530-SW-89-026

STATISTICAL ANALYSIS OF GROUND-WATER MONITORING DATA AT RCRA FACILITIES

INTERIM FINAL GUIDANCE

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

FEBRUARY 1989

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

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 that meet the regulatory requirements 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. Technical support in the preparation of this document was provided by Midwest Research Institute (MRI) under a subcontract to NUS Corporation, the prime contractor with EPA's Office of Solid Waste. MRI staff who assisted with the preparation of the document were Jairus D. Flora, Jr., Ph.D., Principal Statistician, Ms. Karin M. Bauer, Senior Statistician, and Mr. Joseph S. Bartling, Assistant Statistician.

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 recently amended regulations concerning the statistical analysis of ground-water monitoring data at RCRA facilities (53 FR 39720: October 11, 1988), provide a wide variety of statistical methods that may be used to evaluate ground-water quality. To the experienced and inexperienced water quality professional, the choice of which test to use under a particular set of conditions may not be apparent. The reader is referred to Section 4 of this guidance, "Choosing a Statistical Method," for assistance in choosing an appropriate statistical test. For relatively new facilities that have only limited amounts of ground-water monitoring data, it is recommended that a form of hypothesis test (e.g., parametric analysis of variance) be employed to evaluate the data. Once sufficient data are available (after 12 to 24 months or eight background samples), another method of analysis such as the control chart methodology described in Section 7 of the guidance is recommended. Each method of analysis and the conditions under which they will be used can be written in the facility permit. This will eliminate the need for a permit modification each time more information about the hydrogeochemistry is collected, and more appropriate methods of data analysis become apparent.

This guidance was written primarily for the statistical analysis of ground-water monitoring data at RCRA 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 5 of the guidance describes background well (upgradient) vs. 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_o) 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 Section 5.2 (Analysis of Variance), 5.3 (Tolerance Intervals), and 5.4 (Prediction Intervals).

A different situation exists when compliance wells (downgradient) are compared to a fixed standard (e.g., the ground-water protection standard). In that case, Section 6 of the guidance should be consulted. The value to which the constituent concentrations at compliance wells are compared can be any

Preceding page blank

standard established by a Regional Administrator, State or county health official, or another appropriate official.

A note of caution applies to Section 6. The examples used in Section 6 are used to determine whether ground water has been contaminated as a result of a release from a facility. When the lower confidence limit is exceeded, further action or assessment may be warranted. If one wishes to determine whether 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 (Volume 2: Ground Water--Draft), should be consulted. This draft Superfund guidance is a multivolume set that addresses questions regarding the success of air, ground-water, and soil remediation efforts. Information about the availability of this draft guidance, 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 uncontaminated wells or in an intrawell comparison are referred to Section 7 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 this guidance have wide applications in the field of applied statistics, regardless of the intended use or purpose. Section 4.2 and 4.3 provide information on checking distributional assumptions and equality of variance, while Sections 8.1 and 8.2 cover limit of detection problems and outliers. Helpful advice and references for many experiments involving the use of statistics can be found in these sections.

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 nor 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

FIGURES

TABLES

TABLES (continued)

EXECUTIVE SUMMARY

The hazardous waste regulations under the Resource Conservation and Recovery Act (RCRA) require owners and operators of hazardous waste facilities to utilize design features and control measures that prevent the release of hazardous waste into ground water. Further, regulated units (i.e., all surface impoundments, waste piles, land treatment units, and landfills that receive hazardous waste after July 26, 1982) are also subject to the ground-water monitoring and corrective action standards of 40 CFR Part 264, Subpart F. These regulations require that a statistical method and sampling procedure approved by EPA be used to determine whether there are releases from regulated units into ground water.

This document provides guidance to RCRA Facility permit applicants and writers concerning the statistical analysis of ground-water monitoring data at RCRA facilities. Section 1 is an introduction to the guidance; it describes the purpose and intent of the document, and emphasizes the need for site-specific considerations in implementing the Subpart F regulations of 40 CFR Part 264.

Section 2 provides the reader with an overview of the recently promulgated regulations concerning the statistical analysis of ground-water monitoring data (53 FR 39720: October 11, 1988). The requirements of the regulation are reviewed, and the need to consider site specific factors in evaluating data at a hazardous waste facility is emphasized.

Section 3 discusses the important hydrogeologic parameters to consider when choosing a sampling interval. The Darcy equation is used to determine the horizontal component of the average linear velocity of ground water. This parameter provides a good estimate of time of travel for most soluble constituents in ground water, and may be used to determine a sampling interval. Example calculations are provided at the end of the section to further assist the reader.

Section 4 provides guidance on choosing an appropriate statistical method. A flowchart to guide the reader through this section, as well as procedures to test the distributional assumptions of data are presented. Finally, this section outlines procedures to test specifically for equality of variance.

Section 5 covers statistical methods that may be used to evaluate ground-water monitoring data when background wells have been sited hydraulically upgradient from the regulated unit, and a second set of wells are sited hydraulically downgradient from the regulated unit at the point of compliance. The data from these compliance wells are compared to data from the background wells to determine whether a release from a facility has occurred. Parametric and nonparametric analysis of variance, tolerance intervals, and prediction intervals are suggested methods for this type of comparison. Flowcharts, procedures and example calculations are given for each testing method.

Section 6 includes statistical procedures that are appropriate when comparing ground-water constituent concentrations to fixed concentration limits (e.g., alternate concentration limits or maximum concentration limits). The methods applicable to this type of comparison are confidence intervals and tolerance intervals. As in section 5, flowcharts, procedures, and examples explain the calculations necessary for each testing method.

Section 7 presents the case where the level of each constituent within a single, uncontaminated well is being compared to its historic background concentrations. This is known as an intra-well comparison. 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 in this section is a combined Shewhart-CUSUM control chart.

Section 8 contains a variety of special topics that are relatively short and self contained. These topics include methods to deal with data that is below the limit of analytical detection and methods to test for outliers or extreme values in the data.

Finally, the guidance presents appendices that cover general statistical considerations, a glossary of statistical terms, statistical tables, and a listing of references. These appendices provide necessary and ancillary information to aid the user in evaluating ground-water monitoring data.

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 revised 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 recently amended the Subpart F regulations with statistical methods and sampling procedures that are appropriate for evaluating ground-water monitoring data under a variety of situations (53 FR 39720: October 11, 1988). 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 [40 CFR §264.91(a)(1) and (2)].

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, State Directors, 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

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

EPA would like to emphasize the need for site-specific considerations in implementing the Subpart F regulations of 40 CFR Part 264 (especially as amended, 53 FR 39720: October 11, 1988). It has been an ongoing strategy to promulgate regulations that are specific enough to implement, yet flexible enough to accommodate a wide variety of site-specific environmental factors. This is usually achieved by specifying criteria that are appropriate for the majority of monitoring situations, while at the same time allowing alternatives that are also protective of human health and the environment. This philosophy is maintained in the recently promulgated amendments entitled, "Statistical Methods for Evaluating Ground-Water Monitoring Data From Hazardous Waste Facilities" (53 FR 39720: October 11, 1988). The sections that allow for the use of an alternate sampling procedure and statistical method [§264.97(g)(2) and §264.97(h)(5), respectively] are as viable as those that are explicitly referenced [§264.97(g)(1) and §264.97(h)(1-4)], provided they meet the performance standards of §264.97(i). Due consideration to this should be given when preparing and reviewing Part B permits and permit applications.

SECTION 2

REGULATORY OVERVIEW

In 1982, EPA promulgated ground-water monitoring and response standards for permitted facilities in Subpart F of 40 CFR Part 264, for detecting releases of hazardous wastes into ground water from storage, treatment, and disposal units, at permitted facilities (47 FR 32274: July 26, 1982).

The Subpart F regulations required ground-water data to be examined by Cochran's Approximation to the Behrens-Fisher Student's t-test (CABF) 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 this procedure 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 CABF 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 establish 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 Part 264 regulations prior to the October 11, 1988 amendments 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 could occur because the background data, which are often used as the basis of the statistical comparisons, are highly variable due to temporal, spatial, analytical, and sampling effects.

As a result of these concerns, on October 11, 1988 EPA amended both the statistical methods 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. 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 site-specific conditions at the facility, and to ensure that it meets the performance standards outlined below. This demonstration holds for any of the 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 the regulations promulgated on October 11, 1988. 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 prohibit 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 have been collected. This will eliminate the need for a permit modification in the event that data collected during future sampling and analysis events indicate the need to change to a more appropriate statistical method of analysis.

2.3 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 contained in 40 CFR §264.97(i):

1. The statistical test is to be conducted separately for each hazardous constituent in each well [under §264.97(g)]. 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 a 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, then the levels of confidence shall be proposed; in addition, for tolerance intervals, the proportion of the population that the interval must contain (with the proposed confidence) 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 healtn and the environment. Any practical quantitation limit (PQL) approved by the Regional Administrator under §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 available to the facility.

6. If necessary, the statistical method shall include procedures to control or correct for seasonal and spatial variability as well as 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 parameters 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 comment. For individual well comparisons in which an individual compliance well is compared to background, the Type I error level shall be no less than 1% (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 5% (0.05) for each testing period.

In using 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 1% (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 (e.g., Bonferroni t-test).

Setting these levels of significance at 1% and 5%, 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 which are monitored for many constituents may still generate a large number of comparisons during each testing period.

In these particular situations, a determination of whether a release from a facility has occurred may require the Regional Administrator to evaluate the site hydrogeology, geochemistry, climatic factors, and other environmental parameters to determine if a statistically significant result is indicative of an actual release from the facility. In making this determination, 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 is the same relative magnitude as the PQL (practical quantitation limit) or the background concentration level, then a false positive may have occurred, and further sampling and testing may be appropriate. If, however, the background concentration level or an action level is substantially exceeded, then the exceedance is more likely to be indicative of a release from the facility.

2.4 BASIC STATISTICAL METHODS AND SAMPLING PROCEDURES

The October 11, 1988 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 virtually all facilities. 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.

2.4.1 The Five Statistical Methods Outlined in the October 11, 1988 Final Rule

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 sections 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 that Cochran's Approximation to the Behrens-Fisher Student's t-test (CABF) and the accompanying sampling procedures in the past

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	5.2 5.3 5.4
	INTRA-WELL	CONTROL CHARTS	7
ACL/MCL SPECIFIC	FIXED STANDARD	CONFIDENCE INTERVALS TOLERANCE LIMITS	6.2.1 6.2.2
SYNTHETIC	MANY NONDETECTS IN DATA SET	SEE BELOW DETECTION LIMIT TABLE 8-1	8.1

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 statistical methods and procedures consider and control for natural temporal and spatial variation.

2.4.2 Site-Specific Considerations for Sampling

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, an alternate sampling procedure approved by the Regional Administrator may be utilized [40 CFR §264.97(g)(1) and (2)].

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

[40 CFR §264.97(g)]. For example, the use of control charts presumes 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. This approach will help reduce the components of seasonal variation by providing for simultaneous comparisons between background and compliance well information.

The flexibility for establishing 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 [40 CFR §264.97(g)]. 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 parameters as sampling frequency and sample size.

2.4.3 The "Reasonable Confidence" Requirement

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 or her 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.

SECTION 3

CHOOSING A SAMPLING INTERVAL

This section discusses the important hydrogeologic parameters to consider when choosing a sampling interval. The Darcy equation is used to determine the horizontal component of the average linear velocity of ground water. This value provides a good estimate of time of travel for most soluble constituents in ground water, and can be used to determine a sampling interval. Example calculations are provided at the end of the section to further assist the reader.

Section 264.97(g) of 40 CFR Part 264 Subpart F provides the owner or operator of a 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 semiannual 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," Barcelona 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 (V) to effective porosity (Ne), hydraulic gradient (i), and hydraulic conductivity (K):

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

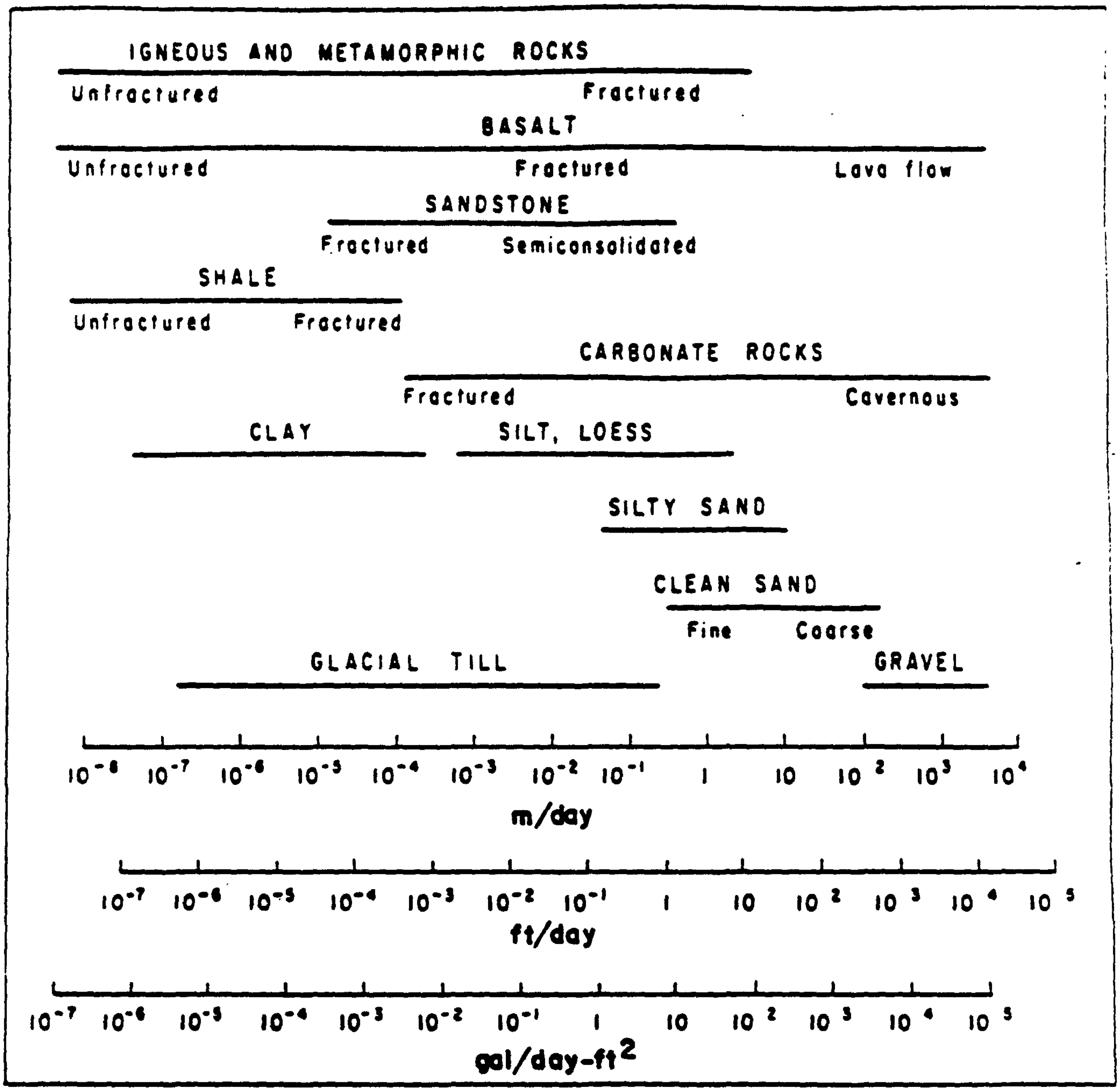
where V_h and V_v are the horizontal and vertical components of the average linear velocity of ground water, respectively; K_h and K_v are the horizontal and vertical components of hydraulic conductivity; i is the head gradient; and Ne is the effective porosity. In applying these equations to ground-water monitoring, the horizontal component of the average linear velocity (V_h) 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 the average linear velocity of ground water (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 Ne. 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 3-1. The conductivities are given in three units: the top line is in meters per day; the middle line, in feet per day, is commonly used; the last line is expressed in gallons per day-foot-squared.

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, Ne, 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 3-1 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 3-2. In the absence of measured values, drainable porosity is often used to approximate effective porosity. Figure 3-2 illustrates representative values of drainable porosity and total porosity as a function of aquifer particle size.

Once the values for K, i, and Ne are determined, the horizontal component of the average linear velocity of ground water 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 horizontal component of the average linear velocity of ground water. This value will represent the <u>minimum</u> time interval required between sampling events that will yield an independent ground-water sample.



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

Figure 3-1. Hydraulic conductivity (in three units) of selected rocks.

TABLE 3-1. DEFAULT VALUES FOR EFFECTIVE POROSITY (Ne) FOR USE IN TIME OF TRAVEL (TOT) ANALYSES

Soil textural classes	Effective porosity of saturation[a]
Unified soil classification system	
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]
USDA soil textural classes	
Clays, silty clays, sandy clays	0.01 (1%)[b]
Silts, silt loams, silty clay loams	0.10 (10%)
All others	0.20 (20%)
Rock units (all)	
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.

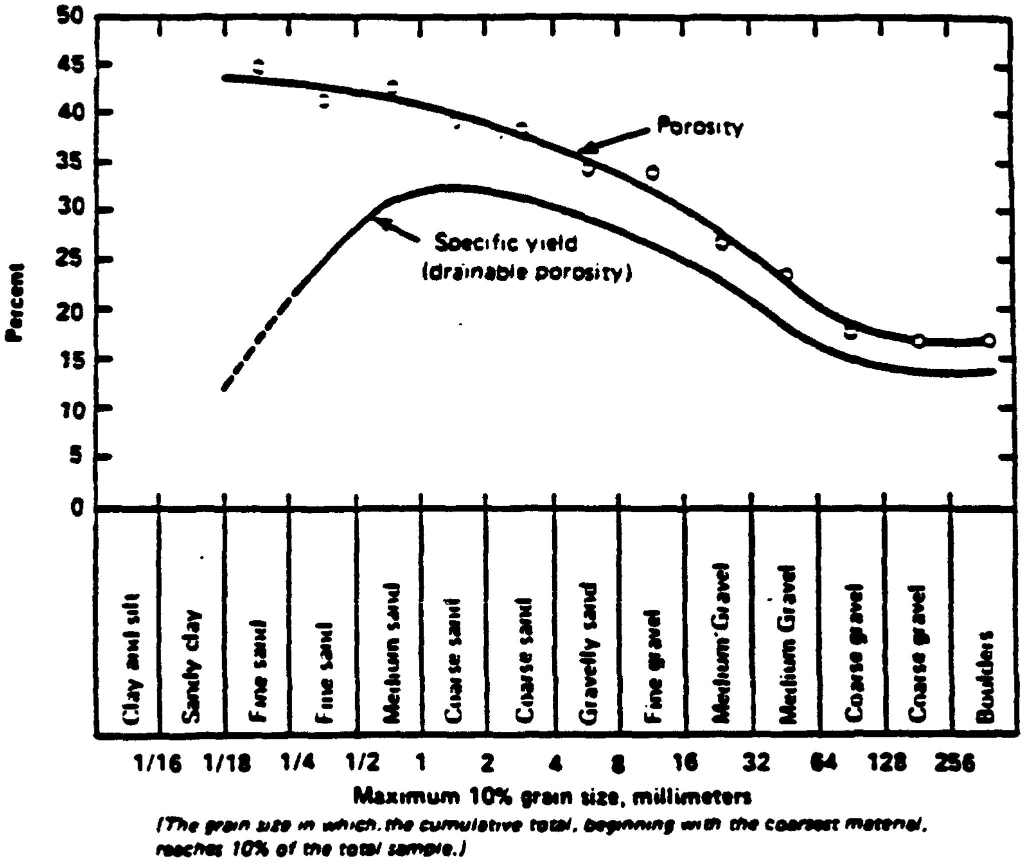
[a] These values are estimates and there may be differences between similar units. For example, recent studies indicate that weathered and unweathered glacial till may have markedly different effective porosities (Barari and Hedges, 1985; Bradbury et al., 1985).

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

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

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 3-2. Total porosity and drainable porosity for typical geologic materials.

(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. 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.

EXAMPLE CALCULATION NO. 1: DETERMINING THE EFFECTIVE POROSITY (Ne)

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

$$Ne = 100\% \times \text{volume of water removed/volume of aquifer dewatered}$$

- Based on a pumping rate of the pump of 50 gal/min and a pumping duration of 30 min, compute the volume of water removed as:

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

- To calculate the volume of aquifer dewatered, use the formula:

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

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

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

Next, converting ft^3 of water to gallons of water,

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

- Substituting the two volumes in the equation for the effective porosity, obtain

$$Ne = 100\% \times 1{,}500/7{,}615 = 19.7\%$$

EXAMPLE CALCULATION NO. 2: DETERMINING THE HYDRAULIC GRADIENT (i)

The hydraulic gradient, i, can be determined from a potentiometric surface map (Figure 3-3 below) as $i = \Delta h/\ell$, where Δh is the difference measured in the gradient at Pz_1 and Pz_2, and ℓ is the orthogonal distance between the two piezometers.

Using the values given in Figure 3-3, obtain

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

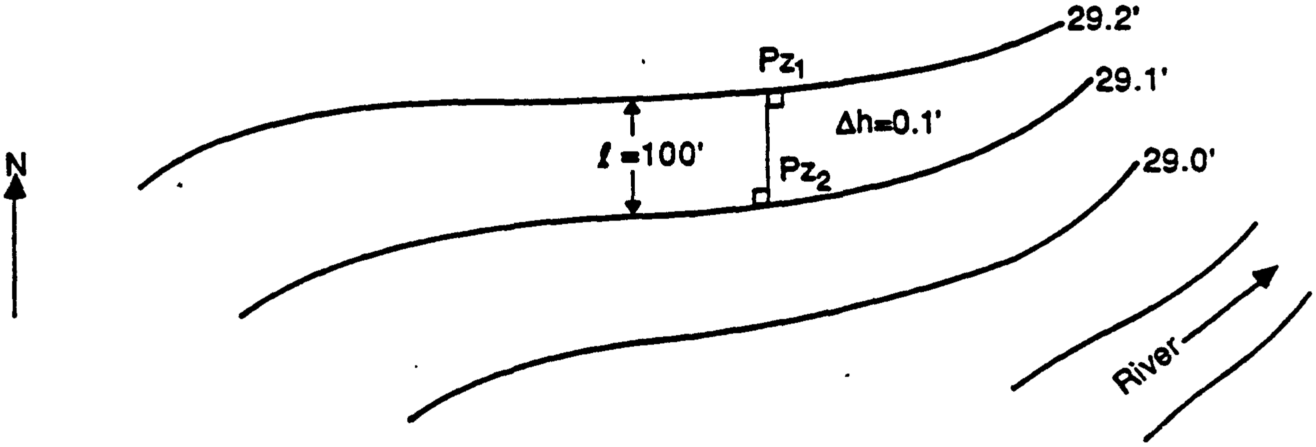


Figure 3-3. Potentiometric surface map for computation of hydraulic gradient.

This method provides only a very general estimate of the natural hydraulic gradient that exists in the vicinity of the two piezometers. Chemical gradients are known to exist and may override the effects of the hydraulic gradient. A detailed study of the effects of multiple chemical contaminants may be necessary to determine the actual average linear velocity (horizontal component) of ground water in the vicinity of the monitoring wells.

EXAMPLE CALCULATION NO. 3: DETERMINING THE HORIZONTAL COMPONENT OF THE AVERAGE LINEAR VELOCITY OF GROUND WATER (V_h)

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 horizontal hydraulic conductivity (K_h) of 15 ft/day and an effective porosity (Ne) of 15%. Using a potentiometric map (as in example 2), the regional hydraulic gradient (i) has been determined to be 0.003 ft/ft.

To estimate the minimum time interval between sampling events that will allow one to obtain an independent sample of ground water proceed as follows.

Calculate the horizontal component of the average linear velocity of ground water (V_h) using the Darcy equation, $V_h = (K_h * i)/Ne$.

With $K_h = 15$ ft/day,

$Ne = 15\%$, and

$i = 0.003$ ft/ft, calculate

$V_h = (15)(0.003)/(15\%) = 0.3$ ft/day, or equivalently

$V_h = (0.3$ ft/day$)(12$ in/ft$) = 3.6$ in/day

Discussion: The horizontal component of the average linear velocity of ground water, V_h, has been calculated and is equal to 3.6 in/day. 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 horizontal component of the average linear velocity of ground water:

Minimum time interval = (4 in)/(3.6 in/day) = 1.1 days

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

Date	Obtain Sample No.
June 1	1
June 8	2
June 15	3
June 22	4

Table 3-3 gives some results for common situations.

TABLE 3-3. DETERMINING A SAMPLING INTERVAL

DETERMINING A SAMPLING INTERVAL

UNIT	K_h (ft/day)	Ne (%)	V_h (in/mo)	SAMPLING INTERVAL
GRAVEL	10^4	19	9.6×10^4	DAILY
SAND	10^2	22	8.3×10^2	DAILY
SILTY SAND	10	14	1.3×10^2	WEEKLY
TILL	10^{-3}	2	9.1×10^{-2}	MONTHLY *
SS (SEMICON)	1	6	30	WEEKLY
BASALT	10^{-1}	8	2.28	MONTHLY *

The horizontal component of the average linear velocities is based on a hydraulic gradient, i, of 0.005 ft/ft.

* Use a Monthly sampling interval or an alternate sampling procedure.

SECTION 4

CHOOSING A STATISTICAL METHOD

This section discusses the choice of an appropriate statistical method. Section 4.1 includes a flowchart to guide this selection. Section 4.2 contains procedures to test the distributional assumptions of statistical methods and Section 4.3 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 (ANOVA). 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, alternate approaches are allowed. These include adjustments for seasonality, use of tolerance intervals, and use of prediction intervals. These methods are discussed in Section 5.

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 Section 6.

When a long history of data from each well is available, intra-well comparisons are appropriate. That is, the data from a single uncontaminated 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 Section 7.

4.1 FLOWCHARTS--OVERVIEW AND USE

The selection and use of a statistical procedure for ground-water monitoring is a detailed process. Because a single flowchart would become too complicated for easy use, a series of flowcharts has been developed. These flowcharts are found at the beginning of each section and are intended to guide the user in the selection and use of procedures in that section. The more detailed flowcharts can be thought of as attaching to the general flowcharts at the indicated points.

Three general types of statistical procedures are presented in the flowchart overview (Figure 4-1): (1) background well to compliance well data comparisons; (2) comparison of compliance well data with a constant limit such as an alternate concentration limit (ACL) or a maximum concentration limit (MCL); and (3) intra-well comparisons. 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 well 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 Section 5. 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 8.1.3) may be recommended. If the constituent occurs in measurable concentrations in background, then analysis of variance (Section 5.2) 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 Section 5. The preferred method is the appropriate analysis of variance method in Section 5.2, or if sufficient data permit, tolerance intervals or control charts. The flow chart in Section 5 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 Section 6, 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. This method of analysis may be used when sufficient data from an individual uncontaminated 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. These methods are presented in Section 7.

FLOWCHART OVERVIEW

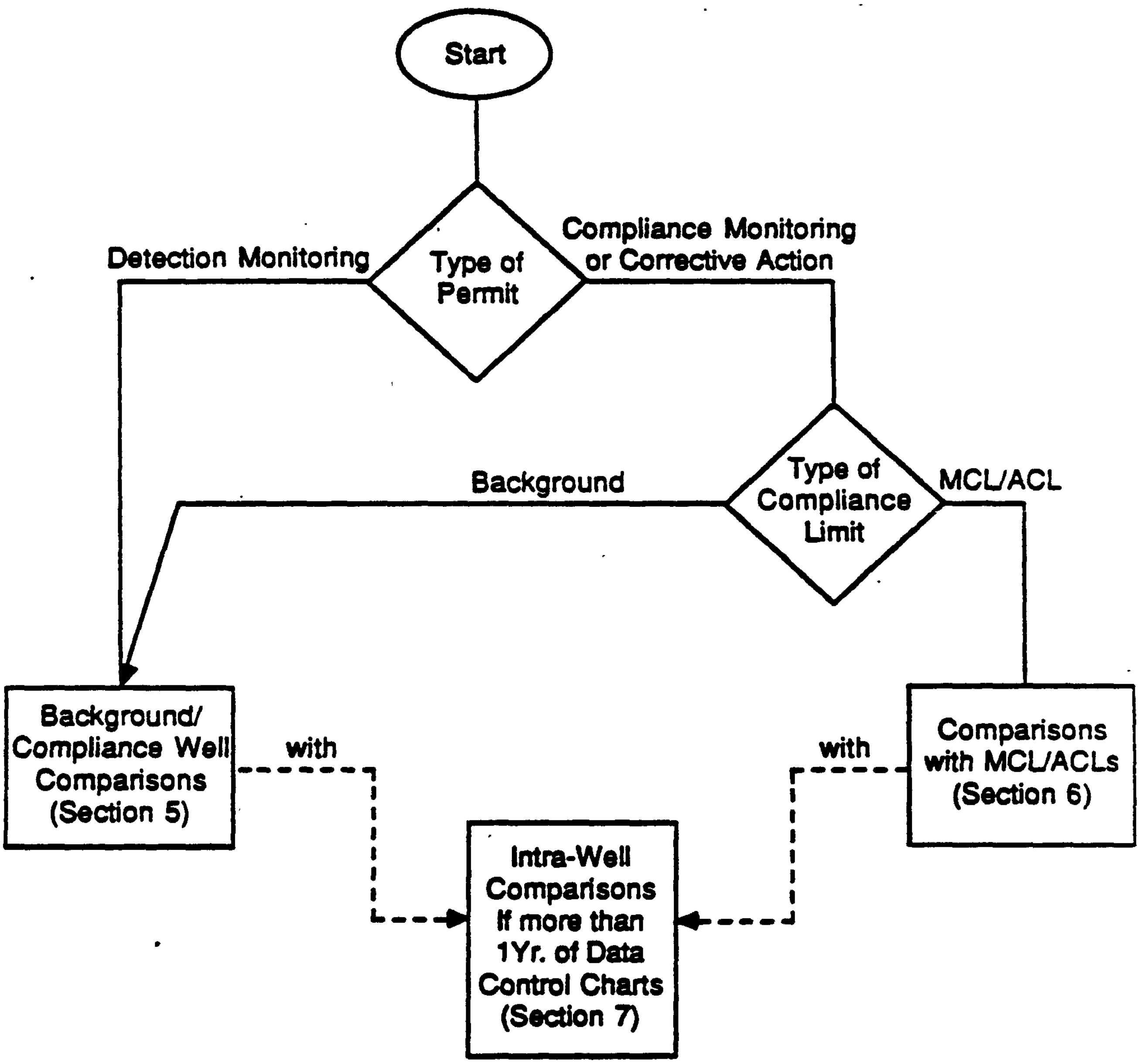


Figure 4-1. Flowchart overview.

4.2 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 in lieu of using a nonparametric method, 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 normal 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 of a normal distribution on a wide scale.

It should be noted that the normal 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. A residual is the difference between the original observation and the value 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 lognormal distribution. That is, the natural logarithm of the original observations has been found to follow the normal distribution. Consequently, if the normal 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 should use a 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).

4.2.1 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.

To justify the use of a literature citation, the owner or operator needs to make sure that the reference cited considers the distribution of data for the specific compound being monitored. In addition, he or she must evaluate the similarity of their site to the site that was discussed in the literature, especially similar hydrogeologic and potential contaminant characteristics. 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.

4.2.2 Coefficient-of-Variation Test

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, $i=1, \ldots, n$.

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

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 4-1 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.

TABLE 4-1. EXAMPLE DATA FOR COEFFICIENT-OF-VARIATION TEST

	Chlordane concentration (ppm)
Dissolved phase	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
Immiscible phase	2.58
	2.69
	2.80
	3.33
	4.50
	6.60

Applying the procedure steps to the data of Table 4-1, 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.

4.2.3 Plotting on Probability Paper

PURPOSE

Probability paper is a visual aid and 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; $X_1, X_2, \ldots, X_i, \ldots, X_n$ 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:

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

Step 2. Compute the cumulative frequency for each distinct value X(i) as (i/(n+1)) x 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*1/(n+1), but the cumulative frequency for X(2) or X(3) is 100*(1+2)/(n+1).

Step 3. Plot the distinct pairs [X(i), (i/n+1)) x 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 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 4-2 lists 22 distinct chlordane concentration values (X) along with their frequencies. These are the same values as those listed in Table 4-1. There is a total of n=24 observations.

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

Step 2. Compute [100 x (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, 100x(i/25)] on probability paper (Figure 4-2).

INTERPRETATION

The points in Figure 4-2 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.

Also, information about the solubility of chlordane in this example is helpful. Chlordane has a solubility (in water) that ranges between 0.0156 and 1.85 mg/L. Because the last six measurements exceed this solubility range, contamination is suspected.

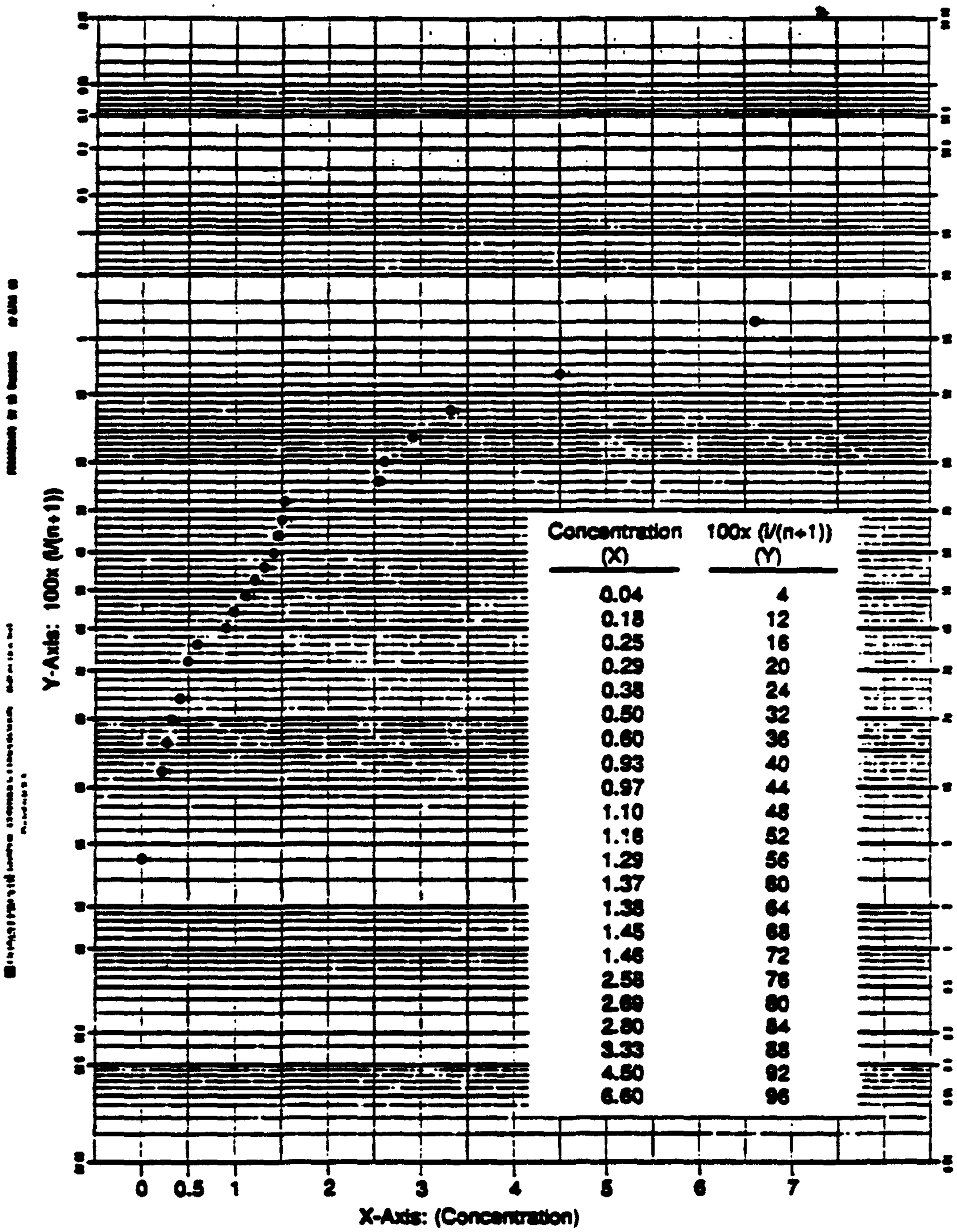
Next, take the natural logarithm of the X-values (ln(X)) (column 5 in Table 4-2). Repeat Step 3 above using the pairs [ln(X), 100x(i/25)]. The resulting plot is shown in Figure 4-3. The points fall approximately on a straight line (hand-drawn) and the hypothesis of lognormality of X, i.e., ln(X) is normally 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.

4.2.4 The Chi-Squared Test

The chi-squared test can be used to test whether a set of data properly fits a specified distribution within a specified probability. 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

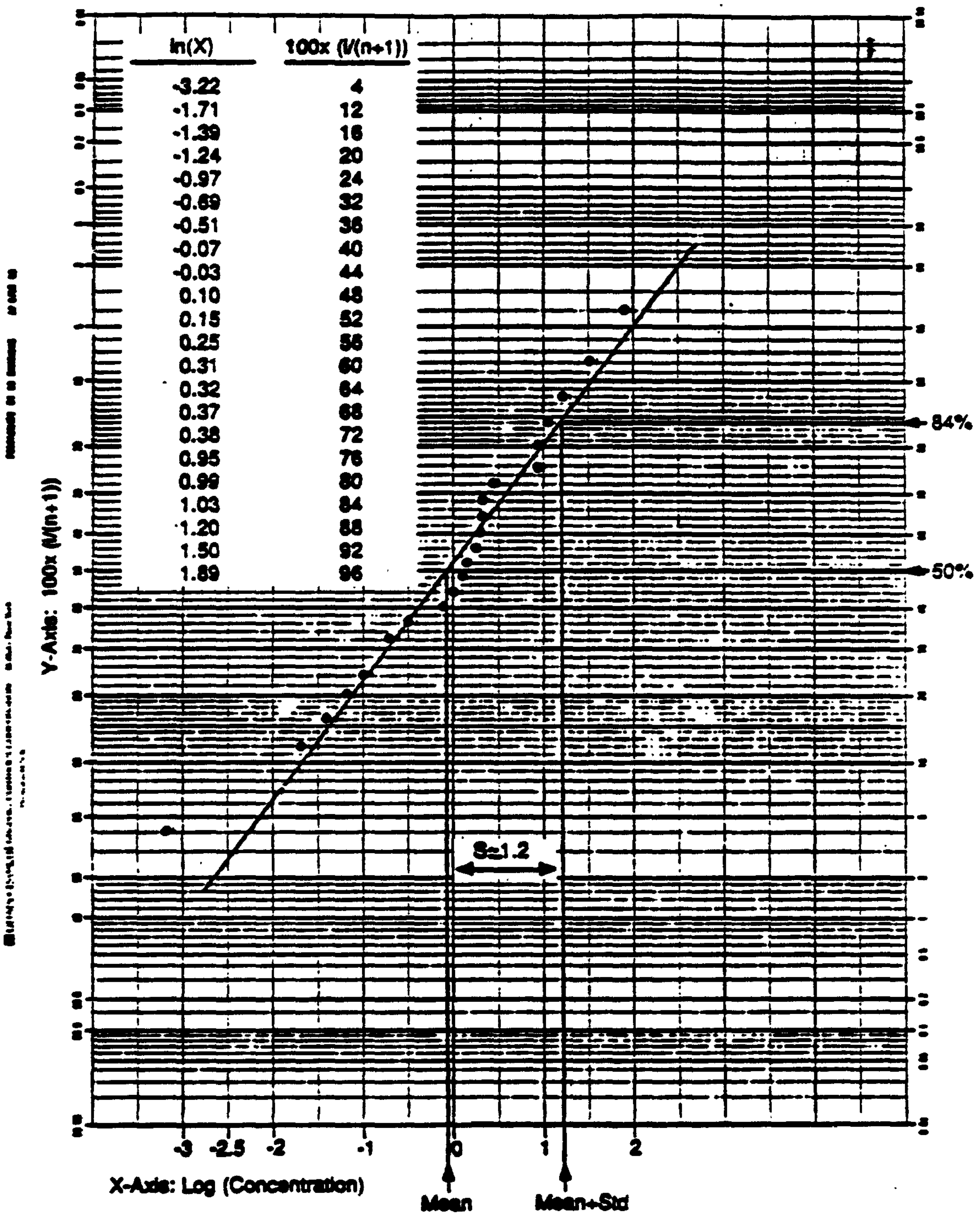
TABLE 4-2. EXAMPLE DATA COMPUTATIONS FOR PROBABILITY PLOTTING

	Concentration X	Absolute frequency	i	100x(i/(n+1))	ln(X)
Dissolved phase	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
Immiscible phase	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



Concentration (X)	100x (I/(n+1)) (Y)
0.04	4
0.18	12
0.25	16
0.29	20
0.38	24
0.50	32
0.60	36
0.93	40
0.97	44
1.10	48
1.16	52
1.29	56
1.37	60
1.38	64
1.45	68
1.46	72
2.58	76
2.69	80
2.80	84
3.33	88
4.50	92
6.60	96

Figure 4-2. Probability plot of raw chlordane concentrations.



ln(X)	100x (I/(n+1))
-3.22	4
-1.71	12
-1.39	16
-1.24	20
-0.97	24
-0.69	32
-0.51	36
-0.07	40
-0.03	44
0.10	48
0.15	52
0.25	56
0.31	60
0.32	64
0.37	68
0.38	72
0.95	76
0.99	80
1.03	84
1.20	88
1.50	92
1.89	96

Figure 4-3. Probability plot of log-transformed chlordane concentrations.

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. Dividing the total number of observations by 4 will guarantee a minimum of four observations necessary for each of the K = N/4 cells. 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 4-3. 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.

TABLE 4-3. CELL BOUNDARIES FOR THE CHI-SQUARED TEST

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

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.

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 4-4 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 4-4.

Step 3. Determine the number of (standardized) observations that fall into the five cells determined from Table 4-3. 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 $21/5 = 4.2$.

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

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.

TABLE 4-4. 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

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.

4.3 CHECKING EQUALITY OF VARIANCE: BARTLETT'S TEST

The analysis of variance procedures presented in Section 5 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. Homogeneity of variances 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 itself 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 $S_1^2,\ldots,S_k^2$. The sample variance, S^2, is the square of the sample standard deviation and is given by the general equation

$$S^2 = \sum_{i=1}^{n} (X_i-\overline{X})^2/(n-1)$$

where $\overline{X}$ is the average of the $X_1,\ldots,X_n$ values. Each variance has associated with it $f_i = n_i-1$ degrees of freedom. Take the natural logarithm of each variance, $\ln(S_1^2),\ldots,\ln(S_k^2)$.

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

$$S_p^2 = \frac{1}{f} \sum_{i=1}^{k} f_i S_i^2$$, is 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, 5%.

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 4-5 below.

TABLE 4-5. 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, $\ln(S_1^2))$,..., $\ln(S_6^2)$, 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 4-5.

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

- Compute S_p^2

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

- Take the natural logarithm of S_p^2: $\ln(S_p^2) = 14$

- Compute $X^2 = 11(14) - 105 = 44$

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 $S_1^2, \ldots, S_6^2$, 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, $n_i=n$, $i=1,\ldots,k$. 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 5

BACKGROUND WELL TO 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 uncontaminated wells can be defined. 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 type of comparison is the general case for detection monitoring. It is also the usual approach for compliance monitoring if the compliance limits are determined by the background well constituent concentration levels. Compounds that are present in background wells (e.g., naturally occurring metals) are most appropriately evaluated using this comparison method.

Section 5.1 provides a flowchart and overview for the selection of methods for comparison of background well and compliance well data. Section 5.2 contains analysis of variance methods. These provide methods for directly comparing background well data to compliance well data. Section 5.3 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 5.4 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 6.

5.1 SUMMARY FLOWCHART FOR BACKGROUND WELL TO COMPLIANCE WELL COMPARISONS

Figure 5-1 is a flowchart 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 8.1 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 alternate 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 5-2 is a flowchart to guide the user if a tolerance limits approach is selected. The first step in using Figure 5-2 is to determine whether the facility is in detection monitoring. If so, much of the data may be below the detection limit. See Section 8.1 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 4.2), 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 5.4 for guidance in performing the procedure.

If analysis of variance is to be used, then continue with Figure 5-1 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 WELL TO COMPLIANCE WELL COMPARISONS

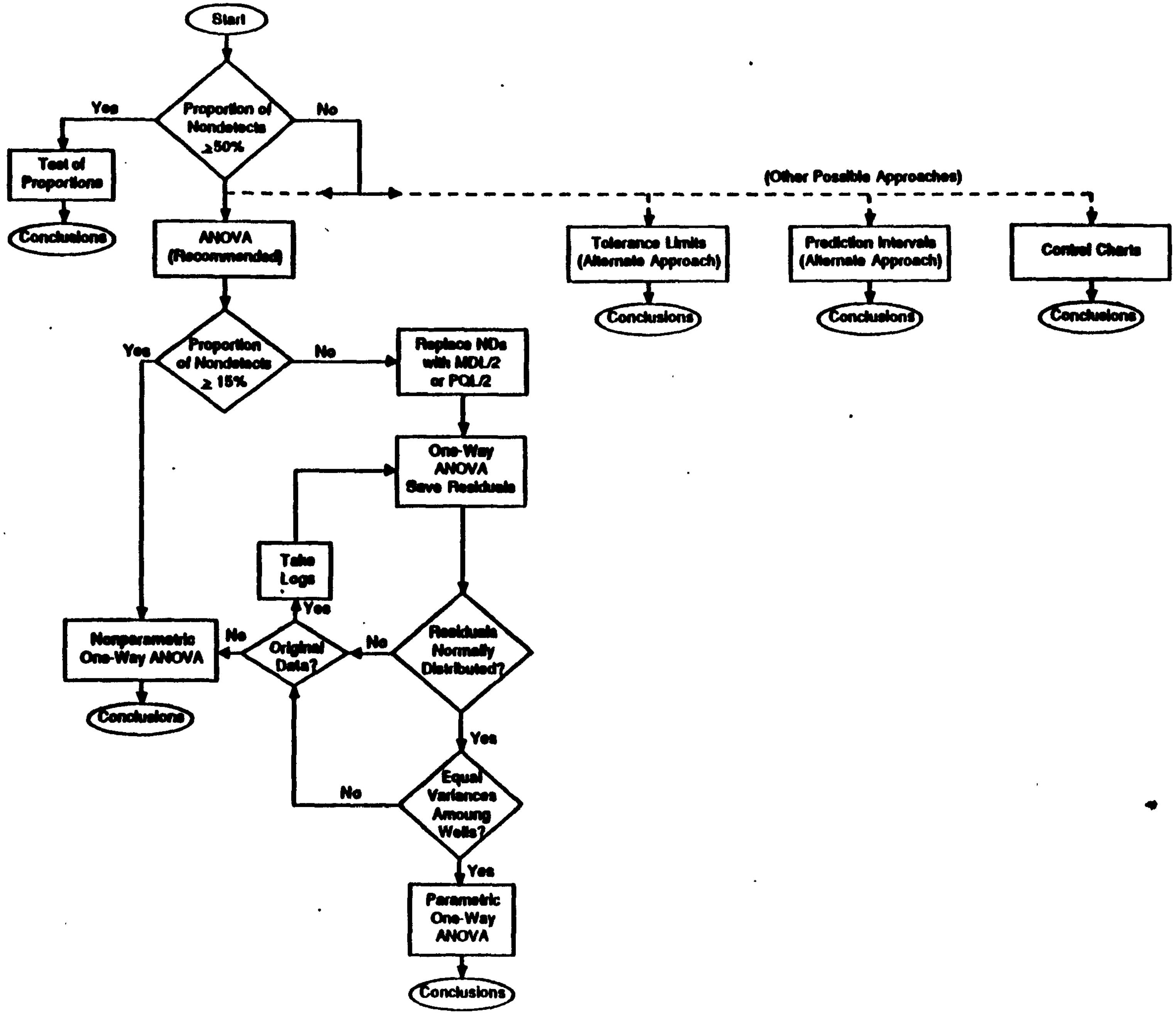


Figure 5-1. Background well to compliance well comparisons.

Tolerance Limits: Alternate Approach to Background Well To Compliance Well Comparisons

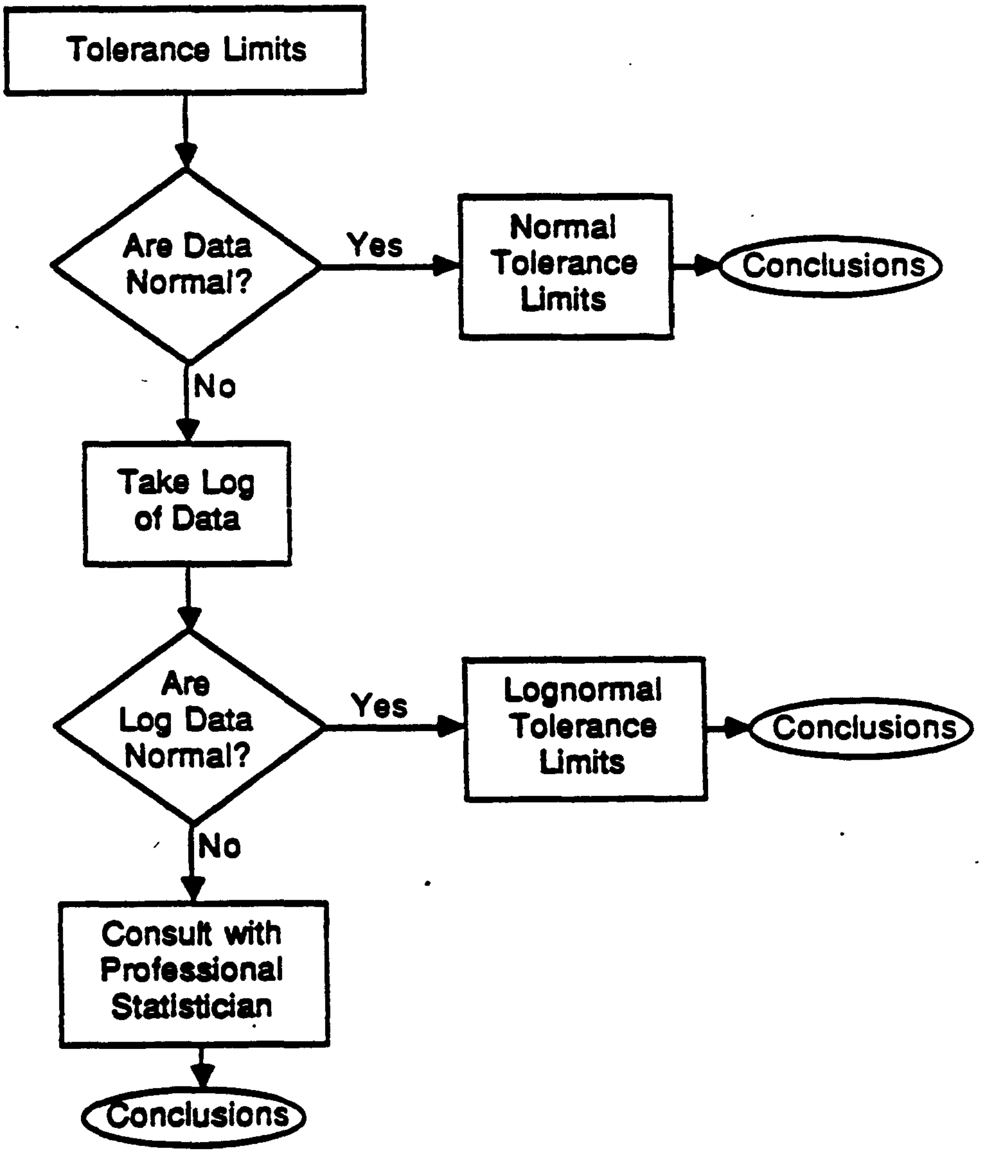


Figure 5-2. Tolerance limits: alternate approach to background well to compliance well comparisons.

5.2 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 4. 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 in Section 5.2.2.

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.

5.2.1 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 ith 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 commercially available 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 (N is the total sample size at this specific regulated unit):

	Observations	Well Total (from Step 1)	Well Mean (from Step 2)
Well No. 1	$X_{11}\cdots\cdots\cdots\cdots X_{1n_1}$	$X_{1.}$	$\bar{X}_{1.}$
2	.		
3	.		
.	.		
u	X_{u1}	$X_{u.}$	$\bar{X}_{u.}$
.	.		
.	.		
p	$X_{p1}\cdots\cdots\cdots\cdots X_{pn_p}$	$X_{p.}$	$\bar{X}_{p.}$
		$X_{..}$	$\bar{X}_{..}$

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

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

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

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

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

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

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

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

(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 <u>between</u> wells.

Step 4. Compute the corrected total sum of squares

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

(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 <u>within</u> 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 <u>within</u> wells.

Step 6. Set up the ANOVA table as shown below in Table 5-1. 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 5-1. 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 5-1). 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 <u>contrasts</u> 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.

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

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

$$\overline{X}_{i.} - \overline{X}_{up}\ , \quad i = 1, \ldots, m$$

- 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 5-1) 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 $\alpha = 0.05$ and (N-p) degrees of freedom.

- 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 at m = 5.

INTERPRETATION

If the difference $\bar{X}_{i.} - \bar{X}_{up}$ exceeds the value D_i, conclude that the ith 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 5-2 below. The wells consist of u=2 background and m=4 compliance wells. (The values in Table 5-2 are actually the natural logarithms of the original lead concentrations.)

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

TABLE 5-2. EXAMPLE DATA FOR ONE-WAY PARAMETRIC ANALYSIS OF VARIANCE

		Natural log of Pb concentrations (µg/L)						
Well No.	Date:	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)
						$X_{..} = 106.08$	$\bar{X}_{..} = 4.42$	

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.

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

with [6 (wells) - 1] = 5 degrees of freedom.

Step 4. Compute the corrected total sum of squares.

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

with [24 (observations) - 1] = 23 degrees of freedom.

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

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

with [24 (observations) - 6 (wells)] = 18 degrees of freedom.

Step 6. Set up the one-way ANOVA as in Table 5-3 below:

TABLE 5-3. 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.

$$SE_i = [0.34\ (1/8 + 1/4)]^{1/2} = 0.357 \text{ for } i = 3,\ldots,6$$

- 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.

$$D_i = SE_i \times t = 0.357 \times 2.43 = 0.868 \text{ for } i = 3, \ldots, 6$$

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.

5.2.2 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 5.2.1) on the ranks. Such an approach can be done with some commercially 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 jth observation in the ith 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 N observations of the groups from least to greatest. Let R_{ij} denote the rank of the jth observation in the ith 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 ith group R_i. Also calculate the average rank for each group, $\bar{R}_i = R_i/n_i$.

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 $\bar{R}_i-\bar{R}_1$ 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. 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.)

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 on the computed test statistic 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). In the present context, the term "negligible" can be more specifically defined as follows. Compute the Kruskal-Wallis statistic without the adjustment for ties. If the test statistic is significant at the 5% level then conclude the test since the statistic with correction for ties will be significant as well. If the test statistic falls between the 10% and the 5% critical values, then proceed with the adjustment for ties as shown below.

ADJUSTMENT FOR TIES

If there are 50% or more observations that fell below the detection limit, then this method for adjustment for ties is inappropriate. The user is referred to Section 8 "Miscellaneous Topics." Otherwise, if there are tied values present in the data, use the following correction for the H statistic

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

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 5-4 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 5-4. 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 5-4.

Step 3. Compute the Kruskal-Wallis statistic

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

ADJUSTMENT FOR TIES

There are four groups of ties in the data of Table 5-4:

$T_1 = (2^3-2) = 6$ for the 2 observations of 1,900.
$T_2 = (2^3-2) = 6$ for the 2 observations of 1,500.
$T_3 = (3^3-3) = 24$ for the 3 observations of 1,300.
$T_4 = (2^3-2) = 6$ for the 2 observations of 0.

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

and $H' = \frac{14.68}{1-(42/(20^3-20))} = \frac{14.68}{0.995} = 14.76$, 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

TABLE 5-4. EXAMPLE DATA FOR ONE-WAY NONPARAMETRIC ANOVA--BENZENE CONCENTRATIONS (ppm)

Date	Background	Compliance wells				
	Well 1	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.

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, $\bar{R}_{i.}$ 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 α = 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:

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

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 5.2.1 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.

5.3 TOLERANCE INTERVALS BASED ON THE NORMAL DISTRIBUTION

An alternate 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 (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 normally distributed. 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 4.2. 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 4.2 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 antilogarithm.

REFERENCE

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

EXAMPLE

Table 5-5 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 1 and 2, 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 statistically significant 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 statistically significant evidence of contamination at Wells 1, 3, and 4.

TABLE 5-5. 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
Mean = 51.4
SD = 16.3

The upper 95% coverage tolerance limit with tolerance coefficient of 95% is 51.4 + (3.188)(16.3) = 103.4

* Indicates contamination

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 statistically significant evidence of contamination.

5.4 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. This approach is algebraically equivalent to the average replicate (AR) test that is presented in the Technical Enforcement Guidance Document (TEGD), September 1986. In ground-water monitoring, a prediction interval approach may be used to make comparisons between background and compliance well data. This method of analysis is similar to that for calculating a tolerance limit, and familiarity with prediction intervals or personal preference would be the only reason for selecting them over the method for tolerance limits. 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.

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 4.2 for tests of normality), but a log transformation results in data that are adequately normal on the log scale, then 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 95% 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

$$[0, \bar{X} + S\sqrt{1/m + 1/n} \quad t_{(n-1,\, K,\, 0.95)}]$$

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, $\nu = n-1$. 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

$$[0, \bar{X} + S\sqrt{1/m + 1/n} \quad t_{(n-1,\ K,\ 0.95)}]$$

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 5-6 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

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

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

TABLE 5-6. 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	July 1, 1986	123
	103		120
	104		116
	85		128
April 1, 1985	120	m =	4
	105	Mean =	122
	104	SD =	5
	108		
July 1, 1985	110	October 1, 1986	116
	95		117
	102		119
	78		101
October 1, 1985	105	m =	4
	94	Mean =	113
	110	SD =	8
	111		
January 1, 1986	80		
	106		
	115		
	105		
April 1, 1986	100		
	93		
	89		
	113		
n =	24		
Mean =	101		
SD =	11		

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.

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 statistically significant 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

Hahn, G. and Wayne Nelson. 1973. "A Survey of Prediction Intervals and Their Applications." *Journal of Quality Technology*. 5:178-188.

SECTION 6

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 5.

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.

6.1 SUMMARY CHART FOR COMPARISON WITH MCLs OR ACLs

Figure 6-1 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 are 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 6.2.1. Section 6.2.2 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 6.2.3.

6.2 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 alternate 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

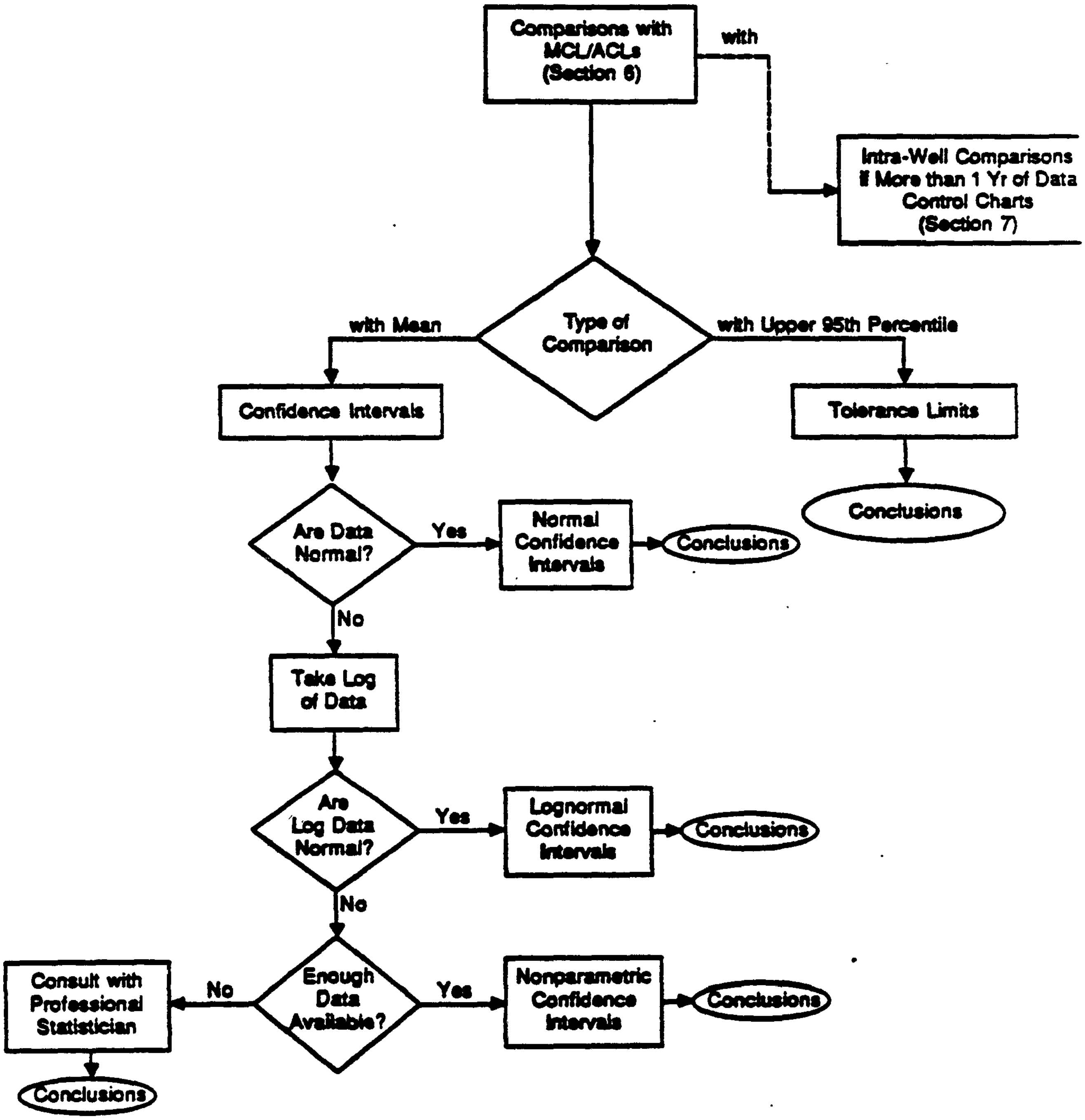


Figure 6-1. Comparisons with MCLs/ACLs.

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

6.2.1 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.

6.2.1.1 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 6-1 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 6-1. 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.

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

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

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 statistically significant 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.

6.2.1.2 Confidence Interval for Log-Normal Data

PURPOSE

The purpose of a confidence interval for the mean concentration of log-normal data 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 normally distributed). 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. Throughout the following procedures and examples, natural logarithms (ln) are used.

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 log-transformed data from each compliance well. (This is Step 1 of the previous section, working now with 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 the lower limit of the confidence interval lies entirely above the logarithm of the compliance limit, there is statistically significant evidence that the unit is out of compliance. Otherwise, the unit is in compliance.

EXAMPLE

Table 6-2 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 natural log-transformation normalized them adequately. The lower part of the table contains the natural logarithms of the concentrations.

TABLE 6-2. EXAMPLE DATA FOR LOG-NORMAL CONFIDENCE INTERVAL--EDB CONCENTRATIONS IN COMPLIANCE WELLS (ppb)

Sampling date	Well 1	Well 2	Well 3
		Concentrations	
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			
		Natural 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
ln (MCL) = 3.00			

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 6-2 to find the values given in the lower part of the table. For example, $\ln(24.2) = 3.19, \ldots, \ln(25.3) = 3.23$. Also, take the logarithm of the MCL to find that $\ln(20) = 3.00$.

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.

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

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

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 3.00. 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:

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

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

Well 3: $(\exp(1.90), \exp(5.44))$ 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 end of the confidence interval exceeds the appropriate compliance limit, then the mean concentration must exceed that compliance limit. These results provide statistically significant evidence of contamination.

6.2.1.3 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). 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. 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 raw 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 (here 99%). 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 $X(1), \ldots, X(n)$, where $X(i)$ is the ith value in the ordered data.

Step 2. Determine the critical values of the order statistics as follows. 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 6-3 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.

TABLE 6-3. 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%

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 statistically significant evidence of contamination. Otherwise, the unit remains in compliance.

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.). For ties, use the average of the tied ranks as in Section 5.2.2, Step 1 of the example. 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 6-4 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

TABLE 6-4. EXAMPLE DATA FOR NONPARAMETRIC CONFIDENCE INTERVAL--SILVEX CONCENTRATIONS (ppm)

Sampling date	Well 1 Concentration (ppm)	Well 1 Rank	Well 2 Concentration (ppm)	Well 2 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)

from each well. The data are not normally distributed, neither as raw data nor when log transformed. Thus, the nonparametric confidence interval is used. The 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 6-4 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.

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

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

Step 3. The approximate 95% confidence limits are given by the 16 + 1 - 14 = 3rd 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 rank-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.

6.2.2 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 end of the tolerance interval (i.e., 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. For example, the owner or operator could suggest a

95% coverage 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 4.2 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, $\overline{X}$, 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

$$[0, \overline{X} + KS]$$

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 statistically significant evidence of contamination.

EXAMPLE

Table 6-5 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 than the permit establishes an ACL of 50 ppm that is not to be exceeded more than 5% of the time.

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

TABLE 6-5. 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 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:

Well 1: 23.1 + 5.145(4.93) = 48.5

Well 2: 24.7 + 5.145(2.28) = 36.4

Well 3: 24.5 + 5.145(2.10) = 35.3

Step 3. Compare the tolerance limits with the ACL of 50 PPM. Since the upper 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 the upper limit of 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.

6.2.3 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 the case of identical concentrations, one should assume that there is some variation in the data, but that the concentrations were rounded and give 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. Assume 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 ±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 any value between 585.0 and 594.99. Thus, 2R is 10 ppm (rounded off), 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 6.2.1) 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. If there is no extrinsic evidence invalidating the data, 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 ±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 ±R is the value R.

SECTION 7

CONTROL CHARTS FOR INTRA-WELL COMPARISONS

The previous sections cover various situations where the compliance well data are compared to the background well data or to specified concentration limits (ACL or MCL) to detect possible contamination. This section discusses the case where the level of each constituent within a single uncontaminated 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 for uncontaminated wells only, 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.

7.1 ADVANTAGES OF PLOTTING DATA

While analyzing the data by means of any of the appropriate statistical procedures discussed in earlier sections is recommended, we also recommend plotting 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 sense of 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 "Detecting and Estimating Trends" and 17 "Trends and Seasonality" 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 using 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.

7.2 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. This could result from variations in chemical concentrations with recharge rates during different seasons throughout the years. 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 ith month during the jth year.

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

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

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

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

$$\bar{X} = \sum_{i=1}^{12} \sum_{j=1}^{N} X_{ij}/N*12 = \sum_{i=1}^{12} \bar{X}_i/12$$

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 7-1 show monthly unadjusted concentrations of a fictitious analyte over a 3-year period.

TABLE 7-1. 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 7-1.

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 7-1.

The reader can check that the average of all 36 adjusted concentrations equals 2.17, the average unadjusted concentration. Figure 7-1 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

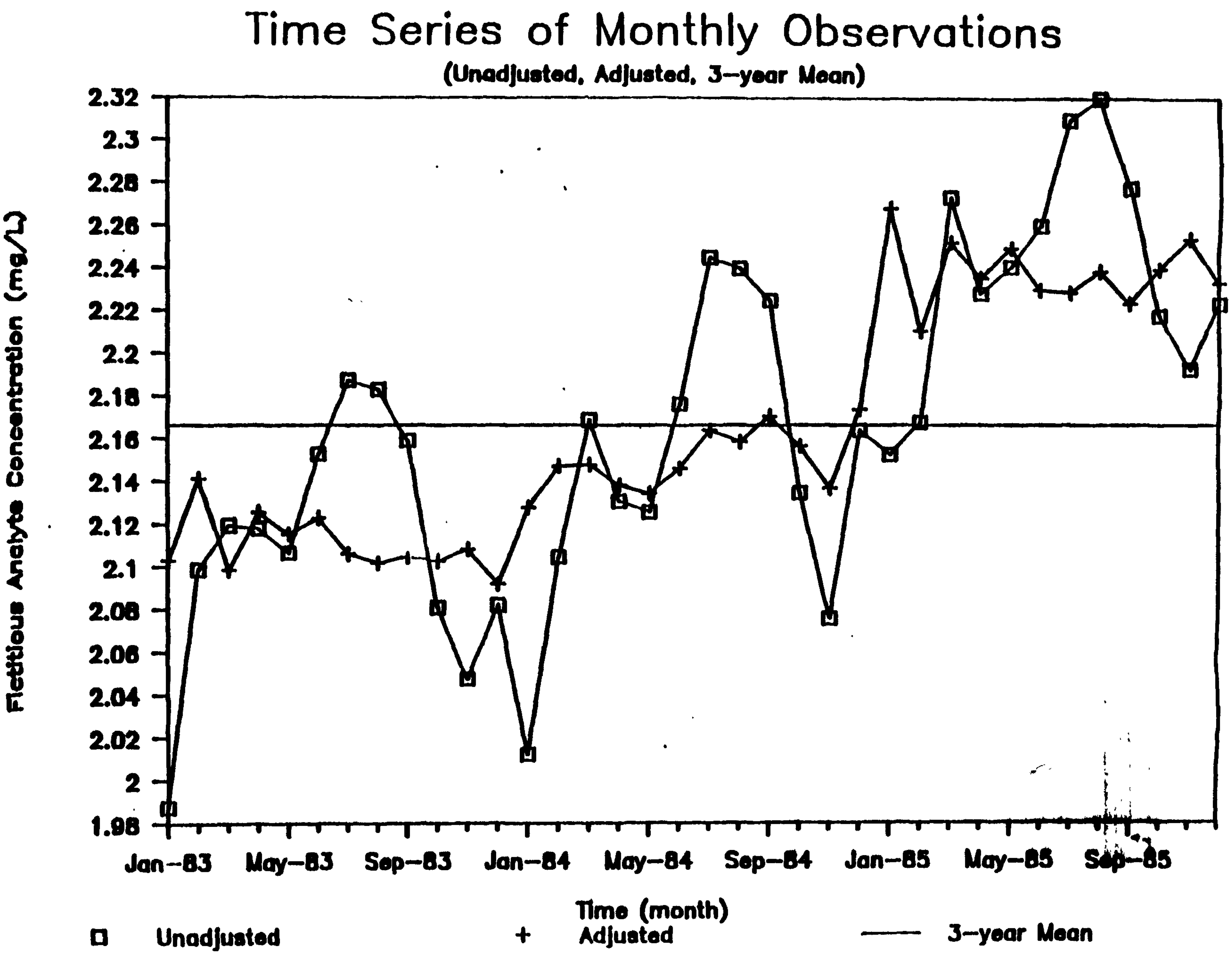
As can be seen in Figure 7-1, seasonal effects were present in the data. After adjusting for monthly effects, the seasonality was removed as can be seen in the adjusted data plotted in the same figure.

7.3 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 within a single well, 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


Time Series of Monthly Observations
(Unadjusted, Adjusted, 3-year Mean)
Fictitious Analyte Concentration (mg/L)
2.32
2.3
2.28
2.26
2.24
2.22
2.2
2.18
2.16
2.14
2.12
2.1
2.08
2.06
2.04
2.02
2
1.98
Jan-83
May-83
Sep-83
Jan-84
May-84
Sep-84
Jan-85
May-85
Sep-85
Time (month)
☐ Unadjusted
+ Adjusted
—— 3-year Mean

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 (see Section 2.4.2) 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, see glossary). 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 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 7.2.

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 internal 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 $X_1, \ldots, X_{n_i}$, are available. Compute their average $\bar{X}_i$.

Step 3. Calculate the standardized mean

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

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. At each time period, T_i, compute the cumulative sum, S_i, as:

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

where max {A, B} is the maximum of A and B, starting with $S_0 = 0$.

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 of two measurements each (n_i = 2 for all i's) are presented in the third column of Table 7-2 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; μ = 5.5 µg/L and σ = 0.4 µg/L.

TABLE 7-2. EXAMPLE DATA FOR COMBINED SHEWHART-CUSUM CHART--
CARBON TETRACHLORIDE CONCENTRATION (µg/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	5.52	0.07	-0.93	0
Feb 3	2	5.60	0.35	-0.65	0
Mar 3	3	5.45	-0.18	-1.18	0
Apr 7	4	5.15	-1.24	-2.24	0
May 5	5	5.95	1.59	0.59	0.59
Jun 2	6	5.54	0.14	-0.86	0.00
Jul 7	7	5.49	-0.04	-1.04	0.00
Aug 4	8	6.08	2.05	1.05	1.05
Sep 1	9	6.91	4.99[a]	3.99	5.04[b]
Oct 6	10	6.78	4.53[a]	3.53	8.56[b]
Nov 3	11	6.71	4.28	3.28	11.84[b]
Dec 1	12	6.65	4.07	3.07	14.91[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 ($Z_i > 4.5$).

[b] CUSUM "out-of-control" signal ($S_i > 5$).

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 the third column of Table 7-2.

Step 3. Standardize the means within each sampling period. These computations are shown in the fourth column of Table 7-2. For example,
$Z_1 = (5.52 - 5.50)*\sqrt{2}/0.4 = 0.07$.

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

$S_1 = \max \{0, -0.93 + 0\} = 0$
$S_2 = \max \{0, -0.65 + 0\} = 0$
•
•
•
$S_5 = \max \{0, 0.59 + S_4\} = \max \{0, 0.59 + 0\} = 0.59$
$S_6 = \max \{0, -0.86 + S_5\} = \max \{0, -0.86 + 0.59) = \max (0, -0.27\} = 0$
•
etc.

These quantities are shown in the last column of Table 7-2.

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 7-2.

The combined chart indicates statistically significant evidence of contamination starting at sampling period T_9. Both the CUSUM scheme and the Shewhart control limit were exceeded by S_9 and Z_9, 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 of two measurements per sampling period. (It should be noted that this method can be used on an individual measurement as well, in which case $n_i = 1$). 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), Hockman and Lucas (1987), and Starks (1988). 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, Hockman and Lucas, and Starks). The user is referred to these articles for further reading.

7.4 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

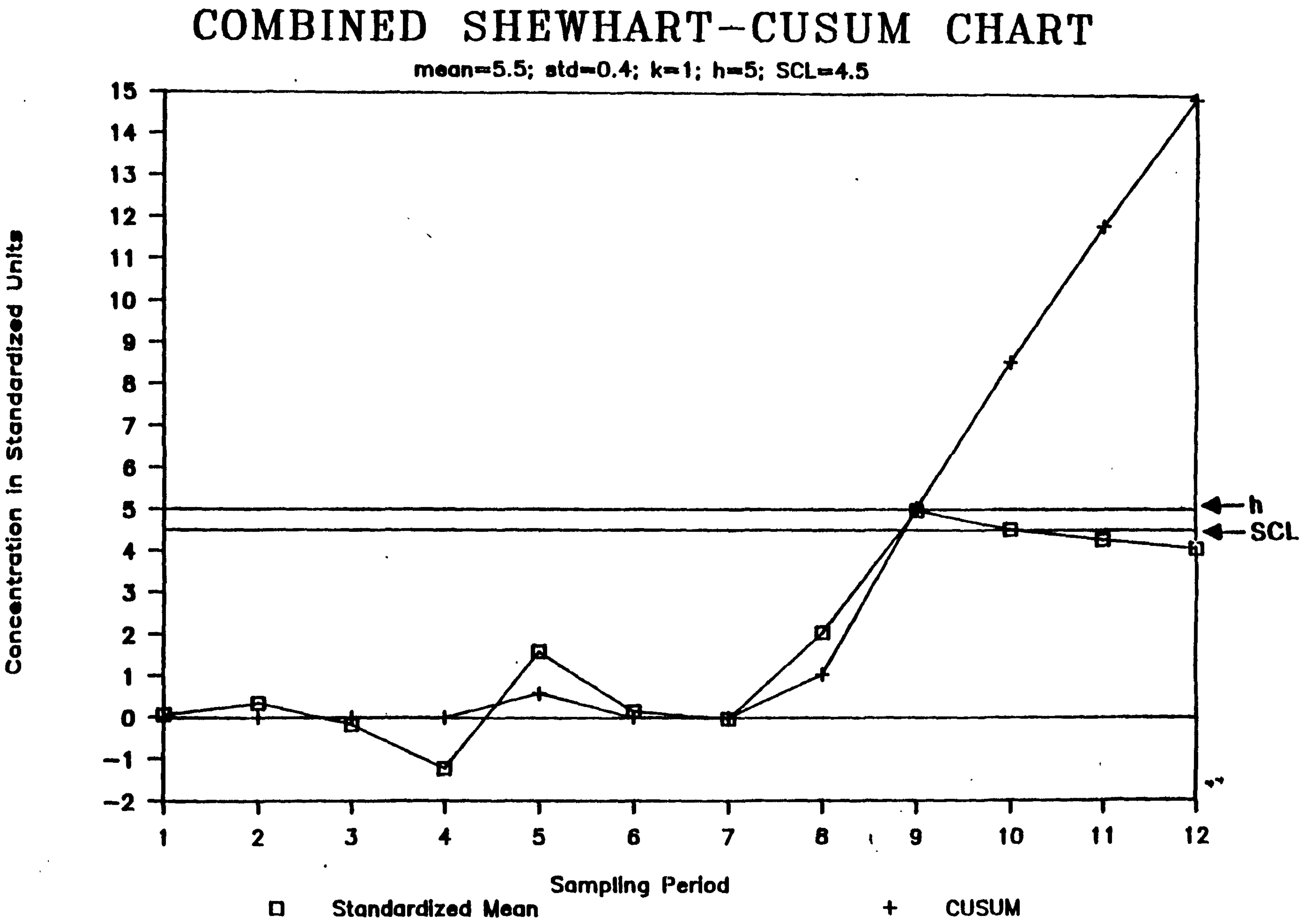


Figure 7-2. Combined Shewhart-CUSUM chart.

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.

7.5 NONDETECTS IN A CONTROL CHART

Regulations require that four independent water samples be taken at each well at a given sampling period. The mean of the four concentration measurements of a particular constituent 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 7.3.

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 7.3.

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 8.1.4. 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.

CAUTIONARY NOTE: Control charts are a useful supplement to other statistical techniques because they are graphical and simple to use. However, it is inappropriate to construct a control chart on wells that have shown evidence of contamination or an increasing trend (see §264.97(a)(1)(i)). Further, contamination may not be present in a well in the form of a steadily increasing concentration profile--it may be present intermittently or may increase in a step function. Therefore, the absence of an increasing trend does not necessarily prove that a release has not occurred.

SECTION 8

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.

8.1 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 only 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 8-1. 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 8-1 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 8-1. METHODS FOR BELOW DETECTION LIMIT VALUES

Percentage of Nondetects in the Data Base	Statistical Analysis Method	Section of Guidance Document
Less than 15%	Replace NDs with MDL/2 or PQL/2, then proceed with parametric procedures:	Section 8.1.1
	• ANOVA	Section 5.2.1
	• Tolerance Units	Section 5.3
	• Prediction Intervals	Section 5.4
	• Control Charts	Section 7
Between 15 and 50%	Use NDs as ties, then proceed with Nonparametric ANOVA	Section 5.2.2
	or	
	use Cohen's adjustment, then proceed with:	Section 8.1.3
	• Tolerance Limits	Section 5.3
	• Confidence Intervals	Section 6.2.1
	• Control Charts	Section 7
More than 50%	Test of Proportions	Section 8.1.2

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 nonquantified 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.

8.1.1 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 5.2.1. 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.

8.1.2. 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 (i.e., between 10% and 50%). 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 determines 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 statistically significant 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 8-2 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 8-2. EXAMPLE DATA FOR A TEST OF PROPORTIONS

Cadmium concentration (µg/L) at background well (24 samples)		Cadmium concentration (µg/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 8-2, 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 in proportions.

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

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 statistically 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.

8.1.3 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 ($\leq$ 50%) 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 ith 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:

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

Step 3. Compute the two parameters, h and γ (lowercase gamma), as follows:

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

and

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

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:

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

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

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

Step 6. Use the corrected values of $\bar{X}$ and S in the procedure for constructing a tolerance interval (Section 5.3) or a confidence interval (Section 6.2.1).

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 8-3 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 8-3. EXAMPLE DATA FOR 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

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

Enter Table 7 of Appendix B at $h = 0.125$ and $\gamma = 0.083$ 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 7 between which the user needs to interpolate are:

γ	$h = 0.10$	$h = 0.15$
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:

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 = $\underline{0.151415}$

On the γ-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 γ-axis is:

$$0.141415 - 0.14683 = 0.004585 \qquad 0.004585 * 0.66 = 0.0030261$$
$$0.14683 + 0.0030261 = \underline{0.14986}$$

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.

8.2 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 $X_1, \ldots, X_n$. 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:

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

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 T_n 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 8-4 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 8-4. 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 $T_{19} = 3.74$ 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 hydrogeology 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 preceding sections should cover the majority of cases. They did 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 significance level or probability of a Type I error.
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 parameter, 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 confidence interval 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 significance level 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 nonparametric. 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 multiple comparisons. 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 of 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.

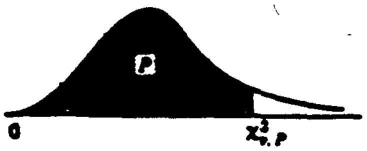
Power	The power of a test is the probability that the test will reject under a specified alternative hypothesis. This is one minus the probability of a <u>Type II error</u>. The power is a measure of the test's ability to detect a difference of specified size from the null hypothesis.
Sample standard deviation	This is the square root of the <u>sample variance</u>.
Sample variance	This is a statistic (computed on a sample of observations rather than on the whole population) that measures the variability or spread of the observations about the sample mean. It is the sum of the squared differences from the sample mean, divided by the number of observations less one.
Serial correlation	This is the correlation of observations spaced a constant interval apart in a series. For example, the first order serial correlation is the correlation between adjacent observations. The first order serial correlation is found by correlating the pairs consisting of the first and second, second and third, third and fourth, etc., observations.
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

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



ν \ P	0.750	0.900	0.950	0.975	0.990	0.995	0.999
1	1.323	2.706	3.841	5.024	6.63	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, $F_{\nu_1,\nu_2,0.95}$

ν_2 \ ν_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	19.5
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	8.5
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	5.6
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	4.3
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	3.6
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	3.2
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	2.9
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	2.7
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	2.5
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	2.4
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	2.3
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	2.2
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	2.1
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	2.0
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	2.0
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	1.9
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	1.9
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	1.8
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	1.8
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	1.8
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	1.7
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	1.7
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	1.7
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	1.7
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	1.6
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	1.6
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	1.6
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	1.6
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	1.6
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	1.5
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	1.3
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	1.2
∞	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	1.0

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(ν, α/m)

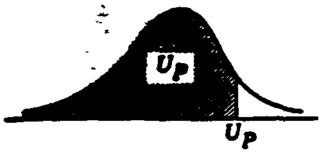
where ν = degrees of freedom associated with the mean squares error
m = number of comparisons
α = 0.05, the experimentwise error level

m / α/m / ν	 0.0	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	1.01	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

<u>SOURCE</u>: For α/m = 0.05, 0.025, and 0.01, the percentiles were extracted from the t-table (Table 6, Appendix B) for values of F=1-α of 0.95, 0.975, and 0.99, respectively.

For α/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

<u>NOTE</u>: 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
3	7.655
4	5.145
5	4.202
6	3.707
7	3.399
8	3.188
9	3.031
10	2.911
11	2.815
12	2.736
13	2.670
14	2.614
15	2.566
16	2.523
17	2.486
18	2.543
19	2.423
20	2.396
21	2.371
22	2.350
23	2.329
24	2.309
25	2.292
30	2.220
35	2.166
40	2.126
45	2.092
50	2.065
75	1.972
100	1.924
125	1.891
150	1.868
175	1.850
200	1.836
225	1.824
250	1.814
275	1.806
300	1.799
325	1.792
350	1.787
375	1.782
400	1.777
425	1.773
450	1.769
475	1.766
500	1.763
525	1.760
550	1.757
575	1.754
600	1.752
625	1.750
650	1.748
675	1.746
700	1.744
725	1.742
750	1.740
775	1.739
800	1.737
825	0.736
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)

n \ F	.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	1.886	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	1.476	2.015	2.571	3.365	4.032	6.8
6	.265	.718	1.440	1.943	2.447	3.143	3.707	5.9
7	.263	.711	1.415	1.895	2.365	2.998	3.499	5.40
8	.262	.706	2.397	1.860	2.306	2.896	3.355	5.04
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	4.015
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

h / γ	.01	.02	.03	.04	.05	.06	.07	.08	.09	.10	.15	.20	h / γ
.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.00
.05	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.05
.10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.10
.15	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.15
.20	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.20
.25	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.25
.30	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.30
.35	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.35
.40	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.40
.45	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.45
.50	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.50
.55	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.55
.60	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.60
.65	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.65
.70	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.70
.75	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.75
.80	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.80
.85	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.85
.90	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.90
.95	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.95
1.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1.00

h / γ	.25	.30	.35	.40	.45	.50	.55	.60	.65	.70	.80	.90	h / γ
.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.00
.05	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.05
.10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.10
.15	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.15
.20	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.20
.25	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.25
.30	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.30
.35	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.35
.40	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.40
.45	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.45
.50	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.50
.55	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.55
.60	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.60
.65	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.65
.70	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.70
.75	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.75
.80	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.80
.85	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.85
.90	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.90
.95	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.95
1.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1.00

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

<u>SOURCE</u>: 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 0.5% Significance Level	Upper 1% Significance Level	Upper 2.5% Significance Level	Upper 5% Significance Level	Upper 10% Significance Level
101	4.089	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	3.764	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	3.688	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	3.138
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 Water in Glacial Till." *Proceedings of the 17th International Congress of t ternational 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.

Gibbons, R. D. 1988. "Statistical Models for the Analysis of Volatile Organic Compounds in Waste Disposal Sites." *Ground Water*. Vol. 26.

Gilbert, R. 1987. *Statistical Methods for Environmental Pollution Monitoring*. Professional Books Series, Van Nos Reinhold.

Hahn, G. and W. Nelson. 1973. "A Survey of Prediction Intervals and Their Applications." *Journal of Quality Technology*. 5:178-188.

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. 245-25

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.

"Statistical Methods for the Attainment of Superfund Cleanup Standards (Volume 2: Ground Water--Draft)."

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*. [illegible] 99-114.

Statistical Software Packages

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

Lotus 1-2-3 Release 2. 1986. Lotus Development Corporation, 55 Cambridge Parkway, Cambridge, Massachusetts 02142.

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.

STATISTICAL ANALYSIS OF GROUND-WATER MONITORING DATA AT RCRA FACILITIES

DRAFT

ADDENDUM TO INTERIM FINAL GUIDANCE

OFFICE OF SOLID WASTE
PERMITS AND STATE PROGRAMS DIVISION
U.S. ENVIRONMENTAL PROTECTION AGENCY
401 M STREET, S.W.
WASHINGTON, D.C. 20460

JULY 1992

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 that meet the regulatory requirements 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.

CONTENTS

ACKNOWLEDGMENT

This document was developed by EPA's Office of Solid Waste under the direction of Mr. James R. Brown of the Permits and State Programs Division. The Addendum was prepared by the joint efforts of Mr. James R. Brown and Kirk M. Cameron, Ph.D., Senior Statistician at Science Applications International Corporation (SAIC). SAIC provided technical support in developing this document under EPA Contract No. 68-W0-0025. Other SAIC staff who assisted in the preparation of the Addendum include Mr. Robert D. Aaron, Statistician.

STATISTICAL ANALYSIS OF
--GROUND-WATER MONITORING DATA
AT RCRA FACILITIES

ADDENDUM TO INTERIM FINAL GUIDANCE

JULY 1992

This Addendum offers a series of recommendations and updated advice concerning the Interim Final Guidance document for statistical analysis of ground-water monitoring data. Some procedures in the original guidance are replaced by alternative methods that reflect more current thinking within the statistics profession. In other cases, further clarification is offered for currently recommended techniques to answer questions and address public comments that EPA has received both formally and informally since the Interim Final Guidance was published.

1. CHECKING ASSUMPTIONS FOR STATISTICAL PROCEDURES

Because any statistical or mathematical model of actual data is an approximation of reality, all statistical tests and procedures require certain assumptions for the methods to be used correctly and for the results to have a proper interpretation. Two key assumptions addressed in the Interim Guidance concern the distributional properties of the data and the need for equal variances among subgroups of the measurements. In the Addendum, new techniques are outlined for testing both assumptions that offer distinct advantages over the methods in the Interim Final Guidance.

1.1 NORMALITY OF DATA

Most statistical tests assume that the data come from a Normal distribution. Its density function is the familiar bell-shaped curve. The Normal distribution is the assumed underlying model for such procedures as parametric analysis of variance (ANOVA), t-tests, tolerance intervals, and prediction intervals for future observations. Failure of the data to follow a Normal distribution at least approximately is not always a disaster, but can lead to false conclusions if the data really follow a more skewed distribution like the Lognormal. This is because the extreme tail behavior of a data distribution is often the most critical factor in deciding whether to apply a statistical test based on the assumption of Normality.

The Interim Final Guidance suggests that one begin by assuming that the original data are Normal prior to testing the distributional assumptions. If the statistical test rejects the model of Normality, the data can be tested for Lognormality instead by taking the natural logarithm of each observation and repeating the test. If the original data are Lognormal, taking the natural logarithm of the observations will result in data that are Normal. As a consequence, tests for Normality can also be used to test for Lognormality by applying the tests to the logarithms of the data.

Unfortunately, all of the available tests for Normality do at best a fair job of rejecting non-Normal data when the sample size is small (say less than 20 to 30 observations). That is, the tests do not exhibit high degrees of statistical power. As such, small samples of untransformed Lognormal data can be accepted by a test of Normality even though the skewness of the data may lead to poor statistical conclusions later. EPA's experience with environmental concentration data, and ground-water data in particular, suggests that a Lognormal distribution is generally more appropriate as a default statistical model than the Normal distribution, a conclusion shared by researchers at the United States Geological Survey (USGS, Dennis Helsel, personal communication, 1991). There also appears to be a plausible physical explanation as to why pollutant concentrations so often seem to follow a Lognormal pattern (Ott, 1990). In Ott's model, pollutant sources are randomly diluted in a multiplicative fashion through repeated dilution and mixing with volumes of uncontaminated air or water, depending on the surrounding medium. Such random and repeated dilution of pollutant concentrations can lead mathematically to a Lognormal distribution.

Because the Lognormal distribution appears to be a better default statistical model than the Normal distribution for most ground-water data, it is recommended that all data first be logged prior to checking distributional assumptions. McBean and Rovers (1992) have noted that "[s]upport for the lognormal distribution in many applications also arises from the shape of the distribution, namely constrained on the low side and unconstrained on the high side.... The logarithmic transform acts to suppress the outliers so that the mean is a much better representation of the central tendency of the sample data."

Transformation to the logarithmic scale is not done to make "large numbers look smaller." Performing a logarithmic or other monotonic transformation preserves the basic ordering within a data set, so that the data are merely rescaled with a different set of units. Just as the physical difference between 80° Fahrenheit and 30° Fahrenheit does not change if the temperatures are rescaled or transformed to the numerically lower Celsius scale, so too the basic statistical relationships between data measurements remain the same whether or not the log transformation is applied. What does change is that the logarithms of Lognormally distributed data are more nearly Normal in character, thus satisfying a key assumption of many statistical procedures. Because of this fact, the same tests used to check Normality, if run on the logged data, become tests for Lognormality.

If the assumption of Lognormality is not rejected, further statistical analyses should be performed on the logged observations, not the original data. If the Lognormal distribution is rejected by a statistical test, one can either test the Normality of the original data, if it was not already done, or use a non-parametric technique on the ranks of the observations.

If no data are initially available to test the distributional assumptions, "referencing" may be employed to justify the use of, say, a Normal or Lognormal assumption in developing a statistical testing regimen at a particular site. "Referencing" involves the use of historical data or data from sites in similar hydrogeologic settings to justify the assumptions applied to currently planned statistical tests. These initial assumptions must be checked when data from the site become available, using the procedures described in this Addendum. Subsequent changes to the initial assumptions should be made if formal testing contradicts the initial hypothesis.

1.1.1 Interim Final Guidance Methods for Checking Normality

The Interim Final Guidance outlines three different methods for checking Normality: the Coefficient-of-Variation (CV) test, Probability Plots, and the Chi-squared test. Of these three, only Probability Plots are recommended within this Addendum. The Coefficient-of-Variation and the Chi-squared test each have potential problems that can be remedied by using alternative tests. These alternatives include the Coefficient of Skewness, the Shapiro-Wilk test, the Shapiro-Francia test, and the Probability Plot Correlation Coefficient.

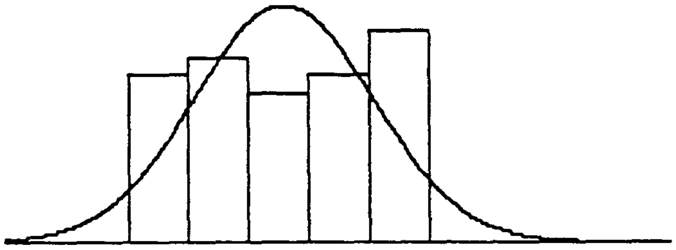
The Coefficient-of-Variation is recommended within the Interim Guidance because it is easy to calculate and is amenable to small sample sizes. To ensure that a Normal model which predicts a significant fraction of negative concentration values is not fitted to positive data, the Interim Final Guidance recommends that the sample Coefficient of Variation be less than one; otherwise this "test" of Normality fails. A drawback to using the sample CV is that for Normally distributed data, one can often get a sample CV greater than one when the true CV is only between 0.5 and 1. In other words, the sample CV, being a random variable, often estimates the true Coefficient of Variation with some error. Even if a Normal distribution model is appropriate, the Coefficient of Variation test may reject the model because the sample CV (but not the true CV) is too large.

The real purpose of the CV is to estimate the skewness of a dataset, not to test Normality. Truly Normal data can have any non-zero Coefficient of Variation, though the larger the CV, the greater the proportion of negative values predicted by the model. As such, a Normal distribution with large CV may be a poor model for positive concentration data. However, if the Coefficient of Variation test is used on the logarithms of the data to test Lognormality, negative logged concentrations will often be expected, nullifying the rationale used to support the CV test in the first place. A better way to estimate the skewness of a dataset is to compute the Coefficient of Skewness directly, as described below.

The Chi-square test is also recommended within the Interim Guidance. Though an acceptable goodness-of-fit test, it is not considered the most sensitive or powerful test of Normality in the current literature (Gan and Koehler, 1990). The major drawback to the Chi-square test can be explained by considering the behavior of parametric tests based on the Normal distribution. Most tests like the t-test or Analysis of Variance (ANOVA), which assume the underlying data to be Normally distributed, give fairly robust results when the Normality assumption fails over the middle ranges of the data distribution. That is, if the extreme tails are approximately Normal in shape even if the middle part of the density is not, these parametric tests will still tend to produce valid results. However, if the extreme tails are non-Normal in shape (e.g., highly skewed), Normal-based tests can lead to false conclusions, meaning that either a transformation of the data or a non-parametric technique should be used instead.

The Chi-square test entails a division of the sample data into bins or cells representing distinct, non-overlapping ranges of the data values (see figure below). In each bin, an expected value is computed based on the number of data points that would be found if the Normal distribution provided an appropriate model. The squared difference between the expected number and observed number is then computed and summed over all the bins to calculate the Chi-square test statistic.

CHI SQUARE GOODNESS OF FIT



If the Chi-square test indicates that the data are not Normally distributed, it may not be clear what ranges of the data most violate the Normality assumption. Departures from Normality in the middle bins are given nearly the same weight as departures from the extreme tail bins, and all the departures are summed together to form the test statistic. As such, the Chi-square test is not as powerful for detecting departures from Normality in the extreme tails of the data, the areas most crucial to the validity of parametric tests like the t-test or ANOVA (Miller, 1986). Furthermore, even if there are departures in the tails, but the middle portion of the data distribution is approximately Normal, the Chi-square test may not register as statistically significant in certain cases where better tests of Normality would. Because of this, four alternative, more sensitive tests of Normality are suggested below which can be used in conjunction with Probability Plots.

1.1.2 Probability Plots

As suggested within the Interim Final Guidance, a simple, yet useful graphical test for Normality is to plot the data on probability paper. The y-axis is scaled to represent probabilities according to the Normal distribution and the data are arranged in increasing order. An observed value is plotted on the x-axis and the proportion of observations less than or equal to each observed value is plotted as the y-coordinate. The scale is constructed so that, if the data are Normal, the points when plotted will approximate a straight line. Visually apparent curves or bends indicate that the data do not follow a Normal distribution (see Interim Final Guidance, pp. 4-8 to 4-11).

Probability Plots are particularly useful for spotting irregularities within the data when compared to a specific distributional model like the Normal. It is easy to determine whether departures from Normality are occurring more or less in the middle ranges of the data or in the extreme tails. Probability Plots can also indicate the presence of possible outlier values that do not follow the basic pattern of the data and can show the presence of significant positive or negative skewness.

If a (Normal) Probability Plot is done on the combined data from several wells and Normality is accepted, it implies that all of the data came from the same Normal distribution. Consequently, each subgroup of the data set (e.g., observations from distinct wells), has the same mean and standard deviation. If a Probability Plot is done on the data residuals (each value minus its subgroup mean) and is not a straight line, the interpretation is more complicated. In this case, either the residuals are not Normal, or there is a subgroup of the data with a Normal distribution but a different mean or standard deviation than the other subgroups. The Probability Plot will indicate a deviation from the underlying Normality assumption either way.

The same Probability Plot technique may be used to investigate whether a set of data or residuals follows the Lognormal distribution. The procedure is the same, except that one first replaces each observation by its natural logarithm. After the data have been transformed to their natural logarithms, the Probability Plot is constructed as before. The only difference is that the natural logarithms of the observations are used on the x-axis. If the data are Lognormal, the Probability Plot (on Normal probability paper) of the logarithms of the observations will approximate a straight line.

Many statistical software packages for personal computers will construct Probability Plots automatically with a simple command or two. If such software is available, there is no need to construct Probability Plots by hand or to obtain special graph paper. The plot itself may be generated somewhat differently than the method described above. In some packages, the observed value is plotted as before on the x-axis. The y-axis, however, now represents the quantile of the Normal distribution (often referred to as the "Normal score of the observation") corresponding to the cumulative probability of the observed value. The y-coordinate is often computed by the following formula:

$$y_i = \Phi^{-1}\left(\frac{i}{n+1}\right)$$

where Φ^{-1} denotes the inverse of the cumulative Normal distribution, n represents the sample size, and i represents the rank position of the ith ordered concentration. Since the computer does these calculations automatically, the formula does not have to be computed by hand.

EXAMPLE 1

Determine whether the following data set follows the Normal distribution by using a Probability Plot.

	Nickel Concentration (ppb)			
Month	Well 1	Well 2	Well 3	Well 4
1	58.8	19	39	3.1
2	1.0	81.5	151	942
3	262	331	27	85.6
4	56	14	21.4	10
5	8.7	64.4	578	637

SOLUTION

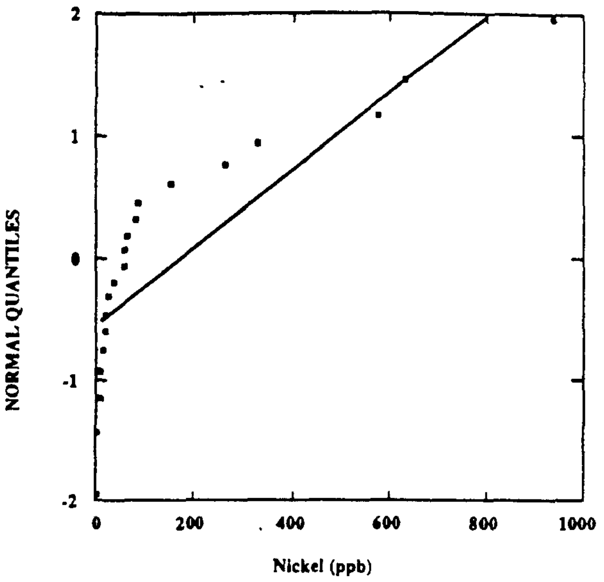
Step 1. List the measured nickel concentrations in order from lowest to highest.

Nickel Concentration (ppb)	Order (i)	Probability 100*(i/(n+1))	Normal Quantile
1	1	5	-1.645
3.1	2	10	-1.28
8.7	3	14	-1.08
10	4	19	-0.88
14	5	24	-0.706
19	6	29	-0.55
21.4	7	33	-0.44
27	8	38	-0.305
39	9	43	-0.176
56	10	48	-0.05
58.8	11	52	0.05
64.4	12	57	0.176
81.5	13	62	0.305
85.6	14	67	0.44
151	15	71	0.55
262	16	76	0.706
331	17	81	0.88
578	18	86	1.08
637	19	90	1.28
942	20	95	1.645

Step 2. The cumulative probability is given in the third column and is computed as 100*(i/(n+1)) where n is the total number of samples (n=20). The last column gives the Normal quantiles corresponding to these probabilities.

Step 3. If using special graph paper, plot the probability versus the concentration for each sample. Otherwise, plot the Normal quantile versus the concentration for each sample, as in the plot below. The curvature found in the Probability Plot indicates that there is evidence of non-Normality in the data.

PROBABILITY PLOT



1.1.3 Coefficient of Skewness

The Coefficient of Skewness (γ_1) indicates to what degree a data set is skewed or asymmetric with respect to the mean. Data from a Normal distribution will have a Skewness Coefficient of zero, while asymmetric data will have a positive or negative skewness depending on whether the right- or left-hand tail of the distribution is longer and skinnier than the opposite tail.

Since ground-water monitoring concentration data are inherently nonnegative, one often expects the data to exhibit a certain degree of skewness. A small degree of skewness is not likely to affect the results of statistical tests based on an assumption of Normality. However, if the Skewness Coefficient is larger than 1 (in absolute value) and the sample size is small (e.g., $n<25$), statistical research has shown that standard Normal theory-based tests are much less powerful than when the absolute skewness is less than 1 (Gayen, 1949).

Calculating the Skewness Coefficient is useful and not much more difficult than computing the Coefficient of Variation. It provides a quick indication of whether the skewness is minimal enough to assume that the data are roughly symmetric and hopefully Normal in distribution. If the original data exhibit a high Skewness Coefficient, the Normal distribution will provide a poor approximation to the data set. In that case, γ_1 can be computed on the logarithms of the data to test for symmetry of the logged data.

The Skewness Coefficient may be computed using the following formula:

$$\gamma_1 = \frac{\frac{1}{n}\Sigma_i (x_i - \bar{x})^3}{\left(\frac{n-1}{n}\right)^{\frac{3}{2}} (SD)^3}$$

where the numerator represents the average cubed residual and SD denotes the standard deviation of the measurements. Most statistics computer packages (e.g., Minitab, GEO-EAS) will compute the Skewness Coefficient automatically via a simple command.

EXAMPLE 2

Using the data in Example 1, compute the Skewness Coefficient to test for approximate symmetry in the data.

SOLUTION

Step 1. Compute the mean, standard deviation (SD), and average cubed residual for the nickel concentrations:

$$\bar{x} = 169.52 \text{ ppb}$$

$$SD = 259.72 \text{ ppb}$$

$$\frac{1}{n}\sum_i (x_i - \bar{x})^3 = 2.98923 * 10^8 \text{ ppb}^3$$

Step 2. Calculate the Coefficient of Skewness using the previous formula to get $\gamma_1 = 1.84$. Since the skewness is much larger than 1, the data appear to be significantly positively skewed. Do not assume that the data follow a Normal distribution.

Step 3. Since the original data evidence a high degree of skewness, one can attempt to compute the Skewness Coefficient on the logged data instead. In that case, the skewness works out to be $|\gamma_1| = 0.24 < 1$, indicating that the logged data values are slightly skewed, but not enough to reject an assumption of Normality in the logged data. In other words, the original data may be Lognormally distributed.

1.1.4 The Shapiro-Wilk Test of Normality ($n \leq 50$)

The Shapiro-Wilk test is recommended as a superior alternative to the Chi-square test for testing Normality of the data. It is based on the premise that if a set of data are Normally distributed, the ordered values should be highly correlated with corresponding quantiles taken from a Normal distribution (Shapiro and Wilk, 1965). In particular, the Shapiro-Wilk test gives substantial weight to evidence of non-Normality in the tails of a distribution, where the robustness of statistical tests based on the Normality assumption is most severely affected. The Chi-square test treats departures from Normality in the tails nearly the same as departures in the middle of a distribution, and so is less sensitive to the types of non-Normality that are most crucial. One cannot tell from a significant Chi-square goodness-of-fit test what sort of non-Normality is indicated.

The Shapiro-Wilk test statistic (W) will tend to be large when a Probability Plot of the data indicates a nearly straight line. Only when the plotted data show significant bends or curves will the test statistic be small. The Shapiro-Wilk test is considered to be one of the very best tests of Normality available (Miller, 1986; Madansky, 1988).

To calculate the test statistic W, one can use the following formula:

$$W = \left[\frac{b}{SD\sqrt{n-1}} \right]^2$$

where the numerator is computed as

$$b = \sum_{i=1}^{k} a_{n-i+1} \left(x_{(n-i+1)} - x_{(i)} \right) = \sum_{i=1}^{k} b_i$$

In this last formula, $x_{(j)}$ represents the jth smallest ordered value in the sample and coefficients a_j depend on the sample size n. The coefficients can be found for any sample size from 3 up to 50 in Table A-1 of Appendix A. The value of k can be found as the greatest integer less than or equal to n/2.

Normality of the data should be rejected if the Shapiro-Wilk statistic is too low when compared to the critical values provided in Table A-2 of Appendix A. Otherwise one can assume the data are approximately Normal for purposes of further statistical analysis. As before, it is recommended that the test first be performed on the logarithms of the original data to test for Lognormality. If the logged data indicate non-Normality by the Shapiro-Wilk test, a re-test can be performed on the original data to test for Normality of the original concentrations.

EXAMPLE 3

Use the data of Example 1 to compute the Shapiro-Wilk test of Normality.

SOLUTION

Step 1. Order the data from smallest to largest and list, as in the following table. Also list the data in reverse order alongside the first column.

Step 2. Compute the differences $x_{(n-i+1)}-x_{(i)}$ in column 3 of the table by subtracting column 1 from column 2.

i	$x_{(i)}$	$x_{(n-i+1)}$	$x_{(n-i+1)}-x_{(i)}$	a_{n-i+1}	b_i
1	1.0	942.0	941.0	.4734	445.47
2	3.1	637.0	633.9	.3211	203.55
3	8.7	578.0	569.3	.2565	146.03
4	10.0	331.0	321.0	.2085	66.93
5	14.0	262.0	248.0	.1686	41.81
6	19.0	151.0	132.0	.1334	17.61
7	21.4	85.6	64.2	.1013	6.50
8	27.0	81.5	54.5	.0711	3.87
9	39.0	64.4	25.4	.0422	1.07
10	56.0	58.8	2.8	.0140	0.04
11	58.8	56.0	-2.8		b=932.88
12	64.4	39.0	-25.4		
13	81.5	27.0	-54.5		
14	85.6	21.4	-64.2		
15	151.0	19.0	-132.0		
16	262.0	14.0	-248.0		
17	331.0	10.0	-321.0		
18	578.0	8.7	-569.3		
19	637.0	3.1	-633.9		
20	942.0	1.0	-941.0		

Step 3. Compute k as the greatest integer less than or equal to n/2. Since n=20, k=10 in this example.

Step 4. Look up the coefficients a_{n-i+1} from Table A-1 and list in column 4. Multiply the differences in column 3 by the coefficients in column 4 and add the first k products to get quantity b. In this case, b=932.88.

Step 5. Compute the standard deviation of the sample, SD=259.72. Then

$$W=\left[\frac{932.88}{259.72\sqrt{19}}\right]^2=0.679.$$

Step 6. Compare the computed value of W=0.679 to the 5% critical value for sample size 20 in Table A-2, namely $W_{.05,20}=0.905$. Since W < 0.905, the sample shows significant evidence of non-Normality by the Shapiro-Wilk test. The data should be transformed using natural logs and rechecked using the Shapiro-Wilk test before proceeding with further statistical analysis (Actually, the logged data should have been tested first. The original concentration data are used in this example to illustrate how the assumption of Normality can be rejected.)

1.1.5 The Shapiro-Francia Test of Normality (n>50)

The Shapiro-Wilk test of Normality can be used for sample sizes up to 50. When the sample is larger than 50, a slight modification of the procedure called the Shapiro-Francia test (Shapiro and Francia, 1972) can be used instead.

Like the Shapiro-Wilk test, the Shapiro-Francia test statistic (W′) will tend to be large when a Probability Plot of the data indicates a nearly straight line. Only when the plotted data show significant bends or curves will the test statistic be small.

To calculate the test statistic W′, one can use the following formula:

$$W' = \frac{\left[\Sigma_i m_i x_{(i)}\right]^2}{(n-1)SD^2 \Sigma_i m_i^2}$$

where $x_{(i)}$ represents the ith ordered value of the sample and where m_i denotes the approximate expected value of the ith ordered Normal quantile. The values for m_i can be approximately computed as

$$m_i = \Phi^{-1}\left(\frac{i}{n+1}\right)$$

where Φ^{-1} denotes the inverse of the standard Normal distribution with zero mean and unit variance. These values can be computed by hand using a Normal probability table or via simple commands in many statistical computer packages.

Normality of the data should be rejected if the Shapiro-Francia statistic is too low when compared to the critical values provided in Table A-3 of Appendix A. Otherwise one can assume the data are approximately Normal for purposes of further statistical analysis. As before, the logged data should be tested first to see if a Lognormal model is appropriate. If these data indicate non-Normality by the Shapiro-Francia test, a re-test can be performed on the original data.

1.1.6 The Probability Plot Correlation Coefficient

One other alternative test for Normality that is roughly equivalent to the Shapiro-Wilk and Shapiro-Francia tests is the Probability Plot Correlation Coefficient test described by Filliben (1975). This test fits in perfectly with the use of Probability Plots, because the essence of the test is to compute the common correlation coefficient for points on a Probability Plot. Since the correlation coefficient is a measure of the linearity of the points on a scatterplot, the Probability Plot Correlation Coefficient, like the Shapiro-Wilk test, will be high when the plotted points fall along a straight line and low when there are significant bends and curves in the Probability Plot. Comparison of the Shapiro-Wilk and Probability Plot Correlation Coefficient tests has indicated very similar statistical power for detecting non-Normality (Ryan and Joiner, 1976).

The construction of the test statistic is somewhat different from the Shapiro-Wilk W, but not difficult to implement. Also, tabled critical values for the correlation coefficient have been derived for sample sizes up to n=100 (and are reproduced in Table A-4 of Appendix A). The Probability Plot Correlation Coefficient may be computed as

$$r = \frac{\sum_{i=1}^{n} X_{(i)} M_i - n\overline{X}\overline{M}}{C_n \times SD\sqrt{n-1}}$$

where $X_{(i)}$ represents the ith smallest ordered concentration value, M_i is the median of the ith order statistic from a standard Normal distribution, and $\overline{X}$ and $\overline{M}$ represent the average values of $X_{(i)}$ and $M_{(i)}$. The ith Normal order statistic median may be approximated as $M_i=\Phi^{-1}(m_i)$, where as before, Φ^{-1} is the inverse of the standard Normal cumulative distribution and m_i can be computed as follows (given sample size n):

$$m_i = \begin{cases} 1-(.5)^{1/n} \text{ for } i=1 \\ (i-.3175)/(n+.365) \quad \text{for } 1<i<n \\ (.5)^{1/n} \text{ for } i=n \end{cases}$$

Quantity C_n represents the square root of the sum of squares of the M_i's minus n times the average value $\overline{M}$, that is

$$C_n = \sqrt{\sum_i M_i^2 - n\overline{M}^2}$$

When working with a complete sample (i.e., containing no nondetects or censored values), the average value $\overline{M}=0$, and so the formula for the Probability Plot Correlation Coefficient simplifies to

$$r = \frac{\sum_i X_{(i)} M_i}{\sqrt{\sum_i M_i^2} \times SD\sqrt{n-1}}$$

EXAMPLE 4

Use the data of Example 1 to compute the Probability Plot Correlation Coefficient test.

SOLUTION

Step 1. Order the data from smallest to largest and list, as in the following table.

Step 2. Compute the quantities m_i from Filliben's formula above for each i in column 2 and the order statistic medians, M_i, in column 3 by applying the inverse Normal transformation to column 2.

Step 3. Since this sample contains no nondetects, the simplified formula for r may be used. Compute the products $X_{(i)}*M_i$ in column 4 and sum to get the numerator of the correlation coefficient (equal to 3,836.81 in this case). Also compute M_i^2 in column 5 and sum to find quantity $C_n^2=17.12$.

i	$x_{(i)}$	m_i	M_i	$X_{(i)}*M_i$	M_i^2
1	1.0	.03406	-1.8242	-1.824	3.328
2	3.1	.08262	-1.3877	-4.302	1.926
3	8.7	.13172	-1.1183	-9.729	1.251
4	10.0	.18082	-0.9122	-9.122	0.832
5	14.0	.22993	-0.7391	-10.347	0.546
6	19.0	.27903	-0.5857	-11.129	0.343
7	21.4	.32814	-0.4451	-9.524	0.198
8	27.0	.37724	-0.3127	-8.444	0.098
9	39.0	.42634	-0.1857	-7.242	0.034
10	56.0	.47545	-0.0616	-3.448	0.004
11	58.8	.52455	0.0616	3.621	0.004
12	64.4	.57366	0.1857	11.959	0.034
13	81.5	.62276	0.3127	25.488	0.098
14	85.6	.67186	0.4451	38.097	0.198
15	151.0	.72097	0.5857	88.445	0.343
16	262.0	.77007	0.7391	193.638	0.546
17	331.0	.81918	0.9122	301.953	0.832
18	578.0	.86828	1.1183	646.376	1.251
19	637.0	.91738	1.3877	883.941	1.926
20	942.0	.96594	1.8242	1718.408	3.328

Step 4. Compute the Probability Plot Correlation Coefficient using the simplified formula for r, where SD=259.72 and C_n=4.1375, to get

$$r = \frac{3836.81}{(4.1375)(259.72)\sqrt{19}} = 0.819$$

Step 5. Compare the computed value of r=0.819 to the 5% critical value for sample size 20 in Table A-4, namely $R_{.05,20}$=0.950. Since $r < 0.950$, the sample shows significant evidence of non-Normality by the Probability Plot Correlation Coefficient test. The data should be transformed using natural logs and the correlation coefficient recalculated before proceeding with further statistical analysis.

EXAMPLE 5

The data in Examples 1, 2, 3, and 4 showed significant evidence of non-Normality. Instead of first logging the concentrations before testing for Normality, the original data were used. This was done to illustrate why the Lognormal distribution is usually a better default model than the Normal. In this example, use the same data to determine whether the measurements better follow a Lognormal distribution.

Computing the natural logarithms of the data gives the table below.

	Logged Nickel Concentrations log (ppb)			
Month	Well 1	Well 2	Well 3	Well 4
1	4.07	2.94	3.66	1.13
2	0.00	4.40	5.02	6.85
3	5.57	5.80	3.30	4.45
4	4.03	2.64	3.06	2.30
5	2.16	4.17	6.36	6.46

SOLUTION

Method 1. Probability Plots

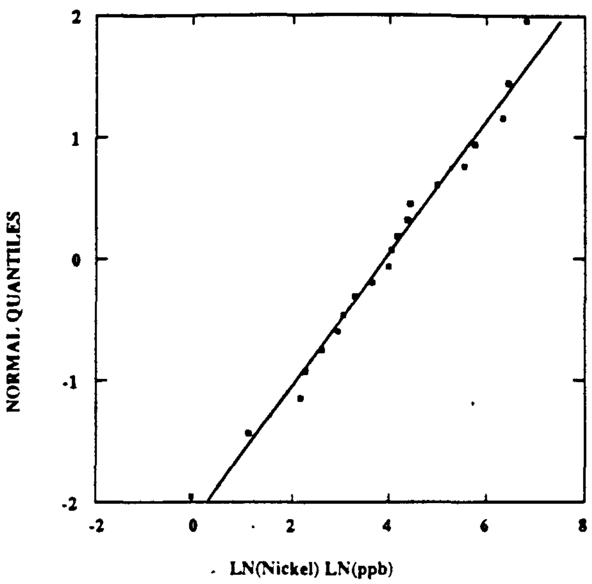
Step 1. List the natural logarithms of the measured nickel concentrations in order from lowest to highest.

Order (i)	Log Nickel Concentration log(ppb)	Probability 100*(i/(n+1))	Normal Quantiles
1	0.00	5	-1.645
2	1.13	10	-1.28
3	2.16	14	-1.08
4	2.30	19	-0.88
5	2.64	24	-0.706
6	2.94	29	-0.55
7	3.06	33	-0.44
8	3.30	38	-0.305
9	3.66	43	-0.176
10	4.03	48	-0.05
11	4.07	52	0.05
12	4.17	57	0.176
13	4.40	62	0.305
14	4.45	67	0.44
15	5.02	71	0.55
16	5.57	76	0.706
17	5.80	81	0.88
18	6.36	86	1.08
19	6.46	90	1.28
20	6.85	95	1.645

Step 2. Compute the probability as shown in the third column by calculating 100*(i/n+1), where n is the total number of samples (n=20). The corresponding Normal quantiles are given in column 4.

Step 3. Plot the Normal quantiles against the natural logarithms of the observed concentrations to get the following graph. The plot indicates a nearly straight line fit (verified by calculation of the Correlation Coefficient given in Method 4). There is no substantial evidence that the data do not follow a Lognormal distribution. The Normal-theory procedure(s) should be performed on the log-transformed data.

PROBABILITY PLOT



Method 2. Coefficient of Skewness

Step 1. Calculate the mean, SD, and average cubed residuals of the natural logarithms of the data.

$$\bar{x} = 3.918 \log(\text{ppb})$$
$$SD = 1.802 \log(\text{ppb})$$

$$\frac{1}{n}\Sigma_i(x_i - \bar{x})^3 = -1.325 \log^3(\text{ppb})$$

Step 2. Calculate the Skewness Coefficient, γ_1.

$$\gamma_1 = \frac{-1.325}{(.95)^{\frac{3}{2}}(1.802)^3} = -0.244$$

Step 3. Compute the absolute value of the skewness, $|\gamma_1|=|-0.244|=0.244$.

Step 4. Since the absolute value of the Skewness Coefficient is less than 1, the data do not show evidence of significant skewness. A Normal approximation to the log-transformed data may therefore be appropriate, but this model should be further checked.

Method 3. Shapiro-Wilk Test

Step 1. Order the logged data from smallest to largest and list, as in following table. Also list the data in reverse order and compute the differences $x_{(n-i+1)}-x_{(i)}$.

i	$LN(x_{(i)})$	$LN(x_{(n-i+1)})$	a_{n-i+1}	b_i
1	0.00	6.85	.4734	3.24
2	1.13	6.46	.3211	1.71
3	2.16	6.36	.2565	1.08
4	2.30	5.80	.2085	0.73
5	2.64	5.57	.1686	0.49
6	2.94	5.02	.1334	0.28
7	3.06	4.45	.1013	0.14
8	3.30	4.40	.0711	0.08
9	3.66	4.17	.0422	0.02
10	4.03	4.07	.0140	0.00
11	4.07	4.03		b=7.77
12	4.17	3.66		
13	4.40	3.30		
14	4.45	3.06		
15	5.02	2.94		
16	5.57	2.64		
17	5.80	2.30		
18	6.36	2.16		
19	6.46	1.13		
20	6.85	0.00		

Step 2. Compute k=10, since n/2=10. Look up the coefficients a_{n-i+1} from Table A-1 and multiply by the first k differences between columns 2 and 1 to get the quantities b_i. Add these 10 products to get b=7.77.

Step 3. Compute the standard deviation of the logged data, SD=1.8014. Then the Shapiro-Wilk statistic is given by

$$W = \left[\frac{7.77}{1.8014\sqrt{19}}\right]^2 = 0.979.$$

Step 4. Compare the computed value of W to the 5% critical value for sample size 20 in Table A-2, namely $W_{.05,20}=0.905$. Since W=0.979>0.905, the sample shows no significant evidence of non-Normality by the Shapiro-Wilk test. Proceed with further statistical analysis using the log-transformed data.

Method 4. Probability Plot Correlation Coefficient

Step 1. Order the logged data from smallest to largest and list below.

Order (i)	Log Nickel Concentration log(ppb)	m_i	M_i	$X_{(i)}*M_i$	M_i^2
1	0.00	.03406	-1.8242	0.000	3.328
2	1.13	.08262	-1.3877	-1.568	1.926
3	2.16	.13172	-1.1183	-2.416	1.251
4	2.30	.18082	-0.9122	-2.098	0.832
5	2.64	.22993	-0.7391	-1.951	0.546
6	2.94	.27903	-0.5857	-1.722	0.343
7	3.06	.32814	-0.4451	-1.362	0.198
8	3.30	.37724	-0.3127	-1.032	0.098
9	3.66	.42634	-0.1857	-0.680	0.034
10	4.03	.47545	-0.0616	-0.248	0.004
11	4.07	.52455	0.0616	0.251	0.004
12	4.17	.57366	0.1857	0.774	0.034
13	4.40	.62276	0.3127	1.376	0.098
14	4.45	.67186	0.4451	1.981	0.198
15	5.02	.72097	0.5857	2.940	0.343
16	5.57	.77007	0.7391	4.117	0.546
17	5.80	.81918	0.9122	5.291	0.832
18	6.36	.86828	1.1183	7.112	1.251
19	6.46	.91738	1.3877	8.965	1.926
20	6.85	.96594	1.8242	12.496	3.328

Step 2. Compute the quantities m_i and the order statistic medians M_i, according to the procedure in Example 4 (note that these values depend only on the sample size and are identical to the quantities in Example 4).

Step 3. Compute the products $X_{(i)}*M_i$ in column 4 and sum to get the numerator of the correlation coefficient (equal to 32.226 in this case). Also compute M_i^2 in column 5 and sum to find quantity $C_n^2=17.12$.

Step 4. Compute the Probability Plot Correlation Coefficient using the simplified formula for r, where SD=1.8025 and $C_n=4.1375$, to get

$$r=\frac{32.226}{(4.1375)(1.8025)\sqrt{19}}=0.991$$

Step 5. Compare the computed value of r=0.991 to the 5% critical value for sample size 20 in Table A-4, namely $R_{.05,20}=0.950$. Since $r > 0.950$, the logged data show no significant evidence of non-Normality by the Probability Plot Correlation Coefficient test. Therefore, Lognormality of the original data could be assumed in subsequent statistical procedures.

1.2 TESTING FOR HOMOGENEITY OF VARIANCE

One of the most important assumptions for the parametric analysis of variance (ANOVA) is that the different groups (e.g., different wells) have approximately the same variance. If this is not the case, the power of the F-test (its ability to detect differences among the group means) is reduced. Mild differences in variance are not too bad. The effect becomes noticeable when the largest and smallest group variances differ by a ratio of about 4 and becomes quite severe when the ratio is 10 or more (Milliken and Johnson, 1984).

The procedure suggested in the EPA guidance document, Bartlett's test, is one way to test whether the sample data give evidence that the well groups have different variances. However, Bartlett's test is sensitive to non-Normality in the data and may give misleading results unless one knows in advance that the data are approximately Normal (Milliken and Johnson, 1984). As an alternative to Bartlett's test, two procedures for testing homogeneity of the variances are described below that are less sensitive to non-Normality.

1.2.1 Box Plots

Box Plots were first developed for exploratory data analysis as a quick way to visualize the "spread" or dispersion within a data set. In the context of variance testing, one can construct a Box Plot for each well group and compare the boxes to see if the assumption of equal variances is reasonable. Such a comparison is not a formal test procedure, but is easier to perform and is often sufficient for checking the group variance assumption.

The idea behind a Box Plot is to order the data from lowest to highest and to trim off 25 percent of the observations on either end, leaving just the middle 50 percent of the sample values. The spread between the lowest and highest values of this middle 50 percent (known as the interquartile range or IQR) is represented by the length of the box. The very middle observation (i.e., the median) can also be shown as a line cutting the box in two.

To construct a Box Plot, calculate the median and upper and lower quantiles of the data set (respectively, the 50th, 25th, and 75th percentiles). To do this, calculate $k=p(n+1)/100$ where n=number of samples and p=percentile of interest. If k is an integer, let the kth ordered or ranked value be an estimate of the pth percentile of the data. If k is not an integer, let the pth percentile be equal to the average of the two values closest in rank position to k. For example, if the data set consists of the 10 values {1, 4, 6.2, 10, 15, 17.1, 18, 22, 25, 30.5}, the position of the median would be found as 50*(10+1)/100=5.5. The median would then be computed as the average of the 5th and 6th ordered values, or (15+17.1)/2=16.05.

Likewise, the position of the lower quartile would be 25*(10+1)/100=2.75. Calculate the average of the 2nd and 3rd ordered observations to estimate this percentile, i.e., (4+6.2)/2=5.1. Since the upper quartile is found to be 23.5, the length of Box Plot would be the difference between the upper and lower quartiles, or (23.5–5.1)=18.4. The box itself should be drawn on a graph with the y-axis representing concentration and the x-axis denoting the wells being plotted. Three horizontal lines are drawn for each well, one line each at the lower and upper quartiles and another at the median concentration. Vertical connecting lines are drawn to complete the box.

Most statistics packages can directly calculate the statistics needed to draw a Box Plot, and many will construct the Box Plots as well. In some computer packages, the Box Plot will also have two "whiskers" extending from the edges of the box. These lines indicate the positions of extreme values in the data set, but generally should not be used to approximate the overall dispersion.

If the box length for each group is less than 3 times the length of the shortest box, the sample variances are probably close enough to assume equal group variances. If, however, the box length for any group is at least triple the length of the box for another group, the variances may be significantly different (Kirk Cameron, SAIC, personal communication). In that case, the data should be further checked using Levene's test described in the following section. If Levene's test is significant, the data may need to be transformed or a non-parametric rank procedure considered before proceeding with further analysis.

EXAMPLE 6

Construct Box Plots for each well group to test for equality of variances.

	Arsenic Concentration (ppm)					
Month	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6
1	22.9	2.0	2.0	7.84	24.9	0.34
2	3.09	1.25	109.4	9.3	1.3	4.78
3	35.7	7.8	4.5	25.9	0.75	2.85
4	4.18	52	2.5	2.0	27	1.2

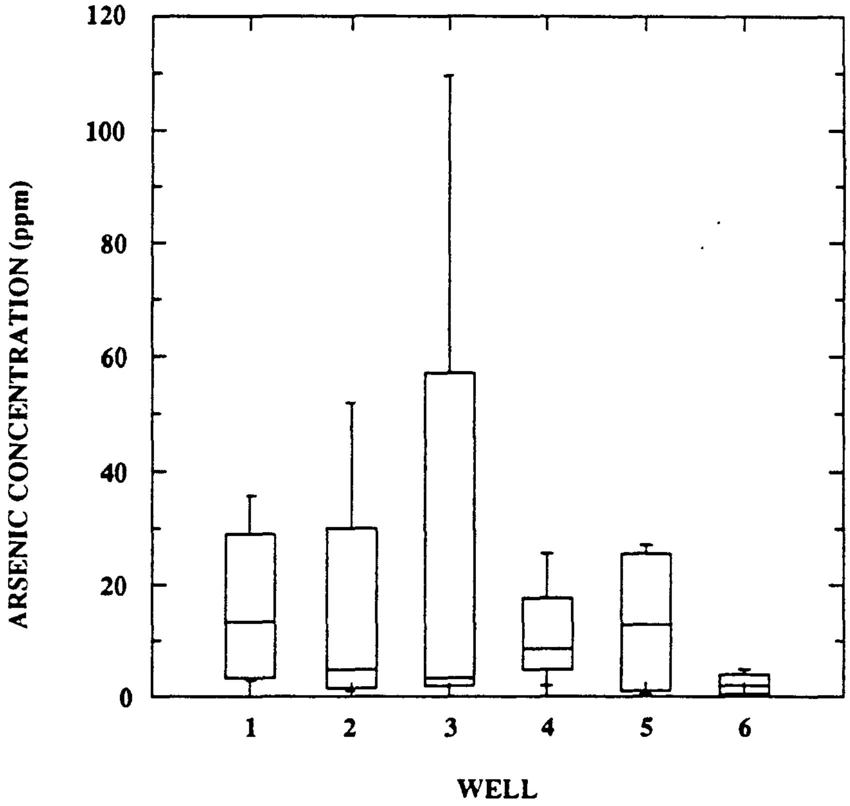
SOLUTION

Step 1. Compute the 25th, 50th, and 75th percentiles for the data in each well group. To calculate the pth percentile by hand, order the data from lowest to highest. Calculate p*(n+1)/100 to find the ordered position of the pth percentile. If necessary, interpolate between sample values to estimate the desired percentile.

Step 2. Using well 1 as an example, n+1=5 (since there are 4 data values). To calculate the 25th percentile, compute its ordered position (i.e., rank) as 25*5/100=1.25. Average the 1st and 2nd ranked values at well 1 (i.e., 3.09 and 4.18) to find an estimated lower quartile of 3.64. This estimate gives the lower end of the Box Plot. The upper end or 75th percentile can be computed similarly as the average of the 3rd and 4th ranked values, or (22.9+35.7)/2=29.3. The median is the average of the 2nd and 3rd ranked values, giving an estimate of 13.14.

Step 3. Construct Box Plots for each well group, lined up side by side on the same axes.

BOX PLOTS OF WELL DATA



Step 4. Since the box length for well 3 is more than three times the box lengths for wells 4 and 6, there is evidence that the group variances may be significantly different. These data should be further checked using Levene's test described in the next section.

1.2.2 Levene's Test

Levene's test is a more formal procedure than Box Plots for testing homogeneity of variance that, unlike Bartlett's test, is not sensitive to non-Normality in the data. Levene's test has been shown to have power nearly as great as Bartlett's test for Normally distributed data and power superior to Bartlett's for non-Normal data (Milliken and Johnson, 1984).

To conduct Levene's test, first compute the new variables

$$z_{ij} = \left| x_{ij} - \bar{x}_i \right|$$

where x_{ij} represents the jth value from the ith well and $\bar{x}_i$ is the ith well mean. The values z_{ij} represent the absolute values of the usual residuals. Then run a standard one-way analysis of variance (ANOVA) on the variables z_{ij}. If the F-test is significant, reject the hypothesis of equal group variances. Otherwise, proceed with analysis of the x_{ij}'s as initially planned.

EXAMPLE 7

Use the data from Example 6 to conduct Levene's test of equal variances.

SOLUTION

Step 1. Calculate the group mean for each well $(\bar{x}_i)$

Well 1 mean = 16.47	Well 4 mean = 11.26
Well 2 mean = 15.76	Well 5 mean = 13.49
Well 3 mean = 29.60	Well 6 mean = 2.29

Step 2. Compute the absolute residuals z_{ij} in each well and the well means of the residuals ($\bar{z}_i$).

	Absolute Residuals					
Month	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6
1	6.43	13.76	27.6	3.42	11.41	1.95
2	13.38	14.51	79.8	1.96	12.19	2.49
3	19.23	7.96	25.1	14.64	12.74	0.56
4	12.29	36.24	27.1	9.26	13.51	1.09
Well Mean ($\bar{z}_i$)	= 12.83	18.12	39.9	7.32	12.46	1.52
Overall Mean ($\bar{z}$)	= 15.36					

Step 3. Compute the sums of squares for the absolute residuals.

$$SS_{TOTAL} = (N-1)\ SDz^2 = 6300.89$$

$$SS_{WELLS} = \sum n_i \bar{z}_i^2 - N\bar{z}^2 = 3522.90$$

$$SS_{ERROR} = SS_{TOTAL} - SS_{WELLS} = 2777.99$$

Step 4. Construct an analysis of variance table to calculate the F-statistic. The degrees of freedom (df) are computed as (#groups–1)=(6–1)=5 df and (#samples–#groups)=(24–6)=18 df.

		ANOVA Table			
Source	Sum-of-Squares	df	Mean-Square	F-Ratio	P
Between Wells	3522.90	5	704.58	4.56	0.007
Error	2777.99	18	154.33		
Total	6300.89	23			

Step 5. Since the F-statistic of 4.56 exceeds the tabulated value of $F_{.05}$=2.77 with 5 and 18 df, the assumption of equal variances should be rejected. Since the original concentration data are used in this example, the data should be logged and retested.

2. RECOMMENDATIONS FOR HANDLING NONDETECTS

The basic recommendations within the Interim Final Guidance for handling nondetect analyses include the following (see p. 8-2): 1) if less than 15 percent of all samples are nondetect, replace each nondetect by half its detection or quantitation limit and proceed with a parametric analysis, such as ANOVA, Tolerance Limits, or Prediction Limits; 2) if the percent of nondetects is between 15 and 50, either use Cohen's adjustment to the sample mean and variance in order to proceed with a parametric analysis, or employ a non-parametric procedure by using the ranks of the observations and by treating all nondetects as tied values; 3) if the percent of nondetects is greater than 50 percent, use the Test of Proportions.

As to the first recommendation, experience at EPA and research at the United States Geological Survey (USGS, Dennis Helsel, personal communication, 1991) has indicated that if less than 15 percent of the samples are nondetect, the results of parametric statistical tests will not be substantially affected if nondetects are replaced by half their detection limits. When more than 15 percent of the samples are nondetect, however, the handling of nondetects is more crucial to the outcome of statistical procedures. Indeed, simple substitution methods tend to perform poorly in statistical tests when the nondetect percentage is substantial (Gilliom and Helsel, 1986).

Even with a small proportion of nondetects, however, care should be taken when choosing between the method detection limit (MDL) and the practical quantitation limit (PQL) in characterizing "nondetect" concentrations. Many nondetects are characterized by analytical laboratories with one of three data qualifier flags: "U," "J," or "E." Samples with a "U" data qualifier represent "undetected" measurements, meaning that the signal characteristic of that analyte could not be observed or distinguished from "background noise" during lab analysis. Inorganic samples with an "E" flag and organic samples with a "J" flag may or may not be reported with an estimated concentration. If no concentration is estimated, these samples represent "detected but not quantified" measurements. In this case, the actual concentration is assumed to be positive, but somewhere between zero and the PQL. Since all of these non-detects may or may not have actual positive concentrations between zero and the PQL, the suggested substitution for parametric statistical procedures is to replace each nondetect by one-half the PQL (note, however, that "E" and "J" samples reported with estimated concentrations should be treated, for statistical purposes, as valid measurements. Substitution of one-half the PQL is not recommended for these samples).

In no case should nondetect concentrations be assumed to be bounded above by the MDL. The MDL is estimated on the basis of ideal laboratory conditions with ideal analyte samples and does not account for matrix or other interferences encountered when analyzing specific, actual field samples. For this reason, the PQL should be taken as the most reasonable upper bound for nondetect concentrations.

It should also be noted that the distinction between "undetected" and "detected but not quantified" measurements has more specific implications for rank-based non-parametric procedures. Rather than assigning the same tied rank to all nondetects (see below and in **Section 3**), "detected but not quantified" measurements should be given larger ranks than those assigned to "undetected" samples. In fact the two types of nondetects should be treated as two distinct groups of tied observations for use in the Wilcoxon and Kruskal-Wallis non-parametric procedures.

2.1 NONDETECTS IN ANOVA PROCEDURES

For a moderate to large percentage of nondetects (i.e., over 15%), the handling of nondetects should vary depending on the statistical procedure to be run. If background data from one or more upgradient wells are to be compared simultaneously with samples from one or more downgradient wells via a t-test or ANOVA type procedure, the simplest and most reliable recommendation is to switch to a non-parametric analysis. The distributional assumptions for parametric procedures can be rather difficult to check when a substantial fraction of nondetects exists. Furthermore, the non-parametric alternatives described in **Section 3** tend to be efficient at detecting contamination when the underlying data are Normally distributed, and are often more powerful than the parametric methods when the underlying data do not follow a Normal distribution.

Nondetects are handled easily in a nonparametric analysis. All data values are first ordered and replaced by their ranks. Nondetects are treated as tied values and replaced by their midranks (see **Section 3**). Then a Wilcoxon Rank-Sum or Kruskal-Wallis test is run on the ranked data depending on whether one or more than one downgradient well is being tested.

The Test of Proportions is not recommended in this Addendum, even if the percentage of nondetects is over 50 percent. Instead, for all two-group comparisons that involve more than 15 percent nondetects, the non-parametric Wilcoxon Rank-Sum procedure is recommended. Although acceptable as a statistical procedure, the Test of Proportions does not account for potentially different magnitudes among the concentrations of detected values. Rather, each sample is treated as a 0 or 1 depending on whether the measured concentration is below or above the detection limit. The Test of Proportions ignores information about concentration magnitudes, and hence is usually less powerful than a non-parametric rank-based test like the Wilcoxon Rank-Sum, even after adjusting for a large fraction of tied observations (e.g., nondetects). This is because the ranks of a dataset preserve additional information about the relative magnitudes of the concentration values, information which is lost when all observations are scored as 0's and 1's.

Another drawback to the Test of Proportions, as presented in the Interim Final Guidance, is that the procedure relies on a Normal probability approximation to the Binomial distribution of 0's and 1's. This approximation is recommended only when the quantities n × (%NDs) and n × (1−%NDs) are no smaller than 5. If the percentage of nondetects is quite high and/or the sample size is fairly small, these conditions may be violated, leading potentially to inaccurate results.

Comparison of the Test of Proportions to the Wilcoxon Rank-Sum test shows that for small to moderate proportions of nondetects (say 0 to 60 percent), the Wilcoxon Rank-Sum procedure adjusted for ties is more powerful in identifying real concentration differences than the Test of Proportions. When the percentage of nondetects is quite high (at least 70 to 75 percent), the Test of Proportions appears to be slightly more powerful in some cases than the Wilcoxon, but the results of the two tests almost always lead to the same conclusion, so it makes sense to simply recommend the Wilcoxon Rank-Sum test in all cases where nondetects constitute more than 15 percent of the samples.

2.2 NONDETECTS IN STATISTICAL INTERVALS

If the chosen method is a statistical interval (Confidence, Tolerance or Prediction limit) used to compare background data against each downgradient well separately, more options are available for handling moderate proportions of nondetects. The basis of any parametric statistical interval limit is the formula $\bar{x} \pm \kappa \cdot s$, where $\bar{x}$ and s represent the sample mean and standard deviation of the (background) data and κ depends on the interval type and characteristics of the monitoring network. To use a parametric interval in the presence of a substantial number of nondetects, it is necessary to estimate the sample mean and standard deviation. But since nondetect concentrations are unknown, simple formulas for the mean and standard deviation cannot be computed directly. Two basic approaches to estimating or "adjusting" the mean and standard deviation in this situation have been described by Cohen (1959) and Aitchison (1955).

The underlying assumptions of these procedures are somewhat different. Cohen's adjustment (which is described in detail on pp. 8-7 to 8-11 of the Interim Final Guidance) assumes that all the data (detects and nondetects) come from the same Normal or Lognormal population, but that nondetect values have been "censored" at their detection limits. This implies that the contaminant of concern is present in nondetect samples, but the analytical equipment is not sensitive to concentrations lower than the detection limit. Aitchison's adjustment, on the other hand, is constructed on the assumption that nondetect samples are free of contamination, so that all nondetects may be regarded as zero concentrations. In some situations, particularly when the analyte of concern has been detected infrequently in background measurements, this assumption may be practical, even if it cannot be verified directly.

Before choosing between Cohen's and Aitchison's approaches, it should be cautioned that Cohen's adjustment may not give valid results if the proportion of nondetects exceeds 50%. In a case study by McNichols and Davis (1988), the false positive rate associated with the use of t-tests based on Cohen's method rose substantially when the fraction of nondetects was greater than 50%. This occurred because the adjusted estimates of the mean and standard deviation are more highly correlated as the percentage of nondetects increases, leading to less reliable statistical tests (including statistical interval tests).

On the other hand, with less than 50% nondetects, Cohen's method performed adequately in the McNichols and Davis case study, provided the data were not overly skewed and that more extensive tables than those included within the Interim Final Guidance were available to calculate Cohen's adjustment parameter. As a remedy to the latter caveat, a more extensive table of Cohen's adjustment parameter is provided in Appendix A (Table A-5). It is also recommended that the data (detected measurements and nondetect detection limits) first be log-transformed prior to computing either Cohen's or Aitchison's adjustment, especially since both procedures assume that the underlying data are Normally distributed.

2.2.1 Censored and Detects-Only Probability Plots

To decide which approach is more appropriate for a particular set of ground water data, two separate Probability Plots can be constructed. The first is called a Censored Probability Plot and is a test of Cohen's underlying assumption. In this method, the combined set of detects and nondetects is ordered (with nondetects being given arbitrary but distinct ranks). Cumulative probabilities or Normal quantiles (see **Section 1.1**) are then computed for the data set as in a regular Probability Plot. However, only the detected values and their associated Normal quantiles are actually plotted. If the shape of the Censored Probability Plot is reasonably linear, then Cohen's assumption that nondetects have been "censored" at their detection limit is probably acceptable and Cohen's adjustment can be made to estimate the sample mean and standard deviation. If the Censored Probability Plot has significant bends and curves, particularly in one or both tails, one might consider Aitchison's procedure instead.

To test the assumptions of Aitchison's method, a Detects-Only Probability Plot may be constructed. In this case, nondetects are completely ignored and a standard Probability Plot is constructed using only the detected measurements. Thus, cumulative probabilities or Normal quantiles are computed only for the ordered detected values. Comparison of a Detects-Only Probability Plot with a Censored Probability Plot will indicate that the same number of points and concentration values are plotted on each graph. However, different Normal quantiles are associated with each detected concentration. If the Detects-Only Probability Plot is reasonably linear, then the assumptions underlying Aitchison's adjustment (i.e., that "nondetects" represent zero concentrations, and that detects and nondetects follow separate probability distributions) are probably reasonable.

If it is not clear which of the Censored or Detects-Only Probability Plots is more linear, Probability Plot Correlation Coefficients can be computed for both approaches (note that the correlations should only involve the points actually plotted, that is, detected concentrations). The plot with the higher correlation coefficient will represent the most linear trend. Be careful, however, to use other, non-statistical judgments to help decide which of Cohen's and Aitchison's underlying assumptions appears to be most reasonable based on the specific characteristics of the data set. It is also likely that these Probability Plots may have to be constructed on the logarithms of the data instead of the original values, if in fact the most appropriate underlying distribution is the Lognormal instead of the Normal.

EXAMPLE 8

Create Censored and Detects-Only Probability Plots with the following zinc data to determine whether Cohen's adjustment or Aitchison's adjustment is most appropriate for estimating the true mean and standard deviation.

	Zinc Concentrations (ppb) at Background Wells				
Sample	Well 1	Well 2	Well 3	Well 4	Well 5
1	<7	<7	<7	11.69	<7
2	11.41	<7	12.85	10.90	<7
3	<7	13.70	14.20	<7	<7
4	<7	11.56	9.36	12.22	11.15
5	<7	<7	<7	11.05	13.31
6	10.00	<7	12.00	<7	12.35
7	15.00	10.50	<7	13.24	<7
8	<7	12.59	<7	<7	8.74

SOLUTION

Step 1. Pool together the data from the five background wells and list in order in the table below.

Step 2. To construct the Censored Probability Plot, compute the probabilities i/(n+1) using the combined set of detects and nondetects, as in column 3. Find the Normal quantiles associated with these probabilities by applying the inverse standard Normal transformation, Φ^{-1}.

Step 3. To construct the Detects-Only Probability Plot, compute the probabilities in column 5 using only the detected zinc values. Again apply the inverse standard Normal transformation to find the associated Normal quantiles in column 6. Note that nondetects are ignored completely in this method.

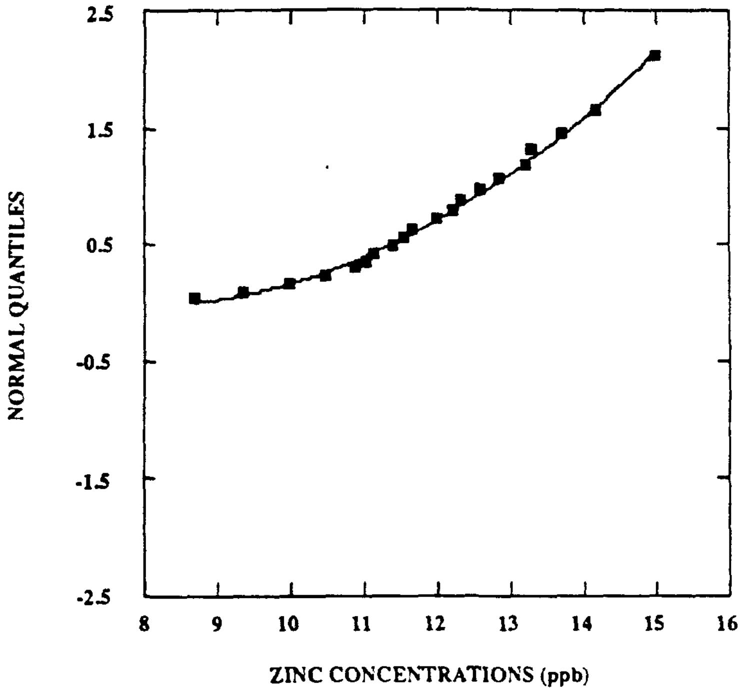
Order (i)	Zinc Conc. (ppb)	Censored Probs.	Normal Quantiles	Detects-Only Probs.	Normal Quantiles
1	<7	.024	-1.971		
2	<7	.049	-1.657		
3	<7	.073	-1.453		
4	<7	.098	-1.296		
5	<7	.122	-1.165		
6	<7	.146	-1.052		
7	<7	.171	-0.951		
8	<7	.195	-0.859		
9	<7	.220	-0.774		
10	<7	.244	-0.694		
11	<7	.268	-0.618		
12	<7	.293	-0.546		
13	<7	.317	-0.476		
14	<7	.341	-0.408		
15	<7	.366	-0.343		
16	<7	.390	-0.279		
17	<7	.415	-0.216		
18	<7	.439	-0.153		
19	<7	.463	-0.092		
20	<7	.488	-0.031		
21	8.74	.512	0.031	.048	-1.668
22	9.36	.537	0.092	.095	-1.309
23	10.00	.561	0.153	.143	-1.068
24	10.50	.585	0.216	.190	-0.876
25	10.90	.610	0.279	.238	-0.712
26	11.05	.634	0.343	.286	-0.566
27	11.15	.659	0.408	.333	-0.431
28	11.41	.683	0.476	.381	-0.303
29	11.56	.707	0.546	.429	-0.180
30	11.69	.732	0.618	.476	-0.060
31	12.00	.756	0.694	.524	0.060
32	12.22	.780	0.774	.571	0.180
33	12.35	.805	0.859	.619	0.303
34	12.59	.829	0.951	.667	0.431
35	12.85	.854	1.052	.714	0.566
36	13.24	.878	1.165	.762	0.712
37	13.31	.902	1.296	.810	0.876
38	13.70	.927	1.453	.857	1.068
39	14.20	.951	1.657	.905	1.309
40	15.00	.976	1.971	.952	1.668

Step 4. Plot the detected zinc concentrations versus each set of probabilities or Normal quantiles, as per the procedure for constructing Probability Plots (see figures below). The nondetect values should not be plotted. As can be seen from the graphs, the Censored Probability Plot indicates a definite curvature in the tails, especially the lower tail. The Detects-Only Probability Plot, however, is reasonably linear. This visual impression is bolstered by calculation of a Probability Plot Correlation Coefficient for each set of

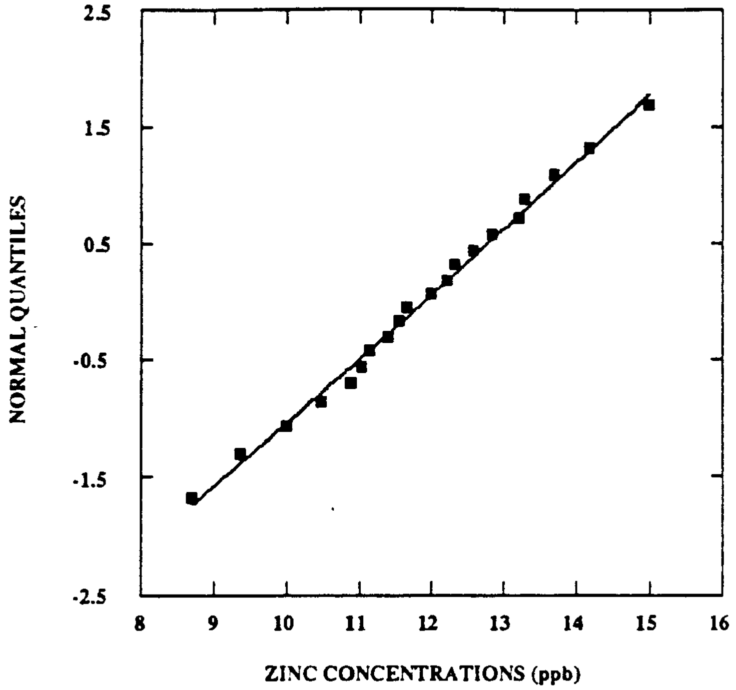
detected values: the Censored Probability Plot has a correlation of r=.969, while the Detects-Only Probability Plot has a correlation of r=.998.

Step 5. Because the Detects-Only Probability Plot is substantially more linear than the Censored Probability Plot, it may be appropriate to consider detects and nondetects as arising from statistically distinct distributions, with nondetects representing "zero" concentrations. Therefore, Aitchison's adjustment may lead to better estimates of the true mean and standard deviation than Cohen's adjustment for censored data.

CENSORED PROBABILITY PLOT



DETECTS-ONLY PROBABILITY PLOT



2.2.2 Aitchison's Adjustment

To actually compute Aitchison's adjustment (Aitchison, 1955), it is assumed that the detected samples follow an underlying Normal distribution. If the detects are Lognormal, compute Aitchison's adjustment on the logarithms of the data instead. Let d=# nondetects and let n=total # of samples (detects and nondetects combined). Then if $\bar{x}^*$ and s^* denote respectively the sample mean and standard deviation of the detected values, the adjusted overall mean can be estimated as

$$\hat{\mu} = \left(1 - \frac{d}{n}\right)\bar{x}^*$$

and the adjusted overall standard deviation may be estimated as the square root of the quantity

$$\hat{\sigma}^2 = \frac{n-(d+1)}{n-1}(s^*)^2 + \frac{d}{n}\left(\frac{n-d}{n-1}\right)(\bar{x}^*)^2$$

The general formula for a parametric statistical interval adjusted for nondetects by Aitchison's method is given by $\hat{\mu} \pm \kappa \cdot \hat{\sigma}$, with κ depending on the type of interval being constructed.

EXAMPLE 9

In Example 8, it was determined that Aitchison's adjustment might lead to more appropriate estimates of the true mean and standard deviation than Cohen's adjustment. Use the data in Example 8 to compute Aitchison's adjustment.

SOLUTION

Step 1. The zinc data consists of 20 nondetects and 20 detected values; therefore d=20 and n=40 in the above formulas.

Step 2. Compute the average $\bar{x}^* = 11.891$ and the standard deviation $s^* = 1.595$ of the set of detected values.

Step 3. Use the formulas for Aitchison's adjustment to compute estimates of the true mean and standard deviation:

$$\hat{\mu} = \left(1 - \frac{20}{40}\right) \times 11.891 = 5.95$$

$$\hat{\sigma}^2 = \left(\frac{40-21}{39}\right)(1.595)^2 + \left(\frac{20}{40}\right)\left(\frac{20}{39}\right)(11.891)^2 = 37.495 \Rightarrow \hat{\sigma} = 6.12$$

If Cohen's adjustment is mistakenly computed on these data instead, with a detection limit of 7 ppb,the estimates become $\hat{\mu} = 7.63$ and $\hat{\sigma} = 4.83$. Thus, the choice of adjustment can have a significant impact on the upper limits computed for statistical intervals.

2.2.3 More Than 50% Nondetects

If more than 50% but less than 90% of the samples are nondetect or the assumptions of Cohen's and Aitchison's methods cannot be justified, parametric statistical intervals should be abandoned in favor of non-parametric alternatives (see **Section 3** below). Nonparametric statistical intervals are easy to construct and apply to ground water data measurements, and no special steps need be taken to handle nondetects.

When 90% or more of the data values are nondetect (as often occurs when measuring volatile organic compounds [VOCs] in ground water, for instance), the detected samples can often be modeled as "rare events" by using the Poisson distribution. The Poisson model describes the behavior of a series of independent events over a large number of trials, where the probability of occurrence is low but stays constant from trial to trial. The Poisson model is similar to the Binomial model in that both models represent "counting processes." In the Binomial case, nondetects are counted as 'misses' or zeroes and detects are counted (regardless of contamination level) as 'hits' or ones; in the case of the Poisson, each particle or molecule of contamination is counted separately but cumulatively, so that the counts for detected samples with high concentrations are larger than counts for samples with smaller concentrations. As Gibbons (1987, p. 574) has noted, it can be postulated

> ...that the number of molecules of a particular compound out of a much larger number of molecules of water is the result of a Poisson process. For example, we might consider 12 ppb of benzene to represent a count of 12 units of benzene for every billion units examined. In this context, Poisson's approach is justified in that the number of units (i.e., molecules) is large, and the probability of the occurrence (i.e., a molecule being classified as benzene) is small.

For a detect with concentration of 50 ppb, the Poisson count would be 50. Counts for nondetects can be taken as zero or perhaps equal to half the detection limit (e.g., if the detection limit were 10 ppb, the Poisson count for that sample would be 5). Unlike the Binomial (Test of Proportions) model, the Poisson model has the ability to utilize the <u>magnitudes</u> of detected concentrations in statistical tests.

The Poisson distribution is governed by the average rate of occurrence, λ, which can be estimated by summing the Poisson counts of all samples in the background pool of data and dividing by the number of samples in the pool. Once the average rate of occurrence has been estimated, the formula for the Poisson distribution is given by

$$\Pr\{X = x\} = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

where x represents the Poisson count and λ represents the average rate of occurrence. To use the Poisson distribution to predict concentration values at downgradient wells, formulas for constructing Poisson Prediction and Tolerance limits are given below.

2.2.4 Poisson Prediction Limits

To estimate a Prediction limit at a particular well using the Poisson model, the approach described by Gibbons (1987b) and based on the work of Cox and Hinkley (1974) can be used. In this case, an upper limit is estimated for an interval that will contain <u>all</u> of k future measurements of an analyte with confidence level $1-\alpha$, given n previous background measurements.

To do this, let T_n represent the sum of the Poisson counts of n background samples. The goal is to predict T_k^*, representing the total Poisson count of the next k sample measurements. As

Cox and Hinkley show, if T_n has a Poisson distribution with mean μ and if no contamination has occurred, it is reasonable to assume that T_k^* will also have a Poisson distribution but with mean $c\mu$, where c depends on the number of future measurements being predicted.

In particular, Cox and Hinckley demonstrate that the quantity

$$\left[T_k^* - \frac{c(T_n + T_k^*)}{(1+c)}\right]^2 \Bigg/ \frac{c(T_n + T_k^*)}{(1+c)^2}$$

has an approximate standard Normal distribution. From this relation, an upper prediction limit for T_k^* is calculated by Gibbons to be approximately

$$T_k^* = cT_n + \frac{ct^2}{2} + ct\sqrt{T_n\left(1+\frac{1}{c}\right) + \frac{t^2}{4}}$$

where $t = t_{n-1,\alpha}$ is the upper $(1-\alpha)$ percentile of the Student's t distribution with (n-1) degrees of freedom. The quantity c in the above formulas may be computed as k/n, where, as noted, k is the number of future samples being predicted.

EXAMPLE 10

Use the following benzene data from six background wells to estimate an upper 99% Poisson Prediction limit for the next four measurements from a single downgradient well.

	Benzene Concentrations (ppb)					
Month	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6
1	<2	<2	<2	<2	<2	<2
2	<2	<2	<2	15.0	<2	<2
3	<2	<2	<2	<2	<2	<2
4	<2	12.0	<2	<2	<2	<2
5	<2	<2	<2	<2	<2	10.0
6	<2	<2	<2	<2	<2	<2

SOLUTION

Step 1. Pooling the background data yields n=36 samples, of which, 33 (92%) are nondetect. Because the rate of detection is so infrequent (i.e., <10%), a Poisson-based Prediction limit may be appropriate. Since four future measurements are to be predicted, k=4, and hence, c=k/n=1/9.

Step 2. Set each nondetect to half the detection limit or 1 ppb. Then compute the Poisson count of the sum of all the background samples, in this case, T_n=33(1)+(12.0+15.0+10.0) = 70.0. To calculate an upper 99% Prediction limit, the upper 99th percentile of the t-distribution with (n-1)=35 degrees of freedom must be taken from a reference table, namely $t_{35,.01}$=2.4377.

Step 3. Using Gibbons' formula above, calculate the upper Prediction limit as:

$$T_k^* = \frac{1}{9}(70) + \frac{(2.4377)^2}{2(9)} + \frac{2.4377}{9}\sqrt{70(1+9) + \frac{(2.4377)^2}{4}} = 15.3\text{ppb}$$

Step 4. To test the upper Prediction limit, the Poisson count of the sum of the next four downgradient wells should be calculated. If this sum is greater than 15.3 ppb, there is significant evidence of contamination at the downgradient well. If not, the well may be regarded as clean until the next testing period.

The procedure for generating Poisson prediction limits is somewhat flexible. The value k above, for instance, need not represent multiple samples from a single well. It could also denote a collection of single samples from k distinct wells, all of which are assumed to follow the same Poisson distribution in the absence of contamination. The Poisson distribution also has the desirable property that the sum of several Poisson variables also has a Poisson distribution, even if the individual components are not identically distributed. Because of this, Gibbons (1987b) has suggested that if several analytes (e.g., different VOCs) can all be modeled via the Poisson distribution, the combined sum of the Poisson counts of all the analytes will also have a Poisson distribution, meaning that a single prediction limit could be estimated for the combined group of analytes, thus reducing the necessary number of statistical tests.

A major drawback to Gibbons' proposal of establishing a combined prediction limit for several analytes is that if the limit is exceeded, it will not be clear which analyte is responsible for "triggering" the test. In part this problem explains why the ground-water monitoring regulations mandate that each analyte be tested separately. Still, if a large number of analytes must be regularly tested and the detection rate is quite low, the overall facility-wide false positive rate may be unacceptably high. To remedy this situation, it is probably wisest to do enough initial testing of background and facility leachate and waste samples to determine those specific parameters present at levels substantially greater than background. By limiting monitoring and statistical tests to a few parameters meeting the above conditions, it should be possible to contain the overall facility-wide false positive rate while satisfying the regulatory requirements and assuring reliable identification of ground-water contamination if it occurs.

Though quantitative information on a suite of VOCs may be automatically generated as a consequence of the analytical method configuration (e.g., SW-846 method 8260 can provide quantitative results for approximately 60 different compounds), it is usually unnecessary to designate all of these compounds as leak detection indicators. Such practice generally aggravates the problem of many comparisons and results in elevated false positive rates for the facility as a whole. This makes accurate statistical testing especially difficult. EPA therefore recommends that the results of leachate testing or the waste analysis plan serve as the primary basis for designating reliable leak detection indicator parameters.

2.2.5 Poisson Tolerance Limits

To apply an upper Tolerance limit using the Poisson model to a group of downgradient wells, the approach described by Gibbons (1987b) and based on the work of Zacks (1970) can be taken. In this case, if no contamination has occurred, the estimated interval upper limit will contain a large fraction of all measurements from the downgradient wells, often specified at 95% or more.

The calculations involved in deriving Poisson Tolerance limits can seem non-intuitive, primarily because the argument leading to a mathematically rigorous Tolerance limit is complicated. The basic idea, however, uses the fact that if each individual measurement follows a common Poisson distribution with rate parameter, λ, the sum of n such measurements will also follow a Poisson distribution, this time with rate $n\lambda$.

Because the Poisson distribution has the property that its true mean is equal to the rate parameter λ, the concentration sum of n background samples can be manipulated to estimate this rate. But since we know that the distribution of the concentration sum is also Poisson, the possible values of λ can actually be narrowed to within a small range with fixed confidence probability (γ).

For each "possible" value of λ in this confidence range, one can compute the percentile of the Poisson distribution with rate λ that would lie above, say, 95% of all future downgradient measurements. By setting as the "probable" rate, that λ which is greater than all but a small percentage α of the most extreme possible λ's, given the values of n background samples, one can compute an upper tolerance limit with, say, 95% coverage and $(1-\alpha)\%$ confidence.

To actually make these computations, Zacks (1970) shows that the most probable rate λ can be calculated approximately as

$$\lambda_{T_n} = \frac{1}{2n}\chi^2_{\gamma}[2T_n + 2]$$

where as before T_n represents the Poisson count of the sum of n background samples (setting nondetects to half the method detection limit), and

$$\chi^2_{\gamma}[2T_n + 2]$$

represents the γ percentile of the Chi-square distribution with $(2T_n+2)$ degrees of freedom.

To find the upper Tolerance limit with $\beta\%$ coverage (e.g., 95%) once a probable rate λ has been estimated, one must compute the Poisson percentile that is larger than $\beta\%$ of all possible measurements from that distribution, that is, the $\beta\%$ quantile of the Poisson distribution with mean rate λ_{Tn}, denoted by $P^{-1}(\beta,\lambda_{Tn})$. Using a well-known mathematical relationship between the Poisson and Chi-square distributions, finding the $\beta\%$ quantile of the Poisson amounts to determining the least positive integer k such that

$$\chi^2_{1-\beta}[2k + 2] \geq 2\lambda_{T_n}$$

where, as above, the quantity [2k+2] represents the degrees of freedom of the Chi-square distribution. By calculating two times the estimated probable rate λ_{Tn} on the right-hand-side of the above inequality, and then finding the smallest degrees of freedom so that the $(1-\beta)\%$ percentile of the Chi-square distribution is bigger than $2\lambda_{Tn}$, the upper tolerance limit k can be determined fairly easily.

Once the upper tolerance limit, k, has been estimated, it will represent an upper Poisson Tolerance limit having approximately $\beta\%$ coverage with $\gamma\%$ confidence in all comparisons with downgradient well measurements.

EXAMPLE 11

Use the benzene data of Example 10 to estimate an upper Poisson Tolerance limit with 95% coverage and 95% confidence probability.

SOLUTION

Step 1. The benzene data consist of 33 nondetects with detection limit equal to 2 ppb and 3 detected values for a total of n=36. By setting each nondetect to half the detection limit as before, one finds a total Poisson count of the sum equal to T_n=70.0. It is also known that the desired confidence probability is γ=.95 and the desired coverage is β=.95.

Step 2. Based on the observed Poisson count of the sum of background samples, estimate the probable occurrence rate λ_{T_n} using Zacks' formula above as

$$\lambda_{T_n} = \frac{1}{2n}\chi_{\gamma}^{2}[2T_n + 2] = \frac{1}{72}\chi_{.95}^{2}[142] = 2.37$$

Step 3. Compute twice the probable occurrence rate as $2\lambda_{T_n}$=4.74. Now using a Chi-square table, find the smallest degrees of freedom (df), k, such that

$$\chi_{.05}^{2}[2k + 2] \geq 4.74$$

Since the 5th percentile of the Chi-square distribution with 12 df equals 5.23 (but only 4.57 with 11 df), it is seen that (2k+2)=12, leading to k=5. Therefore, the upper Poisson Tolerance limit is estimated as k=5 ppb.

Step 4. Because the estimated upper Tolerance limit with 95% coverage equals 5 ppb, any detected value among downgradient samples greater than 5 ppb may indicate possible evidence of contamination.

3. NON-PARAMETRIC COMPARISON OF COMPLIANCE WELL DATA TO BACKGROUND

When concentration data from several compliance wells are to be compared with concentration data from background wells, one basic approach is analysis of variance (ANOVA). The ANOVA technique is used to test whether there is statistically significant evidence that the mean concentration of a constituent is higher in one or more of the compliance wells than the baseline provided by background wells. Parametric ANOVA methods make two key assumptions: 1) that the data residuals are Normally distributed and 2) that the group variances are all approximately equal. The steps for calculating a parametric ANOVA are given in the Interim Final Guidance (pp. 5-6 to 5-14).

If either of the two assumptions crucial to a parametric ANOVA is grossly violated, it is recommended that a non-parametric test be conducted using the ranks of the observations rather than the original observations themselves. The Interim Final Guidance describes the Kruskal-Wallis test when three or more well groups (including background data, see pp. 5-14 to 5-20) are being compared. However, the Kruskal-Wallis test is not amenable to two-group comparisons, say of one compliance well to background data. In this case, the Wilcoxon Rank-Sum procedure (also known as the Mann-Whitney U Test) is recommended and explained below. Since most situations will involve the comparison of at least two downgradient wells with background data, the Kruskal-Wallis test is presented first with an additional example.

3.1 KRUSKAL-WALLIS TEST

When the assumptions used in a parametric analysis of variance cannot be verified, e.g., when the original or transformed residuals are not approximately Normal in distribution or have significantly different group variances, an analysis can be performed using the ranks of the observations. Usually, a non-parametric procedure will be needed when a substantial fraction of the measurements are below detection (more than 15 percent), since then the above assumptions are difficult to verify.

The assumption of independence of the residuals is still required. Under the null hypothesis that there is no difference among the groups, the observations are assumed to come from identical distributions. However, the form of the distribution need not be specified.

A non-parametric ANOVA can be used in any situation that the parametric analysis of variance can be used. However, because the ranks of the data are being used, the minimum sample sizes for the groups must be a little larger. A useful rule of thumb is to require a minimum of three well groups with at least four observations per group before using the Kruskal-Wallis procedure.

Non-parametric procedures typically need a few more observations than parametric procedures for two reasons. On the one hand, non-parametric tests make fewer assumptions concerning the distribution of the data and so more data is often needed to make the same judgment that would be rendered by a parametric test. Also, procedures based on ranks have a discrete distribution (unlike the continuous distributions of parametric tests). Consequently, a larger sample size is usually needed to produce test statistics that will be significant at a specified alpha level such as 5 percent.

The relative efficiency of two procedures is defined as the ratio of the sample sizes needed by each to achieve a certain level of power against a specified alternative hypothesis. As sample sizes get larger, the efficiency of the Kruskal-Wallis test relative to the parametric analysis of variance test approaches a limit that depends on the underlying distribution of the data, but is always at least 86 percent. This means roughly that in the worst case, if 86 measurements are available for a parametric ANOVA, only 100 sample values are needed to have an equivalently powerful Kruskal-Wallis test. In many cases, the increase in sample size necessary to match the power of a parametric ANOVA is much smaller or not needed at all. The efficiency of the Kruskal-Wallis test is 95 percent if the data are really Normal, and can be much larger than 100 percent in other cases (e.g., it is 150 percent if the residuals follow a distribution called the double exponential).

These results concerning efficiency imply that the Kruskal-Wallis test is reasonably powerful for detecting concentration differences despite the fact that the original data have been replaced by their ranks, and can be used even when the data are Normally distributed. When the data are not Normal or cannot be transformed to Normality, the Kruskal-Wallis procedure tends to be more powerful for detecting differences than the usual parametric approach.

3.1.1 Adjusting for Tied Observations

Frequently, the Kruskal-Wallis procedure will be used when the data contain a significant fraction of nondetects (e.g., more than 15 percent of the samples). In these cases, the parametric assumptions necessary for the usual one-way ANOVA are difficult or impossible to verify, making the non-parametric alternative attractive. However, the presence of nondetects prevents a unique ranking of the concentration values, since nondetects are, up to the limit of measurement, all tied at the same value.

To get around this problem, two steps are necessary. First, in the presence of ties (e.g., nondetects), all tied observations should receive the same rank. This rank (sometimes called the midrank (Lehmann, 1975)) is computed as the average of the ranks that would be given to a group of ties if the tied values actually differed by a tiny amount and could be ranked uniquely. For example, if the first four ordered observations are all nondetects, the midrank given to each of these samples would be equal to (1+2+3+4)/4=2.5. If the next highest measurement is a unique detect, its rank would be 5 and so on until all observations are appropriately ranked.

The second step is to compute the Kruskal-Wallis statistic as described in the Interim Final Guidance, using the midranks computed for the tied values. Then an adjustment to the Kruskal-Wallis statistic must be made to account for the presence of ties. This adjustment is described on page 5-17 of the Interim Final Guidance and requires computation of the formula:

$$H' = \frac{H}{1 - \left(\sum_{i=1}^{g} \frac{t_i^3 - t_i}{N^3 - N} \right)}$$

where g equals the number of groups of distinct tied observations and t_i is the number of observations in the ith tied group.

EXAMPLE 12

Use the non-parametric analysis of variance on the following data to determine whether there is evidence of contamination at the monitoring site.

	Toluene Concentration (ppb)				
	Background Wells		Compliance Wells		
Month	Well 1	Well 2	Well 3	Well 4	Well 5
1	<5	<5	<5	<5	<5
2	7.5	<5	12.5	13.7	20.1
3	<5	<5	8.0	15.3	35.0
4	<5	<5	<5	20.2	28.2
5	6.4	<5	11.2	25.1	19.0

SOLUTION

Step 1. Compute the overall percentage of nondetects. In this case, nondetects account for 48 percent of the data. The usual parametric analysis of variance would be inappropriate. Use the Kruskal-Wallis test instead, pooling both background wells into one group and treating each compliance well as a separate group.

Step 2. Compute ranks for all the data including tied observations (e.g., nondetects) as in the following table. Note that each nondetect is given the same midrank, equal to the average of the first 12 unique ranks.

	Toluene Ranks				
	Background Wells		Compliance Wells		
Month	Well 1	Well 2	Well 3	Well 4	Well 5
1	6.5	6.5	6.5	6.5	6.5
2	14	6.5	17	18	21
3	6.5	6.5	15	19	25
4	6.5	6.5	6.5	22	24
5	13	6.5	16	23	20
Rank Sum	$R_b=79$		$R_3=61$	$R_4=88.5$	$R_5=96.5$
Rank Mean	$\overline{R}_b=7.9$		$\overline{R}_3=12.2$	$\overline{R}_4=17.7$	$\overline{R}_5=19.3$

Step 3. Calculate the sums of the ranks in each group (R_i) and the mean ranks in each group ($\overline{R}_i$). These results are given above.

Step 4. Compute the Kruskal-Wallis statistic H using the formula on p. 5-15 of the Interim Final Guidance

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

where N=total number of samples, N_i=number of samples in ith group, and K=number of groups. In this case, N=25, K=4, and H can be computed as

$$H = \frac{12}{25 * 26}\left[\frac{79^2}{10} + \frac{61^2}{5} + \frac{88.5^2}{5} + \frac{96.5^2}{5}\right] - 78 = 10.56.$$

Step 5. Compute the adjustment for ties. There is only one group of distinct tied observations, containing 12 samples. Thus, the adjusted Kruskal-Wallis statistic is given by:

$$H' = \frac{10.56}{1 - \left(\frac{12^3 - 12}{25^3 - 25} \right)} = 11.87.$$

Step 6. Compare the calculated value of H′ to the tabulated Chi-square value with (K-1)= (# groups-1)=3 df, $\chi^2_{3,.05}$=7.81. Since the observed value of 11.87 is greater than the Chi-square critical value, there is evidence of significant differences between the well groups. Post-hoc pairwise comparisons are necessary.

Step 7. Calculate the critical difference for compliance well comparisons to the background using the formula on p. 5-16 of the Interim Final Guidance document. Since the number of samples at each compliance well is four, the same critical difference can be used for each comparison, namely,

$$C_1 = z_{.05/3} \sqrt{\frac{25 \cdot 26}{12} \left(\frac{1}{10} + \frac{1}{5} \right)} = 8.58$$

Step 8. Form the differences between the average ranks of each compliance well and the background and compare these differences to the critical value of 8.58.

Well 3: $\overline{R}_3 - \overline{R}_b = 12.2 - 7.9 = 4.3$

Well 4: $\overline{R}_4 - \overline{R}_b = 17.7 - 7.9 = 9.8$

Well 5: $\overline{R}_5 - \overline{R}_b = 19.3 - 7.9 = 11.4.$

Since the average rank differences at wells 4 and 5 exceed the critical difference, there is significant evidence of contamination at wells 4 and 5, but not at well 3.

3.2 WILCOXON RANK-SUM TEST FOR TWO GROUPS

When a single compliance well group is being compared to background data and a nonparametric test is needed, the Kruskal-Wallis procedure should be replaced by the Wilcoxon Rank-Sum test (Lehmann, 1975; also known as the two-sample Mann-Whitney U test). For most ground-water applications, the Wilcoxon test should be used whenever the proportion of nondetects in the combined data set exceeds 15 percent. However, to provide valid results, do not use the Wilcoxon test unless the compliance well and background data groups both contain at least four samples each.

To run the Wilcoxon Rank-Sum Test, use the following algorithm. Combine the compliance and background data and rank the ordered values from 1 to N. Assume there are n compliance samples and m background samples so that N=m+n. Denote the ranks of the compliance samples by C_i and the ranks of the background samples by B_j. Then add up the ranks of the compliance samples and subtract n(n+1)/2 to get the Wilcoxon statistic W:

$$W = \Sigma_{i=1}^{n} C_i - \frac{1}{2} n(n+1).$$

The rationale of the Wilcoxon test is that if the ranks of the compliance data are quite large relative to the background ranks, then the hypothesis that the compliance and background values came from the same population should be rejected. Large values of the statistic W give evidence of contamination at the compliance well site.

To find the critical value of W, a Normal approximation to its distribution is used. The expected value and standard deviation of W under the null hypothesis of no contamination are given by the formulas

$$E(W) = \frac{1}{2} mn; \qquad SD(W) = \sqrt{\frac{1}{12} mn(N+1)}$$

An approximate Z-score for the Wilcoxon Rank-Sum Test then follows as:

$$Z \approx \frac{W - E(W) - \frac{1}{2}}{SD(W)}.$$

The factor of 1/2 in the numerator serves as a continuity correction since the discrete distribution of the statistic W is being approximated by the continuous Normal distribution.

Once an approximate Z-score has been computed, it may be compared to the upper 0.01 percentile of the standard Normal distribution, $z_{.01}=2.326$, in order to determine the statistical significance of the test. If the observed Z-score is greater than 2.326, the null hypothesis may be rejected at the 1 percent significance level, suggesting that there is significant evidence of contamination at the compliance well site.

EXAMPLE 13

The table below contains copper concentration data (ppb) found in water samples at a monitoring facility. Wells 1 and 2 are background wells and well 3 is a single compliance well suspected of contamination. Calculate the Wilcoxon Rank-Sum Test on these data.

	Copper Concentration (ppb)		
	Background		Compliance
Month	Well 1	Well 2	Well 3
1	4.2	5.2	9.4
2	5.8	6.4	10.9
3	11.3	11.2	14.5
4	7.0	11.5	16.1
5	7.3	10.1	21.5
6	8.2	9.7	17.6

SOLUTION

Step 1. Rank the N=18 observations from 1 to 18 (smallest to largest) as in the following table.

	Ranks of Copper Concentrations		
	Background		Compliance
Month	Well 1	Well 2	Well 3
1	1	2	8
2	3	4	11
3	13	12	15
4	5	14	16
5	6	10	18
6	7	9	17

Step 2. Compute the Wilcoxon statistic by adding up the compliance well ranks and subtracting n(n+1)/2, so that W=85-21=64.

Step 3. Compute the expected value and standard deviation of W.

$$E(W) = \frac{1}{2}mn = 36$$

$$SD(W) = \sqrt{\frac{1}{12}mn(N+1)} = \sqrt{114} = 10.677$$

Step 4. Form the approximate Z-score.

$$Z \approx \frac{W - E(W) - \frac{1}{2}}{SD(W)} = \frac{64 - 36 - 0.5}{10.677} = 2.576$$

Step 5. Compare the observed Z-score to the upper 0.01 percentile of the Normal distribution. Since $Z=2.576>2.326=z_{.01}$, there is significant evidence of contamination at the compliance well at the 1 percent significance level.

3.2.1 Handling Ties in the Wilcoxon Test

Tied observations in the Wilcoxon test are handled in similar fashion to the Kruskal-Wallis procedure. First, midranks are computed for all tied values. Then the Wilcoxon statistic is computed as before but with a slight difference. To form the approximate Z-score, an adjustment is made to the formula for the standard deviation of W in order to account for the groups of tied values. The necessary formula (Lehmann, 1975) is:

$$SD^{*}(W)=\sqrt{\frac{mn(N+1)}{12}\left(1-\Sigma_{i=1}^{g}\frac{t_i^3-t_i}{N^3-N}\right)}$$

where, as in the Kruskal-Wallis method, g equals the number of groups of distinct tied observations and t_i represents the number of tied values in the ith group.

4. STATISTICAL INTERVALS: CONFIDENCE, TOLERANCE, AND PREDICTION

Three types of statistical intervals are often constructed from data: Confidence intervals, Tolerance intervals, and Prediction intervals. Though often confused, the interpretations and uses of these intervals are quite distinct. The most common interval encountered in a course on statistics is a Confidence interval for some parameter of the distribution (e.g., the population mean). The interval is constructed from sample data and is thus a random quantity. This means that each set of sample data will generate a different Confidence interval, even though the algorithm for constructing the interval stays the same every time.

A Confidence interval is designed to contain the specified population parameter (usually the mean concentration of a well in ground-water monitoring) with a designated level of confidence or probability, denoted as $1-\alpha$. The interval will fail to include the true parameter in approximately α percent of the cases where such intervals are constructed.

The usual Confidence interval for the mean gives information about the average concentration level at a particular well or group of wells. It offers little information about the highest or most extreme sample concentrations one is likely to observe over time. Often, it is those extreme values one wants to monitor to be protective of human health and the environment. As such, a Confidence interval generally should be used only in two situations for ground-water data analysis: (1) when directly specified by the permit or (2) in compliance monitoring, when downgradient samples are being compared to a Ground-Water Protection Standard (GWPS) representing the average of onsite background data, as is sometimes the case with an Alternate Contaminant Level (ACL). In other situations it is usually desirable to employ a Tolerance or Prediction interval.

A Tolerance interval is designed to contain a designated proportion of the population (e.g., 95 percent of all possible sample measurements). Since the interval is constructed from sample data, it also is a random interval. And because of sampling fluctuations, a Tolerance interval can contain the specified proportion of the population only with a certain confidence level. Two coefficients are associated with any Tolerance interval. One is the proportion of the population that the interval is supposed to contain, called the coverage. The second is the degree of confidence with which the interval reaches the specified coverage. This is known as the tolerance coefficient. A Tolerance interval with coverage of 95 percent and a tolerance coefficient of 95 percent is constructed to contain, on average, 95 percent of the distribution with a probability of 95 percent.

Tolerance intervals are very useful for ground-water data analysis, because in many situations one wants to ensure that at most a small fraction of the compliance well sample measurements exceed a specific concentration level (chosen to be protective of human health and the environment). Since a Tolerance interval is designed to cover all but a small percentage of the population measurements, observations should very rarely exceed the upper Tolerance limit when testing small sample sizes. The upper Tolerance limit allows one to gauge whether or not too many extreme concentration measurements are being sampled from compliance point wells.

Tolerance intervals can be used in detection monitoring when comparing compliance data to background values. They also should be used in compliance monitoring when comparing compliance data to certain Ground-Water Protection Standards. Specifically, the tolerance interval approach is recommended for comparison with a Maximum Contaminant Level (MCL) or with an ACL if the ACL is derived from health-based risk data.

Prediction intervals are constructed to contain the next sample value(s) from a population or distribution with a specified probability. That is, after sampling a background well for some time and measuring the concentration of an analyte, the data can be used to construct an interval that will contain the next analyte sample or samples (assuming the distribution has not changed). A Prediction interval will thus contain a future value or values with specified probability. Prediction intervals can also be constructed to contain the average of several future observations.

Prediction intervals are probably most useful for two kinds of detection monitoring. The first kind is when compliance point well data are being compared to background values. In this case the Prediction interval is constructed from the background data and the compliance well data are compared to the upper Prediction limits. The second kind is when intrawell comparisons are being made on an uncontaminated well. In this case, the Prediction interval is constructed on past data sampled from the well, and used to predict the behavior of future samples from the same well.

In summary, a Confidence interval usually contains an average value, a Tolerance interval contains a proportion of the population, and a Prediction interval contains one or more future observations. Each has a probability statement or "confidence coefficient" associated with it. For further explanation of the differences between these interval types, see Hahn (1970).

One should note that all of these intervals assume that the sample data used to construct the intervals are Normally distributed. In light of the fact that much ground-water concentration data is better modeled by a Lognormal distribution, it is recommended that tests for Normality be run on the logarithms of the original data before constructing the random intervals. If the data follow the Lognormal model, then the intervals should be constructed using the logarithms of the sample values. In this case, the limits of these intervals should not be compared to the original compliance data or GWPS. Rather, the comparison should involve the logged compliance data or logged GWPS. When neither the Normal or Lognormal models can be justified, a non-parametric version of each interval may be utilized.

4.1 TOLERANCE INTERVALS

In detection monitoring, the compliance point samples are assumed to come from the same distribution as the background values until significant evidence of contamination can be shown. To test this hypothesis, a 95 percent coverage Tolerance interval can be constructed on the background data. The background data should first be tested to check the distributional assumptions. Once the interval is constructed, each compliance sample is compared to the upper Tolerance limit. If any compliance point sample exceeds the limit, the well from which it was drawn is judged to have significant evidence of contamination (note that when testing a large number of samples, the nature of a Tolerance interval practically ensures that a few measurements will be above the upper Tolerance limit, even when no contamination has occurred. In these cases, the offending wells should probably be resampled in order to verify whether or not there is definite evidence of contamination.)

If the Tolerance limit has been constructed using the logged background data, the compliance point samples should first be logged before comparing with the upper Tolerance limit. The steps for computing the actual Tolerance interval in detection monitoring are detailed in the Interim Final Guidance on pp. 5-20 to 5-24. One point about the table of factors κ used to adjust the width of the Tolerance interval is that these factors are designed to provide at least 95% coverage of the population. Applied over many data sets, the average coverage of these intervals will often be close to 98% or more (see Guttman, 1970). To construct a one-sided upper Tolerance interval with average coverage of (1-β)%, the κ multiplier can be computed directly with the aid of a Student's t-distribution table. In this case, the formula becomes

$$\kappa = t_{n-1,1-\beta}\sqrt{1+\frac{1}{n}}$$

where the t-value represents the (1-β)th upper percentile of the t-distribution with (n-1) degrees of freedom.

In compliance monitoring, the Tolerance interval is calculated on the compliance point data, so that the upper one-sided Tolerance limit may be compared to the appropriate Ground-Water Protection Standard (i.e., MCL or ACL). If the upper Tolerance limit exceeds the fixed standard and especially if the Tolerance limit has been constructed to have an average coverage of 95% as described above, there is significant evidence that as much as 5 percent or more of all the compliance well measurements will exceed the limit and consequently that the compliance point wells are in violation of the facility permit. The algorithm for computing Tolerance limits in compliance monitoring is given on pp. 6-11 to 6-15 of the Interim Final Guidance.

EXAMPLE 14

The table below contains data that represent chrysene concentration levels (ppb) found in water samples obtained from the five compliance wells at a monitoring facility. Compute the upper Tolerance limit at each well for an average of 95% coverage with 95% confidence and determine whether there is evidence of contamination. The alternate concentration limit (ACL) is 80 ppb.

	Chrysene Concentration (ppb)				
Month	Well 1	Well 2	Well 3	Well 4	Well 5
1	19.7	10.2	68.0	26.8	47.0
2	39.2	7.2	48.9	17.7	30.5
3	7.8	16.1	30.1	31.9	15.0
4	12.8	5.7	38.1	22.2	23.4
Mean	19.88	9.80	46.28	24.65	28.98
SD	13.78	4.60	16.40	6.10	13.58

SOLUTION

Step 1. Before constructing the tolerance intervals, check the distributional assumptions. The algorithm for a parametric Tolerance interval assumes that the data used to compute the interval are Normally distributed. Because these data are more likely to be Lognormal in distribution than Normal, check the assumptions on the logarithms of the original data given in the table below. Since each well has only four observations, Probability Plots are not likely to be informative. The Shapiro-Wilk or Probability Plot Correlation Coefficient tests can be run, but in this example only the Skewness Coefficient is examined to ensure that gross departures from Lognormality are not missed.

Month	Well 1	Well 2	Well 3	Well 4	Well 5
	Logged Chrysene Concentration [log(ppb)]				
1	2.98	2.32	4.22	3.29	3.85
2	3.67	1.97	3.89	2.87	3.42
3	2.05	2.78	3.40	3.46	2.71
4	2.55	1.74	3.64	3.10	3.15
Mean	2.81	2.20	3.79	3.18	3.28
SD	0.68	0.45	0.35	0.25	0.48

Step 2. The Skewness Coefficients for each well are given in the following table. Since none of the coefficients is greater than 1 in absolute value, approximate Lognormality (that is, Normality of the logged data) is assumed for the purpose of constructing the tolerance intervals.

Well	Skewness	|Skewness|
1	.210	.210
2	.334	.334
3	.192	.192
4	-.145	.145
5	-.020	.020

Step 3. Compute the tolerance interval for each compliance well using the logged concentration data. The means and SDs are given in the second table above.

Step 4. The tolerance factor for a one-sided Normal tolerance interval with an <u>average</u> of 95% coverage with 95% probability and n=4 observations is given by

$$\kappa = t_{3,.05}\sqrt{1+\frac{1}{4}} = 2.631$$

The upper tolerance limit is calculated below for each of the five wells.

Well 1	2.81+2.631(0.68)=	4.61 log(ppb)
Well 2	2.20+2.631(0.45)=	3.38 log(ppb)
Well 3	3.79+2.631(0.35)=	4.71 log(ppb)
Well 4	3.18+2.631(0.25)=	3.85 log(ppb)
Well 5	3.28+2.631(0.48)=	4.54 log(ppb)

Step 5. Compare the upper tolerance limit for each well to the logarithm of the ACL, that is log(80)=4.38. Since the upper tolerance limits for wells 1, 3, and 5 exceed the logged ACL of 4.38 log(ppb), there is evidence of chrysene contamination in wells 1, 3, and 5.

4.1.1 Non-parametric Tolerance Intervals

When the assumptions of Normality and Lognormality cannot be justified, especially when a significant portion of the samples are nondetect, the use of non-parametric tolerance intervals should be considered. The upper Tolerance limit in a non-parametric setting is usually chosen as an order statistic of the sample data (see Guttman, 1970), commonly the maximum value or maybe the second largest value observed. As a consequence, non-parametric intervals should be constructed only from wells that are not contaminated. Because the maximum sample value is often taken as the upper Tolerance limit, non-parametric Tolerance intervals are very easy to construct and use. The sample data must be ordered, but no ranks need be assigned to the concentration values other than to determine the largest measurements. This also means that nondetects do not have to be uniquely ordered or handled in any special manner.

One advantage to using the maximum concentration instead of assigning ranks to the data is that non-parametric intervals (including Tolerance intervals) are sensitive to the actual magnitudes of the concentration data. Another plus is that unless all the sample data are nondetect, the maximum value will be a detected concentration, leading to a well-defined upper Tolerance limit.

Once an order statistic of the sample data (e.g., the maximum value) is chosen to represent the upper tolerance limit, Guttman (1970) has shown that the coverage of the interval, constructed repeatedly over many data sets, has a Beta probability density with cumulative distribution

$$I_t(n-m+1,m)=\int_0^t \frac{\Gamma(n+1)}{\Gamma(n-m+1)\Gamma(m)}u^{n-m}(1-u)^{m-1}du$$

where n=# samples in the data set and m=[(n+1)−(rank of upper tolerance limit value)]. If the maximum sample value is selected as the tolerance limit, its rank is equal to n and so m=1. If the second largest value is chosen as the limit, its rank would be equal to (n−1) and so m=2.

Since the Beta distribution is closely related to the more familiar Binomial distribution, Guttman has shown that in order to construct a non-parametric tolerance interval with at least β% coverage and (1-α) confidence probability, the number of (background) samples must be chosen such that

$$\sum_{t=m}^{n}\binom{n}{t}(1-\beta)^{t}\beta^{n-t} \geq 1-\alpha$$

Table A-6 in Appendix A provides the minimum coverage levels with 95% confidence for various choices of n, using either the maximum sample value or the second largest measurement as the tolerance limit. As an example, with 16 background measurements, the minimum coverage is β=83% if the maximum background value is designated as the upper Tolerance limit and β=74% if the Tolerance limit is taken to be the second largest background value. In general, Table A-6 illustrates that if the underlying distribution of concentration values is unknown, more background samples are needed compared to the parametric setting in order to construct a tolerance interval with sufficiently high coverage. Parametric tolerance intervals do not require as many background samples precisely because the form of the underlying distribution is assumed to be known.

Because the coverage of the above non-parametric Tolerance intervals follows a Beta distribution, it can also be shown that the average (not the minimum as discussed above) level of coverage is equal to 1−[m/(n+1)] (see Guttman, 1970). In particular, when the maximum sample value is chosen as the upper tolerance limit, m=1, and the expected coverage is equal to n/(n+1). This implies that at least 19 background samples are necessary to achieve 95% coverage on average.

EXAMPLE 15

Use the following copper background data to establish a non-parametric upper Tolerance limit and determine if either compliance well shows evidence of copper contamination.

	Copper Concentration (ppb)				
	Background Wells			Compliance Wells	
Month	Well 1	Well 2	Well 3	Well 4	Well 5
1	<5	9.2	<5		
2	<5	<5	5.4		
3	7.5	<5	6.7		
4	<5	6.1	<5		
5	<5	8.0	<5	6.2	<5
6	<5	5.9	<5	<5	<5
7	6.4	<5	<5	7.8	5.6
8	6.0	<5	<5	10.4	<5

SOLUTION

Step 1. Examine the background data in Wells 1, 2, and 3 to determine that the maximum observed value is 9.2 ppb. Set the 95% confidence upper Tolerance limit equal to this value. Because 24 background samples are available, Table A-6 indicates that the minimum coverage is equal to 88% (the expected average coverage, however, is equal to 24/25=96%). To increase the coverage level, more background samples would have to be collected.

Step 2. Compare each sample in compliance Wells 4 and 5 to the upper Tolerance limit. Since none of the measurements at Well 5 is above 9.2 ppb, while one sample from Well 4 is above the limit, conclude that there is significant evidence of copper contamination at Well 4 but not Well 5.

4.2 PREDICTION INTERVALS

When comparing background data to compliance point samples, a Prediction interval can be constructed on the background values. If the distributions of background and compliance point data are really the same, all the compliance point samples should be contained below the upper Prediction interval limit. Evidence of contamination is indicated if one or more of the compliance samples lies above the upper Prediction limit.

With intrawell comparisons, a Prediction interval can be computed on past data to contain a specified number of future observations from the same well, provided the well has not been previously contaminated. If any one or more of the future samples falls above the upper Prediction limit, there is evidence of recent contamination at the well. The steps to calculate parametric Prediction intervals are given on pp. 5-24 to 5-28 of the Interim Final Guidance.

EXAMPLE 16

The data in the table below are benzene concentrations measured at a groundwater monitoring facility. Calculate the Prediction interval and determine whether there is evidence of contamination.

Background Well Data		Compliance Well Data	
Sampling Date	Benzene Concentration (ppb)	Sampling Date	Benzene Concentration (ppb)
Month 1	12.6	Month 4	48.0
	30.8		30.3
	52.0		42.5
	28.1		15.0
Month 2	33.3		
	44.0		n=4
	3.0		Mean=33.95
	12.8		SD=14.64

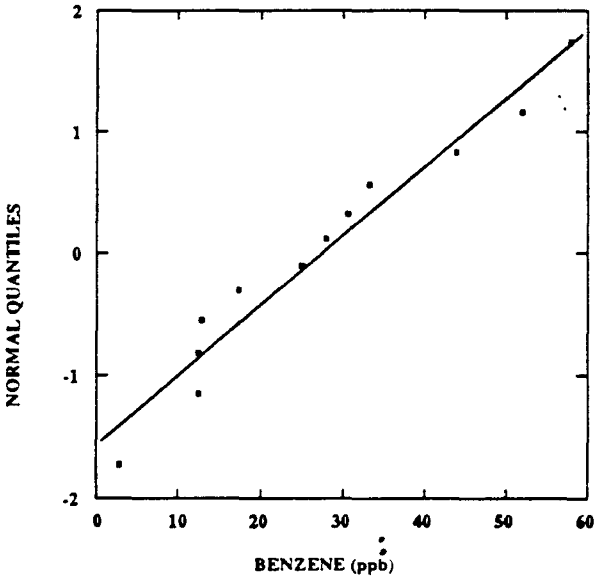
Month 3	58.1	Month 5	47.6
--	12.6		3.8
	17.6		2.6
	25.3		51.9
	n=12		n=4
	Mean=27.52		Mean=26.48
	SD=17.10		SD=26.94

SOLUTION

Step 1. First test the background data for approximate Normality. Only the background data are included since these values are used to construct the Prediction interval.

Step 2. A Probability Plot of the 12 background values is given below. The plot indicates an overall pattern that is reasonably linear with some modest departures from Normality. To further test the assumption of Normality, run the Shapiro-Wilk test on the background data.

PROBABILITY PLOT



Step 3. List the data in ascending and descending order as in the following table. Also calculate the differences $x_{(n-i+1)}-x_{(i)}$ and multiply by the coefficients a_{n-i+1} taken from Table A-1 to get the components of vector b_i used to calculate the Shapiro-Wilk statistic (W).

i	$x_{(i)}$	$x_{(n-i+1)}$	a_{n-i+1}	b_i
1	3.0	58.1	0.548	30.167
2	12.6	52.0	0.333	13.101
3	12.6	44.0	0.235	7.370
4	12.8	33.3	0.159	3.251
5	17.6	30.8	0.092	1.217
6	25.3	28.1	0.030	0.085
7	28.1	25.3		b=55.191
8	30.8	17.6		
9	33.3	12.8		
10	44.0	12.6		
11	52.0	12.6		
12	58.1	3.0		

Step 4. Sum the components b_i in column 5 to get quantity b. Compute the standard deviation of the background benzene values. Then the Shapiro-Wilk statistic is given as

$$W = \left[\frac{b}{SD\sqrt{n-1}}\right]^2 = \left[\frac{55.191}{17.101\sqrt{11}}\right]^2 = 0.947.$$

Step 5. The critical value at the 5% level for the Shapiro-Wilk test on 12 observations is 0.859. Since the calculated value of W=0.947 is well above the critical value, there is no evidence to reject the assumption of Normality.

Step 6. Compute the Prediction interval using the original background data. The mean and standard deviation of the 12 background samples are given by 27.52 ppb and 17.10 ppb, respectively.

Step 7. Since there are two future months of compliance data to be compared to the Prediction limit, the number of future sampling periods is k=2. At each sampling period, a mean of four independent samples will be computed, so m=4 in the prediction interval formula (see Interim Final Guidance, p. 5-25). The Bonferroni t-statistic, $t_{(11,2,.95)}$, with k=2 and 11 df is equivalent to the usual t-statistic at the .975 level with 11 df, i.e., $t_{11,.975}=2.201$.

Step 8. Compute the upper one-sided Prediction limit (UL) using the formula:

$$\overline{X} + t_{(n-1,k,.95)} S\sqrt{\frac{1}{m}+\frac{1}{n}}$$

Then the UL is given by:

$$UL = 27.52 + (17.10)(2.201)\sqrt{\frac{1}{4}+\frac{1}{12}} = 49.25\,\text{ppb}.$$

Step 9. Compare the UL to the compliance data. The means of the four compliance well observations for months 4 and 5 are 33.95 ppb and 26.48 ppb, respectively. Since the mean concentrations for months 4 and 5 are below the upper Prediction limit, there is no evidence of recent contamination at the monitoring facility.

4.2.1 Non-parametric Prediction Intervals

When the parametric assumptions of a Normal-based Prediction limit cannot be justified, often due to the presence of a significant fraction of nondetects, a non-parametric Prediction interval may be considered instead. A non-parametric upper Prediction limit is typically constructed in the same way as a non-parametric upper Tolerance limit, that is, by estimating the limit to be the maximum value of the set of background samples.

The difference between non-parametric Tolerance and Prediction limits is one of interpretation and probability. Given n background measurements and a desired confidence level, a non-parametric Tolerance interval will have a certain coverage percentage. With high probability, the Tolerance interval is designed to miss only a small percentage of the samples from downgradient wells. A Prediction limit, on the other hand, involves the confidence probability that the next future sample or samples will definitely fall below the upper Prediction limit. In this sense, the Prediction limit may be thought of as a 100% coverage Tolerance limit for the next k future samples.

As Guttman (1970) has indicated, the confidence probability associated with predicting that the next single observation from a downgradient well will fall below the upper Prediction limit -- estimated as the maximum background value -- is the <u>same</u> as the <u>expected coverage</u> of a similarly constructed upper Tolerance limit, namely $(1-\alpha)=n/(n+1)$. Furthermore, it can be shown from Gibbons (1991b) that the probability of having k future samples all fall below the upper non-parametric Prediction limit is $(1-\alpha)=n/(n+k)$. Table A-7 in Appendix A lists these confidence levels for various choices of n and k. The false positive rate associated with a single Prediction limit can be computed as one minus the confidence level.

Balancing the ease with which non-parametric upper Prediction limits are constructed is the fact that, given fixed numbers of background samples and future sample values to be predicted, the maximum confidence level associated with the Prediction limit is also fixed. To increase the level of confidence, the only choices are to 1) decrease the number of future values to be predicted at any testing period, or 2) increase the number of background samples used in the test. Table A-7 can be used along these lines to plan an appropriate sampling strategy so that the false positive rate can be minimized and the confidence probability maximized to a desired level.

EXAMPLE 17

Use the following arsenic data from a monitoring facility to compute a non-parametric upper Prediction limit that will contain the next 2 monthly measurements from a downgradient well and determine the level of confidence associated with the Prediction limit.

Arsenic Concentrations (ppb)				
	Background Wells			Compliance
Month	Well 1	Well 2	Well 3	Well 4
1	<5	7	<5	
2	<5	6.5	<5	
3	8	<5	10.5	
4	<5	6	<5	
5	9	12	<5	8
6	10	<5	9	14

SOLUTION

Step 1. Determine the maximum value of the background data and use this value to estimate the upper Prediction limit. In this case, the Prediction limit is set to the maximum value of the n=18 samples, or 12 ppb. As is true of non-parametric Tolerance intervals, only uncontaminated wells should be used in the construction of Prediction limits.

Step 2. Compute the confidence level and false positive rate associated with the Prediction limit. Since two future samples are being predicted and n=18, the confidence level is found to be n/(n+k)=18/20=90%. Consequently, the Type I error or false positive rate is equal to (1-.90)=10%. If a lower false positive rate is desired, the number of background samples used in the test must be enlarged.

Step 3. Compare each of the downgradient samples against the upper Prediction limit. Since the value of 14 ppb for month 2 exceeds the limit, conclude that there is significant evidence of contamination at the downgradient well at the 10% level of significance.

4.3 CONFIDENCE INTERVALS

Confidence intervals should only be constructed on data collected during compliance monitoring, in particular when the Ground-Water Protection Standard (GWPS) is an ACL computed from the average of background samples. Confidence limits for the average concentration levels at compliance wells should not be compared to MCLs. Unlike a Tolerance interval, Confidence limits for an average do not indicate how often individual samples will exceed the MCL. Conceivably, the lower Confidence limit for the mean concentration at a compliance well could fall below the MCL, yet 50 percent or more of the individual samples might exceed the

MCL. Since an MCL is designed to set an upper bound on the acceptable contamination, this would not be protective of human health or the environment.

When comparing individual compliance wells to an ACL derived from average background levels, a lower one-sided 99 percent Confidence limit should be constructed. If the lower Confidence limit exceeds the ACL, there is significant evidence that the true mean concentration at the compliance well exceeds the GWPS and that the facility permit has been violated. Again, in most cases, a Lognormal model will approximate the data better than a Normal distribution model. It is therefore recommended that the initial data checking and analysis be performed on the logarithms of the data. If a Confidence interval is constructed using logged concentration data, the lower Confidence limit should be compared to the logarithm of the ACL rather than the original GWPS. Steps for computing Confidence intervals are given on pp. 6-3 to 6-11 of the Interim Final Guidance.

5. STRATEGIES FOR MULTIPLE COMPARISONS

5.1 BACKGROUND OF PROBLEM

Multiple comparisons occur whenever more than one statistical test is performed during any given monitoring or evaluation period. These comparisons can arise as a result of the need to test multiple downgradient wells against a pool of upgradient background data or to test several indicator parameters for contamination on a regular basis. Usually the same statistical test is performed in every comparison, each test having a fixed level of confidence $(1-\alpha)$, and a corresponding false positive rate, α.

The false positive rate (or Type I error) for an individual comparison is the probability that the test will falsely indicate contamination, i.e., that the test will "trigger," though no contamination has occurred. If ground-water data measurements were always constant in the absence of contamination, false positives would never occur. But ground-water measurements typically vary, either due to natural variation in the levels of background concentrations or to variation in lab measurement and analysis.

Applying the same test to each comparison is acceptable if the number of comparisons is small, but when the number of comparisons is moderate to large the false positive rate associated with the testing network as a whole (that is, across all comparisons involving a separate statistical test) can be quite high. This means that if enough tests are run, there will be a significant chance that at least one test will indicate contamination, even if no actual contamination has occurred. As an example, if the testing network consists of 20 separate comparisons (some combination of multiple wells and/or indicator parameters) and a 99% confidence level Prediction interval limit is used on each comparison, one would expect an overall network-wide false positive rate of over 18%, even though the Type I error for any single comparison is only 1%. This means there is nearly 1 chance in 5 that one or more comparisons will falsely register potential contamination even if none has occurred. With 100 comparisons and the same testing procedure, the overall network-wide false positive rate jumps to more than 63%, adding additional expense to verify the lack of contamination at falsely triggered wells.

To lower the network-wide false positive rate, there are several important considerations. As noted in Section 2.2.4, only those constituents that have been shown to be reliable indicators of potential contamination should be statistically tested on a regular basis. By limiting the number of tested constituents to the most useful indicators, the overall number of statistical comparisons that must be made can be reduced, lowering the facility-wide false alarm rate. In addition, depending on the hydrogeology of the site, some indicator parameters may need to be tested only at one (or a few adjacent) regulated waste units, as opposed to testing across the entire facility, as long as the permit specifies a common point of compliance, thus further limiting the number of total statistical comparisons necessary.

One could also try to lower the Type I error applied to each individual comparison. Unfortunately, for a given statistical test in general, the lower the false positive rate, the lower the power of the test to detect real contamination at the well. If the statistical power drops too much, real contamination will not be identified when it occurs, creating a situation not protective of the environment or human health. Instead, alternative testing strategies can be considered that specifically account for the number of statistical comparisons being made during any evaluation period. All alternative testing strategies should be evaluated in light of two basic goals:

1. Is the network-wide false positive rate (across all constituents and wells being tested) acceptably low? and
2. Does the testing strategy have adequate statistical power to detect real contamination when it occurs?

To establish a standard recommendation for the network-wide overall false positive rate, it should be noted that for some statistical procedures, EPA specifications mandate that the Type I error for any individual comparison be at least 1%. The rationale for this minimum requirement is motivated by statistical power. For a given test, if the Type I error is set too low, the power of the test will dip below "acceptable" levels. EPA was not able to specify a minimum level of acceptable power within the regulations because to do so would require specification of a minimum difference of environmental concern between the null and alternative hypotheses. Limited current knowledge about the health and/or environmental effects associated with incremental changes in concentration levels of Appendix IX constituents greatly complicates this task. Therefore, minimum false positive rates were adopted for some statistical procedures until more specific guidance could be recommended. EPA's main objective, however, as in the past, is to approve tests that have adequate statistical power to detect real contamination of ground water, and not to enforce minimum false positive rates.

This emphasis is evident in §264.98(g)(6) for detection monitoring and §264.99(i) for compliance monitoring. Both of these provisions allow the owner or operator to demonstrate that the statistically significant difference between background and compliance point wells or between compliance point wells and the Ground-Water Protection Standard is an artifact caused by an error in sampling, analysis, statistical evaluation, or natural variation in ground-water chemistry. To make the demonstration that the statistically significant difference was caused by an error in sampling, analysis, or statistical evaluation, re-testing procedures that have been approved by the Regional Administrator can be written into the facility permit, provided their statistical power is comparable to the EPA Reference Power Curve given below.

For large monitoring networks, it is almost impossible to maintain a low network-wide overall false positive rate if the Type I errors for individual comparisons must be kept above 1%. As will be seen, some alternative testing strategies can achieve a low network-wide false positive rate while maintaining adequate power to detect contamination. EPA therefore recommends hat instead of the 1% criterion for individual comparisons, the overall network-wide false positive rate (across all wells and constituents) of any alternative testing strategy should be kept to approximately 5% for each monitoring or evaluation period, while maintaining statistical power comparable to the procedure below.

The other goal of any testing strategy should be to maintain adequate statistical power for detecting contamination. Technically, power refers to the probability that a statistical testing procedure will register and identify evidence of contamination when it exists. However, power is typically defined with respect to a single comparison, not a network of comparisons. Since some testing procedures may identify contamination more readily when several wells in the network are contaminated as opposed to just one or two, it is suggested that all testing strategies be compared on the following more stringent, but common, basis. Let the effective power of a testing procedure be defined as the probability of detecting contamination in the monitoring network when one and only one well is contaminated with a single constituent. Note that the effective power is a conservative measure of how a testing regimen will perform over the network, because the test must uncover one contaminated well among many clean ones (i.e., like "finding a needle in a haystack").

To establish a recommended standard for the statistical power of a testing strategy, it must be understood that the power is not single number, but rather a function of the level of contamination actually present. For most tests, the higher the level of contamination, the higher the statistical power; likewise, the lower the contamination level, the lower the power. As such, when increasingly contaminated ground water passes a particular well, it becomes easier for the statistical test to distinguish background levels from the contaminated ground water; consequently, the power is an increasing function of the contamination level.

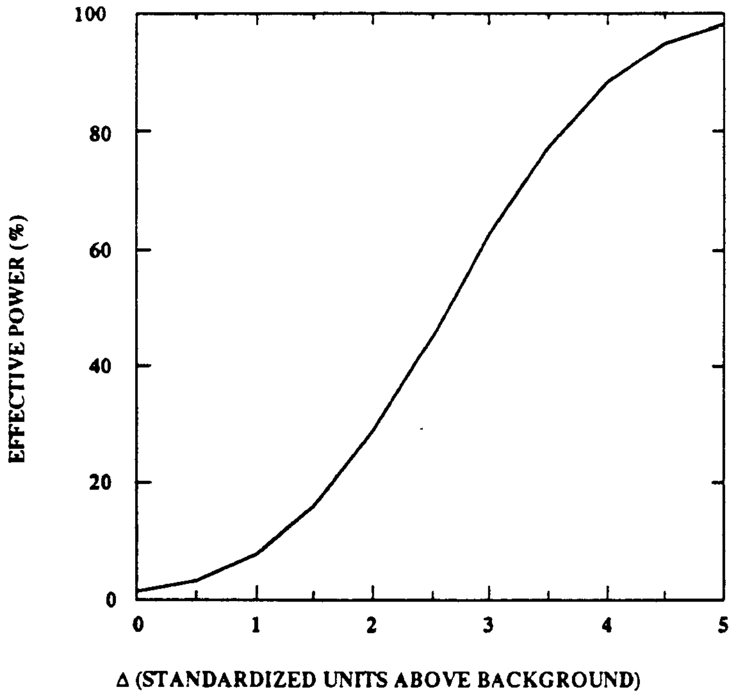
Perhaps the best way to describe the power function associated with a particular testing procedure is via a graph, such as the example below of the power of a standard Normal-based upper Prediction limit with 99% confidence. The power in percent is plotted along the y-axis against the standardized mean level of contamination along the x-axis. The standardized contamination levels are in units of standard deviations above the baseline (estimated from background data), allowing different power curves to be compared across indicator parameters, wells, and so forth. The standardized units, Δ, may be computed as

$$\Delta = \frac{\text{(Mean Contamination Level)} - \text{(Mean Background Level)}}{\text{(SD of Background Data)}}$$

In some situations, the probability that contamination will be detected by a particular testing procedure may be difficult if not impossible to derive analytically and will have to be simulated on a computer. In these cases, the power is typically estimated by generating Normally-distributed random values at different mean levels and repeatedly simulating the test procedure. With enough repetitions a reliable power curve can be plotted (e.g., see figure below).

EPA REFERENCE POWER CURVE

(16 Background Samples)



Notice that the power at $\Delta=0$ represents the false positive rate of the test, because at that point no contamination is actually present and the curve is indicating how often contamination will be "detected" anyway. As long as the power at $\Delta=0$ is approximately 5% (except for tests on an individual constituent at an individual well where the false positive rate should approximate 1%) and the rest of the power curve is acceptably high, the testing strategy should be adequately comparable to EPA standards.

To determine an acceptable power curve for comparison to alternative testing strategies, the following EPA Reference Power Curve is suggested. For a given and fixed number of background measurements, and based on Normally-distributed data from a single downgradient well generated at various mean levels above background, the EPA Reference Power Curve will represent the power associated with a 99% confidence upper prediction limit on the next single future sample from the well (see figure above for n=16).

Since the power of a test depends on several factors, including the background sample size, the type of test, and the number of comparisons, a different EPA Reference Power Curve will be associated with each distinct number of background samples. Power curves of alternative tests should only be compared to the EPA Reference Power Curve using a comparable number of background measurements. If the power of the alternative test is at least as high as the EPA reference, while maintaining an approximate 5% overall false positive rate, the alternative procedure should be acceptable.

With respect to power curves, keep in mind three important considerations: 1) the power of any testing method can be increased merely by relaxing the false positive rate requirement, letting α become larger than 5%. This is why an approximate 5% alpha level is suggested as the standard guidance, to ensure fair power comparisons among competing tests and to limit the overall network-wide false positive rate. 2) The simulation of alternative testing methods should incorporate every aspect of the procedure, from initial screens of the data to final decisions concerning the presence of contamination. This is especially applicable to strategies that involve some form of retesting at potentially contaminated wells. 3) When the testing strategy incorporates multiple comparisons, it is crucial that the power be gauged by simulating contamination in one and only one indicator parameter at a single well (i.e., by measuring the effective power). As noted earlier, EPA recommends that power be defined conservatively, forcing any test procedure to find "the needle in the haystack."

5.2 POSSIBLE STRATEGIES

5.2.1 Parametric and Non-parametric ANOVA

As described in the Interim Final Guidance, ANOVA procedures (either the parametric method or the Kruskal-Wallis test) allow multiple downgradient wells (but not multiple constituents) to be combined into a single statistical test, thus enabling the network-wide false positive rate for any single constituent to be kept at 5% regardless of the size of the network. The ANOVA method also maintains decent power for detecting real contamination, though only for small to moderately-sized networks. In large networks, even the parametric ANOVA has a difficult time finding the "needle in a haystack." The reason for this is that the ANOVA F-test combines all downgradient wells simultaneously, so that "clean" wells are mixed together with the single contaminated well, potentially masking the test's ability to detect the source of contamination.

Because of these characteristics, the ANOVA procedure may have poorer power for detecting a narrow plume of contamination which affects only one or two wells in a much larger network (say 20 or more comparisons). Another drawback is that a significant ANOVA test result will not indicate which well or wells is potentially contaminated without further post-hoc testing. Furthermore, the power of the ANOVA procedure depends significantly on having at least 3 to 4 samples per well available for testing. Since the samples must be statistically independent, collection of 3 or more samples at a given well may necessitate a several-month wait if the natural ground-water velocity at that well is low. In this case, it may be tempting to look for other strategies (e.g., Tolerance or Prediction intervals) that allow statistical testing of each new ground water sample as it is collected and analyzed. Finally, since the simple one-way ANOVA procedure outlined in the Interim Final Guidance is not designed to test multiple constituents simultaneously, the overall false positive rate will be approximately 5% per constituent, leading to a potentially high overall network-wide false positive rate (across wells and constituents) if many constituents need to be tested.

5.2.2 Retesting with Parametric Intervals

One strategy alternative to ANOVA is a modification of approaches suggested by Gibbons (1991a) and Davis and McNichols (1987). The basic idea is to adopt a two-phase testing strategy. First, new samples from each well in the network are compared, for each designated constituent parameter, against an upper Tolerance limit with pre-specified average coverage (Note that the upper Tolerance limit will be different for each constituent). Since some constituents at some wells in a large network would be expected to fail the Tolerance limit even in the absence of contamination, each well that triggers the Tolerance limit is resampled and only those constituents that "triggered" the limit are retested via an upper Prediction limit (again differing by constituent). If one or more resamples fails the upper Prediction limit, the specific constituent at that well failing the test is deemed to have a concentration level significantly greater than background. The overall strategy is effective for large networks of comparisons (e.g., 100 or more comparisons), but also flexible enough to accommodate smaller networks.

To design and implement an appropriate pair of Tolerance and Prediction intervals, one must know the number of background samples available and the number of comparisons in the network. Since parametric intervals are used, it is assumed that the background data are either Normal or can be transformed to an approximate Normal distribution. The tricky part is to choose an average coverage for the Tolerance interval and confidence level for the Prediction interval such that the twin goals are met of keeping the overall false positive rate to approximately 5% and maintaining adequate statistical power.

To derive the overall false positive rate for this retesting strategy, assume that when no contamination is present each constituent and well in the network behaves independently of other constituents and wells. Then if A_i denotes the event that well i is triggered falsely at some stage of the testing, the overall false positive rate across m such comparisons can be written as

$$\text{total } \alpha = \Pr\{A_1 \text{ or } A_2 \text{ or } \ldots \text{ or } A_i \text{ or } \ldots \text{ or } A_m\} = 1 - \prod_{i=1}^{m} \Pr\{\overline{A}_i\}$$

where $\overline{A}_i$ denotes the complement of event A_i. Since $P\{\overline{A}_i\}$ is the probability of <u>not</u> registering a false trigger at uncontaminated well i, it may be written as

$$\Pr\{\overline{A}_i\} = \Pr\{X_i \le TL\} + \Pr\{X_i > TL\} \times \Pr\{Y_i \le PL \mid X_i > TL\}$$

where X_i represents the original sample at well i, Y_i represents the concentrations of one or more resamples at well i, TL and PL denote the upper Tolerance and Prediction limits respectively, and the right-most probability is the conditional event that all resample concentrations fall below the Prediction limit when the initial sample fails the Tolerance limit.

Letting $x=\Pr\{X_i \le TL\}$ and $y=\Pr\{Y_i \le PL \mid X_i > TL\}$, the overall false positive rate across m constituent-well combinations can be expressed as

$$\text{total } \alpha = 1 - [x + (1 - x) \cdot y]^m$$

As noted by Guttman (1970), the probability that any random sample will fall below the upper Tolerance limit (i.e., quantity x above) is equal to the expected or average coverage of the Tolerance interval. If the Tolerance interval has been constructed to have average coverage of 95%, x=0.95. Then given a predetermined value for x, a fixed number of comparisons m, and a desired overall false positive rate α, we can solve for the conditional probability y as follows:

$$y = \frac{\sqrt[m]{1-\alpha} - x}{1 - x}$$

If the conditional probability y were equal to the probability that the resample(s) for the ith constituent-well combination falls below the upper Prediction limit, one could fix α at, say, 5%, and construct the Prediction interval to have confidence level y. In that way, one could guarantee an expected network-wide false positive rate of 5%. Unfortunately, whether or not one or more resamples falls below the Prediction limit depends partly on whether the initial sample for that comparison eclipsed the Tolerance limit. This is because the same background data are used to construct both the Tolerance limit and the Prediction limit, creating a statistical dependence between the tests.

The exact relationship between the conditional probability y and the unconditional probability $\Pr\{Y_i \leq PL\}$ is not known; however, simulations of the testing strategy suggest that when the confidence level for the Prediction interval is equated to the above solution for y, the overall network-wide false positive rate turns out to be higher than 5%. How much higher depends on the number of background samples and also the number of downgradient comparisons. Even with a choice of y that guarantees an expected facility-wide false positive rate of 5%, the power characteristics of the resulting testing strategy are not necessarily equivalent to the EPA Reference Power Curve, again depending on the number of background samples and the number of monitoring well-constituent combinations in the network.

In practice, to meet the selection criteria of 1) establishing an overall false positive rate of approximately 5% and 2) maintaining adequate statistical power, the confidence level chosen for the upper Prediction limit should be somewhat higher than the solution y to the preceding equation. The table below provides recommended choices of expected coverage and confidence levels for the Tolerance interval-Prediction interval pair when using specific combinations of numbers of downgradient comparisons and background samples. In general, one should pick lower coverage

Tolerance limits for smaller networks and higher coverage Tolerance limits for larger networks. That way (as can be seen in the table), the resulting Prediction limit confidence levels will be low enough to allow the construction of Prediction limits with decent statistical power.

PARAMETRIC RETESTING STRATEGIES

# COMPARISONS	# BG SAMPLES	TOLERANCE COVERAGE (%)	PREDICTION LEVEL (%)	RATING
5	8	95	90	**
	16	95	90	**
	16	95	85	*
	24	95	85	**
	24	95	90	*
20	8	95	98	**
	16	95	97	**
	24	95	97	**
50	16	98	97	**
	16	99	92	*
	24	98	95	**
	24	99	90	**
100	16	98	98	*
	24	99	95	*
	24	98	98	*

Note: ** = strongly recommended
* = recommended

Only strategies that approximately met the selection criteria are listed in the table. It can be seen that some, but not all, of these strategies are strongly recommended. Those that are merely "recommended" failed in the simulations to fully meet one or both of the selection criteria. The performance of all the recommended strategies, however, should be adequate to correctly identify contamination while maintaining a modest facility-wide false positive rate.

Once a combination of coverage and confidence levels for the Tolerance-Prediction interval pair is selected, the statistical power of the testing strategy should be estimated in order to compare with the EPA Reference Power Curve (particularly if the testing scenario is different from those computed in this Addendum). Simulation results have suggested that the above method for choosing a two-phase testing regimen can offer statistical power comparable to the EPA Reference for almost any sized monitoring network (see power curves in Appendix B).

Several examples of simulated power curves are presented in Appendix B. The range of downgradient wells tested is from 5 to 100 (note that the number of wells could actually represent the number of constituent-well combinations if testing multiple parameters), and each curve is based on either 8, 16, or 24 background samples. The y-axis of each graph measures the effective power of the testing strategy, i.e., the probability that contamination is detected when one and only one constituent at a single well has a mean concentration higher than background level. For each case, the EPA Reference Power Curve is compared to two different two-phase testing strategies. In the first case, wells that trigger the initial Tolerance limit are resampled once. This single resample is compared to a Prediction limit for the next future sample. In the second case, wells that trigger the Tolerance limit are resampled twice. Both resamples are compared to an upper Prediction limit for the next two future samples at that well.

The simulated power curves suggest two points. First, with an appropriate choice of coverage and prediction levels, the two-phase retesting strategies have comparable power to the EPA Reference Power Curve, while maintaining low overall network-wide false positive rates. Second, the power of the retesting strategy is slightly improved by the addition of a second resample at wells that fail the initial Tolerance limit, because the sample size is increased.

Overall, the two-phase testing strategy defined above--i.e., first screening the network of wells with a single upper Tolerance limit, and then applying an upper Prediction limit to resamples from wells which fail the Tolerance interval--appears to meet EPA's objectives of maintaining adequate statistical power for detecting contamination while limiting network-wide false positive rates to low levels. Furthermore, since each compliance well is compared against the interval limits separately, a narrow plume of contamination can be identified more efficiently than with an ANOVA procedure (e.g., no post-hoc testing is necessary to finger the guilty wells, and the two-phase interval testing method has more power against the "needle-in-a-haystack" contamination hypothesis).

5.2.3 Retesting with Non-parametric Intervals

When parametric intervals are not appropriate for the data at hand, either due to a large fraction of nondetects or a lack of fit to Normality or Lognormality, a network of individual comparisons can be handled via retesting using non-parametric Prediction limits. The strategy is to establish a non-parametric prediction limit for each designated indicator parameter based on background samples that accounts for the number of well-constituent comparisons in the overall network.

In order to meet the twin goals of maintaining adequate statistical power and a low overall rate of false positives, a non-parametric strategy must involve some level of retesting at those wells which initially indicate possible contamination. Retesting can be accomplished by taking a specific number of additional, independent samples from each well in which a specific constituent triggers the initial test and then comparing these samples against the non-parametric prediction limit for that parameter.

Because more independent data is added to the overall testing procedure, retesting of additional samples, in general, enables one to make more powerful and more accurate determinations of possible contamination. Retesting does, however, involve a trade-off. Because the power of the test increases with the number of resamples, one must decide how quickly resamples can be collected to ensure 1) quick identification and confirmation of contamination and yet, 2) the statistical independence of successive resamples from any particular well. Do not forget that the performance of a non-parametric retesting strategy depends substantially on the independence of the data from each well.

Two basic approaches to non-parametric retesting have been suggested by Gibbons (1990 and 1991b). Both strategies define the upper Prediction limit for each designated parameter to be the maximum value of that constituent in the set of background data. Consequently, the background wells used to construct the limits must be uncontaminated. After the Prediction limits have been calculated, one sample is collected from each downgradient well in the network. If any sample constituent value is greater than its upper prediction limit, the initial test is "triggered" and one or more resamples must be collected at that downgradient well on the constituent for further testing.

At this point, the similarity between the two approaches ends. In his 1990 article, Gibbons computes the probability that at least one of m independent samples taken from each of k downgradient wells will be below (i.e., pass) the prediction limit. The m samples include both the initial sample and (m-1) resamples. Because retesting only occurs when the initial well sample fails the limit, a given well fails the overall test (initial comparison plus retests) only if all (m-1) resamples are above the prediction limit. If any resample passes the prediction limit, that well is regarded as showing no significant evidence of contamination.

Initially, this first strategy may not appear to be adequately sensitive to mild contamination at a given downgradient well. For example, suppose two resamples are to be collected whenever the initial sample fails the upper prediction limit. If the initial sample is above the background maximum and one of the resamples is also above the prediction limit, the well can still be classified as "clean" if the other resample is below the prediction limit. Statistical power simulations (see Appendix B), however, suggest that this strategy will perform adequately under a number of monitoring scenarios. Still, EPA recognizes that a retesting strategy which might classify a well as "clean" when the initial sample and a resample both fail the upper Prediction limit could offer problematic implications for permit writers and enforcement personnel.

A more stringent approach was suggested by Gibbons in 1991. In that article (1991b), Gibbons computes, as "passing behavior," the probability that all but one of m samples taken from each of k wells pass the upper prediction limit. Under this definition, if the initial sample fails the upper Prediction limit, all (m-1) resamples must pass the limit in order for well to be classified as "clean" during that testing period. Consequently, if any single resample falls above the background maximum, that well is judged as showing significant evidence of contamination.

Either non-parametric retesting approach offers the advantage of being extremely easy to implement in field testing of a large downgradient well network. In practice, one has only to determine the maximum background sample to establish the upper prediction limit against which all other comparisons are made. Gibbons' 1991 retesting scheme offers the additional advantage of requiring less overall sampling at a given well to establish significant evidence of contamination. Why? If the testing procedure calls for, say, two resamples at any well that fails the initial prediction limit screen, retesting can end whenever either one of the two resamples falls above the prediction limit. That is, the well will be designated as potentially contaminated if the first resample fails the prediction limit even if the second resample has not yet been collected.

In both of his papers, Gibbons offers tables that can be used to compute the overall network-wide false positive rate, given the number of background samples, the number of downgradient comparisons, and the number of retests for each comparison. It is clear that there is less flexibility in adjusting a non-parametric as opposed to a parametric prediction limit to achieve a certain Type I error rate. In fact, if only a certain number of retests are feasible at any given well (e.g., in order to maintain independence of successive samples), the only recourse to maintain a low false positive rate is to collect a larger number of background samples. In this way, the inability to make parametric assumptions about the data illustrates why non-parametric tests are on the whole less efficient and less powerful than their parametric counterparts.

Unfortunately, the power of these non-parametric retesting strategies is not explored in detail by Gibbons. To compare the power of both Gibbons' strategies against the EPA Reference Power

Curve, Normally distributed data were simulated for several combinations of numbers of background samples and downgradient wells (again, if multiple constituents are being tested, the number of wells in the simulations may be regarded as the number of constituent-well combinations). Up to three resamples were allowed in the simulations for comparative purposes. EPA recognizes, however, that it will be feasible in general to collect only one or two independent resamples from any given well. Power curves representing the results of these simulations are given in Appendix B. For each scenario, the EPA Reference Power Curve is compared with the simulated powers of six different testing strategies. These strategies include collection of no resamples, one resample, two resamples under Gibbons' 1990 approach (designated as A on the curves) and his 1991 approach (labelled as B), and three resamples (under approaches A and B). Under the one resample strategy, a potentially contaminated compliance well is designated as "clean" if the resample passes the retest and "contaminated" otherwise.

The following table lists the best-performing strategies under each scenario. As with the use of parametric intervals for retesting, the criteria for selecting the best-performing strategies required 1) an approximate 5% facility-wide false positive rate and 2) power equivalent to or better than the EPA Reference Power Curve. Because Normal data were used in these power simulations, more realistically skewed data would likely result in greater advantages for the non-parametric retesting strategies over the EPA Reference test.

Examination of the table and the power curves in Appendix B shows that the number of background samples has an important effect on the recommended testing strategy. For instance, with 8 background samples in a network of at least 20 wells, the best performing strategies all involve collection of 3 resamples per "triggered" compliance well (EPA regards such a strategy as impractical for permitting and enforcement purposes at most RCRA facilities). It tends to be true that as the number of available background samples grows, fewer resamples are needed from each potentially contaminated compliance well to maintain adequate power. If, as is expected, the number of feasible, independent retests is limited, a facility operator may have to collect additional background measurements in order to establish an adequate retesting strategy.

NON-PARAMETRIC RETESTING STRATEGIES

# WELLS	# BG SAMPLES	STRATEGY	REFERENCE	RATING
5	8	1 Resample		*
	8	2 Resamples (A)	Gibbons, 1990	**
	16	1 Resample		**
	16	2 Resamples (B)	Gibbons, 1991	**
	24	2 Resamples (B)	Gibbons, 1991	**
20	8	2 Resamples (A)	Gibbons, 1990	*
	16	1 Resample		*
	16	2 Resamples (A)	Gibbons, 1990	*
	24	1 Resample		**
	24	2 Resamples (B)	Gibbons, 1991	*
	32	1 Resample		*
	32	2 Resamples (B)	Gibbons, 1991	**
50	16	2 Resamples (A)	Gibbons, 1990	**
	24	1 Resample		*
	24	2 Resamples (A)	Gibbons, 1990	*
	32	1 Resample		**
100	16	2 Resamples (A)	Gibbons, 1990	**
	24	2 Resamples (A)	Gibbons, 1990	*
	32	1 Resample		*

Note: ** = very good performance * = good performance

6. OTHER TOPICS

6.1 CONTROL CHARTS

Control Charts are an alternative to Prediction limits for performing either intrawell comparisons or comparisons to historically monitored background wells during detection monitoring. Since the baseline parameters for a Control Chart are estimated from historical data, this method is only appropriate for initially uncontaminated compliance wells. The main advantage of a Control Chart over a Prediction limit is that a Control Chart allows data from a well to be viewed graphically over time. Trends and changes in the concentration levels can be seen easily, because all sample data is consecutively plotted on the chart as it is collected, giving the data analyst an historical overview of the pattern of contamination. Prediction limits allow only point-in-time comparisons between the most recent data and past information, making long-term trends difficult to identify.

More generally, intrawell comparison methods eliminate the need to worry about spatial variability between wells in different locations. Whenever background data is compared to compliance point measurements, there is a risk that any statistically significant difference in concentration levels is due to spatial and/or hydrogeological differences between the wells rather than contamination at the facility. Because intrawell comparisons involve but a single well, significant changes in the level of contamination cannot be attributed to spatial differences between wells, regardless of whether the method used is a Prediction limit or Control Chart.

Of course, past observations can be used as baseline data in an intrawell comparison only if the well is known to be uncontaminated. Otherwise, the comparison between baseline data and newly collected samples may negate the goal in detection monitoring of identifying evidence of contamination. Furthermore, without specialized modification, Control Charts do not efficiently handle truncated data sets (i.e., those with a significant fraction of nondetects), making them appropriate only for those constituents with a high frequency of occurrence in monitoring wells. Control Charts tend to be most useful, therefore, for inorganic parameters (e.g., some metals and geochemical monitoring parameters) that occur naturally in the ground water.

The steps to construct a Control Chart can be found on pp. 7-3 to 7-10 of the Interim Final Guidance. The way a Control Chart works is as follows. Initial sample data is collected (from the specific compliance well in an intrawell comparison or from background wells in comparisons of compliance data with background) in order to establish baseline parameters for the chart, specifically, estimates of the well mean and well variance. These samples are meant to characterize the concentration levels of the uncontaminated well, before the onset of detection monitoring. Since the estimate of well variance is particularly important, it is recommended that at least 8 samples be collected (say, over a year's time) to estimate the baseline parameters. Note that none of these 8 or more samples is actually plotted on the chart.

As future samples are collected, the baseline parameters are used to standardize the data. At each sampling period, a standardized mean is computed using the formula below, where m represents the baseline mean concentration and s represents the baseline standard deviation.

$$Z_i = \sqrt{n_i}(\bar{x} - m) / s$$

A cumulative sum (CUSUM) for the ith period is also computed, using the formula $S_i = \max\{0, (Z_i\text{-}k)+S_{i-1}\}$, where Z_i is the standardized mean for that period and k represents a pre-chosen Control Chart parameter.

Once the data have been standardized and plotted, a Control Chart is declared out-of-control if the sample concentrations become too large when compared to the baseline parameters. An out-of-control situation is indicated on the Control Chart when either the standardized means or CUSUMs cross one of two pre-determined threshold values. These thresholds are based on the rationale that if the well remains uncontaminated, new sample values standardized by the original baseline parameters should not deviate substantially from the baseline level. If contamination does occur, the old baseline parameters will no longer accurately represent concentration levels at the well and, hence, the standardized values should significantly deviate from the baseline levels on the Control Chart.

In the combined Shewhart-cumulative sum (CUSUM) Control Chart recommended by the Interim Final Guidance (Section 7), the chart is declared out-of-control in one of two ways. First, the standardized means (Z_i) computed at each sampling period may cross the Shewhart control limit (SCL). Such a change signifies a rapid increase in well concentration levels among the most recent sample data. Second, the cumulative sum (CUSUM) of the standardized means may become too large, crossing the "decision internal value" (h). Crossing the h threshold can mean either a sudden rise in concentration levels or a gradual increase over a longer span of time. A gradual increase or trend is particularly indicated if the CUSUM crosses its threshold but the standardized mean Z_i does not. The reason for this is that several consecutive small increases in Z_i will not trigger the SCL threshold, but may trigger the CUSUM threshold. As such, the Control Chart can indicate the onset of either sudden or gradual contamination at the compliance point.

As with other statistical methods, Control Charts are based on certain assumptions about the sample data. The first is that the data at an uncontaminated well (i.e., a well process that is "in control") are Normally distributed. Since estimates of the baseline parameters are made using initially collected data, these data should be tested for Normality using one of the goodness-of-fit techniques described earlier. Better yet, the logarithms of the data should be tested first, to see if a Lognormal model is appropriate for the concentration data. If the Lognormal model is not rejected, the Control Chart should be constructed solely on the basis of logged data.

The methodology for Control Charts also assumes that the sample data are independently distributed from a statistical standpoint. In fact, these charts can easily give misleading results if the consecutive sample data are not independent. For this reason, it is important to design a sampling plan so that distinct volumes of water are analyzed each sampling period and that duplicate sample analyses are not treated are independent observations when constructing the Control Chart.

The final assumption is that the baseline parameters at the well reflect current background concentration levels. Some long-term fluctuation in background levels may be possible even though contamination has not occurred at a given well. Because of this possibility, if a Control Chart remains "in control" for a long period of time, the baseline parameters should be updated to include more recent observations as background data. After all, the original baseline parameters will often be based only on the first year's data. Much better estimates of the true background mean and variance can be obtained by including more data at a later time.

To update older background data with more recent samples, a two-sample t-test can be run to compare the older concentration levels with the concentrations of the proposed update samples. If the t-test does not show a significant difference at the 5 percent significance level, proceed to re-estimate the baseline parameters by including more recent data. If the t-test does show a significant difference, the newer data should not be characterized as background unless some specific factor can be pinpointed explaining why background levels on the site have naturally changed.

EXAMPLE 18

Construct a control chart for the 8 months of data collected below.

μ=27 ppb
σ=25 ppb

	Nickel Concentration (ppb)	
Month	Sample 1	Sample 2
1	15.3	22.6
2	41.1	27.8
3	17.5	18.1
4	15.7	31.5
5	37.2	32.4
6	25.1	32.5
7	19.9	27.5
8	99.3	64.2

SOLUTION

Step 1. The three parameters necessary to construct a combined Shewhart-CUSUM chart are h=5, k=1, and SCL=4.5 in units of standard deviation (SD).

Step 2. List the sampling periods and monthly means, as in the following table.

Month	T_i	Mean (ppb)	Z_i	Z_i - k	S_i
1	1	19.0	-0.45	-1.45	0.00
2	2	34.5	0.42	-0.58	0.00
3	3	17.8	-0.52	-1.52	0.00
4	4	23.6	-0.19	-1.19	0.00
5	5	34.8	0.44	-0.56	0.00
6	6	28.8	0.10	-0.90	0.00
7	7	23.7	-0.19	-1.19	0.00
8	8	81.8	3.10	2.10	2.10

Step 3. Compute the standardized means Z_i and the quantities S_i. List in the table above. Each S_i is computed for consecutive months using the formula on p. 7-8 of the EPA guidance document.

$$S_1 = \max\{0, -1.45 + 0\} = 0.00$$

$$S_2 = \max\{0, -0.58 + 0\} = 0.00$$

$$S_3 = \max\{0, -1.52 + 0\} = 0.00$$

$$S_4 = \max\{0, -1.19 + 0\} = 0.00$$

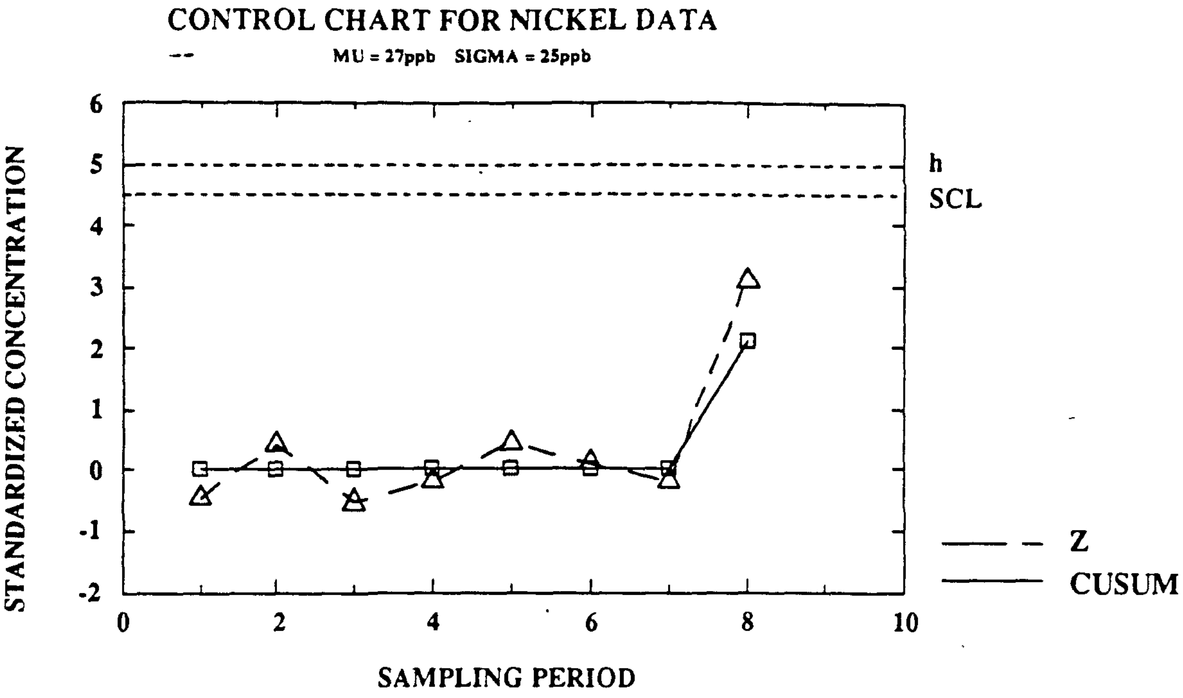
$$S_5 = \max\{0, -0.56 + 0\} = 0.00$$

$$S_6 = \max\{0, -0.90 + 0\} = 0.00$$

$$S_7 = \max\{0, -1.19 + 0\} = 0.00$$

$$S_8 = \max\{0, 2.10 + 0\} = 2.10$$

Step 4. Plot the control chart as given below. The combined chart indicates that there is no evidence of contamination at the monitoring facility because neither the standardized mean nor the CUSUM statistic exceeds the Shewhart control limits for the months examined.



Note: In the above Control Chart, the CUSUMs are compared to threshold h, while the standardized means (Z) are compared to the SCL threshold.

6.2 OUTLIER TESTING

Formal testing for outliers should be done only if an observation seems particularly high (by orders of magnitude) compared to the rest of the data set. If a sample value is suspect, one should run the outlier test described on pp. 8-11 to 8-14 of the EPA guidance document. It should be cautioned, however, that this outlier test assumes that the rest of the data values, except for the suspect observation, are Normally distributed (Barnett and Lewis, 1978). Since Lognormally distributed measurements often contain one or more values that appear high relative to the rest, it is recommended that the outlier test be run on the logarithms of the data instead of the original observations. That way, one can avoid classifying a high Lognormal measurement as an outlier just because the test assumptions were violated.

If the test designates an observation as a statistical outlier, the sample should not be treated as such until a specific reason for the abnormal measurement can be determined. Valid reasons may, for example, include contaminated sampling equipment, laboratory contamination of the sample, or errors in transcription of the data values. Once a specific reason is documented, the sample should be excluded from any further statistical analysis. If a plausible reason cannot be found, the sample should be treated as a true but extreme value, not to be excluded from further analysis.

EXAMPLE 19

The table below contains data from five wells measured over a 4-month period. The value 7066 is found in the second month at well 3. Determine whether there is statistical evidence that this observation is an outlier.

Carbon Tetrachloride Concentration (ppb)				
Well 1	Well 2	Well 3	Well 4	Well 5
1.69	302	16.2	199	275
3.25	35.1	7066	41.6	6.5
7.3	15.6	350	75.4	59.7
12.1	13.7	70.14	57.9	68.4

SOLUTION

Step 1. Take logarithms of each observation. Then order and list the logged concentrations.

Order	Concentration (ppb)	Logged Concentration
1	1.69	0.525
2	3.25	1.179
3	6.5	1.872
4	7.3	1.988
5	12.1	2.493
6	13.7	2.617
7	15.6	2.747
8	16.2	2.785
9	35.1	3.558
10	41.6	3.728
11	57.9	4.059
12	59.7	4.089
13	68.4	4.225
14	70.1	4.250
15	75.4	4.323
16	199	5.293
17	275	5.617
18	302	5.710
19	350	5.878
20	7066	8.863

Step 2. Calculate the mean and SD of all the logged measurements. In this case, the mean and SD are 3.789 and 1.916, respectively.

Step 3. Calculate the outlier test statistic T_{20} as

$$T_{20} = \frac{X_{(20)} - \overline{X}}{SD} = \frac{8.863 - 3.789}{1.916} = 2.648.$$

Step 4. Compare the observed statistic T_{20} with the critical value of 2.557 for a sample size n=20 and a significance level of 5 percent (taken from Table 8 on p. B-12 of the Interim Final Guidance). Since the observed value T_{20}=2.648 exceeds the critical value, there is significant evidence that the largest observation is a statistical outlier. Before excluding this value from further analysis, a valid explanation for this unusually high value should be found. Otherwise, treat the outlier as an extreme but valid concentration measurement.

REFERENCES

Aitchison, J. (1955) On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of American Statistical Association*, **50**(272): 901-8.

Barnett, V. and Lewis, T. (1978) Outliers in statistical data. New York: John Wiley & Sons.

Cohen, A.C., Jr. (1959) Simplified estimators for the normal distribution when samples are single censored or truncated. *Technometrics*, **1**:217-37.

Cox, D.R. and Hinkley, D.V. (1974) Theoretical statistics. London: Chapman & Hall.

Davis, C.B. and McNichols, R.J. (1987) One-sided intervals for at least p of m observations from a normal population on each of r future occasions. *Technometrics*, **29**(3):359-70.

Filliben, J.J. (1975) The probability plot correlation coefficient test for normality. *Technometrics*, **17**:111-7.

Gan, F.F. and Koehler, K.J. (1990) Goodness-of-fit tests based on p-p probability plots. *Technometrics*, **32**(3):289-303.

Gayen, A.K. (1949) The distribution of "Student's" t in random samples of any size drawn from non-normal universes. *Biometrika*, **36**:353-69.

Gibbons, R.D. (1987a) Statistical prediction intervals for the evaluation of ground-water quality. *Ground Water*, **25**(4):455-65.

Gibbons, R.D. (1987b) Statistical models for the analysis of volatile organic compounds in waste disposal sites. *Ground Water*, **25**(5):572-80.

Gibbons, R.D. (1990) A general statistical procedure for ground-water detection monitoring at waste disposal facilities. *Ground Water*, **28**(2):235-43.

Gibbons, R.D. (1991a) Statistical tolerance limits for ground-water monitoring. *Ground Water*, **29**(4):563-70.

Gibbons, R.D. (1991b) Some additional nonparametric prediction limits for ground-water detection monitoring at waste disposal facilities. *Ground Water*, **29**(5):729-36.

Gilliom, R.J. and Helsel, D.R. (1986) Estimation of distributional parameters for censored trace level water quality data: part 1, estimation techniques. *Water Resources Research*, **22**(2):135-46.

Guttman, I. (1970) Statistical tolerance regions: classical and bayesian. Darien, Connecticut: Hafner Publishing.

Hahn, G.J. (1970) Statistical intervals for a normal population: part 1, tables, examples, and applications. *Journal of Quality Technology*, **2**(3):115-25.

Lehmann, E.L. (1975) Nonparametrics: statistical methods based on ranks. San Francisco: Holden Day, Inc.

Madansky, A. (1988) Prescriptions for working statisticians. New York: Springer-Verlag.

McBean, E.A. and Rovers, F.A. (1992) Estimation of the probability of exceedance of contaminant concentrations. *Ground Water Monitoring Review*, Winter, 115-9.

McNichols, R.J. and Davis, C.B. (1988) Statistical issues and problems in ground water detection monitoring at hazardous waste facilities. *Ground Water Monitoring Review*, Fall.

Miller, R.G., Jr. (1986) Beyond ANOVA, basics of applied statistics. New York: John Wiley & Sons.

Milliken, G.A. and Johnson, D.E. (1984) Analysis of messy data: volume 1, designed experiments. Belmont, California: Lifetime Learning Publications.

Ott, W.R. (1990) A physical explanation of the lognormality of pollutant concentrations. *Journal of Air Waste Management Association*, **40**:1378-83.

Ryan, T.A., Jr. and Joiner, B.L. (1990) Normal probability plots and tests for normality. *Minitab Statistical Software: Technical Reports*, November, 1–1 to 1–14.

Shapiro, S.S. and Wilk, M.B. (1965) An analysis of variance test for normality (complete samples). *Biometrika*, **52**:591-611.

Shapiro, S.S. and Francia, R.S. (1972) An approximate analysis of variance test for normality. *Journal of American Statistical Association*, **67**(337):215-6.

Zacks, S. (1970) Uniformly most accurate upper tolerance limits for monotone likelihood ratio families of discrete distributions. *Journal of American Statistical Association*, **65**(329):307-16.

TABLE A-1.

COEFFICIENTS {A_{N-I+1}} FOR W TEST OF NORMALITY, FOR N=2(1)50

i/n	2	3	4	5	6	7	8	9	10
1	0.7071	0.7071	0.6872	0.6646	0.6431	0.6233	0.6052	0.5888	0.5739
2	—	.0000	.1677	.2413	.2806	.3031	.3164	.3244	.3291
3	—	—	—	.0000	.0875	.1401	.1743	.1976	.2141
4	—	—	—	—	—	.0000	.0561	.0947	.1224
5	—	—	—	—	—	—	—	.0000	.0399

i/n	11	12	13	14	15	16	17	18	19	20
1	0.5601	0.5475	0.5359	0.5251	0.5150	0.5056	0.4968	0.4886	0.4808	0.4734
2	.3315	.3325	.3325	.3318	.3306	.3290	.3273	.3253	.3232	.3211
3	.2260	.2347	.2412	.2460	.2495	.2521	.2540	.2553	.2561	.2565
4	.1429	.1586	.1707	.1802	.1878	.1939	.1988	.2027	.2059	.2085
5	.0695	.0922	.1099	.1240	.1353	.1447	.1524	.1587	.1641	.1686
6	0.0000	0.0303	0.0539	0.0727	0.0880	0.1005	0.1109	0.1197	0.1271	0.1334
7	—	—	.0000	.0240	.0433	.0593	.0725	.0837	.0932	.1013
8	—	—	—	—	.0000	.0196	.0359	.0496	.0612	.0711
9	—	—	—	—	—	—	.0000	.0163	.0303	.0422
10	—	—	—	—	—	—	—	—	.0000	.0140

i/n	21	22	23	24	25	26	27	28	29	30
1	0.4643	0.4590	0.4542	0.4493	0.4450	0.4407	0.4366	0.4328	0.4291	0.4254
2	.3185	.3156	.3126	.3098	.3069	.3043	.3018	.2992	.2968	.2944
3	.2578	.2571	.2563	.2554	.2543	.2533	.2522	.2510	.2499	.2487
4	.2119	.2131	.2139	.2145	.2148	.2151	.2152	.2151	.2150	.2148
5	.1736	.1764	.1787	.1807	.1822	.1836	.1848	.1857	.1864	.1870
6	0.1399	0.1443	0.1480	0.1512	0.1539	0.1563	0.1584	0.1601	0.1616	0.1630
7	.1092	.1150	.1201	.1245	.1283	.1316	.1346	.1372	.1395	.1415
8	.0804	.0878	.0941	.0997	.1046	.1089	.1128	.1162	.1192	.1219
9	.0530	.0618	.0696	.0764	.0823	.0876	.0923	.0965	.1002	.1036
10	.0263	.0368	.0459	.0539	.0610	.0672	.0728	.0778	.0822	.0862
11	0.0000	0.0122	0.0228	0.0321	0.0403	0.0476	0.0540	0.0598	0.0650	0.0697
12	—	—	.0000	.0107	.0200	.0284	.0358	.0424	.0483	.0537
13	—	—	—	—	.0000	.0094	.0178	.0253	.0320	.0381
14	—	—	—	—	—	—	.0000	.0084	.0159	.0227
15	—	—	—	—	—	—	—	—	.0000	.0076

i/n	31	32	33	34	35	36	37	38	39	40
1	0.4220	0.4188	0.4156	0.4127	0.4096	0.4068	0.4040	0.4015	0.3989	0.3964
2	.2921	.2898	.2876	.2854	.2834	.2813	.2794	.2774	.2755	.2737
3	.2475	.2463	.2451	.2439	.2427	.2415	.2403	.2391	.2380	.2368
4	.2145	.2141	.2137	.2132	.2127	.2121	.2116	.2110	.2104	.2098
5	.1874	.1878	.1880	.1882	.1883	.1883	.1883	.1881	.1880	.1878
6	0.1641	0.1651	0.1660	0.1667	0.1673	0.1678	0.1683	0.1686	0.1689	0.1691
7	.1433	.1449	.1463	.1475	.1487	.1496	.1503	.1513	.1520	.1526
8	.1243	.1265	.1284	.1301	.1317	.1331	.1344	.1356	.1366	.1376
9	.1066	.1093	.1118	.1140	.1160	.1179	.1196	.1211	.1225	.1237
10	.0899	.0931	.0961	.0988	.1013	.1036	.1056	.1075	.1092	.1108

TABLE A-1. (CONTINUED)

COEFFICIENTS $\{A_{N-I+1}\}$ FOR W TEST OF NORMALITY, FOR N=2(1)50

i/n	31	32	33	34	35	36	37	38	39	40
11	0.0739	0.0777	0.0812	0.0844	0.0873	0.0900	0.0924	0.0947	0.0967	0.0986
12	.0585	.0629	.0669	.0706	.0739	.0770	.0798	.0824	.0848	.0870
13	.0435	.0485	.0530	.0572	.0610	.0645	.0677	.0706	.0733	.0759
14	.0289	.0344	.0395	.0441	.0484	.0523	.0559	.0592	.0622	.0651
15	.0144	.0206	.0262	.0314	.0361	.0404	.0444	.0481	.0515	.0546
16	0.0000	0.0068	0.0131	0.0187	0.0239	0.0287	0.0331	0.0372	0.0409	0.0444
17	—	—	.0000	.0062	.0119	.0172	.0220	.0264	.0305	.0343
18	—	—	—	—	.0000	.0057	.0110	.0158	.0203	.0244
19	—	—	—	—	—	—	.0000	.0053	.0101	.0146
20	—	—	—	—	—	—	—	—	.0000	.0049

i/n	41	42	43	44	45	46	47	48	49	50
1	0.3940	0.3917	0.3894	0.3872	0.3850	0.3830	0.3808	0.3789	0.3770	0.3751
2	.2719	.2701	.2684	.2667	.2651	.2635	.2620	.2604	.2589	.2574
3	.2357	.2345	.2334	.2323	.2313	.2302	.2291	.2281	.2271	.2260
4	.2091	.2085	.2078	.2072	.2065	.2058	.2052	.2045	.2038	.2032
5	.1876	.1874	.1871	.1868	.1865	.1862	.1859	.1855	.1851	.1847
6	0.1693	0.1694	0.1695	0.1695	0.1695	0.1695	0.1695	0.1693	0.1692	0.1691
7	.1531	.1535	.1539	.1542	.1545	.1548	.1550	.1551	.1553	.1554
8	.1384	.1392	.1398	.1405	.1410	.1415	.1420	.1423	.1427	.1430
9	.1249	.1259	.1269	.1278	.1286	.1293	.1300	.1306	.1312	.1317
10	.1123	.1136	.1149	.1160	.1170	.1180	.1189	.1197	.1205	.1212
11	0.1004	0.1020	0.1035	0.1049	0.1062	0.1073	0.1085	0.1095	0.1105	0.1113
12	.0891	.0909	.0927	.0943	.0959	.0972	.0986	.0998	.1010	.1020
13	.0782	.0804	.0824	.0842	.0860	.0876	.0892	.0906	.0919	.0932
14	.0677	.0701	.0724	.0745	.0775	.0785	.0801	.0817	.0832	.0846
15	.0575	.0602	.0628	.0651	.0673	.0694	.0713	.0731	.0748	.0764
16	0.0476	0.0506	0.0534	0.0560	0.0584	0.0607	0.0628	0.0648	0.0667	0.0685
17	.0379	.0411	.0442	.0471	.0497	.0522	.0546	.0568	.0588	.0608
18	.0283	.0318	.0352	.0383	.0412	.0439	.0465	.0489	.0511	.0532
19	.0188	.0227	.0263	.0296	.0328	.0357	.0385	.0411	.0436	.0459
20	.0094	.0136	.0175	.0211	.0245	.0277	.0307	.0335	.0361	.0386
21	0.0000	0.0045	0.0087	0.0126	0.0163	0.0197	0.0229	0.0259	0.0288	0.0314
22	—	—	.0000	.0042	.0081	.0118	.0153	.0185	.0215	.0244
23	—	—	—	—	.0000	.0039	.0076	.0111	.0143	.0174
24	—	—	—	—	—	—	.0000	.0037	.0071	.0104
25	—	—	—	—	—	—	—	—	.0000	.0035

TABLE A-2.

PERCENTAGE POINTS OF THE W TEST FOR N=3(1)50

n	0.01	0.05
3	0.753	0.767
4	.687	.748
5	.686	.762
6	0.713	0.788
7	.730	.803
8	.749	.818
9	.764	.829
10	.781	.842
11	0.792	0.850
12	.805	.859
13	.814	.866
14	.825	.874
15	.835	.881
16	0.844	0.887
17	.851	.892
18	.858	.897
19	.863	.901
20	.868	.905
21	0.873	0.908
22	.878	.911
23	.881	.914
24	.884	.916
25	.888	.918
26	0.891	0.920
27	.894	.923
28	.896	.924
29	.898	.926
30	.900	.927
31	0.902	0.929
32	.904	.930
33	.906	.931
34	.908	.933
35	.910	.934

TABLE A-2. (CONTINUED)

PERCENTAGE POINTS OF THE W TEST FOR N=3(1)50

n	0.01	0.05
36	0.912	0.935
37	.914	.936
38	.916	.938
39	.917	.939
40	.919	.940
41	0.920	0.941
42	.922	.942
43	.923	.943
44	.924	.944
45	.926	.945
46	0.927	0.945
47	.928	.946
48	.929	.947
49	.929	.947
50	.930	.947

TABLE A-3.

PERCENTAGE POINTS OF THE W′ TEST FOR N≥35

n	.01	.05
35	0.919	0.943
50	.935	.953
51	0.935	0.954
53	.938	.957
55	.940	.958
57	.944	.961
59	.945	.962
61	0.947	0.963
63	.947	.964
65	.948	.965
67	.950	.966
69	.951	.966
71	0.953	0.967
73	.956	.968
75	.956	.969
77	.957	.969
79	.957	.970
81	0.958	0.970
83	.960	.971
85	.961	.972
87	.961	.972
89	.961	.972
91	0.962	0.973
93	.963	.973
95	.965	.974
97	.965	.975
99	.967	.976

TABLE A-4.

PERCENT POINTS OF THE NORMAL PROBABILITY PLOT CORRELETION COEFFICIENT FOR N=3(1)50(5)100

n	.01	.025	.05
3	.869	.872	.879
4	.822	.845	.868
5	.822	.855	.879
6	.835	.868	.890
7	.847	.876	.899
8	.859	.886	.905
9	.868	.893	.912
10	.876	.900	.917
11	.883	.906	.922
12	.889	.912	.926
13	.895	.917	.931
14	.901	.921	.934
15	.907	.925	.937
16	.912	.928	.940
17	.912	.931	.942
18	.919	.934	.945
19	.923	.937	.947
20	.925	.939	.950
21	.928	.942	.952
22	.930	.944	.954
23	.933	.947	.955
24	.936	.949	.957
25	.937	.950	.958
26	.939	.952	.959
27	.941	.953	.960
28	.943	.955	.962
29	.945	.956	.962
30	.947	.957	.964
31	.948	.958	.965
32	.949	.959	.966
33	.950	.960	.967
34	.951	.960	.967
35	.952	.961	.968
36	.953	.962	.968
37	.955	.962	.969
38	.956	.964	.970
39	.957	.965	.971
40	.958	.966	.972

TABLE A-4. (CONTINUED)

PERCENT POINTS OF THE NORMAL PROBABILITY PLOT CORRELETION COEFFICIENT FOR N=3(1)50(5)100

n	.01	.025	.05
41	.958	.967	.973
42	.959	.967	.973
43	.959	.967	.973
44	.960	.968	.974
45	.961	.969	.974
46	.962	.969	.974
47	.963	.970	.975
48	.963	.970	.975
49	.964	.971	.977
50	.965	.972	.978
55	.967	.974	.980
60	.970	.976	.981
65	.972	.977	.982
70	.974	.978	.983
75	.975	.979	.984
80	.976	.980	.985
85	.977	.981	.985
90	.978	.982	.985
95	.979	.983	.986
100	.981	.984	.987

TABLE A-5.

VALUES OF LAMBDA FOR COHEN'S METHOD

γ	Percentage of Non-detects										
	.01	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
.01	.0102	.0530	.1111	.1747	.2443	.3205	.4043	.4967	.5989	.7128	.8403
.05	.0105	.0547	.1143	.1793	.2503	.3279	.4130	.5066	.6101	.7252	.8540
.10	.0110	.0566	.1180	.1848	.2574	.3366	.4233	.5184	.6234	.7400	.8703
.15	.0113	.0584	.1215	.1898	.2640	.3448	.4330	.5296	.6361	.7542	.8860
.20	.0116	.0600	.1247	.1946	.2703	.3525	.4422	.5403	.6483	.7678	.9012
.25	.0120	.0615	.1277	.1991	.2763	.3599	.4510	.5506	.6600	.7810	.9158
.30	.0122	.0630	.1306	.2034	.2819	.3670	.4595	.5604	.6713	.7937	.9300
.35	.0125	.0643	.1333	.2075	.2874	.3738	.4676	.5699	.6821	.8060	.9437
.40	.0128	.0657	.1360	.2114	.2926	.3803	.4755	.5791	.6927	.8179	.9570
.45	.0130	.0669	.1385	.2152	.2976	.3866	.4831	.5880	.7029	.8295	.9700
.50	.0133	.0681	.1409	.2188	.3025	.3928	.4904	.5967	.7129	.8408	.9826
.55	.0135	.0693	.1432	.2224	.3073	.3987	.4976	.6051	.7225	.8517	.9950
.60	.0137	.0704	.1455	.2258	.3118	.4045	.5046	.6133	.7320	.8625	1.0070
.65	.0140	.0715	.1477	.2291	.3163	.4101	.5114	.6213	.7412	.8729	1.0188
.70	.0142	.0726	.1499	.2323	.3206	.4156	.5180	.6291	.7502	.8832	1.0303
.75	.0144	.0736	.1520	.2355	.3249	.4209	.5245	.6367	.7590	.8932	1.0416
.80	.0146	.0747	.1540	.2386	.3290	.4261	.5308	.6441	.7676	.9031	1.0527
.85	.0148	.0756	.1560	.2416	.3331	.4312	.5370	.6515	.7761	.9127	1.0636
.90	.0150	.0766	.1579	.2445	.3370	.4362	.5430	.6586	.7844	.9222	1.0743
.95	.0152	.0775	.1598	.2474	.3409	.4411	.5490	.6656	.7925	.9314	1.0847
1.00	.0153	.0785	.1617	.2502	.3447	.4459	.5548	.6725	.8005	.9406	[illegible]
1.05	.0155	.0794	.1635	.2530	.3484	.4506	.5605	.6793	.8084	.9496	[illegible]
1.10	.0157	.0803	.1653	.2557	.3521	.4553	.5662	.6860	.8161	.9584	1.1152
1.15	.0159	.0811	.1671	.2584	.3557	.4598	.5717	.6925	.8237	.9671	1.1250
1.20	.0160	.0820	.1688	.2610	.3592	.4643	.5771	.6990	.8312	.9756	1.1347
1.25	.0162	.0828	.1705	.2636	.3627	.4687	.5825	.7053	.8385	.9841	1.1443
1.30	.0164	.0836	.1722	.2661	.3661	.4730	.5878	.7115	.8458	.9924	1.1537
1.35	.0165	.0845	.1738	.2686	.3695	.4773	.5930	.7177	.8529	1.0006	1.1629
1.40	.0167	.0853	.1754	.2710	.3728	.4815	.5981	.7238	.8600	1.0087	1.1721
1.45	.0168	.0860	.1770	.2735	.3761	.4856	.6031	.7298	.8670	1.0166	1.1812
1.50	.0170	.0868	.1786	.2758	.3793	.4897	.6081	.7357	.8738	1.0245	1.1901
1.55	.0171	.0876	.1801	.2782	.3825	.4938	.6130	.7415	.8806	1.0323	1.1989
1.60	.0173	.0883	.1817	.2805	.3856	.4977	.6179	.7472	.8873	1.0400	1.2076
1.65	.0174	.0891	.1832	.2828	.3887	.5017	.6227	.7529	.8939	1.0476	1.2162
1.70	.0176	.0898	.1846	.2851	.3918	.5055	.6274	.7585	.9005	1.0551	1.2248
1.75	.0177	.0905	.1861	.2873	.3948	.5094	.6321	.7641	.9069	1.0625	1.2332
1.80	.0179	.0913	.1876	.2895	.3978	.5132	.6367	.7696	.9133	1.0698	1.2415
1.85	.0180	.0920	.1890	.2917	.4007	.5169	.6413	.7750	.9196	1.0771	1.2497
1.90	.0181	.0927	.1904	.2938	.4036	.5206	.6458	.7804	.9259	1.0842	1.2579
1.95	.0183	.0933	.1918	.2960	.4065	.5243	.6502	.7857	.9321	1.0913	1.2660

TABLE A-5. (CONTINUED)

VALUES OF LAMBDA FOR COHEN'S METHOD

	Percentage of Non-detects										
γ	.01	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
2.00	.0184	.0940	.1932	.2981	.4093	.5279	.6547	.7909	.9382	1.0984	1.2739
2.05	.0186	.0947	.1945	.3001	.4122	.5315	.6590	.7961	.9442	1.1053	1.2819
2.10	.0187	.0954	.1959	.3022	.4149	.5350	.6634	.8013	.9502	1.1122	1.2897
2.15	.0188	.0960	.1972	.3042	.4177	.5385	.6676	.8063	.9562	1.1190	1.2974
2.20	.0189	.0967	.1986	.3062	.4204	.5420	.6719	.8114	.9620	1.1258	1.3051
2.25	.0191	.0973	.1999	.3082	.4231	.5454	.6761	.8164	.9679	1.1325	1.3127
2.30	.0192	.0980	.2012	.3102	.4258	.5488	.6802	.8213	.9736	1.1391	1.3203
2.35	.0193	.0986	.2025	.3122	.4285	.5522	.6844	.8262	.9794	1.1457	1.3278
2.40	.0194	.0992	.2037	.3141	.4311	.5555	.6884	.8311	.9850	1.1522	1.3352
2.45	.0196	.0998	.2050	.3160	.4337	.5588	.6925	.8359	.9906	1.1587	1.3425
2.50	.0197	.1005	.2062	.3179	.4363	.5621	.6965	.8407	.9962	1.1651	1.3498
2.55	.0198	.1011	.2075	.3198	.4388	.5654	.7005	.8454	1.0017	1.1714	1.3571
2.60	.0199	.1017	.2087	.3217	.4414	.5686	.7044	.8501	1.0072	1.1777	1.3642
2.65	.0201	.1023	.2099	.3236	.4439	.5718	.7083	.8548	1.0126	1.1840	1.3714
2.70	.0202	.1029	.2111	.3254	.4464	.5750	.7122	.8594	1.0180	1.1902	1.3784
2.75	.0203	.1035	.2123	.3272	.4489	.5781	.7161	.8639	1.0234	1.1963	1.3854
2.80	.0204	.1040	.2135	.3290	.4513	.5812	.7199	.8685	1.0287	1.2024	1.3924
2.85	.0205	.1046	.2147	.3308	.4537	.5843	.7237	.8730	1.0339	1.2085	1.3993
2.90	.0206	.1052	.2158	.3326	.4562	.5874	.7274	.8775	1.0392	1.2145	1.4061
2.95	.0207	.1058	.2170	.3344	.4585	.5905	.7311	.8819	1.0443	1.2205	1.4129
3.00	.0209	.1063	.2182	.3361	.4609	.5935	.7348	.8863	1.0495	1.2264	1.4197
3.05	.0210	.1069	.2193	.3378	.4633	.5965	.7385	.8907	1.0546	1.2323	1.4264
3.10	.0211	.1074	.2204	.3396	.4656	.5995	.7422	.8950	1.0597	1.2381	1.4330
3.15	.0212	.1080	.2216	.3413	.4679	.6024	.7458	.8993	1.0647	1.2439	1.4396
3.20	.0213	.1085	.2227	.3430	.4703	.6054	.7494	.9036	1.0697	1.2497	1.4462
3.25	.0214	.1091	.2238	.3447	.4725	.6083	.7529	.9079	1.0747	1.2554	1.4527
3.30	.0215	.1096	.2249	.3464	.4748	.6112	.7565	.9121	1.0796	1.2611	1.4592
3.35	.0216	.1102	.2260	.3480	.4771	.6141	.76	.9163	1.0845	1.2668	1.4657
3.40	.0217	.1107	.2270	.3497	.4793	.6169	.7635	.9205	1.0894	1.2724	1.4720
3.45	.0218	.1112	.2281	.3513	.4816	.6197	.7670	.9246	1.0942	1.2779	1.4784
3.50	.0219	.1118	.2292	.3529	.4838	.6226	.7704	.9287	1.0990	1.2835	1.4847
3.55	.0220	.1123	.2303	.3546	.4860	.6254	.7739	.9328	1.1038	1.2890	1.4910
3.60	.0221	.1128	.2313	.3562	.4882	.6282	.7773	.9369	1.1086	1.2945	1.4972
3.65	.0222	.1133	.2324	.3578	.4903	.6309	.7807	.9409	1.1133	1.2999	1.5034
3.70	.0223	.1138	.2334	.3594	.4925	.6337	.7840	.9449	1.1180	1.3053	1.5096
3.75	.0224	.1143	.2344	.3609	.4946	.6364	.7874	.9489	1.1226	1.3107	1.5157
3.80	.0225	.1148	.2355	.3625	.4968	.6391	.7907	.9529	1.1273	1.3160	1.5218
3.85	.0226	.1153	.2365	.3641	.4989	.6418	.7940	.9568	1.1319	1.3213	1.5279
3.90	.0227	.1158	.2375	.3656	.5010	.6445	.7973	.9607	1.1364	1.3266	1.5339
3.95	.0228	.1163	.2385	.3672	.5031	.6472	.8006	.9646	1.1410	1.3318	1.5399

TABLE A-5. (CONTINUED)

VALUES OF LAMBDA FOR COHEN'S METHOD

	Percentage of Non-detects										
γ	.01	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
4.00	.0229	.1168	.2395	.3687	.5052	.6498	.8038	.9685	1.1455	1.3371	1.5458
4.05	.0230	.1173	.2405	.3702	.5072	.6525	.8070	.9723	1.1500	1.3423	1.5518
4.10	.0231	.1178	.2415	.3717	.5093	.6551	.8102	.9762	1.1545	1.3474	1.5577
4.15	.0232	.1183	.2425	.3732	.5113	.6577	.8134	.9800	1.1590	1.3526	1.5635
4.20	.0233	.1188	.2435	.3747	.5134	.6603	.8166	.9837	1.1634	1.3577	1.5693
4.25	.0234	.1193	.2444	.3762	.5154	.6629	.8198	.9875	1.1678	1.3627	1.5751
4.30	.0235	.1197	.2454	.3777	.5174	.6654	.8229	.9913	1.1722	1.3678	1.5809
4.35	.0236	.1202	.2464	.3792	.5194	.6680	.8260	.9950	1.1765	1.3728	1.5866
4.40	.0237	.1207	.2473	.3806	.5214	.6705	.8291	.9987	1.1809	1.3778	1.5924
4.45	.0238	.1212	.2483	.3821	.5234	.6730	.8322	1.0024	1.1852	1.3828	1.5980
4.50	.0239	.1216	.2492	.3836	.5253	.6755	.8353	1.0060	1.1895	1.3878	1.6037
4.55	.0240	.1221	.2502	.3850	.5273	.6780	.8384	1.0097	1.1937	1.3927	1.6093
4.60	.0241	.1225	.2511	.3864	.5292	.6805	.8414	1.0133	1.1980	1.3976	1.6149
4.65	.0241	.1230	.2521	.3879	.5312	.6830	.8445	1.0169	1.2022	1.4024	1.6205
4.70	.0242	.1235	.2530	.3893	.5331	.6855	.8475	1.0205	1.2064	1.4073	1.6260
4.75	.0243	.1239	.2539	.3907	.5350	.6879	.8505	1.0241	1.2106	1.4121	1.6315
4.80	.0244	.1244	.2548	.3921	.5370	.6903	.8535	1.0277	1.2148	1.4169	1.6370
4.85	.0245	.1248	.2558	.3935	.5389	.6928	.8564	1.0312	1.2189	1.4217	1.6425
4.90	.0246	.1253	.2567	.3949	.5407	.6952	.8594	1.0348	1.2230	1.4265	1.6479
4.95	.0247	.1257	.2576	.3963	.5426	.6976	.8623	1.0383	1.2272	1.4312	1.65
5.00	.0248	.1262	.2585	.3977	.5445	.7000	.8653	1.0418	1.2312	1.4359	1.
5.05	.0249	.1266	.2594	.3990	.5464	.7024	.8682	1.0452	1.2353	1.4406	1.6641
5.10	.0249	.1270	.2603	.4004	.5482	.7047	.8711	1.0487	1.2394	1.4453	1.6694
5.15	.0250	.1275	.2612	.4018	.5501	.7071	.8740	1.0521	1.2434	1.4500	1.6747
5.20	.0251	.1279	.2621	.4031	.5519	.7094	.8768	1.0556	1.2474	1.4546	1.6800
5.25	.0252	.1284	.2629	.4045	.5537	.7118	.8797	1.0590	1.2514	1.4592	1.6853
5.30	.0253	.1288	.2638	.4058	.5556	.7141	.8825	1.0624	1.2554	1.4638	1.6905
5.35	.0254	.1292	.2647	.4071	.5574	.7164	.8854	1.0658	1.2594	1.4684	1.6958
5.40	.0255	.1296	.2656	.4085	.5592	.7187	.8882	1.0691	1.2633	1.4729	1.7010
5.45	.0255	.1301	.2664	.4098	.5610	.7210	.8910	1.0725	1.2672	1.4775	1.7061
5.50	.0256	.1305	.2673	.4111	.5628	.7233	.8938	1.0758	1.2711	1.4820	1.7113
5.55	.0257	.1309	.2682	.4124	.5646	.7256	.8966	1.0792	1.2750	1.4865	1.7164
5.60	.0258	.1313	.2690	.4137	.5663	.7278	.8994	1.0825	1.2789	1.4910	1.7215
5.65	.0259	.1318	.2699	.4150	.5681	.7301	.9022	1.0858	1.2828	1.4954	1.7266
5.70	.0260	.1322	.2707	.4163	.5699	.7323	.9049	1.0891	1.2866	1.4999	1.7317
5.75	.0260	.1326	.2716	.4176	.5716	.7346	.9077	1.0924	1.2905	1.5043	1.7368
5.80	.0261	.1330	.2724	.4189	.5734	.7368	.9104	1.0956	1.2943	1.5087	1.7418
5.85	.0262	.1334	.2732	.4202	.5751	.7390	.9131	1.0989	1.2981	1.5131	1.7468
5.90	.0263	.1338	.2741	.4215	.5769	.7412	.9158	1.1021	1.3019	1.5175	1.7518
5.95	.0264	.1342	.2749	.4227	.5786	.7434	.9185	1.1053	1.3057	1.5218	1.7568
6.00	.0264	.1346	.2757	.4240	.5803	.7456	.9212	1.1085	1.3094	1.5262	1.7617

TABLE A-6.

MINIMUM COVERAGE (BETA) OF 95% CONFIDENCE NON-PARAMETRIC UPPER TOLERANCE LIMITS

N	β(maximum)	β(2nd largest)
1	5.0	----
2	22.4	2.6
3	36.8	13.6
4	47.3	24.8
5	54.9	34.2
6	60.7	41.8
7	65.2	48.0
8	68.8	53.0
9	71.7	57.0
10	74.1	60.6
11	76.2	63.6
12	77.9	66.2
13	79.4	68.4
14	80.7	70.4
15	81.9	72.0
16	82.9	73.6
17	83.8	75.0
18	84.7	76.2
19	85.4	77.4
20	86.1	78.4
21	86.7	79.4
22	87.3	80.2
23	87.8	81.0
24	88.3	81.8
25	88.7	82.4
26	89.1	83.0
27	89.5	83.6
28	89.9	84.2
29	90.2	84.6
30	90.5	85.2
31	90.8	85.6
32	91.1	86.0
33	91.3	86.4
34	91.6	86.8
35	91.8	87.2
36	92.0	87.4
37	92.2	87.8
38	92.4	88.2
39	92.6	88.4
40	92.8	88.6

TABLE A-6. (CONTINUED)

MINIMUM COVERAGE (BETA) OF 95% CONFIDENCE NON-PARAMETRIC UPPER TOLERANCE LIMITS

N	β(maximum)	β(2nd largest)
41	93.0	89.0
42	93.1	89.2
43	93.3	89.4
44	93.4	89.6
45	93.6	89.8
46	93.7	90.0
47	93.8	90.2
48	93.9	90.4
49	94.1	90.6
50	94.2	90.8
55	94.7	91.6
60	95.1	92.4
65	95.5	93.0
70	95.8	93.4
75	96.1	93.8
80	96.3	94.2
85	96.5	94.6
90	96.7	94.8
95	96.9	95.0
100	97.0	95.4

TABLE A-7.

CONFIDENCE LEVELS FOR NON-PARAMETRIC PREDICTION LIMITS FOR N=1(1)100

	NUMBER OF FUTURE SAMPLES							
N	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
1	50.0	33.3	25.0	20.0	16.7	14.3	12.5	11.1
2	66.7	50.0	40.0	33.3	28.6	25.0	22.2	20.0
3	75.0	60.0	50.0	42.9	37.5	33.3	30.0	27.3
4	80.0	66.7	57.1	50.0	44.4	40.0	36.4	33.3
5	83.3	71.4	62.5	55.6	50.0	45.5	41.7	38.5
6	85.7	75.0	66.7	60.0	54.5	50.0	46.2	42.9
7	87.5	77.8	70.0	63.6	58.3	53.8	50.0	46.7
8	88.9	80.0	72.7	66.7	61.5	57.1	53.3	50.0
9	90.0	81.8	75.0	69.2	64.3	60.0	56.3	52.9
10	90.9	83.3	76.9	71.4	66.7	62.5	58.8	55.6
11	91.7	84.6	78.6	73.3	68.8	64.7	61.1	57.9
12	92.3	85.7	80.0	75.0	70.6	66.7	63.2	60.0
13	92.9	86.7	81.3	76.5	72.2	68.4	65.0	61.9
14	93.3	87.5	82.4	77.8	73.7	70.0	66.7	63.6
15	93.8	88.2	83.3	78.9	75.0	71.4	68.2	65.2
16	94.1	88.9	84.2	80.0	76.2	72.7	69.6	66.7
17	94.4	89.5	85.0	81.0	77.3	73.9	70.8	68.0
18	94.7	90.0	85.7	81.8	78.3	75.0	72.0	69.2
19	95.0	90.5	86.4	82.6	79.2	76.0	73.1	70.4
20	95.2	90.9	87.0	83.3	80.0	76.9	74.1	71.4
21	95.5	91.3	87.5	84.0	80.8	77.8	75.0	72.4
22	95.7	91.7	88.0	84.6	81.5	78.6	75.9	73.3
23	95.8	92.0	88.5	85.2	82.1	79.3	76.7	74.2
24	96.0	92.3	88.9	85.7	82.8	80.0	77.4	75.0
25	96.2	92.6	89.3	86.2	83.3	80.6	78.1	75.8
26	96.3	92.9	89.7	86.7	83.9	81.3	78.8	76.5
27	96.4	93.1	90.0	87.1	84.4	81.8	79.4	77.1
28	96.6	93.3	90.3	87.5	84.8	82.4	80.0	77.8
29	96.7	93.5	90.6	87.9	85.3	82.9	80.6	78.4
30	96.8	93.8	90.9	88.2	85.7	83.3	81.1	78.9
31	96.9	93.9	91.2	88.6	86.1	83.8	81.6	79.5
32	97.0	94.1	91.4	88.9	86.5	84.2	82.1	80.0
33	97.1	94.3	91.7	89.2	86.8	84.6	82.5	80.5
34	97.1	94.4	91.9	89.5	87.2	85.0	82.9	81.0
35	97.2	94.6	92.1	89.7	87.5	85.4	83.3	81.4
36	97.3	94.7	92.3	90.0	87.8	85.7	83.7	81.8
37	97.4	94.9	92.5	90.2	88.1	86.0	84.1	82.2
38	97.4	95.0	92.7	90.5	88.4	86.4	84.4	82.6
39	97.5	95.1	92.9	90.7	88.6	86.7	84.8	83.0
40	97.6	95.2	93.0	90.9	88.9	87.0	85.1	83.3

TABLE A-7. (CONTINUED)

CONFIDENCE LEVELS FOR NON-PARAMETRIC PREDICTION LIMITS FOR N=1(1)100

N	NUMBER OF FUTURE SAMPLES k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
41	97.6	95.3	93.2	91.1	89.1	87.2	85.4	83.7
42	97.7	95.5	93.3	91.3	89.4	87.5	85.7	84.0
43	97.7	95.6	93.5	91.5	89.6	87.8	86.0	84.3
44	97.8	95.7	93.6	91.7	89.8	88.0	86.3	84.6
45	97.8	95.7	93.8	91.8	90.0	88.2	86.5	84.9
46	97.9	95.8	93.9	92.0	90.2	88.5	86.8	85.2
47	97.9	95.9	94.0	92.2	90.4	88.7	87.0	85.5
48	98.0	96.0	94.1	92.3	90.6	88.9	87.3	85.7
49	98.0	96.1	94.2	92.5	90.7	89.1	87.5	86.0
50	98.0	96.2	94.3	92.6	90.9	89.3	87.7	86.2
51	98.1	96.2	94.4	92.7	91.1	89.5	87.9	86.4
52	98.1	96.3	94.5	92.9	91.2	89.7	88.1	86.7
53	98.1	96.4	94.6	93.0	91.4	89.8	88.3	86.9
54	98.2	96.4	94.7	93.1	91.5	90.0	88.5	87.1
55	98.2	96.5	94.8	93.2	91.7	90.2	88.7	87.3
56	98.2	96.6	94.9	93.3	91.8	90.3	88.9	87.5
57	98.3	96.6	95.0	93.4	91.9	90.5	89.1	87.7
58	98.3	96.7	95.1	93.5	92.1	90.6	89.2	87.9
59	98.3	96.7	95.2	93.7	92.2	90.8	89.4	88.1
60	98.4	96.8	95.2	93.8	92.3	90.9	89.6	88.2
61	98.4	96.8	95.3	93.8	92.4	91.0	89.7	88.4
62	98.4	96.9	95.4	93.9	92.5	91.2	89.9	88.6
63	98.4	96.9	95.5	94.0	92.6	91.3	90.0	88.7
64	98.5	97.0	95.5	94.1	92.8	91.4	90.1	88.9
65	98.5	97.0	95.6	94.2	92.9	91.5	90.3	89.0
66	98.5	97.1	95.7	94.3	93.0	91.7	90.4	89.2
67	98.5	97.1	95.7	94.4	93.1	91.8	90.5	89.3
68	98.6	97.1	95.8	94.4	93.2	91.9	90.7	89.5
69	98.6	97.2	95.8	94.5	93.2	92.0	90.8	89.6
70	98.6	97.2	95.9	94.6	93.3	92.1	90.9	89.7
71	98.6	97.3	95.9	94.7	93.4	92.2	91.0	89.9
72	98.6	97.3	96.0	94.7	93.5	92.3	91.1	90.0
73	98.6	97.3	96.1	94.8	93.6	92.4	91.3	90.1
74	98.7	97.4	96.1	94.9	93.7	92.5	91.4	90.2
75	98.7	97.4	96.2	94.9	93.8	92.6	91.5	90.4
76	98.7	97.4	96.2	95.0	93.8	92.7	91.6	90.5
77	98.7	97.5	96.3	95.1	93.9	92.8	91.7	90.6
78	98.7	97.5	96.3	95.1	94.0	92.9	91.8	90.7
79	98.8	97.5	96.3	95.2	94.0	92.9	91.9	90.8
80	98.8	97.6	96.4	95.2	94.1	93.0	92.0	90.9

TABLE A-7. (CONTINUED)

CONFIDENCE LEVELS FOR NON-PARAMETRIC PREDICTION LIMITS FOR N=1(1)100

	NUMBER OF FUTURE SAMPLES							
N	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
81	98.8	97.6	96.4	95.3	94.2	93.1	92.0	91.0
82	98.8	97.6	96.5	95.3	94.3	93.2	92.1	91.1
83	98.8	97.6	96.5	95.4	94.3	93.3	92.2	91.2
84	98.8	97.7	96.6	95.5	94.4	93.3	92.3	91.3
85	98.8	97.7	96.6	95.5	94.4	93.4	92.4	91.4
86	98.9	97.7	96.6	95.6	94.5	93.5	92.5	91.5
87	98.9	97.8	96.7	95.6	94.6	93.5	92.6	91.6
88	98.9	97.8	96.7	95.7	94.6	93.6	92.6	91.7
89	98.9	97.8	96.7	95.7	94.7	93.7	92.7	91.8
90	98.9	97.8	96.8	95.7	94.7	93.8	92.8	91.8
91	98.9	97.8	96.8	95.8	94.8	93.8	92.9	91.9
92	98.9	97.9	96.8	95.8	94.8	93.9	92.9	92.0
93	98.9	97.9	96.9	95.9	94.9	93.9	93.0	92.1
94	98.9	97.9	96.9	95.9	94.9	94.0	93.1	92.2
95	99.0	97.9	96.9	96.0	95.0	94.1	93.1	92.2
96	99.0	98.0	97.0	96.0	95.0	94.1	93.2	92.3
97	99.0	98.0	97.0	96.0	95.1	94.2	93.3	92.4
98	99.0	98.0	97.0	96.1	95.1	94.2	93.3	92.5
99	99.0	98.0	97.1	96.1	95.2	94.3	93.4	92.5
100	99.0	98.0	97.1	96.2	95.2	94.3	93.5	92.6

I. CONSTRUCTION OF POWER CURVES

To construct power curves for each of the parametric and non-parametric retesting strategies, random standard Normal deviates were generated on an IBM mainframe computer using SAS. The background level mean concentration was set to zero, while the alternative mean concentration level was incremented in steps of $\Delta=0.5$ standardized units above the background level. At each increment, 5000 iterations of the retesting strategy were simulated; the proportion of iterations indicating contamination at any one of the wells in the downgradient monitoring network was designated as the effective power of the retesting strategy (for that Δ and configuration of background samples and monitoring wells).

Power values for the EPA Reference Power Curves were not simulated, but represent analytical calculations based on the non-central t-distribution with non-centrality parameter Δ. SAS programs for simulating the effective power of any of the parametric or non-parametric retesting strategies are presented below.

```
//*********************************************************************;
//*   DESCRIPTION:  *** PARAMETRIC SIMULATIONS ***
//*
//*   This program produces power curves for 35 different curve
//*   simulations (refer to the %LET statements below).  Delta ranges
//*   from 0 to 5 by 0.5.  The variable list is as follows for the
//*   input parameters:
//*
//*   BG = Background
//*   WL = Well
//*   TL = Tolerance Limit
//*   PL = Prediction Limit
//*
//*********************************************************************;
//    EXEC SAS
//    OUTSAS DD DSN=XXXXXXX.GWT03000.SJA3092.CURVES,
//    DISP=OLD
//    SYSIN DD *

OPTIONS LS=132 PS=57;
%LET ISTART=1;
%LET CURVENUM=35;
%LET RSEED=2020;
%LET REPEAT=5000;
%LET ITPRINT=1000;

%LET BG1 =24;     %LET WL1 =5;      %LET TL1 =0.95;    %LET PL1 =0.80;
%LET BG2 =24;     %LET WL2 =5;      %LET TL2 =0.95;    %LET PL2 =0.85;
%LET BG3 =8;      %LET WL3 =5;      %LET TL3 =0.95;    %LET PL3 =0.80;
%LET BG4 =8;      %LET WL4 =5;      %LET TL4 =0.95;    %LET PL4 =0.85;
%LET BG5 =24;     %LET WL5 =20;     %LET TL5 =0.95;    %LET PL5 =0.95;
%LET BG6 =24;     %LET WL6 =20;     %LET TL6 =0.95;    %LET PL6 =0.97;
%LET BG7 =8;      %LET WL7 =20;     %LET TL7 =0.95;    %LET PL7 =0.95;
%LET BG8 =8;      %LET WL8 =20;     %LET TL8 =0.95;    %LET PL8 =0.97;
%LET BG9 =24;     %LET WL9 =50;     %LET TL9 =0.95;    %LET PL9 =0.98;
%LET BG10=24;     %LET WL10=50;     %LET TL10=0.95;    %LET PL10=0.99;
```

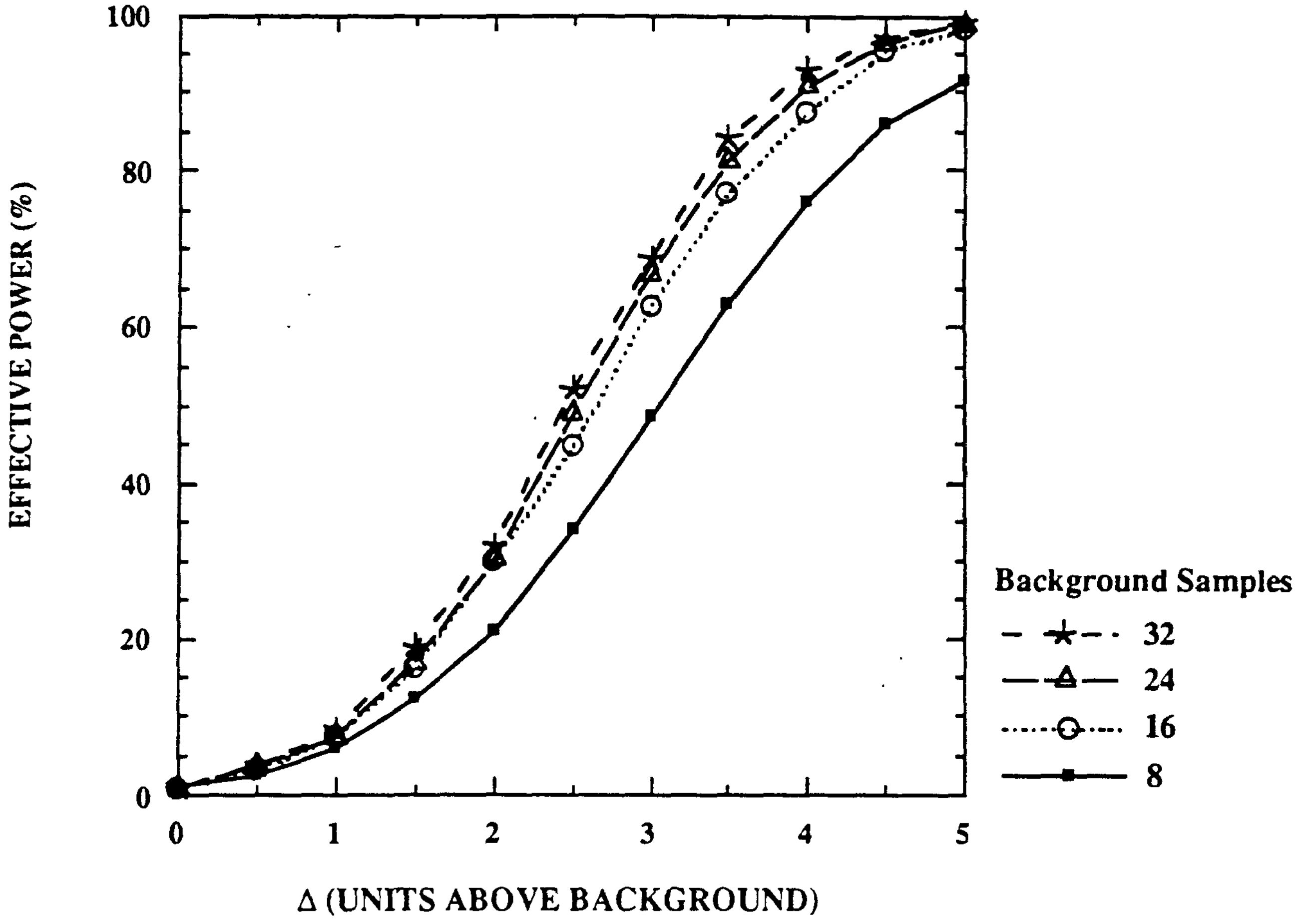
```
%LET BG11=24;       %LET WL11=50;       %LET TL11=0.99;     %LET PL11=0.90;
%LET BG12=24;       %LET WL12=50;       %LET TL12=0.99;     %LET PL12=0.93;
%LET BG13=24;       %LET WL13=50;       %LET TL13=0.99;     %LET PL13=0.94;
%LET BG14=24;       %LET WL14=50;       %LET TL14=0.98;     %LET PL14=0.95;
%LET BG15=24;       %LET WL15=50;       %LET TL15=0.98;     %LET PL15=0.97;
%LET BG16=24;       %LET WL16=100;      %LET TL16=0.98;     %LET PL16=0.97;
%LET BG17=24;       %LET WL17=100;      %LET TL17=0.98;     %LET PL17=0.99;
%LET BG18=24;       %LET WL18=100;      %LET TL18=0.99;     %LET PL18=0.95;
%LET BG19=24;       %LET WL19=100;      %LET TL19=0.99;     %LET PL19=0.97;
%LET BG20=24;       %LET WL20=100;      %LET TL20=0.99;     %LET PL20=0.98;
%LET BG21=8;        %LET WL21=20;       %LET TL21=0.95;     %LET PL21=0.98;
%LET BG22=8;        %LET WL22=5;        %LET TL22=0.95;     %LET PL22=0.90;
%LET BG23=16;       %LET WL23=5;        %LET TL23=0.95;     %LET PL23=0.85;
%LET BG24=16;       %LET WL24=5;        %LET TL24=0.95;     %LET PL24=0.90;
%LET BG25=24;       %LET WL25=5;        %LET TL25=0.95;     %LET PL25=0.90;
%LET BG26=16;       %LET WL26=20;       %LET TL26=0.95;     %LET PL26=0.95;
%LET BG27=16;       %LET WL27=20;       %LET TL27=0.95;     %LET PL27=0.97;
%LET BG28=16;       %LET WL28=50;       %LET TL28=0.98;     %LET PL28=0.95;
%LET BG29=16;       %LET WL29=50;       %LET TL29=0.98;     %LET PL29=0.97;
%LET BG30=16;       %LET WL30=50;       %LET TL30=0.99;     %LET PL30=0.90;
%LET BG31=16;       %LET WL31=50;       %LET TL31=0.99;     %LET PL31=0.92;
%LET BG32=24;       %LET WL32=100;      %LET TL32=0.98;     %LET PL32=0.98;
%LET BG33=16;       %LET WL33=100;      %LET TL33=0.98;     %LET PL33=0.98;
%LET BG34=16;       %LET WL34=100;      %LET TL34=0.99;     %LET PL34=0.95;
%LET BG35=16;       %LET WL35=100;      %LET TL35=0.99;     %LET PL35=0.96;

%MACRO PARSIM;
DATA ITERATE;
*** Set changing simulation variable to common variable names;
      BG=&&BG&I;
      WL=&&WL&I;
      TL=&&TL&I;
      PL=&&PL&I;

DO DELTA=0 TO 5 BY 0.5;
*** Initialize TP0, TP1 & TP2 to 0 before entering simulation;
      TP0=0;
      TP1=0;
      TP2=0;

DO J=1 TO &REPEAT;
*** Initialize CNT0, CNT1 & CNT2 to 0;
      CNT0=0;
      CNT1=0;
      CNT2=0;

XB=RANNOR(&RSEED)/SQRT(BG);
SB=SQRT(2*RANGAM(&RSEED,(BG-1)/2)/(BG-1));

PL2=XB+SB*SQRT(1+1/BG)*TINV((1-(1-PL)/2),(BG-1));
PL1=XB+SB*SQRT(1+1/BG)*TINV((1-(1-PL)),(BG-1));
PL0=XB+SB*SQRT(1+1/BG)*TINV((1-(1-TL)),(BG-1));
TLIM=XB+SB*SQRT(1+1/BG)*TINV((1-(1-TL)),(BG-1));

DO K=1 TO WL;
      IF K<WL THEN DO;
      X1=RANNOR(&RSEED);
      X2=RANNOR(&RSEED);
      X3=RANNOR(&RSEED);
      END;
      ELSE DO;
      X1=RANNOR(&RSEED)+DELTA;
      X2=RANNOR(&RSEED)+DELTA;
```

```
      X3=RANNOR(&RSEED)+DELTA;
      END;
      IF X1>TLIM THEN DO;
      CNT0=CNT0+1;
      IF X2>PL1 THEN CNT1=CNT1+1;
      IF X2>PL2 OR X3>PL2 THEN CNT2=CNT2+1;
      END;
END;

IF CNT0>0 THEN TP0=TP0+100/&REPEAT;
IF CNT1>0 THEN TP1=TP1+100/&REPEAT;
IF CNT2>0 THEN TP2=TP2+100/&REPEAT;

*** Print iteration information every 100 iterations;
I=&I;
IF MOD(J,&ITPRINT)=0 THEN
   PUT '>>> CURVE ' I ', ITERATION ' J ', ' BG= ', ' WL= ', ' TL= ', '
      PL= ', ' DELTA= ', ' TP0= ', ' TP1= ', ' TP2= '<<<';
END;
OUTPUT;
END;
RUN;

DATA OUTSAS.PCURVE&I; SET ITERATE(KEEP=BG WL TL PL TP0 TP1 TP2 DELTA);
RUN;

PROC PRINT DATA=OUTSAS.PCURVE&I;
 FORMAT TP0 TP1 TP2 8.4;
 TITLE1"TEST PRINT OF PARAMETRIC SIMULATION PCURVE&I";
 TITLE2"NUMBER OF ITERATIONS = &REPEAT";
RUN;

%MEND PARSIM;
 %MACRO CURVE;
  %DO I=&ISTART %TO &CURVENUM;
   %PARSIM
  %END;
 %MEND CURVE;
%CURVE


//********************************************************************;
//*   DESCRIPTION:  *** NON-PARAMETRIC SIMULATION ***
//*
//*   This program produces power curves for 15 different curve
//*   simulations (refer to the %LET statements below).  Delta ranges
//*   from 0 to 5 by 0.5.  The variable list is as follows for the
//*   input parameters:
//*
//*   BG = Background
//*   WL = Well
//*
//********************************************************************;
//    EXEC SAS
//    OUTSAS DD DSN=XXXXXXX.GWT03000.SJA3092.CURVES,DISP=OLD
//    SYSIN DD *

OPTIONS LS=132 PS=57;
%LET ISTART=1;
%LET CURVENUM=15;
%LET RSEED=3030;
%LET REPEAT=5000;
%LET ITPRINT=1000;
```

```
%LET BG1 =8;       %LET WL1 =5;
%LET BG2 =16;      %LET WL2 =5;
%LET BG3 =24;      %LET WL3 =5;
%LET BG4 =8;       %LET WL4 =20;
%LET BG5 =16;      %LET WL5 =20;
%LET BG6 =24;      %LET WL6 =20;
%LET BG7 =8;       %LET WL7 =50;
%LET BG8 =16;      %LET WL8 =50;
%LET BG9 =24;      %LET WL9 =50;
%LET BG10=8;       %LET WL10=100;
%LET BG11=16;      %LET WL11=100;
%LET BG12=24;      %LET WL12=100;
%LET BG13=32;      %LET WL13=100;
%LET BG14=32;      %LET WL14=20;
%LET BG15=32;      %LET WL15=50;

%MACRO NPARSIM;
DATA ITERATE;
 *** Set changing simulation variable to common variable names;
 BG=&&BG&I;
 WL=&&WL&I;

 DO DELTA=0 TO 5 BY 0.5;
      *** Initialize PLx variables to 0 before entering simulation;
      PL0=0;
      PL1=0;
      PL2A=0;
      PL2B=0;
      PL3A=0;
      PL3B=0;

 DO J=1 TO &REPEAT;
      *** Initialize CNTx variables to 0;
      CNT0=0;
      CNT1=0;
      CNT2=0;
      CNT3=0;
      CNT4=0;
      CNT5=0;

 DO K=1 TO BG;
      TEST=RANNOR(&RSEED);
      IF K=1 THEN MAX=TEST;
       ELSE IF TEST>MAX THEN MAX=TEST;
 END;

 DO L=1 TO WL;
      IF L<WL THEN DO;
      X1=RANNOR(&RSEED);
      X2=RANNOR(&RSEED);
      X3=RANNOR(&RSEED);
      X4=RANNOR(&RSEED);
      END;
      ELSE DO;
      X1=RANNOR(&RSEED)+DELTA;
      X2=RANNOR(&RSEED)+DELTA;
      X3=RANNOR(&RSEED)+DELTA;
      X4=RANNOR(&RSEED)+DELTA;
END;
IF X1>MAX THEN DO;
      CNT0=CNT0+1;
      IF X2>MAX THEN CNT1=CNT1+1;
```

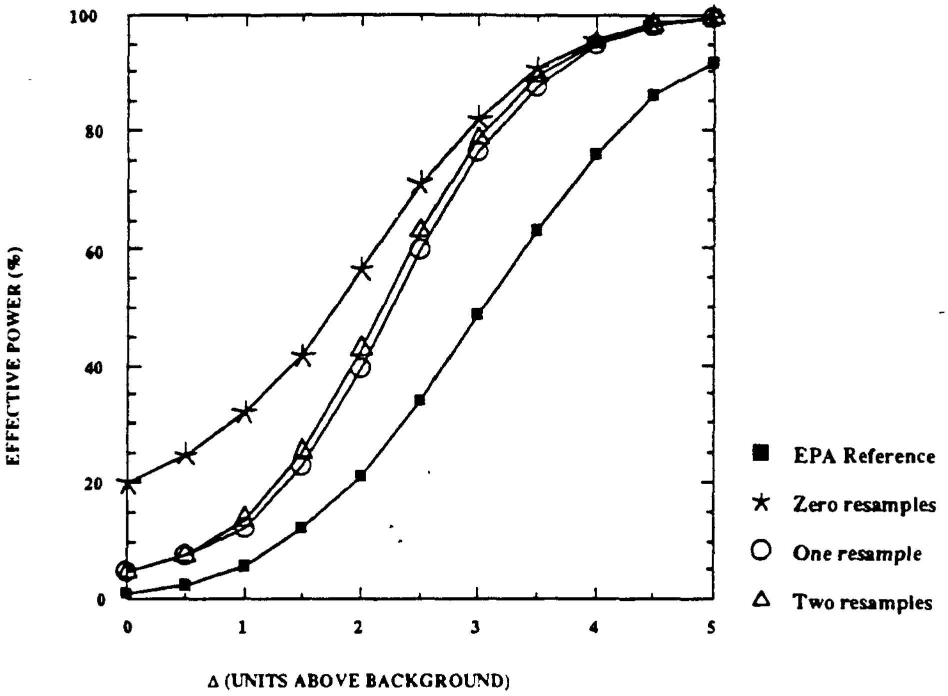
```
       IF X2>MAX & X3>MAX THEN CNT2=CNT2+1;
       IF X2>MAX OR X3>MAX THEN CNT3=CNT3+1;
       IF X2>MAX & X3>MAX & X4>MAX THEN CNT4=CNT4+1;
       IF X2>MAX_OR X3>MAX OR X4>MAX THEN CNT5=CNT5+1;
END;

IF CNT0>0 THEN PL0=PL0+100/&REPEAT;
IF CNT1>0 THEN PL1=PL1+100/&REPEAT;
IF CNT2>0 THEN PL2A=PL2A+100/&REPEAT;
IF CNT3>0 THEN PL2B=PL2B+100/&REPEAT;
IF CNT4>0 THEN PL3A=PL3A+100/&REPEAT;
IF CNT5>0 THEN PL3B=PL3B+100/&REPEAT;

*** Print iteration information every X iterations;
I=&I;
IF MOD(J,&ITPRINT)=0 THEN
 PUT '>>> CURVE ' I ', ITERATION ' J ', ' BG= ', ' WL= ', ' DELTA=
      ', ' PL0= ', ' PL1= ', ' PL2A= ', ' PL2B= ', ' PL3A= ', ' PL3B= '<<<';
END;
OUTPUT;
END;
RUN;

DATA OUTSAS.NCURVE&I; SET ITERATE(KEEP=BG WL PL0 PL1 PL2A PL2B PL3A PL3B DELTA);
RUN;

PROC PRINT DATA=OUTSAS.NCURVE&I;
 FORMAT PL0 PL1 PL2A PL2B PL3A PL3B 8.4;
 TITLE1"TEST PRINT OF NON-PARAMETRIC SIMULATION NCURVE&I";
 TITLE2"NUMBER OF ITERATIONS = &REPEAT";
RUN;

%MEND NPARSIM;
 %MACRO CURVE;
  %DO I=&ISTART %TO &CURVENUM;
   %NPARSIM
  %END;
 %MEND CURVE;
%CURVE
```

EPA REFERENCE POWER CURVES



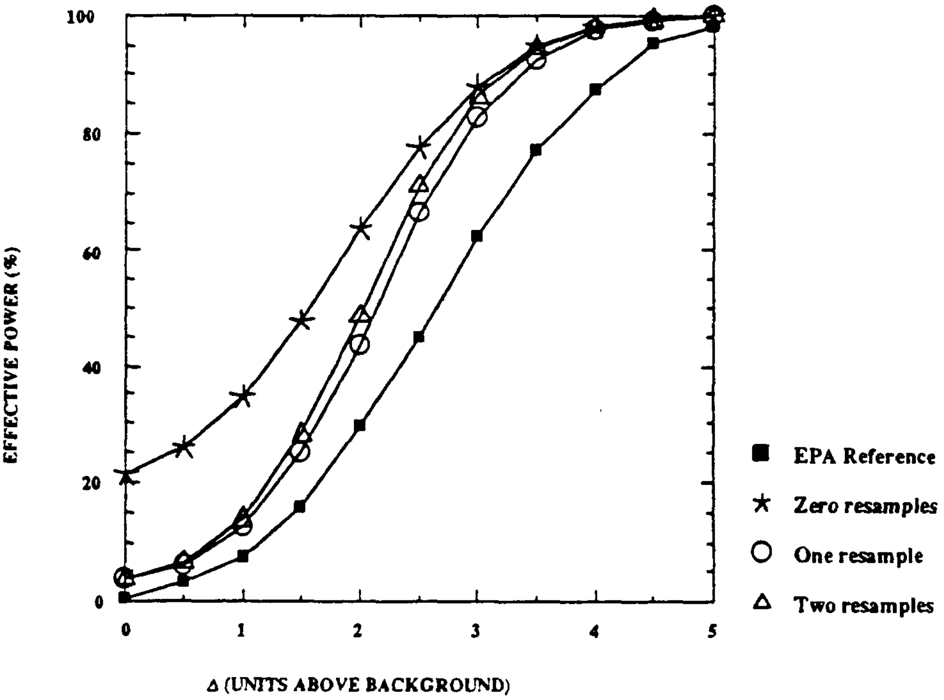
POWER CURVE FOR 95% TOLERANCE AND 90% PREDICTION LIMIT

(8 Background Samples; 5 wells)



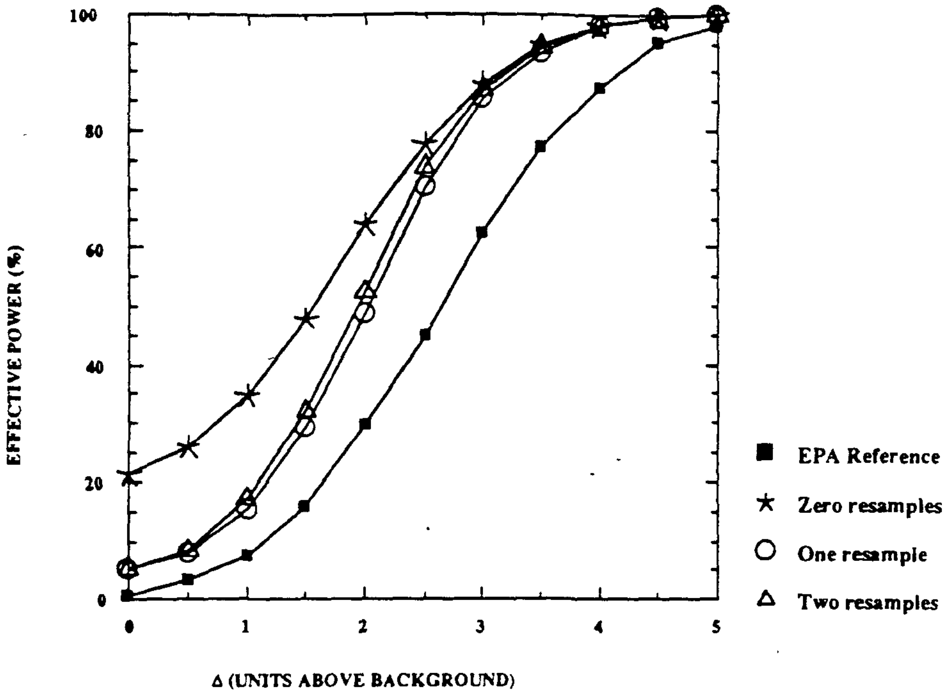
POWER CURVE FOR 95% TOLERANCE AND 90% PREDICTION LIMIT

(16 Background Samples; 5 wells)



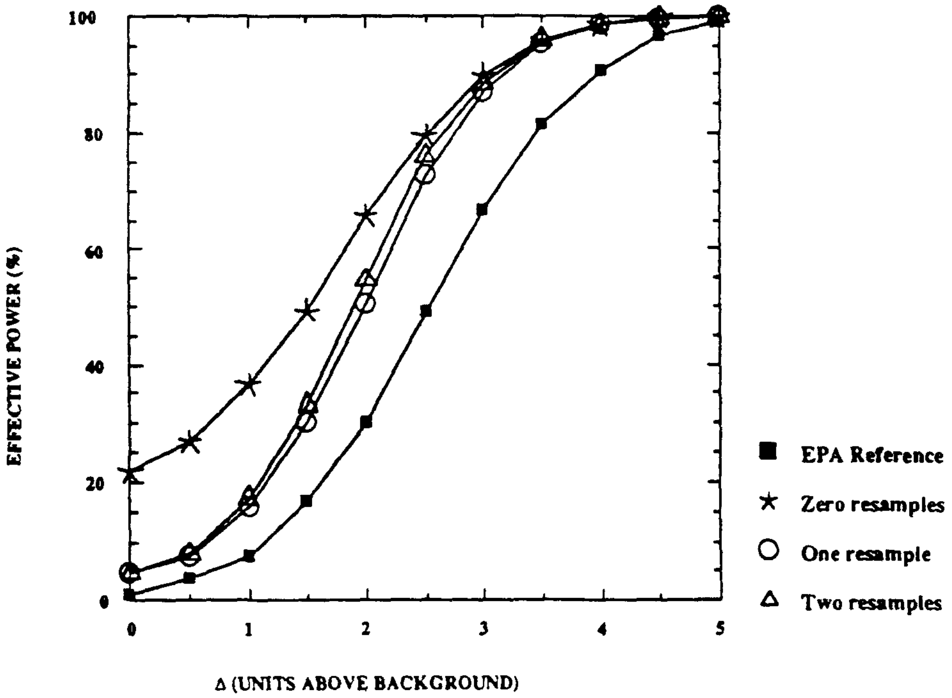
POWER CURVE FOR 95% TOLERANCE AND 85% PREDICTION LIMIT

(16 Background Samples; 5 wells)



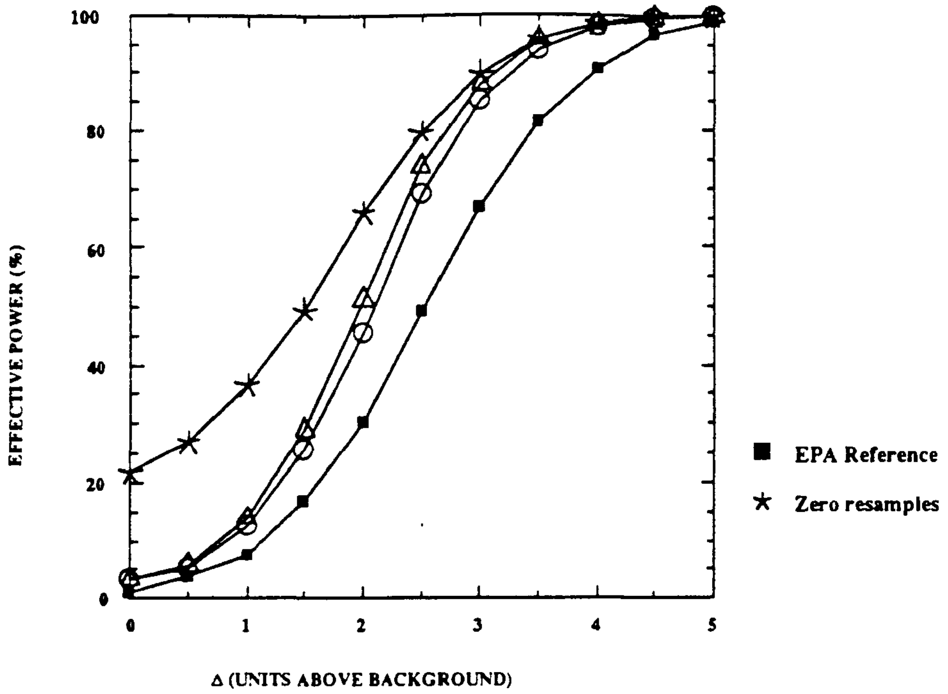
POWER CURVE FOR 95% TOLERANCE AND 85% PREDICTION LIMIT

(24 Background Samples; 5 wells)



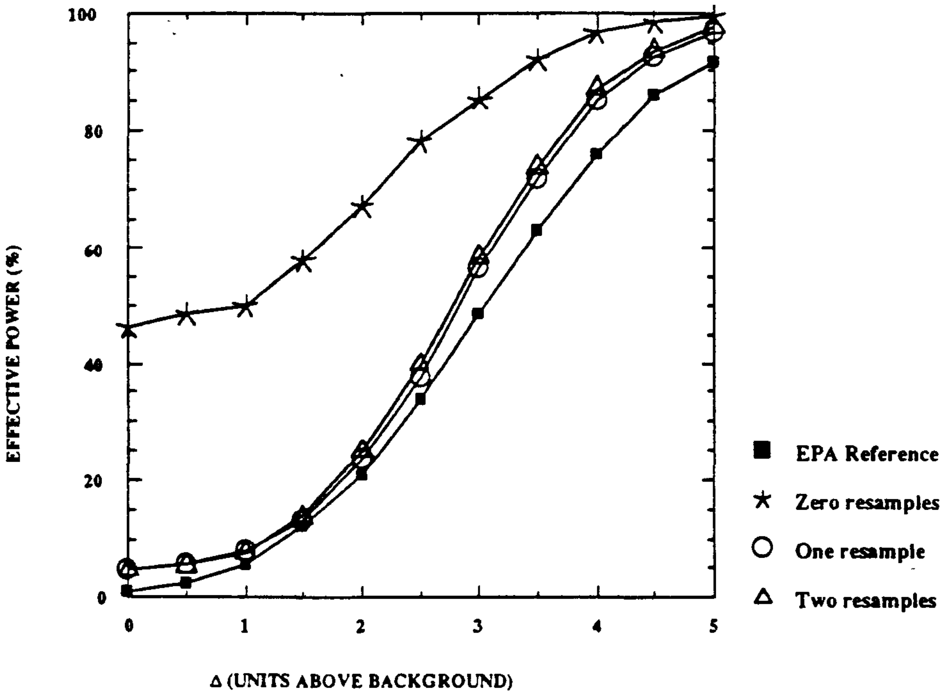
POWER CURVE FOR 95% TOLERANCE AND 90% PREDICTION LIMIT

(24 Background Samples; 5 wells)



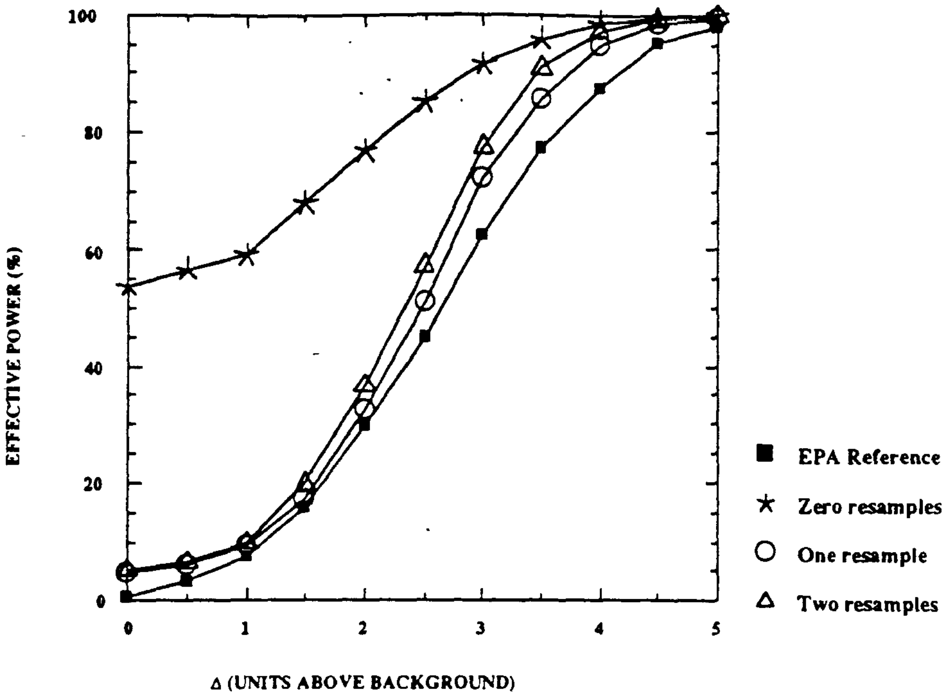
POWER CURVE FOR 95% TOLERANCE AND 98% PREDICTION LIMIT

(8 Background Samples; 20 wells)



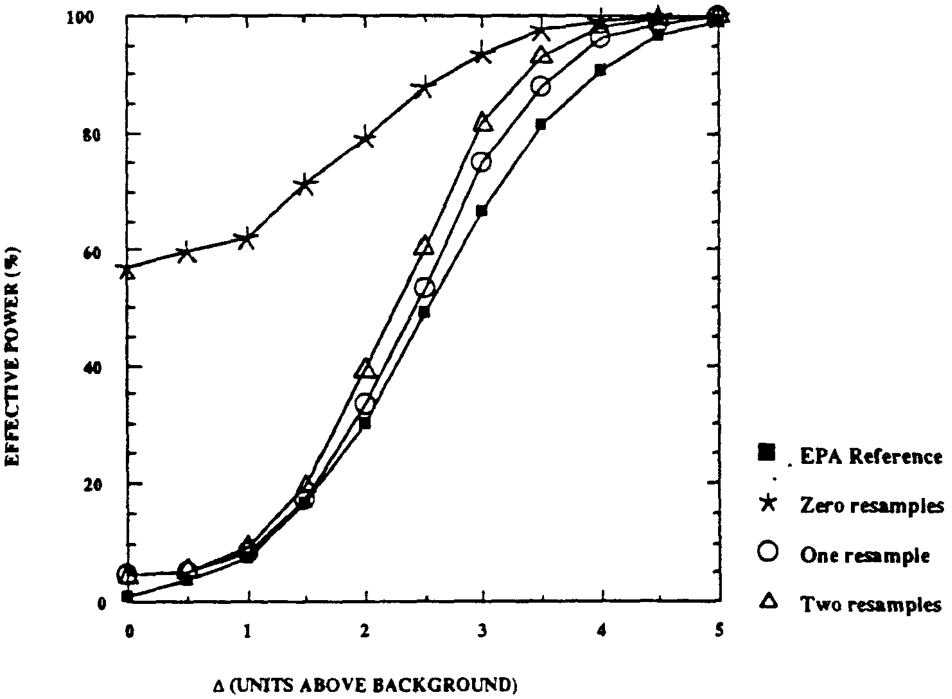
POWER CURVE FOR 95% TOLERANCE AND 97% PREDICTION LIMIT

(16 Background Samples; 20 wells)



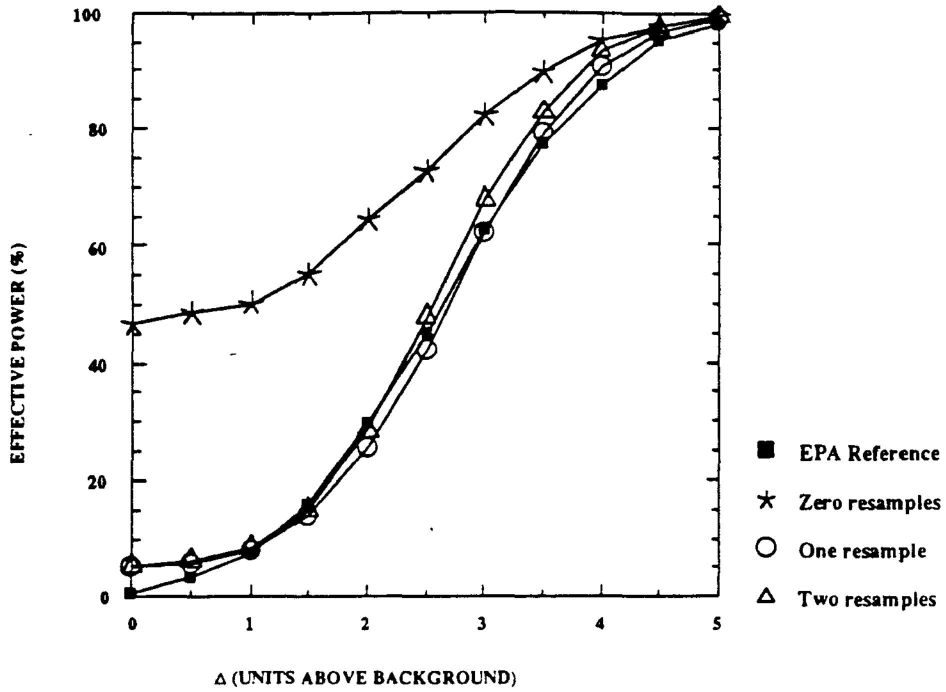
POWER CURVE FOR 95% TOLERANCE AND 97% PREDICTION LIMIT

(24 Background Samples; 20 wells)



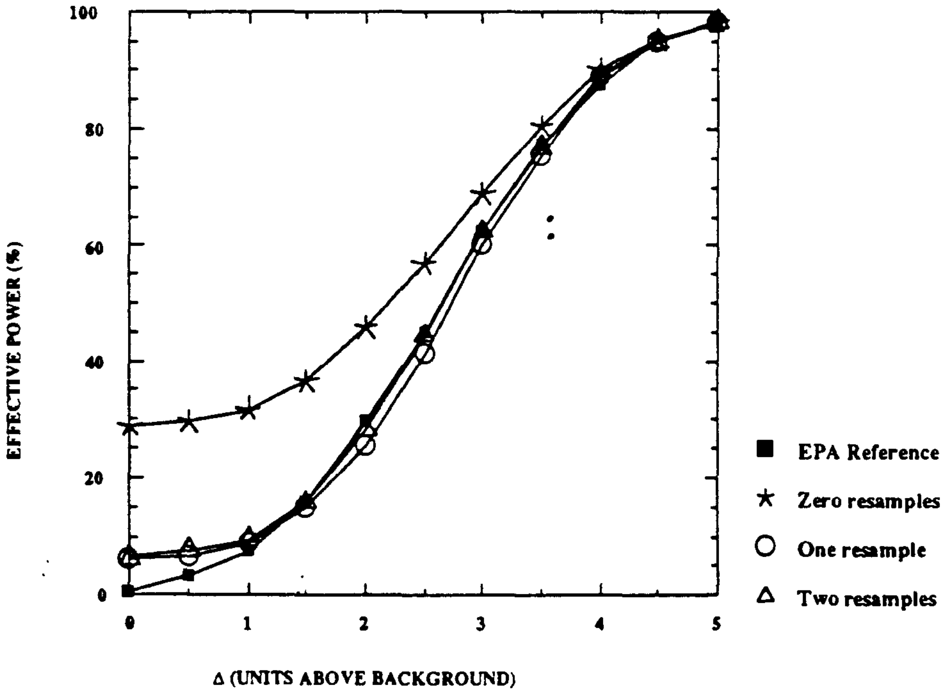
POWER CURVE FOR 98% TOLERANCE AND 97% PREDICTION LIMIT

(16 Background Samples; 50 wells)



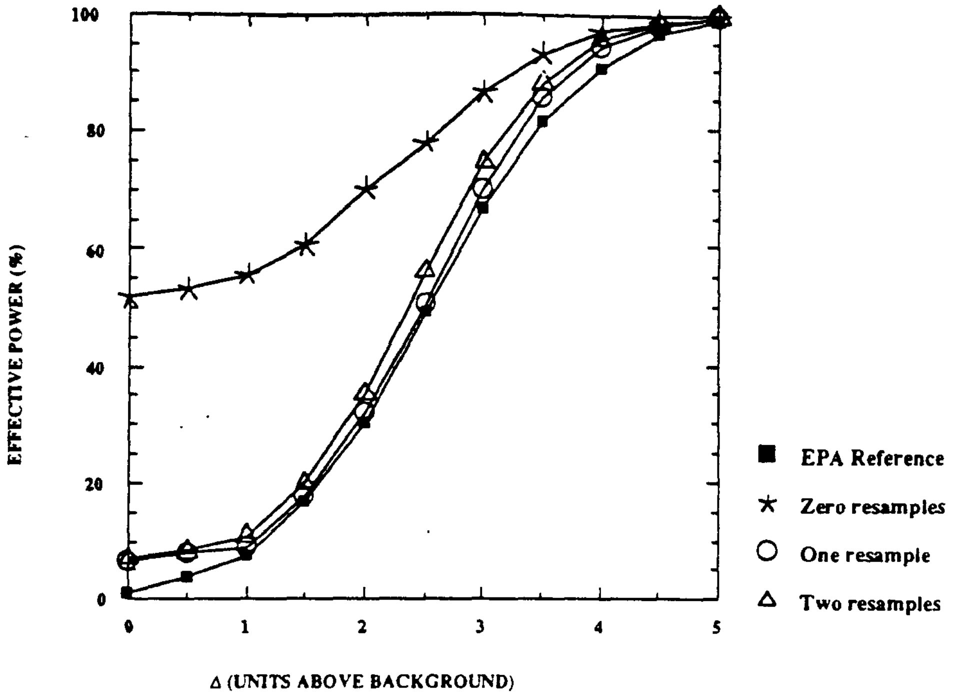
POWER CURVE FOR 99% TOLERANCE AND 92% PREDICTION LIMIT

(16 Background Samples; 50 wells)



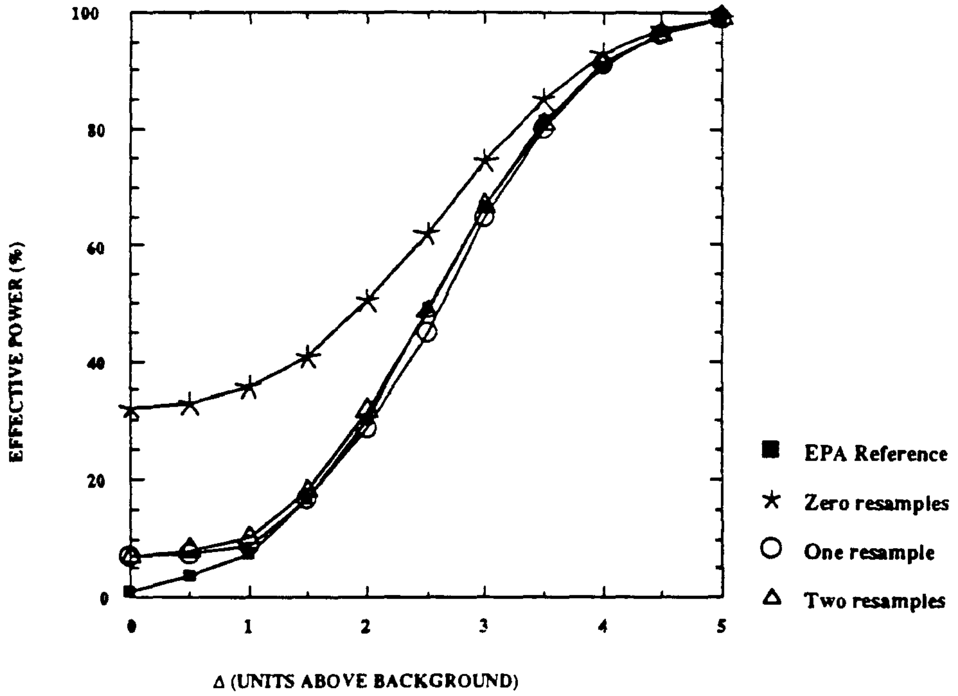
POWER CURVE FOR 98% TOLERANCE AND 95% PREDICTION LIMIT

(24 Background Samples; 50 wells)



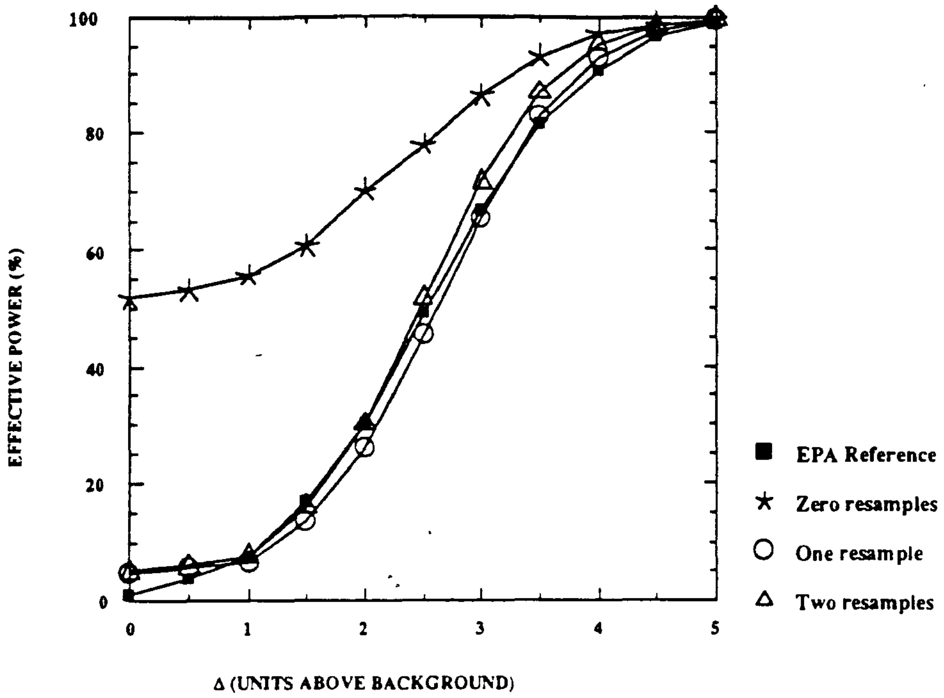
POWER CURVE FOR 99% TOLERANCE AND 90% PREDICTION LIMIT

(24 Background Samples; 50 wells)



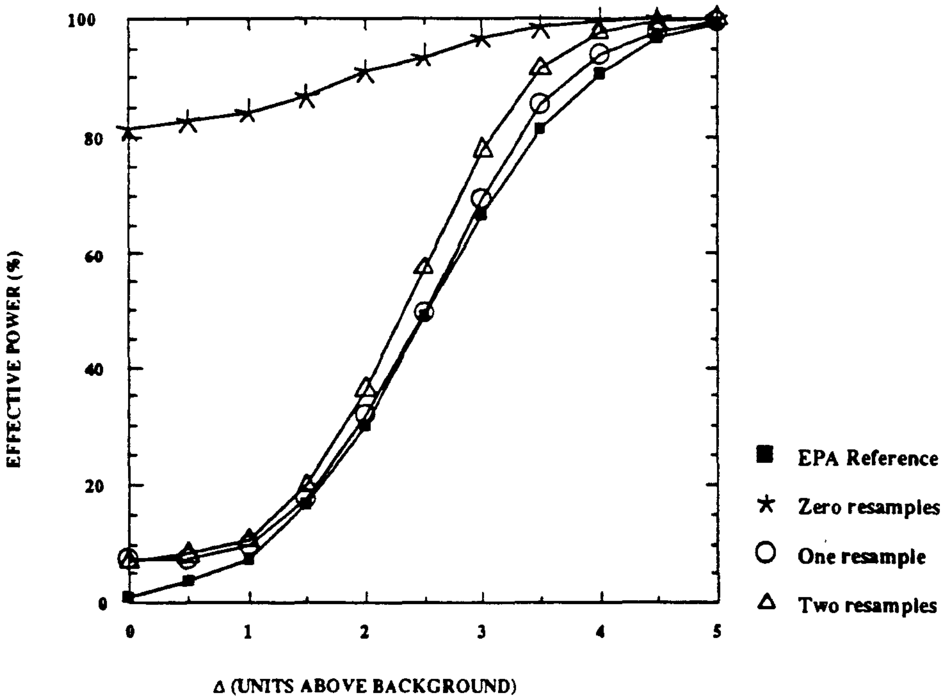
POWER CURVE FOR 98% TOLERANCE AND 97% PREDICTION LIMIT

(24 Background Samples; 50 wells)



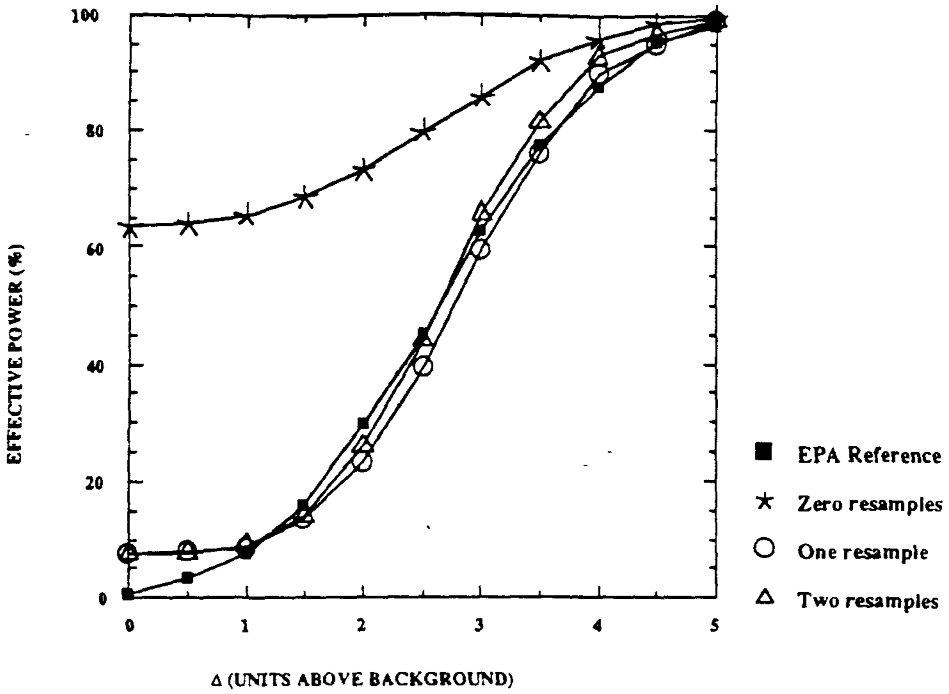
POWER CURVE FOR 95% TOLERANCE AND 98% PREDICTION LIMIT

(24 Background Samples; 50 wells)



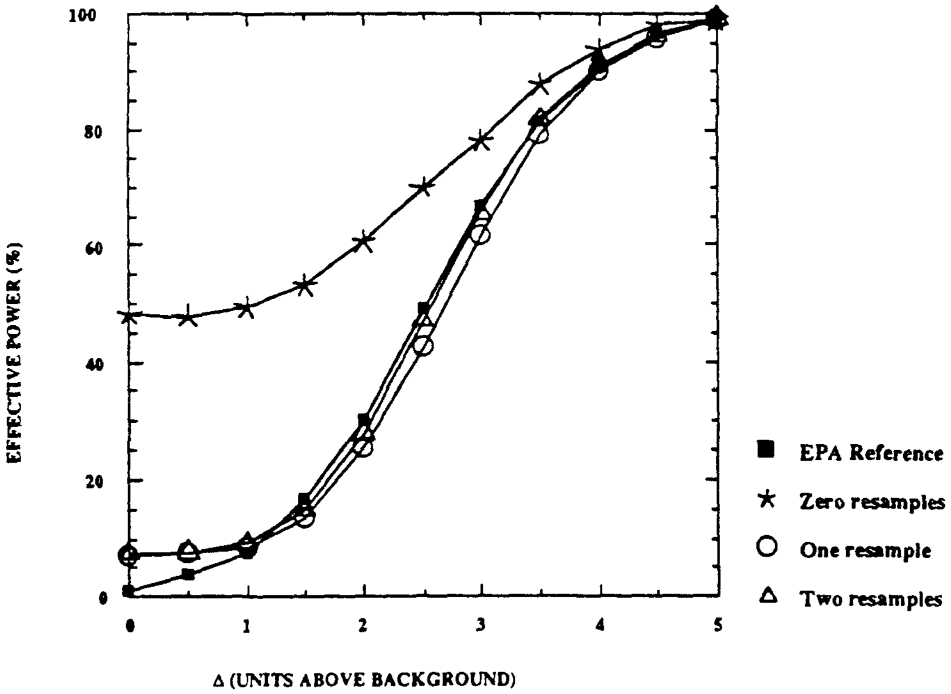
POWER CURVE FOR 98% TOLERANCE AND 98% PREDICTION LIMIT

(16 Background Samples; 100 wells)



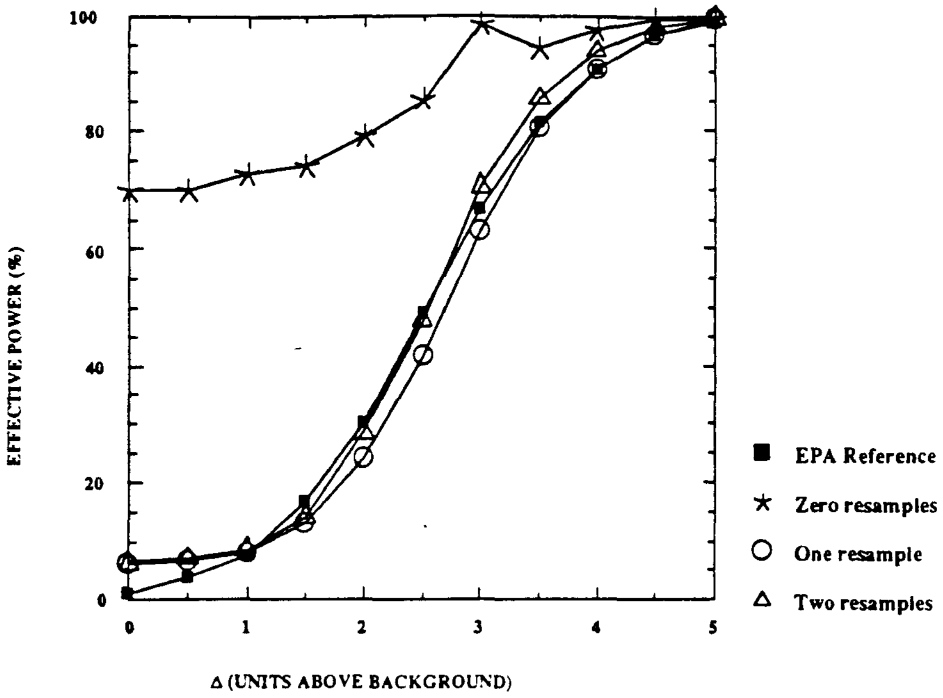
POWER CURVE FOR 99% TOLERANCE AND 95% PREDICTION LIMIT

(24 Background Samples; 100 wells)

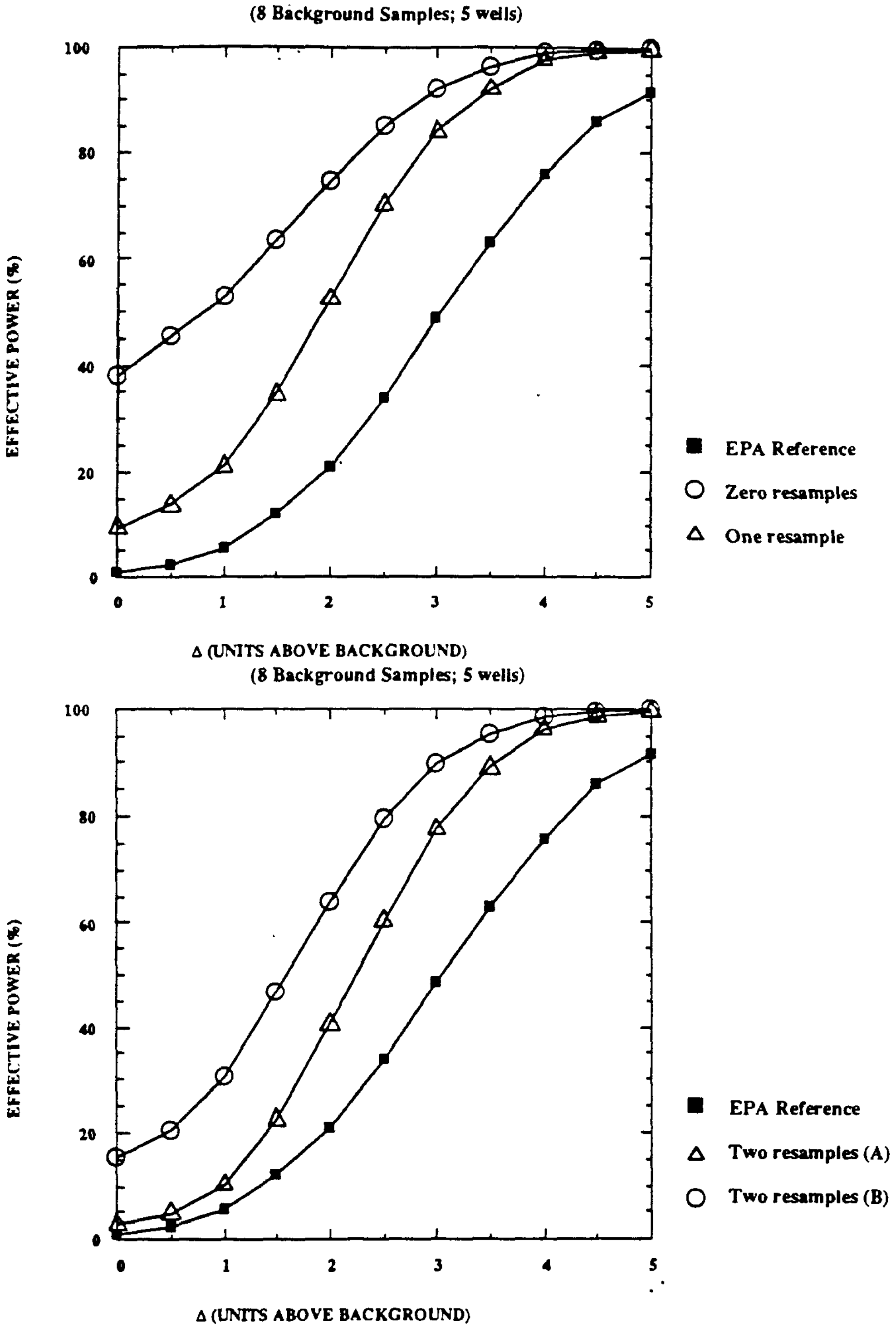


POWER CURVE FOR 98% TOLERANCE AND 98% PREDICTION LIMIT

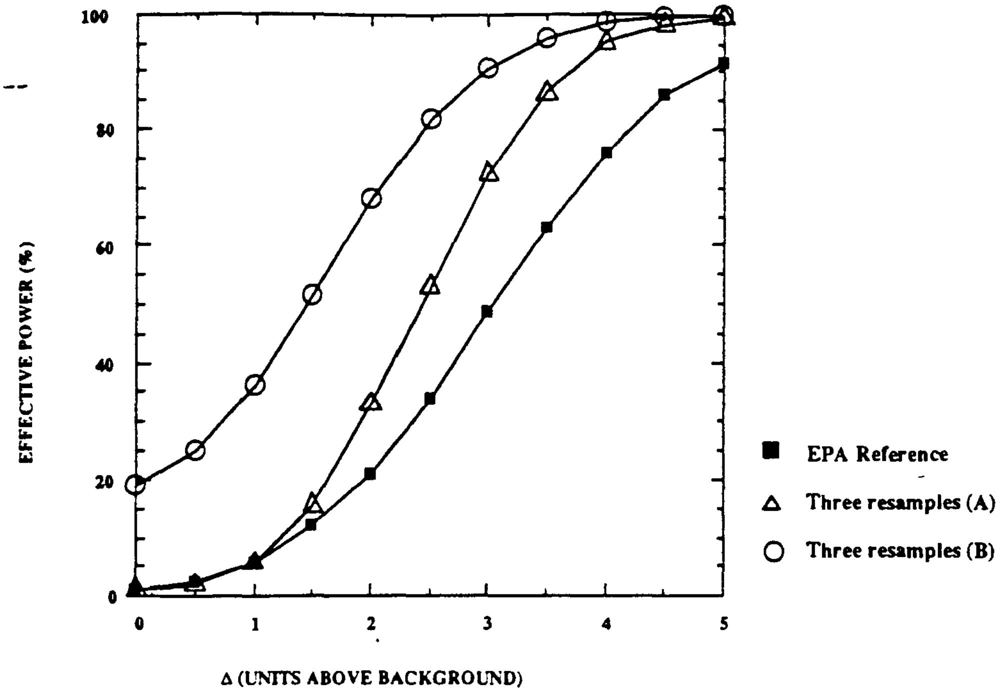
(24 Background Samples; 100 wells)



POWER CURVE FOR NON-PARAMETRIC PREDICTION LIMITS

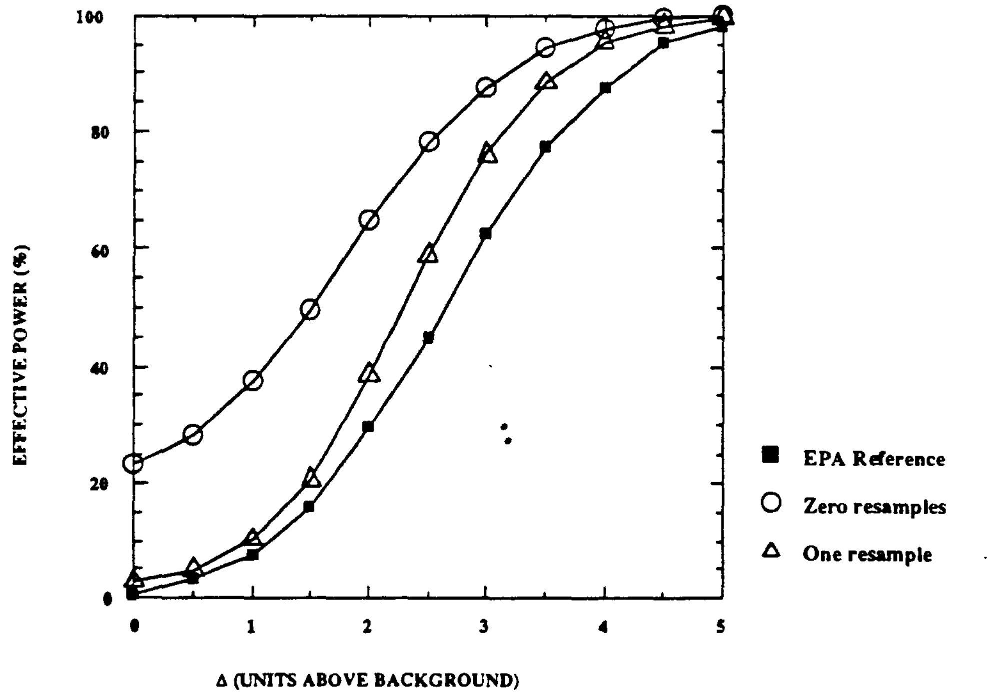


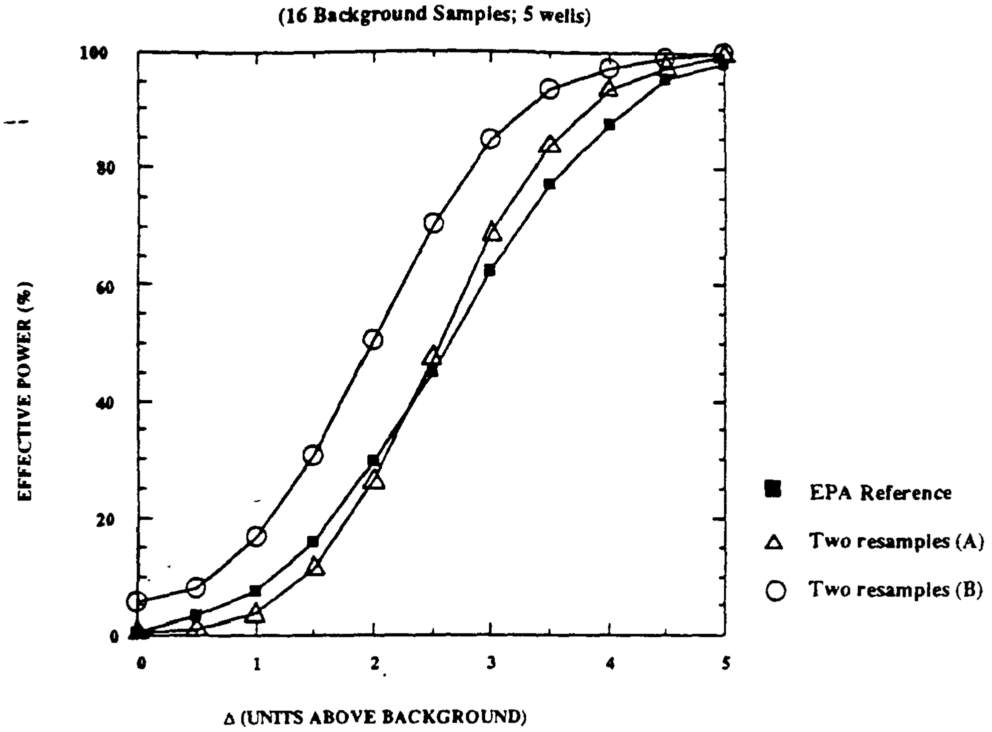
(8 Background Samples; 5 wells)

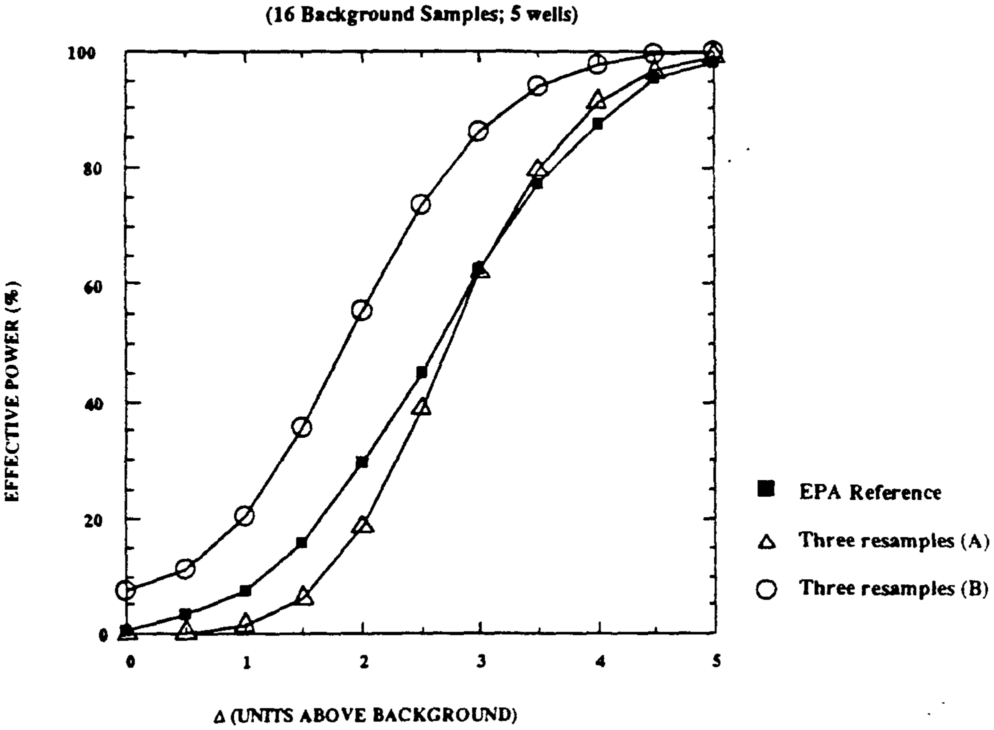


POWER CURVE FOR NON-PARAMETRIC PREDICTION LIMITS

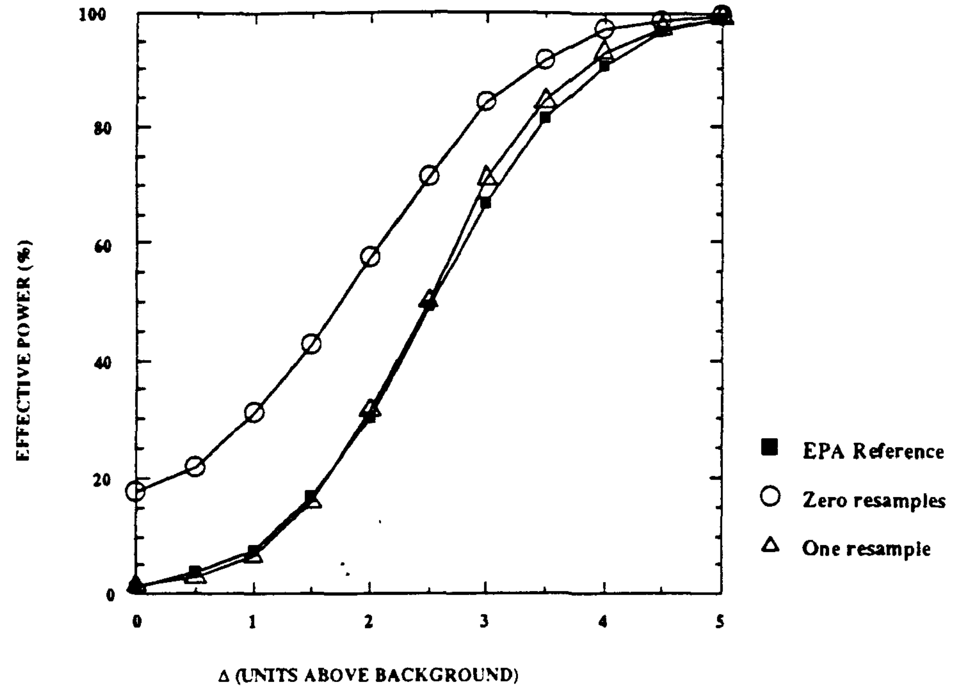
(16 Background Samples; 5 wells)

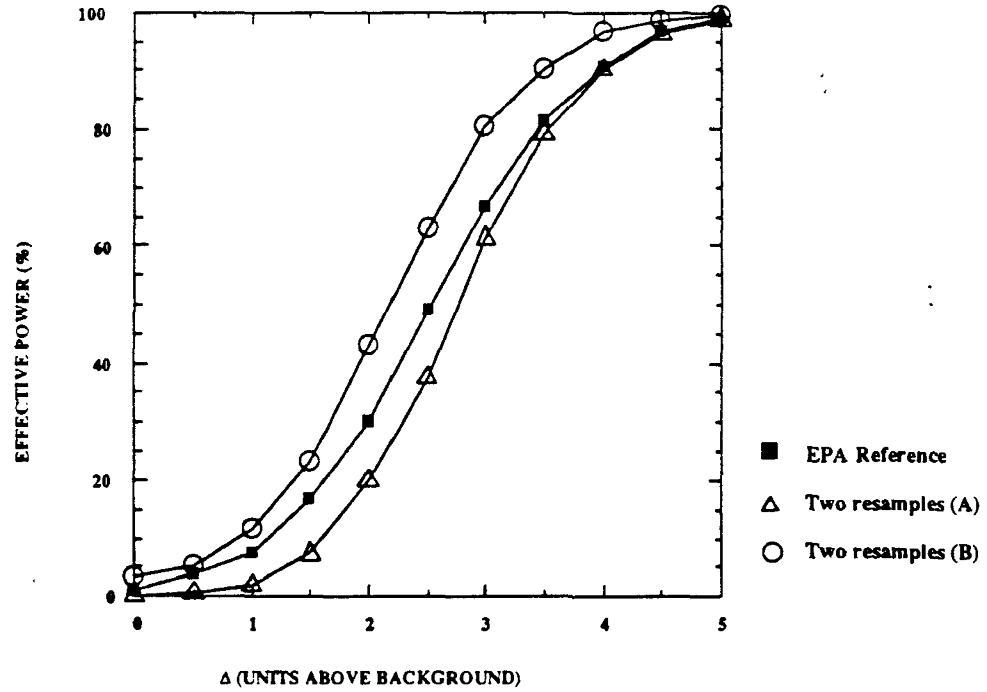



(16 Background Samples; 5 wells)
EFFECTIVE POWER (%)
100
80
60
40
20
0
0
1
2
3
4
5
Δ (UNITS ABOVE BACKGROUND)
■ EPA Reference
△ Two resamples (A)
○ Two resamples (B)

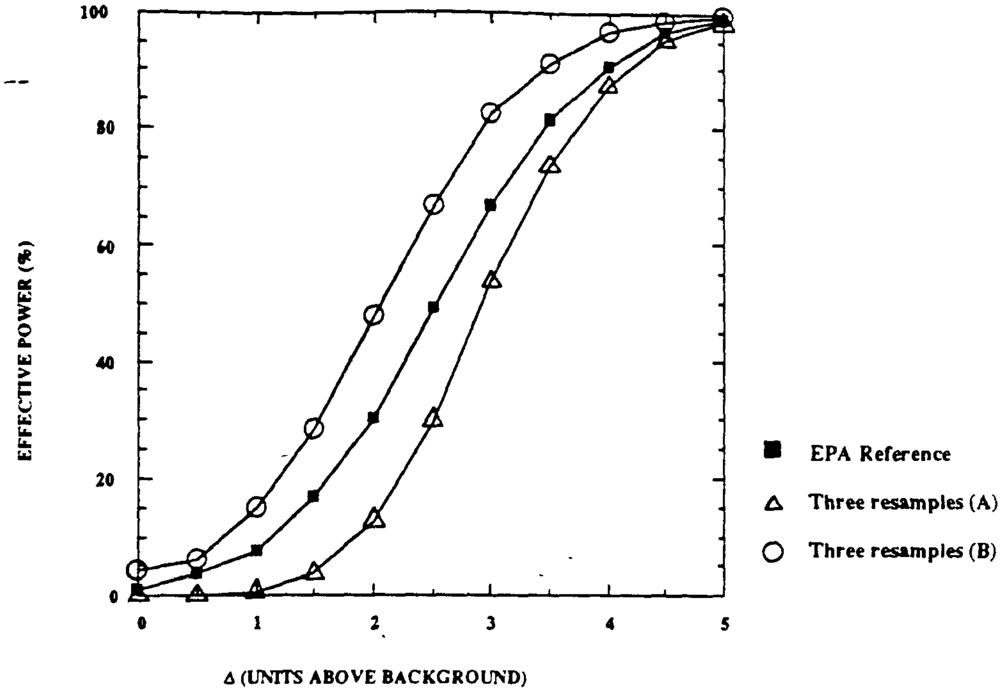

(16 Background Samples; 5 wells)
EFFECTIVE POWER (%)
100
80
60
40
20
0
0
1
2
3
4
5
Δ (UNITS ABOVE BACKGROUND)
■ EPA Reference
△ Three resamples (A)
○ Three resamples (B)

POWER CURVE FOR NON-PARAMETRIC PREDICTION LIMITS


(24 Background Samples; 5 wells)

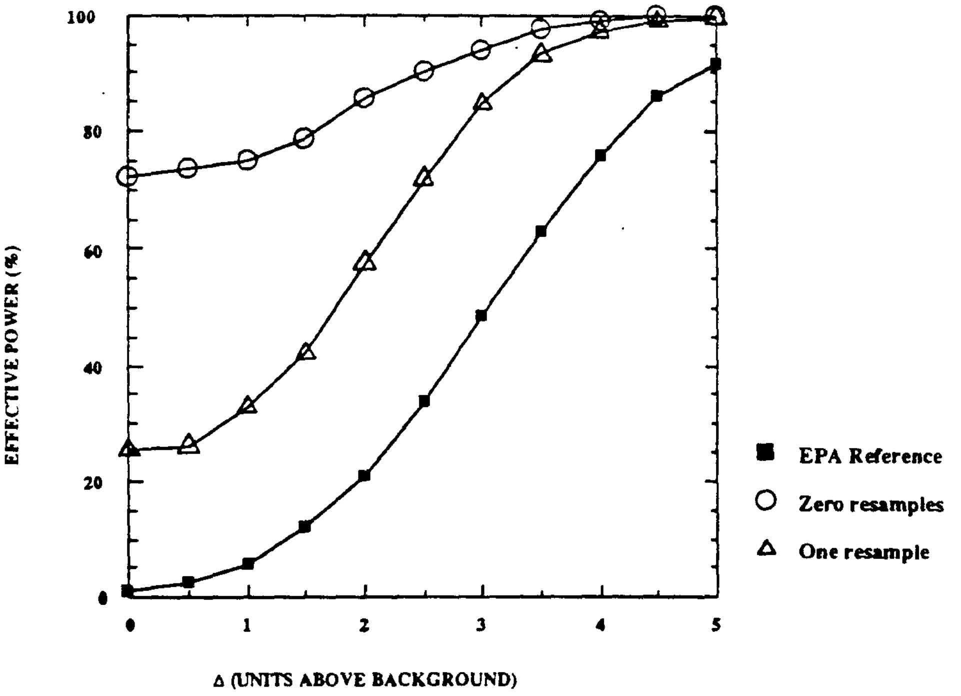

(24 Background Samples; 5 wells)

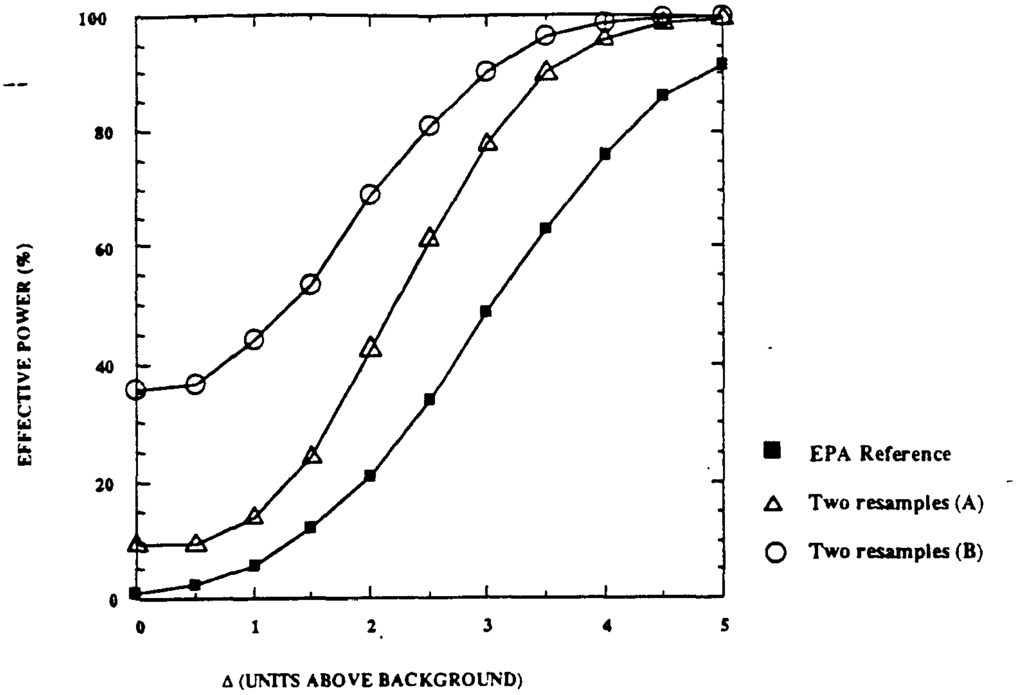
(24 Background Samples; 5 wells)

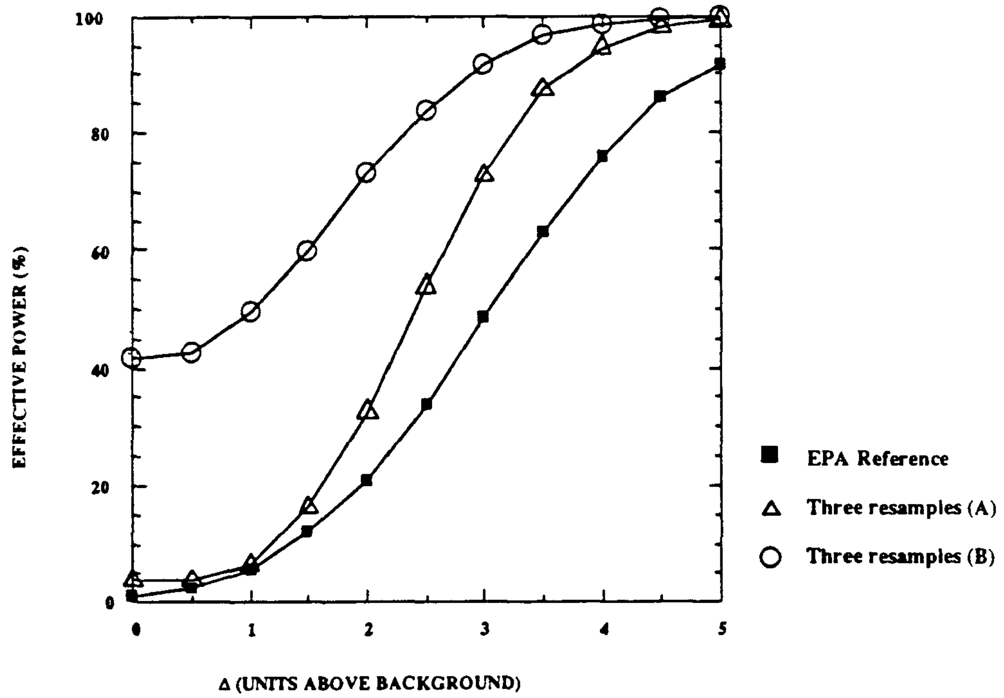


POWER CURVE FOR NON-PARAMETRIC PREDICTION LIMITS

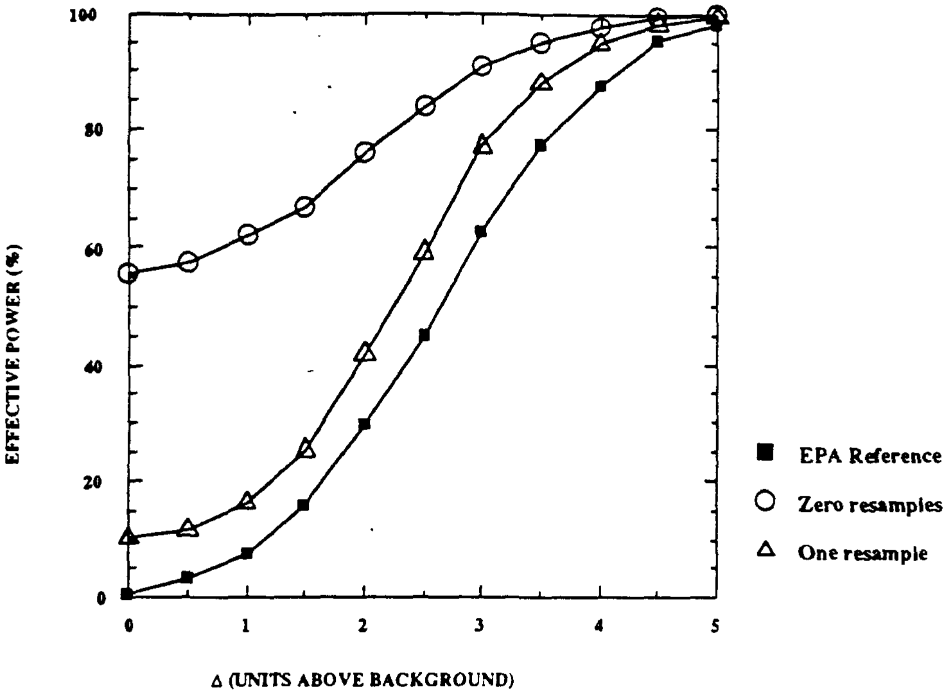
(8 Background Samples; 20 wells)

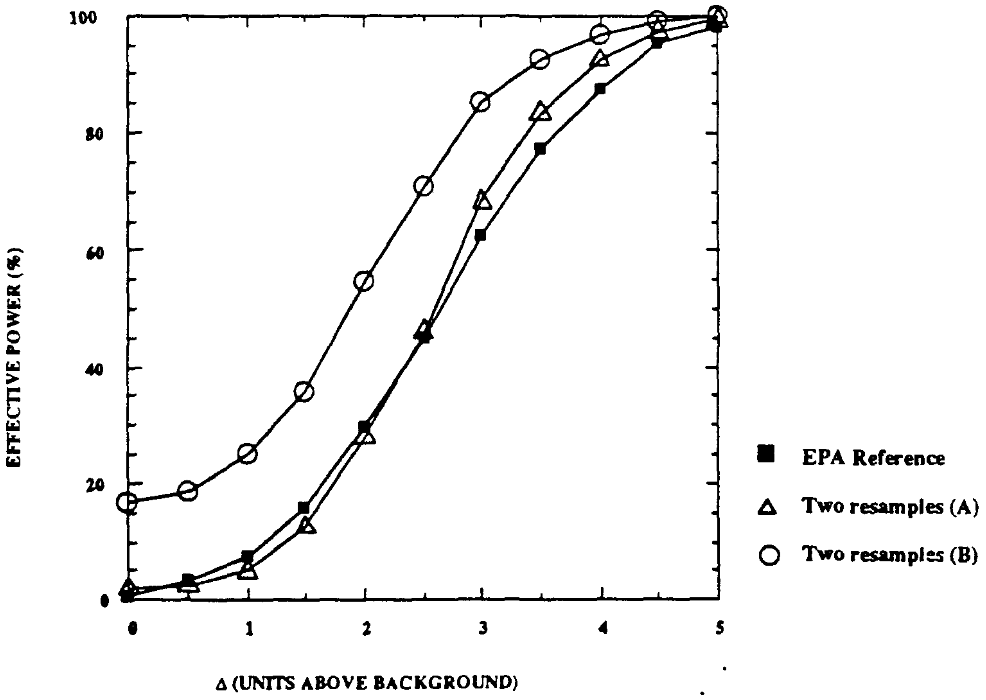



(8 Background Samples; 20 wells)

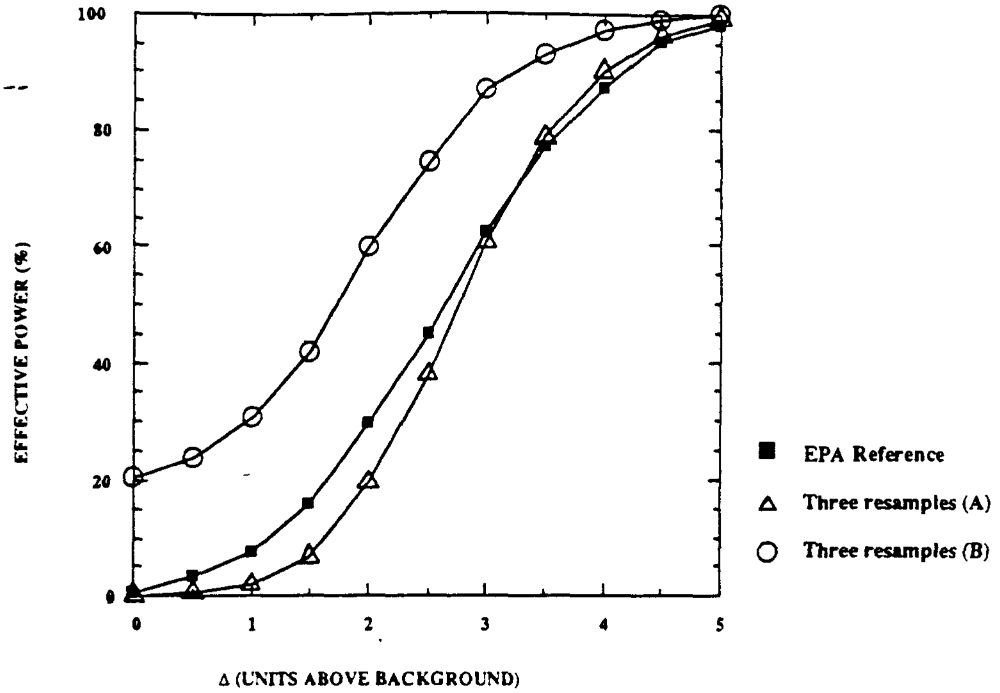

(8 Background Samples; 20 wells)

POWER CURVE FOR NON-PARAMETRIC PREDICTION LIMITS


(16 Background Samples; 20 wells)

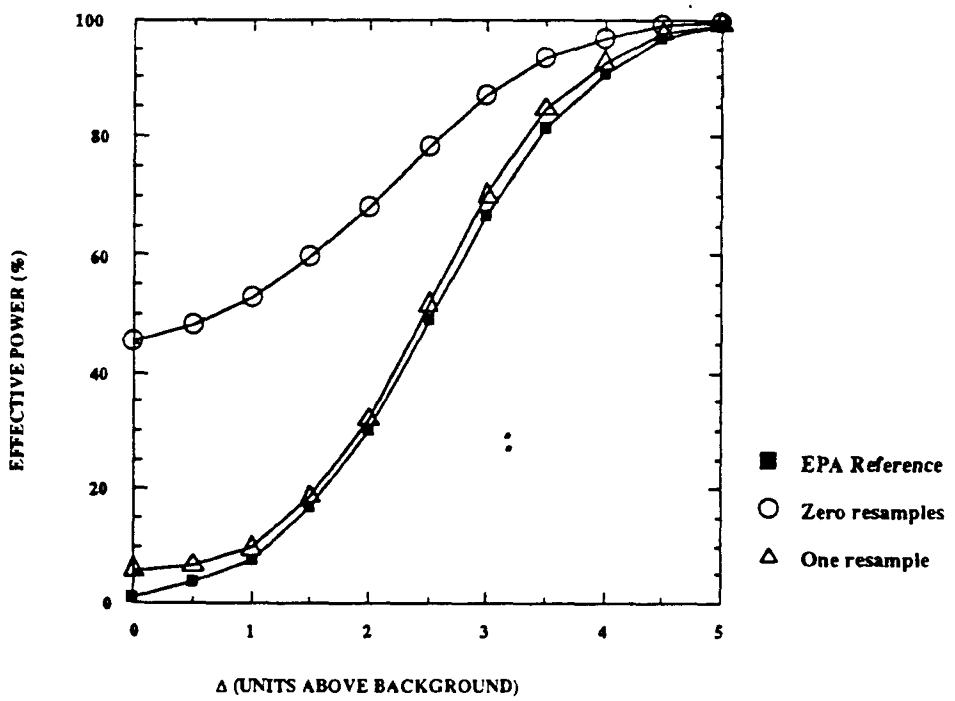

(16 Background Samples; 20 wells)

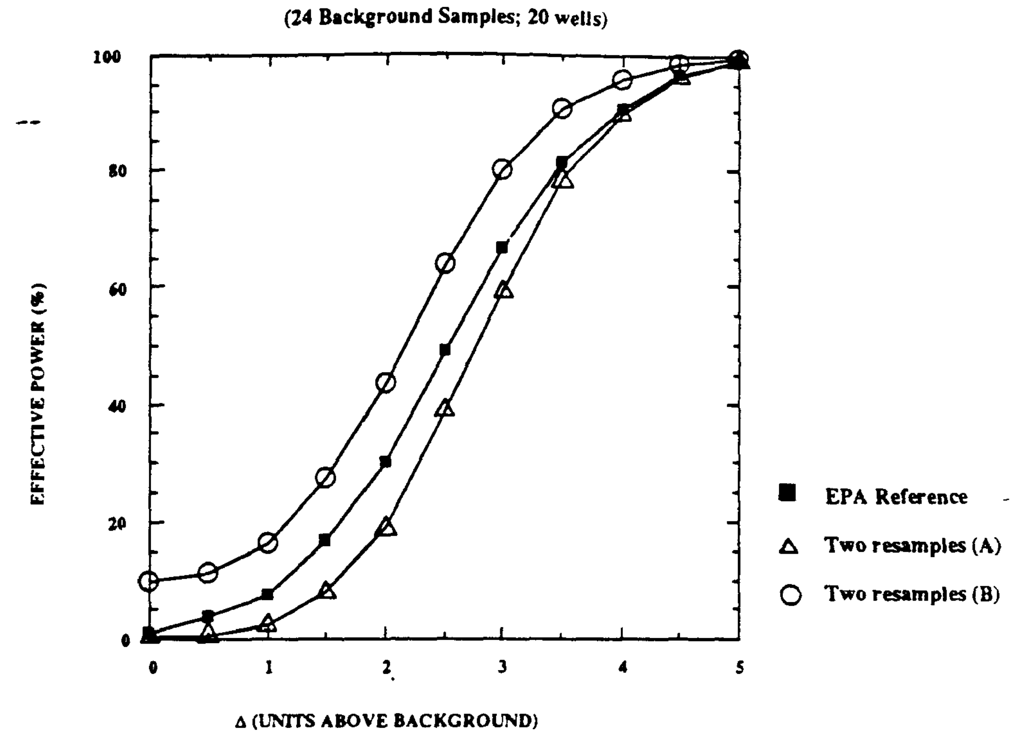
(16 Background Samples; 20 wells)

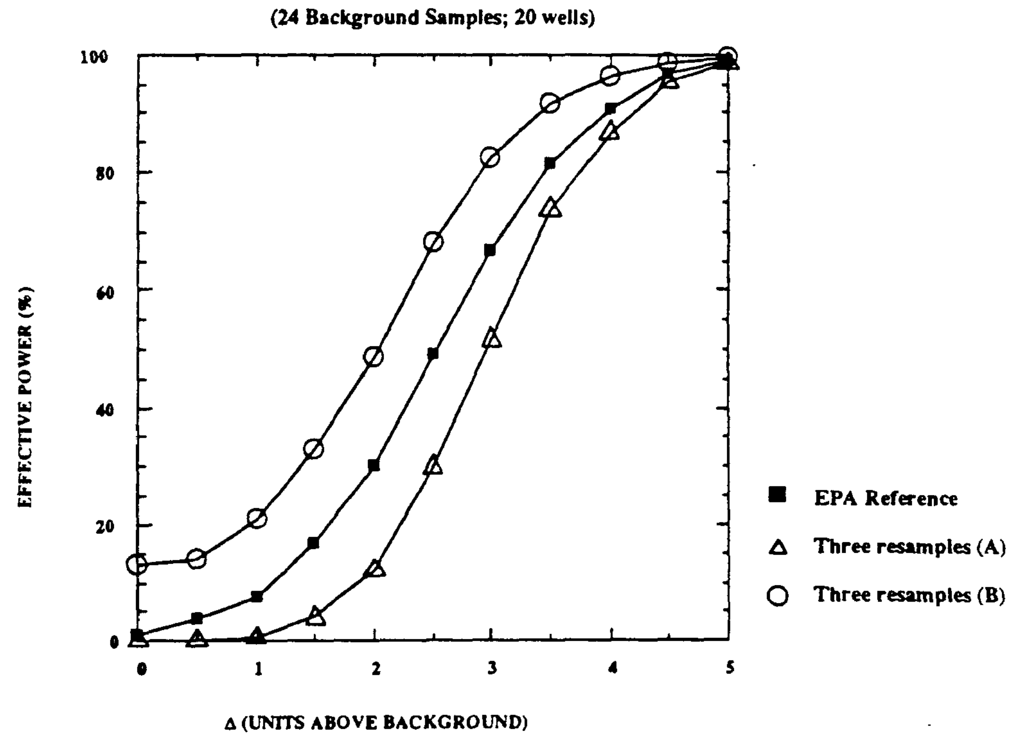


POWER CURVE FOR NON-PARAMETRIC PREDICTION LIMITS

(24 Background Samples; 20 wells)

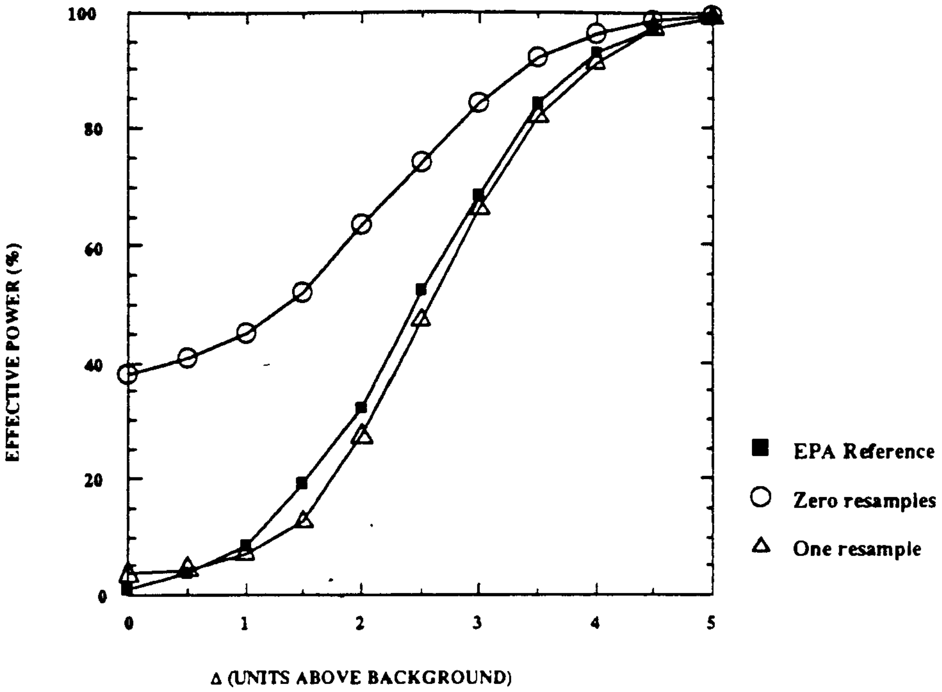



(24 Background Samples; 20 wells)
EFFECTIVE POWER (%)
100
80
60
40
20
0
0
1
2
3
4
5
Δ (UNITS ABOVE BACKGROUND)
■ EPA Reference
△ Two resamples (A)
○ Two resamples (B)

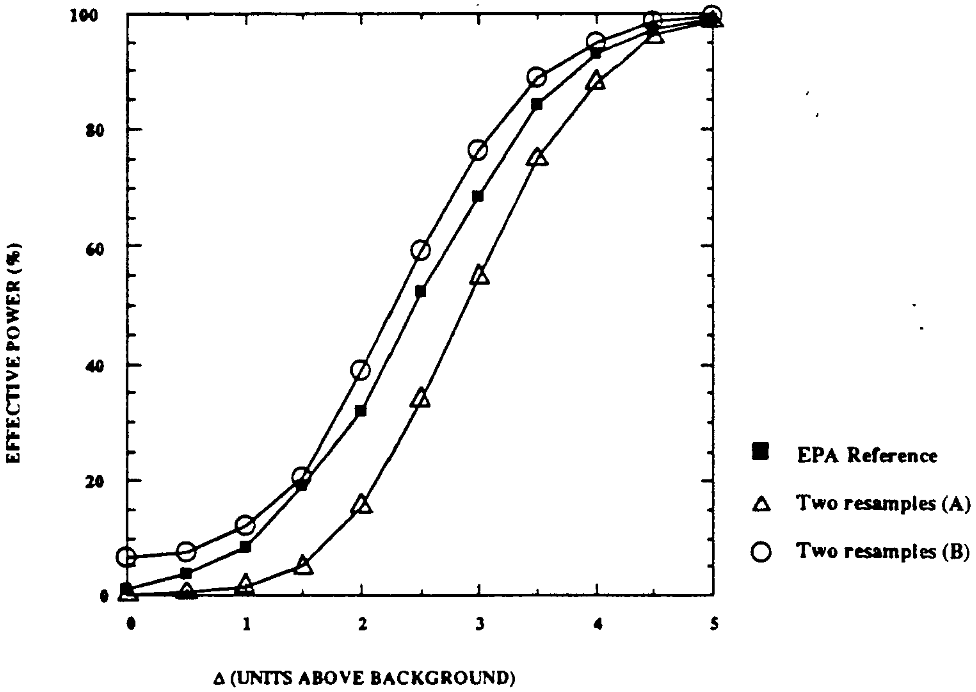

(24 Background Samples; 20 wells)
EFFECTIVE POWER (%)
100
80
60
40
20
0
0
1
2
3
4
5
Δ (UNITS ABOVE BACKGROUND)
■ EPA Reference
△ Three resamples (A)
○ Three resamples (B)

POWER CURVE FOR NON-PARAMETRIC PREDICTION LIMITS

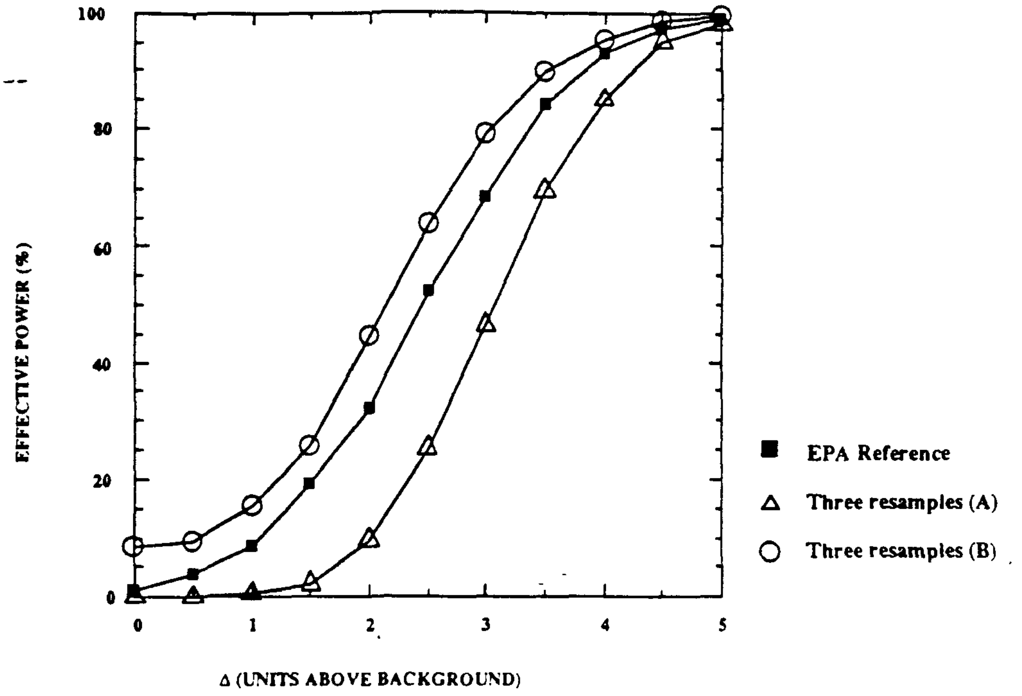
(32 Background Samples; 20 wells)



(32 Background Samples; 20 wells)

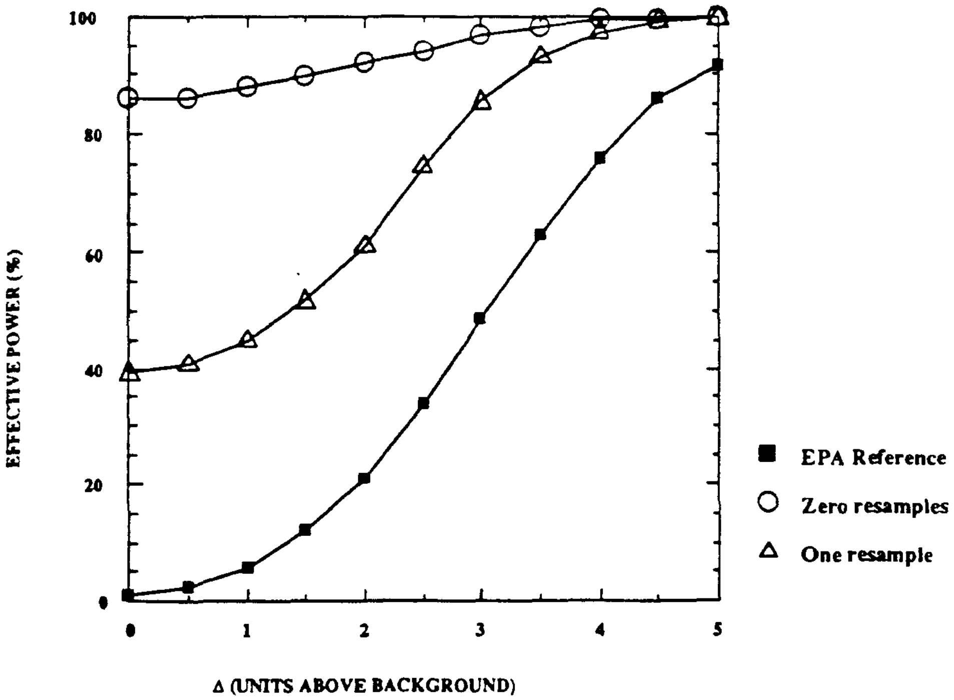


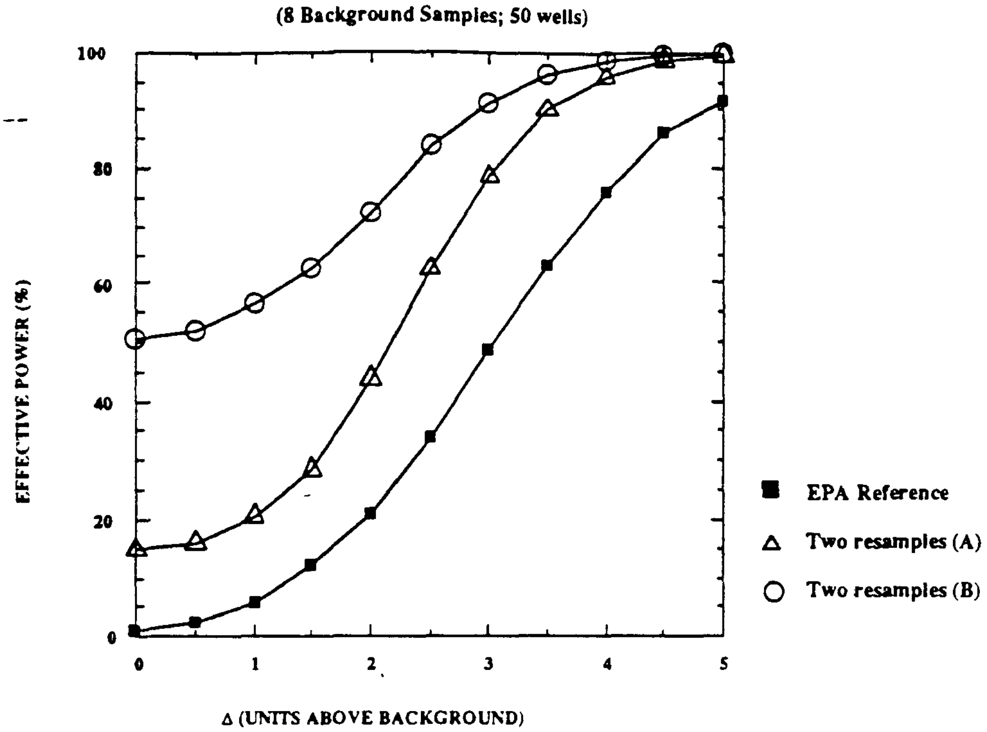
(32 Background Samples; 20 wells)

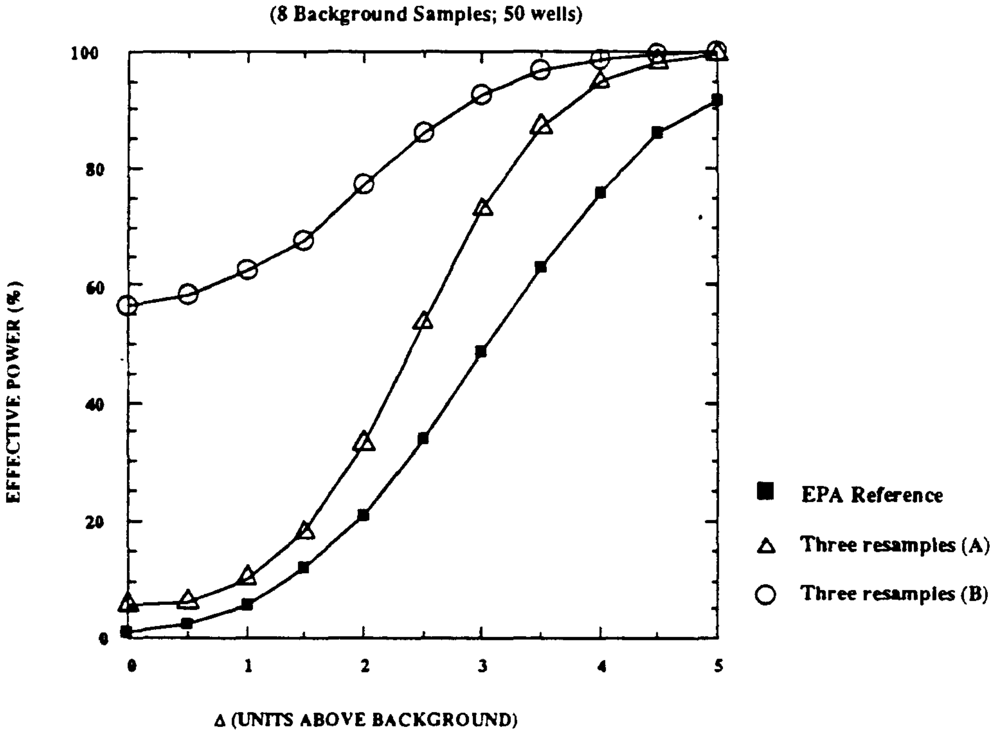


POWER CURVE FOR NON-PARAMETRIC PREDICTION LIMITS

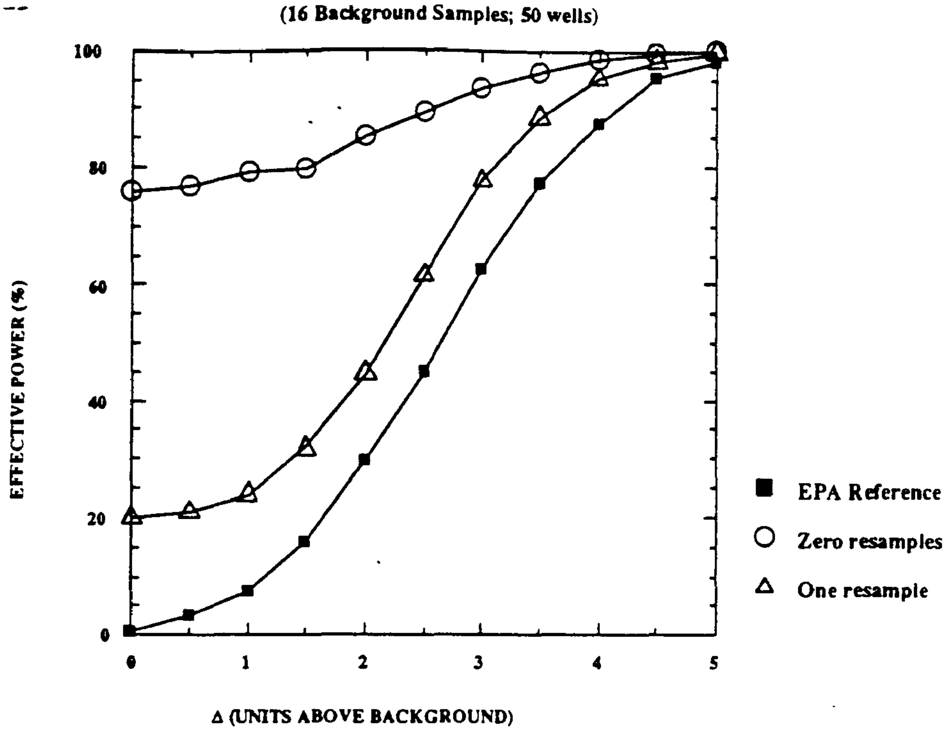
(8 Background Samples; 50 wells)

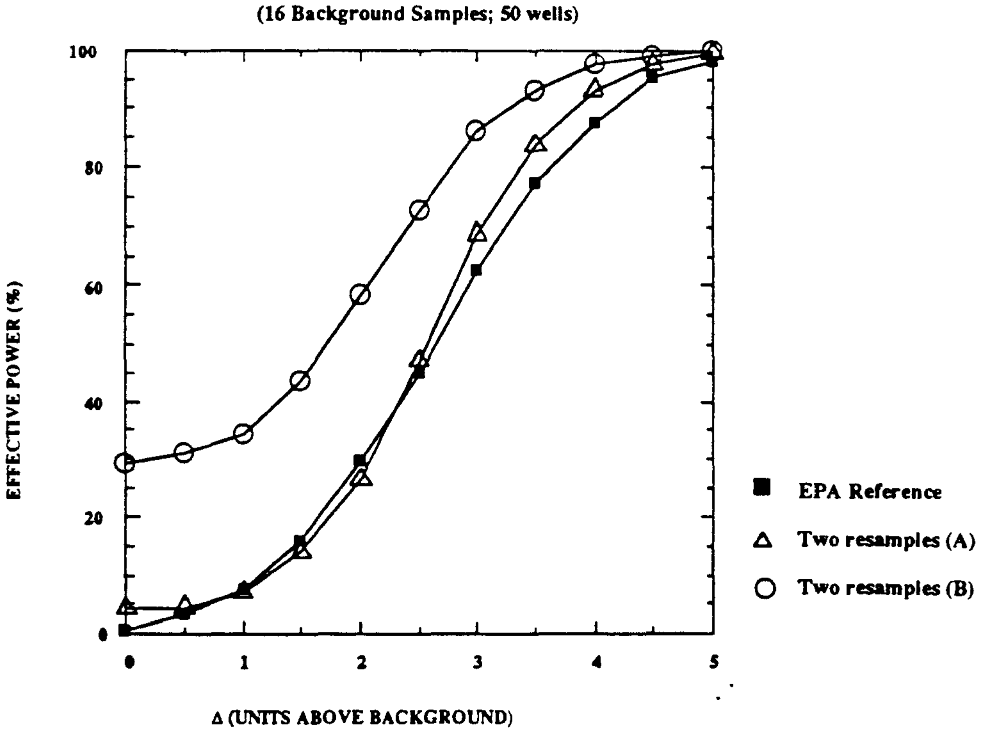


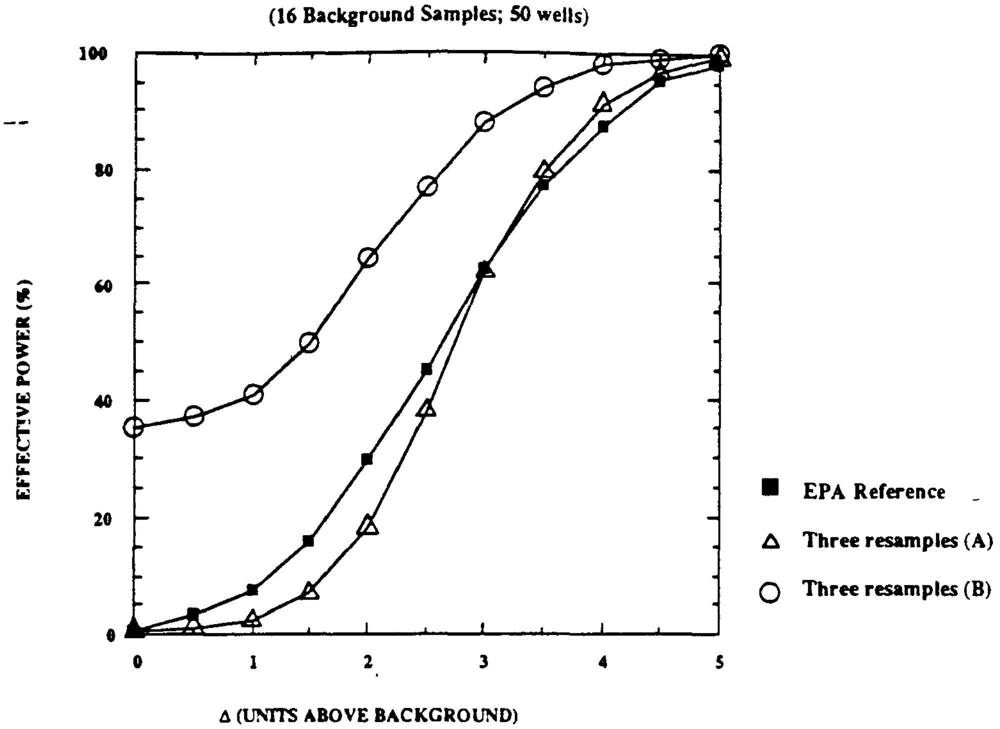

(8 Background Samples; 50 wells)
EFFECTIVE POWER (%)
0
20
40
60
80
100
0
1
2
3
4
5
Δ (UNITS ABOVE BACKGROUND)
■ EPA Reference
△ Two resamples (A)
○ Two resamples (B)


(8 Background Samples; 50 wells)
EFFECTIVE POWER (%)
0
20
40
60
80
100
0
1
2
3
4
5
Δ (UNITS ABOVE BACKGROUND)
■ EPA Reference
△ Three resamples (A)
○ Three resamples (B)

POWER CURVE FOR NON-PARAMETRIC PREDICTION LIMITS

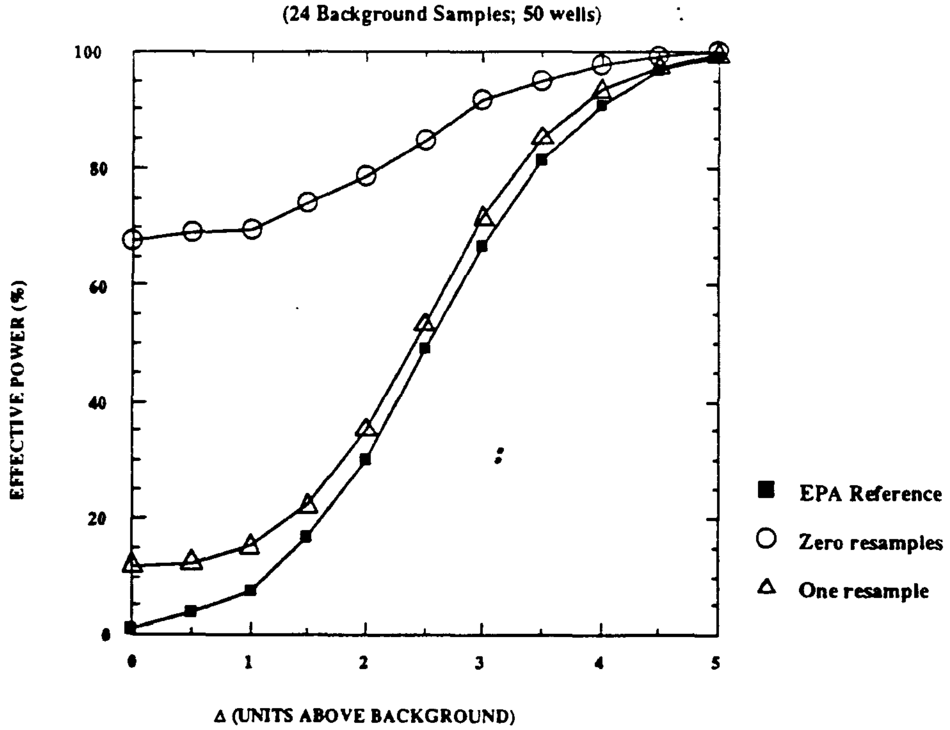




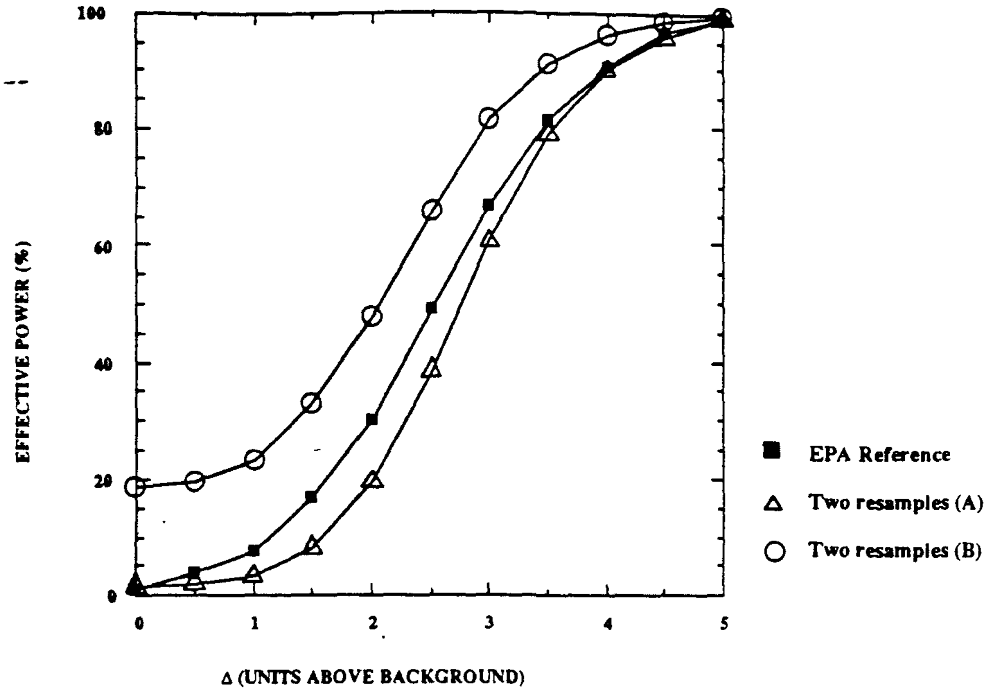


(16 Background Samples; 50 wells)

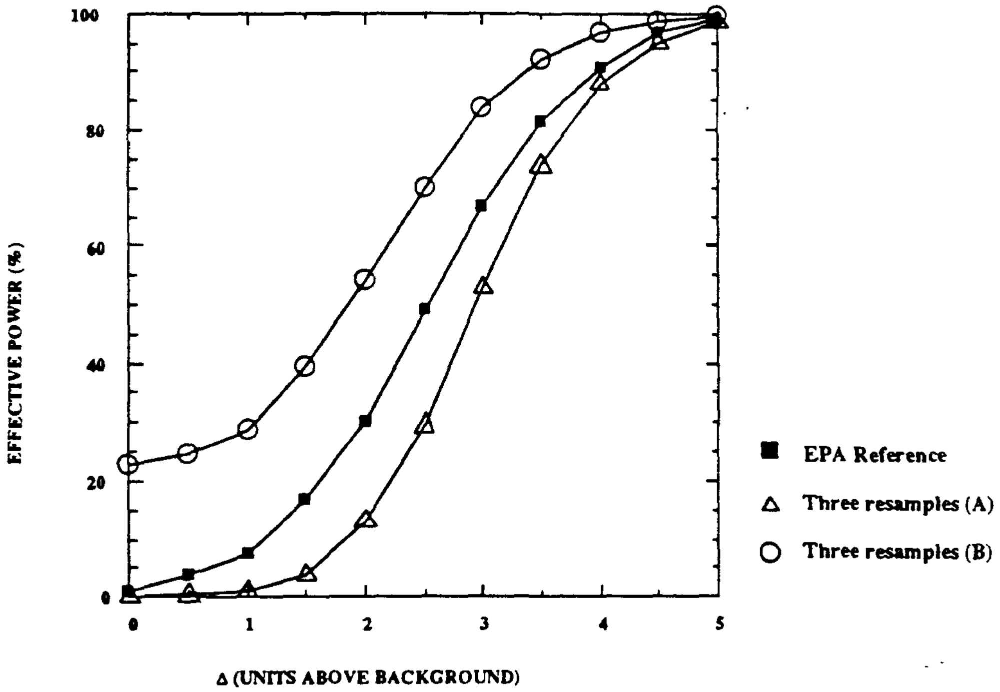
POWER CURVE FOR NON-PARAMETRIC PREDICTION LIMITS



(24 Background Samples; 50 wells)

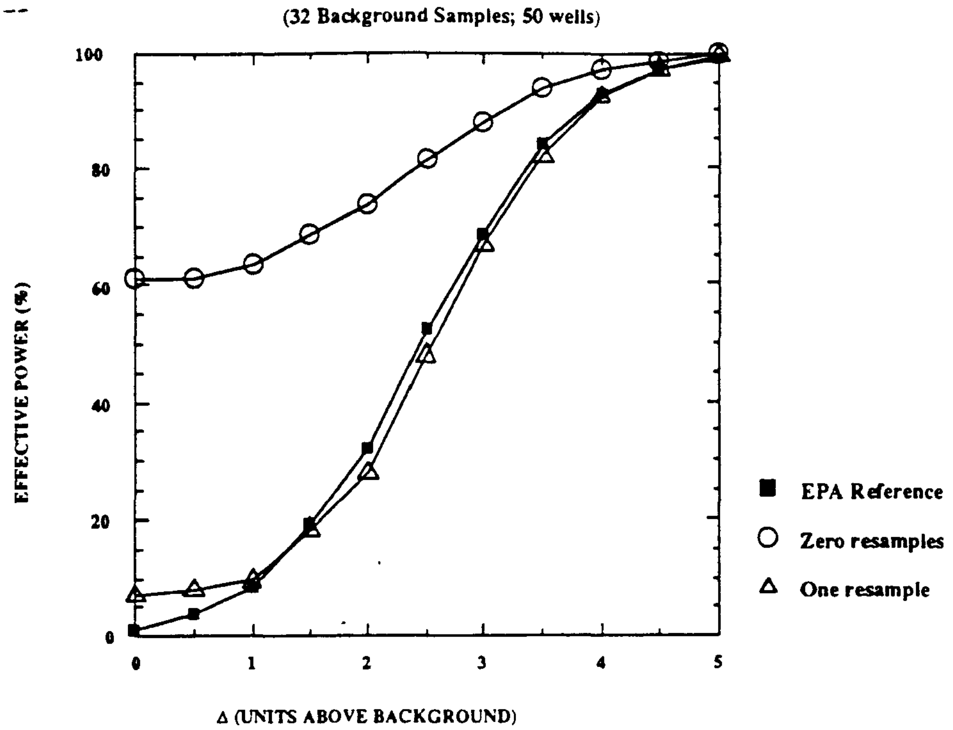


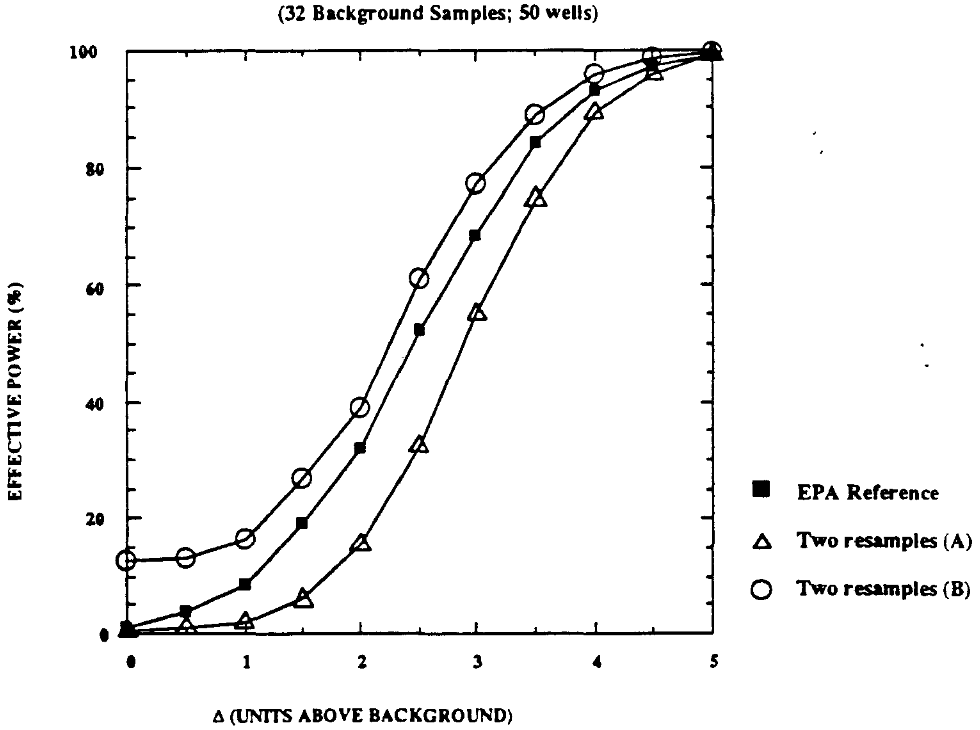
(24 Background Samples; 50 wells)



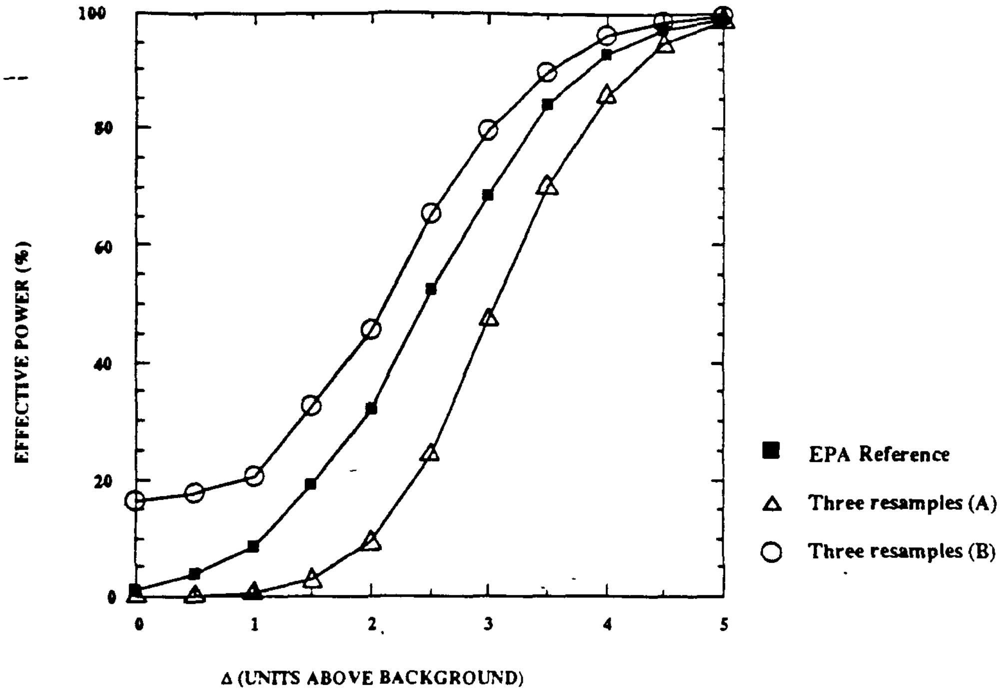
(24 Background Samples; 50 wells)

POWER CURVE FOR NON-PARAMETRIC PREDICTION LIMITS



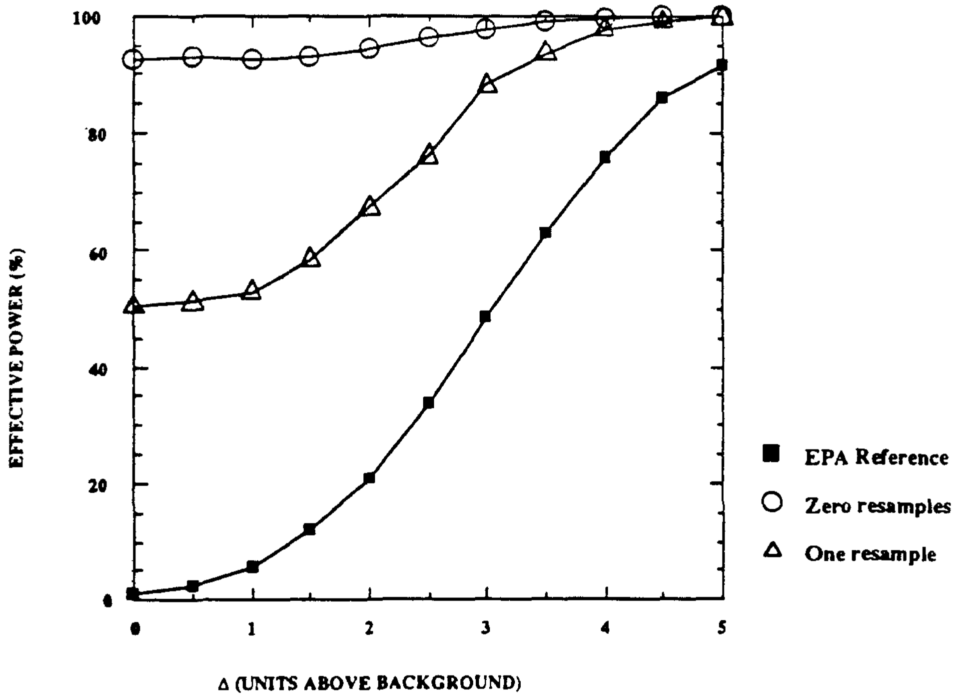


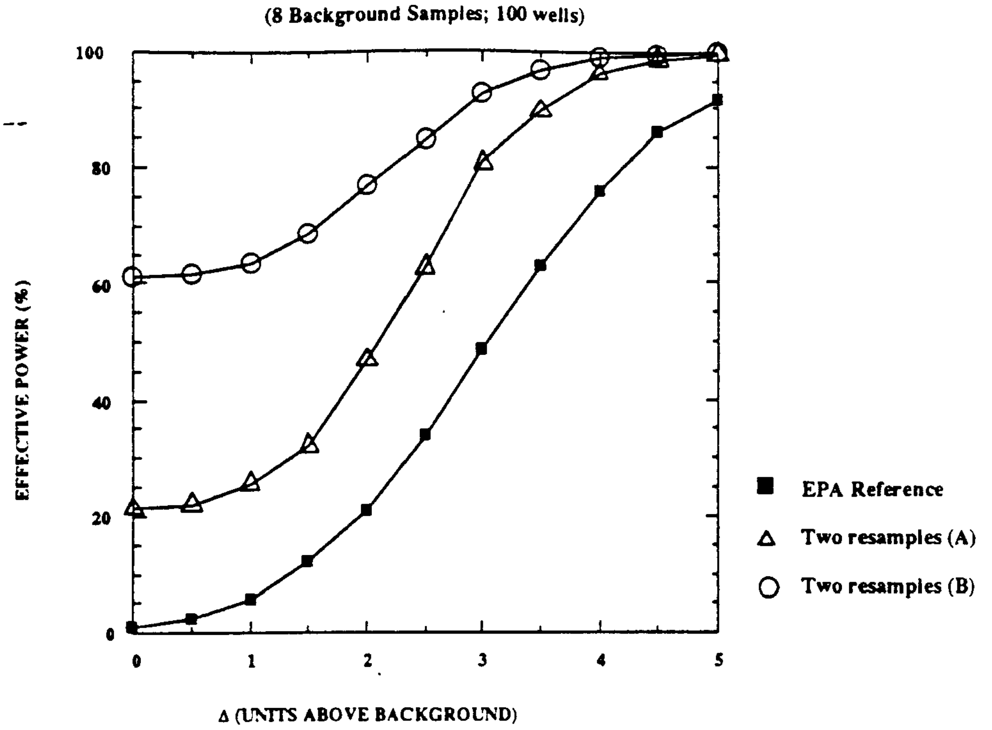
(32 Background Samples; 50 wells)

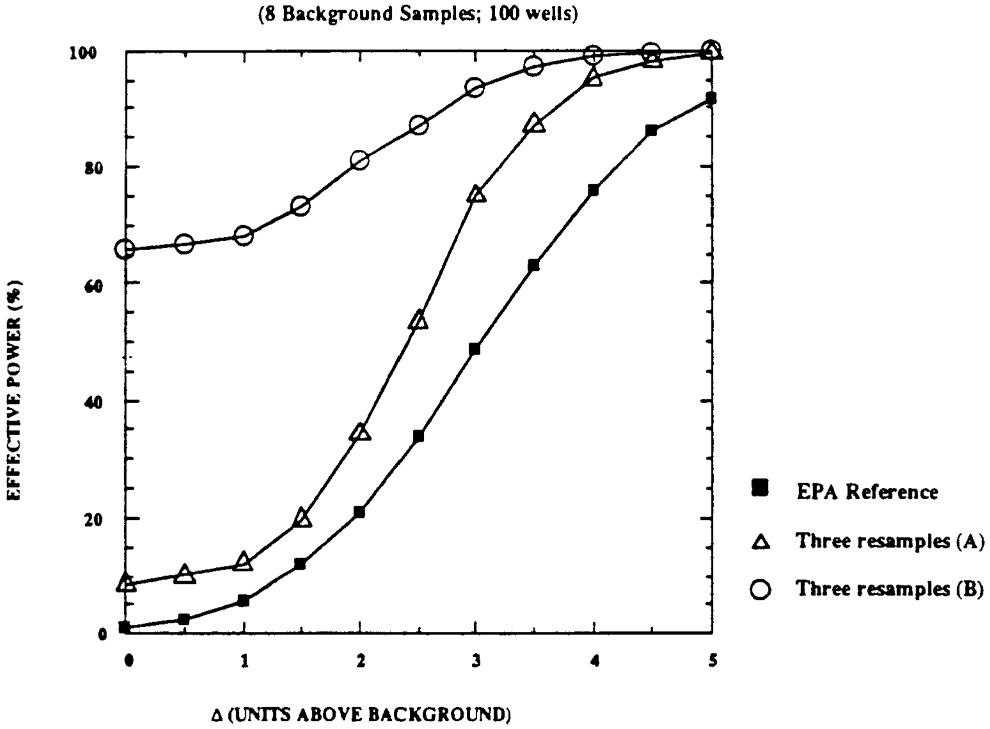


POWER CURVE FOR NON-PARAMETRIC PREDICTION LIMITS

(8 Background Samples; 100 wells)

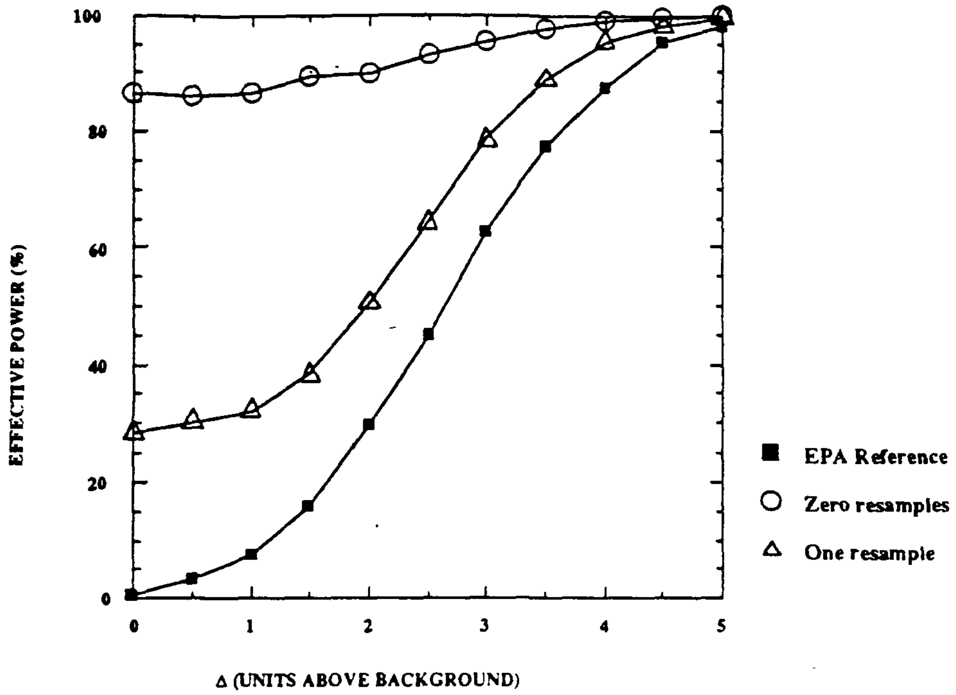



(8 Background Samples; 100 wells)
EFFECTIVE POWER (%)
0
20
40
60
80
100
0
1
2
3
4
5
Δ (UNITS ABOVE BACKGROUND)
■ EPA Reference
△ Two resamples (A)
○ Two resamples (B)

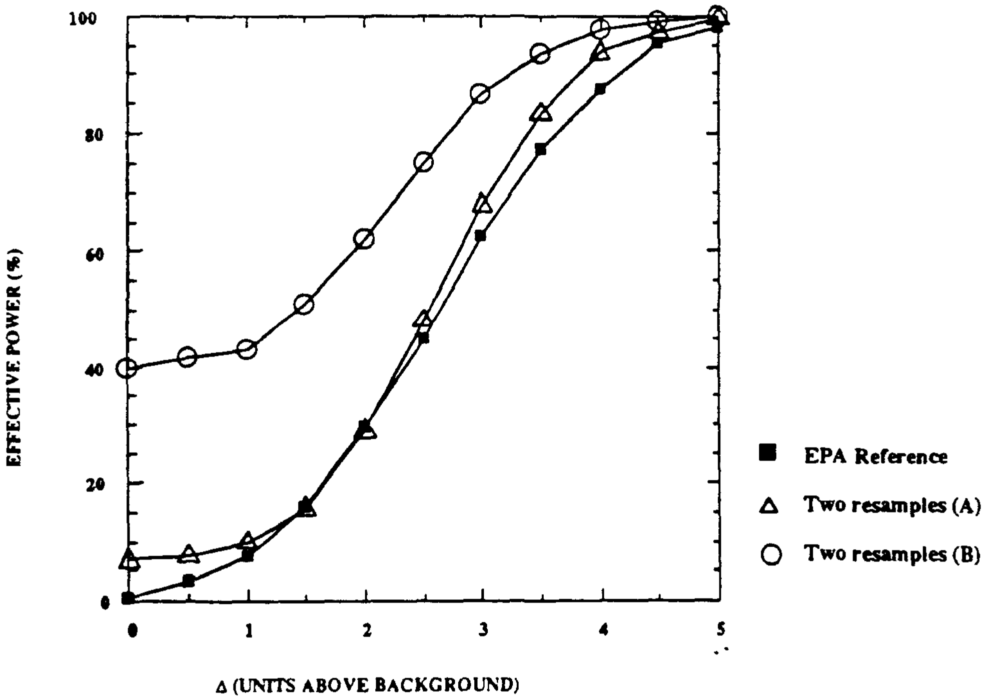

(8 Background Samples; 100 wells)
EFFECTIVE POWER (%)
0
20
40
60
80
100
0
1
2
3
4
5
Δ (UNITS ABOVE BACKGROUND)
■ EPA Reference
△ Three resamples (A)
○ Three resamples (B)

POWER CURVE FOR NON-PARAMETRIC PREDICTION LIMITS

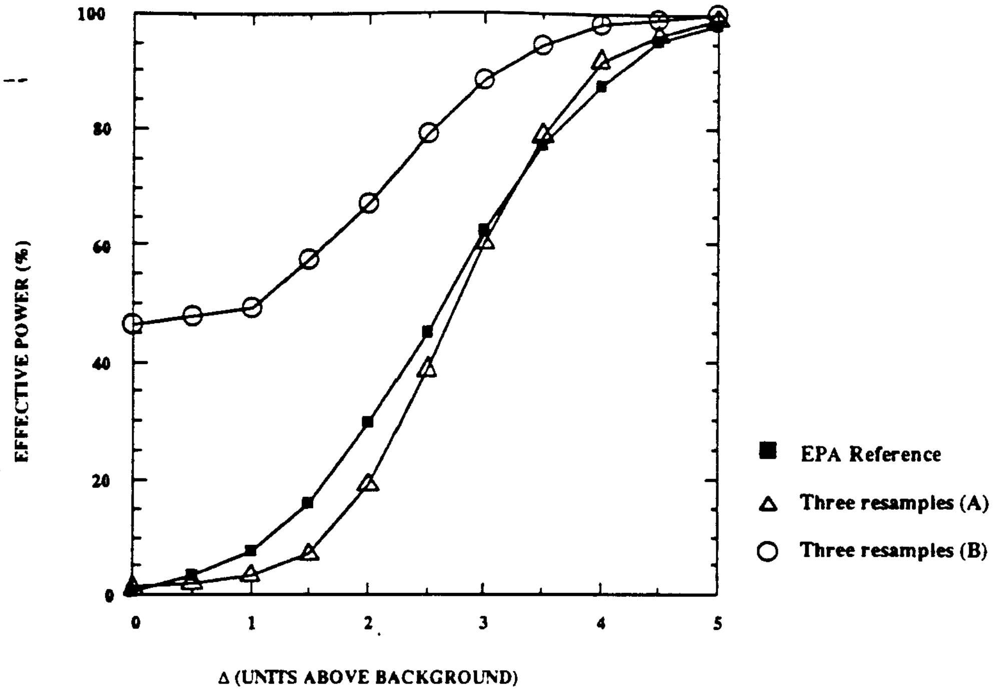
(16 Background Samples; 100 wells)



(16 Background Samples; 100 wells)

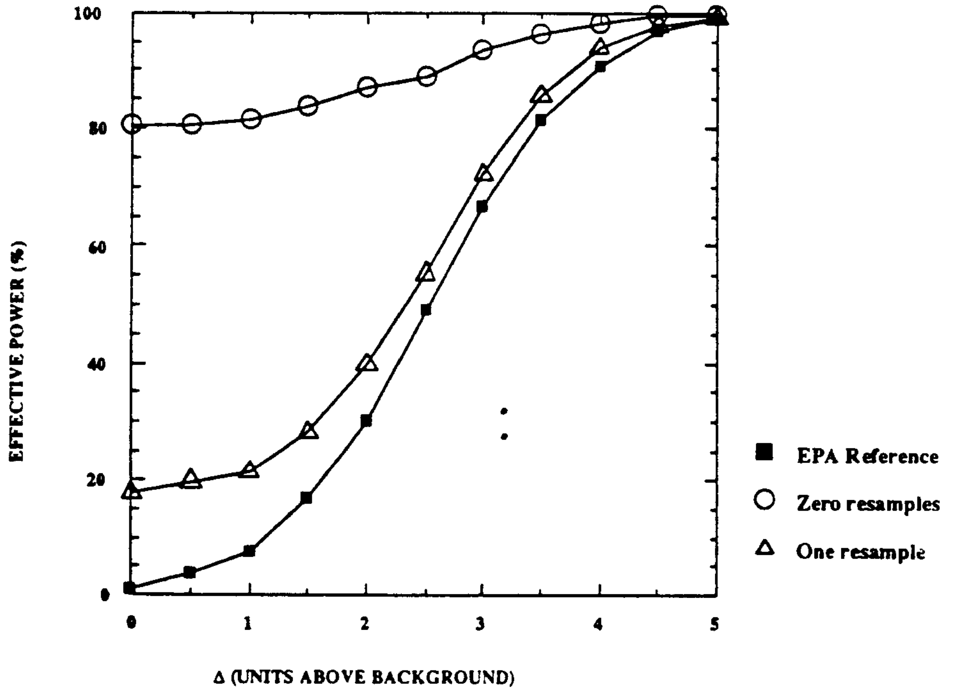


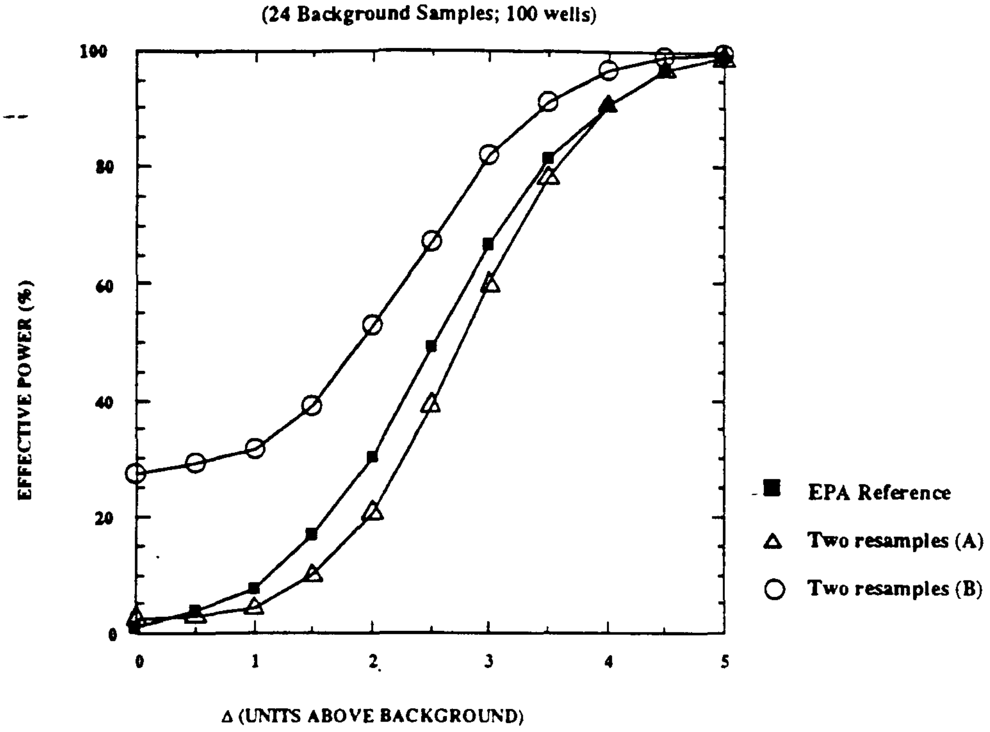
(16 Background Samples; 100 wells)

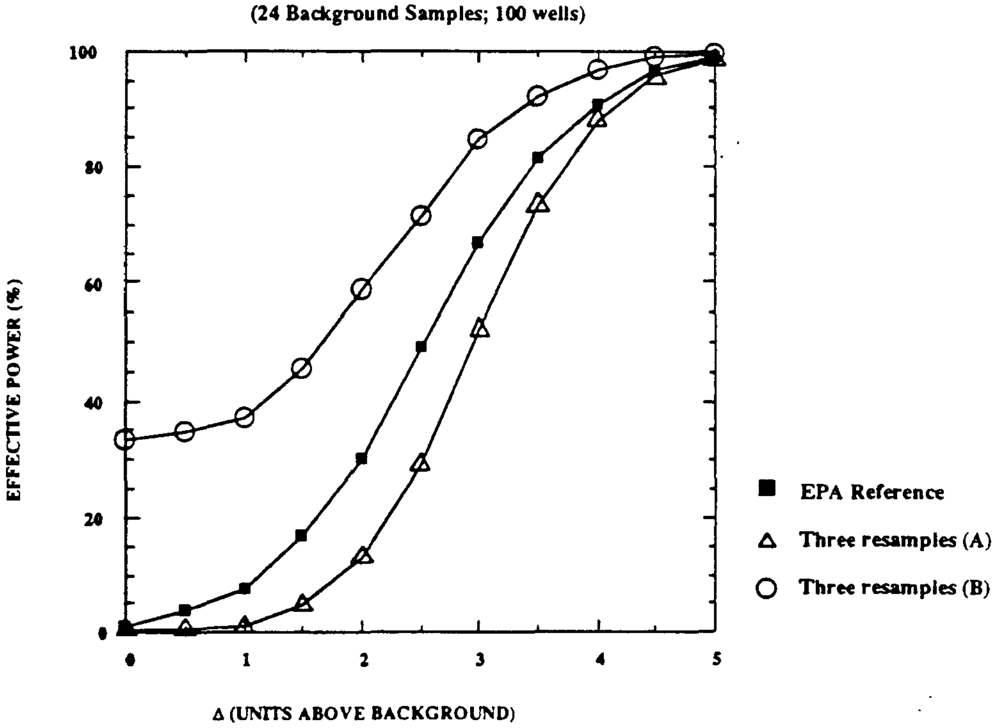


POWER CURVE FOR NON-PARAMETRIC PREDICTION LIMITS

(24 Background Samples; 100 wells)

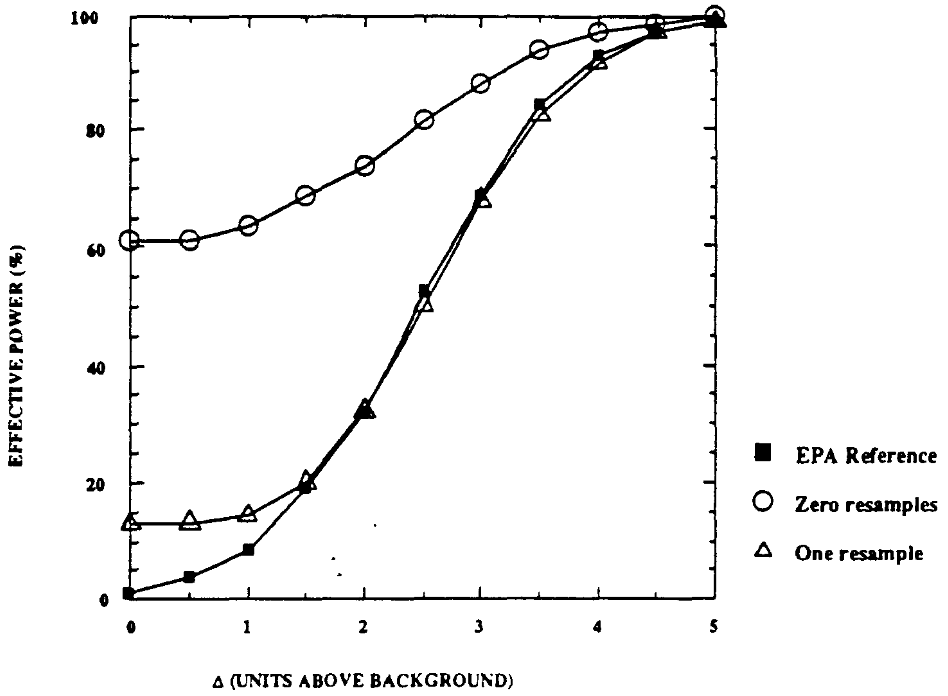



(24 Background Samples; 100 wells)

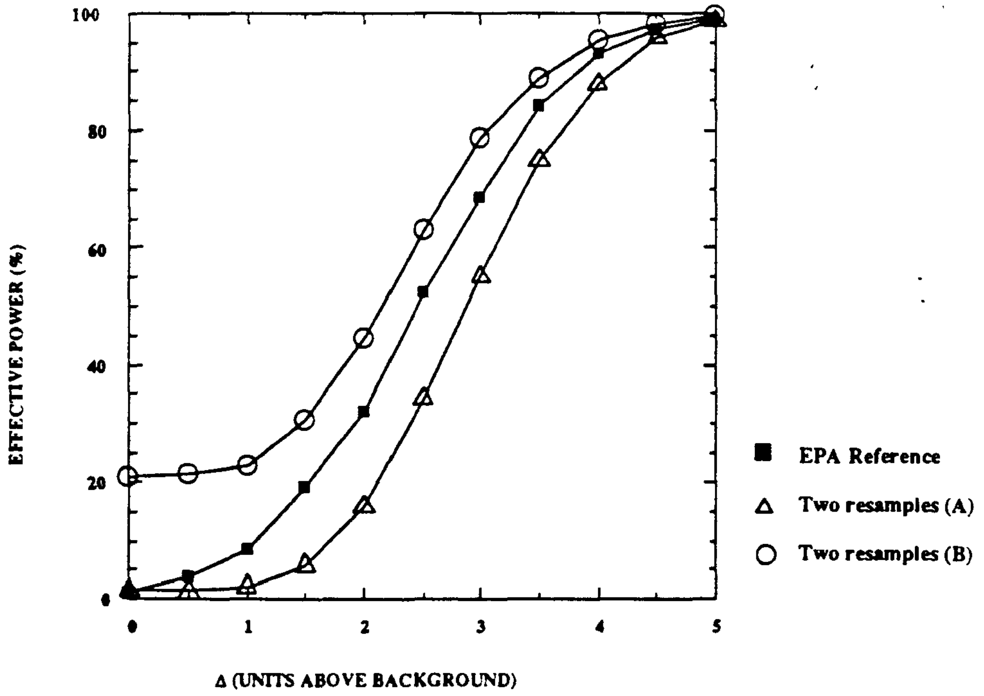

(24 Background Samples; 100 wells)

POWER CURVE FOR NON-PARAMETRIC PREDICTION LIMITS

(32 Background Samples; 100 wells)



(32 Background Samples; 100 wells)



(32 Background Samples; 100 wells)