b_1 and b_0 arc referred to **as least squares estimates.** If the four critical assumptions given in Section 6.1.1. hold for the simple linear regression model in Box 6.1, b_1 and b_0 will be unbiased **estimates of β_1 and β_0, and the precision of the estimates can be determined.**

The estimated regression line (or, more generally, the **fitted curve)** under the model is represented by equation (6.4) in Box 6.3.

Box 6.2
Calculating Least Square Estimates

$$b_1 = \frac{\sum\limits_{i=1}^{N} x_i y_i - \dfrac{\sum\limits_{i=1}^{N} x_i \sum\limits_{i=1}^{N} y_i}{N}}{\sum\limits_{i=1}^{N} x_i^2 - \dfrac{\left(\sum\limits_{i=1}^{N} x_i\right)^2}{N}} = \frac{\sum\limits_{i=1}^{N} x_i y_i - N\bar{x}\,\bar{y}}{\sum\limits_{i=1}^{N} x_i^2 - N\bar{x}^2} \qquad (6.2)$$

$$b_0 = \frac{\sum\limits_{i=1}^{N} y_i}{N} - b_1 \frac{\sum\limits_{i=1}^{N} x_i}{N} = \bar{y} - b_1 \bar{x} \qquad (6.3)$$

Box 6.3
Estimated Regression Line

$$\hat{y}_i = b_0 + b_1 x_i \qquad (6.4)$$

The calculated value of $\hat{y}_i$ is called the **predicted value** under the model corresponding to the value of the independent variable, x_i. The difference between the predicted value, $\hat{y}_i$, and the observed value, y_i, is called the residual. The equation for calculating the residuals is shown in Box 6.4. If the model provides a good prediction of the data, we would expect the predicted values, $\hat{y}_i$, to be close to the observed values, y_i. Thus, the sum of the squared differences $(y_i - \hat{y}_i)^2$ provides a measure of how well the model fits the data and is a basic quantity necessary for assessing the model.

Box 6.4
Calculation of Residuals

$$e_i = y_i - \hat{y}_i \qquad\qquad (6.5)$$

Formally, we define the **sum of squares due to error (SSE)** and the corresponding mean square error **(MSE)** by formulas (6.5) and (6.6). respectively, in **Box 6.5.**

Box 6.5
Sum of Squares Due to Error and the Mean Square Error

$$SSE = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \qquad\qquad (6.6)$$

$$MSE = \frac{SSE}{N-2}. \qquad\qquad (6.7)$$

As seen in the formulas in Box 6.2, the analysis of a simple linear regression model requires the computation of certain sums and sums of cross products of the observed data values. Therefore, it is convenient to define the five basic regression quantitics in Box 6.6.

The estimated model parameters and SSE can be computed from these terms using the formulas in Box 6.7.

Box 6.6
Five Basic Quantities for Use in Simple Linear Regression Analysis

$$S_x = \sum_{i=1}^{N} x_i \qquad (6.8)$$

$$S_y = \sum_{i=1}^{N} y_i \qquad (6.9)$$

$$S_{xx} = \sum_{i=1}^{N} x_i^2 - \frac{S_x^2}{N} \qquad (6.10)$$

$$S_{yy} = \sum_{i=1}^{N} y_i^2 - \frac{S_y^2}{N} \qquad (6.11)$$

$$S_{yx} = \sum_{i=1}^{N} y_i x_i - \frac{S_y S_x}{N} \qquad (6.12)$$

Box 6.7
Calculation of the Estimated Model Parameters ad SSE

$$b_1 = \frac{S_{yx}}{S_{xx}} \qquad (6.13)$$

$$b_0 = \frac{S_y}{N} - b_1 \frac{S_x}{N} \qquad (6.14)$$

$$SSE = S_{yy} - \frac{S_{yx}^2}{S_{xx}} \qquad (6.15)$$

An example of these basic regression calculations is presented in Box 6.8.

Box 6.8
Example of Basic Calculations for Linear Regression

Table 6.1 gives hypothetical water contamination levels for each of 15 consecutive months. A plot of the data is shown in Figure 6.3. Using the formulas in Box 6.5, the following quantities were calculated:

$$S_x = 120 \qquad S_y = 137.4 \qquad S_{xx} = 280 \qquad S_{yy} = 11.801$$

$$S_{yx} = -51.05 \qquad \bar{y} = 9.16 \qquad \bar{x} = 8$$

The estimated regression coefficients are then calculated as:

$$\hat{b}_1 = -0.1823 \qquad\qquad \hat{b}_0 = 10.62$$

Therefore the fitted model is

$$\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_i = 10.62 - .1832 x_i$$

and, the corresponding mean square error is

$$MSE = SSE/(N - 2) = \frac{2.4935}{13} = .1918.$$

The straight line in Figure 6.4 is a plot of the fitted model.

Table 6.1   Hypothetical Data for the Regression Example in F'igure 6.3

Time (Month)	Contamination (PPM)
1	10.6
2	10.4
3	9.5
4	9.6
5	10.0
6	9.5
7	8.9
8	9.5
9	9.6
10	9.4
11	8.75
12	7.8
13	7.6
14	8.25
15	8.0

Figure 6.3   Plot of data for from Table 6.1
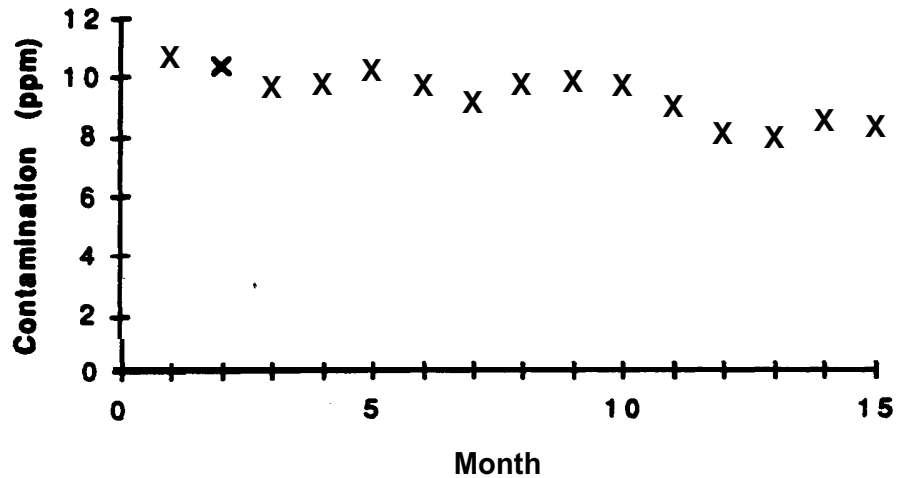
Figure 6.4   Plot of data and predicted values for from Table 6.1
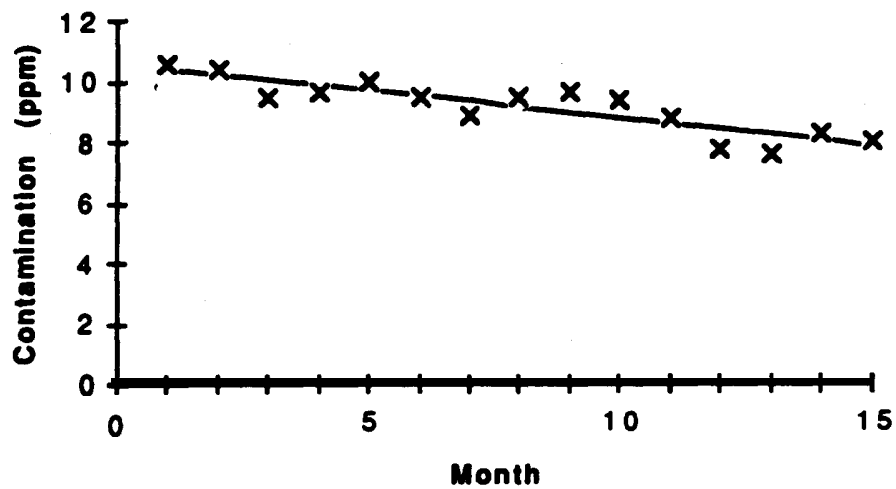
6.1.3   Assessing the Fit of the Model

It is important to note that the computational procedures given in Section 6.1.2 can-always be applied to a set of data, regardless of whether the assumed model is true. That is, it is always possible to fit a line (or curve) to a set of data. Whether the fitted model provides an adequate description of the observed pattern of data is a question that must be answered through examination of the "residuals." The residuals are the difference between the observed and predicted values for the dependent variable (see Box 6.4). If the model does not provide an adequate description of the data, examination of the residuals can provide clues on how to modify the model.

In a regression analysis, a residual is the difference between the observed concentration measurement, y_i and the corresponding fitted (predicted) value, $\hat{y}_i$ (Box 6.3). Recall that $\hat{y}_i = b_0 + b_1 x_i$, where b_0 and b_1 are the least squares estimates given by equations (6.3) and (6.2), respectively.

Since the residuals, e_i, estimate the underlying error, ε_i, the patterns exhib-ited by the residuals should be consistent with the assumptions given in Section 6.1.1 if the fitted model is correct. This means that the residuals should be randomly and approximately normally distributed around zero, independent, and have constant variance. Some graphical checks of these assumptions, arc indicated below. An example of an analysis of residuals is presented in Box 6.17.

1.   To check for model fit, lot the residuals against the time index or the time variable, x_i:. The appearance of cyclical of curvilinear patterns (see Figure 6.5, plots b and c) indicate lack of fit or inadequacy of the model (see Section 6.2.1 for a discussion of corrective measures).

2.   To check for constancy of variance, examine the plot of the residuals against x_i and the plot of the residuals against the predicted value, $\hat{y}_i$. **For both plots, the residuals should be confined within a** horizontal band such as illustrated in Figure 6.5a If the variability in the residuals increases such as in Figure 6.5d, the assumption of constant variance is violated (see Section 6.2.4 for a discussion of corrective measures in the presence of nonconstant variances).

Figure 6.5   Examples of Residual Plots (source: adapted from figures in Draper and
Smith, 1966, page 89)
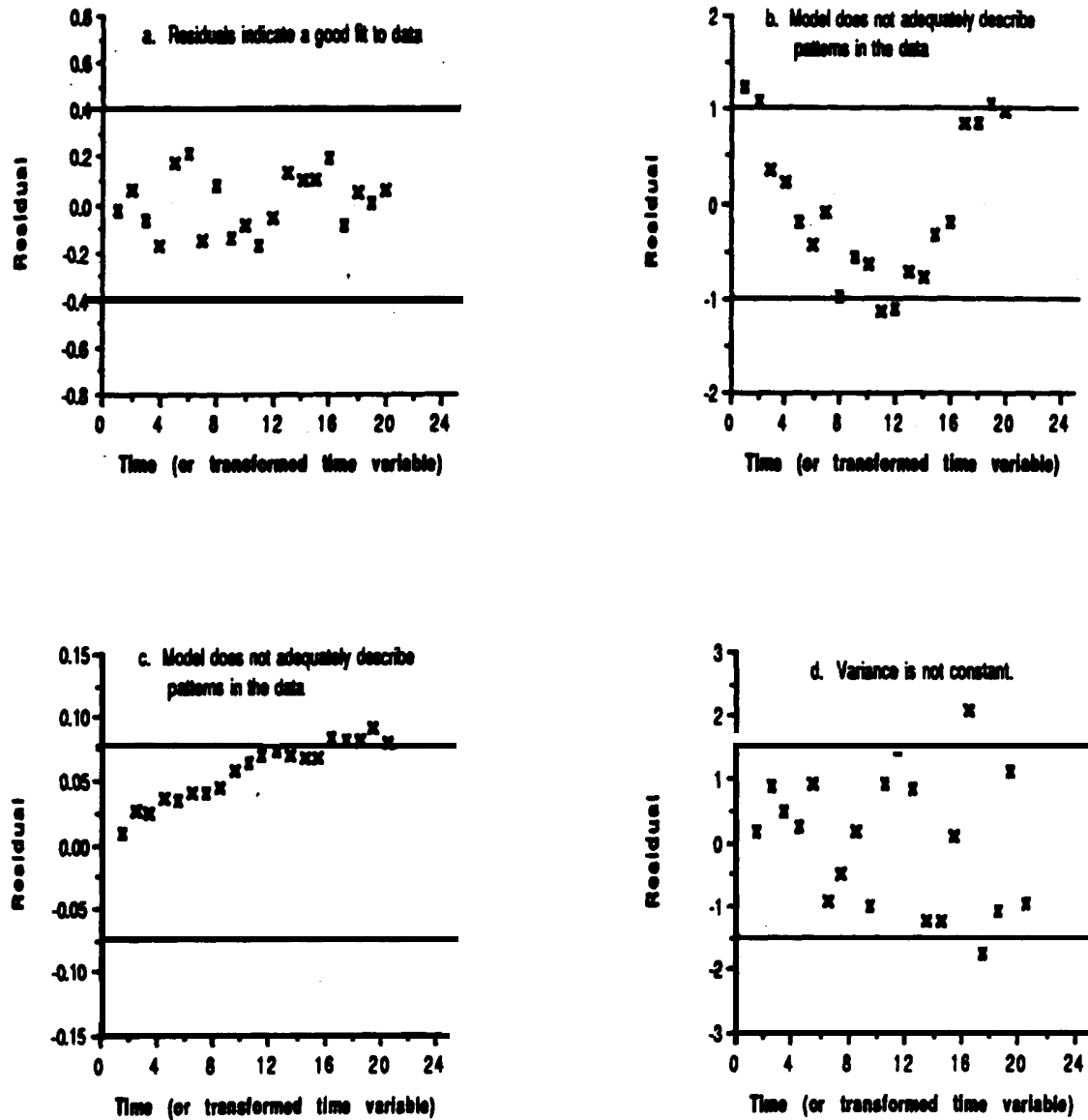
3.   To check for normality of the residuals, plot the <u>ordered</u> residuals (from smallest to largest) against their expected values under normality, EV_i using the procedures **of Section 5.7.2. Note that in** this case, the formula for computing EV_i is given by equation (5.24) **with s_{res} replaced by $\sqrt{MSE}$.**

4.   To test for independence of the error terms, compute the serial correlation of the residuals and perform the Durbin-Watson test (or the approximate large-sample test) described in Section 5.6.

It may happen that one or more of the underlying assumptions for linear regression is violated. Corrective measures are discussed in Section 6.2. Figure 6.6 shows the residuals for the analysis discussed in Box 6.8. These residuals can be compared to the examples in Figure 6.5.

Figure 6.6   Plot of residuals for from Table 6.1
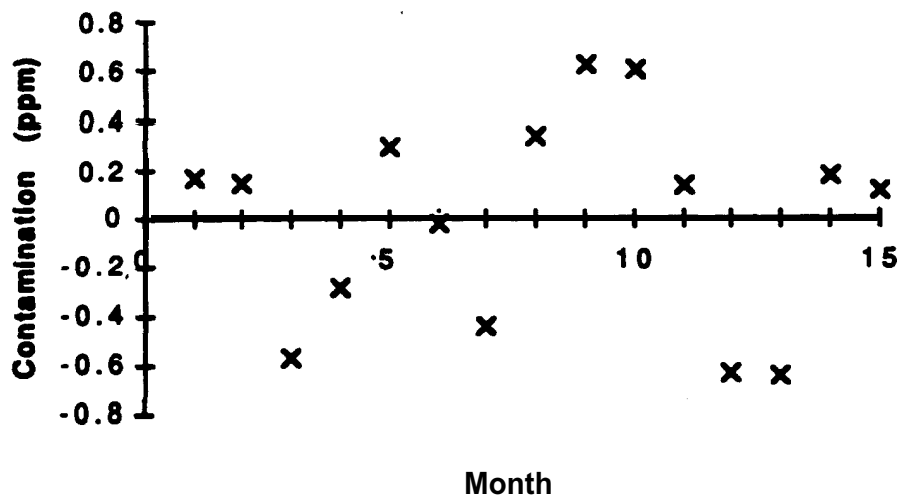
6.1.4   Inference in Regression

As mentioned earlier, two important goals of a regression analysis on ground water remediation arc the determination of significant trends in the concentration measurements and the prediction of future concentration levels. Assuming that the hypothesized model is correct, the mean square error (MSE) defined by equation (6.6) plays an

important role in malting inferences from regression models. The MSE is an estimate of that portion of the variance of the concentration measurements that is not explained by the model. It provides information about the precision of the estimated regression coefficients and predicated values, as well as the overall fit of the model.

6.1.4.1   Calculating the Coefficient of Determination

The coefficient of determination, denoted by R^2, is a descriptive statistic that provides a measure of the overall fit of the model and is defined in Box 6.9.

Box 6.9
Coefficient of Determination

$$R^2 = 1 - \frac{SSE}{S_{yy}},$$
(6.16)

where SSE is given by equation (6.6) and S_{yy} is given by equation (6.11).

R^2 is always a number between 0 and 1 and can be interpreted as the proportion of the total variance in the y_1's that is accounted for by the regression model. If R_2 is close to 1 then the regression model provides a much better prediction of individual observations than does the mean of the observations. If R_2 is close to 0 then using the regression equation to predict future observations is not much better than using the mean of the y_1's to predict future observations. A perfect fit (i.e., when all of the observed data points fall on the fitted regression line) would be indicated by an R_2 equal to 1. In practice, a value of R_2 of 0.6 or greater is usually considered to be high and thus an indicator that the model can be reasonably used for predicting future observations; however, it is not a guarantee. A plot of the predicted values from the model and the corresponding observed **values** should be examined to assess the usefulness of the model.

Figure 6.7 shows the R_2 values for several hypothetical data sets. Notice that the data in the middle of the chart (represented by the symbol "x") exhibit a pronounced

downward linear and, and this is reflected in a high R^2 of .93. On the other hand,-the set of data in the top of the chart (represented by "diamonds") exhibits no and in concentrations, and this is reflected in a low R_2 of .02. Finally, we note that the R^2 for the set of data at the bottom of the chart is fairly low (about 0.5). even though there appears to be a fairly strong (nonlinear) trend. This is because R^2 measures the linear trend over time (months). For these data, the and in the concentrations is not linear, thus the corresponding R^2 is fairly low. If the time axis were transformed to the reciprocal of time, the resulting R^2 for the third data set would be close to 0.90.

Figure 6.7   Examples of R-Square for Selected Data Sets
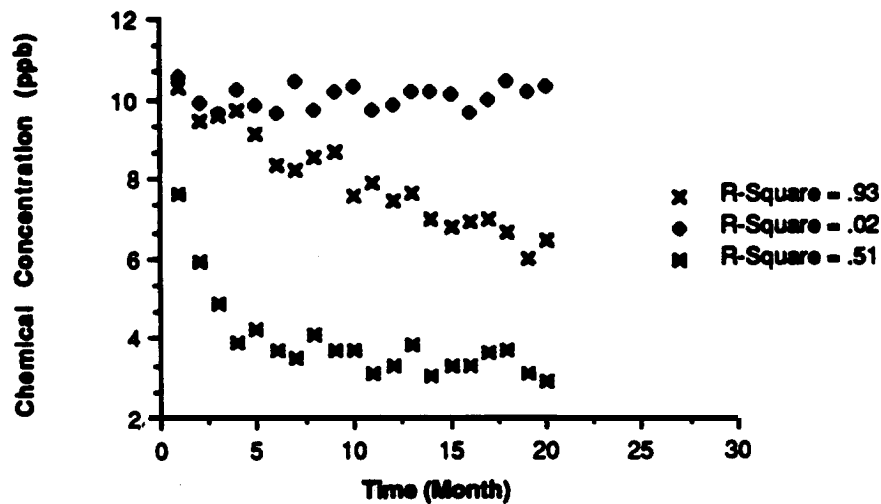
While R^2 is a useful indicator of the fit of a model and the usefulness of the model for predicting individual observations, it is not definitive. If the model is used to predict the mean concentration rather than an individual observation or if the trend in the concentrations is of interest, other measures of the model fit arc more useful. These arc addressed in the following sections.

6.1.4.2 Calculating the Standard Error of the Estimated Slope

In a simple linear regression, the slope of the fitted regression line gives the magnitude and direction of the underlying trend (if any). Because different sets of samples would provide different estimates of the slope, the estimated slope given by equation (6.2) 'is subject to sampling variability. Even if the form of the assumed model (6.1) were known to be true, it would still not be possible to determine the slope of the true relationship exactly. However, it is possible to estimate, with a specified degree of confidence, a range within which the true slope is expected to fall.

The standard error of b_1 provides a measure of the variability of the estimated slope. It is denoted by $s(b_1)$ and is defined in Box 6.10.

Box 6.10
Calculating the Standard Error of the Estimated Slope

$$s(b_1) = \sqrt{\frac{MSE}{S_{xx}}} \qquad (6.17)$$

The standard error can be used to construct a confidence interval around the true slope of the regression line. The formula for a $100(1-\alpha)$ percent confidence interval is given by equation (6.17) in Box 6.11.

Box 6.11
Calculating a Confidence Interval Around the Slope

$$b_1 \pm t_{1-\alpha/2;N-2} \, s(b_1) \qquad (6.18)$$

where $t_{1-\alpha/2;N-2}$ is the upper $1-\frac{\alpha}{2}$ percentage point of a t distribution with N-2 degrees of freedom (see Appendix Table A.1).

The confidence interval provides a measure of reliability for the estimated value b_1. The narrower the interval, the greater is the precision of the estimate b_1. Because **the confidence interval provides a range of likely values of β_1 when the model holds, it can** be used to test hypotheses concerning the significant of the observed trend.

6.1.4.3 Decision Rule for Identifying Significant Trends

If the confidence interval given by equation (6.17) contains the value zero, there is insufficient evidence (at the a significance level) to conclude that them is a trend

On the other hand, if the confidence interval includes only negative (or only positive) values, we would conclude that there is a significant negative (or positive) trend.

An example in which the above decision rule is used to identify a significant trend is given in Box 6.12.

6.1.4.4 Predicting Future Observations

If the fitted model is appropriate, then an unbiased prediction of the concentration level at time h is $\hat{y}_h = b_0 + b_1 x_h$, where x_h is the value of the time variable at time h. The standard error of the estimate is given by equation (6.18), and the corresponding $100(1 - \alpha)$ percent confidence limits around the predicted value at time h are given by formula (6.19) in Box 6.13.

Note that if the fitted regression model is based on data collected during the cleanup period, the confidence limits given by formula (6.20) may not strictly apply after treatment is terminated. Consequently, confidence limits based on data from the treatment period which are used to draw inferences about the post-treatment period should be interpreted with caution. Further discussion of the use of predicted values in ground water monitoring studies is given in Section 6.2.

Box 6.12
Using the Confidence Interval for the Slope to Identify a Significant Trend

For the data in Table 6.1, the estimated regression line was determined to be $\hat{y}_i = b_0 + b_1 x_i = 10.62 - .1823 x_i$.

The coefficient of determination for the fitted model is $R^2 = 1 - \frac{SSE}{S_{yy}}$
$= 1 - (2.49/11.8) = .79$. That is, 79 percent of the variability in the contamination measurements is explained by the regression model provided that the model is correct.

Using equation (6.16), the standard error of the estimated slope is $s(b_1) = \sqrt{\frac{MSE}{S_{xx}}} = \sqrt{\frac{.191807}{280}} = .02617$; and the corresponding 95 percent confidence limits for β_1 are given by $-.1823 \pm (2.101)(.02617)$ or $-.2373$ to $-.1273$. (Note that $\alpha = .05$, $1 - \frac{\alpha}{2} = .975$, $N = 15$, and $N-2 = 13$; thus, $t_{1-\alpha/2,N-2} = t_{.925,13} = 2.101$ from Appendix Table A.1.)

Since the interval $(-.2373, -.1273)$ does not include zero, we can conclude that the observed downward trend is significant at the $a = .05$ level. That is, we have high confidence that the observed downward trend is real and not just due to sample variability.

Box 6.13
Calculating the Standard Error and Confidence Intervals for Predicted Values

$$s(\hat{y}_h) = \sqrt{MSE \{1 + \frac{1}{N} + \frac{(x_h - \bar{x})^2}{S_{xx}}\}} \qquad (6.19)$$

$$\hat{y}_h \pm t_{1-\alpha/2;N-2} \, s(\hat{y}_h) \qquad (6.20)$$

An example in which the regression model is used to predict future values is presented in Box 6.14.

Box 6.14
Using the Simple Regression Model to Predict Future Values

Continuing the example in Box 6.11, suppose that the site manager is interested in predicting the contaminant concentration for month 16*. The predicted concentration level for month 16, assuming that the model holds, is

$$\hat{y}_{16} = b_0 + b_1 x_{16} = 10.62 - .1823(16) = 7.703.$$

The standard error of the predicted value is

$$s(\hat{y}_{16}) = \sqrt{MSE \left\{ 1 + \frac{1}{N} + \frac{(x_{16} - \bar{x})^2}{S_{xx}} \right\}}$$

$$= \sqrt{.1918 \left(1 + \frac{1}{15} + \frac{(16 - 8)^2}{280} \right)} = .4984.$$

Therefore, if the model holds, 99 percent confidence limits around the predicted value [see formula (6.20)] are given by $7.703 \pm 2.878 (.4984)$ or from 6.269 to 9.137.

---

* Again, it should be emphasized that whenever a regression model is used to make predictions about concentrations outside the range of the sampling period, extreme caution should be used in interpreting the results. In particular, the regression results should not be used alone, but should be combined with other sources of information (see discussion in Section 6.3).

6.1.4.5   Predicting Future Mean Concentrations

If the fitted model is appropriate, then an unbiased prediction of the mean concentration level at time h is $\hat{y}_h = b_0 + b_1 x_h$, where x_h is the value of the time variable at time h. Although the predicted mean and the predicted value for an individual observation are the same, the prediction error of the predicted mean is less than that for an individual predicted value. The standard error of the predicted mean is given by equation (6.21), and the corresponding $100(1 - a)$ percent confidence limits around the predicted mean at time h are given by formula (6.22) in Box 6.15

Box 6.15
Calculating the Standard Error and Confidence Interval a Predicted Mean

$$s(\hat{y}_h) = \sqrt{MSE \left\{ \frac{1}{N} + \frac{(x_h - \bar{x})^2}{S_{xx}} \right\}} \qquad (6.21)$$

$$\hat{y}_h \pm t_{1-\alpha/2;N-2} \, s(\hat{y}_h) \qquad (6.22)$$

Note that if the fitted regression model is based on data collected during the cleanup period, the confidence limits given by formula (6.19) may not strictly apply after treatment is terminated. Consequently, confidence limits based on data from the treatment period which are used to draw inferences about the post-treatment period should be interpreted with caution. Further discussion of the use of predicted values in ground water monitoring studies is given in Section 6.2.

6.1.4.6   Example of a "Nonlinear" Regression

Applying regression analysis is not always as straightforward as the examples in Boxes 6.8, and 6.12 indicate. To show some of the possible complexities and to help fix some of the ideas presented, we will do a regression analysis on the data in Table 6.2. As shown in Figure 6.8, these data are not linear with respect to time and hence a transformation of the independent variable was employed (More information about the use of transformations is given later in Section 6.2.3.) The analysis is summarized in Box 6.16 and the fitted model is plotted in Figure 6.9.

CHAPTER 6: DECIDING TO TERMINATE TREATMENT USING REGRESSION ANALYSIS

Table 6.2   Hypothetical concentration measurement for mercury (Hg) in ppm for 20 ground water samples taken at monthly intends

Month	Year	Coded month (i)	Concentration (y)	Reciprocal of month (x)
January	1986	1	0.401	1.0000
February	1986	2	0.380	0.5000
March	1986	3	0.352	0.3333
April	1986	4	0.343	0.2500
May	1986	5	0.354	0.2000
June	1986	6	0.350	0.1667
July	1986	7	0.343	0.1429
August	1986	8	0.333	0.1250
September	1986	9	0.325	0.1111
October	1986	10	0.325	0.1000
November	1986	11	0.327	0.0909
December	1986	12	0.329	0.0833
January	1987	13	0.324	0.0769
February	1987	14	0.325	0.0714
March	1987	15	0.319	0.0667
April	1987	16	0.323	0.0625
May	1987	17	0.316	0.0588
June	1987	18	0.318	0.0556
July	1987	19	0.321	0.0526
August	1987	20	0.331	0.0500

Figure 6.8   Plot of Mercury Measurements as a Function of Time (See Box 6.16)

6-21
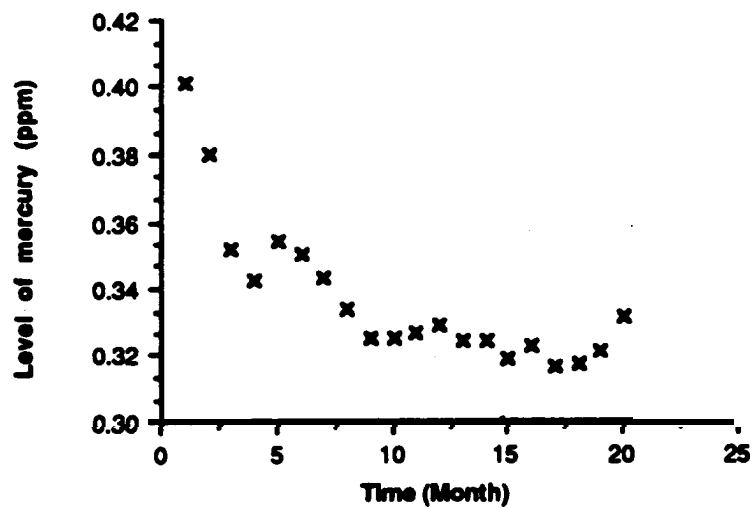
Box 6.16
Example of Basic Regression Calculations

Table 6.2 shows mercury concentrations for 20 ground water samples taken from January 1986 to August 1987. A plot of the concentration measurements as a function of time is shown in Figure 6.8. Because the data exhibited a nonlinear trend, it was decided to consider the model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where $x_i = 1/i$. The values of the reciprocals of time are shown in the last column of the table.

For these data, the following quantities were calculated: $S_x = 3.598$; $S_y = 6.739$; $S_{xx} = .949$; $S_{yy} = .00909$; $S_{yx} = .0866$, $y = .337$, $\bar{y} = .337$, $\bar{x} = .180$.

The estimated regression coefficients were then calculated as: $b_1 = .0866/.949 = .0913$; and $b_0 = .337 - (.0913)(.180) = .321$. The fitted model is therefore

$$\hat{y}_i = b_0 + b_1 x_1 = .321 + \frac{.0913}{i}$$

and the associated mean square error is

$$MSE = \frac{SSE}{18} = \frac{.00909 - \frac{.0866^2}{.949}}{18} = .000066.$$

Figure 6.9 shows a plot of the fitted model against the observed concentration values.

Figure 6.9   Comparison of Observed Mercury Measurements and Predicted Values under the Fitted Model (See Box 6.16)
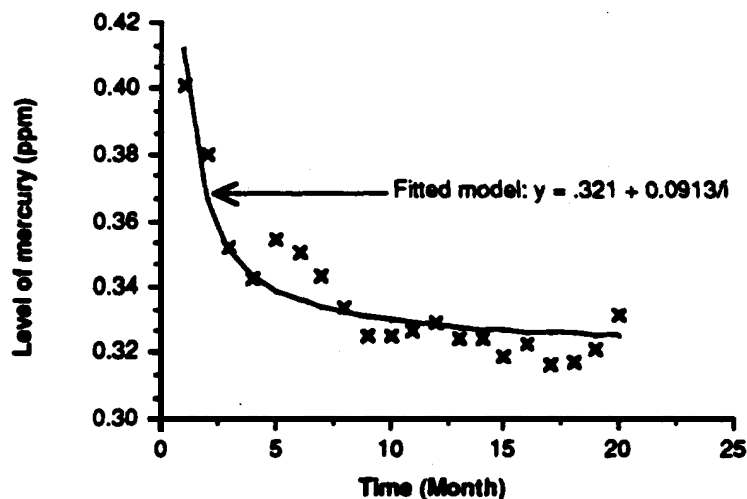
Box 6.17
Analysis of Residuals for Mercury Example

Figure 6.10 shows a plot of the residuals for the mercury data in Table 6.2 based on the fitted model, $\hat{y}_i$ = .321 + 0.0913/i (see Box 6.16). The residual plot indicates some lack of fit of the model. In particular, it appears that the fitted model tends to underestimate concentrations at the earlier times while overestimating concentrations at the later times. (Since the residuals represent the differences between the actual and predicted values, the positive values of the residuals in the earlier months indicate that the actual values tend to be larger than the predicted values then. Hence, the model underestimates the earlier concentrations.)

To see whether the fit could be improved by using a different transformation of i, the following alternative model was considered: $y_i = \beta_0 + \beta_1/\sqrt{i} + \varepsilon_i$. For this model, the estimated regression coefficients are b_0 = .2957 and b_1 = .1087, and the coefficient of determination is R^2 = .927 (compared to .89 for the earlier model). This indicates a somewhat better fit when $1/\sqrt{i}$ is used as the independent variable (see Figure 6.11). The residual plot under the new model (see Figure 6.12) seems to support this conclusion. Moreover, the standard error of b_1 is $s(b_1)$ = .0072, and hence 95 percent confidence limits around the true slope are given by .1087 ± (2.101)(.0072), or .094 to .124. Since the interval does not include zero, we further conclude that the trend is significant.

Finally, Figure 6.13 shows a normal probability plot of the ordered residuals based on the revised model, where the expected values, EV_i were computed using formula (5.24) with $s_{res} = \sqrt{MSE}$. There is a nonlinear pattern in the residuals which suggests that the normality assumption may not be appropriate for this model. If a formal test indicates the lack of normality is significant, nonlinear regression procedures should be considered.

Figure 6.10   Plot of Residuals Against Time for Mercury Example (see Box 6.17)
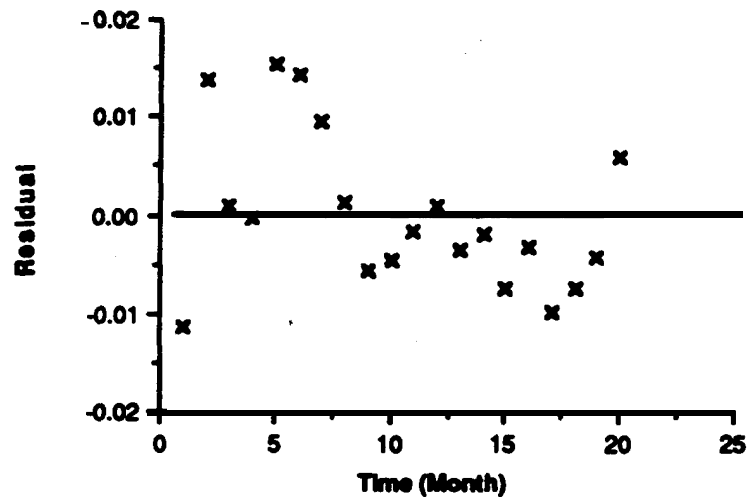
Figure 6.11   Plot of Mercury Concentrations Against $x = 1/\sqrt{i}$, and Alternative Fitted Model (see Box 6.17)
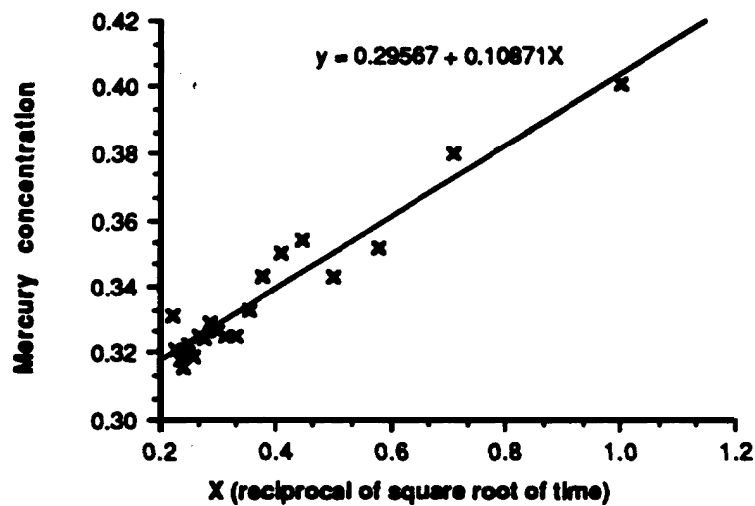
Figure 6.12   Plot of Residuals Based on Alternative Model (see Box 6.17)
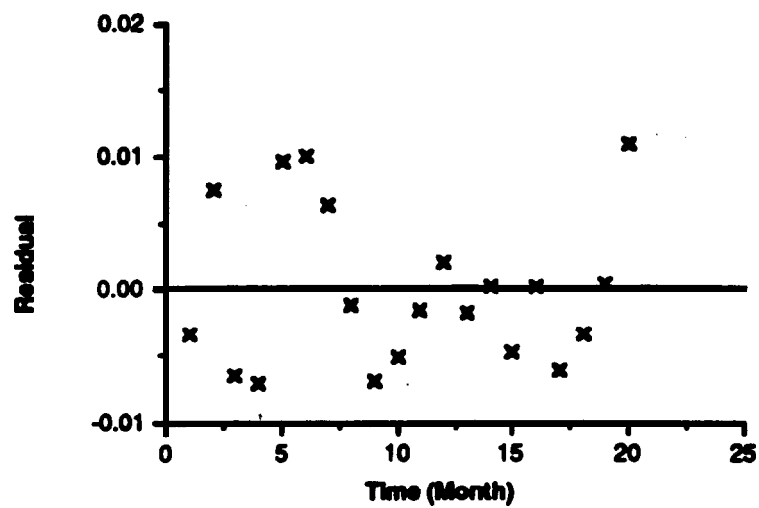
Figure **6.13   Plot of Ordered Residuals Versus Expected Values for Alternative Model
(see Box 6.17)**
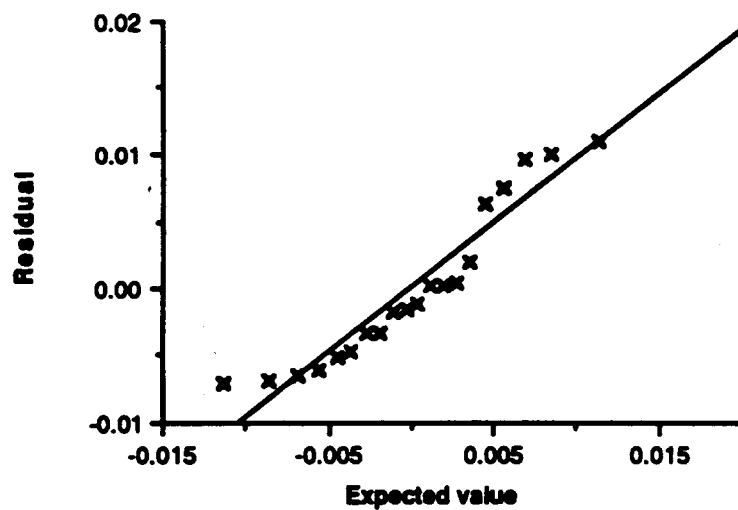
To summarize, if the data are originally linear (such as the data in Table 6.l), then we may fit the simple linear regression model of Box 6.1. If the data are more complex (e.g. the data in Table 6.2). then a transformation may be used as was done in Box 6.16. One can transform either the independent (i.e., the explanatory) variable or the dependent (i.e., the outcome) variable, or both. Finding the appropriate transformation is as much an art as it is a science. Consultation with a statistician is recommend in order to help identify useful transformations and to help interpret the model based on the transformed data.

6.2 Using Regression to Model the Progress of Ground Water Remediation

As samples arc collected and analyzed during the cleanup period, trends or other patterns in the concentration levels may become evident. **As** illustrated in Figure 6.14, a variety of patterns are possible. In situation 1, regression might be used to determine the slope for observations beyond time 20 to infer if the treatment is effective. If not, a decision might be made to consider a different remedial program. For Situation 2, the concentration measurements have decreased below the cleanup standard, and regression might be used to investigate whether the concentrations can be expected to stay below the cleanup standard. For Situation 3 in Figure 6.14, which could arise from factors such as interruptions or changes in the treatment technology or fluctuating environmental conditions, regression can be used to assess trends. However, due to the highly erratic nature of the data any p&dictions of trends of future concentrations arc likely to be very inaccurate. Additional data collection will be necessary before conclusions can be reached. Where appropriate, regression analysis can be useful in estimating and assessing the significance of observed trends and in predicting expected levels of contaminant concentrations at future points in time.

Figure 6.15 summarizes the steps for implementing a simple linear regression analysis at Superfund sites. These steps are described in detail in the sections that follow.

Figure 6.14   Examples of Contaminant Concentrations that Could Be Observed During Cleanup
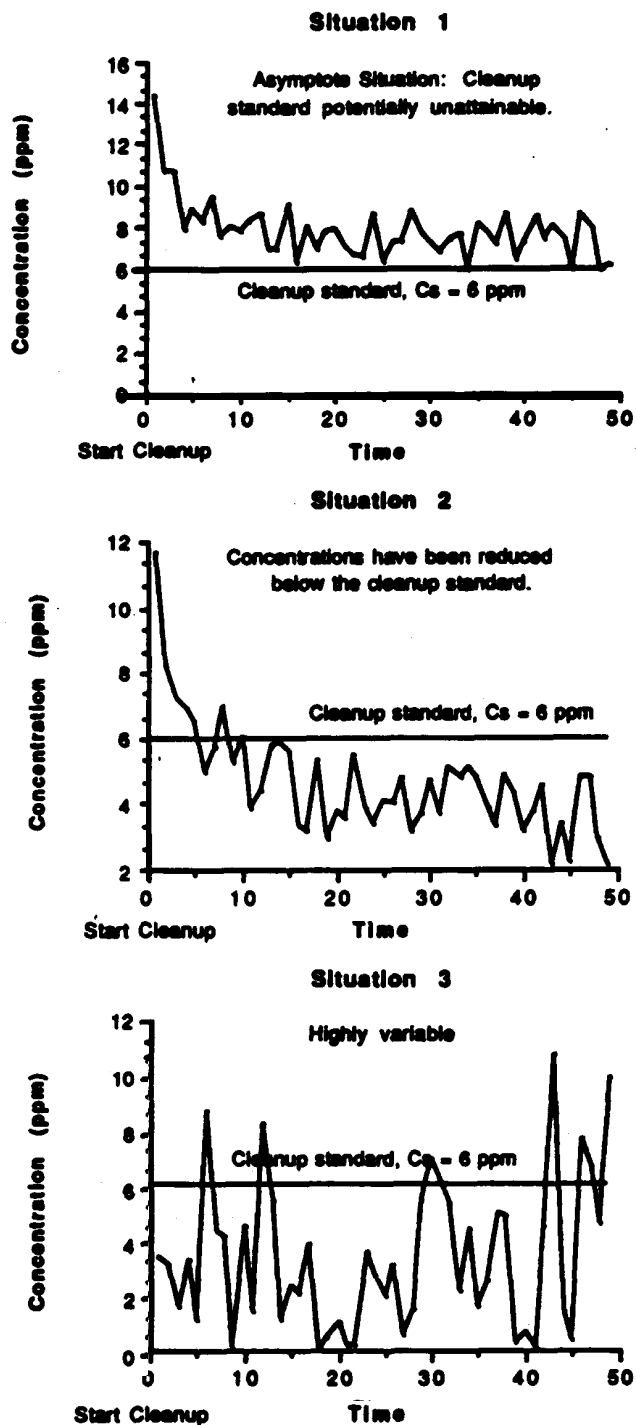
Figure 6.15   Steps for Implementing Regression Analysis at Superfund Sites
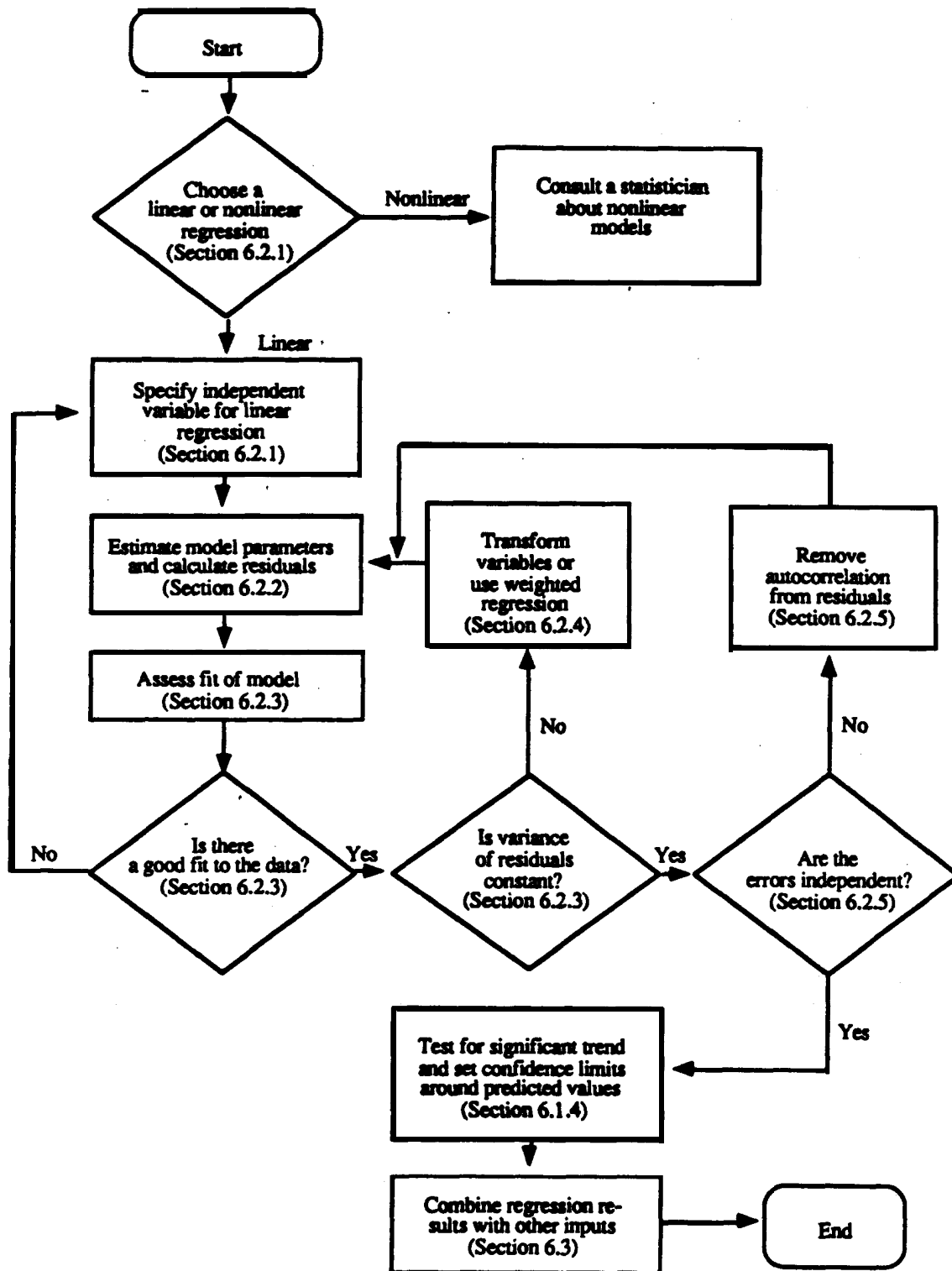
6.2.1   Choosing a Linear or Nonlinear Regression

      The first step in a regression analysis is to decide whether a linear or nonlinear model is appropriate. An initial choice can often be made by observing a plot of the sample data over time. For example, for the data of Figure 6.2, the relationship between concentration measurements and time is apparently linear. In this case, the regression model (6.1) with $x_i = i$ would be appropriate. However, for the data displayed in Figure 6.16. some sort of nonlinear model would be appropriate.

      Sometimes it is possible to model a nonlinear relationship such as that shown in Figure 6.16 with linear regression techniques by transforming either the dependent or independent variable. In some cases, theoretical considerations of ground water flows and the type of treatment applied may lead to the formulaton of a particular nonlinear model such as "exponential decay." This, in turn, may lead to consideration of a particular type of transformation (e.g., logarithmic or inverse transformations). However, these a priori considerations do not preclude testing the model for adequacy of fit. Choosing the appropriate transformation may require the assistance of a statistician; however, if the (nonlinear) relationship is not too complicated, some relatively simple transformations may be sufficient to "linearize" the model, and the procedures given in Section 6.1 may be used. On the other hand, after analysis of the residuals (as described below in Section 6.2.3), if none of the given transformations appears to be adequate, nonlinear regression methods should be used (see Draper and Smith, 1966; Neter, Wasserman, and Kutner, 1985). A statistician should be consulted about these methods.

      Figure 6.17 shows examples of two general types of curves that might reasonably approximate the relationship between observed contaminant levels and time. If a plot of the concentration measurements versus time exhibits one of these patterns, the transformations listed below in Box 6.18 may be helpful in making the model linear. Since the initial choice of transformation may not provide a "good" fit, the process of determining the appropriate transformation may require several iterations. The procedures described in Section 6.2.3 can be used to assess the fit of a particular model. Box 6.18 contains some suggested transformations for the two types of curves shown in Figure 6.17 (source: Neter, Wasserman, and Kutner, 1985).

---

[1]Although a model such as $y = \beta_0 + \beta_1 \left(\frac{1}{x}\right)$ is a nonlinear equation; it is called a linear regression model because the coefficients, β_0 and β_1, occur in a linear form (as opposed to say $y = \beta_0 + x^{\beta_1}$).

Figure 6.16   Example of a Nonlinear Relationship Between Chemical Concentration Measurements and Time
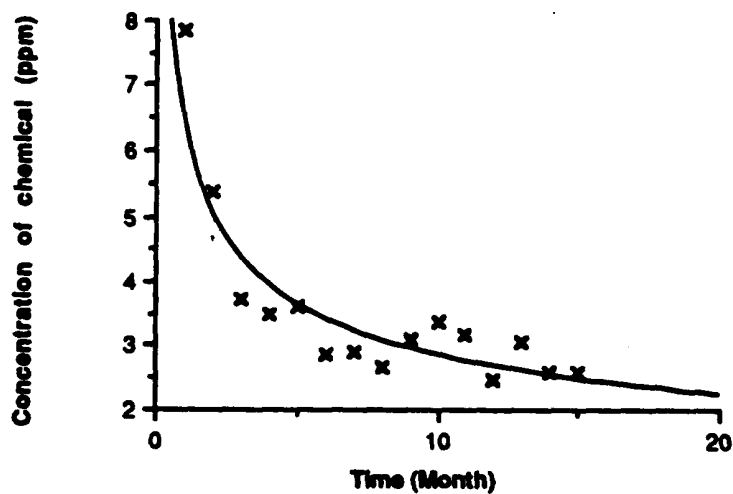
Figure 6.17   Examples of Nonlinear Relationships
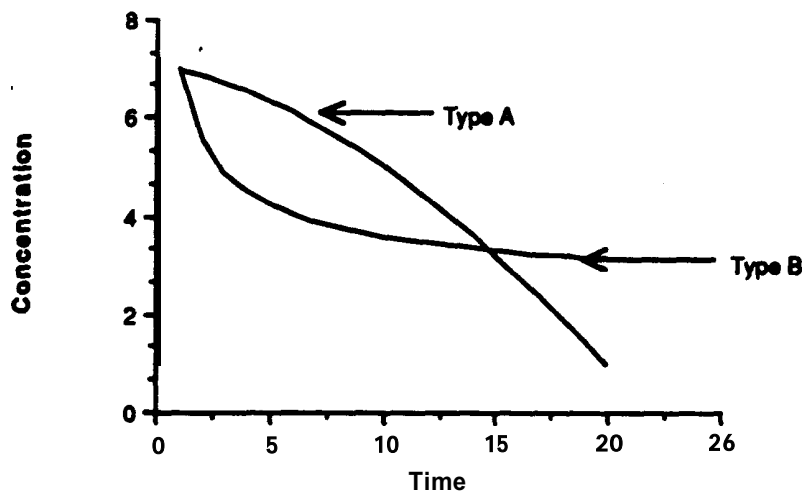
Box 6.18
Suggested Transformations

Type A: Contaminant concentrations following this pattern decrease slowly at tit and then more rapidly later on. A useful transformation to consideris

$$x_i = i^p$$

where p is a constant greater than l. If the decline in concentrations is very steep, set $p = 2$, initially, and then try alternative values, if necessary, to obtain a good fit.

Type B: Contaminant concentrations following this pattern decrease rapidly at first and then more slowly later on. Useful transformations to consider in this case are

- $$x_i = \frac{1}{i}$$

- $$x_i = 1/\sqrt{i}$$

- $$x_i = \log(i)$$

- $$x_i = \sqrt{i}.$$

Alternatively, one can also consider transforming y_i; e.g., use the transformed variable

- $$y_i' = \sqrt{y_i}$$

- $$y_i' = \log(y_i)$$

- $$y_i' = 1/y_i$$

either in lieu of or together with the transformed time variable, whichever appears to be appropriate.

There is no guarantee that using transformations will help; and its effectiveness must be determined by checking the fit of the model and examining the residuals. Consultation with a statistician is recommended to help identify useful transformations and to interpret the model based on the transformed measurements.

6.2.2 Fitting the Model

In a regression analysis, the process of "fitting the model" refers to the process of estimating the regression parameters and associated sampling errors from the observed data. With these estimates, it is then possible to (1) determine whether the model provides an adequate description of the observed chemical measurements; (2) test whether there is a significant trend in the chemical measurements over time; and (3) obtain estimates of concentration levels at future points in time.

Given a set of concentration measurements, y_i, $i = 1, 2, \ldots, N$, and corresponding time values, x_i, the estimated slope and interpret of the fitted regression line can be computed from the equations in Section 6.1.2. For the fitted model, the error sum of squares, SSE, and coefficient of determination should also be computed,

Note that the model fitting will, in general, be an iterative process. If the fitted model is inadequate for any of the reasons indicated below in Section 6.2.3, it may be possible to obtain a better fitting model by considering transformations of the data.

6.2.3 Regression in the Presence of Nonconstant Variances

If the residuals for a fitted model exhibit a pattern such as that shown in Figure 6.14d the assumption of constant variance is violated, and corrective steps must be taken. The two most common corrective measures arc: (1) transform the <u>dependent</u> variable to stabilize the variance; or (2) perform a "weighted least squares regression" (Neter, Wasserman, and Kutner, 1985).

Transformations of the dependent variable that are useful for stabilizing variances are the square root transformation, the logarithmic transformation, and the inverse transformation. Which transformation to use in a particular situation depends on the way the variance increases. To determine this relationship, it is useful to divide the data into four or five groups based on the time at which observations were made. For example, the first group might consist of the first four observations, the second group might consist of the next four observations, and so on. For the gth group, compute the mean of the observed concentrations, $\bar{y}_g$, and the standard deviation of the concentrations, s_g (Section 5.1). If a plot of s_g^2 versus $\bar{y}_g$ is approximately a straight line, use $\sqrt{y_i}$, the square root

transformation, in the regression analysis; if a plot of s_g versus $\bar{y}_g$ is approximately a straight line, use $\log(y_i)$, the logarithmic transformation, in the analysis; and, finally, if a plot of $\sqrt{s_g}$ versus $\bar{y}_g$ is approximately a straight line, use $\dfrac{1}{y_i}$, the inverse transformation, in the analysis (Neter, Wasserman, and Kutner, 1985).

The other major method for dealing with nonconstant variance is weighted least squares regression. Weighted least squares analysis provides a formal way of **accommodati**ng nonconstant variance in regression. To apply this method, the form of the underlying variance structure must be known or estimated from the data. This method is described elsewhere; e.g., Draper and Smith (1966). A statistician should be consulted when applying these methods.

6.2.4 Correcting for Serial Correlation

It is sometimes possible to remove the serial correlation in the residuals by transforming the dependent and independent variables. Applied Linear Statistical Models by Neter, Wasserman, and Kutner (1985), amplifies the following iterative procedure.

6.2.4.1 Fitting the Model

The four steps for fitting the model to remove serial correlations arc discussed below.

(1) Calculate the serial correlation of the residuals, $\hat{\phi}_{obs}$, using the formula in Box 5.14.

(2) For $i = 2, 3, \ldots, N$, transform both the dependent and independent variables using equation (6.23) in Box 6.19. Perform an ordinary least squares regression on the transformed variables. That is, using the procedures of Section 6.1.2, fit the "new" model given by equation (6.24).

Box 6.19
Transformation to "New" Model

Transform <u>both</u> the dependent and independent variables using the formulas:

$$y_i' = y_i - \hat{\phi}_{obs} y_{i-1} \quad \text{and} \quad x_i' = x_i - \hat{\phi}_{obs} x_{i-1}, \qquad (6.23)$$

Fit the following model using the transformed variables:

$$y_i' = \beta_0' + \beta_1' x_i' + \epsilon_i . \qquad (6.24)$$

Note that one observation is lost in the transformed measurements because (6.26) cannot be determined for i = 1.

Denote the least squares estimates of the parameters of the new (transformed) model by b_0' and b_1' and denote the fitted model for the transformed variables by equation (6.25) in Box 6.20.

Box 6.20
"New" Fitted Model for Transformed Variables

$$\hat{y}_i' = b_0' + b_1' x_i' \qquad (6.25)$$

Calculate the residuals for the new model: $e_i' = y_i' - (b_0' + b_1' x_i')$. Note that the fitted model (6.25) is expressed in terms of the transformed variables and not the original variables.

(3) Perform the Durbin-Watson test (or approximate test if the sample size is large) on the residuals of the model fitted in step (2). If the test indicates that the serial correlation is not significant, go to step (4). Otherwise, terminate the process and consult a statistician far alternative methods of correcting for serial correlation.

(4) In terms of the <u>original variables,</u> the slope and the intercept of the fitted regression line are provided in Box 6.21.

Box 6.21
Slope and Intercept of Fitted Regression Line in Terms of Original Variables

$$b_1 = b_1' \text{ and } b_0 = \frac{b_0'}{1 - \hat{\phi}_{obs}} \qquad (6.26)$$

where $\hat{\phi}_{obs}$ is the estimated autocorrelation determined by using the residuals obtained from fitting the untransformed data, and b_0' and b_1' are least squares estimates obtained from the transformed data.

The approach given above has the effect of adjusting the estimates of variance to account for the presence of autocorrelation. Typically, the variance of the estimated regression coefficients is larger when the errors are correlated, as compared with uncorrelated errors. An example of the use of this technique is given in Box 6.22.

6.2.4.2 **Determining Whether the Slope is Significant**

The standard error of the slope of the original model is simply the standard error of the slope, b_1, obtained from the regression analysis performed on the transformed data defined in Box 6.21. The formulas given in Section 6.1.4 can be used to compute the standard error of b_1'. The decision rule in Section 6.1.4.3 can be used to identify whether the trend is statistically significant Note that for the transformed data, the total number of observations is N- 1.

Box 6.22
Correcting far Serial Correlation

Table 6.3 shows the concentration of benzene in 15 quarterly ground water samples taken from a monitoring well at a former manufacturing site. It appeared from a plot of the data (see Figure 6.18) that a simple linear model of the form: $y_i = \beta_0 + \beta_1 i + \epsilon_i$ might be appropriate in describing the relationship between concentrations and time.

A regression analysis was performed on the data with the following results: (a) the fitted model was estimated to be $\hat{y}_i = 29.20 - .478i$; (b) $R^2 = 0.73$; (c) 95 percent confidence limits around the slope of the line were calculated to be $-0.478 \pm (2.16)(.082)$, or -0.66 to -0.30; and (d) the Durbin-Watson statistic was computed to be D = .795.

For N = 15 and p-1=1 (there are two parameters in the model), the critical value for the Durbin-Watson test is $d_U = 1.36$ at the .05 significance level. Since D < 1.36, it was concluded that there was a significant autocorrelation. Although the calculated confidence interval for the slope of the line apparently indicated that the observed downward trend was significant, it was recognized that the presence of autocorrelations could lead to erroneous conclusions. Therefore, the data were re-analyzed using the method of transformations described earlier in this section.

First, the serial correlation was computed from the residuals as $\hat{\phi}_{obs} = .57$. Then the observed concentrations and time variable were transformed as follows: $y_i' = y_i - .57y_{i-1}$; and $x_i' = i - .57(i-1)$. A regression of y_i' on x_i' resulted in least squares estimates of $b_1' = -.34$ and $b_0' = 11.89$ for the transformed variables, with $s(b_1') = .17$. Therefore, using equation (6.26), estimates of the slope and intercept for the original data were calculated as

$$b_1 = b_1' = -.34, \text{ and } b_0 = \frac{b_0'}{1-.57} = \frac{11.89}{.43} = 27.65.$$ Note that the revised estimates are close to the original estimates, except that now the standard error of b_1 is much larger that it was before the effect of the autocorrelations was taken into account in the analysis (.17 vs. .082). Because of this increase in variance, 95 percent confidence limits around the true slope are now given by $-.34 \pm (2.179)(.17)$, or -.71 to .03. In this case, the interval includes zero, and therefore at the five percent significance level, we cannot conclude that the observed trend is significant.

Table 6.3   Benzene concentrations in 15 quarterly samples (see Box 6.22)

Year	Quarter	Coded quarter (i)	Concentration in ppb (y)
1985	First	1	30.02
	Second	2	29.32
	Third	3	28.12
	Fourth	4	28.32
1986	First	5	27.01
	Second	6	24.78
	Third	7	24.00
	Fourth	8	23.78
1987	First	9	24.25
	Second	10	23.24
	Third	11	21.98
	Fourth	12	25.00
1988	First	13	24.10
	Second	14	23.75
	Third	15	23.00

Figure 6.18   Plot of Benzene Data and Fitted Model (see Box 6.22)
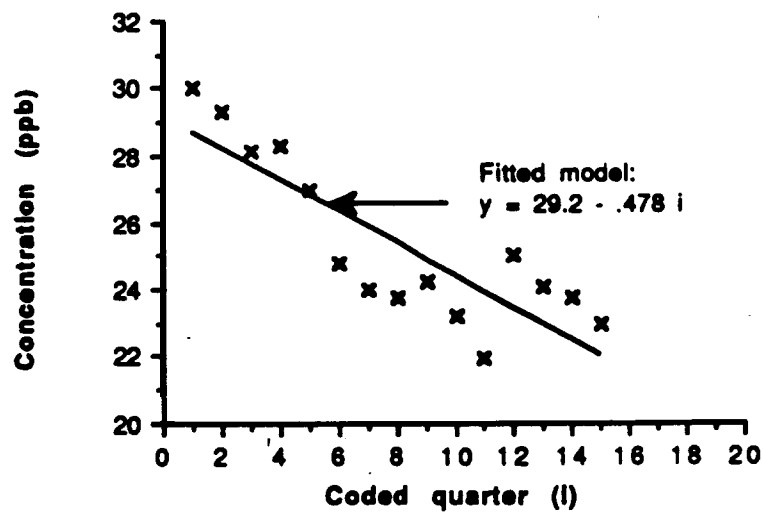
6.2.4.3 Calculating the Confidence Interval for a Predicted Value

The general procedures in Section 6.1.4 can also be used to develop confidence limits for the predicted concentration at arbitrary time h (as shown in Box 6.23).

Box 6.23
Constructing Confidence Limits around an Expected Transformed Value

Referring to the fitted model (6.28), use equation (6.19) to construct confidence limits around the expected transformed value at time h:

$$U_h' = \hat{y}_h' + t_{1-\alpha/2;N-3}\, s(\hat{y}_h') \qquad (6.27)$$

and

$$L_h' = \hat{y}_h' - t_{1-\alpha/2;N-3}\, s(\hat{y}_h'). \qquad (6.28)$$

where, $\hat{y}_h' = b_0' + b_1'x_h'$; x_h = the value of the time variable at time h; and $s(\hat{y}_h')$ is the standard error of $\hat{y}_h'$ as computed from equation (6.18) using the transformed data. Note that the "t value" used in the confidence interval is based on N-3 (instead of N-2) degrees of freedom because we are estimating and additional parameter (the serial correlation) from the data.

Since the limits given in equations (6.27) and (6.28) are in the transformed scale, the upper- and lower-confidence limits in the original scale arc given by:

$$y_{h,upper} = U_h' + \hat{\phi}_{obs}y_h \qquad (6.29)$$

and

$$y_{h,lower} = L_h' + \hat{\phi}_{obs}y_h. \qquad (6.30)$$

6.3 Combining Statistical Information with Other Inputs to the Decision Process

The statistical techniques presented in this chapter can be used to (1) determine whether contaminate concentrations are decreasing over time, and/or (2) predict future concentrations if present trends continue. Other factors must be used in combination with

these statistical results to decide whether the remedial effort has been successful, and when treatment should be terminated Several factors to consider are:

·   Expert knowledge of the ground Water at this site and experience with other remedial efforts at similar sites;

·   The results of mathematical models of ground water flow and chemistry with sensitivity analysis and assessment of the accuracy of the modeling results; and

·   cost and scheduling considerations.

The sources of information **above** can be used to answer the following questions:

·   How long will it take for the ground water system to reach steady state before the sampling for the attainment decision can begin?

·   What is the chance that the ground water concentrations will substantially exceed the cleanup standard before the ground **water** reaches steady state?

·   What are the chances that the final assessment will conclude that the site attains the cleanup standard?

·   what are the costs of (1) continuing treatment, **(2)** performing the assessment, and (3) planning for and initiating additional treatment if it is decided that the site does not attain the cleanup standard?

The answers to these questions should be made in consultation with both statistical and ground water experts, managers of the remediation effort and the regulatory agencies.

6.4   Summary

This chapter discussed the use of regression methods for helping to decide when to stop treatment. In particular, procedures were given for estimating the and in contamination levels and predicting contamination levels at future points in time. General methods for fitting simple linear models and assessing the adequacy of the model were also discussed.

In deciding when to terminate treatment, the chapter emphasized that:

- **Interpreting the data is usually a multiple-step process of refining the model and understanding the data;**

- **Models are a useful but imperfect description of the data. The usefulness of a model can be evaluated by examining how well the assumptions fit the data, including an analysis of the residuals;**

- **Correlation between observations collected over time can be important and must be considered in the model;**

- **Changes in treatment over time can result in changes in variation, and correlation and can produce anomalous behavior which must be understood to make correct conclusions from the data; and**

- **Consultation with a ground water expert is advisable to help interpret the results and to decide when to terminate treatment.**

Deciding when to terminate treatment should be based on a combination of statistical results, expert knowledge, and policy decisions. Note that regression is only one of various statistical methods that may be used to decide when treatment should be terminated. Regression analysis was discussed in this document because of its relative simplicity and wide range of applicability; however, this does not constitute an endorsement of regression as a method of choice.

7. ISSUES TO BE CONSIDERED BEFORE STARTING ATTAINMENT SAMPLING

After terminating treatment and before collecting water samples to assess attainment, a period of time must pass to ensure that any transient effects of treatment on the ground water system have sufficiently decayed. This period is represented by the unshaded portion in the figure below. This chapter discusses considerations for deciding when the sampling for the attainment decision can begin and provides statistical tests, which can be easily applied, to guide this decision. The decision on whether the ground water has reached steady state is based on a combination of statistical calculations, ground water modeling, and expert advice from hydrogeologists familiar with the site.

Figure 7.1 **Example Scenario for Contaminant Measurements During Successful Remedial Action**
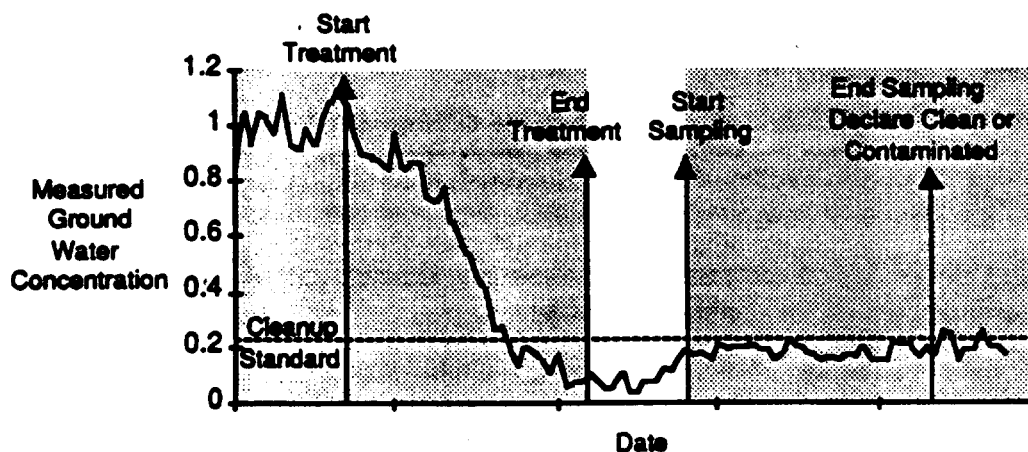
The degree to which remediation efforts affect the ground water system at a site is difficult to determine and depends on the physical conditions of the site and the treatment technologies used. **As** previously discussed, the ground water can only be judged to attain the cleanup standard if both present and future contaminant concentrations are acceptable. Changes in the ground water system due to treatment will affect the contaminant concentrations in the sampling wells. For example, while remediation is in progress pumping can alter water levels, water flow, and thus the level of contamination being measured at monitoring wells. To adequately determine whether the cleanup standard has been attained, the ground water conditions for sampling must approximate the expected conditions in the

future. Consequently, it is important to establish when the residual effects of the treatment process (or any other temporary intervention) on the ground water appear to be negligible. When this point is reached, sampling to assess attainment can be started and inferences on attainment can be drawn. We will define the state of the ground water when temporary influences no longer affect it as a "steady state." "Steady state," although sometimes defined in the precise technical sense, is used here in a less formal manner as indicated in Section 7.1.

7.1 The Notion of "Steady State"

The notion of "steady state" may be characterized by the following components:

1.a.   After treatment, the water levels and water flow, and the corresponding variability associated with these parameters (e.g., seasonal patterns), should be essentially the same as for those from comparable periods of time prior to the remediation effort.

or

1.b.   In cases where the treatment technology has resulted in permanent changes in the ground water system, such as the placement of slurry wells, the hydrologic conditions may not return to their previous state. Nevertheless, they should achieve a state of stability which is likely to reflect future conditions expected at the site. For this steady state, the residual effects of the treatment will be small compared to seasonal changes.

2.   The pollutant levels should have statistical characteristics (e.g., a mean and standard deviation) which will be similar to those of future periods.

The first component implies that it is important to establish estimates of the ground water levels and flows prior to remediation or to predictively model the effect of structures or other features which may have permanently affected the ground water. Variables such as the level of ground water should be measured at the monitoring wells for a reasonable period of time prior to remediation, so that the general behavior and characteristics of the ground water at the site are understood.

The second component is more judgmental. Projections must be made as to the future characteristics of the ground water and the source(s) of contamination, based

on available, current information. Of course, such projections cannot be made with certainty, but reasonable estimates about the likelihood of events may be established.

The importance of identifying when ground water has reached a steady state is related to the need to make inferences about the future. Conclusions drawn from tests assessing the attainment of cleanup standards assume that the current state of the ground water will persist into future. There must be confidence that once a site is judged clean, it will remain clean. Achieving a steady state gives credence to future projections derived from current data.

7.2 Decisions to be Made in Determining When a Steady State is Reached

Immediately after remediation efforts have ended. the major concern is determining when ground water achieves steady state. In order to keep expenditures of time and money to a minimum, it is desirable to begin collecting data to assess attainment as soon as one is confident that the ground water has reached a steady state.

When sampling to determine whether the ground water system is at steady state, three decisions arc possible:

·   The ground water has reached steady state and sampling for assessing attainment can begin;

·   The measurements of contaminant concentrations during this period indicate that the contaminant(s) are unlikely to attain the cleanup standard and further treatment must be considered; or

·   More time and sampling must occur before it can be confidently assumed that the ground water has reached steady state.

Next, various criteria will be considered that can be used in determining whether a steady state has been reached

7.3 Determining When a Steady State Has Been Achieved

In the following sections, qualitative and quantitative criteria involved in making the decision as to whether the ground water has returned to a steady state following

remediation are discussed. Some of these criteria are based on a comparison of present ground water levels with comparable levels before treatment. Others are based solely on measurements and conditions after treatment has terminated. To a certain extent, the decision as to when steady state has been reached is judgmental. It is not possible to prove that a ground water system has achieved steady state. Thus, it is important to examine data obtained from the ground water system to see if there are patterns which suggest that steady state has not been achieved. If there are no such patterns (e.g., in the water level or speed and direction of water flow), it may be reasonable to conclude that a steady state has been reached.

Any data on the behavior of the ground water prior to the undertaking of remediation may serve as a useful baseline, indicating what "steady state" for that system had been and, thus, to what it might return. However, the actions of remediation and the resulting physical changes in the area may change the characteristics of steady state. In this case, such a comparison may be less useful. When it seems clear that steady state characteristics have changed after remediation efforts, it is usually prudent to allow mom time for remediation effects to decay.

Collection of data to determine whether steady state has been achieved should begin at the various monitoring wells at the site after remediation has been terminated. The variables for which data will be obtained should include measures related to the contaminant levels, the ground water levels, the speed and direction of the flow, and any other measures that will aid in determining if the ground water has returned to a steady state. The frequency of data collection will depend on the correlation among consecutively obtained values (it is desirable to have a low correlation). A period of three months between data collection activities at the wells may be appropriate if them appears to be some correlation between observations. With little or no correlation, monthly observations may prove useful. If the serial correlation seems to be high, the time interval between data collection efforts should be lengthened. With little or no information about seasonal patterns or serial correlations in the data, at least six observations per year are recommended. After several years of data collection, this number of observations will allow an assessment of seasonal patterns, trends, and serial correlation. It may be useful to consult with a statistician if there is some concern about the appropriate sampling frequency.

All data collected should be plotted over time in order to permit a visual analysis of the extent to which a steady state exists for the ground water. In Section 7.4,

7-4

the chatting of data and the construction of plots arc discussed. Section 7.4.3 provides illustrations of such plots and their interpretation. In Section 7.4.4, statistical tests that can be employed for identifying departures from randomness (e.g., trends) in the data are indicated. Suggestions far seasonally adjusting data prior to plotting are provided, and graphical methods are discussed.

7.3.1   Rough Adjustment of Data for Seasonal Effects

One concern in applying graphical techniques is that the data points being plotted are assumed to be independent of each other. Even if the serial correlation between observations is low, there may be a seasonal effect on the observations. For example, concentrations may be typically higher than the overall average in the spring and lower in the fall. To adjust for seasonal effects, one may subtract a measure of the "seasonal" average from each data value and then add back the overall average (Box 7.1). The addition of the overall average will bring the adjusted values back to the original levels of the variable to maintain the same reference frame as the original data.

Box 7.1
Adjusting for Seasonal Effects

Suppose we let x_{jk} be the jth individual data observation in year k, $\bar{x}_j$ be the average for period j obtained from the baseline period prior to treatment for period j, and $\bar{\bar{x}}$ be the overall average for all data collected for the baseline period. For example, if six data values per year have been collected bimonthly for each of three years during the baseline period, six $\bar{x}_j$ values would be computed, each based on three data points taken from the three different years for which data were collected. The value $\bar{\bar{x}}$ would be computed over all 18 data values. The adjusted jth data observation in year k, x'_{jk}, can then be computed from:

$$x'_{jk} = x_{jk} - \bar{x}_j + \bar{\bar{x}} \qquad (7.1)$$

If there are missing values, calculate $\bar{x}_j$ as in Box 5.4.

Plot the values of x_{jk} versus time. In examining these plots, checks for runs and trends can be made for the adjusted values.

7.4 Charting the Data

In general, it is useful to plot the data collected from a monitoring program. Such plots are similar to "control charts" often used to monitor industrial processes, except control limits will not appear on the charts discussed here. Use the horizontal, or X-axis, to indicate the time at which the observation was taken; and use the vertical, or Y-axis, to indicate the value of the variable of interest (e.g., the contaminant level or water table level or the value of other variables after adjustment for seasonal effects). Figure 7.2 gives an example of a plot which may be used to assess stability during the period immediately following treatment.

Notice that in Figure 7.2, the "prior average" has also been placed on the plot. This line represents the average of the baseline data collected before remediation efforts began. For example, this value could be the average of eight points collected quarterly over a two-year period. It may also be useful to plot separately the individual observations gathered to serve as the baseline data, so that information reflecting seasonal variability and the degree of serial correlation associated with the baseline period can be readilyexamined.

Figure 7.2   Example of Time Chart for Use in Assessing Stability
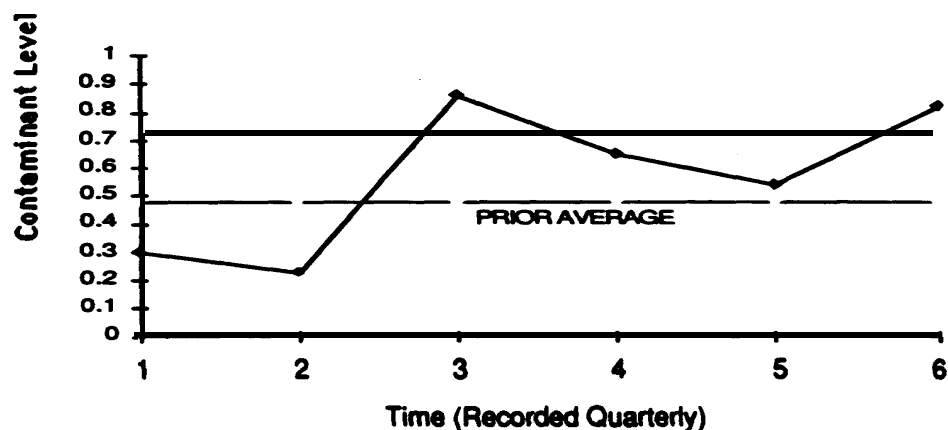
7.4.1 A Test for Change of Levels Based on Charts

If the ground water conditions after remediation are expected to be compa-rable to the prior conditions, we would expect that the behavior of water levels and flows to resemble that of those same variables prior to the remediation effort in terms of average and variability. One indication that a steady state may <u>not</u> have been reached is the presence of a string of measurements from the post treatment period which arc consistently above or below the average prior to beginning remediation. A common rule of thumb used in indus-trial Statistical Process Control (SPC) is that if eight consecutive points are above or below the average (often called a "run" in SPC terminology), the data are likely to come from a different process than that from which the average was obtained (Grant and Leavenworth, 1980). This rule is based on the assumption that the observations are independent. This assumption is not strictly applicable in ground water studies since there is likely to be serial correlation between observations as well as seasonal variability. Assuming independent observations, an eight-point run is associated with a 1 in 128 chance of concluding that the mean of the variable of interest has changed when, in fact, there has been no change in the mean.

The above discussion suggests that for the purpose of deciding whether the ground water has achieved steady state, a string of 7 to 10 consecutive points above or below the prior average might serve as evidence indicating that the state of the ground water is different from that in the baseline period. If it is suspected that a high degree of serial conelation exists, it would be appropriate to require a larger number of consecutive points.

7.4.2 A Test for Trends Based on Charts

The charts described here provide a simple way of identifying trends. If six consecutive data points arc increasing (or decreasing)[1] -- sometimes stated as "5 consecutive intervals of data" so that it is understood that the first point in the string is to be counted -- then there is evidence that the variable being monitored (e.g., water levels or flows, or contaminant concentrations) has changed (exhibits a trend). Again, independence

---

[1]This rule of 6 is based on the assumption that all 720 orderings of the points are equally likely. This is **not always true. Hence such rules are to be considered only as quick but reasonable approximations.**

of the observations is assumed. A group of consecutive points that increase in value is sometimes referred to as a "run up," while a group of consecutive points that decrease in value is referred to as a "run down."

With the rule of six consecutive data points described above, the chance of erroneously concluding that a trend exists is only 1 in 360, or about 0.3 percent. In contrast, a rule based on five consecutive points has a 1 in 60 chance (1.6 percent) of erroneously concluding that there is a trend, while a rule based on seven consecutive points would have a corresponding 1 in 2,520 chance (0.04 percent) of erroneously concluding that there is a trend. Thus, depending on the degree of serial correlation expected, a "and" of 5 to 7 points may suggest that the ground water levels and flows are not at steady state.

In practice, data for many ground water samples may be collected before any significant runs are identified. For example, in a set of 30 monthly ground water flow rate measurements, there may be a run up of seven points and several shorter runs. Such patterns of runs can be analyzed by examining the length or number of runs in the series. Formal statistical procedures for analyzing trends in a time series are given by Gilbert (1987).

A quick check for a general trend over a long period of time can be accomplished as follows. Divide the total number of data points available, N, by 6. Take the closest integer smaller than N/6 and call it I. Then select the Ith data value over time, the $2(I^{th})$, the $3(^{Ith})$, etc. For example, if N = 65, then I = 10, and we would select the 10th, 20th, etc., points over time. If there are six consecutive points increasing or decreasing over time, there is evidence of a trend. This test will partially compensate for serial con-elation.

7.4.3   Illustrations and Interpretation

Once the plotting of data has begun, there are various patterns that may appear. Figures 7.3 through 7.8 represent six charts which indicate possible patterns that may be encountered. Evidence of departures from stability is being sought. The first five charts, except Figure 7.4, indicate evidence of instability (or in the cases of Figures 7.5 and 7.6, suspicions of possible instability), i.e., changes in characteristics over time. Figure 7.3 shows "sudden" apparent outliers or spikes that indicate unexpected variability

in the variable being monitored. Figure 7.4 illustrates a six-point trend in the variable being monitored. Figures 7.5 and 7.6 suggest that a trend may exist but there is insufficient evidence to substantiate it. Attention should be paid to the behavior of subsequent dam in these cases. (In particular, the data in Figure 7.5 could indicate a general trend using the "quick check" discus& in the previous section depending on the randomly selected set of points included in the test.) Figure 7.7 reflects a change (around observation 15) in both variability (the spread of the data becomes much greater) and average (the average appears to have increased). Figure 7.8 indicates a variable that appears to be stable.

In interpreting the plots, the return to a steady state will generally be indicated by a random scattering of data points about the prior average. The existence of patterns such as runs or trends suggests instability. Pattern, associated with seasonality and serial correlation should be consistent with those seen prior to remediation. At the very least, the average value for levels of contaminants after remediation should be lower than that prior to remediation. A run below the prior average fur contaminant level measures would certainly not be evidence that the ground water is not at steady state, since the whole point of the remediation effort is to reduce the level of contamination. A trend downwards in contamination levels may be an indication that a steady state has not been reached. Nevertheless, if substantial evidence suggests that this decline or an eventual leveling off will be the future state of that contaminant on the site, tests for attainment of the cleanup standards would be appropriate.

On the other hand, if it seems that the average contamination level after remediation will be above the prior average or that there is a consistent trend upwards in contamination levels, it may be decided that the previous remediation efforts were not totally successful, and further remediation efforts must be undertaken. This may be done with a minimal amount 'of data, if, based on the data available, it appears unlikely that the cleanup standard will be met. However, what should be taken into account is the relative cost of making the wrong decision. Two costs should be weighed against each other: the cost of obtaining further observations from the monitoring wells if it turns out that the decision to resume remediation is made at a later date (the loss here is in terms of time and the cost of monitoring up to the time that remediation actually is resumed) against the cost of resuming remediation when in fact a steady state would eventually have been achieved (the loss here is in terms of the cost of unnecessary cleanup effort and time). In addition, the likelihood of making each of these wrong decisions, as estimated based on the available information, should be incorporated into the decision process.

Figure 7.3 Example of Apparent Outliers
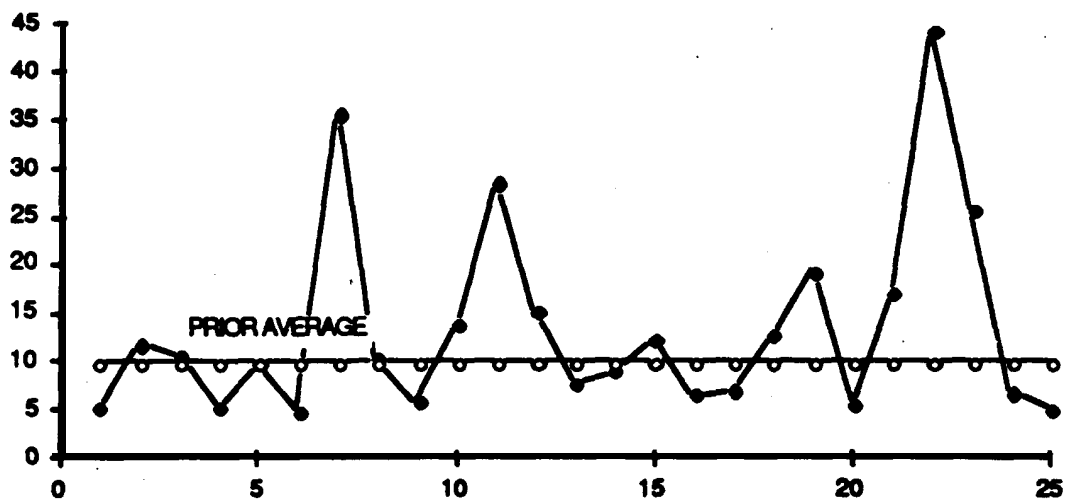
Figure 7.4   Example of a Six-point Upward Trend in the Dam
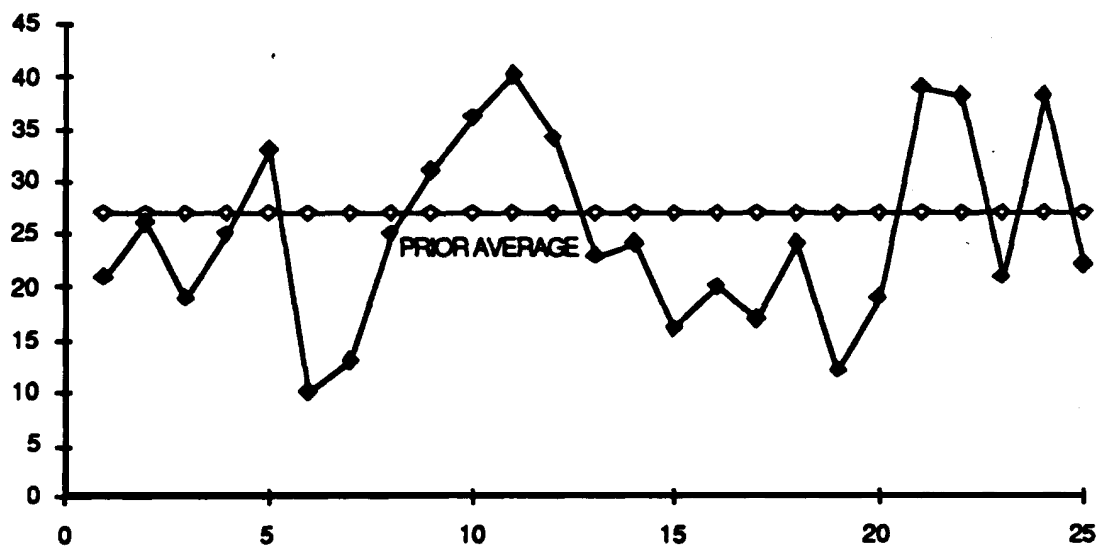
7.4.4　　　Assessing Trends via Statistical Tests

The discussions- in Section 7.4.3 considered graphical techniques for exploring the possible existence of trends in the dam. Regression techniques discussed in Chapter 6 provide a formal statistical procedure for considering possible trends in the **data.**

Other formal procedures for testing for trends also exist. Gilbert (1987) discusses several of them, such as the Seasonal Kendall Test, Sen's Test for Trend, and a Test for Global Trends (the original articles in which these tests ate described were: Hirsh and Slack, 1984; Hirsch, Slack, and Smith, 1982; Farrell, 1980; and van Belle and Hughes, 1984).

The Seasonal Kendall Test provides a test for trends removes seasonal effects. It has been shown to be applicable in cases where monthly observations have been gathered for at least three years. The degree to which critical values obtained from a normal table approximate the true critical values apparently has nut been established for other time intervals of data collection--e.g., quarterly or semi-annually. This test would have to be carried out for each monitoring well separately at a site. Sen's Test for Trend is a more sensitive test far detecting monotonic trends if seasonal effects exist, but requires more complicated computations if there are missing data. The Test for Global Trends provides the capability for looking at differences between seasons and between monitoring wells, at season-well interactions, and also provides an overall trend test. All three of these tests (the Seasonal Kendall, Sen's, and the Global tests) require the assumption of independent observations. (Extensions of these tests allowing for serial correlations require that much more data be collected--for example, roughly 10 years worth of monthly data for the Seasonal Kendall test extension.) If this assumption is violated these tests tend to indicate that a trend exists at a higher rate than specified by the chosen a level when it actually does not. Thus, these tests may provide useful tools for detecting trends, but the finding of a trend via such a test may not necessarily represent conclusive evidence that **a** trend exists. Gilbert provides a detailed discussion of all three tests as well as computer code that can be used for implementing the tests. However, this discussion does not consider the power of these trend tests, i.e., the likelihood that such tests identify a trend when a trend actually

exists is not addressed. If the power of these tests is low, existing trends may not be detected in a timely fashion.

7.4.5   Considering the Location of Wells

In addition to assessing the achievement of steady state in a well over time, it is also useful to consider the comparison of water and contamination levels across wells at given points in time. This can readily be done by constructing either (1) a scatter plot with water or contamination levels on the vertical axis and the various monitoring wells indicated on the horizontal axis, or (2) constructing a contour plot of concentrations or water levels across the site and surrounding area. Commercial computer programs are available for preparing contour plots. In particular, see the discussion in Volume 1 (Chapter 10) on kriging. If there are, large, unexpected differences in water or contamination levels between wells, this may suggest that steady state has not yet been reached.

7.5   Summary

Finding that the ground water has returned to a steady state after terminating remediation efforts is an essential step in the establishment of a meaningful test of whether or not the cleanup standard have been attained. There arc uncertainties in the process, and to some extent it is 'judgmental. However, if an adequate amount of data are carefully gathered prior to beginning remediation and after ceasing remediation, reasonable decisions can be made as to whether or not the ground water can be considered to have reached a state of stability.

The decision on whether the ground water has reached steady state will be based on a combination of statistical calculations, plots of data, ground water modeling, use of predictive models, and expert advice from hydrogeologists familiar with the site

Figure 7.5   Example of a Pattern in the Data that May Indicate an Upward Trend
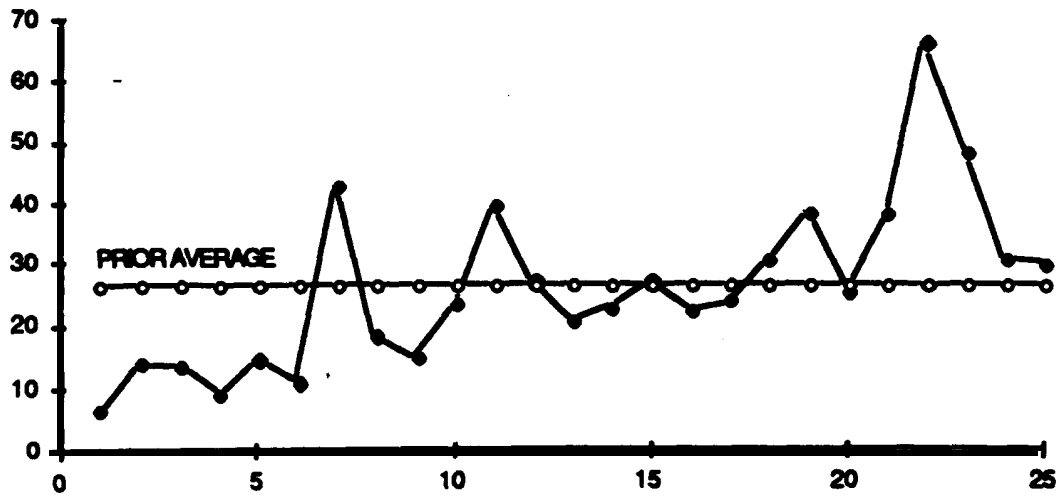
Figure 7.6   Example of a Pattern in the Data that May Indicate a Downward Trend
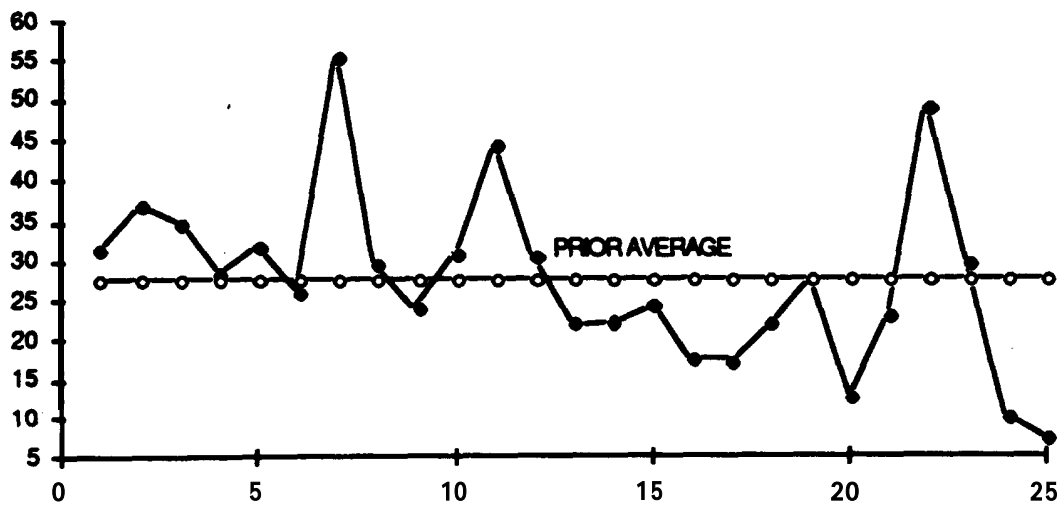
Figure 7.7   Example of Changing Variability in the Data Over Time
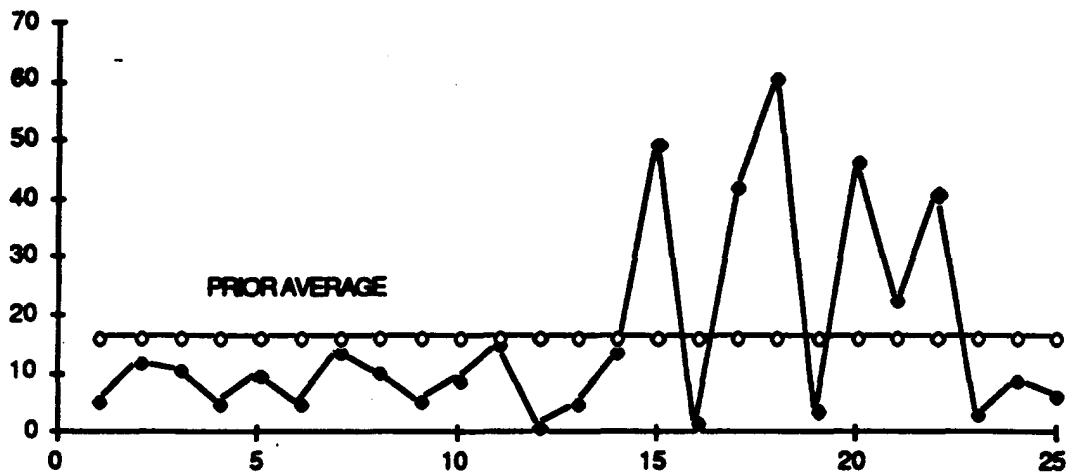
Figure 7.8   Example of a Stable Situation with Constant Average and Variation
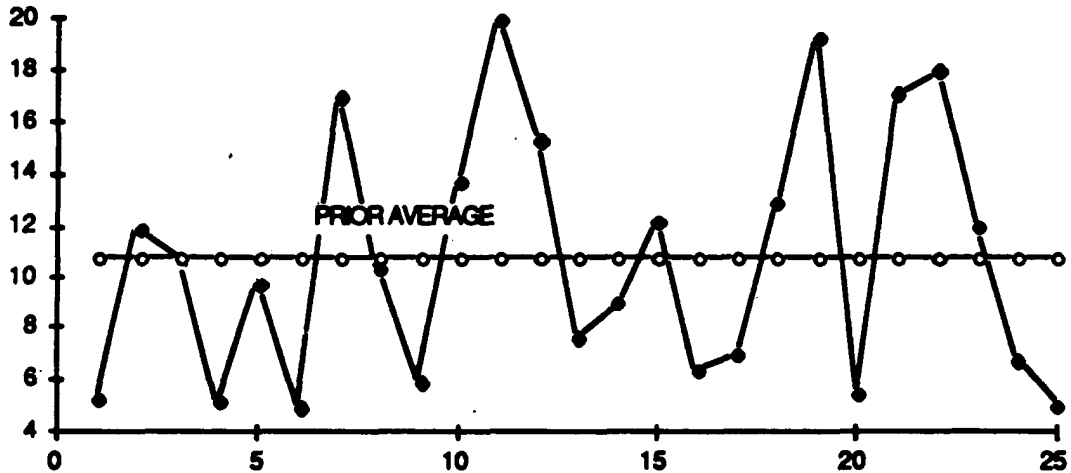
8. ASSESSING ATTAINMENT USING FIXED SAMPLE SIZE TESTS

After the remediation effort and after the ground water has achieved steady state, water samples can be collected to determine whether the contaminant concentrations attain the relevant cleanup standards. The sampling and evaluation period for making this attainment decision is represented by the unshaded portion in the figure below.

Figure 8.1 **Example Scenario for Contaminant Measurements During Successful Remedial Action**
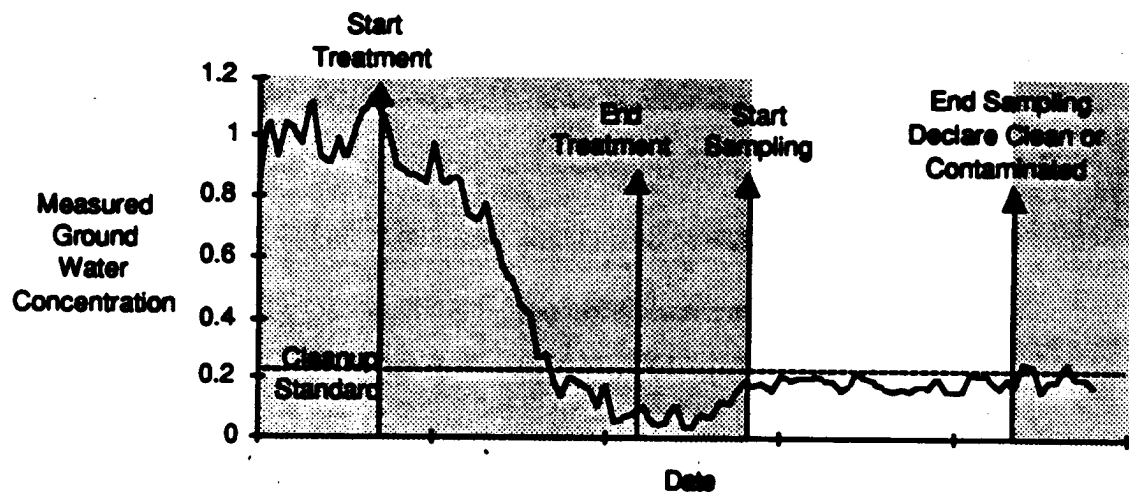
In this chapter statistical procedures are present for assessing the attainment of cleanup standards for ground water at Superfund sites. As discussed previously, the procedures presented arc suitable for assessing the time series of chemical concentrations measured in individual wells relative to a cleanup standard. Note that attainment objectives, as discussed in Chapter 3, must be specified by those managing the site remediation <u>before</u> the sampling for assessing attainment begins.

The collection of samples for assessing attainment of the cleanup standards will occur after the remedial action at the site has been completed and after a subsequent period has passed to allow transient affects due to the remediation to dissipate. This will allow the ground water concentrations, flows, and water table levels to reach equilibrium with the surrounding environment. It will be important to continue to chart the ground

water dam to monitor the possibility of unexpected departures from an apparent steady state. Some such departures are illustrated in Figures 7.3 through 7.7.

The attainment decision is an assessment of whether the post-cleanup contaminant concentrations are acceptable compared to the cleanup standard and whether they are likely to remain acceptable. To assess whether the contaminant concentrations are likely to remain acceptable, the statistical procedures provide methods for determining whether or not a long-term average concentration or a long-term percentage of the well water concentration measurements are below the established cleanup standards.

It is assumed in this chapter that the periodic or seasonal patterns in the data repeat on a yearly cycle. It may be that another, perhaps shorter, period of time would be appropriate. In such a case, the reference to "yearly" averages may be adjusted by the reader to reflect the appropriate period of time for the site under consideration. In the text, mention of alternative "seasonal cycles or periods" indicates where such adjustments may be appropriate.

This chapter presents statistical procedures for determining whether:

· **The mean concentration is below the Cleanup standards; or**

· A selected percentile of all samples is below the cleanup standard (e.g., does the 90th percentile of the distribution of concentrations fall below the cleanup standard?).

Many different statistical procedures can be used to assess the attainment of the cleanup standard. The procedures presented here have been selected to provide reasonable results with a small sample size in the presence of correlated dam. They require minimal statistical background and expertise. If other procedures are considered, consultation with a statistician is recommended. In particular, in the unlikely event that the measurements are not serially correlated, the methods presented in chapter 5 which assume a random sample can be used.

The procedures presented arc of two types: fixed sample size tests are discussed in this chapter, and sequential tests arc discus& in Chapter 9. Figure 8.2 is a flow chart outlining the steps involved in the cleanup process when using a fixed sample size test. Section 8.6 discusses testing for trends if the levels of contaminants are acceptable.

Figure 8.2   Steps in the Cleanup Process When Using a Fixed Sample Size Test
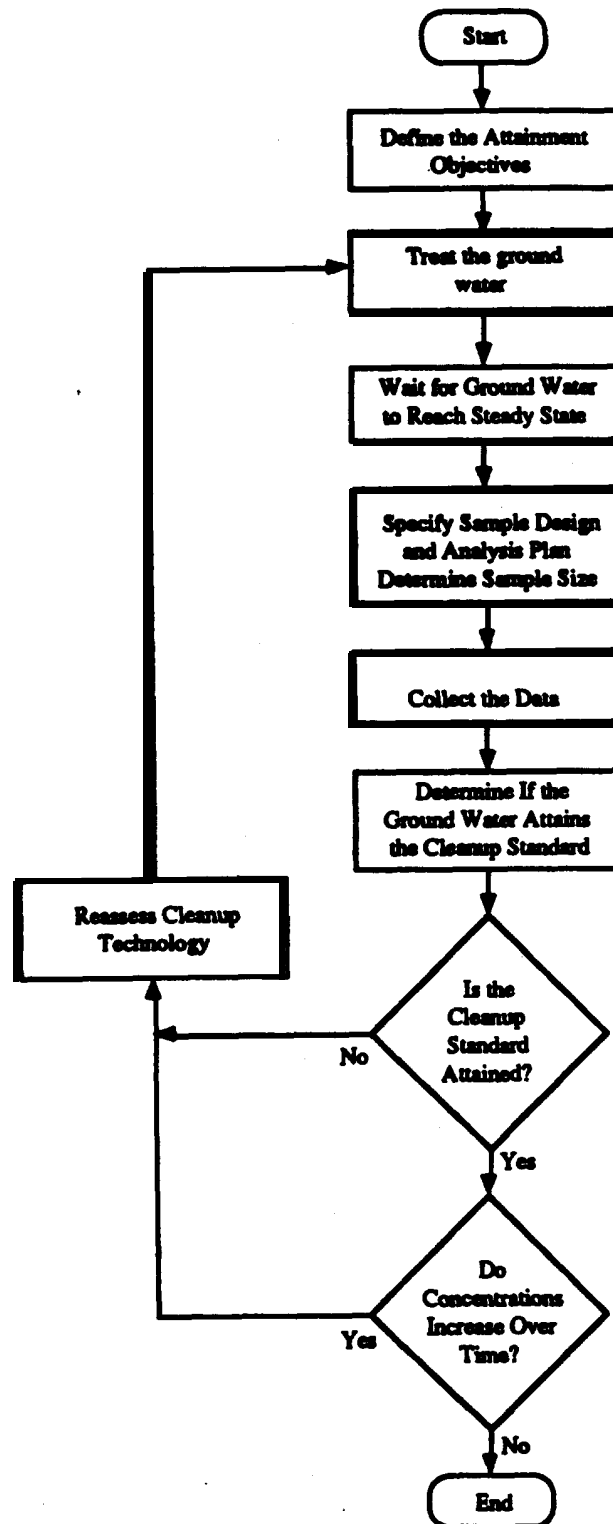
CHAPTER 8: ASSESSING ATTAINMENT USING FIXED SAMPLE SIZE TESTS

8.1   Fixed Sample Size Tests

This chapter discusses assessing the attainment of cleanup standards using a test based on-a predetermined sample size. For a fixed sample size test, the ground water samples are collected on a regular schedule, such as every two months, for a predetermined **number** of years. After all the data have been collected, the data are analyzed to determine whether the concentrations in the ground water attain the cleanup standard. Even if the initial measurements suggest that the ground water may attain the cleanup standard, all samples must be collected before the statistical test can be performed. An advantage of this approach is that the number of samples required to perform the statistical test will be known before the sampling begins, making some budgeting and planning tasks easier than when **using a sequential test (Chapter 9).**

Three procedures are presented for testing the mean when using fixed sample size tests. The first and second procedures use yearly averages concentrations. The first method, based on the assumption that the yearly means have a normal distribution, is recommended when there are missing values in the data and the missing values are not distributed evenly throughout the year. The second procedure assumes that the distribution of the yearly average is skewed, similar to a lognormal distribution, rather than symmetric. If there are few or no missing values, the second method using the log transformed yearly averages is recommended even if the data are not highly skewed. The third method requires calculation of seasonal effects and serial correlations to determine the variance of the mean. Because the third method is sensitive to the skewness of the data, it is recommended only if the distribution of the residuals is reasonably symmetric. Regardless of the procedure used, the sample size for assessing the mean should be determined using the steps described in Section 8.2.1.

8.2   Determining Sample Size and Sampling Frequency

Whether the calculation procedures used for assessing attainment use yearly averages or individual measurements, the formulas presented below for determining the required sample size use the characteristics of the individual observations. In the unlikely event that many years of observations are available for estimating the variance of yearly average, the number of years of sampling (using the same sample frequency as in the available data) can also be determined from the yearly averages using equation (5.35). The

following sections discuss the calculation of sample size for testing the mean and testing proportions.

8.2.1   Sample Size for Testing Means

The equations for determining sample size require the specification of the following quantities: Cs, μ_1, α, and β (see Sections 3.6 and 3.7) for each chemical under investigation. In addition, estimates of the serial correlation ϕ between monthly observations and the standard deviation σ of the measurements are required. For sample size determination, these quantities need not be precise. The procedures described in Section 5.10 and 5.3 may be used to obtain rough estimates of σ and the serial correlation.

The total number of samples to collect and analyze from each well is determined by selecting the frequency of sampling within a year or seasonal period and then determining the number of years or seasonal periods through which data must be collected. Given the values for Cs, μ_1, α, and β, the steps for determining sample size are provided in Box 8.1 and are discussed below in more detail.

Using previous data to estimate the serial correlation between observations separated by a month is discussed in Section 5.3. Since these estimates will not be exact, they will require the following adjustment before calculating the sample size: If the estimated correlation is less than or equal to 0.1, a serial correlation between monthly observations of 0.1 should be assumed when determining the frequency of sampling. The higher the serial correlation, the larger will be the recommended time interval between samples.

From cost records or budget projections, estimate the ratio of the annual overhead cost of maintaining sampling operations at the site to the unit cost of collecting, processing, and analyzing one ground water sample. Call this ratio $R. This ratio will be used to obtain a preliminary estimate of the sampling frequency.

Box 8.1
Steps for Determining Sample Size for Testing the Mean

(1) Determine the estimates of σ and ϕ which describe the data. Denote these estimates by $\hat{\sigma}$ and $\hat{\phi}$.

(2) Estimate the ratio of the annual overhead cost of maintaining sampling operations at the site to the unit cost of collecting process-ing, and analyzing one ground water sample. Call this ratio $\$_R$.

(3) **Based on the values of $\$_R$ and $\hat{\phi}$, use Appendix Table A.4 to deter-mine the approximate number, n_p, of samples to collect per year or seasonal period. The value n_p may be modified based on site-specific considerations, as discussed in the text.**

(4) The sampling frequency (i.e., the number of samples to be taken per year) is n_p or 4, whichever is larger. Denote this sampling frequency as n. Note that, under this rule, at least four samples per year per sampling well will be collected.

(5) For given values of n and $\hat{\phi}$, determine a "variance factor" from Appendix Table A.5. Denote this factor by F. For example, for $\hat{\phi} = 0.4$ and n = 12, the factor is F = 5.23.

(6) A preliminary estimate of the required number of years to sample, m_d, is

$$m_d = \frac{\hat{\sigma}^2}{F}\left\{\frac{z_{1-\beta} + z_{1-\alpha}}{Cs-\mu_1}\right\}^2 + 2 \qquad (8.1)$$

where $z_{1-\beta}$ and $z_{1-\alpha}$ are the critical values from the normal distribu-tion with probabilities of $1-\alpha$ and $1-\beta$ (Table A.2).

(7) The number of years of data will be denoted by m and will be determined by rounding m_d to the next highest integer. The total number of samples per well will be N=nm.

Appendix Table A.4 shows the approximate number of observations per year (or period) which will result in the minimum overall cost for the assessment (see Appendix F for the basis for Table A.4). Note that the sampling frequencies given in Table A.4 are approximate and are based on numerous assumptions which may only approximate the situation and costs at a particular Superfund site. Using the table requires knowledge of the serial cotrelations between observations separated by one month (or one-twelfth of the

seasonal cycle) and the cost of extending the sampling period for one more year relative to taking an additional ground water sample.

Find the column in Table A.4 that is closest to the estimate of $\$_R$ being used. Find the row which most closely corresponds to ϕ. Denote the tabulated value by n_p. For example, suppose that the cost ratio is estimated to be 25 and $\phi = 0.3$. Then from Table A.4 under the fifth column (ratio = 20), $n_p = 9$. Since the costs and serial correlations will not be known exactly, the sample frequencies in Table A.4 should be considered as suggested frequencies. They should be modified to a sampling frequency which can be reasonably implemented in the field. For example, if collecting a sample every month and a half ($n_p = 8$) will allow easy coordination of schedules, n_p can be changed from 9 to 8.

For determination of sample frequency, these quantities need not be precise. If there arc several compounds to be measured in each sample, calculate the sample frequency for each compound. Use the average sample frequency for the various compounds

It is recommended that at least four samples per year (or seasonal period) be collected to reasonably reflect the variability in the measured concentration within the year. Therefore, the sampling frequency (i.e., number of samples to be taken per year) is the maximum of four and n_p. Denote the sampling frequency by n. Note that, under this rule, at least four samples per year per sampling well will be collected.

As more observations per year arc collected, the number of years of sampling required for assessing attainment can be reduced. However, there arc limits to how much the sampling time can be reduced by increasing the number of observations per year. If the cost of collecting, processing, and analyzing the ground water samples is very small compared to the cost of maintaining the overall sampling effort many samples can be collected each year and the primary cost of the assessment sampling will be associated with maintaining the assessment effort until a decision is reached. On the other hand, if the cost of each sample is very large and a monitoring effort is to be maintained at the site regardless of the attainment decision, the costs of waiting for a decision may be minimal and the sampling frequency should be specified so as to minimize the sample collection, handling, and analysis costs. It should be noted that it is assumed that the ground water remains in steady state throughout the period of data collection.

The frequency of sampling discussed in this document is the simplest and most straightforward to implement: determine a single time interval between samples and select a sample at all wells of interest after that period of time has elapsed (e.g., once every month, once every six weeks, once a quarter, etc.). However, there are other approaches to determining sampling frequency, for example, site specific data may suggest that time intervals should vary among wells or groups of wells in order to achieve approximately the same precision for each well. Considering such approaches is beyond the scope of this document, but the interested reader may reference such articles as Ward, Loftis, Nielsen, and Anderson (1979), and Sanders and Adrian (1978). It should be noted that these articles arc oriented around issues related to sampling surface rather than ground water but many of the general principlesapply to both. In general, consultation with a statistician is recommended when establishing sampling procedures.

Use the sample frequency per year, the estimated serial correlation between monthly observations, and Appendix Table A.5 to determine a "variance factor" for estimating the required sample size. For the given values of n and ϕ, determine the variance factor in Table A.5. Denote this factor by F. For example, for $\phi = 0.4$ and n = 12, the factor is F = 5.23. For values of ϕ and n not listed in Table A.5, interpolation between listed values may be used to determine F. Alternatively, if a conservative approach is desired (i.e., to take a larger sample of data), take the smaller value of F associated with listed values of ϕ and n. For values outside the range of values covered in Table A.5, see Appendix F.

A preliminary estimate of the required number of years of sampling, m_d is given by equation (8.1). The first ratio in this equation is the estimated variance of the yearly average, $\hat{\sigma}_{\bar{x}}^2 = \frac{\hat{\sigma}^2}{F}$. The final addition of 2 to the sample size estimate improves the estimate with small sample sizes (see Appendix F).

Because the statistical tests require a full year's worth of data, the number of years of data collection, m_d, is rounded to the next highest integer, m. Thus, n samples will be collected in each of m years, for a total number of samples per well of N where N is the product m*n. An example of using these procedures to calculate sample size for testing the mean is provided in Box 8.2.

Box 8.2
Example of Sample Size Calculation for Testing the Mean

Suppose that, for $\alpha = .01$, it is desired to detect a difference of .2 ppm from the cleanup standard of .5 ppm (for example: $Cs = .5$, $\mu_1 = .3$) with a power of .80 (i.e., $\beta = .20$). Also suppose that the ratio of annual overhead costs to per-unit sampling and analysis costs ($\$_R$) is close to 10. Further, it is estimated that $\delta = .43$ and $\phi = .20$. Then for $\phi = .20$ and cost($\$_R$) = 10, Table A.4 gives $n_p = 9$. For $n_p = 9$ and $\phi = .20$, F = 7.17 from Table A.5. Further, using equation (8.1):

$$m_d = \frac{\delta^2}{F} \left\{ \frac{z_{1-\beta} + z_{1-\alpha}}{Cs-\mu_1} \right\}^2 + 2$$

to determine the number of years, m_d, to collect data, we find

$$m_d = \frac{.43^2}{7.17} \left\{ \frac{.842 + 2.326}{.5 - .3} \right\}^2 + 2 = 8.47,$$

where $z_{1-\beta} = .842$ and $z_{1-\alpha} = 2.326$, as can be found from Table A.2 or any normal probability table.

Rounding up gives a sampling duration of nine years and a total sample size of 9*9= 81 samples.

8.2.2   Sample Size for Testing Proportions

The testing of proportions is similar to the testing of means in that the average coded observation (e.g., the proportion of samples fop which the cleanup standard has been exceeded) is compared to a specified proportion. The method for determining sample size described below works well when there is a low con-elation between observations and no or small seasonal patterns in the data. If the correlation between monthly observations is high or there arc large seasonal changes in the measurements, then consultation with a statistician is recommended. If the parameter to be tested is the proportion of contaminated samples from either one well or an array of wells, one can determine the sample size for a fixed sample size test using the procedures in Box 8.3. These procedures for determining sample size require the specification of the following quantities: α, β, P_0, and P_1 (see Section 3.7 and Section 5.4.1). In general, many samples are required for testing when testing small proportions.

Box 8.3
Determining Sample Size for Testing Proportions

(1) Compute the estimates of σ and ϕ which describe the measurements (not the coded values). Denote this estimates by $\hat{\sigma}$ and $\hat{\phi}_m$. Let $\hat{\phi} = \frac{\hat{\phi}_m}{2.5}$, ($\hat{\phi}$ is the estimated correlation between the coded observations).

(2) Estimate the ratio of the annual overhead cost of maintaining sampling operations at the site to the unit cost of collecting, processing and analyzing one ground water Sample. Call this ratio $\$_R$.

(3) Based on the values of $\$_R$ and $\hat{\phi}$, use Table A.4 to determine the approximate number, n_p, of samples to collect per year or seasonal period. Based on site-specific considerations, the value n_p may be modified to a number which is administratively convenient.

(4) The sampling frequency (i.e., the number of samples to be taken per year) is n_p or 4, whichever is larger. Denote this sampling frequency as n. Note that, under this rule, at least four samples per year per sampling well will be collected.

(5) For given values of n and $\hat{\phi}$, determine a "variance factor" from Table A.5. Denote this factor by F.

(6) For given values of F, α, β, P_0, and P_1 a preliminary estimate of the number of years to sample is

$$m_d = \frac{1}{F} \left\{ \frac{z_{1-\beta}\sqrt{P_1(1-P_1)} + z_{1-\alpha}\sqrt{P_0(1-P_0)}}{P_0 - P_1} \right\}^2 \qquad (8.2)$$

where $z_{1-\beta}$ and $z_{1-\alpha}$ are critical values from the normal distribution associated with probabilities of $1-\alpha$ and $1-\beta$ (Appendix Table A.2). If m_d is less than $\frac{10}{n*P_0}$, use $m_d = \frac{10}{n*P_0}$ instead. Equation (8.2) is an adaptation of (8.1), using equation (5.25) of Chapter 5.

(7) The number of years of data will be denoted by m, and will be determined by rounding m_d to the next highest integer. The total number N of samples per well will be N=nm.

8.2.3   **An AlternativeMethod for Determining Maximum Sampling Frequency**

The maximum sampling frequency can be determined using the hydrogeo-logic parameters of ground water wells. The Darcy equation (Box 8.4) using the hydraulic conductivity, hydraulic gradient, and effective porosity of the aquifer. can be used to determine the horizontal component of the average linear velocity of ground water. This method is useful for determining the sampling frequency that allows sufficient time to pass between sampling events to ensure, to the greatest extent technically feasible, that there is a complete exchange of the water in the sampling well between collection of water samples. Although samples collected at the maximum sampling frequency may be independent in the physical sense, statistical independence is unlikely. Other factors such as the effect of contamination history, remediation, and seasonal influences can also result in correlations over time periods greater than that required to flush the well. As a result, we recommend that the sampling frequency be less than the maximum frequency based on Darcy's equation. Use of the maximum frequency can be approached only if estimated correlations based on ground-water samples are close to zero and the cost ratio, $\$_R$, is high. A detailed discussion of the hydrogeologic components of this procedure is beyond the scope of this document. For further information refer to <u>Practical Guide for Ground-Water Sampling</u> (Barcelona et al., 1985) or <u>Statistical Analysis of Ground-Water Monitoring Data a</u>t RCRA <u>Facilitie</u>s (U.S. EPA, 1989b).

Box 8.4
Choosing a Sampling Interval Using the Darcy Equation

The sampling frequency can be based on estimates using the average linear velocity of ground water. The Darcy equation relates ground water velocity (V) to effective porosity (NC), hydraulic gradient (i), and hydraulic conductivity (k):

$$V = \frac{(k*i)}{Ne} \qquad\qquad (8.3)$$

The values for k, i, and Ne can be determined from a well's hydrogeologic characteristics. The time required for ground water to pass through the well diameter can be determined by dividing the monitoring well diameter by the average linear velocity of ground water (V). This value represents the minimum   time interval required between sampling events which might yield an independent ground water sample.

8.3   Assessing Attainment of the Mean Using Yearly Averages

When using yearly averages for the analysis, the effects of serial correlation can generally be ignored (except for extreme **conditions** unlikely to be encountered in ground water). For the procedures discussed in this section, the variance of the observed yearly averages is used to estimate the variance of the ova-all average concentration. First, data are collected using the guidelines indicated in Chapter 4. Values recorded below the detection limit should be recorded according to the procedures in Section 2.3.7. Wells can be tested individually or a group of wells can be tested jointly. In the latter case, the data for the individual wells at each point in time are used to produce a summary measure (e.g., the mean or maximum) for the group as a whole.

Two calculation procedures for assessing attainment are described below. Both procedures use the yearly average concentrations. The first is based on the assumption that the yearly averages can be described by a (symmetric) normal distribution. This is based on a standard t-test described in many statistics books. The second procedure uses the log transformed yearly averages and is based on the assumption that the distribution of the yearly averages can be described by a (skewed) lognormal distribution. Because the second procedure performs well even when the data have a symmetric distribution, the second method is recommended in most situations. Only when there arc missing data values for which the sampling dates axe not evenly distributed throughout the year and there is also an apparent seasonal pattern in the data is the first procedure recommended.

The calculations and procedures when using the untransformed yearly averages are described below and summarized in Box 8.5. This procedure is appropriate in all situations but is not preferred, particularly if the data axe highly skewed. The calculations can be used (with some minor loss in efficiency) if a some observations are missing. If the proportion of missing observations varies considerably from season to season and there are differences in the average measurements among seasons, consultation with a statistician is recommended. If there arc few missing values and the data arc highly skewed, the procedures described in Box 8.12 which use the log transformed yearly averages are recommended.

Box 8.5
Steps far Assessing Attainment Using Yearly Averages

(1) Calculate the yearly averages (see Box 8.6)

(2) Calculate the mean, $\bar{\bar{x}}_m$, and variance, $s_{\bar{x}}^2$, of the yearly averages (see Box 8.7)

(3) If there are no missing observations, set

$$\bar{x} = \bar{\bar{x}}_m \qquad\qquad (8.4)$$

Otherwise, if there are missing observations calculate the seasonal averages and the mean of the seasonal averages, $\bar{\bar{x}}_{ms}$, (Box 8.8) and set

$$\bar{x} = \bar{\bar{x}}_{ms} \qquad\qquad (8.5)$$

where $\bar{x}$ is the mean to be compared to the cleanup standard.

(4) Calculate the upper 1-α percent one-sided confidence interval for the mean, $\bar{x}$. (Box 8.9)

(5) Decide whether the ground water attains the cleanup standards (Box 8.10).

Use the formulas in Box 8.6 for calculating the yearly averages. If there ate missing observations within a year, average the non-missing observations. Using the yearly averages for the statistical analysis, calculate the mean and variance of the yearly averages using the equations in Box 8.7. The variance will have degrees of freedom equal to one less than the number of years over which the data was collected.

If there are no missing observations, the mean of the yearly averages, $\bar{\bar{x}}_m$, will be compared to the cleanup standard for assessing attainment. If however, there are missing observations, the mean of the yearly averages may provide a biased estimate of the average concentration during the sample period. This will be true if the missing observations occur mostly at times when the concentrations are generally higher or lower than throughout most of the year. To correct for this bias, the average of the seasonal averages will be compared to the cleanup standard when there are missing observations. Box 8.8 provides equations for calculating the seasonal averages and $\bar{\bar{x}}_{ms}$, the mean of the seasonal averages. Using $\bar{x}$ to designate the mean which is to be compared to the cleanup standard; set $\bar{x} = \bar{\bar{x}}_m$ if there are no missing observations, otherwise set $\bar{x} = \bar{\bar{x}}_{ms}$.

Box 8.6
Calculation of the Yearly Averages

Let x_{jk} = the measurements from an individual well or a combined measure from a group of wells obtained for testing whether the mean attains the cleanup standard; x_{jk} represents the concentration for season j (the jth sample collection time out of n) in year k (where data is collected for m years.

For each year, the yearly average is the average of all of the observations taken within the year. If the results for one or more sample times within a year are missing, calculate the average of the non-missing observations.

If there are n_k ($n_k \leq n$) non-missing observations in year k, the yearly average, $\bar{x}_k$, is:

$$\bar{x}_k = \frac{1}{n_k} \sum_{j=1}^{n_k} x_{jk}.$$ (8.6)

where the summation is over all non-missing observations within the year.

Box 8.7
Calculation of the Mean and Variance of the Yearly Averages

The mean of the yearly averages, $\bar{\bar{x}}_m$ is:

$$\bar{\bar{x}}_m = \frac{1}{m} \sum_{k=1}^{m} \bar{x}_k$$ (8.7)

where $\bar{x}_k$ is the yearly average for year k and the summation covers m years.

The variance of the yearly averages, $s_{\bar{x}}^2$, can be calculated using either of the two equivalent equations below:

$$s_{\bar{x}}^2 = \frac{\sum_{k=1}^{m} \bar{x}_k^2 - \frac{1}{m}\left(\sum_{k=1}^{m} \bar{x}_k\right)^2}{(m-1)} = \frac{\sum_{k=1}^{m}\left(\bar{x}_k - \bar{\bar{x}}_m\right)^2}{(m-1)}$$ (8.8)

This variance estimate has m-1 degrees of freedom.

Box 8.8
Calculation of Seasonal Averages and the Mean of the Seasonal Averages

For the n   sample collection times within the year, the jth seasonal average is the average of all the measurements taken at the jth collection time. If there is a missing observation, the measurement from the jth sample collection time may be different from the jth sequential measurement within the year. Note that observations below the detection limits should be replaced by the detection limit and are not counted as missing observations.

For all collection times j, from 1 to n, within each year, calculate the seasonal average, $\bar{x}_j$, where the number of observations at the jth collection time is $m_j \leq m$. If there are missing observations, sum over the m_j non-missing observations.

$$\bar{x}_j = \frac{1}{m_j} \sum_{k=1}^{m_j} x_{jk}. \qquad\qquad (8.9)$$

The mean of n seasonal averages is:

$$\bar{\bar{x}}_{ms} = \frac{1}{n} \sum_{j=1}^{n} \bar{x}_j \qquad\qquad (8.10)$$

Using the mean which is to be compared to the cleanup standard, $\bar{x}$, and the standard deviation of the mean calculated from the yearly averages, calculate the upper one sided 1-α percent confidence interval for the mean using equation 8.11 in Box 8.9. The standard deviation is the square root of the variance calculated from equation (8.7). Calculation of the upper confidence interval requires use of α, specified in the attainment objectives, and the degrees of freedom for the standard deviation, the number of years of data minus one, to determine the relevant t-statistic from Table A.1 in Appendix A. If the lower one-sided confidence limit is desired, replace the plus sign in equation (8.11) with a minus sign.

Finally, if the upper one-sided confidence interval is less than the cleanup standard and if the concentrations are not increasing over time, decide that the tested ground water attains the cleanup standard. If the ground water from all wells or groups of wells attains the cleanup standard then conclude that the ground water at the site attains the cleanup standard. The steps in deciding attainment of the cleanup standard are shown in Box 8.10.

Box 8.9
Calculation of Upper One-sided Confidence Limit for the Mean

The upper one-sided confidence limit is:

$$\mu_{U\alpha} = \bar{x} + t_{1-\alpha,m-1} \frac{s_{\bar{x}}}{\sqrt{m}} = \bar{x} + t_{1-\alpha,m-1}\, s_{\bar{x}_m} \qquad (8.11)$$

where $\bar{x}$ is the mean level of contamination, and $s_{\bar{x}}$ is the square root of the variance of the yearly means. The degrees of freedom associated with $s_{\bar{x}}$ is $m-1$, and the appropriate value of $t_{1-\alpha,m-1}$ can be obtained from Table A.1.

Box 8.10
Deciding if the Tested Ground Water Attains the Cleanup Standard

If $\mu_{U\alpha} < C_s$, conclude that the average ground water concentration in the well (or group of wells) attains the cleanup standard.

If the average ground water concentration in the wells is less than the cleanup standard, perform a trend test using the regression techniques described in Chapter 6 to determine if there is a statistically significant increasing trend to the yearly averages over the sampling period (also see Section 8.6). Note that at least 3 years' worth of data are required to identify a trend. If there is not a statistically significant increasing trend conclude that the ground water attains the cleanup standard (and possibly initiate a follow-up monitoring program). If a significant trend does exist, resume sampling or reconsider treatment effectiveness.

If $\mu_{U\alpha} \geq C_s$, conclude that the average ground water concentration in the wells does not attain the cleanup standard.

When the data are noticeably skewed, the calculation procedures in Box 8.12 (using the log transformed yearly averages) are recommended over those in Box 8.5. Because the procedures in Box 8.12 also perform well when the data have a symmetric **distribution, these procedures are** generally recommended in all cases where there are no missing data. There is no easy adjustment for missing data when using the log transformed yearly averages. Therefore, if the number of observations per season (month etc.) is not the same for all seasons and if there is any seasonal pattern in the data, use of the procedures in Box 8.5 is recommended

Box 8.11
Sample of Assessing Attainment of the Mean Using Yearly Averages

To test whether the cleanup standard (Cs = 0.50) has been attained for a particular chemical, 48 ground water samples were collected for four years at monthly intervals. All 48 ground water samples were collected and analyzed, and three values which were below the detection level were replaced in the analysis by the detection limit. Based on the sample data, the overall mean concentration was determined to be .330 ppb. The corresponding yearly means were computed as: $\bar{x}_1 = .31$; $\bar{x}_2 = .32$; $\bar{x}_3 = .34$; and $\bar{x}_4 = .35$. The variance of the yearly means is $s_{\bar{x}}^2 = .000333$.

The one-sided 99 percent confidence interval extends from zero to

$$\mu_{U\alpha} = \bar{x} + t_{1-\alpha,3} \frac{s_{\bar{x}}}{\sqrt{m}} = .33 + 4.541 \frac{.018}{\sqrt{4}} = .37 \text{ ppb.}$$

Since the cleanup standard is Cs = 0.5 ppm the average is significantly less than the cleanup standard. However, the yearly averages are consistently increasing and regression analysis indicates that the trend is statistically significant at the 5 percent level (p = .0101). Therefore, it cannot be concluded that the attainment objectives have been achieved. If the present trend continues, the concentrations would exceed the cleanup standard in about 10 years. Possible options include continued monitoring to determine if the trend will continue or to reassess the treatment effectiveness and why the upward trend exists.

The calculations when using the log transformed yearly averages are slightly more difficult than when using the untransformed yearly averages. After calculating the yearly averages, the natural log is used to transform the data. The transformed averages are then used in the subsequent analysis. The upper confidence interval for the mean concentration is based on the mean and variance of the log transformed yearly averages. The formulas are based on the assumption that the yearly averages have a log normal distribution.

Box 8.12
Steps far Assessing Attainment Using the Log Transformed Yearly Averages

(1) ⁻ Calculate the yearly averages (see Box 8.6)

(2) Calculate the natural log of the yearly averages (see Box 8.13)

(3) Calculate the mean, $\bar{\bar{x}}_m$, and variance, $s_{\bar{x}}^2$, of the log transformed yearly averages (see Box 8.14)

(4) Calculate the upper 1-α percent one-sided confidence interval for the overall mean. (Box 8.15)

(5) Decide whether the ground water attains the cleanup standards (Box 8.10).

Use the formulas in Box 8.6 for calculating the yearly averages. If there are missing observations within a year, average the non-missing observations. Calculate the log transformed yearly averages using equation (8.12) in Box 8.13. The natural log transformation is available on many calculators and computers, usually designated as "LN", "1n", or "log$_e$." Although the equations could be changed to use the base 10 logarithms, use only the base e logarithms when using the equations in Boxes 8.13 through 8.15. Calculate the mean and variance of the log transformed yearly averages using the equations in Box 8.14. The variance will have degrees of freedom equal to one less than the number of years over which the data was collected.

Box 8.13
Calculation of the Natural Logs of the Yearly Averages

The natural log of the yearly average is:

$$y_k = \ln(\bar{x}_k) \tag{8.12}$$

Box 8.14
Calculation of Mean and Variance of the Natural Logs of the Yearly

The average of the m log transformed yearly averages, $\bar{y}_m$:

$$\bar{y}_m = \frac{1}{m} \sum_{k=1}^{m} y_k \tag{8.13}$$

The variance of the log transformed yearly averages, $s_{\bar{y}}^2$:

$$s_{\bar{y}}^2 = \frac{\sum_{k=1}^{m} y_k^2 - \frac{1}{m}\left(\sum_{k=1}^{m} y_k\right)^2}{(m-1)} = \frac{\sum_{k=1}^{m}(y_k - \bar{y}_m)^2}{(m-1)} \tag{8.14}$$

This variance estimate has m-1 degrees of freedom.

Calculate the upper one sided **l-a percent** confidence interval for the mean using equation 8.x in Box 8.15. **Calculation of the upper confidence interval requires use** of a, specified in the attainment objectives, and the degrees of freedom for the standard **deviation, the number of years of data minus one, to determine the relevant t-statistic from** Table A.2 in Appendix A. If the lower one-sided confidence limit is desired, replace the second plus sign in equation (8.15) with a minus sign.

Finally, if the upper one-sided confidence interval is less than the cleanup **standard and if the log transformed concentrations arc not increasing over time, decide that** **the tested ground water attains the cleanup standard If the ground water from all wells or** **groups of wells attains the cleanup standard then conclude that the ground water at the site** **attains the cleanup standard. The steps in deciding attainment of the cleanup standard are** shown in Box 8.10.

Box 8.15
Calculation of the Upper Confidence Limit for the Mean Based on Log Transformed Yearly **Averages**

The upper one-sided confidence limit for the mean is:

$$\mu_{U\alpha} = \exp\left(\bar{y}_m + \frac{s_{\bar{y}}^2}{2} + t_{1-\alpha;m-1} \sqrt{\frac{s_{\bar{y}}^2}{Df+1} + \frac{s_{\bar{y}}^4}{2\,Df}} \right) \quad (8.15)$$

where the degrees of freedom (Df) associated with $s_{\bar{y}}$ is m-1, and the appropriate value of $t_{1-\alpha,m-1}$ can be obtained from Table A.1. The term under the square root is the variance of $\bar{y}_m + \frac{s_{\bar{y}}^2}{2}$ and was calculated from the variance of the two terms, which are independent if the data have a lognormal distribution.

8.4 Assessing Attainment of the Mean After Adjusting for Seasonal Variation

This section provides an alternative procedure for testing the mean concentration. It is expected to provide more accurate results with large sample sizes, correlated data, and data which is not skewed. Because this procedure is sensitive to skewed data, it is recommended only, if the distribution of the residuals is reasonably symmetric.

After the data have been collected using the guidelines indicated in Chapter 4, wells can be tested individually or a group of wells can be tested jointly. In the latter case, the data for the individual wells at each point in time arc used to produce a summary measure for the group as a whole. This summary measure may be an average, maximum, or some other measure (see Section 2.35). These summary measures will be averaged over the entire sampling period. The tests for attainment and the corresponding calculations required when removing seasonal averages arc described next.

The calculations and procedures when using the mean adjusted fop seasonal variation arc described below and summarized in Box 8.16. This procedure is not recommended if the data are noticeably skewed. The following calculations and procedures are appropriate if the number of observations per year is the same far all years. However, they

can still be used (with some minor loss in efficiency) if a few observations are lost as long as the loss is not concentrated in a particular season (note example in Section 8.3). If the proportion of observations varies considerably from season to season, consultation with a statistician is recommended. If the data are obviously skewed, the procedures described in Box 8.15 which use the log transformed yearly averages are recommended.

Box 8.16
Steps for Assessing Attainment Using the Mean After Adjusting for Seasonal Variation

(1) Calculate the seasonal averages and the mean of the seasonal averages, $\bar{x}_{ms}$, (Box 8.8)

(2) Calculate the deviations from the seasonal averages (residuals) (Box 8.17)

(3) Calculate the variance, s_x^2 of the residuals (see Box 8.18)

(4) Calculate the lag 1 serial correlation of the residuals using equation (8.18) in Box 8.19. Denote the computed serial correlation by $\hat{\phi}_{obs}$.

(5) Calculate the upper 1-α percent one-sided confidence interval for the mean, $\bar{x}$. (Box 8.20)

(6) Decide whether the ground water attains the cleanup standards (Box 8.10).

Use the formulas in Box 8.8 for calculating the seasonal averages and the mean of the seasonal averages. If there are missing observations within a season, average the non-missing observations. Calculate the residuals, the deviations of the measurements from the respective seasonal means using equation (8.16) in Box 8.17. Box 8.18 shows how to calculate the variance of the residuals. The variance will have degrees of freedom equal to the number of measurements less the number of seasons. Calculate the serial correlation of the residuals using equation (8.18) in Box 8.19. If the serial correlation is less than zero, use zero when calculating the confidence interval.

Box 8.17
Calculation of the Residuals

From each sample observation, subtract the corresponding seasonal mean. That is, compute the e_{jk}, the deviation from the mean:

$$e_{jk} = x_{jk} - \bar{x}_{j}. \qquad (8.16)$$

Box 8.18
Calculation of the Variance of the Residuals

Calculate the variance of the residuals e_{jk} after adjustments for possible seasonal differences:

$$s^2 = \frac{1}{N-m} \sum_{k=1}^{m} \sum_{j=1}^{n} e_{jk}^2. \qquad (8.17)$$

Alternatively, the ANOVA approach described in Appendix D can be used to compute the required variance.

Box 8.19
Calculating the Serial correlation from the Residuals After Removing Seasonal Averages

The sample estimate of the serial correlation of the residuals is:

$$\hat{\phi}_{obs} = \frac{\sum_{i=2}^{N} e_i e_{i-1}}{\sum_{i=1}^{N} e_i^2} \qquad (8.18)$$

Where e_i, $i = 1, 2, ...,N$ are the residuals after removing seasonal averages, in the time order in which the samples were collected.

Using the mean of the seasonal averages and the standard deviation of the mean, calculated from the residuals, calculate the upper one sided 1-α percent confidence interval for the mean using equation (8.19) in Box 8.20. The standard deviation is the square root of the variance calculated from equation (8.17). If the observed serial correlation is less than zero, use zero in equation (8.19). Calculation of the upper confidence interval requires use of α, specified in the attainment objectives, and the degrees of freedom for the standard deviation, the number of years of data minus one, to determine the relevant t-statistic from Table A.2 in Appendix A. If the lower one-sided confidence limit is desired, replace the plus sign in equation (8.19) with a minus sign.

Box 8.20
Calculation of the Upper Confidence Limit for the Mean After Adjusting for Seasonal Variation

Calculation of the Upper One-Sided Confidence Limit

$$\mu_{U\alpha} = \bar{X} + t_{1-\alpha,Df}\frac{s}{\sqrt{N}}\sqrt{\frac{1+\hat{\phi}_{obs}}{1-\hat{\phi}_{obs}}} \qquad (8.19)$$

where $\bar{X}$ is the computed mean level of contamination computed from equation (8.8), and s is the square root of the variance of the observations taking into account possible seasonal variation as computed from equation (8.17). The degrees of freedom, Df, associated with s is $Df = \frac{N-m}{3}$, and the appropriate value of $t_{1-\alpha,Df}$ can be obtained from Table A.1. If $\hat{\phi}_{obs}$ is less than zero, set $\hat{\phi}_{obs}$ to zero. For the derivation of the term under the square root, see Appendix F.

Box 8.21
Example calculation of confidence Intervals

Table 8.1 and Figure 8.3 show hypothetical arsenic measurements for ground water samples taken at quarterly intervals for four years. For these data, the four seasonal (quarterly) means are: $\bar{x}_1 = 6.688$; $\bar{x}_2 = 6.013$; $\bar{x}_3 = 5.078$; and $\bar{x}_4 = 5.878$, and the overall mean is $\bar{x} = 5.914$ ppb. The adjusted arsenic measurements labeled "residuals," shown in the last column of the table, are obtained by subtracting the seasonal means from the original observations.

The estimated variance of the data, taking into account possible seasonal differences, is $s^2 = \frac{1.95}{12} = .163$ (equation (8.11)) with 4 $\left(\text{i.e.} \ \frac{N-m}{3} = \frac{16-4}{3} \right)$ degrees of freedom, and the corresponding auto correlation is $\hat{\phi}_{obs} = .37$ (eq. 8.18).

The upper one-sided 90 percent confidence interval extends from zero to

$$5.914 + 1.533 \frac{\sqrt{.163}}{\sqrt{16}} \sqrt{\frac{1+.37}{1-.37}} = 6.142 \text{ ppb.}$$

If the cleanup standard were 6 ppb, it would be concluded that the ground water has not attained the cleanup standard.

Figure 8.3   Plot of Arsenic Measurements for 16 Ground Water Samples (see Box 8.21)
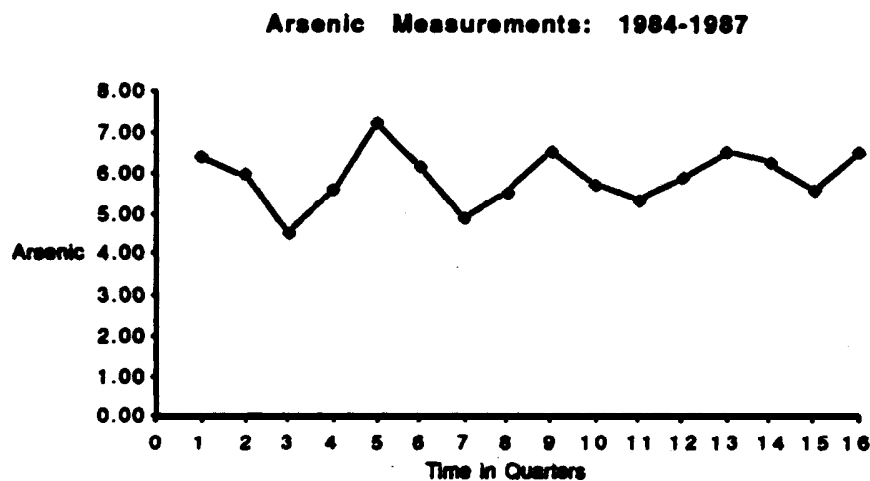
Table 8.1   Arsenic measurements (ppb) for 16 ground water samples (see Box 8.21)

Year	Quarter	Arsenic Measurement	Residual
1984	1	6.40	-.288
1984	2	5.91	-.103
1984	3	4.51	-.568
1984	4	5.57	-.308
1985	1	7.21	.522
1985	2	6.19	.177
1985	3	4.89	-.188
1985	4	5.51	-.368
1986	1	6.57	-.118
1986	2	5.70	-.313
1986	3	5.32	.242
1986	4	5.87	-.008
1987	1	6.57	-.118
1987	2	6.25	.237
1987	3	5.59	.512
1987	4	6.56	.682

8.5   Fixed Sample Size Tests for Proportions

If the parameter to be tested is the porportion of contaminated samples from either one well or an array of wells, the sample collection and analysis procedures are the same as those outlined above for testing the mean with the following changes:

·   To apply this nonparametric test, each measurement is either coded "1" (the actual measurement was equal to or above the relevant cleanup standard Cs), or "0" (below Cs). The statistical analysis is based on the resulting coded variable of O's and 1's.

·   Only the analysis procedure which used yearly averages, in Box 8.6 is appropriate for the calculations. Do not use the calculation procedures which correct for the seasonal pattern in the data and the serial correlation of the residuals or which use the log transformed data.

·   See Section 8.22 for procedures far estimating the sample size.

8.6   Checking for Trends in Contaminant Levels After Attaining the Cleanup Standard

Once a fixed sample size statistical test indicates that the cleanup standard for the site has been met, there remains one final concern. The model we have used assumes that ground water at the site has reached **a** steady state and that there is no reason to believe that contaminant levels will rise above the cleanup standard in the future. We need to check this assumption. Regression models, as discussed in Chapter 6, can be used to do so. By establishing a simple regression model with the contaminant measure as the dependent variable and time as the independent variable, a test of significance can be made as to whether or not the estimated slope of the resulting linear model is positive (see Section 6.1.3). Scatter plots of the data will prove useful in assessing the model. When using the yearly averages, the regression can be performed without adjusting for serial correlation.

To minimize the chance of incorrectly concluding that the concentrations are increasing over time, we recommend that the alpha level for testing the slope (and selecting the t statistic in Box 6.11) be set at a small value, such as 0.01 (one percent). If, on the basis of the test, there is not significant evidence that the slope is positive, then the evidence is consistent with the preliminary conclusion that the ground water in the well(s) attains the cleanup standard. If the slope is significantly greater than zero, then the concern that contaminant levels may later exceed the cleanup standard still exists and the assumption of a steady state is called into question. In this case, further consideration must be given to the reasons for this apparent increase and, perhaps, to additional remediation efforts.

8.7   Summary

This chapter presented the procedures for assessing attainment of the cleanup standards for ground water measurements using a fixed sample size test. The testing procedures can be applied to samples from either individual wells or wells tested as a group. These procedures are used after the ground water has achieved steady state. Both parametric and nonparametric methods for evaluating attainment arc discussed. If the ground water at the site is judged to attain the cleanup standards because the concentrations are not increasing and the long-term average is significantly less than the cleanup standard, follow-up monitoring is recommended to check that the steady state assumption holds.

9. ASSESSING ATTAINMENT USING SEQUENTIAL TESTS

After the remediation effort has been terminated and the ground water has achieved steady state, ground water samples can be collected to determine whether the resulting concentrations of contaminants attain the relevant cleanup standard The sampling and evaluation period making this attainment decision is represented by the unshaded portion in the figure below.

Figure 9.1   Example for Contaminant Measurements During Successful
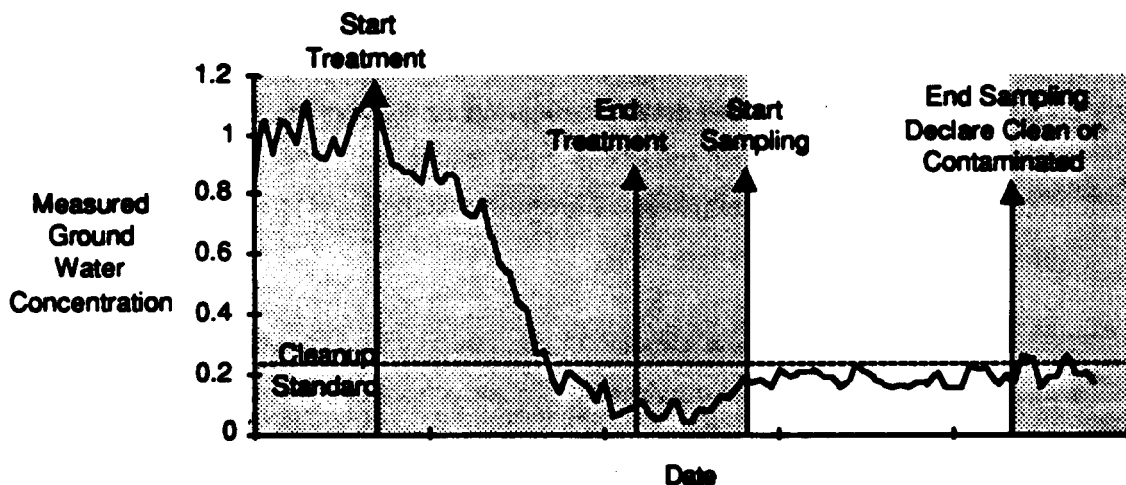
In this chapter statistical procedures are presented for assessing the attainment of cleanup standard for ground water at Superfund sites using sequential statistical tests. Note that attainment **objectives,**as discussed in Chapter 3, must be specified before the sampling for assessing attainment begins.

The collection of samples for assessing attainment of the cleanup standards will occur after the remedial action at the site has been completed and after a subsequent period has passed to allow transient affects due to the remediation to dissipate. The attainment decision is an assessment of whether the remaining contaminant concentrations are acceptable compared to the cleanup standard and whether they are likely to remain accept-

able. To assess whether the contaminant concentrations are likely to remain acceptable, the statistical procedures provide methods for determining whether or not a long-term average concentration or a long-term percentage of the well water concentration measurements are below the established cleanup standards. In particular, in the unlikely event that the measurements are not serially correlated, the methods presented in chapter 5, which assume a random sample, can be used and consultation with a statistician is recommended. If sequential tests are being considered, note that on the average, the sequential tests will require fewer samples than the fixed sample size tests in Chapter 8 or, if applicable, those in chapter 5.

This chapter discusses assessing the attainment of cleanup standards using a sequential statistical test. For a sequential test, the ground water samples are collected on a regular schedule, such as every two months. Starting after the collection of three years of data, a statistical test is performed every year to determine whether (1) the ground water being sampled attains the cleanup standard, or (2) the ground water does not attain the cleanup standard, or (3) more data are required to make a decision. If more data are required, another year's worth of data is collected before the next statistical test is performed. Figure 9.2 is a flow chart outlining the steps involved in the cleanup process when using a sequential statistical test.

Unlike the fixed sample size test, the number of samples required to reach a decision using the sequential test is not known at the beginning of the sampling period. On the average, the sequential tests will require fewer samples and a corresponding shorter time to make the attainment decision than for the tests in Chapter 8. If the ground water clearly attains the cleanup standard, the sequential test will almost always require fewer samples than a fixed sample size test. Only when the contaminant concentrations are less than the cleanup standard and greater than the mean for the alternate hypothesis might the sequential test be likely to require more samples than the fixed sample size test.

This chapter presents statistical procedures for determining whether:

·   The mean concentration is below the cleanup standard; or

·   A selected percentile of all samples is below the cleanup standard (e.g., does the 90th percentile of the distribution of concentrations fall below the cleanup standard?).

Figure 9.2   Steps in the Cleanup Process When Using a Sequential Statistical Test
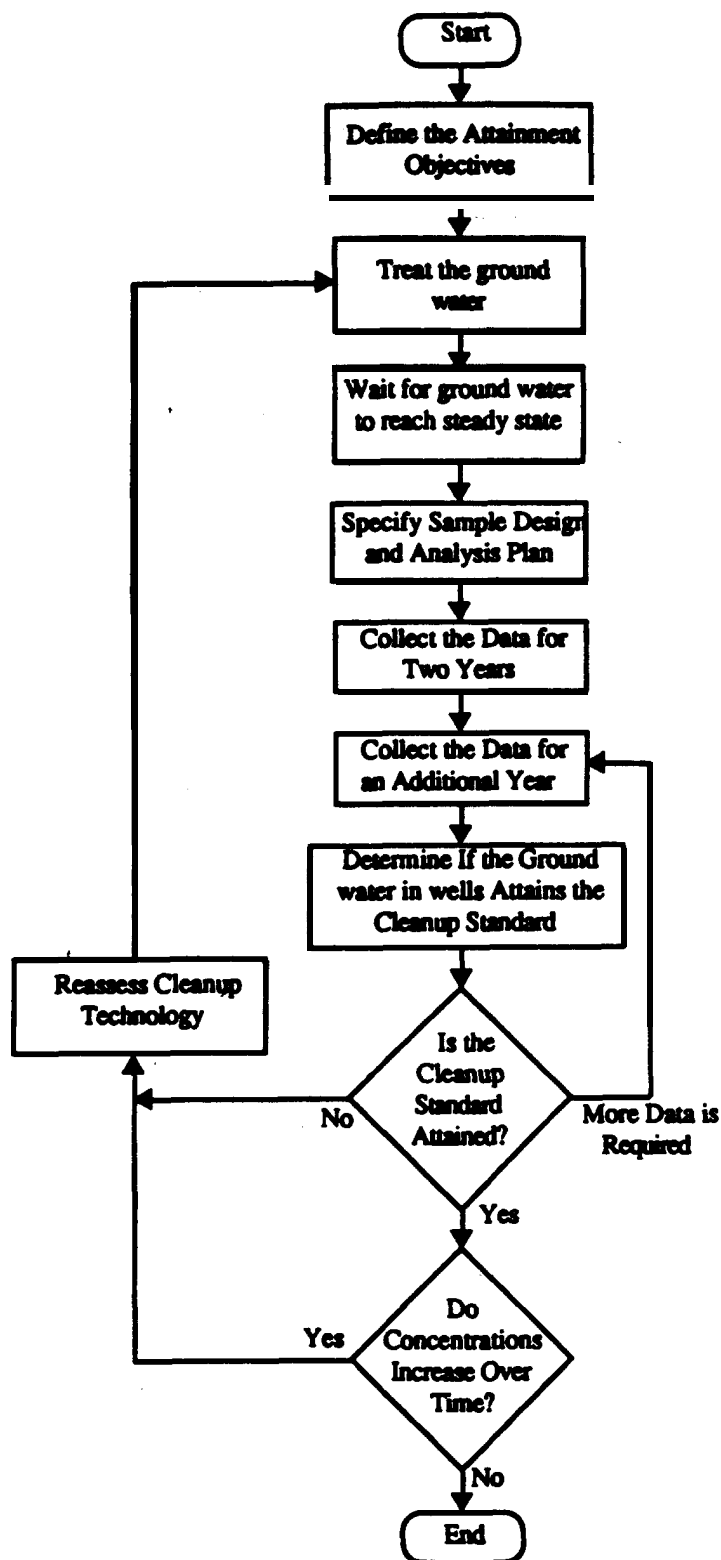
The measured ground water concentrations may fluctuate over time due to many factors including:

· Seasonal and short-term weather patterns affecting the ground water levels and flows;

· Variation in ground water concentrations due to historical fluctuations in the contamination introduced the ground water; and

· Sampling errors and laboratory measurement errors and fluctuations.

The effects of periodic seasonal fluctuations in concentration can be eliminated from the analysis, resulting in a more precise statistical test, by either averaging the measurements over a year or correcting for any seasonal patterns found in the data These two statistical analysis procedures arc presented in sections 9.3 and 9.4, respectively. The method of using yearly averages is, in general, easier to implement and preferred. Correcting for the seasonal pattern may provide more precise statistical tests in situations where large correlations exist between measurements and when the measurement errors have a symmetric distribution.

Three procedures are presented for testing the mean when using sequential tests. The first and second procedures use yearly average concentrations. The first method, based on the assumption that the yearly means have a normal distribution, is recommended when there are missing values in the data and the missing values are not distributed evenly throughout the year. The second procedure assumes that the distribution of the yearly average is skewed, similar to a lognormal distribution, rather than symmetric. If there are no missing values, the second method using the log transformed yearly averages is recommended even if the data are not highly skewed. The third method requires calculation of seasonal effects and serial correlations to determine the variance of the mean. Because the third method is sensitive to the skewness of the data, it is recommended only if the distribution of the residuals is reasonably symmetric. Regardless of the procedure used, the sample frequency for assessing the mean should be determined using the steps described in Section 9.1.

These sequential procedures arc an adaptation of Wald's sequential probability ratio test, specifically a version of the sequential t-test They assume that the data is normally distributed or can be made so by a log transformation. See Hall (1962). Hayre (1983). and Appendix F for details.

9.1 Determining Sampling Frequency for Sequential Tests

The ground water samples will be collected at regular intervals using a systematic sample with a random start as described in Chapter 4. An important part of determining the sample collection procedures is to select the time interval between samples or the number of samples to collect per seasonal period usually per year. Asdiscussed in Chapter 8, the term "year" will be used to mean a full seasonal cycle, which in most cases can be considered a calendar year.

The steps for determining sample frequency when testing the mean are provided in Box 9.1 and are discussed in Section 8.2 in more detail. The procedures for determining sample frequency require the specification of the serial correlation, ϕ, and the measurement error, σ, for the chemical under investigation. The procedures described in Section 5.3 may be used to obtain rough estimates of the serial correlation. Denote these estimates by $\hat{\phi}$. An example of calculating sample frequency is presented in Box 9.3.

Box 9.1
Steps for Determining Sample Frequency for Testing the Mean

(1) Determine the estimates of σ and ϕ which describe the data. Denote these estimates by $\hat{\sigma}$ and $\hat{\phi}$.

(2) Estimate the ratio of the annual overhead **cost** of maintaining sampling operations at the site to the unit cost of collecting processing, and analyzing one ground water sample. Call this ratio $\$_R$.

(3) Based on the values of $\$_R$ and $\hat{\phi}$, use Appendix Table A.4 to determine the approximate number, n_p, of samples to collect per year or seasonal period. The value n_p may be modified based on site-specific considerations, as discussed in the text.

(4) The sampling frequency (i.e., the number of samples to be taken per year) is n_p or 4, whichever is larger. Denote this sampling frequency as n. Note that, under this rule, at least four samples per year per sampling well will be collected.

The steps for determining sample frequency when testing a proportion are provided in Box 9.2 and are discussed in Section 8.2 in more detail.

Box 9.2
Steps for Determining Sample Frequency far Testing a Proportion

(1) Compute the estimates of σ and ϕ which describe the measurements (not the coded values). Denote this estimates by $\hat{\sigma}$ and $\hat{\phi}_m$.

Let $\hat{\phi} = \dfrac{\hat{\phi}_m}{2.5}$, ($\hat{\phi}$ is the estimated correlation between the coded observations, the constant 2.5 was determined from simulations).

(2) Estimate the ratio of the annual overhead cost of maintaining sampling operations at the site to the unit cost of collecting, processing, and analyzing one ground water sample. Call this ratio $\$_R$.

(3) Based on the values of $\$_R$ and $\hat{\phi}$, use Appendix Table A.4 to determine the approximate number, n_p, of samples to collect per year or seasonal period. The value n_p may be modified based on site-specific considerations, as discussed in the text.

(4) The sampling frequency (i.e., the number of samples to be taken per year) is n_p or 4, whichever is larger. Denote this sampling frequency as n. Note that, under this rule, at least four samples per year per sampling well will be collected.

Box 9.3
Example of Sample Frequency Calculations

In Box 8.2, an example of determining the sample frequency is provided for a fixed sample size test. The determination of the number of samples to be taken per year is required for sequential sampling also. In that example, it was found that $n_p = 9$, so that 9 samples per year (practically speaking, once every 1.5 months) should be collected. This is all that is needed for sequential sampling. Samples will then be collected until a decision can be made. Note that in Box 8.2, a further calculation was done (computing m_d) to determine the number of years for which data are to be collected for the fixed sample size approach. After this period of time (eight years in the example) a statistical test would be made to determine whether the ground water could be considered clean or not. On average, a sequential test will require a shorter time period to reach a decision than a fixed sample size test, but this is not guaranteed.

9.2   Sequential Procedures for Sample Collection and Data Handling

The samples are assumed to be collected using a systematic sample as discussed Chapter 4.

The sample collection and analysis procedures require the following limitations on the quantity and frequency of data collected:

·   To provide the minimal amount of data required for the statistical tests, at least three years of data must be collected before any statistical test can be performed.

·   It is strongly recommended that at least four samples be collected in each period or year to capture any seasonal differences or variation within a year or period.

·   The statistical tests are performed only on data representing a complete year of data collection. Thus, the first statistical test would be performed after three full years of data collection, and the second after four full years of data collection, etc.

·   If the proportion of contaminated samples is required to be below a specified value of P_0 collect at least a number of samples N' such that **N'*$P_0 \geq 4$ before doing the first sequential test.**

Handling of outliers and measurements below the detection limit is discussed in Section 2.3.7.

9.3   Assessing Attainment of the Mean Using Yearly Averages

As noted in Chapter 8, the approach of using yearly averages substantially reduces the effects of any serial correlation in the measurements. For the procedures discussed in this section, the variance of the observed yearly averages is used to estimate the variance of the overall average concentration.   Wells can be tested individually or a group of wells can be tested jointly. In the latter case, the data for the individual wells at each point in time are used to produce a summary measure for the group as a whole. This may be an average, a maximum, or some other measure for all data values collected at a particular point in time (see Sections 2.3.5). These summary measures will be averaged over the yearly period.

Two calculation procedures for assessing attainment are described in this section. Both procedures use the yearly average concentrations. The first is based on the assumption that the yearly averages can be described by a symmetric normal distribution. The second procedure uses the log transformed yearly averages and is based on the assumption that the distribution of the yearly averages can be described by a (skewed) lognormal distribution. Because the second procedure performs well even when the data have a symmetric distribution, the second method is recommended in most situations. Only when there are missing data values which are not evenly distributed throughout the year and there is also an apparent seasonal pattern in the data is the first procedure recommended.

The calculations and procedures when using the untransformed yearly averages are described below and summarized in Box 9.4. This procedure is appropriate in most situations but is not preferred particularly if the data are highly skewed. The calculations can-be used (with some minor loss in efficiency) if some observations (are missing. If **the proportion of missing observations varies** considerably from season to season and there are differences in the average measurements among seasons, consultation with a statistician is recommended. If the data are highly skewed, the procedures described in Box 9.12 which use the log transformed yearly averages are recommended unless the data exhibit both a seasonal pattern and missing observations.

Use the formulas in Box 9.5 for calculating the yearly averages for the m years of data collect&l so far. If there are missing observations within a year, average the non-missing observations. Calculate the mean and variance of the yearly averages using the equations in Box 9.6. The variance will have degrees of freedom equal to m-l, one less than the number of years over which the data was collected.

If there are no missing observations, the mean of the yearly averages, $\overline{\overline{x}}_m$, will he compared to the cleanup standard for assessing attainment. If however, there are missing observations, the mean of the yearly averages may provide a biased estimate of the average concentration during the sample period. This will be true if the missing observations occur mostly at times when the concentrations arc generally higher or lower than the mean concentration. To correct for this bias, the mean of the seasonal averages will be compared to the cleanup standard when there are missing observations. Box 9.7 provides equations for calculating the seasonal averages and $\overline{\overline{x}}_{ms}$, the mean of the seasonal averages.

Using $\bar{x}$ to designate the mean value which is to be compared to the cleanup standard, set $\bar{x}$ $= \bar{\bar{x}}_m$ if there are no missing observations, otherwise set $\bar{x} = \bar{\bar{x}}_{ms}$.

Box 9.4
Steps for Assessing Attainment Using Yearly Averages

(1) Calculate the yearly averages for the m years of data collected so far (see Box 9.5)

(2) Calculate the mean, $\bar{\bar{x}}_m$, and variance, $s_{\bar{x}}^2$, of the yearly averages (see Box 9.6)

(3) If there are no missing observations, set

$$\bar{x} = \bar{\bar{x}}_m \qquad\qquad (9.1)$$

Otherwise, if there are missing observations calculate the seasonal averages and the mean of the seasonal averages, $\bar{\bar{x}}_{ms}$, (Box 9.7) and set

$$\bar{x} = \bar{\bar{x}}_{ms} \qquad\qquad (9.2)$$

where $\bar{x}$ is the mean to be compared to the cleanup standard.

(4) Calculate the t and δ for the likelihood ratio. (Box 9.8)

(5) Calculate the likelihood ratio for the statistical test. (Box 9.9)

(6) Decide whether the ground water attains the cleanup standards (Box 9.10).

(7) If more data are required, collect an additional years samples and repeat the procedures in this Box.

Box 9.5
Calculation of the Yearly Averages

Let x_{jk} the measurements from an individual well or a combined measure from a group of wells obtained for testing whether the mean attains the cleanup standard, x_{jk} represents the concentration for season j (the jth sample collection time out of n) in year k (where data has been collected for m years).

The yearly average is the average of all of the observations taken within the year. If the results for one or more sample times within a year are missing, calculate the average of the non-missing observations. If there are n_k ($n_k \le$ n) non-missing observations in year k, the yearly average, $\bar{x}_k$, is:

$$\bar{x}_k = \frac{1}{n_k} \sum_{j=1}^{n_k} x_{jk}. \qquad (9.3)$$

where the summation is over all non-missing observations within the year. Calculate the yearly average for all m years.

Box 9.6
Calculation of the Mean and Variance of the Yearly Averages

The mean of the m yearly averages, $\bar{\bar{x}}_m$ is:

$$\bar{\bar{x}}_m = \frac{1}{m} \sum_{k=1}^{m} \bar{x}_k \qquad (9.4)$$

where $\bar{x}_k$ is the yearly average for year k.

The variance of the yearly averages, $s_{\bar{x}}^2$, can be calculated using either of the two equivalent formulas below:

$$s_{\bar{x}}^2 = \frac{\sum_{k=1}^{m} \bar{x}_k^2 - \frac{1}{m}\left(\sum_{k=1}^{m} \bar{x}_k\right)^2}{(m-1)} = \frac{\sum_{k=1}^{m}\left(\bar{x}_k - \bar{\bar{x}}_m\right)^2}{(m-1)} \qquad (9.5)$$

This variance estimate has m-l degrees of freedom.

Box 9.7

Calculation of Seasonal Averages and the Mean of the seasonal Averages

Far the n sample collection times within the year, the jth seasonal average is the average of all the measurements taken at the jth collection time. Note that if there is a missing observation at one collection time, the measurement from the jth sample collection time may be different than the jth sequential measurement within the year.

For all collection times j, from 1 to n, within each year, calculate the seasonal average, $\bar{x}_j$. The number of observations at the jth collection time is $m_j \leq m$. If there are missing observations, sum over the m_j non-missing observations.

$$\bar{x}_j = \frac{1}{m_j} \sum_{j=1}^{m_j} x_{jk}.$$
(9.6)

The mean of n seasonal averages is:

$$\bar{\bar{x}}_{ms} = \frac{1}{n} \sum_{j=1}^{n} \bar{x}_j$$
(9.7)

The total number of observations is:

$$N = \sum_{j=1}^{n} m_j$$
(9.8)

Using the mean $\bar{x}$, and the standard deviation of the mean calculated from the yearly averages, $s_{\bar{x}}$, calculate t and δ using equations (9.9) and (9.10) in Box 9.8. These values are used in the calculation of the likelihood ratio. The standard deviation is the square root of the variance calculated from equation (9.5). The t-statistic used here is slightly different from that used in the standard t-test. Use of this definition of t makes calculation of the likelihood ratio easier.

Use equation (9.11) in Box 9.9 to calculate the likelihood ratio for the sequential test 'This equation provides a good approximation to the actual likelihood ratio which is difficult to calculate exactly. For references and more details about this approximation, see Appendix F.

Box 9.8

Calculation of t and δ When Using the Untransformed Yearly Averages

$$t = \frac{\bar{x} - \frac{Cs + \mu_1}{2}}{\sqrt{\frac{s_{\bar{x}}^2}{m}}} \qquad (9.9)$$

$$\delta = \frac{\mu_1 - Cs}{\sqrt{\frac{s_{\bar{x}}^2}{m}}} \qquad (9.10)$$

where $\bar{x}$ is the mean level of contamination, and $s_{\bar{x}}$ is the square root of the variance of the yearly means. The degrees of freedom associated with $s_{\bar{x}}$ is m-1

Box 9.9

Calculation of the Likelihood Ratio for the Sequential Test

The likelihood ratio is:

$$LR = \exp\left(\delta \frac{m-2}{m} t \sqrt{\frac{m}{m-1+t^2}} \right) \qquad (9.11)$$

where m is the number of years of data collected so far and t and δ are calculated from the m years of data.

Finally, the likelihood ratio, α, and β are used to decide if the average concentration is less than the cleanup standard. If the average is less than the cleanup standard and if the concentrations are not increasing over time (see Section 9.7), conclude that the tested ground water attains the cleanup standard. If the ground water from all wells or groups of wells attains the cleanup standard then conclude that the ground water at the site attains the cleanup standard. If the average concentration is not less than the cleanup standard or if the concentrations are increasing over time, conclude that the ground water in the well does not attain the cleanup standard. The steps in deciding attainment of the cleanup standard are shown in Box 9.10.

Box 9.10
Deciding if the Tested Ground Water Attains the Cleanup Standard

Calculate:

$$A = \frac{\beta}{(1-\alpha)} \quad \text{and} \quad B = \frac{(1-\beta)}{\alpha} \qquad (9.12)$$

If LR ≤ A, conclude that the ground water in the wells <u>does not attain</u> the cleanup standard. ,

If LR > B, conclude that the average ground water concentration in the well (or group of wells) is less than the cleanup standard. Perform a trend test using the regression techniques described in Chapter 6 to determine if there is a statistically significant increasing trend in the yearly averages over the sampling period (also see Section 9.7).

If there is not a statistically significant increasing trend, conclude that the ground water <u>attains</u> the cleanup standard (and possibly initiate a follow-up monitoring program). If a significant trend does exist, conclude that the ground water in the wells <u>does not attain</u> the cleanup standard and resume sampling or reconsider treatment effectiveness.

If A < LR ≤ B then collect an additional years worth of data before performing the hypothesis test again.

Box 9.11
Example Attainment Decision Based on a Sequential Test

In this example we will use the arsenic measurements appearing in Table 8.1. Suppose we wish to compare the cleanup standard ($Cs = 6$) with a targeted cleanup average (μ_1) of 5.72 (μ_1 is the value for which the false negative rate β is to be controlled). Box 8.21 indicates the four yearly means $\bar{x}_k$ and the overall average $\bar{x}_m = 5.914$. Using equation (9.5), the value of $s_{\bar{x}}^2 = .0706$ for $m = 4$. Thus,

$$t = \frac{\bar{x} - \frac{Cs + \mu_1}{2}}{\sqrt{\frac{s_{\bar{x}}^2}{m}}} = \frac{5.914 - \frac{6 + 5.72}{2}}{\sqrt{\frac{.0706}{4}}} = .406 \text{ and}$$

$$\delta = \frac{\mu_1 - Cs}{\sqrt{\frac{s_{\bar{x}}^2}{m}}} = \frac{5.72 - 6}{\sqrt{\frac{.0706}{4}}} = -2.108$$

$$LR = \exp\left(\delta \frac{m-2}{m} t \sqrt{\frac{m}{m-1+t^2}}\right) = \exp\left(-2.108 \frac{4-2}{4} .406 \sqrt{\frac{4}{4-1+.406^2}}\right) = 0.618$$

With $\alpha = .1$ and $\beta = .1$, then $A = .111$, $B = 9.0$. Since 0.618 is neither less than A or greater than B, we have insufficient data to conclude that the cleanup standard has been either attained or not attained. Thus, more data must be gathered.

Suppose data continue to be collected for seven more years without a decision being reached. At that time, the overall average $\bar{x}_m = 5.77$ and $s_{\bar{x}}^2 = .1024$ for $m = 11$. Thus,

$$t = \frac{5.77 - \frac{6 + 5.72}{2}}{\sqrt{\frac{.1024}{11}}} = -.933 \text{ and } \delta = \frac{5.72 - 6}{\sqrt{\frac{.1024}{11}}} = -2.902$$

$$LR = \exp\left(-2.902 \frac{11-2}{11} (-.933) \sqrt{\frac{11}{11-1+.933^2}}\right) = 9.29$$

Since $LR = 9.38 > 9.0$, we conclude that the mean ground water concentrations are less than the cleanup standard.

When the data are noticeably skewed, the calculation procedures using the log transformed yearly averages (Box 9.12) arc recommended over those in Box 9.4. Because the procedures in Box 9.12 also perform well when the data have a symmetric **distribution, these procedures arc generally recommended, in all cases where there are no missing data. There is no easy adjustment for missing data when using the log transformed yearly averages. Therefore, if the number of observations per season (month etc.) is not** the same for all seasons and if there is any seasonal pattern in the data, use of the procedures in Box 9.4 is recommended.

The calculations procedure when using the log transformed yearly averages is described below and summarized in Box 9.12. The calculations arc slightly more difficult than when using the transformed yearly averages. After calculating the yearly **averages, take the natural log is used to transform the data. The transformed averages are** then used in the subsequent analysis. The upper **confidence interval for the mean concen-** tration is based on the mean and variance of the log transformed yearly averages. The formulas are based on the assumption that the yearly averages have a log normal **distribution.**

Box 9.12
Steps for Assessing Attainment Using the Log Transformed Yearly Averages

(1)   Calculate the yearly averages (see Box 9.5)

(2)   Calculate the natural log of the yearly averages (see Box 9.13)

(3)   Calculate the mean, $\bar{y}_m$, and variance, $s_{\bar{y}}^2$, of the log transformed yearly averages (see Box 9.14)

(4)   Calculate the t and δ for the likelihood ratio. (Box 9.15)

(5)   Calculate the likelihood ratio (Box 9.9)

(6)   Decide whether the ground water attains the cleanup standards (Box 9.10).

(7)   If more data are required, collect an additional years samples and repeat the procedures in this Box.

Use the formulas in Box 9.5 for calculating the yearly averages. **If there are missing observations within a year, average the non-missing observations. Calculate the** log transformed yearly averages using equation (9.13) in Box 9.13. The natural log transformation is available on many calculators and computers, usually designated as **"LN", "1n", or "log$_e$."** Although the equations could be changed to use the base 10 logarithms, use only the base e logarithms when using the equations in Boxes 9.13 through **9.15. Calculate the mean and variance of the log transformed yearly averages using the** equations in Box 9.14. The variance will have degrees of freedom equal to one less than **the number of years over which the data was collected.**

Box 9.13
Calculation of the Natural Logs of the Yearly Averages

The natural log of the yearly average is:

$$y_k = \ln(\bar{x}_k) \tag{9.13}$$

Box 9.14
Calculation of the Mean and Variance of the Natural Logs of the Yearly Averages

The average of the m log transformed yearly averages, $\bar{y}_m$:

$$\bar{y}_m = \frac{1}{m} \sum_{k=1}^{m} y_k \tag{9.14}$$

The variance of the log transformed yearly averages, $s_{\bar{y}}^2$:

$$s_{\bar{y}}^2 = \frac{\sum_{k=1}^{m} y_k^2 - \frac{1}{m} \left(\sum_{k=1}^{m} y_k \right)^2}{(m-1)} = \frac{\sum_{k=1}^{m} (y_k - \bar{y})^2}{(m-1)} \tag{9.15}$$

This variance estimate has m-l degrees of freedom.

Using the mean $\bar{y}_m$, and the variance of the mean calculated from the log transformed yearly averages, $s_{\bar{y}}^2$, calculate t and δ using equations (9.16) and (9.17) in Box 9.15. These values are used in the calculation of the likelihood ratio.

$$\text{Box } 9.15$$

Calculation of t and δ When Using the Log Transformed Yearly Averages

$$t = \frac{\bar{y}_m + \frac{s_{\bar{y}}^2}{2} - \frac{\ln(Cs)+\ln(\mu_1)}{2}}{\sqrt{\frac{s_{\bar{y}}^2}{m} + \frac{s_{\bar{y}}^4}{2\,Df}}} \qquad (9.16)$$

$$\delta = \frac{\ln(\mu_1) - \ln(Cs)}{\sqrt{\frac{s_{\bar{y}}^2}{m} + \frac{s_{\bar{y}}^4}{2\,Df}}} \qquad (9.17)$$

where the degrees of freedom (Df) associated with $s_{\bar{y}}^2$ is m-1

Use equation (9.11) in Box 9.9 to calculate the likelihood ratio for the sequential test. Finally, the likelihood ratio, α, and β are used to decide if the average concentration is less than the cleanup standard. If the average is less than the cleanup standard and if the concentrations are not increasing over time, conclude that the tested ground water attains the cleanup standard. If the ground water from all wells or groups of wells attains the cleanup standard then conclude that the ground water at the site attains the cleanup standard. If the average concentration is not less than the cleanup standard or if the concentrations are increasing over time, conclude that the ground water in the well does not attain the cleanup standard. The steps in deciding attainment of the cleanup standard are shown in Box 9.10.

9.4   **Assessing Attainment of the Mean After Adjusting for Seasonal Variation**

This section provides an alternative procedure for testing if the mean concentration is less than the cleanup standard. It is expected to provide more accurate results when there are many samples per year and the data is both serially correlated and the distribution of the data is not skewed. Because this procedure is sensitive to skewness in the data, it is recommended only if the distribution of the measurement errors is reasonably symmetric.

After the data have been collected using the guidelines indicated in Chapter 4, wells can be tested individually or **a group** of wells can be tested jointly. In the latter case, the data for the individual wells at each point in time are used to produce a summary measure for the group as **a** whole. This summary measure may be an average, maximum, or some other measure (see Chapter 2). These summary measures will be averaged over the entire sampling period. The steps involved for incorporating seasonal adjustments and serial correlations into the calculations associated with the statistical tests arc discussed.

The calculations and procedures for assessing the mean after adjusting for seasonal variation arc described below and summarized in Box 9.16. An example is provided in Box 9.21. The calculations can be used (with some minor loss in efficiency) if some observations are missing. With a large proportion of missing observations in any season, consultation with a statistician is recommended. If the data are obviously skewed, the procedures described in Box 9.12 which use the log transformed yearly averages arc recommended.

Use the formulas in Box 9.7 for calculating the seasonal averages and the mean of the seasonal averages. If there are missing observations within a season, average the non-missing observations. Calculate the residuals, the deviations of the measurements from the respective seasonal means, using equation (9.18) in Box 9.17. Box 9.18 shows how to calculate the variance of the residuals. The variance will have degrees **of freedom** equal to the number of measurements less the number of seasons.

Box 9.16
Steps for Assessing Attainment Using the Mean After Adjusted for Seasonal Variation

(1)   Calculate the seasonal averages and the mean of the seasonal averages, $\bar{x}_{ms}$, (Box 9.7)

(2)   Calculate the residuals, the differences between the observations and the corresponding seasonal averages (Box 9.17)

(3)   Calculate the variance, s^2, of the residuals (see Box 9.18)

(4)   Calculate the lag 1 serial correlation of the residuals using equation (9.20) in Box 9.19. Denote the computed serial correlation by $\hat{\phi}_{obs}$.

(5)   Calculate the t statistic based on the mean, $\bar{x}_{ms}$, the standard deviation s, and $\hat{\phi}_{obs}$. (Box 9.20)

(6)   Calculate the likelihood ratio (Box 9.21)

(7)   Decide whether the ground water attains the cleanup standards (Box 9.10).

Box 9.17
Calculation of the Residuals

From each sample observation, subtract the corresponding seasonal mean. That is, compute the, e_{jk} the deviation from the seasonal mean:

$$e_{jk} = x_{jk} - \bar{x}_{j.} \qquad (9.18)$$

Using the mean of the seasonal averages and the variance of the residuals, s^2, calculate t and δ using equations (9.21) and (9.22) in Box 9.20. These values are used in the calculation of the likelihood ratio.

Box 9.18
Calculation of the Variance of the Residuals

Calculate the variance of the observations e_{jk} reflecting adjustments for possible seasonal differences using the equation in Box 8.12.

$$s^2 = \frac{1}{N-m} \sum_{j=1}^{n} \sum_{k=1}^{m_j} e_{jk}^2. \qquad (9.19)$$

Alternatively, the ANOVA approach described in Appendix D can be used to compute the required variance.

Box 9.19
Calculating the Serial Correlation from the Residuals After Removing Seasonal Averages

The sample estimate of the serial correlation of the residuals is:

$$\hat{\phi}_{obs} = \frac{\sum_{i=2}^{N} e_i e_{i-1}}{\sum_{i=1}^{N} e_i^2} \qquad (8.18)$$

Where e_i, $i = 1, 2, ...,N$ are the residuals after removing seasonal averages, in the time order in which the samples were collected.

Box 9.20

Calculation of t and δ When Using the Mean Corrected for Seasonal Variation

$$t = \frac{\bar{x}_{ms} - \frac{Cs + \mu_1}{2}}{\sqrt{\frac{s^2}{N} \frac{1 + \hat{\phi}_{obs}}{1 - \hat{\phi}_{obs}}}} \qquad (9.20)$$

$$\delta = \frac{\mu_1 - Cs}{\sqrt{\frac{s^2}{N} \frac{1 + \hat{\phi}_{obs}}{1 - \hat{\phi}_{obs}}}} \qquad (9.21)$$

where $\bar{x}_{ms}$ is the mean level of contamination computed from equation (9.7), and s^2 is the variance of the observations computed from equation (9.16). The degrees of freedom, Df, associated with these estimates is $Df = \frac{N - m}{3}$.

Use the formula in Box 9.21 to calculate the likelihood ratio for the sequential test Although this formula for calculating the likelihood ratio looks different than when using the yearly averages (see Box 9.9). the two formulas are equivalent

---

Box 9.21

Calculation of the Likelihood Ratio for the Sequential Test When Adjusting for Serial correlation

The likelihood ratio is:

$$LR = \exp\left(\delta \frac{Df-1}{Df+1} t \sqrt{\frac{Df+1}{Df+t^2}} \right) \qquad (9.22)$$

where Df is the degrees of freedom for s^2.

Box 9.22
Example Calculation of Sequential Test Statistics after Adjustments for
Seasonal Effects and serial Correlation

In Box 8.21, a test was performed for a fixed sample size after adjusting for seasonal effects and seasonal correlation. We will use the same data (from Table 8.1) to conduct the corresponding sequential test after four years of data collection. From Box 8.21 we have $\bar{x} = 5.914$, $s^2 = .163$, $\hat{\phi}_{obs} = .37$, $c_s = 6.0$, $m = 4$, and $N = 16$. We will stipulate that $\alpha = .1$, $\beta = .1$, and $\mu_1 = 5.72$. Thus,

$$t = \frac{\bar{\bar{x}}_{ms} - \frac{c_s + \mu_1}{2}}{\sqrt{\frac{s^2}{N} \frac{1 + \hat{\phi}_{obs}}{1 - \hat{\phi}_{obs}}}} = \frac{5.914 - \frac{6 + 5.72}{2}}{\sqrt{\frac{.0706}{16} \frac{1 + .37}{1 - .37}}} = 0.551$$

$$\delta = \frac{\mu_1 - c_s}{\sqrt{\frac{s^2}{N} \frac{1 + \hat{\phi}_{obs}}{1 - \hat{\phi}_{obs}}}} = \frac{5.72 - 6}{\sqrt{\frac{.0706}{16} \frac{1 + .37}{1 - .37}}} = -2.858$$

and

$$Df = \frac{N - m}{3} = \frac{16 - 4}{3} = 4$$

$$LR = \exp\left(\delta \frac{Df - 1}{Df + 1} t \sqrt{\frac{Df + 1}{Df + t^2}}\right) = \exp\left(-2.858 \frac{4 - 1}{4 + 1} 0.551 \sqrt{\frac{4 + 1}{4 + 0.551^2}}\right) = .347$$

With $\alpha = .1$ and $\beta = .1$, then $A = .111$, $B = 9.0$. Since 0.746 is neither less than A or greater than B, we have insufficient data to conclude that the cleanup standard has been either attained or not attained. Thus, more data must be gathered.

9.5 Sequential Tests for Proportions

In general, sequential procedures for testing proportions require that more samples be collected before starting the fast test of hypothesis than when testing the mean. If the parameter to be tested is the proportion of contaminated samples from either one well or an array of wells, the sample collection and analysis procedures arc the same as those outlined above for testing the mean, with the following changes:

· To apply this test, each ground water sample measurement is either coded "1" (the **actual** measurement was equal to or above the cleanup standard Cs), or **"0"** (below Cs). The statistical analysis is based on the resulting coded variable of O's and 1's.

· Only the analysis procedure which used yearly averages is appropriate for the calculations (Box 9.4). Do not use either of the calculation procedures in Boxes 9.12 or 9.16.

• A total of at least $\frac{4}{P_0}$ samples should be collected before using the statistical procedures to determine, on a yearly basis, whether sampling can be stopped and a decision can be made.

9.6 A Further Note on Sequential Testing

It should be noted that sequential testing, as discussed in this chapter, has a small chance of continuing for a very long time if the data gathered provide insufficient evidence for making a clear-cut determination. A stopping rule, such as the following can be implemented to handle such cases: determine the sample size necessary for a fixed sample test for the specified values of Cs, μ_1, α, and β (data collected during the sampling for assessing attainment can be used to estimate the variance so the sample size can be computed). Call this sample size m_{fixed}. If the number of years of sample collection exceeds twice m_{fixed}, determine the likelihood ratio. If the likelihood ratio is less than 1.0, conclude that the ground water does not attain the cleanup standard. If the likelihood ratio is greater than 1.0 conclude that the mean concentration is less than the cleanup standard and test if there is a significant positive slope in the data.[1]

9.1 Checking for Trends in Contaminant Levels After Attaining the Cleanup Standard

Once a fixed sample size statistical test indicates that the cleanup standard for the site has been met, there remains one final concern. The model we have used assumes that ground water at the site has reached a steady state and that there is no reason to believe that contaminant levels will rise above the cleanup standard in the future. We need to check this assumption. Regression models, as discussed in Chapter 6, can be used

---

[1] A likelihood ratio of one occurs when the sample mean is at the mid-point between the cleanup standard and the mean for the alternate hypothesis.

to do so. By establishing a simple regression model with the contaminant measure as the dependent variable and time as the independent variable, a test of significance can be made as to whether or not the estimated slope of the resulting linear model is positive (see Section 6.1.3). Scatter plots of the data will prove useful in assessing the model. When using the yearly averages, the regression can be performed without adjusting for serial correlation.

To minimize the chance of incorrectly concluding that the concentrations are increasing over time, we recommend that the alpha level for testing the slope (and selecting the t statistic in Box 6.11) be set at a small value, such as 0.01 (one percent). If, on the basis of the test, there is not significant evidence that the slope is positive, then the evidence is consistent with the **preliminary** conclusion that the ground water in the well(s) attains the cleanup standard. If the slope is significantly greater than zero, then the concern that contaminant levels may later exceed the cleanup standard still exists and the assumption of a steady state is called into question. In this case, further consideration must be given to the reasons for this apparent increase and, perhaps, to additional remediation efforts.

9.8   Summary

This chapter presented the procedures for assessing attainment of the cleanup standard for ground water measurements using a sequential statistical test. For most statistical tests or procedures, the analysis is performed after the entire sample has been collected and the laboratory results are complete. However, in sequential testing, the samples are analyzed as they are collected. A statistical analysis of the data collected so far is used to determine whether another years worth of samples should be collected or whether the analysis should terminate.

We presented three alternate procedures for assessing attainment using sequential tests. Two procedures use the yearly average concentrations, one assumes the yearly average has a normal distribution, the other assumes a log normal distribution. The third procedure uses the individual observations and makes a correction for seasonal patterns and serial correlations. In general, the method which assumes the yearly averages have a log normal distribution is recommended.

These testing procedures can be applied to samples from either individual wells or wells tested as a group. These procedures are used after the ground water has

achieved steady state. If the ground water at the site is judged to attain the cleanup standards because the concentrations arc not increasing and the long-term average is **significantly less than the cleanup standard, follow-up monitoring is recommended to check that the steady state assumption holds.**

blank page

BIBLIOGRAPHY

Abraham, B. and Ledolter, J., 1983, Statistical Methods for Forecasting. New York John Wiley and Sons, Inc.

Albers, W., 1978, "One Sample Rank Tests Under Autoregressive Dependence," Annals *of Statistics,* Vol. 6, No. 4: 836-845.

Albers, W., 1978, 'Testing the Mean of a Normal Population Under Dependence," Annals *of Statistics,* Vol. 6, No. 6: 1337-1344.

Armitage, P. 1947, "Some Sequential Tests of **Student's** Hypothesis," Journal *of the Royal Statistical Society, Series* B, Vol. 9 : 250-263.

Armitage, P. 1957, "Restricted Sequential Procedures," *Biometrika,* Vol. 44 : 9-26.

Barcelona, M., Gibb, J., and Miller, R., 1983, A Guide to the Selection of Materials For Monitoring Well Construction and Ground-Water Sampling. Illinois State Water Survey, Champaign, Illinois, US EPA- RSKERL, EPA

Barcelona, M., Gibb, J., Helfrich, J., and Garske, E., 1985, Practical Guide for Ground-Water Sampling, Illinois State Water Survey, Champaign, Illinois, USEPA-RSKERL, EPA 600/2-85/104.

Barnett, V., and Lewis, T., 1984, Outliers in Statistical Data. New York: John Wiley and Sons, Inc.

Bartels, R., 1982 "The Rank Version of von Newmann's Ratio Test for Randomness," *Journal of the American Statistical Association,* Vol. 77, No. 377: 40-46.

Bauer, P., and Hackl, P., 1978, "The Use of MOSUMS for Quality Control," Technometrics, Vol. 20, No. 4: 431--436.

Bell, C., and Smith, E., 1986, "Inference for Non-Negative Autoregressive Schemes," *Communications in Statistics: Theory and* Methods, Vol. 18, No. 8: 2267-2293.

Berthouex, P., Hunter, W., and Pallisen, L., 1978, "Monitoring Sewage Treatment Plants: Some Quality Control Aspects," *Journal of Quality Technology,* Vol. 10, No. 4: 139-149.

Bisgaard, S., and Hunter, W. G., 1986, Report No. 7, Studies in Quality Improvement: Designing Environmental Regulations,. Center-for Quality and Productivity Improvement. University of Wisconsin-Mason, (February

Simulations

Preliminary simulations using lognormally distributed data and a factorial design with 100 simulations for each set of parameters was used to determine which factors affected the power of the sequential tests. The factors in the simulations were: scale factor; proportion of the random variance which is correlated versus independent; lag 1 correlation; presence of a seasonal pattern; proportion of the observations which were censored; number of samples per year; and μ. Analysis of the factorial design clearly indicated that the skewness and scale factor were most important in determining the power of the test. The serial correlation and censoring were also important. The presence of a cyclical component (which resulted in significant changes in the variance throughout the year) did not significantly affect the power of the test.

As a result of these preliminary simulations, further simulations were run using scale factors ranging from 1.6 to 4.8, $\alpha = \beta = .05$, $\mu = \mu_0$ or μ_1, and the following distributions and sampling designs:

(1)   Normal distribution with independent errors and 4 samples per year;
(2)   Lognormal distribution with coefficient of variation of 0.5, independent errors and 4 samples per year. This is the basic distribution. The following simulations all are based on changes to the basic distribution.
(3)   The basic distribution with 12 observations per year;
(4)   The basic distribution but more skewed, with a coefficient of variation of 1.5;
(5)   The basic distribution with censoring of 30% of the data (censored values were set equal to the detection limit);
(6)   The basic distribution with correlated errors, the serial correlation between log transformed monthly observations is 0.8; and
(7)   Data which are both skewed and correlated, with coefficient of variation of 1.5 and serial correlation between log transformed monthly observations is 0.8. For this set of simulations, the random error was the sum of two components, one random, representing random measurement error, and the second correlated, reflecting correlations in the the groundwater concentrations. The correlated error made up 75% of the total error variance.

For each test and each set of simulations with the same distributional assumptions, Figure 8 shows the range in the false positive rate across simulations. Figure 9 shows similar information for the false negative rate.

As can be seen from Figure 8, the false positive rate for the tests are close to the nominal level of 0.05 when the data have a normal distribution, as desired. For skewed and correlated data, the false positive rate generally exceeds the nominal level.

For skewed and correlated data, the false positive rate for the standard sequential t-test exceeds the nominal value for all simulations. The performance of the modified test and the modified test with adjustments for seasonal patterns and serial correlations had similar false positive rates. Both of these tests are sensitive to correlated and skewed data. The false positive rate for the modified test adjusted for skewness is lower than for the other three tests. Only for correlated data does this test have a false positive rate consistently greater than the nominal level. Censoring resulted in a relative decrease in the false positive rate. Of the tests based on the modified sequential t-test, the test with adjustments for skewness had the lowest average sample sizes and lowest false positive rates.

Based on both the average sample sizes and false positive rates from the simulations, the modified test adjusted for skewness is preferred over the other sequential tests. To the extent that the false positive rate exceeds the nominal level for skewed and correlated data, the power can be improved

by using two year averages instead of one year averages. Results for the skewed and correlated data using two year averages are also shown in Figure 8.

As shown in Figure 9, the false negative rate for all tests was generally similar to or less than the nominal level. The false negative rate for the standard sequential t-test exceeded that for the **procedures based** on the modified test For all tests, the false negative rate increased greatly in the presence of censoring. Procedures based on the modified test, the modified test adjusted for skewness had a false negative rate closest to the nominal level under the simulated conditions. Although the average sample sizes for the tests were similar, the test adjusted for skewness had highest average sample sizes. At the alternate hypothesis no one calculation procedure is clearly preferred, however, the modified test has false negative rates lower than the nominal value for all but censored observations and is the simplest to calculate.

The sample sizes for the skewed data were similar to those for the normally distributed data for which the sequential test required fewer samples, on the average, than the equivalent fixed sample size test. Therefore, it is likely that the sequential tests would also have lower average sample size than for a fixed sample size test where the sample size calculations accounted for the skewed and/or correlated nature of the data.

6.   Conclusions and Discussion

For assessing attainment of Superfund cleanup standards based on the mean contaminant levels using sequential tests, the conclusions from this simulation study are:

- Given the situations found at Superfund sites, a sequential test can reduce the number of samples compared to the that for an equivalent fixed sample size test;

- The standard sequential t-test can have false negative rates greater than the nominal value.

- An adjustment factor can be used to improve the power performance of the sequential t-test without greatly increasing the sample sizes. Different criteria will result in the selection of different adjustment factors, however. all of the adjustment factors considered improved the performance of the test. In this paper, the adjustment factor (n-2)/n was evaluated.

- Use of a simple approximation to the likelihood ratio performs well compared to that based on the non-central t distribution;

- Sampling rules which terminate the sequential test if the number of samples exceeds twice the sample size for the equivalent fixed sample size test are likely to have little effect on the power of the sequential t-test;

- A modified sequential t-test with an adjustment for skewness has the lowest false positive rate among the tests considered and has acceptable false negative rates and sample sizes relative to the other tests; and

- All test procedures were sensitive to censored data.

The procedures used here set censored values equal to the detection limit. Other possible approaches place censored values at half the detection limit or at zero. Further work is required to determine how the sequential tests perform using different rules for handling values below the detection limit The decision rule which places censored values at the detection level was chosen to protect human health and the environment when assessing attainment at Superfund sites.

The problem of testing multiple wells and contaminants is particularly troublesome when the decision rule requires that all wells and all contaminants must attain the relevant cleanup standards. Even if all concentrations are below the cleanup standard, the probability of a false negative on any one of several statistical tests increases the probability of falsely concluding that additional cleanup is required. The false negative rate for the modified sequential tests considered in this paper are generally lower than the nominal value for all but censored data. Therefore, use of these tests will generally not contribute, beyond that planned for in the sample and analysis plan, to incorrectly concluding that the ground water attains the Cleanup standard unless the data are censored.

All of the power curves are based on the assumption that the standard deviation will remain constant as the mean changes. Another possible assumption is that the coefficient of variation will remain constant as the mean changes. While the assumption about how the standard deviation changes as the mean changes does not affect the conclusions presented, the actual shape of the power curves will depend on the assumptions made.

Finally, these modified sequential t-tests can also be used when the alternate hypothesis is *greater* than the null hypothesis. The results above can be applied if the false negative and false positive labels are reversed. For compliance monitoring, i.e., to answer the question: do the concentrations exceed an action level?, all of the modified sequential tests perform well if the data arc not censored. With censored data, alternate rules for handling the observations below the detection level should be considered.

Bibliography

Ghosh, B. K., 1970, Sequential Tests of Statistical Hypotheses, Reading **MA, Addison** Wesley.

Hall, W. J., 1962, "Some Sequential Analogs of Stein's Two Stage Test," *Biometrika*, Vol 49,: 367-378.

Hayre, L. S,, 1983, "An Alternative to the Sequential T-Test." *Sankhya : The Indian Journal of Statistics, 45, Series* A, Pt. 3, 288-300

Liebetrau, A. M., 1979, "Water Quality Sampling: Some Statistical Considerations," *Water Resources Research,* Vol. 16, No. 6: 1717-1725.

Loftis, J., Montgomery, R., Harris, J., Nettles, D., Porter, P., Ward, R., and Sanders, T., 1986, "Monitoring Strategies for Ground Water Quality Management," Prepared for the United States Geological Service by Colorado State University, Fort Collins Colorado.

Rushton, S., 1950, "On a Sequential t-Test," *Biometrika,* Vol. 37: 326-333.

Wald, A., 1947, Sequential Analysis. New York: Dover Publications.

Figure 1   Example of Simulated Monthly Ground Water Data
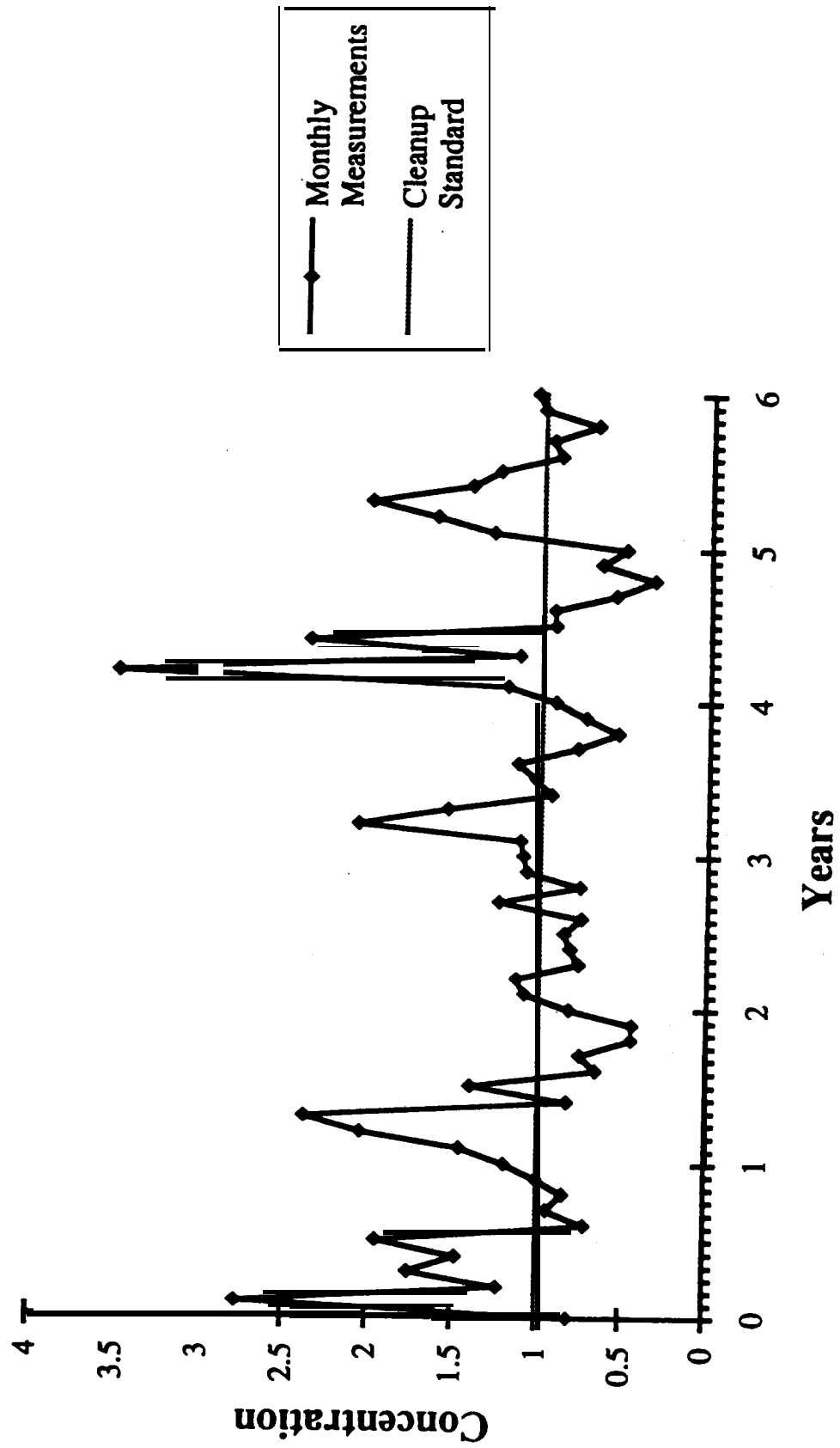
Figure 2 Power Curve and Average Sample Size for a Sequential t-Test
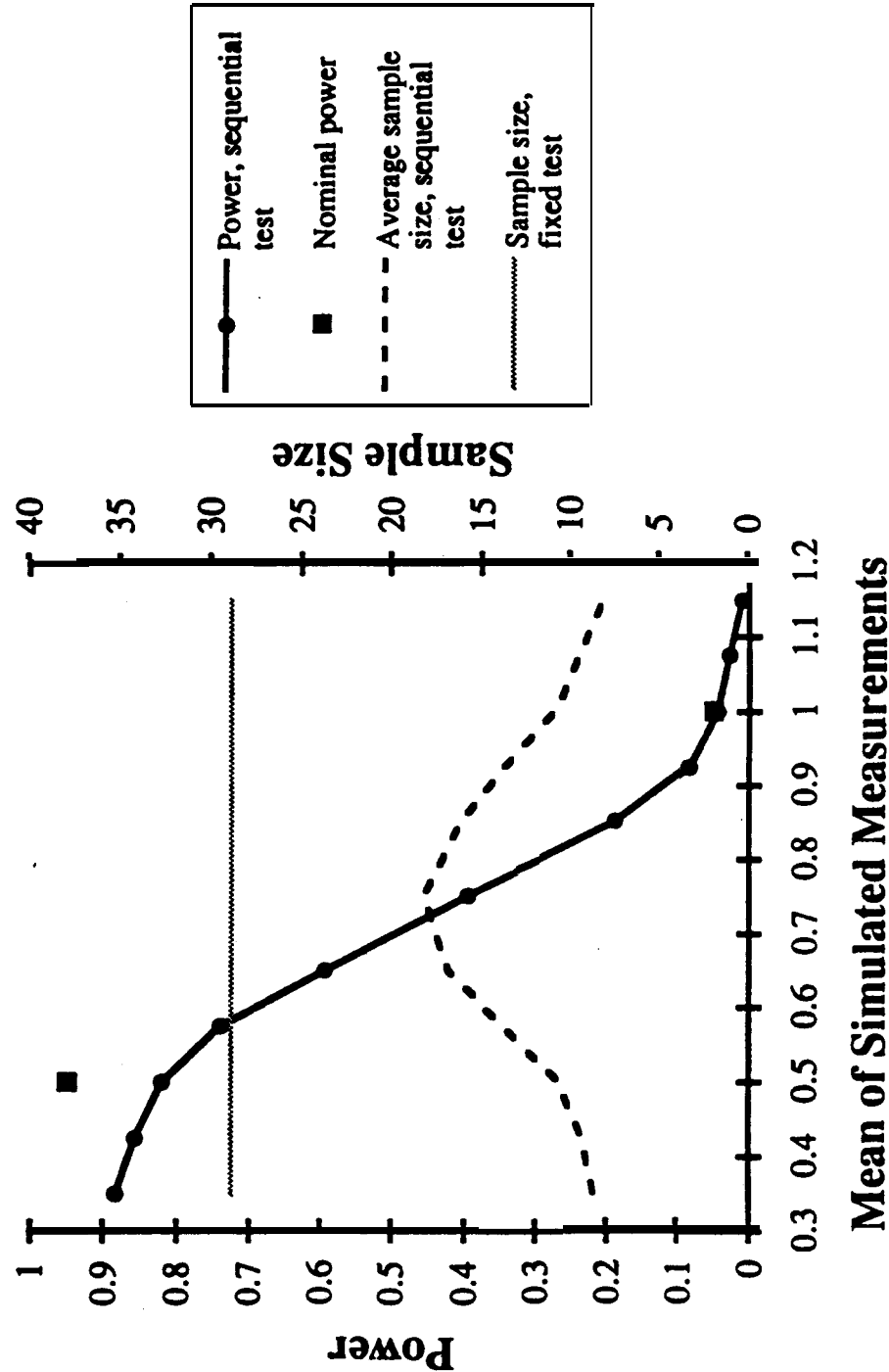
Figure 3   **False Decision Rate and Sample Size versus Scale Factor (Centered test)**
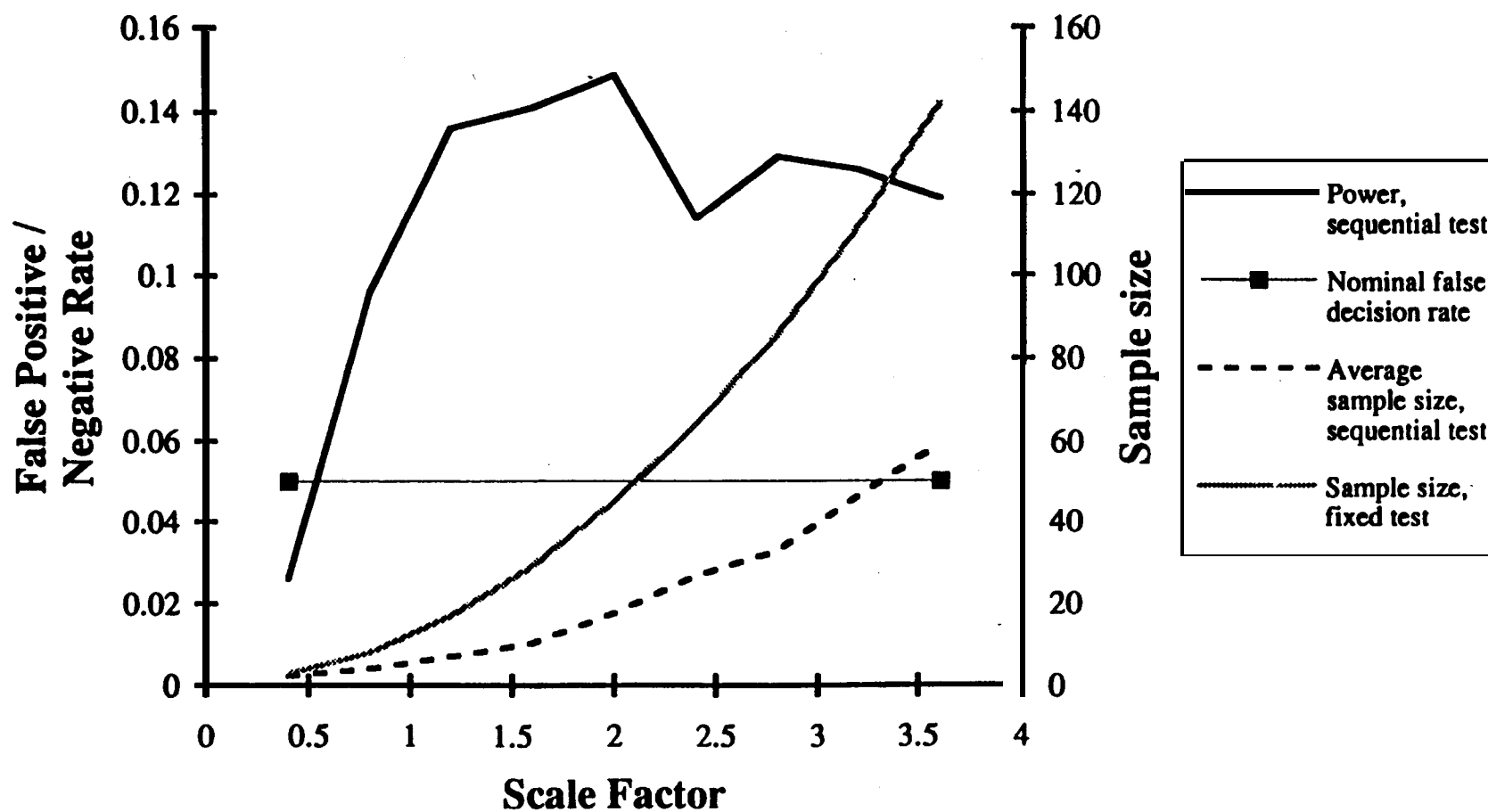
Figure 4 Distribution of Sample Sizes for the Centered and Modified Sequential t-test, by Test Result

True mean = alternate hypothesis
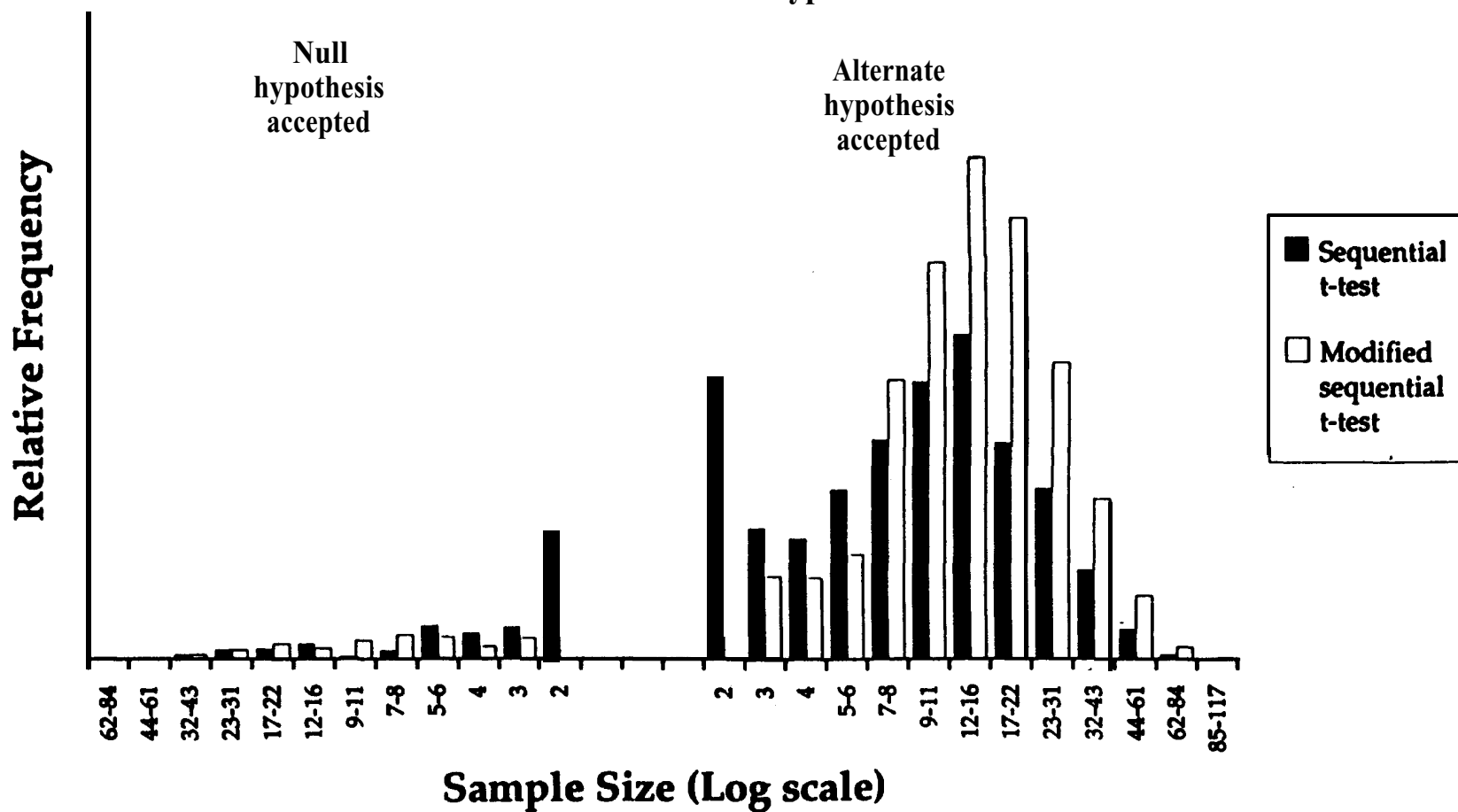
Figure 5   False Decision Rate for Three Alternate Adjustment Factors (F) for the Sequential t-Test.
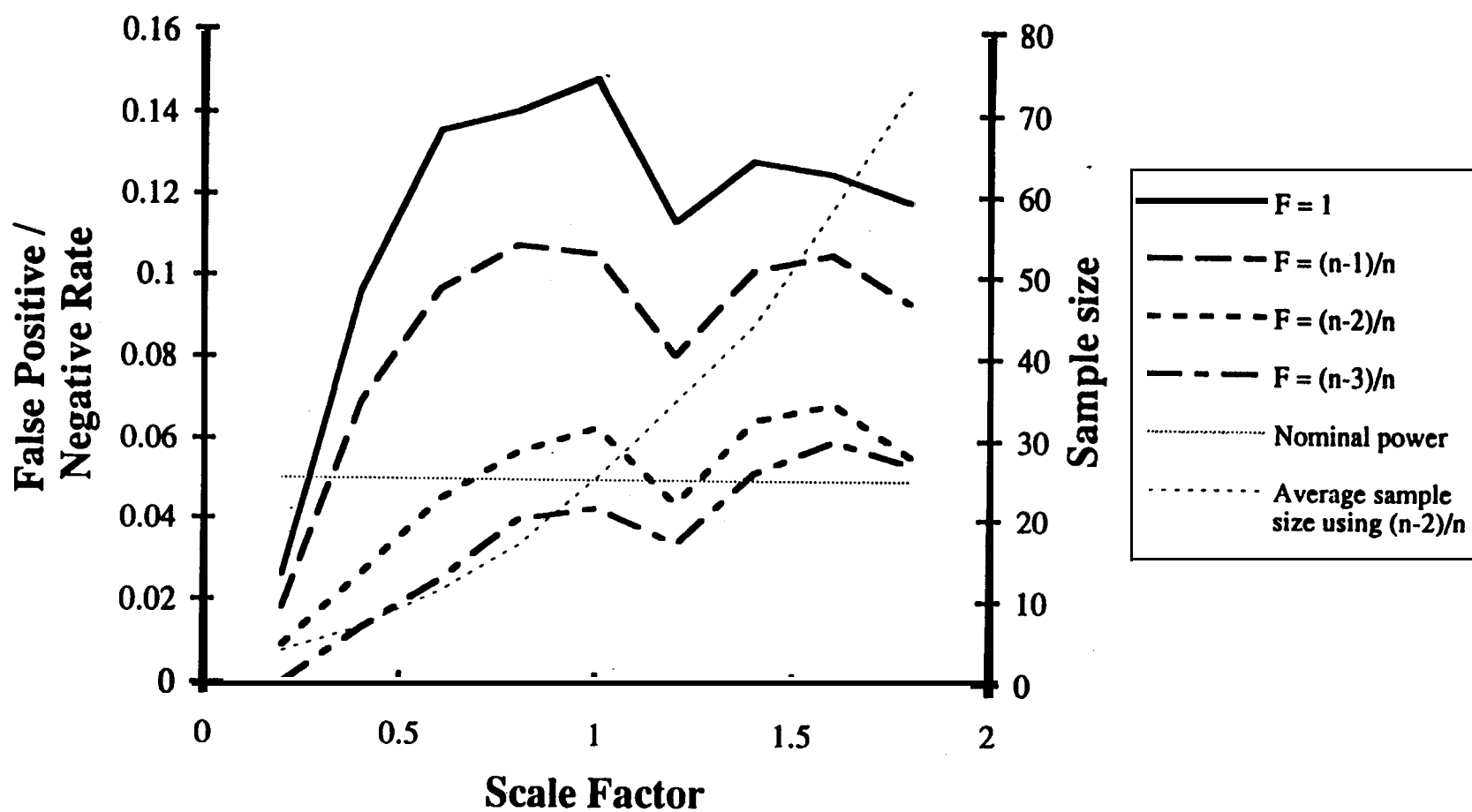
Figure 6   Power Curve and Average Sample Size for the Modified Sequential *t*-Test
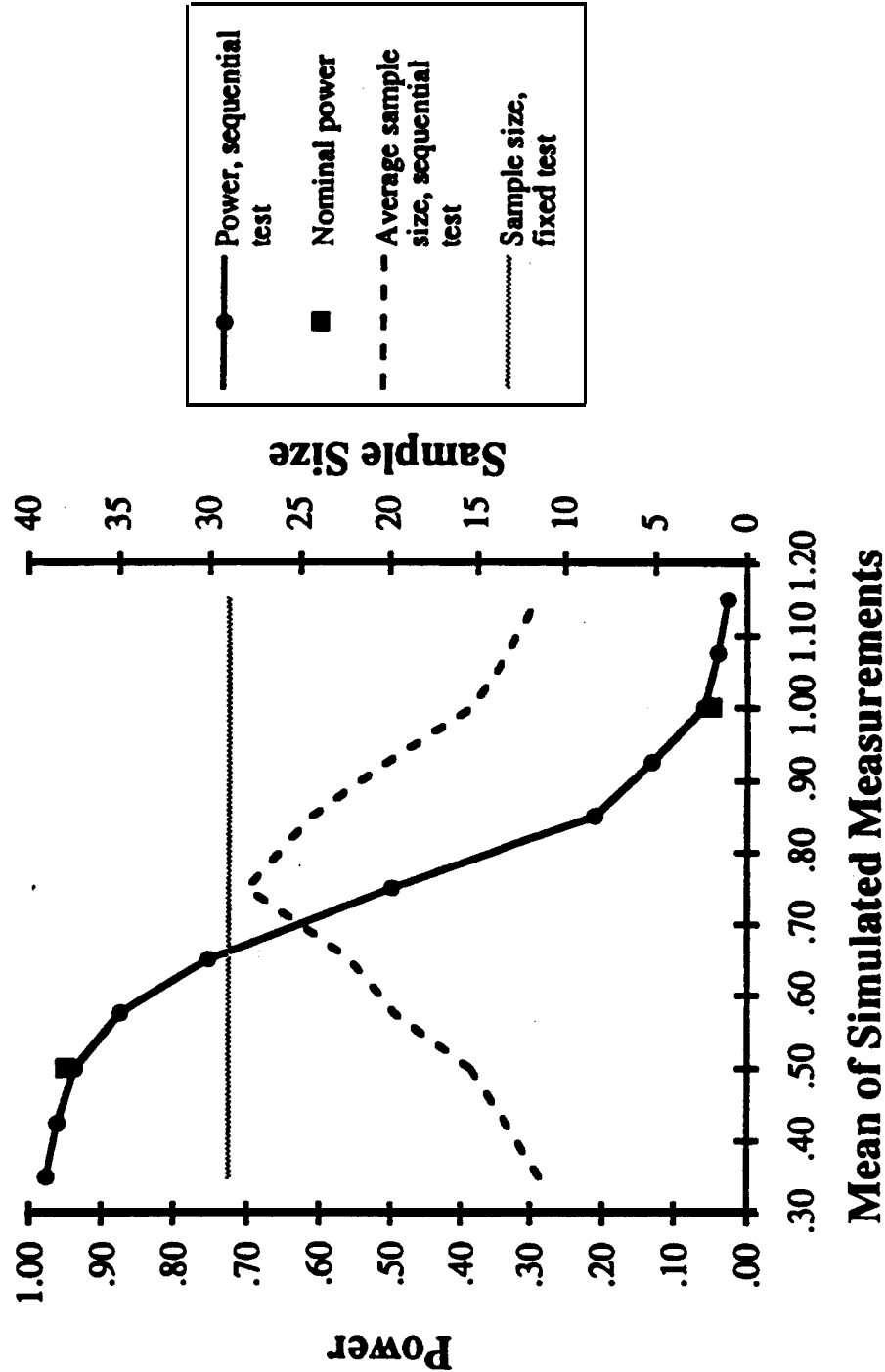
Figure 7   Sample Size Distribution for Modified Sequential t-Test versus Mean
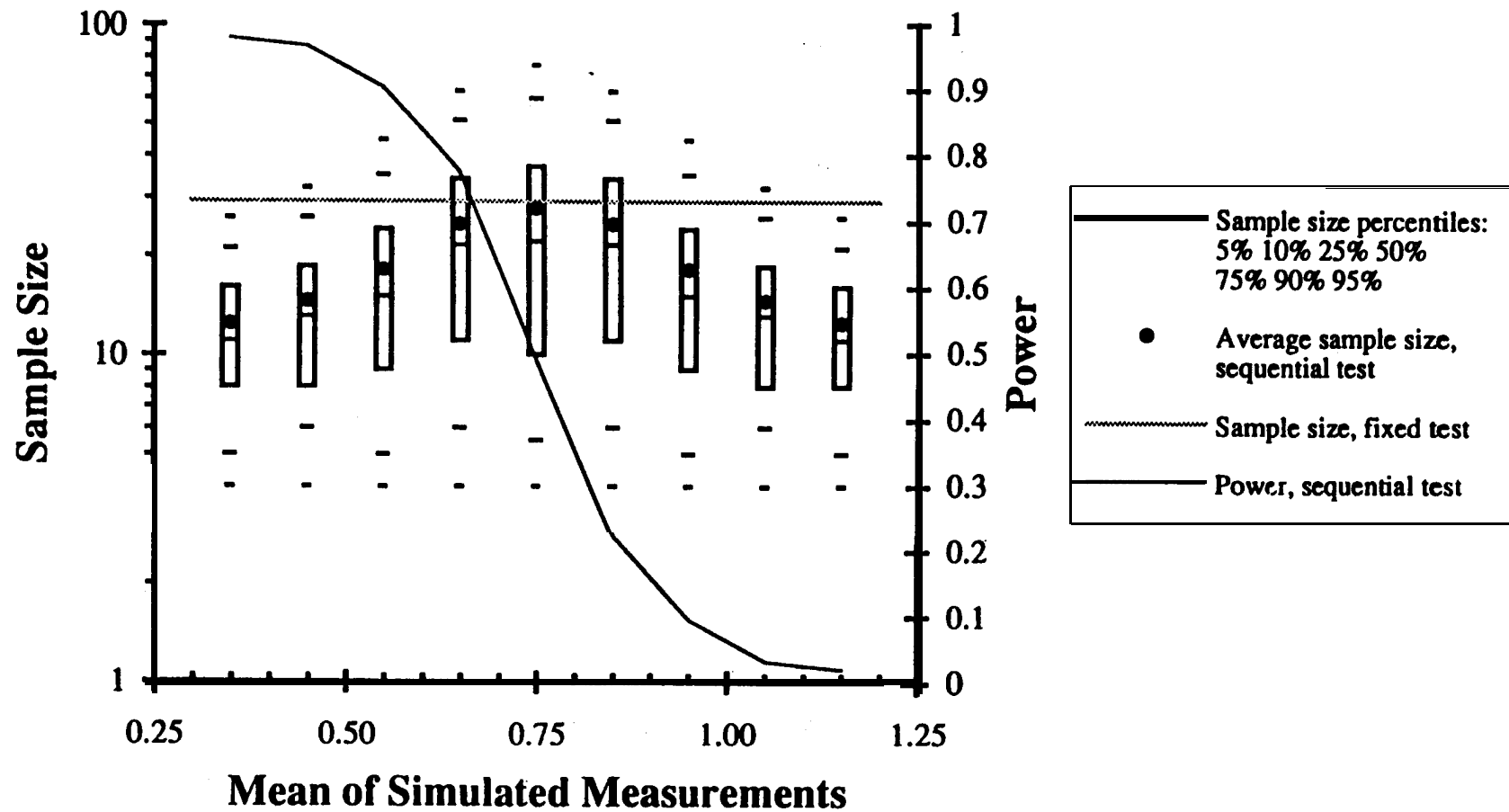
Bishop, T., 1985, Statistical View of Detection and Determination Limits in Chemical Analyses prepared for the Committee on Applications of Statistical Techniques in Chemical Problems. Columbus, Ohio: Battelle Columbus Laboratories, [January 5, 1982].

Box, **G. E. P., and Jenkins, G.M.,** 1970, Time Series Analysis Forecasting and Control. San Francisco Holden-Day.

Box, G. E. P., Hunter W. G., and Hunter, J. S., 1987, Statistics For Experimenters. New York: John Wiley and Sons, Inc.

Bross, Irwin D., 1985, "Why Proof of Safety is Much More Difficult Than Proof of Hazard," *Biometrics,* Vol. 41: 785-793.

Brown, G. H, and Fisher, N. **I.,** 1972, "Subsampling a Mixture of Sampled *Material," Technometrics,* Vol. 14, No. 3: 663-668.

Brown, M. B., and Wolfe, R. A., 1983b, "Estimation of the Variance of Percentile Estimates," Computational Stat&&s and *Data* Analysis, Vol. l: 167-174.

Brow&, K. A., 1965, Statistical Theory and Methodology in Science and Engineering, 2nd. New York: John **Wiley and Sons, Inc.**

Cantor, L. W., and Knox, **R.** C., 1986, Groundwater Pollution Control. Chelsea, Michigan: Lewis Publishers.

Cantor, L. W., Knox, R. C., and Fairchild, D. M., 1987, Groundwater Quality Protection. Chelsea, Michigan: Lewis Publishers.

Casey, D., Nemetz, P: N. and Uyeno, D., 1985, "Efficient Search Procedures for Extreme Pollutant Values," *Environmental Monitoring and Assessment,* Vol. 5: 165-176.

Clayton, C. A., Hines, J. W., Hartwell, T. D., and Burrows, P. M., 1986, Demonstration of a Technique for Estimating Detection Limits with Specified Assurance Probabilities. Washington, D.C.: EPA, [March 1986].

Cochran, W., 1977, Sampling Techniques. New York: John Wiley and Sons, Inc.

Cohen, A. C., 1961, "Tables for Maximum Likelihood Estimates: Singly Truncated and Singly Censored Samples," *Technometrics,* Vol.3, No.4: 535-541.

Conover, W. J., 1980, Practical Nonparametric Statistics. New York: John Wiley and Sons, Inc.

D'Agostino, R. B., 1970, "A Simple Portable Test of Normality: Geary's Test Revisited," *Psychological Bulletin,* Vol. 74, No. 2: 138-140.

Draper, N., and Smith, H., 1966, Applied Regression Analysis. New York: John Wiley and Sons, Inc.

DuMouchel, W.H., Govindarajulu, Z., and Rothman, E., 1973, "Note on Estimating the Variance of the Sample Mean in Stratified Sampling," *Canadian Journal of Statistics,* Vol. 1, No.2: 267-274.

Duncan, A, 1974, Quality Control and Industrial Statistics, Fourth Edition. Homewood IL: Richard Irwin, Inc.

Elder, R S., Thompson, W. O., and Myers, R. H., 1980, "Properties of Composite Sampling Procedures," *Technometrics,* Vol. 22, No. 2: 179-186.

Environ Corporation, 1985a, principles of Risk Assessment: A Nontechnical Review. EPA Workshop on Risk Assessment. Easton, Md., March 17-18, 1985.

Environ Corporation; Jellinek, Schwartz, Connolly, and Freshman; and Temple, Barker, and Sloan, Inc., 1985c, Case Study on Risk Assessment: Part I. EPA Workshop on Risk Assessment. Easton, Md., March 17-18, 1985.

Environ Corporation; Jellinek, Schwartz, Connolly, and Freshman; and Temple, Barker, and Sloan, Inc., 1985b, Additional Data on the Risk Assessment Case: Part II. EPA Workshop on Risk Management. Easton, Md., March 17-18, 1985.

Fairbanks, K., and Madsen, R., 1982, "P Values for Tests Using a Repeated Significance Test Design," *Biometrika,* Vol. 69, No. 1: 69-74.

Farrell, R., 1980, Methods for Classifying Changes in Environmental Conditions, Technical Report **VRF-EPA7.** 4-FR8O-1, Vector Research Inc., Ann Arbor, Michigan.

Filliben, J J. 1975, "Probability Plot Correlation Coefficient Test for Normality," *Technomerics,* Vol. 17, No.1: 111-117.

Ford, P., **Turina, P., and** GCA **Corporation,** 1985, Characterization of Hazardous Waste Sites-A Methods Manual, Volume I-Site Investigations. Las Vegas, Nevada: EPA Environmental Monitoring Systems Laboratory, [April 1985].

Fuller. F., and Tsokos, C., 1971, 'Time Series Analysis of Water Pollution Data," *Biometrics,* Vol. 27: 1017-1034.

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

Gastwirth, J.L., and Rubin, H., 1971, "Effect of Dependence on the Level of Some One-Sample Tests," *Journal of the American Statistical Association,* Vol. 66: 816-820

Geraghty and Miller, Inc., 1984, "Annual Report, August 1984, Rollins Environmental Services, Baton Rouge, Louisiana," Baton Rouge, Louisiana.

Ghosh, B. K. 1970, Sequential Tests of Statistical Hypotheses, Reading MA, Addison

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

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

Gilliom, R. J., and Helsel, D. R., 1986, "Estimation of Distributional Parameters for Censored Trace Level Water Quality Data, 1. Estimation Techniques," *Water Resources Research,* Vol. 22, No. 2 : 135-146.

Gleit, A., 1985, "Estimation for Small Normal Data Sets with Detection Limits," *Environmental Science and Technology,* Vol. 19, No. 12: 1201-1206.

Goldstein, B., 1985, Elements of Risk Assessment. EPA Risk Assessment Conference. Easton, Md., March 18, 1985.

Goodman, I., 1987, "Graphical and Statistical Methods to Assess the Effect of Landfills on Groundwater Quality," Land Resources Program, University of Wisconsin-Madison.

Grant, E. L., and Leavenworth, R. S., 1980, Statistical Quality Control Fifth Edition. New York: McGraw-Hill.

Groeneveld, L., and Duval, R., 1985, "Statistical Procedures and Considerations for Environmental Management (SPACEMAN)," Prepared for the Florida Department of Environmental Regulation, Tallahassee, Florida.

Grubbs, F. E., 1969, "Procedures for Detecting Outlying Observations in Samples," *Technometrics,* Vol. 11, No.1: l-21.

Guttman, I., 1970, Statistical Tolerance Regions: Classical and Bayesian. (Being Number Twenty-Six of Griffin's Statistical Monographs and Courses edited by Alan Stuart.) Darien, Conn.: Hafner Publishing.

Hall, W. J., 1962, "Some Sequential Analogs of Stein's Two Stage Test," *Biometrika,* Vol 49,: 367-378.

Hansen, M., Hurwitz, W, and Madow, W., 1953, Sample Survey Methods and Theory, Volume 1. New York: John Wiley and Sons, Inc.

Hayre, L. S., 1983, *An Alternative to the Sequential T-Test." *Sunkhya : The Indian Journal of Statistics, 45, Series* A, Pt. 3, 288-300

Hazardous Materials Control Research Institute, 1985, 6th National Conference on Management of Uncontrolled Hazardous Waste Sites. Washington, D.C.: HMCRI, [November 4-6, 1985].

Hazardous Materials Control Research Institute, 1986, 7th National Conference on Management of Uncontrolled Hazardous Waste Sites. Washington, D.C.: HMRCI, [December l-3, 1986].

Hazardous Materials Control Research Institute, 1988, 9th National Conference on Management of Uncontrolled Hazardous Waste Sites. Washington, D.C.: HMRCI, [November 28-30, 1988].

Helsel, D.R., and Cohn, T.A., 1988, "Estimation of Descriptive Statistics for Multiply *Censored* Water Quality Data," *Water Resources Research,* Vol. 24, No. 12: 1997-2004.

Helsel, D. R., and Gilliom, R., 1986, "Estimation of Distributional Parameters for Censored Trace Level Water Quality Data, 2. Verification and Applications," *Water Resources Research,* **Vol. 22, No. 2: 147-155.**

Hem, J.D., 1989, Study and Interpretation of the Chemical Characteristics of Natural Water, Third Edition, U.S. Geological Survey Water-Supply Paper 2254.

Hipel, K., Lennox, W., Unny, T., and McLeod, A., 1975, "Intervention Analysis in Water Resources," *Water Resources Research,* Vol. 11, No. 3: 567-575.

Hipel, K., McLeod, A., and Lennox, W, 1977, "Advance in Box-Jenkins Modeling, 1. Model Construction," *Water Resources Research,* **Vol. 12, No. 6: 855-861.**

Hirsch, R. M., and Slack, J. R., 1984, "A Nonparametric Trend Test for Seasonal Data with Serial Dependence," *Water Resources Research,* Vol. 20, No. 6: 727-732.

Hirsch, R. M., Slack, J. R. and Smith R.A., 1982, 'Techniques for Trend **Analysis** for Monthly Water Quality Data," *Water Resources Research,* Vol. 18, No. 1: 107-121.

Hoaglin, D.C., Mosteller, F., and Tukey, J.W., 1983, Understanding Robust and Exploratory Data **Analysis. New York:** John Wiley and Sons, Inc.

Johnson, N. L., and Kotz, S., 1970, Distributions in Statistics: Continuous Univariate Distributions - 2, Houghton Mifflin Co.

Johnson, Norman L. and Leone, F.C., 1977, *Statistics and Experimental Design in Engineering and the Physical Sciences.* Vol. I, Second Edition. Johns Wiley and Sons, Inc.

Joiner, B.L., and Rosenblatt, J.R., 1975, "Some Properties of the Range in Samples from Tukey's Symmetric Lambda Distributions," Journal *of the* American Statistical Association Vol. 66: 394.

Kedem, B., 1980, "Estimation of Parameters in Stationary Autoregressive Processes After Hard Limiting," *Journal **of the American** Statistical Association,* Vol. 75, No. 369: 146-153.

Land, C. E., 1971, "Confidence Intervals for Linear Functions of the Normal **Mean** and Variance," *Annals of Mathematical Statistics,* Vol. 42 No.4 1187-1205.

Land, C. E., 1975, Tables of Confidence for Linear Functions of the Normal Mean and Variance. ***Selected Tables in Mathematical Statistics,*** Vol. III, pp. 385-419. Providence, R.I.: American Mathematical Society.

Lehmann, EL, 1975, Nonparametrics: Statistical Methods -Based on Ranks. San Francisco: Holden-Day.

Lettenmaier, D., 1976, "Detection of Trends in Water Quality Data From Records **With** Dependent Observations," ***Water Resources Research,*** Vol. 12, No. 5: 1037-1046.

Liebetrau, A. M., 1979, "Water Quality Sampling: Some Statistical Considerations," ***Water Resources Research,*** Vol. 16, No. 6: 1717-1725.

Liggett, W., 1985, "Statistical Designs for Studying Sources of Contamination," Quality Assurance for Environmental Measurements, ASTM STP 867, J. K. Taylor and T. W. Stanley Eds. American Society for Testing Materials, Philadelphia, 22-40.

Locks, M. O., Alexander, M. J., and Byars, B. J., 1963, "New Tables of the Noncentral t-Distribution," Report ARL63-19, Wright-Patterson Air Force Base.

Loftis, J. and Ward, R., 1980, "Sampling Frequency for Regulatory Water Quality Monitoring," ***Water Resources Bulletin,*** Vol. 16, No. 3: 501-507.

Loftis, J. and Ward, R., 1980, "Water Quality Monitoring-Some Practical Sampling Frequency Considerations," ***Environmental Management,*** Vol. 4, No. 6: 521-526.

Loftis, J., Montgomery, R., Harris, J., Nettles, D., **Porter, P., Ward, R., and Sanders, T.,** 1986, "Monitoring Strategies for Ground Water Quality Management," Prepared for the United States Geological Service by Colorado State University, Fort Collins Colorado.

Madow, W. C., and Madow, L. H., 1944, "On the Theory of Systematic Sampling" ***Annals of Mathematical Statistics*** Vol. 15: l-24.

Mage, D. T., 1982, "Objective Graphical Method for Testing Normal Distributional Assumptions Using. Probability Plots," ***American Statistician,*** Vol. 36, No.2: 116-120.

McLeod, **A.,** Hipel, K., **and Comancho, F.,** 1983, "Trend Assessment of Water Quality Time Series," ***Water Resources Bulletin,*** Vol. 19, No. 4: 537-547.

McLeod, **A.,** Hipel, K., Lennox, W., 1977, "Advance in Box-Jenkins Modeling, 1. Applications," ***Water Resources Research,*** Vol. 13, No. 3: 577-586.

Mee, Robert W., 1984, 'Tolerance Limits and Bounds for Proportions Based on Data Subject to Measurement Error," ***Journal of Quality Technology.*** Vol.16, No.2: 74-80.

Mee, **Robert W., Owen, D.B., and Shyu, Jyh-Cherng.,** 1986, "Confidence Bounds for Misclassification Probabilities Based on Data Subject to Measurement Error"' ***Journal of*** Quality ***Technology,*** Vol. 18, No. 1: 29-40.

Mendenhall, W., and Ott, L., 1980, Understanding Statistics. N. Scituate, Mass.: Duxbury Press.

Millard, S., Yearsley, J., and Lettenmaier, D., 1985, "Space-Time Correlation and Its Effects on Methods for Detecting Aquatic Ecological Change"' CAN. J. FISH. AQUAT. SCI, Vol. 42: 1391-1400.

Montgomery, R., and Loftis, J., 1987, "Applicability of the t-test for Detecting Trends in Water Quality Variables"' ***Water Resources Bulletin,*** Vol. 23, No. 4: 6 5 3 - 6 6 2 .

Montgomery, R., and Reckhow, H., 1984, "Techniques for Detecting Trends in Lake Water Quality," ***Water Resources Bulletin,*** Vol. 20, No. 1: 43-52.

Natrella, M., 1963, Experimental Statistics. Washington, D.C.: U.S. Department of Commerce, National Bureau of Standards.

Nelson, J., and Ward, R., 1981, "Statistical Considerations and Sampling Techniques for Ground-Water Quality Monitoring," Ground Water, Vol. 19, No. 6: 617-625.

Neter, J., Wasserman, W., and Kutner, M., 1985, Applied Linear Statistical Models. Homewood Illinois: Irwin.

Neter, J., Wasserman, W., and Whitmore, G., 1982, Applied Statistics. Boston: AlIyn-Bacon.

Noether, C., 1956, "Two Sequential Tests Against Trend"' ***Journal of the American Statistical Association,*** September 1956: 440-450.

Ness, **R.,** 1985, "Groundwater Quality," ***Journal-Water Pollution Control Fe&ration,*** Vol. 57, No. 6: 642-649.

Nyer, Evan K., 1985, Groundwater Treatment Technology. New York: Van Nostrand Reinhold Co.

Oak Ridge National Laboratory, 1984, "Results of the Groundwater Monitoring Performed at the Former **St.** Louis Airport Storage Site for the Period January 1981 Through January 1983," ORNL/TM-8879, Oak Ridge, Tennessee.

Owen, D.B., 1963, Factors for One-Sided Tolerance Limits and for Variables Sampling Plans. Albuquerque, N.M.: Sandia Corporation, [March 1963].

Patel, J.K., 1986, "Tolerance Limits-A Review," ***Communications in Statistics: Theory and Methods,*** Vol. 15, No. 9: 2719-2762.

BIBLIOGRAPHY

Pederson, G.L., and Smith, M.M., 1989, U.S. Geological Survey Second National Symposium of Water Quality; Abstracts of the Technical Sessions. Orlando, Florida, November 1989.

Pettyjohn, W., 1976, "Monitoring Cyclic Fluctuations in Ground-Water Quality," *Ground Water,* Vol. 14, No. 6: 472-480.

Pucci, A., and Murashige, J., 1987, "Applications of Universal Kriging to an Aquifer Study in New Jersey," *Ground Water,* Vol. 25, No. 6: 672-678.

Rendu, J. **M.,** 1979, "Normal and Lognormal Estimation," *Mathematical Geology,* Vol. 11, No.4: 407-422.

Resnikoff, G. J., and Lieberman, G. **J.,** 1957, Tables of the Non-central t-distribution. Stanford: Stanford University Press.

'Rockwell International, 1979, "Hanford Groundwater Modeling-Statistical Methods for Evaluating Uncertainty and Assessing Sampling Effectiveness," Rockwell Hanford Operations, Energy Systems Group, Richland Washington, RHO-C-18.

Rohde, C. **A.,** 1976, "Composite Sampling," *Biometrics,* Vol. 32: 278-282.

Rohde, C. **A.,** 1979, "Batch, Bulk, and Composite Sampling," *Sampling Biological Popularions,* pp. 365-367. Edited by R.M. Cormack. Fairland, Md.: International Cooperative Pub. House.

Rushton, S.. 1950, "On a Sequential t-Test," *Biometrika,* Vol. 37: 326-333.

Rushton, S., 1952, "On a Two-Sided Sequential t-Test," *Biometrika,* Vol. 39: 302-308.

Sanders, T. and Adrian, D., 1978, "Sampling Frequency for River Quality Monitoring," *Water Resources Research,* Vol. 14, No. 4: 569-576.

SAS Institute, 1985, Sas Users Guide: Statistics. Cary, North Carolina.

Schaeffer, D. and Kerster, H., 1988, "Quality Control Approach to NPDES Compliance Determination," *Journal-Water Pollution Control Fe&ration,* Vol. 60: 1436-1438.

Scheaffer, R. L., Mendenhall, W., and Ott, L., 1979, Elementary Survey Sampling, Second Edition. Boston: Duxbury Press.

Schmid, C. **F.,** 1983, Statistical Graphics: Design Principles and Practices. New York: John Wiley and Sons, Inc.

Schmidt, K., 1977, "Water Quality Variations for Pumping Wells," Ground *Water,* Vol. 15, No. 2: 130-137.

Schwartz, J. E., 1985, "Neglected Problem of Measurement Error in Categorical Data," *Sociological Methods and Research,* Vol. 13, No. 4: 435-466.

BIBLIOGRAPHY

Schweitzer, G. E., and Santolucito, J. A., editors, 1984, Environmental Sampling for Hazardous Wastes. **ACS** Symposium Series 267. Washington, D.C.: American Chemical Society.

Shapiro, S.S., and Wilk, M.B., 1965, "Analysis of Variance Test for Normality (Complete Samples)," *Biometrika,* Vol. 52: 591-611.

Sharpe, K., 1970, "Robustness of Normal Tolerance Intervals," *Biometrika.* Vol. 57, No.1: 71-78.

Siegmund, D., 1985, Sequential Analysis: Tests and Confidence Intervals. New York: Springer-Verlag.

Sirjaev, A.N., 1973, Statistical Sequential Analysis. Providence, R.I.: American Mathematical society.

Size, W. B., editor, 1987, Use and Abuse of Statistical Methods in the Earth Sciences. New York: Oxford University Press.

Snedecor, G. W., and Cochran, W. G., 1980, Statistical **Method.** Seventh Edition Ames Iowa: The Iowa State Press.

Sokal, R. R., and Rohlf, F. J., 1981, Biometry: The Principles and Practice of Statistics in Biological Research. Second Edition. New York: W. I-L Freeman.

Stoline M., and Cook, R., 1986, "A Study of Statistical Aspects of the Love Canal Environmental Monitoring Study," American Statistician, Vol. 40, No. 2: 172-

Switzer, P., 1983, When Will a Pollutant Standard Be Exceeded: Model Prediction and Uncertainty. Technical Report No. 67. New Canaan, Ct.: SIMS, [January

Temple, Barker, and Sloan, Inc. 1986, Case Study on Risk Management. EPA Workshop on Risk Management. Easton, Md., April 13-14, 1986.

Tomqvist, L., 1963, "Theory of Replicated Systematic Cluster Sampling With Random *Stan," Reviewof the International Statistical Institute,*ol. 31, No. 1: 11-23.

Tukey, J. W., 1977, Exploratory Data Analysis. Reading, Mass.: Addison-Wesley.

U.S. Congress. Office of Technical Assessment, 1985, Superfund Strategy. Washington, D.C.: G.P.O.

U.S. Department of Energy, 1985a How Clean is Clean: **A** Review of Superfund Cleanups. Washington D. C. (GJ/TMC-08-ED.2).

U.S. Department of Energy, 1985b, Procedures for Collections and Preservation of Groundwater and Surface Water Samples and For the Installation of Monitoring Wells. Washington D. C. (CONF-87 1075-21).

U.S. Environmental Protection Agency, 1982, The Handbook for Sampling and Sample Preservation of Water and Wastewater. Washington, D. C., September 1982   (EPA-600/4-82-029).

U.S. Environmental Protection Agency, 1984, Sampling Procedures for Ground Water Quality Investigations. Washington D.C., May 1984 (EPA-600/D-84-137).

U.S. Environmental Protection Agency, 1985a, Data Quality Objectives for the RI/FS Process: Accuracy Testing Definitions, Appendix F [draft]. Washington, D.C., [November 5, 1985].

U.S. Environmental Protection Agency, 1985b, EPA Guide for Minimizing the Adverse Environmental Effects of Cleanup of Uncontrolled Hazardous Wastes Sites. Washington, D.C., June 1985 (EPA/600/8-85/008).

U.S. Environmental Protection Agency, 1986a, Guidance **Document** for Cleanup of Surface Impoundment Sites. Washington D. C., June 1986.

U.S. Environmental Protection Agency, 1986b, Resource Conservation and Recovery Act (RCRA) Ground-Water Monitoring Technical Enforcement Guidance Document. Washington D.C., September 1986 (OSWER-9950.1).

U.S. **Environmental Protection Agency,** 1986c, Superfund Public Health Evaluation Manual. Washington D. C.: EPA [October 1986].

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

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

U.S. **Environmental Protection Agency,** 1987c, EPA Journal, The New Superfund: Protecting People and the Environment. Washington D. C., Vol. 13, No. 1.

U.S. **Environmental Protection Agency,** 1987d, Surface Impoundment Clean Closure Guidance Manual [draft]. Washington D. C., March 1987.

U.S. Environmental Protection Agency, 1987e, Using Models in Ground-Water Protection Programs. Washington D.C., January 1987 (EPA/600/8-87/003).

U.S. Environmental Protection Agency, 1988, Guidance on Remedial Actions for Contaminated Ground Water at Superfund Sites [Interim Final]. Washington D. C.

BIBLIOGRAPHY

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

U.S. Environmental Protection Agency, 1989b, Statistical Analysis of Ground-
Water Monitoring Data at RCRA Facilities. Office of Solid Waste,
Washington, D.C., April 1989

van Belle, G., and Hughes, J. **P.,** 1984, "Nonparametric Tests for Trend in Water
Quality," *Water Resources Research,* Vol. 20, No. 1: 127-136.

Wald, A., 1947, Sequential Analysis. New York Dover Publications.

Ward, C., Loftis, J., Nielsen, K., and Anderson, R., 1979, "Statistical
Evaluation of Sampling Frequencies in Monitoring Networks," *Journal-Water
Pollution Control Federation,* **Vol. 51, No. 9: 2292-2300.**

Wesolowsky, G.O., 1976, Multiple Regression and Analysis of Variance. New York:
John Wiley.

Wetherill, G. **B.,** 1975, Sequential Methods in Statistics. New York: Halsted Press.

Wilson, J., 1982, Ground Water: A Non-Technical Guide. Academy of Natural
Sciences, Philadelphia, Pa.

Wolter, K. M., 1984, "Investigation of Some Estimators of Variance for Systematic
Sampling," *Journal of* the *American Statistical Association,* Vol. 79, No.388:
781-790.

Wolter, Kirk M.,, 1985, Introduction to Variance Estimation. New York: Springer-
Verlag.

Wood, E. F., Ferrara, R. A., **Gray, W. G., and Pinder, G. F.,** 1984,
Groundwater Contamination From Hazardous Wastes. Englewood Cliffs, New
Jersey: Prentice-Hall.

Table A.1   Tables of t for selected alpha and degrees of freedom

Use alpha to determine which column to use based on the desired parameter, $t_{1-\alpha,Df}$ or $t_{1-\alpha/2,Df}$.
Use the degrees of freedom to determine which row to use. The t value will be found at the
intersection of the row and column. For values of degrees of freedom not in the table, interpolate
between those values provided.

		When determining $t_{1-\alpha,Df}$ for α specified as:							
		.25	.10	.05	.025	.01	.005	.0025	.001
		When determining $t_{1-\alpha/2,Df}$ for α specified as:							
		.50	.20	.10	.05	.02	.01	.005	.002
	Df								
	1	1.000	3.078	6.314	12.706	31.821	63.657	127.321	318.309
	2	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327
Degrees of	3	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.215
Freedom	4	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173
Df	5	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893
	6	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208
	7	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785
	8	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501
	9	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297
	10	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144
	11	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025
	12	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930
	13	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852
	14	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787
	15	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733
	16	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686
	17	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646
	18	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610
	19	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579
	20	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552
	21	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527
	22	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505
	23	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485
	24	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467
	25	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450
	26	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435
	27	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421
	28	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408
	29	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396
	30	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385
	40	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307
	60	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232
	120	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160
	400	0.675	1.284	1.649	1.966	2.336	2.588	2.823	3.111
	infinite	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090

Table A.2   Tables of z for selected alpha

Use alpha to determine which column to read. Use the desired parameter, $z_{1-\alpha}$ or $z_{1-\alpha/2}$, to determine which row to use. Read the z value at the intersection of the row and column,

$z_{1-\alpha}$	$z_{1-\alpha/2}$
0.674	1.150
.842	1.282
1.282	1.654
1.645	1.960
1.960	2.326
2.326	2.576
2.576	2.807
2.807	3.090
3.090	3.29

Table A.3   Tables of k for selected alpha, P_0, and sample size for use in a tolerance interval test

Use alpha to determine which table to read. The value k is found at the intersection of the column with the specified P_0 and the row with the sample size n. When testing tolerance intervals, let $T = \bar{x} + ks$. If T is less than the cleanup standard, the sample area attains the cleanup standard based on the statistical test.

Alpha = 0.10   (i.e., 10%)

n	P_0			
	0.25	0.1	0.05	0.01
2	5.842	0.253	3.090	8.500
3	2.603	4.258	5.311	7.340
4	1.972	3.188	3.957	5.438
5	1.698	2.742	3.400	4.666
6	1.540	2.494	3.092	4.243
7	1.435	2.333	2.894	3.972
8	1.360	2.219	2.754	3.783
9	1.302	2.133	2.650	3.641
10	1.257	2.066	2.568	3.532
11	1.219	2.011	2.503	3.443
12	1.188	1.966	2.448	3.371
13	1.162	1.928	2.402	3.309
14	1.139	1.895	2.363	3.257
15	1.119	1.867	2.329	3.212
16	1.101	1.842	2.299	3.172
17	1.085	1.819	2.272	3.137
18	1.071	1.800	2.249	3.105
19	1.058	1.782	2.227	3.077
20	1.046	1.765	2.208	3.052
21	1.035	1.750	2.190	3.028
22	1.025	1.737	2.174	3.007
23	1.016	1.724	2.159	2.987
24	1.007	1.712	2.145	2.969
25	1.000	1.702	2.132	2.952
26	0.992	1.691	2.120	2.937
27	0.985	1.682	2.109	2.922
28	0.979	1.673	2.099	2.909
29	0.973	1.665	2.089	2.896
30	0.967	1.657	2.080	2.884
35	0.942	1.624	2.041	2.833
40	0.923	1.598	2.010	2.793
50	0.894	1.559	1.965	2.735
70	0.857	1.511	1.909	2.662
100	0.825	1.470	1.861	2.601
200	0.779	1.411	1.793	2.514
500	0.740	1.362	1.736	2.442
infinity	0.674	1.282	1.645	2.326

Table A.3   Tables of k for selected alpha, P0, and sample size far use in a toleranace interval test
(Continued)

Alpha = 0.05 (i.e., 5%)

n	P0			
	0.25	0.1	0.05	0.01
2	1.763	20.581	26.260	37.094
3	3.806	6.155	7.656	10.553
4	2.618	4.162	5.144	7.042
5	2.150	3.407	4.203	5.741
6	1.895	3.006	3.708	5.062
7	1.732	2.755	3.399	4.642
8	1.618	2.582	3.187	4.354
9	1.532	2.454	3.031	4.143
10	1.465	2.355	2.911	3.981
11	1.411	2.275	2.815	3.852
12	1.366	2.210	2.736	3.747
13	1.328	2.155	2.671	3.659
14	1.296	2.109	2.614	3.585
15	1.268	2.068	2.566	3.520
16	1.243	2.033	2.524	3.464
17	1.220	2.002	2.486	3.414
18	1.201	1.974	2.453	3.370
19	1.183	1.949	2.423	3.331
20	1.166	1.926	2.396	3.295
21	1.152	1.905	2.371	3.263
22	1.138	1.886	2.349	3.233
23	1.125	1.869	2.328	3.206
24	1.114	1.853	2.309	3.181
25	1.103	1.838	2.292	3.158
26	1.093	1.824	2.275	3.136
27	1.083	1.811	2.260	3.116
28	1.075	1.799	2.246	3.098
29	1.066	1.788	2.232	3.080
30	1.058	1.777	2.220	3.064
35	1.025	1.732	2.167	2.995
40	0.999	1.697	2.125	2.941
50	0.960	1.646	2.065	2.862
70	0.911	1.581	1.990	2.765
100	0.870	1.527	1.927	2.684
200	0.809	1.450	1.837	2.570
500	0.758	1.385	1.763	2.475
infinity	0.674	1.282	1.645	2.326

Table A.3   Tables of k for selected alpha, P$_0$, and sample size far use in a tolerance interval test (Continued)

Alpha = 0.01 (i.e., 1%)

n	P$_0$			
	0.25	0.1	0.05	0.01
2	58.939	103.029	131.426	185.61
3	8.728	13.995	17.370	23.896
4	4.715	7.380	9.083	12.387
5	3.454	5.362	6.578	8.939
6	2.848	4.411	5.406	7.335
7	2.491	3.859	4.728	6.412
8	2.253	3.497	4.258	5.812
9	2.083	3.240	3.972	5.389
10	1.954	3.048	3.738	5.074
11	1.853	2.898	3.556	4.829
12	1.771	2.777	3.410	4.633
13	1.703	2.677	3.290	4.472
14	1.645	2.593	3.189	4.337
15	1.595	2.521	3.102	4.222
16	1.552	2.459	3.028	4.123
17	1.514	2.405	2.963	4.037
18	1.481	2.357	2.905	3.960
19	1.450	2.314	2.854	3.892
20	1.423	2.276	2.808	3.832
21	1.399	2.241	2.766	3.777
22	1.376	2.209	2.729	3.727
23	1.355	2.180	2.694	3.681
24	1.336	2.154	2.662	3.640
25	1.319	2.129	2.633	3.601
26	1.303	2.105	2.606	3.566
27	1.287	2.085	2.581	3.533
28	1.273	2.065	2.558	3.502
29	1.260	2.047	2.536	3.473
30	1.247	2.030	2.515	3.447
35	1.195	1.957	2.430	3.334
40	1.154	1.902	2.364	3.249
50	1.094	1.821	2.269	3.125
70	1.020	1.722	2.153	2.974
100	0.957	1.639	2.056	2.850
200	0.868	1.524	1.923	2.679
500	0.794	1.430	1.814	2.540
infinity	0.674	1.282	1.645	2.326

Table A.4   Recommended number of samples per seasonal period (n_p) to minimize total cost for assessing attainment

		Estimated Lag 1 serial correlation between monthly observations										
		0.05	0.1	0.15	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	1	8	7	6	5	4	4	4	4	4	4	4
	2	10	8	7	6	5	4	4	4	4	4	4
	5	12	10	9	8	6	5	4	4	4	4	4
Cost ratio $\$_R$	10	15	12	10	9	8	6	5	4	4	4	4
	20	18	15	13	11	9	8	6	5	4	4	4
Yearly cost	50	23	20	17	15	13	10	9	7	6	4	4
Sample cost	100	30	24	21	19	16	13	11	9	7	5	4
	200	36	30	26	24	20	16	14	11	9	6	4
	1000	61	52	46	40	34	28	23	19	15	11	7
	2000	73	61	61	52	40	36	30	24	19	14	8
	5000	91	91	73	73	61	46	40	32	25	19	11
	10000	183	91	91	91	73	61	52	40	32	23	14

Table A.5   Variance factors F for determining sample size

		Estimated Lag 1 serial correlation between monthly observations										
		0.05	0.1	0.15	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	4	4.00	3.99	3.97	3.94	3.80	3.53	3.13	2.61	1.99	1.31	0.64
	5	4.99	4.96	4.90	4.80	4.49	4.03	3.44	2.77	2.05	1.33	0.64
	6	5.97	5.89	5.75	5.55	5.04	4.38	3.64	2.87	2.09	1.34	0.64
	7	6.92	6.74	6.50	6.19	5.46	4.64	3.78	2.93	2.11	1.35	0.64
	8	7.83	7.53	7.15	6.73	5.80	4.83	3.88	2.97	2.13	1.35	0.64
	9	8.69	8.23	7.71	7.17	6.05	4.97	3.95	3.00	2.14	1.35	0.64
	10	9.48	8.85	8.19	7.53	6.26	5.08	4.00	3.03	2.15	1.36	0.64
	11	10.22	9.40	8.60	7.83	6.42	5.16	4.04	3.04	2.15	1.36	0.64
Samples	12	10.89	9.88	8.95	8.09	6.55	5.23	4.07	3.06	2.16	1.36	0.64
per year or	13	11.51	10.30	9.24	8.30	6.66	5.28	4.09	3.07	2.16	1.36	0.64
seasonal	14	12.07	10.67	9.50	8.47	6.75	5.32	4.11	3.07	2.16	1.36	0.64
period	15	12.57	11.00	9.72	8.62	6.82	5.35	4.13	3.08	2.17	1.36	0.64
	16	13.03	11.28	9.90	8.75	6.88	5.38	4.14	3.09	2.17	1.36	0.64
	17	13.44	11.53	10.07	8.86	6.93	5.41	4.15	3.09	2.17	1.36	0.64
	18	13.81	11.75	10.21	8.96	6.97	5.43	4.16	3.09	2.17	1.36	0.64
	19	14.15	11.95	10.33	9.04	7.01	5.45	4.17	3.10	2.17	1.36	0.64
	20	14.45	12.12	10.44	9.11	7.05	5.46	4.18	3.10	2.17	1.36	0.64
	21	14.72	12.27	10.54	9.17	7.07	5.47	4.18	3.10	2.17	1.36	0.64
	22	14.97	12.41	10.62	9.23	7.10	5.49	4.19	3.10	2.18	1.36	0.64
	23	15.20	12.53	10.70	9.28	7.12	5.50	4.19	3.11	2.18	1.36	0.64
	24	15.41	12.65	10.77	9.32	7.14	5.50	4.20	3.11	2.18	1.36	0.64
	25	15.59	12.75	10.83	9.36	7.16	5.51	4.20	3.11	2.18	1.36	0.64
	26	15.76	12.84	10.88	9.39	7.17	5.52	4.20	3.11	2.18	1.36	0.64
	28	16.06	12.99	10.98	9.45	7.20	5.53	4.21	3.11	2.18	1.36	0.64
	30	16.32	13.12	11.05	9.50	7.22	5.54	4.21	3.11	2.18	1.36	0.64
	32	16.53	13.23	11.12	9.54	7.24	5.55	4.22	3.12	2.18	1.36	0.64
	34	16.71	13.32	11.17	9.58	7.25	5.56	4.22	3.12	2.18	1.36	0.64
	36	16.87	13.40	11.21	9.60	7.26	5.56	4.22	3.12	2.18	1.36	0.64
	40	17.13	13.52	11.29	9.65	7.28	5.57	4.22	3.12	2.18	1.36	0.64
	46	17.40	13.66	11.36	9.70	7.30	5.58	4.23	3.12	2.18	1.36	0.64
	52	17.59	13.75	11.42	9.73	7.32	5.58	4.23	3.12	2.18	1.37	0.64
	61	17.79	13.84	11.47	9.76	7.33	5.59	4.23	3.12	2.18	1.37	0.64
	73	17.95	13.91	11.51	9.79	7.34	5.60	4.24	3.12	2.18	1.37	0.64
	91	18.08	13.98	11.54	9.81	7.35	5.60	4.24	3.12	2.18	1.37	0.64
	183	18.27	14.06	11.59	9.84	7.36	5.60	4.24	3.13	2.18	1.37	0.64
	365	18.31	14.08	11.60	9.85	7.37	5.61	4.24	3.13	2.18	1.37	0.64

APPENDIX B: EXAMPLE WORKSHEETS

The worksheets in this appendix have been completed to serve as an example in understanding the forms and making the necessary calculations.

Please note that to maintain adquate precision in doing the computations appearing in the worksheets, (particularly in the calculations of estimated variances, standard deviations, or standard errors), the number of decimal places retained should be as high as possible, with a minimum of four.

A Scenario

To help understand how to use the worksheets provided, a scenario has been constructed with associated data concerning a site for which a cleanup effort has been undertaken. In order that undue time is not spent on data manipulation and data entry, parameters were set in such a way that the number of years for which data needed to be collected in the example was kept artificially low. For example, in Worksheet 3, α and β were set higher than will generally be the case in practice while μ_1 and δ were set relatively low. As a consequence, the number of years required for a fixed sample size test was limited to three years, which is highly unlikely to be the case in practice.

The scenario involves a Superfund site with a treatment well and 5 monitoring wells. Two of the' monitoring wells are close to the source of contamination and have been monitored individually (involving Worksheets 2 through 7b). The remaining three wells are relatively far from the source of contamination and have been analyzed as a group (Worksheets 8 through 14b). Two chemicals were of interest in monitoring for cleanup. The example worksheets have been provided for one of the two chemicals for one of the two wells being monitored individually and for the group of three wells. For illustrative purposes, for the single well being examined, both a fixed sample test and a sequential test have been carried out. However, in practice, a decision would be made before hand about which of the two approaches would be used, and only that test would be employed. It is interesting to note that, for the example data set, it turns out that the fixed sample size test indicates that the site is clean while the sequential test indicates that more data are needed before a decision can be reached. On average, the sequential test will yield a result more quickly, but since the parameters were specified so as to require only-three years for the fixed sample test, which is the minimum amount of time required

for a sequential test, it is not altogether surprising that a decision could not be made via the sequential test

Worksheets 15 and 16 have been filled out with data independent of the five well example. They were used simply to indicate how a serial correlation could be estimated via the worksheets. The number of observations on which the estimated serial correlation is based, twelve, is fewer than should normally be used in practice.

The number of samples per year used in the example was six. Note that in Worksheet 3 the estimated serial correlation between monthly data was .2, so that the correlation between observations obtained between two-month periods would be estimated to be $.2^2=.04$. Since .04 represents a rather low correlation between Observations, data could be reasonably gathered on a bimonthly schedule without great concern about a lack of independence between observations.

Worksheets 1R and 2R present the computation of regression coefficients and related tests of significance using the three sample means obtained during the three years of data collection for the test of the single well to serve as the three data observations from which a linear model was to be constructed. Since the fixed sample test indicated that the cleanup effort was successful, it is desirable to examine the trend of the data over time to make sure that there is no evidence that the cleanup standard could be exceeded in the future. This could be indicated by evidence of a statistically significant positive slope for the sample data (in this case, the three yearly averages). Three observations is a rather small sample on which to base such decisions, but again the chief purpose of these example worksheets is illustrative. The reader can more quickly determine how the regression estimates were computed using a small data set. In practice, it is quite likely that the number of years' worth of data resulting in a decision that the site is clean will exceed three by several years.

Table B.l Summary of Notation Used in Appendix B

Symbol	Definition
m	The number of years for which data were collected (usually the analysis will be performed with full years worth of data)
n	The number of sample measurements per year (for monthly data, $n = 12$; for quarterly data, $n = 4$). This is also referred to as the number of "seasons" per year
N	The total number of sample measurements (if there are no missing observations, $N = mn$)
index i	Indicates the order in which the ground-water samples are collected
index k	Indicates the year in which the ground-water samples are collected
index j	Indicates the season or time within the year at which the groundwater samples are collected
index c	Indicates the chemical analyzed
index w	Indicates the well sampled
x_i	Contaminant measurement for the ith ground-water sample
x_{jk}	An alternative way of denoting a contaminant measurement, where $k = 1, 2, ..., m$ denotes the year; and $j = 1, 2, ..., n$ denotes the sampling period (season) within the year. The subscript for x_{jk} is related to the subscript for x_i in the following manner: $i = (k-1)n + j$.
$\bar{x}_k$	The mean (or average) of the contaminant measurements for year k (see Boxes 8.5 and 9.4)
$\bar{x}_m$	The mean of the yearly averages for years $k = 1$ to m.
$s_{\bar{x}}$	The standard deviation of the yearly average contaminant concentrations from m years of sample collection (see Boxes 8.7 and 9.6)
$s_{\bar{x}_m}$	The standard error of the mean of the yearly means (see Boxes 8.9 and 9.8)
Cs	The designated clean up standard
Df	The degrees of freedom associated with the standard error of an estimate (see Boxes 8.7 and 9.6)
d_j	The distance of the monitoring well from the treatment well.

WORKSHEET *1* *Sampling Wells*

See Section 3.2 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE: Site ABC

Sample Well Number w	Describe each sampling well to be used to assess attainment
1	monitoring well d_1 feet northeast of treatment well
2	monitoring well d_2 feet west of treatment well
3	monitoring well d_3 feet north of treatment well
4	monitoring well d_4 feet southwest of treatment well
5	monitoring well d_5 feet southeast of treatment well
	wells 1 and 2 will be assessed individually
	wells 3, 4, and 5 will be assessed as a group

Decision Criteria: Wells assessed (Checked one) Individually ☑ As a Group ☑

Use the Sampling Well Number (w) to refer on subsequent sheets to the sampling wells described above.

Attach a map showing the sampling wells within the waste site.

Date Completed: _EXAMPLE____   Completed by ____EXAMPLE_____

Use additional sheets if necessary.   Page ____ of ____

Continue to WORKSHEET 2 if wells are assessed individually.
Continue to WORKSHEET 8 if wells arc assessed as a group.

WORKSHEET 2 *Attainment Objectives for Assessing Individual Wells*

See Chapter 3 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:	Site ABC

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

(for purposes of illustration, both methods will be used)

Sample Design (Check one): Fixed Sample Size ☒   Sequential Sampling ☒

Probability of mistakenly declaring the well(s) clean = α = | .1 |

Probability of mistakenly declaring the well(s) contaminated = β = | .2 |

chemical Number c	Chemical Name	Cleanup Standard (with units) Cs	Parameter to test: Check one	If Mean, Enter: alternate hypothesis mean μ_1	If %tile, Enter: Critical proportion for alternate/null hypothesis null P_0	If %tile, Enter: Critical proportion for alternate/null hypothesis alternate P_1
1	Hazardous #1	100	Mean ☒ %tile ☐	75		
2	Hazardous #2	60	Mean ☒ %tile ☐	30		
			Mean ☐ %tile ☐			
			Mean ☐ %tile ☐			

Sample Collection Procedures to be used (attach separate sheet if necessary):

Not specified for this example

Secondary Objectives/Other purposes for which the data is to be collected

Use the Chemical Number (c) to refer on other sheets to the chemical described above.
Attach documentation describing the lab analysis procedure for each chemical.

Date Completed: __EXAMPLE__   Completed by __EXAMPLE__

Use additional sheets if necessary.   Page ____ of ____

Continue to **WORKSHEET 3** if a fixed sample size test is used; or
Continue to **WORKSHEET 4** if a sequential sample test is used.

WORKSHEET 3 *Sample Size When Using a Fixed Sample Size Test for Assessing Individual Wells*

See Sections 8.2 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:	Site ABC

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

From Table A.2, Appendix A

Probability of mistakenly declaring the site clean [2] = α = $\boxed{.1}$ $z_{1-\alpha}$ = $\boxed{1.282}$

Probability of mistakenly declaring the site contaminated [2] = β = $\boxed{.2}$ $z_{1-\beta}$ = $\boxed{.842}$

Number of samples per year = n = $\boxed{6}$ (based on calculations described in Section 8.2)

Variance factor from Table A.5, Appendix A = F^1 = $\boxed{5.55}$

For testing the mean concentration

Chemical Number [2] c	Cleanup Standard[2] Cs	[2] μ_1	Standard Deviation of yearly mean $\hat{\delta}$	Calculate: $B = \left(\dfrac{Cs-\mu_1}{z_{1-\alpha}+z_{1-\beta}}\right)^2$	$m_d = \dfrac{\hat{\delta}^2}{F*B} + 2$
1	100	75	23	138.53	2.69
2	60	30	6	199.50	2.03

For testing the proportion of contaminated wells or samples

Chemical Number [2] c	Cleanup Standard[2] Cs	[2] P_0	[2] P_1	Calculate: $B = \left(Z_{1-\beta}\sqrt{P_1(1-P_1)} + Z_{1-\alpha}\sqrt{P_0(1-P_0)}\right)^2$	$m_d = \dfrac{B}{F(P_0-P_1)^2}$

Column Maximum, (Maximum of m_d values) = C = $\boxed{2.69}$

Round C to next largest integer=Number of years of sample collection= m= $\boxed{3}$

Total number of samples = nm = N = $\boxed{18}$

Date Completed: __EXAMPLE__   Completed by __EXAMPLE__

Use additional sheets if necessary.   Page ____ of ____
Continue to **WORKSHEET 4**

---

[1] An estimate of ϕ, the serial correlation, is necessary to determine the appropriate value of F. Worksheets 15 and 16 can be used to estimate ϕ. ϕ = .2 was assumed for this example.

WORKSHEET 4 *Data Records and Calculations When Assessing Individual Wells; by Chemical, Well, and Year*

See Chapter 8 or 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	Site ABC
CHEMICAL: NUMBER(c) AND DESCRIPTION [2]	1. Hazardous #1
WELL: NUMBER(w) AND DESCRIPTION [1]	#1. d_1 ft. northeast of treatment well
YEAR: NUMBER(K)	1988, K = 1

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Sample Design (Check one): Fixed Sample Size ☑   Sequential Sampling ☑
For purposes of illustration, both methods are used.

Parameter to be tested [2] (Check one) =	Mean ☑ %tile ☐
Number of samples per year [3] = n =	6
Number of samples with nonmissing data in year = n_k =	6
Cleanup standard[2] = Cs=	100
Concentration used for observations below the detection limit =	10

"Season" Number: j within this k^{th} year	Sample ID	Sample Collection date/time	Reported Concentration	Concentration Corrected for Detection Limit A	Is A Greater than Cs? 1 = Yes 0 = No B	Data for analysis x_{jk} = A if Mean √ x_{jk} = B if %tile
1	11	Feb. 18, '88	88	88		88
2	21	April 12, '88	123	123		123
3	31	June 16, '88	98	98		98
4	41	Aug. 15, '88	78	78		78
5	51	Oct. 12, '88	89	89		89
6	61	Dec. 11, '88	65	65		65

Total of x_{jk} for this year = C = | 541 |

Mean of x_{jk} for this k^{th} year = $\dfrac{C}{n_k}$ = $\bar{x}_k$ = | 90.17 |

Date Completed: __EXAMPLE__   Completed by __EXAMPLE__

Use additional sheets if necessary.   Page __1__ of __3__

Complete **WORKSHEET 4** for other chemicals, years, and wells; otherwise,
Continue to **WORKSHEET 5** if a fixed sample size test is used; or
Continue to **WORKSHEET 7** if a sequential sample test is used.

WORKSHEET 4 *Data Records and Calculations When Assessing Individual Wells; by Chemical, Well, and Year*

See Chapter 8 or 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	· Site ABC
CHEMICAL: NUMBER(c) AND DESCRIPTION [2]	1. Hazardous #1
WELL: NUMBER(w) AND DESCRIPTION [1]	1. d_1 ft. northeast of treatment well
YEAR: NUMBER(K)	1989, K = 2

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Sample Design (Check one): Fixed Sample Size ☑   Sequential Sampling ☑
For purposes of illustration, both methods are used.

Parameter to be tested [2] (Check one) =	Mean ☑ %tile ☐
Number of samples per year [3] = n =	6
Number of samples with nonmissing data in year = n_k =	6
Cleanup standard[2] = Cs =	100
Concentration used for observations below the detection limit =	10

"Season" Number: j within this k^{th} year	Sample ID	Sample Collection date/time	Reported Concentration	Concentration Corrected for Detection Limit A	Is A Greater than Cs? 1 = Yes 0 = No B	Data for analysis x_{jk} = A if Mean √ x_{jk} = B if %tile
1	12	Feb. 15, '89	89	89		89
2	22	April 17, '89	72	72		72
3	32	June 14, '89	105	105		105
4	42	Aug. 18, '89	77	77		77
5	52	Oct. 15, '89	63	63		63
6	62	Dec. 13, '89	92	92		92

Total of x_{jk} for this year = C = [498]

Mean of x_{jk} for this k^{th} year = $\frac{C}{n_k}$ = $\bar{x}_k$ = [83.00]

Date Completed: __EXAMPLE__   Completed by __EXAMPLE__

Use additional sheets if necessary.   Page _2_ of _3_

Complete **WORKSHEET 4** for other chemicals, years, and wells; otherwise,
Continue to **WORKSHEET 5** if a fixed sample size test is used; or
Continue to **WORKSHEET 7** if a sequential sample test is used.

WORKSHEET 4 Data Records and Calculations When Assessing Individual Wells: by Chemical, Well, and Year

See Chapter 8 or 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	Site ABC
CHEMICAL: NUMBER(c) AND DESCRIPTION [2]	1. Hazardous #1
WELL: NUMBER(w) AND DESCRIPTION [1]	1. d_1 ft. northeast of treatment well
YEAR: NUMBER(K)	1990, K = 3

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Sample Design (Check one): Fixed Sample Size ☑   Sequential Sampling ☑
For purposes of illustration, both methods are used.

Parameter to be tested [2] (Check one) =	Mean ☑ %tile ☐
Number of samples per year [3] = n =	6
Number of samples with nonmissing data in year = n_k =	6
Cleanup standard[2] = Cs=	100
Concentration used for observations below the detection limit =	10

"Season" Number: j within this k^{th} year	Sample ID	Sample Collection date/time	Reported Concen- tration	Concentration Corrected for Detection Limit A	Is A Greater than Cs? 1 = Yes 0 = No B	Data for analysis x_{jk} = A if Mean x_{jk} = B if %tile √
1	13	Feb. 16, '90	71	71		71
2	23	April 14, '90	62	62		62
3	33	June 14, '90	88	88		88
4	43	Aug. 17, '90	43	43		43
5	53	Oct. 15, '90	62	62		62
6	63	Dec. 15, '90	73	73		73

Total of x_{jk} for this year = C = | 399

Mean of x_{jk} for this k^{th} year = $\dfrac{C}{n_k}$ = $\bar{x}_k$ = | 66.50

Date Completed: __EXAMPLE__   Completed by __EXAMPLE__

Use additional sheets if necessary.   Page _3_ of _3_

Complete **WORKSHEET 4** for other chemicals, years, and wells; otherwise,
Continue to **WORKSHEET 5** if a fixed sample size test is used; or
Continue to **WORKSHEET 7** if a sequential sample test is used.

WORKSHEET 5 *Data Calculations for a Fixed Sample Size Test When Assessing Individual Wells; by*
Chemical and Well

See Chapter 8 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:	Site ABC
CHEMICAL: NUMBER(c) AND DESCRIPTION [2]	1. Hazardous #1
WELL: NUMBER(w) AND DESCRIPTION [1]	1. d_1 ft. northeast of treatment well

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Year Number k	Mean for the year [4] $\bar{x}_k$	$(\bar{x}_k)^2$
1	90.17	8,130.63
2	83.00	6,889.00
3	66.50	4,422.25

Total from previous page
(if more than one Worksheet
5 used)

Column Totals:

A 239.67	B 19,441.88

$$A = \sum (\bar{x}_k) \qquad B = \sum (\bar{x}_k)^2$$

Date Completed: __EXAMPLE__   Completed by __EXAMPLE__

Use additional sheets if necessary.   Page ____ of ____

Complete **WORKSHEET 5** for other chemicals and wells or continue to **WORKSHEET 6**

WORKSHEET 6 Inference for Fixed Sample Sizes Tests When Assessing Individual Wells, by Chemical and Well

See Chapter 8 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:	Site ABC
CHEMICAL: NUMBER(c) AND DESCRIPTION [2]	1. Hazardous #1
WELL: NUMBER(w) AND DESCRIPTION [1]	1. d_1 ft. northeast of treatment well

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

[2] α = .1

[2] Cs = 100

Number of Years [3] = m = 3

Sum of the yearly means [5] = $\sum \bar{x}_k$ = A = 239.67

Sum of the squared yearly means [5] = $\sum (\bar{x}_k)^2$ = B = 19,441.88

Overall mean concentration = $\dfrac{A}{m}$ = $\bar{x}$ = 79.89

Standard Deviation of the yearly means = $\sqrt{\dfrac{B - m(\bar{x})^2}{m-1}}$ = $s_{\bar{x}}$ = 12.1376

Degrees of Freedom for $s_{\bar{\bar{x}}}$ = m-1 = Df = 2

Critical value from table of the t-distribution
(Table A.1) for specified values of (1-α) and Df = t = 1.886

Standard Error for the overall mean = $\dfrac{s_{\bar{x}}}{\sqrt{m}}$ = $s_{\bar{x}_m}$ = 7.0076

Upper One Sided Confidence Interval = $\bar{x} + t\, s_{\bar{x}_m}$ = $\mu_{U\alpha}$ = 93.1063

If $\mu_{U\alpha}$ < Cs then circle Clean, otherwise circle Contaminated:
Based on the mean concentration, the sampling well is: **Clean** Contaminated √

Date Completed: __EXAMPLE__   Completed by __EXAMPLE__

Page ____ of ____

Complete **WORKSHEET 6** for other chemicals and wells

WORKSHEET 7a Data Calculations for a Sequential Sample When Assessing Wells Individually; by Chemical and Well

See Chapter 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:	**Site ABC**	
CHEMICAL:	NUMBER(c) AND DESCRIPTION [2]	1. Hazardous #1
WELL:	NUMBER(w) AND DESCRIPTION [1]	1. d_1 ft. northeast of treatment well

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Cleanup standard[2] = Cs = ____100____

Alternate mean = μ_1 = ____75____

Probability of mistakenly declaring the well(s) clean [2] = α = ____.1____

Probability of mistakenly declaring the well(s) contaminated [2] = β = ____.2____

Year Number [4] k or m	Yearly Average [4] $\bar{x}_k$	Cumulative Sum of $\bar{x}_k$ ($A_0 = 0$) $A_k = A_{k-1}+\bar{x}_k$	Cumulative Sum of $\bar{x}_k^2$ ($B_0 = 0$) $B_k = B_{k-1}+\bar{x}_k^2$	Mean (average of yearly averages) $\bar{x}_m = \frac{A_k}{k}$	Standard Error of Mean $s_{\bar{x}_m} = \sqrt{\frac{B_k-k(\bar{x}_m)^2}{(k-1)k}}$
1	90.17	90.17	8,130.63	90.1700	–
2	83.00	173.17	15,019.63	86.5950	3.4622
3	66.50	239.67	19,441.88	79.8900	7.0077
		Carry as many significant figures as possible			

Date Completed: __EXAMPLE__   Completed by __EXAMPLE__

Use additional sheets if necessary.   Page ____ of ____

Complete **WORKSHEETS 7a and 7b** for other chemicals and wells

WORKSHEET 7b Data Calculations for a Sequential Sample When Assessing Wells Individually; by Chemical and Well

See Chapter 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:	Site ABC
CHEMICAL: NUMBER(c) AND DESCRIPTION [2]	1. Hazardous #1
WELL: NUMBER(w) AND DESCRIPTION [1]	1. d_1 ft. northeast of treatment well

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Year Number [4] m	$\delta = \dfrac{\mu_1 - Cs}{s\bar{x}_m}$	$t = \dfrac{\bar{x}_m - \dfrac{Cs + \mu_1}{2}}{s\bar{x}_m}$	Likelihood ratio LR*	Critical value: clean $A = \dfrac{\beta}{1-\alpha}$	Critical value: contaminated $B = \dfrac{1-\beta}{\alpha}$	Decision: clean LR > B, contaminated LR ≤ A, or no decision A < LR ≤ B
1						no test until year 3
2						no test until year 3
3	-3.5675	-1.086	3.506	0.222	8.00	no decision

$$* LR = \exp\left(\delta \frac{m-2}{m} t \sqrt{\frac{m}{m-1+t^2}}\right)$$

If "no decision", collect another years' allotment of samples and test the hypothesis again.

Date Completed: __EXAMPLE__   Completed by __EXAMPLE__

Use additional sheets if necessary.   Page ____ of ____

Complete **WORKSHEETS 7a and 7b** for other chemicals and wells

WORKSHEET 8 Attainment Objectives When Assessing Wells as a Group
See Chapter 3 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:	Site ABC

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Sample Design (Check one): Fixed Sample Size ☐   Sequential Sampling ☒

Probability of mistakenly declaring the well(s) **clean** = α = [.1]

Probability of mistakenly declaring the well(s) **contaminated** = β = [.2]

Chemical to be tested number c	Chemical name	Cleanup standard (with units) Cs	Parameter to test: Check one	If mean, enter the alternate hypothesis μ_1	If mean, enter the alternate hypothesis Max$_1$
1	Hazardous #1	100	Mean ☒ %tile ☐	75	
2	Hazardous #2	60	Mean ☒ %tile ☐	30	
			Mean ☐ Max ☐		
			Mean ☐ Max ☐		

Sample Collection Procedures to be used (attach separate sheet if necessary):

Not specified for this example

Secondary Objectives/ Other purposes for which the data is to be collected:

Use the Chemical Number (c) to refer on other sheets to the chemical described above.
Attach documentation d scribing the lab analysis procedure for each chemical.

Date Completed: __EXAMPLE__   Completed by ___EXAMPLE___

Use additional sheets if necessary.   Page ____ of ____

Continue to **WORKSHEET 9** if a fixed sample size test is used; or
Continue to **WORKSHEET 10** if a sequential sample test is used.

WORKSHEET 9 Sample Size When Using a Fixed Sample Size Test for Assessing Wells as a Group

See Sections 8.2 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:	Site ABC

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

From Table A.2, Appendix A

Probability of mistakenly declaring the site clean [8] = α = | .1 | $z_{1-\alpha}$ = | 1.282 |

Probability of mistakenly declaring the site contaminated [8] = β = | .2 | $z_{1-\beta}$ = | .842 |

Number of samples per year = n = | 6 | (based on calculations described in Section 8.2)

Variance factor from Table A.5, Appendix A = F^1 = | 5.55 |

For testing the mean concentration

Chemical Number [8] c	Cleanup Standard[8] Cs	[8] μ_1	Standard Deviation of mean $\hat{\sigma}$	Calculate: $B=\left(\dfrac{Cs-\mu_1}{z_{1-\alpha}+z_{1-\beta}}\right)^2$	$m_d=\dfrac{\hat{\sigma}^2}{F*B}+2$
1	100	75	23	138.53	2.69
2	60	30	6	199.50	2.03

For testing the maximum concentration across all wells

Chemical Number [8] c	Cleanup Standard[8] Cs	[8] Max_1	Standard Deviation of yearly mean $\hat{\sigma}$	Calculate: $B=\left(\dfrac{Cs-Max_1}{z_{1-\alpha}+z_{1-\beta}}\right)^2$	$m_d=\dfrac{\hat{\sigma}^2}{F*B}+2$

Column Maximum, (Maximum of m_d values) = C = | 2.69 |

Round C to next largest integer=Number of years of sample collection= m= | 3 |

Total number of samples = nm = N = | 18 |

Date Completed: __EXAMPLE__   Completed by _____EXAMPLE_____

Use additional sheets if necessary.   Page ___ of ___

Continue to **WORKSHEET 10**

---

[1] An estimate of ϕ, the serial correlation, is necessary to determine the appropriate value of F. Worksheets 15 and 16 can be used to estimate ϕ. ϕ was assumed to be .20 for this example.

WORKSHEET 10 Data Records for an Individual Well and Calculations When Assessing Wells as a Group; by Chemical, Well and Year

See Chapter 8 or 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	Site ABC
CHEMICAL: NUMBER(c) AND DESCRIPTION [8]	1. Hazardous #1
WELL: NUMBER(w) AND DESCRIPTION [1]	3. d₃ ft. north of treatment well
YEAR: NUMBER(k)	1988, k = 1

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Parameter to be tested (Check one) = | Mean ☑ Max ☐

Number of samples per year = n = | 6

Concentration used for observations below the detection limit = | 10

"Season" Number j	Sample ID	Sample Collection time	Reported Concen-tration	Concentration Corrected for Detection Limit x_{jk}
1	31	Feb. 18, '88	88.71	88.71
2	32	Apr. 12, '88	89.38	89.38
3	33	June 16. '88	74.92	74.92
4	34	Aug. 15, '88	80.03	80.03
5	35	Oct. 12, '88	89.98	89.98
6	36	Dec. 11, '88	91.34	91.34

Date Completed: __EXAMPLE__   Completed by __EXAMPLE__

Use additional sheets if necessary.   Page _1_ of _9_

Complete **WORKSHEET 10** for other chemicals, years, and wells or continue to **WORKSHEET 11**

WORKSHEET *10* Data Records for an Individual Well and Calculations When Assessing Wells as a Group; by Chemical, Well and Year

See Chapter 8 or 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:		Site ABC
CHEMICAL:	NUMBER(c) AND DESCRIPTION [8]	1. Hazardous #1
WELL:	NUMBER(w) AND DESCRIPTION [1]	4. d4 ft. southwest of treatment well
YEAR:	NUMBER(K)	1988, k = 1

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Parameter to be tested (Check one) =	Mean ☑ Max ☐
Number of samples per year = n =	6
Concentration used for observations below the detection limit =	10

"Season" Number j	Sample ID	Sample Collection time	Reported Concentration	Concentration Corrected for Detection Limit x_{jk}
1	41	Feb. 18, '88	76.50	76.50
2	42	Apr. 12, '88	71.28	71.28
3	43	June 16, '88	93.77	93.77
4	44	Aug. 15, '88	73.60	73.60
5	45	Oct. 12, '88	120.94	120.94
6	46	Dec. 11, '88	82.56	82.56

Date Completed: __EXAMPLE__   Completed by ___EXAMPLE___

Use additional sheets if necessary.   Page _2_ of _9_

Complete **WORKSHEET 10** for other chemicals, years, and wells or continue to **WORKSHEET 11**

WORKSHEET 10 *Data Records for an Individual Well and Calculations When Assessing Wells as a Group; by Chemical, Well and Year*

See Chapter 8 or 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	Site ABC
CHEMICAL: NUMBER(c) AND DESCRIPTION [8]	1. Hazardous #1
WELL: NUMBER(w) AND DESCRIPTION [1]	5. d₅ ft. southeast of treatment well
YEAR: NUMBER(k)	1988, k = 1

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Parameter to be tested (Check one) = Mean ☑ / Max ☐

Number of samples per year = n = 6

Concentration used for observations below the detection limit = 10

"Season" Number j	Sample ID	Sample Collection time	Reported Concentration	Concentration Corrected for Detection Limit x_{jk}
1	51	Feb. 18, '88	62.68	62.68
2	52	Apr. 12, '88	92.49	92.49
3	53	June 16, '88	80.94	80.94
4	54	Aug. 15, '88	103.38	103.38
5	55	Oct. 12, '88	95.39	95.39
6	56	Dec. 11, '88	99.04	99.04

Date Completed: EXAMPLE   Completed by EXAMPLE

Use additional sheets if necessary.   Page 3 of 9

Complete **WORKSHEET 10** for other chemicals, years, and wells or continue to **WORKSHEET 11**

WORKSHEET 10 *Data Records for an Individual Well and Calculations When Assessing Wells as a Group; by Chemical, Well and Year*

See Chapter 8 or 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	Site ABC
CHEMICAL: NUMBER(c) AND DESCRIPTION [8]	1. Hazardous #1
WELL: NUMBER(w) AND DESCRIPTION [1]	3. d_3 ft. north of treatment well
YEAR: NUMBER(k)	1989, k = 2

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Parameter to be tested (Check one) = Mean ☑ Max ☐

Number of samples per year = n = 6

Concentration used for observations below the detection limit = 10

"Season" Number j	Sample ID	Sample Collection time	Reported Concentration	Concentration Corrected for Detection Limit x_{jk}
1	31	Feb. 15, '89	87.11	87.11
2	32	Apr. 17, '89	78.38	78.38
3	33	June 14, '89	80.61	80.61
4	34	Aug. 18, '89	73.51	73.51
5	35	Oct. 15, '89	89.16	89.16
6	36	Dec. 13, '89	100.26	100.26

Date Completed: EXAMPLE   Completed by EXAMPLE

Use additional sheets if necessary.   Page 4 of 9

Complete **WORKSHEET 10** for other chemicals, years, and wells or continue to **WORKSHEET 11**

WORKSHEET 10 Data Records for an Individual Well and Calculations When Assessing Wells as a Group; by Chemical, Well and Year

See Chapter 8 or 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	Site ABC
CHEMICAL: NUMBER(c) AND DESCRIPTION [8]	1. Hazardous #1
WELL: NUMBER(w) AND DESCRIPTION [1]	4. d_4 ft. southwest of treatment well
YEAR: NUMBER(k)	1989, k = 2

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Parameter to be tested (Check one) = Mean ☑ / Max ☐

Number of samples per year = n = 6

Concentration used for observations below the detection limit = 10

"Season" Number J	Sample ID	Sample Collection time	Reported Concentration	Concentration Corrected for Detection Limit x_{jk}
1	41	Feb. 15, '89	82.34	82.34
2	42	Apr. 17, '89	85.69	85.69
3	43	June 14, '89	96.72	96.72
4	44	Aug. 18, '89	108.61	108.61
5	45	Oct. 15, '89	95.75	95.75
6	46	Dec. 13, '89	66.77	66.77

Date Completed: EXAMPLE   Completed by EXAMPLE

Use additional sheets if necessary.   Page 5 of 9

Complete **WORKSHEET 10** for other chemicals, years, and wells or continue to **WORKSHEET 11**

WORKSHEET 10 Data Records for an Individual Well and Calculations When Assessing Wells as a Group; by Chemical, Well and Year

See Chapter 8 or 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	Site ABC	
CHEMICAL:	NUMBER(c) AND DESCRIPTION [8]	1. Hazardous #1
WELL:	NUMBER(w) AND DESCRIPTION [1]	5. d5 ft. southeast of treatment well
YEAR:	NUMBER(k)	1989, k = 2

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Parameter to be tested (Check one) = Mean☑ Max☐

Number of samples per year = n = 6

Concentration used for observations below the detection limit = 10

"Season" Number j	Sample ID	Sample Collection time	Reported Concentration	Concentration Corrected for Detection Limit x_{jk}
1	51	Feb. 15, '89	80.05	80.05
2	52	Apr. 17, '89	81.44	81.44
3	53	June 14, '89	92.89	92.89
4	54	Aug. 18, '89	93.87	93.87
5	55	Oct. 15, '89	95.82	95.82
6	56	Dec. 13, '89	78.39	78.39

Date Completed: EXAMPLE   Completed by EXAMPLE

Use additional sheets if necessary.   Page 6 of 9

Complete **WORKSHEET 10** for other chemicals, years, and wells or continue to **WORKSHEET 11**

B-21

WORKSHEET 10 Data Records for an Individual Well and Calculations When Assessing Wells as a Group; by Chemical, Well and Year

See Chapter 8 or 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	Site ABC
CHEMICAL: NUMBER(c) AND DESCRIPTION [8]	1. Hazardous #1
WELL: NUMBER(w) AND DESCRIPTION [1]	3. d_3 ft. north of treatment well
YEAR: NUMBER(K)	1990, k = 3

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Parameter to be tested (Check one) =	Mean ☑ Max ☐
Number of samples per year = n =	6
Concentration used for observations below the detection limit =	10

"Season" Number j	Sample ID	Sample Collection time	Reported Concen- tration	Concentration Corrected for Detection Limit x_{jk}
1	31	Feb. 16, '90	76.86	76.86
2	32	Apr. 14, '90	76.38	76.38
3	33	June 14, '90	87.46	87.46
4	34	Aug. 17, '90	80.84	80.84
5	35	Oct. 15, '90	71.65	71.65
6	36	Dec. 15, '90	57.28	57.28

Date Completed: __EXAMPLE__   Completed by ____EXAMPLE____

Use additional sheets if necessary.   Page _7_ of _9_

Complete **WORKSHEET 10** for other chemicals, years, and wells or continue to **WORKSHEET 11**

WORKSHEET *10* Data Records for an Individual Well and Calculations When Assessing Wells as a Group; by Chemical, Well and Year

See Chapter 8 or 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:		Site ABC
CHEMICAL:	NUMBER(c) AND DESCRIPTION [8]	1. Hazardous #1
WELL:	NUMBER(w) AND DESCRIPTION [1]	4. d_4 ft. southwest of treatment well
YEAR:	NUMBER(k)	1990, k = 3

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Parameter to be tested (Check one) = | Mean ☑ / Max ☐ |

Number of samples per year = n = | 6 |

Concentration used for observations below the detection limit = | 10 |

"Season" Number j	Sample ID	Sample Collection time	Reported Concentration	Concentration Corrected for Detection Limit x_{jk}
1	41	Feb. 16, '90	87.85	87.85
2	42	Apr. 14, '90	87.08	87.08
3	43	June 14, '90	97.84	97.84
4	44	Aug. 17, '90	105.95	105.95
5	45	Oct. 15, '90	81.58	81.58
6	46	Dec. 15, '90	87.76	87.76

Date Completed: __EXAMPLE__   Completed by ____EXAMPLE____

Use additional sheets if necessary.   Page _8_ of _9_

Complete **WORKSHEET 10** for other chemicals, years, and wells or continue to **WORKSHEET 11**

APPENDIX B: EXAMPLE WORKSHEETS

WORKSHEET 10 Data Records for an Individual Well and Calculations When Assessing Wells as a Group; by Chemical, Well and Year

See Chapter 8 or 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	Site ABC
CHEMICAL: NUMBER(c) AND DESCRIPTION [8]	1. Hazardous #1
WELL: NUMBER(w) AND DESCRIPTION [1]	5. d5 ft. southeast of treatment well
YEAR: NUMBER(K)	1990, k = 3

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Parameter to be tested (Check one) = Mean ☑ / Max ☐

Number of samples per year = n = 6

Concentration used for observations below the detection limit = 10

"Season" Number j	Sample ID	Sample Collection time	Reported Concentration	Concentration Corrected for Detection Limit x_{jk}
1	51	Feb. 16, '90	79.70	79.70
2	52	Apr. 14, '90	59.32	59.32
3	53	June 14, '90	66.64	66.64
4	54	Aug. 17, '90	52.48	52.48
5	55	Oct. 15, '90	91.63	91.63
6	56	Dec. 15, '90	35.08	35.08

Date Completed: EXAMPLE   Completed by EXAMPLE

Use additional sheets if necessary.   Page 9 of 9

Complete **WORKSHEET 10** for other chemicals, years, and wells or continue to **WORKSHEET 11**

B-24

APPENDIX B: EXAMPLE WORKSHEETS

WORKSHEET *11* *Data Records and Calculations When Assessing Wells as a Group; by Chemical and Year*

See Chapter 8 or 9 in "Statistical Methods for Evaluating the Attainment of Superfund Cleanup Standards", Vol. 2

SITE:	Site ABC
CHEMICAL: NUMBER(c) AND DESCRIPTION [8]	1. Hazardous #1
YEAR: NUMBER(K)	1988, k = 1

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Sample Design (Check one): Fixed Sample Size ☐   Sequential Sampling ☑

Parameter to be tested (Check one) = Mean ☑ Max ☐

Number of samples per year [9] = n = 6

"Season" Number[10] j	Well #3 [10] x_{jk}	Well #4 [10] x_{jk}	Well #5 [10] x_{jk}	Well #_ [10] x_{jk}	Well #_ [10] x_{jk}	Measure for analysis (row maximum or row mean) √ x_j
1	88.71	76.50	62.68			75.96
2	89.38	71.28	92.49			84.38
3	74.92	93.77	80.94			83.21
4	80.03	73.60	103.38			85.67
5	89.98	120.94	95.39			102.10
6	91.34	82.56	99.04			90.98

Total of x_j for this year = A = 522.30

Mean of x_{jk} for this year = $\bar{x}_k = \dfrac{A}{n}$ = 87.05

Date Completed: ___EXAMPLE___   Completed by ___EXAMPLE___

Use additional sheets if necessary.   Page __1_ of _3__

Complete **WORKSHEET 11** for other chemicals; otherwise,
Continue to **WORKSHEET 12** if a fixed sample size test is used; or
Continue to **WORKSHEET 14** if a sequential sample test is used.

WORKSHEET *11* *Data Records and Calculations When Assessing Wells as a Group; by Chemical and Year*

See Chapter 8 or 9 in "Statistical Methods for Evaluating the Attainment of Superfund Cleanup Standards", Vol. 2

SITE:	Site ABC
CHEMICAL: NUMBER(c) AND DESCRIPTION [8]	1. Hazardous #1
YEAR: NUMBER(K)	1989, k = 2

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Sample Design (Check one): Fixed Sample Size ☐   Sequential Sampling ☑

Parameter to be tested (Check one) = Mean ☑ Max ☐

Number of samples per year [9] = n = 6

"Season" Number[10] j	Well #3 [10] x_{jk}	Well #4 [10] x_{jk}	Well #5 [10] x_{jk}	Well #_ [10] x_{jk}	Well #_ [10] x_{jk}	Measure for analysis (row maximum or row mean) √ x_j
1	87.11	82.34	80.05			83.17
2	78.38	85.69	81.44			81.84
3	80.61	96.72	92.89			90.07
4	73.51	108.61	93.87			92.00
5	89.16	95.75	95.82			93.58
6	100.26	66.77	78.39			81.81

Total of x_j for this year = A = $\boxed{522.47}$

Mean of x_{jk} for this year = $\bar{x}_k = \dfrac{A}{n} = \boxed{87.08}$

Date Completed: E X A M P L E _

Use additional sheets if necessary.

Completed by ___EXAMPLE___

Page _2_ of _3_

Complete **WORKSHEET 11** for other chemicals; otherwise,
Continue to **WORKSHEET 12** if a fixed sample size test is used; or
Continue to **WORKSHEET 14** if a sequential sample test is used.

WORKSHEET 11 *Data Records and Calculations When Assessing Wells as a Group; by Chemical and Year*

See Chapter 8 or 9 in "Statistical Methods for Evaluating the Attainment of Superfund Cleanup Standards", Vol. 2

SITE:	Site ABC
CHEMICAL: NUMBER(c) AND DESCRIPTION [8]	1. Hazardous #1
YEAR: NUMBER(k)	1990, k = 3

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Sample Design (Check one): Fixed Sample Size ☐   Sequential Sampling ☑

Parameter to be tested (Check one) =

Mean ☑
Max ☐

Number of samples per year [9] = n = 6

"Season" Number[10] j	Well #3 [10] x_{jk}	Well #4 [10] x_{jk}	Well #5 [10] x_{jk}	Well #_ [10] x_{jk}	Well #_ [10] x_{jk}	Measure for analysis (row maximum or row mean) √ x_j
1	76.86	87.85	79.70			81.47
2	76.38	87.08	59.32			74.26
3	87.46	97.84	66.64			83.98
4	80.84	105.95	52.48			79.76
5	71.65	81.58	91.63			81.62
6	57.28	87.76	35.08			60.04

Total of x_j for this year = A = 461.13

Mean of x_{jk} for this year = $\bar{x}_k = \dfrac{A}{n}$ = 76.86

Date Completed: ___EXAMPLE___   Completed by ___EXAMPLE___

Use additional sheets if necessary.   Page __3_ of _3__

Complete **WORKSHEET 11** for other chemicals; otherwise,
Continue to **WORKSHEET 12** if a fixed sample size test is used; or
Continue to **WORKSHEET 14** if a sequential sample test is used.

WORKSHEET *14a* Data Calculations for a Sequential Sample When Assessing Wells as a Group; by Chemical

See Chapter 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:	Site ABC
CHEMICAL:	NUMBER(c) AND DESCRIPTION [8] 1. Hazardous #1

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Cleanup standard[8] = C_s = | 100

Alternate mean = μ_1 = | 75

Probability of mistakenly declaring the well(s) **clean** [8] = α = | .1

Probability of mistakenly declaring the well(s) **contaminated** [8] = β = | .2

Year Number [11] k	Yearly Average [11] $\bar{x}_k$	Cumulative Sum of $\bar{x}_k$ ($A_0 = 0$) $A_k = A_{k-1} + \bar{x}_k$	Cumulative Sum of $\bar{x}_k^2$ ($B_0 = 0$) $B_k = B_{k-1} + \bar{x}_k^2$	Mean (average of yearly averages) $\bar{x}_m = \dfrac{A_k}{k}$	Standard Deviation of Mean $s_{\bar{x}_m} = \sqrt{\dfrac{B_k - k(\bar{x}_m)^2}{(k-1)k}}$
1	87.05	87.05	7,577.70	87.0500	–
2	87.08	174.13	15,160.63	87.0650	–
3	76.86	250.99	21,068.09	83.6633	3.402
			Carry as many significant figures as possible		

Date Completed: __EXAMPLE__   Completed by ____EXAMPLE____

Use additional sheets if necessary.   Page ___ of ___

Complete **WORKSHEET 14a** and **14b** for other chemicals and groups of wells

WORKSHEET 14b *Data Calculations for a Sequential Sample When Assessing Wells as a Group; by Chemical*

See Chapter 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:	Site ABC
CHEMICAL: NUMBER(c) AND DESCRIPTION [8]	1. Hazardous #1

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Year Number [4] m	$\delta_1 = \dfrac{\mu_1 - Cs}{s_{\bar{x}_m}}$	$t = \dfrac{\bar{x}_m - \dfrac{Cs + \mu_1}{2}}{s_{\bar{x}_m}}$	Likelihood ratio LR*	Critical value: clean $A = \dfrac{\beta}{1-\alpha}$	Critical value: contaminated $B = \dfrac{1-\beta}{\alpha}$	Decision: clean LR > B, contaminated LR ≤ A, or no decision A < LR ≤ B
1						no test until year 3
2						no test until year 3
3	-7.349	-1.128	14.094	0.222	8.00	CLEAN

$$* \; LR = \exp\left(\delta \frac{m-2}{m} \cdot t \sqrt{\frac{m}{m-1+t^2}} \right)$$

If "no decision", collect another years' allotment of samples and test the hypothesis again.

Date Completed: __EXAMPLE__   Completed by ____EXAMPLE____

Use additional sheets if necessary.   Page ____ of ____

Complete **WORKSHEET 14a** and **14b** for other chemicals and groups of wells

WORKSHEET 15 *Removing Seasonal Patterns in the Data (Use as First Step in Computing Serial Correlations)*

See Sections 8.4 and 9.4 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	Site DEF *(data independent of five-well example)*
CHEMICAL: NUMBER(c) AND DESCRIPTION [2 OR 8]	1. Chemical #1
WELL: NUMBER(w) AND DESCRIPTION [1]	1. d_1 ft. south of treatment well

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

"Season" Number j	Measurements for each "season" for year k Yr=_1_ x_{jk}	Yr=_2_ x_{jk}	Yr=___ x_{jk}	Yr=___ x_{jk}	Yr=___ x_{jk}	Number of years with Data m_j	Row Total $T_j = \sum_k x_{jk}$	Row Mean $\bar{x}_j = \dfrac{T_j}{m_j}$
1	120	133				2	253	126.5
2	163	117				2	280	140
3	128	113				2	241	120.5
4	150	126				2	276	138
5	125	114				2	239	119.5
6	110	145				2	255	127.5

Corrected measurements with seasonal patterns removed

"Season" Corrected Measurements for each "season" for year k

Number j	Yr=_1_ $x_{jk} - \bar{x}_j$	Yr=_2_ $x_{jk} - \bar{x}_j$	Yr=___ $x_{jk} - \bar{x}_j$	Yr=___ $x_{jk} - \bar{x}_j$	Yr=___ $x_{jk} - \bar{x}_j$
1	-6.5	6.5			
2	23	-23			
3	7.5	-7.5			
4	12	-12			
5	5.5	-5.5			
6	-17.5	17.5			

Date Completed: __EXAMPLE__   Completed by __EXAMPLE__

Use additional sheets if necessary.   Page ____ of ____

Complete **WORKSHEET 15** for other chemicals
Continue to **WORKSHEET 16** if serial correlations are being computed.

WORKSHEET 16 Calculating Serial Correlations

See Sections 8.4 and 9.4 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	Site DEF *(data independent of five-well example)*
CHEMICAL: NUMBER(c) AND DESCRIPTION [2 OR 8]	1. Chemical #1
WELL: NUMBER(w) AND DESCRIPTION [1]	1. d_1 ft. south of treatment well

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Year = k = $\boxed{1}$

Period between well samples in months = t = $\boxed{2}$

Data Numbers jk (season within year k)	Residual $e_{jk} = x_{jk} - \bar{x}_j$ [15]	Product $d_{jk} = e_{jk} * e_{j-1\,k}$	e_{jk}^2
11	-6.5		42.25
21	23.0	-149.5	529.00
31	7.5	172.5	56.25
41	12.0	90.00	144.00
51	5.5	66.00	30.25
61	-17.5	-96.25	306.25

Totals from previous page = $\boxed{}$ $\boxed{}$
(if more than one
Worksheet 16 is used)

Column Totals = $\boxed{A\ 82.75}$ $\boxed{B\ 1,108}$

Estimated Serial Correlation based on the data $= \dfrac{A}{B} = \hat{\phi}_{obs} = \boxed{\quad - \quad}$

Serial Correlation between monthly observations $= \hat{\phi} = \left(\hat{\phi}_{obs}\right)^{\frac{1}{t}} = \boxed{\quad - \quad}$

Date Completed: __EXAMPLE__   Completed by __EXAMPLE__

Use additional sheets if necessary.   Page _1_ of _2_

Complete **WORKSHEET 16** for other chemicals

WORKSHEET 16 *Calculating Serial Correlations*
See Sections 8.4 and 9.4 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	Site DEF *(data independent of five-well example)*
CHEMICAL: NUMBER(C) AND DESCRIPTION [2 OR 8]	1. Chemical #1
WELL: NUMBER(W) AND DESCRIPTION [1]	1. d_1 ft. south of treatment well

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Year $= k =$ [2]

Period between well samples in months $= t =$ [2]

Data Numbers jk (season within year k)	Residual $e_{jk} = x_{jk} - \bar{x}_j$ [15]	Product $d_{jk} = e_{jk} * e_{j-1\,k}$	e_{jk}^2
12	6.5		42.25
22	-23.0	-149.50	529.00
32	-7.5	172.50	56.25
42	-12.0	90.00	144.00
52	-5.5	66.00	30.25
62	17.5	-96.25	306.25

Totals from previous page = (if more than one Worksheet 16 is used)   [82.75]   [1,108]

Column Totals = | A 165.5 | | B 2,216 |

Estimated Serial Correlation based on the data $= \dfrac{A}{B} = \hat{\phi}_{obs} =$ [.0747]

Serial Correlation between monthly observations $= \hat{\phi} = \left(\hat{\phi}_{obs}\right)^{\frac{1}{t}} =$ [.2733]

Date Completed: __EXAMPLE__   Completed by __EXAMPLE__

Use additional sheets if necessary.   Page _2_ of _2_

Complete **WORKSHEET 16** for other chemicals

WORKSHEET $1R$ *Basic Calculations for a Simple Linear Regression*

See Section 6.1 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	Site ABC	
CHEMICAL: NUMBER(c) AND DESCRIPTION [2 OR 8]		1. Hazardous #1
WELL: NUMBER(w) AND DESCRIPTION [1]		1. d_1 ft. northeast of treatment well

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Concentration used when no concentration is reported = **10**

Number of collectable samples = N = **3**

Sample Number n	Concentration Corrected for Detection Limit y_n	y_n^2	Transformed Time Variable x_n	x_n^2	$x_n y_n$
1	90.17	8,130.63	1	1	90.17
2	83.00	6,889.00	2	4	166.00
3	66.50	4,422.25	3	9	199.50

Totals from previous page(s):

Column Totals:

	A 239.67	B 19,441.88	C 6	D 14	E 455.67

$$A = \Sigma y_n \qquad B = \Sigma y_n^2 \qquad C = \Sigma x_n \qquad D = \Sigma x_n^2 \qquad E = \Sigma y_n x_n$$

Corrected Sum of Squares and Cross Products:

79.89	2	294.64		2	-23.67

$$\bar{y} = \frac{A}{N} \qquad \bar{x} = \frac{C}{N} \qquad S_{yy} = B - \frac{A^2}{N} \qquad\qquad S_{xx} = D - \frac{C^2}{N} \qquad S_{yx} = E - \frac{AC}{N}$$

Date Completed: __EXAMPLE__   Completed by __EXAMPLE__

Use additional sheets if necessary.   Page ____ of ____

Complete **WORKSHEET 1R** for other chemicals or continue to **WORKSHEET 2R**.

WORKSHEET 2R *Inference in a Simple Linear Regression*

See Section 6.1 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	Site ABC
CHEMICAL: NUMBER(c) AND DESCRIPTION [2 OR 8]	1. Hazardous #1
WELL: NUMBER(w) AND DESCRIPTION [1]	1. d_1 ft. northeast of treatment well

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Estimating Regression Coefficients

S_{yy} [1R] = $\boxed{294.64}$ Number of collectable samples [1R] = N = $\boxed{3}$

S_{xx} [1R] = $\boxed{2}$ Mean of y_t [1R] = $\bar{y}$ = $\boxed{79.89}$

S_{yx} [1R] = $\boxed{-23.67}$ Mean of x_t [1R] = $\bar{x}$ = $\boxed{2}$

Type 1 error probability α = $\boxed{.1}$ Estimated slope [1R], b_1 = $\dfrac{S_{yx}}{S_{xx}}$ = $\boxed{-11.84}$

Estimated Intercept [1R], b_0 = $\bar{y} - (b_1 * \bar{x})$ = $\boxed{103.57}$

Sum of squares due to error [1R], SSE = $S_{yy} - \left(\dfrac{S_{yx}^2}{S_{xx}}\right)$ = $\boxed{14.51}$

Degrees of freedom, Df = N-2 = $\boxed{1}$

Critical value from table of t-distribution (Appendix A.1)
for specified values of $(1 - \frac{\alpha}{2})$ and Df = t = $\boxed{6.314}$

Mean Square Error, MSE = $\dfrac{SSE}{N-2}$ = $\boxed{14.51}$

Standard Error of the Slope, $s(b_1)$ = $\sqrt{\dfrac{MSE}{S_{xx}}}$ = $\boxed{2.69}$

Upper Two Sided Confidence Interval
for Slope: $b_1 + t * s(b_1)$ = $\boxed{5.14}$

Lower Two Sided Confidence
Interval for Slope: $b_1 - t * s(b_1)$ = $\boxed{-28.82}$

Calculating Prediction Limits

Value of x_t at which concentration is to be predicted = $\boxed{2.5}$

Predicted value, $\hat{y}$ = $b_0 + b_1 x_t$ = $\boxed{73.97}$

Standard Error of Predicted Value = $S\hat{y}$ = $\sqrt{MSE\{1 + \dfrac{1}{N} + \dfrac{(x_t - \bar{x})^2}{S_{xx}}\}}$ = $\boxed{4.6000}$

Upper Two Sided Confidence Interval for Prediction = $\hat{y} + t * S\hat{y}$ = $\boxed{\text{Upper} \quad 103.01}$

Lower Two Sided Confidence Interval for Prediction = $\hat{y} - t * S\hat{y}$ = $\boxed{\text{Lower} \quad 44.93}$

Date Completed: __EXAMPLE__   Completed by ____EXAMPLE____

Use additional sheets if necessary.   Page ____ of ____

Complete **WORKSHEET 2R** for other chemicals

APPENDIX C: BLANK WORKSHEETS

The worksheets in this appendix can be photocopied when needed. Then the copies may be us&i in their current form or modified, as appropriate. They may be employed to document the objectives and decisions, record data, and make calculations to determine if the ground water at the site attains the cleanup standard. These worksheets refer to in the main text of this document. Appendix B provides examples of how to fill out the worksheets.

The initial appearance of a "Bold" letter in a worksheet represents an intermediate computation, the result of which will be used in a later computation and will also be signified by the letter in "Bold" script.

To maintain adequate precision in doing the computations appearing in the worksheets, (particularly in the calculation of estimated variances, standard deviations, or standard errors), the number of decimal places retained should be as high as possible, with a minimum of four.

Table C.1 Summary of Notation Used in Appendix C

Symbol	Definition
m	The number of years for which data were collected (usually the analysis will be performed with full years worth of data)
n	The number of sample measurements per year (for monthly data, n = 12; for quarterly data, n = 4). This is also referred to as the number of "seasons" per year
N	The total number of sample measurements (if there are no missing observations, N = mn)
index i	Indicates the order in which the ground-water samples are collected
index k	Indicates the year in which the ground-water samples are collected
index j	Indicates the season or time within the year at which the groundwater samples are collected
index c	Indicates the chemical analyzed
index w	Indicates the well sampled
x_i	Contaminant measurement for the ith ground-water sample
x_{jk}	An alternative way of denoting a contaminant measurement, where k = 1, 2, ..., m denotes the year; and j = 1, 2, ..., n denotes the sampling period (season) within the year. The subscript for x_{jk} is related to the subscript for x_i in the following manner: i = (k-1)n + j.
$\bar{x}_k$	The mean (or average) of the contaminant measurements for year k (see Boxes 8.5 and 9.4)
$\bar{x}_m$	The mean of the yearly averages for years k = 1 to m.
$s_{\bar{x}}$	The standard deviation of the yearly average contaminant concentrations from m years of sample collection (see Boxes 8.7 and 9.6)
$s_{\bar{x}_m}$	The standard error of the mean of the yearly means (see Boxes 8.9 and 9.8)
Cs	The designated clean up standard
Df	The degrees of freedom associated with the standard error of an estimate (see Boxes 8.7 and 9.6)

WORKSHEET 1 Sampling Wells

See Section 32 in "Statistical Methods for Evaluating the Attainment of cleanup Standards". Volume 2

SITE:

Sample -
Well
Number   Describe each sampling well to be used to assess attainment
w

Decision Criteria: Wells assessed (Checked one)   Individually ☐ As a Group ☐

Use the Sampling Well Number (w) to refer on subsequent sheets to the sampling wells described above.

Attach a map showing the sampling wells within the waste site.

Date Completed: _____   Completed by _____

Use additional sheets if necessary.   Page ____ of ____

Continue to **WORKSHEET 2** if wells are assessed individually.
Continue to **WORKSHEET 8** if wells are assessed as a group.

APPENDIX C: BLANK WORKSHEETS

WORKSHEET 2 *Attainment Objectives for Assessing Individual Wells*
See Chapter 3 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Sample Design (Check one): Fixed Sample Size ☐ Sequential Sampling ☐

Probability of mistakenly declaring the well(s) **clean** = α = [   ]

Probability of mistakenly declaring the well(s) **contaminated** = β = [   ]

Chemical Number c	Chemical Name	Cleanup Standard (with units) Cs	Parameter to test: Check one	If Mean, Enter: alternate hypothesis mean μ_1	If %tile, Enter: Critical proportion for alternate/null hypothesis null P_0	alternate P_1
			Mean ☐ %tile ☐			
			Mean ☐ %tile ☐			
			Mean ☐ %tile ☐			
			Mean ☐ %tile ☐			

Sample Collection Procedures to be used (attach separate sheet if necessary):

Secondary Objectives/ Other purposes for which the data is to be collected:

Use the Chemical Number (c) to refer on other sheets to the chemical described above.
Attach documentation describing the lab analysis procedure for each chemical.

Date Completed: _____   Completed by _____

Use additional sheets if necessary.   Page ____ of ____

Continue to **WORKSHEET 3** if a fixed sample size test is used; or
Continue to **WORKSHEET 4** if a sequential sample test is used.

WORKSHEET 3 Sample Size When Using a Fixed Sample Size Test for Assessing Individual Wells

See Sections 8.2 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

From Table A.2, Appendix A

Probability of mistakenly declaring the site clean [2] = α = [   ] $z_{1-\alpha}$ = [   ]

Probability of mistakenly declaring the site contaminated [2] = β = [   ] $z_{1-\beta}$ = [   ]

Number of samples per year = n = [   ] (based on calculations described in Section 8.2)

Variance factor from Table A.5, Appendix A = F^1 = [   ]

For testing the mean concentration

Chemical Number [2] c	Cleanup Standard[2] Cs	[2] μ_1	Standard Deviation of yearly mean $\hat{\sigma}$	Calculate: $B = \left(\dfrac{Cs-\mu_1}{z_{1-\alpha}+z_{1-\beta}}\right)^2$	$m_d = \dfrac{\hat{\sigma}^2}{F*B} + 2$

For testing the proportion of contaminated wells or samples

Chemical Number [2] c	Cleanup Standard[2] Cs	[2] P_0	[2] P_1	Calculate: $B = \left(Z_{1-\beta}\sqrt{P_1(1-P_1)} + Z_{1-\alpha}\sqrt{P_0(1-P_0)}\right)^2$	$m_d = \dfrac{B}{F(P_0-P_1)^2}$

Column Maximum, (Maximum of m_d values) = C = [   ]

Round C to next largest integer=Number of years of sample collection= m= [   ]

Total number of samples = nm = N = [   ]

Date Completed: _____ Completed by _____

Use additional sheets if necessary. Page ____ of ____

Continue to **WORKSHEET 4**

---

[1] An estimate of ϕ, the serial correlation, is necessary to determine the appropriate value of F. Worksheets 15 and 16 can be used to estimate ϕ.

WORKSHEET 4 Data Records and Calculations When Assessing Individual Wells; by Chemical, Well, and Year

See Chapter 8 or 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	
CHEMICAL:	NUMBER(c) AND DESCRIPTION [2]
WELL:	NUMBER(w) AND DESCRIPTION [1]
YEAR:	NUMBER(K)

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Sample Design (Check one): Fixed Sample Size ☐   Sequential Sampling ☐

Parameter to be tested [2] (Check one) =

Mean ☐
%tile ☐

Number of samples per year [3] = n =

Number of samples with nonmissing data in year = n_k =

Cleanup standard[2] = Cs =

Concentration used for observations below the detection limit =

"Season" Number: j within this k^{th} year	Sample ID	Sample Collection date/time	Reported Concentration	Concentration Corrected for Detection Limit A	Is A Greater than Cs? 1 = Yes 0 = No B	Data for analysis x_{jk} = A if Mean x_{jk} = B if %tile

Total of x_{jk} for this year = C =

Mean of x_{jk} for this k^{th} year = $\dfrac{C}{n_k}$ = $\bar{x}_k$ =

Date Completed: _____   Completed by _____

Use additional sheets if necessary.   Page ____ of ____

Complete **WORKSHEET 4** for other chemicals, years, and wells; otherwise,
Continue to **WORKSHEET 5** if a fixed sample size test is used; or
Continue to **WORKSHEET 7** if a sequential sample test is used.

WORKSHEET 5 Data Calculations for a Fixed Sample Size Test When Assessing Individual Wells; by Chemical and Well

See Chapter 8 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:	
CHEMICAL:	NUMBER(c) AND DESCRIPTION [2]
WELL:	NUMBER(w) AND DESCRIPTION [1]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Year Number k	Mean for the year [4] $\bar{x}_k$	$(\bar{x}_k)^2$

Total from previous page (if more than one Worksheet 5 used)

Column Totals:

A	B

$$A = \sum (\bar{x}_k) \qquad B = \sum (\bar{x}_k)^2$$

Date Completed: _____   Completed by _____

Use additional sheets if necessary.   Page ____ of ____

Complete **WORKSHEET 5** for other chemicals and wells or continue to **WORKSHEET 6**

WORKSHEET 6 Inference for Fixed Sample Sizes Tests When Assessing Individual Wells, by Chemical and Well

See Chapter 8 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:	
CHEMICAL:	NUMBER(c) AND DESCRIPTION [2]
WELL:	NUMBER(w) AND DESCRIPTION [1]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

[2] $\alpha =$

[2] $Cs =$

Number of Years [3] = m =

Sum of the yearly means [5] = $\sum \bar{x}_k$ = A =

Sum of the squared yearly means [5] = $\sum (\bar{x}_k)^2$ = B =

Overall mean concentration = $\dfrac{A}{m}$ = $\bar{\bar{x}}$ =

Standard Deviation of the yearly means = $\sqrt{\dfrac{B - m(\bar{\bar{x}})^2}{m-1}}$ = $s_{\bar{x}}$

Degrees of Freedom for $s_{\bar{\bar{x}}}$ = $m - 1$ = Df =

Critical value from table of the t-distribution
(Table A.1) for specified values of (1-α) and Df = t =

Standard Error for the overall mean = $\dfrac{s_{\bar{x}}}{\sqrt{m}}$ = $s_{\bar{x}_m}$ =

Upper One Sided Confidence Interval = $\bar{\bar{x}} + t\, s_{\bar{x}_m}$ = $\mu_{U\alpha}$ =

If $\mu_{U\alpha} < Cs$ then circle Clean, otherwise circle Contaminated:
Based on the mean concentration, the sampling well is: **Clean   Contaminated**

Date Completed: _____   Completed by _____

Page ____ of ____

Complete **WORKSHEET 6** for other chemicals and wells

WORKSHEET 7a Data Calculations for a Sequential Sample When Assessing Wells Individually; by Chemical and Well

See Chapter 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:	
CHEMICAL:	NUMBER(c) AND DESCRIPTION [2]
WELL:	NUMBER(w) AND DESCRIPTION [1]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Cleanup standard[2] = Cs = ☐

Alternate mean = μ_1 = ☐

Probability of mistakenly declaring the well(s) clean [2] = α = ☐

Probability of mistakenly declaring the well(s) contaminated [2] = β = ☐

Year Number [4] k or m	Yearly Average [4] $\bar{x}_k$	Cumulative Sum of $\bar{x}_k$ ($A_0 = 0$) $A_k = A_{k-1} + \bar{x}_k$	Cumulative Sum of $\bar{x}_k^2$ ($B_0 = 0$) $B_k = B_{k-1} + \bar{x}_k^2$	Mean (average of yearly averages) $\bar{x}_m = \dfrac{A_k}{k}$	Standard Error of Mean $s_{\bar{x}_m} = \sqrt{\dfrac{B_k - k(\bar{x}_m)^2}{(k-1)k}}$

Date Completed: _____   Completed by _____

Use additional sheets if necessary.   Page ____ of ____

Complete WORKSHEETS 7a and 7b for other chemicals and wells

WORKSHEET 7b *Data Calculations for a Sequential Sample When Assessing Wells Individually; by Chemical and Well*

See Chapter 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:	
CHEMICAL:	NUMBER(c) AND DESCRIPTION [2]
WELL:	NUMBER(w) AND DESCRIPTION [1]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Year Number [4] m	$\delta = \dfrac{\mu_1 - Cs}{s\bar{x}_m}$	$t = \dfrac{\bar{x}_m - \dfrac{Cs + \mu_1}{2}}{s\bar{x}_m}$	Likelihood ratio LR*	Critical value: clean $A = \dfrac{\beta}{1-\alpha}$	Critical value: contaminated $B = \dfrac{1-\beta}{\alpha}$	Decision: clean LR > B, contaminated LR ≤ A, or no decision A < LR ≤ B

$$* \ LR = \exp\left(\delta \, \frac{m-2}{m} \, t \, \sqrt{\frac{m}{m-1+t^2}} \right)$$

If "no decision", collect another years' allotment of samples and test the hypothesis again.

Date Completed: _____   Completed by _____

Use additional sheets if necessary.   Page ____ of ____

Complete WORKSHEETS 7a and 7b for other chemicals and wells

WORKSHEET 8 *Attainment Objectives When Assessing Wells as a Group*
See Chapter 3 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Sample Design (Check one): Fixed Sample Size ☐   Sequential Sampling ☐

Probability of mistakenly declaring the well(s) clean = α = [   ]

Probability of mistakenly declaring the well(s) contaminated = β = [   ]

Chemical to be tested number c	Chemical name	Cleanup standard (with units) Cs	Parameter to test: Check one	If mean, enter the alternate hypothesis μ_1	If mean, enter the alternate hypothesis Max_1
			Mean ☐ Max ☐		
			Mean ☐ Max ☐		
			Mean ☐ Max ☐		
			Mean ☐ Max ☐		

Sample Collection Procedures to be used (attach separate sheet if necessary):

Secondary Objectives/ Other purposes for which the data is to be collected:

Use the Chemical Number (c) to refer on other sheets to the chemical described above.
Attach documentation describing the lab analysis procedure for each chemical.

Date Completed: _____   Completed by _____

Use additional sheets if necessary.   Page ____ of ____

Continue to **WORKSHEET 9** if a fixed sample size test is used; or
Continue to **WORKSHEET 10** if a sequential sample test is used.

WORKSHEET 9 *Sample Size When Using a Fixed Sample Size Test for Assessing Wells as a Group*

See Section 8.2 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

From Table A.2, Appendix A

Probability of mistakenly declaring the site clean [8] = α = ☐ $z_{1-\alpha}$ = ☐

Probability of mistakenly declaring the site contaminated [8] = β = ☐ $z_{1-\beta}$ = ☐

Number of samples per year = n = ☐ (based on calculations described in Section 8.2)

Variance factor from Table A.5, Appendix A = F^1 = ☐

For testing the mean concentration

Chemical Number [8] c	Cleanup Standard[8] Cs	[8] μ_1	Standard Deviation of mean $\hat{\sigma}$	Calculate: $B = \left(\dfrac{Cs-\mu_1}{z_{1-\alpha}+z_{1-\beta}}\right)^2$	$m_d = \dfrac{\hat{\sigma}^2}{F*B} + 2$

For testing the maximum concentration across all wells

Chemical Number [8] c	Cleanup Standard[8] Cs	[8] Max_1	Standard Deviation of yearly mean $\hat{\sigma}$	Calculate: $B = \left(\dfrac{Cs-Max_1}{z_{1-\alpha}+z_{1-\beta}}\right)^2$	$m_d = \dfrac{\hat{\sigma}^2}{F*B} + 2$

Column Maximum, (Maximum of m_d values) = C = ☐

Round C to next largest integer=Number of years of sample collection= m= ☐

Total number of samples = nm = N = ☐

Date Completed: _____   Completed by _____

Use additional sheets if necessary.   Page ____ of ____
Continue to **WORKSHEET 10**

---

[1] An estimate of ϕ, the serial correlation, is necessary to determine the appropriate value of F. Worksheets 15 and 16 can be used to estimate ϕ.

WORKSHEET 10 *Data Records for an Individual Well and Calculations When Assessing Wells as a Group; by Chemical, Well and Year*

See Chapter 8 or 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	
CHEMICAL:	NUMBER(c) AND DESCRIPTION [8]
WELL:	NUMBER(w) AND DESCRIPTION [1]
YEAR:	NUMBER(K)

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Parameter to be tested (Check one) =	Mean ☐ Max ☐
Number of samples per year = n =	
Concentration used for observations below the detection limit =	

"Season" Number j	Sample ID	Sample Collection time	Reported Concentration	Concentration Corrected for Detection Limit x_{jk}

Date Completed: _____   Completed by _____

Use additional sheets if necessary.   Page ____ of ____

Complete WORKSHEET 10 for other chemicals, years, and wells or continue to WORKSHEET 11

APPENDIX C: BLANK WORKSHEETS

WORKSHEET 11 *Data Records and Calculations When Assessing Wells as a Group; by Chemical and Year*

See Chapter 8 or 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	
	NUMBER(c) AND DESCRIPTION [8]
CHEMICAL:	
	NUMBER(k)
YEAR:	

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Sample Design (Check one): Fixed Sample Size ☐ Sequential Sampling ☐

Parameter to be tested (Check one) = | Mean ☐ |
| Max ☐ |

Number of samples per year [9] = n = [   ]

"Season" Number[10] j	Well #__ [10] x_{jk}	Well #__ [10] x_{jk}	Well #__ [10] x_{jk}	Well #__ [10] x_{jk}	Well #__ [10] x_{jk}	Measure for analysis (row maximum or row mean) x_j
			•			

Total of x_j for this year = A = [   ]

Mean of x_{jk} for this year = $\bar{x}_k = \dfrac{A}{n}$ = [   ]

Date Completed: _____   Completed by _____

Use additional sheets if necessary.

Complete WORKSHEET 11 for other chemicals; otherwise,
Continue to WORKSHEET 12 if a fixed sample size test is used; or
Continue to WORKSHEET 14 if a sequential sample test is used.

WORKSHEET 12 *Data Calculations for a Fixed Sample Size Test When Assessing Wells as a Group; by Chemical*

See Chapter 8 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:	
	NUMBER(C) AND DESCRIPTION [8]
CHEMICAL:	

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Year Number k	Mean for the year [11] $\bar{x}_k$	$(\bar{x}_k)^2$

Total from previous page (if more than one copy of Worksheet 12 is necessary)

Column Totals:

A	B

$$A = \sum (\bar{x}_k) \qquad B = \sum (\bar{x}_k)^2$$

Date Completed: _____   Completed by _____

Use additional sheets if necessary.   Page ____ of ____

Complete **WORKSHEET 12** for other chemicals or continue to **WORKSHEET 13**

WORKSHEET 13 *Inference for Fixed Sample Sizes Tests When Assessing Wells as a Group; by Chemical*

See Chapter 8 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:
NUMBER(c) AND DESCRIPTION [8]
CHEMICAL:

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

[8] α =

[8] C_s =

Number of Years [9] =   m =

Sum of the yearly means [12] = $\sum \bar{x}_k$ = A =

Sum of the squared yearly means [12] = $\sum (\bar{x}_k)^2$ = B =

Overall mean concentration = $\dfrac{A}{m}$ = $\bar{x}$ =

Standard Deviation of the yearly means = $\sqrt{\dfrac{B - m(\bar{x})^2}{m-1}}$ = $s_{\bar{x}}$

Degrees of Freedom for $s_{\bar{x}}$ = $m - 1$ = Df =

Value from table of T-distribution (Appendix A.1)
for specified values of $(1 - \alpha)$ and Df =   t =

Standard Error for the overall mean = $\dfrac{s_{\bar{x}}}{\sqrt{m}}$ = $s_{\bar{x}_m}$ =

Upper One Sided Confidence Interval = $\bar{x} + t\, s_{\bar{x}_m}$ = $\mu_{U\alpha}$ =

If $\mu_{U\alpha}$ < Cs then circle Clean, otherwise circle Contaminated:
Based on the mean concentration, the sampling well is: **Clean   Contaminated**

Date Completed: _____   Completed by _____

Page ____ of ____

Complete **WORKSHEET 13** for other chemicals

WORKSHEET 14a Data Calculations for a Sequential Sample When Assessing Wells as a Group; by Chemical

See Chapter 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:
NUMBER(c) AND DESCRIPTION [8]
CHEMICAL:

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Cleanup standard[8] = Cs = [ ]

Alternate mean = μ_1 = [ ]

Probability of mistakenly declaring the well(s) clean [8] = α = [ ]

Probability of mistakenly declaring the well(s) contaminated [8] = β = [ ]

Year Number [11]	Yearly Average [11]	Cumulative Sum of $\bar{x}_k$ ($A_0 = 0$)	Cumulative Sum of $\bar{x}_k^2$ ($B_0 = 0$)	Mean (average of yearly averages)	Standard Error of Mean
k or m	$\bar{x}_k$	$A_k = A_{k-1} + \bar{x}_k$	$B_k = B_{k-1} + \bar{x}_k^2$	$\bar{x}_m = \dfrac{A_k}{k}$	$s\bar{x}_m = \sqrt{\dfrac{B_k - k(\bar{x}_m)^2}{(k-1)k}}$

Date Completed: _____   Completed by _____

Use additional sheets if necessary.   Page ____ of ____

Complete WORKSHEET 14a and 14b for other chemicals and groups of wells

WORKSHEET 14b *Data Calculations for a Sequential Sample When Assessing Wells as a Group; by Chemical*

See Chapter 9 in "Methods for Evaluating the Attainment of Cleanup Standards", Volume 2

SITE:
CHEMICAL: NUMBER(c) AND DESCRIPTION [8]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Year Number [4] m	$\delta_1 = \dfrac{\mu_1 - Cs}{s_{\bar{x}_m}}$	$t = \dfrac{\bar{x}_m - \dfrac{Cs + \mu_1}{2}}{s_{\bar{x}_m}}$	Likelihood ratio LR* ·	Critical value: clean $A = \dfrac{\beta}{1-\alpha}$	Critical value: contaminated $B = \dfrac{1-\beta}{\alpha}$	Decision: clean LR > B, contaminated LR ≤ A, or no decision A < LR ≤ B

$$* \; LR = \exp\left(\delta \, \frac{m-2}{m} \, t \, \sqrt{\frac{m}{m-1+t^2}} \right)$$

If "no decision", collect another years' allotment of samples and test the hypothesis again.

Date Completed: _____   Completed by _____

Use additional sheets if necessary.   Page ____ of ____

Complete **WORKSHEET 14a** and **14b** for other chemicals and groups of wells

WORKSHEET 15 Removing Seasonal Patterns in the Data (Use as First Step in Computing Serial Correlations)

See Sections 8.4 and 9.4 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	
CHEMICAL:	NUMBER(c) AND DESCRIPTION [2 OR 8]
WELL:	NUMBER(w) AND DESCRIPTION [1]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

"Season" Number j	Measurements for each "season" for year k Yr=___ x_{jk}	Yr=___ x_{jk}	Yr=___ x_{jk}	Yr=___ x_{jk}	Yr=___ x_{jk}	Number of years with Data m_j	Row Total $T_j = \sum_k x_{jk}$	Row Mean $\bar{x}_j = \dfrac{T_j}{m_j}$

Corrected measurements with seasonal patterns removed

"Season" Corrected Measurements for each "season" for year k

Number j	Yr=___ $x_{jk} - \bar{x}_j$	Yr=___ $x_{jk} - \bar{x}_j$	Yr=___ $x_{jk} - \bar{x}_j$	Yr=___ $x_{jk} - \bar{x}_j$	Yr=___ $x_{jk} - \bar{x}_j$

Date Completed: _____   Completed by _____

Use additional sheets if necessary.   Page ____ of ____

Complete **WORKSHEET 15** for other chemicals
Continue to **WORKSHEET 16** if serial correlations are being computed.

WORKSHEET 16 Calculating Serial Correlations

See Sections 8.4 and 9.4 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	
CHEMICAL:	NUMBER(c) AND DESCRIPTION [2 OR 8]
WELL:	NUMBER(w) AND DESCRIPTION [1]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Year = k =

Period between well samples in months = t =

Data Numbers	Residual	Product	
jk (season within year k)	$e_{jk} = x_{jk} - \bar{x}_j$ [15]	$d_{jk} = e_{jk}*e_{j-1\,k}$	e_{jk}^2

Totals from previous page =
(if more than one
Worksheet 16 is used)

Column Totals =   **A**   **B**

Estimated Serial Correlation based on the data $= \dfrac{A}{B} = \hat{\phi}_{obs} =$

Serial Correlation between monthly observations $= \hat{\phi} = \left(\hat{\phi}_{obs}\right)^{\frac{1}{t}} =$

Date Completed: _____   Completed by _____

Use additional sheets if necessary.   Page ____ of ____

Complete **WORKSHEET 16** for other chemicals

WORKSHEET 1R Basic Calculations for a Simple Linear Regression

See Section 6.1 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	
CHEMICAL:	NUMBER(c) AND DESCRIPTION [2 OR 8]
WELL:	NUMBER(w) AND DESCRIPTION [1]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Concentration used when no concentration is reported = ▢

Number of collectable samples = N = ▢

Sample Number n	Concentration Corrected for Detection Limit y_n	y_n^2	Transformed Time Variable x_n	x_n^2	$x_n y_n$

Totals from previous page(s):

Column Totals:

	A	B	C	D	E

$$A = \Sigma y_n \qquad B = \Sigma y_n^2 \qquad C = \Sigma x_n \qquad D = \Sigma x_n^2 \qquad E = \Sigma y_n x_n$$

Corrected Sum of Squares and Cross Products:

$$\bar{y} = \frac{A}{N} \qquad \bar{x} = \frac{C}{N} \qquad S_{yy} = B - \frac{A^2}{N} \qquad S_{xx} = D - \frac{C^2}{T} \qquad S_{yx} = E - \frac{AC}{N}$$

Date Completed: _____   Completed by _____

Use additional sheets if necessary.   Page ____ of ____

Complete **WORKSHEET 1R** for other chemicals or continue to **WORKSHEET 2R.**

WORKSHEET $2R$ *Inference in a Simple Linear Regression*

See Section 6.1 in "Methods for Evaluating the Attainment of Cleanup Standards", Vol. 2

SITE:	
CHEMICAL:	NUMBER(c) AND DESCRIPTION [2 OR 8]
WELL:	NUMBER(w) AND DESCRIPTION [1]

Numbers in square brackets [] refer to the Worksheet from which the information may be obtained.

Estimating Regression Coefficients

S_{yy} [1R] = [   ]   Number of collectable

samples [1R] = N = [   ]

S_{xx} [1R] = [   ]   Mean of y_t [1R] = $\bar{y}$ = [   ]

S_{yx} [1R] = [   ]   Mean of x_t [1R] = $\bar{x}$ = [   ]

Type 1 error probability α = [   ]   Estimated slope [1R], b_1 = $\dfrac{S_{yx}}{S_{xx}}$ = [   ]

Estimated Intercept [1R], b_0 = $\bar{y} - (b_1 * \bar{x})$ = [   ]

Sum of squares due to error [1R], SSE = $S_{yy} - \left(\dfrac{S_{yx}^2}{S_{xx}}\right)$ = [   ]

Degrees of freedom, Df = $N-2$ = [   ]

Critical value from table of t-distribution (Appendix A.1)
for specified values of $(1 - \frac{\alpha}{2})$ and Df = t = [   ]

Mean Square Error, MSE = $\dfrac{SSE}{N-2}$ = [   ]

Standard Error of the Slope, $s(b_1)$ = $\sqrt{\dfrac{MSE}{S_{xx}}}$ = [   ]

Upper Two Sided Confidence Interval
for Slope: $b_1 + t * s(b_1)$ = [   ]

Lower Two Sided Confidence
Interval for Slope: $b_1 - t * s(b_1)$ = [   ]

Calculating Prediction Limits

Value of x_t at which concentration is to be predicted = [   ]

Predicted value, $\hat{y}$ = $b_0 + b_1 x_t$ = [   ]

Standard Error of Predicted Value = $S\hat{y}$ = $\sqrt{MSE\{1 + \dfrac{1}{N} + \dfrac{(x_t - \bar{x})^2}{S_{xx}}\}}$ = [   ]

Upper Two Sided Confidence Interval for Prediction = $\hat{y} + t * S\hat{y}$ = [Upper   ]

Lower Two Sided Confidence Interval for Prediction = $\hat{y} - t * S\hat{y}$ = [Lower   ]

Date Completed: _____   Completed by _____

Use additional sheets if necessary.   Page ____ of ____

Complete **WORKSHEET 2R** for other chemicals

APPENDIX D: MODELING THE DATA

A model is a mathematical description of the process or phenomenon from which the data arc collected. A model provides a framework for extrapolating from the measurements obtained during the data collection period to other periods of time and describing the important characteristics of the data. Perhaps most importantly, a model serves as a formal description of the assumptions which are being made about the data. The choice of statistical method used to analyze the data depends on the nature of these assumptions.

The results of the statistical analysis may be sensitive to the degree to which the data adhere to the assumptions of the analysis. If the statistical results arc quite insensitive to the validity of a particular assumption, the statistical methods arc said to be "robust" to departures from that assumption. On the other hand, if the results are sensitive to an assumption so that the results may be substantially incorrect if the assumption does not hold, the validity of that assumption should be checked before the results of the analysis arc used or given credence.

After steady state conditions have been reached, the model assumed to describe the ground water data is the equation in Box D.1.

The laboratory measurement, x_{tcw}, will be expressed in measurement units selected by either the lab or the management of the cleanup effort. All terms in the model equation must have the same units. The samples on which the measurements are made can be identified by the time and location of collection. In the model above, the location is indicated by the well identifier w. For wells in which samples are collected at different depths or by different sampling equipment, a more extensive set of identifiers and subscripts will be required. If the parameter being tested represents a group of wells (e.g., an average concentration in several wells), x_{tcw} represents the combined measure and w refers to the group of wells.

Box D.1
Modeling the Data

The model assumed to describe ground water data after steady-state conditions have been reached is:

$$x_{tcw} = \mu_{cw} + S_{u(t)cw} + z_{tcw} + \varepsilon_{tcw} \qquad (D.1)$$

where

x_{tcw} =   lab measurement of chemical c for the sample collected at time t for well w.

μ_{cw} =   long-term (or short-term) average concentration for chemical c in well w.

$S_{u(t)cw}$ =   a seasonal pattern in the data for concentration of chemical c in well w, assumed to repeat on a regular cycle. The subscript u(t) designates the point in time within the cycle when the sample was collected. In most situations the term $S_{u(t)cw}$ will correspond to a yearly cycle associated with yearly patterns in temperature and precipitation.

z_{tcw} =   serially correlated normal error following an auto-regressive model of order one (Box and Jenkins, 1970). (Note: seasonal auto-correlations are assumed to be negligible after the seasonal cycles ($S_{u(t)cw}$) have been removed). The correlation, ρ, between two measurements separated by time t (in months) is assumed to be $\rho = R^t$ where R is the correlation for measurements separated by one month.

ε_{tcw} = independent normal errors.

This model for the data assumes that the average level of contamination is constant over the period, of concern (either a short or very long period). However, the actual measurements may fluctuate around that level due to seasonal differences, lab measurement errors, or serially correlated fluctuations (described below). The purpose of the statistical test is to decide if there is sufficient evidence to conclude that μ_{cw} is less than the cleanup standard in the presence of this variability.

APPENDIX D: MODELING THE DATA

Because the primary cyclical force affecting the ground water system is climatic, in most situations the seasonal term will have a period of one year. In some climates there are two rainy seasons and two dry seasons, possibly resulting in a seasonal pattern of a half year. The connection between the seasonal pattern in the ground water concentrations and the climatic changes may be-complex such that both patterns may have the same period; however, the shape of the patterns, the relative times of maximum rainfall or the maximum or minimum concentration may differ.

Ground water concentrations at points close together in time or space are likely to be more similar than observations taken far apart in time or space. There are several physical reasons why this may be the case. In statistical terms, observations taken close together are said to be more correlated than observations taken far apart

The serial correlation of observations separated by a time difference oft can be denoted by $\rho(t)$, where ρ is the Greek letter rho (ρ). A plot of the serial correlation between two observations versus the time separating the two observations is called an auto-correlation function. The model above assumes that the autocorrelation function has the shape shown in Figure D.1, which is described by the equation in Box D.2.

Box D.2
Auto&relation Function

$$\rho(t) = R^t \qquad\qquad (D.2)$$

where R is the serial correlation for measurements separated by a month, and t is the time between observations in months.

If the serial correlation of the measurements is zero, the data behave as if they were collected randomly. As the correlation increases, the similarity of measurements taken close together relative to all other measurements becomes more pronounced. Figure D.2 shows simulated data with serial correlations of 0.0, 0.4 and 0.8. Serial correlations are always between -1 and 1. However, for most environmental data, serial correlations are usually between 0 and 1, indicating that measurements taken close together in time will be more alike than measurements taken far apart.

Figure D.1   Theoretical Autocorrelation Function Assumed in the Model of the Ground Water Data

The between observations
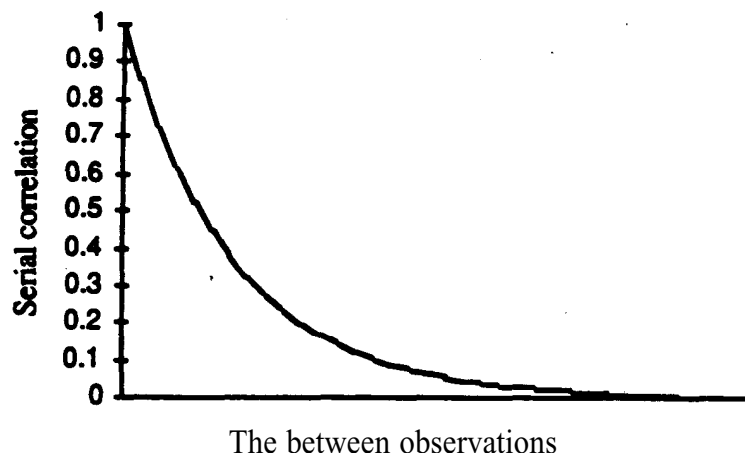
Many common statistical procedures will provide incorrect conclusions if an existing correlation in the data is not properly accounted for. For example, the variability in the data may be inappropriately estimated. Proper selection of a simple random sample for estimating the mean guarantees that the errors are uncorrelated. However, when using a systematic sample (such as for ground water samples collected at regular intervals), the formulae based on a random sample provide a good estimate of the standard error of the mean only if there is no serial correlation. With serial correlation, a correction term is required. For the autocorrelation function assumed above, the correction term increases the standard error of the long-term mean and decreases it for the short-term mean.

The autocorrelation function can have many different shapes; however, in general, correlations will decrease as the time between observations increases. If the samples are taken farther apart in time, the correction becomes less important.

The error term, ε_{tcw}, represents errors resulting from lab measurement error and other factors associated with the environment being sampled and the sample handling procedures.

Figure D.2   Examples of Data with Serial Correlations of 0, 0.4, and 0.8. The higher
the serial correlation the more the distribution dampens out
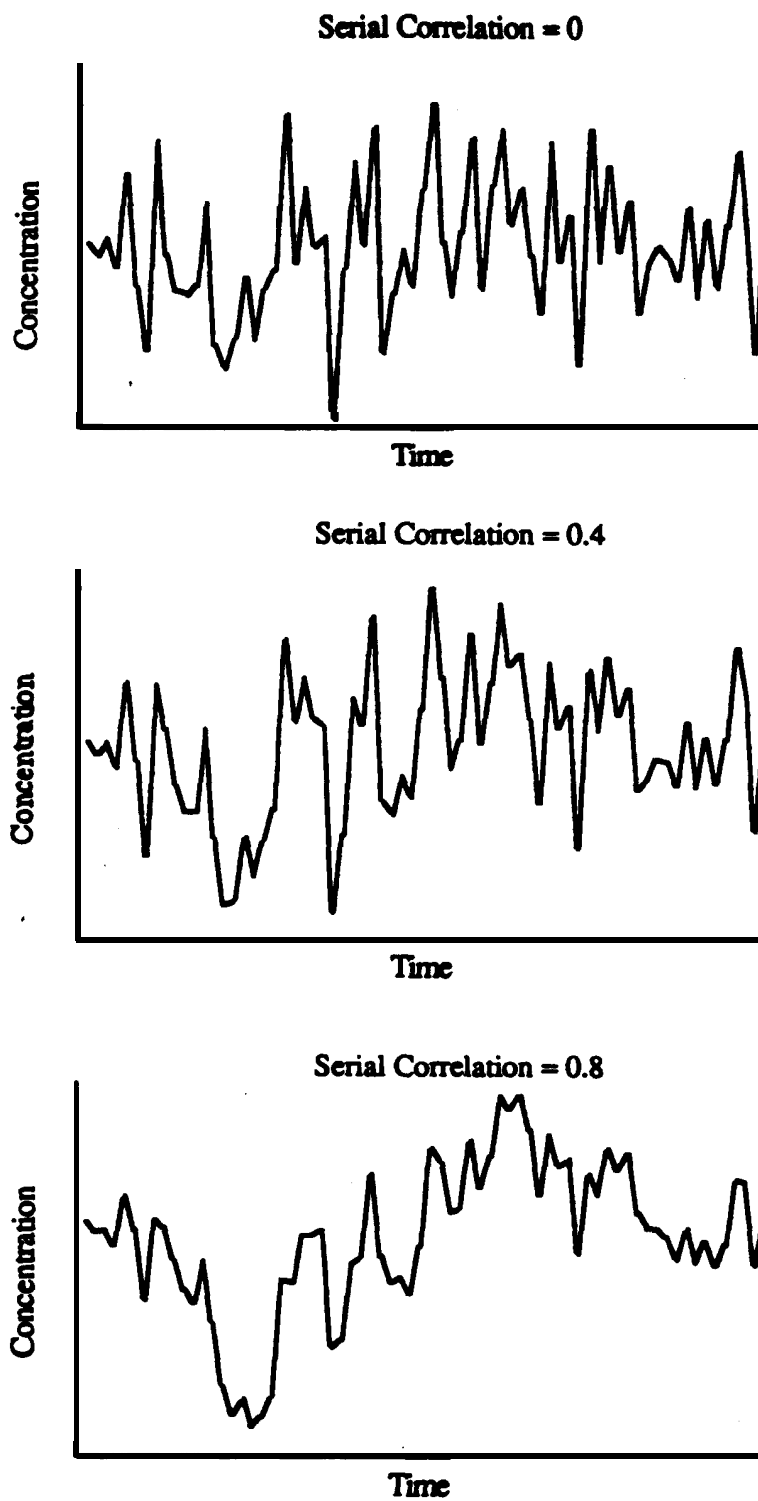
APPENDIX D: MODELING THE DATA

Different models may be used to describe the data collected during the treatment phase and the post-treatment assessment phase because either (1) the characteristics of the data will be different, or (2) different information about the measured concentrations is of interest. The statistical procedures discussed in Chapter 6 to be used during treatment are therefore different from those discussed in Chapters 8 and 9 for assessing attainment of the cleanup standards.

There are two terms which have been excluded from the model above and could be used to model ground water concentrations in some situations. These are a slope (or trend) term and a spatial correlation term.

In many situations it is reasonable to assume that the general level of contamination is either gradually decreasing or gradually increasing. It may be desirable to assume a functional form for this change in concentration. For example, the concentration may be considered to be decreasing linearly a exponentially. A revised model with a linear trend term is presented in Box D.3.

If the slope is not zero, as in the model in Box D.3, then the ground water is not at steady state. If the slope is positive, the concentrations are increasing over time. If the slope is negative, the concentrations are decreasing over time. If concentrations are below the cleanup standard and are increasing over time, the ground water may be judged to attain the cleanup standard; however the cleanup standard may not be attained in the future as concentrations increase. Therefore, the ground water in the sampled wells will be judged to attain the cleanup standard only if (1) the selected parameter is significantly less than the cleanup standard, and (2) the concentrations are not increasing. This decision criteria is presented in Table D.1.

The model in Box D.3 does not include spatial correlation. In this guidance, it is assumed that the results from different wells (or different depths in the same well) are combined using criteria developed based on expert knowledge of the site rather than by fitting statistical models. For this reason a spatial correlation has not been included.

Box D.3
Revised Model for Ground Water Data

A revised model with a linear trend term would be:

$$x_{tcw} = \alpha_{cw} + \beta_{cw}t + S_{u(t)cw} + z_{tcw} + \varepsilon_{tcw} \qquad (D.3)$$

where

β_{cw} = the change in concentration over time for measurements of chemical c in well w.

α_{cw} = the concentration of chemical c in well w at time zero, usually at the beginning of sampling. Note that $\alpha_{cw} = \mu_{cw}$ if $\beta_{cw} = 0$.

Table D.1   Decision criteria for determining whether the ground water concentrations attain the cleanup standard

Test for parameter (mean or percentile) less than the cleanup standard (Equation D.2)	Test for significant slope β_{cw} (Equation D.3)	
	β_{cw} significantly greater than zero	β_{cw} **not** significantly greater than zero
Parameter is significantly less than the cleanup standard	Ground water is contaminated	Ground water from the tested wells attains the cleanup standard
Parameter is **not** significantly less than the cleanup standard	Ground water is contaminated	Ground water is contaminated

APPENDIX E: CALCULATING RESIDUALS AND SERIAL CORRELATIONS
USING SAS[1]

Several statistical programs can be used-to make the calculations outlined in this guidance document. Although these programs can be used to perform the required calculations, they were not specifically designed for the application addressed in this document. Therefore, they can only be used as a partial aid for the procedures presented here. Only one of the many available statistical packages, SAS, will be discussed below in the example. This example makes no attempt to thoroughly introduce the SAS system, and no endorsement of SAS is implied. Help from a statistician or programmer familiar with any software being used is strongly recommended.

The basic quantities discussed in the Sections 5.2.3 and 5.2.4 can be calculated using one of several statistical procedures available in SAS. Among them ate PROC GLM, PROC ANOVA, and PROC REG (see SAS Users Guide: *Statistics,* SAS Institute, 1985). All of these procedures require specifying a linear model and requesting certain options in the MODEL statement A SAS data set containing the data to be used in the analysis should first be created (see *SAS* Users Guide: *Basics,* SAS Institute, 1985). In the data set, the observations should be listed or sorted in time order, otherwise the calculated serial correlations will be meaningless.

Given below is an example of a SAS program using PROC REG that will subtract seasonal means from the observed concentration measurements and calculate the required first order serial cotrelation of the residuals.

```
PROC REG DATA = CHEMl;
        MODEL CONC = SEAS1 SEAS2 SEAS3 SEAS4/NOINT,DW;
```

In the program, CHEM1 is the SAS data set containing the following variables: CONC, the concentration measurement of the ground water sample; TIME, a sequence number indicating the time at which the sample was drawn; YEAR, the year the sample was drawn, and PER, the period within the year in which the sample was drawn. For this illustration, data were collected quarterly so that PER = 1, 2, 3, or 4. The variables SEAS 1 through SEAS4 are indicator variables defined at a previous DATA step. For each observation, these indicator variables are defined as follows: SEAS1 = 1 if PER = 1, and is 0, otherwise; SEAS2 = 1 if PER = 2, and is 0

---

otherwise; SEAS3 = 1 if PER = 3, and is 0, otherwise; and SEAS4 = 1 if PER = 4, and is 0, otherwise. Creation of these indicator or "dummy" variables is required if PROC REG is used. On the other hand, dummy variables arc not required for PROC ANOVA or PROC GLM. Note that in this example, the variable TIME is not included as an independent variable in the model.

The model statement specifics the form of the linear model to be fitted. In the example, CONC is the dependent variable and SEAS1 through SEAS4 are the independent variables. The reason for specifying this particular model is to have the seasonal means subtracted from the observed concentrations. NOINT is an option that specifics that a "no-intercept model" is to be estimated. Other models can also be used to produce the required residuals, but they will not be discussed here. Finally, DW is the "Durbin-Watson" option, which requests that the Durbin-Watson test (see Section 5.6.1) and the serial correlation of the residuals be calculated. The output from the above computer run will look like:

DEP Variable: CONC

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	4	580.455	145.114	1051.355	0.000
ERROR	12	1.656	0.138		

ROOT MSE	0.3715	R-SQUARE	0.997	
DEP MEAN	5.995	ADJ R-SQ	0.996	
C.V.	6.197			

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR HO: PARAMETER=0	PROB>|T|
SEAS1	1	6.778	0.186	36.490	0.000
SEAS2	1	6.025	0.186	36.490	0.000
SEAS3	1	5.134	0.186	36.490	0.000
SEAS4	1	6.042	0.186	36.490	0.000

DURBIN-WATSON D	2.280
1ST ORDER AUTOCORRELATION	-.184

The first part of the output (identified by the heading SOURCE, DF, SUM OF SQUARES, etc.) is referred to as the "analysis of variance table." In the "MEAN SQUARE" column of the table corresponding to the row titled "ERROR," is the mean square error, s_e^2. In the example output, $s_e^2 = 0.138$.

The second part of the output gives the "PARAMETER ESTIMATES" for each of the four indicator variables, SEAS1 to SEAS4. Because of the way these variables were defined, the parameter estimates are actually the seasonal means, $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$, and $\bar{x}_4$, respectively. These seasonal means are used to calculate the residuals, e_t, as defined in equation (5.8). The last line of the output shows the serial correlation of the residuals as computed from equation (5.14), viz., $\hat{\phi}_{obs} = -.184$. From Neter, Wasserman, and Kutner (1985), $d_u = 1.73$, for $N = 16$ (16 observations) and $p - 1 = 3$ (where p is the number of variables in the model). Since $D = 2.28 > 1.73$, it can be assumed that there is no autocorrelation in the error terms of the model.

As mentioned earlier, PROC GLM or PROC ANOVA can also be used to compute the required statistical quantities. The interested reader should refer to the SAS users manual for more information.

APPENDIX F: DERIVATIONS AND EQUATIONS

This appendix provides background fop several equations presented in the document. This background is provided only far equations which cannot be easily verified in a standard statistical text. A simulation study provides the background for the sequential tests presented in Chapter 9. The simulation study was supported by Westat. The last section of this appendix incorporates a technical paper prepared for publication which summarizes the simulations.

F.1   Derivation of Tables A.4 and A.5

This section outlines the derivation of Table A.4 for determining a recommended number of samples to take per year and Table A.5 for obtaining variance factors for use in determining sample size. Table A.4 is based on the assumption that the number of samples per year will be chosen to minimize the total sampling costs while still achieving the desired precision. The assumptions on which the derivation is based arc explained below. The values in Table A.5 follow directly from the calculations used to obtain Table A.4.

For a fixed sample size test, the cost of the sampling program can be approximated by:

$$C = E + (Y+nS)m \qquad (F.1)$$

where

c = the total cost of the sampling program;
E = the cost to establish the sampling program;
Y = the yearly cost to maintain the program;
S = the incremental cost to collect each sample;
n = the number of samples per year and
m = the number of years of sampling

This can also be written as:

$$C = E + S(R+n)m \qquad (F.2)$$

Where $R = \frac{1}{c}$. Since E and S are constants, the total sampling cost can be minimized by minimizing $(R + n)m$ subject to the constraint that the choices of n and m achieve the desired precision. The total number of samples collected is:

$$N = nm \qquad \text{(F.3)}$$

Consider the hypothesis test where a mean is being compared to a standard and assume that 1) the measurements are independent and 2) a normal approximation can be used. Then the following equation can be used to determine the required sample size:

$$N_{eff} = \sigma^2 \left\{ \frac{z_{1-\beta} + z_{1-\alpha}}{Cs-\mu_1} \right\}^2 \qquad \text{(F-4)}$$

Where:

σ^2 = variance of the individual measurements;

Cs = the cleanup standard to which the mean is being compared;

μ_1 = the concentration on which the alternate hypothesis and β are based;

α = the probability of a false positive decision if the true mean is Cs;

β = the probability of a false negative decision if the true mean is μ_1;

$z_{1-\alpha}$ = the 1-α percentile point of the normal distribution; and

N_{eff} = the required number of independent observations.

Noting that $\dfrac{\sigma^2}{N_{eff}}$ is the standard error of the mean based on independent measurements, equation (F.4) can be rewritten as:

$$N_{eff} = \frac{\sigma^2}{\sigma_{nm}^2} = \sigma^2 \left\{ \frac{z_{1-\beta} + z_{1-\alpha}}{Cs-\mu_1} \right\}^2 \qquad \text{(F.5)}$$

Where: σ_{nm}^2 = the standard error of the mean when taking n samples per year over m years (for correlated observations, the variance of the mean depends on the individual values of n and m rather than just the total number of samples).

The problem is to select the combination of n and m such that equation (F.5) is satisfied and the sampling costs are minimized.

The values of n and m which satisfy equation (F.5) depend only slightly on the values of α, β, Cs, μ_1, and σ^2. For the purposes of estimating the values in Table A.4 and A.5, the following assumptions were used: $\alpha = .10$, $\beta = .10$, Cs $= 1$, $\mu_1 = .5$, and $\sigma^2 = 1.0$, resulting in $N_{eff} = 26.3$.

The following equation (derived in section F.2) can be used for N_{eff} for the mean of n observations per year collected over m years with a lag 1 serial correlation of ϕ.

$$N_{eff} = \frac{N}{\left(\frac{1+\phi}{1-\phi}\right)\left(1 - \frac{2\phi(1-\phi^N)}{N(1-\phi^2)}\right)} \qquad (F.6)$$

Note that the serial correlation in equation (F.6) is the serial correlation between successive observations. As the number of observations per year changes, ϕ will also change. If Φ is the serial correlation between monthly observations, then $\phi = \sqrt[n]{\Phi^{12}}$.

The values in Tables A.4 and A.5 were calculated using the following procedures:

(1)   For selected values of Φ and n, calculate ϕ and use a successive approximation procedure to determine m such that the criteria in equation (F.6) are met.

(2)   The values in Table A.5 are $\frac{N_{eff}}{m}$, or the effective number of samples per year;

(3)   For each calculation in step (1) and for selected values of R, calculate the sampling cost using equation (F.2).

(4)   Using all the sampling costs calculated for the selected values of Φ, n, and R, determine the value of n which has the minimum sampling cost. Show this value in Table A.4.

F.2   Derivation of Equation (F.6)

A series of periodic ground water measurements following an autoregressive (AR(1)) process can be described by the following equation (see Box and Jenkins (1970) for details):

$$x_t = \mu + \sum_{i=0}^{\infty} \phi^i a_{t-i} = \mu + z_t \tag{F.7}$$

where:

x_t = the measurement at time t;

μ = the long-term (attainment) mean concentration ;

ϕ = the serial correlation between successive measurements;

a_t = a random change from the measurement at time t-i to time t such that $x_t - \phi x_{t-1} = a_t$. The a_t are assumed to be independent and have a mean of zero and a variance of ϵ^2; and

z_t = the difference between the mean being estimated and the measurement at time t. The values z_t will have a mean of zero.

The mean of N successive observations is

$$\bar{x} = \frac{1}{N} \sum_{k=0}^{N-1} x_{t-k} = \mu + \frac{1}{N} \sum_{k=0}^{N-1} z_{t-k} = \mu + \bar{z} . \tag{F.8}$$

The variance of z_t and $\bar{z}$ are derived below. Note that the variance of x_t and z_t are the same, written $V(x_t) = V(z_t)$; also, $V(\bar{x}) = V(\bar{z})$.

The following relationships are used in the derivation of the variance:

$$\frac{1}{1-\phi} = 1 + \phi + \phi^2 + \phi^3 + ... \tag{F.9}$$

and

$$\frac{1-\phi^N}{1-\phi} = 1 + \phi + \phi^2 + ... + \phi^{N-2} + \phi^{N-1} \tag{F.10}$$

F.2.1   Variance of z_t

The variance of z_t is:

$$V(z_t) = E\left[z_t^2 - E[z_t]^2\right] \tag{F.11}$$

Here E[ ] indicates the expected value of the term inside the brackets.

Since $E[z_t]$ is zero, the variance can be written as;

$$V(z_t) = E\left[z_t^2\right] \tag{F.12}$$

$$= E\left[\left[\sum_{i=0}^{\infty} \phi^i a_{t-i}\right]^2\right] \tag{F.13}$$

$$= E\left[\sum_{i=0}^{\infty} \phi^{2i} a_{t-i}^2\right] \tag{F.14}$$

Since the expected value of all the cross product terms are zero (i.e., $E[z_t z_{t-i}]=0$, for $i \neq 0$), they have been dropped from the summation.

$$V(z_t) = \sum_{i=0}^{\infty} E[\phi^{2i} a_{t-i}^2] \tag{F.15}$$

$$V(z_t) = \sum_{i=0}^{\infty} \phi^{2i} E[a_{t-i}^2] \tag{F.16}$$

Since $E[a_{t-i}^2] = \varepsilon^2$,

$$V(z_t) = \varepsilon^2 \sum_{i=0}^{\infty} \phi^{2i} = \varepsilon^2 (1 + \phi^2 + \phi^4 + ..) \tag{F.17}$$

Using equation (F.9):

$$V(z_t) = \frac{\epsilon^2}{1-\phi^2} = \sigma^2 \qquad\qquad (F.\ 18)$$

F.2.2   Variance of $\bar{z}$

Note that $\bar{z}$ can be expressed as

$$\bar{z} = \frac{1}{N} \sum_{k=0}^{N-1} z_{t-k} = \frac{1}{N} \sum_{k=0}^{N-1} \sum_{i=0}^{\infty} \phi^i a_{t-i-k} \qquad\qquad (F.19)$$

$$\bar{z} = \frac{1}{N} \sum_{i=0}^{N-1} \left(\frac{1-\phi^{i+1}}{1-\phi} \right) a_{t-i} + \frac{1}{N} \sum_{i=N}^{\infty} \left(\frac{(1-\phi^N)\phi^{i-N+1}}{1-\phi} \right) a_{t-i}. \qquad (F.20)$$

This last relationship is illustrated in the Table F.1 for the case where N = 3.

The variance of $\bar{z}$ is:

$$V(\bar{z}) = E \left[\bar{z}^2 - E\left[\bar{z}\right]^2 \right] \qquad\qquad (F.21)$$

Since $\dot{E}\left[\bar{z}\right]$ is zero, the variance can be written as;

$$V(\bar{z}) = E \left[\bar{z}^2 \right] \qquad\qquad (F.22)$$

$$V(\bar{z}) = E \left[\left[\frac{1}{N} \sum_{i=0}^{N-1} \left(\frac{1-\phi^{i+1}}{1-\phi} \right) a_{t-i} + \frac{1}{N} \sum_{i=N}^{\infty} \left(\frac{(1-\phi^N)\phi^{i-N+1}}{1-\phi} \right) a_{t-i} \right]^2 \right] \qquad (F.23)$$

Table F.1   Coefficients for the terms a_t, a_{t-1}, etc., in the sum of three successive correlated observations

observation	term						
	a_t	a_{t-1}	a_{t-2}	a_{t-3}	a_{t-4}	a_{t-5}	...
$z_t =$	1	ϕ	ϕ^2	ϕ^3	ϕ^4	ϕ^5	...
$z_{t-1} =$	1	ϕ	ϕ^2	ϕ^3	ϕ^4		...
$z_{t-2} =$	1	ϕ	ϕ^2	ϕ^3	...		
$\bar{z} =$	$\dfrac{1}{3}$	$\dfrac{1-\phi^2}{3(1-\phi)}$	$\dfrac{1-\phi^3}{3(1-\phi)}$	$\dfrac{(1-\phi^3)\phi}{3(1-\phi)}$	$\dfrac{(1-\phi^3)\phi^2}{3(1-\phi)}$	$\dfrac{(1-\phi^3)\phi^3}{3(1-\phi)}$	...

Since the expected value of all the cross product terms are zero (i.e., $E[z_t z_{t-i}]=0$, for $i \neq 0$) they have been dropped from the summation.

$$V(\bar{z}) = E\left[\frac{1}{N^2} \sum_{i=0}^{N-1} \left(\frac{1-\phi^{i+1}}{1-\phi}\right)^2 a_{t-i}^2 + \frac{1}{N^2} \sum_{i=N}^{\infty} \left(\frac{(1-\phi^N)\phi^{i-N+1}}{1-\phi}\right)^2 a_{t-i}^2 \right] \quad (F.24)$$

$$V(\bar{z}) = \frac{1}{N^2} \sum_{i=0}^{N-1} \left(\frac{1-\phi^{i+1}}{1-\phi}\right)^2 E\left[a_{t-i}^2\right] + \frac{1}{N^2} \sum_{i=N}^{\infty} \left(\frac{(1-\phi^N)\phi^{i-N+1}}{1-\phi}\right)^2 E\left[a_{t-i}^2\right] \quad (F.25)$$

Since $E\left[a_{t-i}^2\right] = \varepsilon^2$,

$$V(\bar{z}) = \frac{\varepsilon^2}{N^2(1-\phi)^2} \left(\sum_{i=0}^{N-1} \left(1-\phi^{i+1}\right)^2 + \sum_{i=N}^{\infty} \left((1-\phi^N)\phi^{i-N+1}\right)^2 \right) \quad (F.26)$$

$$V(\bar{z}) = \frac{\varepsilon^2}{N^2(1-\phi)^2} \left(\sum_{i=0}^{N-1} \left(1-2\phi^{i+1}+\phi^{2i+2}\right) + \sum_{i=0}^{\infty} (1-\phi^N)^2 \phi^{2i+2} \right) \quad (F.27)$$

Using equations (F.9) and (F.10):

$$V(\bar{z}) = \frac{\varepsilon^2}{N^2(1-\phi)^2} \left(N - 2\phi\frac{(1-\phi^N)}{(1-\phi)} + \phi^2\frac{(1-\phi^{2N})}{(1-\phi^2)} + (1-\phi^N)^2\frac{\phi^2}{(1-\phi^2)} \right) \quad (F.28)$$

This can be simplified to:

$$V(\bar{z}) = \sigma_{nm}^2 = \frac{\varepsilon^2}{N(1-\phi)^2}\left(1 - \frac{2\phi(1-\phi^N)}{N(1-\phi^2)}\right) \qquad \text{(F.29)}$$

Combining equations (F.5), (F.18). and (F.29):

$$N_{eff} = \frac{\sigma^2}{\sigma_{nm}^2} = \frac{V(z_i)}{V(\bar{z})} = \frac{N}{\left(\dfrac{1+\phi}{1-\phi}\right)\left(1 - \dfrac{2\phi(1-\phi^N)}{N(1-\phi^2)}\right)} \qquad \text{(F.30)}$$

Note that the denominator in equation (F.30) has the term $\left(\dfrac{1+\phi}{1-\phi}\right)$ multiplied by a "correction term" which is usually close to 1.0 and approaches 1.0 as the sample size increases.

F.3 Derivation of the Sample Size Equation

When the variance is known, the sample size for a hypothesis test of the mean is shown in equation (F.4). When the variance, σ^2, is to be estimated from the data, use of the t statistic is recommended, as shown below, where $\hat{\sigma}^2$ is the estimate of σ^2:

$$N = \hat{\sigma}^2 \left\{\frac{t_{df;1-\beta} + t_{df;1-\alpha}}{Cs - \mu_1}\right\}^2 \qquad \text{(F.31)}$$

To use this equation, the recommended procedure is to substitute the normal statistic for the t statistic (e.g., $z_{1-\beta}$ for $t_{N-1;1-\beta}$), calculate a preliminary sample size from which the degrees of freedom can be estimated, and use this to determine t and a new estimate of the sample size. For small sample sizes, a third or fourth estimate of the sample size may be required.

Using equation (F.31) the exact sample size satisfies the following equation:

$$\text{Sample size (t)} = N_t = \hat{\sigma}^2 \left\{\frac{t_{N_t-1;1-\beta} + t_{N_t-1;1-\alpha}}{Cs - \mu_1}\right\}^2 \qquad \text{(F.32)}$$

Using the conditions which satisfy equation (F.32), the calculated sample size using (F.4) would be:

$$\text{Sample size (z)} = N_z = \frac{N_t \left\{ z_{1-\beta} + z_{1-\alpha} \right\}^2}{\left\{ t_{N-1;1-\beta} + t_{N-1;1-\alpha} \right\}^2} \qquad (F.33)$$

The difference between these two sample size estimates where $\alpha = .10$ and $\beta = .10$ is shown in figure F.1.

Figure F.1   Differences in Sample Size Using Equations Based on a Normal Distribution (Known Variance) or a t Statistic, Assuming $\alpha = .10$ and $\beta = .10$
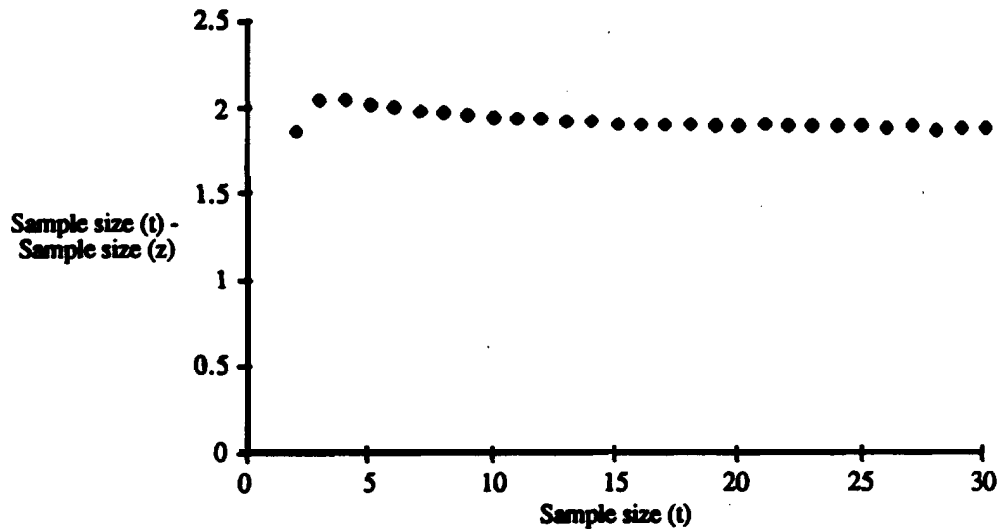
Note that the difference in the sample sizes using equations (F.4) and (F.31) is fairly constant over a wide range of possible sample sizes. This property can be used to estimate the samples size based on equation (F.31) from equation (F.4). Thus:

$$N_t \approx \delta^2 \left\{ \frac{z_{1-\beta} + z_{1-\alpha}}{C_s - \mu_1} \right\}^2 + K \qquad (F.34)$$

where K is a constant which will depend on on α and β. Table F.2 tabulates K at a sample size of 20, for selected values of α and β.

- The equations for sample size in the text use equation (F.34) with K = 2.

Table F.2   Differences between the calculated sample sizes using a t distribution and a normal distribution when the samples size based on the t distribution is 20, for selected values of α (Alpha) and β (Beta)

Beta	Alpha				
	.25	.10	.05	.025	.01
.25	0.8	1.2	1.6	2.1	2.7
.10	1.2	1.4	1.7	2.0	2.6
.05	1.6	1.7	1.9	2.2	2.7
.025	2.1	2.0	2.2	2.5	2.9
.01	2.7	2.6	2.7	2.9	3.2

F.4   Effective Df for the Mean from an AR1 Process

The following formula is appropriate for estimating the variance of the mean of n observations from an AR1 series, assuming a large sample size:

$$s^2_{mean} = \frac{s^2}{n} \frac{(1+ \phi)}{(1- \phi)} \qquad (F.35)$$

if the serial correlation is assumed to be zero then, s^2, the estimated variance of the data, has a scaled chi-square distribution with n-1 degrees of freedom. The mean of a chi-square distribution is v, the degrees of freedom, with a variance of $2v$. Thus, the coefficient of variation squared is $cv^2 = \frac{2v}{v^2} = \frac{2}{v}$.

With zero serial correlation, $\hat{\phi}$ will have a mean of zero and variance of $\frac{1}{n}$ (Box and Jenkins, 1970). The term $\frac{(1+\hat{\phi})}{(1-\hat{\phi})} \approx 1 + 2\hat{\phi}$ (for small ϕ) has a mean of roughly 1 and a variance of approximately $\frac{4}{n}$. The cv^2 is also approximately $\frac{4}{n}$ since the mean ≈ 1.

Assuming a large sample size, the cv of the product of two estimates is equal to the square root of the sums of the squares of the cv's for each term if the terms are independent (which will be true if the serial correlation is zero). Thus, the cv^2 of s^2_{mean} is roughly the sum of two cv^2's:1) the chi-square distribution, and 2) the correction term based on $\hat{\phi}$. Thus the

$$cv^2 \text{ of } s^2_{mean} \approx \frac{2}{(n-1)} + \frac{4}{n} \approx \frac{2}{(n-1)} + \frac{4}{(n-1)} = \frac{6}{(n-1)} \qquad (F.36)$$

Assuming that the distribution of s^2_{mean} is roughly chi-square, then the effective number of degrees of freedom for s^2_{mean} is v' where $\frac{2}{v'} = \frac{6}{(n-1)}$, or $v' = \frac{(n-1)}{3}$.

Simulations appear to be consistent with this result when $\phi = 0$, and suggest that the number of degrees of freedom drops further when $\phi > 0$.

F.4   Sequential Tests for Assessing Attainment

The following paper, prepared by Westat, has been included in this appendix as it was submitted for publication.

Assessing Attainment of Ground Water Cleanup Standards Using Modified Sequential t-Tests

By John Rogers, Westat, Rockville Maryland[1]

Assessing the attainment of Superfund cleanup standards in ground water can be complex due to measurements with skewed distributions, seasonal or periodic patterns, high variability, serial correlations, and censoring of observations below the laboratory detection limit. The attainment decision is further complicated by trends and transient changes in the concentrations as a result of the cleanup effort. EPA contracted Westat to prepare a guidance document recommending statistical procedures for assessing the attainment of ground water cleanup standards. The recommended statistical procedures were to require a minimum of statistical training. The recommended procedures included a sequential t-test based on yearly average concentrations.

Further research and simulations by Westat indicate that modifications of the sequential t-test have better performance and are easier to use than the originally proposed sequential t-test, particularly with highly skewed data. This paper presents three modified sequential tests with simulation results showing how the sequential t-test and the modifications perform under a variety of situations similar to those found in the field. The modified tests use an easy-tocalculate approximation for the log likelihood ratio and an adjustment to improve the power of the test for small sample sires. Using the log transformed yearly averages improves the test performance with skewed data. Expected sample sizes and practical considerations for application of these tests are also discussed.

Key words: Sequential t-test, Simulations, Ground water, Superfund.

1. Introduction

EPA contracted Westat to prepare a draft guidance document recommending sampling and statistical methods for evaluating the attainment of ground-water cleanup standards at Superfund sites. The recommended statistical methods were to be applicable to a variety of site conditions and be able to be implemented by technical staff with a minimum of statistical training.

The draft document included an introduction to basic statistical procedures and recommended a variety of statistical methods including a sequential t-test. Although the sequential t-test has several advantages for testing ground water, one significant disadvantage is the relative complexity of the calculations, requiring use of the noncentral t distribution. Additional research was undertaken by Westat to find an alternative to the standard sequential t-test which is easier to implement. As part of this research, simulations have been used to evaluate the performance of the sequential t-test and several modifications of it.

This paper presents these simulation results showing how the sequential t-test and the modified tests perform under a variety of situations similar to those found in the field.

The Problem of Assessing Ground Water at a Superfund Site

The history of contamination and cleanup at a Superfund site will result in ground water contaminant concentrations which generally (1) increase during periods of contamination, (2)

---

decrease. during remediation, and (3) settle into dynamic equilibrium with the surrounding environment after remediation, at which point the success of the remediation can be determined.

Specifying the attainment objectives and assessing attainment of cleanup standards can be complicated by many site specific factors, including: multiple wells, multiple contaminants, and data which have seasonal patterns, serial correlations, significant lab measurement variation, non-constant variance, skewed distributions, long-term trends, and censored values below the detection limits. The general characteristics of ground water quality data have been discussed by Loftis et al. (1986). All of these factors complicate the specification of an appropriate statistical test. Figure 1 illustrates the variation which might be found in monthly ground water measurements, using simulated **observations.**

The Statistical Problem to be Discussed

The following statistical problem is addressed in this paper. Suppose remediation is complete and any transient effects of the remediation on the ground water levels and flows have dissipated. We then wish to determine if the mean concentration of a contaminant, μ, is less than the relevant cleanup standard, μ_0. The ground water will be judged to attain the cleanup standard if the null hypothesis, $H_0: \mu \geq \mu_0$, can be rejected based on a statistical test. The power of the test, the probability of rejecting the null hypothesis, is to be α when $\mu = \mu_0$. For a specified alternate hypothesis, $H_1: \mu = \mu_1$ $(0 < \mu_1 < \mu_0)$ the power is to be $1-\beta$, where β is the probability of a false negative decision (the probability of incorrectly accepting the null hypothesis).

The statistical tests considered in this paper are the sequential t-test for comparing means and modifications of this test. Using a sequential procedure, a test of hypothesis is performed after each sample, or set of samples, is collected. The test of hypothesis results in three possible outcomes, (1) accept the null hypotheisis, (2) reject the null hypotheisis, or (3) continue sampling. The hypothesis is tested based on the n ground water samples, x_l to x_n, collected prior to the test of hypothesis. The sample size at the termination of the test is a random variable. The power and sample size distribution of the sequential tests were evaluated using monte carlo simulations. For the simulations the following parameters were varied: the mean, standard deviation, detection limit, proportion of the variation which is serially cotrelated versus independent, lag 1 serial correlation, alpha and beta, distribution (normal or lognormal), and μ_1. For all simulations μ_0 is set at 1.0. 1000 simulations were made for each set of parameters *tested,* unless otherwise noted. Simulations were performed using SAS version 6.

Section 2 reviews and compares the fixed sample size and sequential t-tests. Sections 3 and 4 discuss the performance of the t-test and several modifications when applied to normally distributed and independent observations. The performance of the sequential tests when applied to simulated ground water data is evaluated in Section 5. Section 6 discusses the results and presents the conclusions.

2. Fixed Versus Sequential Tests

The fixed sample size test and sequential t-test are reviewed briefly below, emphasizing factors which are relevant to the development of a modified test and for selecting a test for assessing ground water.

Fixed Sample Size t-Test

The fixed sample size t-test, familiar to many users of statistics, requires the following steps:

(1) Estimate the variance of the future measurements, δ^2, based available data;

(2) Determine sample size n, such that,

$$n = \left(\frac{(t_{\alpha,n-1} + t_{\beta,n-1})\,\delta}{\mu_1 - \mu_0}\right)^2 ; \qquad \text{eq. (1)}$$

where $t_{\alpha,n-1}$ is the α percentile of the t distribution with n-1 degrees of freedom.

(3) Collect n samples and measure the contaminant concentrations;

(4) Calculate the test statistic t, with n-1 degrees of freedom,

$$\text{where}: t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}}, \quad \bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} \text{ and } s_{\bar{x}} = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n\,(n-1)}}; \text{ and}$$

(5) Conclude that the ground water attains the cleanup standard if $t < t_{\alpha,n-1}$ otherwise, accept the null hypothesis that the ground water does not attain the cleanup standard.

The t-test does well to preserve the power of the test at the null hypothesis when the data have a roughly normal distribution. However ,the power at the alternate hypothesis depends on the the accuracy of the initial variance estimate, δ^2. Thus the fixed sample size test fixes α and n, leaving β variable.

Standard Sequential t-Test

With normally distributed independent observations and known σ^2, an optimal sequential test is the sequential probability ratio test (SPRT) (Wald 1947). When σ^2 is unknown, as here, one approach is provided by the sequential t-test which states the null hypothesis in terms of the unknown standard deviation (Rushton 1950, Ghosh 1970, and others). For testing hypotheses about means, an alternative heuristic solution replaces the unknown variance by the sample estimate at each step in the sequential test (Hall 1962, Hayre 1983). This second version of the sequential t-test can be used to compare the mean to an established cleanup standard. Liebetrau (1979) discussed the application of this test to water quality sampling.

The steps in implementing the sequential t-test for comparing the mean to a standard are:

(1) Collect k-1 samples without, testing the hypothesis.

(2) Collect one additional sample for a total of n samples collected so far and calculate:

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}}, \quad \delta_0 = 0, \quad \delta_1 = \frac{\mu_1 - \mu_0}{s_{\bar{x}}}; \qquad \text{eq.(2)}$$

(3) Calculate the likelihood ratio:

$$L = \frac{f_{n-1}(t \mid \delta = \delta_1)}{f_{n-1}(t \mid \delta = \delta_0)} \qquad \text{(eq. 3)}$$

where $f_{n-1}(t \mid \delta)$ is the density of the noncentral t distribution with n-1 degrees of freedom, and noncentrality parameter δ.

(4)   If $L > \dfrac{1-\beta}{\alpha}$ then reject the null hypothesis and conclude that the ground water attains the cleanup standard

if $L < \dfrac{\beta}{1-a}$ then accept the null hypothesis that the groundwater does not attain the cleanup

Standard,

otherwise, return to step (2) and collect additional samples until a decision is reached.

Unlike the fixed sample size test, for the sequential test, α and β are fixed and n is variable.

Comparison of the Sequential and Fixed Sample Size Test

Table 1 compares the sequential and fixed sample size tests based on several characteristics. The choice of which test to use depends on the circumstances in which the test is to be applied.

Table 1   Comparison of the fixed sample size and sequential t-test.

Characteristic	Sequential t-Test	Fixed Sample Size t-Test
Power	Fixed at the null and alternate hypothesis	Fixed at the null hypothesis. Power at the alternate hypothesis depends on the estimate of measurement variance used for calculating sample size.
Sample Size	Subject to variation, often less than for a fixed sample size test with the same power	Fixed
Sampling	Works well if the time between collection of samples is long relative to the analysis time.	Works well if the sample collection period is short relative to the analysis time.
Estimate of the mean	Biased	Unbiased
Ease of Calculation	Standard test requires tables of the non-central t distribution which are not generally available. Modified test reported here can be easily calculated.	Uses widely available tables

Application of the Sequential Test to Ground Water Data

For testing contaminant concentrations against a cleanup standard, the sequential t-test has some distinct advantages: (1) ground water sample collection is sequential with sample analysis time often short compared to the sample collection period, (2) a good estimate of measurement variance for calculating the sample size for the fixed test may not be available, (3) for assessing attainment, the objective is to test a hypothesis rather than to obtain an unbiased estimate of the mean or construct a confidence interval, (4) reducing sample size can be important when the cost of laboratory sample analysis is high, and (5) if the concentrations at the site are indeed below the cleanup standard, maintaining the power at the alternate hypothesis can protect against incorrectly concluding that additional costly cleanup is required. For many users, the main disadvantage of using the standard sequential t-test is the relative complexity of the calculations.

3. Power and Sample Sizes for the Sequential t-Test with Normally Distributed Data

For the purpose of describing the simulation results used to determine the power of the sequential t-test, define the **scale factor** as the ratio of the standard deviation of the measurements to the difference between the means for the null and alternate hypotheses:

$$\text{Scale factor} = \frac{\sigma}{\mu_0 - \mu_1}.$$

Also let n_{fixed} designate the sample size for a fixed sample size test with the same nominal power as the sequential test being discussed, where n_{fixed} is calculated using the known variance, σ^2, set in the simulation. For the fixed sample size test, the scale factor is proportional to the square root of the sample size, n_{fixed}. As will be shown later, the scale factor is also roughly proportional to the square root of the average sample size for the sequential test.

Although the power of the sequential t-test approaches the nominal levels (α at μ_0 and $1-\beta$ at μ_1) for large sample sizes, the power curve at small sample sizes depends on α, β, and the scale factor. Figure 2 shows the power and sample size of the sequential t-test using normally distributed data with $\alpha = 0.05$, $\beta = 0.05$, and a scale factor of 1.6. Also shown are the nominal power at the null and alternate hypothesis and the sample size for the equivalent fixed sample size test, n_{fixed}. The power at the null hypothesis is close to the nominal level of α. At the alternate hypothesis, the power is significantly lower than the nominal level of $1-\beta$. The average sample size reaches a maximum when the true mean is mid-way between the null and alternate hypotheses.

Table 2 presents the false positive and false negative rates for the sequential t-test for values of (α,β) of $(0.1, 0.1)$, $(0.01, 0.1)$ and $(0.01, 0.01)$, μ at the null and alternate hypothesis, and the scale factor $= 0.4$ and 3.0. The false positive rates are less than (i.e. conservative) or similar to the nominal levels. However, the false negative rates are significantly higher than the nominal level.

Table 2   Simulated power of the sequential t-test

Scale factor $\frac{\sigma}{\mu_0 - \mu_1}$	Corresponding Fixed sample size	False positive rate, α		False negative rate, β	
		Nominal	Simulated	Nominal	Simulated
0.4	3	0.10	.005	0.10	.188
0.4	5	0.01	.000	0.10	.265
0.4	7	0.01	.000	0.01	.249
3.0	61	0.10	.113	0.10	.232
3.0	120	0.01	.016	0.10	.201
3.0	196	0.01	.011	0.01	.103

Several modifications, which are discussed below, were considered to improve the power of the test. Note that equations (2) above can be rewritten as:

$$t = \frac{\bar{x} - h_0}{s_{\bar{x}}}, \quad \delta_0 = \frac{\mu_0 - h_0}{s_{\bar{x}}}, \quad \delta_1 = \frac{\mu_1 - h_0}{s_{\bar{x}}}; \qquad \text{eq. (4)}$$

where $h_0 = \mu_0$. For the sequential t-test, the nominal probability of accepting the hypothesis H: $\mu = \mu_1$ is the same for both of the following tests:

H_0: $\mu = \mu_0$ against H_1: $\mu = \mu_1$, power at $\mu_0 = \alpha$, $\mu_1 = 1-\beta$ (i.e. $h_0 = \mu_0$); and

H_0: $\mu = \mu_1$ against H_1: $\mu = \mu_0$, power at $\mu_0 = \alpha$, $\mu_1 = 1-\beta$ (i.e. $h_0 = \mu_1$).

Based on this symmetry, the nominal power of the sequential t-test is the same whether $h_0 = \mu_0$ or $h_0 = \mu_1$. In practice, h_0 serves as the **zero point** around which the parameters for the non-central t distribution are calculated rather than the mean value at which the power is maintained, as in the fixed sample size test. If the equations for the sequential test are modified to put the zero point mid-way between μ_0 and μ_1, then (1) $\delta_1 = -\delta_0$, (2) only one non-central t distribution needs to be evaluated, and (3) the power of the test is symmetric around h_0 when $\alpha = \beta$, i.e. the false positive and false negative rates are equal. Although Rushton (1950) considered null hypotheses other than zero and $h_0 = \mu_0$, in this paper h_0 is called the zero point rather than the null hypothesis. To avoid confusion, the terms null and alternate hypothesis will be used as defined in Section 1, reflecting the intentions of those performing the test.

Define the **centered sequential t-test** by replacing equations (2) by equations (4) and setting the zero point for the calculations mid-way between μ_0 and μ_1, i.e.:

$$h_0 = \frac{\mu_0 + \mu_1}{2}.$$ eq. (5)

This centered test is used in the following simulations to determine the relationship between power and sample size.

Changes in Power with Increasing Sample Size

Figure 3 shows the false decision rate (false positive or false negative rate) and average sample size for the centered sequential t-test with α and β set at .05, and the scale factor ranging from 0.4 to 3.6. For this symmetric test, the false positive and false negative rates are equal. The false decision rate at very low sample sizes is smaller than the nominal level of .05. As the scale factor increases, resulting in increasing sample sizes, the false decision rate increases to a maximum of roughly three times the nominal level and then decreases slowly. The average sample size is roughly half of that for the corresponding fixed sample size test except at very low sample sizes. Similar patterns were seen in the false negative rates when the zero point was set at the null hypothesis.

The good performance of the test at low samples sizes is in part due to the discrete nature of the sampling. From the sample just before the termination of the test to the sample which terminates the test, the likelihood ratio jumps from inside the decision limits to outside. With small sample sizes, the likelihood ratio may be considerably beyond the decision limits on the last sample. This is equivalent to having more information than is necessary to make the decision, resulting in improved performance.

Distribution of Sample Sizes

Simulations were used to look at the distribution of sample sizes at the termination of the test, for selected values of μ and scale factors of 1.0 and 3.0. Figure 4 shows the distribution of sample sizes, using a log scale, when $\mu = \mu_1$ and the scale factor equals 1.0. The sample sizes are displayed separately for simulations which rejected the null hypothesis (correct decision) and those which did not. For both decisions a relatively large proportion of the simulations terminate at a sample size of two. The false decision rate is greater than the nominal value by roughly the proportion of simulations terminating with only two samples. The modified sequential test, for which the distribution of samples sizes is also shown in Figure 4, is discussed in the next section.

The general characteristics of the sample size distributions are the same regardless of the conditions simulated. Samples sizes for the sequential t-test are highly skewed. For many simulations, the test terminated with two samples. For those simulations not terminating with two or three samples, the distribution of sample sizes was roughly log-normal.

4. Modifications to Simplify the Calculations and Improve the Power

The poor performance of the centered sequential t-test at the alternate and null hypotheses and the observation that many of the simulations which terminate at two samples contribute to the large false decision rates, suggest that a modification to the test might improve the performance. Other authors have noted this problem and suggested alternate procedures. In particular, Hayre (1983) suggested changing the test boundaries. Hayre's suggestion is equivalent to multiplying the the log likelihood ratio by the adjustment factor $(nd)/(n+c)$ where $d < k$ and $c \geq -d$. Based on heuristic arguments, Hayre concluded that k, the minimum number of samples, should be at least 5 if a large sample size is expected.

When small sample sizes are expected, requiring as many as 5 samples before the first test of hypothesis can result in an overly conservative test. In this research decision rules requiring a minimum of 2, 3, or 4 samples were considered. In addition, the performance of the centered sequential t-test was simulated using adjustment factors of: 1, $(n-1)/n$, $(n-2)/n$, $(n-3)/n$. The simulations used α and β set at 0.10, 0.05, and 0.01.

The false decision rates for the four adjustment factors, with $(\alpha,\beta) = (0.05,0.05)$, are shown in Figure 5. All of the adjustment factors improved the performance of the test by reducing the maximum probability of a false decision to values closer to the nominal value. The selection of an optimal adjustment factor requires specification of the conditions under which the test is to be used. One adjustment factor might be chosen if small sample sizes are expected, another if large sample sizes are expected. In all cases, the test is conservative for low sample sizes, possibly liberal for intermediate sample sizes, and approaches the nominal values for large sample sizes. Over the range of the scale factor considered in the simulations, the average false decision rate for the adjustment factor $(n-2)/n$ was closest to the nominal value. Therefore, this adjustment factor, $(n-2)/n$, with k=3 was chosen for evaluation in subsequent simulations.

Approximation for Non-central t

Calculation of the likelihood ratio using the noncentral t-distribution is difficult because the tables are not generally available and are difficult to use. The use of the sequential t-test can therefore be simplified by using an approximation to the log likelihood ratio of the two non-central t-distributions. Rushton (1950) published three approximations for the log of the likelihood ratio. Westat's analysis showed that the approximations performed well, particularly when the zero point for the test was set mid-way between the null and alternate hypotheses. Using Rushton's simplest approximation and the adjustment factor selected above, the equations for the **modified sequential t-test** become:

$$h_0 = \frac{\mu_0 + \mu_1}{2}, \quad t = \frac{\bar{x} - h_0}{s_{\bar{x}}}, \quad \delta = \frac{\mu_1 - \mu_0}{s_{\bar{x}}} \quad \text{and} \qquad \text{eq. (5)}$$

$$L = \exp\left(t \, \delta \, \frac{n-2}{n} \sqrt{\frac{n}{n-1+t^2}} \right). \qquad \text{eq. (6)}$$

Figure 4 shows the distribution of samples sizes for the modified test compared to that for the standard sequential t-test. Figure 6 shows the power curve and average sample sizes for the modified test with $\alpha=\beta$ and scale factor = 1.6. Figure 6 can be compared directly with Figure 2 for the standard sequential t-test.

Termination of the Test Before a Decision Has Been Reached

Figure 7 shows the distribution of sample sizes for selected values of μ, the mean of the simulated measurements, using the modified test with scale factor of 1.6. As noted before, the distribution of the sample sizes is roughly log-normal. The minimum sample size is 3 because a minimum of three samples are required before the first test of hypothesis. The mean sample size is generally similar to or less than n_{fixed}. The 95th percentile of the sample sizes is less than three times n_{fixed} and, for values of μ close to the null and alternate hypothesis, is generally similar to or less than N_{fixed}.

Several authors, including Wald, have suggested that, for practical purposes, the sequential test can be terminated after some fixed large number of samples if the test has not otherwise terminated, with the decision going to which ever hypothesis is more favored at termination. Figure 7 suggests that a decision rule terminating the test with a maximum sample size of three times n_{fixed} is reasonable because very few tests would be terminated early when the true mean is close to the null or alternate hypothesis. When the mean is mid-way between the null and alternate hypothesis, acceptance of the null hypothesis is essentially random, and early termination will not affect the power of the test.

Simulations were performed to evaluate different termination rules. One hundred simulations were run for all combinations of: termination at 1, 2, 3, 4, and 5 times n_{fixed}; four scale factors from .4 to 3.6; $\alpha = \beta = 0.1, 0.05, 0.01$; and $\mu_1 = 0.5$. In addition, 100 simulations were run for all combinations of: 11 values of μ from .35 to 1.15; termination at 1, 2, 3, and 4 times the fixed sample size; scale factor = 1.6; and $\alpha = \beta = 0.05$. The differences in the power due to early termination were not statistically significant. Early termination resulted in a decrease in the average sample size with μ mid-way between the null and alternate hypotheses; however, with μ at the null or alternate hypothesis, changes in the average sample size were, practically speaking, insignificant.

These results indicate that early termination of the sequential test will have little effect on the power of the test. Because the fixed sample size is estimated from δ^2 based on data available before sampling and is therefore subject to error, it is recommend that sequential tests not be terminated until the samples size is at least twice the estimated sample size for an equivalent fixed sample size test. For the simulations reported in other sections of this paper, the sequential tests were terminated if the sample size exceeded 5 times n_{fixed}.

5. Application to Ground Water Data from Superfund Sites

The modified sequential t-test performs well with normally distributed data, having average sample sizes below those for equivalent fixed sample size tests and power close to the nominal power. However, ground water measurements may be skewed, serially correlated, censored, and have seasonal patterns. How well does the modified test perform with ground water data? Simulations were used to determine how four sequential tests performed when assessing ground-water data.

For all statistical tests, the following sequential sample design is assumed: m ground water samples are collected at periodic intervals throughout the year, with at least 4 samples per year. The samples are analyzed and the test of hypothesis is performed once per year starting after three years of data are collected. The number of years of data collection is n.

The four statistical tests evaluated using the simulations are:

1)   Standard sequential t-test described in section 2 using the yearly averages;

2)   **Modified sequential t-test using the yearly averages;**

3)   **Modified sequential t-test with adjustments for seasonal variation and serial correlation:**

   Remove seasonal patterns from the data using one-way analysis of variance. Calculate the standard error, s_{ε}, and the lag1 serial correlation of the residuals, r. Estimate the standard error of the mean as:

$$s_{\overline{x}} = \sqrt{s_{\varepsilon}^2 \, \frac{1+r}{1-r}} \ \ \text{with Df} = \frac{m(n-1)}{3}.$$

   The effective sample size is assumed to be one more than the number of degrees of freedom. Therefore:

$$L = \exp\left(t \ \delta \ \frac{Df\text{-}1}{Df\text{+}1} \sqrt{\frac{Df\text{+}1}{Df\text{+}t^2}} \ \right).$$

4)   Modified sequential t-test with an adjustment for skewness:

   Calculate y = ln(yearly average). Estimate the log transformed mean and its standard error using the following equations:

$$\ln(\overline{x}) = \overline{y} + \frac{s_y^2}{2} \ \text{and} \ \ s_{\ln(\overline{x})} = \sqrt{\frac{s_y^2}{n} + \frac{s_y^4}{2\,Df}}, \ \text{where} \ s_y^2 = \sum_{i=1}^{n} \frac{(y_i - \overline{y})^2}{n-1} \ \text{and} \ \overline{y} = \sum_{i=1}^{n} \frac{y_i}{n}.$$

   The test statistic for the sequential t-test uses:

$$h_0 = \frac{\ln(\mu_0) + \ln(\mu_1)}{2}, \ t = \frac{\ln(\overline{x}) - h_0}{s_{\ln(\overline{x})}}, \text{and} \ \ \delta = \frac{\ln(\mu_1) - \ln(\mu_0)}{s_{\ln(\overline{x})}}$$

The first, second and fourth tests use the yearly average concentrations, averaging across the within year seasonal patterns. The serial correlation between the yearly averages is less than between individual observations, reducing the influence of correlation on the test results. The third test removes the seasonal patterns. The standard error of the mean is adjusted by a factor which accounts for the serial correlation, assuming an AR(1) model and many observations per year. Although this assumption may not be correct, the lag 1 correlation is expected to dominate the correlations for higher lags, making the AR(1) model a reasonable approximation to the data. The effective degrees of freedom for the standard error is based on asymptotic approximations. The fourth test is based on the assumption that the yearly averages have a log normal distribution. For highly skewed data this assumption is more reasonable than assuming a normal distribution. The mean and standard error of the mean are first order approximations based on a lognormal distribution.

The second test was expected to perform well with data which has an approximately normal distribution. The third test was expected to perform best with highly skewed data. The fourth test was expected to perform best with data with significant correlation and little skewness. Simulations were performed to test these assumptions.

Figure 8   Range of False Positive Rates for Scale Factors from
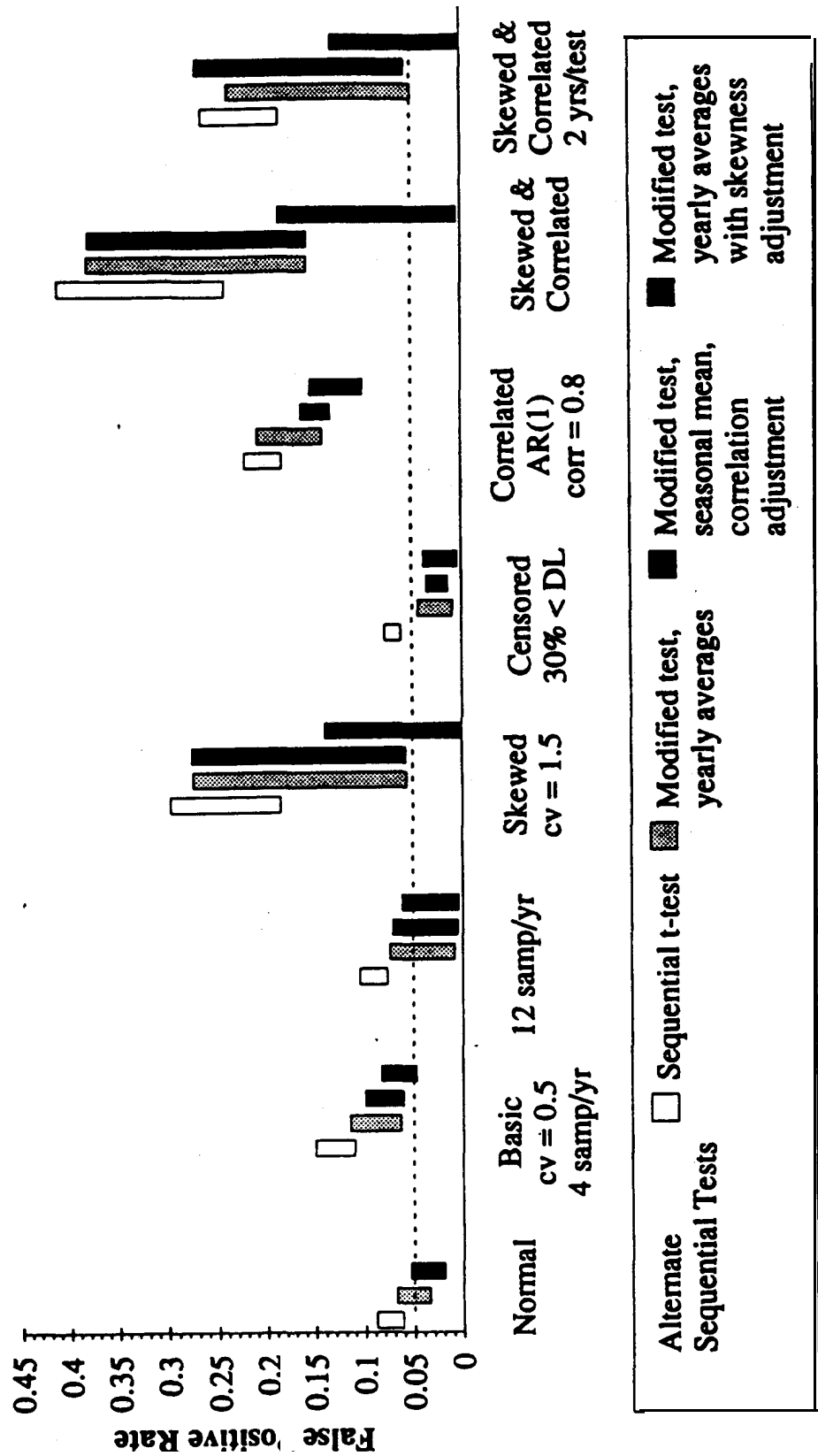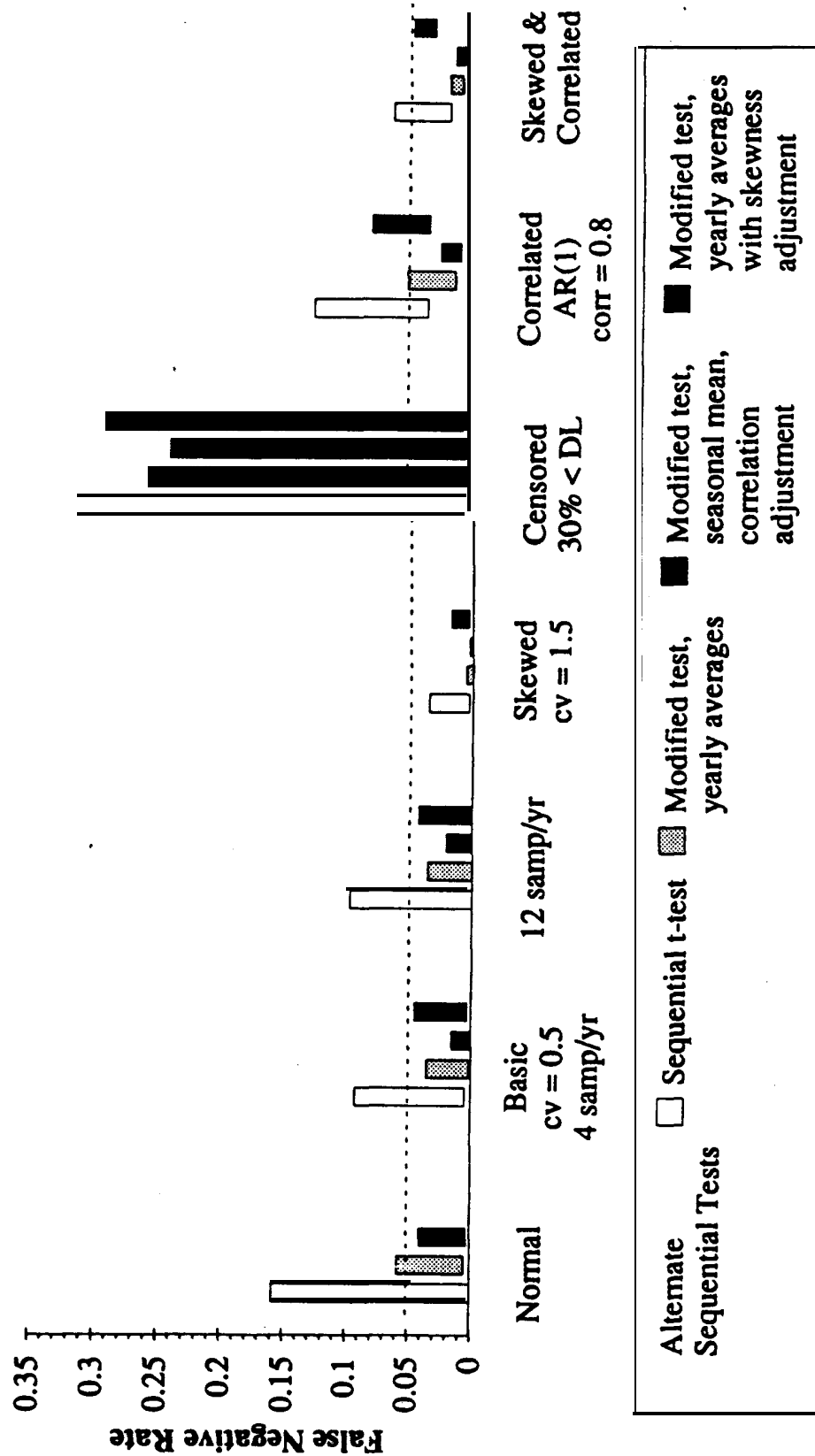1.6 to 4.8 for Four Sequential Tests, by Data Type

Figure 9 Range of False Negative Rates for Scale Factors from 1.6 to 4.8 for Four Sequential Tests, by Data Type

APPENDIX G: GLOSSARY

Alpha (a) -In the context of a statistical test, a is probability of a Type I **error.**

Alternative Hypothesis See *hypothesis.*

Analysis Plan The plan that specifics how the data are to be analyzed once they have been collected, includes what estimates are to be made **from the** data, how the estimates are to be calculated, and how the results of the analysis will be r e p o r t e d .

Autocorrelation See *serial correlation*

Attainment This term by itself refers to the successful achievement of the attainment objectives. In brief, attainment means that site contamination has been reduced to or below the level of the cleanup standard.

Attainment Objectives The attainment objectives refer to a set of site descriptors and parameters together with standards as to what the desired level should be for the parameters. These are usually decided upon by the courts and the responsible parties. For example, these objectives usually include the chemicals to be tested, the cleanup standards to be attained, the measures or parameters to be compared to the cleanup standard, and the level of confidence required if the environment and human health are to be protected (Chapter 3).

Beta (β) In the context of a statistical test, β is the probability of a Type II error.

Binomial Distribution A probability distribution used to describe the number of occurrences of a specified event in n independent trials. In this manual, the binomial distribution is used to develop statistical tests concerned with testing the proportion of ground water samples that have excessive concentrations of a contaminant (see Chapters 8 and 9). For example, suppose the parameter of interest is the portion (or percent) of the ground water wells that exceed a level specified by the cleanup standard, Cs. Then one might estimate that portion by taking a sample of 10 wells and counting the number of wells that exceed the Cs. Such a sampling process results in a binomial distribution. For additional details about the binomial distribution, consult Conover (1980).

Central Limit Theorem If X has a distribution with the mean μ and variance σ^2, then the sample mean $\bar{X}$, based on a random sample of size n has an approximately normal distribution with mean μ and variance $\frac{\sigma^2}{n}$. The approximation becomes increasingly good as n increases. In other words, no matter what the original distribution of X (so long as it has a finite mean and variance), the distribution of $\bar{X}$ from a large sample can be approximated by a normal distribution. This fact is very important since knowing the approximate distribution of X allows us to make corresponding approximate probabilistic estimates. For example, reasonably good estimates for confidence intervals on X can frequently be given even though the underlying probabilistic structure of Y is unknown.

Chain of Custody Procedures Procedures for documenting who has **custody** of and the condition of samples from the point of collection to the analysis at the laboratory. Chain of custody procedures are used to insure that the samples are not lost, tampered with, or improperly stored or handled.

Clean Attains the cleanup standard. That is, a judgment has been made that the site has been cleaned or processed to the point that in the attainment objectives, as defined above, have been met.

Cleanup Standard (Cs) The criterion set by EPA against which the measured concentrations are compared to determine whether the ground water at the Super-fund site is acceptable or not (Sections 2.2.4 and 3.4). For example, the Cs might be set at 5 parts per million (5 ppm) for a site chemical. Hence, any water that tests out at greater than 5 ppm is not acceptable.,

Coefficient of Determination (R^2) A descriptive statistic, $R^2 = 1 - \frac{SSE}{S_{yy}}$ and $0 \leq R^2 \leq 1$, that provides a rough measure of the overall fit of the model. A perfect fit; i.e., all of the observed data points fall on the fitted regression line, would be indicated by an R^2 equal to 1. Low values of R^2 can indicate either a relatively poor fit of the model or no relationship between the concentration levels and time. R^2 is just the square of the well-known correlation coefficient. For more information, see any standard text book.

Coefficient of Variation (cv) The ratio of the standard deviation to the mean ($\frac{\sigma}{\mu}$) for a set of data or distribution. For data which can only have positive values, such as concentration measurements, the coefficient of variation provides a crude measure of skewness. Data with larger cv's usually are more skewed to the right. The cv provides a relative measure of variation (i.e., relative with respect to the mean). As such, it can be used as a rough measure of precision. It is useful to know if the cv is relatively constant over the range of the variable of interest.

Comparison-wise Alpha 'For an individual statistical decision on one compound or well, the maximum probability of a false positive decision.

Compositing Physically mixing several samples into one larger sample, called a composite sample. Then either the entire composite is measured or one or more random subsamples from the composite are measured Generally the individual samples which are composited must be the same size or volume, and the composite sample must be completely mixed. Composite samples can be useful for estimating the mean concentration. If appropriate, compositing can result in substantial savings where the cost of analyzing individual samples is high.

Confidence Interval A sample-based estimate of a population parameter which is expressal as a range or interval of values which will include the true parameter value with a known probability or confidence. For example, instead of giving an estimate of the population mean, say x = 15.3, we can give a 95 percent confidence interval, say [x-3, x+3] or [12.3 to 18.3] that we are 95 percent confident contains the population mean.

Confidence Level The degree of confidence associated with an interval estimate. For example, with a 95 percent confidence interval, we would be 95 percent certain that the interval contains the true value being estimated. By this, we mean that 95 percent of independent 95 percent confidence intervals will contain the population mean. In the context of a statistical test, the confidence level is equal to 1 minus the Type I error (false positive rate). In this case, the confidence level represents the probability of correctly concluding that the null hypothesis is true.

Conservative Test A statistical test for which the Type I error rate (false positive rate) is actually less than that specified for the test. For a conservative test there will be a greater tendency to accept the null hypothesis when it is not true than for a non-conservative test. In the context of this volume, a conservative test errs on the side of protecting the public health. That is to say, the mistake (i.e.. error) of wrongly deciding that the site is clean will be less than the stated Type I Error **Rate.**

Contaminated A site is called contaminated if it does not attain the cleanup standards. In other words, the contamination level on the site is higher than that allowed by the cleanup standard.

Degrees of Freedom (Df) The degrees of freedom of an estimate of variance, standard deviation, or standard error is a measure of the amount of information on which the estimate is based or the precision of the estimate. Usually, high degrees of freedom are associated with a large sample size and a corresponding increase in accuracy of an estimation.

Dependent Variable (y_i) An outcome whose variation is explained by the influence of independent variables. For example, the contamination level in ground water (i.e., the dependent variable y) may depend on the distance (i.e., the independent variable x) from the site incinerator.

Detection Limit The level below which concentration measurements cannot be reliably determined (see Section 2.3.7). Technically, the lowest concentration of a specified contaminant which is unlikely to be obtained when analyzing a sample with none of the contaminant

Distribution The frequencies (either relative or absolute) with which measurements in a data set fall within specified classes. **A** graphical display of a distribution is referred to as a histogram. Formally, a distribution is defined in terms of the underlying probability function. For example, the distribution of x, say Fx(t), may be defined as the probability that x is less than t (i.e., $P(x<t)$). For the purposes of this volume, the frequency interpretation is adequate.

Durbin-Watson Test This a test for serial correlation (specifically it is a test for first-order autoregression). If the Durbin-Watson test statistic, $\hat{\phi}_{obs}$, given in the

test is "statistically" large then the decision rule is to declare that we do not believe that serial correlation is present. If $\hat{\phi}_{obs}$ is "statistically" small, then the decision rule is to declare we believe the serial correlation is present.

Estimate Any numerical quantity computed from a sample of data. For example, a sample mean is an estimate of the corresponding population mean.

Estimated Regression Line The fitted curve which estimates the linear regression model. The regression is simple if there is only one independent variable and it is represented by $\hat{y}_i = b_0 + b_1 x_i$.

Experiment-wise Alpha See *overall alpha*

Explanatory Variable See *independent variable.*

False Positive Rate The probability of mistakenly concluding that the ground water is clean when it is contaminated. It is the probability of making a Type I error.

False Negative Rate The probability of mistakenly concluding that the ground water is contaminated when it is clean. It is the probability of making a Type II error.

Ground Water Sample See *physical sample.*

Histogram A graphical display of a frequency distribution. This is usually given by a collection of bars in multiple intervals, where the height of a bar in its interval is proportional to the frequency of occurrence of the variable in that interval.

Hypothesis An assumption about a property or characteristic of a population under study. A major theme of statistical inference is to decide which of two complementary hypotheses is likely to be true. In the context of this document, the *null* hypothesis is the hypothesis that the ground water is "contaminated," and the *alternative* hypothesis is the hypothesis that the ground water is "clean."

Hypothesis Test A basic statistical technique for deciding which of two hypotheses is to be accepted, based on measurements which have measurement error. The null hypothesis is rejected in favor of the alternate hypothesis if the measurements are improbable when the null hypothesis is we. Otherwise, the null hypothesis is accepted in favor of the alternate hypothesis.

Independent Variable (x_i) The characteristic being observed or measured that is hypothesized to influence an event (the dependent variable) within the defined area of relationships under study. The independent variable is not influenced by the event but may cause it or contribute to its variation.

Inference The process of generalizing (extrapolating) results from a sample to a larger population. **More** generally, statistical inference is the art of evaluating information (such as samples) in order to draw reliable conclusions about the phenomena under study. This usually means drawing conclusions about the distribution of some variable.

Interquartile Range The difference between the 75th and 25th percentiles of the distribution.

Judgment Sample A sample of data selected according to non-probabilistic methods; usually based on expert judgment.

Kriging Kriging is the name given to the least squares prediction of spatial processes. It is a form of curve fitting using a variety of techniques from regression and time series. Statistically, kriging is best linear unbiased estimation using generalized least squares. This statistical technique can be used to model the contours of water and contaminant levels across wells at given points in time (see Chapter 7 of this guidance and Volume I, Chapter 10). Kriging is not appropriate for assessing attainment in ground water.

Laboratory Error See *measurement error.*

Lag 1 Serial Correlation See *serial correlation.*

Least Squares **Estimates** This is a common estimation technique. In regression, the purpose is to find estimates for the regression curve fit. The estimates are chosen so that the regression curve is "close" to the plotted sample data in the sense that the square of their distances is minimized (i.e., the least). For example, the estimates β_0 and β_1 of the y-intercept β_0 and the slope β_1 are least square estimates (see Section 6.1.2).

Less-than-Detection Limit A concentration value that is reported to be below the detection limit with now measured concentration provided by the lab. It is

generally recommended that these values be included in the analysis as values at the detection limit.

Lognormal Distribution A family of positive-valued, skewed distributions commonly used in environmental work. See Gilbert (1987) for a detailed discussion of lognormal distributions.

Mean **The arithmetic** average **of** a set of data values. Specifically, the mean of a data set, $x_1, x_2, ..., x_n,$ is defined by $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}.$

Mean Square Error (MSE) The sum of squares due to error divided by the appropriate degrees of freedom which provides an estimate of the variance about the regression.

Measurement Error Error or variation in laboratory measurements resulting from unknown factors in the handling and laboratory analysis procedures.

Median The values which separates the lowest 50 percent of the observations from the upper 50 percent of the observations. Equivalently, the "middle" value of a set of data, after the values have been arranged in ascending order. If the number of data points is even, the median is defined to be the average of the two middle values.,

Mode The value with the greatest probability, i.e., the value which occurs more often than any other.

Model A mathematical description of the process or phenomenon by which the data arc generated and collected.

Non-Central t-Distribution Similar to the t-distribution with the exception that the numerator is a normal variate with mean equal to something other than zero (see *also t-distribution).*

Nonparametric Test A test based on relatively few assumptions about the underlying process generating the data. In particular, no assumptions arc made about the exact form of the underlying probability distribution. **As** a consequence, nonparametric tests are valid for a fairly broad class of distributions.

Normal Distribution A family of "bell-shaped" distributions described by the mean and variance, μ and σ^2. Refer to a statistical text (e.g., Sokal and Rohlf, 1973) for a formal definition. The standard normal distribution has $\mu = 0$ and $\sigma^2 = 1$.

Normal Probability Plot A plot of the ordered residuals against their expected values under normality (see Section 5.6.2).

Normality See *normal distribution* (see also Section 5.6).

Null Hypothesis See *hypothesis.*

Outlier Measurements that are (1) very large or small relativeto the rest of the data, or (2) suspected of being unrepresentative of the true concentration at the sample location.

Overall Alpha When multiple chemicals or wells are being assessed, the probability that all chemicals in all wells are judged to attain-the cleanup standard when in reality, the concentrations for at least one well or chemical do not attain the cleanup standard.

Parameter A statistical property or characteristic of a population of values. Statistical quantities such as means, standard deviations, percentiles, etc. are parameters if they refer to a population of values, rather than to a sample of values.

Parameters of the Model See regression *coefficients.*

Parametric Test A test based on assumptions about the underlying process generating the data. For example, most parametric tests assume that the underlying data are normally distributed. Although parametric tests are strictly not valid unless the underlying assumptions are met, in many cases parametric tests perform well over a range of conditions found in the field. In particular, with reasonably large sample sizes the distribution of the mean will be approximately normal. See *robust* test. and *Central Limit Theorem.*

Percentile The specific value of a distribution that divides the set of measurements in such a way that P percent of the measurements fall below (or equal) this value, and 1-P percent of the measurements exceed this value. For specificity, a percentile is described by the value of P (expressed as a percentage). For

example, the 95th percentile (P=0.95) is that value X such that 95 percent of the data have values less than X, and 5 percent have values exceeding X. By definition, the median is the 50th percentile.

Physical Sample A portion of ground water collected from a well at the waste site and used to make measurements. This may also be called a **water sample.** A water sample may be mixed, subsampled, or otherwise handled to obtain the lab sample of ground water which is sent for laboratory analysis.

Point Estimate See *estimate.*

Population The totality of ground water samples in a well for which inferences regarding attainment of cleanup standards are to be made.

Population Mean Concentration The concentration which is the arithmetic average for the totality of ground water units (see also *mean and population).*

Population Parameters See *parameter.*

Power The probability that a statistical test will result in rejecting the null hypothesis when the null hypothesis is false. Power = $1 - \beta$, where β is the Type II error rate associated with the test. The term "power function" is more accurate because it reflects the fact that power is a function of a particular value of the parameter of interest under the alternative hypothesis.

Precision Recision refers to the degree to which repeated measurements are similar to one another. It measures the agreement (reproducibility) among individual measurements, obtained under prescribed similar conditions. Measurements which are precise are in close agreement. To use an analogy from archery, precise archers have all of their arrows land very close together. However, the arrows of a precise archer may or may not land on (or even near) the bull's-eye.

Predicted Value In regression analysis, the calculated value of $\hat{y}_i$, under the estimated regression line, for a particular value of x_i.

Proportion The number of ground water samples in a set of ground water samples that have a specified characteristic, divided by the total number of ground water samples in the set.

Random Error (ε_i) Represents "random" fluctuations of the observed chemical measurements around the hypothesized mean or regression model.

Random Sample A sample of ground water units selected using the simple random sampling procedures described in Section 4.1.

Range The difference between the maximum and minimum values of measurements in a data set.

Regression Analysis The process of finding the "best" mathematical model (within some restricted class of models) to describe the dependent variable, y_i, as a function of the independent variable, x_i, or to predict y_i from x_i. The most common form is the linear model.

Regression Coefficients The constants β_0 and β_1 in the simple linear regression model which represent the y-intercept and slope of the model.

Residual In regression analysis, the difference between the observed value of the concentration measurement y_i and the corresponding fitted (predicted) value, $\hat{y}_i$, from the estimated regression line.

Response Variable See *dependent variable.*

Robust Test A statistical test which is approximately valid under a wide range of conditions.

Sample Any collection of ground water samples taken from a well.

Sample Design The procedures used to select the ground water samples.

Sample Mean See *mean.*

Sample Residual See *residual.*

Sample Size The number of lab samples (i.e., the size of the statistical sample). Thus, a sample of size 10 consists of the measurements taken on 10 ground water samples or composite samples.

Sample Standard Deviation *See standard deviation.*

APPENDIX G: GLOSSARY

Sample Statistics Numerical quantities which summarize the properties of a data set

Sampling Error Variability in sample statistics between different samples that is used to characterize the precision of sample-based estimates

Sampling Frequency (n) The number of samples to be taken per year or seasonal period.

Sampling Plan See *sample design..*

Sampling Variability See *sampling error.*

Sequential Test A statistical test in which the decision to accept or reject the null hypothesis is made in a sequential fashion. Sequential tests are described in Chapters 4, 8, and 9 of this manual.

Serial Correlation A measure of the extent to which successive observations are **related.**

Significance Level The probability of a Type I error associated with a statistical test. In the context of the statistical tests presented in this manual, it is the probability that the ground water from a well or group of wells is declared to be clean when it is contaminated. The significance level is often denoted by the symbol a (Greek letter alpha).

Simple Linear Regression A regression analysis where there is only one independent variable and the equation for the model is of the form $y_i = \beta_0 + \beta_1 x_i$, where β_0 is the intercept and β_1 is the slope of the regression (see Section 6.1).

Simple Linear Regression Model A linear model relating the concentration measurements (or some other parameter) to time (see Section 6.1).

Size of **the Physical Sample** The volume of a physical ground water sample.

Skewness A measure of the extent to which a distribution is symmetric or asymmetric.

Skewed Distribution Any asymmetric distribution.

Standard Deviation A measure of dispersion of a set of data. Specifically, given a set of measurements, $x_1, x_2, ..., x_n$, the standard deviation is defined to be the

$$\text{quantity, } s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \text{ , where } \bar{x} \text{ is the sample mean.}$$

Standard Error A measure of the variability (or precision) of a sample estimate. Standard errors are often used to construct confidence intervals.

Statistical Sample A collection of chemical concentration measurements reported by the lab for one or more lab samples where the lab samples were collected using statistical sampling methods, Collection of a statistical sample allows estimation of precision and confidence intervals.

Statistical Test A formal statistical procedure and decision rule for deciding whether the ground water in a well attains the specified cleanup standard.

Steady State A state at which the residual effects of the treatment process (or any other temporary intervention) on general ground water characteristics appear to be negligible (see Section 7.1).

Sum of Squares Due to Error (SSE) A measure of how well the model fits the data necessary for assessing the adequacy of the model. If the SSE is small, the fit is good; if it is large, the fit is poor.

Symmetric Distribution A distribution of measurements for which the two sides of its overall shape are mirror images of each other about a center line.

Systematic Sample Ground water samples that are collected at equally-spaced intervals of time.

t-Distribution The distribution of a quotient of independent random variables, the numerator of which is a standardized normal variate with mean equal to zero and variance equal to one, and the denominator of which is the positive square root of the quotient of a chi-square distributed variate and its number of degrees of freedom. For additional details about the t-distribution, consult Resnikoff and Lieferman (1957) and Locks, Alexander, and Byars (1943).

Tolerance Interval A confidence interval around a percentile of a distribution of concentrations.

Transformation A manipulation of either the dependent of independent variable, or both, to normalize a distribution or linearize a model. Useful transformations include logarithmic, inverse, square root, etc.

Trends A general increase or decrease in concentrations over time which is persistent and unlikely to be due to random variation.

True Population Mean The actual, unknown arithmetic average contaminant level for all ground water samples in the population (see also mean and population).

Type I Error The error made when the ground water in a well is declared to be clean based on a statistical test when it is actually contaminated. This is also referred to as a *false positive.*

Type II Error The error made when the ground water in a well is declared to be contaminated when it is actually clean. This is also referred to as a *false negative.*

Variance The square of the standard deviation.

Waste Site The entire area being investigated for contamination.

Z Value Percentage point of a standard normal distribution. Z values are tabulated in Table A.2 of Appendix A.