

**STATISTICAL ANALYSIS
OF
GROUND WATER
AT
RCRA FACILITIES
-DRAFT GUIDANCE-**

**OFFICE OF SOLID WASTE
WASTE MANAGEMENT DIVISION
U.S. ENVIRONMENTAL PROTECTION AGENCY
401 M STREET, S.W.
WASHINGTON, D.C. 20460**

OCTOBER 1988

STATISTICAL ANALYSIS OF GROUND WATER

AT

RCRA FACILITIES

DRAFT GUIDANCE

For

**U.S. ENVIRONMENTAL PROTECTION AGENCY
OFFICE OF SOLID WASTE
WASTE MANAGEMENT DIVISION
401 M Street, S.W.
Washington, D.C. 20460**

DISCLAIMER

This document is intended to assist Regional and State personnel in evaluating ground-water monitoring data from RCRA facilities. Conformance with this guidance is expected to result in statistical methods and sampling procedures that meet the regulatory standard of protecting human health and the environment. However, EPA will not in all cases limit its approval of statistical methods and sampling procedures to those that comport with the guidance set forth herein. This guidance is not a regulation (i.e., it does not establish a standard of conduct which has the force of law) and should not be used as such. Regional and State personnel should exercise their discretion in using this guidance document as well as other relevant information in choosing a statistical method and sampling procedure for evaluating ground-water monitoring data from RCRA facilities.

This document has been reviewed by the Office of Solid Waste, U.S. Environmental Protection Agency, Washington, D.C., and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U.S. Environmental Protection Agency, nor does mention of trade names, commercial products, or publications constitute endorsement or recommendation for use.

ACKNOWLEDGMENT

This document was developed by EPA's Office of Solid Waste under the direction of Dr. Vernon Myers, Chief of the Ground-Water Section of the Waste Management Division. The document was prepared by the joint efforts of Dr. Vernon B. Myers, Mr. James R. Brown of the Waste Management Division, Mr. James Craig of the Office of Policy Planning and Information, and Mr. Barnes Johnson of the Office of Policy, Planning, and Evaluation.

PREFACE

This guidance document has been developed primarily for evaluating ground-water monitoring data at RCRA (Resource Conservation and Recovery Act) facilities. The statistical methodologies described in this document can be applied to both hazardous (Subtitle C of RCRA), and municipal (Subtitle D of RCRA) waste land disposal facilities.

The guidance has wider applications however, if one examines the spatial relationships involved between the monitoring wells and the potential contaminant source. For example, Section 4 of the guidance describes background well (upgradient) versus compliance well (downgradient) comparisons. This scenario can be applied to other "non-RCRA" situations involving the same spatial relationships and the same null hypothesis. The explicit null hypothesis (H_0) for testing contrasts between means, or where appropriate between medians, is that the means between groups (here monitoring wells) are equal (i.e., no release has been detected), or that the group means are below a prescribed action level (e.g., the ground-water protection standard). Statistical methods that can be used to evaluate these conditions are described in Sections 4.2 (Analysis of Variance), 4.3 (Tolerance Intervals), and 4.4 (Prediction Intervals).

When compliance wells (downgradient) are compared to a fixed standard (e.g., the ground-water protection standard), Section 5 of the guidance should be consulted. The value to which the compliance wells are compared can be any standard established by a Regional Administrator, State or county health official, or another appropriate official.

A note of caution applies to Section 5. The examples are used to determine whether or not ground-water has been contaminated as a result of a release from a facility. When the upper confidence limit is exceeded, further action or assessment may be warranted. If one wishes to determine whether or not a cleanup standard has been attained for a Superfund site or a RCRA facility in corrective action, another EPA guidance document entitled, "Statistical Methods for the Attainment of Superfund Cleanup Standards - Draft," should be consulted. This draft Superfund guidance is a multi-volume set that addresses questions regarding the success of air, ground-water, and soil remediation efforts. Information about the availability of this draft guidance that is currently being developed can be obtained by calling the RCRA/Superfund Hotline, telephone (800) 424-9346 or (202) 382-3000. Those interested in evaluating individual wells or an intra-well comparison are referred to Section 6 of the guidance which describes the use of Shewhart-CUSUM control charts and trend analysis.

Municipal water supply engineers for example, who wish to monitor water quality parameters in supply wells may find this section useful.

Other sections of the guidance have wide applications in the field of applied statistics, regardless of the intended use or purpose. Sections 3.2 and 3.3 provide information on checking distributional assumptions and equality of variance, while Sections 7.1 and 7.2 cover limit of detection problems and outliers. Any experiment involving the use of statistics may consult these sections for helpful advice and references.

Finally, it should be noted that this guidance is not intended to be the final chapter on the statistical analysis of ground-water monitoring data, nor should it be used as such. 40 CFR Part 264 Subpart F offers an alternative (§264.97(h)(5)) to the methods suggested and described in this guidance document. In fact, the guidance recommends a procedure (confidence intervals) for comparing monitoring data to a fixed standard that is not mentioned in the Subpart F regulations. This is neither contradictory or inconsistent, but rather epitomizes the complexities of the subject matter and exemplifies the need for flexibility due to the site specific monitoring requirements of the RCRA program.

CONTENTS

1.	Introduction.....	1-1
2.	Regulatory Overview.....	2-1
	2.1 Background.....	2-1
	2.2 Overview of Methodology.....	2-3
	2.3 General Performance Standards.....	2-3
	2.4 Basic Statistical Methods and Sampling Procedures.....	2-6
	2.5 Choosing a Sampling Interval.....	2-10
	2.6 Example Calculations.....	2-16
3.	Choosing a Statistical Method.....	3-1
	3.1 Flow Charts--Overview and Use.....	3-1
	3.2 Checking Distributional Assumptions.....	3-3
	3.3 Checking Equality of Variance: Bartlett's Test.....	3-16
4.	Background Well/Compliance Well Comparisons.....	4-1
	4.1 Summary Flow Chart for Background Well/Compliance Well Comparisons.....	4-2
	4.2 Analysis of Variance.....	4-5
	4.3 Tolerance Intervals Based on the Normal Distribution.....	4-18
	4.4 Prediction Intervals.....	4-21
5.	Comparisons with MCLs or ACLs.....	5-1
	5.1 Summary Chart for Comparison with MCLs or ACLs.....	5-1
	5.2 Statistical Procedures.....	5-1
6.	Control Charts for Intra-Well Comparisons.....	6-1
	6-1 Advantages of Plotting Data.....	6-1
	6-2 Correcting for Seasonality.....	6-2
	6.3 Combined Shewhart-Cusum Control Charts for Each Well and Constituent.....	6-5
	6.4 Update of a Control Chart.....	6-11
	6.5 Nondetects in a Control Chart.....	6-12
7.	Miscellaneous Topics.....	7-1
	7.1 Limit of Detection.....	7-1
	7.2 Outliers.....	7-10
Appendices		
A.	General Statistical Considerations and Glossary of Statistical Terms.....	A-1
B.	Statistical Tables.....	B-1
C.	General Bibliography.....	C-1

FIGURES

<u>Number</u>		<u>Page</u>
2-1	Hydraulic conductivity of selected rocks.....	2-12
2-2	Total porosity and drainable porosity for typical geologic materials.....	2-15
2-3	Potentiometric surface map for computation of hydraulic gradient.....	2-17
3-1	Flowchart overview.....	3-4
3-2	Probability plot of raw chlordane concentrations.....	3-11
3-3	Probability plot of log-transformed chlordane concentrations..	3-13
4-1	Background well/compliance well comparisons.....	4-3
4-2	Tolerance limits: alternate approach to background well/compliance well comparisons.....	4-4
5-1	Comparisons with MCLs/ACLs.....	5-2
6-1	Plot of unadjusted and seasonally adjusted monthly observations.....	6-6
6-2	Combined Shewhart-CUSUM chart.....	6-2

TABLES

<u>Number</u>		<u>Page</u>
2-1	Summary of Statistical Methods.....	2-7
2-2	Default Values for Effective Porosity for Use in Time of Travel (TOT) Analyses.....	2-13
2-3	Specific Yield Values for Selected Rock Units.....	2-14
2-4	Determining a Sampling Interval.....	2-19
3-1	Example Data for Coefficient of Variation Test.....	3-8
3-2	Example Data Computations for Probability Plotting.....	3-10
3-3	Cell Boundaries for the Chi-Squared Test.....	3-17
3-5	Example Data for Bartlett's Test.....	3-18
4-1	One-Way Parametric ANOVA Table.....	4-8
4-2	Example Data for One-Way Parametric Analysis of Variance.....	4-10
4-3	Example Computations in One-Way Parametric ANOVA Table.....	4-11
4-4	Example Data for One-Way Nonparametric ANOVA--Benzene Concentrations (in ppm).....	4-16
4-5	Example Data for Normal Tolerance Interval.....	4-22
4-6	Example Data for Prediction Interval--Chlordane Levels.....	4-25
5-1	Example Data for Normal Confidence Interval--Aldicarb Concentrations in Compliance Wells (ppb).....	5-4
5-2	Example Data for Log-Normal Confidence Interval--EDB Concentrations in Compliance Wells (ppb).....	5-6
5-3	Values of M and n+1-M and Confidence Coefficients for Small Samples.....	5-9
5-4	Example Data for Nonparametric Confidence Interval--Silvex Concentrations (ppm).....	5-10
5-5	Example Data for a Tolerance Interval Compared to an ACL.....	5-12

TABLES (continued)

<u>Number</u>		<u>Page</u>
6-1	Example Computation for Deseasonalizing Data.....	6-5
6-2	Example Data for Combined Shewhart-Cusum Chart--Carbon Tetrachloride Concentration ($\mu\text{g/L}$).....	6-9
7-1	Methods for Below Detection Limit Values.....	7-2
7-2	Example Data for a Test of Proportions.....	7-5
7-3	Example Data for Testing Cohen's Test.....	7-8
7-4	Example Data for Testing for an Outliers.....	7-12

SECTION 1

INTRODUCTION

The U.S. Environmental Protection Agency (EPA) promulgated regulations for detecting contamination of ground water at hazardous waste land disposal facilities under the Resource Conservation and Recovery Act (RCRA) of 1976. The statistical procedures specified for use to evaluate the presence of contamination have been criticized and require improvement. Therefore, EPA has proposed to revise those statistical procedures in 40 CFR Part 264, "Statistical Methods for Evaluating Ground-Water Monitoring Data from Hazardous Waste Facilities."

In 40 CFR Part 264, EPA has proposed statistical procedures that are appropriate for evaluating ground-water monitoring data under a variety of situations. The purpose of this document is to provide guidance in determining which situation applies and consequently which statistical procedure may be used. In addition to providing guidance on selection of an appropriate statistical procedure, this document provides instructions on carrying out the procedure and interpreting the results.

The regulations provide three levels of monitoring for a regulated unit: detection monitoring; compliance monitoring; and corrective action. The regulations define conditions for a regulated unit to be changed from one level of monitoring to a more stringent level of monitoring (e.g., from detection monitoring to compliance monitoring). These conditions are that there is statistically significant evidence of contamination.

The regulations allow the benefit of the doubt to reside with the current stage of monitoring. That is, a unit will remain in its current monitoring stage unless there is convincing evidence to change it. This means that a unit will not be changed from detection monitoring to compliance monitoring (or from compliance monitoring to corrective action) unless there is statistically significant evidence of contamination (or contamination above the compliance limit).

The main purpose of this document is to guide owners, operators, regional administrators, and other interested parties in the selection, use, and interpretation of appropriate statistical methods for monitoring the ground water at each specific regulated unit. Topics to be covered include sampling needed, sample sizes, selection of appropriate statistical design, matching analysis of data to design, and interpretation of results. Specific recommended methods are detailed and a general discussion of evaluation of alternate methods is provided. Statistical concepts are discussed in an

appendix. References for suggested procedures are provided as well as references to alternative procedures and general statistics texts. Situations calling for external consultation are mentioned as well as sources for obtaining expert assistance when needed.

SECTION 2

REGULATORY OVERVIEW

EPA promulgated ground-water monitoring and response standards for permitted facilities in 1982 (47 FR 32274, July 26, 1982), for detecting releases of hazardous wastes into ground water from storage, treatment, and disposal units.

The Subpart F regulations required ground-water data to be examined by statistical procedures to determine whether there was a significant exceedance of background levels, or other allowable levels, of specified chemical parameters and hazardous waste constituents. One concern was that the procedure in the regulations could result in a high rate of "false positives" (Type I error), thus requiring an owner or operator unnecessarily to advance into a more comprehensive and expensive phase of monitoring. More importantly, another concern was that the procedure could result in a high rate of "false negatives" (Type II error), i.e., instances where actual contamination would go undetected.

As a result of these concerns, EPA amended the procedure with five different statistical methods that are more appropriate for ground-water monitoring (53 FR 39720; October 11, 1988). These amendments also outline sampling procedures and performance standards that are designed to help minimize the event that a statistical method will indicate contamination when it is not present (Type I error), and fail to detect contamination when it is present (Type II error).

2.1 BACKGROUND

Subtitle C of the Resource Conservation Recovery Act of 1976 (RCRA) creates a comprehensive program for the safe management of hazardous waste. Section 3004 of RCRA requires owners and operators of facilities that treat, store, or dispose of hazardous waste to comply with standards established by EPA that are "necessary to protect human health and the environment." Section 3005 provides for implementation of these standards under permits issued to owners and operators by EPA or authorized States. Section 3005 also provides that owners and operators of existing facilities that apply for a permit and comply with applicable notice requirements may operate until a permit

determination is made. These facilities are commonly known as "interim status" facilities. Owners and operators of interim status facilities also must comply with standards set under Section 3004.

EPA promulgated ground-water monitoring and response standards for permitted facilities in 1982 (47 FR 32274, July 26, 1982), codified in 40 CFR Part 264, Subpart F. These standards established programs for protecting ground water from releases of hazardous wastes from treatment, storage, and disposal units. Facility owners and operators were required to sample ground water at specified intervals and to use a statistical procedure to determine whether or not hazardous wastes or constituents from the facility are contaminating ground water. As explained in more detail below, the Subpart F regulations regarding statistical methods used in evaluating ground-water monitoring data that EPA promulgated in 1982 have generated criticism.

The former Part 264 regulations provided that the Cochran's Approximation to the Behrens Fisher Student's t-test (CABF) or an alternate statistical procedure approved by EPA be used to determine whether there is a statistically significant exceedance of background levels, or other allowable levels, of specified chemical parameters and hazardous waste constituents. Although the regulations have always provided latitude for the use of an alternate statistical procedure, concerns were raised that the CABF statistical procedure in the regulations was not appropriate. It was pointed out that: (1) the replicate sampling method is not appropriate for the CABF procedure, (2) the CABF procedure does not adequately consider the number of comparisons that must be made, and (3) the CABF does not control for seasonal variation. Specifically, the concerns were that the CABF procedure could result in "false positives" (Type I error), thus requiring an owner or operator unnecessarily to collect additional ground-water samples, to further characterize ground-water quality, and to apply for a permit modification, which is then subject to EPA review. In addition, there was concern that CABF may result in "false negatives" (Type II error), i.e., instances where actual contamination goes undetected. This occurred because the background data, which is often used as the basis of the statistical comparisons, was highly variable due to temporal, spatial, analytical, and sampling effects.

As a result of these concerns, EPA amended both the statistical method and the sampling procedures of the regulations, by requiring (if necessary) that owners or operators more accurately characterize the hydrogeology and potential contaminants at the facility, and by including in the regulations performance standards that all the statistical methods and

sampling procedures must meet (53 FR 39720; October 11, 1988). Statistical methods and sampling procedures meeting these performance standards would have a low probability of indicating contamination when it is not present, and of failing to detect contamination that actually is present. The facility owner or operator would have to demonstrate that a procedure is appropriate for the conditions at the facility and to ensure that it meets the performance standards outlined below. This demonstration holds for any of the four statistical methods and sampling procedures outlined in this regulation as well as any alternate methods or procedures proposed by facility owners and operators.

EPA recognizes that the selection of appropriate monitoring parameters is also an essential part of a reliable statistical evaluation. The Agency addressed this issue in a previous Federal Register notice (52 FR 25942, July 9, 1987).

2.2 OVERVIEW OF METHODOLOGY

EPA has elected to retain the idea of general performance requirements that the regulated community must meet. This approach allows for flexibility in designing statistical methods and sampling procedures to site-specific considerations. EPA has tried to bring a measure of certainty to these methods, while accommodating the unique nature of many of the regulated units in question. Consistent with this general strategy, the Agency is establishing several options for the sampling procedures and statistical methods to be used in detection monitoring and, where appropriate, in compliance monitoring. The owner or operator shall submit, for each of the chemical parameters and hazardous constituents listed in the facility permit one or more of the statistical methods and sampling procedures described in today's regulations. In deciding which statistical test is appropriate, he or she will consider the theoretical properties of the test, the data available, the site hydrogeology, and the fate and transport characteristics of potential contaminants at the facility. The Regional Administrator will review, and if appropriate approve the proposed statistical methods and sampling procedures when issuing the facility permit.

The Agency recognizes that there may be situations where any one statistical test may not be appropriate. This is true of new facilities with little or no ground-water monitoring data. If insufficient data prohibits the owner or operator from specifying a statistical method of analysis, then contingency plans containing several methods of data analysis and the conditions under which the method can be used will be specified by the Regional Administrator in the permit. In many cases, the parametric ANOVA can be performed after six months of data has

been collected. The owner or operator may propose modifying the permit at a later date when more data is available and he wishes to use a specific method of analysis.

2.2 GENERAL PERFORMANCE STANDARDS

EPA's basic concern in establishing these performance standards for statistical methods is to achieve a proper balance between the risk that the procedures will falsely indicate that a regulated unit is causing background values or concentration limits to be exceeded (false positives), and the risk that the procedures will fail to indicate that background values or concentration limits are being exceeded (false negatives). EPA's approach is designed to address that concern directly. Thus, any statistical method or sampling procedure, whether specified here or as an alternative to those specified, should meet the following performance standards:

1. The statistical test is to be conducted separately for each hazardous constituent in each well. If the distribution of the chemical parameters or constituents is shown by the owner or operator to be inappropriate for a normal theory test, then the data should be transformed or distribution-free theory test should be used. If the distributions for the constituents differ, more than one statistical method may be needed.

2. If an individual well comparison procedure is used to compare an individual compliance well constituent concentration with background constituent concentrations or a ground-water protection standard, the test shall be done at a Type I error level of no less than 0.01 for each testing period. If a multiple comparisons procedure is used the Type I experimentwise error rate shall be no less than 0.05 for each testing period, however the Type I error of no less than 0.01 for individual well comparisons must be maintained. This performance standard does not apply to control charts, tolerance intervals, or prediction intervals unless they are modeled after hypothesis testing procedures that involve setting significance levels.

3. If a control chart approach is used to evaluate ground-water monitoring data, the specific type of control chart and its associated parameters shall be proposed by the owner or operator and approved by the Regional Administrator if he or she finds it to be protective of human health and the environment.

4. If a tolerance interval or a prediction interval is used to evaluate ground-water monitoring data, the levels of confidence, and for tolerance intervals the percentage of the population that the interval must contain, shall be proposed by the owner or operator and approved by the Regional Administrator if he or she

finds these parameters to be protective of human health and the environment. These parameters will be determined after considering the number of samples in the background data base, the distribution of the data, and the range of the concentration values for each constituent of concern.

5. The statistical method will include procedures for handling data below the limit of detection with one or more procedures that are protective of human health and the environment. Any practical quantification limit (pql) approved by the Regional Administrator under section 264.97(h) that is used in the statistical method shall be the lowest concentration level that can be reliably achieved within specified limits of precision and accuracy during routine laboratory operating conditions that are available to the facility.

6. The statistical method will consider, and if necessary control or correct for, seasonal and spatial variability and temporal correlation in the data.

In referring to "statistical methods", EPA means to emphasize that the concept of "statistical significance" must be reflected in several aspects of the monitoring program. This involves not only the choice of a level of significance, but also the choice of a statistical test, the sampling requirements, the number of samples, and the frequency of sampling. Since all of these interact to determine the ability of the procedure to detect contamination. The statistical methods, like a comprehensive ground-water monitoring program, must be evaluated in their entirety, not by individual components. Thus a systems approach to ground-water monitoring is endorsed.

The second performance standard requires further delineation. For individual well comparisons in which an individual compliance well is compared to background, the Type I error level shall be no less than 0.01 for each testing period. In other words, the probability of the test resulting in a false positive is no less than 1 in 100. EPA believes that this significance level is sufficient in limiting the false positive rate while at the same time controlling the false negative (missed detection) rate.

Owners and operators of facilities that have an extensive network of ground-water monitoring wells may find it more practical to use a multiple well comparisons procedure. Multiple comparisons procedures control the experimentwise error rate for comparisons involving multiple upgradient and downgradient wells. If this method is used, the Type I experimentwise error rate for each constituent shall be no less than 0.05 for each testing period.

In conducting a multiple well comparisons procedure, if the owner or operator chooses to use a t-statistic rather than an

F-statistic, the individual well Type I error level must be maintained at no less than 0.01. This provision should be considered if a facility owner or operator wishes to use a procedure that distributes the risk of a false positive evenly throughout all monitoring wells and monitoring parameters (eg. Bonferroni t-test).

Setting these levels of significance at 0.01 and 0.05 respectively raises an important question in how the false positive rate will be controlled at facilities with a large number of ground-water monitoring wells and monitoring constituents. The Agency set these levels of significance on the basis of a single testing period, and not on the entire operating life of the facility. Further, large facilities can reduce the false positive rate by implementing a unit-specific monitoring approach. Nonetheless it is evident that facilities with an extensive number of ground-water monitoring wells and that are monitoring for many constituents will still generate a large number of comparisons during each testing period. At these facilities, it may be difficult to keep the false positive error rate at an acceptable level.

Such cases may require the Regional Administrator to use discretion in deciding if a statistically significant result is indicative of an actual release from the facility, or if it is a false positive. In making this decision, the Regional Administrator may note the relative magnitude of the concentration of the constituent(s). If the exceedance is based on an observed compliance well value that has the same relative magnitude as the pql (practical quantification limit), or the background concentration level, then a false positive is more likely to be observed, and further sampling and testing may be appropriate. If however the background concentration level or an action level is exceeded by an order of magnitude, then the exceedance is more likely to be indicative of a release from the facility.

2.4 BASIC STATISTICAL METHODS AND SAMPLING PROCEDURES

The Final Rule specifies five types of statistical methods to detect contamination in ground water. EPA believes that at least one of these types of procedures will be appropriate for a wide variety of situations. To address situations where these methods may not be appropriate, EPA has included a provision for the owner or operator to select an alternate method which is subject to approval by the Regional Administrator.

1. A parametric analysis of variance (ANOVA) followed by multiple comparison procedures to identify specific sources of difference. The procedures will include estimation and testing of the contrasts between the mean of each compliance well and the background mean for each constituent.

2. An analysis of variance (ANOVA) based on ranks followed by multiple comparison procedures to identify specific sources of difference. The procedure will include estimation and testing of the contrasts between the median of each compliance well and the median background levels for each constituent.

3. A procedure in which a tolerance interval or a prediction interval for each constituent is established from the background data, and the level of each constituent in each compliance well is compared to its upper tolerance or prediction limit.

4. A control chart approach which will give control limits for each constituent. If any compliance well has a value or a sequence of values that lie outside the control limits for that constituent, it may constitute statistically significant evidence of contamination.

5. Another statistical method submitted by the owner or operator and approved by the Regional Administrator.

A summary of these statistical methods and their applicability is presented in Table 2-1. The table lists types of comparisons and the recommended procedure and refers the reader to the appropriate section where a discussion and example can be found.

EPA is specifying multiple statistical methods and sampling procedures and has allowed for alternatives because no one method or procedure is appropriate for all circumstances. EPA believes that the suggested methods and procedures are appropriate for the site-specific design and analysis of data from ground-water monitoring systems, and that they can account for more of the site-specific factors than Cochran's Approximation to the Behrens Fisher Student's t-test (CABF) and the accompanying sampling procedures in the past regulations. The statistical methods specified here address the multiple comparison problems and provide for documenting and accounting for sources of natural variation. EPA believes that the specified methods and

TABLE 2-1. SUMMARY OF STATISTICAL METHODS

SUMMARY OF STATISTICAL METHODS			
COMPOUND	TYPE OF COMPARISON	RECOMMENDED METHOD	SECTION OF GUIDANCE DOCUMENT
ANY COMPOUND IN BACKGROUND	BACKGROUND VS COMPLIANCE WELL	ANOVA TOLERANCE LIMITS PREDICTION INTERVALS	4.2 4.3 4.4
	INTRA-WELL	CONTROL CHARTS	6
ACL/MCL SPECIFIC	FIXED STANDARD	CONFIDENCE INTERVALS TOLERANCE LIMITS	5.2.1 5.2.2
SYNTHETIC	MANY NONDETECTS IN DATA SET	SEE BELOW DETECTION LIMIT TABLE 7-1	7.1

procedures consider and control for natural temporal and spatial variation. The decision on the number of wells needed in a monitoring system will be made on a site-specific basis by the Regional Administrator and will consider the statistical method being used, the site hydrogeology, the fate and transport characteristics of potential contaminants, and the sampling procedure. The number of wells must be sufficient to ensure a high probability of detecting contamination when it is present. To determine which sampling procedure should be used, the owner or operator shall consider existing data and site characteristics, including the possibility of trends and seasonality. These sampling procedures are:

1. Obtain a sequence of at least four samples taken at an interval that ensures, to the greatest extent technically feasible, that an independent sample is obtained, by reference to the uppermost aquifer's effective porosity, hydraulic conductivity, and hydraulic gradient, and the fate and transport characteristics of potential contaminants. The sampling interval that is proposed must be approved by the Regional Administrator.
2. An alternate sampling procedure proposed by the owner or operator and approved by the Regional Administrator if he or she finds it to be protective of human health and the environment.

EPA believes that the above sampling procedures will allow the use of statistical methods that will accurately detect contamination. These sampling procedures may be used to replace the sampling method present in the former Subpart F regulations. Rather than taking a single ground-water sample and dividing it into four replicate samples, a sequence of at least four samples taken at intervals far enough apart in time (daily, weekly, or monthly, depending on rates of ground-water flow and contaminant fate and transport characteristics) will help ensure the sampling of a discrete portion (i.e., an independent sample) of ground water. In hydrogeologic environments where the ground-water velocity prohibits one from obtaining four independent samples on a semiannual basis, the replicate sampling method described in the former Subpart F regulations or an alternate sampling procedure that is approved by the Regional Administrator may be utilized.

The Regional Administrator shall approve an appropriate sampling procedure and interval submitted by the owner or operator after considering the effective porosity, hydraulic conductivity, and hydraulic gradient in the uppermost aquifer under the waste management area, and the fate and transport characteristics of potential contaminants. Most of this information is already required to be submitted in the facility's Part B permit application under §270.14(c) and may be used by the owner or operator to make this determination. Further, the number and kinds of samples collected to establish background concentration levels should be appropriate to the form of statistical test employed, following generally accepted statistical principles. For example, the use of control charts presume a well defined background of at least eight samples per well. By contrast, ANOVA alternatives might require only four samples per well.

It seems likely that most facilities will be sampling monthly over four consecutive months, twice a year. In order to maintain a complete annual record of ground-water data, the facility owner or operator may find it desirable to obtain a sample each month of the year. This will help identify seasonal trends in the data and permit evaluation of the effects of autocorrelation and seasonal variation if present in the samples.

The concentrations of a constituent determined in these samples are intended to be used in one-point-in-time comparisons between background and compliance wells. Some facility owners or operators may want to use the concentrations to establish a "moving average" in the background well data base for comparison to the compliance well values at the frequency required in the facility permit. Using several background values to establish a "moving average" is an acceptable method of analysis; however, the number of degrees of freedom will be increased, making this method more sensitive to changes in constituent concentrations. Further, this method does not account for seasonal variation as effectively as one-point-in-time comparison procedures. Therefore, most owners and operators will find one-point-in-time comparisons to be the preferred method of analysis. This approach will help reduce the components of seasonal variation by providing for simultaneous comparisons between background and compliance well information.

The different sampling intervals were chosen to allow for the unique nature of the hydrogeologic systems beneath hazardous waste sites. This sampling scheme will give proper consideration to the temporal variation of and autocorrelation among the ground-water constituents. The specified procedure requires sampling data from background wells, at the compliance point, and according to a specific test protocol. The owner or operator should use a background value determined from data collected under this scenario if a test approved by the Regional Administrator requires it or if a concentration limit in compliance monitoring is to be based upon background data.

EPA recognizes that there may be situations where the owner or operator can devise alternate statistical methods and sampling procedures that are more appropriate to the facility and that will provide reliable results. Therefore, today's regulations allow the Regional Administrator to approve such procedures if he or she finds that the procedures balance the risk of false positives and false negatives in a manner comparable to that provided by the above specified tests and that they meet specified performance standards. In examining the comparability of the procedure to provide a reasonable balance between the risk of false positives and false negatives, the owner or operator will specify in the alternate plan such things as the sampling frequency and the sample size.

The methods indicate that the procedure must provide reasonable confidence that the migration of hazardous constituents from a regulated unit into and through the aquifer will be detected. (The reference to hazardous constituents does not mean that this option applies only to compliance monitoring; the procedure also applies to monitoring parameters and constituents in the detection monitoring program since they are surrogates indicating the presence of hazardous constituents.) The protocols for the specific tests, however, will be used as general benchmark to define "reasonable confidence"

in the proposed procedure. If the owner or operator shows that his suggested test is comparable in its results to one of the specified tests, then it is likely to be acceptable under the "reasonable confidence" test. There may be situations, however, where it will be difficult to directly compare the performance of an alternate test to the protocols for the specified tests. In such cases the alternate test will have to be evaluated on its own merits.

2.5 CHOOSING A SAMPLING INTERVAL

Section 264.97(g) of 40 CFR Part 264 Subpart F provides the owner or operator of an RCRA facility with a flexible sampling schedule that will allow him or her to choose a sampling procedure that will reflect site-specific concerns. This section specifies that the owner or operator shall, on a semi-annual basis, obtain a sequence of at least four samples from each well, based on an interval that is determined after evaluating the uppermost aquifer's effective porosity, hydraulic conductivity, and hydraulic gradient, and the fate and transport characteristics of potential contaminants. The intent of this provision is to set a sampling frequency that allows sufficient time to pass between sampling events to ensure, to the greatest extent technically feasible, that an independent ground-water sample is taken from each well. For further information on ground-water sampling, refer to the EPA "Practical Guide for Ground-Water Sampling," Carcelona et al., 1985.

The sampling frequency of the four semiannual sampling events required in Part 264 Subpart F can be based on estimates using the average linear velocity of ground water. Two forms of the Darcy equation stated below relate ground-water velocity to effective porosity (N_e), hydraulic gradient (i), and hydraulic conductivity (K):

$$V_h = (K_h * i) / N_e \quad \text{and} \quad V_v = (K_v * i) / N_e$$

where V_h and V_v are the horizontal and vertical components of the average linear velocity of ground water, respectively; K_h and K_v are the horizontal and vertical components of hydraulic conductivity; i is the head gradient; and N_e is the effective porosity. In applying these equations to ground-water monitoring, the average linear horizontal velocity can be used to determine an appropriate sampling interval. Usually, field investigations will yield bulk values for hydraulic conductivity. In most cases, the bulk hydraulic conductivity determined by a pump test, tracer test, or a slug test will be sufficient for these calculations. The vertical component of velocity (V_v), however, should be considered in estimating flow velocities in areas with significant components of vertical velocity such as recharge and discharge zones.

To apply the Darcy equation to ground-water monitoring, one needs to determine the parameters K , i , and N_e . The hydraulic conductivity, K , is the volume of water at the existing kinematic viscosity that will move in unit time under a unit hydraulic gradient through a unit area measured at right angles to the direction of flow. The reference to "existing kinematic viscosity" relates to the fact that hydraulic conductivity is not only determined by the media (aquifer), but also by fluid properties (ground water or potential contaminants). Thus, it is possible to have several hydraulic conductivity values for many different chemical substances that are present in the

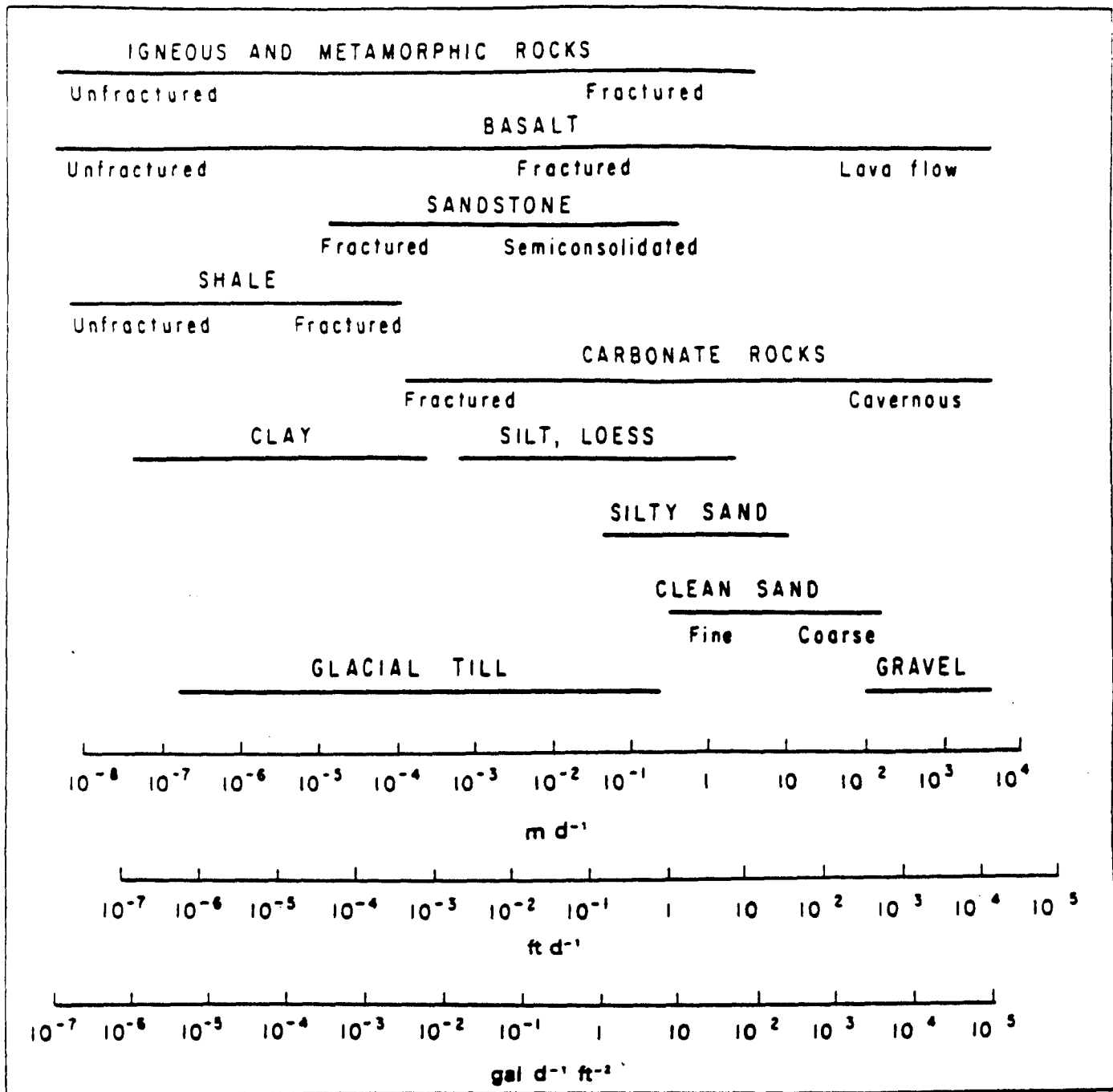
same aquifer. In either case it is advisable to use the greatest value for velocity that is calculated using the Darcy equation to determine sampling intervals. This will provide for the earliest detection of a leak from a hazardous waste facility and expeditious remedial action procedures. A range of hydraulic conductivities (the transmitted fluid is water) for various aquifer materials is given in Figure 2-1. The top line has units m/d; the middle line ft/d is commonly used. The bottom line has volume units per ft² of area.

The hydraulic gradient, i , is the change in hydraulic head per unit of distance in a given direction. It can be determined by dividing the difference in head between two points on a potentiometric surface map by the orthogonal distance between those two points (see example calculation). Water level measurements are normally used to determine the natural hydraulic gradient at a facility. However, the effects of mounding in the event of a leak from a waste disposal facility may produce a steeper local hydraulic gradient in the vicinity of the monitoring well. These local changes in hydraulic gradient should be accounted for in the velocity calculations.

The effective porosity, N_e , is the ratio, usually expressed as a percentage, of the total volume of voids available for fluid transmission to the total volume of the porous medium dewatered. It can be estimated during a pump test by dividing the volume of water removed from an aquifer by the total volume of aquifer dewatered (see example calculation). Table 2-2 presents approximate effective porosity values for a variety of aquifer materials. In cases where the effective porosity is unknown, specific yield may be substituted into the equation. Specific yields of selected rock units are given in Table 2-3. In the absence of measured values, drainable porosity is often used to approximate effective porosity. Figure 2-2 illustrates representative values of drainable porosity and total porosity as a function of aquifer particle size.

Once the values for K , i , and N_e are determined, the average linear ground-water velocity can be calculated. Using the Darcy equation, we can determine the time required for ground water to pass through the complete monitoring well diameter by dividing the monitoring well diameter by the average linear velocity of ground water. This value will represent the minimum time interval required between sampling events that will yield an independent ground-water sample. (Three-dimensional mixing of ground water in the vicinity of the monitoring well will occur when the well is purged before sampling, which is one reason why this method only provides an estimation of travel time).

In determining these sampling intervals, one should note that many chemical compounds will not travel at the same velocity as ground water. Chemical characteristics such as adsorptive potential, specific gravity, and molecular size will influence the way chemicals travel in the subsurface. Large molecules, for example, will tend to travel slower than the average linear velocity of ground water because of matrix interactions. Compounds that exhibit a strong adsorptive potential will undergo a similar fate that will dramatically change time of travel predictions using the Darcy equation. In some cases chemical interaction with the matrix material will alter the matrix structure and its associated hydraulic conductivity that may result in an increase in contaminant mobility.



Source: Heath, R. C. 1983. Basic Ground-Water Hydrology. U.S. Geological Survey Water Supply Paper, 2220, 84 pp.

Figure 2-1. Hydraulic conductivity of selected rocks.

TABLE 2-2. DEFAULT VALUES FOR EFFECTIVE POROSITY FOR USE
IN TIME OF TRAVEL (TOT) ANALYSES

Soil textural classes	Effective porosity of saturation ^a
<u>Unified soil classification system</u>	
GS, GP, GM, GC, SW, SP, SM, SC	0.20 (20%)
ML, MH	0.15 (15%)
CL, OL, CH, OH, PT	0.01 (1%) ^b
<u>USDA soil textural classes</u>	
Clays, silty clays, sandy clays	0.01 (1%) ^b
Silts, silt loams, silty clay loams	0.10 (10%)
All others	0.20 (20%)
<u>Rock units (all)</u>	
Porous media (nonfractured rocks such as sandstone and some carbonates)	0.15 (15%)
Fractured rocks (most carbonates, shales, granites, etc.)	0.0001 (0.01%)

Source: Barari, A., and L. S. Hedges. 1985. Movement of Water in Glacial Till. *Proceedings of the 17th International Congress of the International Association of Hydrogeologists*, pp. 129-134.

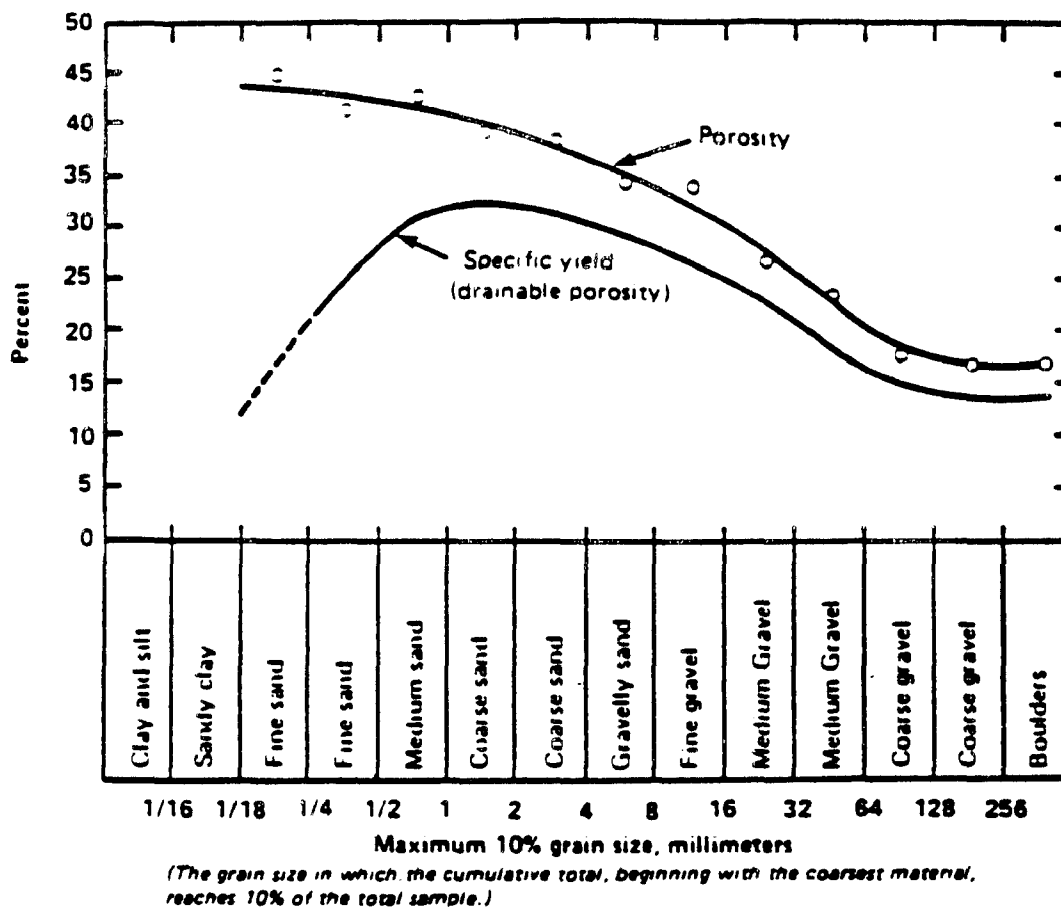
^a These values are estimates and there may be differences between similar units. For example, recent studies indicate that weathered and unweathered glacial till may have markedly different effective porosities (Barari and Hedges, 1985; Bradbury et al., 1985).

^b Assumes *de minimus* secondary porosity. If fractures or soil structure are present, effective porosity should be 0.001 (0.1%).

TABLE 2-3. SPECIFIC YIELD VALUES FOR
SELECTED ROCK UNITS

Rock type	Specific yield (%)
Clay	2
Sand	22
Gravel	19
Limestone	18
Sandstone (semiconsolidated)	6
Granite	0.09
Basalt (young)	8

Source: Heath, R. C. 1983. Basic Ground-Water Hydrology. U.S. Geological Survey, Water Supply Paper 2220, 84 pp.



Source: Todd, D. K. 1980. Ground Water Hydrology. John Wiley and Sons, New York. 534 pp.

Figure 2-2. Total porosity and drainable porosity for typical geologic materials.

This effect has been observed with certain organic solvents in clay units (see Brown and Andersen, 1981). Contaminant fate and transport models may be useful in determining the influence of these effects on movement in the subsurface. A variety of these models are available on the commercial market for private use.

2.6 EXAMPLE CALCULATIONS

2.6.1 Example 1: Determining the Effective Porosity (Ne)

The effective porosity, Ne, can be determined during a pump test using the following method:

$$Ne = \text{Volume of water removed} / \text{Volume of aquifer dewatered}$$

Volume of water removed:

Pumping rate of pump: 50 gpm
Pumping duration: 30 min

$$50 \text{ gpm} \times 30 \text{ min} = 1,500 \text{ gal}$$

Volume of aquifer dewatered:

$$V = (1/3)\pi r^2 h$$

where r is the radius of area affected by pumping and h is the drop in the water level. If, for example, h = 3 ft and r = 18 ft, then:

$$V = (1/3) \times 3.14 \times 18^2 \times 3 = 1,018 \text{ ft}^3$$

Next,

$$(1,018 \text{ ft}^3)(7.48 \text{ gal/ft}^3) = 7,615 \text{ gal}$$

from which:

$$Ne = 1,500 \text{ gal} / 7,615 \text{ gal} = 19.7\%$$

2.6.2 Example 2: Determining the Hydraulic Gradient (i)

The hydraulic gradient can be determined from a potentiometric surface map (Figure 2-3 below) using the following method.

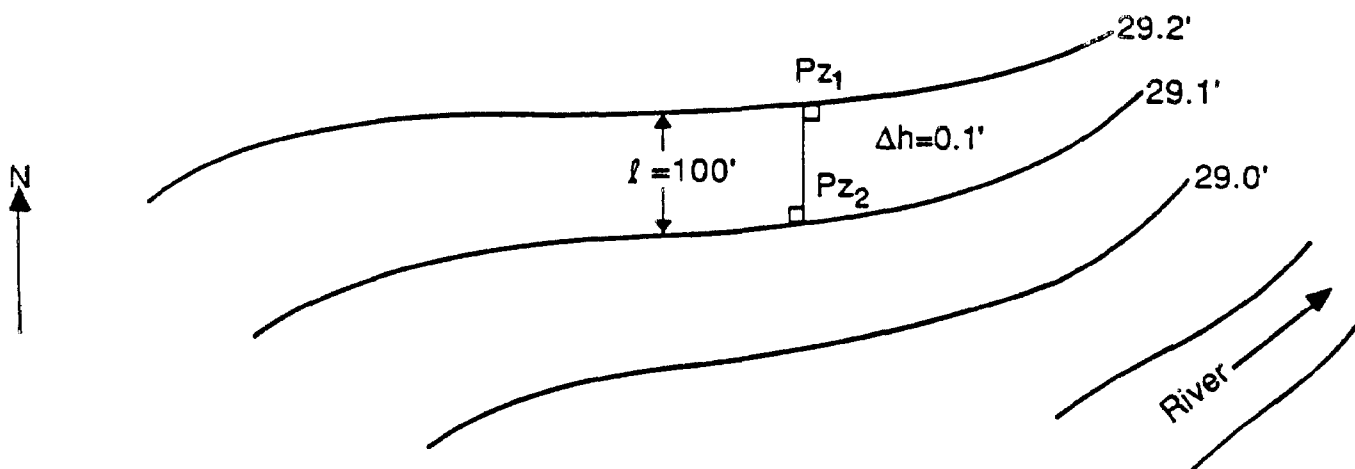


Figure 2-3. Potentiometric surface map for computation of hydraulic gradient.

$$i = \Delta h / l = (29.2 \text{ ft} - 29.1 \text{ ft}) / 100 \text{ ft} = 0.001 \text{ ft/ft}$$

Here, Δh is the difference in gradient measured at P_{z1} and P_{z2} , and l is the distance between the two points.

2.6.3 Example 3: Determining the Average Linear Velocity of Ground Water (Vh)

A land disposal facility has ground-water monitoring wells that are screened in an unconfined silty sand aquifer. Slug tests, pump tests, and tracer tests conducted during a hydrogeologic site investigation have revealed that the aquifer has a hydraulic conductivity (K_h) of 15 ft/d and an effective porosity (N_e) of 15%. Using a potentiometric map (as in example 2), the regional hydraulic gradient has been determined to be 0.003 ft/ft.

What is the minimum time interval between sampling events that will allow one to obtain an independent sample of ground water?

Calculate the average linear horizontal component of ground water (V_h):

$$K_h = 15 \text{ ft/d}$$

$$N_e = 0.15$$

$$i = 0.003 \text{ ft/ft}$$

$$V_h = iKh/Ne = (0.003)(15)/(0.15) = 0.3 \text{ ft/d}$$

$$(0.3 \text{ ft/d})(12 \text{ in/ft}) = 3.6 \text{ in/d}$$

Discussion: The average linear horizontal velocity of ground water has been calculated and is equal to 3.6 in/d. Monitoring well diameters at this particular facility are 4 in. We can determine the minimum time interval between sampling events that will allow one to obtain an independent sample of ground water by dividing the monitoring well diameter by the average linear velocity.

$$(4 \text{ in})/(3.6 \text{ in/d}) = 1.1 \text{ d}$$

Based on the above calculations, the owner or operator could sample every other day. However, because the velocity can vary with recharge rates seasonally, a weekly sampling interval would be advised.

Suggested Sampling Interval

<u>Date</u>	<u>Obtain Sample No.</u>
June 1	1
June 8	2
June 15	3
June 22	4

Table 2-4 gives some results for common situations.

TABLE 2-4. DETERMINING A SAMPLING INTERVAL

DETERMINING A SAMPLING INTERVAL				
UNIT	Kh (ft/d)	Ne (%)	Vh (in/mo)	SAMPLING INTERVAL
GRAVEL	EE 4	19	9.5 EE 4	DAILY
SAND	EE 2	22	8.2 EE 2	DAILY
SILT SAND	EE 1	14	1.3 EE 2	WEEKLY
TILL	EE -3	2	3.6 EE -2	MONTHLY *
CLAY (SEMICON)	EEO	6	3.0 EE 1	WEEKLY
BASALT	EE -1	8	2.25	MONTHLY *

The average linear velocities assume $I=0.005$.

* Use a Monthly sampling interval or an alternate sampling procedure

SECTION 3

CHOOSING A STATISTICAL METHOD

This chapter discusses the choice of an appropriate statistical method. Section 3.1 includes a flow chart to guide this selection. Section 3.2 contains procedures to test the distributional assumptions of statistical methods and Section 3.3 has procedures to test specifically for equality of variances.

The choice of an appropriate statistical test depends on the type of monitoring and the nature of the data. The proportion of values in the data set that are below detection is one important consideration. If most of the values are below detection, a test of proportions is suggested.

One set of statistical procedures is suggested when the monitoring consists of comparisons of water sample data from the background (hydraulically upgradient) well with the sample data from compliance (hydraulically downgradient) wells. The recommended approach is analysis of variance. Also, for a facility with limited amounts of data, it is advisable to initially use the ANOVA method of data evaluation, and later, when sufficient amounts of data are collected, to change to a tolerance interval or a control chart approach for each compliance well. However, alternative approaches are allowed. These include adjustments for seasonality, use of tolerance intervals, and use of prediction intervals. These methods are discussed in Chapter 4.

When the monitoring objective is to compare the concentration of a hazardous constituent to a fixed level such as a maximum concentration limit (MCL), a different type of approach is needed. This type of comparison commonly serves as a basis of compliance monitoring. Control charts may be used, as may tolerance or confidence intervals. Methods for comparison with a fixed level are presented in Chapter 5.

When a long history of data from each well is available, intra-well comparisons are appropriate. That is, the data from a single well are compared over time to detect shifts in concentration, or gradual trends in concentration that may indicate contamination. Methods for this situation are presented in Chapter 6.

3.1 FLOW CHARTS--OVERVIEW AND USE

The selection and use of a statistical procedure for ground-water monitoring is a detailed process. Because a single flow chart would become too complicated for easy use, a series of flow charts has been developed. These flow charts are found at the beginning of each chapter and are intended to guide the user in the selection and use of procedures in that chapter. The

more detailed flow charts can be thought of as attaching to the general flow charts at the indicated points.

There are three general types of statistical procedures discussed. One type of procedure is appropriate for comparisons among wells; that is, background well to compliance well data comparisons. The second type of procedure is to compare compliance well data with a constant limit such as an alternate concentration limit (ACL) or a maximum concentration limit (MCL). The third type of procedure is designed for intra-well comparisons. This method of analysis may be used when sufficient data from an individual well exist and the data allow for the identification of trends. A recommended control chart procedure (Starks, 1988) suggests that a minimum background sample of eight observations is needed. Thus an intra-well control chart approach could begin after the first complete year of data collection.

The first question to be asked in determining the appropriate statistical procedure is the type of monitoring program specified in facility permit. The type of monitoring program may determine if the appropriate comparison is among wells, comparison of downgradient data to a constant, intra-well comparisons, or a special case.

If the facility is in detection monitoring, the appropriate comparison is between wells that are hydraulically upgradient from the facility and those that are hydraulically downgradient. The statistical procedures for this type of monitoring are presented in Chapter 4. In detection monitoring, it is likely that many of the monitored constituents may result in few quantified results (i.e., much of the data are below the limit of analytical detection). If this is the case, then the test of proportions (Section 7.1.3) may be recommended. If the constituent occurs in measurable concentrations in background, then analysis of variance (Section 4.2) is recommended. This method of analysis is preferred when the data lack sufficient quantity to allow for the use of tolerance intervals or control charts.

If the facility is in compliance monitoring, the permit will specify the type of compliance limit. If the compliance limit is determined from the background, the statistical method is chosen from those that compare background well to compliance well data. Statistical methods for this case are presented in Chapter 4. The preferred method is the appropriate analysis of variance method in Section 4.2, or if sufficient data permit, tolerance intervals or control charts. The flow chart in Chapter 4 aids in determining which method is applicable.

If a facility in compliance monitoring has a constant maximum concentration limit (MCL) or alternate concentration limit (ACL) specified, then the appropriate comparison is with a constant. Methods for comparison with MCLs or ACLs are presented in Chapter 5, which contains a flow chart to aid in determining which method to use.

Finally, when more than one year of data have been collected from each well, the facility owner or operator may find it useful to perform intra-well comparisons over time to supplement the other methods. This is not a regulatory requirement, but it could provide the facility owner or operator with

information about the site hydrogeology. These methods are presented in Chapter 6.

The user should refer to Figure 3-1 to determine the type of comparison required. Initially this is either a comparison of background well to compliance well data or a comparison of compliance well data with an MCL or ACL. This leads the user to either Chapter 4 or 5. Eventually, there will be sufficient data to add the intra-well comparison presented in Chapter 6.

3.2 CHECKING DISTRIBUTIONAL ASSUMPTIONS

The purpose of this section is to provide users with methods to check the distributional assumptions of the statistical procedures recommended for ground-water monitoring. It is emphasized that one need not do an extensive study of the distribution of the data unless a nonparametric method of analysis is used to evaluate the data. If the owner or operator wishes to transform the data, it must first be shown that the untransformed data are inappropriate for a normal theory test. Similarly, if the owner or operator wishes to use nonparametric methods, he or she must demonstrate that the data do violate normality assumptions.

EPA has adopted this approach because most of the statistical procedures that meet the criteria set forth in the regulations are robust with respect to departures from many of the distributional assumptions. That is, only extreme violations of assumptions will result in an incorrect outcome of a statistical test. Moreover, it is only in situations where it is unclear whether contamination is present that departures from assumptions will alter the outcome of a statistical test. EPA therefore believes that it is protective of the environment to adopt the approach of not requiring testing of assumptions on a wide scale.

It should be noted that the distributional assumptions for statistical procedures apply to the errors of the observations. Application of the distributional tests to the observations themselves may lead to the conclusion that the distribution does not fit the observations. In some cases this lack of fit may be due to differences in means for the different wells or some other cause. The tests for distributional assumptions are best applied to the residuals from a statistical analysis. The residuals are the differences between the original observations and the values predicted by a model. For example, in analysis of variance, the predicted values are the group means and the residual is the difference between each observation and its group mean.

If the conclusion from testing the assumptions is that the assumptions are not adequately met, then a transformation of the data may be used or a nonparametric statistical procedure selected. Many types of concentration data have been reported in the literature to be adequately described by a log-normal distribution. That is, the natural logarithm of the original observations has been found to follow the normal distribution. Consequently, if the distributional assumptions are found to be violated for the original data, a transformation by taking the natural logarithm of each observation is suggested. This assumes that the data are all positive. If the log transformation does not adequately normalize the data or stabilize the variance, one

OVERVIEW FLOWCHART

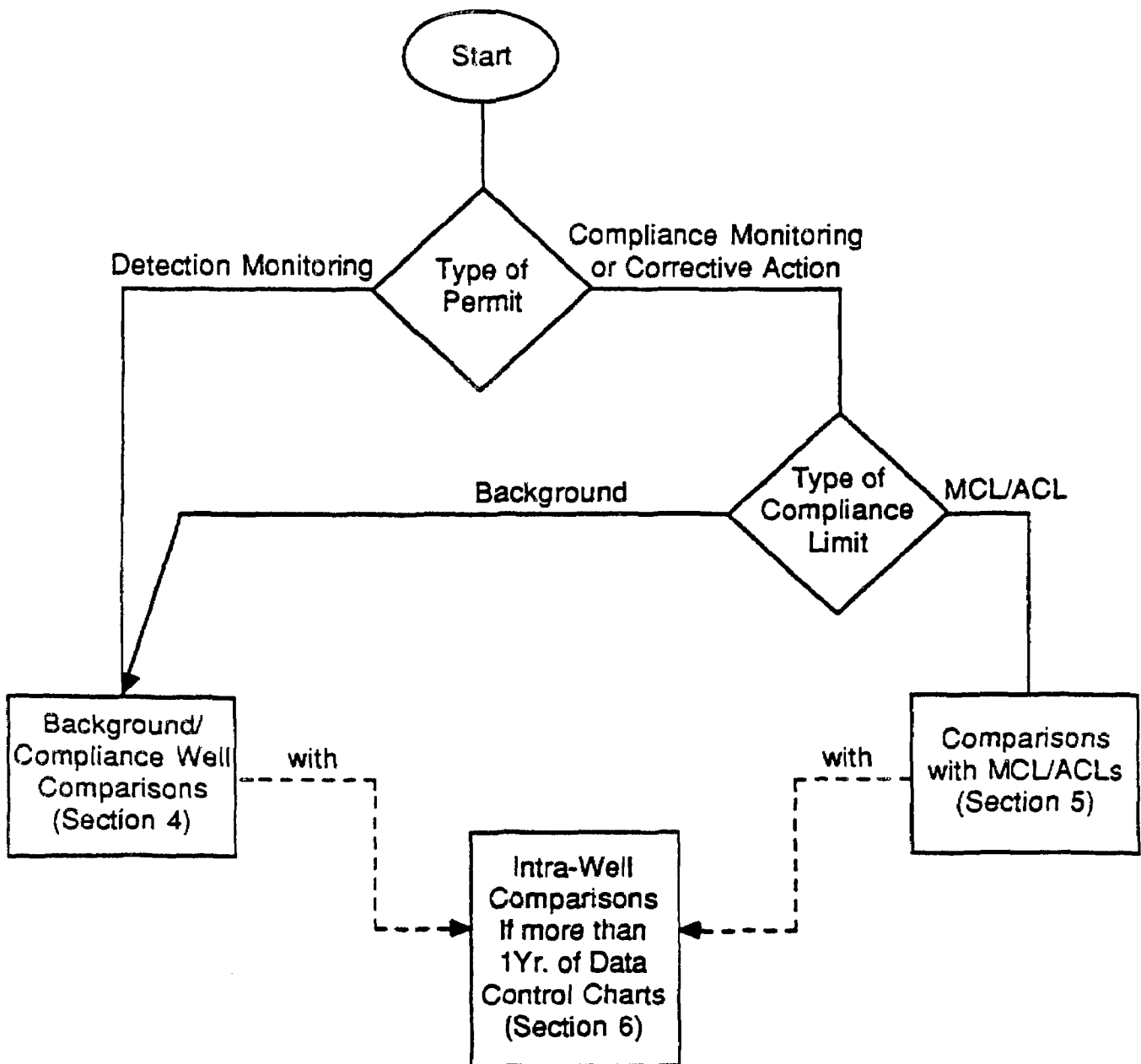


Figure 3-1

should use the nonparametric procedure or seek the consultation of a professional statistician to determine an appropriate statistical procedure.

The following sections present four selected approaches to check for normality. The first option refers to literature citation, the other three are statistical procedures. The choice is left to the user. The availability of statistical software and the user's familiarity with it will be a factor in the choice of a method. The coefficient of variation method, for example, requires only the computation of the mean and standard deviation of the data. Plotting on probability paper can be done by hand but becomes tedious with many data sets. However, the commercial Statistical Analysis System (SAS) software package provides a computerized version of a probability plot in its PROC UNIVARIATE procedure. SYSTAT, a package for PCs also has a probability plot procedure. The chi-squared test is not readily available through commercial software but can be programmed on a PC (for example in LOTUS 1-2-3) or in any other (statistical) software language with which the user is familiar. The amount of data available will also influence the choice. All tests of distributional assumptions require a fairly large sample size to detect moderate to small deviations from normality. The chi-squared test requires a minimum of 20 samples for a reasonable test.

Other statistical procedures are available for checking distributional assumptions. The more advanced user is referred to the Kolmogorov-Smirnov test (see, for example, Lindgren, 1976) which is used to test the hypothesis that data come from a specific (that is, completely specified) distribution. The normal distribution assumption can thus be tested for. A minimum sample size of 50 is recommended for using this test.

A modification to the Kolmogorov-Smirnov test has been developed by Lilliefors who uses the sample mean and standard deviation from the data as the parameters of the distribution (Lilliefors, 1967). Again, a sample size of at least 50 is recommended.

Another alternative to testing for normality is provided by the rather involved Shapiro-Wilk's test. The interested user is referred to the relevant article in *Biometrika* by Shapiro and Wilk (1965).

3.2.1 Literature Citation

PURPOSE

An owner or operator may wish to consult literature to determine what type of distribution the ground-water monitoring data for a specific constituent are likely to follow. This may avoid unnecessary computations and make it easier to determine whether there is statistically significant evidence of contamination.

PROCEDURE

One simple way to select a procedure based on a specific statistical distribution, is by citing a relevant published reference. The owner or operator may find papers that discuss data resulting from sampling ground water and

conclude that such data for a particular constituent follow a specified distribution. Citing such a reference may be sufficient justification for using a method based on that distribution, provided that the data do not show evidence that the assumptions are violated.

A literature citation should consider the distribution of data for the specific compound being monitored. In addition, it should consider sites with similar hydrogeologic characteristics to the extent possible. However, because many of the compounds may not be studied in the literature, extrapolations to compounds with similar chemical characteristics and to sites with similar hydrogeologic conditions are also acceptable. Basically, the owner or operator needs to provide some reason or justification for choosing a particular distribution.

3.2.2 Coefficient of Variation

Many statistical procedures assume that the data are normally distributed. The concentration of a hazardous constituent in ground water is inherently nonnegative, while the normal distribution allows for negative values. However, if the mean of the normal distribution is sufficiently above zero, the distribution places very little probability on negative observations and is still a valid approximation.

One simple check that can rule out use of the normal distribution is to calculate the coefficient of variation of the data. The use of this method was required by the former Part 264 Subpart F regulations pursuant to Section 264.97(h)(1). Because most owners and operators as well as regional personnel are already familiar with this procedure, it will probably be used frequently. The coefficient of variation, CV, is the standard deviation of the observations, divided by their mean. If the normal distribution is to be a valid model, there should be very little probability of negative values. The number of standard deviations by which the mean exceeds zero determines the probability of negative values. For example, if the mean exceeds zero by one standard deviation, the normal distribution will have less than 0.159 probability of a negative observation.

Consequently, one can calculate the standard deviation of the observations, calculate the mean, and form the ratio of the standard deviation divided by the mean. If this ratio exceeds 1.00, there is evidence that the data are not normal and the normal distribution should not be used for those data. (There are other possibilities for nonnormality, but this is a simple check that can rule out obviously nonnormal data.)

PURPOSE

This test is a simple check for evidence of gross nonnormality in the ground-water monitoring data.

PROCEDURE

To apply the coefficient of variation check for normality proceed as follows.

Step 1. Calculate the sample mean, \bar{X} , of n observations X_i , $i=1, \dots, n$.

$$\bar{X} = (\sum_{i=1}^n X_i)/n$$

Step 2. Calculate the sample standard deviation, S.

$$S = \left[\sum_{i=1}^n (X_i - \bar{X})^2 / (n - 1) \right]^{1/2}$$

Step 3. Divide the sample standard deviation by the sample mean. This ratio is the CV.

$$CV = S/\bar{X}.$$

Step 4. Determine if the result of Step 3 exceeds 1.00. If so, this is evidence that the normal distribution does not fit the data adequately.

EXAMPLE

Table 3-1 is an example data set of chlordane concentrations in 24 water samples from a fictitious site. The data are presented in order from least to greatest.

Applying the procedure steps to the data of Table 3-1, we have:

Step 1. $\bar{X} = 1.52$

Step 2. $S = 1.56$

Step 3. $CV = 1.56/1.52 = 1.03$

Step 4. Because the result of Step 3 was 1.03, which exceeds 1.00, we conclude that there is evidence that the data do not adequately follow the normal distribution. As will be discussed in other sections one would then either transform the data, use a nonparametric procedure, or seek professional guidance.

NOTE. The owner or operator may choose to use parametric tests since 1.03 is so close to the limit but should use a transformation or a nonparametric test if he or she believes that the parametric test results would be incorrect due to the departure from normality.

TABLE 3-1. EXAMPLE DATA FOR COEFFICIENT
OF VARIATION TEST

Chlordane concentration (ppm)
0.04
0.18
0.18
0.25
0.29
0.38
0.50
0.50
0.60
0.93
0.97
1.10
1.16
1.29
1.37
1.38
1.45
1.46
2.58
2.69
2.80
3.33
4.50
6.60

3.2.3 Plotting on Probability Paper

PURPOSE

Probability paper is a visual aid and a diagnostic tool in determining whether a set of data follows a normal distribution. Also, approximate estimates of the mean and standard deviation of the distribution can be read from the plot.

PROCEDURE

Let X be the variable; $X_1, X_2, \dots, X_i, \dots, X_n$ the set of n observations. The values of X can be raw data, residuals, or transformed data.

Step 1. Rearrange the observations in ascending order:

$$X(1), X(2), \dots, X(n).$$

Step 2. Compute the cumulative frequency for each distinct value $X(i)$ as $(i/(n+1)) \times 100\%$. The divisor of $(n+1)$ is a plotting convention to avoid cumulative frequencies of 100% which would be at infinity on the probability paper.

If a value of X occurs more than once, then the corresponding value of i increases appropriately. For example, if $X(2) = X(3)$, then the cumulative frequency for $X(1)$ is $100 \times 1/(n+1)$, but the cumulative frequency for $X(2)$ or $X(3)$ is $100 \times (1+2)/(n+1)$.

Step 3. Plot the distinct pairs $[X(i), (i/(n+1)) \times 100]$ values on probability paper (this paper is commercially available) using an appropriate scale for X on the horizontal axis. The vertical axis for the cumulative frequencies is already scaled from 0.01 to 99.99%.

If the points fall roughly on a straight line (the line can be drawn by hand with a ruler), then one can conclude that the underlying distribution is approximately normal. Also, an estimate of the mean and standard deviation can be made from the plot. The horizontal line drawn through 50% cuts the plotted line at the mean of the X values. The horizontal line going through 84% cuts the line at a value corresponding to the mean plus one standard deviation. By subtraction, one obtains the standard deviation.

REFERENCE

Dixon, W. J., and F. J. Massey, Jr. *Introduction to Statistical Analysis*. McGraw-Hill, Fourth Edition, 1983.

EXAMPLE

Table 3-2 lists 22 distinct chlordane concentration values (X) along with their frequencies. These are the same values as those listed in Table 3-1. There is a total of $n=24$ observations.

Step 1. Sort the values of X in ascending order (column 1).

Step 2. Compute $[100 \times (i/25)]$, column 4, for each distinct value of X , based on the values of i (column 2).

Step 3. Plot the pairs $[X_i, 100 \times (i/25)]$ on probability paper (Figure 3-2).

INTERPRETATION

The points in Figure 3-2 do not fall on a straight line; therefore, the hypothesis of an underlying normal distribution is rejected. However, the shape of the curve indicates a lognormal distribution. This is checked in the next step.

TABLE 3-2. EXAMPLE DATA COMPUTATIONS FOR
PROBABILITY PLOTTING

Concentration •X	Absolute frequency	i	$100 \times (i/(n+1))$	$\ln(X)$
0.04	1	1	4	-3.22
0.18	2	3	12	-1.71
0.25	1	4	16	-1.39
0.29	1	5	20	-1.24
0.38	1	6	24	-0.97
0.50	2	8	32	-0.69
0.60	1	9	36	-0.51
0.93	1	10	40	-0.07
0.97	1	11	44	-0.03
1.10	1	12	48	0.10
1.16	1	13	52	0.15
1.29	1	14	56	0.25
1.37	1	15	60	0.31
1.38	1	16	64	0.32
1.45	1	17	68	0.37
1.46	1	18	72	0.38
2.58	1	19	76	0.95
2.69	1	20	80	0.99
2.80	1	21	84	1.03
3.33	1	22	88	1.20
4.50	1	23	92	1.50
6.60	1	24	96	1.89

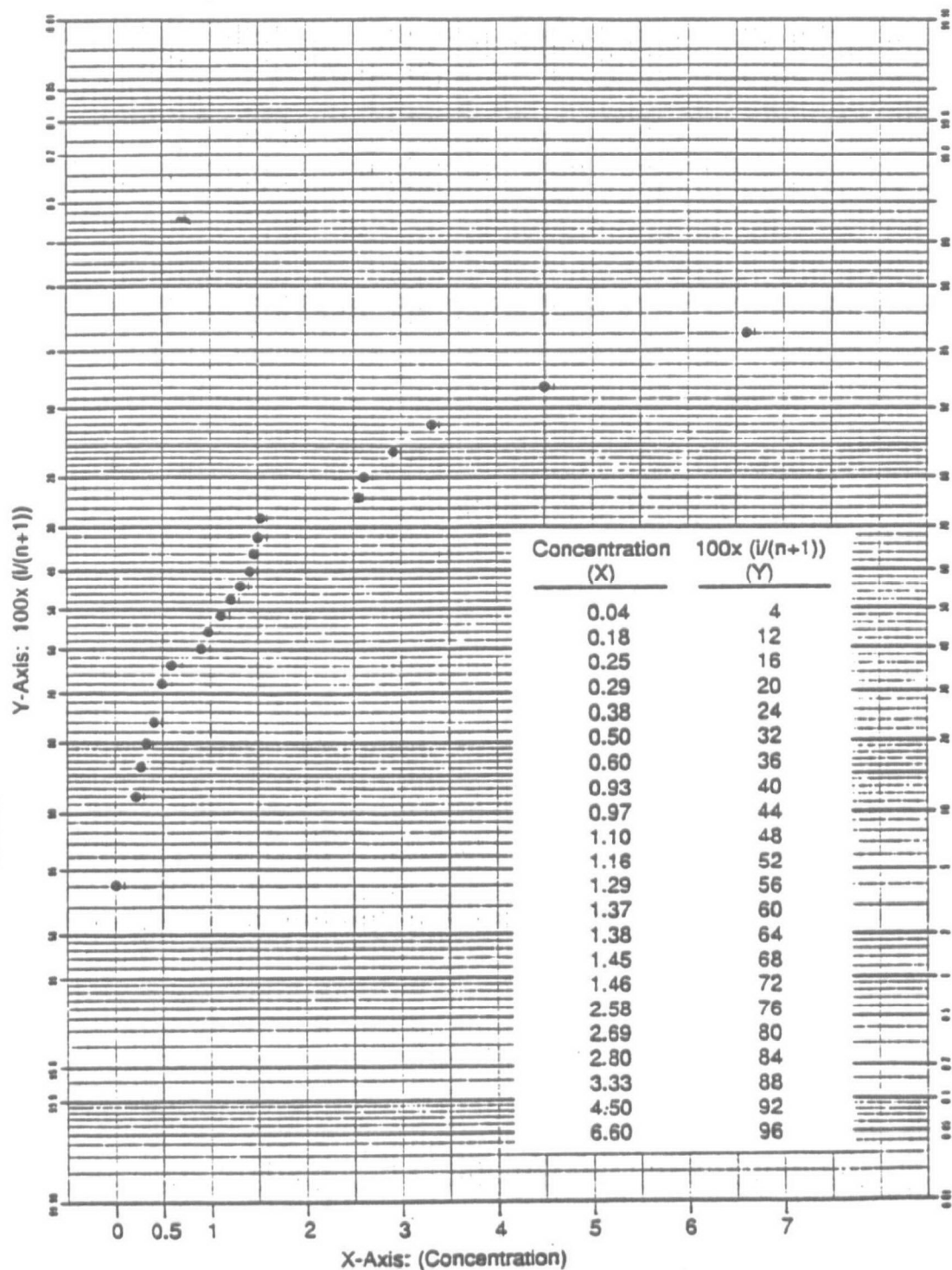


Figure 3-2. Probability plot of raw chlordane concentrations.

Next, take the natural logarithm of the X-values ($\ln(X)$), column 5 in Table 3-2). Repeat Step 3 above using the pairs $[\ln(X), 100 \times (i/25)]$. The resulting plot is shown in Figure 3-3. The points fall approximately on a straight line (hand-drawn) and the hypothesis of lognormality of X, i.e., $\ln(X)$ is normally distributed, can be accepted. The mean can be estimated at slightly below 0 and the standard deviation at about 1.2 on the log scale.

3.2.4 The Chi-Squared Test

The chi-squared test can be used to test whether a set of data fits a specified distribution. Most introductory courses in statistics explain the chi-squared test, and its familiarity among owners and operators as well as regional personnel may make it a frequently used method of analysis. In this application the assumed distribution is the normal distribution, but other distributions could also be used. The test consists of defining cells or ranges of values and determining the expected number of observations that would fall in each cell according to the hypothesized distribution. The actual number of data points in each cell is compared with that predicted by the distribution to judge the adequacy of the fit.

PURPOSE

The chi-squared test is used to test the adequacy of the assumption of normality of the data.

PROCEDURE

Step 1. Determine the appropriate number of cells, K. This number usually ranges from 5 to 10. Divide the number of observations, N, by 4. Use the largest whole number of this result, using 10 if the result exceeds 10.

Step 2. Standardize the data by subtracting the sample mean and dividing by the sample standard deviation:

$$Z_i = (X_i - \bar{X})/S$$

Step 3. Determine the number of observations that fall in each of the cells defined according to Table 3-3. The expected number of observations for each cell is N/K , where N is the total number of observations and K is the number of cells. Let N_i denote the observed number in cell i (for i taking values from 1 to K) and let E_i denote the expected number of observations in cell i. Note that in this case the cells are chosen to make the E_i 's equal.

Step 4. Calculate the chi-squared statistic by the formula below:

$$\chi^2 = \sum_{i=1}^K \frac{(N_i - E_i)^2}{E_i}$$

Step 5. Compare the calculated result to the table of the chi-squared distribution with K-3 degrees of freedom (Table 1, Appendix B). Reject the hypothesis of normality if the calculated value exceeds the tabulated value.

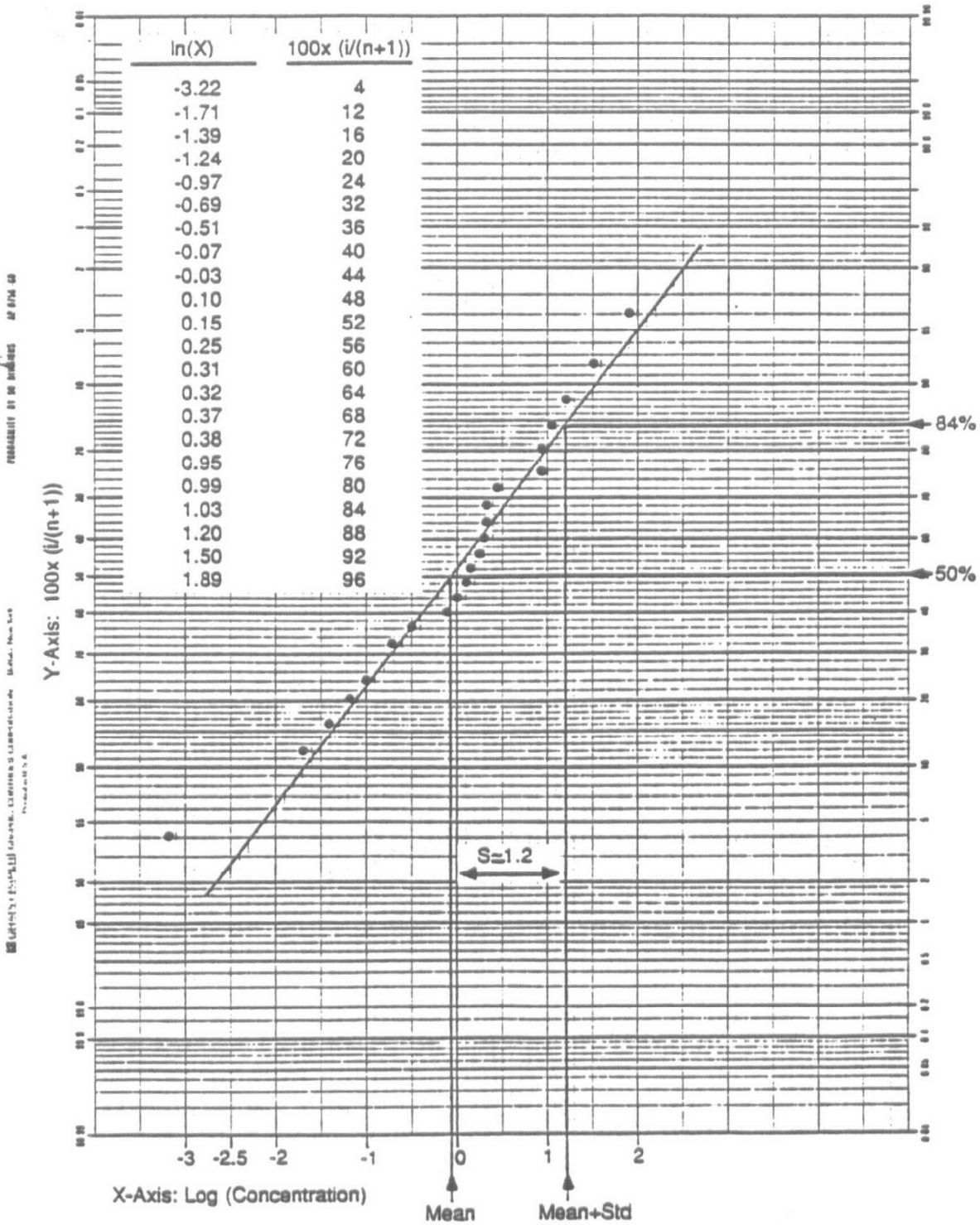


Figure 3-3. Probability plot of log-transformed chlordane concentrations.

TABLE 3-3. CELL BOUNDARIES FOR THE CHI-SQUARED TEST

	Number of cells (K)					
	5	6	7	8	9	10
Cell boundaries	-0.84	-0.97	-1.07	-1.15	-1.22	-1.28
for equal ex-	-0.25	-0.43	-0.57	-0.67	-1.08	-0.84
pected cell	0.25	0.00	-0.18	-0.32	-0.43	-0.52
sizes with the	0.84	0.43	0.18	0.00	-0.14	-0.25
normal distri-		0.97	0.57	0.32	0.14	0.00
bution			1.07	0.67	0.43	0.25
				1.15	1.08	0.52
					1.22	0.84
						1.28

REFERENCE

Remington, R. D., and M. A. Schork. *Statistics with Applications to the Biological and Health Sciences*. Prentice-Hall, 1970. 235-236.

EXAMPLE

The data in Table 3-4 are $N = 21$ residuals from an analysis of variance on dioxin concentrations. The analysis of variance assumes that the errors (estimated by the residuals) are normally distributed. The chi-squared test is used to check this assumption.

Step 1. Divide the number of observations, 21, by 4 to get 5.25. Keep only the integer part, 5, so the test will use $K = 5$ cells.

Step 2. The sample mean and standard deviation are calculated and found to be: $\bar{X} = 0.00$, $S = 0.24$. The data are standardized by subtracting the mean (0 in this case) and dividing by S . The results are also shown in Table 3-4.

Step 3. Determine the number of (standardized) observations that fall into the five cells determined from Table 3-3. These divisions are: (1) less than or equal to -0.84, (2) greater than -0.84 and less than or equal to -0.25, (3) greater than -0.25 and less than or equal to +0.25, (4) greater than 0.25 and less than or equal to 0.84, and (5) greater than 0.84. We find 4 observations in cell 1, 6 in cell 2, 2 in cell 3, 4 in cell 4, and 5 in cell 5.

Step 4. Calculate the chi-squared statistic. The expected number in each cell is N/K or $21/5 = 4.2$.

$$\chi^2 = \frac{(4 - 4.2)^2}{4.2} + \dots + \frac{(5 - 4.2)^2}{4.2} = 2.10$$

TABLE 3-4. EXAMPLE DATA FOR CHI-SQUARED
TEST

Observation	Residual	Standardized residual
1	-0.45	-1.90
2	-0.35	-1.48
3	-0.35	-1.48
4	-0.22	-0.93
5	-0.16	-0.67
6	-0.13	-0.55
7	-0.11	-0.46
8	-0.10	-0.42
9	-0.10	-0.42
10	-0.06	-0.25
11	-0.05	-0.21
12	0.04	0.17
13	0.11	0.47
14	0.13	0.55
15	0.16	0.68
16	0.17	0.72
17	0.20	0.85
18	0.21	0.89
19	0.30	1.27
20	0.34	1.44
21	0.41	1.73

Step 5. The critical value at the 5% level for a chi-squared test with 2 ($K-3 = 5-3 = 2$) degrees of freedom is 5.99 (Table 1, Appendix B). Because the calculated value of 2.10 is less than 5.99 there is no evidence that these data are not normal.

INTERPRETATION

The cell boundaries are determined from the normal distribution so that equal numbers of observations should fall in each cell. If there are large differences between the number of observations in each cell and that predicted by the normal distribution, this is evidence that the data are not normal. The chi-squared statistic is a nonnegative statistic that increases as the difference between the predicted and observed number of observations in each cell increases.

If the calculated value of the chi-squared statistic exceeds the tabulated value, there is statistically significant evidence that the data do not follow the normal distribution. In that case, one would need to do a transformation, use a nonparametric procedure, or seek consultation before interpreting the results of the test of the ground-water data. If the calculated value of the chi-squared statistic does not exceed the tabulated critical value, there is no significant lack of fit to the normal distribution and one can proceed assuming that the assumption of normality is adequately met.

Remark: The chi-squared statistic can be used to test whether the residuals from an analysis of variance or other procedure are normal. In this case the degrees of freedom are found by (number of cells minus one minus the number of parameters that have been estimated). This may require more than the suggested 10 cells. The chi-squared test does require a fairly large sample size in that there should be generally at least four observations per cell.

3.3 CHECKING EQUALITY OF VARIANCE: BARTLETT'S TEST

The analysis of variance procedures presented in Chapter 4 are often more sensitive to unequal variances than to moderate departures from normality. The procedures described in this section allow for testing to determine whether group variances are equal or differ significantly. Often in practice unequal variances and nonnormality occur together. Sometimes a transformation to stabilize or equalize the variances also produces a distribution that is more nearly normal. This sometimes occurs if the initial distribution was positively skewed with variance increasing with the number of observations. Only Bartlett's test for checking equality, or homogeneity, of variances is presented here. It encompasses checking equality of more than two variances with unequal sample sizes. Other tests are available for special cases. The F-test is a special situation when there are only two groups to be compared. The user is referred to classical textbooks for this test (e.g., Snedecor and Cochran, 1980). In the case of equal sample sizes but more than two variances to be compared, the user might want to use Hartley's or maximum F-ratio test (see Nelson, 1987). This test provides a quick procedure to test for variance homogeneity.

PURPOSE

Bartlett's test is a test of homogeneity of variances. In other words, it is a means of testing whether a number of population variances of normal distributions are equal. This is an assumption made in analysis of variance when comparing concentrations of constituents between background and compliance wells, or among compliance wells. It should be noted that Bartlett's test is sensitive to nonnormality in the data. With long-tailed distributions the test too often rejects equality (homogeneity) of the variances.

PROCEDURE

Assume that data from k wells are available and that there are n_i data points for well i .

Step 1. Compute the k sample variances S_1^2, \dots, S_k^2 . Each variance has associated with it $f_i = n_i - 1$ degrees of freedom. Take the natural logarithm of each variance, $\ln(S_1^2), \dots, \ln(S_k^2)$.

Step 2. Compute the test statistic

$$X^2 = f \ln(S_p^2) - \sum_{i=1}^k f_i \ln(S_i^2)$$

$$\text{where } f = \sum_{i=1}^k f_i = \left(\sum_{i=1}^k n_i \right) - k$$

thus f is the total sample size minus the number of wells (groups); and

$$S_p^2 = \frac{1}{f} \sum_{i=1}^k f_i S_i^2, \text{ the pooled variance across wells.}$$

Step 3. Using the chi-squared table (Table 1, Appendix B), find the critical value for X^2 with $(k-1)$ degrees of freedom at a predetermined significance level, for example, 0.05.

INTERPRETATION

If the calculated value X^2 is larger than the tabulated value, then conclude that the variances are not equal at that significance level.

REFERENCE

Johnson N. L., and F. C. Leone. *Statistics and Experimental Design in Engineering and the Physical Sciences*. Vol. I, John Wiley and Sons, New York, 1977.

EXAMPLE

Manganese concentrations are given for k=6 wells in Table 3-5 below.

TABLE 3-5. EXAMPLE DATA FOR BARTLETT'S TEST

Sampling date	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6
January 1	50	46	272	34	48	68
February 1	73	77	171	3,940	54	991
March 1	244		32			54
April 1	202		53			
$n_i =$	4	2	4	2	2	3
$f_i = n_i - 1 =$	3	1	3	1	1	2
$S_i =$	95	22	112	2,762	3	537
$S_i^2 =$	9,076	481	12,454	7,628,418	8	288,349
$f_i * S_i^2 =$	27,229	481	37,362	7,628,418	8	576,698
$\ln(S_i^2) =$	9	6	9	16	2	13
$f_i * \ln(S_i^2) =$	27	6	28	16	2	25

Step 1. • Compute the six sample variances and take their natural logarithm, $\ln(S_1^2), \dots, \ln(S_6^2)$, as 9, 6, ..., 13, respectively.

Step 2. • Compute $\sum_{i=1}^6 f_i * \ln(S_i^2) = 105$,

This is the sum of the last line in Table 3-5.

• Compute $f = \sum_{i=1}^6 f_i = 3 + 1 + \dots + 2 = 11$

• Compute S_p^2

$$S_p^2 = \frac{1}{11} \sum_{i=1}^6 f_i * S_i^2 = \frac{1}{11} (27,299 + \dots + 576,698) = \frac{1}{11} * 8,270,195 = 751,836$$

- Take the natural logarithm of S_p^2 : $\ln(S_p^2) = 14$
- Compute $X^2 = 11 \cdot 14 - 105 = 44$

Step 3. The critical X^2 value with $6-1 = 5$ degrees of freedom at the 5% significance level is 11.1 (Table 1 in Appendix B). Since 44 is larger than 11.1, we conclude that the six variances S_1^2, \dots, S_6^2 , are not homogeneous at the 5% significance level.

INTERPRETATION

The sample variances of the data from the six wells were compared by means of Bartlett's test. The test was significant at the 5% level, suggesting that the variances are significantly unequal (heterogeneous). A log-transform of the data can be done and the same test performed on the transformed data. Generally, if the data followed skewed distribution, this approach resolves the problem of unequal variances and the user can proceed with an ANOVA for example.

On the other hand, unequal variances among well data could be a direct indication of well contamination, since the individual data could come from different distributions (i.e., different means and variances). Then the user may wish to test which variance differs from which one. The reader is referred here to the literature for a gap test of variance (Tukey, 1949; David, 1956; or Nelson, 1987).

NOTE

- In the case of $k=2$ variances, the test of equality of variances is the F-test (Snedecor and Cochran, 1980).

- Bartlett's test simplifies in the case of equal sample sizes, $n_i = n$, $i=1, \dots, k$. The test used then is Cochran's test. Cochran's test focuses on the largest variance and compares it to the sum of all the variances. Hartley introduced a quick test of homogeneity of variances that uses the ratio of the largest over the smallest variances. Technical aids for the procedures under the assumption of equal sample sizes are given by L. S. Nelson in the *Journal of Quality Technology*, Vol. 19, 1987, pp. 107 and 165.

SECTION 4

BACKGROUND WELL/COMPLIANCE WELL COMPARISONS

There are many situations in ground-water monitoring that call for the comparison of data from different wells. The assumption is that a set of wells can be defined that are not contaminated. Generally these are background wells and have been sited to be hydraulically upgradient from the regulated unit. A second set of wells are sited hydraulically downgradient from the regulated unit and are otherwise known as compliance wells. The data from these compliance wells are compared to the data from the background wells to determine whether there is any evidence of contamination in the compliance wells that would presumably result from a release from the regulated unit.

If the owner or operator of a hazardous waste facility does not have reason to suspect that the test assumptions of equal variance or normality will be violated, then he or she may simply choose the parametric analysis of variance as a default method of statistical analysis. In the event that this method indicates a statistically significant difference between the groups being tested, then the test assumptions should be evaluated.

This situation, where the relevant comparison is between data from background wells and data from compliance wells, is the topic of this section. Comparisons between background well data and compliance well data may be called for in all phases of monitoring. This is the general case for detection monitoring. It is also the case for compliance monitoring if the compliance limits are determined by the background well constituent concentration levels. It may also be the case in corrective action. Compounds that are present in background wells (e.g., naturally occurring metals) are most appropriately evaluated using this comparison method.

The procedures described in this section are applicable whenever the relevant comparison is between background well data and compliance well data. Section 4.1 provides a flow chart and overview for the selection of methods for comparison of background well and compliance well data. Section 4.2 contains analysis of variance methods. These provide methods for directly comparing background well data to compliance well data. Section 4.3 describes a tolerance interval approach, where the background well data are used to define the tolerance limits for comparison with the compliance well data. Section 4.4 contains an approach based on prediction intervals, again using the background well data to determine the prediction interval for comparison with the compliance well data. Methods for comparing data to a fixed compliance limit (an MCL or ACL) will be described in Section 5.

4.1 SUMMARY FLOW CHART FOR BACKGROUND WELL/COMPLIANCE WELL COMPARISONS

Figure 4-1 is a flow chart to aid in selecting the appropriate statistical procedure for background well to compliance well comparisons. The first step is to determine whether most of the observations are quantified (that is, above the detection limits) or not. Generally, if more than 50% of the observations are below the detection limit (as might be the case with detection or compliance monitoring for volatile organics) then the appropriate comparison is a test of proportions. The test of proportions compares the proportion of detected values in the background wells to those in the compliance wells. See Section 7.1 for a discussion of dealing with data below the detection limit.

If the proportion of detected values is 50% or more, then an analysis of variance procedure is the first choice. Tolerance limits or prediction intervals are acceptable alternative choices that the user may select. The analysis of variance procedures give a more thorough picture of the situation at the facility, however, the tolerance limit or prediction interval approach is acceptable and requires less computation in many situations.

Figure 4-2 is a flow chart to guide the user if a tolerance limits approach is selected. The first step in using Figure 4-2 is to determine whether the facility is in detection monitoring. If so, much of the data may be below the detection limit. See Section 7.1 for a discussion of this case, which may call for consulting a statistician. If most of the data are quantified, then follow the flow chart to determine if normal tolerance limits can be used. If the data are not normal (as determined by one of the procedures in Section 3.2), then the logarithm transformation may be done and the transformed data checked for normality. If the log data are normal, the lognormal tolerance limit should be used. If neither the original data nor the log-transformed data are normal, seek consultation with a professional statistician.

If a prediction interval is selected as the method of choice, see Section 4.4 for guidance in performing the procedure.

If analysis of variance is to be used, then continue with Figure 4-1 to select the specific method that is appropriate. A one-way analysis of variance is recommended. If the data show evidence of seasonality (observed, for example, in a plot of the data over time), a trend analysis or perhaps a two-way analysis of variance may be the appropriate choice. These instances may require consultation with a professional statistician.

If the one-way analysis of variance is appropriate, the computations are performed, then the residuals are checked to see if they meet the assumptions of normality and equal variance. If so, the analysis concludes. If not, a logarithm transformation may be tried and the residuals from the analysis of variance on the log data are checked for assumptions. If these still do not adequately satisfy the assumptions, then a one-way nonparametric analysis of variance may be done, or professional consultation may be sought.

BACKGROUND/COMPLIANCE WELL COMPARISONS

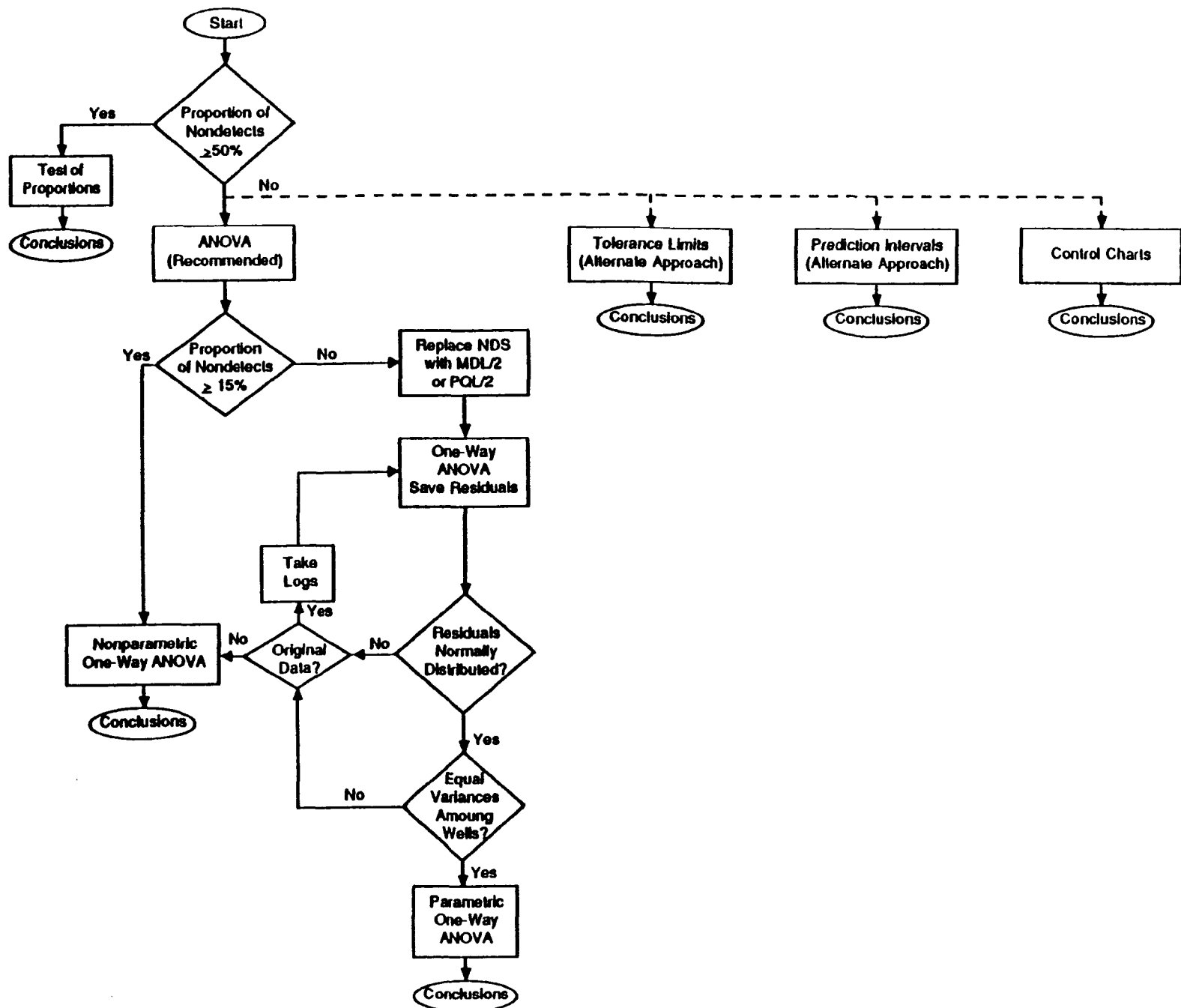


Figure 4-1. Background well/compliance well comparisons.

Tolerance Limits: Alternate Approach to Background/Compliance Well Comparisons

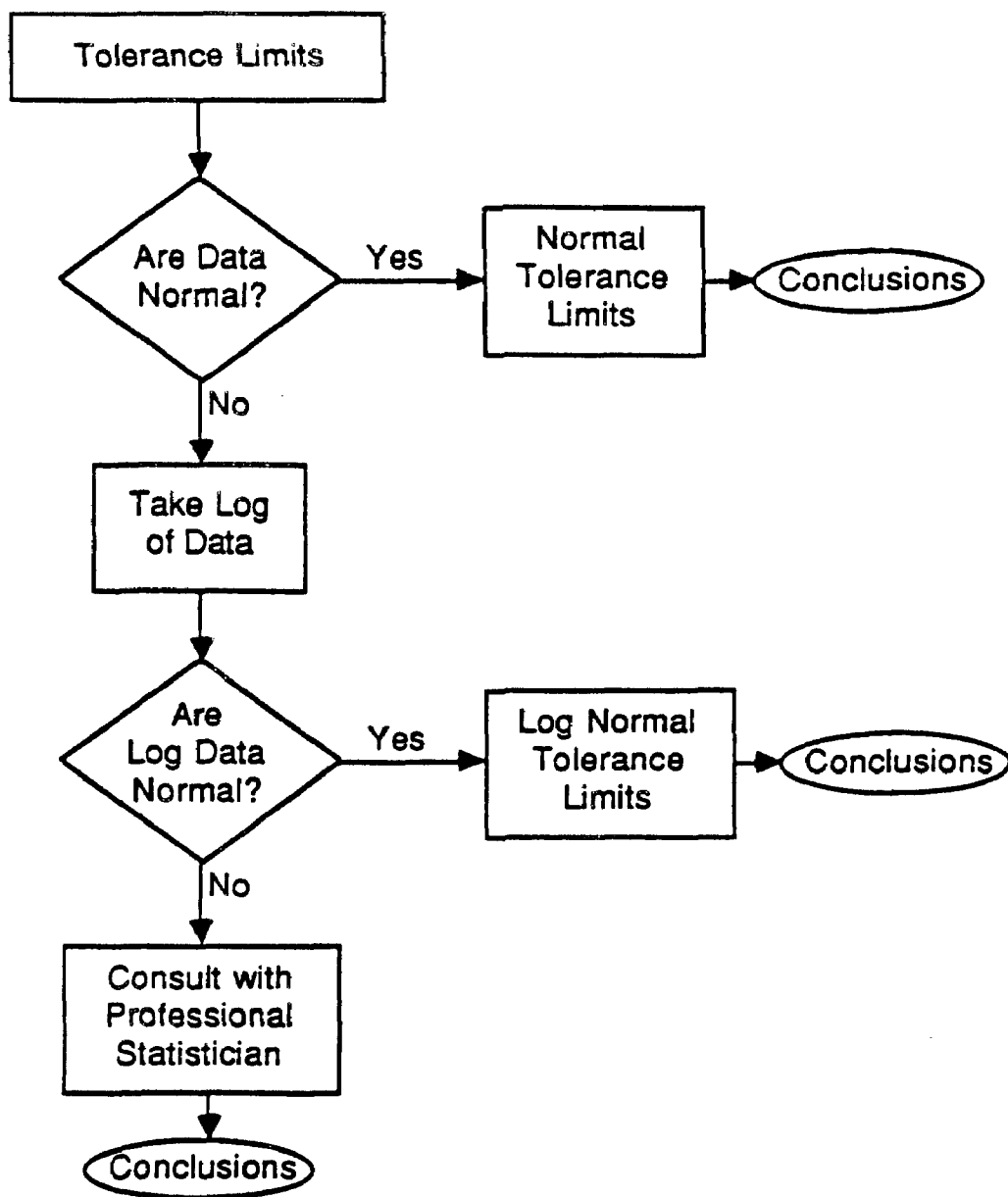


Figure 4-2. Tolerance limits: alternate approach to background well/compliance well comparisons.

4.2 ANALYSIS OF VARIANCE

If contamination of the ground water occurs from the waste disposal facility and if the monitoring wells are hydraulically upgradient and hydraulically downgradient from the site, then contamination is unlikely to change the levels of a constituent in all wells by the same amount. Thus, contamination from a disposal site can be seen as differences in average concentration among wells, and such differences can be detected by analysis of variance.

Analysis of variance (ANOVA) is the name given to a wide variety of statistical procedures. All of these procedures compare the means of different groups of observations to determine whether there are any significant differences among the groups, and if so, contrast procedures may be used to determine where the differences lie. Such procedures are also known in the statistical literature as general linear model procedures.

Because of its flexibility and power, analysis of variance is the preferred method of statistical analysis when the ground-water monitoring is based on a comparison of background and compliance well data. Two types of analysis of variance are presented: parametric and nonparametric one-way analyses of variance. Both methods are appropriate when the only factor of concern is the different monitoring wells at a given sampling period.

The hypothesis tests with parametric analysis of variance usually assume that the errors (residuals) are normally distributed with constant variance. These assumptions can be checked by saving the residuals (the difference between the observations and the values predicted by the analysis of variance model) and using the tests of assumptions presented in Section 3. Since the data will generally be concentrations and since concentration data are often found to follow the lognormal distribution, the log transformation is suggested if substantial violations of the assumptions are found in the analysis of the original concentration data. If the residuals from the transformed data do not meet the parametric ANOVA requirements, then nonparametric approaches to analysis of variance are available using the ranks of the observations. A one-way analysis of variance using the ranks is presented.

When several sampling periods have been used and it is important to consider the sampling periods as a second factor, then two-way analysis of variance, parametric or nonparametric, is appropriate. This would be one way to test for and adjust the data for seasonality. Also, trend analysis (e.g., time series) may be used to identify seasonality in the data set. If necessary, data that exhibit seasonal trends can be adjusted. Usually, however, seasonal variation will affect all wells at a facility by nearly the same amount, and in most circumstances, corrections will not be necessary. Further, the effects of seasonality will be substantially reduced by simultaneously comparing aggregate compliance well data to background well data. Situations that require an analysis procedure other than a one-way ANOVA should be referred to a professional statistician.

4.2.1 One-Way Parametric Analysis of Variance

In the context of ground-water monitoring, two situations exist for which a one-way analysis of variance is most applicable:

- * Data for a water quality parameter are available from several wells but for only one time period (e.g., monitoring has just begun).
- * Data for a water quality parameter are available from several wells for several time periods. However, the data do not exhibit seasonality.

In order to apply a parametric one-way analysis of variance, a minimum number of observations is needed to give meaningful results. At least $p \geq 2$ groups are to be compared (i.e., two or more wells). It is recommended that each group (here, wells) have at least three observations and that the total sample size, N , be large enough so that $N-p \geq 5$. A variety of combinations of groups and number of observations in groups will fulfill this minimum. One sampling interval with four independent samples per well and at least three wells would fulfill the minimum sample size requirements. The wells should be spaced so as to maximize the probability of intercepting a plume of contamination. The samples should be taken far enough apart in time to guard against autocorrelation.

PURPOSE

One-way analysis of variance is a statistical procedure to determine whether differences in mean concentrations among wells, or groups of wells, are statistically significant. For example, is there significant contamination of one or more compliance wells as compared to background wells?

PROCEDURE

Suppose the regulated unit has p wells and that n_i data points (concentrations of a constituent) are available for the i th well. These data can be from either a single sampling period or from more than one. In the latter case, the user could check for seasonality before proceeding by plotting the data over time. Usually the computation will be done on a computer using a commercial program. However, the procedure is presented so that computations can be done using a desk calculator, if necessary.

Step 1. Arrange the $N = \sum_{i=1}^p n_i$ data points in a data table as follows:

	Observations	Well Total	Well Mean
Well No. 1	$X_{11} \dots \dots \dots X_{1n_1}$	$X_{1.}$	$\bar{X}_{1.}$
2	\vdots		
3	\vdots		
\vdots	\vdots		
\vdots	\vdots		
p	$X_{p1} \dots \dots \dots X_{pn_p}$	$X_{p.}$	$\bar{X}_{p.}$
		$X_{..}$	$\bar{X}_{..}$

Step 2. Compute well totals and well means as follows:

$$X_{i.} = \sum_{j=1}^{n_i} X_{ij}, \text{ total of all } n_i \text{ observations at well } i$$

$$\bar{X}_{i.} = \frac{1}{n_i} X_{i.}, \text{ average of all } n_i \text{ observations at well } i$$

$$X_{..} = \sum_{i=1}^p \sum_{j=1}^{n_i} X_{ij}, \text{ grand total of all } n_i \text{ observations}$$

$$\bar{X}_{..} = \frac{1}{N} X_{..}, \text{ grand mean of all observations}$$

These totals and means are shown in the last two columns of the table above.

Step 3. Compute the sum of squares of differences between well means and the grand mean:

$$SS_{\text{Wells}} = \sum_{i=1}^p n_i (\bar{X}_{i.} - \bar{X}_{..})^2 = \sum_{i=1}^p \frac{1}{n_i} X_{i.}^2 - \frac{1}{N} X_{..}^2$$

(The formula on the far right is usually most convenient for calculation.) This sum of squares has $(p-1)$ degrees of freedom associated with it and is a measure of the variability between wells.

Step 4. Compute the corrected total sum of squares

$$SS_{Total} = \sum_{i=1}^p \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{..})^2 = \sum_{i=1}^p \sum_{j=1}^{n_i} X_{ij}^2 - (X_{..}^2/N)$$

(The formula on the far right is usually most convenient for calculation.) This sum of squares has (N-1) degrees of freedom associated with it and is a measure of the variability in the whole data set.

Step 5. Compute the sum of squares of differences of observations within wells from the well means. This is the sum of squares due to error and is obtained by subtraction:

$$SS_{Error} = SS_{Total} - SS_{Wells}$$

It has associated with it (N-p) degrees of freedom and is a measure of the variability within wells.

Step 6. Set up the ANOVA table as shown below in Table 4-1. The sums of squares and their degree of freedom were obtained from Steps 3 through 5. The mean square quantities are simply obtained by dividing each sum of squares by its corresponding degrees of freedom.

TABLE 4-1. ONE-WAY PARAMETRIC ANOVA TABLE

Source of Variation	Sums of squares	Degrees of freedom	Mean squares	F
Between wells	SS_{Wells}	p-1	MS_{Wells} $= SS_{Wells}/(p-1)$	$F = \frac{MS_{Wells}}{MS_{Error}}$
Error (within wells)	SS_{Error}	N-p	MS_{Error} $= SS_{Error}/(N-p)$	
Total	SS_{Total}	N-1		

Step 7. To test the hypothesis of equal means for all p wells, compute $F = MS_{Wells}/MS_{Error}$ (last column in Table 4-1). Compare this statistic to the tabulated F statistic with (p-1) and (N-p) degrees of freedom (Table 2, Appendix B) at the 5% significance level. If the calculated F value exceeds the tabulated value, reject the hypothesis of equal well means. Otherwise, conclude that there is no significant difference between the concentrations at the p wells and thus no evidence of well contamination.

In the case of a significant F (calculated F greater than tabulated F in Step 7), the user will conduct the next few steps to determine which compliance well(s) is (are) contaminated. This will be done by comparing each compliance well with the background well(s). Concentration differences between a pair of background wells and compliance wells or between a compliance well and a set of background wells are called contrasts in the ANOVA and multiple comparisons framework.

Step 8. Determine if the significant F is due to differences between background and compliance wells (computation of Bonferroni t statistics)

Assume that of the p wells, u are background wells and m are compliance wells (thus $u + m = p$). Then m differences--m compliance wells each compared with the average of the background wells--need to be computed and tested for statistical significance. If there are more than five downgradient wells, the individual comparisons are done at the comparisonwise significance level of 1%, which may make the experimentwise significance level greater than 5%.

- Obtain the total sample size of all u background wells.

$$n_{up} = \sum_{i=1}^u n_i$$

- Compute the average concentration from the u background wells.

$$\bar{X}_{up} = \frac{1}{n_{up}} \sum_{i=1}^u X_i.$$

- Compute the m differences between the average concentrations from each compliance well and the average background wells.

$$X_{i.} - \bar{X}_{up}, \quad i = 1, \dots, m$$

- Compute the standard error of each difference as

$$SE_i = [MS_{Error} (1/n_{up} + 1/n_i)]^{1/2}$$

where MS_{Error} is determined from the ANOVA table (Table 4-1) and n_i is the number of observations at well i.

- Obtain the t-statistic $t = t_{(N-p), (1-\alpha/m)}$ from Bonferroni's t-table (Table 3, Appendix B) with $\alpha = 0.05$.
- Compute the m quantities $D_i = SE_i \times t$ for each compliance well i. If $m > 5$ use the entry for $t_{(N-p), (1-0.01)}$. That is, use the entry $m = 5$.

INTERPRETATION

If the difference $\bar{X}_i - \bar{X}_{up}$ exceeds the value D_i , conclude that the i th compliance well has significantly higher concentrations than the average background wells. Otherwise conclude that the well is not contaminated. This exercise needs to be performed for each of the m compliance wells individually. The test is designed so that the overall experimentwise error is 5% if there are no more than five compliance wells.

CAUTIONARY NOTE

Should the regulated unit consist of more than five compliance wells, then the Bonferroni t-test should be modified by doing the individual comparisons at the 1% level so that the Part 264 Subpart F regulatory requirement pursuant to §264.97(i)(2) will be met. Alternately, a different analysis of contrasts, such as Scheffe's, may be used. The more advanced user is referred to the second reference below for a discussion of multiple comparisons.

REFERENCES

Johnson, Norman L., and F. C. Leone. 1977. *Statistics and Experimental Design in Engineering and the Physical Sciences*. Vol. II, Second Edition, John Wiley and Sons, New York.

Miller, Ruppert G., Jr. 1981. *Simultaneous Statistical Inference*. Second Edition, Springer-Verlag, New York.

EXAMPLE

Four lead concentration values at each of six wells are given in Table 4-2 below. The wells consist of $u=2$ background and $m=4$ compliance wells. (The values in Table 4-2 are actually the natural logarithms of the original lead concentrations.)

Step 1. Arrange the $4 \times 6 = 24$ observations in a data table as follows:

TABLE 4-2. EXAMPLE DATA FOR ONE-WAY PARAMETRIC ANALYSIS OF VARIANCE

Well No.	Date:	Log of Pb concentrations ($\mu\text{g/L}$)						
		Jan 1	Feb 1	Mar 1	Apr 1	Well total ($X_{i.}$)	Well mean ($\bar{X}_{i.}$)	Well std. dev.
1	Background wells	4.06	3.99	3.40	3.83	15.28	3.82	0.295
2		3.83	4.34	3.47	4.22	15.86	3.96	0.398
3	Compliance wells	5.61	5.14	3.47	3.97	18.18	4.55	0.996 (max)
4		3.53	4.54	4.26	4.42	16.75	4.19	0.456
5		3.91	4.29	5.50	5.31	19.01	4.75	0.771
6		5.42	5.21	5.29	5.08	21.01	5.25	0.142 (min)

$$X_{..} = 106.08 \quad \bar{X}_{..} = 4.42$$

Step 2. The calculations are shown on the right-hand side of the data table above. Sample standard deviations have been computed also.

Step 3. Compute the between-well sum of squares.

$$SS_{\text{Wells}} = \frac{1}{4} (15.28^2 + \dots + 21.01^2) - \frac{1}{24} \times 106.08^2 = 5.76$$

with $[6 \text{ (wells)} - 1] = 5$ degrees of freedom.

Step 4. Compute the corrected total sum of squares.

$$SS_{\text{Total}} = 4.06^2 + 3.99^2 + \dots + 5.08^2 - \frac{1}{24} \times 106.08^2 = 11.94$$

with $[24 \text{ (observations)} - 1] = 23$ degrees of freedom.

Step 5. Obtain the within-well or error sum of squares by subtraction.

$$SS_{\text{Error}} = 11.94 - 5.76 = 6.18$$

with $[24 \text{ (observations)} - 6 \text{ (wells)}] = 18$ degrees of freedom.

Step 6. Set up the one-way ANOVA as in Table 4-3 below:

TABLE 4-3. EXAMPLE COMPUTATIONS IN ONE-WAY PARAMETRIC ANOVA TABLE

Source of variation	Sums of squares	Degrees of freedom	Mean squares	F
Between wells	5.76	5	$5.76/5 = 1.15$	$1.15/0.34 = 3.38$
Error (within wells)	6.18	18	$6.18/18 = 0.34$	
Total	11.94	23		

Step 7. The calculated F statistic is 3.38. The tabulated F value with 5 and 18 degrees of freedom at the $\alpha = 0.05$ level is 2.77 (Table 2, Appendix B). Since the calculated value exceeds the tabulated value, the hypothesis of equal well means must be rejected, and post hoc comparisons are necessary.

Step 8. Computation of Bonferroni t statistics.

- Note that there are four compliance wells, so $m = 4$ comparisons will be made
- $n_{up} = 8$ total number of samples in background wells
- $\bar{X}_{up} = 3.89$ average concentration of background wells
- Compute the differences between the four compliance wells and the average of the two background wells:

$$\bar{X}_3 - \bar{X}_{up} = 4.55 - 3.89 = 0.66$$

$$\bar{X}_4 - \bar{X}_{up} = 4.19 - 3.89 = 0.3$$

$$\bar{X}_5 - \bar{X}_{up} = 4.75 - 3.89 = 0.86$$

$$\bar{X}_6 - \bar{X}_{up} = 5.25 - 3.89 = 1.36$$

- Compute the standard error of each difference. Since the number of observations is the same for all compliance wells, the standard errors for the four differences will be equal.

$$SE_i = [0.34 (1/8 + 1/4)]^{1/2} = 0.357$$

- From Table 3, Appendix B, obtain the critical t with $(24 - 6) = 18$ degrees of freedom, $m = 4$, and for $\alpha = 0.05$. The approximate value is 2.43 obtained by linear interpolation between 15 and 20 degrees of freedom.
- Compute the quantities D_i . Again, due to equal sample sizes, they will all be equal.

$$D_i = SE_i \times t = 0.357 \times 2.43 = 0.868$$

INTERPRETATION

The F test was significant at the 5% level. The Bonferroni multiple comparisons procedure was then used to determine for which wells there was statistically significant evidence of contamination. Of the four differences $\bar{X}_i - \bar{X}_{up}$, only $\bar{X}_6 - \bar{X}_{up} = 1.36$ exceeds the critical value of 0.868. From this it is concluded that there is significant evidence of contamination at Well 6. Well 5 is right on the boundary of significance. It is likely that Well 6 has intercepted a plume of contamination with Well 5 being on the edge of the plume.

All the compliance well concentrations were somewhat above the mean concentration of the background levels. The well means should be used to indicate the location of the plume. The findings should be reported to the Regional Administrator.

4.2.2 One-Way Nonparametric Analysis of Variance

This procedure is appropriate for interwell comparisons when the data or the residuals from a parametric ANOVA have been found to be significantly different from normal and when a log transformation fails to adequately normalize the data. In one-way nonparametric ANOVA, the assumption under the null hypothesis is that the data from each well come from the same continuous distribution and hence have the same median concentrations of a specific hazardous constituent. The alternatives of interest are that the data from some wells show increased levels of the hazardous constituent in question.

The procedure is called the Kruskal-Wallis test. For meaningful results, there should be at least three groups with a minimum sample size of three in each group. For large data sets use of a computer program is recommended. In the case of large data sets a good approximation to the procedure is to replace each observation by its rank (its numerical place when the data are ordered from least to greatest) and perform the (parametric) one-way analysis of variance (Section 4.2.1) on the ranks. Such an approach can be done with some commercial statistical packages such as SAS.

PURPOSE

The purpose of the procedure is to test the hypothesis that all wells (or groups of wells) around regulated units have the same median concentration of a hazardous constituent. If the wells are found to differ, *post-hoc* comparisons are again necessary to determine if contamination is present.

Note that the wells define the groups. All wells will have at least four observations. Denote the number of groups by K and the number of observations in each group by n_i , with n being the total number of all observations. Let X_{ij} denote the j th observation in the i th group, where j runs from 1 to the number of observations in the group, n_i , and i runs from 1 to the number of groups, K .

PROCEDURE

Step 1. Rank all observations from least to greatest. Let R_{ij} denote the rank of the j th observation in the i th group. As a convention, denote the background well(s) as group 1.

Step 2. Add the ranks of the observations in each group. Call the sum of the ranks for the i th group R_i . Also calculate the average rank for each group, $R_i = R_i/n_i$.

Step 3. Compute the Kruskal-Wallis statistic:

$$H = \left[\frac{12}{n(n+1)} \sum_{i=1}^K \frac{R_i^2}{n_i} \right] - 3(n+1)$$

Step 4. Compare the calculated value H to the tabulated chi-squared value with $(K-1)$ degrees of freedom, where K is the number of groups (Table 1, Appendix B). Reject the null hypothesis if the computed value exceeds the tabulated critical value.

Step 5. If the computed value exceeds the value from the chi-squared table, compute the critical difference for well comparisons to the background, assumed to be group 1:

$$C_i = Z_{(\alpha/(K-1))} \left[\frac{n(n+1)}{12} \right]^{1/2} \left[\frac{1}{n_1} + \frac{1}{n_i} \right],$$

for i taking values 2, ..., K ,

where $Z_{(\alpha/(K-1))}$ is the upper $(\alpha/(K-1))$ percentile from the standard normal distribution found in Table 4, Appendix B. Note: If there are more than five compliance wells at the regulated unit ($k > 6$), use $Z_{.01}$, the upper one-percentile from the standard normal distribution.

Step 6. Form the differences of the average ranks for each group to the background and compare these with the critical values found in step 5 to determine which wells give evidence of contamination. That is, compare $R_i - R_1$ to C_i for i taking the values 2 through K . (Recall that group 1 is the background.)

While the above steps are the general procedure, some details need to be specified further to handle special cases. First, it may happen that two or more observations are numerically equal or tied. When this occurs, determine the ranks that the tied observations would have received if they had been slightly different from each other, but still in the same places with respect to the rest of the observations. Add these ranks and divide by the number of observations tied at that value to get an average rank. This average rank is used for each of the tied observations. This same procedure is repeated for any other groups of tied observations.

The effect of tied observations is to increase the value of the statistic, H . Unless there are many observations tied at the same value, the effect of ties is negligible (in practice, the effect of ties can probably be neglected unless some group contains 10 percent of the observations all tied, which is most likely to occur for concentrations below detection limit). If the statistic is significant without any correction for ties, it will also be significant if corrected for ties.

Second, if there are any values below detection, consider all values below detection as tied at zero. (It is irrelevant what number is assigned to nondetected values as long as all such values are assigned the same number, and it is smaller than any detected or quantified value.)

ADJUSTMENT FOR TIES

If there are 50% or more observations that fell below the detection limit, then this method is inappropriate. The user is referred to Section 7, "Miscellaneous Topics." Otherwise, if there are tied values present in the data, use the following correction for the H statistic

$$H' = \frac{H}{1 - \left(\frac{\sum_{i=1}^g T_i^3 / (n^3 - n)}{n} \right)}$$

where g = the number of groups of distinct tied observations and $T_i = (t_i^3 - t_i)$, where t_i is the number of observations in the tied group i . Note that unique observations can be considered groups of size 1, with the corresponding $T_i = (1^3 - 1) = 0$.

REFERENCE

Hollander, Myles, and D. A. Wolfe. 1973. *Nonparametric Statistical Methods*. John Wiley and Sons, New York.

EXAMPLE

The data in Table 4-4 represent benzene concentrations in water samples taken at one background and five compliance wells.

Step 1. The 20 observations have been ranked from least to greatest. The limit of detection was 1.0 ppm. Note that two values in Well 4 were below detection and were assigned value zero. These two are tied for the smallest value and have consequently been assigned the average of the two ranks 1 and 2, or 1.5. The ranks of the observations are indicated in parentheses after the observation in Table 4-4. Note that there are 3 observations tied at 1.3 that would have had ranks 4, 5, and 6 if they had been slightly different. These three have been assigned the average rank of 5 resulting from averaging 4, 5, and 6. Other ties occurred at 1.5 (ranks 7 and 8) and 1.9 (ranks 11 and 12).

Step 2. The values of the sums of ranks and average ranks are indicated at the bottom of Table 4-4.

Step 3. Compute the Kruskal-Wallis statistic

$$H = \frac{12}{20(20-1)} (34^2/4 + \dots + 35.5^2/3) - 3(20+1) = 14.68$$

TABLE 4-4. EXAMPLE DATA FOR ONE-WAY NONPARAMETRIC ANOVA--BENZENE CONCENTRATIONS (in ppm)

Date	Background Well 1	Compliance wells				
		Well 2	Well 3	Well 4	Well 5	Well 6
Jan 1	1.7 (10)	11.0 (20)	1.3 (5)	0 (1.5)	4.9 (17)	1.6 (9)
Feb 1	1.9 (11.5)	8.0 (18)	1.2 (3)	1.3 (5)	3.7 (16)	2.5 (15)
Mar 1	1.5 (7.5)	9.5 (19)	1.5 (7.5)	0 (1.5)	2.3 (14)	1.9 (11.5)
Apr 1	1.3 (5)			2.2 (13)		

	$n_1 = 4$	$n_2 = 3$	$n_3 = 3$	$n_4 = 4$	$n_5 = 3$	$n_6 = 3$
Sum of ranks:	$R_1 = 34$	$R_2 = 57$	$R_3 = 15.5$	$R_4 = 21$	$R_5 = 47$	$R_6 = 35.5$
Average rank:	$\bar{R}_1 = 8.5$	$\bar{R}_2 = 19$	$\bar{R}_3 = 5.17$	$\bar{R}_4 = 5.25$	$\bar{R}_5 = 15.67$	$\bar{R}_6 = 11.83$

$K = 6$, the number of wells

$n = \sum_{i=1}^6 n_i = 20$, the total number of observations.

ADJUSTMENT FOR TIES

There are four groups of ties in the data of Table 4-4:

$$\begin{aligned}T_1 &= (2^3-2) = 6 && \text{for the 2 observations of 1,900.} \\T_2 &= (2^3-2) = 6 && \text{for the 2 observations of 1,500.} \\T_3 &= (3^3-3) = 24 && \text{for the 3 observations of 1,300.} \\T_4 &= (2^3-2) = 6 && \text{for the 2 observations of 0.}\end{aligned}$$

$$\text{Thus } \sum_{i=1}^4 T_i = 6+6+24+6 = 42$$

$$\text{and } H' = \frac{14.68}{1-(42/(20^3-20))} = \frac{14.68}{0.995} = 14.76, \text{ a negligible change from 14.68.}$$

Step 4. To test the null hypothesis of no contamination, obtain the critical chi-squared value with $(6-1) = 5$ degrees of freedom at the 5% significance level from Table 1, Appendix B. The value is 11.07. Compare the calculated value, H' , with the tabulated value. Since 14.76 is greater than 11.07, reject the hypothesis of no contamination at the 5% level. If the site was in detection monitoring it should move into compliance monitoring. If the site was in compliance monitoring it should move into corrective action. If the site was in corrective action it should stay there.

In the case where the hydraulically upgradient wells serve as the background against which the compliance wells are to be compared, comparisons of each compliance well with the background wells should be performed in addition to the analysis of variance procedure. In this example, data from each of the compliance wells would be compared with the background well data. This comparison is accomplished as follows. The average ranks for each group, R_i , are used to compute differences. If a group of compliance wells for a regulated unit have larger concentrations than those found in the background wells, the average rank for the compliance wells at that unit will be larger than the average rank for the background wells.

Step 5. Calculate the critical values to compare each compliance well to the background well.

In this example, $K=6$, so there are 5 comparisons of the compliance wells with the background wells. Using an experimentwise significance level of $\alpha = 0.05$, we find the upper $0.05/5 = 0.01$ percentile of the standard normal distribution to be 2.33 (Table 4, Appendix B). The total sample size, n , is 20. The approximate critical value, C_2 , is computed for compliance Well 2, which has the largest average rank, as:

$$C_2 = 2.32 \left[\frac{20(21)}{12} \right]^{1/2} \left[\frac{1}{4} + \frac{1}{3} \right]^{1/2} = 10.5$$

The critical values for the other wells are: 10.5 for Wells 3, 5, and 6; and 9.8 for Well 4.

Step 6. Compute the differences between the average rank of each compliance well and the average rank of the background well:

<u>Differences</u>	<u>Critical values</u>
$D_2 = 19.0 - 8.5 = 10.5$	$C_2 = 10.5$
$D_3 = 5.17 - 8.5 = -3.33$	$C_3 = 10.5$
$D_4 = 5.25 - 8.5 = -3.25$	$C_4 = 9.8$
$D_5 = 15.67 - 8.5 = 7.17$	$C_5 = 10.5$
$D_6 = 11.83 - 8.5 = 3.13$	$C_6 = 10.5$

Compare each difference with the corresponding critical difference. $D_2 = 10.5$ equals the critical value of $C_2 = 10.5$. We conclude that the concentration of benzene averaged over compliance Well 2 is significantly greater than that at the background well. None of the other compliance well concentration of benzene is significantly higher than the average background value. Based upon these results, only compliance Well 2 can be singled out as being contaminated.

For data sets with more than 30 observations, the parametric analysis of variance performed on the rank values is a good approximation to the Kruskal-Wallis test (Quade, 1966). If the user has access to SAS, the PROC RANK procedure is used to obtain the ranks of the data. The analysis of variance procedure detailed in Section 4.2.1 is then performed on the ranks. Contrasts are tested as in the parametric analysis of variance.

INTERPRETATION

The Kruskal-Wallis test statistic is compared to the tabulated critical value from the chi-squared distribution. If the test statistic does not exceed the tabulated value, there is no statistically significant evidence of contamination and the analysis would stop and report this finding. If the test statistic exceeds the tabulated value, there is significant evidence that the hypothesis of no differences in compliance concentrations from the background level is not true. Consequently, if the test statistic exceeds the critical value, one concludes that there is significant evidence of contamination. One then proceeds to investigate where the differences lie, that is, which wells are indicating contamination.

The multiple comparisons procedure described in steps 5 and 6 compares each compliance well to the background well. This determines which compliance wells show statistically significant evidence of contamination at an experimentwise error rate of 5 percent. In many cases, inspection of the mean or median concentrations will be sufficient to indicate where the problem lies.

4.3 TOLERANCE INTERVALS BASED ON THE NORMAL DISTRIBUTION

An alternative approach to analysis of variance to determine whether there is statistically significant evidence of contamination is to use tolerance intervals. A tolerance interval is constructed from the data on

(uncontaminated) background wells. The concentrations from compliance wells are then compared with the tolerance interval. With the exception of pH, if the compliance concentrations do not fall in the tolerance interval, this provides statistically significant evidence of contamination.

Tolerance intervals are most appropriate for use at facilities that do not exhibit high degrees of spatial variation between background wells and compliance wells. Facilities that overlie extensive, homogeneous geologic deposits (for example, thick, homogeneous lacustrine clays) that do not naturally display hydrogeochemical variations may be suitable for this statistical method of analysis.

A tolerance interval establishes a concentration range that is constructed to contain a specified proportion of at least $P\%$ of the population with a specified confidence coefficient, Y . The proportion of the population included, P , is referred to as the coverage. The probability with which the tolerance interval includes the proportion $P\%$ of the population is referred to as the tolerance coefficient.

A coverage of 95% is recommended. If this is used, random observations from the same distribution as the background well data would exceed the upper tolerance limit less than 5% of the time. Similarly, a tolerance coefficient of 95% is recommended. This means that one has a confidence level of 95% that the upper 95% tolerance limit will contain at least 95% of the distribution of observations from background well data. These values were chosen to be consistent with the performance standards described in Section 2. The use of these values corresponds to the selection of α of 5% in the multiple well testing situation.

The procedure can be applied with as few as three observations from the background distribution. However, doing so would result in a large upper tolerance limit. A sample size of eight or more results is an adequate tolerance interval. The minimum sampling schedule called for in the regulations would result in at least four observations from each background well. Only if a single background well is sampled at a single point in time is the sample size so small as to make use of the procedure questionable.

Tolerance intervals can be constructed assuming that the data or the transformed data are normal. Tolerance intervals can also be constructed assuming other distributions. It is also possible to construct nonparametric tolerance intervals using only the assumption that the data came from some continuous population. However, the nonparametric tolerance intervals require such a large number of observations to provide a reasonable coverage and tolerance coefficient that they are impractical in this application.

The range of the concentration data in the background well samples should be considered in determining whether the tolerance interval approach should be used, and if so, what distribution is appropriate. The background well concentration data should be inspected for outliers and tests of normality applied before selecting the tolerance interval approach. Tests of normality were presented in Section 3.2. Note that in this case, the test of normality would be applied to the background well data that are used to construct the

tolerance interval. These data should all be from the same normal distribution.

In this application, unless pH is being monitored, a one-sided tolerance interval or an upper tolerance limit is desired, since contamination is indicated by large concentrations of the hazardous constituents monitored. Thus, for concentrations, the appropriate tolerance interval is (0, TL), with the comparison of importance being the larger limit, TL.

PURPOSE

The purpose of the tolerance interval approach is to define a concentration range from background well data, within which a large proportion of the monitoring observations should fall with high probability. Once this is done, data from compliance wells can be checked for evidence of contamination by simply determining whether they fall in the tolerance interval. If they do not, this is evidence of contamination.

In this case the data are assumed to be approximately normally distributed. Section 3.2 provided methods to check for normality. If the data are not normal, take the natural logarithm of the data and see if the transformed data are approximately normal. If so, this method can be used on the logarithms of the data. Otherwise, seek the assistance of a professional statistician.

PROCEDURE

Step 1. Calculate the mean, \bar{X} , and the standard deviation, S , from the background well data.

Step 2. Construct the one-sided upper tolerance limit as

$$TL = \bar{X} + K S,$$

where K is the one-sided normal tolerance factor found in Table 5, Appendix B.

Step 3. Compare each observation from compliance wells to the tolerance limit found in Step 2. If any observation exceeds the tolerance limit, that is statistically significant evidence that the well is contaminated. Note that if the tolerance interval was constructed on the logarithms of the original background observations, the logarithms of the compliance well observations should be compared to the tolerance limit. Alternatively the tolerance limit may be transferred to the original data scale by taking the anti-logarithm.

REFERENCE

Lieberman, Gerald J. 1958. "Tables for One-sided Statistical Tolerance Limits." *Industrial Quality Control*. Vol. XIV, No. 10.

EXAMPLE

Table 4-5 contains example data that represent lead concentration levels in parts per million in water samples at a hypothetical facility. The background well data are in columns one and two, while the other four columns represent compliance well data.

Step 1. The mean and standard deviation of the $n = 8$ observations have been calculated for the background well. The mean is 51.4 and the standard deviation is 16.3.

Step 2. The tolerance factor for a one-sided normal tolerance interval is found from Table 5, Appendix B as 3.188. This is for 95% coverage with probability 95% and for $n = 8$. The upper tolerance limit is then calculated as $51.4 + (3.188)(16.3) = 103.4$.

Step 3. The tolerance limit of 103.3 is compared with the compliance well data. Any value that exceeds the tolerance limit indicates evidence of contamination. Two observations from Well 1, two observations from Well 3, and all four observations from Well 4 exceed the tolerance limit. Thus there is evidence of contamination at Wells 1, 3, and 4.

INTERPRETATION

A tolerance limit with 95% coverage gives an upper bound below which 95% of the observations of the distribution should fall. The tolerance coefficient used here is 95%, implying that at least 95% of the observations should fall below the tolerance limit with probability 95%, if the compliance well data come from the same distribution as the background data. In other words, in this example, we are 95% certain that 95% of the background lead concentrations are below 104 ppm. If observations exceed the tolerance limit, this is evidence that the compliance well data are not from the same distribution, but rather are from a distribution with higher concentrations. This is interpreted as evidence of contamination.

4.4 PREDICTION INTERVALS

A prediction interval is a statistical interval calculated to include one or more future observations from the same population with a specified confidence. In ground-water monitoring, a prediction interval approach may be used to make comparisons between background and compliance well data. The concentrations of a hazardous constituent in the background wells are used to establish an interval within which K future observations from the same population are expected to lie with a specified confidence. Then each of K future observations of compliance well concentrations is compared to the prediction interval. The interval is constructed to contain all of K future observations with the stated confidence. If any future observation exceeds the prediction interval, this is statistically significant evidence of contamination. In application, the number of future observations to be collected, K , must be specified. Thus, the prediction interval is constructed for a specified time period in the future. One year is suggested. The interval can be constructed either to contain all K individual observations with a specified probability, or to contain the K means observed at the K sampling periods.

TABLE 4-5. EXAMPLE DATA FOR NORMAL TOLERANCE INTERVAL

Lead concentrations (ppm)						
Date	Background well		Compliance wells			
	A	B	Well 1	Well 2	Well 3	Well 4
Jan 1	58.0	46.1	273.1*	34.1	49.9	225.9*
Feb 1	54.1	76.7	170.7*	93.7	73.0	183.1*
Mar 1	30.0	32.1	32.1	70.8	244.7*	198.3*
Apr 1	46.1	68.0	53.0	83.1	202.4*	160.8*

n = 8 The upper 95% coverage tolerance limit
 Mean = 51.4 with tolerance coefficient of 95% is
 SD = 16.3 $51.4 + (3.188)(16.3) = 103.4$

* Indicates contamination

The prediction interval presented here is constructed assuming that the background data all follow the same normal distribution. If that is not the case (See Section 3.2 for tests of normality), but a log transformation results in data that are adequately normal on the log scale the interval may still be used. In this case, use the data after transforming by taking the logarithm. The future observations need to also be transformed by taking logarithms before comparison to the interval. (Alternatively, the end points of the interval could be converted back to the original scale by taking their anti-logarithms.)

PURPOSE

The prediction interval is constructed so that K future compliance well observations can be tested by determining whether they lie in the interval or not. If not, evidence of contamination is found. Note that the number of future observations, K, for which the interval is to be used, must be specified in advance. In practice, an owner or operator would need to construct the prediction interval on a periodic (at least yearly) basis, using the most recent background data. The interval is described using the 99% confidence factor appropriate for individual well comparisons. It is recommended that a one-sided prediction interval be constructed for the mean of the four observations from each compliance well at each sampling period.

PROCEDURE

Step 1. Calculate the mean, \bar{X} , and the standard deviation, S, for the background well data (used to form the prediction interval).

Step 2. Specify the number of future observations for a compliance well to be included in the interval, K. Then the interval is given by

$$\bar{X} + S \sqrt{1/m + 1/n} \quad t_{(n-1, K, 0.95)}$$

where it is assumed that the mean of the m observations taken at the K sampling periods will be used. Here n is the number of observations in the background data, and $t_{(n-1, K, 0.95)}$ is found from Table 3 in Appendix B. The

table is entered with K as the number of future observations, and degrees of freedom, $v = n-1$. If $K > 5$, use the column for $K = 5$.

Step 3. Once the interval has been calculated, at each sampling period, the mean of the m compliance well observations is obtained. This mean is compared to see if it falls in the interval. If it does, this is reported and monitoring continues. If a mean concentration at a sampling period does not fall in the prediction interval, this is statistically significant evidence of contamination. This is also reported and the appropriate action taken.

REMARK

For a single future observation, t is given by the t -distribution found in Table 6 of Appendix B. In general, the interval to contain K future means of sample size m each is given by

$$\bar{X} + S \sqrt{1/m + 1/n} \quad t_{(n-1, K, 0.95)}$$

where t is as before from Table 3 of Appendix B and where m is the number of observations in each mean. Note that for K single observations, $m=1$, while for the mean of four samples from a compliance well, $m=4$.

Note, too, that the prediction intervals are one-sided, giving a value that should not be exceeded by the future observations. The 5% experimentwise significance level is used with the Bonferroni approach. However, to ensure that the significance level for the individual comparisons does not go below 1%, α/K is restricted to be 1% or larger. If more than K comparisons are used, the comparisonwise significance level of 1% is used, implying that the comparisonwise level may exceed 5%.

EXAMPLE

Table 4-6 contains chlordane concentrations measured at a hypothetical facility. Twenty-four background observations are available and are used to develop the prediction interval. The prediction interval is applied to $K=2$ sampling periods with $m=4$ observations at a single compliance well each.

Step 1. Find the mean and standard deviation of the 24 background well measurements. These are 101 and 11, respectively.

Step 2. There are $K = 2$ future observations of means of 4 observations to be included in the prediction interval. Entering Table 3 of Appendix B at $K = 2$ and 20 degrees of freedom (the nearest entry to the 23 degrees of freedom), we find $t_{(20, 2, 0.95)} = 2.09$. The interval is given by

$$101 + (11)2.09(1/4 + 1/24)^{1/2} = (0, 113.4).$$

Step 3. The mean of each of the four compliance well observations at sampling period one and two is found and compared with the interval found in Step 2. The mean of the first sampling period is 122 and that for the second sampling period is 113. Comparing the first of these to the prediction interval for two means based on samples of size 4, we find that the mean exceeds the upper limit of the prediction interval. This is statistically significant evidence of contamination and should be reported to the Regional Administrator. Since the second sampling period mean is within the prediction interval, the Regional Administrator may allow the facility to remain in its current stage of monitoring.

TABLE 4-6. EXAMPLE DATA FOR PREDICTION INTERVAL--CHLORDANE LEVELS

Background well data--Well 1		Compliance well data--Well 2	
Sampling date	Chlordane concentration (ppb)	Sampling date	Chlordane concentration (ppb)
January 1, 1985	97 103 104 85	July 1, 1986	123 120 116 <u>128</u>
April 1, 1985	120 105 104 108	n = 4 Mean = 122 SD = 5	
July 1, 1985	110 95 102 78	October 1, 1986	116 117 119 <u>101</u>
October 1, 1985	105 94 110 111	n = 4 Mean = 113 SD = 8	
January 1, 1986	80 106 115 105		
April 1, 1986	100 93 89 <u>113</u>		
	n = 24 Mean = 101 SD = 11		

INTERPRETATION

A prediction interval is a statistical interval constructed from background sample data to contain a specified number of future observations from the same distribution with specified probability. That is, the prediction interval is constructed so as to have a 95% probability of containing the next K sampling period means, provided that there is no contamination. If the future observations are found to be in the prediction interval, this is evidence that there has been no change at the facility and that no contamination is occurring. If the future observation falls outside of the prediction interval, this is statistical evidence that the new observation does not come from the same distribution, that is, from the population of uncontaminated water samples previously sampled. Consequently, if the observation is a concentration above the prediction interval's upper limit, it is evidence of contamination.

The prediction interval could be constructed in several ways. It can be developed for means of observations at each sampling period, or for each individual observation at each sampling period.

It should also be noted that the estimate of the standard deviation, S , that is used should be an unbiased estimator. The usual estimator, presented above, assumes that there is only one source of variation. If there are other sources of variation, such as time effects, or spatial variation in the data used for the background, these should be included in the estimate of the variability. This can be accomplished by use of an appropriate analysis of variance model to include the other factors affecting the variability. Determination of the components of variance in complicated models is beyond the scope of this document and requires consultation with a professional statistician.

REFERENCE

Gibbons, Robert D. 1987. "Statistical Prediction Intervals for the Evaluation of Ground-Water Quality." *Ground Water*. Vol. 25, pp. 455-465.

SECTION 5

COMPARISONS WITH MCLs OR ACLs

This section includes statistical procedures appropriate when the monitoring aims at determining whether ground-water concentrations of hazardous constituents are below or above fixed concentration limits. In this situation the maximum concentration limit (MCL) or alternate concentration limit (ACL) is a specified concentration limit rather than being determined by the background well concentrations. Thus the applicable statistical procedures are those that compare the compliance well concentrations estimated from sampling with the prespecified fixed limits. Methods for comparing compliance well concentrations to a (variable) background concentration were presented in Section 4.

The methods applicable to the type of comparisons described in this section include confidence intervals and tolerance intervals. A special section deals with cases where the observations exhibit very small or no variability.

5.1 SUMMARY CHART FOR COMPARISON WITH MCLs OR ACLs

Figure 5-1 is a flow chart to aid the user in selecting and applying a statistical method when the permit specifies an MCL or ACL.

As with each type of comparison, a determination is made first to see if there is enough data for intra-well comparisons. If so, these should be done in parallel with the other comparisons.

Here, whether the compliance limit is a maximum concentration limit (MCL) or an alternate concentration limit (ACL), the recommended procedure to compare the mean compliance well concentration against the compliance limit is the construction of a confidence interval. This approach is presented in Section 5.2.1. Section 5.2.2 adds a special case of limited variance in the data. If the permit requires that a compliance limit is not to be exceeded more than a specified fraction of the time, then the construction of tolerance limits is the recommended procedure, discussed in Section 5.2.3.

5.2 STATISTICAL PROCEDURES

This section presents the statistical procedures appropriate for comparison of ground-water monitoring data to a constant compliance limit, a fixed standard. The interpretation of the fixed compliance limit (MCL or ACL) is that the mean concentration should not exceed this fixed limit. An alternative interpretation may be specified. The permit could specify a compliance limit as a concentration not to be exceeded by more than a small, specified

Comparisons with MCL/ACLs

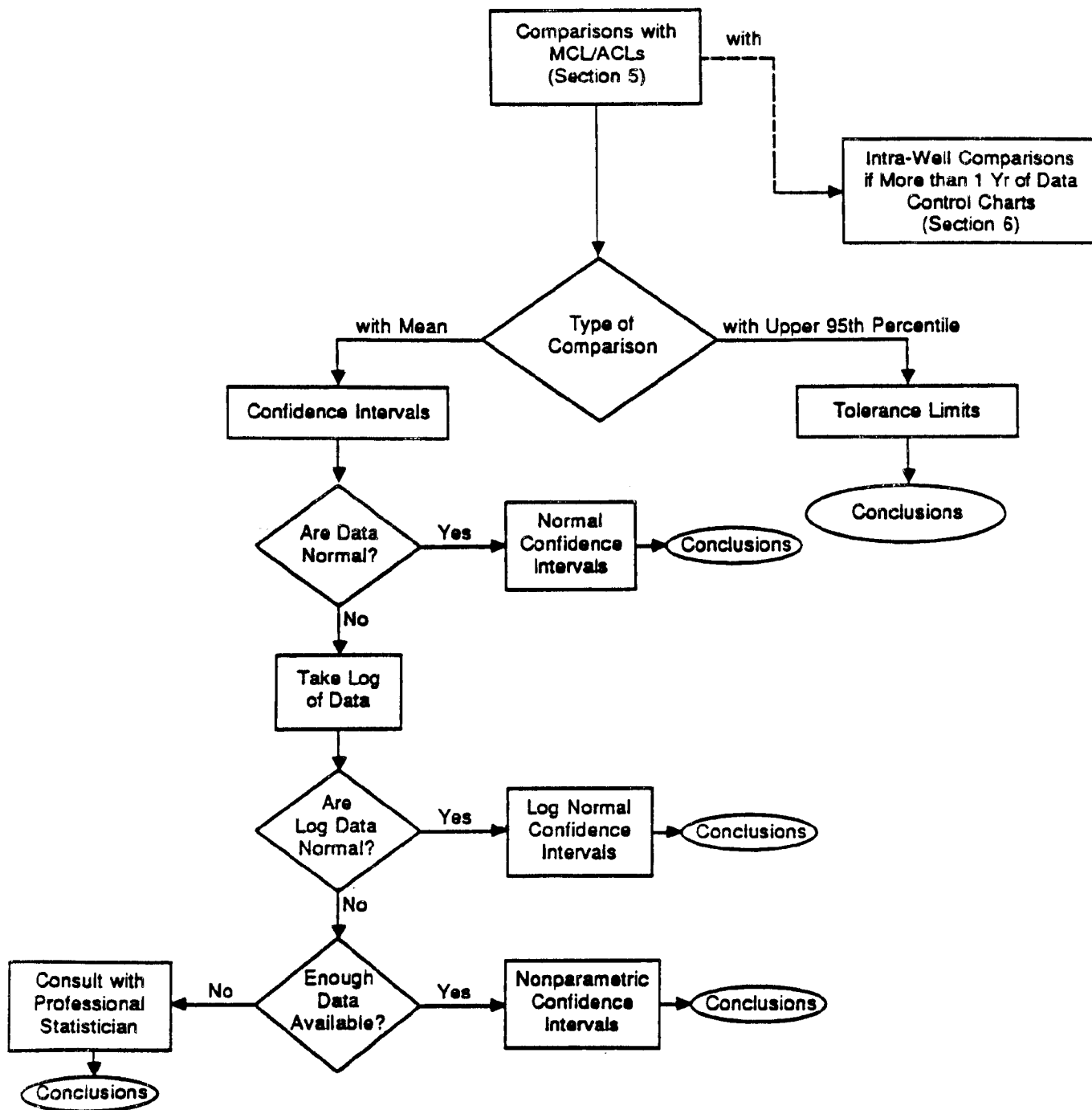


Figure 5-1. Comparisons with MCLs/ACLs.

proportion of the observations. A tolerance interval approach for such a situation is also presented.

5.2.1 Confidence Intervals

When a regulated unit is in compliance monitoring with a fixed compliance limit (either an MCL or an ACL), confidence intervals are the recommended procedure pursuant to §264.97(h)(5) in the Subpart F regulations. The unit will remain in compliance monitoring unless there is statistically significant evidence that the mean concentration at one or more of the downgradient wells exceeds the compliance limit. A confidence interval for the mean concentration is constructed from the sample data for each compliance well individually. These confidence intervals are compared with the compliance limit. If the entire confidence interval exceeds the compliance limit, this is statistically significant evidence that the mean concentration exceeds the compliance limit.

Confidence intervals can generally be constructed for any specified distribution. General methods can be found in texts on statistical inference some of which are referenced in Appendix C. A confidence limit based on the normal distribution is presented first, followed by a modification for the log-normal distribution. A nonparametric confidence interval is also presented.

5.2.1.1 Confidence Interval Based on the Normal Distribution

PURPOSE

The confidence interval for the mean concentration is constructed from the compliance well data. Once the interval has been constructed, it can be compared with the MCL or ACL by inspection to determine whether the mean concentration significantly exceeds the MCL or ACL.

PROCEDURE

Step 1. Calculate the mean, \bar{X} , and standard deviation, S , of the sample concentration values. Do this separately for each compliance well.

Step 2. For each well calculate the confidence interval as

$$\bar{X} \pm t_{(0.99, n-1)} S/\sqrt{n}$$

where $t_{(0.99, n-1)}$ is obtained from the t-table (Table 6, Appendix B). Generally, there will be at least four observations at each sampling period, so t will usually have at least 3 degrees of freedom.

Step 3. Compare the intervals calculated in Step 2 to the compliance limit (the MCL or ACL, as appropriate). If the compliance limit is contained in the interval or is above the upper limit, the unit remains in compliance.

If any well confidence interval's lower limit exceeds the compliance limit, this is statistically significant evidence of contamination.

REMARK

The 99th percentile of the t-distribution is used in constructing the confidence interval. This is consistent with an alpha (probability of Type I error) of 0.01, since the decision on compliance is made by comparing the lower confidence limit to the MCL or ACL. Although the interval as constructed with both upper and lower limits is a 98% confidence interval, the use of it is one-sided, which is consistent with the 1% alpha level of individual well comparisons.

EXAMPLE

Table 5-1 lists hypothetical concentrations of Aldicarb in three compliance wells. For illustration purposes, the MCL for Aldicarb has been set at 7 ppb. There is no evidence of nonnormality, so the confidence interval based on the normal distribution is used.

TABLE 5-1. EXAMPLE DATA FOR NORMAL CONFIDENCE INTERVAL--ALDICARB CONCENTRATIONS IN COMPLIANCE WELLS (ppb)

Sampling date	Well 1	Well 2	Well 3
Jan. 1	19.9	23.7	5.6
Feb. 1	29.6	21.9	3.3
Mar. 1	18.7	26.9	2.3
Apr. 1	24.2	26.1	6.9
$\bar{X} =$	23.1	24.6	4.5
$S =$	4.9	2.3	2.1

MCL = 7 ppb

Step 1. Calculate the mean and standard deviation of the concentrations for each compliance well. These statistics are shown in the table above.

Step 2. Obtain the 99th percentile of the t-distribution with $(4-1) = 3$ degrees of freedom from Table 6, Appendix B as 4.541. Then calculate the confidence interval for each well's mean concentration.

$$\text{Well 1: } 23.1 \pm 4.541(4.9)/\sqrt{4} = (12.0, 34.2)$$

$$\text{Well 2: } 24.6 \pm 4.541(2.3)/\sqrt{4} = (19.4, 29.8)$$

$$\text{Well 3: } 4.5 \pm 4.541(2.1)/\sqrt{4} = (-0.3, 9.3)$$

where the usual convention of expressing the upper and lower limits of the confidence interval in parentheses separated by a comma has been followed.

Step 3. Compare each confidence interval to the MCL of 7 ppb. When this is done, the confidence interval for Well 1 lies entirely above the MCL of 7, indicating that the mean concentration of Aldicarb in Well 1 significantly exceeds the MCL. Similarly, the confidence interval for Well 2 lies entirely above the MCL of 7. This is significant evidence that the mean concentration in Well 2 exceeds the MCL. However, the confidence interval for Well 3 is mostly below the MCL. Thus, there is no evidence that the mean concentration in Well 3 exceeds the MCL.

INTERPRETATION

The confidence interval is an interval constructed so that it should contain the true or population mean with specified confidence (98% in this case). If this interval does not contain the compliance limit, then the mean concentration must differ from the compliance limit. If the lower end of the interval is above the compliance limit, then the mean concentration must be significantly greater than the compliance limit, indicating noncompliance.

5.2.1.2 Confidence Interval for Log-Normal Data

PURPOSE

The purpose of a confidence interval for the mean concentration is to determine whether there is statistically significant evidence that the mean concentration exceeds a fixed compliance limit. The interval gives a range that includes the true mean concentration with confidence 98%. The lower limit will be below the true mean with confidence 99%, corresponding to an alpha of 1%.

PROCEDURE

This procedure is used to construct a confidence interval for the mean concentration from the compliance well data when the data are log-normal (that is, when the logarithms of the data are normal). Once the interval has been constructed, it can be compared with the MCL or ACL by inspection to determine whether the mean concentration significantly exceeds the MCL or ACL.

Step 1. Take the natural logarithm of each data point (concentration measurement). Also, take the natural logarithm of the compliance limit.

Step 2. Calculate the sample mean and standard deviation of the data (on the log scale) from each compliance well. (This is Step 1 of the previous section, working now on the logarithms.)

Step 3. Form the confidence intervals for each compliance well as

$$\bar{X} \pm t_{(0.99, n-1)} S/\sqrt{n}$$

where $t(0.99, n-1)$ is from the t-distribution in Table 6 of Appendix B. Here t will typically have 3 degrees of freedom.

Step 4. Compare the confidence intervals found in Step 3 to the logarithm of the compliance limit found in Step 1. If any of the intervals lies entirely above the logarithm of the compliance limit, there is evidence that the unit is out of compliance. Otherwise, the unit is in compliance.

EXAMPLE

Table 5-2 contains EDB concentration data from three compliance wells at a hypothetical site. The MCL is assumed to be 20 ppb. For demonstration purposes, the data are assumed not normal; a log-transformation normalized them adequately. The lower part of the table contains the natural logarithms of the concentrations.

Step 1. The logarithms of the data are used to calculate a confidence interval. Take the natural log of the concentrations in the top part of Table 5-2 to find the values given in the lower part of the table. For example, $\ln(24.2) = 3.19$, . . . , $\ln(25.3) = 3.23$. Also, take the logarithm of the MCL to find that $\ln(20) = 2.99$.

TABLE 5-2. EXAMPLE DATA FOR LOG-NORMAL CONFIDENCE INTERVAL--EDB CONCENTRATIONS IN COMPLIANCE WELLS (ppb)

Sampling date	Well 1	Well 2	Well 3
Jan. 1	24.2	39.7	55.7
Apr. 1	10.2	75.7	17.0
Jul. 1	17.4	60.2	97.8
Oct. 1	39.7	10.9	25.3
$\bar{X} =$	22.9	46.6	49.0
$S =$	12.6	28.0	36.6

MCL = 20 ppb

Log concentrations

Jan. 1	3.19	3.68	4.02
Apr. 1	2.32	4.33	2.84
Jul. 1	2.85	4.10	4.58
Oct. 1	3.68	2.39	3.23
$\bar{X} =$	3.01	3.62	3.67
$S =$	0.57	0.86	0.78

Log (MCL) = 2.99

Step 2. Calculate the mean and standard deviation of the log concentrations for each compliance well. These are shown in the table.

Step 3. Form the confidence intervals for each compliance well.

$$\text{Well 1: } 3.01 \pm 4.541(0.57)/\sqrt{4} = (1.72, 4.30)$$

$$\text{Well 2: } 3.62 \pm 4.541(0.86)/\sqrt{4} = (1.67, 5.57)$$

$$\text{Well 3: } 3.67 \pm 4.541(0.78)/\sqrt{4} = (1.90, 5.44)$$

where 4.541 is the value obtained from the t-table (Table 6 in Appendix B) as in the previous example.

Step 4. Compare the individual well confidence intervals with the MCL (expressed on the log scale). The natural log of the MCL of 20 ppm is 2.99. None of the individual well confidence intervals for the mean has a lower limit that exceeds this value, so none of the individual well mean concentrations is significantly different from the MCL.

Note: The lower and upper limits of the confidence interval for each well's mean concentration could be converted back to the original scale by taking antilogs. For example, on the original scale, the confidence intervals would be:

$$\text{Well 1: } (\exp(1.72), \exp(4.30)) \text{ or } (5.58, 73.70)$$

$$\text{Well 2: } (\exp(1.67), \exp(5.51)) \text{ or } (5.31, 262.43)$$

$$\text{Well 3: } (\exp(1.90), \exp(5.44)) \text{ or } (6.69, 230.44)$$

These limits could be compared directly with the MCL of 20 ppb. It is generally easier to take the logarithm of the MCL rather than the antilogarithm of all of the intervals for comparison.

INTERPRETATION

If the original data are not normal, but the log-transformation adequately normalizes the data, the confidence interval (on the log scale) is an interval constructed so that the lower confidence limit should be less than the true or population mean (on the log scale) with specified confidence (99% in this case). If the lower limit exceeds the appropriate compliance limit, then the mean concentration must exceed that compliance limit. If the lower end of the interval is above the compliance limit, the mean concentration must be significantly greater (at the 1% level) than the compliance limit.

5.2.1.3 Nonparametric Confidence Interval

If the data do not adequately follow the normal distribution even after the logarithm transformation, a nonparametric confidence interval can be constructed. This interval is for the median concentration (which equals the mean if the distribution is symmetric). It requires a minimum of seven (7) observations in order to construct an interval with a two-sided confidence coefficient of 98%, corresponding to a one-sided confidence coefficient of 99%. Consequently, it is applicable only for the pooled concentration of compliance wells at a single point in time or for special sampling to produce a minimum of seven observations at a single well during the sampling period.

PURPOSE

The nonparametric confidence interval is used when the data have been found to violate the normality assumption, a log-transformation fails to normalize the data, and no other specific distribution is assumed. It produces a simple confidence interval that is designed to contain the true or population median concentration with specified confidence. If this confidence interval contains the compliance limit, it is concluded that the median concentration does not differ significantly from the compliance limit. If the interval's lower limit exceeds the compliance limit, this is statistically significant evidence that the concentration exceeds the compliance limit and the unit is out of compliance.

PROCEDURE

Step 1. Within each compliance well, order the n data from least to greatest, denoting the ordered data by $X(1), \dots, X(n)$, where $X(i)$ is the i th value in the ordered data.

Step 2. Determine the critical values of the order statistics. If the minimum seven observations is used, the critical values are 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 $n+1-M$ together with the exact confidence coefficient for sample sizes from 4 to 11. For larger samples, take as an approximation the nearest integer value to

$$M = n/2 + 1 + Z_{0.99} \sqrt{n/4}$$

where $Z_{0.99}$ is the 99th percentile from the normal distribution (Table 4, Appendix B) and equals 2.33.

Step 3. Once M has been determined in Step 2, find $n+1-M$ 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 $X(7)$.)

Step 4. Compare the confidence limits found in Step 3 to the compliance limit. If the lower limit, $X(M)$ exceeds the compliance limit, there is evidence of contamination. Otherwise, the unit remains in compliance.

TABLE 5-3. VALUES OF M AND $n+1-M$ AND CONFIDENCE COEFFICIENTS FOR SMALL SAMPLES

n	M	$n+1-M$	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%

REMARK

The nonparametric confidence interval procedure requires at least seven observations in order to obtain a (one-sided) significance level of 1% (confidence of 99%). This means that data from two (or more) wells or sampling periods would have to be pooled to achieve this level. If only the four observations from one well taken at a single sampling period were used, the one-sided significance level would be 6.25%. This would also be the false alarm rate.

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.). The ordered statistics are found by counting positions up from the bottom of the list as before. Multiple values from separate observations are counted separately.

EXAMPLE

Table 5-4 contains concentrations of Silvex in parts per million from two hypothetical compliance wells. The data are assumed to consist of four samples taken each quarter for a year, so that sixteen observations are available from each well. The data are neither normal nor log-normal, so that the nonparametric confidence interval is used. The MCL is taken to be 25 ppm.

Step 1. Order the 16 measurements from least to greatest within each well separately. The numbers in parentheses beside each concentration in Table 5-4 are 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.

TABLE 5-4. EXAMPLE DATA FOR NONPARAMETRIC CONFIDENCE INTERVAL--SILVEX CONCENTRATIONS (ppm)

Sampling date	Well 1		Well 2	
	Concentration (ppm)	Rank	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)
Apr. 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)
Jul. 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	(19)

Step 2. The sample size is large enough so that the approximation is used to find M.

$$M = 16/2 + 1 + 2.33 \sqrt{(16/4)} = 13.7 \approx 14$$

Step 3. The approximate 95% confidence limits are given by the $16 + 1 - 14 = 3$ rd largest observation and the 14th largest observation. For Well 1 the 3rd observation is 3.39 and the 14th largest observation is 10.25. Thus the confidence limits for Well 1 are (3.39, 10.25). Similarly for Well 2, the 3rd largest observation and the 14th largest observation are found to give the confidence interval (2.20, 11.02). Note that for Well 2 there were two values below detection. These were assigned a value of zero and received the two smallest ranks. Had there been three or more values below the limit of 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.

Step 4. Neither of the two confidence intervals' lower limit exceeds the MCL of 25. In fact, the upper limit is less than the MCL, implying that the concentration in each well is significantly below the MCL.

INTERPRETATION

The order statistics used to form the confidence interval in the nonparametric confidence interval procedure will contain the population median with confidence coefficient of 98%. The population median equals the mean whenever the distribution is symmetric. The nonparametric confidence interval is generally wider and requires more data than the corresponding normal distribution interval, and so the normal or log-normal distribution interval should be used whenever it is appropriate.

If the confidence interval contains the compliance limit (either MCL or ACL), then it is reasonable to conclude that the median compliance well concentration does not differ significantly from the compliance limit. If the lower end of the confidence interval exceeds the compliance limit, this is statistically significant evidence at the 1% level that the median compliance well concentration exceeds the compliance limit and the unit is out of compliance.

5.2.2 Tolerance Intervals for Compliance Limits

In some cases a permit may specify that a compliance limit (MCL or ACL) is not to be exceeded more than a specified fraction of the time. Since limited data will be available from each monitoring well, these data can be used to estimate a tolerance interval for concentrations from that well. If the upper tolerance limit is less than the compliance limit, the data indicate that the unit is in compliance. That is, concentrations should be less than the compliance limit at least a specified fraction of the time. If the upper tolerance limit of the interval exceeds the compliance limit, then the concentration of the hazardous constituent could exceed the compliance limit more than the specified proportion of the time.

This procedure compares an upper tolerance limit to the MCL or ACL. With small sample sizes the upper tolerance limit can be fairly large, particularly if large coverage with high confidence is desired. If the owner or operator wishes to use a tolerance limit in this application, he/she should suggest values for the parameters of the procedure subject to the approval of the Regional Administrator. To exemplify the procedure a 95% coverage is used with 95% confidence. This means that the upper tolerance limit is a value which, with 95% confidence, will be exceeded less than 5% of the time.

PURPOSE

The purpose of the tolerance interval approach is to construct an interval that should contain a specified fraction of the concentration measurements from compliance wells with a specified degree of confidence. In this application it is generally desired to have the tolerance interval contain 95% of the measurements of concentration with confidence at least 95%.

PROCEDURE

It is assumed that the data used to construct the tolerance interval are approximately normal. The data may consist of the concentration measurements themselves if they are adequately normal (see Section 3.2 for tests of normality), or the data used may be the natural logarithms of the concentration data. It is important that the compliance limit (MCL or ACL) be expressed in the same units (either concentrations or logarithm of the concentrations) as the observations.

Step 1. Calculate the mean, \bar{X} , and the standard deviation, S , of the compliance well concentration data.

Step 2. Determine the factor, K , from Table 5, Appendix B, for the sample size, n , and form the one-sided tolerance interval

$$\bar{X} + KS$$

Table 5, Appendix B contains the factors for a 95% coverage tolerance interval with confidence factor 95%.

Step 3. Compare the upper limit of the tolerance interval computed in Step 2 to the compliance limit. If the upper limit of the tolerance interval exceeds that limit, this is evidence of contamination.

EXAMPLE

Table 5-5 contains Aldicarb concentrations at a hypothetical facility in compliance monitoring. The data are concentrations in parts per million (PPM) and represent observations at three compliance wells. Assume that the permit establishes an ACL of 50 PPM that is not to be exceeded more than 5% of the time.

TABLE 5-5. EXAMPLE DATA FOR A TOLERANCE
INTERVAL COMPARED TO AN ACL

Sampling date	Aldicarb concentrations (ppm)		
	Well 1	Well 2	Well 3
Jan. 1	19.9	23.7	25.6
Feb. 1	29.6	21.9	23.3
Mar. 1	18.7	26.9	22.3
Apr. 1	24.2	26.1	26.9
Mean =	23.1	24.7	24.5
SD =	4.93	2.28	2.10

ACL = 50 ppm

Step 1. Calculate the mean and standard deviation of the observations from each well. These are given in the table.

Step 2. For $n = 4$, the factor, K , in Table 5, Appendix B, is found to be 5.145. Form the upper tolerance interval limits as:

$$\text{Well 1: } 23.1 + 5.145(4.93) = 48.5$$

$$\text{Well 2: } 24.7 + 5.145(2.28) = 36.4$$

$$\text{Well 3: } 24.5 + 5.145(2.10) = 35.3$$

Step 3. Compare the tolerance limits with the ACL of 50 PPM. Since the tolerance limits are below the ACL, there is no statistically significant evidence of contamination at any well. The site remains in detection monitoring.

INTERPRETATION

It may be desirable in a permit to specify a compliance limit that is not to be exceeded more than 5% of the time. A tolerance interval constructed from the compliance well data provides an estimated interval that will contain 95% of the data with confidence 95%. If this interval is below the selected compliance limit, concentrations measured at the compliance wells should exceed the compliance limit less than 5% of the time. If the upper limit of the tolerance interval exceeds the compliance limit, then more than 5% of the concentration measurements would be expected to exceed the compliance limit.

5.2.3 Special Cases with Limited Variance

Occasionally, all four concentrations from a compliance well at a particular sampling period could be identical. If this is the case, the formula for estimating the standard deviation at that specific sampling period would give zero, and the methods for calculating parametric confidence intervals would give the same limits for the upper and lower ends of the intervals, which is not appropriate.

In this case it is assumed that there is some variation in the data, but that the concentrations were rounded and gave the same values after rounding. To account for the variability that was present before rounding, take the least significant digit in the reported concentration as having resulted from rounding. It is assumed that rounding results in a uniform error on the interval centered at the reported value with the interval ranging up or down one half unit from the reported value. This assumed rounding is used to obtain a nonzero estimate of the variance for use in cases where all the measured concentrations were found to be identical.

PURPOSE

The purpose of this procedure is to obtain a nonzero estimate of the variance when all observations from a well during a given sampling period gave identical results. Once this modified variance is obtained, its square root

is used in place of the usual sample standard deviation, S , to construct confidence intervals or tolerance intervals.

PROCEDURE

Step 1. Determine the least significant value of any data point. That is, determine whether the data were reported to the nearest 10 ppm, nearest 1 ppm, nearest 100 ppm, etc. Denote this value by $2R$.

Step 2. The data are assumed to have been rounded to the nearest $2R$, so each observation is actually the reported value $\pm R$. Assuming that the observations were identical because of rounding, the variance is estimated to be $R^2/3$, assuming the uniform distribution for the rounding error. This gives the estimated standard deviation as

$$S' = R/\sqrt{3}$$

Step 3. Take this estimated value from Step 2 and use it as the estimate of the standard deviation in the appropriate parametric procedure. That is, replace S by S' .

EXAMPLE

In calculating a confidence interval for a single compliance well, suppose that four observations were taken during a sampling period and all resulted in 590 ppm. There is no variance among the four values 590, 590, 590, and 590.

Step 1. Assume that each of the values 590 came from rounding the concentration to the nearest 10 ppm. That is, 590 could actually be anything from 585.0 to 594.99, or any value between 585 and 595. Thus, $2R$ is 10 ppm, so R is 5 ppm.

Step 2. The estimate of the standard deviation is

$$S' = 5/\sqrt{3} = 5/1.732 = 2.89 \text{ ppm}$$

Step 3. Use $S' = 2.89$ and $\bar{X} = 590$ to calculate the confidence interval (see Section 5.2.1) for the mean concentration from this well. This gives

$$590 \pm (4.541)(2.89/\sqrt{4}) = (583.4, 596.6)$$

as the 98% confidence interval of the average concentration. Note that 4.541 is the 99th percentile from the t -distribution (Table 6, Appendix B) with 3 degrees of freedom since the sample size was 4.

INTERPRETATION

When identical results are obtained from several different samples, the interpretation is that the data are not reported to enough significant figures to show the random differences. The data are regarded as having resulted from rounding more precise results to the reported observations. The rounding is assumed to result in variability that follows the uniform distribution on the range $\pm R$, where $2R$ is the smallest unit of reporting. This assumption is used to calculate a standard deviation for the observations that otherwise appear to have no variability.

REMARK

Assuming that the data are reported correctly to the units indicated, other distributions for the rounding variability could be assumed. The maximum standard deviation that could result from rounding when the observation is $\pm R$ is the value R .

SECTION 6

CONTROL CHARTS FOR INTRA-WELL COMPARISONS

The previous sections covered various situations where the compliance well data were compared to the background data to detect possible contamination. This section discusses the case where the level of each constituent within a single well is being monitored over time. In essence, the data for each constituent in each well are plotted on a time scale and inspected for obvious features such as trends or sudden changes in concentration levels. The method suggested here is a combined Shewhart-CUSUM control chart for each well and constituent.

The control chart method is recommended when data comprising at least eight independent samples over a one-year period are available. This requirement is specified under current RCRA regulations and applies to each constituent in each well.

As discussed in Section 2, a common sampling plan will obtain four independent samples from each well on a semi-annual basis. With this plan a control chart can be implemented when one year's data are available. As a result of Monte Carlo simulations, Starks (1988) recommended at least four sampling periods at a unit of eight or more wells, and at least eight sampling periods at a unit with fewer than four wells.

The use of control charts can be an effective technique for monitoring the levels of a constituent at a given well over time. It also provides a visual means of detecting deviations from a "state of control." It is clear that plotting of the data is an important part of the analysis process. Plotting is an easy task, although time-consuming if many data sets need to be plotted. Advantage should be taken of graphics software, since plotting of time series data will be an ongoing process. New data points will be added to the already existing data base each time new data are available. The following few sections will discuss, in general terms, the advantages of plotting time series data; the corrective steps one could take to adjust when seasonality in the data is present; and finally, the detailed procedure for constructing a Shewhart-CUSUM control chart, along with a demonstration of that procedure, is presented.

6.1 ADVANTAGES OF PLOTTING DATA

While analyzing the data by means of any of the appropriate statistical procedures discussed in earlier sections is recommended, it is also recommended to plot the data. Each data point should be plotted against time using a time scale (e.g., month, quarter). A plot should be generated for each

constituent measured in each well. For visual comparison purposes, the scale should be kept identical from well to well for a given constituent.

Another important application of the plotting procedure is for detecting possible trends or drifts in the data from a given well. Furthermore, when visually comparing the plots from several wells within a unit, possible contamination of one rather than all downgradient wells could be detected which would then warrant a closer look at that well. In general, graphs can provide highly effective illustrations of the time series, allowing the analyst to obtain a much greater feeling for the data. Seasonal fluctuations or sudden changes, for example, may become quite evident, thereby supporting the analyst in his/her decision of which statistical procedure to use. General upward or downward trends, if present, can be detected and the analyst can follow-up with a test for trend, such as the nonparametric Mann-Kendall test (Mann, 1945; Kendall, 1975). If, in addition, seasonality is suspected, the user can perform the seasonal Kendall test for trend developed by Hirsch et al. (1982). The reader is also referred to Chapters 16 and 17 of Gilbert's "Statistical Methods for Environmental Pollution Monitoring," 1987. In any of the above cases, the help of a professional statistician is recommended.

Another important use of data plots is that of identifying unusual data points (e.g., outliers). These points should then be investigated for possible QC problems, data entry errors, or whether they are truly outliers.

Many software packages are available for computer graphics, developed for mainframes, mini-, or microcomputers. For example, SAS features an easy-to-use plotting procedure, PROC PLOT; where the hardware and software are available, a series of more sophisticated plotting routines can be accessed through SAS GRAPH. On microcomputers, almost everybody has his or her favorite graphics software that they use on a regular basis and no recommendation will be made as to the most appropriate one. The plots shown in this document were generated in LOTUS 1-2-3.

Once the data for each constituent and each well are plotted, the plots should be examined for seasonality and a correction is recommended should seasonality be present. A fairly simple-to-use procedure for deseasonalizing data is presented in the following paragraphs.

6.2 CORRECTING FOR SEASONALITY

A necessary precaution before constructing a control chart is to take into account seasonal variation of the data to minimize the chance of mistaking seasonal effect for evidence of well contamination. If seasonality is present, then deseasonalizing the data prior to using the combined Shewhart-CUSUM control chart procedure is recommended.

Many approaches to deseasonalize data exist. If the seasonal pattern is regular, it may be modeled with a sine or cosine function. Moving averages can be used, or differences (of order 12 for monthly data for example) can be used. However, time series models may include rather complicated methods for deseasonalizing the data. Another simpler method exists which should be adequate for the situations described in this document. It has the advantage of

being easy to understand and apply, and of providing natural estimates of the monthly or quarterly effects via the monthly or quarterly means. The method proposed here can be applied to any seasonal cycle--typically an annual cycle for monthly or quarterly data.

NOTE

Corrections for seasonality should be used with great caution as they represent extrapolation into the future. There should be a good scientific explanation for the seasonality as well as good empirical evidence for the seasonality before corrections are made. Larger than average rainfalls for two or three Augusts in a row does not justify the belief that there will never be a drought in August, and this idea extends directly to groundwater quality. In addition, the quality (bias, robustness, and variance) of the estimates of the proper corrections must be considered even in cases where corrections are called for. If seasonality is suspected, the user might want to seek the help of a professional statistician.

PURPOSE

When seasonality is known to exist in a time series of concentrations, then the data should be deseasonalized prior to constructing control charts in order to take into account seasonal variation rather than mistaking seasonal effects for evidence of contamination.

PROCEDURE

The following instructions to adjust a time series for seasonality are based on monthly data with a yearly cycle. The procedure can be easily modified to accommodate a yearly cycle of quarterly data.

Assume that N years of monthly data are available. Let x_{ij} denote the unadjusted observation for the i th month during the j th year.

Step 1. Compute the average concentration for month i over the N -year period:

$$\bar{X}_i = (X_{i1} + \dots + X_{iN})/N$$

This is the average of all observations taken in different years but during the same month. That is, calculate the mean concentrations in January, then the mean for February, and so on for each of the 12 months.

Step 2. Calculate the grand mean, \bar{X} , of all $N*12$ observations.

Step 3. Compute the adjusted concentrations,

$$Z_{ij} = x_{ij} - (\bar{X}_i - \bar{X})$$

Computing $X_{ij} - \bar{X}_i$ removes the average effect of month i from the monthly data, and adding \bar{X} , the overall mean, places the adjusted z_{ij} values about the same mean, \bar{X} . It follows that the overall mean adjusted observation, \bar{Z} , equals the overall mean unadjusted value, \bar{X} .

EXAMPLE

Columns 2 through 4 of Table 6-1 show monthly unadjusted concentrations of a fictitious analyte over a 3-year period.

TABLE 6-1. EXAMPLE COMPUTATION FOR DESEASONALIZING DATA

	Unadjusted concentrations			3-Month average	Monthly adjusted concentrations		
	1983	1984	1985		1983	1984	1985
January	1.99	2.01	2.15	2.05	2.10	2.13	2.27
February	2.10	2.10	2.17	2.12	2.14	2.15	2.21
March	2.12	2.17	2.27	2.19	2.10	2.15	2.25
April	2.12	2.13	2.23	2.16	2.13	2.14	2.24
May	2.11	2.13	2.24	2.16	2.12	2.13	2.25
June	2.15	2.18	2.26	2.20	2.12	2.15	2.23
July	2.19	2.25	2.31	2.25	2.11	2.16	2.23
August	2.18	2.24	2.32	2.25	2.10	2.16	2.24
September	2.16	2.22	2.28	2.22	2.11	2.17	2.22
October	2.08	2.13	2.22	2.14	2.10	2.16	2.24
November	2.05	2.08	2.19	2.11	2.11	2.14	2.25
December	2.08	2.16	2.22	2.16	2.09	2.17	2.23

Overall 3-year average = 2.17

Step 1. Compute the monthly averages across the 3 years. These values are shown in the fifth column of Table 6-1.

Step 2. The grand mean over the 3-year period is calculated to be 2.17.

Step 3. Within each month and year, subtract the average monthly concentration for that month and add the grand mean. For example, for January 1983, the adjusted concentration becomes

$$1.99 - 2.05 + 2.17 = 2.11$$

The adjusted concentrations are shown in the last three columns of Table 6-1.

The reader can check that the average of all 36 adjusted concentrations equals 2.17, the average unadjusted concentration. Figure 6-1 shows the plot of the unadjusted and adjusted data. The raw data clearly exhibit seasonality as well as an upwards trend which is less evident by simply looking at the data table.

INTERPRETATION

As can be seen in Figure 6-1, seasonal effects were present in the data. After adjusting for monthly effects, the seasonality was removed as can be seen in the adjusted data plotted in the same figure.

6.3 COMBINED SHEWHART-CUSUM CONTROL CHARTS FOR EACH WELL AND CONSTITUENT

Control charts are widely used as a statistical tool in industry as well as research and development laboratories. The concept of control charts is relatively simple, which makes them attractive to use. From the population distribution of a given variable, such as concentrations of a given constituent, repeated random samples are taken at intervals over time. Statistics, for example the mean of replicate values at a point in time, are computed and plotted together with upper and/or lower predetermined limits on a chart where the x-axis represents time. If a result falls outside these boundaries, then the process is declared to be "out of control"; otherwise, the process is declared to be "in control." The widespread use of control charts is due to their ease of construction and the fact that they can provide a quick visual evaluation of a situation, and remedial action can be taken, if necessary.

In the context of ground water monitoring, control charts can be used to monitor the inherent statistical variation of the data collected and to flag anomalous results. Further investigation of data points lying outside the established boundaries will be necessary before any direct action is taken.

A control chart that can be used on a real time basis must be constructed from a data set large enough to characterize the behavior of a specific well. It is recommended that data from a minimum of eight samples within a year be collected for each constituent at each well to permit an evaluation of the consistency of monitoring results with the current concept of the hydrogeology of the site. Starks (1988) recommends a minimum of four sampling periods at a unit with eight or more wells and a minimum of eight sampling periods at a unit with less than four wells. Once the control chart for the specific constituent at a given well is acceptable, then subsequent data points can be plotted on it to provide a quick evaluation as to whether the process is in control.

The standard assumptions in the use of control charts are that the data generated by the process, when it is in control, are independently and normally distributed with a fixed mean μ and constant variance σ^2 . The most important assumption is that of independence; control charts are not robust with respect to departure from independence (e.g., serial correlation). In general, the sampling scheme will be such that the possibility of obtaining serially correlated results is minimized, as noted in Section 2. The assumption of normality is of somewhat less concern, but should be investigated

Time Series of Monthly Observations

(Unadjusted, Adjusted, 3-year Mean)

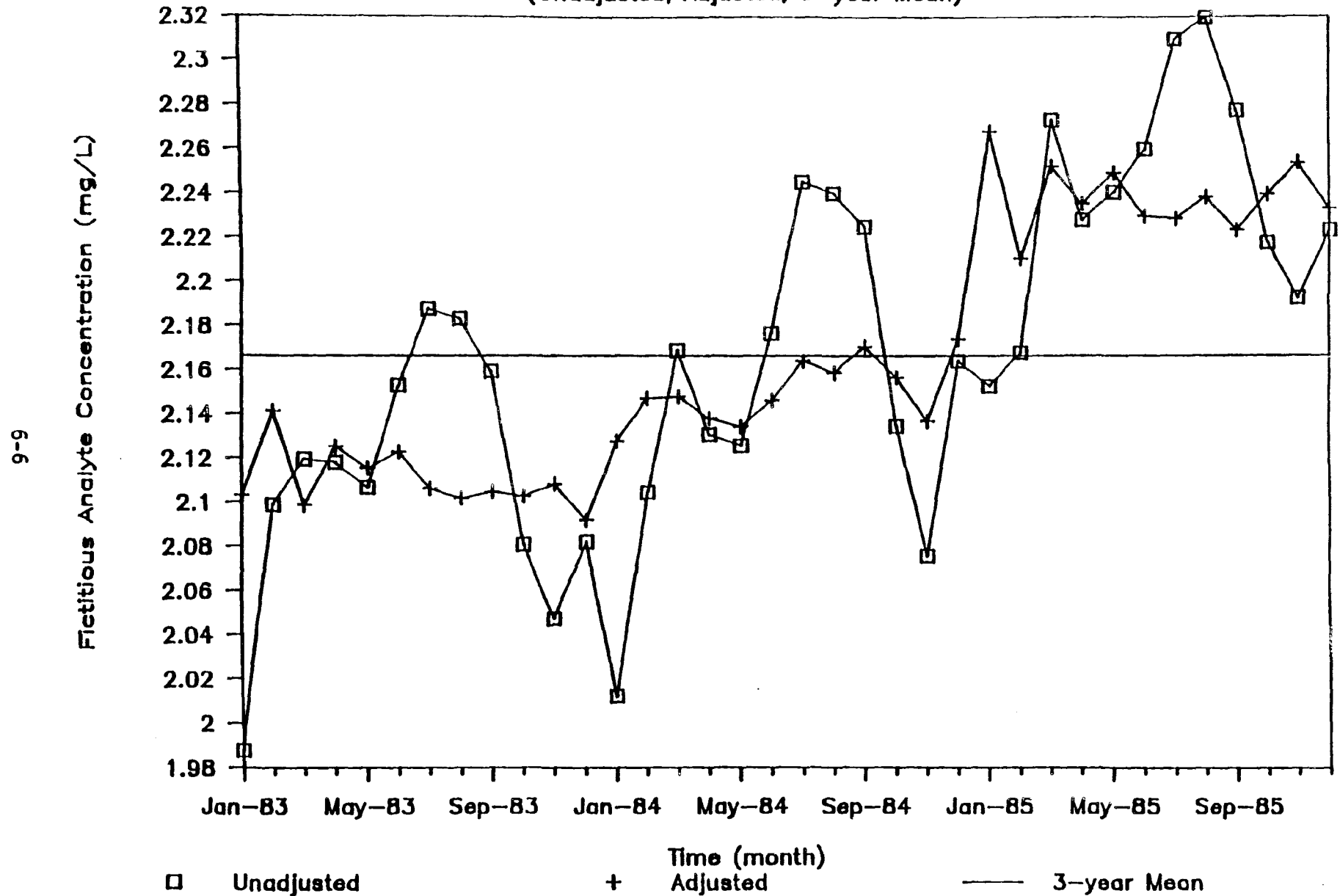


Figure 6-1. Plot of unadjusted and seasonally adjusted monthly observations.

before plotting the charts. A transformation (e.g., log-transform, square root transform) can be applied to the raw data so as to obtain errors normally distributed about the mean. An additional situation which may decrease the effectiveness of control charts is seasonality in the data. The problem of seasonality can be handled by removing the seasonality effect from the data, provided that sufficient data to cover at least two seasons of the same type are available (e.g., 2 years when monthly or quarterly seasonal effect). A procedure to correct a time series for seasonality was shown above in Section 6.2.

PURPOSE

Combined Shewhart-cumulative sum (CUSUM) control charts are constructed for each constituent at each well to provide a visual tool of detecting both trends and abrupt changes in concentration levels.

PROCEDURE

Assume that data from at least eight independent samples of monitoring are available to provide reliable estimates of the mean, μ , and standard deviation, σ , of the constituent's concentration levels in a given well.

Step 1. To construct a combined Shewhart-CUSUM chart, three parameters need to be selected prior to plotting:

h - a decision interval value

k - a reference value

SCL - Shewhart control limit (denoted by U in Starks (1988))

The parameter k of the CUSUM scheme is directly obtained from the value, D , of the displacement that should be quickly detected; $k = D/2$. It is recommended to select $k = 1$, which will allow a displacement of two standard deviations to be detected quickly.

When k is selected to be 1, the parameter h is usually set at values of 4 or 5. The parameter h is the value against which the cumulative sum in the CUSUM scheme will be compared. In the context of groundwater monitoring, a value of $h = 5$ is recommended (Starks, 1988; Lucas, 1982).

The upper Shewhart limit is set at $SCL = 4.5$ in units of standard deviation. This combination of $k = 1$, $h = 5$, and $SCL = 4.5$ was found most appropriate for the application of combined Shewhart-CUSUM charts for groundwater monitoring (Starks, 1988).

Step 2. Assume that at time period T_i , n_i concentration measurements X_{i1}, \dots, X_{in_i} , are available. Compute their average \bar{X}_i .

Step 3. Calculate the standardized mean

$$Z_i = (\bar{X}_i - \mu) / \sigma$$

where μ and σ are the mean and standard deviation obtained from prior monitoring at the same well (at least four sampling periods in a year).

Step 4. Compute the quantity

$$S_i = \max \{0, Z_i - k + S_{i-1}\}$$

where $\max \{A, B\}$ is the maximum of A and B. At sampling period T_1 , $S_0 = 0$.

Step 5. Plot the values of S_i versus T_i on a time chart for this combined Shewhart-CUSUM scheme. Declare an "out-of-control" situation at sampling period T_i if for the first time, $S_i \geq h$ or $Z_i \geq SCL$. This will indicate probable contamination at the well and further investigations will be necessary.

REFERENCES

Lucas, J. M. 1982. "Combined Shewhart-CUSUM Quality Control Schemes." *Journal of Quality Technology*. Vol. 14, pp. 51-59.

Starks, T. H. 1988 (Draft). "Evaluation of Control Chart Methodologies for RCRA Waste Sites."

Hockman, K. K., and J. M. Lucas. 1987. "Variability Reduction Through Subvessel CUSUM Control." *Journal of Quality Technology*. Vol. 19, pp. 113-121.

EXAMPLE

The procedure is demonstrated on a set of carbon tetrachloride measurements taken monthly at a compliance well over a 1-year period. The monthly means are presented in the second column of Table 6-2 below. Estimates of μ and σ , the mean and standard deviation of carbon tetrachloride measurements at that particular well were obtained from a preceding monitoring period at that well; $\mu = 5.5 \mu\text{g/L}$ and $\sigma = 0.4 \mu\text{g/L}$.

Step 1. The three parameters necessary to construct a combined Shewhart-CUSUM chart were selected as $h = 5$; $k = 1$; $SCL = 4.5$ in units of standard deviation.

Step 2. The monthly means are presented in Table 6-2.

Step 3. Standardize the mean observations within each sampling period. These computations are shown in the third column of Table 6-2. For example, $Z_1 = (4.28 - 5.50)/0.4 = -3.05$.

TABLE 6-2. EXAMPLE DATA FOR COMBINED SHEWHART-CUSUM CHART--
CARBON TETRACHLORIDE CONCENTRATION ($\mu\text{g/L}$)

Date	Sampling period T_i	Mean concentration, X_i	Standardized \bar{X}_i , Z_i	$Z_i - k$	CUSUM, S_i
Jan 6	1	4.28	-3.05	-4.05	0
Feb 3	2	5.21	-0.73	-1.73	0
Mar 3	3	4.77	-1.83	-2.83	0
Apr 7	4	4.71	-1.98	-2.98	0
May 5	5	6.16	1.65	0.65	0.65
Jun 2	6	5.84	0.85	-0.15	0.50
Jul 7	7	6.53	2.58	1.58	2.08 ^b
Aug 4	8	7.38	4.70 ^a	3.70	5.78 ^b
Sep 1	9	6.43	2.32	1.32	7.10 ^b
Oct 6	10	6.29	1.98	0.98	8.08 ^b
Nov 3	11	6.21	1.78	0.77	8.85 ^b
Dec 1	12	6.20	1.75	0.75	9.60 ^b

Parameters: Mean = 5.50; std = 0.4; $k = 1$; $h = 5$; SCL = 4.5.

^a Indicates "out-of-control" process via Shewhart control limit (SCL).

^b CUSUM "out-of-control" signal.

Step 4. Compute the quantities S_i , $i = 1, \dots, 12$. For example,

$$S_1 = \max \{0, -4.05 + 0\} = 0$$

$$S_2 = \max \{0, -1.73 + 0\} = 0$$

.

.

$$S_5 = \max \{0, 0.65 + S_4\} = \max \{0, 0.65 + 0\} = 0.65$$

$$S_6 = \max \{0, -0.15 + 0.65\} = \max \{0, 0.50\} = 0.50$$

etc.

These quantities are shown in the last column of Table 6-2.

Step 5. Construct the control chart. The y-axis is in units of standard deviations. The x-axis represent time, or the sampling periods. For each sampling period, T_i , record the value of X_i and S_i . Draw horizontal lines at values $h = 5$ and SCL = 4.5. These two lines represent the upper control limits for the CUSUM scheme and the Shewhart control limit, respectively. The chart for this example data set is shown in Figure 6-2.

COMBINED SHEWHART-CUSUM CHART

mean=5.5; std=0.4; k=1; h=5; SCL=4.5

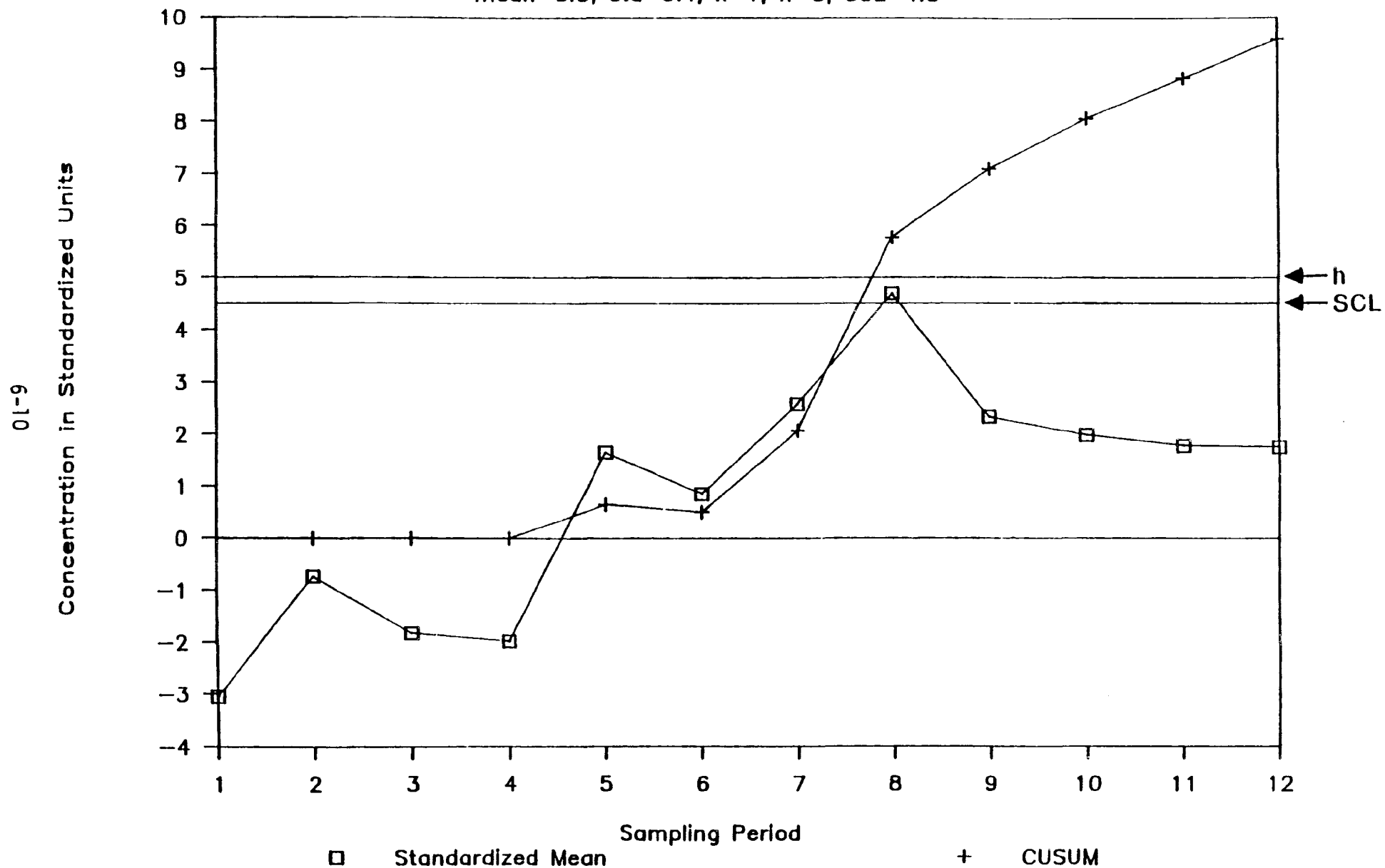


Figure 6-2

The combined chart indicates possible contamination starting at sampling period T_0 . Both the CUSUM scheme and the Shewhart control limit were exceeded by S_0 and Z_0 , respectively. Investigation of the situation should begin to confirm contamination and action should be required to bring the variability of the data back to its previous level.

INTERPRETATION

The combined Shewhart-CUSUM control scheme was applied to an example data set of carbon tetrachloride measurements taken on a monthly basis at a well. The statistic used in the construction of the chart was the mean measurement per sampling period. (It should be noted that this method can be used on an individual measurement as well.) Estimates of the mean and standard deviation of the measurements were available from previous data collected at that well over at least four sampling periods.

The parameters of the combined chart were selected to be $k = 1$ unit, the reference value or allowable slack for the process; $h = 5$ units, the decision interval for the CUSUM scheme; and $SCL = 4.5$ units, the upper Shewhart control limit. All parameters are in units of σ , the standard deviation obtained from the previous monitoring results. Various combinations of parameter values can be selected. The particular values recommended here appear to be the best for the initial use of the procedure from a review of the simulations and recommendations in the references. A discussion on this subject is given by Lucas (1982), Starks (1988), and Hockman and Lucas (1987). The choice of the parameters h and k of a CUSUM chart is based on the desired performance of the chart. The criterion used to evaluate a control scheme is the average number of samples or time periods before an out-of-control signal is obtained. This criterion is denoted by ARL or average run length. The ARL should be large when the mean concentration of a hazardous constituent is near its target value and small when the mean has shifted too far from the target. Tables have been developed by simulation methods to estimate ARLs for given combinations of the parameters (Lucas, Starks, Hockman and Lucas). The user is referred to these articles for further reading.

6.4 UPDATE OF A CONTROL CHART

The control chart is based on preselected performance parameters as well as on estimates of μ and σ , the parameters of the distribution of the measurements in question. As monitoring continues and the process is found to be in control, these parameters need periodic updating so as to incorporate this new information into the control charts. Starks (1988) has suggested that in general, adjustments in sample means and standard deviations be made after sampling periods 4, 8, 12, 20, and 32, following the initial monitoring period recommended to be at least eight sampling periods. Also, the performance parameters h , k , and SCL would need to be updated. The author suggests that $h = 5$, $k = 1$, and $SCL = 4.5$ be kept at those values for the first 12 sampling periods following the initial monitoring plan, and that k be reduced to 0.75 and SCL to 4.0 for all subsequent sampling periods. These values and sampling period numbers are not mandatory. In the event of an out-of-control state or a trend, the control chart should not be updated.

6.5 NONDETECTS IN A CONTROL CHART

Regulations require that four independent water samples be taken at each well at a given sampling period. It is the mean of the four concentration measurements of a particular constituent that is used in the construction of a control chart. Now situations will arise when the concentration of a constituent is below detection limit for one or more samples. The following approach is suggested for treating nondetects when plotting control charts.

If only one of the four measurements is a nondetect, then replace it with one half of the detection limit ($MDL/2$) or with one half of the practical quantitation limit ($PQL/2$) and proceed as described in Section 6.3.

If either two or three of the measurements are nondetects, use only the quantitated values (two or one, respectively) for the control chart and proceed as discussed earlier in Section 6.3.

If all four measurements are nondetects, then use one half of the detection limit or practical quantitation limit as the value for the construction of the control chart. This is an obvious situation of no contamination of the well.

In the event that a control chart requires updating and a certain proportion of the measurements is below detection limit, then adjust the mean and standard deviation necessary for the control chart by using Cohen's method described in Section 7.1.4. In that case, the proportion of nondetects applies to the pool of data available at the time of the updating and would include all nondetects up to that time, not just the four measurements taken at the last sampling period.

SECTION 7

MISCELLANEOUS TOPICS

This chapter contains a variety of special topics that are relatively short and self contained. These topics include methods to deal with data below the limit of detection and methods to check for, and deal with outliers or extreme values in the data.

7.1 LIMIT OF DETECTION

In a chemical analysis some compounds may be below the detection limit (DL) of the analytical procedure. These are generally reported as not detected (rather than as zero or not present) and the appropriate limit of detection is usually given. Data that include not detected results are a special case referred to as censored data in the statistical literature. For compounds not detected, the concentration of the compound is not known. Rather it is known that the concentration of the compound is less than the detection limit.

There are a variety of ways to deal with data that include values below detection. There is no general procedure that is applicable in all cases. However there are some general guidelines that usually prove adequate. If these do not cover a specific situation, the user should consult a professional statistician for the most appropriate way to deal with the values below detection.

A summary of suggested approaches to deal with data below the detection limit is presented as Table 7-1. The method suggested depends on the amount of data below the detection limit. For small amounts of below detection values, simply replacing a "ND" (not detected) report with a small number, say the detection limit divided by two, and proceeding with the usual analysis is satisfactory. For moderate amounts of below detection limit data, a more detailed adjustment is appropriate, while for large amounts one may need to only consider whether a compound was detected or not as the variable of analysis.

The meaning of small, moderate, and large above is subject to judgment. Table 7-1 contains some suggested values. It should be recognized that these values are not hard and fast rules, but are based on judgment. If there is a question about how to handle values below detection, consult a statistician.

TABLE 7-1. METHODS FOR BELOW DETECTION LIMIT VALUES

Percentage of Nondetects in the Data Base	Statistical Analysis Method	Section of Guidance Document
Less than 15%	<p>Replace NDs with MDL/2 or PQL/2, then proceed with parametric procedures:</p> <ul style="list-style-type: none"> • ANOVA • Tolerance Units • Prediction Intervals • Control Charts 	<p>Section 7.1.1</p> <p>Section 4.2.1 Section 4.3 Section 4.4 Section 6</p>
Between 15 and 50%	<p>Use NDs as ties, then proceed with Nonparametric ANOVA or use Cohen's adjustment, then proceed with:</p> <ul style="list-style-type: none"> • Tolerance Limits • Confidence Intervals • Control Charts 	<p>Section 4.2.2</p> <p>Section 7.1.3</p> <p>Section 4.3 Section 5.2.1 Section 6</p>
More than 50%	Test of Proportions	Section 7.1.2

It should be noted that the nonparametric methods presented earlier automatically deal with values below detection by regarding them as all tied at a level below any quantitated results. The nonparametric methods may be used if there is a moderate amount of data below detection. If the proportion of non-quantified values in the data exceeds 25%, these methods should be used with caution. They should probably not be used if less than half of the data consists of quantified concentrations.

7.1.1 The DL/2 Method

The amount of data that are below detection plays an important role in selecting the method to deal with the limit of detection problem. If a small proportion of the observations are not detected, these may be replaced with a small number, usually the method detection limit divided by 2 (MDL/2), and the usual analysis performed. This is the recommended method for use with the analysis of various procedure of Section 4.2.1. Seek professional help if in doubt about dealing with values below detection limit. The results of the analysis are generally not sensitive to the specific choice of the replacement number.

As a guideline, if 15% or fewer of the values are not detected, replace them with the method detection limit divided by two and proceed with the appropriate analysis using these modified values. Practical quantitation limits (PQL) for Appendix IX compounds were published by EPA in the Federal Register (Vol 52, No 131, July 9, 1987, pp 25947-25952). These give practical quantitation limits by compound and analytical method that may be used in replacing a small amount of nondetected data with the quantitation limit divided by 2. If approved by the Regional Administrator, site specific PQL's may be used in this procedure. If more than 15% of the values are reported as not detected, it is preferable to use a nonparametric method or a test of proportions.

7.1.2. Test of Proportions

If more than 50% of the data are below detection but at least 10% of the observations are quantified, a test of proportions may be used to compare the background well data with the compliance well data. Clearly, if none of the background well observations were above the detection limit, but all of the compliance well observations were above the detection limit, one would suspect contamination. In general the difference may not be as obvious. However, a higher proportion of quantitated values in compliance wells could provide evidence of contamination. The test of proportions is a method to determine whether a difference in proportion of detected values in the background well observations and compliance well observations provides statistically significant evidence of contamination.

The test of proportions should be used when the proportion of quantified values is small to moderate. If very few quantified values are found, a method based on the Poisson distribution may be used as an alternative approach. A method based on a tolerance limit for the number of detected compounds and the maximum concentration found for any detected compound has been proposed by Gibbons (1988). This alternative would be appropriate when

the number of detected compounds is quite small relative to the number of compounds analyzed for as might occur in detection monitoring.

PURPOSE

The test of proportions tests whether the proportion of compounds detected in the compliance well data differs significantly from the proportion of compounds detected in the background well data. If there is a significant difference, this is evidence of contamination.

PROCEDURE

The procedure uses the normal distribution approximation to the binomial distribution. This assumes that the sample size is reasonably large. Generally, if the proportion of detected values is denoted by P , and the sample size is n , then the normal approximation is adequate, provided that nP and $n(1-P)$ both are greater than or equal to 5.

Step 1. Determine X , the number of background well samples in which the compound was detected. Let n be the total number of background well samples analyzed. Compute the proportion of detects:

$$\hat{P}_U = x/n$$

Step 2. Determine Y , the number of compliance well samples in which the compound was detected. Let M be the total number of compliance well samples analyzed. Compute the proportion of detects:

$$\hat{P}_D = y/m$$

Step 3. Compute the standard error of the difference in proportions:

$$S_D = \{[(x+y)/(n+m)][1 - (x+y)/(n+m)][1/n + 1/m]\}^{1/2}$$

and form the statistic:

$$Z = (\hat{P}_U - \hat{P}_D)/S_D$$

Step 4. Compare the absolute value of Z to the 97.5th percentile from the standard normal distribution, 1.96. If the absolute value of Z exceeds 1.96, this provides statistically significant evidence at the 5% significance level that the proportion of compliance well samples where the compound was detected exceeds the proportion of background well samples where the compound was detected. This would be interpreted as evidence of contamination. (The two-sided test is used to provide information about differences in either direction.)

EXAMPLE

Table 7-2 contains data on cadmium concentrations measured in background well and compliance wells at a facility. In the table, "BDL" is used for below detection limit.

TABLE 7-2. EXAMPLE DATA FOR A TEST OF PROPORTIONS

Cadmium concentration ($\mu\text{g/L}$) at background well (24 samples)		Cadmium concentration ($\mu\text{g/L}$) at compliance wells (64 samples)		
0.1	BDL	0.12	BDL	0.024
0.12	BDL	0.08	BDL	BDL
BDL*	BDL	BDL	BDL	BDL
0.26	BDL	0.2	0.11	BDL
BDL		BDL	0.06	BDL
0.1		0.1	BDL	BDL
BDL		BDL	0.23	0.1
0.014		0.012	BDL	0.04
BDL		BDL	0.11	BDL
BDL		BDL	BDL	BDL
BDL		BDL	0.031	0.1
BDL		BDL	BDL	BDL
BDL		BDL	BDL	0.01
0.12		0.12	BDL	BDL
BDL		0.07	BDL	BDL
0.21		BDL	BDL	BDL
BDL		0.19	0.12	BDL
0.12		BDL	0.08	BDL
BDL		0.1	BDL	
BDL		BDL	0.26	
		0.01	BDL	
		BDL	0.02	
		BDL	BDL	

*BDL means below detection limit.

Step 1. Estimate the proportion above detection in the background wells. As shown in Table 7-2, there were 24 samples from background wells analyzed for cadmium, so $n = 24$. Of these, 16 were below detection and $x = 8$ were above detection, so $P_U = 8/24 = 0.333$.

Step 2. Estimate the proportion above detection in the compliance wells. There were 64 samples from compliance wells analyzed for cadmium, with 40 below detection and 24 detected values. This gives $m = 64$, $y = 24$, so $P_d = 24/64 = 0.375$.

Step 3. Calculate the standard error of the difference.

$$S_D = \{[(8+24)/(24+64)][1-(8+24)/(24+64)](1/24 + 1/64)\}^{1/2} = 0.115$$

Step 4. Form the statistic Z and compare it to the normal distribution.

$$Z = \frac{0.375 - 0.333}{0.115} = 0.37$$

which is less in absolute value than the value from the normal distribution, 1.96. Consequently, there is no significant evidence that the proportion of samples with cadmium levels above the detection limit differs in the background well and compliance well samples.

INTERPRETATION

Since the proportion of water samples with detected amounts of cadmium in the compliance wells was not significantly different from that in the background wells, the data are interpreted to provide no evidence of contamination. Had the proportion of samples with detectable levels of cadmium in the compliance wells been significantly higher than that in the background wells this would have been evidence of contamination. Had the proportion been significantly higher in the background wells, additional study would have been required. This could indicate that contamination was migrating from an off-site source, or it could mean that the hydraulic gradient had been incorrectly estimated or had changed and that contamination was occurring from the facility, but the ground-water flow was not in the direction originally estimated. Mounding of contaminants in the ground water near the background wells could also be a possible explanation of this observance.

7.1.3 Cohen's Method

If a confidence interval or a tolerance interval based upon the normal distribution is being constructed, a technique presented by Cohen (1959) specifies a method to adjust the sample mean and sample standard deviation to account for data below the detection limit. The only requirements for the use of this technique is that the data are normally distributed and that the detection limit be always the same. This technique is demonstrated below.

PURPOSE

Cohen's method provides estimates of the sample mean and standard deviation when some observations are below detection. These estimates can then be used to construct tolerance, confidence, or prediction intervals.

PROCEDURE

Let n be the total number of observations, m represent the number of data points above the detection limit (DL), and X_i represent the value of the i th constituent value above the detection limit.

Step 1. Compute the sample mean \bar{x}_d from the data above the detection limit as follows:

$$\bar{x}_d = \frac{1}{m} \sum_{i=1}^m x_i$$

Step 2. Compute the sample variance S_d^2 from the data above the detection limit as follows:

$$S_d^2 = \frac{\sum_{i=1}^m (x_i - \bar{x})^2}{m-1} = \frac{\sum_{i=1}^m x_i^2 - \frac{1}{m} (\sum_{i=1}^m x_i)^2}{m-1}$$

Step 3. Compute the two parameters, h and τ , as follows:

$$h = \frac{(n-m)}{n}$$

and

$$\tau = \frac{S_d^2}{(\bar{x} - DL)^2}$$

where n is the total number of observations (i.e., above and below the detection limit), and where DL is equal to the detection limit.

These values are then used to determine the value of the parameter $\hat{\lambda}$ from Table 7 in Appendix B.

Step 4. Estimate the corrected sample mean, which accounts for the data below detection limit, as follows:

$$\bar{X} = \bar{x}_d - \lambda(\bar{x}_d - DL)$$

Step 5. Estimate the corrected sample standard deviation, which accounts for the data below detection limit, as follows:

$$S = (S_d^2 + \lambda (\bar{x}_d - DL)^2)^{1/2}$$

Step 6. Use the modified values of \bar{X} and S in the procedure for constructing a tolerance interval (Section 4.3) or a confidence interval (Section 5.2.1).

REFERENCE

Cohen, A. C., Jr. 1959. "Simplified Estimators for the Normal Distribution When Samples are Singly Censored or Truncated." *Technometrics*. 1:217-237.

EXAMPLE

Table 7-3 contains data on sulfate concentrations. Three observations of the 24 were below the detection limit of 1,450 mg/L and are denoted by "< 1,450" in the table.

TABLE 7-3. EXAMPLE DATA FOR TESTING COHEN'S TEST

Sulfate concentration (mg/L)
1,850
1,760
< 1,450
1,710
1,575
1,475
1,780
1,790
1,780
< 1,450
1,790
1,800
< 1,450
1,800
1,840
1,820
1,860
1,780
1,760
1,800
1,900
1,770
1,790
1,780
> DL = 1,450 mg/L

Note: A symbol "<" before a number indicates that the value is not detected. The number following is then the limit of detection.

Step 1. Calculate the mean from the $m = 21$ values above detection

$$\bar{x}_d = 1,771.9$$

Step 2. Calculate the sample variance from the 21 quantified values

$$S_d^2 = 8,593.69$$

Step 3. Determine

$$h = (24-21)/24 = 0.125$$

and

$$\tau = 8593.69 / (1771.9 - 1450)^2 = 0.083$$

Enter Table 7 of Appendix B at $h = 0.125$ and $\tau = 0.083$ to determine the value of λ . Since the table does not contain these entries exactly, double linear interpolation was used to estimate $\lambda = 0.14986$.

REMARK

For the interested reader, the details of the double linear interpolation are provided.

The values from Table 14 between which the user needs to interpolate are:

	<u>$h = 0.10$</u>	<u>$h = 0.15$</u>
<u>τ</u>		
0.05	0.11431	0.17935
0.10	0.11804	0.18479

There are 0.025 units between 0.01 and 0.125 on the h -scale. There are 0.05 units between 0.10 and 0.15. Therefore, the value of interest (0.125) lies $(0.025/0.05 * 100) = 50\%$ of the distance along the interval between 0.10 and 0.15. To linearly interpolate between the tabulated values on the h axis, the range between the values must be calculated, the value that is 50% of the distance along the range must be computed and then that value must be added to the lower point on the tabulated values. The result is the interpolated value. The interpolated points on the h -scale for the current example are:

$$\begin{array}{ll} 0.17935 - 0.11431 = 0.06504 & 0.06504 * 0.50 = 0.03252 \\ 0.11431 + 0.03252 = \underline{0.14683} & \\ \\ 0.18479 - 0.11804 = 0.06675 & 0.06675 * 0.50 = 0.033375 \\ 0.11804 + 0.033375 = \underline{0.151415} & \end{array}$$

On the τ -axis there are 0.033 units between 0.05 and 0.083. There are 0.05 units between 0.05 and 0.10. The value of interest (0.083) lies

$(0.0330.05 * 100) = 66\%$ of the distance along the interval between 0.05 and 0.10. The interpolated point on the τ -axis is:

$$\begin{aligned} 0.141415 - 0.14683 &= 0.004585 & 0.004585 * 0.66 &= 0.0030261 \\ 0.14683 + 0.0030261 &= \underline{0.14986} \end{aligned}$$

Thus, $\hat{\lambda} = 0.14986$.

Step 5. The corrected sample mean and standard deviation are then estimated as follows:

$$\bar{X} = 1,771.9 - 0.14986 (1,771.9 - 1,450) = 1,723.66$$

$$S = [8,593.69 + 0.14986(1,771.9 - 1,450)^2]^{1/2} = 155.31$$

Step 6. These modified estimates of the mean, $\bar{X} = 1723.66$, and of the standard deviation, $S = 155.31$, would be used in the tolerance or confidence interval procedure. For example, if the sulfate concentrations represent background at a facility, the upper 95% tolerance limit becomes

$$1723.7 + (155.3)(2.309) = 2082.3 \text{ mg/L}$$

Observations from compliance wells in excess of 2,082 mg/L would give statistically significant evidence of contamination.

INTERPRETATION

Cohen's method provides maximum likelihood estimates of the mean and variance of a censored normal distribution. That is, of observations that follow a normal distribution except for those below a limit of detection, which are reported as "not detected." The modified estimates reflect the fact that the not detected observations are below the limit of detection, but not necessarily zero. The large sample properties of the modified estimates allow for them to be used with the normal theory procedures as a means of adjusting for not detected values in the data. Use of Cohen's method in more complicated calculations such as those required for analysis of variance procedures, requires special consideration from a professional statistician.

7.2 OUTLIERS

A ground-water constituent concentration value that is much different from most other values in a data set for the same ground-water constituent concentration can be referred to as an "outlier." Possible reasons for outliers can be:

- A catastrophic unnatural occurrence such as a spill;
- Inconsistent sampling or analytical chemistry methodology that may result in laboratory contamination or other anomalies;
- Errors in the transcription of data values or decimal points; and

- True but extreme ground-water constituent concentration measurements.

There are several tests to determine if there is statistical evidence that an observation is an outlier. The reference for the test presented here is ASTM paper E178-75.

PURPOSE

The purpose of a test for outliers is to determine whether there is statistical evidence that an observation that appears extreme does not fit the distribution of the rest of the data. If a suspect observation is identified as an outlier, then steps need to be taken to determine whether it is the result of an error or a valid extreme observation.

PROCEDURE

Let the sample of observations of a hazardous constituent of ground water be denoted by X_1, \dots, X_n . For specificity, assume that the data have been ordered and that the largest observation, denoted by X_n , is suspected of being an outlier. Generally, inspection of the data suggests values that do not appear to belong to the data set. For example, if the largest observation is an order of magnitude larger than the other observations, it would be suspect.

Step 1. Calculate the mean, \bar{X} and the standard deviation, S , of the data including all observations.

Step 2. Form the statistic

$$T_n = (X_n - \bar{X})/S$$

Note that T_n is the difference between the largest observation and the sample mean, divided by the sample standard deviation.

Step 3. Compare the statistic T_n to the critical value given the sample size, n , in Table 8 in Appendix B. If the statistic exceeds the critical value from the table, this is evidence that the suspect observation, X_n , is a statistical outlier.

Step 4. If the value is identified as an outlier, one of the actions outlined below should be taken. (The appropriate action depends on what can be learned about the observation.) The records of the sampling and analysis of the sample that led to it should be investigated to determine whether the outlier resulted from an error that can be identified.

- If an error (in transcription, dilution, analytical procedure, etc.) can be identified and the correct value recovered, the observation should be replaced by its corrected value and the appropriate statistical analysis done with the corrected value.

- If it can be determined that the observation is in error, but the correct value cannot be determined, then the observation should be deleted

from the data set and the appropriate statistical analysis performed. The fact that the observation was deleted and the reason for its deletion should be reported when reporting the results of the statistical analysis.

- If no error in the value can be documented then it must be assumed that the observation is a true but extreme value. In this case it must not be altered. It may be desirable to obtain another sample to confirm the observation. However, analysis and reporting should retain the observation and state that no error was found in tracing the sample that led to the extreme observation.

EXAMPLE

Table 7-4 contains 19 values of total organic carbon (TOC) that were obtained from a monitoring well. Inspection shows one value which at 11,000 mg/L is nearly an order of magnitude larger than most of the other observations. It is a suspected outlier.

Step 1. Calculate the mean and standard deviation of the data.

$$\bar{X} = 2300 \text{ and } S = 2325.9$$

TABLE 7-4. EXAMPLE DATA FOR TESTING FOR AN OUTLIER

Total organic carbon (mg/L)	
	1,700
	1,900
	1,500
	1,300
	11,000
	1,250
	1,000
	1,300
	1,200
	1,450
	1,000
	1,300
	1,000
	2,200
	4,900
	3,700
	1,600
	2,500
	1,900

Step 2. Calculate the statistic T_{19} .

$$T_{19} = (11000 - 2300) / 2325.9 = 3.74$$

Step 3. Referring to Table 8 of Appendix B for the upper 5% significance level, with $n = 19$, the critical value is 2.532. Since the value of the statistic $T = 3.74$ is greater than 2.532, there is statistical evidence that the largest observation is an outlier.

Step 4. In this case, tracking the data revealed that the unusual value of 11,000 resulted from a keying error and that the correct value was 1,100. This correction was then made in the data.

INTERPRETATION

An observation that is 4 or 5 times as large as the rest of the data is generally viewed with suspicion. An observation that is an order of magnitude different could arise by a common error of misplacing a decimal. The test for an outlier provides a statistical basis for determining whether an observation is statistically different from the rest of the data. If it is, then it is a statistical outlier. However, a statistical outlier may not be dropped or altered just because it has been identified as an outlier. The test provides a formal identification of an observation as an outlier, but does not identify the cause of the difference.

Whether or not a statistical test is done, any suspect data point should be checked. An observation may be corrected or dropped only if it can be determined that an error has occurred. If the error can be identified and corrected (as in transcription or keying) the correction should be made and the corrected values used. A value that is demonstrated to be incorrect may be deleted from the data. However, if no specific error can be documented, the observation must be retained in the data. Identification of an observation as an outlier but with no error documented could be used to suggest resampling to confirm the value.

APPENDIX A

GENERAL STATISTICAL CONSIDERATIONS AND
GLOSSARY OF STATISTICAL TERMS

GENERAL STATISTICAL CONSIDERATIONS

FALSE ALARMS OR TYPE I ERRORS

The statistical analysis of data from ground-water monitoring at RCRA sites has as its goal the determination of whether the data provide evidence of the presence of, or an increase in the level of, contamination. In the case of detection monitoring, the goal of the statistical analysis is to determine whether statistically significant evidence of contamination exists. In the case of compliance monitoring, the goal is to determine whether statistically significant evidence of concentration levels exceeding compliance limits exists. In monitoring sites in corrective action, the goal is to determine whether levels of the hazardous constituents are still above compliance limits or have been reduced to at or below the compliance limit.

These questions are addressed by the use of hypothesis tests. In the case of detection monitoring, it is hypothesized that a site is not contaminated; that is, the hazardous constituents are not present in the ground water. Samples of the ground water are taken and analyzed for the constituents in question. A hypothesis test is used to decide whether the data indicate the presence of the hazardous constituent. The test consists of calculating one or more statistics from the data and comparing the calculated results to some prespecified critical levels.

In performing a statistical test, there are four possible outcomes. Two of the possible outcomes result in the correct decision: (a) the test may correctly indicate that no contamination is present or (b) the test may correctly indicate the presence of contamination. The other two possibilities are errors: (c) the test may indicate that contamination is present when in fact it is not or (d) the test may fail to detect contamination when it is present.

If the stated hypothesis is that no contamination is present (usually called the null hypothesis) and the test indicates that contamination is present when in fact it is not, this is called a Type I error. Statistical hypothesis tests are generally set up to control the probability of Type I error to be no more than a specified value, called the significance level, and usually denoted by α . Thus in detection monitoring, the null hypothesis would be that the level of each hazardous constituent is zero (or at least below detection). The test would reject this hypothesis if some measure of concentration were too large, indicating contamination. A Type I error would be a false alarm or a triggering event that is inappropriate.

In compliance monitoring, the null hypothesis is that the level of each hazardous constituent is less than or equal to the appropriate compliance

limit. For the purpose of setting up the statistical procedure, the simple null hypothesis that the level is equal to the compliance limit would be used. As in detection monitoring, the test would indicate contamination if some measure of concentration is too large. A false alarm or Type I error would occur if the statistical procedure indicated that levels exceed the appropriate compliance limits when, in fact, they do not. Such an error would be a false alarm in that it would indicate falsely that compliance limits were being exceeded.

PROBABILITY OF DETECTION AND TYPE II ERROR

The other type of error that can occur is called a Type II error. It occurs if the test fails to detect contamination that is present. Thus a Type II error is a missed detection. While the probability of a Type I error can be specified, since it is the probability that the test will give a false alarm, the probability of a Type II error depends on several factors, including the statistical test, the sample size, and the significance level or probability of Type I error. In addition, it depends on the degree of contamination present. In general, the probability of a Type II error decreases as the level of contamination increases. Thus a test may be likely to miss low levels of contamination, less likely to miss moderate contamination, and very unlikely to miss high levels of contamination.

One can discuss the probability of a Type II error as the probability of a missed detection, or one can discuss the complement (one minus the probability of Type II error) of this probability. The complement, or probability of detection, is also called the power of the test. It depends on the magnitude of the contamination so that the power or probability of detecting contamination increases with the degree of contamination.

If the probability of a Type I error is specified, then for a given statistical test, the power depends on the sample size and the alternative of interest. In order to specify a desired power or probability of detection, one must specify the alternative that should be detected. Since generally the power will increase as the alternative differs more and more from the null hypothesis, one usually tries to specify the alternative that is closest to the null hypothesis, yet enough different that it is important to detect.

In the detection monitoring situation, the null hypothesis is that the concentration of the hazardous constituent is zero (or at least below detection). In this case the alternative of interest is that there is a concentration of the hazardous constituent that is above the detection limit and is large enough so that the monitoring procedure should detect it. Since it is a very difficult problem to select a concentration of each hazardous constituent that should be detectable with specified power, a more useful approach is to determine the power of a test at several alternatives and decide whether the procedure is acceptable on the basis of this power function rather than on the power against a single alternative.

In order to increase the power, a larger sample must be taken. This would mean sampling at more frequent intervals. There is a limit to how much can be achieved, however. In cases with limited water flow, it may not be possible to sample wells as frequently as desired. If samples close together

in time prove to be correlated, this correlation reduces the information available from the different samples. The additional cost of sampling and analysis will also impose practical limitations on the sample size that can be used.

Additional wells could also be used to increase the performance of the test. The additional monitoring wells would primarily be helpful in ensuring that a plume would not escape detection by missing the monitoring wells. However, in some situations the additional wells would contribute to a larger sample size and so improve the power.

In compliance monitoring the emphasis is on determining whether additional contamination has occurred, raising the concentration above a compliance limit. If the compliance limit is determined from the background well levels, the null hypothesis is that the difference between the background and compliance well concentrations is zero. The alternative of interest is that the compliance well concentration exceeds the background concentration. This situation is essentially the same for power considerations as that of the detection monitoring situation.

If compliance monitoring is relative to a compliance limit (MCL or ACL), specified as a constant, then the situation is different. Here the null hypothesis is that the concentration is less than or equal to the compliance limit, with equality used to establish the test. The alternative is that the concentration is above the compliance limit. In order to specify power, a minimum amount above the compliance limit must be established and power specified for that alternative or the power function evaluated for several possible alternatives.

SAMPLE DESIGNS AND ASSUMPTIONS

As discussed in Section 2, the sample design to be employed at a regulated unit will primarily depend on the hydrogeologic evaluation of the site. Wells should be sited to provide multiple background wells hydraulically upgradient from the regulated unit. The background wells allow for determination of natural spatial variability in ground-water quality. They also allow for estimation of background levels with greater precision than would be possible from a single upgradient well. Compliance wells should be sited hydraulically downgradient to each regulated unit. The location and spacing of the wells, as well as the depth of sampling, would be determined from the hydrology to ensure that at least one of the wells should intercept a plume of contamination of reasonable size.

Thus the assumed sample design is for a sample of wells to include a number of background wells for the site, together with a number of compliance wells for each regulated unit at the site. In the event that a site has only a single regulated unit, there would be two groups of wells, background and compliance. If a site has multiple regulated units, there would be a set of compliance wells for each regulated unit, allowing for detection monitoring or compliance monitoring separately at each regulated unit.

Data from the analysis of the water at each well are initially assumed to follow a normal distribution. This is likely to be the case for detection

monitoring of analytes in that levels should be near zero and errors would likely represent instrument or other sampling and analysis variability. If contamination is present, then the distribution of the data may be skewed to the right, giving a few very large values. The assumption of normality of errors in the detection monitoring case is quite reasonable, with deviations from normality likely indicating some degree of contamination. Tests of normality are recommended to ensure that the data are adequately represented by the normal distribution.

In the compliance monitoring case, the data for each analyte will again initially be assumed to follow the normal distribution. In this case, however, since there is a nonzero concentration of the analyte in the ground water, normality is more of an issue. Tests of normality are recommended. If evidence of nonnormality is found, the data should be transformed or a distribution-free test be used to determine whether statistically significant evidence of contamination exists.

The standard situation would result in multiple samples (taken at different times) of water from each well. The wells would form groups of background wells and compliance wells for each regulated unit. The statistical procedures recommended would allow for testing each compliance well group against the background group. Further, tests among the compliance wells within a group are recommended to determine whether a single well might be intercepting an isolated plume. The specific procedures discussed and recommended in the following sections should cover the majority of cases. They will not cover all of the possibilities. In the event that none of the procedures described and illustrated appears to apply to a particular case at a given regulated site, consultation with a statistician should be sought to determine an appropriate statistical procedure.

The following approach is recommended. If a regulated unit is in detection monitoring, it will remain in detection monitoring until or unless there is statistically significant evidence of contamination, in which case it would be placed in compliance monitoring. Likewise, if a regulated unit is in compliance monitoring, it will remain in compliance monitoring unless or until there is statistically significant evidence of further contamination, in which case it would move into corrective action.

In monitoring a regulated unit with multiple compliance wells, two types of significance levels are considered. One is an experimentwise significance level and the other is a comparisonwise significance level. When a procedure such as analysis of variance is used that considers several compliance wells simultaneously, the significance is an experimentwise significance. If individual well comparisons are made, each of those comparisons is done at a comparisonwise significance level.

The fact that many comparisons will be made at a regulated unit with multiple compliance wells can make the probability that at least one of the comparisons will be incorrectly significant too high. To control the false positive rate, multiple comparisons procedures are allowed that control the experimentwise significance level to be 5%. That is, the probability that one or more of the comparisons will falsely indicate contamination is controlled

at 5%. However, to provide some assurance of adequate power to detect real contamination, the comparisonwise significance level for comparing each individual well to the background is required to be no less than 1%.

Control of the experimentwise significance level via multiple comparisons procedures is allowed for comparisons among several wells. However, use of an experimentwise significance level for the comparisons among the different hazardous constituents is not permitted. Each hazardous constituent to be monitored for in the permit must be treated separately.

GLOSSARY OF STATISTICAL TERMS
(underlined terms are explained subsequently)

Alpha (α)	A greek letter used to denote the <u>significance level</u> or probability of a <u>Type I error</u> .
Alpha-error	Sometimes used for Type I error.
Alternative hypothesis	An alternative hypothesis specifies that the underlying distribution differs from the null hypothesis. The alternative hypothesis usually specifies the value of a <u>parameter</u> , for example the mean concentration, that one is trying to detect.
Arithmetic average	The arithmetic average of a set of observations is their sum divided by the number of observations.
Confidence coefficient	The confidence coefficient of a <u>confidence interval</u> for a parameter is the probability that the random interval constructed from the sample data contains the true value of the parameter. The confidence coefficient is related to the <u>significance level</u> of an associated hypothesis test by the fact that the significance level (in percent) is one hundred minus the confidence coefficient (in percent).
Confidence interval	A confidence interval for a parameter is a random interval constructed from sample data in such a way that the probability that the interval will contain the true value of the parameter is a specified value.
Cumulative distribution function	Distribution function.
Distribution-free	This is sometimes used as a synonym for <u>nonparametric</u> . A statistic is distribution-free if its distribution does not depend upon which specific distribution function (in a large class) the observations follow.

Distribution function	The distribution function for a random variable, X , is a function that specifies the probability that X is less than or equal to t , for all real values of t .
Experimentwise error rate	This term refers to <u>multiple comparisons</u> . If a total of n decisions are made about comparisons (for example of compliance wells to background wells) and x of the decisions are wrong, then the experimentwise error rate is x/n .
Hypothesis	This is a formal statement about a parameter of interest and the distribution of a statistic. It is usually used as a null hypothesis or an alternative hypothesis. For example, the null hypothesis might specify that ground water had a zero concentration of benzene and that analytical errors followed a normal distribution with mean zero and standard deviation 1 ppm.
Independence	A set of events are independent if the probability of the joint occurrence of any subset of the events factors into the product of the probabilities of the events. A set of observations is independent if the joint distribution function of the random errors associated with the observations factors into the product of the distribution functions.
Mean	Arithmetic average.
Median	This is the middle value of a sample when the observations have been ordered from least to greatest. If the number of observations is odd, it is the middle observation. If the number of observations is even, it is customary to take the midpoint between the two middle observations. For a distribution, the median is a value such that the probability is one-half that an observation will fall above or below the median.
Multiple comparison procedure	This is a statistical procedure that makes a large number of decisions or comparisons on one set of data. For example, at a sampling period, several compliance well concentrations may be compared to the background well concentration.

Nonparametric statistical procedure	A nonparametric statistical procedure is a statistical procedure that has desirable properties that hold under mild assumptions regarding the data. Typically the procedure is valid for a large class of distributions rather than for a specific distribution of the data such as the normal.
Normal population, normality	The errors associated with the observations follow the normal or Gaussian distribution function.
Null hypothesis	A null hypothesis specifies the underlying distribution of the data completely. Often the null distribution specifies that there is no difference between the mean concentration in background well water samples and compliance well water samples.
One-sided test	A one-sided test is appropriate if concentrations higher than those specified by the null hypothesis are of concern. A one-sided test only rejects for differences that are large and in a prespecified direction.
One-sided tolerance limit	This is an upper limit on observations from a specified distribution.
One-sided confidence limit	This is an upper limit on a parameter of a distribution.
Order statistics	The sample values observed after they have been arranged in increasing order.
Outlier	An outlier is an observation that is found to lie an unusually long way from the rest of the observations in a series of replicate observations.
Parameter	A parameter is an unknown constant associated with a population. For example, the mean concentration of a hazardous constituent in ground water is a parameter of interest.
Percentile	A percentile of a distribution is a value below which a specified proportion or percent of the observations from that distribution will fall.

Power	The power of a test is the probability that the test will reject under a specified alternative hypothesis. This is one minus the probability of a <u>Type II error</u> . The power is a measure of the test's ability to detect a difference of specified size from the null hypothesis.
Significance level	Sometimes referred to as the alpha level, the significance level of a test is the probability of falsely rejecting a true null hypothesis. The probability of a Type I error.
Type I error	A Type I error occurs when a true null hypothesis is rejected erroneously. In the monitoring context a Type I error occurs when a test incorrectly indicates contamination or an increase in contamination at a regulated unit.
Type II error	A Type II error occurs when one fails to reject a null hypothesis that is false. In the monitoring context, a Type II error occurs when monitoring fails to detect contamination or an increase in a concentration of a hazardous constituent.

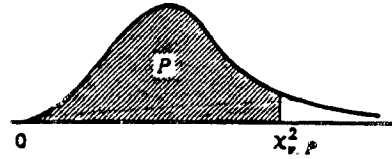
APPENDIX B

STATISTICAL TABLES

CONTENTS

<u>Table</u>		<u>Page</u>
1	Percentiles of the χ^2 Distribution With ν Degrees of Freedom, $\chi^2_{\nu,p}$	B-3
2	95th Percentiles of the F-Distribution With ν_1 and ν_2 Degrees of Freedom, $F_{\nu_1,\nu_2,0.95}$	B-4
3	95th Percentiles of the Bonferroni t-Statistics, $t(\nu, \alpha/m)$	B-5
4	Percentiles of the Standard Normal Distribution, U_p	B-6
5	Tolerance Factors (K) for One-Sided Normal Tolerance Intervals With Probability Level (Confidence Factor) $Y = 0.95$ and Coverage $P = 95\%$	B-8
6	Percentiles of Student's t-Distribution.....	B-9
7	Values of the Parameter $\hat{\lambda}$ for Cohen's Estimates Adjusting for Nondetected Values.....	B-10
8	Critical Values for T_n (One-Sided Test) When the Standard Deviation Is σ_n Calculated From the Same Sample...	B-11

TABLE 1. PERCENTILES OF THE χ^2 DISTRIBUTION WITH
 ν DEGREES OF FREEDOM, $\chi^2_{\nu, P}$



$\nu \backslash P$	0.750	0.900	0.950	0.975	0.990	0.995	0.999
1	1.323	2.706	3.841	5.024	6.635	7.879	10.83
2	2.773	4.605	5.991	7.378	9.210	10.60	13.82
3	4.108	6.251	7.815	9.348	11.34	12.84	16.27
4	5.385	7.779	9.488	11.14	13.28	14.86	18.47
5	6.626	9.236	11.07	12.83	15.09	16.75	20.52
6	7.841	10.64	12.59	14.45	16.81	18.55	22.46
7	9.037	12.02	14.07	16.01	18.48	20.28	24.32
8	10.22	13.36	15.51	17.53	20.09	21.96	26.12
9	11.39	14.68	16.92	19.02	21.67	23.59	27.88
10	12.55	15.99	18.31	20.48	23.21	25.19	29.59
11	13.70	17.28	19.68	21.92	24.72	26.76	31.26
12	14.85	18.55	21.03	23.34	26.22	28.30	32.91
13	15.98	19.81	22.36	24.74	27.69	29.82	34.53
14	17.12	21.06	23.68	26.12	29.14	31.32	36.12
15	18.25	22.31	25.00	27.49	30.58	32.80	37.70
16	19.37	23.54	26.30	28.85	32.00	34.27	39.25
17	20.49	24.77	27.59	30.19	33.41	35.72	40.79
18	21.60	25.99	28.87	31.53	34.81	37.16	42.31
19	22.72	27.20	30.14	32.85	36.19	38.58	43.82
20	23.83	28.41	31.41	34.17	37.57	40.00	45.32
21	24.93	29.62	32.67	35.48	38.93	41.40	46.80
22	26.04	30.81	33.92	36.78	40.29	42.80	48.27
23	27.14	32.01	35.17	38.08	41.64	44.18	49.73
24	28.24	33.20	36.42	39.36	42.98	45.56	51.18
25	29.34	34.38	37.65	40.65	44.31	46.93	52.62
26	30.43	35.56	38.89	41.92	45.64	48.29	54.05
27	31.53	36.74	40.11	43.19	46.96	49.64	55.48
28	32.62	37.92	41.34	44.46	48.28	50.99	56.89
29	33.71	39.09	42.56	45.72	49.59	52.34	58.30
30	34.80	40.26	43.77	46.98	50.89	53.67	59.70
40	45.62	51.80	55.76	59.34	63.69	66.77	73.40
50	56.33	63.17	67.50	71.42	76.15	79.49	86.66
60	66.98	74.40	79.08	83.30	88.38	91.95	99.61
70	77.58	85.53	90.53	95.02	100.4	104.2	112.3
80	88.13	96.58	101.9	106.6	112.3	116.3	124.8
90	98.65	107.6	113.1	118.1	124.1	128.3	137.2
100	109.1	118.5	124.3	129.6	135.8	140.2	149.4

SOURCE: Johnson, Norman L. and F. C. Leone. 1977. *Statistics and Experimental Design in Engineering and the Physical Sciences*. Vol. I. Second Edition. John Wiley and Sons, New York.

TABLE 2. 95th PERCENTILES OF THE F-DISTRIBUTION WITH
 ν_1 AND ν_2 DEGREES OF FREEDOM, $F_{\nu_1, \nu_2, 0.95}$



$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

NOTE: ν_1 : Degrees of freedom for numerator
 ν_2 : Degrees of freedom for denominator

SOURCE: Johnson, Norman L. and F. C. Leone. 1977. *Statistics and Experimental Design in Engineering and the Physical Sciences*. Vol. I. Second Edition. John Wiley and Sons, New York.

TABLE 3. 95th PERCENTILES OF THE BONFERRONI
t-STATISTICS, $t(v, \alpha/m)$

where v = degrees of freedom associated with the mean
squares error

m = number of comparisons

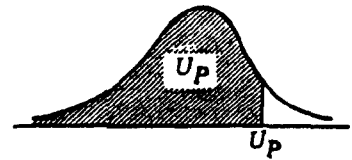
α = 0.05, the experimentwise error level

$\begin{array}{c} m \\ \alpha/m \\ v \end{array}$	1 0.05	2 0.025	3 0.0167	4 0.0125	5 0.01
4	2.13	2.78	3.20	3.51	3.75
5	2.02	2.57	2.90	3.17	3.37
6	1.94	2.45	2.74	2.97	3.14
7	1.90	2.37	2.63	2.83	3.00
8	1.86	2.31	2.55	2.74	2.90
9	1.83	2.26	2.50	2.67	2.82
10	1.81	2.23	2.45	2.61	2.76
15	1.75	2.13	2.32	2.47	2.60
20	1.73	2.09	2.27	2.40	2.53
30	1.70	2.04	2.21	2.34	2.46
∞	1.65	1.96	2.13	2.24	2.33

SOURCE: For $\alpha/m = 0.05, 0.025, \text{ and } 0.01$, the percentiles were extracted from the t-table (Table 6, Appendix B) for values of $F=1-\alpha$ of 0.95, 0.975, and 0.99, respectively.

For $\alpha/m = 0.05/3$ and $0.05/4$, the percentiles were estimated using "A Nomograph of Student's t" by Nelson, L. S. 1975. *Journal of Quality Technology*, Vol. 7, pp. 200-201.

TABLE 4. PERCENTILES OF THE STANDARD NORMAL DISTRIBUTION, U_p



P	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
0.50	0.0000	0.0025	0.0050	0.0075	0.0100	0.0125	0.0150	0.0175	0.0201	0.0226
0.51	0.0251	0.0276	0.0301	0.0326	0.0351	0.0376	0.0401	0.0426	0.0451	0.0476
0.52	0.0502	0.0527	0.0552	0.0577	0.0602	0.0627	0.0652	0.0677	0.0702	0.0728
0.53	0.0753	0.0778	0.0803	0.0828	0.0853	0.0878	0.0904	0.0929	0.0954	0.0979
0.54	0.1004	0.1030	0.1055	0.1080	0.1105	0.1130	0.1156	0.1181	0.1206	0.1231
0.55	0.1257	0.1282	0.1307	0.1332	0.1358	0.1383	0.1408	0.1434	0.1459	0.1484
0.56	0.1510	0.1535	0.1560	0.1586	0.1611	0.1637	0.1662	0.1687	0.1713	0.1738
0.57	0.1764	0.1789	0.1815	0.1840	0.1866	0.1891	0.1917	0.1942	0.1968	0.1993
0.58	0.2019	0.2045	0.2070	0.2096	0.2121	0.2147	0.2173	0.2198	0.2224	0.2250
0.59	0.2275	0.2301	0.2327	0.2353	0.2378	0.2404	0.2430	0.2456	0.2482	0.2508
0.60	0.2533	0.2559	0.2585	0.2611	0.2637	0.2663	0.2689	0.2715	0.2741	0.2767
0.61	0.2793	0.2819	0.2845	0.2871	0.2898	0.2924	0.2950	0.2976	0.3002	0.3029
0.62	0.3055	0.3081	0.3107	0.3134	0.3160	0.3186	0.3213	0.3239	0.3266	0.3292
0.63	0.3319	0.3345	0.3372	0.3398	0.3425	0.3451	0.3478	0.3505	0.3531	0.3558
0.64	0.3585	0.3611	0.3638	0.3665	0.3692	0.3719	0.3745	0.3772	0.3799	0.3826
0.65	0.3853	0.3880	0.3907	0.3934	0.3961	0.3989	0.4016	0.4043	0.4070	0.4097
0.66	0.4125	0.4152	0.4179	0.4207	0.4234	0.4261	0.4289	0.4316	0.4344	0.4372
0.67	0.4399	0.4427	0.4454	0.4482	0.4510	0.4538	0.4565	0.4593	0.4621	0.4649
0.68	0.4677	0.4705	0.4733	0.4761	0.4789	0.4817	0.4845	0.4874	0.4902	0.4930
0.69	0.4959	0.4987	0.5015	0.5044	0.5072	0.5101	0.5129	0.5158	0.5187	0.5215
0.70	0.5244	0.5273	0.5302	0.5330	0.5359	0.5388	0.5417	0.5446	0.5476	0.5505
0.71	0.5534	0.5563	0.5592	0.5622	0.5651	0.5681	0.5710	0.5740	0.5769	0.5799
0.72	0.5828	0.5858	0.5888	0.5918	0.5948	0.5978	0.6008	0.6038	0.6068	0.6098
0.73	0.6128	0.6158	0.6189	0.6219	0.6250	0.6280	0.6311	0.6341	0.6372	0.6403
0.74	0.6433	0.6464	0.6495	0.6526	0.6557	0.6588	0.6620	0.6651	0.6682	0.6713

NOTE: For values of P below 0.5, obtain the value of $U_{(1-p)}$ from Table 4 and change its sign. For example, $U_{0.45} = -U_{(1-0.45)} = -U_{0.55} = -0.1257$.

(Continued)

TABLE 4 (Continued)

P	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
0.75	0.6745	0.6776	0.6808	0.6840	0.6871	0.6903	0.6935	0.6967	0.6999	0.7031
0.76	0.7063	0.7095	0.7128	0.7160	0.7192	0.7225	0.7257	0.7290	0.7323	0.7356
0.77	0.7388	0.7421	0.7454	0.7488	0.7521	0.7554	0.7588	0.7621	0.7655	0.7688
0.78	0.7722	0.7756	0.7790	0.7824	0.7858	0.7892	0.7926	0.7961	0.7995	0.8030
0.79	0.8064	0.8099	0.8134	0.8169	0.8204	0.8239	0.8274	0.8310	0.8345	0.8381
0.80	0.8416	0.8452	0.8488	0.8524	0.8560	0.8596	0.8633	0.8669	0.8705	0.8742
0.81	0.8779	0.8816	0.8853	0.8890	0.8927	0.8965	0.9002	0.9040	0.9078	0.9116
0.82	0.9154	0.9192	0.9230	0.9269	0.9307	0.9346	0.9385	0.9424	0.9463	0.9502
0.83	0.9542	0.9581	0.9621	0.9661	0.9701	0.9741	0.9782	0.9822	0.9863	0.9904
0.84	0.9945	0.9986	1.0027	1.0069	1.0110	1.0152	1.0194	1.0237	1.0279	1.0322
0.85	1.0364	1.0407	1.0450	1.0494	1.0537	1.0581	1.0625	1.0669	1.0714	1.0758
0.86	1.0803	1.0848	1.0893	1.0939	1.0985	1.1031	1.1077	1.1123	1.1170	1.1217
0.87	1.1264	1.1311	1.1359	1.1407	1.1455	1.1503	1.1552	1.1601	1.1650	1.1700
0.88	1.1750	1.1800	1.1850	1.1901	1.1952	1.2004	1.2055	1.2107	1.2160	1.2212
0.89	1.2265	1.2319	1.2372	1.2426	1.2481	1.2536	1.2591	1.2646	1.2702	1.2759
0.90	1.2816	1.2873	1.2930	1.2988	1.3047	1.3106	1.3165	1.3225	1.3285	1.3346
0.91	1.3408	1.3469	1.3532	1.3595	1.3658	1.3722	1.3787	1.3852	1.3917	1.3984
0.92	1.4051	1.4118	1.4187	1.4255	1.4325	1.4395	1.4466	1.4538	1.4611	1.4684
0.93	1.4758	1.4833	1.4909	1.4985	1.5063	1.5141	1.5220	1.5301	1.5382	1.5464
0.94	1.5548	1.5632	1.5718	1.5805	1.5893	1.5982	1.6072	1.6164	1.6258	1.6352
0.95	1.6449	1.6546	1.6646	1.6747	1.6849	1.6954	1.7060	1.7169	1.7279	1.7392
0.96	1.7507	1.7624	1.7744	1.7866	1.7991	1.8119	1.8250	1.8384	1.8522	1.8663
0.97	1.8808	1.8957	1.9110	1.9268	1.9431	1.9600	1.9774	1.9954	2.0141	2.0335
0.98	2.0537	2.0749	2.0969	2.1201	2.1444	2.1701	2.1973	2.2262	2.2571	2.2904
0.99	2.3263	2.3656	2.4089	2.4573	2.5121	2.5758	2.6521	2.7478	2.8782	3.0902

SOURCE: Johnson, Norman L. and F. C. Leone. 1977. *Statistics and Experimental Design in Engineering and the Physical Sciences*. Vol. I, Second Edition. John Wiley and Sons, New York.

TABLE 5. TOLERANCE FACTORS (K) FOR ONE-SIDED NORMAL TOLERANCE INTERVALS WITH PROBABILITY LEVEL (CONFIDENCE FACTOR) $Y = 0.95$ AND COVERAGE $P = 95\%$

n	K	n	K
3	7.655	75	1.972
4	5.145	100	1.924
5	4.202	125	1.891
6	3.707	150	1.868
7	3.399	175	1.850
8	3.188	200	1.836
9	3.031	225	1.824
10	2.911	250	1.814
11	2.815	275	1.806
12	2.736	300	1.799
13	2.670	325	1.792
14	2.614	350	1.787
15	2.566	375	1.782
16	2.523	400	1.777
17	2.486	425	1.773
18	2.543	450	1.769
19	2.423	475	1.766
20	2.396	500	1.763
21	2.371	525	1.760
22	2.350	550	1.757
23	2.329	575	1.754
24	2.309	600	1.752
25	2.292	625	1.750
30	2.220	650	1.748
35	2.166	675	1.746
40	2.126	700	1.744
45	2.092	725	1.742
50	2.065	750	1.740
		775	1.739
		800	1.737
		825	1.736
		850	1.734
		875	1.733
		900	1.732
		925	1.731
		950	1.729
		975	1.728
		1000	1.727

SOURCE: (a) for sample sizes ≤ 50 : Lieberman, Gerald F. 1958. "Tables for One-sided Statistical Tolerance Limits." *Industrial Quality Control*. Vol. XIV, No. 10. (b) for sample sizes ≥ 50 : K values were calculated from large sample approximation.

TABLE 6. PERCENTILES OF STUDENT'S t-DISTRIBUTION

(F = 1- α ; n = degrees of freedom)

α \ P	.60	.75	.90	.95	.975	.99	.995	.9995
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	636.619
2	.289	.816	1.888	2.920	4.303	6.965	9.925	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	12.941
4	.271	.741	1.533	2.132	2.776	3.747	4.604	8.610
5	.267	.727	1.478	2.015	2.571	3.365	4.032	6.859
6	.265	.718	1.440	1.943	2.447	3.143	3.707	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	5.405
8	.262	.706	1.397	1.860	2.306	2.898	3.355	5.041
9	.261	.703	1.383	1.833	2.262	2.821	3.250	4.781
10	.260	.700	1.372	1.812	2.228	2.764	3.169	4.587
11	.260	.697	1.363	1.796	2.201	2.718	3.106	4.437
12	.259	.695	1.356	1.782	2.179	2.681	3.055	4.318
13	.259	.694	1.350	1.771	2.160	2.650	3.012	4.221
14	.258	.692	1.345	1.761	2.145	2.624	2.977	4.140
15	.258	.691	1.341	1.753	2.131	2.602	2.947	4.073
16	.258	.690	1.337	1.746	2.120	2.583	2.921	4.016
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.965
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.922
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.883
20	.257	.687	1.325	1.725	2.086	2.528	2.845	3.850
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.819
22	.256	.686	1.321	1.717	2.074	2.508	2.819	3.792
23	.256	.685	1.319	1.714	2.069	2.500	2.807	3.767
24	.256	.685	1.318	1.711	2.064	2.492	2.797	3.745
25	.256	.684	1.316	1.708	2.060	2.485	2.787	3.725
26	.256	.684	1.315	1.706	2.056	2.479	2.779	3.707
27	.256	.684	1.314	1.703	2.052	2.473	2.771	3.690
28	.256	.683	1.313	1.701	2.048	2.467	2.763	3.674
29	.256	.683	1.311	1.699	2.045	2.462	2.756	3.659
30	.256	.683	1.310	1.697	2.042	2.457	2.750	3.646
40	.255	.681	1.303	1.684	2.021	2.423	2.704	3.551
60	.254	.679	1.296	1.671	2.000	2.390	2.660	3.460
120	.254	.677	1.289	1.658	1.980	2.358	2.617	3.373
∞	.253	.674	1.282	1.645	1.960	2.326	2.576	3.291

SOURCE: *CRC Handbook of Tables for Probability and Statistics*. 1966.
W. H. Beyer, Editor. Published by the Chemical Rubber Company. Cleveland,
Ohio.

TABLE 7. VALUES OF THE PARAMETER $\hat{\lambda}$ FOR COHEN'S ESTIMATES
ADJUSTING FOR NONDETECTED VALUES

λ	.01	.02	.03	.04	.05	.06	.07	.08	.09	.10	.15	.20	λ
.00	.010109	.020409	.030903	.041363	.052367	.063327	.074963	.086488	.09824	.11020	.17343	.24266	.00
.05	.010381	.021294	.032233	.043330	.054670	.066186	.077909	.089834	.10197	.11431	.17933	.25023	.05
.10	.010930	.022082	.033398	.044903	.056619	.068482	.080566	.092823	.10534	.11804	.18479	.25741	.10
.15	.011310	.022798	.034466	.046318	.058336	.070586	.083009	.095629	.10843	.12146	.18983	.26403	.15
.20	.011642	.023486	.035433	.047629	.059990	.072539	.085286	.098216	.11133	.12469	.19440	.27031	.20
.25	.011982	.024076	.036377	.048886	.061622	.074572	.087743	.10103	.11456	.12832	.19910	.27626	.25
.30	.012343	.024688	.037249	.050018	.062999	.076186	.089483	.10289	.11647	.13026	.20338	.28193	.30
.35	.012720	.025211	.038077	.051170	.064345	.077736	.091285	.10491	.11864	.13253	.20747	.28727	.35
.40	.013104	.025738	.038846	.052173	.065460	.078932	.092513	.10621	.12000	.13389	.21139	.29260	.40
.45	.013491	.026283	.039524	.053182	.066621	.080243	.093986	.10786	.12177	.13567	.21517	.29768	.45
.50	.013879	.026838	.040232	.054193	.067813	.081531	.095367	.10921	.12308	.13696	.21892	.30283	.50
.55	.014271	.027394	.040938	.055204	.068936	.082767	.096617	.11048	.12436	.13824	.22268	.30795	.55
.60	.014667	.027950	.041503	.056215	.070049	.083980	.097930	.11180	.12568	.13956	.22644	.31307	.60
.65	.015066	.028507	.042069	.057226	.071060	.085091	.099141	.11301	.12689	.14077	.23020	.31819	.65
.70	.015467	.029064	.042631	.058237	.072071	.086202	.100352	.11422	.12810	.14198	.23396	.32331	.70
.75	.015870	.029621	.043193	.059248	.073082	.087313	.101463	.11533	.12921	.14309	.23772	.32843	.75
.80	.016275	.030178	.043755	.060259	.074093	.088424	.102574	.11644	.13032	.14420	.24148	.33355	.80
.85	.016681	.030735	.044317	.061270	.075104	.089535	.103685	.11755	.13143	.14531	.24524	.33867	.85
.90	.017088	.031292	.044879	.062281	.076115	.090646	.104796	.11866	.13254	.14642	.24900	.34379	.90
.95	.017495	.031849	.045441	.063292	.077126	.091757	.105907	.11977	.13365	.14753	.25276	.34891	.95
1.00	.017902	.032406	.046003	.064303	.078137	.092868	.107018	.12109	.13497	.14885	.25652	.35403	1.00

λ	.25	.30	.35	.40	.45	.50	.55	.60	.65	.70	.80	.90	λ
.00	.21882	.4021	.5941	.7961	.9986	.1188	.1388	1.145	1.338	1.561	2.176	3.283	.00
.05	.22793	.4130	.6066	.8101	.9986	.1246	.1446	1.151	1.344	1.567	2.181	3.314	.05
.10	.23683	.4233	.6184	.8234	.9986	.1304	.1504	1.164	1.357	1.580	2.195	3.345	.10
.15	.24563	.4336	.6306	.8367	.9986	.1362	.1562	1.177	1.370	1.593	2.209	3.376	.15
.20	.25443	.4439	.6429	.8499	.9986	.1420	.1620	1.190	1.383	1.606	2.223	3.407	.20
.25	.26323	.4542	.6552	.8632	.9986	.1478	.1678	1.203	1.396	1.619	2.237	3.438	.25
.30	.27203	.4645	.6675	.8765	.9986	.1536	.1736	1.216	1.409	1.632	2.251	3.469	.30
.35	.28083	.4748	.6798	.8898	.9986	.1594	.1794	1.229	1.422	1.645	2.265	3.500	.35
.40	.28963	.4851	.6921	.9031	.9986	.1652	.1852	1.242	1.435	1.658	2.279	3.531	.40
.45	.29843	.4954	.7044	.9164	.9986	.1710	.1910	1.255	1.448	1.671	2.293	3.562	.45
.50	.30723	.5057	.7167	.9297	.9986	.1768	.1968	1.268	1.461	1.684	2.307	3.593	.50
.55	.31603	.5160	.7290	.9430	.9986	.1826	.2026	1.281	1.474	1.697	2.321	3.624	.55
.60	.32483	.5263	.7413	.9563	.9986	.1884	.2084	1.294	1.487	1.710	2.335	3.655	.60
.65	.33363	.5366	.7536	.9696	.9986	.1942	.2142	1.307	1.500	1.723	2.349	3.686	.65
.70	.34243	.5469	.7659	.9829	.9986	.2000	.2200	1.320	1.513	1.736	2.363	3.717	.70
.75	.35123	.5572	.7782	.9962	.9986	.2058	.2258	1.333	1.526	1.749	2.377	3.748	.75
.80	.36003	.5675	.7905	.9962	.9986	.2116	.2316	1.346	1.539	1.762	2.391	3.779	.80
.85	.36883	.5778	.8028	.9962	.9986	.2174	.2374	1.359	1.552	1.775	2.405	3.810	.85
.90	.37763	.5881	.8151	.9962	.9986	.2232	.2432	1.372	1.565	1.788	2.419	3.841	.90
.95	.38643	.5984	.8274	.9962	.9986	.2290	.2490	1.385	1.578	1.801	2.433	3.872	.95
1.00	.39523	.6087	.8397	.9962	.9986	.2348	.2548	1.398	1.591	1.814	2.447	3.903	1.00

For all values $0 \leq \lambda \leq 1$, $\lambda(0, \gamma) = 0$.

SOURCE: Cohen, A. C., Jr. 1961. "Tables for Maximum Likelihood Estimates: Singly Truncated and Singly Censored Samples." *Technometrics*.

TABLE 8. CRITICAL VALUES FOR T_n (ONE-SIDED TEST) WHEN THE STANDARD DEVIATION IS CALCULATED FROM THE SAME SAMPLE

Number of Observations, n	Upper 0.1% Significance Level	Upper 0.5% Significance Level	Upper 1% Significance Level	Upper 2.5% Significance Level	Upper 5% Significance Level	Upper 10% Significance Level
3	1.155	1.155	1.155	1.155	1.153	1.148
4	1.499	1.496	1.492	1.481	1.463	1.425
5	1.780	1.764	1.749	1.715	1.672	1.602
6	2.011	1.973	1.944	1.887	1.822	1.729
7	2.201	2.139	2.097	2.020	1.938	1.828
8	2.358	2.274	2.221	2.126	2.032	1.909
9	2.492	2.387	2.323	2.215	2.110	1.977
10	2.606	2.482	2.410	2.290	2.176	2.036
11	2.705	2.564	2.485	2.355	2.234	2.088
12	2.791	2.636	2.550	2.412	2.285	2.134
13	2.867	2.699	2.607	2.462	2.331	2.175
14	2.935	2.755	2.659	2.507	2.371	2.213
15	2.997	2.806	2.705	2.549	2.409	2.247
16	3.052	2.852	2.747	2.585	2.443	2.279
17	3.103	2.894	2.785	2.620	2.475	2.309
18	3.149	2.932	2.821	2.651	2.504	2.335
19	3.191	2.968	2.854	2.681	2.532	2.361
20	3.230	3.001	2.884	2.709	2.557	2.385
21	3.266	3.031	2.912	2.733	2.580	2.408
22	3.300	3.060	2.939	2.758	2.603	2.429
23	3.332	3.087	2.963	2.781	2.624	2.448
24	3.362	3.112	2.987	2.802	2.644	2.467
25	3.389	3.135	3.009	2.822	2.663	2.486
26	3.415	3.157	3.029	2.841	2.681	2.502
27	3.440	3.178	3.049	2.859	2.698	2.519
28	3.464	3.199	3.068	2.876	2.714	2.534
29	3.486	3.218	3.085	2.893	2.730	2.549
30	3.507	3.236	3.103	2.908	2.745	2.563
31	3.528	3.253	3.119	2.924	2.759	2.577
32	3.546	3.270	3.135	2.938	2.773	2.591
33	3.565	3.286	3.150	2.952	2.786	2.604
34	3.582	3.301	3.164	2.965	2.799	2.616
35	3.599	3.316	3.178	2.979	2.811	2.628
36	3.616	3.330	3.191	2.991	2.823	2.639
37	3.631	3.343	3.204	3.003	2.835	2.650
38	3.646	3.356	3.216	3.014	2.846	2.661
39	3.660	3.369	3.228	3.025	2.857	2.671
40	3.673	3.381	3.240	3.036	2.866	2.682
41	3.687	3.393	3.251	3.046	2.877	2.692
42	3.700	3.404	3.261	3.057	2.887	2.700
43	3.712	3.415	3.271	3.067	2.896	2.710
44	3.724	3.425	3.282	3.075	2.905	2.719
45	3.736	3.435	3.292	3.085	2.914	2.727
46	3.747	3.445	3.302	3.094	2.923	2.736
47	3.757	3.455	3.310	3.103	2.931	2.744
48	3.768	3.464	3.319	3.111	2.940	2.753
49	3.779	3.474	3.329	3.120	2.948	2.760
50	3.789	3.483	3.336	3.128	2.956	2.768

(Continued)

TABLE 8 (Continued)

Number of Observations, n	Upper 0.1% Significance Level	Upper 0.5% Significance Level	Upper 1% Significance Level	Upper 2.5% Significance Level	Upper 5% Significance Level	Upper 10% Significance Level
51	3.798	3.491	3.345	3.136	2.964	2.775
52	3.808	3.500	3.353	3.143	2.971	2.783
53	3.816	3.507	3.361	3.151	2.978	2.790
54	3.825	3.516	3.368	3.158	2.986	2.798
55	3.834	3.524	3.376	3.166	2.992	2.804
56	3.842	3.531	3.383	3.172	3.000	2.811
57	3.851	3.539	3.391	3.180	3.006	2.818
58	3.858	3.546	3.397	3.186	3.013	2.824
59	3.867	3.553	3.405	3.193	3.019	2.831
60	3.874	3.560	3.411	3.199	3.025	2.837
61	3.882	3.566	3.418	3.205	3.032	2.842
62	3.889	3.573	3.424	3.212	3.037	2.849
63	3.896	3.579	3.430	3.218	3.044	2.854
64	3.903	3.586	3.437	3.224	3.049	2.860
65	3.910	3.592	3.442	3.230	3.055	2.866
66	3.917	3.598	3.449	3.235	3.061	2.871
67	3.923	3.605	3.454	3.241	3.066	2.877
68	3.930	3.610	3.460	3.246	3.071	2.883
69	3.936	3.617	3.466	3.252	3.076	2.888
70	3.942	3.622	3.471	3.257	3.082	2.893
71	3.948	3.627	3.476	3.262	3.087	2.897
72	3.954	3.633	3.482	3.267	3.092	2.903
73	3.960	3.638	3.487	3.272	3.098	2.908
74	3.965	3.643	3.492	3.278	3.102	2.912
75	3.971	3.648	3.496	3.282	3.107	2.917
76	3.977	3.654	3.502	3.287	3.111	2.922
77	3.982	3.658	3.507	3.291	3.117	2.927
78	3.987	3.663	3.511	3.297	3.121	2.931
79	3.992	3.669	3.516	3.301	3.125	2.935
80	3.998	3.673	3.521	3.305	3.130	2.940
81	4.002	3.677	3.525	3.309	3.134	2.945
82	4.007	3.682	3.529	3.315	3.139	2.949
83	4.012	3.687	3.534	3.319	3.143	2.953
84	4.017	3.691	3.539	3.323	3.147	2.957
85	4.021	3.695	3.543	3.327	3.151	2.961
86	4.026	3.699	3.547	3.331	3.155	2.966
87	4.031	3.704	3.551	3.335	3.160	2.970
88	4.035	3.708	3.555	3.339	3.163	2.973
89	4.039	3.712	3.559	3.343	3.167	2.977
90	4.044	3.716	3.563	3.347	3.171	2.981
91	4.049	3.720	3.567	3.350	3.174	2.984
92	4.053	3.725	3.570	3.355	3.179	2.989
93	4.057	3.728	3.575	3.358	3.182	2.993
94	4.060	3.732	3.579	3.362	3.186	2.996
95	4.064	3.736	3.582	3.365	3.189	3.000
96	4.069	3.739	3.586	3.369	3.193	3.003
97	4.073	3.744	3.589	3.372	3.196	3.006
98	4.076	3.747	3.593	3.377	3.201	3.011
99	4.080	3.750	3.597	3.380	3.204	3.014
100	4.084	3.754	3.600	3.383	3.207	3.017

(Continued)

TABLE 8 (Continued)

Number of Observations, n	Upper 0.1% Significance Level	Upper 0.35 Significance Level	Upper 1% Significance Level	Upper 2.5% Significance Level	Upper 5% Significance Level	Upper 10% Significance Level
101	4.084	3.757	3.603	3.386	3.210	3.021
102	4.092	3.760	3.607	3.390	3.214	3.024
103	4.095	3.765	3.610	3.393	3.217	3.027
104	4.098	3.768	3.614	3.397	3.220	3.030
105	4.102	3.771	3.617	3.400	3.224	3.033
106	4.105	3.774	3.620	3.403	3.227	3.037
107	4.109	3.777	3.623	3.406	3.230	3.040
108	4.112	3.780	3.626	3.409	3.233	3.043
109	4.116	3.784	3.629	3.412	3.236	3.046
110	4.119	3.787	3.632	3.415	3.239	3.049
111	4.122	3.790	3.636	3.418	3.242	3.052
112	4.125	3.793	3.639	3.422	3.245	3.055
113	4.129	3.796	3.642	3.424	3.248	3.058
114	4.132	3.799	3.645	3.427	3.251	3.061
115	4.135	3.802	3.647	3.430	3.254	3.064
116	4.138	3.805	3.650	3.433	3.257	3.067
117	4.141	3.808	3.653	3.435	3.259	3.070
118	4.144	3.811	3.656	3.438	3.262	3.073
119	4.146	3.814	3.659	3.441	3.265	3.075
120	4.150	3.817	3.662	3.444	3.267	3.078
121	4.153	3.819	3.665	3.447	3.270	3.081
122	4.156	3.822	3.667	3.450	3.274	3.083
123	4.159	3.824	3.670	3.452	3.276	3.086
124	4.161	3.827	3.672	3.455	3.279	3.089
125	4.164	3.831	3.675	3.457	3.281	3.092
126	4.166	3.833	3.677	3.460	3.284	3.095
127	4.169	3.836	3.680	3.462	3.286	3.097
128	4.173	3.838	3.683	3.465	3.289	3.100
129	4.175	3.840	3.686	3.467	3.291	3.102
130	4.178	3.843	3.689	3.470	3.294	3.104
131	4.180	3.845	3.690	3.473	3.296	3.107
132	4.183	3.848	3.693	3.475	3.298	3.109
133	4.185	3.850	3.695	3.478	3.302	3.112
134	4.188	3.853	3.697	3.480	3.304	3.114
135	4.190	3.856	3.700	3.482	3.306	3.116
136	4.193	3.858	3.702	3.484	3.309	3.119
137	4.196	3.860	3.704	3.487	3.311	3.122
138	4.198	3.863	3.707	3.489	3.313	3.124
139	4.200	3.865	3.710	3.491	3.315	3.126
140	4.203	3.867	3.712	3.493	3.318	3.129
141	4.205	3.869	3.714	3.497	3.320	3.131
142	4.207	3.871	3.716	3.499	3.322	3.133
143	4.209	3.874	3.719	3.501	3.324	3.135
144	4.212	3.876	3.721	3.503	3.326	3.136
145	4.214	3.879	3.723	3.505	3.328	3.140
146	4.216	3.881	3.725	3.507	3.331	3.142
147	4.219	3.883	3.727	3.509	3.334	3.144

SOURCE: ASTM Designation E178-75, 1975. "Standard Recommended Practice for Dealing With Outlying Observations."

APPENDIX C

GENERAL BIBLIOGRAPHY

The following list provides the reader with those references directly mentioned in the text. It also includes, for those readers desiring further information, references to literature dealing with selected subject matters in a broader sense. This list is in alphabetical order.

ASTM Designation: E178-75. 1975. "Standard Recommended Practice for Dealing with Outlying Observations."

ASTM Manual on Presentation of Data and Control Chart Analysis. 1976. ASTM Special Technical Publication 15D.

Barari, A., and L. S. Hedges. 1985. "Movement of Water in Glacial Till." *Proceedings of the 17th International Congress of the International Association of Hydrogeologists.* pp. 129-134.

Barcelona, M. J., J. P. Gibb, J. A. Helfrich, and E. E. Garske. 1985. "Practical Guide for Ground-Water Sampling." Report by Illinois State Water Survey, Department of Energy and Natural Resources for USEPA. EPA/600/2-85/104.

Bartlett, M. S. 1937. "Properties of Sufficiency and Statistical Tests." *Journal of the Royal Statistical Society, Series A.* 160:268-282.

Box, G. E. P., and J. M. Jenkins. 1970. *Time Series Analysis.* Holden-Day, San Francisco, California.

Brown, K. W., and D. C. Andersen. 1981. "Effects of Organic Solvents on the Permeability of Clay Soils." EPA 600/2-83-016, Publication No. 83179978, U.S. EPA, Cincinnati, Ohio.

Cohen, A. C., Jr. 1959. "Simplified Estimators for the Normal Distribution When Samples Are Singly Censored or Truncated." *Technometrics.* 1:217-237.

Cohen, A. C., Jr. 1961. "Tables for Maximum Likelihood Estimates: Singly Truncated and Singly Censored Samples." *Technometrics.* 3:535-541.

Conover, W. J. 1980. *Practical Nonparametric Statistics.* Second Edition, John Wiley and Sons, New York, New York.

CRC Handbook of Tables for Probability and Statistics. 1966. William H. Beyer (ed.). The Chemical Rubber Company.

Current Index to Statistics. Applications, Methods and Theory. Sponsored by American Statistical Association and Institute of Mathematical Statistics. Annual series providing indexing coverage for the broad field of statistics.

David, H. A. 1956. "The Ranking of Variances in Normal Populations." *Journal of the American Statistical Association.* Vol. 51, pp. 621-626.

Davis, J. C. 1986. *Statistics and Data Analysis in Geology.* Second Edition. John Wiley and Sons, New York, New York.

- Dixon, W. J., and F. J. Massey, Jr. 1983. *Introduction to Statistical Analysis*. Fourth Edition. McGraw-Hill, New York, New York.
- Gibbons, R. D. 1987. "Statistical Prediction Intervals for the Evaluation of Ground-Water Quality." *Ground Water*. Vol. 25, pp. 455-465.
- Gilbert, R. 1987. *Statistical Methods for Environmental Pollution Monitoring*. Professional Books Series, Van Nos Reinhold.
- Heath, R. C. 1983. *Basic Ground-Water Hydrology*. U.S. Geological Survey Water Supply Paper. 2220, 84 p.
- Hirsch, R. M., J. R. Slack, and R. A. Smith. 1982. "Techniques of Trend Analysis for Monthly Water Quality Data." *Water Resources Research*. Vol. 18, No. 1, pp. 107-121.
- Hockman, K. K., and J. M. Lucas. 1987. "Variability Reduction Through Subvessel CUSUM Control." *Journal of Quality Technology*. Vol. 19, pp. 113-121.
- Hollander, M., and D. A. Wolfe. 1973. *Nonparametric Statistical Methods*. John Wiley and Sons, New York, New York.
- Huntsberger, D. V., and P. Billingsley. 1981. *Elements of Statistical Inference*. Fifth Edition. Allyn and Bacon, Inc., Boston, Massachusetts.
- Johnson, N. L., and F. C. Leone. 1977. *Statistics and Experimental Design in Engineering and the Physical Sciences*. 2 Vol., Second Edition. John Wiley and Sons, New York, New York.
- Kendall, M. G., and A. Stuart. 1966. *The Advanced Theory of Statistics*. 3 Vol. Hafner Publication Company, Inc., New York, New York.
- Kendall, M. G., and W. R. Buckland. 1971. *A Dictionary of Statistical Terms*. Third Edition. Hafner Publishing Company, Inc., New York, New York.
- Kendall, M. G. 1975. *Rank Correlation Methods*. Charles Griffin, London.
- Langley, R. A. 1971. *Practical Statistics Simply Explained*. Second Edition. Dover Publications, Inc., New York, New York.
- Lehmann, E. L. 1975. *Nonparametric Statistical Methods Based on Ranks*. Holsten Day, San Francisco, California.
- Lieberman, G. J. 1958. "Tables for One-Sided Statistical Tolerance Limits." *Industrial Quality Control*. Vol. XIV, No. 10.
- Lilliefors, H. W. 1967. "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown." *Journal of the American Statistical Association*. 64:399-402.
- Lingren, B. W. 1976. *Statistical Theory*. Third Edition. McMillan.

- Lucas, J. M. 1982. "Combined Shewhart-CUSUM Quality Control Schemes." *Journal of Quality Technology*. Vol. 14, pp. 51-59.
- Mann, H. B. 1945. "Non-parametric Tests Against Trend." *Econometrica*. Vol. 13, pp. 245-259.
- Miller, R. G., Jr. 1981. *Simultaneous Statistical Inference*. Second Edition. Springer-Verlag, New York, New York.
- Nelson, L. S. 1987. "Upper 10%, 5%, and 1%--Points of the Maximum F-Ratio." *Journal of Quality Technology*. Vol. 19, p. 165.
- Nelson, L. S. 1987. "A Gap Test for Variances." *Journal of Quality Technology*. Vol. 19, pp. 107-109.
- Noether, G. E. 1967. *Elements of Nonparametric Statistics*. Wiley, New York.
- Pearson, E. S., and H. O. Hartley. 1976. *Biometrika Tables for Statistician*. Vol. 1, Biometrika Trust, University College, London.
- Quade, D. 1966. "On Analysis of Variance for the K-Sample Problem." *Annals of Mathematical Statistics*. 37:1747-1748.
- Remington, R. D., and M. A. Schork. 1970. *Statistics with Applications to the Biological and Health Sciences*. Prentice-Hall, pp. 235-236.
- Shapiro, S. S., and M. R. Wilk. 1965. "An Analysis of Variance Test for Normality (Complete Samples)." *Biometrika*. Vol. 52, pp. 591-611.
- Snedecor, G. W., and W. G. Cochran. 1980. *Statistical Methods*. Seventh Edition. The Iowa State University Press, Ames, Iowa.
- Starks, T. H. 1988 (Draft). "Evaluation of Control Chart Methodologies for RCRA Waste Sites." Report by Environmental Research Center, University of Nevada, Las Vegas, for Exposure Assessment Research Division, Environmental Monitoring Systems Laboratory-Las Vegas, Nevada. CR814342-01-3.
- Steel, R. G. D., and J. H. Torrie. 1980. *Principles and Procedures of Statistics, A Biometrical Approach*. Second Edition. McGraw-Hill Book Company, New York, New York.
- Todd, D. K. 1980. *Ground Water Hydrology*. John Wiley and Sons, New York, 534 p.
- Tukey, J. W. 1949. "Comparing Individual Means in the Analysis of Variance." *Biometrics*. Vol. 5, pp. 99-114.

Statistical Software Packages:

BMDP Statistical Software. 1983. 1985 Printing. University of California Press, Berkeley.

SAS: Statistical Analysis System, SAS Institute, Inc.

SAS® User's Guide: Basics, Version 5 Edition, 1985.

SAS® User's Guide: Statistics, Version 5 Edition, 1985.

SPSS: Statistical Package for the Social Sciences. 1982. McGraw-Hill.

SYSTAT: Statistical Software Package for the PC. Systat, Inc., 1800 Sherman Avenue, Evanston, Illinois 60201.