United States
Environmental Protection
Agency

Office of Environmental
Information
Washington, DC 20460

EPA/240/B-06/003
February 2006

Data Quality Assessment: Statistical Methods for Practitioners

EPA QA/G-9S

FOREWORD

This document, *Data Quality Assessment: Statistical Methods for Practitioners,* provides general guidance to organizations on assessing data quality criteria and performance specifications. The Environmental Protection Agency (EPA) has developed the Data Quality Assessment (DQA) Process for project managers and planners to determine whether the type, quantity, and quality of data needed to support Agency decision making has been achieved. This guidance is the culmination of experiences in the design and statistical analyses of environmental data in different Program Offices at the EPA. Many elements of prior guidance, statistics, and scientific planning have been incorporated into this document.

This document is one of a series of quality management guidance documents that the EPA Quality Staff has prepared to assist users in implementing the Agency-wide Quality System. Other related documents include:

EPA QA/G-4	*Guidance on Systematic Planning Using the Data Quality Objectives Process*
EPA QA/G-5S	*Guidance on Choosing a Sampling Design for Environmental Data Collection*
EPA QA/G-9R	*Data Quality Assessment: A Reviewer's Guide*

This document provides guidance to EPA program managers and planning teams as well as to the general public as appropriate. It does not impose legally binding requirements and may not apply to a particular situation based on the circumstances. EPA retains the discretion to adopt approaches on a case-by-case basis that differ from this guidance where appropriate.

This guidance is one of the *U.S Environmental Protection Agency Quality System Series* documents. These documents describe the EPA policies and procedures for planning, implementing, and assessing the effectiveness of the Quality System. These documents are updated periodically to incorporate new topics and revision or refinements to existing procedures. Comments received on this, the 2006 version, will be considered for inclusion in subsequent versions. Please send your comments to:

Quality Staff (2811R)
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, NW
Washington, DC 20460
Phone: (202) 564-6830
Fax: (202) 565-2441
E-mail: quality@epa.gov

Copies of the EPA Quality System documents may be downloaded from the Quality Staff Home Page: www.epa.gov.quality.

PREFACE

Data Quality Assessment: Statistical Methods for Practitioners describes the statistical methods used in Data Quality Assessment (DQA) in evaluating environmental data sets. DQA is the scientific and statistical evaluation of environmental data to determine if they meet the planning objectives of the project, and thus are of the right type, quality, and quantity to support their intended use. This guidance applies DQA to environmental decision-making (e.g., compliance determinations) and also addresses DQA in environmental estimation (e.g., monitoring programs).

This document is distinctly different from other guidance documents in that it is not intended to be read in a linear or continuous fashion. Instead, it is intended to be used as a "tool-box" of useful techniques in assessing the quality of data. The overall structure of the document will enable the analyst to investigate many problems using a systematic methodology. Each statistical technique examined in the text is demonstrated separately in the form of a series of steps to be taken. The technique is then illustrated with a practical example following the steps described.

This document is intended for all EPA and extramural organizations that have quality systems based on EPA policies and specifications, and that may periodically assess these quality systems (or have them assessed by EPA) for compliance to the specifications. In addition, this guidance may be used by other organizations that assess quality systems applied to specific environmental programs.

The guidance provided herein is non-mandatory and is intended to help personnel who have minimal experience with statistical terminology to understand how a technique works and how it may be applied to a problem. An explanation of DQA in plain English may be found in the companion guidance document, Data Quality Assessment: A Reviewer's Guide (EPA QA/G-9R) (U.S. EPA, 2006b).

TABLE OF CONTENTS

If applicable, a list of tables, figures, and boxes is listed at the beginning of each chapter.

INTRODUCTION

Data Quality Assessment (DQA) is the scientific and statistical evaluation of environmental data to determine if they meet the planning objectives of the project, and thus are of the right type, quality, and quantity to support their intended use. DQA is built on a fundamental premise: data *quality* is meaningful only when it relates to the *intended use* of the data. This guidance describes the technical aspects of DQA in evaluating environmental data sets. A conceptual presentation of the DQA process is contained in *Data Quality Assessment: A Reviewer's Guide* (EPA QA/G-9R) (U.S. EPA 2004).

By using DQA, a reviewer can answer four important questions:

1. Can a decision (or estimate) be made with the desired level of certainty, given the quality of the data?

2. How well did the sampling design perform?

3. If the same sampling design strategy is used again for a similar study, would the data be expected to support the same intended use with the desired level of certainty?

4. Is it likely that sufficient samples were taken to enable the reviewer to see an effect if it was really present?

The first question addresses the reviewer's immediate needs. For example, if the data are being used for decision-making and provide evidence strongly in favor of one course of action over another, then the decision maker can proceed knowing that the decision will be supported by unambiguous data. However, if the data do not show sufficiently strong evidence to favor one alternative, then the data analysis alerts the decision maker to this uncertainty. The decision maker now is in a position to make an informed choice about how to proceed (such as collect more or different data before making the decision, or proceed with the decision despite the relatively high, but tolerable, chance of drawing an erroneous conclusion).

The second question addresses how robust this sampling design is with respect to slightly changing conditions. If the design is very sensitive to potentially disturbing influences, then interpretation of the results may be difficult. By addressing the second question, the reviewer guards against the possibility of a spurious result arising from a unique set of circumstances.

The third question addresses the problem of whether this could be considered a unique situation where the results of this DQA only apply only to this situation and cannot be extrapolated to other situations. It also addresses the suitability of using this data collection design again for future projects. For example, if it is intended to use a certain sampling design at a different location from where the design was first used, it should be determined how well the design can be expected to perform given that the outcomes and environmental conditions of this sampling event will be different from those of the original event. As environmental conditions

will vary from one location or one time to another, the adequacy of the sampling design should be evaluated over a broad range of possible outcomes and conditions.

The final question addresses the issue of whether sufficient resources were used in the study. For example, in an epidemiological investigation, was it likely the effect of interest could be reliably observed given the limited number of samples actually obtained.

The data life cycle comprises three steps: planning, implementation, and assessment. During the planning phase, the Data Quality Objectives (DQO) Process (or any other systematic planning procedure) is used to define criteria for determining the number, location, and timing of samples to be collected in order to produce a result with the desired level of certainty. This, along with the sampling methods, analytical procedures, and appropriate quality assurance (QA) and quality control (QC) procedures, is documented in the QA Project Plan. The implementation phase consists of collecting data following the QA Project Plan specifications. At the outset of the assessment phase, the data are validated and verified to ensure that the sampling and analysis protocols specified in the QA Project Plan were followed, and that the measurement systems performed in accordance with the criteria specified in the QA Project Plan. Then, DQA completes the data life cycle through determining if the performance and acceptance criterion from the DQO planning objectives were achieved and to draw conclusions from the data.

DQA involves five steps that begin with a review of the planning documentation and end with an answer to the question posed during the planning phase of the study. These steps roughly parallel the actions of an environmental statistician when analyzing a set of data. The five steps, which are described in more detail in the following chapters of this guidance, are:

1. *Review the project objectives and sampling design.*
2. *Conduct a preliminary data review.*
3. *Select the statistical method.*
4. *Verify the assumptions of the statistical method*.
5. *Draw conclusions from the data.*

These five steps are presented in a linear sequence, but DQA is actually an iterative process. For example, if the preliminary data review reveals patterns or anomalies in the data set that are inconsistent with the project objectives, then some aspects of the study planning may have to be reconsidered in Step 1. Likewise, if the underlying assumptions of the statistical method are not supported by the data, then previous steps of the DQA may have to be revisited.

This guidance is written for a broad audience of potential data users, data analysts, and data generators. Data users (such as project managers or risk assessors who are responsible for making decisions or producing estimates regarding environmental characteristics) should find this guidance useful for understanding and directing the technical work of others who produce and analyze data. Data analysts should find this guidance to be a convenient compendium of basic assessment tools. Data generators (such as analytical chemists or field sampling specialists responsible for collecting and analyzing environmental samples and reporting the resulting data values) should find this guidance useful for understanding how their work will be used and for

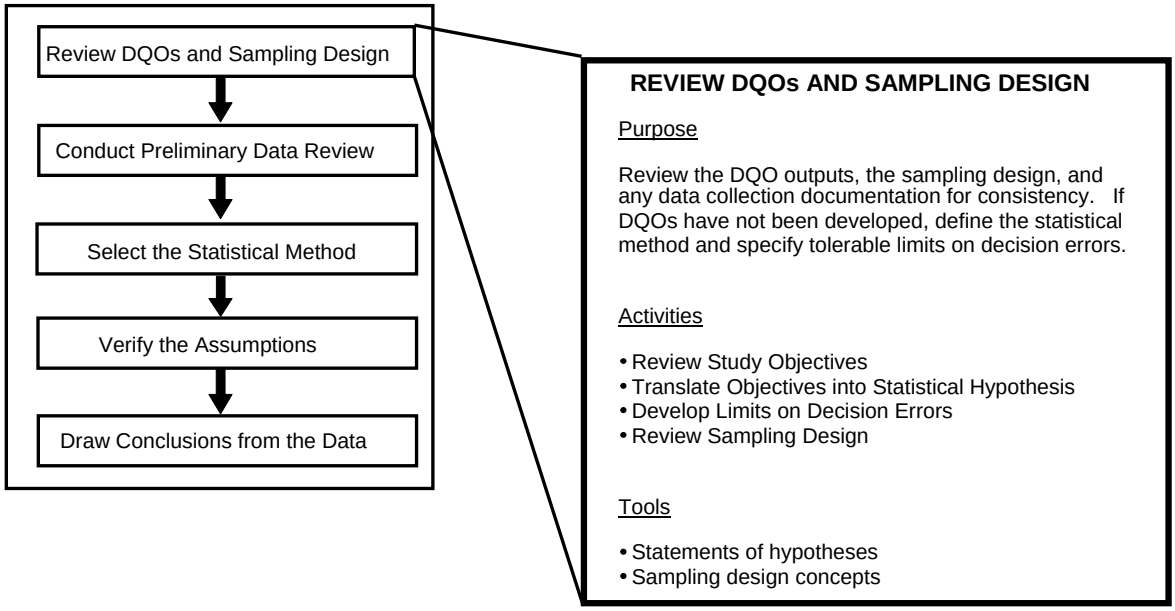
providing a foundation for improving the efficiency and effectiveness of the data generation process.

This guidance presents background information and statistical tools for performing DQA; a non-statistical discussion of DQA is to be found in the companion document *Data Quality Assessment: A Reviewer's Guide* (EPA QA/G-9R) (U.S. EPA 2004). Each chapter corresponds to a step in the DQA and begins with an overview of the activities to be performed for that step. Following the overviews in Chapters 1, 2, 3, and 4, specific graphical or statistical tools are described and step-by-step procedures are provided along with examples. Chapter 5 gives some advice on the interpretation of statistical tests. Appendix A contains statistical tables, and Appendix B provides references and useful publications for in-depth statistical analyses.

CHAPTER 1

STEP 1: REVIEW DQOs AND THE SAMPLING DESIGN

THE DATA QUALITY ASSESSMENT PROCESS



Step 1: Review DQOs and Sampling Design

- Review the objectives of the study.
  - If DQOs were developed, then review the outputs from the DQO Process.
  - If DQOs have not been developed, then ascertain what these objectives were.

- Translate the data user's objectives into a statement of the primary statistical hypothesis.
  - If DQOs were developed, translate them into a statement of the primary hypothesis.
  - If DQOs have not been developed, then ascertain what hypotheses or estimates were developed.

- Translate the data user's objectives into limits on Type I or Type II decision errors.
  - If DQOs have not been developed, document the data user's probable tolerable limits on decision errors, width of gray region, and estimated preliminary values.
  - If DQOs were developed, confirm the limits on decision errors.

- Review the sampling design and note any special features or potential problems.
  - Review the sampling design for any potentially serious deviations.

CHAPTER 1

STEP 1: REVIEW DQOs AND THE SAMPLING DESIGN

1.1 OVERVIEW AND ACTIVITIES

DQA begins by reviewing the key outputs from the planning phase of the data life cycle such as the Data Quality Objectives, the QA Project Plan, and any associated documents. The study objectives provide the context for understanding the purpose of the data collection effort and establish the qualitative and quantitative basis for assessing the quality of the data set for the intended use. The sampling design (documented in the QA Project Plan) provides important information about how to interpret the data. By studying the sampling design, the analyst can gain an understanding of the assumptions under which the design was developed, as well as the relationship between these assumptions and the objectives.

In the unfortunate instances when project objectives have not been developed during the planning phase of the study, it is necessary to develop statements of the data user's objectives prior to conducting the DQA. The primary purpose of stating the data user's objectives prior to analyzing the data is to establish appropriate criteria for evaluating the quality of the data with respect to their intended use. Analysts who are not familiar with the DQO Process should refer to the *Guidance for the Data Quality Objectives Process (EPA QA/G-4)* (U.S. EPA 2000), a book on statistical planning and analysis, or a consulting statistician. The seven steps of the DQO Process are illustrated in Figure 1-1.

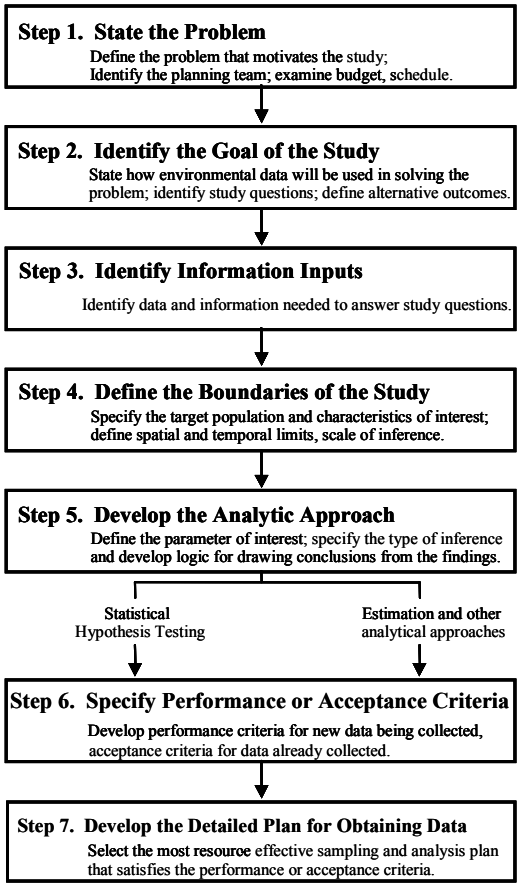


Figure 1-1. The DQO Process

If the project has been framed as a hypothesis test, then the uncertainty limits can be expressed as the data user's tolerance for committing false rejection (Type I or a false positive) or false acceptance (Type II or a false negative) decision errors. A false rejection error occurs when the null hypothesis is rejected when it is in fact true. A false acceptance error occurs when the null hypothesis is not rejected when it is in fact false. Other related phrases in common use include "level of significance" which is equal to the probability of a Type I error, and "power" which is equal to 1 minus probability of a Type II error. Statistical power really is a function that describes a "power curve" over a range of Type II errors. The characteristics of the power curve can be a great importance in choosing the appropriate statistical test. For detailed information on how to develop false rejection and false acceptance

decision error rates, see Chapter 6 of *Guidance on Data Quality Objectives* EPA QA/G-4 (U.S. EPA 2000).

If the project has been framed in terms of confidence intervals, then uncertainty is expressed as a combination of two interrelated terms: the width of the confidence interval (smaller intervals correspond to a smaller degree of uncertainty) or the confidence level that the true value of the parameter of interest lies within the interval (a higher confidence level represents a smaller degree of uncertainty).

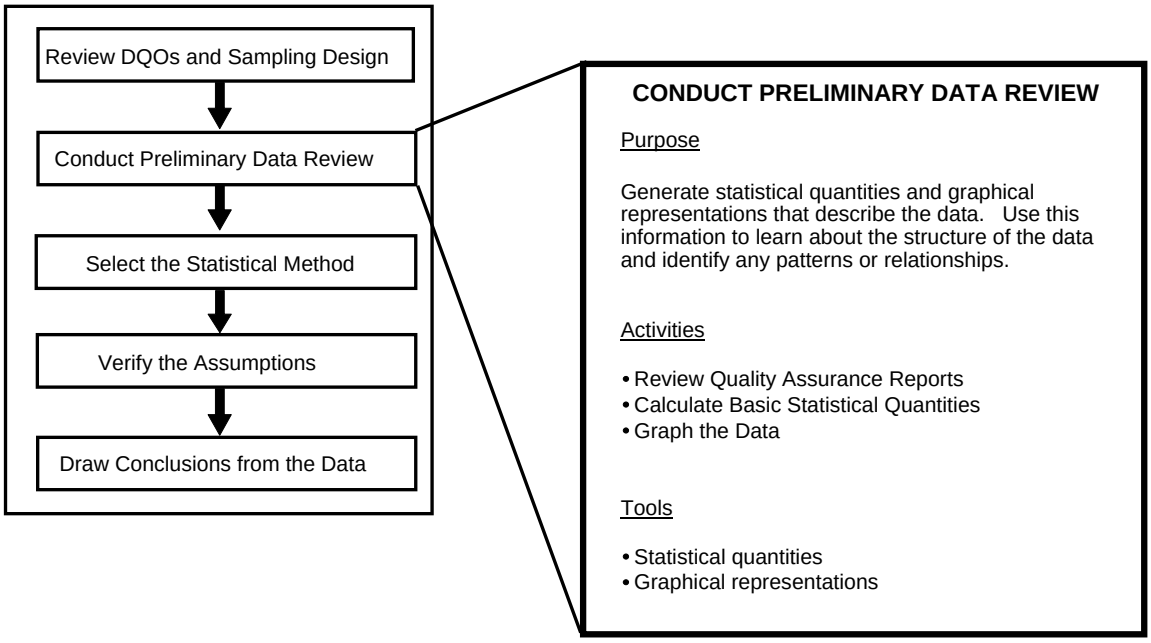
For the review of sampling design, recall that the key distinction in sampling design is between authoritative or judgmental sampling (in which sample numbers and locations are selected based on expert opinion) and probability sampling (in which sample numbers and locations are selected based on randomization and each member of the target population has a known probability of being included in the sample). Judgmental sampling should be considered only when the objectives of the investigation are not of a statistical nature (for example, when the objective of a study is to identify specific locations of leaks, or when the study is focused solely on the sampling locations themselves). Generally, conclusions drawn from authoritative samples apply only to the individual samples; aggregation may result in severe bias due to lack of representativeness and lead to highly erroneous conclusions. Judgmental sampling also precludes the use of the sample for any purpose other than the original one. If judgmental sample data are used, great care should be taken in interpreting any statistical statements concerning the conclusions to be drawn. Using some probabilistic statement with and judgmental sample is incorrect and should be avoided as it gives an illusion of correctness when there is none. The further the judgmental sample is from a truly random sample, the more questionable the conclusions. *Guidance on Choosing a Sampling Design for Environmental Data Collection* (EPA QA/G-5S) (U.S. EPA 2002) provides extensive information on sampling design issues and their implications for data interpretation.

The analyst should review the sampling design documentation with the data user's objectives in mind. Look for design features that support or contradict those objectives. For example, if the data user is interested in making a decision about the mean level of contamination in an effluent stream over time, then composite samples may be an appropriate sampling approach. On the other hand, if the data user is looking for hot spots of contamination at a hazardous waste site, compositing should be used with caution, to avoid "averaging away" hot spots. Also, look for potential problems in the implementation of the sampling design. For example, if simple random sampling has been used verify that each point in space (or time) had an equal probability of being selected for a simple random sampling design. Small deviations from a sampling plan may have minimal effect on the conclusions drawn from the data set. Significant or substantial deviations should be flagged and their potential effect carefully considered throughout the entire DQA. The most important point is to verify that the collected data are consistent with how the QA Project Plan, Sampling and Analysis Plan, or overall objectives of the study stated them to be.

CHAPTER 2

STEP 2: CONDUCT A PRELIMINARY DATA REVIEW

THE DATA QUALITY ASSESSMENT PROCESS



Step 2: Conduct a Preliminary Data Review

- Review quality assurance reports.
  - Look for problems or anomalies in the implementation of the sample collection and analysis procedures.
  - Examine QC data for information to verify assumptions underlying the Data Quality Objectives, the Sampling and Analysis Plan, and the QA Project Plans.
- Calculate the statistical quantities.
  - Consider calculating appropriate percentiles (Section 2.2.1).
  - Select measures of central tendency (Section 2.2.2) and dispersion (Section 2.2.3).
  - If the data involve two variables, calculate the correlation coefficient (Section 2.2.4).
- Display the data using graphical representations.
  - Select graphical representations (Section 2.3) that illuminate the structure of the data set and highlight assumptions underlying the Data Quality Objectives, the Sampling and Analysis Plan, and the QA Project Plans.
  - Use a variety of graphical representations that examine different features of the set.

List of Boxes

List of Figures

CHAPTER 2

STEP 2: CONDUCT A PRELIMINARY DATA REVIEW

2.1 OVERVIEW AND ACTIVITIES

In this step of DQA, the analyst conducts a preliminary evaluation of the data set, calculates some basic statistical quantities, and examines the data using graphical representations The first activity in conducting a preliminary data review is to review any relevant QA reports giving particular attention should be paid to information that can be used to check assumptions made in the DQO Process. Of great importance are apparent anomalies in recorded data, missing values, deviations from standard operating procedures, and the use of nonstandard data collection methodologies.

Graphs can be used to identify patterns and trends, to quickly confirm or disprove hypotheses, to discover new phenomena, to identify potential problems, and to suggest corrective measures. Since no single graphical representation will provide a complete picture of the data set, the analyst should choose different graphical techniques to illuminate different features of the data. Section 2.3 provides descriptions and examples of common graphical representations.

For a more extensive discussion of the overview and activities of this step, see *Data Quality Assessment: A Reviewer's Guide* (EPA QA/G-9R) (U.S. EPA 2004).

2.2 STATISTICAL QUANTITIES

2.2.1 Measures of Relative Standing

Sometimes the analyst is interested in knowing the relative position of one or several observations in relation to all of the observations. Percentiles or quantiles are measures of relative standing that are useful for summarizing data. A percentile is the data value that is greater than or equal to a given percentage of the data values. Stated in mathematical terms, the p^{th} percentile is a data value that is greater than or equal to p% of the data values and is less than or equal to (1-p)% of the data values. Therefore, if 'x' is the p^{th} percentile, then p% of the values in the data set are less than or equal to x, and (100-p)% of the values are greater than x. A sample percentile may fall between a pair of observations. For example, the 75^{th} percentile of a data set of 10 observations is not uniquely defined. Therefore, there are several methods for computing sample percentiles, the most common of which is described in Box 2-1.

Important percentiles usually reviewed are the quartiles of the data, the 25^{th}, 50^{th}, and 75^{th} percentiles. The 50^{th} percentile is also called the sample median (Section 2.2.2), and the 25^{th} and 75^{th} percentile are used to estimate the dispersion of a data set (Section 2.2.3). Also important for environmental data are the 90^{th}, 95^{th}, and 99^{th} percentiles where a decision maker would like to be sure that 90%, 95%, or 99% of the contamination levels are below a fixed risk level.

Box 2-1: Directions for Calculating Percentiles with an Example

Let X_1, X_2, ..., X_n represent the n data points. To compute the p^{th} percentile, $y(p)$, first rank the data from smallest to largest and label these points $X_{(1)}$, $X_{(2)}$, . . ., $X_{(n)}$. The p^{th} percentile is:

$$y(p) = (1-f)\cdot X_{(i)} + f \cdot X_{(i+1)}$$

where $r = (n\text{-}1)p + 1$, i = floor(r), $f = r - i$, and $X_{(n+1)} = X_{(n)}$. Note that floor(r) means calculate r and then discard all decimals.

Example: The 90th and 95th percentile will be computed for the following 10 data points (ordered from smallest to largest) : 4, 4, 4, 5, 5, 6, 7, 7, 8, and 10 ppb.

For the 95th percentile, $r = (10\text{-}1)\times 0.95 + 1 = 9.55$, i = floor(9.55) = 9, and $f = 9.55 - 9 = 0.55$. Therefore, the 95th percentile is $y(0.95) = 0.45\times 8 + 0.55\times 10 = 9.1$.

For the 90th percentile, $r = (10\text{-}1)\times 0.9 + 1 = 9.1$, i = floor(9.1) = 9, and $f = 9.1 - 9 = 0.1$. Therefore, the 90th percentile is $y(0.9) = 0.9\times 8 + 0.1\times 10 = 8.2$.

2.2.2 Measures of Central Tendency

Measures of central tendency characterize the center of a data set. The three most common estimates are the mean, median, and the mode. Directions for calculating these quantities are contained in Box 2-2; examples are provided in Box 2-3.

The most commonly used measure of the center of a data set is the sample mean, denoted by $\overline{X}$. The sample mean can be thought of as the "center of gravity" of the data set. The sample mean is an arithmetic average for simple sampling designs; however, for complex sampling designs, such as stratification, the sample mean is a weighted arithmetic average. The sample mean is influenced by extreme values (large or small) and the treatment of non-detects (see Section 4.7).

The sample median, is the second most popular measure of the center of the data. This value falls directly in the middle of the ordered data set. This means that ½ of the data are smaller than the sample median and ½ of the data are larger than the sample median. The median is another name for the 50th percentile (Section 2.2.1). The median is not influenced by extreme values and can easily be used if non-detects are present.

Another method of measuring the center of the data is the sample mode. The sample mode is the value that occurs with the greatest frequency. Since the sample mode may not exist or be unique, it is the least commonly used measure of center. However, the mode is useful for qualitative data.

Box 2-2: Directions for Calculating the Measures of Central Tendency

Let X_1, X_2, ..., X_n represent the n data points.

Sample Mean: The sample mean, $\overline{X}$, is the sum of the data points divided by the sample size, n:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Sample Median: The sample median, $\tilde{X}$, is the center of the ordered data set. To compute the sample median, sort the data from smallest to largest and label these points $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$. Then,

$$\tilde{X} = \begin{cases} \frac{1}{2}\left[X_{(n/2)} + X_{(n/2+1)}\right] & \text{if } n \text{ is even} \\ X_{((n+1)/2)} & \text{if } n \text{ is odd} \end{cases}$$

Sample Mode: The sample mode is the value in the sample that occurs with the greatest frequency. The sample mode may not exist or be unique. Count the number of times each value occurs. The sample mode is the value that occurs most frequently.

Box 2-3: Example Calculations of the Measures of Central Tendency

The following is an example of computing the sample mean, median, and mode for the 10 data points: 4, 4, 7, 7, 4, 10, 4, 3, 7, and 8.

Sample mean:

$$\overline{X} = \frac{4 + 4 + 7 + 7 + 4 + 10 + 4 + 3 + 7 + 8}{10} = \frac{58}{10} = 5.8$$

Sample median: The ordered data are: 3, 4, 4, 4, 4, 7, 7, 7, 8, and 10. Since n = 10 is even, the sample median is:

$$\text{median} = \frac{1}{2}\left[X_{(10/2)} + X_{(10/2+1)}\right] = \frac{1}{2}\left[X_{(5)} + X_{(6)}\right] = \frac{1}{2}\left[4 + 7\right] = 5.5$$

Sample mode: Computing the number of times each value occurs yields:

4 appears 4 times; 5 appears 0 times; 6 appears 1 time; 7 appears 3 times; 8 appears 1 time; and 10 appears 1 time.

As the value of 4 occurs most often, it is the sample mode of this data set.

2.2.3 Measures of Dispersion

Measures of central tendency are more meaningful if accompanied by a measure of the spread of values about the center. Measures of dispersion in a data set include the range, variance, sample standard deviation, coefficient of variation, and the interquartile range. Directions for computing these measures are given in Box 2-4; examples are given in Box 2-5.

Box 2-4: Directions for Calculating the Measures of Dispersion

Let $X_1, X_2, \ldots, X_n$ represent the n data points.

Sample Range: The sample range, R, is the difference between the largest and smallest values of the data set, i.e., $R = \max(X_i) - \min(X_i)$.

Sample Variance: To compute the sample variance, s^2, compute: $$s^2 = \frac{\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2}{n-1}$$

Sample Standard Deviation: The sample standard deviation, s, is the square root of the sample variance, i.e., $s = \sqrt{s^2}$

Coefficient of Variation: The coefficient of variation (CV) is the sample standard deviation divided by the sample mean (Section 2.2.2), i.e., $CV = s / \overline{X}$. The CV is often expressed as a percentage.

Interquartile Range: The interquartile range (IQR) is the difference between the 75th and the 25th percentiles, i.e., IQR = $y(75) - y(25)$.

Box 2-5: Example Calculations of the Measures of Dispersion

The directions in Box 2-4 and the following 10 data points (in ppm): 4, 5, 6, 7, 4, 10, 4, 5, 7, and 8, are used to calculate the measures of dispersion. From Box 2-3, $\overline{X} = 6$ ppm.

Sample Range: $R = \max(X_i) - \min(X_i) = 10 - 4 = 6$ ppm

Sample Variance: $$s^2 = \frac{\left(4^2 + 5^2 + \cdots + 8^2\right) - \frac{1}{10}\left(4 + 5 + \cdots + 8\right)^2}{10-1} = \frac{396 - \frac{1}{10}\cdot 60^2}{9} = 4\ \text{ppm}^2$$

Sample Standard Deviation: $s = \sqrt{s^2} = \sqrt{4} = 2$ ppm

Coefficient of Variation: $CV = s / \overline{X} = 2\ \text{ppm} / 6\ \text{ppm} = 0.33 = 33\%$

Interquartile Range: Using the directions in Section 2.2.1 to compute the 25th and 75th percentiles of the data are: $y(25) = X_{(3)} = 4.25$ ppm and $y(75) = X_{(8)} = 7$ ppm. The interquartile range (IQR) is the difference between these values: IQR = $y(75) - y(25) = 7 - 4.25 = 2.75$ ppm

The easiest measure of dispersion to compute is the sample range. For small samples, the range is easy to interpret and may adequately represent the dispersion of the data. For large

samples, the range is not very informative because it only considers extreme values and is therefore greatly influenced by outliers.

Generally speaking, the sample variance measures the average squared distance of data points from the sample mean. A large sample variance implies the data are not clustered close to the mean. A small sample variance (relative to the mean) implies most of the data are near the mean. The sample variance is affected by extreme values and by a large number of non-detects. The sample standard deviation is the square root of the sample variance and has the same unit of measure as the data.

The coefficient of variation (CV) is a measure having no units that allows the comparison of dispersion across several sets of data. The CV (also known as the relative standard deviation) is often used in environmental applications because variability (when expressed as a standard deviation) is often proportional to the mean.

When extreme values are present, the interquartile range may be more representative of the dispersion of the data than the standard deviation. This statistical quantity is the difference of the 75^{th} and 25^{th} percentiles and therefore, is not influenced by extreme values.

2.2.4 Measures of Association

Data sets often include measurements of several characteristics (variables) for each sampling point. There may be interest in understanding the relationship or level of association between two or more of these variables. One of the most common measures of association is the correlation coefficient. The correlation coefficient measures the relationship between two variables, such as a linear relationship between two sets of measurements. Note that the correlation coefficient does not imply cause and effect. The analyst may say the correlation between two variables is high and the relationship is strong, but may not say an increase or decrease in one variable causes the other variable to increase or decrease without further evidence and strong statistical controls.

2.2.4.1 Pearson's Correlation Coefficient

The Pearson correlation coefficient measures the strength of the linear relationship between two variables. A linear association implies that as one variable increases, the other increases or decreases linearly. Values of the correlation coefficient close to +1 (positive correlation) imply that as one variable increases, the other increases nearly linearly. On the other hand, a correlation coefficient close to –1 implies that as one variable increases, the other decreases nearly linearly. Values close to 0 imply little linear correlation between the variables. When data are truly independent, the correlation between data points is zero (note, however, that a correlation of 0 does not necessarily imply independence). Directions and an example for calculating Pearson's correlation coefficient are contained in Box 2-6.

Box 2-6: Directions for Calculating Pearson's Correlation Coefficient with an Example

Let $X_1, X_2, \ldots, X_n$ represent one variable of the n data points and let $Y_1, Y_2, \ldots, Y_n$ represent a second variable of the n data points. The Pearson correlation coefficient, r, between X and Y is:

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{\sqrt{\left[\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2\right] \cdot \left[\sum_{i=1}^{n} Y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2\right]}}$$

Example: Consider the following data set (in ppb):

Sample No.	1	2	3	4
Arsenic	8.0	6.0	2.0	1.0
Lead	8.0	7.0	7.0	6.0

$$\sum_{i=1}^{4} X_i = 17, \sum_{i=1}^{4} Y_i = 28, \sum_{i=1}^{4} X_i^2 = 105, \sum_{i=1}^{4} Y_i^2 = 198, \sum_{i=1}^{4} X_i Y_i = 8 \cdot 8 + 6 \cdot 7 + 2 \cdot 7 + 1 \cdot 6 = 126$$

$$\text{and } r = \frac{126 - \frac{17 \cdot 28}{4}}{\sqrt{\left(105 - \frac{17 \cdot 17}{4}\right)\left(198 - \frac{28 \cdot 28}{4}\right)}} = 0.865$$

Since r is close to 1, there is a strong linear relationship between these two contaminants.

The correlation coefficient does not detect nonlinear relationships so it should be used only in conjunction with a scatter plot (Section 2.3.7.2). A scatter plot can be used to determine if the correlation coefficient is meaningful or if some measure of nonlinear relationships should be used. The correlation coefficient can be significantly influenced by extreme values so a scatter plot should be used first to identify such values.

An important property of the correlation coefficient is that it is unaffected by changes in location of the data (adding or subtracting a constant from all of the X measurements or all the Y measurements) or by changes in scale of the data (multiplying or dividing all X or all Y values by a positive constant). Thus, linear transformations on the Xs and Ys do not affect the correlation of the measurements. This is reasonable since the correlation reflects the degree to which linearity between X and Y measurements occur and the degree of linearity is unaffected by changes in location or scale. For example, if a variable was temperature in Celsius, then the correlation would not change if Celsius was converted to Fahrenheit.

On the other hand, if nonlinear transformations of the X and/or Y measurements are made, then the Pearson correlation between the transformed values will differ from the correlation of the original measurements. For example, if X and Y, respectively, represent PCB and dioxin concentrations in soil, and $x = \log(X)$ and $y = \log(Y)$, then the Pearson correlations between X and

Y, *X* and *y*, *x* and *Y*, and *x* and *y*, will all be different, in general, since the logarithmic transformation is a nonlinear transformation.

Pearson's correlation may be sensitive to the presence of one or two extreme values, especially when sample sizes are small. Such values may result in a high correlation, suggesting a strong linear trend, when only moderate trend is present. This may happen, for instance, if a single (*X*,*Y*) pair has very high values for both *X* and *Y* while the remaining data values are uncorrelated. Extreme values may also lead to low correlations between *X* and *Y*, thus tending to mask a strong linear trend. This may happen if all the (*X*,*Y*) pairs except one (or two) tend to cluster tightly about a straight line, and the exceptional point has a very large *X* value paired with a moderate or small *Y* value (or vice versa). As influences of extreme values can be important, it is wise to use a scatter plot (Section 2.3.7.2) in conjunction with a correlation coefficient.

2.2.4.2 Spearman's Rank Correlation

An alternative to the Pearson correlation is Spearman's rank correlation coefficient. It is calculated by first replacing each *X* value by its rank (i.e., 1 for the smallest *X* value, 2 for the second smallest *X* value, etc.) and each *Y* value by its rank. These pairs of ranks are then treated as the (*X*,*Y*) data and Spearman's rank correlation is calculated using the same formulae as for Pearson's correlation (Box 2-6). Directions and an example for calculating the Spearman's Rank correlation coefficient are contained in Box 2-7.

Since meaningful (i.e., monotonic increasing) transformations of the data will not alter the ranks of the respective variables (e.g., the ranks for log (*X*) will be the same for the ranks for *X*), Spearman's correlation will not be altered by nonlinear increasing transformations of the *X*s or the *Y*s. For instance, the Spearman correlation between PCB and dioxin concentrations (*X* and *Y*) in soil will be the same as the correlations between their logarithms (*x* and *y*). This desirable property, and the fact that Spearman's correlation is less sensitive to extreme values, makes Spearman's correlation a good alternative or complement to Pearson's correlation coefficient.

2.2.4.3 Serial Correlation

For a sequence of data points taken serially in time, or one-by-one in a row, the serial correlation coefficient can be calculated by replacing the sequencing variable by the numbers 1 through *n* and calculating Pearson's correlation coefficient with *x* being the actual data values, and *y* being the numbers 1 through *n*. For example, for a sequence of data collected at a waste site along a straight transit line, the distances on the transit line of the data points are replaced by the numbers 1 through *n*, e.g., first 10-foot sample point = 1, the 20-foot sample point = 2, the 30-foot sample point = 3, etc., for samples taken at 10-foot intervals. Directions for the Serial correlation coefficient, along with an example, are given in Box 2-8.

Box 2-7: Directions for Calculating Spearman's Correlation Coefficient with an Example

Let $X_1, X_2, \ldots, X_n$ represent a set of ranks of the n data points of one data set and let $Y_1, Y_2, \ldots, Y_n$ represent a set of ranks of a second variable of the n data points. The Spearman Rank correlation coefficient, r, between X and Y is computed by:

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{\sqrt{\left[\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2\right] \cdot \left[\sum_{i=1}^{n} Y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2\right]}}$$

Example: Concentrations of arsenic and lead are taken at 4 sample locations. The following table gives the data set (in ppb) ranked according to the arsenic values. Ranks are given in parentheses. Note that any tied values are assigned average rank.

Sample No.	4	3	2	1
Arsenic	1.0 (1)	2.0 (2)	6.0 (3)	8.0 (4)
Lead	6.0 (1)	7.0 (2.5)	7.0 (2.5)	8.0 (4)

Now, $\sum_{i=1}^{4} X_i = 10, \sum_{i=1}^{4} Y_i = 10, \sum_{i=1}^{4} X_i^2 = 30, \sum_{i=1}^{4} Y_i^2 = 29.5, \sum_{i=1}^{4} X_i Y_i = 1 \cdot 1 + 2 \cdot 2.5 + 3 \cdot 2.5 + 4 \cdot 4 = 29.5$

$$\text{and } r = \frac{29.5 - \frac{10 \cdot 10}{4}}{\sqrt{\left(30 - \frac{10 \cdot 10}{4}\right)\left(29.5 - \frac{10 \cdot 10}{4}\right)}} = 0.9487$$

Since r is close to 1, there is a strong linear relationship between these two contaminants when ranked.

Box 2-8: Directions for Estimating the Serial Correlation Coefficient with a Example

Directions: Let $X_1, X_2, \ldots, X_n$ represent the data values collected in sequence over equally spaced periods of time. Label the periods of time 1, 2, ..., n to match the data values. Use the directions in Box 2-6 to calculate the Pearson's Correlation Coefficient between the data X and the time periods Y.

Example: The following are hourly readings from a discharge monitor.

Time	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00	24:00
Reading	6.5	6.6	6.7	6.4	6.3	6.4	6.2	6.2	6.3	6.6	6.8	6.9	7.0
Time Periods	1	2	3	4	5	6	7	8	9	10	11	12	13

Using Box 2-6, with the readings being the X values and the Time Periods being the Y values gives a serial correlation coefficient of 0.4318.

2.3 GRAPHICAL REPRESENTATIONS

2.3.1 Histogram

A histogram is a visual representation of the data collected into groups. This graphical technique provides a visual method of identifying the underlying distribution of the data. The data range is divided into several bins or classes and the data is sorted into the bins. A histogram is a bar graph conveying the bins and the frequency of data points in each bin. Other forms of the histogram use a normalization of the bin frequencies for the heights of the bars. The two most common normalizations are relative frequencies (frequencies divided by sample size) and densities (relative frequency divided by the bin width). Figure 2-1 is an example of a histogram using frequencies and Figure 2-2 is a histogram of densities.

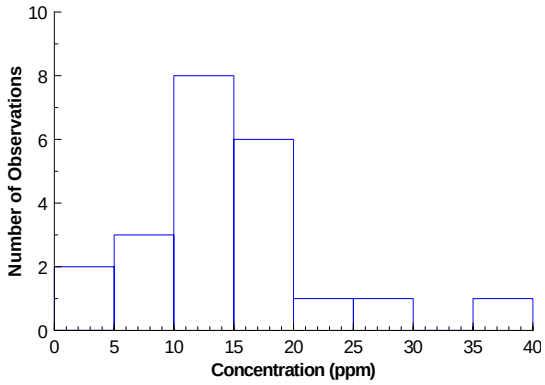


Figure 2-1. A Histogram of Concentration Frequencies

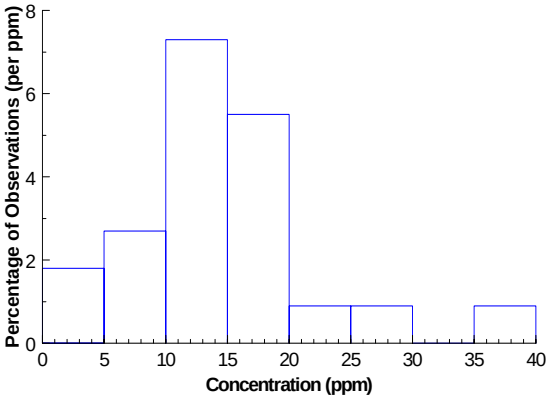


Figure 2-2. A Histogram of Concentration Densities

Histograms provide a visual method of accessing location, shape and spread of the data. Also, extreme values and multiple modes can be identified. The details of the data are lost, but an overall picture of the data is obtained. The stem-and-leaf plot described in the next section offers the same insights into the data as a histogram, but the data values are retained. Therefore, stem-and-leaf plots can be more informative than histograms for smaller data sets.

The visual impression of a histogram is sensitive to the choice of bins. A large number of bins will increase data detail while fewer bins will increase the smoothness of the histogram. A good starting point when choosing the number of bins is the square-root of the sample size. Another factor in choosing bins is the choice of endpoints. Using simple bin endpoints can improve the readability of the histogram. Simple bin endpoints include multiples of 5k units for some integer *k*. For example, 0 to <5, 5 to <10, etc. (Figure 2-1), or 1 to <1.5, 1.5 to <2, etc. Finally, when plotting a histogram for a continuous variable, e.g., concentration, it is necessary to decide on an endpoint convention; that is, what to do with data points that fall on the boundary of a bin. With discrete variables, (e.g., family size) the intervals can be centered in between the variables. For the family size data, the intervals can span between 1.5 and 2.5, 2.5 and 3.5, and so on. Then the whole numbers that relate to the family size can be centered within the box. Directions for generating a histogram are contained in Box 2-9 and an example is contained in Box 2-10.

Box 2-9: Directions for Generating a Histogram

STEP 1: Partition the data range into several bins. Partitioning means the bins should be connected, but not overlapping. For example, 0 to <5, 5 to <10, etc. In almost all cases, the bin widths should be the same. When selecting bins, first choose a number of bins. A good rule of thumb is the square root of the sample size. Next, select bin endpoints that are simple (for example, multiples of 5k for some integer *k*) and will produce approximately the number of desired bins.

STEP 2: Place each data point into the proper bin. This creates a summary of the data called a frequency table. If desired, compute one or both types of normalizations of the frequencies. The relative frequency is the frequency divided by the sample size. The density is the relative frequency divided by the bin width.

STEP 3: Determine the horizontal axis based on the range of the frequencies or normalized frequencies.

STEP 4: A histogram is a bar graph of the frequencies or normalized frequencies. For each bin, draw a bar using the bin endpoints on the x-axis as the width and the frequency or normalized frequency on the y-axis as the height.

Box 2-10: Example of Generating a Histogram

Consider the following 22 measurements of a contaminant concentration (in ppm): 17.7, 17.4, 22.8, 35.5, 28.6, 17.2, 19.1, <4, 7.2, <4, 15.2, 14.7, 14.9, 10.9, 12.4, 12.4, 11.6, 14.7, 10.2, 5.2, 16.5, and 8.9.

STEP 1: With 22 data points, a rough estimate of the number of bins is $\sqrt{22} = 4.69$ or 5. Since the data ranges from 0 to 40, the suggested bins are 0 to <8, 8 to <16, etc. This is a little awkward and using multiples of 5 to create simpler bins leads to 0 to <5, 5 to <10, etc. This choice leads to 8 bins which is close to suggested number and are the bins that will be used.

STEP 2: Column 2 of the frequency table below shows the frequency or number of observations within each bin defined in Step 1. The third column shows the relative frequencies which is the frequency divided by the sample size. The final column of the table gives the densities or the relative frequencies divided by the bin widths.

STEP 3: The horizontal axis for the data is from 0 to 40 ppm. The vertical axis for the histogram of frequencies is from 0 - 10 and the vertical axis for the histogram of densities is from 0% - 10%.

STEP 4: The histogram of frequencies is shown in Figure 2-1 and the histogram of densities is shown in Figure 2-2.

Bin	Frequency	Relative Frequency	Density
0 - 5 ppm	2	9.09	1.8
5 - 10 ppm	3	13.64	2.7
10 - 15 ppm	8	36.36	7.3
15 - 20 ppm	6	27.27	5.5
20 - 25 ppm	1	4.55	0.9
25 - 30 ppm	1	4.55	0.9
30 - 35 ppm	0	0.00	0.0
35 - 40 ppm	1	4.55	0.9

2.3.2 Stem-and-Leaf Plot

The stem-and-leaf plot is used to show both the data values and visual information about the distribution of the data. Like a histogram, a stem-and-leaf plot is visual representation

of the data collected into groups. However, data detail is retained in a stem-and-leaf. As with histograms, stem-and-leaf plots provide an understanding of the shape of the data (and the underlying distribution), i.e., location, shape, spread, number of modes and identification of potential outliers. Stem-and-leaf is also sensitive to bin selection. A stem-and-leaf plot can be more useful than a histogram in analyzing small data sets because it not only allows a visualization of the data distribution, but enables the data to be reconstructed.

First, bins that divide the data range are selected in a similar manner as for a histogram; these are the stems. Since the number of data points is typically small the number of bins or stems should be approximately 4 to 8. Data points are then sorted into the proper stem. Each observation in the stem-and-leaf plot consists of two parts: the stem and the leaf. The stem is generally made up of the leading digit or digits of the data values while the leaf is made up of the trailing digit or digits. The stems are displayed on the vertical axis and the trailing digits of the data points make up the leaves. Changing the stem can be accomplished by increasing or decreasing the leading digits that are used, dividing the groupings of one stem (i.e., all numbers which start with the numeral 6 can be divided into smaller groupings), or multiplying the data by a constant factor (i.e., multiply the data by 10 or 100). Non-detects can be placed at the detection limit with the leaf indicating the observation was actually a nondetect. Directions for constructing a stem-and-leaf plot are given in Box 2-11 and an example is contained in Box 2-12.

Box 2-11: Directions for Generating a Stem and Leaf Plot

Let $X_1, X_2, \ldots, X_n$ represent the n data points. To develop a stem-and-leaf plot, complete the following steps:

STEP 1: Arrange the observations in ascending order. The ordered data is labeled (from smallest to largest) $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$.

STEP 2: Choose either one or more of the leading digits to be the stem values. For example, if data ranges from 0 to 30, then the best choice for stems would be the values of the tens column, 0, 1, 2, 3.

STEP 3: List the stem values from smallest to largest along the vertical axis. Enter the trailing digits for each data point as the leaf. The leaves extend to the right of the stems. Continuing the example from above, the values of the ones column (and possible one decimal place) would be used as the leaf.

Box 2-12: Example of Generating a Stem and Leaf Plot

Consider the following 22 samples of trifluorine (in ppm): 17.7, 17.4, 22.8, 35.5, 28.6, 17.2, 19.1, <4, 7.2, <4, 15.2, 14.7, 14.9, 10.9, 12.4, 12.4, 11.6, 14.7, 10.2, 5.2, 16.5, and 8.9.

STEP 1: Arrange the observations in ascending order: <4, <4, 5.2, 7.7, 8.9, 10.2, 10.9, 11.6, 12.4, 12.4, 14.7, 14.7, 14.9, 15.2, 16.5, 17.2, 17.4, 17.7, 19.1, 22.8, 28.6, 35.5.

STEP 2: The data ranges from 0 to 40 so the best choice for the stems is the tens column. Initially, this gives 4 stems. We can divide the stems in half to give stems of 0 to <5, 5 to <10, etc. This gives 8 stems.

STEP 3: List the stem values along the vertical axis from smallest to largest. Enter the leaf (the remaining digits) values in order from lowest to highest to the right of the stem.

```
0 | <4  <4
0 | 5.2  7.7  8.9
1 | 0.2  0.9  1.6  2.4  2.4  4.7  4.7  4.9
1 | 5.2  6.5  7.2  7.4  7.7  9.1
2 | 2.8
2 | 8.6
3 |
3 | 5.5
```

Inspection of the stem-and-leaf indicates the data values are centered in the low teens with a slightly right-skewed distribution. Summary statistics can be calculated to strengthen these visual conclusions.

2.3.3 Box-and-Whiskers Plot

Box and Whisker plots, also known as box-plots, are useful in situations where a picture of the distribution is desired, but it is not necessary or feasible to portray all the details of the data. A box-plot (see Figure 2-3) displays several percentiles of the data set. It is a simple plot, yet provides insight into the location, shape, and spread of the data and underlying distribution. A simple box-plot contains only the 0th (minimum data value), 25th, 50th, 75th and 100th (maximum data value) percentiles. A more complex version includes identification of the mean with a plus-sign and potential outliers (identified by stars). Since the box-plot is compact (essentially one-dimensional), several can be placed on to a single graph. This allows a simple method to compare the locations, spreads and shapes of several data sets or different groups within a single data set. In this situation, the width of the box can be proportional to the sample size of each data set.

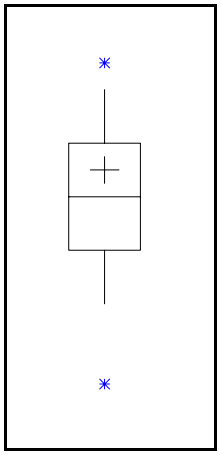


Figure 2-3. Example of a Box-Plot

A box-plot divides the data into 4 sections, each containing 25% of the data. The length of the central box indicates the spread of the data (the central 50%) while the length of the whiskers shows the breadth of the tails of the distribution. The box-plot also demonstrates the shape of the data in the following manner. If the upper box and whisker are approximately the same length as the lower box and whisker, then the data are distributed symmetrically. If the upper box and whisker are longer than the lower box and whisker, then the data are right-

skewed. If the upper box and whisker are shorter than the lower box and whisker, then the data are left-skewed. Directions for generating a box and whiskers plot are contained in Box 2-13, and an example is contained in Box 2-14.

If the mean is added to the box-plot, then recall comparing the mean and median provides another method of identifying the shape of the data. If the mean is approximately equal to the median, then the data are distributed symmetrically. If the mean is greater than the median, then the data are right-skewed; if the mean is less than the median, then the data are left-skewed.

Box 2-13: Directions for Generating a Box and Whiskers Plot

STEP 1: Compute the 0^{th} (minimum value), 25th, 50^{th} (median), 75th and 100^{th} (maximum value) percentiles.

STEP 2: Plot these points either vertically or horizontally. If more than one box-plot is drawn, then the other axis can be used to identify the box-plots of the different groups or data sets. Draw a box around the 25th and 75th percentiles and add a line through the box at the 50th percentile. Optionally, make the width of the box proportional to the sample size.

STEP 3: If desired, compute the mean and indicate this value on the box-plot with a plus-sign. Also, identify potential outliers if desired. A potential outlier is a value at a distance greater than 1.5×IQR from the closest end of the box.

STEP 4: Draw the whiskers from each end of the box to the furthest data point that has not been identified as an outlier. Plot the potential outliers using asterisks.

Box 2-14: Example of a Box and Whiskers Plot

Consider the following 22 samples of trifluorine (in ppm) listed in order from smallest to largest: 4.0, 6.1, 9.8, 10.7, 10.8, 11.5, 11.6, 12.4, 12.4, 14.6, 14.7, 14.7, 16.5, 17.0, 17.5, 20.6, 20.8, 25.7, 25.9, 26.5, 32.0, and 35.5.

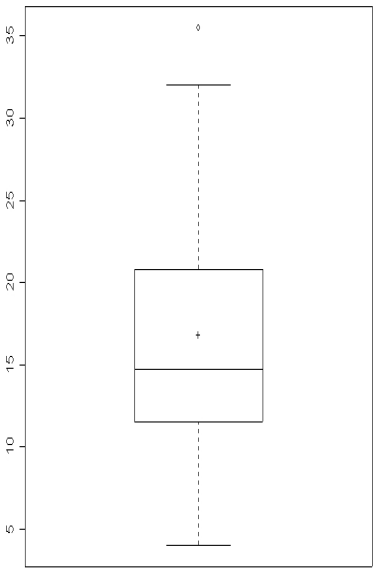
STEP 1: The minimum value is 4.0, $y(25) = 11.5$, the median is 14.7, $y(75) = 20.8$, and the maximum value is 35.5.

STEP 2: The simple box-plot is drawn using the summary statistics calculated in step 1

STEP 3: The mean of the data is 16.9. The only potential outlier is 35.5 because 35.5 – 20.8 = 14.7 which is greater than 1.5×IQR = 13.95.

STEP 4: The lower whisker extends from 11.5 to 4.0 and the upper whisker extends from 20.8 to 32.0

Examining the box-plot, it is easy to see that the data are centered at 15 ppm and 50% of the data lie between approximately 12 ppm and 21 ppm. Also, since the upper box and whisker are longer than the lower box and whisker, the distribution of the data is right-skewed.



2.3.4 Quantile Plot and Ranked Data Plots

The quantile plot and the ranked data plot are two further methods of visualizing the location, spread and shape of the data. Both plots are displays of the ranked data and differ only in the horizontal axis labels. No subjective decisions such as bin size are necessary and all of the data is plotted rather than summaries.

A ranked data plot is a display of the data from smallest to largest at evenly spaced intervals. A quantile plot is a graph of the ranked data versus the fraction of data points it exceeds. Therefore, a quantile plot can be used to read the quantile information such as the median, quartiles, and the interquartile range. This additional information can aid in the interpretation of the shape of the data.

Using either the quantile plot or the ranked data plot, the spread of the data may be visualized by examining the slope of the graph. The closer the general slope is to 0, the lower the variability of the data set.

Also using either plot, the shape of the data maybe determined by inspecting the tails of the graph. If the left and right tails have approximately the same curvature, then the data are distributed symmetrically. If the curvature of the right-tail is greater than the curvature of the left-tail, then the data are right-skewed. This is the case of the data plotted in Figure 2-4. If the curvature of the left-tail is greater than the curvature of the right-tail, then the data are left-skewed. Finally, the degree of curvature in the tails of either plot is proportional to the length of the tail of the data. In Figure 2-4, the plot rises slowly to a point, then the slope increases dramatically. This implies there is a large concentration of small data values and relatively few large data values. In this case, the left-tail of the data is very short and the right-tail is relatively long.

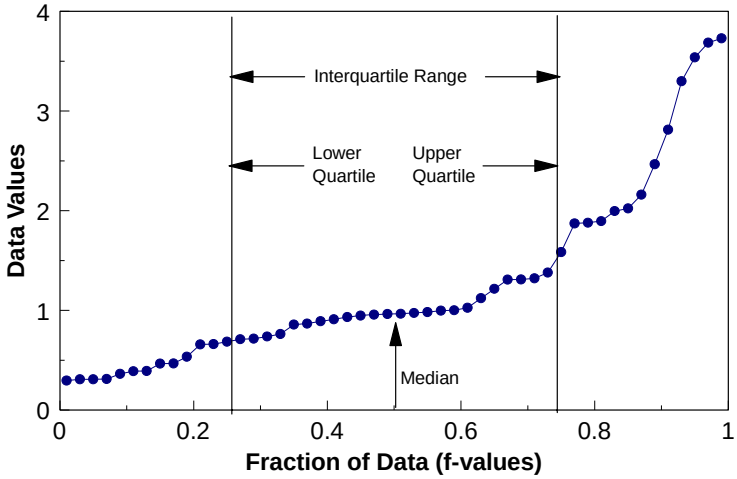


Figure 2-4. Example of a Quantile Plot of Right-Skewed Data

Directions for developing a quantile plot are given in Box 2-15 and an example is given in Box 2-16.

Box 2-15: Directions for Generating a Quantile Plot and a Ranked Data Plot

Let $X_1, X_2, \ldots, X_n$ represent the n data points. and $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be the ordered data from smallest to largest . Compute the fractions $f_i = (i - 0.5)/n$ for $i = 1,\ldots,n$. The quantile plot is a plot of the pairs $(f_i, X_{(i)})$, with straight lines connecting consecutive points. If desired, add vertical lines indication the quartiles, median, or other quantiles of interest.

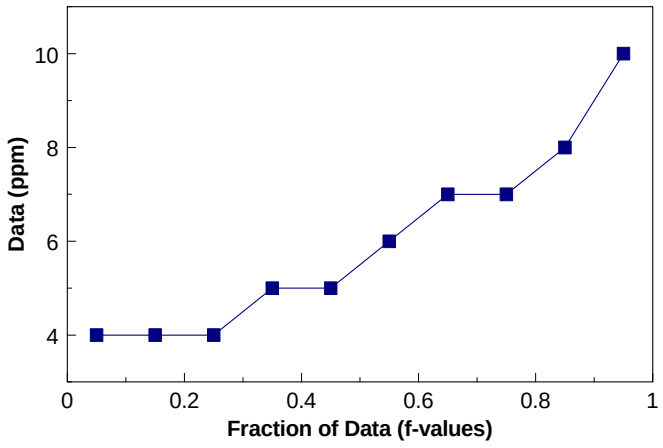
Alternatively, a ranked data plot can be made by simply plotting the ordered X values at equally spaced intervals along the horizontal axis.

Box 2-16: Example of Generating a Quantile Plot

Consider the following 10 data points (in ppm): 4, 5, 6, 7, 4, 10, 4, 5, 7, and 8. The data ordered from smallest to largest, $X_{(i)}$, are shown in the second row of the table below and the third row displays the values f_i for each i where $f_i = (i - 0.5)/n$.

i	1	2	3	4	5	6	7	8	9	10
$X_{(i)}$	4	4	4	5	5	6	7	7	8	10
f_i	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95

The pairs $(f_i, X_{(i)})$ are then plotted to yield the following quantile plot:



Note that the graph curves upward. This indicates that the data are skewed to the right.

2.3.5 Quantile-Quantile Plots and Probability Plots

There are two types of quantile-quantile plots or q-q plots. One is an empirical q-q plot which involves plotting the quantiles of two data sets against each other. This is a technique to determine if the data sets were generated by the same underlying distribution and is discussed in Section 2.3.6.3. The other type of q-q plot involves graphing the quantiles of a data set against the quantiles of a specific probability distribution. This is a technique to determine if the data set was generated by the theoretical distribution. The following section will focus on the most common of these plots for environmental data, the normal probability plot (normal q-q plot). However, the discussion holds for other probability distributions. The normal probability plot is

a visual method to roughly determine how well the data set is modeled by a normal distribution. Formal tests are contained in Chapter 4, Section 2. Directions for developing a normal probability plot are given in Box 2-17 and an example is given in Box 2-18. A discussion of the normal distribution is contained in Section 2.4.

A normal probability plot is the graph of the quantiles of a data set against the quantiles of the standard normal distribution. This can be accomplished by using a software package, plotting the sample quantiles against standard normal quantiles, or plotting the sample quantiles on normal probability paper. If the graph is approximately linear (the correlation coefficient being fairly high excluding outliers), then this is an indication that the data are normally distributed and a formal test should be performed. If the graph is not linear, then the departures from linearity give important information about how the data distribution deviates from a normal distribution.

A nonlinear normal probability plot may be used to interpret the shape and tail behavior of the data. First, add the quartile line, the line through the first and third quartiles, to the plot. Next, examine the relationship of the tails of the normal probability plot to the quartile line.

Relationship of the q-q plot to the quartile line		*Distribution of the data in relation to the normal distribution*	
upper-tail	**lower-tail**	**shape**	**tail behavior**
above	below	symmetric	heavy
below	above	symmetric	light
above	above	right-skewed	-
below	below	left-skewed	-

A normal probability plot can also be used to identify potential outliers. A data value (or a few data values) much larger or much smaller will appear distant from the other values, which will be concentrated in the center of the graph.

As a final note, using a simple natural log transformation of the data, the normal probability plot can be used to determine if the data are well modeled by a lognormal distribution. The lognormal is an important probability distribution when analyzing environmental data where normality cannot be assumed.

2.3.6 Plots for Two or More Variables

Data often consist of measurements of several characteristics (variables) for each sample point in the data set. For example, a data set may consist of measurements of lead, mercury, and chromium for each soil sample or may consist of daily concentration readings for several monitoring sites. In this case, graphs may be used to compare and contrast different variables.

Box 2-17: Directions for Constructing a Normal Probability Plot

Let $X_1, X_2, \ldots, X_n$ represent the n data points. Many software packages will produce normal probability plots and this is the recommended method. If one of these packages is unavailable, then use the following procedure.

STEP 1: Compute the fractions $f_i = (i - 0.5)/n$ for $i = 1,\ldots,n$.

STEP 2: Compute the normal quantiles for the f_i. These are $Y_i = \Phi^{-1}(f_i)$, where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative density function. The Y_i can be found using Table A-1: locate an f_i in the body of the table, then the Y_i is the found on the edges. For example, if $f_i = 0.975$, then Y_i is 1.96.

STEP 3: Plot the pairs (Y_i, X_i), for $i = 1, \ldots,n$.

STEP 4: If desired, then add the quartile line through the first and third quartiles. This acts as a guide to determine if the points follow a straight line, which may be fitted to the data.

If the graph of these pairs approximately form a straight line, then the data may be well modeled by a normally distribution. Otherwise, use the table above to determine the shape and tail behavior.

Box 2-18: Example of Normal Probability Plot

Consider the following 15 data points: 5, 5, 6, 6, 8, 8, 9, 10, 10, 10, 10, 10, 12, 14, and 15.

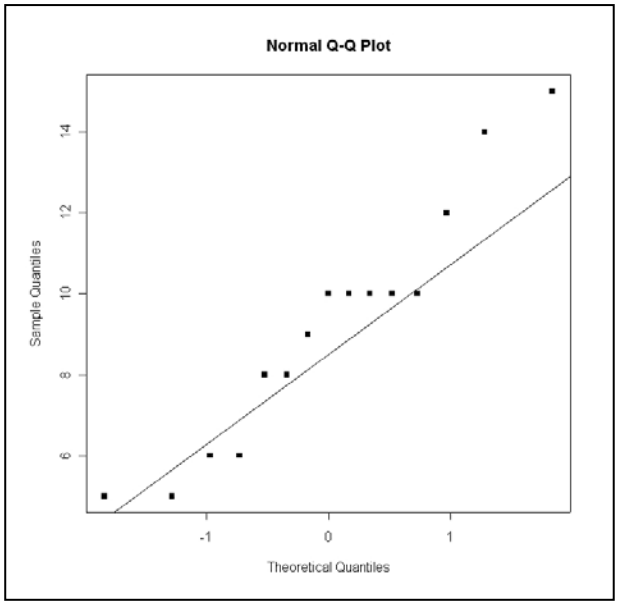
STEP 1: The computed fractions, f_i, are: 0.0333, 0.1000, 0.1667, 0.2333,…, 0.9667.

STEP 2: Working inside-out in Table A-1, the normal quantiles for the f_i are: -1.83, -1.28, -0.97, -0.73, -0.52, -0.34, -0.17, 0.00, 0.17, 0.34, 0.52, 0.73, 0.97, 1.28, 1.83.

STEP 3: The data and normal quantile pairs are plotted in the graph below.

STEP 4: Using Box 2-1, the first and third quartiles are 7 and 10, respectively. The first and third quartiles of the standard normal distribution are –0.67 and 0.67, respectively. The quartile line is the line through the points (-0.67, 7) and (0.67, 10), and has been added to the plot.

The plot looks approximately linear, but deviates from the quartile line in the upper-tail. However, no definite conclusion should be drawn from a plot of 15 points. A formal test of normality (section 4.2) should be performed.



To compare and contrast several variables, collections of the single variable displays described in previous sections are useful. For example, the analyst may generate box-plots or histograms for each variable using the same axis for all of the variables. Separate plots for each variable may be overlaid on one graph, such as overlaying quantile plots for each variable. Another useful technique for comparing two variables is to place histograms (sometimes called bi-histograms) or stem-and-leaf plots back to back. Additional plots to display two or more variables are described in Sections 2.3.6.1 through 2.3.6.3. In many software packages, three dimensional scatter plots can be rotated and compared in order to find hidden relationships and anomalies.

2.3.6.1 Scatterplot

For data sets consisting of multiple observations per sampling point, a scatterplot is one of the most powerful graphical tools for analyzing the relationship between two or more variables. Scatterplots are easy to construct for two variables (Figure 2-5) and many software packages can construct 3-dimensional scatterplots. Directions for constructing a 2-dimensional scatterplot are given in Box 2-19 along with an example.

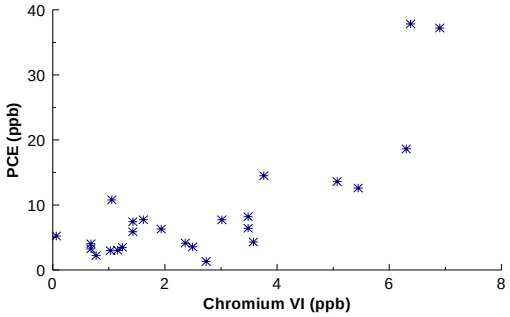


Figure 2-5. Example of a Scatterplot

Box 2-19: Directions for Generating a Scatterplot and an Example

Let $X_1, X_2, \ldots, X_n$ represent one variable of the n data points and let $Y_1, Y_2, \ldots, Y_n$ represent a second variable of the n data points. The paired data can be written as (X_i, Y_i) for $i = 1, \ldots, n$. To construct a scatterplot, plot one variable along the horizontal axis and the other variable along the vertical axis.

Example: PCE values are displayed on the vertical axis and Chromium VI values are displayed on the horizontal axis of Figure 2-5.

Obs	PCE (ppb)	Chromium VI (ppb)
1	14.49	3.76
2	37.21	6.92
3	10.78	1.05
4	18.62	6.30
5	7.44	1.43
6	37.84	6.38
7	13.59	5.07
8	4.31	3.56

Obs	PCE (ppb)	Chromium VI (ppb)
9	2.23	0.77
10	3.51	1.24
11	6.42	3.48
12	2.98	1.02
13	3.04	1.15
14	12.60	5.44
15	3.56	2.49
16	7.72	3.01

Obs	PCE (ppb)	Chromium VI (ppb)
17	4.14	2.36
18	3.26	0.68
19	5.22	0.65
20	4.02	0.68
21	6.30	1.93
22	8.22	3.48
23	1.32	2.73
24	7.73	1.61
25	5.88	1.42

A scatterplot can clearly show the relationship between two variables if the data range is sufficiently large. Truly linear relationships can always be identified in scatter plots, but truly nonlinear relationships may appear linear (or some other form) if the data range is relatively small. Scatterplots of linearly correlated variables cluster about a straight line. As an example of a nonlinear relationship, consider two variables where one variable is approximately equal to the square of the other. With an adequate range in the data, a scatterplot of this data would display a

partial parabolic curve. Other important modeling relationships that may appear are exponential or logarithmic. Two additional uses of scatterplots are the identification of potential outliers for a single variable or for the paired variables and the identification of clustering in the data.

2.3.6.2 Extensions of the Scatterplot

It is easy to construct a 2-dimensional scatterplot by hand and many software packages can construct a useful 3-dimensional scatterplot. However, with more than 3 variables, it is difficult to construct and interpret a scatterplot. Therefore, several graphical representations have been developed that extend the idea of a scatterplot for data consisting of more than 2 variables.

The simplest of these graphical techniques is a coded scatterplot. All possible two-way combinations are given a code and the pairs of data are plotted on one scatterplot. The coded scatterplot does not provide information on three-way or higher interactions between the variables. If the data ranges for the variables are comparable, then a single set of axes may suffice, if greater data ranges, different scales will be required. As an example, consider a data set of 3 variables, A, B, and C and assume the all of the data ranges are similar. An analyst may choose to display the pairs (A_i, B_i) using a small circle, the pairs (A_i, C_i) using a plus-sign, and the pairs (B_i, C_i) using a square on one scatterplot. The completed coded scatterplot is given in Figure 2-6.

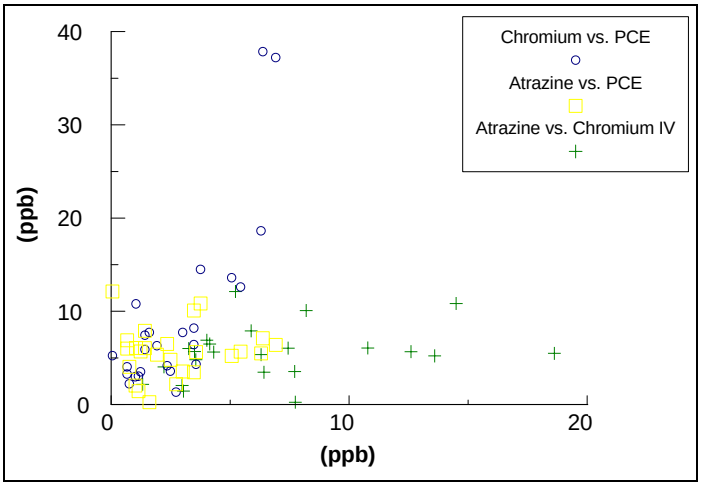


Figure 2-6. Example of a Coded Scatterplot

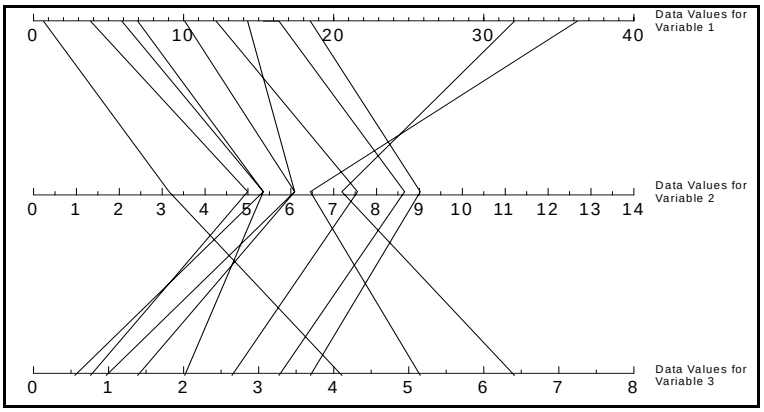


Figure 2-7. Example of a Parallel-Coordinate Plot

The parallel-coordinates method employs a scheme where coordinate axes are drawn in parallel (instead of perpendicular). Figure 2-7 is an example of a parallel-coordinate plot. Consider a sample point X consisting of values X_1 for measurement 1, X_2 for measurement 2, through X_p for measurement p. A parallel-coordinate plot is constructed by placing an axis for each of the p measurements in parallel and plotting X_1 on axis 1, X_2 on axis 2, and so on. The points are then

joined with a broken line. This graph contains all of the information of a scatterplot in addition to information concerning 3-way and higher interactions. However, with *p* measurements, one must construct several parallel-coordinate plots in order to display all possible pairs.

A scatterplot matrix is another useful method of extending scatterplots to higher dimensions. In this case, a scatterplot is created for all possible pairs of measurements which are then displayed in a matrix format. This method is easy to implement and provides a concise method of displaying the individual scatterplots. Interpretation proceeds as with a simple or coded scatterplot. As in those cases, this method does not provide information about 3-way or higher interactions. An example of a scatterplot matrix is contained in Figure 2-8.

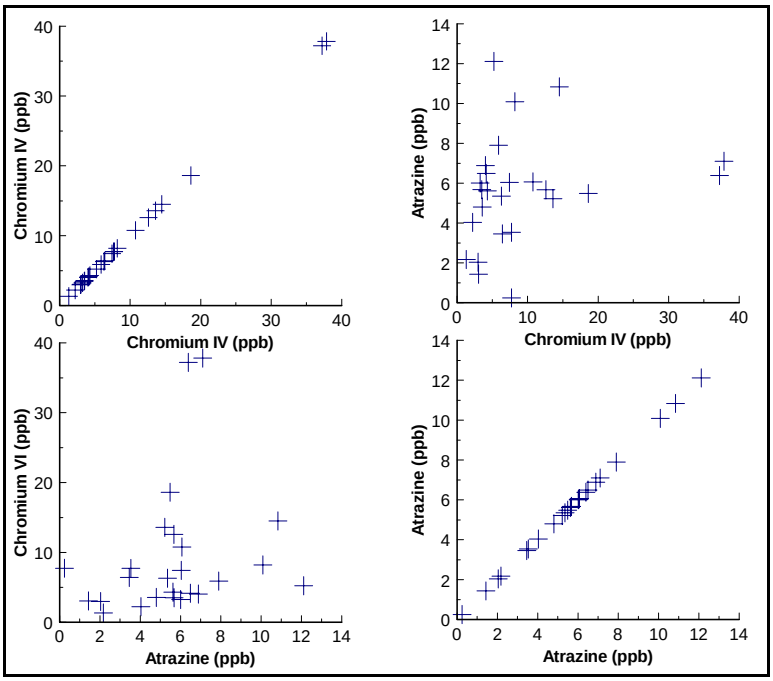


Figure 2-8. Example of a Scatterplot Matrix

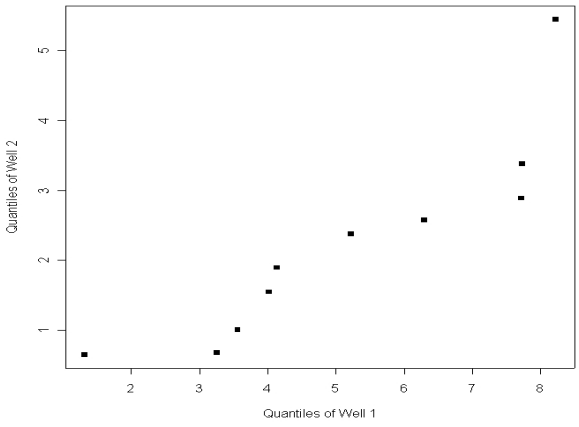


Figure 2-9. Example of an Empirical Q-Q Plot

2.3.6.3 Empirical Quantile-Quantile Plot

An empirical quantile-quantile (q-q) plot is a plot of the quantiles (Section 2.2.1) of two data sets against each other and is similar to the normal probability plot discussed in Section 2.3.5. This plot (Figure 2-9) is used to compare the distributions of two measurements; for example, the distributions of lead and iron concentrations in a drinking water well. This technique is used to determine if the data sets were generated from the same underlying distribution. If the graph is approximately linear, then the distributions are roughly the same. If the graph is not linear, then the departures from linearity give important information about how the two data distributions differ. The interpretation is similar to that of a normal probability plot. Directions for constructing an empirical q-q plot with an example are given in Box 2-20.

Box 2-20: Directions for Constructing an Empirical Q-Q Plot with an Example

Let $X_1, X_2,..., X_n$ represent n data points of one variable and let $Y_1, Y_2,..., Y_m$ represent a second variable of m data points. Let $X_{(1)}, X_{(2)},..., X_{(n)}$ and $Y_{(1)}, Y_{(2)},..., Y_{(m)}$ be the ordered data sets.

If $n = m$, then an empirical q-q plot of the two variables is simply a plot of the ordered values of the variables, i.e., a scatterplot of the pairs $(X_{(1)}, Y_{(1)})$, $(X_{(2)}, Y_{(2)})$, ..., $(X_{(n)}, Y_{(n)})$.

Suppose $n > m$. Then the empirical quantile-quantile plot will consist of m pairs of points with the ordered Y values plotted against the m evenly spaced quantiles of X. Using Box 2-1, compute the quantiles that correspond to the fractions $q_j = (j-1)/(m-1)$ for $j = 1,\ldots,m$.

Example: Consider contaminant readings from two separate drinking water wells at the same site.

well 1: 1.32, 3.26, 3.56, 4.02, 4.14, 5.22, 6.30, 7.72, 7.73, 8.22
well 2: 0.65, 0.68, 0.68, 1.42, 1.61, 1.93, 2.36, 2.49, 2.73, 3.01, 3.48, 5.44.

Since the sample sizes are not equal, we need to compute the 10 evenly space quantiles for well 2.

$q_1 = 0$, so the first quantile is $y(q_1) = 0.65$.
$q_2 = 1/9$. So $r = (n-1)\cdot q_2 + 1 = 11\cdot 1/9 + 1 = 20/9$, $i = \text{floor}(r) = 2$, and $f = r - i = 2/9$.
Therefore, the second quantile is $y(q_2) = (7/9)\cdot X_{(2)} + (2/9)\cdot X_{(3)} = 0.68$.

Continuing this process for $j = 2,\ldots,10$ yields the following 10 quantiles:

0.650, 0.680, 1.009, 1.547, 1.894, 2.374, 2.570, 2.886, 3.376, 5.440.

These, paired with the well 1 data, are plotted in Figure 2-9.

2.3.7 Plots for Temporal Data

Data collected over specific time intervals (e.g., monthly, biweekly, or hourly) have a temporal component. For example, air monitoring measurements of a pollutant may be collected once a minute or once a day; water quality monitoring measurements of a contaminant level may be collected weekly or monthly. An analyst examining temporal data may be interested in the trends over time, correlation among time periods, or cyclical patterns. Some graphical techniques specific to temporal data are the time plot, lag plot, correlogram, and variogram.

A data sequence collected at regular time intervals is called a time series. Time series data analysis is beyond the scope of this guidance. It is recommended that the interested reader consult a statistician. The graphical representations presented in this section are recommended for any data set that includes a temporal component regardless of the decision to perform a time series analysis. The graphical techniques described below will help identify temporal patterns that need to be accounted for in any analysis of the data.

The analyst examining temporal environmental data may be interested in seasonal trends, directional trends, serial correlation, or stationarity. Seasonal trends are patterns in the data that repeat over time, i.e., the data rise and fall regularly over one or more time periods. Seasonal trends may be large scale, such as a yearly cycle where the data show the same pattern of rising and falling from year to year, or the trends may be small scale, such as a daily cycle. Directional

trends are positive or negative trends in the data which is of importance to environmental applications where contaminant levels may be increasing or decreasing. Serial correlation is a measure of the strength of the linear relationship of successive observations. If successive observations are related, statistical quantities calculated without accounting for the serial correlation may be biased. Finally, another item of interest for temporal data is stationarity (cyclical patterns). Stationary data look the same over all time periods. Directional trends or a change in the variability in the data imply non-stationarity.

2.3.7.1 Time Plot

A time plot (also known as a time series plot) is simply a plot of the data over time. This plot makes it easy to identify lack of randomness, changes in location, change in scale, small-scale trends, or large-scale trends over time. Small-scale trends are displayed as fluctuations over smaller time periods. For example, ozone levels over the course of one day typically rise until the afternoon, then decrease, and this process is repeated every day. Larger scale trends, such as seasonal fluctuations, appear as regular rises and drops in the graph. For example, ozone levels tend to be higher in the summer than in the winter so ozone data tend to show both a daily trend and a seasonal trend. A time plot can also show directional trends or changing variability over time.

A time plot (Figure 2-10) is constructed by plotting the measurements on the vertical axis versus the actual time of observation or the order of observation on the horizontal axis. The points plotted may be connected by lines, but this may create an unfounded sense of continuity. It is important to use the actual time or number at which the observation was made. This can create discontinuities in the plot but are needed as the data that should have been collected now appear as "missing values" but do not disturb the integrity of the plot. Plotting the data at equally spaced intervals when in reality there were different time periods between observations is not advised.

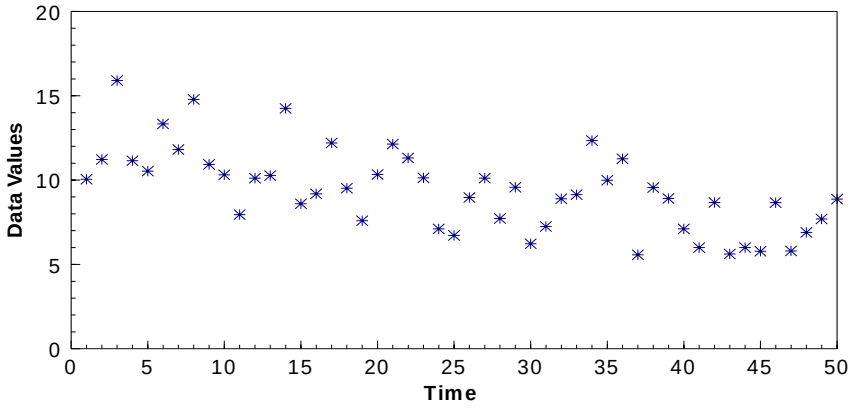


Figure 2-10. Example of a Time Plot

The scaling of the vertical axis of a time plot is of some importance. A wider scale tends to emphasize large-scale trends, whereas a narrower scale tends to emphasize small-scale trends. Using the ozone example above, a wide scale would emphasize the seasonal component of the data, whereas a smaller scale would tend to emphasize the daily fluctuations. Directions for constructing a time plot are contained in Box 2-21 along with an example.

Box 2-21: Directions for Generating a Time Plot and an Example

Let $X_1, X_2, \ldots, X_n$ represent n data points listed in order by time, i.e., the subscript represents the ordered time interval. A plot of the pairs (i, X_i) is a time plot of the data.

Example: Consider the following 50 daily observations listed in order by day:

10.05, 11.22, 15.90, 11.15, 10.53, 13.33, 11.81, 14.78, 10.93, 10.31,
7.95, 10.11, 10.27, 14.25, 8.60, 9.18, 12.20, 9.52, 7.59, 10.33,
12.13, 11.31, 10.13, 7.11, 6.72, 8.97, 10.11, 7.72, 9.57, 6.23,
7.25, 8.89, 9.14, 12.34, 9.99, 11.26, 5.57, 9.55, 8.91, 7.11,
6.04, 8.67, 5.62, 5.99, 5.78, 8.66, 5.80, 6.90, 7.70, 8.87.

A time plot is constructed by plotting the pairs (i, X_i) where i represents the number of the day and X_i represents the concentration level. The plot is shown in Figure 2-10. Note the slight negative trend.

2.3.7.2 Lag Plot

A lag plot is another method to determine if the data set or time series is random. Nonrandom structure in the lag plot implies nonrandomness in the data. Examples of nonrandom structure are linear patterns or elliptical patterns. A linear pattern implies the data contain a directional trend while an elliptical pattern implies the data contain a seasonal component.

If we have data points $X_1, X_2, \ldots, X_n$, then a lag plot is a scatterplot of the points (X_t, X_{t-k}) for some integer lag k, the most common being lags 1, 2, or 3. Figure 2-11 is a 1-lag plot for the data in the example from Box 2-21. Notice that there is a light linear structure suggestion a possible directional trend in the data. See Section 2.3.7.3 for higher lags.

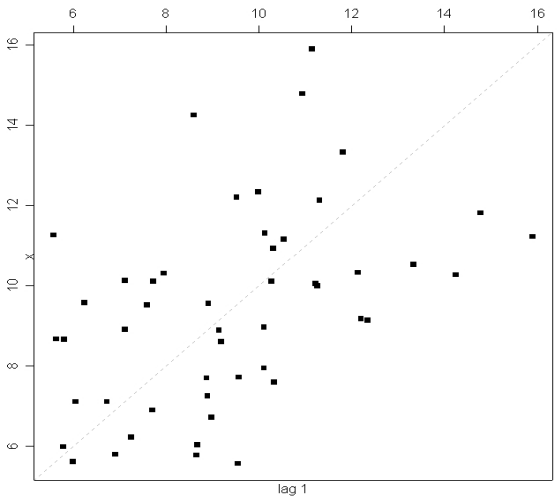


Figure 2-11. Example of a Lag Plot

2.3.7.3 Plot of the Autocorrelation Function (Correlogram)

Serial correlation is a measure of the strength of the relationship between successive observations. If successive observations are related, then the data must be transformed or the relationship must be accounted for in the data analysis. The correlogram is a plot that is used to display serial correlations when the data are collected at equally spaced time intervals. The autocorrelation function is a summary of the serial correlations of data. The 1st sample autocorrelation coefficient, r_1, is the correlation between points at lag 1 (points that are 1 time unit apart); r_2 is the correlation between points at lag 2; etc. A correlogram (Figure 2-12) is a plot of the sample autocorrelation coefficients, r_k, versus k.

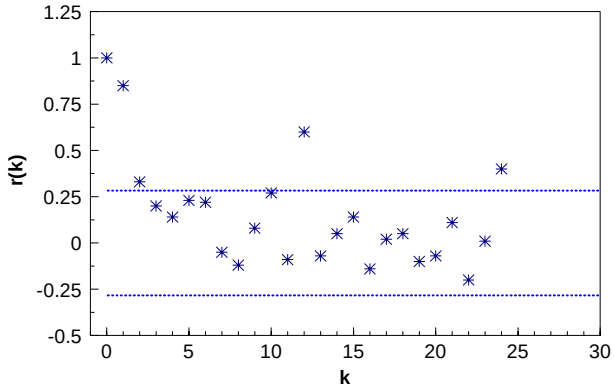


Figure 2-12. Example of a Correlogram

For a large independent data sequence of n time points, the autocorrelations are approximately normally distributed with mean zero and variance $1/n$. Therefore, to determine if the time points are independent, first plot the approximate 95% confidence lines $\pm 2/\sqrt{n}$ (shown as dashed lines in Figure 2-12) on the correlogram. If any of the autocorrelations lie outside the confidence lines, then there is evidence of serial correlation and we conclude that the time points are not independent.

In examining Figure 2-12, there are 4 time points that lie outside the 95% confidence bounds. These are at lags 1, 2, 12 and 24. This demonstrates strong evidence that the sequence is serially correlated (most likely containing a positive trend along with an annual component).

In application, the correlogram is only useful for data at equally spaced intervals and for irregular intervals a text on the geostatistical use of a variogram is recommended. Directions for constructing a correlogram are contained in Box 2-22; example calculations are contained in Box 2-23. For large sample sizes, a correlogram is tedious to construct by hand; therefore, statistical software should be used.

Box 2-22: Directions for Constructing a Correlogram

Let $X_1, X_2, \ldots, X_n$ represent the data points ordered by time for equally spaced time points, i.e., X_1 was collected at time 1, X_2 was collected at time 2, and so on. To construct a correlogram, first compute the sample autocorrelation coefficients, r_k. So for k = 0, 1, ..., compute

$$r_k = \frac{g_k}{g_0} \text{ and } g_k = \frac{1}{n}\sum_{t=k+1}^{n}\left(X_t - \bar{X}\right)\left(X_{t-k} - \bar{X}\right)$$

Once the r_k have been computed, a correlogram is the graph (k, r_k) for k = 0, 1, . . . , and so on. As an approximation, compute up to $k = n/6$. Also, note that $r_0 = 1$. Finally, place horizontal lines at $\pm 2/\sqrt{n}$.

2.3.7.4 Multiple Observations in a Time Period

Many times in environmental data collection, multiple samples are taken at each time point. For example, the data collection design may specify collecting 5 samples from a drinking well every Wednesday for three months. In this case, the time plot described in Section 2.3.7.1 may be used to display the complete data set, display the mean weekly level, display a confidence interval for each mean, or display a confidence interval for each mean with the individual data values. A time plot of all the data will allow the analyst to determine if the variability for the different collection periods changes. A time plot of the means will allow the

analyst to determine if the means are possibly changing between the collection periods. In addition, each collection period may be treated as a distinct variable and the methods described in Section 2.3.6 may be applied.

Box 2-23: Example Calculations for Generating a Correlogram

A correlogram will be constructed using the following four hourly data points: hour 1: 4.5 ppm, hour 2: 3.5 ppm, hour 3: 2.5 ppm, and hour 4: 1.5 ppm. Only four data points are used so all computations may be shown for illustrative purposes. Therefore, the guideline of computing no more than $n/6$ autocorrelation coefficients will be ignored. The first step to constructing a correlogram is to compute the sample mean (Box 2-2), which is 3. Then,

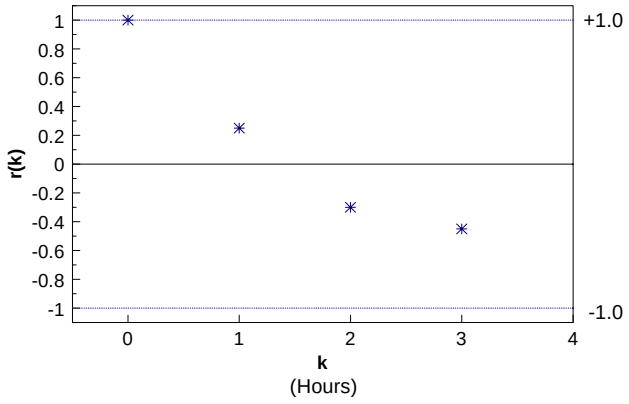
$$g_0 = \frac{1}{4}\sum_{t=1}^{4}\left(X_t - \overline{X}\right)\left(X_{t-0} - \overline{X}\right) = \frac{1}{4}\sum_{t=1}^{4}\left(X_t - \overline{X}\right)^2 = \frac{1}{4}\left[(4.5-3)^2 + (3.5-3)^2 + (2.5-3)^2 + (1.5-3)^2\right] = 1.25$$

$$g_1 = \frac{1}{4}\sum_{t=2}^{4}\left(X_t - \overline{X}\right)\left(X_{t-1} - \overline{X}\right) = \frac{1}{4}\left[(3.5-3)(4.5-3) + (2.5-3)(3.5-3) + (1.5-3)(2.5-3)\right] = 0.3125$$

Similarly, $g_2 = -0.375$ and $g_3 = -0.5625$.

So $r_1 = \frac{g_1}{g_0} = \frac{0.3125}{1.25} = 0.25$, $r_2 = \frac{g_2}{g_0} = \frac{-0.375}{1.25} = -0.3$ and $r_2 = \frac{g_3}{g_0} = \frac{-0.5625}{1.25} = -0.45$.

Thus, the correlogram of these data is a plot of (0, 1) (1, 0.25), (2, -0.3) and (3, -0.45) with two confidence lines at $\pm 2/\sqrt{n} = \pm 1$. This graph is shown below.



2.3.7.5 Four-Plot

A four-plot is a collection of 4 specific plots that provide different visual methods for illustrating a measurement process. There are 4 basic assumptions that underlie all measurement processes, there should be: random data, a fixed distribution, with a fixed location and variance. If all of these assumptions hold, then the measurement process or data set is considered to be stable. The fixed distribution we discuss here is a normal distribution, but any probability distribution of interest could be used. A four-plot consists of a time plot, a lag plot, a histogram and a normal probability plot. The data are random if the lag plot is structureless. The data are

from a normal distribution if the histogram is symmetric and bell-shaped and the normal probability plot is linear. If the time plot is flat and non-drifting, then the data have a fixed location. If the time plot has a constant spread, the data have a fixed variance. Figure 2-13 is a four-plot for the data contained in the example of Box 2-21. In this particular example, note that the data are not quite normal (deviations from the straight line on the plot), does not have a fixed location (a downward trend in the time plot), and possibly has serial correlation present (the tendency of the lag plot to be increasing from left to right).

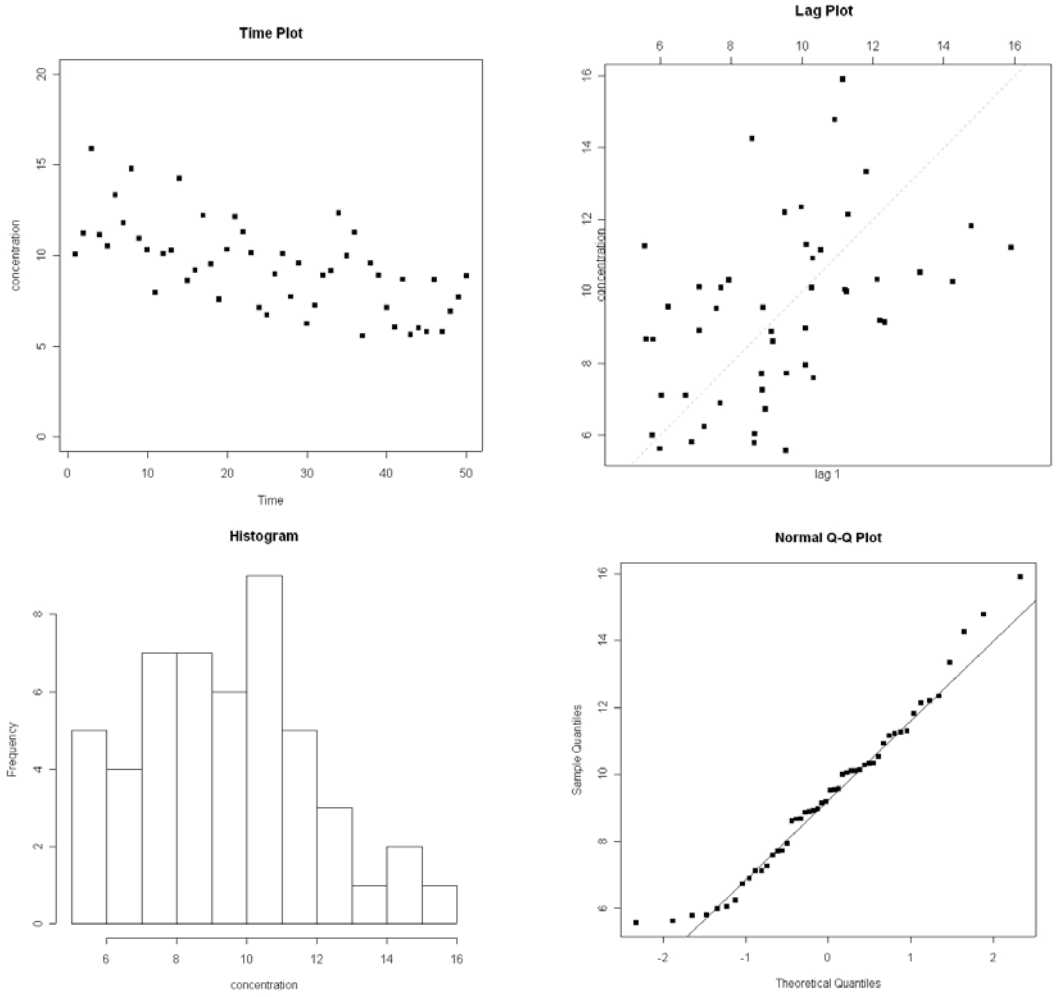


Figure 2-13. Example of a Four-Plot

2.3.8 Plots for Spatial Data

The graphical representations of the preceding sections may be useful for exploring spatial data. However, an analyst examining spatial data may be interested in the location of extreme values, overall spatial trends, and the degree of continuity among neighboring locations. Graphical representations for spatial data include postings, symbol plots, correlograms, h-scatter plots, and contour plots.

The graphical representations presented in this section are recommended for all spatial data regardless of whether or not geostatistical methods will be used to analyze the data. The graphical representations described below will help identify spatial patterns that need to be accounted for in the analysis of the data.

2.3.8.1 Posting Plots

A posting plot (Figure 2-14) is a map of data locations along with the corresponding data values. Data posting may reveal obvious errors in data location and identify data values that may be in error. The graph of the sampling locations gives the analyst an idea of where the data were collected (i.e., the sampling design), areas that may have been inaccessible, and areas of special interest to the decision maker which may have been heavily sampled. It is often useful to mark the highest and lowest values of the data to see if there are any obvious trends. If all of the highest concentrations fall in one region of the plot, the analyst may consider some method such as post-stratifying the data (stratification after the data are collected and analyzed) to account for this fact in the analysis. Directions for generating a posting plot are contained in Box 2-24.

Figure 2-14. Example of a Posting Plot

Box 2-24: Directions for Generating a Posting Plot and a Symbol Plot with an Example

On a map of the site, plot the location of each sample. At each location, either indicate the value of the data point (a posting plot) or indicate the data value by a symbol (a symbol plot) or circle (a bubble plot). The circles on a bubble plot are equal to the square roots of the data values.

Example: The spatial data displayed in the table below contains both a location (Northing and Easting) and a concentration level ([c]). The chosen symbols are floor (concentration/5) and range from '0' to '7.'. The data values with corresponding symbols are:

N	E	[c]	Sym
15	0	4.0	0
15	5	11.6	2
15	10	14.9	2
15	15	17.4	3
15	20	17.7	3
15	25	12.4	2
10	0	28.6	5
10	5	7.7	1

N	E	[c]	Sym
10	10	15.2	3
10	15	35.5	7
10	20	14.7	2
10	25	16.5	3
5	0	8.9	1
5	5	14.7	2
5	10	10.9	2

N	E	[c]	Sym
5	15	12.4	2
5	20	22.8	4
5	25	19.1	3
0	10	10.2	2
0	15	5.2	1
0	20	4.9	0
0	25	17.2	3

The posting plot of this data is displayed in Figure 2-14, the symbol plot is displayed in Figure 2-15, and the bubble plot is displayed in Figure 2-16.

2.3.8.2 Symbol Plots and Bubble Plots

For large amounts of data, a posting plot may be visual unappealing. In this case, a symbol plot (Figure 2-15) or bubble plot (Figure 2-16) can be used. Rather than plotting the actual data values, symbols or bubbles that represent the data values are displayed. A symbol plot breaks the data range into bins and plots a bin label corresponding to the data value. A bubble plot consists of circles, with radii equal to the square roots of the data values. So unlike posting and symbol plots, a bubble plot gives a visual impression of the size of the data values. Directions for generating a symbol or bubble plot are contained in Box 2-24.

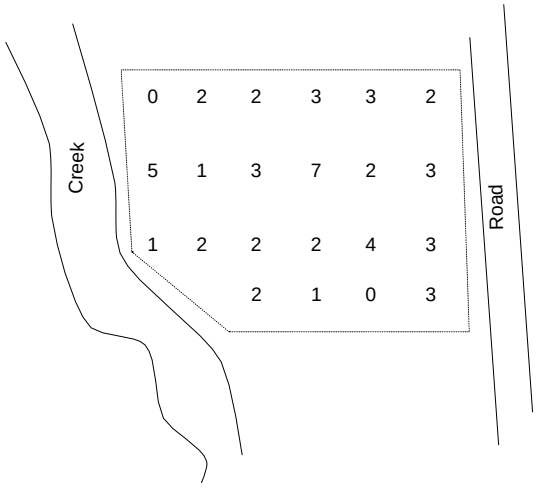


Figure 2-15. Example of a Symbol Plot

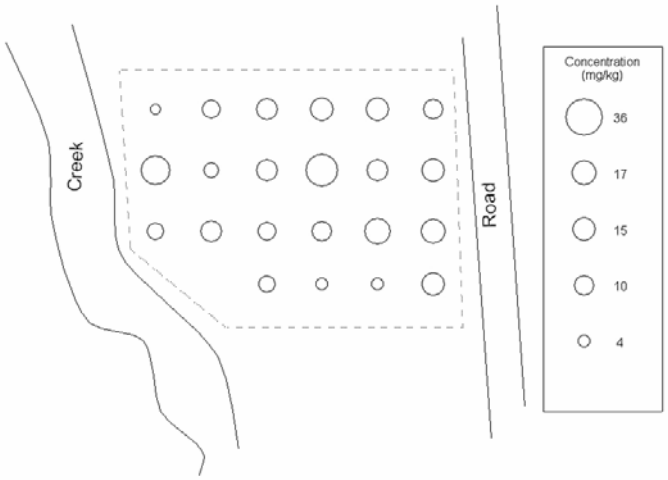


Figure 2-16. Example of a Bubble Plot

2.3.8.3 Other Spatial Graphical Representations

The three plots described in Sections 2.3.8.1 and 2.3.8.2 give information on the location of extreme values and spatial trends. The graphs below provide another item of interest to the data analyst, continuity of the spatial data. The graphical representations are not described in detail because they are used more for preliminary geostatistical analysis. These graphical representations can be difficult to develop and interpret. For more information on these representations, consult a statistician.

An h-scatterplot is a plot of all possible pairs of data whose locations are separated by a fixed distance in a fixed direction (indexed by h). For example, an h-scatterplot could be based on all the pairs whose locations are 1 meter apart in a southerly direction. An h-scatterplot is similar in appearance to a scatterplot (Section 2.3.6.1). The shape of the spread of the data in an h-scatterplot indicates the degree of continuity among data values a certain distance apart in particular direction. If all the plotted values fall close to a fixed line, then the data values at locations separated by a fixed distance in a fixed location are very similar. As data values become less and less similar, the spread of the data around the fixed line increases outward. The data analyst may construct several h-scatterplots with different distances to evaluate the change in continuity in a fixed direction.

A spatial correlogram is a plot of the correlations of the h-scatterplots. As the h-scatterplot only displays the correlation between the pairs of data whose locations are separated by a fixed distance in a fixed direction, it is useful to have a graphical representation of how these correlations change for different separation distances in a fixed direction. The spatial correlogram is such a plot which allows the analyst to evaluate the change in continuity in a fixed direction as a function of the distance between two points. A spatial correlogram is similar in appearance to a temporal correlogram (Section 2.3.8.2).

Contour plots are used to reveal overall spatial trends in the data by interpolating data values between sample locations. Most contour procedures depend on the density of the grid covering the sampling area (higher density grids usually provide more information than lower densities). A contour plot gives one of the best overall pictures of the important spatial features. However, contouring often requires that the actual fluctuations in the data values are smoothed so that many spatial features of the data may not be visible. The contour map should be used with other graphical representations of the data and requires expert judgment to interpret.

2.4 PROBABILITY DISTRIBUTIONS

2.4.1 The Normal Distribution

Data, especially measurements, often occur in natural patterns that can be considered to come from a distribution of values. In most instances, the data values will be grouped around some measure of central tendency such as the mean or median. The spread of the data is called the variance (the square root of this is called the standard deviation). A distribution with a large variance will be more spread out than one with a small variance (Figure 2-17). If a histogram of the data has a bell-shape (symmetric pattern about the mean with rapidly tapering tails), then the underlying distribution is often a normal distribution.

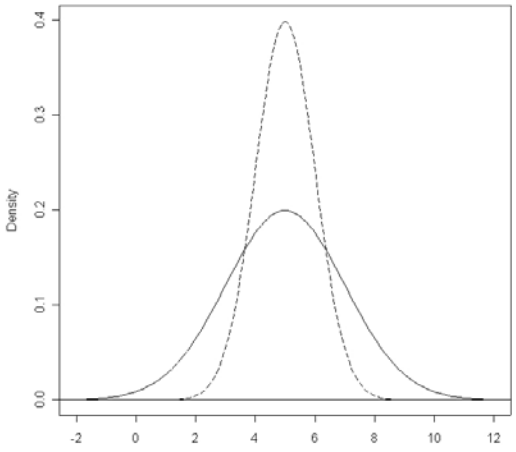


Figure 2-17. Two Normal Curves, Common Mean, Different Variances

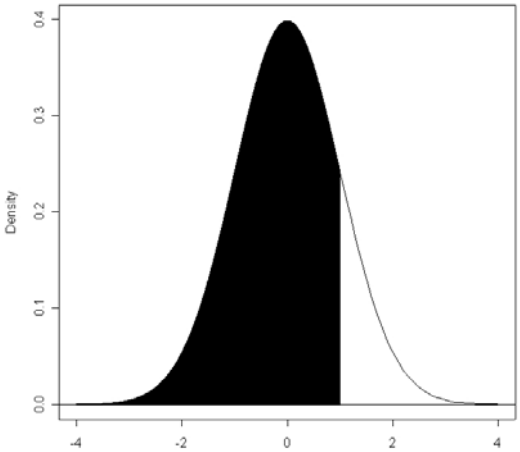


Figure 2-18. The Standard Normal Curve, Centered on Zero

If it is known or assumed that the underlying distribution is normal, then this is usually written as 'distributed N(μ,σ^2)' where μ is the mean and σ^2 is the variance. By subtracting μ and dividing by σ, any normal distribution can be transformed to a standard normal distribution, N(0,1). A standard normal random variable is most often denoted by *Z*. A plot of the standard normal is given in Figure 2-18. It is frequently necessary to refer to the percentiles of a standard normal and

in this guidance document, the subscript to a quoted *Z*-value will denote the percentile (or area under the curve, cumulative from the left), see Figure 2-18 (showing the area, 0.8413 corresponding to a Z-value of 1.00, and written as $Z_{0.8413}$ =1.00). Although common practice to use this subscript notation, some text tables and software programs use a different nomenclature; the user is advised to verify the meaning of any statistic encountered.

2.4.2 The *t*-Distribution

The standard normal curve is used when exact information on the mean and variance are available, but when only estimates from a sample are available, a different type of sampling distribution applies. When only information from a random sample on sample mean and sample variance is known for decision making purposes, a Student's-*t* distribution is appropriate. It resembles a standard normal but is lower in the center and fatter in the tails. The degree of fatness in the tails is a function of the degrees of freedom available, which in turn is related to sample size. As the sample size increases, the estimates of the mean and variance improve. As a result, the Student's-*t* distribution more closely resembles the standard normal distribution.

2.4.3 The Lognormal Distribution

Another commonly used distribution in environmental work is the lognormal distribution. The lognormal distribution is bounded on the left by 0, has a fatter right-tail than the normal distribution, and has a right-skewed shape. These characteristics are shown in Figure 2-19. The lognormal and normal distributions are related by a simple transformation: if X is distributed lognormally, then $Y = \ln(X)$ is distributed normally. However, log-transforming lognormal data to perform statistical procedures requiring normal data is a practice that should be done with care

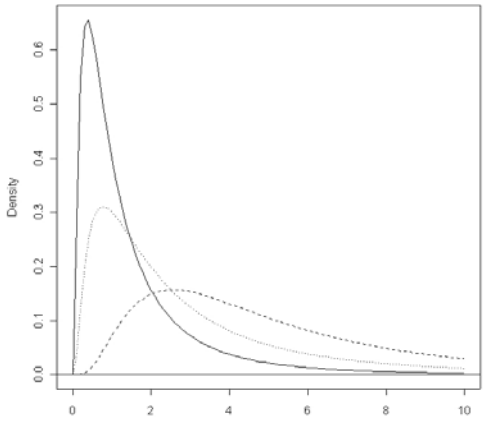


Figure 2-19. Three Different Lognormal Distributions

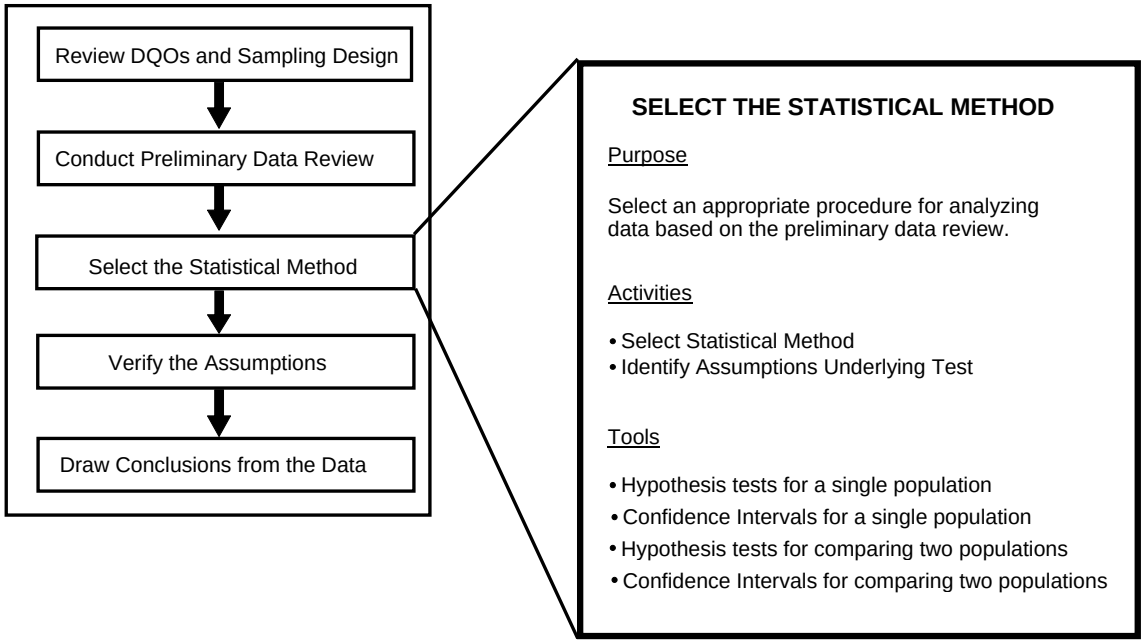
2.4.4 Central Limit Theorem

In many testing and estimation situations in environmental work, the focus of the investigation centers on the mean of a population. It is rare that true normality of the observations can be assumed. In most cases, the normally-based statistical tests are not overly affected by the lack of normality since tests are very robust and perform tolerably well unless gross non-normality is present. In addition, many tests become increasingly tolerant of deviations from normality as the number of observations increase. In simple terms, the underlying distribution of the sample mean more closely resembles a normal distribution as the number of observations increases. This occurs no matter what the underlying distribution of an individual observation. This phenomenon is called the Central Limit Theorem and is the basis of many statistical procedures.

CHAPTER 3

STEP 3: SELECT THE STATISTICAL METHOD

THE DATA QUALITY ASSESSMENT PROCESS



Step 3: Select the Statistical Method

- Select the statistical method based on the data user's objectives and the results of the preliminary data review.
  - If the problem involves comparing study results to a fixed threshold, such as a regulatory standard, consider the methods in Section 3.2.
  - If the problem involves comparing two populations, such as comparing data from two different locations or processes, then consider the hypothesis tests in Section 3.3.
- Identify the assumptions underlying the statistical method.
  - List the key underlying assumptions of the statistical procedure, such as distributional form, dispersion, independence, etc.
  - Note any sensitive assumptions where relatively small deviations could jeopardize the validity of the results.

List of Boxes

CHAPTER 3

STEP 3: SELECT THE STATISTICAL METHOD

3.1 OVERVIEW AND ACTIVITIES

This chapter provides an overview of issues associated with selecting an appropriate statistical method that will be used to draw conclusions from the data. There are two important outputs from this step: (1) the chosen method, and (2) the assumptions underlying the method. If a particular statistical procedure has been specified either in the DQO Process, the QA Project Plan, or the particular program or study, the analyst should use the results of the preliminary data review to determine if it is appropriate for the data collected. If a particular procedure has not been specified, then the analyst should select one based upon the data user's objectives and the preliminary data review.

One division in the methods of this section is between parametric and nonparametric hypothesis tests. Parametric tests typically concern the population mean or quantile, use the actual data values, and assume data values follow a specific probability distribution. Nonparametric tests typically concern the population mean or median, use data ranks, and don't assume a specific probability distribution. Parametric tests will have more power than a nonparametric counterpart if the assumptions are met. However, the distributional assumptions are often strict or undesirable for the parametric tests and deviations can lead to misleading results.

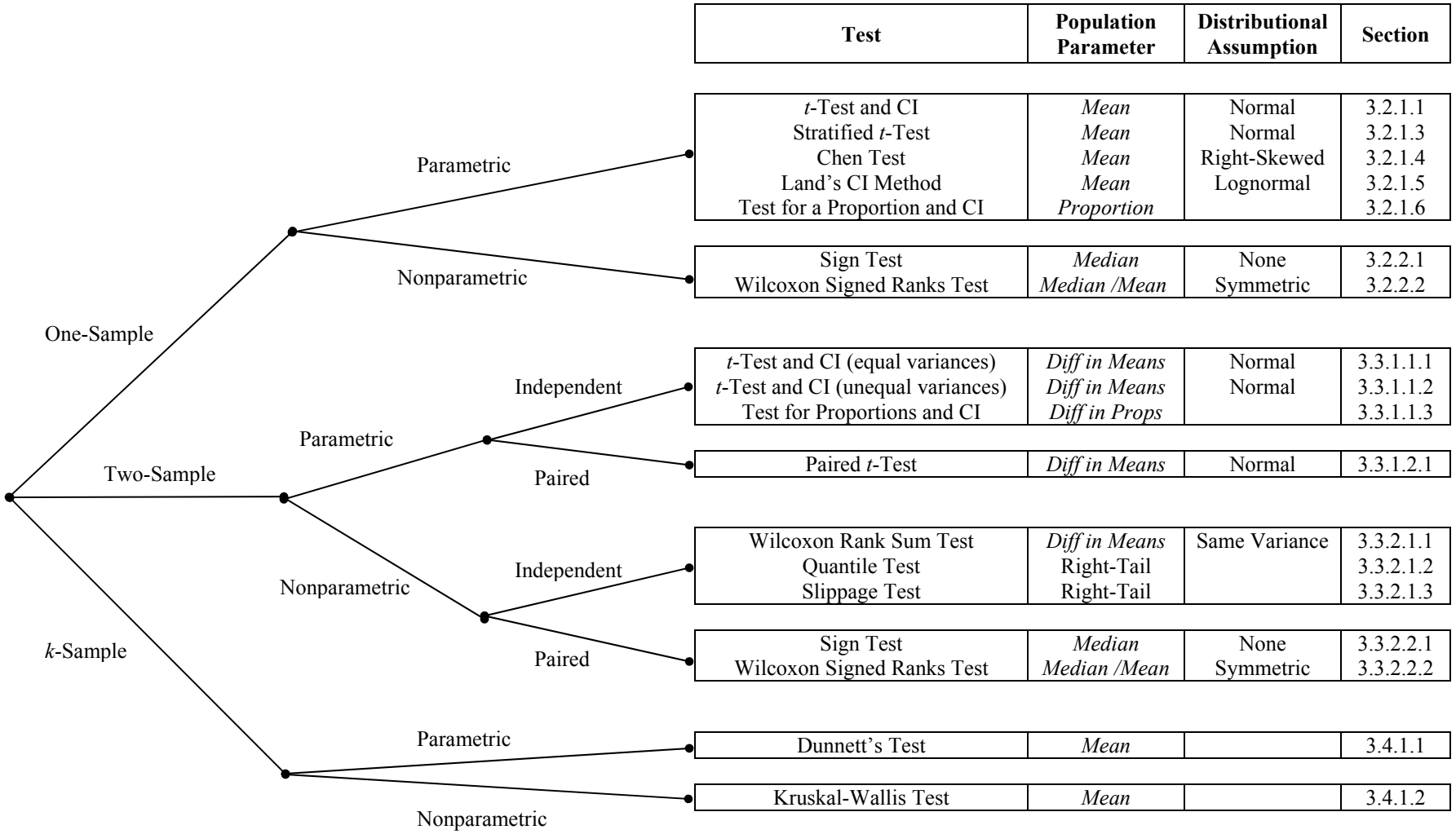
3.2 METHODS FOR A SINGLE POPULATION

The methods of this section concern comparing a single population parameter to a regulatory value (i.e. a fixed number) or the estimation of the population parameter. If the regulatory or action-value was estimated, then a one-sample method is not appropriate and a two-sample test should be selected.

An example of a one-sample test would be to determine if 95% of all companies emitting sulfur dioxide into the air are below a fixed discharge level. For this example, the population parameter is a proportion and the threshold value is 95% (0.95). Comparing the mean contaminant concentration of a hazardous site to the mean concentration of a background area would be a considered a two-sample test.

The hypothesis tests discussed in this section may be used to determine if there is evidence that $\theta < \theta_0$, $\theta > \theta_0$, or $\theta \neq \theta_0$ where θ represents the population mean, median, proportion, or quantile, and θ_0 represents the threshold value. There are also confidence/tolerance interval procedures to estimate θ. Section 3.2.1 discusses parametric hypothesis tests and confidence/tolerance intervals for a population mean or a population proportion. Section 3.2.2 discusses nonparametric hypothesis tests for the population median or population mean.

Decision Tree for Selecting the Specific Method



Test	Population Parameter	Distributional Assumption	Section
t-Test and CI	*Mean*	Normal	3.2.1.1
Stratified *t*-Test	*Mean*	Normal	3.2.1.3
Chen Test	*Mean*	Right-Skewed	3.2.1.4
Land's CI Method	*Mean*	Lognormal	3.2.1.5
Test for a Proportion and CI	*Proportion*		3.2.1.6
Sign Test	*Median*	None	3.2.2.1
Wilcoxon Signed Ranks Test	*Median /Mean*	Symmetric	3.2.2.2
t-Test and CI (equal variances)	*Diff in Means*	Normal	3.3.1.1.1
t-Test and CI (unequal variances)	*Diff in Means*	Normal	3.3.1.1.2
Test for Proportions and CI	*Diff in Props*		3.3.1.1.3
Paired *t*-Test	*Diff in Means*	Normal	3.3.1.2.1
Wilcoxon Rank Sum Test	*Diff in Means*	Same Variance	3.3.2.1.1
Quantile Test	Right-Tail		3.3.2.1.2
Slippage Test	Right-Tail		3.3.2.1.3
Sign Test	*Median*	None	3.3.2.2.1
Wilcoxon Signed Ranks Test	*Median /Mean*	Symmetric	3.3.2.2.2
Dunnett's Test	*Mean*		3.4.1.1
Kruskal-Wallis Test	*Mean*		3.4.1.2

3.2.1 Parametric Methods

These methods rely on the knowing the specific distribution of the population or of the statistic of interest.

3.2.1.1 The One-Sample *t*-test and Confidence Interval or Limit

Purpose: Test for a difference between a population mean and a fixed threshold or to estimate a population mean.

Data: A simple or systematic random sample, $x_1,\ldots,x_n$, from the population of interest.

Assumptions: The data are independent and come from an approximately normal distribution or the sample size is large ($n \geq 30$).

Limitations and Robustness: One-sample *t* methods are robust against the population distribution deviating moderately from normality. However, they are not robust against outliers and have difficulty dealing with non-detects. The nonparametric methods of section 3.2.2 are an alternative. The substitution or adjustment procedures of chapter 4 can be used for non-detects.

Directions for the one-sample t-test are contained in Box 3-1, with an example in Box 3–3. Directions for the one-sample confidence interval are contained in Box 3-2, with an example in Box 3-4.

3.2.1.2 The One-Sample Tolerance Interval or Limit

Purpose: A tolerance interval specifies a region that contains a certain proportion of the population with a certain confidence. For example, "the 99% tolerance interval for 90% of the population is 7.5 to 9.9", is interpreted as, "I can be 99% certain that the interval 7.5 to 9.9 captures 90% of the population."

Data: A simple or systematic random sample, $x_1,\ldots,x_n$, from the population of interest. The sample may or may not contain compositing.

Assumptions: The data are independent and approximately normally distributed.

Limitations and Robustness: Tolerance intervals are robust against the population distribution deviating moderately from normality. However, they are not robust against outliers.

Directions for the one-sample tolerance interval are contained in Box 3-5, with an example in Box 3-6.

Box 3-1: Directions for the One-Sample *t*-Test

COMPUTATIONS: Compute the mean, $\overline{X}$, and standard deviation, *s*, of the data set.

STEP 1. Null Hypothesis: $H_0: \mu = C$

STEP 2. Alternative Hypothesis:
i) $H_A: \mu > C$ (upper-tail test)
ii) $H_A: \mu < C$ (lower-tail test)
iii) $H_A: \mu \neq C$ (two-tail test)

STEP 3. Test Statistic: $t_0 = \dfrac{\overline{X} - C}{s/\sqrt{n}}$

STEP 4. a) Critical Value: Use Table A-2 to find:
i) $t_{n-1,1-\alpha}$
ii) $-t_{n-1,1-\alpha}$
iii) $t_{n-1,1-\alpha/2}$

STEP 4. b) p-value: Use Table A-2 to find:
i) $P(t_{n-1} > t_0)$
ii) $P(t_{n-1} < t_0)$
iii) $2 \cdot P(t_{n-1} > |t_0|)$, where $|t_0|$ is the absolute value of t_0.

STEP 5. a) Conclusion:
i) If $t_0 > t_{n-1,1-\alpha}$, then reject the null hypothesis that the true population mean is equal to the threshold C.
ii) If $t_0 < -t_{n-1,1-\alpha}$, then reject the null hypothesis.
iii) If $|t_0| > t_{n-1,1-\alpha/2}$, then reject the null hypothesis.

STEP 5. b) Conclusion: If p-value < α, then reject the null hypothesis that the true population mean is equal to the threshold C.

STEP 6. If the null hypothesis was not rejected, there is only one false acceptance error rate (β at μ_1), and if $n \geq \dfrac{s^2\left(z_{1-\alpha'} + z_{1-\beta}\right)^2}{(\mu_1 - C)^2} + \dfrac{z^2_{1-\alpha'}}{2}$, then the sample size was probably large enough to achieve the DQOs.
The value of α' is α for a one-sided test and $\alpha/2$ for a two-sided test.

Box 3-2: Directions for Computing a One-Sample *t* Confidence Interval or Limit

COMPUTATIONS: Compute the sample mean, $\overline{X}$, and sample standard deviation, *s*, of the data set.

A 100(1 - α)% confidence interval for μ is $\overline{X} \pm t_{n-1,1-\alpha/2} \cdot s/\sqrt{n}$, where Table A-2 is used to find $t_{n-1,1-\alpha/2}$.

A 100(1 - α)% upper confidence limit (UCL) for μ is $\overline{X} + t_{n-1,1-\alpha} \cdot s/\sqrt{n}$.

A 100(1 - α)% lower confidence limit (LCL) for μ is $\overline{X} - t_{n-1,1-\alpha} \cdot s/\sqrt{n}$.

Box 3-3: A One-Sample *t*-Test Example

Consider the following 9 random data values (in ppm):

82.39, 103.46, 104.93, 105.52, 98.37, 113.23, 86.62, 91.72, 108.21.

This data will be used to test: H_0: $\mu \leq 95$ ppm vs. H_A: $\mu > 95$ ppm. The decision maker has specified a 5% false rejection error rate (α) at 95 ppm (C), and a 20% false acceptance error rate (β) at 105 ppm (μ_1).

COMPUTATIONS: The mean is $\overline{X} = 99.38$ ppm and the standard deviation is $s = 10.41$ ppm.

STEP 1. Null Hypothesis: H_0: $\mu \leq 95$

STEP 2. Alternative Hypothesis: H_A: $\mu > 95$ (upper-tail test)

STEP 3. Test Statistic: $t_0 = \dfrac{\overline{X} - C}{s/\sqrt{n}} = \dfrac{99.38 - 95}{10.41/\sqrt{9}} = 1.26$

STEP 4. a) Critical Value: Using Table A-2, $t_{n-1,1-\alpha} = t_{8,0.95} = 1.86$

STEP 4. b) p-value Using Table A-2, 0.10 < p-value < 0.15. Using statistical software, p - value $= P(t_{n-1} > t_0) = P(t_8 > 1.26) = 0.1216$.

STEP 5. a) Conclusion: Since 1.26 < 1.86, we fail to reject the null hypothesis that the true population mean is at most 95 ppm.

STEP 5. b) Conclusion: Since p-value = 0.1216 > 0.05 = significance level, we fail to reject the null hypothesis that the true population mean is at most 95 ppm.

STEP 6. Since the null hypothesis was not rejected and there is only one false acceptance error rate, it is possible to use the sample size formula to determine if the error rate has been satisfied. Since,

$$n \geq \frac{s^2\left(z_{1-\alpha'} + z_{1-\beta}\right)^2}{\left(\mu_1 - C\right)^2} + \frac{z_{1-\alpha'}^2}{2} = \frac{10.41^2(1645 + 0.842)^2}{(95 - 105)^2} + \frac{1645^2}{2} = 8.049,$$

the false acceptance error rate has probably been satisfied.

Box 3-4: An Example of a One-Sample *t* Upper Confidence Limit for a Population Mean

The effluent from a discharge point in a plating manufacturing plant was sampled 7 times over the course of 4 days for the presence of Arsenic with the following results (in ppm): 8.1, 7.9, 7.9, 8.2, 8.2, 8.0, 7.9. A 95% upper confidence limit for the population mean will be computed.

COMPUTATIONS: The sample mean is $\overline{X}$ = 8.03 ppm and the sample standard deviation is s = 0.138 ppm.

A 95% upper confidence limit for μ is: $\overline{X} + t_{n-1,\,1-\alpha} \cdot s/\sqrt{n}$ *i.e.* $8.03 \pm 1.943 \cdot 0.138/\sqrt{7}$ or 8.131 ppm.

Box 3-5: Directions for Computing a One-Sample Tolerance Interval or Limit

COMPUTATIONS: Compute the sample mean, $\overline{X}$, and sample standard deviation, s, of the data set.

A 100(1 – α)% tolerance interval for (1 – p)% of the population is $\overline{X} \pm k_2 \cdot s$, where $k_2 = z_{1-p/2} \cdot \sqrt{\frac{n^2 - 1}{n \cdot \chi^2_{n-1,\alpha}}}$.

Table A-1 is used to find $z_{1-p/2}$ and Table A-9 is used to find $\chi^2_{n-1,\alpha}$.

A 100(1 – α)% upper tolerance limit for (1 – p)% of the population is $\overline{X} + k_1 \cdot s$ and
a 100(1 – α)% lower tolerance limit for (1 – p)% of the population is $\overline{X} - k_1 \cdot s$, where

$$k_1 = \frac{z_{1-p} + \sqrt{z_{1-p}^2 - ab}}{a},\ a = 1 - \frac{z_{1-\alpha}^2}{2 \cdot (n-1)},\ \text{and}\ b = z_{1-p}^2 - \frac{z_{1-\alpha}^2}{n}.$$

Box 3-6: An Example of a One-Sample Upper Tolerance Limit

The effluent from a discharge point in a plating manufacturing plant was sampled 7 times over the course of 4 days for the presence of Arsenic with the following results (in ppm): 8.1, 7.9, 7.9, 8.2, 8.2, 8.0, 7.9. A 95% upper tolerance limit for 90% of the population will be computed.

COMPUTATIONS: The sample mean is $\overline{X}$ = 8.03 ppm and the sample standard deviation is s = 0.138 ppm.

Also, $a = 1 - \frac{1.65^2}{2 \cdot 6} = 0.7731$, $b = 1.28^2 - \frac{1.65^2}{7} = 1.2495$, and $k_1 = \frac{1.28 + \sqrt{1.28^2 - 0.7731 \cdot 1.2495}}{0.7731} = 2.716$.

A 95% upper tolerance limit for 90% of the population is: $\overline{X} + k_1 \cdot s$ *i.e.* $8.03 + 2.716 \cdot 0.138$ or 8.405 ppm.
So we are 95% confident that at least 90% of the population is less than 8.405 ppm.

3.2.1.3 Stratified Random Sampling

This section provides a brief introduction of stratified random sampling and directions for a one-sample *t*-test with stratified data. For a more in-depth discussion of stratified random sampling, see *Guidance on Choosing a Sampling Design for Environmental Data Collection*, EPA QA/G-5S (U.S. EPA 2002).

A stratified random sample occurs when a population is split into subpopulations (called strata) and simple random samples are taken from the subpopulations. Reasons for implementing stratified random sampling include administrative convenience and a gain in the precision in the estimates if a heterogeneous population can be split into homogeneous subpopulations.

Suppose the cost of taking a sample is $C = c_0 + \Sigma c_h n_h$, where c_0 is the overhead cost, c_h is the cost of sampling in stratum h, and n_h is the sample size taken from stratum h. Then the variance of the sample mean is minimized for given cost C and the cost is minimized for a specified variance if n_h is proportional to $W_h\sigma_h / \sqrt{c_h}$, where W_h is the stratum weight (the size of the subpopulation in relation to the entire population) and σ_h is the population standard deviation for stratum h. Therefore, a larger sample is taken in a given stratum when the stratum is large, the stratum variance is large, or sampling is cheaper in the stratum. If the cost of sampling is equal across all strata, then the variance of the sample mean is minimized for a fixed total sample size n when

$$n_h = n \cdot \frac{W_h\sigma_h}{\sum W_h\sigma_h}.$$

This type of sample size selection is called Neyman allocation. In practice, the unknown σ_h is replaced by the observed sample standard deviation.

Purpose: Test for a difference between a population mean and a fixed threshold or to estimate the population mean.

Data: A stratified random sample, $x_{h1},\ldots,x_{hn_h}$, for $h = 1,\ldots,L$, where the n_h are not necessarily equal. Also, let the total sample size be $n = n_1 + \cdots + n_L$.

Assumptions: The data in each stratum comes from an approximately normal distribution.

Limitations and Robustness: Defining the strata requires previous knowledge of the population. Further, determining samples for each stratum requires knowledge or assumptions about the variance in each stratum. Finally, if strata are chosen poorly, then observations may appear to be outliers when they are actually observations from a different stratum.

Directions for a one-sample t-test for a stratified sample are contained in Box 3-7, with an example in Box 3-8.

3.2.1.4 The Chen Test

Purpose: Test for a difference between a population mean and a fixed threshold when the underlying distribution is right-skewed, i.e., the right-tail of the distribution is longer than the left-tail. The standard one-sample *t*-test works best when the underlying distribution is normal or at least symmetric. Often times though, environmental data has an underlying distribution that is right-skewed. Therefore, the Chen test may be more appropriate when most of the data values are relatively small, but there are also a few relatively large values.

Data: A simple or systematic random sample, $x_1,\ldots,x_n$, from the population of interest.

Assumptions: A simple or systematic random sample, $x_1,\ldots,x_n$, from the population of interest which is believed to be right-skewed. This can be verified by inspection of a histogram of the data or a sample skewness that is greater than one.

Box 3-7: Directions for a One-Sample *t*-Test for a Stratified Random Sample

COMPUTATIONS: Calculate the stratum weights, $W_h = V_h \Big/ \sum_{h=1}^{L} V_h$, where V_h is the surface area of stratum h multiplied by the depth of sampling in stratum h. For each stratum, calculate the sample stratum mean and the sample stratum standard deviation,

$$\overline{X}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} X_{hi} \text{ and } s_h^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}\left(X_{hi} - \overline{X}_h\right)^2.$$

Compute the overall sample mean and the sample variance of the mean, $\overline{X}_{st} = \sum_{h=1}^{L} W_h \overline{X}_h$ and $s_{st}^2 = \sum_{h=1}^{L} \frac{W_h s_n^2}{n_h}$.

Finally, calculate the approximate degrees of freedom, $df = \dfrac{s_{st}^4}{\sum_{h=1}^{L} \dfrac{W_h^4 s_h^4}{n_h^2(n_h - 1)}}$ (round up to the next integer).

STEP 1. Null Hypothesis: $H_0 : \mu = C$

STEP 2. Alternative Hypothesis:
i) $H_A : \mu > C$ (upper-tail test)
ii) $H_A : \mu < C$ (lower-tail test)
iii) $H_A : \mu \neq C$ (two-tail test)

STEP 3. Test Statistic: $t_0 = \dfrac{\overline{X}_{st} - C}{s_{st}}$

STEP 4. a) Critical Value: Use Table A-2 t find:
i) $t_{df,1-\alpha}$
ii) $-t_{df,1-\alpha}$
iii) $t_{df,1-\alpha/2}$

STEP 4. b) p-value: Use Table A-2 to find:
i) $P(t_{df} > t_0)$
ii) $P(t_{df} < t_0)$
iii) $2 \cdot P(t_{df} > |t_0|)$

STEP 5. a) Conclusion:
i) If $t_0 > t_{df,1-\alpha}$, then reject the null hypothesis that the true population mean is equal to the threshold C.
ii) If $t_0 < -t_{df,1-\alpha}$, then reject the null hypothesis.
iii) If $|t_0| > t_{df,1-\alpha/2}$, then reject the null hypothesis.

STEP 5. b) Conclusion: If p-value < α, then reject the null hypothesis that the true population mean is equal to the threshold C.

STEP 6. If the null hypothesis was not rejected, there is only one false acceptance error rate (β at μ_1), and if $n \geq \dfrac{s_{st}^2\left(z_{1-\alpha'} + z_{1-\beta}\right)^2}{(\mu_1 - C)^2} + \dfrac{z_{1-\alpha'}^2}{2}$, then the sample size was probably large enough to achieve the DQOs.
The value of α' is α for a one-sided test and $\alpha/2$ for a two-sided test.

Box 3-8: An Example of a One-Sample *t*-Test for a Stratified Random Sample

Consider a stratified random sample consisting of two strata where stratum 1 comprises 25% of the total site surface area and stratum 2 comprises the other 75%. Suppose 40 samples were collected from stratum 1, and 60 samples were collected from stratum 2. This information will be used to test the null hypothesis that the overall site mean is 40 ppm versus the lower-tail alternative. The decision maker has specified a 1% false rejection decision limit (α) at 40 ppm and a 20% false acceptance decision error limit (β) at 35 ppm (μ_1).

COMPUTATIONS: The stratum weights are $W_1 = 0.25$, $W_2 = 0.75$. For stratum 1, the sample mean is 31 ppm and the sample standard deviation is 18.2 ppm. For stratum 2, the sample mean is 35 ppm, and the sample standard deviation is 20.5 ppm.

The sample overall mean concentration is $\overline{X}_{st} = \sum_{h=1}^{L} W_h \overline{X}_h = 0.25 \cdot 31 + 0.75 \cdot 35 = 34$ and the sample variance of the mean is

$$s_{st}^2 = \sum_{h=1}^{L} \frac{W_h^2 s_h^2}{n_h} = \frac{0.25^2 \cdot 18.2^2}{40} + \frac{0.75^2 \cdot 20.5^2}{60} = 4.46 .$$

The approximate degrees of freedom is:

$$df = \frac{s_{st}^4}{\sum_{h=1}^{L} \frac{W_h^4 s_h^4}{n_h^2 (n_h - 1)}} = \frac{4.46^2}{\frac{0.25^4 \cdot 18.2^4}{40^2 \cdot 39} + \frac{0.75^4 \cdot 20.5^4}{60^2 \cdot 59}} = 73.7 \rightarrow 74 .$$

STEP 1. Null Hypothesis: $H_0 : \mu = 40$

STEP 2. Alternative Hypothesis: $H_A : \mu < 40$ (lower-tail test)

STEP 3. Test Statistic: $t_0 = \frac{\overline{X}_{st} - C}{s_{st}} = \frac{34 - 40}{\sqrt{4.46}} = -2.841$

STEP 4. a) Critical Value: Using Table A-2, $-t_{df,1-\alpha} = -t_{74,\,0.99} = -2.378$.

STEP 4. b) p-value: Using Table A-2, p-value < 0.005. (the exact value is $P(t_{df} < t_0) = P(t_{74} < -2.841) = 0.0029$).

STEP 5. a) Conclusion: Since test statistic = -2.841 < -2.378 = critical value, we reject the null hypothesis that the true population mean is equal to 40.

STEP 5. b) Conclusion: Since p-value = 0.0029 < 0.01 = significance level, we reject the null hypothesis that the true population mean is equal to 40.

Limitations and Robustness: Chen's test is a generalization of the one-sample *t*-test. Like the *t*-test, this Chen's test can have some difficulties in dealing with non-detects, especially if there are a large number of them. For a moderate amount of non-detects, a substitution method (e.g.,½ of the detection limit) will suffice if the threshold level C is much larger than the detection limit, otherwise refer to the methods discussed in Chapter 4.

Directions for the Chen test are contained in Box 3-9, with an example in Box 3-10.

Box 3-9: Directions for the Chen Test

COMPUTATIONS: Visually check the assumption of right-skewness by inspecting a histogram. If at most 15% of the data points are below the detection limit (DL) and C is much larger than the DL, then replace values below the DL with DL/2. Compute the mean, $\overline{X}$, and the standard deviation, *s* of the data set. Then compute

$$b = \frac{n}{(n-1)(n-2)s^3}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^3 \text{ (sample skewness), } a = \frac{b}{6\sqrt{n}}\text{, and } t_0 = \frac{\overline{X} - C}{s/\sqrt{n}}.$$

NOTE: The sample skewness should be greater than 1 to indicate the underlying distribution is right-skewed.

STEP 1. Null Hypothesis: $H_0: \mu = C$

STEP 2. Alternative Hypothesis:
i) $H_A: \mu > C$ (upper-tail test)
ii) $H_A: \mu < C$ (lower-tail test)
iii) $H_A: \mu \neq C$ (two-tail test)

STEP 3. Test Statistic: $T = t_0 + a\left(1 + 2t_0^2\right) + 4a^2\left(t_0 + 2t_0^3\right)$

STEP 4. a) Critical Value: Use Table A-1to find:
i) $z_{1-\alpha}$
ii) z_{α}
iii) $z_{1-\alpha/2}$

STEP 4. b) p-value: Use Table A-1 to find:
i) $P(Z > T)$
ii) $P(Z < T)$
iii) $2 \cdot P(Z > |T|)$

STEP 5. a) Conclusion:
i) If $T > z_{1-\alpha}$, then reject the null hypothesis that the true population mean is equal to the threshold value C.
ii) If $T < z_{\alpha}$, then reject the null hypothesis.
iii) If $|T| > z_{1-\alpha/2}$, then reject the null hypothesis.

STEP 5. b) Conclusion: If p-value < α, then reject the null hypothesis that the true population mean is equal to the threshold value C.

Box 3-10: Example of the Chen Test

Consider the following sample of a contaminant concentration measurements (in ppm):

2.0, 2.0, 5.0, 5.2, 5.9, 6.6, 7.4, 7.4, 9.7, 9.7, 10.2, 11.5,
12.4, 12.7, 14.1, 15.2, 17.7, 18.9, 22.8, 28.6, 30.5, 35.5.

We want to test the null hypothesis that the mean μ is less than 10 ppm versus the alternative that it exceeds 10 ppm. A significance level of 0.05 is to be used.

COMPUTATIONS: A histogram of the 22 data points indicates a right-skewed distribution. It is found that $\overline{X} = 13.227$ ppm and s = 9.173 ppm. Also,

$$b = \frac{22}{21 \cdot 20 \cdot 9.173^3} \sum_{i=1}^{22} (X_i - 13.227)^3 = 1.078\,, \quad a = \frac{1.078}{6\sqrt{22}} = 0.0383\,, \text{ and } t_0 = \frac{\overline{X} - C}{s/\sqrt{n}} = \frac{13.227 - 10}{9.173/\sqrt{22}} = 1.650\,.$$

A skewness value of 1.078 indicates the data are right-skewed.

STEP 1. Null Hypothesis: H_0: $\mu \le 10$

STEP 2. Alternative Hypothesis: H_A: $\mu > 10$ (upper-tail test)

STEP 3. Test Statistic: $T = t_0 + a \cdot \left(1 + 2t_0^2\right) + 4a^2 \cdot \left(t_0 + 2t_0^3\right) = 1.96\,.$

STEP 4. a) Critical Value: Using Table A-1, $z_{1-\alpha} = z_{0.95} = 1.645$.

STEP 4. b) p-value: Using Table A-1, $\mathrm{P}(Z > 1.96) = 1 - 0.9750 = 0.0250\,.$

STEP 5. a) Conclusion: Since the test statistic = 1.96 > 1.645 = the critical value, we reject the null hypothesis.

STEP 5. b) Conclusion: Since the p-value = 0.0250 < 0.05 = the significance level, we reject the null hypothesis.

3.2.1.5 Land's Method for Lognormally Distributed Data

Purpose: Estimate the mean of a lognormally distributed population using confidence limits. Another alternative is to log-transform the data to make it approximately normal, then use the *t* confidence interval described in Box 3-2, and finally transforming the confidence bounds back to the original scale. This method produces a biased interval estimate and should be avoided, unless a median is being estimated in which case this method produces an unbiased estimate.

Data: A simple or systematic random sample, $x_1,\ldots,x_n$, from the population of interest. The sample may or may not contain compositing.

Assumptions: The underlying distribution of the data is approximately lognormal.

Limitations and Robustness: Land's method is extremely sensitive to outliers since the mean and standard deviation are not robust against outliers unless a preventative limit on variance is used.

Directions for Land's Method are contained in Box 3-11, with an example in Box 3-12.

Box 3-11: Directions for Computing Confidence Limits for the Population Mean of a Lognormal Distribution Using Land's Method

COMPUTATIONS: Transform the data: $y_i = \ln x_i$, $i = 1,\ldots,n$. Next, compute the sample mean, $\bar{y}$, and sample variance, s_y^2, of the transformed data. The values for $H_{1-\alpha}$ and H_α come from Table A-17.

An upper one-sided 100(1 - α)% confidence limit for the population mean is $\exp\left(\bar{y} + \frac{s_y^2}{2} + \frac{s_y H_{1-\alpha}}{\sqrt{n-1}}\right)$.

A lower one-sided 100α% confidence limit for population mean is $\exp\left(\bar{y} + \frac{s_y^2}{2} + \frac{s_y H_\alpha}{\sqrt{n-1}}\right)$.

Box 3-12: An Example Using Land's Method

A random sample of 15 concentrations from a monitoring process (assumed to be lognormal) is reported:

8.12, 7.32, 4.82, 6.52, 7.80, 11.89, 12.94, 7.51, 18.14, 4.09, 5.70, 15.57, 6.68, 8.15, 5.56.

Compute an upper one-sided 95% confidence limit for the population mean of the process.

COMPUTATIONS: The log-transformed data set is:

2.09, 1.99, 1.57, 1.87, 2.05, 2.48, 2.56, 2.02, 2.90, 1.41, 1.74, 2.75, 1.90, 2.10, 1.72

The sample mean and sample standard deviation of the transformed data are $\bar{y} = 2.0767$ and $s_y = 0.4272$. The value of $H_{0.95}$ is found by interpolation in Table A-17.

An upper one-sided 95% confidence limit for the population mean is

$$\exp\left(\bar{y} + \frac{s_y^2}{2} + \frac{s_y H_{0.95}}{\sqrt{n-1}}\right) = \exp\left(2.0767 + \frac{0.4272^2}{2} + \frac{0.4272 \cdot 1.9939}{\sqrt{14}}\right) = 10.97$$

3.2.1.6 The One-Sample Proportion Test and Confidence Interval

Purpose: Test for a difference between a population proportion, P, and a fixed threshold (P_0) or to estimate the population proportion. If $P_0 = 0.5$, this test is equivalent to the Sign test.

Data: A simple or systematic random sample, $x_1,\ldots,x_n$, from the population of interest.

Assumptions: The data constitutes an independent random sample from the population.

Limitations and Robustness: Both nP_0 and $n(1\text{-}P_0)$ must be at least 5 to apply the normal approximation. Otherwise, exact tests must be used and a statistician should be consulted.

Directions for the one-sample test for proportions are contained in Box 3-13, with an example in Box 3-15. Directions for computing a confidence interval for a proportion are contained in Box 3-14.

Box 3-13: Directions for the One-Sample Test for Proportions

COMPUTATIONS: Compute p, the sample proportion of data values that fit the desired characteristic.

STEP 1. Null Hypothesis: $H_0 : P = P_0$

STEP 2. Alternative Hypothesis:
i) $H_A : P > P_0$ (upper-tail test)
ii) $H_A : P < P_0$ (lower-tail test)
iii) $H_A : P \neq P_0$ (two-tail test)

STEP 3. Test Statistic: $z_0 = \dfrac{p + c - P_0}{\sqrt{P_0(1-P_0)/n}}$, where $c = \begin{cases} -0.5/n & \text{if } p > P_0 \\ 0.5/n & \text{if } p < P_0 \end{cases}$

STEP 4. a) Critical Value: Use Table A-1 to find:
i) $z_{1-\alpha}$
ii) z_{α}
iii) $z_{1-\alpha/2}$

STEP 4. b) p-value: Use Table A-1 to find:
i) $P(Z > z_0)$
ii) $P(Z < z_0)$
iii) $2 \cdot P(Z > |z_0|)$

STEP 5. a) Conclusion:
i) If $z_0 > z_{1-\alpha}$, then reject the null hypothesis that the true proportion is equal to the threshold value P_0.
ii) If $z_0 < z_{\alpha}$, then reject H_0.
iii) If $|z_0| > z_{1-\alpha/2}$, then reject H_0.

STEP 5. b) Conclusion: If p-value < α, then reject the null hypothesis that the true proportion is equal to the threshold value P_0.

STEP 6. If the null hypothesis was not rejected, there is only one false acceptance error rate (β at P_1), and if $n \geq \left[\dfrac{z_{1-\alpha'}\sqrt{P_0(1-P_0)} + z_{1-\beta}\sqrt{P_1(1-P_1)}}{P_1 - P_0} \right]^2$, then the sample size was large enough to achieve the DQOs. The value of α' is α for a one-sided test and $\alpha/2$ for a two-sided test.

Box 3-14: Directions for Computing a Confidence Interval for a Population Proportion

COMPUTATIONS: Compute p, the sample proportion of data values that fit the desired characteristic.

A 100(1 - α)% confidence interval for P is $p \pm z_{1-\alpha/2} \cdot \sqrt{\dfrac{p(1-p)}{n}}$, where Table A-1 is used to find $z_{1-\alpha/2}$.

Box 3-15: An Example of the One-Sample Test for Proportions

Consider 85 concentration samples of which 11 have are greater than the clean-up standard. Test the null hypothesis that the population proportion of concentrations greater than the clean-up standard is 0.2 versus the lower-tailed alternative. The decision maker has specified a 5% false rejection rate (α) for $P_0 = 0.2$, and a false acceptance rate (β) of 20% for $P_1 = 0.15$.

COMPUTATIONS: Since both $nP_0 = 85 \cdot 0.2 = 17$ and $n(1- P_0) = 85 \cdot (1-0.2) = 68$ are at least 5, the normal approximation can be applied. From the data, the sample proportion is $p = 11/85 = 0.1294$

STEP 1. Null Hypothesis: H_0: $P \geq 0.20$

STEP 2. Alternative Hypothesis: H_A: $P < 0.20$

STEP 3. Test Statistic: $z_0 = \dfrac{p + 0.5/n - P_0}{\sqrt{P_0(1-P_0)/n}} = \dfrac{0.1294 + 0.5/85 - 0.2}{\sqrt{0.2(1-0.2)/85}} = -1.49$

STEP 4. a) Critical Value: Using Table A-1, $z_\alpha = z_{0.05} = -1.645$.

STEP 4. b) p-value: Using Tables A-1, $P(Z < -1.49) = 1 - 0.9319 = 0.0681$.

STEP 5. a) Conclusion: Since the test statistic = -1.49 > -1.645, we fail to reject the null hypothesis that the true population proportion is at least 0.20.

STEP 5. b) Conclusion: Since p-value = 0.0681 > 0.05 = significance level, we fail to reject the null hypothesis that the true population proportion is at least 0.20.

STEP 6. Since the null hypothesis was not rejected and there is only one false acceptance error rate, it is possible to use the sample size formula to determine if the error rate has been satisfied. Since,

$$n < \left[\frac{z_{1-\alpha'}\sqrt{P_0(1-P_0)} + z_{1-\beta}\sqrt{P_1(1-P_1)}}{P_1 - P_0} \right]^2 = \left[\frac{1.64\sqrt{0.2(1-0.2)} + 0.84\sqrt{0.15(1-0.15)}}{0.15 - 0.2} \right]^2 = 365.53 ,$$

the sample size was not large enough to achieve the DQOs. So the null hypothesis was not rejected, but the false acceptance error rate was not satisfied. Therefore, there is insufficient evidence that the proportion is less than 0.20, but this conclusion is uncertain because the sample size was too small.

3.2.2 Nonparametric Methods

These methods rely on the relative rankings of data values. Knowledge of the precise form of the population distribution is not necessary.

3.2.2.1 The Sign Test

Purpose: Test for a difference between the population median, and a fixed threshold.

Data: A simple or systematic random sample, $x_1,\ldots,x_n$, from the population of interest. The sample may or may not contain compositing.

Assumptions: The Sign test can be used for any underlying population distribution.

Limitations and Robustness: The Sign test has less power than the one-sample *t*-test or the Wilcoxon Signed Rank test. However, the Sign test makes no distributional assumptions like the other two tests and it can handle non-detects if the detection limit is below the threshold.

Directions for conducting a sign test are contained in Box 3-16, with an example in Box 3-17.

3.2.2.2 The Wilcoxon Signed Rank Test

Purpose: Test for a difference between the true location (mean or median) of a population and a fixed threshold. If the underlying population distribution is approximately normal, then the one-sample *t*-test will have more power (chance of rejecting the null hypothesis when it is false) than the Wilcoxon Signed Rank test. For symmetric distributions, the Wilcoxon Signed Rank test will have more power than the Sign test. If the sample size is small and the data is not approximately symmetric nor normally distributed, then the Sign test should be used.

Data: A simple or systematic random sample, $x_1,\ldots,x_n$, from the population of interest.

Assumptions: The data set comes from an approximately symmetric distribution.

Limitations and Robustness For large sample sizes ($n > 50$), the one-sample *t*-test is more robust against violations of its assumptions than the Wilcoxon Signed Rank test. The Wilcoxon signed rank test may produce misleading results if there are many tied data values. If possible, measurements should be recorded with sufficient accuracy so that a large number of tied values do not occur. Estimated concentrations should be reported for data below the detection limit, even if these estimates are negative, as their relative magnitude is of importance. If this is not possible, substitute the value DL/2 for each value below the detection limit providing all the data have the same detection limit. When different detection limits are present, all data could be censored at the highest detection limit but this will substantially weaken the test. A statistician should be consulted on the potential use of Gehan ranking.

Directions for the Wilcoxon signed rank test are contained in Box 3-18, with an example in Box 3-19.

Box 3-16: Directions for the Sign Test (One-Sample)

COMPUTATIONS: Let C be the threshold. Compute the deviations $d_i = x_i - C$. If any of the deviations are zero delete them and correspondingly reduce the sample size. Finally, compute B, the number of times $d_i > 0$.

STEP 1: Null Hypothesis: H_0: median = C

STEP 2: Alternative Hypothesis:
i) H_A: median > C (upper-tail test)
ii) H_A: median < C (lower-tail test)
iii) H_A: median $\neq$ C (two-tail test)

STEP 3: Test Statistic: If $n \leq 20$, then the test statistic is B.

If $n > 20$, then the test statistic is $z_0 = \frac{B - n/2}{\sqrt{n/4}}$.

STEP 4 a): Critical Value: If $n \leq 20$, then use Table A-18 to find
i) $B_{upper}(n, 2\alpha)$
ii) $B_{lower}(n, 2\alpha)$ - 1
iii) $B_{lower}(n, \alpha) - 1$ and $B_{upper}(n, \alpha)$

If $n > 20$, then use Table A-1 to find
i) $z_{1-\alpha}$
ii) z_{α}
iii) $z_{1-\alpha/2}$

STEP 4 b): p-value: If $n \leq 20$, then let $p(i) = \binom{n}{i}\frac{1}{2^n}$ and the p-value is
i) $\sum_{i=B}^{n} p(i)$
ii) $\sum_{i=0}^{B} p(i)$
iii) $2 \cdot \min\left\{ \sum_{i=0}^{B} p(i),\ \sum_{i=B}^{n} p(i) \right\}$

If $n > 20$, then use Table A-1 to find
i) $P(Z > z_0)$
ii) $P(Z < z_0)$
iii) $2 \cdot P(Z > |z_0|)$

STEP 5 a): Conclusion: If $n \leq 20$, then
i) If $B \geq B_{upper}(n, 2\alpha)$, then reject the null hypothesis that the true population median is equal to the threshold value C.
ii) If $B \leq B_{lower}(n, 2\alpha)$ - 1, then reject the null hypothesis.
iii) If $B \geq B_{upper}(n, \alpha)$ or $B \leq B_{lower}(n, \alpha)$ - 1, then reject the null hypothesis.

If $n > 20$, then
i) If $z_0 > z_{1-\alpha}$, then reject the null hypothesis that the true population median is equal to the threshold value C.
ii) If $z_0 < z_{\alpha}$, then reject the null hypothesis.
iii) If $|z_0| > z_{1-\alpha/2}$, then reject the null hypothesis.

STEP 5 b): Conclusion: If p-value < α, then reject the null hypothesis that the true population median is equal to the threshold value C.

Box 3-17: An Example of the Sign Test (One-Sample)

The following 10 data points (in ppb) will be used to test the null hypothesis that the population median is at least 1000 ppb versus the lower-tail alternative. The decision maker has specified a 10% false rejection error rate (α) at 1000 ppb (C), and a 20% false acceptance error rate (β) at 900 ppb ($\tilde{\mu}_1$).

COMPUTATIONS: The table below displays the data values and the deviations:

x_i	974	1044	1093	897	879	1161	839	824	796	<750 (DL)
d_i	-26	44	93	-103	-121	161	-161	-176	-204	-625

Therefore, B = number of $d_i > 0 = 3$.

STEP 1: Null Hypothesis: H_0: median $\geq$ 1000 ppb

STEP 2: Alternative Hypothesis: H_A: median < 1000 ppb

STEP 3: Test Statistic: Since $n \leq 20$, the test statistic is $B = 3$.

STEP 4 a): Critical Value: Since $n \leq 20$, Table A-18 is used to find $B_{lower}(n, 2\alpha) - 1 = 2$.

STEP 4 b): p-value: Since $n \leq 20$, p-value = $\sum_{i=0}^{3} \binom{10}{i} \frac{1}{2^{10}} = 0.1719$.

STEP 5 a): Conclusion: Since test statistic = 3 > 2 = critical value, we fail to reject the null hypothesis that the true median is at least 1000 ppb.

STEP 5 b): Conclusion: Since p-value = 0.1719 > 0.10 = significance level, we fail to reject the null hypothesis that the true median is at least 1000 ppb.

3.3 COMPARING TWO POPULATIONS

The two-sample methods of this section concern the comparison of two population parameters (means, medians, or proportions). During Step 1: Review DQOs and Sampling Design, the population parameters were specified. In environmental applications, the two populations to be compared may be a potentially contaminated area with a background area or concentration levels from an upgradient and a downgradient well. Another example is comparing site concentration levels before and after clean-up to test for significant improvement.

The populations of interest can be one of two types, independent or paired. With independent populations, measurements are taken separately from the two groups of sampling units; for example, concentrations from a contaminated site and a background site. The observations from paired populations are correlated. Here, two measurements are taken upon one set of sampling units at separate instances; for example, measurements before and after clean-up or two labs making separate measurements on a single set of samples. Parametric and nonparametric methods are described for both independent and paired populations.

Box 3-18: Directions for the Wilcoxon Signed Rank Test (One-Sample)

COMPUTATIONS: Let C be the threshold. Compute the deviations $d_i = x_i$ - C. If any of the deviations are zero, then delete them and correspondingly reduce the sample size. Rank the absolute deviations, $|d_i|$, from smallest to largest. If there are tied observations, then assign the average rank. Let R_i be the signed rank of $|d_i|$, where the sign of R_i is determined by the sign of d_i.

STEP 1. Null Hypothesis: H_0: location = C

STEP 2. Alternative Hypothesis:
i) H_A: location > C (upper-tail test)
ii) H_A: location < C (lower-tail test)
iii) H_A: location ≠ C (two-tail test)

STEP 3. Test Statistic: If $n \le 20$, then $T^+ = \sum_{\{i:\, R_i > 0\}} R_i$, the sum of the positive signed ranks.

If $n > 20$, then $z_0 = \dfrac{T^+ - n(n+1)/4}{\sqrt{\text{var}(T^+)}}$,

where $\text{var}(T^+) = \dfrac{n(n+1)(2n+1)}{24} - \dfrac{1}{48}\sum_{j=1}^{g} t_j(t_j^2 - 1)$, g is the number of tied d_i groups and t_j is the size of group j.

STEP 4. a) Critical Value:

If $n \le 20$, then use Table A-7 to find:	If $n > 20$, then use Table A-1 to find:
i) $n(n+1)/2 - w_\alpha$	i) $z_{1-\alpha}$
ii) w_α	ii) z_α
iii) $w_{\alpha/2}$	iii) $z_{1-\alpha/2}$

STEP 4. b) p-value:

If $n \le 20$, then use Table A-7 to find:	If $n > 20$, then use Table A-1 to find:
i) $P(W \le n(n+1)/2 - T^+)$	i) $P(Z > z_0)$
ii) $P(W \le T^+)$	ii) $P(Z < z_0)$
iii) $2 \cdot \min\{P(W \le n(n+1)/2 - T^+),\ P(W \le T^+)\}$	iii) $2\,P(Z > \lvert z_0 \rvert)$

STEP 5. a) Conclusion:

If $n \le 20$, then:	If $n > 20$, then:
i) If $T^+ \ge n(n+1)/2 - w_\alpha$, then reject the null	i) If $z_0 > z_{1-\alpha}$, then reject the null
ii) If $T^+ \le w_\alpha$, then reject the null hypothesis.	ii) If $z_0 < z_\alpha$, then reject the null
iii) If $T^+ \ge n(n+1)/2 - w_{\alpha/2}$ or $T^+ \le w_\alpha$, then reject the null	iii) If $\lvert z_0 \rvert > z_{1-\alpha/2}$, then reject the null

STEP 5. b) Conclusion: If p-value < α, then reject the null hypothesis that the true population location is equal to the threshold C.

STEP 6. If the null hypothesis was not rejected and n>20, then the sample size necessary to achieve the DQOs assuming only one false acceptance error rate (β at μ_1) has been specified is

$$n \ge 1.16 \cdot \left\{ \frac{\text{var}(T^+)(z_{1-\alpha'} + z_{1-\beta})^2}{(\mu_1 - C)^2} + \frac{z_{1-\alpha'}^2}{2} \right\}.$$

If this is true then the false acceptance error rate has probably been satisfied (the value of α' is α for a one-sided test and $\alpha/2$ for a two-sided test).

Box 3-19: An Example of the Wilcoxon Signed Rank Test (One-Sample)

The following 10 data points (in ppb) will be used to test the null hypothesis that the population median is at least 1000 ppb versus the lower-tail alternative. The decision maker has specified a 10% false rejection error rate (α) at 1000 ppb (C), and a 20% false acceptance error rate (β) at 900 ppb).

974, 1044, 1093, 897, 879, 1161, 839, 824, 796, <750 (detection limit)

COMPUTATIONS: For this example, the only option is to assign the value 375 ppb (DL/2) to the nondetect. The table below displays the remaining computations.

x_i	974	1044	1093	897	879	1161	839	824	796	375
d_i	-26	44	93	-103	-121	161	-161	-176	-204	-625
$\|d_i\|$	26	44	93	103	121	161	161	176	204	625
rank	1	2	3	4	5	6.5	6.5	8	9	10
R_i	-1	2	3	-4	-5	6.5	-6.5	-8	-9	-10

STEP 1. Null Hypothesis: H_0: median $\geq$ 1000 ppb

STEP 2. Alternative Hypothesis: H_A: median < 1000 ppb

STEP 3. Test Statistic: Since $n \leq 20$, compute $T^+ = \sum_{\{i: R_i > 0\}} R_i = 2 + 3 + 6.5 = 11.5$

STEP 4. a) Critical Value: Since $n \leq 20$, Table A-7 is used to find $w_{0.10} = 14$.

STEP 4. b) p-value: Since $n \leq 20$, Table A-7 is used to find the p-value is between 0.05 and 0.075. Using software, the p-value is found to be 0.0527.

STEP 5. a) Conclusion: Since test statistic = $T^+ = 11.5 \leq 14$ = critical value, we reject the null hypothesis that the true population median is at least 1000.

STEP 5. b) Conclusion: Since p-value = 0.0527 < 0.10 = significance level, we reject the null hypothesis that the true population median is at least 1000.

NOTE: The Sign test failed to reject the null hypothesis for this example. Recall that the Wilcoxon Signed Rank test has more power than the Sign test if the distribution of the population is symmetric.

The hypothesis tests in this section may be used to determine if there is evidence that $\theta_1 - \theta_2 < \delta_0$, $\theta_1 - \theta_2 > \delta_0$, or $\theta_1 - \theta_2 \neq \delta_0$, where θ_1 and θ_2 represent population means, medians, or proportions and δ_0 represents the threshold value. Also, there are confidence interval procedures to estimate $\theta_1 - \theta_2$. Section 3.3.1.1 covers parametric methods for comparing two independent populations, while Section 3.3.1.2 covers parametric methods for comparing two paired populations. Section 3.3.2 describes nonparametric counterparts to the parametric methods.

3.3.1 Parametric Methods

These methods rely on the knowing the specific distribution of the populations or of the statistics of interest.

3.3.1.1 Independent Samples

3.3.1.1.1 The Two-Sample *t*-Test and Confidence Interval (Equal Variances)

Purpose: Test for a difference or estimate the difference between two population means when it can be assumed the population variances are approximately equal.

Data: A simple or systematic random sample $x_1, x_2, \ldots, x_m$ from one population, and an independent simple or systematic random sample $y_1, y_2, \ldots, y_n$ from the second population.

Assumptions: The two populations are independent. If not, then it is possible that a paired method could be used. Both are approximately normally distributed or the sample sizes are large (m and n both at least 30). If this is not the case, then a nonparametric procedure is an alternative. Finally, the variances of both populations are approximately equal. If the population variances are not equal (tests are available in Section 4.5), then use the methods of the next section.

Limitations and Robustness: The two-sample *t*-test with equal variances is robust to moderate violations of the assumption of normality, but not to large inequalities of variances. An alternative is the parametric methods for unequal variances described in the next section. The *t*-test is not robust against outliers because sample means and standard deviations are sensitive to outliers.

Directions for the two-sample t-test with equal variances are contained in Box 3-20, with an example in Box 3-22. Directions for a two-sample confidence interval with equal variances are contained in Box 3-21.

3.3.1.1.2 The Two-Sample *t*-Test and Confidence Interval (Unequal Variances)

Purpose: Test for a difference or estimate the difference between two population means when it is suspected the population variances are not equal.

Data: A simple or systematic random sample $x_1, x_2, \ldots, x_m$ from the one population, and an independent simple or systematic random sample $y_1, y_2, \ldots, y_n$ from the second population.

Assumptions: The two populations are independent. If not, then it is possible that a paired method could be used. Both are approximately normally distributed or the sample sizes are large (m and n both at least 30). If this is not the case, then a nonparametric procedure is an alternative.

Limitations and Robustness: The two-sample *t*-test with unequal variances is robust to moderate violations of the assumption of normality. The *t*-test is also not robust to outliers because sample means and standard deviations are sensitive to outliers.

Directions for the two-sample t-test with unequal variances are contained in Box 3-23, with an example in Box 3-25. Directions for a two-sample confidence interval with unequal variances are contained in Box 3-24.

Box 3-20: Directions for the Two-Sample *t*-Test (Equal Variances)

COMPUTATIONS: Calculate the sample means, $\overline{X}$ and $\overline{Y}$, and the sample variances, s_X^2 and s_Y^2 of the two populations. Also compute the pooled standard deviation estimate, $s_p = \sqrt{\frac{(m-1)\cdot s_X^2 + (n-1)\cdot s_Y^2}{m+n-2}}$

STEP 1. Null Hypothesis: $H_0 : \mu_X - \mu_Y = \delta_0$

STEP 2. Alternative Hypothesis:
i) $H_A : \mu_X - \mu_Y > \delta_0$ (upper-tail test)
ii) $H_A : \mu_X - \mu_Y < \delta_0$ (lower-tail test)
iii) $H_A : \mu_X - \mu_Y \neq \delta_0$ (two-tail test)

STEP 3. Test Statistic: $t_0 = \frac{\left(\overline{X} - \overline{Y}\right) - \delta_0}{s_p\sqrt{\frac{1}{m} + \frac{1}{n}}}$

STEP 4. a) Critical Value: Use Table A-2 to find:
i) $t_{m+n-2,1-\alpha}$
ii) $-t_{m+n-2,1-\alpha}$
iii) $t_{m+n-2,1-\alpha/2}$

STEP 4. b) p-value: Use Table A-2 to find:
i) $P\left(t_{m+n-2} > t_0\right)$
ii) $P\left(t_{m+n-2} < t_0\right)$
iii) $2 \cdot P\left(t_{m+n-2} > |t_0|\right)$

STEP 5. a) Conclusion:
i) If $t_0 > t_{m+n-2,1-\alpha}$, then reject the null hypothesis that the true difference between population means is equal to the threshold δ_0.
ii) If $t_0 < -t_{m+n-2,1-\alpha}$, then reject the null hypothesis.
iii) If $|t_0| > t_{m+n-2,1-\alpha/2}$, then reject the null hypothesis.

STEP 5. b) Conclusion: If p-value < α, then reject the null hypothesis that the true difference between population means is equal to the threshold δ_0.

STEP 6. If the null hypothesis was not rejected, there is only one false acceptance error rate (β at δ_1), and both *m* and *n* are at least $\frac{2s_p^2\left(z_{1-\alpha'} + z_{1-\beta}\right)^2}{\left(\delta_1 - \delta_0\right)^2} + \frac{z_{1-\alpha'}^2}{4}$, then the sample sizes were probably large enough to achieve the DQOs. The value of α' is α for a one-sided test and $\alpha/2$ for a two-sided test.

Box 3-21: Directions for a Two-Sample *t* Confidence Interval (Equal Variances)

COMPUTATIONS: Calculate the sample means, $\overline{X}$ and $\overline{Y}$, and the sample variances, s_X^2 and s_Y^2 of the two populations. Also compute the pooled standard deviation estimate, $s_p = \sqrt{\frac{(m-1)\cdot s_X^2 + (n-1)\cdot s_y^2}{m+n-2}}$.

A 100(1 - α)% confidence interval for $\mu_X - \mu_Y$ is $\left(\overline{X} - \overline{Y}\right) \pm t_{m+n-2,\,1-\alpha/2} \cdot s_p \sqrt{\frac{1}{m} + \frac{1}{n}}$, where Table A-2 is used to find $t_{m+n-2,\,1-\alpha/2}$.

Box 3-22: An Example of a Two-Sample *t*-Test (Equal Variances)

At a hazardous waste site, an area cleaned using an in-situ methodology (area 1) was compared with a similar, but relatively uncontaminated reference area (area 2). If the in-situ methodology worked, then the average contaminant levels at the two sites should be approximately equal. If the methodology did not work, then area 1 should have a higher average than the reference area. Suppose 7 random samples were taken from area 1, and 8 were taken from area 2. Methods described in Section 4.5 were used to determine that the variances were essentially equal. The false rejection error rate was set at 5% and the false acceptance error rate was set at 20% (β) if the difference between the areas is 2.5 ppm (δ_1).

COMPUTATIONS: The sample means and sample variances are $\overline{X}_1 = 7.8$, $\overline{X}_2 = 6.6$, $s_1^2 = 2.1$, and $s_2^2 = 2.2$.

The pooled standard deviation is: $s_p = \sqrt{\frac{(7-1)\cdot 2.1 + (8-1)\cdot 2.2}{7+8-2}} = 14676$

STEP 1. Null Hypothesis: $H_0: \mu_1 - \mu_2 = 0$

STEP 2. Alternative Hypothesis: $H_A: \mu_1 - \mu_2 > 0$

STEP 3. Test Statistic: $t_0 = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \delta_0}{s_p\sqrt{\frac{1}{m} + \frac{1}{n}}} = \frac{7.8 - 6.6 - 0}{14676\sqrt{\frac{1}{7} + \frac{1}{8}}} = 15799$

STEP 4. a) Critical Value: Using Table A-2, $t_{m+n-2,1-\alpha} = t_{13,\,0.95} = 1.771$

STEP 4. b) p-value: Using Table A-2, 0.05 < p-value < 0.10. (The exact p - value $= P\left(t_{m+n-2} > t_0\right) = P\left(t_{13} > 1.5799\right) = 0.0691$)

STEP 5. a) Conclusion: Since test statistic = 1.5799 < 1.771 = critical value, we fail to reject the null hypothesis of no difference between population means.

STEP 5. b) Conclusion: Since p-value = 0.0691 > 0.05 = significance level, we fail to reject the null hypothesis of no difference between population means.

STEP 6. Since the null hypothesis was not rejected and there is only one false acceptance error rate, we can compute the sample sizes necessary to achieve the DQOs. Each sample size must be at least

$$\frac{2\cdot 1.4676^2(1.645 + 0.842)^2}{(2.5-0)^2} + 0.25\cdot 1.645^2 = 4.94.$$

Since *m* and *n* are both greater than 4.94, the false acceptance error rate has probably been satisfied.

Box 3-23: Directions for the Two-Sample *t*-Test (Unequal Variances)

COMPUTATIONS: Calculate the sample means, $\overline{X}$ and $\overline{Y}$, and the sample variances, s_X^2 and s_Y^2 of the two populations. Also, compute the degrees of freedom (Satterthwaite's approximation) for the test:

$$df = \frac{\left(\frac{s_X^2}{m} + \frac{s_Y^2}{n}\right)^2}{\frac{s_X^2}{m^2(m-1)} + \frac{s_Y^2}{n^2(n-1)}}$$, rounded down to the next integer.

STEP 1. Null Hypothesis: $H_0 : \mu_X - \mu_Y = \delta_0$

STEP 2. Alternative Hypothesis:
i) $H_A : \mu_X - \mu_Y > \delta_0$ (upper-tail test)
ii) $H_A : \mu_X - \mu_Y < \delta_0$ (lower-tail test)
iii) $H_A : \mu_X - \mu_Y \neq \delta_0$ (two-tail test)

STEP 3. Test Statistic: $$t_0 = \frac{\left(\overline{X} - \overline{Y}\right) - \delta_0}{\sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}}}$$

STEP 4. a) Critical Value: Use Table A-2 to find:
i) $t_{df,1-\alpha}$
ii) $-t_{df,1-\alpha}$
iii) $t_{df,1-\alpha/2}$

STEP 4. b) p-value: Use Table A-2 to find:
i) $P(t_{df} > t_0)$
ii) $P(t_{df} < t_0)$
iii) $2 \cdot P(t_{df} > |t_0|)$

STEP 5. a) Conclusion:
i) If $t_0 > t_{df,1-\alpha}$, then reject the null hypothesis that the difference between the population means is equal to δ_0.
ii) If $t_0 < -t_{df,1-\alpha}$, then reject H_0.
iii) If $|t_0| > t_{df,1-\alpha/2}$, then reject H_0.

STEP 5. b) Conclusion: If p-value < α, then reject the null hypothesis that the difference between the population means is equal to δ_0.

STEP 6. If the null hypothesis was not rejected, then there is no simple method to determine if the sample sizes were large enough to achieve the DQOs.

Box 3-24: Directions for a Two-Sample *t* Confidence Interval (Unequal Variances)

COMPUTATIONS: Calculate the sample means, $\overline{X}$ and $\overline{Y}$, and the sample variances, s_X^2 and s_Y^2 of the two populations. Also, compute the degrees of freedom (Satterthwaite's approximation) for the confidence interval:

$$df = \frac{\left(\frac{s_X^2}{m} + \frac{s_Y^2}{n}\right)^2}{\frac{s_X^4}{m^2(m-1)} + \frac{s_Y^4}{n^2(n-1)}}$$, rounded down to the next integer.

A 100(1 - α)% confidence interval for $\mu_X - \mu_Y$ is $\left(\overline{X} - \overline{Y}\right) \pm t_{df,1-\alpha/2} \cdot \sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}}$, where Table A-2 is used to find $t_{m+n-2,1-\alpha/2}$.

Box 3-25: An Example of the Two-Sample *t*-Test (Unequal Variances)

At a hazardous waste site, an area cleaned using a new methodology (area 1) was compared with a similar area cleaned with the standard technology (area 2). If the new methodology worked, then the two sites should be approximately equal in average contaminant levels. If the new methodology did not work, then area 1 should have a higher average than the reference area. Suppose 7 random samples were taken from area 1 and 8 were taken from area 2. As the contaminant concentrations in the two areas are supposedly equal, we will test no difference in population means versus the upper-tail alternative. Using Section 4.5, it was determined that the variances of the two populations were not equal, therefore using Satterthwaite's method is appropriate. The false rejection error rate was set at 5% and the false acceptance error rate was set at 20% (β) if the difference between the areas is 2.5 ppm.

COMPUTATIONS: The sample means and sample variances are $\overline{X}_1 = 9.2$, $\overline{X}_2 = 6.1$, $s_1^2 = 1.3$, and $s_2^2 = 5.7$. Satterthwaite's approximation of the degrees of freedom for the test is:

$$df = \frac{\left(\frac{s_X^2}{m} + \frac{s_Y^2}{n}\right)^2}{\frac{s_X^4}{m^2(m-1)} + \frac{s_Y^4}{n^2(n-1)}} = \frac{\left(\frac{1.3}{7} + \frac{5.7}{8}\right)^2}{\frac{1.3^2}{7^2(7-1)} + \frac{5.7^2}{8^2(8-1)}} = 10.3 \rightarrow 10$$

STEP 1. Null Hypothesis: $H_0: \mu_1 - \mu_2 = 0$

STEP 2. Alternative Hypothesis: $H_A: \mu_1 - \mu_2 > 0$

STEP 3. Test Statistic: $t_0 = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \delta_0}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}} = \frac{9.2 - 6.1 - 0}{\sqrt{\frac{1.3}{7} + \frac{5.7}{8}}} = 3.271$

STEP 4. a) Critical Value: Using Table A-2, $t_{df,1-\alpha} = t_{10,0.95} = 1.812$

STEP 4. b) p-value: Using Table A-2, p-value < 0.005. (The exact p - value $= P(t_{df} > t_0) = P(t_{10} > 3.271) = 0.0042$)

STEP 5. a) Conclusion: Since test statistic = 3.271 > 1.812 = critical value, we reject the null hypothesis that there is no difference between the population means.

STEP 5. b) Conclusion: Since p-value = 0.0042 < 0.05 = significance level, we reject the null hypothesis that there is no difference between the population means.

3.3.1.1.3 Two-Sample Test for Population Proportions

Purpose: Test for a difference between two population proportions, P_1 and P_2. A simple or systematic random sample, $x_1,\ldots,x_n$, from one population and an independent simple or systematic random sample, $y_1, y_2, \ldots, y_n$, from second population of interest.

Assumptions: The data constitutes independent simple or systematic random samples from the populations.

Limitations and Robustness: To ensure the normal approximation is appropriate, compute mp_1, $m(1\text{-}p_1)$, np_2, and $n(1\text{-}p_2)$, where m and n are the sample sizes and p_1 and p_2 are the sample proportions. If all of the products are at least 5, then the normal approximation may be used. Otherwise, seek the assistance from a statistician as exact tests must be used. Since data positioning is used rather than actual data values, the procedures are robust to outliers.

Directions for the two-sample test for proportions are contained in Box 3-26, with an example in Box 3-28. Directions for a two-sample confidence interval for the difference between proportions are contained in Box 3-27.

3.3.1.2 Paired Samples

Observations from paired populations are correlated. The general set up for this teat involves taken two measurements upon one group of sampling units at separate instances; for example, measurements before and after clean-up, or two labs making separate measurements on a single set of objects.

3.3.1.2.1 The Paired *t*-Test and Confidence Interval

Purpose: Test for or estimate the difference between two paired population means.

Data: Two paired data sets $x_1,\ldots,x_n$ and $y_1,\ldots,y_n$.

Assumptions: The two data sets come from approximately normal distributions or $n \geq 30$.

Limitations and Robustness: Since there is really only one sample, the limitations for the paired *t*-test are the same as those for the one-sample *t*-test. These methods are robust against the population distribution deviating moderately from normality. However, they are not robust against outliers. In either case, the nonparametric methods of section 3.3.2.2 offer an alternative. Finally, these methods have difficulty dealing with non-detects. The substitution or adjustment procedures of chapter 4 are an alternative, but it is best to use a nonparametric method.

Directions for the paired t-test are given in Box 3-29, with an example in Box 3-31. Directions for a paired t confidence interval are given in Box 3-30.

Box 3-26: Directions for a Two-Sample Test for Proportions

COMPUTATIONS: Let p_i, i = 1,2, denote the sample proportion of data values from population i that fit the characteristic of interest. Also, calculate the pooled sample proportion p, which is the total number of data values that fit the characteristic divided by the total sample size, $m+n$.

To ensure the normal approximation is appropriate, compute mp_1, $m(1-p_1)$, np_2, $n(1-p_2)$. If all of these values are at least 5, then the normal approximation may be used. Otherwise, seek assistance from a statistician.

STEP 1. Null Hypothesis: $H_0 : P_1 - P_2 = \delta_0$

STEP 2. Alternative Hypothesis:
i) $H_A : P_1 - P_2 > \delta_0$ (upper-tail test)
ii) $H_A : P_1 - P_2 < \delta_0$ (lower-tail test)
iii) $H_A : P_1 - P_2 \neq \delta_0$ (two-tail test)

STEP 3. Test Statistic: $$z_0 = \frac{(p_1 - p_2) - \delta_0}{\sqrt{p(1-p)\cdot\left(\frac{1}{m}+\frac{1}{n}\right)}}.$$

STEP 4. a) Critical Value: Use Table A-1 to find:
i) $z_{1-\alpha}$
ii) z_{α}
iii) $z_{1-\alpha/2}$

STEP 4. b) p-value: Use Table A-1 to find:
i) $P(Z > z_0)$
ii) $P(Z < z_0)$
iii) $2 \cdot P(Z > |z_0|)$

STEP 5. a) Conclusion:
i) If $z_0 > z_{1-\alpha}$, then reject the null hypothesis that the difference in population proportions is equal to δ_0.
ii) If $z_0 < z_{\alpha}$, then reject the null hypothesis.
iii) If $|z_0| > z_{1-\alpha/2}$, then reject the null hypothesis.

STEP 5. b) Conclusion: If p-value < α, then reject the null hypothesis that the difference in population proportions is equal to δ_0.

STEP 6. If the null hypothesis was not rejected, there is only one false acceptance error rate (β at P_1-P_2), and both m and n are at least

$$\frac{2\left(z_{1-\alpha'} + z_{1-\beta}\right)^2 \cdot \bar{p}(1-\bar{p})}{(P_2 - P_1)^2} \text{ where } \bar{p} = \frac{p_1 + p_2}{2},$$

then the sample sizes were probably large enough to achieve the DQOs. The value of α' is α for a one-sided test and $\alpha/2$ for a two-sided test. If only one of m or n meets the sample size criterion, then use statistical software to calculate the power of the test, assuming that the true values for the proportions P_1 and P_2 are those obtained in the sample. If the estimated power is below 1-β, the false acceptance error rate has not been satisfied.

Box 3-27: Directions for Computing a Confidence Interval for the Difference Between Population Proportions

COMPUTATIONS: Let p_i, i = 1,2, denote the sample proportion of data values from population i that fit the characteristic of interest. Also, calculate the pooled sample proportion p, which is the total number of data values that fit the characteristic divided by the total sample size, $m+n$.

To ensure the normal approximation is appropriate, compute mp_1, $m(1-p_1)$, np_2, $n(1-p_2)$. If all of these values are at least 5, then the normal approximation may be used. Otherwise, seek the assistance from a statistician as exact tests must be used.

A 100(1 - α)% confidence interval for the difference between population proportions, $P_1 - P_2$, is

$$(p_1 - p_2) \pm z_{1-\alpha/2} \cdot \sqrt{p(1-p)\left(\frac{1}{m}+\frac{1}{n}\right)}$$, where Table A-1 is used to find $z_{1-\alpha/2}$

Box 3-28: An Example of a Two-Sample Test for Proportions

At a hazardous waste site, investigators must determine whether an area suspected to be contaminated with dioxin needs to be remediated. The possibly contaminated area (area 1) will be compared to a reference area (area 2) to see if dioxin levels in area 1 are greater than dioxin levels in the reference area. An inexpensive surrogate probe was used to determine if each individual sample is either "contaminated," i.e., over the health standard of 1 ppb, or "clean." The decision maker is willing to accept a false rejection decision error rate of 10% (α) and a false-negative decision error rate of 5% (β) when the difference in proportions between areas exceeds 0.1. A team collected 92 readings from area 1 (of which 12 were contaminated) and 80 from area 2, the reference area, (of which 10 were contaminated).

COMPUTATIONS: The sample proportion for area 1 is p_1 = 12/92 = 0.130, the sample proportion for area 2 is p_2 = 10/80 = 0.125, and the pooled sample proportion is p = (12 + 10) / (92 + 80) = 0.128. Since mp_1 = 12, $m(1-p_1)$ = 80, np_2 = 10, $n(1-p_2)$ = 70 are all at least 5, the normal approximation is appropriate.

STEP 1. Null Hypothesis: $H_0 : P_1 - P_2 = 0$

STEP 2. Alternative Hypothesis: $H_A : P_1 - P_2 > 0$ (upper-tail test)

STEP 3. Test Statistic: $$z_0 = \frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{m}+\frac{1}{n}\right)}} = \frac{0.13 - 0.125}{\sqrt{0.128(1-0.128)\left(\frac{1}{92}+\frac{1}{80}\right)}} = 0.10.$$

STEP 4. a) Critical Value: Using Table A-1, $z_{1-\alpha} = z_{0.90} = 1.282$

STEP 4. b) p-value: Using Table A-1, $P(Z > z_0) = P(Z > 0.10) = 1 - 0.5398 = 0.4602$

STEP 5. a) Conclusion: Since test statistic = 0.10 < 1.282 = critical value, we fail to reject the null hypothesis of no difference between population proportions.

STEP 5. b) Conclusion: Since p-value = 0.4602 > 0.10 = significance level, we fail to reject the null hypothesis of no difference between population proportions.

STEP 6. Since the null hypothesis was not rejected and there is only one false acceptance error rate (β = 0.05 at a difference of P_1 - P_2 = 0.1) has been specified, it is possible to calculate the sample sizes that achieve the DQOs. So m and n must be of size at least

$$\frac{2(1.282 + 1.645)^2 0.1275(1-0.1275)}{0.1^2} = 190.6,$$

since $\overline{P} = \frac{0.13 + 0.125}{2} = 0.1275$. As both m and n are less than 190.6, the false acceptance error rate has not been satisfied. Therefore, the null hypothesis was not rejected, but the samples sizes were not large enough to ensure adequate power in the test.

Box 3-29: Directions for the Paired *t*-Test

COMPUTATIONS: Compute the differences $d_i = y_i - x_i$ for $i = 1,\ldots,n$ and the sample mean, $\overline{D}$, and sample standard deviation, s_D, of the differences.

STEP 1. Null Hypothesis: $H_0 : \mu_D = \delta_0$

STEP 2. Alternative Hypothesis:
i) $H_A : \mu_D > \delta_0$ (upper-tail test)
ii) $H_A : \mu_D < \delta_0$ (lower-tail test)
iii) $H_A : \mu_D \neq \delta_0$ (two-tail test)

STEP 3. Test Statistic: $t_0 = \dfrac{\overline{D} - \delta_0}{s_D / \sqrt{n}}$

STEP 4. a) Critical Value: Use Table A-2 to find:
i) $t_{n-1,1-\alpha}$
ii) $-t_{n-1,1-\alpha}$
iii) $t_{n-1,1-\alpha/2}$

STEP 4. b) p-value: Use Table A-2 to find:
i) $P(t_{n-1} > t_0)$
ii) $P(t_{n-1} < t_0)$
iii) $2 \cdot P(t_{n-1} > |t_0|)$

STEP 5. a) Conclusion:
i) If $t_0 > t_{n-1,1-\alpha}$, then reject the null hypothesis that there is no difference between the population means.
ii) If $t_0 < -t_{n-1,1-\alpha}$, then reject the null hypothesis.
iii) If $|t_0| > t_{n-1,1-\alpha/2}$, then reject the null hypothesis.

STEP 5. b) Conclusion: If p-value < α, then reject that the difference between the population means is δ_0.

STEP 6. If the null hypothesis was not rejected, there is only one false acceptance error rate (β at μ_1), and $n \geq \dfrac{s_D^2 \left(z_{1-\alpha'} + z_{1-\beta}\right)^2}{(\mu_1 - C)^2} + \dfrac{z_{1-\alpha'}^2}{2}$, then the sample size was probably large enough to achieve the DQOs.
The value of α' is α for a one-sided test and $\alpha/2$ for a two-sided test.

Box 3-30: Directions for Computing the Paired *t* Confidence Interval

COMPUTATIONS: Compute the differences $d_i = y_i - x_i$ for $i = 1,\ldots,n$ and the sample mean, $\overline{D}$, and sample standard deviation, s_D, of the differences.

A 100(1 - α)% confidence interval for μ_D is $\overline{D} \pm t_{n-1,1-\alpha/2} \cdot s_D / \sqrt{n}$, where Table A-2 is used to find $t_{n-1,1-\alpha/2}$.

Box 3-31: An Example of the Paired *t*-Test

Consider the following 9 pairs of data points (in ppb):

x_i	178	52	161	245	164	184	157	308	130
y_i	92	67	62	206	106	126	108	314	126

This data will be used to test the null hypothesis that there is no difference in means versus the alternative that the mean of the first population is greater than the mean of the second population. The decision maker has specified a 5% false rejection error rate and a 10% false acceptance error rate at a difference of 50 ppb.

COMPUTATIONS: The table below displays the computations.

d_i	-86	15	-99	-39	-58	-58	-49	6	-4

The sample mean of the differences is -41.33 and the sample standard deviation is 39.92.

STEP 1. Null Hypothesis: $H_0: \mu_D = 0$

STEP 2. Alternative Hypothesis: $H_A: \mu_D < 0$ (lower-tail test)

STEP 3. Test Statistic: $t_0 = \dfrac{\overline{D} - \delta_0}{s_D/\sqrt{n}} = \dfrac{-41.33 - 0}{39.92/\sqrt{9}} = -3.11.$

STEP 4. a) Critical Value: Using Table A-2, $-t_{n-1,1-\alpha} = -t_{8,\,0.95} = -1.860$.

STEP 4. b) p-value: Using Table A-2, 0.005 < p-value < 0.01. (The exact p - value $= \mathrm{P}(t_{n-1} < t_0) = \mathrm{P}(t_8 < -3.11) = 0.0072$).

STEP 5. a) Conclusion: Since test statistic = 3.11 < 1.860 = critical value, we reject the null hypothesis of no difference between population means.

STEP 5. b) Conclusion: Since p-value = 0.0072 < 0.05 = significance level, we reject the null hypothesis of no difference between population means.

3.3.2 Nonparametric Methods

These methods rely on the relative rankings of data values. Knowledge of the precise form of the population distributions is not necessary.

3.3.2.1 Independent Samples

3.3.2.1.1 The Wilcoxon Rank Sum Test

Purpose: Test for a difference between two population means. The Wilcoxon Rank Sum test applied with the Quantile test, provides a powerful combination for detecting true differences between two population distributions.

Data: A random sample $x_1, x_2, \ldots, x_m$ from one population, and an independent random sample $y_1, y_2, \ldots, y_n$ from the second population.

Assumptions: The validity of the random sampling and independence assumptions should be verified by review of the procedures used to select the sampling points. The two underlying distributions are assumed to have approximately the same shape (variance) and that the only difference between them is a shift in location. A qualitative test of this assumption can be done by comparing histograms.

Limitations and Robustness: The Wilcoxon signed rank test may produce misleading results if there are many tied data values. When many ties are present, their relative ranks are the same, and this has the effect of diluting the statistical power of the Wilcoxon test. If possible, results should be recorded with sufficient accuracy so that a large number of tied values do not occur. Estimated concentrations should be reported for data below the detection limit, even if these estimates are negative, as their relative magnitude to the rest of the data is of importance. If this is not possible, substitute the value DL/2 for each value below the detection limit providing all the data have the same detection limit. When different detection limits are present, all data could be censored at the highest detection limit but this will substantially weaken the test. A statistician should be consulted on the potential use of Gehan ranking.

Directions for the Wilcoxon Rank Sum test are given in Box 3-32, with an example in Box 3-33. Directions for the large sample approximation for the Wilcoxon Ranked Sum test are given in the example in Box 3-34.

3.3.2.1.2 The Quantile Test

Purpose: Test for a shift to the right in the right-tail of population 1 versus population 2 This may be regarded as being equivalent to detecting if the values in the right-tail of population 1 distribution are generally larger than the values in the right-tail of the population 2 distribution.

Data: A simple or systematic random sample, $x_1, x_2, \ldots, x_n$, from the site population and an independent simple or systematic random sample, $y_1, y_2, \ldots, y_m$, from the background population.

Assumptions: The validity of the random sampling and independence assumptions is assured by using proper randomization procedures, which can be verified by reviewing the procedures used to select the sampling points.

Limitations and Robustness: Since the Quantile test focuses on the right-tail, large outliers will bias results. Also, the Quantile test says nothing about the center of the two distributions. Therefore, this test should be used in combination with a location test like the *t*-test (if the data are normally distributed) or the Wilcoxon Rank Sum test.

Box 3-32: Directions for the Wilcoxon Rank Sum Test

COMPUTATIONS: Rank the pooled data from smallest to largest assigning average rank to ties. Sum the ranks of the first population and denote this by R_1. Then compute

$$W_0 = R_1 - \frac{m(m+1)}{2}$$

STEP 1. Null Hypothesis: $H_0: \mu_X - \mu_Y = 0$ (no difference between population means)

STEP 2. Alternative Hypothesis:
i) $H_A: \mu_X - \mu_Y > 0$ (upper-tail test)
ii) $H_A: \mu_X - \mu_Y < 0$ (lower-tail test)
iii) $H_A: \mu_X - \mu_Y \neq 0$ (two-tail test)

STEP 3. Test Statistic: If *m* and *n* are at most 20, then the test statistic is W_0.

If *m* and *n* are greater than 20, then compute $z_0 = \frac{W_0 - mn/2}{\sqrt{\text{var}(W_0)}}$,

where $\text{var}(W_0) = \frac{mn(m+n+1)}{12} - \left\{ \frac{mn}{12(m+n)(m+n-1)} \sum_{j=1}^{g} t_j\left(t_j^2 - 1\right) \right\}$,

g is the number of tied groups, and t_j is the number of ties in the j^{th} group.

STEP 4. a) Critical Value: to find:

If *m* and *n*< =20, use Table A-8 to find:
i) $mn - w_\alpha$
ii) w_α
iii) $w_{\alpha/2}$

If *m* and *n* >20 use Table A-1
i) $z_{1-\alpha}$
ii) z_α
iii) $z_{1-\alpha/2}$

STEP 4. b) p-value:

If *m* and *n* <=20, use Table A-8 to find:
i) $P(W_{rs} < mn - W_0)$
ii) $P(W_{rs} < W_0)$
iii) $2 \cdot \min\{P(W_{rs} < mn - W_0), P(W_{rs} < W_0)\}$

If *m* and *n* > 20 use Table A-1 to find:
i) $P(Z > z_0)$
ii) $P(Z < z_0)$
iii) $2 \cdot P(Z > |z_0|)$

STEP 5. a) Conclusion:

If *m* and *n* are at most 20, then
i) If $W_0 \geq mn - w_\alpha$, then reject the null hypothesis.
ii) If $W_0 \leq w_\alpha$, then reject the null hypothesis.
iii) If $W_0 \geq mn - w_{\alpha/2}$ or $W_0 \leq w_{\alpha/2}$, then reject the null hypothesis.

If *m* and *n* are greater than 20, then
i) If $z_0 > z_{1-\alpha}$, then reject the null hypothesis of no difference between population means.
ii) If $z_0 < z_\alpha$, then reject the null hypothesis.
iii) If $|z_0| > z_{1-\alpha/2}$, then reject the null hypothesis.

STEP 5. b) Conclusion: If p-value < significance level, then reject the null hypothesis

STEP 6. If the null hypothesis was not rejected, then the sample sizes necessary to achieve the DQOs should be computed. If the sample sizes are large, only one false acceptance error rate (β at δ_1) has been specified, then the false acceptance error rate has probably been satisfied if both *m* and *n* are at least

$$1.16 \cdot \left\{ \frac{2 \cdot \text{var}(W_0) \cdot \left(z_{1-\alpha'} + z_{1-\beta}\right)^2}{\delta_1^2} + \frac{z_{1-\alpha'}^2}{4} \right\}.$$

NOTE: The value of α' is α for a one-sided test and $\alpha/2$ for a two-sided test. The large sample normal approximation is adequate as long as min(*m*, *n*) > 10.

Box 3-33: An Example of the Wilcoxon Rank Sum Test

At a hazardous waste site, an area cleaned (area 1) was compared with a relatively uncontaminated reference area (area 2). If the methodology worked, then the two sites should be approximately equal in average contaminant levels. If the methodology did not work, then area 1 should have a higher average than the reference area. The false rejection error rate was set at 10% (α) and the false acceptance error rate was set at 20% (β) if the difference between the areas is 2.5 ppb. Seven random samples were taken from area 1 and 8 samples were taken from area 2:

Area 1: 17, 23, 26, 5, 13, 13, 12
Area 2: 16, 20, 5, 4, 8, 10, 7, 3

COMPUTATIONS: The ordered pooled data and their ranks are (Area 1 denoted by *):

Pooled data	3	4	5*	5	7	8	10	12*	13*	13*	16	17*	20	23*	26*
Rank	1	2	3.5*	3.5	5	6	7	8*	9.5*	9.5*	11	12*	13	14*	15*

The sum of the ranks of area 1 is $R_1 = 3.5 + 8 + 9.5 + 9.5 + 12 + 14 + 15 = 71.5$ and $W_0 = 71.5 - 7(7 + 1)/2 = 43.5$

STEP 1. Null Hypothesis: $H_0 : \mu_X - \mu_Y = 0$

STEP 2. Alternative Hypothesis: $H_A : \mu_X - \mu_Y > 0$

STEP 3. Test Statistic: Since m and n are less than 20, the test statistic is W_0.

STEP 4. a) Critical Value: Using Table A-8, $mn - w_{0.10} = 56 – 16 = 40$

STEP 4. b) p-value: Using Table A-8, $mn - w_{0.05} < W_0 < mn - w_{0.025}$, which implies 0.05 < p-value < 0.025. Using statistical software, p-value = 0.0410.

STEP 5. a) Conclusion: Since Test Statistic = 43.5 > 40 = Critical Value, we reject the null hypothesis of no difference between population means.

STEP 5. b) Conclusion: Since p-value = 0.0410 < 0.10 (significance level), we reject the null hypothesis

Directions for the Quantile test are contained in Box 3-35, with an example in Box 3-36.

3.3.2.1.3 The Slippage Test

Purpose: Test for a shift to the right in the extreme right-tail of population 1 (site) versus population 2 (background). This is equivalent to asking if a set of the largest values of the site distribution are larger than the maximum value of the background distribution.

Data: A simple or systematic random sample, $x_1, x_2, \ldots, x_m$, from the site population and an independent simple or systematic random sample, $y_1, y_2, \ldots, y_n$, from the background population.

Box 3-34: A Large Sample Example of the Wilcoxon Rank Sum Test

Arsenic concentrations (in ppm) from a site are to be compared to a reference area. The null hypothesis is that the means of the two areas are equal versus the upper-tail alternative. The false rejection error rate was set at 5% (α) and the false acceptance error rate was set at 20% (β) if the difference between the areas is 2.5 ppm.

Site concentrations (m = 22): 11.2, 11.3, 12.2, 13.2, 14.2, 15.9, 16.3, 17.1, 18.6, 19.2, 21.5, 22.3, 22.4, 22.7, 22.8, 23.3, 24.1, 25.8, 30.2, 30.7, 31.4, 37.1

background concentrations (n = 21): 6.1, 8.5, 11.1, 11.3, 12.6, 12.8, 13.6, 15.0, 15.2, 15.3, 16.1, 16.2, 17.0, 17.1, 17.6, 19.2, 19.2, 19.6, 21.1, 22.2, 25.0

COMPUTATIONS: The ordered pooled data and their ranks are (site concentrations are denoted by *):

Value	Rank
6.1	1
8.5	2
11.1	3
11.2*	4*
11.3	5.5
11.3*	5.5*
12.2*	7*
12.6	8
12.8	9
13.2*	10*
13.6	11

Value	Rank
14.2*	12*
15.0	13
15.2	14
15.3	15
15.9*	16*
16.1	17
16.2	18
16.3*	19*
17.0	20
17.1	21.5
17.1*	21.5*

Value	Rank
17.6	23
18.6*	24*
19.2	26
19.2	26
19.2*	26*
19.6	28
21.1	29
21.5*	30*
22.2	31
22.3*	32*
22.4*	33*

Value	Rank
22.7*	34*
22.8*	35*
23.3*	36*
24.1*	37*
25.0	38
25.8*	39*
30.2*	40*
30.7*	41*
31.4*	42*
37.1*	43*

The sum of the site ranks is

R_1 = 4 + 5.5 + 7 + 10 + 12 + 16 + 19 + 21.5 + 24 + 26 + 30 + 32 + 33 + 34 + 35 + 36 + 37 + 39 + 40 + 41 + 42 + 43 = 587

and $W_0 = 587 - 22(22 + 1)/2 = 334$.

STEP 1. Null Hypothesis: $H_0 : \mu_X - \mu_Y = 0$

STEP 2. Alternative Hypothesis: $H_A : \mu_X - \mu_Y > 0$

STEP 3. Test Statistic: Since m and n are greater than 20, the test statistic is $z_0 = \dfrac{W_0 - mn/2}{\sqrt{\text{var}(W_0)}}$. To compute var($W_0$), we need to determine the number of tied groups and the number of values in each group. There are 3 tied groups, so g = 3. The number of tied values in the groups are 2 (at 11.3), 2 (at 17.1), and 3 (at 19.2). Therefore,

$$\text{var}(W_0) = \frac{22 \cdot 21 \cdot (22 + 21 + 1)}{12} - \left\{ \frac{22 \cdot 21}{12 \cdot (22 + 21)(22 + 21 - 1)} \left[2 \cdot \left(2^2 - 1\right) + 2 \cdot \left(2^2 - 1\right) + 3 \cdot \left(3^2 - 1\right) \right] \right\} = 1693.23 .$$

Finally, $z_0 = \dfrac{334 - 22 \cdot 21/2}{\sqrt{1693.23}} = 2.50$.

STEP 4. a) Critical Value: Using Table A-1, $z_{0.95}$ = 1.645.

STEP 4. b) p-value: Using Table A-1, $P(Z > 2.50) = 1 - 0.9938 = 0.0062$.

STEP 5. a) Conclusion: Since Test Statistic = 2.50 > 1.645 = Critical Value, we reject the null hypothesis of no difference between population means.

STEP 5. b) Conclusion: Since p-value = 0.0062 < 0.05 = significance level, we reject the null hypothesis of no difference between population means.

Box 3-35: Directions for the Quantile Test

COMPUTATIONS: Select a quantile, $0.5 \le b_1 < 1$. Rank the pooled data from smallest to largest. Find the number of pooled data points larger than the b_1^{th} quantile,

$c = m + n - \text{floor}\{(m+n-1)\cdot b_1\} - 1$, where floor means to calculate the value and discard all decimals.

STEP 1. Null Hypothesis: H_0: The right-tails of the two population distributions are the same.

STEP 2. Alternative Hypothesis: H_A: The right-tail of the distribution of Population 1 (site) is shifted to the right, i.e., the values in the right-tail of the site distribution are larger than the values in the right-tail of the background distribution.

STEP 3. Test Statistic: s = number of site samples greater than the b_1^{th} quantile.

STEP 4. a) Critical Value: If m and n are both at most 20, then use Table A-19 to find q_α.

STEP 4. b) p-value: If m and n are both at least 15, then use Table A-1 to find

$$1 - \text{P}\left(Z < \frac{s - 0.5 - \mu}{\sigma}\right), \text{ where } \mu = n \cdot \frac{c}{m+n} \text{ and } \sigma = \sqrt{n \cdot \frac{c}{m+n} \cdot \left(1 - \frac{c}{m+n}\right) \cdot \frac{m}{m+n-1}}$$

STEP 5. a) Conclusion: If $s \ge q_\alpha$, then reject the null hypothesis that the right tails of the two population distributions are the same.

STEP 5. b) Conclusion: If p-value < α, then reject the null hypothesis that the right tails of the two population distributions are the same.

Box 3-36: A Example of the Quantile Test

At a hazardous waste site a new, cheaper, in-site methodology was compared against an existing methodology by remediating separate areas of the site using each method. It will be assumed that the new methodology works as well as the old and we will test for evidence that the new method leaves higher concentrations. A Quantile Test with a significance level of 0.05 will be used to make this determination based on 12 samples from the area remediated using the new methodology and 7 samples from the area remediated using the standard methodology.

New Methodology: 7, 18, 2, 4, 6, 11, 5, 9, 10, 2, 3, 3
Standard Methodology: 17, 8, 20, 4, 6, 5, 4

COMPUTATIONS: The 0.75 quantile is selected. The ranked pooled data set is:

2*, 2*, 3*, 3*, 4, 4, 4*, 5*, 5, 6, 6*, 7*, 8, 9*, 10*, 11*, 17, 18*, 20,

where * denoted samples from the new methodology portion of the site. The number of values larger than the 0.75 quantile is:

$$c = m + n - \text{floor}\{(m+n-1)\cdot b_1\} - 1 = 7 + 12 - \text{floor}\{(7+12-1)\cdot 0.75\} - 1 = 5$$

STEP 1. Null Hypothesis: H_0: The right-tails of the two population distributions are the same.

STEP 2. Alternative Hypothesis: H_A: The right-tail of the distribution of Population 1 (site) is shifted to the right, i.e., the values in the right-tail of the site distribution are larger than the values in the right-tail of the background distribution.

STEP 3. Test Statistic: s = number of site samples greater than the 0.75 quantile = 3

STEP 4. a) Critical Value: Since m and n are both less than 20, we use Table A-19 to find $q_{0.10} = 5$

STEP 4. b) p-value: Since m and n are both less than 15, we can't use the normal approximation for the p-value.

STEP 5. a) Conclusion: Since $s = 3 < 5 = q_{0.10}$, we fail to reject the null hypothesis that the right-tails of the two-population distributions are the same.

Assumptions: The validity of the random sampling and independence assumptions is assured by using proper randomization procedures, which can be verified by reviewing the procedures used to select the sampling points.

Limitations and Robustness: Since the Slippage test focuses on the right-tail, large outliers will bias results. Also, the Slippage test says nothing about the center of the two distributions. Therefore, this test should be used in combination with a location test like the *t*-test (if the data are normally distributed) or the Wilcoxon Rank Sum test.

Directions for the slippage test are contained in Box 3-37, with an example in Box 3-38.

Box 3-37: Directions for the Slippage Test

COMPUTATIONS: Let $x_1,\ldots,x_m$ be the site data and $y_1,\ldots,y_n$ be the background data. Order the samples separately.

STEP 1. Null Hypothesis:	H_0: The right-tails of the two population distributions are the same.
STEP 2. Alternative Hypothesis:	H_A: The extreme right-tail of the distribution of Population 1 (site) is shifted to the right, i.e., the largest values of the site distribution are larger than the largest values of the background distribution.
STEP 3. Test Statistic:	s = number of site values greater than the maximum background value.
STEP 4. a) Critical Value:	Use Table A-14 to find S_α. Note that α can be 0.01, 0.05, or 0.10.
STEP 4. b) p-value:	$\frac{n \cdot m!}{(m+n)!} \cdot \sum_{i=s}^{m} \frac{(m+n-i-1)!}{(m-i)!} = 1 - \frac{n \cdot m!}{(m+n)!} \cdot \sum_{i=0}^{s-1} \frac{(m+n-i-1)!}{(m-i)!}$
	Where "n!' = nx(n-1)x(n-2)x(n-3)x…3x2x1
STEP 5. a) Conclusion:	If $s \geq S_\alpha$, then reject the null hypothesis that the right-tails of the two population distributions are the same.
STEP 5. b) Conclusion:	If p-value < α, then reject the null hypothesis that the right-tails of the two population distributions are the same.

3.3.2.2 Paired Samples

Recall that the observations from paired populations are correlated. The general setting involves taken two measurements upon one group of sampling units at separate instances; for example, measurements before and after clean-up or two labs making separate measurements on a single set of objects.

3.3.2.2.1 The Sign Test

Purpose: Test for a difference between the medians of two paired populations. This test is very similar to the one sample version presented in section 3.2.2.1.

Box 3-38: A Example of the Slippage Test

At a hazardous waste site a new, cheaper, in-site methodology was compared against an existing methodology by remediating separate areas of the site using each method. It will be assumed that the new methodology works as well as the old and we will test for evidence that the new method leaves higher concentrations. The Slippage Test with a significance level of 0.05 will be used to make this determination based on 12 samples from the area remediated using the new methodology and 7 samples from the area remediated using the standard methodology.

New Methodology: 7, 18, 2, 4, 6, 11, 5, 9, 10, 2, 3, 3
Standard Methodology: 17, 8, 20, 4, 6, 5, 4

COMPUTATIONS: The ordered samples are:

New Methodology: 2, 2, 3, 3, 4, 5, 6, 7, 9, 10, 11, 18
Standard Methodology: 4, 4, 5, 6, 8, 17, 20

STEP 1. Null Hypothesis: H_0: The right-tails of the two population distributions are the same.

STEP 2. Alternative Hypothesis: H_A: The extreme right-tail of the distribution of Population 1 (site) is shifted to the right, i.e., the largest values of the site distribution are larger than the largest values of the background distribution.

STEP 3. Test Statistic: s = number of site values greater than the maximum background value = 0.

STEP 4. a) Critical Value: Using Table A-14, $S_{0.05} = 6$.

STEP 4. b) p-value: $$1 - \frac{n \cdot m!}{(m+n)!} \cdot \sum_{i=0}^{0-1} \frac{(m+n-i-1)!}{(m-i)!} = 1 - 0 = 1$$

STEP 5. a) Conclusion: Since test statistic = 0 < 6 = $S_{0.05}$, we fail to reject the null hypothesis that the right-tails of the two population distributions are the same.

STEP 5. b) Conclusion: Since p-value = 1 > 0.05 = significance level, we fail to reject the null hypothesis that the right-tails of the two population distributions are the same.

Data: Two paired data sets $x_1,\ldots,x_n$ and $y_1,\ldots,y_n$ selected randomly or systematically.

Assumptions: The Sign test can be used no matter what the underlying distributions may be.

Limitations and Robustness: The Sign test has less power than the two-sample *t*-test or the Wilcoxon Signed Rank Test. However, the Sign test makes no distributional assumptions like the other two tests and it can handle non-detects (if the detection limit is below the threshold).

Directions for the paired populations sign test are contained in Box 3-39, with an example in Box 3-40.

3.3.2.2.2 The Wilcoxon Signed Rank Test

Purpose: Test for a difference between the locations (means or medians) of two paired populations. This test is very similar to the one sample version presented in Section 3.2.2.2.

Box 3-39: Directions for the Sign Test (Paired Samples)

COMPUTATIONS: Compute the deviations $d_i = x_i - y_i$. If any of the deviations are zero delete them and correspondingly reduce the sample size. Finally, compute B = number of differences greater than zero.

STEP 1. Null Hypothesis: H_0: median 1 = median 2

STEP 2. Alternative Hypothesis:
i) H_A: median 1 > median 2 (upper-tail test)
ii) H_A: median1 < median 2 (lower-tail test)
iii) H_A: median 1 ≠ median 2 (two-tail test)

STEP 3. Test Statistic: If $n \le 20$, then the test statistic is B.

If $n > 20$, then the test statistic is $z_0 = \dfrac{B - n/2}{\sqrt{n/4}}$.

STEP 4. a) Critical Value:

If $n \le 20$, then use Table A-18 to find:	If $n > 20$, then use Table A-1 to find:
i) $B_{upper}(n, 2\alpha)$	i) $z_{1-\alpha}$
ii) $B_{lower}(n, 2\alpha) - 1$	ii) z_{α}
iii) $B_{lower}(n, \alpha)$ and $B_{lower}(n, \alpha)$	iii) $z_{1-\alpha/2}$

STEP 4. b) p-value: If $n \le 20$, then let $p(i) = \binom{n}{i}\frac{1}{2^n}$ where $\binom{n}{i} = \dfrac{n!}{(n-i)!i!}$

i) $\sum_{i=B}^{n} p(i)$

ii) $\sum_{i=o}^{B} p(i)$

iii) $2 \cdot \min\left\{\sum_{i=0}^{B} p(i), \sum_{i=B}^{n} p(i)\right\}$

If $n > 20$, then use Table A-1 to find:

i) $P(Z > z_0)$

ii) $P(Z < z_0)$

iii) $2 \cdot P(Z > |z_0|)$

STEP 5. a) Conclusion: If $n \le 20$, then

i) If $B \ge B_{upper}(n, 2\alpha)$, then reject the null hypothesis that the two population medians are equal.
ii) If $B \le B_{lower}(n, 2\alpha) - 1$, then reject the null hypothesis.
iii) If $B \ge B_{upper}(n, \alpha)$ or $B \le B_{lower}(n, \alpha) - 1$, then reject the null hypothesis.

If $n > 20$, then

i) If $z_0 > z_{1-\alpha}$, then reject the null hypothesis that the two population medians are equal.
ii) If $z_0 < z_{\alpha}$, then reject the null hypothesis.
iii) If $|z_0| > z_{1-\alpha/2}$, then reject the null hypothesis.

STEP 5. b) Conclusion: If p-value < α, then reject the null hypothesis that the two population medians are equal.

Box 3-40: An Example of the Sign Test (Paired Samples)

Consider the following 9 pairs of data points (in ppb):

x_i	178	52	161	245	164	184	157	308	130
y_i	92	67	62	206	106	126	108	314	126

This data will be used to test the null hypothesis that there is no difference in medians versus the alternative that the median of the first population is greater than the median of the second population. The decision maker has specified a 5% false rejection error rate and a 10% false acceptance error rate at a difference of 50 ppb.

COMPUTATIONS: The table below displays the computations.

d_i	86	-15	99	39	58	58	49	-6	4

Therefore, B = the number of differences greater than zero = 7.

STEP 1. Null Hypothesis: H_0: $\tilde{\mu}_1 = \tilde{\mu}_2$

STEP 2. Alternative Hypothesis: H_A: $\tilde{\mu}_1 > \tilde{\mu}_2$

STEP 3. Test Statistic: Since $n \leq 20$, the test statistic is $B = 7$.

STEP 4. a) Critical Value: Since $n \leq 20$, Table A-18 is used to find $B_{upper}(n, 2\alpha) = 8$.

STEP 4. b) p-value: Since $n \leq 20$, p-value = $\sum_{i=7}^{9} \binom{9}{i} \frac{1}{2^9} = 0.0898$.

STEP 5. a) Conclusion: Since test statistic = 7 < 8 = critical value, we fail to reject H_0.

STEP 5. b) Conclusion: Since p-value = 0.0898 > 0.05 = significance level, we fail to reject H_0.

Data: Two paired data sets $x_1,\ldots,x_n$ and $y_1,\ldots,y_n$ selected randomly or systematically.

Assumptions: The data sets come from a approximately symmetric distributions.

Limitations and Robustness: For large sample sizes ($n > 50$), the paired t-test is more robust to violations of its assumptions than the Wilcoxon signed rank test. For small sample sizes, if the data are not approximately symmetric or normally distributed, the sign test should be used.

The Wilcoxon signed rank test may produce misleading results if there are many tied data values. Ties have the effect of diluting the statistical power of the Wilcoxon test. If possible, results should be recorded with sufficient accuracy so that a large number of tied values do not occur. Estimated concentrations should be reported for data below the detection limit, even if these estimates are negative, as their relative magnitude to the rest of the data is of importance. If this is not possible, substitute the value DL/2 for each value below the detection limit providing all the data have the same detection limit. When different detection limits are present, all data could be censored at the highest detection limit but this will substantially weaken the test. A statistician should be consulted on the potential use of Gehan ranking.

Directions for the paired sample Wilcoxon signed rank test are contained in Box 3-41, with an example in Box 3-42. Directions for the large sample version of the test are contained in the example in Box 3-43.

Box 3-41: Directions for the Wilcoxon Signed Rank Test (Paired Samples)

COMPUTATIONS: Compute the deviations $d_i = x_i - y_i$. If any of the deviations are zero delete them and correspondingly reduce the sample size. Rank the absolute deviations, $|d_i|$, from smallest to largest. If there are tied observations, then assign the average rank. Let R_i be the signed rank of $|d_i|$, where the sign of R_i is determined by the sign of d_i.

STEP 1. Null Hypothesis: H_0: $\text{location}_1 = \text{location}_2$

STEP 2. Alternative Hypothesis:
i) H_A: $\text{location}_1 > \text{location}_2$ (upper-tail test)
ii) H_A: $\text{location}_1 < \text{location}_2$ (lower-tail test)
iii) H_A: $\text{location}_1 \neq \text{location}_2$ (two-tail test)

STEP 3. Test Statistic: If $n \leq 20$, then $T^+ = \sum_{\{i: R_i > 0\}} R_i$, the sum of the positive signed ranks.

If $n > 20$, then $z_0 = \dfrac{T^+ - n(n+1)/4}{\sqrt{\text{var}(T^+)}}$,

where $\text{var}(T^+) = \dfrac{n(n+1)(2n+1)}{24} - \dfrac{1}{48}\sum_{j=1}^{g} t_j\left(t_j^2 - 1\right)$, g is the number of tied d_i groups and t_j is the size of group j.

STEP 4. a) Critical Value:

If $n \leq 20$, use Table A-7 to find:	If $n > 20$, use Table A-1 to find:
i) $n(n+1)/2 - w_\alpha$	i) $z_{1-\alpha}$
ii) w_α	ii) z_α
iii) $w_{\alpha/2}$	iii) $z_{1-\alpha/2}$

STEP 4. b) p-value:

If $n \leq 20$, use Table A-7 to find:	If $n > 20$, use Table A-1 to find:
i) $P\left(W \leq n(n+1)/2 - T^+\right)$	i) $P(Z > z_0)$
ii) $P\left(W \leq T^+\right)$	ii) $P(Z < z_0)$
iii) $2 \cdot \min\left\{P\left(W \leq n(n+1)/2 - T^+\right), P\left(W \leq T^+\right)\right\}$	iii) $2 \cdot P(Z > \lvert z_0 \rvert)$

STEP 5. a) Conclusion:

If $n \leq 20$,	If $n > 20$,
i) If $T^+ \geq n(n+1)/2 - w_\alpha$, then reject the null	i) If $z_0 > z_{1-\alpha}$, then reject the null
ii) If $T^+ \leq w_\alpha$, then reject the null hypothesis.	ii) If $z_0 < z_\alpha$, then reject the null
iii) If $T^+ \geq n(n+1)/2 - w_{\alpha/2}$ or $T^+ \leq w_\alpha$, then reject the null	iii) If $\lvert z_0 \rvert > z_{1-\alpha/2}$, then reject the null

STEP 5. b) Conclusion: If p-value < α, then reject the null hypothesis that the two population locations are equal.

STEP 6. If the null hypothesis was not rejected, then the sample sizes necessary to achieve the DQOs should be computed. If the sample size is large, only one false acceptance error rate (β at δ_1) has been specified, and

$$n \geq 1.16 \cdot \left\{ \frac{\text{var}\left(T^+\right)\left(z_{1-\alpha} + z_{1-\beta}\right)^2}{\delta_1^2} + \frac{z_{1-\alpha}^2}{2} \right\},$$

then the false acceptance error rate has probably been satisfied. The value of α' is α for a one-sided test and $\alpha/2$ for a two-sided test.

Box 3-42: An Example of the Wilcoxon Signed Rank Test (Paired Samples)

Consider the following 9 pairs of data points (in ppb):

x_i	178	52	161	245	164	184	157	308	130
y_i	92	67	62	206	106	126	108	314	126

This data will be used to test the null hypothesis that there is no difference in medians versus the alternative that the median of the first population is greater than the median of the second population. The decision maker has specified a 5% false rejection error rate and a 10% false acceptance error rate at a difference of 50 ppb.

COMPUTATIONS: The table below displays the computations.

d_i	86	-15	99	39	58	58	49	-6	4
$\|d_i\|$	86	15	99	39	58	58	49	6	4
rank	8	3	9	4	6.5	6.5	5	2	1
R_i	8	-3	9	4	6.5	6.5	5	-2	1

STEP 1. Null Hypothesis: H_0: median1 = median 2

STEP 2. Alternative Hypothesis: H_A: median 1< median 2 (lower-tail test)

STEP 3. Test Statistic: Since $n \le 20$, compute $T^+ = \sum_{\{i: R_i > 0\}} R_i = 1 + 4 + 5 + 6.5 + 6.5 + 8 + 9 = 40$

STEP 4. a) Critical Value: Since $n \le 20$, Table A-7 is used to find $n(n+1)/2 - w_{0.05} = 45 - 8 = 37$.

STEP 4. b) p-value: Since $n \le 20$, Table A-7 is used to find the approximate p-value of 0.025. Using software, the p-value is found to be 0.0195.

STEP 5. a) Conclusion: Since test statistic = $40 \ge 37$ = critical value, we reject the null hypothesis of equal population medians.

STEP 5. b) Conclusion: Since p-value = 0.0195 < 0.05 = significance level, we reject the null hypothesis of equal population medians.

NOTE: The Sign test failed to reject the null hypothesis for this example. The Wilcoxon Signed Rank test has more power than the Sign test if the distribution of the population is symmetric.

3.4 COMPARING SEVERAL POPULATIONS SIMULTANEOUSLY

This section describes procedures for comparing several population means simultaneously. The comparison is made between several treatment populations versus a single control population. For example, we can simultaneously test for a difference between the concentrations at several sites versus a single background area.

The methods in section compare the several populations while controlling the overall significance level. If individual two-sample *t*-tests are performed at significance level α, then the overall significance level is higher than α. Possible much higher if there are many experimental groups. For example, comparing two experimental groups to a control using two two-sample *t*-tests at significance level 0.05 results in an overall significance level of 1-(1-0.05)(1-0.05) = 0.0975. The tests in this section are more powerful in detecting differences between the experimental groups and the control than other multiple comparison methods that compare all possible pairs of population means, e.g., the ANOVA *F*-test and the Kruskal-Wallis test.

Box 3-43: A Large Sample Example of the Wilcoxon Signed Rank Test (Paired Samples)

A hazardous waste site has recently gone through remediation. To determine if the remediation method was effective, 24 paired samples (before and after clean-up) will be compared. The following data will be used to test the null hypothesis that there is no difference in medians versus the alternative that the median concentration before is greater than the median concentration after clean-up. The decision maker has specified a 5% false rejection error rate and a 10% false acceptance error rate at a difference of 30 ppb.

before: 331, 351, 259, 323, 305, 336, 196, 233, 336, 349, 352, 341, 172, 253, 275, 285, 212, 349, 301, 343, 368, 332, 374, 311
after: 246, 270, 229, 326, 295, 238, 278, 302, 331, 264, 267, 249, 288, 272, 270, 313, 337, 284, 271, 253, 295, 271, 289, 281

COMPUTATIONS: The table below displays the computation of the signed ranks.

d_i	$\|d_i\|$	rank	R_i
85	85	17.5	17.5
81	81	14	14
30	30	8	8
-3	3	1	-1
10	10	4	4
98	98	22	22
-82	82	15	-15
-69	69	12	-12

d_i	$\|d_i\|$	rank	R_i
5	5	2.5	2.5
85	85	17.5	17.5
85	85	17.5	17.5
93	93	21	21
-117	117	23	-23
-19	19	5	-5
5	5	2.5	2.5
-28	28	6	-6

d_i	$\|d_i\|$	rank	R_i
-125	125	24	-24
65	65	11	11
30	30	8	8
90	90	20	20
73	73	13	13
60	60	10	10
85	85	17.5	17.5
30	30	8	8

STEP 1. Null Hypothesis: H_0: median 1 = median 2

STEP 2. Alternative Hypothesis: H_A: median 1 > median 2 (upper-tail test)

STEP 3. Test Statistic: Since $n > 20$, compute $z_0 = \dfrac{T^+ - n(n+1)/4}{\sqrt{\text{var}(T^+)}}$. First,

$T^+ = \sum_{\{i:\, R_i > 0\}} R_i = 17.5 + 14 + \cdots + 8 = 214$. To compute $\text{var}(T^+)$, we need to identify the number of tied groups and the number of values in each of the tied groups. There are 3 tied groups so $g = 3$. The number of values in the tied groups are 2 (at 2.5), 3 (at 30), and 4 (at 85). Therefore,

$$\text{var}(T^+) = \frac{24 \cdot (24+1)(2 \cdot 24+1)}{24} - \frac{1}{48}\left\{2 \cdot \left(2^2 - 1\right) + 3 \cdot \left(3^2 - 1\right) + 4 \cdot \left(4^2 - 1\right)\right\} = 1223.125 .$$

Finally, $z_0 = \dfrac{214 - 24 \cdot (24+1)/4}{\sqrt{1223.125}} = 1.83$

STEP 4. a) Critical Value: Since $n > 20$, Table A-1 is used to find $z_{0.95} = 1.645$.

STEP 4. b) p-value: Since $n > 20$, Table A-1 is used to find $P(Z > z_0) = 1 - 0.9664 = 0.0336$.

STEP 5. a) Conclusion: Since test statistic = 1.83 > 1.645 = critical value, we reject the null hypothesis of equal population medians.

STEP 5. b) Conclusion: Since p-value = 0.0336 < 0.05 = significance level, we reject the null hypothesis of equal population medians.

The Dunnett test of Section 3.4.1.1 is a parametric test that compares several treatment means to a control mean. Section 3.4.2.1 presents a nonparametric alternative to the Dunnett test.

3.4.1 Parametric Methods

These methods rely on the knowing the specific distribution of the populations.

3.4.1.1 Dunnett's Test

Purpose: Test simultaneously for a difference between several population means and the mean of a control population. A typical application would involve comparing different potentially cleaned areas of a hazardous waste site to an uncontaminated reference area.

Data: A set of k-1 independent experimental random samples, $x_{i1},\ldots,x_{in_i}$, i = 1,…, k-1 and an independent random sample from the control population, $x_{c1},\ldots,x_{cn_c}$.

Assumptions: The Dunnett test is similar to the two-sample t-test so the populations need to be approximately normal or the sample sizes need to be large (≥30). If this is not the case, then the nonparametric Fligner-Wolfe test can be used. However, that test simple detects a difference between all population means. Both tests assume that the k populations have equal variances.

Limitations and Robustness: The Dunnett critical values (Table A-14) are for the case of equal number of samples in the control and experimental groups, but are approximately correct provided the number of samples from the investigated groups are more than half but less than double the size of the control group. Also, Table A-14 is for one-tailed tests only.

Directions for Dunnett's test are contained in Box 3-44, with an example in Box 3-45.

Box 3-44: Directions for Dunnett's Test

COMPUTATIONS: Calculate the sample mean, $\overline{X}_i$, and the sample variance, s_i^2 for each of the k-1 populations along with $\overline{X}_c$ and s_c^2. Also compute $\text{SSE} = (n_c - 1)s_c^2 + \sum_{i=1}^{k-1}(n_i - 1)s_i^2$

STEP 1. Null Hypothesis:	H_0: $\mu_I = \mu_c$, for i = 1, …,k-1
STEP 2. Alternative Hypotheses:	i) H_A: At least one $\mu_I > \mu_c$, for i = 1, …,k-1. ii) H_A: At least one $\mu_I < \mu_c$, for i = 1, …,k-1.
STEP 3. Test Statistic:	For each of the k-1 experimental populations, compute: $t_i = \dfrac{\left\lvert \overline{X}_i - \overline{X}_c \right\rvert}{\sqrt{\dfrac{\text{SSE}}{N\text{-}k}\left(\dfrac{1}{n_i} + \dfrac{1}{n_c}\right)}}$, where N is the total sample size.
STEP 4. a) Critical Value:	Use Table A-15 to determine the critical value $T_D(\alpha)$ where the degrees of freedom are N-k.
STEP 4. b) p-value:	Too complex for this guidance.
STEP 5. a) Conclusion:	For each i, reject the corresponding null hypothesis if $t_i > T_D(\alpha)$.
STEP 5. b) Conclusion:	Use the critical value approach.

Box 3-45: An Example of Dunnett's Test

At a hazardous work site, 6 designated areas previously identified as 'problems' have been cleaned. In order for these areas to be admitted to the overall work site abatement program these areas should be sown to be the same as the reference area. The means of these areas will be compared to mean of a reference area located on the site using Dunnett's test. The null hypothesis is no difference between the means of the 'problem' areas and the mean of the reference area and the alternative hypotheses are that the 'problem' means are greater than the reference area mean. The significance level is 0.05. Summary statistics for the data are given in the table below.

	Reference	IAK3	ZBF6	3BG5	4GH2	5FF3	6GW4
n_i	7	6	5	6	7	8	7
$\overline{X}_i$	10.3	11.4	12.2	10.2	11.4	11.9	12.1
s_i^2	2.5	2.6	3.3	3.0	3.2	2.6	2.8
n_c/n_i		1.16	1.4	1.16	1	0.875	1
t_i		1.18	1.93	0.11	1.22	1.84	2.00

Since all of the sample size ratios fall between 0.5 and 2.0, Dunnett's test may be used.

COMPUTATIONS: The sample means and sample variances are displayed in rows 2 and 3 above. The t_i row is the collection of test statistics that are generated in Step 3 below. The final computation is
$\text{MSE} = (7-1)\cdot 2.5 + (6-1)\cdot 2.6 + \cdots + (7-1)\cdot 2.8 = 110.4$.

STEP 1. Null Hypothesis: H_0: $\mu_I = \mu_c$, for i = 1, …,6

STEP 2. Alternative Hypotheses: H_A: $\mu_I > \mu_c$, for i = 1, …,6

STEP 3. Test Statistic: For each of the 6 'problem' areas, t_i was computed. For example, $t_1 = \dfrac{|11.4 - 10.3|}{\sqrt{\dfrac{110.4}{46-7}\cdot\left(\dfrac{1}{6}+\dfrac{1}{7}\right)}} = 1.18$. The results are displayed in the 5th row of the table.

STEP 4. a) Critical Value: Using Table A-15 of Appendix A with N-k = 46-7 = 39 degrees of freedom, the critical value $T_D(0.95)$ = 2.37.

STEP 5. a) Conclusion: Since $t_i < 2.37 = T_D(0.95)$ for all i, we fail to reject the null hypothesis. We conclude that none of the 'problem' areas have contamination levels significantly higher than the reference area. Therefore, these areas may be admitted to the work site abatement program.

3.4.2 Nonparametric Methods

These methods rely on the relative rankings of data values. Knowledge of the precise form of the population distributions is not necessary.

3.4.2.1 The Fligner-Wolfe Test

Purpose: Test simultaneously for a difference between several population locations (means or medians) and the location of a control population. A typical application would involve

comparing different potentially cleaned areas of a hazardous waste site to an uncontaminated reference area. This test is similar to the Wilcoxon Rank Sum test.

Data: A set of k-1 independent experimental simple or systematic random samples, $x_{i1},\ldots,x_{in_i}$, i = 1,…, k-1 and an independent simple or systematic random sample from the control population, $x_{c1},\ldots,x_{cn_c}$. Let $N^* = \sum_{i=1}^{k-1} n_i$ and $N = N^* + n_c$.

Assumptions: All data sets come from distributions with similar shapes.

Limitations and Robustness: The alternative hypothesis is quite restrictive as it demands all of the experimental groups have population means that are either at least as large as the control mean (or at most the control mean). They are not appropriate tests when some experimental group means may be higher than the control mean and some might be lower.

Directions for the Fligner-Wolfe test are contained in Box 3-46, with an example in Box 3-47.

Box 3-46: Directions of the Fligner-Wolfe Test

COMPUTATIONS: Rank the N pooled data points from smallest to largest assigning average rank to ties. Let r_{ij} denote the rank of data point X_{ij}. Compute

$$FW_0 = \sum_{j=1}^{k-1}\sum_{i=1}^{n_i} r_{ij} - \frac{N^*\left(N^*+1\right)}{2}.$$

Note that the first term of FW is the sum of the ranks from all experimental groups.

STEP 1. Null Hypothesis: H_0 : location$_i$ = location$_c$, for $i = 1,\ldots,k-1$

STEP 2. Alternative Hypothesis: i) H_A : location$_i \geq$ location$_c$, for $i = 1,\ldots,k-1$ (at least one strict inequality)

ii) H_0 : location$_i \leq$ location$_c$, for $i = 1,\ldots,k-1$ (at least one strict inequality)

STEP 3. Test Statistic: If min(n_c, N^*) ≤ 20, then the test statistic is FW_0.

If min(n_c, N^*) > 20, then the test statistic is $z_0 = \dfrac{FW_0 - n_c N^*/2}{\sqrt{\text{var}\left(FW_0\right)}}$, where

$\text{var}_0\left(FW\right) = \dfrac{n_c N^*\left(N+1\right)}{12} - \dfrac{n_c N^*}{12N\left(N-1\right)}\sum_{j=1}^{g} t_j\left(t_j^2-1\right)$, g is the number of tied groups and t_j is the number of tied values in the j^{th} group.

STEP 4. a) Critical Value: If min(n_c, N^*) ≤ 20, then use Table A-8 to find w_α.

If min(n_c, N^*) > 20, then use Table A-1 to find:

i) $z_{1-\alpha}$

ii) z_α

STEP 4. b) p-value: If min(n_c, N^*) ≤ 20, then use Table A-8 to find:

i) $P\left(W_{rs} < n_c N^* - FW_0\right)$

ii) $P\left(W_{rs} < FW_0\right)$

If min(n_c, N^*) > 20, then use Table A-1to find:

i) $P\left(Z > z_0\right)$

ii) $P\left(Z < z_0\right)$

STEP 5. a) Conclusion: If min(n_c, N^*) ≤ 20, then

i) If $FW_0 \geq n_c N^* - w_\alpha$, then reject the null hypothesis of no difference between population locations.

ii) If $FW_0 \leq w_\alpha$, then reject the null hypothesis.

If min(n_c, N^*) > 20, then

i) If $z_0 > z_{1-\alpha}$, then reject the null hypothesis of no difference between population locations.

ii) If $z_0 < z_\alpha$, then reject the null hypothesis.

STEP 5. b) Conclusion: If p-value < significance level, then reject the null hypothesis of no difference between population locations.

NOTE: The large sample normal approximation is adequate as long as min(n_c, N^*) > 10.

Box 3-47: An Example of the Fligner-Wolfe Test

4 contaminated ponds are to be combined with a fifth (reference) pond before further work can commence. If the ponds are approximately equal, the proposed remediation method will be acceptable. The assumption of normality cannot be made but all ponds were produced by the same waste process, they should exhibit similar characteristics. The significance level is 0.05. Data values with ranks in parentheses are given in the table below.

Reference Pond	North	East	South	West
10 (1)	15 (3)	20 (8)	12 (2)	19 (5.5)
19 (5.5)	25 (14)	26 (15)	16 (4)	20 (8)
20 (8)	32 (19.5)	34 (21.5)	22 (10)	28 (16)
23 (11.5)	39 (23)	43 (24.5)	23 (11.5)	31 (18)
24 (13)	43 (24.5)	52 (28)	34 (21.5)	49 (27)
30 (17)				
32 (19.5)				
46 (26)				

COMPUTATIONS: Note that $n_c = 8$ and $N^* = 20$ since $n_i = 5$ for $i = 1,\ldots,4$. Compute

$$FW_0 = \sum_{j=1}^{k-1}\sum_{i=1}^{n_i} r_{ij} - \frac{N^*\left(N^*+1\right)}{2} = 304.5 - 210 = 94.5 .$$

STEP 1. Null Hypothesis: $\mathrm{H}_0 : \mathrm{pond}_i = \mathrm{pond}_c, \text{ for } i = 1,\ldots,4$

STEP 2. Alternative Hypothesis: $\mathrm{H}_\mathrm{A} : \mathrm{pond}_i \geq \mathrm{pond}_c, \text{ for } i = 1,\ldots,4$ (at least one strict inequality)

STEP 3. Test Statistic: Since $\min(n_c, N^*) = 8 \leq 20$, then the test statistic is $FW_0 = 94.5$.

STEP 4. a) Critical Value: Since $\min(n_c, N^*) \leq 20$, Table A-8 is used to find $w_{0.05} = 47$.

STEP 4. b) p-value: Since $\min(n_c, N^*) \leq 20$, using Table A-8 gives a p-value of $\mathrm{P}\left(W_{rs} < 8 \cdot 20 - 94.5\right) = \mathrm{P}\left(W_{rs} < 65.5\right) > 0.10$ (the exact p-value is 0.2380.)

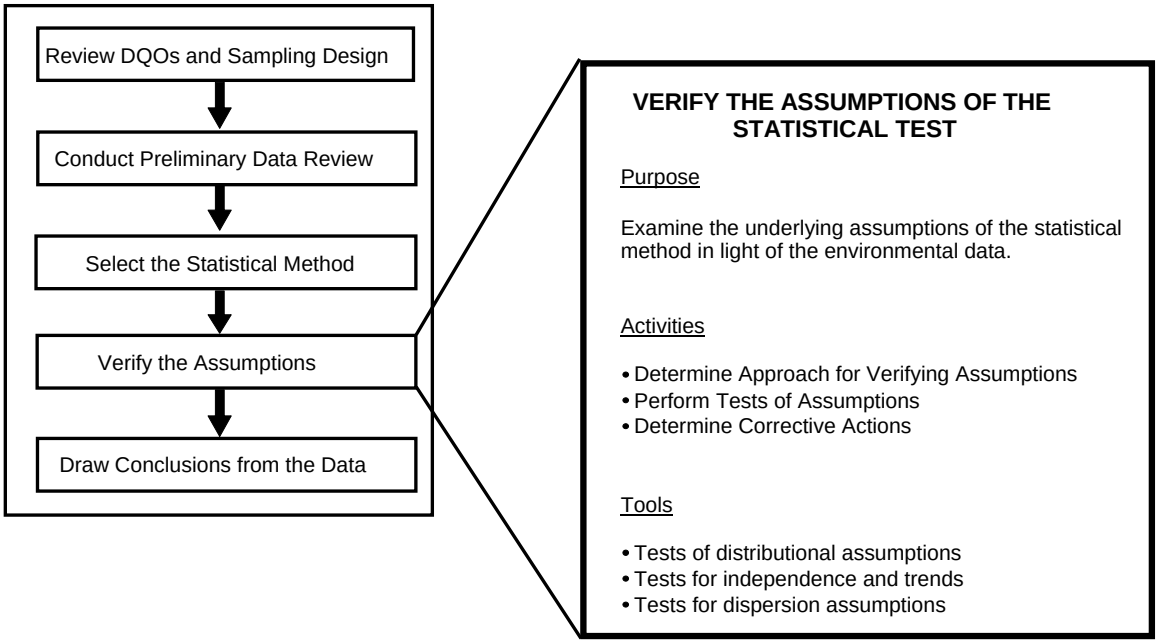
STEP 5. a) Conclusion: Since $\min(n_c, N^*) \leq 20$, and $FW_0 = 94.5 < 113 = 8 \cdot 20 - 47 = n_c N^* - w_\alpha$, we fail to reject the null hypothesis of no difference between ponds.

STEP 5. b) Conclusion: Since p-value = 0.2380 > 0.05 = significance level, we fail to reject the null hypothesis of no difference between population ponds.

CHAPTER 4

STEP 4: VERIFY THE ASSUMPTIONS OF THE STATISTICAL METHOD

THE DATA QUALITY ASSESSMENT PROCESS



Step 4: Verify the Assumptions of the Statistical Test

- Determine approach for verifying assumptions.
  - Identify any strong graphical evidence from the preliminary data review.
  - Review (or develop) the statistical model for the data.
  - Select the methods for verifying assumptions.
- Perform tests of assumptions.
  - Adjust for distributional assumption if warranted.
  - Perform the calculations required for the tests.
- If necessary, determine corrective actions.
  - Determine whether data transformations will correct the problem.
  - If data are missing, explore the feasibility of using theoretical justification or of collecting new data.
  - Consider robust procedures or nonparametric hypothesis tests.

List of Boxes

List of Tables

CHAPTER 4

STEP 4: VERIFY THE ASSUMPTIONS OF THE STATISTICAL METHOD

4.1 OVERVIEW AND ACTIVITIES

In this step, the analyst should assess the validity of the statistical method chosen in step 3 by examining its underlying assumptions or determine that the data support the underlying assumptions necessary for the selected method, or if a different statistical method should be used.

If it is determined that one or more of the assumptions is not met, then an alternative action is required. Typically, this means the selection of a different statistical method. Each method of Chapter 3 provides a detailed list of alternative methods.

Parametric tests also have difficulty dealing with outliers and non-detects. If either is found in the data, then a corrective action would be to use the corresponding nonparametric method. In general, nonparametric methods handle outliers and non-detects better than parametric methods. If a trend in the data is detected or the data are found to be dependent, then the methods of Chapter 3 should not be applied. Time series or geostatistical methods may be required and a statistician should be consulted. For a more extensive discussion of the overview and activities of this step, see *Data Quality Assessment: A Reviewer's Guide* (EPA QA/G-9R) (U.S.EPA 2004).

4.2 TESTS FOR DISTRIBUTIONAL ASSUMPTIONS

Many statistical tests and models are only appropriate for data that follow a particular distribution. This section will aid in determining if a distributional assumption of a statistical test is satisfied, in particular, the assumption of normality. Two of the most important distributions for tests involving environmental data are the normal distribution and the lognormal distribution, both of which are discussed in this section. To test if the data follow a distribution other than the normal distribution or the lognormal distribution, apply the chi-square test discussed in Section 4.2.5 or consult a statistician.

There are many methods available for verifying the assumption of normality ranging from simple to complex. This section discusses methods based on graphs, sample moments (kurtosis and skewness), sample ranges, the Shapiro-Wilk test and closely related tests, and goodness-of-fit tests. Discussions for the simplest tests contain step-by-step directions and examples based on the data in Table 4-1. These tests are summarized in Table 4-2. This section ends with a comparison of the tests to help the analyst select a test for normality.

Table 4-1. Data for Examples in Section 4.2

These are 10 observations of dust from commercial air filters in ppm

15.63	11.00	11.75	10.45	13.18	10.37	10.54	11.55	11.01	10.23	$\overline{X}$ =11.57 s = 1.677 n = 10

Table 4-2. Tests for Normality

Test	Section	Sample Size	Recommended Use
Shapiro Wilk *W* Test	4.2.2	≤ 5000	Highly recommended.
Filliben's Statistic	4.2.2	≤ 100	Highly recommended, especially when used in conjunction with a normal probability plot.
Skewness and Kurtosis Tests	4.2.3	> 50	Useful for large sample sizes.
Studentized Range Test	4.2.4	≤ 1000	Highly recommended (with some conditions).
Geary's Test	4.2.4	> 50	Useful when tables for other tests are not available.
Chi-Square Test	4.2.5	Large[a]	Useful for grouped data and when the comparison distribution is known.
Lilliefors Kolmogorov- Smirnoff Test	4.2.5	> 50	Useful when tables for other tests are not available.

[a] The necessary sample size depends on the number of groups formed when implementing this test. Each group should contain at least 5 observations.

The normal distribution is one of the most common probability distributions in the analysis of environmental data. It can often be used to approximate other probability distributions and in some instances more complex distributions can be transformed to an appropriate normal distribution. Additionally, as the sample size becomes larger, the sample mean has an approximate normal distribution hence the common assumption associated with parametric tests is that the data follows a normal distribution.

The graph of a normally distributed random variable is bell-shaped (see Figure 4-1) with the highest point located at the mean which is equal to the median. A normal curve is symmetric about the mean, hence the part to the left of the mean is a mirror image of the part to the right. In environmental data, random errors occurring during the measurement process may be normally distributed.

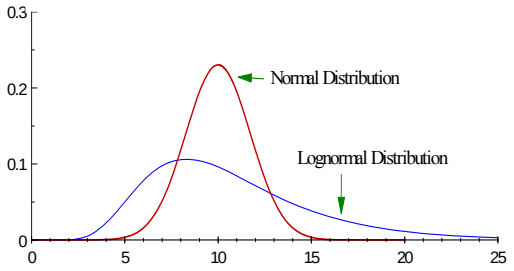


Figure 4-1. Density Plots for the Normal and Lognormal Distributions

Environmental data commonly exhibit frequency distributions that are non-negative and skewed with heavy or long right-tails. Several standard parametric probability models have these properties, including the Weibull, gamma, and lognormal distributions. The lognormal distribution (Figure 4-1) is a commonly used distribution for modeling environmental contaminant data. The advantage to this distribution is that a simple (logarithmic) transformation will transform a lognormal distribution into a normal distribution. Therefore, the methods for testing for normality described in this section can be used to test for lognormality if a logarithmic transformation has been used. It should be noted that as the shape parameter of the lognormal approaches zero, the lognormal approaches a normal. When the shape parameter is large, the lognormal becomes more positively skewed.

4.2.1 Graphical Methods

Graphical methods (Section 2.3) present detailed information about data sets that may not be apparent from a test statistic. Histograms, stem-and-leaf plots, and normal probability plots are some graphical methods that are useful for determining whether or not data follow a normal curve. Both the histogram and stem-and-leaf plot of a normal distribution are bell-shaped. The normal probability plot of a normal distribution follows a straight line. For non-normally distributed data, there will be large deviations in the tails or middle of a normal probability plot.

Using a plot to decide if the data are normally distributed is a qualitative judgment. For extremely non-normal data, it is easy to make this determination; however, in many cases the decision is not straightforward. Therefore, formal test procedures are usually necessary to test the assumption of normality.

4.2.2 Normal Probability Plot Tests

One of the most powerful tests for normality is the Shapiro-Wilk W test. This test is similar to computing a correlation between the quantiles of the standard normal distribution and the ordered values of a data set. If the normal probability plot is approximately linear (i.e., the data follow a normal curve), the test statistic will be relatively high. If the normal probability plot is nonlinear, then the test statistic will be relatively low.

The W test is recommended in several EPA guidance documents and in many statistical texts. Tables of critical values for sample sizes up to 50 have been developed for determining the significance of the test statistic. However, many software packages can perform the W test for data sets with sample sizes as large as 5000. This test is difficult to compute by hand since it requires two different sets of tabled values and a large number of summations and multiplications. Therefore, directions for implementing this test are not given in this document.

Several tests related to the Shapiro-Wilk test have been proposed. D'Agostino's test for sample sizes between 50 and 1000 and Royston's test for sample sizes up to 2000 are two such tests that approximate some of the key quantities or parameters of the W test.

Another related test is the Filliben statistic, also called the probability plot correlation coefficient. This test measures the linearity of the points on the normal probability plot. Similar to the W test, if the normal probability plot is approximately linear, then the correlation coefficient will be relatively high. Although easier to compute that the W test, the Filliben statistic is still difficult to compute by hand. Therefore, directions for implementing this test are not given in this guidance.

4.2.3 Coefficient of Skewness/Coefficient of Kurtosis Tests

The degree of symmetry (or asymmetry) displayed by a data set is measured by the coefficient of skewness (g_3). The coefficient of kurtosis, g_4, measures the degree of flatness of a probability distribution near its center. Several test methods have been proposed using these coefficients to test for normality. One method tests for normality by adjusting the coefficients of

skewness and kurtosis to approximate a standard normal distribution for sample sizes greater than 50.

Two other tests based on these coefficients include a combined test based on a chi-squared (χ^2) distribution and Fisher's cumulant test. Fisher's cumulant test computes the exact sampling distribution of g_3 and g_4; therefore, it is more powerful than previous methods which assume that the distributions of the two coefficients are normal. Fisher's cumulant test requires a table of critical values, and these tests require a sample size of greater than 50. Tests based on skewness and kurtosis are rarely used as they are less powerful than many alternatives.

4.2.4 Range Tests

Nearly 100% of the area of a normal curve lies within ±5 standard deviations from the mean. Tests for normality have been developed based on this fact. Two such tests are the Studentized Range test and Geary's test. Both of these tests use a ratio of an estimate of the sample range to the sample standard deviation. Very large and very small values of the ratio imply that the data are not well modeled by a normal distribution.

4.2.4.1 The Studentized Range Test

This test compares the sample range to the sample standard deviation. Tables of critical values for sample sizes up to 1000 (Table A-3 of Appendix A) are available for determining whether the absolute value of this ratio is significantly large. Directions for implementing this method are given in Box 4-1 along with an example. The studentized range test does not perform well if the data are asymmetric or if the tails of the data are heavier than the normal distribution. In addition, this test may be sensitive to extreme values. Many environmental data sets are positively skewed (have a long tail of high values) and are similar to a lognormal distribution.. If the data appear to be lognormally distributed, then this test should not be used. In most cases, the studentized range test performs as well as the Shapiro-Wilk test and is easier to apply.

Box 4-1: Directions for Studentized Range Test

COMPUTATIONS: Using Section 2.2.3, calculate sample range, *w*, and sample standard deviation, *s*.

STEP 1. Null Hypothesis:	H_0: The underlying distribution of the data is a normal distribution.
STEP 2. Null Hypothesis:	H_A: The underlying distribution of the data is not a normal distribution.
STEP 3. Test Statistic:	$R = w/s$
STEP 4. a) Critical Value:	Use Table A-3 to find the critical values *a* and *b*.
STEP 5. a) Conclusion:	If *R* falls outside the two critical values, then reject the null hypothesis that the underlying distribution is normal.

Directions for the studentized range test are contained in Box 4-1, with an example in Box 4-2.

Box 4-2: Example of the Studentized Range Test

Using a significance level of 0.01, determine if the underlying distribution of the data from Table 4-1 can be modeled using a normal distribution.

COMPUTATIONS: $w = X_{(10)} - X_{(1)} = 15.63 - 10.23 = 5.40$ and $s = 1.677$.

STEP 1. Null Hypothesis:	H_0: The underlying distribution of the data is a normal distribution.
STEP 2. Null Hypothesis:	H_A: The underlying distribution of the data is not a normal distribution.
STEP 3. Test Statistic:	$R = w/s = 5.4 / 1.677 = 3.22$.
STEP 4. a) Critical Value:	Using Table A-3, the critical values *a* and *b* are 2.51 and 3.88.
STEP 5. a) Conclusion:	Since R falls between the two critical values, we fail to reject the null hypothesis that the underlying distribution is normal.

4.2.4.2 Geary's Test

Geary's test compares the sum of the absolute deviations from the mean to the sum of the squares. If the ratio is too large or too small, then the underlying distribution of the data should not be modeled as a normal distribution. Directions for implementing this method are given in Box 4-3 and an example is given in Box 4-4. This test does not perform as well as the Shapiro-Wilk test or the Studentized Range test.

Directions for Geary's test are given in Box 4-3, with an example in Box 4-4.

4.2.5 Goodness-of-Fit Tests

Goodness-of-fit tests are used to test whether data follow a specific distribution, i.e., how "good" a specified distribution fits the data. In verifying assumptions of normality, one would compare the data to a normal distribution with a specified mean and variance.

4.2.5.1 Chi-Square Test

One classic goodness-of-fit test is the chi-square test which involves breaking the data into groups and comparing these groups to the expected groups from the known distribution. There are no fixed methods for selecting these groups and this test also requires a large sample size since at least 5 observations per group is needed to implement this test. In addition, the chi-square test does not have the power of the Shapiro-Wilk test or some of the other tests mentioned above. However, it is more flexible since the data can be compared to probability distributions other than the normal but the application of goodness-of-fit tests to non-normal data is beyond the scope of this guidance.

Box 4-3: Directions for Geary's Test

COMPUTATIONS: Calculate the sample mean, $\overline{X}$, the sample sum of squares, SSS, and the sum of absolute deviations, SAD:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i\,,\ SSS = \sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2,\ \text{and}\ SAD = \sum_{i=1}^{n}\left|X_i - \overline{X}\right|.$$

STEP 1. Null Hypothesis: H_0: The underlying distribution of the data is a normal distribution.

STEP 2. Null Hypothesis: H_A: The underlying distribution of the data is not a normal distribution.

STEP 3. Test Statistic: $z_0 = \dfrac{a - 0.7979}{0.2123/\sqrt{n}}$, where $a = \dfrac{SAD}{\sqrt{n \cdot SSS}}$.

STEP 4. a) Critical Value: Use Table A-1 to find $z_{1-\alpha/2}$.

STEP 4. b) p-value: Use Table A-1 to find $2 \cdot P(Z > |z_0|)$.

STEP 5. a) Conclusion: If $|z_0| > z_{1-\alpha/2}$, then reject the null hypothesis that the underlying distribution is normal.

STEP 5. b) Conclusion: If p-value < α, then reject the null hypothesis that the underlying distribution is normal.

Box 4-4: Example of Geary's Test

Using a significance level of 0.05, determine if the underlying distribution of the data from Table 4-1 can be modeled using a normal distribution.

COMPUTATIONS: $\overline{X} = 11.571$, $SAD = 11.694$, and $SSS = 25.298$.

STEP 1. Null Hypothesis: H_0: The underlying distribution of the data is a normal distribution.

STEP 2. Null Hypothesis: H_A: The underlying distribution of the data is not a normal distribution.

STEP 3. Test Statistic: $z_0 = \dfrac{0.7352 - 0.7979}{0.2123/\sqrt{10}} = -0.9339$, since $a = \dfrac{11694}{\sqrt{10 \cdot 25.298}} = 0.7352$.

STEP 4. a) Critical Value: Using Table A-1, $z_{0.975} = 1.96$.

STEP 4. b) p-value: Using Table A-1, $2 \cdot P(Z > 0.9339) = 2 \cdot (1 - 0.8238) = 0.3524$.

STEP 5. a) Conclusion: Since $|z_0| = 0.9339 < 1.96 = z_{1-\alpha}$, we fail to reject the null hypothesis that the underlying distribution is normal.

STEP 5. b) Conclusion: Since p-value = 0.3524 > 0.05 = α, we fail to reject the null hypothesis that the underlying distribution is normal.

4.2.5.2 Tests Based on the Empirical Distribution Function

Various methods have been used to measure the discrepancy between the empirical distribution function and the theoretical cumulative distribution function (cdf). These measures are referred to as empirical distribution function statistics. The best known is the Kolmogorov-Smirnov (K-S) statistic. The K-S approach is appropriate if the sample size exceeds 50 or if the theoretical cumulative density function is a specific distribution with known parameters. A modification to the test, called the Lilliefors K-S test, can be used to test that the data ($n > 50$) comes from a normal distribution with mean and variance equal to the sample values.

Unlike the K-S type statistics, most empirical distribution function statistics are based on integrated or average values between the empirical distribution function and hypothesized cumulative distribution functions. The two most powerful are the Cramer-von Mises and Anderson-Darling statistics. Extensive simulations show that the Anderson-Darling empirical distribution function statistic is as good as any test, including the Shapiro-Wilk test, when testing for normality. However, the Shapiro-Wilk test is applicable only for the normal distribution, while the Anderson-Darling method applies to other distributions.

4.2.6 Recommendations

Tests for normality with small samples have very little statistical power. Therefore, for small sample sizes it is common for a nonparametric statistical test be selected during Step 3 of the DQA in order to avoid incorrectly assuming the data are normally distributed when there is simply not enough information.

This guidance recommends using the Shapiro-Wilk *W* test, wherever practicable. The Shapiro-Wilk *W* test is one of most powerful tests for normality. This test is difficult to implement by hand but can be applied easily using a statistical software package. If the Shapiro-Wilk *W* test is not feasible, then this guidance recommends using either Filliben's statistic together with inspection of the normal probability plot, or the studentized range test. Filliben's statistic performs similarly to the Shapiro-Wilk test. The studentized range is a simple test to perform, but, it is not applicable for non-symmetric data with large tails. If the data are not highly skewed and the tails are not significantly large compared to a normal distribution, then the studentized range provides a simple and powerful test that can be calculated by hand. The Lilliefors Kolmogorov-Smirnoff test is statistically powerful but is also more difficult to apply and uses specialized tables not readily available.

4.3 TESTS FOR TRENDS

4.3.1 Introduction

This section presents statistical tools for detecting and estimating trends in environmental data. The detection and estimation of temporal or spatial trends are important for many environmental studies or monitoring programs. In cases where temporal or spatial patterns are strong, simple procedures such as time plots or linear regression over time can reveal trends. In more complex situations, sophisticated statistical models and procedures may be needed. For

example, the detection of trends may be complicated by the overlaying of long- and short-term trends, cyclical effects (e.g., seasonal or weekly systematic variations), autocorrelations, or impulses or jumps (e.g., due to interventions or procedural changes).

The graphical representations of Section 2.3.7 are recommended as the first step to identify possible trends. A Time Plot and/or Lag Plot is recommended for temporal data as it may reveal long-term trends, seasonal behavior, or impulses. A posting or bubble plot is recommended for spatial data to reveal spatial trends such as areas of high concentration or inaccessible areas.

For most of the statistical tools presented below, the focus is on monotonic long-term trends (i.e., a trend that is exclusively increasing or decreasing), as well as other sources of systematic variation, such as seasonality. The investigations of trend in this section are limited to one-dimensional domains, e.g., trends in a pollutant concentration over time. The current edition of this document does not address spatial trends or trends over space and time, which may involve sophisticated geostatistical techniques such as kriging and require the assistance of a statistician. Section 4.3.2 discusses estimating and testing for trends using regression techniques. Section 4.3.3 discusses nonparametric trend estimation procedures, and Section 4.3.4 discusses hypothesis tests for detecting trends under several types of situations.

4.3.2 Regression-Based Methods for Estimating and Testing for Trends

4.3.2.1 Estimating a Trend Using the Slope of the Regression Line

The classic procedures for assessing linear trends involve regression. Linear regression is a commonly used procedure in which calculations are performed on a data set containing pairs of observations (X_i, Y_i), so as to obtain the slope and intercept of a line that best fits the data. For temporal data, the X_i values represent time and the Y_i values represent the observations. An estimate of the magnitude of trend can be obtained by performing a regression of the data versus time and using the slope of the regression line as the measure of the strength of the trend.

Regression procedures are easy to apply. All statistical software packages and spreadsheet programs will calculate the slope and intercept of the best fitting line, as well as the correlation coefficient r (see Section 2.2.4). However, regression entails several limitations and assumptions. First of all, simple linear regression (the most commonly used method) is designed to detect linear relationships between two variables; other types of regression models are generally needed to detect non-linear relationships such as cyclical or non-monotonic trends. Regression is very sensitive to outliers and presents difficulties in handling data below the detection limit, which are commonly encountered in environmental studies. Hypothesis testing for linear regression also relies on two key assumptions: normally distributed errors, and constant variance. It may be difficult or burdensome to verify these assumptions in practice, so the accuracy of the slope estimate may be suspect. Moreover, the analyst must ensure that time plots of the data show no cyclical patterns, outlier tests show no extreme data values, and data validation reports indicate that nearly all the measurements were above detection limits. Due to these drawbacks, linear regression is not recommended as a general tool for estimating and

detecting trends, although it may be useful as an informal and quick screening tool for identifying strong linear trends.

4.3.2.2 Testing for Trends Using Regression Methods

For simple linear regression, the statistical test of whether the slope is significantly different from zero is equivalent to testing if the correlation coefficient is significantly different from zero. This test assumes a linear relation between Y and X with independent normally distributed errors and constant variance across all the X values. Non-detects and outliers may, however, invalidate the test.

Directions for this test are given in Box 4-5 along with an example.

4.3.3 General Trend Estimation Methods

4.3.3.1 Sen's Slope Estimator

Sen's Slope Estimate is a nonparametric alternative for estimating a slope. This approach involves computing slopes for all the pairs of time points and then using the median of these slopes as an estimate of the overall slope. As such, it is insensitive to outliers and can handle a moderate number of values below the detection limit and missing values. Assume that there are n time points, and let X_i denote the data value for the i^{th} time point. If there are no missing data, there will be $n(n\text{-}1)/2$ possible pairs of time points (i, j) in which $i < j$. The slope for such a pair is $b_{ij} = (X_j - X_i) / (j - i)$. Sen's slope estimator is then the median of the $n(n\text{-}1)/2$ pairwise slopes. If there is no underlying trend, there would be an approximately equal number of positive and negative slopes, and thus the median would be near zero.

4.3.3.2 Seasonal Kendall Slope Estimator

If the data exhibit cyclic trends, then Sen's slope estimator can be modified to account for the cycles. For example, if data are available for each month for a number of years, 12 separate sets of slopes would be determined; similarly, if daily observations exhibit weekly cycles, seven sets of slopes would be determined. In these estimates, the above pairwise slope is calculated for each time period and the median of all of the slopes is an estimator of the slope for a long-term trend. This is known as the seasonal Kendall slope estimator.

4.3.4 Hypothesis Tests for Detecting Trends

Most of the trend tests treated in this section involve the Mann-Kendall test or extensions of it. The Mann-Kendall test does not assume any particular distributional form and accommodates values below the detection limit by assigning them a common value. The test can also be modified to deal with multiple observations, multiple sampling locations, and seasonality.

Box 4-5: Directions for the Test for a Correlation Coefficient and an Example

COMPUTATIONS: Calculate the correlation coefficient, *r* (Section 2.2.4).

STEP 1. Null Hypothesis: H_0: The correlation coefficient is zero.

STEP 2. Alternative Hypothesis: H_A: The correlation coefficient is different from zero.

STEP 3. Test Statistic: $t_0 = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$.

STEP 4. a) Critical Value: Use Table A-2 to find $t_{n-2,1-\alpha/2}$.

STEP 4. b) p-value: Use Table A-2 to find $2 \cdot P\left(t_{n-2} > |t_0|\right)$.

STEP 5. a) Conclusion: If $|t_0| > t_{n-2,1-\alpha/2}$, then reject the null hypothesis that the correlation coefficient is zero.

STEP 5. b) Conclusion: If p-value < α, then reject the null hypothesis that the correlation coefficient is zero.

Example: Using a significance level of 0.10, determine if the following data set (in ppb) has significant correlation between it's two variables:

Sample Number	1	2	3	4
Arsenic	8.0	6.0	2.0	1.0
Lead	8.0	7.0	7.0	6.0

COMPUTATIONS: In Section 2.2.4, the correlation coefficient *r* for this data was calculated to be 0.865.

STEP 1. Null Hypothesis: H_0: The correlation coefficient is zero.

STEP 2. Alternative Hypothesis: H_A: The correlation coefficient is different from zero.

STEP 3. Test Statistic: $t_0 = \frac{0.865}{\sqrt{\frac{1-0.865^2}{4-2}}} = 2.438$.

STEP 4. a) Critical Value: Using Table A-2, $t_{2,0.975} = 4.303$.

STEP 4. b) p-value: Using Table A-2, 0.10< p-value < 0.20 (the exact p-value = 0.1350)

STEP 5. a) Conclusion: Since $|t_0| < t_{2,0.975}$, we fail to reject the null hypothesis that the correlation coefficient is zero.

STEP 5. b) Conclusion: Since p-value = 0.1350 > 0.10 (α) we fail to reject the null hypothesis that the correlation coefficient is zero.

4.3.4.1 One Observation per Time Period for One Sampling Location

The Mann-Kendall test involves computing a statistic S, which is the difference between the number of pairwise differences that are positive minus the number that are negative. If S is a large positive value, then there is evidence of an increasing trend in the data. If S is a large negative value, then there is evidence of a decreasing trend in the data. The null hypothesis or baseline condition for this test is that there is no temporal trend in the data values. The alternative hypothesis is that of either an upward trend or a downward trend.

Box 4-6: Upper Triangular Computations for Basic Mann-Kendall Trend Test with a Single Measurement at Each Time Point

Original Time	t_1	t_2	t_3	t_4	. . .	t_{n-1}	t_n		
Measurement	X_1	X_2	X_3	X_4	. . .	X_{n-1}	X_n		
X_1		Y_{12}	Y_{13}	Y_{32}	. . .	$Y_{1,n-1}$	Y_{1n}	#(plusses)	#(minuses)
X_2			Y_{23}	Y_{32}	. . .	$Y_{2,n-1}$	Y_{2n}	#(plusses)	#(minuses)
.						.	.	.	.
.						.	.	.	.
.						.	.	.	.
X_{n-2}						$Y_{n-2,n-1}$	$Y_{n-2,n}$	#(plusses)	#(minuses)
X_{n-1}							$Y_{n-1,n}$	#(plusses)	#(minuses)
								Total #(plusses)	Total #(minuses)

where Y_{ij} = sign $(X_j - X_i)$ = + if $X_j > X_i$
= 0 if $X_j = X_i$
= - if $X_j < X_i$

Box 4-7: Directions for the Mann-Kendall Trend Test for Small Sample Sizes

COMPUTATIONS: Create the upper triangular table of pairwise differences as described in Box 4-6.

STEP 1. Null Hypothesis:	H_0: There is no trend.
STEP 2. Null Hypothesis:	i) H_A: There is a downward trend. ii) H_A: There is an upward trend.
STEP 3. Test Statistic:	S = Total #(plusses) – Total #(minuses)
STEP 4. a) Critical Value:	Use Table A-12a to find the critical value.
STEP 4. b) p-value:	Use Table A-12b to find the p-value.
STEP 5. a) Conclusion:	If $\|S\| \geq$ the critical value, then reject the null hypothesis of no trend.
STEP 5. b) Conclusion:	If p-value < α, then reject the null hypothesis of no trend.

The computations for the basic Mann-Kendall trend test are depicted in Box 4-6. Assign a value of DL/2 to all non-detects. The test statistic is the difference between the number of strictly positive differences and the number of strictly negative differences. Differences of zero are not included in the test statistic (and should be avoided, if possible, by recording data to sufficient accuracy). The steps for conducting the Mann-Kendall test for small sample sizes ($n < 10$) are contained in Box 4-7 and an example is contained in Box 4-8. For sample sizes greater than 10, there is a normal approximation for the Mann-Kendall test. Directions for this approximation are contained in Box 4-9 with an example given in Box 4-10.

Box 4-8: An Example of Mann-Kendall Trend Test for Small Sample Sizes

Consider 5 measurements ordered by the time of their collection: 5, 6, 11, 8, and 10. This data will be used to test for an upward trend at a significance level of 0.05.

COMPUTATIONS: A triangular table was used to construct the pairwise differences.

Time	1	2	3	4	5	No. of	No. of -
Data	5	6	11	8	10	+ Signs	Signs
5		+	+	+	+	4	0
6			+	+	+	3	0
11				-	-	0	2
8					+	1	0
						8	2

STEP 1. Null Hypothesis: H_0: There is no trend.

STEP 2. Null Hypothesis: H_A: There is an upward trend.

STEP 3. Test Statistic: S = Total #(plusses) – Total #(minuses) = 8 - 2 = 6

STEP 4. a) Critical Value: Using Table A-12a, the critical value is 8.

STEP 4. b) p-value: Using Table A-12b, the p-value is 0.117.

STEP 5. a) Conclusion: As $S = 6 < 8$ (critical value), fail to reject the null hypothesis of no trend

STEP 5. b) Conclusion: Since p-value = 0.117 > 0.05(significance level), we fail to reject the null hypothesis of no trend.

Box 4-9: Directions for the Mann-Kendall Test Using a Normal Approximation

COMPUTATIONS: If the sample size is 10 or more, a normal approximation to the Mann-Kendall procedure may be used. Compute S as described in Box 4-7 above.

STEP 1. Null Hypothesis: H_0: There is no trend.

STEP 2. Null Hypothesis: i) H_A: There is a downward trend.
ii) H_A: There is an upward trend.

STEP 3. Test Statistic: $z_0 = \frac{S - \text{sign}(S)}{\sqrt{V(S)}}$, where $V(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{j=1}^{g} t_j(t_j-1)(2t_j+5) \right]$,

g is the number of tied groups, and t_j is the number of points in the j^{th} group. Note that sign(S) = 1 if $S > 0$, 0 if $S = 0$ and –1 if $S < 0$.

STEP 4. a) Critical Value: Use Table A-1 to find $z_{1-\alpha}$.

STEP 4. b) p-value: Use Table A-1 to find $P(Z > |z_0|)$.

STEP 5. a) Conclusion: If $|z_0| > z_{1-\alpha}$, then reject the null hypothesis of no trend.

STEP 5. b) Conclusion: If p-value < α, then reject the null hypothesis of no trend.

Box 4-10: An Example of Mann-Kendall Trend Test by Normal Approximation

A test for an upward trend with α = 0.05 will be based on the 11 weekly measurements is shown below.

COMPUTATIONS: Using Box 4-6, a triangular table was constructed of the pairwise slopes. "0" indicates a tie.

Data	1	2	3	4	5	6	7	8	9	10	11	No. of	No. of
10	10	10	10	5	10	20	18	17	15	24	15	+ Signs	- Signs
10		0	0	-	0	+	+	+	+	+	+	6	1
10			0	-	0	+	+	+	+	+	+	6	1
5				-	0	+	+	+	+	+	+	6	1
10					+	+	+	+	+	+	+	7	0
20						+	+	+	+	+	+	6	0
18							-	-	-	+	-	1	4
17								-	-	+	-	1	3
15									-	+	-	1	2
24										+	0	1	0
											-	0	1
												35	13

S = (sum of + signs) - (sum of - signs) = 35 - 13 = 22.

STEP 1. Null Hypothesis: H_0: There is no trend.

STEP 2. Null Hypothesis: H_A: There is an upward trend.

STEP 3. Test Statistic: There are several observations tied at 10 and 15. Thus, the formula for tied values will be used. In this formula, g = 2, t_1 = 4 for tied values of 10, and t_2 = 2 for tied values of 15. Therefore,

$$V(S) = \frac{1}{18}\{11(11-1)(2\cdot 11+5) - [4(4-1)(2\cdot 4+5) + 2(2-1)(2\cdot 2+5)]\} = 155.33$$

$$\text{and } z_0 = \frac{S - \text{sign}(S)}{\sqrt{V(S)}} = \frac{22-1}{\sqrt{155.33}} = 1.685 .$$

STEP 4. a) Critical Value: Using Table A-1, $z_{0.95}$ = 1.645.

STEP 4. b) p-value: Using Table A-1, p-value = $P(Z > 1.685)$ = 0.046.

STEP 5. a) Conclusion: Since test statistic = 1.685 > 1.645 = critical value, we reject the null hypothesis of no trend.

STEP 5. b) Conclusion: Since p-value = 0.046 < 0.05 = α, we reject the null hypothesis of no trend.

4.3.4.2 Multiple Observations per Time Period for One Sampling Location

Often, more than one sample is collected for each time period. There are two ways to deal with multiple observations per time period. One method is to compute a summary statistic, such as the median, for each time period and to apply one of the Mann-Kendall trend tests of Section 4.3.4.1 to the summary statistic. Therefore, instead of using the individual data points in the triangular table, the summary statistic would be used. Then the steps given in Box 4-7 or Box 4-9 could be applied to the summary statistics.

An alternative approach is to consider all the multiple observations within a given time period as being essentially taken at the same time within that period. The *S* statistic is computed as before with *n* being the total of all observations. The variance of the *S* statistic (calculated in step 2 of Box 4-9) is changed to:

$$\mathrm{VAR}(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{j=1}^{g} t_j(t_j-1)(2t_j+5) - \sum_{k=1}^{h} u_k(u_k-1)(2u_k+5) \right]$$

$$+ \frac{\sum_{j=1}^{g} t_j(t_j-1)(t_j-2) \cdot \sum_{k=1}^{h} u_k(u_k-1)(u_k-2)}{9n(n-1)(n-2)} + \frac{\sum_{j=1}^{g} t_j(t_j-1) \cdot \sum_{k=1}^{h} u_k(u_k-1)}{2n(n-1)}$$

where *g* represents the number of tied groups, t_j represents the number of data points in the j^{th} group, *h* is the number of time periods which contain multiple data, and u_k is the sample size in the k^{th} time period.

The preceding variance formula assumes that the data are not correlated. If correlation within single time periods is suspected, it is preferable to use a summary statistic (e.g., the median) for each time period and then apply either Box 4-7 or Box 4-9 to the summary statistics.

4.3.4.3 Multiple Sampling Locations with Multiple Observations

The preceding methods involve a single sampling location (station). However, environmental data often consist of sets of data collected at several sampling locations (see Box 4-11). For example, data are often systematically collected at several fixed sites on a lake or river, or within a region or basin. The data collection plan (or experimental design) must be systematic in the sense that approximately the same sampling times should be used at all locations. In this situation, it is desirable to express the results by an overall regional summary statement across all sampling locations. However, there must be consistency in behavioral characteristics across sites over time in order for a single summary statement to be valid across all sampling locations. A useful plot to assess the consistency requirement is a single time plot (Section 2.3.7.1) of the measurements from all stations where a different symbol is used to represent each station.

If the stations exhibit approximate trends in the same direction with comparable slopes, then a single summary statement across stations is valid and this implies two relevant sets of hypotheses should be investigated:

Comparability of stations. H_0: Similar dynamics affect all *K* stations vs. H_A: At least two stations exhibit different dynamics.

Testing for overall monotonic trend. H_0*: Contaminant levels do not change over time vs. H_A*: There is an increasing (or decreasing) trend consistent across all stations.

Therefore, the analyst must first test for homogeneity of stations, and then, if homogeneity is confirmed, test for an overall monotonic trend. Directions for the test are contained in Box 4-11 and ideally, the stations in Box 4-11 should have equal sample sizes. However, the numbers of observations at the stations can differ slightly, because of isolated missing values, but the overall time periods spanned must be similar. This guidance recommends that for less than 3 time periods, an equal number of observations (a balanced design) are required. For 4 or more time periods, up to 1 missing value per sampling location may be tolerated.

Box 4-11: Data for Multiple Times and Multiple Stations

Let i = 1, 2, ..., n represent time, k = 1, 2, ..., K represent sampling locations, and X_{ik} represent the measurement at time i for location k. This data can be summarized in matrix form, as shown below.

		Stations			
		1	2	. . .	K
Time	1	X_{11}	X_{12}	. . .	X_{1K}
	2	X_{21}	X_{22}	. . .	X_{2K}
	.	.	.	. . .	.
	.	.	.	. . .	.
	.	.	.	. . .	.
	n	X_{n1}	X_{n2}	. . .	X_{nK}
		S_1	S_2	. .	S_K
		$V(S_1)$	$V(S_2)$	.	$V(S_K)$
		Z_1	Z_2	. .	Z_K
				.	
				. .	
				.	

where S_k = Mann-Kendall statistic for station k (see STEP 3, Box 4-7),

$V(S_k)$ = variance for S statistic for station k (see STEP 3, Box 4-9), and

$$z_k = \frac{S_k - \text{sign}(S_k)}{\sqrt{V(S_k)}}$$

a. One Observation per Time Period. When only one measurement is taken for each time period for each station, a generalization of the Mann-Kendall statistic can be used to test the above hypotheses. This procedure is described in Box 4-12.

b. Multiple Observations per Time Period. If multiple measurements are taken at some times and stations, then the previous approaches are still applicable. However, the variance of the statistic S_k must be calculated using the equation for calculating V(S) given in Section 4.3.4.2. Note that S_k is computed for each station, so n, t_j, g, h, and u_k are all station-specific.

Box 4-12: Testing for Comparability of Stations and an Overall Monotonic Trend

Let i = 1, 2, ..., n represent time, k = 1, 2, ..., K represent sampling locations, and X_{ik} represent the measurement at time i for location k. Let α represent the significance level for testing homogeneity and α^* represent the significance level for testing for an overall trend.

COMPUTATIONS: Calculate the Mann-Kendall statistic S_k and its variance $V(S_k)$ for each of the K stations using the methods of Box 4-9. Now, calculate

$$Z_k = \frac{S_k - \text{sign}(S_k)}{\sqrt{V(S_k)}}, \text{ for } k = 1,\ldots,K \text{ and } \overline{Z} = \frac{1}{K}\sum_{k=1}^{K} Z_k .$$

Test of Homogeneity

STEP 1. Null Hypothesis: H_0: Similar dynamics affect all K stations.

STEP 2. Null Hypothesis: H_A: At least two stations exhibit different dynamics.

STEP 3. Test Statistic: $\chi_h^2 = \sum_{k=1}^{K} Z_k^2 - K\overline{Z}^2$.

STEP 4. a) Critical Value: Use Table A-9 to find $\chi^2_{K-1,1-\alpha}$.

STEP 4. b) p-value: Use Table A-9 to find $P\left(\chi^2_{K-1} > \chi_h^2\right)$.

STEP 5. a) Conclusion: If $\chi_h^2 > \chi^2_{K-1,1-\alpha}$, then reject the null hypothesis that similar dynamics affect all K stations.

STEP 5. b) Conclusion: If p-value < α, then reject the null hypothesis that similar dynamics affect all K stations.

If H_0 is not rejected, then proceed to test of overall trend. Otherwise, individual α^*-level Mann-Kendall tests should be conducted using the methods presented in Section 4.3.4.1.

Test of Overall Trend

STEP 1. Null Hypothesis: H_0^*: Contaminant levels do not change over time.

STEP 2. Null Hypothesis: H_A^*: There is an increasing (or decreasing) trend consistently exhibited across all stations.

STEP 3. Test Statistic: $\chi_o^2 = K \cdot \overline{Z}^2$.

STEP 4. a) Critical Value: Use Table A-9 to find $\chi^2_{1,1-\alpha^*}$.

STEP 4. b) p-value: Use Table A-9 to find $P\left(\chi_1^2 > \chi_o^2\right)$.

STEP 5. a) Conclusion: If $\chi_h^2 > \chi^2_{1,1-\alpha^*}$, then reject the null hypothesis that contaminant levels do not change over time.

STEP 5. b) Conclusion: If p-value < α, then reject the null hypothesis .

4.3.4.4 One Observation for One Station with Multiple Seasons

Temporal data are often collected over extended periods of time. Within the time variable, data may exhibit periodic cycles, which are patterns in the data that repeat over time. For example, temperature and humidity may change with the season or month, and may affect environmental measurements. (For more information on seasonal cycles, see Section 2.3.7). In the following discussion, the term season represents one time point in the periodic cycle, such as a month within a year or an hour within a day.

If seasonal cycles are anticipated, then two approaches for testing for trends are the seasonal Kendall test and Sen's test for trends. The seasonal Kendall test may be used for large sample sizes, and Sen's test for trends may be used for small sample sizes. If different seasons manifest similar slopes (rates of change) but possibly different intercepts, then the Mann-Kendall technique of Section 4.3.4.3 is applicable, replacing time by year and replacing station by season.

The seasonal Kendall test, which is an extension of the Mann-Kendall test, involves calculating the Mann-Kendall test statistic, *S*, and its variance separately for each season. The sum of the *S*'s and the sum of their variances are then used to form an overall test statistic that is assumed to be approximately normally distributed for larger size samples.

For data at a single site, collected at multiple seasons within multiple years, the techniques of Section 4.3.4.3 can be applied to test for homogeneity of time trends across seasons. The methodology follows Boxes 4-11 and 4-12 exactly except that 'station' is replaced by 'season' and the inferences refer to seasons.

4.3.5 A Discussion on Tests for Trends

This section discusses some further considerations for choosing among the many tests for trend. Mann-Kendall type nonparametric trend tests and estimates use ordinal time (ranks) rather than cardinal time (actual time values) and this restricts the interpretation of measured trends. All of the Mann-Kendall Trend Tests presented are based on certain pairwise differences in measurements at different time points. The only information about these differences that is used in the Mann-Kendall calculations is the sign, and can be regarded as generalizations of the signed rank test. However, since information about the magnitudes of the differences is not used, this can adversely affect the statistical power when only limited amounts of data are available.

There are nonparametric methods based on ranks that take such magnitudes into account and still retain the benefit of robustness to outliers. These procedures can be thought of as replacing the data with their ranks and then conducting parametric analyses. These include the Wilcoxon Rank Sum test and its many generalizations. These methods are more resistant to outliers than parametric methods. Rank-based methods which make fuller use of the information in the data than the Mann-Kendall methods are not as robust with respect to outliers as the signed rank test and the Mann-Kendall tests, but they have more statistical power. This kind of tradeoff

between power and robustness shows the need for an evaluation process leading to the selection of the best statistical procedure for the situation.

4.3.6 Testing for Trends in Sequences of Data

There are cases where it is desirable to see if a sequence of data (for example, readings from a monitoring station) could be considered random variation or correlated in some way. One test to make this determination is the Wald-Wolfowitz test. This test can only be used if the data are binary, i.e., there are only two potential values. For example, the data could be either 'Yes/No' or an investigation of persistent violation of a permitted limit by a pollution control process where a violation equals 1, with 0 for not in violation. Directions for the Wald-Wolfowitz test are given in Box 4-13 and an example in Box 4-14.

Box 4-13: Directions for the Wald-Wolfowitz Runs Test

Consider a sequence of binary values. Let m and n denote the number of observations for the two values with $n<m$. This test is used to test the null hypothesis that the sequence is random against the alternative hypothesis that the values in the sequence are correlated or may come from different populations.

COMPUTATIONS: List the data in the order collected and compute T, the number of runs in the sequence. For example, if the data sequence is AAABAABBBBBBBABB, then $T = 6$.

STEP 1. Null Hypothesis:	H_0: data sequence is random.
STEP 2. Alternative Hypothesis:	H_A: data sequence is non-random.
STEP 3. Test Statistic:	If either m or n is less than 10, then T. If both m and n are at least 10, then $z_0 = \frac{T - \mu_r}{\sigma_r}$, where $\mu_r = \frac{2mn}{m+n} + 1$ and $\sigma_r = \sqrt{\frac{2mn(2mn-m-n)}{(m+n)^2(m+n-1)}}$
STEP 4. a) Critical Value:	If either m or n is less than 10, then use Table A-13 to find $w_{\alpha/2}$. If both m and n are both at least 10, then use Table A-1 to find $z_{1-\alpha/2}$.
STEP 4. a) p-value:	If both m and n are both at least 10, then use Table A-1 to find $P(Z > \lvert z_0 \rvert)$.
STEP 5. a) Conclusion:	If either m or n is less than 10 and if $T < w_{\alpha/2}$ or $T > 2\ \mu_r - w_{\alpha/2}$, then reject the null hypothesis that the data sequence is random. If both m and n are at least 10 and if $\lvert z_0 \rvert > z_{1-\alpha/2}$, then reject the null hypothesis that the data sequence is random.
STEP 5. b) Conclusion:	If p-value < α, the reject the null hypothesis that the data sequence is random.

Box 4-14: An Example of the Wald-Wolfowitz Runs Test

The main discharge station at a chemical manufacturing plant is under a monitoring program. The permit states that the discharge should have a pH of 7.0 and should never be less than 5.0. So the plant manager has decided to use a pH of 6.0 to an indicate potential problems. In a four-week period the following values were recorded:

6.5	6.6	6.4	6.2	5.9	5.8	5.9	6.2	6.2	6.3	6.6	6.6	6.7	6.4
6.2	6.3	6.2	5.8	5.9	5.8	6.1	5.9	6.0	6.2	6.3	6.2		

Since the plant manager has decided that a pH of 6.0 will indicate trouble the data have been replaced with a binary indicator. If the value is greater than 6.0, the value will be replaced by a 1; otherwise the value will be replaced by a 0. So the binary sequence is:

1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 0 0 0 1 0 0 1 1 1

As there are 8 values of '0'and 19 values of '1', n = 8 and m = 19 and the number of runs is T = 7. Test at a significance level of 0.10 whether the data sequence is random.

STEP 1. Null Hypothesis: H_0: data sequence is random.

STEP 2. Alternative Hypothesis: H_A: data sequence is non-random.

STEP 3. Test Statistic: Since n is less than 10, the test statistic is T = 7.

STEP 4. a) Critical Value: Since n is less than 10, Table A-13 is used to find $w_{0.05}$ = 9.

STEP 5. a) Conclusion: Since n is less than 10 and T =7 < 9 = $w_{0.05}$, we reject H_0.

4.4 OUTLIERS

4.4.1 Background

Potential outliers are measurements that are extremely large or small relative to the rest of the data and, therefore, are suspected of misrepresenting the population from which they were collected. Potential outliers may result from transcription errors, data-coding errors, or measurement system problems. However, outliers may also represent true extreme values of a distribution (for instance, hot spots) and indicate more variability in the population than was expected. Failure to remove true outliers or the removal of false outliers both lead to a distortion of estimates of population parameters and if it is recommended that the QA Project Plan or Sampling and Analysis Plan be reviewed for anomalies that could account for the potential outlier.

Statistical outlier tests give the analyst probabilistic evidence that an extreme value does not "fit" with the distribution of the remainder of the data and is therefore a statistical outlier. These tests should only be used to *identify* data points that require further investigation. The tests alone cannot determine whether a statistical outlier should be discarded or corrected within a data set. This decision should be based on judgmental or scientific grounds.

There are 5 steps involved in treating extreme values or outliers:

1. Identify extreme values that may be potential outliers;
2. Apply statistical test;
3. Scientifically review statistical outliers and decide on their disposition;
4. Conduct data analyses with and without statistical outliers; and
5. Document the entire process.

Potential outliers may be identified through the graphical representations of Chapter 2 (step 1 above). Graphs such as the box and whisker plot, ranked data plot, normal probability plot, and time plot can all be used to identify observations that are much larger or smaller than the rest of the data. If potential outliers are identified, the next step is to apply one of the statistical tests described in the following sections. Section 4.4.2 provides recommendations on selecting a statistical test for outliers.

If a data point is found to be an outlier, the analyst may either: 1) correct the data point; 2) discard the data point from analysis; or 3) use the data point in all analyses. This decision should be based on scientific reasoning *in addition to* the results of the statistical test. For instance, data points containing transcription errors should be corrected, whereas data points collected while an instrument was malfunctioning may be discarded. Discarding an outlier from a data set should be done with extreme caution, particularly for environmental data sets, which often contain legitimate extreme values. If an outlier is discarded from the data set, all statistical analysis of the data should be applied to both the full and truncated data set so that the effect of discarding observations may be assessed. If scientific reasoning does not explain the outlier, it should not be discarded from the data set.

If any data points are found to be statistical outliers through the use of a statistical test, this information will need to be documented along with the analysis of the data set, regardless of whether any data points are discarded. If no data points are discarded, document the identification of any statistical outliers by documenting the statistical test performed and the possible scientific reasons investigated. If any data points are discarded, document each data point, the statistical test performed, the scientific reason for discarding each data point, and the effect on the analysis of deleting the data points. This information is critical for effective peer review.

4.4.2 Selection of a Statistical Test for Outliers

There are several statistical tests for determining whether or not one or more observations are statistical outliers. Step by step directions for implementing some of these tests are described in Sections 4.4.3 through 4.4.6. Section 4.4.7 describes statistical tests for multivariate outliers.

If the data are approximately normally distributed, this guidance recommends Rosner's test when the sample size is greater than 25 and the Extreme Value test when the sample size is less than 25. If only one outlier is suspected, then the Discordance test may be substituted for either of these tests. If the data are not normally distributed, or if the data cannot be transformed

so that the transformed data are normally distributed, then the analyst should either apply a nonparametric test (such as Walsh's test) or consult a statistician. A summary of these recommendations is contained in Table 4-3.

Table 4-3. Recommendations for Selecting a Statistical Test for Outliers

Sample Size	Test	Section	Assumes Normality	Multiple Outliers
$n \leq 25$	Extreme Value Test	4.4.3	Yes	No/Yes
$n \leq 50$	Discordance Test	4.4.4	Yes	No
$n \geq 25$	Rosner's Test	4.4.5	Yes	Yes
$n \geq 50$	Walsh's Test	4.4.6	No	Yes

4.4.3 Extreme Value Test (Dixon's Test)

Dixon's Extreme Value test can be used to test for statistical outliers when the sample size is less than or equal to 25. This test considers both extreme values that are much smaller than the rest of the data (case 1) and extreme values that are much larger than the rest of the data (case 2). This test assumes that the data without the suspected outlier are normally distributed; therefore, it is necessary to perform a test for normality on the data without the suspected outlier before applying this test. If the data are not normally distributed, then either transform the data, apply a different test, or consult a statistician. Directions for the Extreme Value test are contained in Box 4-15; an example of this test is contained in Box 4-16.

This guidance recommends using this test when only one outlier is suspected in the data. If more than one outlier is suspected, the Extreme Value test may lead to masking where two or more outliers close in value "hide" one another. Therefore, if the analyst decides to use the Extreme Value test for multiple outliers, apply the test to the least extreme value first.

4.4.4 Discordance Test

The Discordance test can be used to test if one extreme value is an outlier. The Discordance test assumes that the data without the suspected outlier are approximately normally distributed. Therefore, it is necessary to check for normality before applying this test. Directions and an example of the Discordance test are contained in Box 4-17 and Box 4-18.

Box 4-15: Directions for the Extreme Value Test (Dixon's Test)

Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ represent the data ordered from smallest to largest. Check that the data without the suspect outlier are normally distributed, using one of the methods of Section 4.2. If normality fails, either transform the data or apply a different outlier test.

STEP 1. Null Hypothesis: H_0: There are no outliers in the data.

STEP 2. Alternative Hypothesis: i) H_A: $X_{(1)}$ is an outlier.
ii) H_A: $X_{(n)}$ is an outlier.

STEP 3. Test Statistic: i) Compute the test statistic C, where

$3 \le n \le 7$	$8 \le n \le 10$	$11 \le n \le 13$	$14 \le n \le 25$
$C = \frac{X_{(2)} - X_{(1)}}{X_{(n)} - X_{(1)}}$	$C = \frac{X_{(2)} - X_{(1)}}{X_{(n-1)} - X_{(1)}}$	$C = \frac{X_{(3)} - X_{(1)}}{X_{(n-1)} - X_{(1)}}$	$C = \frac{X_{(3)} - X_{(1)}}{X_{(n-2)} - X_{(1)}}$

ii) Compute the test statistic C, where

$3 \le n \le 7$	$8 \le n \le 10$	$11 \le n \le 13$	$14 \le n \le 25$
$C = \frac{X_{(n)} - X_{(n-1)}}{X_{(n)} - X_{(1)}}$	$C = \frac{X_{(n)} - X_{(n-1)}}{X_{(n)} - X_{(2)}}$	$C = \frac{X_{(n)} - X_{(n-2)}}{X_{(n)} - X_{(2)}}$	$C = \frac{X_{(n)} - X_{(n-2)}}{X_{(n)} - X_{(3)}}$

STEP 4. a) Critical Value: Use Table A-4 to find d_α.

STEP 5. a) Conclusion: If $C > d_\alpha$, then reject the null hypothesis that there are no outliers in the data.

Box 4-16: An Example of the Extreme Value Test (Dixon's Test)

The data in order of magnitude from smallest to largest are (in ppm):

82.39, 86.62, 91.72, 98.37, 103.46, 104.93, 105.52, 108.21, 113.23, 150.55.

As the value 150.55 is much larger than the other values, it is suspected that this data point might be an outlier. The Studentized Range test (Section 4.2.6) shows that there is no reason to suspect that the data are not normally distributed.

STEP 1. Null Hypothesis: H_0: There are no outliers in the data.

STEP 2. Alternative Hypothesis: H_A: $X_{(n)}$ is an outlier.

STEP 3. Test Statistic: Since n = 10, $C = \frac{X_{(n)} - X_{(n-1)}}{X_{(n)} - X_{(2)}} = \frac{150.55 - 113.23}{150.55 - 86.62} = 0.584$.

STEP 4. a) Critical Value: Using Table A-4, $d_{0.05} = 0.477$.

STEP 5. a) Conclusion: Since $C = 0.584 > 0.477 = d_{0.05}$, we reject the null hypothesis of no outliers in the data.

Box 4-17: Directions for the Discordance Test

Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ represent the data ordered from smallest to largest. Check that the data without the suspect outlier are normally distributed, using one of the methods of Section 4.2. If normality fails, either transform the data or apply a different outlier test.

COMPUTATIONS: Compute the sample mean, $\overline{X}$, and the sample standard deviation, s.

STEP 1. Null Hypothesis: H_0: There are no outliers in the data.

STEP 2. Alternative Hypothesis: i) H_A: $X_{(1)}$ is an outlier.
ii) H_A: $X_{(n)}$ is an outlier.

STEP 3. Test Statistic: i) $D = \dfrac{\overline{X} - X_{(1)}}{s}$ ii) $D = \dfrac{X_{(n)} - \overline{X}}{s}$

STEP 4. a) Critical Value: Use Table A-5 to find d_α.

STEP 5. a) Conclusion: If $D > d_\alpha$, then reject the null hypothesis that there are no outliers in the data.

Box 4-18: An Example of the Discordance Test

The data in order of magnitude from smallest to largest are (in ppm):

82.39, 86.62, 91.72, 98.37, 103.46, 104.93, 105.52, 108.21, 113.23, 150.55.

It is suspected that the data point 150.55 might be an outlier. The Studentized Range test (Section 4.2.6) shows that there is no reason to suspect that the data are not normally distributed.

COMPUTATIONS: $\overline{X}$ = 104.5 ppm and s = 18.922 ppm.

STEP 1. Null Hypothesis: H_0: There are no outliers in the data.

STEP 2. Alternative Hypothesis: H_A: $X_{(n)}$ is an outlier.

STEP 3. Test Statistic: $D = \dfrac{X_{(n)} - \overline{X}}{s} = \dfrac{150.55 - 104.5}{18.922} = 2.43$

STEP 4. a) Critical Value: Using Table A-5, $d_{0.05} = 2.176$.

STEP 5. a) Conclusion: Since $D = 2.43 > 2.176 = d_\alpha$, we reject the null hypothesis of no outliers.

4.4.5 Rosner's Test

A parametric test developed by Rosner can be used to detect up to 10 outliers for sample sizes of 25 or more. This test assumes that the data without the suspected outliers are normally distributed. Therefore, it is necessary to perform a test for normality before applying this test. If the data are not normally distributed, then either transform the data, apply a different test, or consult a statistician. Directions for Rosner's test are contained in Box 4-19 and an example is contained in Box 4-20.

Box 4-19: Directions for Rosner's Test for Outliers

Let $X_1, X_2, \ldots, X_n$ represent the ordered data points. By inspection, identify the maximum number of possible outliers, $1 \le r_0 \le 10$. Check that the data are normally distributed without the suspected outlier(s) using one of the methods of Section 4.2.

COMPUTATIONS: Using the entire data set, compute the sample mean, $\overline{X}^{(0)}$, and the sample standard deviation, $s^{(0)}$. Determine the observation farthest from $\overline{X}^{(0)}$ and label it $y^{(0)}$. Delete $y^{(0)}$ from the data and compute the sample mean, $\overline{X}^{(1)}$, and the sample standard deviation, $s^{(1)}$. Now determine the observation farthest from $\overline{X}^{(1)}$ and label it $y^{(1)}$. Delete $y^{(1)}$ from the data and compute the sample mean, $\overline{X}^{(2)}$, and the sample standard deviation, $s^{(2)}$. Continue this process until r_0 extreme values have been eliminated. After this process the analyst should have:

$$\left\{\overline{X}^{(0)}, s^{(0)}, y^{(0)}\right\}, \left\{\overline{X}^{(1)}, s^{(1)}, y^{(1)}\right\}, \ldots, \left\{\overline{X}^{(r_0-1)}, s^{(r_0-1)}, y^{(r_0-1)}\right\}$$

STEP 1. Null Hypothesis: H_0: There are no outliers in the data.

STEP 2. Alternative Hypothesis: H_A: There are as many as r_0 outliers in the data.

Steps 3 through 5 of this test are iterative. First, test if there are r_0 outliers. If not, then test if there are r_0 - 1 outliers. Continue, until it is determined that either there are a certain number of outliers or that there are no outliers.

STEP 3. Test Statistic: $R_r = \dfrac{\left|y^{(r-1)} - \overline{X}^{(r-1)}\right|}{s^{(r-1)}}$, where r starts at r_0 and runs through 1.

STEP 4. a) Critical Value: Use Table A-6 to find λ_r.

STEP 5. a) Conclusion: If $R_r > \lambda_r$, then conclude that there are r outliers. Otherwise, return to step 3 to test for r-1 outliers.

4.4.6 Walsh's Test

A nonparametric test was developed by Walsh to detect multiple outliers in a data set. This test requires a large sample size: $n > 220$ for a significance level of $\alpha = 0.05$, and $n > 60$ for a significance level of $\alpha = 0.10$. However, since the test is a nonparametric test, it may be used when the data is not normally distributed. It should be noted that this test is used infrequently with environmental data. Directions for the test for large sample sizes are given in Box 4-21.

4.4.7 Multivariate Outliers

Multivariate analysis, such as factor analysis and principal components analysis, involves the analysis of several variables simultaneously. Outliers in multivariate analysis are then values that are extreme in relationship to either one or more variables. As the number of variables increases, identifying potential outliers using graphical representations becomes more difficult. In addition, special procedures are required to test for multivariate outliers and details of these procedures are beyond the scope of this guidance.

Box 4-20: An Example of Rosner's Test for Outliers

Consider the following 32 ordered data points (in ppm): 2.07, 40.55, 84.15, 88.41, 98.84, 100.54, 115.37, 121.19, 122.08, 125.84, 129.47, 131.90, 149.06, 163.89, 166.77, 171.91, 178.23, 181.64, 185.47, 187.64, 193.73, 199.74, 209.43, 213.29, 223.14, 225.12, 232.72, 233.21, 239.97, 251.12, 275.36, 395.67.

A normal probability plot excluding the suspected outliers shows that there is no reason to suspect that the data is not normally distributed. In addition, this graph identified four potential outliers: 2.07, 40.55, 275.36, and 395.67. Rosner's test at a significance level of 0.05 will be applied to see if there are $r_0 = 4$ or fewer outliers.

COMPUTATIONS: The summary statistics and suspected outliers are listed in the following table:

i	$\overline{X}^{(i)}$	$s^{(i)}$	$y^{(i)}$
0	169.923	75.133	395.67
1	162.640	63.872	2.07
2	167.993	57.460	40.55
3	172.387	53.099	275.36

STEP 1. Null Hypothesis: H_0: There are no outliers in the data.

STEP 2. Alternative Hypothesis: H_A: There are as many as 4 outliers in the data.

STEP 3. Test Statistic: $$R_4 = \frac{\left|y^{(3)} - \overline{X}^{(3)}\right|}{s^{(3)}} = \frac{|275.36 - 172.387|}{53.099} = 1.939$$

STEP 4. a) Critical Value: Using Table A-6, $\lambda_4 = 2.89$.

STEP 5. a) Conclusion: Since $R_4 = 1.939 < 2.89 = \lambda_4$, there aren't 4 outliers. Therefore, test if there are 3 outliers by computing

STEP 3. Test Statistic: $$R_3 = \frac{\left|y^{(2)} - \overline{X}^{(2)}\right|}{s^{(2)}} = \frac{|40.55 - 167.993|}{57.46} = 2.218$$

STEP 4. a) Critical Value: Using Table A-6, $\lambda_3 = 2.91$.

STEP 5. a) Conclusion: Since $R_3 = 2.218 < 2.91 = \lambda_3$, there aren't 3 outliers. Therefore, test if there are 2 outliers by computing

STEP 3. Test Statistic: $$R_2 = \frac{\left|y^{(1)} - \overline{X}^{(1)}\right|}{s^{(1)}} = \frac{|2.07 - 162.64|}{63.872} = 2.514$$

STEP 4. a) Critical Value: Using Table A-6, $\lambda_2 = 2.92$.

STEP 5. a) Conclusion: Since $R_2 = 2.514 < 2.92 = \lambda_2$, there aren't 2 outliers. Therefore, test if there are 1 outlier by computing

STEP 3. Test Statistic: $$R_1 = \frac{\left|y^{(0)} - \overline{X}^{(0)}\right|}{s^{(0)}} = \frac{|395.67 - 169.923|}{75.133} = 3.005$$

STEP 4. a) Critical Value: Using Table A-6, $\lambda_1 = 2.94$.

STEP 5. a) Conclusion: Since $R_1 = 3.005 > 2.94 = \lambda_1$, we conclude there is one outlier, 395.67.

Box 4-21: Directions for Walsh's Test for Large Sample Sizes

Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ represent the ordered data. If $n \le 60$, do not apply this test. If $60 < n \le 220$, then $\alpha = 0.10$. If $n > 220$, then $\alpha = 0.05$.

COMPUTATIONS: Identify the number of possible outliers, r. Compute

$$c = \text{ceiling}\left(\sqrt{2n}\right),\ k = r + c,\ b^2 = 1/\alpha, \text{ and } a = \frac{1 + b\sqrt{\left(c - b^2\right)/(c-1)}}{c - b^2 - 1}$$

where ceiling( ) indicates rounding the value to the next largest integer.

STEP 1. The r smallest points are outliers (with an α level of significance) if

$$X_{(r)} - (1+a)X_{(r+1)} + aX_{(k)} < 0$$

STEP 2. The r largest points are outliers (with an α level of significance) if

$$X_{(n+1-r)} - (1+a)X_{(n-r)} + aX_{(n+1-k)} > 0$$

STEP 3. If both of the inequalities are true, then both small and large outliers are indicated.

4.5 TESTS FOR DISPERSIONS

Many statistical tests make assumptions about the dispersion (variance) of data. This section considers some of the most commonly used statistical tests for comparing variances. Section 4.5.1 constructs a confidence interval for a population variance. Section 4.5.2 provides a test for comparing two population variances. Section 4.5.3 (Bartlett's test) and Section 4.5.4 (Levene's test) describe tests that compare two or more population variances. The analyst should be aware that many statistical tests only require the approximate equality of variances and that many of the tests remain valid unless there is gross inequality in the population variances.

4.5.1 Confidence Intervals for a Single Variance

This section discusses confidence intervals for a single variance or standard deviation. The method described in Box 4-22 can be used to find a two-sided 100(1-α)% confidence interval. The upper end point of a two-sided 100(1-α)% confidence *interval* is also a 100(1-α/2)% upper confidence *limit*, and the lower end point is also a 100(1-α/2)% lower confidence *limit*. Since the standard deviation is the square root of the variance, a confidence interval for the variance can be converted to a confidence interval for the standard deviation by taking the square roots of the endpoints of the interval. The confidence interval procedure assumes the data are a random sample from a normally distributed population and can be highly sensitive to outliers or to departures from normality.

Box 4-22: Directions for Constructing Confidence Intervals and Confidence Limits for the Sample Variance and Standard Deviation with an Example

Let $X_1, X_2, \ldots, X_n$ represent the n data points.

COMPUTATIONS: Calculate the sample variance s^2 (Section 2.2.3).

A 100(1-α)% confidence interval for the population variance, σ^2, is $\left\{ \frac{(n-1)s^2}{\chi^2_{n-1,\,1-\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1,\,\alpha/2}} \right\}$, where Table A-9 is used to find $\chi^2_{n-1,\,\alpha/2}$ and $\chi^2_{n-1,\,1-\alpha/2}$. A 100(1-$\alpha$)% confidence interval for the population standard deviation is the square root of the interval above.

Example: Ten samples were analyzed for lead (in ppb): 46.4, 46.1, 45.8, 47, 46.1, 45.9, 45.8, 46.9, 45.2, 46.

COMPUTATIONS: The sample variance is $s^2 = 0.286$.

A 95% confidence interval for the population variance, σ^2, is

$$\left\{ \frac{(10-1)\cdot 0.286}{\chi^2_{9,\,0.975}}, \frac{(10-1)\cdot 0.286}{\chi^2_{9,\,0.025}} \right\} = \left\{ \frac{9\cdot 0.286}{19.02}, \frac{9\cdot 0.286}{2.70} \right\} = \{0.135,\ 0.953\}$$

4.5.2 The *F*-Test for the Equality of Two Variances

An *F*-test may be used to compare two population variances. The assumptions underlying the *F*-test are that the two samples are independent random samples from two normal populations. The *F*-test for equality of variances is highly sensitive to departures from normality. Directions for implementing an *F*-test with an example are given in Box 4-23.

4.5.3 Bartlett's Test for the Equality of Two or More Variances

Bartlett's test is a means of testing whether two or more population variances of normal distributions are equal. In the case of only two variances, Bartlett's test is equivalent to the *F*-test.

Often in practice unequal variances and non-normality occur together and Bartlett's test is itself sensitive to departures from normality. With long-tailed distributions, the test too often rejects equality of the variances. Rejecting the equality of variances does not mean that one or more is significantly different from the others. It simply implies the variances are unequal as a group. Directions for Bartlett's test are given in Box 4-24 and an example is given in Box 4-25.

4.5.4 Levene's Test for the Equality of Two or More Variances

Levene's test provides an alternative to Bartlett's test for comparing population variances. Levene's test is less sensitive to departures from normality than Bartlett's test and has greater power than Bartlett's for non-normal data. In addition, Levene's test has power nearly as great as Bartlett's test for normally distributed data. Directions and an example of Levene's test are contained in Box 4-26 and Box 4-27, respectively.

Box 4-23: Directions for an *F*-Test to Compare Two Population Variances with an Example

Let $X_1, X_2, \ldots, X_m$ represent the m data points from population 1 and $Y_1, Y_2, \ldots, Y_n$ represent the n data points from population 2.

COMPUTATIONS: Calculate the sample variances s_X^2 and s_Y^2 (Section 2.2.3).

STEP 1. Null Hypothesis: H_0: $\sigma_X^2 = \sigma_Y^2$.

STEP 2. Alternative Hypothesis: H_A: $\sigma_X^2 \neq \sigma_Y^2$.

STEP 3. Test Statistic: $F_0 = \max(F_X, F_Y) = \max\left(\frac{s_X^2}{s_Y^2}, \frac{s_Y^2}{s_X^2}\right)$. If $F_0 = F_X$, then let $k = m - 1$ and $q = n - 1$. If $F_0 = F_Y$, then let $k = n - 1$ and $q = m - 1$.

STEP 4. a) Critical Value: Use Table A-10 to find $F_{k, q, 1-\alpha/2}$.

STEP 4. b) p-value: Use Table A-10 to find $2 \cdot P(F_{k, q} > F_0)$.

STEP 5. a) Conclusion: If $F_0 > F_{k, q, 1-\alpha/2}$, then reject the null hypothesis of equal population variances.

STEP 5. a) Conclusion: If p-value < α, then reject the null hypothesis of equal population variances.

Example: Manganese concentrations (in ppm) were collected from 2 wells. A 0.05-level *F*-test will be used to test if the population variances are equal.

well *X*: 50, 73, 244, 202
well *Y*: 272, 171, 32, 250, 53

COMPUTATIONS: The sample variances are s_X^2 = 9076 and s_Y^2 = 12125.

STEP 1. Null Hypothesis: H_0: $\sigma_X^2 = \sigma_Y^2$.

STEP 2. Alternative Hypothesis: H_A: $\sigma_X^2 \neq \sigma_Y^2$.

STEP 3. Test Statistic: $F_0 = \max(F_X, F_Y) = \max\left(\frac{s_X^2}{s_Y^2}, \frac{s_Y^2}{s_X^2}\right) = \max\left(\frac{9076}{12125}, \frac{12125}{9076}\right) = 1.336$.

Also, $k = 5 - 1 = 4$ and $q = 4 - 1 = 3$.

STEP 4. a) Critical Value: Using Table A-10, $F_{4, 3, 0.975} = 15.1$.

STEP 4. b) p-value: Using Table A-10, p-value = $2 \cdot P(F_{4, 3} > 1.336)$ which exceeds 0.20. (Using software, the exact p-value is 0.8454)

STEP 5. a) Conclusion: Since $F_0 = 1.336 < 15.1 = F_{4, 3, 0.975}$, we fail to reject the null hypothesis of equal population variances.

STEP 5. a) Conclusion: Since p-value = 0.8454 > 0.05 = α, we fail to reject the null hypothesis of equal population variances.

Box 4-24: Directions for Bartlett's Test

Consider k independent random samples with a sample size of n_i for the i^{th} group and let $N = n_1 + \cdots + n_k$.

COMPUTATIONS: For each of the k groups, calculate the sample variance, s_i^2 (Section 2.2.3). Also, compute the pooled variance: $s_p^2 = \frac{1}{N-k}\sum_{i=1}^{k}(n_i - 1)s_i^2$.

STEP 1. Null Hypothesis: H_0: $\sigma_1^2 = \cdots = \sigma_k^2$ (The variances are equal)

STEP 2. Alternative Hypothesis: H_A: The variances are not equal.

STEP 3. Test Statistic: $B_0 = (N-k)\cdot \ln\left(s_p^2\right) - \sum_{i=1}^{k}(n_i - 1)\cdot \ln\left(s_i^2\right)$.

STEP 4. a) Critical Value: Use Table A-9 to find $\chi^2_{k-1,1-\alpha}$.

STEP 4. b) p-value: Use Table A-9 to find $P\left(\chi^2_{k-1} > B_0\right)$.

STEP 5. a) Conclusion: If $B_0 > \chi^2_{k-1,1-\alpha}$, then reject the null hypothesis of equal variances.

STEP 5. a) Conclusion: If p-value < α, then reject the null hypothesis of equal variances.

Box 4-25: An Example of Bartlett's Test

Manganese concentrations were collected from 6 wells over a 4 month period. It is important to determine if the variances of the six wells are equal. Bartlett's test at a significance level of 0.05 will be used.

Sampling Date	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6
January 1	50		272			
February 1	73		171			68
March 1	244	46	32	34	48	991
April 1	202	77	53	3940	54	54
n_i (N=17)	4	2	4	2	2	3
$\overline{X}_i$	142.25	61.50	132	1987	51	371
s_i^2	9076.25	480.5	12454	7628418	18	288349

COMPUTATIONS: The pooled variance is $s_p^2 = \frac{1}{17-6}\left[(4-1)\cdot 9076.25 + \cdots + (3-1)\cdot 288349\right] = 751836.84$

STEP 1. Null Hypothesis: H_0: $\sigma_1^2 = \cdots = \sigma_k^2$.

STEP 2. Alternative Hypothesis: H_A: The variances are not equal.

STEP 3. Test Statistic: $B_0 = (17-6)\cdot \ln(751836.84) - \left[(4-1)\cdot \ln(9076.25) + \cdots + (3-1)\cdot \ln(288349)\right]$
$= 43.15$

STEP 4. a) Critical Value: Using Table A-9, $\chi^2_{5,\,0.95} = 11.07$.

STEP 4. b) p-value: Using Table A-9, p-value = $P\left(\chi_5^2 > 43.15\right)$ < 0.005.
(Using statistical software, the exact p-value is almost 0).

STEP 5. a) Conclusion: Since B_0 = 43.15 > 11.07 = $\chi^2_{k-1,1-\alpha}$, we reject the null hypothesis of equal population variances.

STEP 5. a) Conclusion: Since p-value = almost 0 < 0.05 = α, we reject the null hypothesis of equal population variances.

Box 4-26: Directions for Levene's Test

Consider k independent random samples with a sample size of n_i for the i^{th} group and let $N = n_1 + \cdots + n_k$.

COMPUTATIONS: For each of the k groups, calculate the group mean, $\overline{X}_i$. Then compute the absolute residuals, $z_{ij} = \left| X_{ij} - \overline{X}_i \right|$, where X_{ij} represents the j^{th} value of the i^{th} group. For each group, calculate the mean absolute residual, $\overline{z}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} z_{ij}$. Next, calculate the overall mean absolute residual,

$$\overline{z} = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} z_{ij} = \frac{1}{N} \sum_{i=1}^{k} n_i \overline{z}_i .$$

Finally, compute the following sums of squares for the absolute residuals:

$$SS_{TOTAL} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} z_{ij}^2 - N \cdot \overline{z}^2 , \; SS_{GROUPS} = \sum_{i=1}^{k} n_i \overline{z}_i^2 - N \cdot \overline{z}^2 , \text{ and } SS_{ERROR} = SS_{TOTAL} - SS_{GROUPS} .$$

STEP 1. Null Hypothesis:	H_0: $\sigma_1^2 = \cdots = \sigma_k^2$.
STEP 2. Alternative Hypothesis:	H_A: The variances are not equal.
STEP 3. Test Statistic:	$F_0 = \frac{SS_{GROUPS} / (k-1)}{SS_{ERROR} / (N-k)}$.
STEP 4. a) Critical Value:	Use Table A-10 to find $F_{k\text{-}1,\, N\text{-}k,\, 1\text{-}\alpha}$.
STEP 4. b) p-value:	Use Table A-10 to find $P(F_{k\text{-}1,\, N\text{-}k} > F_0)$.
STEP 5. a) Conclusion:	If $F_0 > F_{k\text{-}1,\, N\text{-}k,\, 1\text{-}\alpha}$, then reject the null hypothesis of equal population variances.
STEP 5. a) Conclusion:	If p-value < α, then reject the null hypothesis of equal population variances.

4.6 TRANSFORMATIONS

Most statistical tests and procedures contain assumptions about the data. For example, some common assumptions are that the data is normally distributed; variance components of a statistical model are additive; two independent data sets have equal variance; and a data set has no trends over time or space. If the data do not satisfy such assumptions, then the results of a statistical procedure or test may be biased or incorrect. Fortunately, data that do not satisfy statistical assumptions may often be converted or transformed mathematically into a form that allows standard statistical tests to perform adequately.

It is not recommended to transform data for estimation purposes. Transforming, estimating, and then transforming the estimate back to the original domain will, in general, lead to biased estimates. A better approach to estimation here is to use a nonparametric procedure.

Box 4-27: An Example of Levene's Test

Four months of data on arsenic concentration (ppm) were collected from six wells at a Superfund site. This data set is shown in the table below. Before analyzing this data, it is important to determine if the variances of the six wells are equal. Levene's test at a significance level of 0.05 will be used to make this determination.

COMPUTATIONS: The table below contains the data values, the absolute residuals (labeled res), the sample means and the absolute residual means.

Month	well 1		well 2		well 3		well 4		well 5		well 6	
	value	res	value	res	value	res	value	res	value	res	value	res
1	22.90	6.43	2.00	13.76	2.0	27.6	7.84	3.42	24.90	11.41	0.34	1.95
2	3.09	13.38	1.25	14.51	109.4	79.8	9.30	1.96	1.30	12.19	4.78	2.49
3	35.70	19.23	7.80	7.96	4.5	25.1	25.90	14.64	0.75	12.74	2.85	0.56
4	4.18	12.29	52.00	36.24	2.5	27.1	2.00	9.26	27.00	13.51	1.20	1.09
	$\overline{X}_1 = 16.47$		$\overline{X}_2 = 15.76$		$\overline{X}_3 = 29.60$		$\overline{X}_4 = 11.26$		$\overline{X}_5 = 13.49$		$\overline{X}_6 = 2.29$	
	$\overline{z}_1 = 12.83$		$\overline{z}_2 = 18.12$		$\overline{z}_3 = 39.90$		$\overline{z}_4 = 7.32$		$\overline{z}_5 = 12.46$		$\overline{z}_6 = 1.52$	

The overall absolute residual mean is $\overline{z}$ = (12.83 + 18.12 + 39.9 + 7.32 + 12.46 + 1.52)/6 = 15.36.

The sum of squares are: $SS_{TOTAL} = 6300.89$, $SS_{WELLS} = 3522.90$, and $SS_{ERROR} = 2777.99$.

STEP 1. Null Hypothesis: H_0: $\sigma_1^2 = \cdots = \sigma_6^2$.

STEP 2. Alternative Hypothesis: H_A: The variances are not equal.

STEP 3. Test Statistic: $$F_0 = \frac{SS_{GROUPS}/(k-1)}{SS_{ERROR}/(N-k)} = \frac{3522.9/(6-1)}{2777.99/(24-6)} = 4.56 .$$

STEP 4. a) Critical Value: Using Table A-10, $F_{5, 18, 0.95} = 2.77$.

STEP 4. b) p-value: Using Table A-10, 0.01 < p-value < 0.001.
Using software, $P(F_{5, 18} > 4.56) = 0.0073$.

STEP 5. a) Conclusion: Since $F_0 = 4.56 > 2.77 = F_{5, 18, 0.95}$, we reject the null hypothesis of equal population variances.

STEP 5. a) Conclusion: Since p-value = 0.0073 < 0.05 = α, we reject the null hypothesis of equal population variances.

4.6.1 Types of Data Transformations

Any mathematical function that is applied to every point in a data set is called a transformation. Some commonly used transformations include:

Logarithmic (*Log X or Ln X*): This transformation is best suited for data that is right-skewed which may occur when the measurement follows a lognormal distribution. The transformation is also helpful when the variance at each level of the data is proportional to the square of the mean of the data points at that level. For example, if the variance of data collected around 50 ppm is approximately 250, but the variance of data collected around 100 ppm is approximately 1000, then a logarithmic transformation may be useful. This situation is often

characterized by having a constant coefficient of variation (standard deviation divided by the mean) over all possible data values.

If some of the original values are zero or negative, it is customary to add a small quantity to make the data value non-zero since the logarithm of zero or a negative number does not exist. The size of the small quantity depends on the magnitude of the non-zero data. As a working point, a value of one tenth the smallest non-zero value could be selected. It does not matter whether a natural (ln) or base 10 (log) transformation is used because the two transformations are related by the expression $\ln(X) = 2.303 \log(X)$. Directions for applying a logarithmic transformation with an example are given in Box 4-28.

Square Root ($\sqrt{X}$): This transformation may be used when dealing with small whole numbers, such as bacteriological counts, or the occurrence of rare events, such as violations of a standard over the course of a year. The underlying assumption is that the original data follow a Poisson-like distribution in which case the mean and variance of the data are equal. It should be noted that the square root transformation overcorrects when very small values and zeros appear in the original data. In these cases, $\sqrt{X+1}$ is often used as a transformation.

Inverse Sine (Arcsine X): This transformation may be used for binomial proportions based on count data to achieve stability in variance. The resulting transformed data are expressed in radians (angular degrees). Special tables must be used to transform the proportions into degrees.

Box-Cox Transformations: This transformation is a complex power transformation that takes the original data and raises each data observation to the power lambda (λ). A logarithmic transformation is a special case of the Box-Cox transformation. The rationale is to find λ such that the transformed data have the best possible additive model for the variance structure, the errors are normally distributed, and the variance is as constant as possible over all possible concentration values. The Maximum Likelihood technique is used to find λ such that the residual error from fitting the theorized model is minimized. In practice, the exact value of λ is often rounded to a convenient value for ease in interpretation (for example, $\lambda = 0.45$ would be rounded to 0.5 as it would then have the interpretation of a square root transform). One of the drawbacks of the Box-Cox transformation is the difficulty in physically interpreting the transformed data.

4.6.2 Reasons for Transforming Data

By transforming the data, assumptions that are not satisfied in the original data can be satisfied by the transformed data. For instance, a right-skewed distribution can be transformed to be approximately Gaussian (normal) by using a logarithmic or square-root transformation. Then the normal-theory procedures can be applied to the transformed data. If data are lognormally distributed, then apply procedures to logarithms of the data. However, selecting the correct transformation may be difficult. If standard transformations do not apply, it is suggested that the data user consult a statistician.

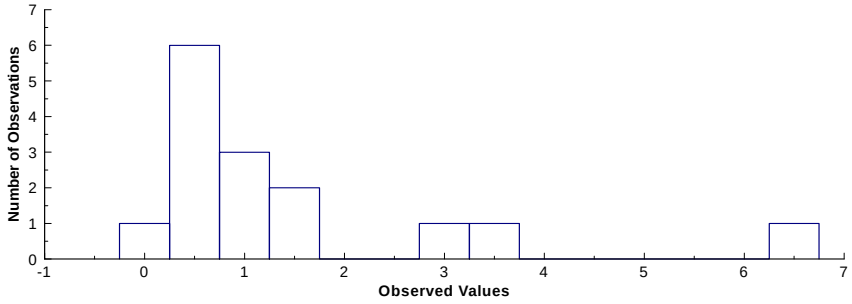
Box 4-28: Directions for Transforming Data and an Example

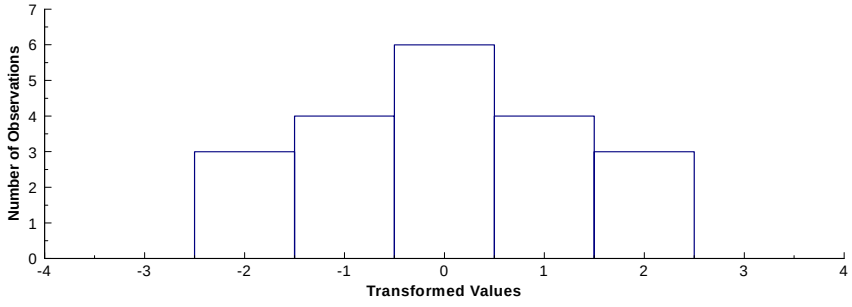
Let $X_1, X_2, \ldots, X_n$ represent the n data points. To apply a transformation, simply apply the transforming function to each data point. When a transformation is implemented to make the data satisfy some statistical assumption, it will need to be verified that the transformed data satisfy this assumption.

Example: Transforming Lognormal Data

A logarithmic transformation is particularly useful for pollution data. Pollution data are often right-skewed, thus the log-transformed data will tend to be symmetric. Consider the data set shown below with 15 data points. A histogram of this data below shows that the data are possibly lognormally distributed. The transformed data are shown in column 2. A histogram of the transformed data below shows that the transformed data appear to be normally distributed.

Observed X	→	Transformed $\ln(X)$	Observed X	→	Transformed $\ln(X)$
0.22	→	-1.51	0.47	→	-0.76
3.48	→	1.25	0.67	→	-0.40
6.67	→	1.90	0.75	→	-0.29
2.53	→	0.93	0.60	→	-0.51
1.11	→	0.10	0.99	→	-0.01
0.33	→	-1.11	0.90	→	-0.11
1.64	→	0.50	0.26	→	-1.35
1.37	→	0.31			





While transformations are useful for dealing with data that do not satisfy statistical assumptions, they can also be used for various other purposes. For example, transformations are useful for consolidating data that may be spread out or that have several extreme values. In addition, transformations can be used to derive a linear relationship between two variables on the newly transformed data, so that linear regression analysis can be applied. They can also be used to efficiently estimate quantities such as the mean and variance of a lognormal distribution.

Transformations may also make the analysis of data easier by changing the scale into one that is more familiar or easier to work with.

Once the data have been transformed, all statistical analysis should be performed on the transformed data. Rarely should an attempt made to transform the data back to the original form because this can lead to biased estimates. For example, estimating quantities such as means, variances, confidence limits, and regression coefficients in the transformed scale typically leads to biased estimates when transformed back into original scale. However, it may be difficult to understand or apply results of statistical analysis expressed in the transformed scale. Therefore, if the transformed data do not give noticeable benefits to the analysis, it is better to use the original data. There is no point in working with transformed data unless it adds value to the analysis.

4.7 VALUES BELOW DETECTION LIMITS

Data generated from chemical analysis may fall below the detection limit (DL) of the analytical procedure. These measurement data are generally described as non-detects rather than as zero or not present and the appropriate limit of detection is usually reported. In cases where measurement data are described as non-detects, the concentration of the chemical is unknown although it lies somewhere between zero and the detection limit. Data that includes both detected and non-detected results are called censored data in the statistical literature.

There are a variety of ways to evaluate data that includes values below the detection limit. However, there are no general procedures that are applicable in all cases. Some general guidelines are presented in Table 4-4. Although these guidelines are usually adequate, they should be implemented cautiously.

Table 4-4. Guidelines for Analyzing Data with Non-Detects

Approximate Percentage of Non-Detects	Section	Statistical Analysis Method
< 15%	4.7.1	Replace non-detects with 0, DL/2, DL., Cohen's Method.
15% - 50%	4.7.2	Trimmed mean, Cohen's Method, Winsorized mean and standard deviation.
> 50% - 90%	4.7.3	Tests for proportions (Section 3.2.1.5)

All of the suggested procedures for analyzing data with non-detects depend upon the amount of data below the detection limit. For relatively small amounts below detection limit values, replacing the non-detects with a small number and proceeding with the usual analysis may be satisfactory depending on the purpose of the analysis. For moderate amounts of data below the detection limit, a more detailed adjustment is appropriate. In situations where relatively large amounts of data fall below the detection limit, one may need only to consider whether or not the chemical was detected above some level. Table 4-4 provides percentages to

assist the user in evaluating their particular situation. However, it should be recognized that these percentages are not hard and fast rules.

In addition, sample size influences which procedures should be used to evaluate the data. For example, the case where 1 sample out of 4 is not detected should be treated differently from the case where 25 samples out of 100 are non-detects. Therefore, this guidance suggests that the data analyst consult a statistician for the most appropriate way to evaluate data containing values below the detection level.

4.7.1 Approximately less than 15% Non-detects - Substitution Methods

If a small proportion of the observations are non-detects, then these may be replaced with a small number, usually DL/2, and the usual analysis performed. Alternative substitution values are 0 (see Aitchison's Method below) or the detection limit. It should be noted that Cohen's Method (section 4.7.2.1) will also work with small amounts of non-detects.

4.7.1.1 Aitchison's Method

Later adjustments to the mean and variance assume that the data values really were present but could not be recorded since they were below the detection limit. However, there are cases where the data values are below the detection limit because they are actually zero, i.e., the contaminant or chemical of concern being entirely absent. Such data sets typically contain a mixture of zero values and present, but nondetected values. Aitchison's Method is simply adjustment formulas for the mean and variance if 0 values are substituted for non-detects. Directions for Aitchison's method are contained in Box 4-29 with an example in Box 4-30.

Box 4-29: Directions for Aitchison's Method to Adjust Means and Variances

Let $X_1, X_2, \ldots, X_m, X_{m+1}, \ldots, X_n$ represent the data points where the first m values are above the detection limit and the remaining n-m data points are below the detection limit.

COMPUTATIONS: Using the data above the detection level, compute the sample mean and sample variance:

$$\overline{X}_d = \sum_{i=1}^{m} X_i \text{ and } s_d^2 = \frac{1}{m-1}\left\{\sum_{i=1}^{m} X_i^2 - \frac{1}{m}\left(\sum_{i=1}^{m} X_i\right)^2\right\}.$$

Compute the adjusted sample mean and sample variance, $\overline{X} = \frac{m}{n}\overline{X}_d$ and $s^2 = \frac{m-1}{n-1}s_d^2 + \frac{m(n-m)}{n(n-1)}\overline{X}_d^2$.

Box 4-30: An Example of Aitchison's Method

The following data consist of 10 Methylene Chloride samples: 1.9,1.3, <1, 2.0, 1.9, <1, <1, <1, 1.6, 1.7. There are 6 values above the detection limit and 4 below, so $m = 6$ and $n\text{-}m = 4$. Aitchison's method will be used to estimate the mean and sample variance of this data.

COMPUTATIONS: Compute the mean and variance for the 6 values above the detection limit

$$\overline{X}_d = 1.733 \text{ and } s_d^2 = 0.0667 .$$

The adjusted sample mean and sample variance are: $\overline{X} = \frac{6}{10} \cdot 1.733 = 1.04$ and

$$s^2 = \frac{6-1}{10-1} \cdot 0.0667 + \frac{6 \cdot 4}{10 \cdot (10-1)} \cdot 1.7333^2 = 0.8382 .$$

4.7.2 Between Approximately 15% - 50% Non-detects

4.7.2.1 Cohen's Method

Cohen's method provides adjusted estimates of the sample mean and standard deviation that accounts for data below the detection level. The adjusted estimates are based on the statistical technique of maximum likelihood estimation of the mean and variance so that the non-detects are accounted for. Care has to be taken when using the adjusted mean and variance in statistical tests. If the percentage of data below the detection limit is relatively small (e.g., less than 20%), the significance level and power of the statistical test are approximately correct. As the proportion of data below detection increases, power declines and the true significance level increases dramatically. This is mainly attributable to the lack of independence between the adjusted mean and adjusted variance. If more than 50% of the observations are not detected, Cohen's method should not be used. In addition, this method requires that the data without the non-detects be normally distributed and that the detection limit is always the same. Directions for Cohen's method are contained in Box 4-31 with an example in Box 4-32.

4.7.2.2 Selecting Between Aitchison's Method and Cohen's Method

Cohen's underlying model is that the population contains a normal distribution, but we cannot see the values below the censoring point. Aitchison's underlying model is that the population consists of a proportion following a normal distribution together with a proportion of values at zero. The difference in concepts becomes relevant depending on the types of inferences made. For example, in estimating upper quantiles, the analyst may use only the normal portion for the statistics, adjusting the quantile to account for the estimated proportion at zero. If a confidence interval for the mean was required a simple substitution of zero for all data below detection would suffice. To determine if a data set is better adjusted by Cohen's method or Aitchison's method, a simple graphical procedure using a Normal Probability Plot (Section 2.3.5) can be used. Directions for this procedure are given in Box 4-34 with an example in Box 4-35.

Box 4-31: Directions for Cohen's Method

Let $X_1, X_2, \ldots, X_m, X_{m+1}, \ldots, X_n$ represent the data points where the first m values are above the detection limit (DL) and the remaining $n-m$ data points are below the detection limit.

COMPUTATIONS: Using the data above the detection level, compute the sample mean and sample variance:

$$\overline{X}_d = \frac{1}{m}\sum_{i=1}^{m} X_i \text{ and } s_d^2 = \frac{1}{m-1}\left\{\sum_{i=1}^{m} X_i^2 - \frac{1}{m}\left(\sum_{i=1}^{m} X_i\right)^2\right\}.$$

Compute $h = \frac{n-m}{n}$ and $\gamma = \frac{s_d^2}{\left(\overline{X}_d - DL\right)^2}$. Use h, γ, and Table A-11 to determine $\hat{\lambda}$. If the exact values of h and γ do not appear in the table, use double linear interpolation (Box 4-33) to estimate $\hat{\lambda}$.

Estimate the corrected sample mean, $\overline{X}$, and sample variance, s^2:

$$\overline{X} = \overline{X}_d - \hat{\lambda}\left(\overline{X}_d - DL\right)$$

$$s^2 = s_d^2 + \hat{\lambda}\left(\overline{X}_d - DL\right)^2.$$

Box 4-32: An Example of Cohen's Method

Sulfate concentrations (mg/L) were measured for 24 data points with 3 values falling below the detection limit of 1450 mg/L. The 24 values are:

1850, 1760, <1450, 1710, 1575, 1475, 1780, 1790, 1780, <1450, 1790, 1800,
<1450, 1800, 1840, 1820, 1860, 1780, 1760, 1800, 1900, 1770, 1790, 1780.

Cohen's Method will be used to adjust the sample mean and sample variance for use in a t-test to determine if the mean is greater than 1600 mg/L.

COMPUTATIONS: The sample mean and sample variance of the $m = 21$ values above the detection level are

$$\overline{X}_d = 1771.9 \text{ and } s_d^2 = 8593.69.$$

The values of h and γ are: $h = \frac{24-21}{24} = 0.125$ and $\gamma = \frac{8593.69}{(1771.9-1450)^2} = 0.083$. Table A-11 does not contain the exact entries for h and γ, double linear interpolation was used to estimate $\hat{\lambda} = 0.149839$ (see Box 4-33).

The adjusted sample mean and sample variance are:

$$\overline{X} = \overline{X}_d - \hat{\lambda}\left(\overline{X}_d - DL\right) = 1771.9 - 0.149839 \cdot (1771.9 - 1450) = 1723.67 = \overline{X}$$

$$s^2 = s_d^2 + \hat{\lambda}\left(\overline{X}_d - DL\right)^2 = 8593.69 + 0.149839 \cdot (1771.9 - 1450)^2 = 24119.95 = s^2$$

Box 4-33: Example of Double Linear Interpolation

The details of the double linear interpolation are provided to assist in the use of Table A-11. The desired value for $\hat{\lambda}$ corresponds to $\gamma = 0.083$ and $h = 0.125$ from Box 4-32. The values from Table A-11 used for interpolation are:

γ	h	
	$c_1 = 0.10$	$c_2 = 0.15$
$r_1 = 0.05$	$x_{11} = 0.11431$	$x_{12} = 0.17925$
$r_2 = 0.10$	$x_{21} = 0.11804$	$x_{22} = 0.18479$

We first interpolate between columns:

$$y_1 = x_{11} + \frac{c - c_1}{c_2 - c_1} \cdot (x_{12} - x_{11}) = 0.11431 + \frac{0.125 - 0.10}{0.15 - 0.10} \cdot (0.17925 - 0.11431) = 0.14678$$

$$y_2 = x_{21} + \frac{c - c_1}{c_2 - c_1} \cdot (x_{22} - x_{21}) = 0.11804 + \frac{0.125 - 0.10}{0.15 - 0.10} \cdot (0.18479 - 0.11804) = 0.151415$$

Now we interpolate between the rows:

$$\hat{\lambda} = y_1 + \frac{r - r_1}{r_2 - r_1} \cdot (y_2 - y_1) = 0.14678 + \frac{0.083 - 0.05}{0.10 - 0.05} \cdot (0.151415 - 0.14678) = 0.149839$$

Box 4-34: Directions for Selecting Between Cohen's Method or Aitchison's Method

Let $X_1, X_2, \ldots, X_m, \ldots, X_n$ represent the data points with the first m values are above the detection limit (DL) and the remaining n-m data points are below the DL.

STEP 1: Use Box 2-17 to construct a Normal Probability Plot of all the data but only plot the values belonging to those above the detection level. This is called the Censored Plot.

STEP 2: Use Box 2-17 to construct a Normal Probability Plot of only those values above the detection level. This called the Detects only Plot.

STEP 3: If the Censored Plot is more linear than the Detects Only Plot, use Cohen's Method to estimate the sample mean and variance. If the Detects Only Plot is more linear than the Censored Plot, then use Aitchison's Method to estimate the sample mean and variance.

4.7.3 Greater than Approximately 50% Non-detects - Test of Proportions

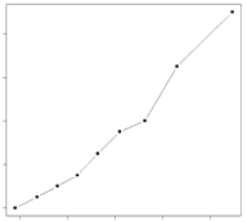
If more than 50% of the data are below the detection limit but at least 10% of the observations are quantified, then the best option is a test of proportions. Thus, if the parameter of interest is a mean, consider switching the parameter of interest to some percentile greater than the percent of data below the detection limit. For example, if 67% of the data are below the DL, consider switching the parameter of interest to the 75th percentile. Then the method described in 3.2.1.5 can be applied to test a hypothesis concerning the 75th percentile. It is important to note that the tests of proportions may not be applicable for composite samples. In this case, the data analyst should consult a statistician before proceeding with analysis.

Box 4-35: Example of Determining Between Cohen's Method and Aitchison's Method

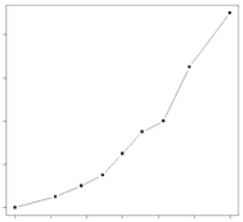
Readings of Chlorobenzene were obtained from a monitoring well:

<1, <1, <1, 1.2, 1.25, 1.3, 1.45, 1.35, 1.55, 1.6, 1.85, 2.1

Step 1: Using the directions in Box 2-17 the following is the Censored Plot:



STEP 2: Using the directions in Box 2-17 the following is the Detects only Plot:



STEP 3: Since the Censored Plots is more linear than the Detects Only Plot, Cohen's Method should be used to estimate the sample mean and variance.

4.7.4 Greater than Approximately 90% Non-detects

If very few quantified values are found, a method based on the Poisson distribution may be used as an alternative approach. However, with a large proportion of non-detects in the data, the data analyst should consult with a statistician before proceeding with analysis.

4.7.5 Recommendations

If the number of sample observations is small ($n < 20$), Cohen's and other maximum likelihood methods can produce biased results since it is difficult to assure that the underlying distribution is appropriate and the solutions to the likelihood equation are statistically consistent only if the number of samples is large. Additionally, most methods will yield estimated parameters with large estimation variance, which reduces the power to detect import differences from standards or between populations. While these methods can be applied to small data sets, the user should be cautioned that they will only be effective in detecting large departures from the null hypothesis.

If the degree of censoring is relatively low, reasonably good estimates of means, variances and upper percentiles can be obtained. However, if the rate of censoring is very high (greater than 50%) then little can be done statistically except to focus on some upper quantile of

the contaminant distribution, or on some proportion of measurements above a certain critical level that is at or above the censoring limit.

When the numerical standard is at or below one of the censoring levels and a one-sample test is used, the most useful statistical method is to test whether the proportion of a population that is above (below) the standard is too large, or to test whether and upper quantile of the population distribution is above the numerical standard. Table 4-5 gives some recommendation on which statistical parameter to use when censoring is present in data sets for different sizes of the coefficient of variation.

Table 4-5. Guidelines for Recommended Parameters for Different Coefficient of Variations and Censoring

Approximate Coefficient of Variation (CV)	Approximate Proportion of Data Below the Detection Limit	
	Low (<50%)	High (>50%)
Large: CV > 1.5	Mean or Upper Percentile	Upper Percentile
Medium: 0.5 < CV < 1.5	Mean or Upper Percentile	Upper Percentile
Small: CV < 0.5	Mean or Median	Median

When comparing two data sets with different censoring levels (i.e., different detection limits), it is recommended that all data be censored at the highest censoring value present and a nonparametric test such as the Wilcoxon Rank Sum Test (Section 3.3.2.1.1) used to compare the two data sets. There is a corresponding loss of statistical power but to a certain extent this can be minimized through the use of large sample sizes.

4.8 INDEPENDENCE

When data are truly independent, the correlation between data points is by definition zero and the selected statistical tests attains the desired decision error rates (given the appropriate assumptions have been satisfied). When correlation exists, the effectiveness of statistical tests is diminished. Environmental data are particularly susceptible to correlation problems due to the fact that such environmental data are collected under a spatial pattern or sequentially over time.

If observations are positively correlated over time or space, then the effective sample size for a test tends to be smaller than the actual sample size—i.e., each additional observation does not provide as much 'new' information because its value is partially determined by the value of adjacent observations. This smaller effective sample size means that the degrees of freedom for the test statistic is smaller, or equivalently, the test is not as powerful as originally thought. In addition to affecting the false acceptance error rate, applying the usual tests to correlated data tends to result in a test whose actual significance level is larger than the nominal error rate.

When observations are correlated, the estimate of the variance in the test statistic formula is often understated. For example, consider the mean of a series of n temporally-ordered observations. If these observations are independent, then the variance of the mean is σ^2/n, where σ^2 is the variance of an individual observation. However, if the observations are not independent and the correlation between successive observations is ρ, then the variance of the mean is

$$\text{var}\left(\overline{X}\right) = \sigma^2\left(1+q\right) \text{ where } q = \frac{2}{n}\sum_{k=1}^{n-1}\left(n-k\right)\rho^k ,$$

which will tend to be larger than σ^2/n if the correlations are positive. If one conducts a t-test at a certain significance level using the usual formula for the estimated variance, then the actual significance level can be as much as double what was expected even for low values of ρ.

One of the most effective ways to determine statistical independence is through use of the Rank von Neumann Test. Directions for this test are given in Box 4-36 with an example in Box 4-37. Compared to other tests of statistical independence, the Rank von Neumann test has been shown to be more powerful over a wide variety of cases. It is also a reasonable test when the data follow a normal distribution.

Box 4-36: Directions for the Rank von Neumann Test

Let $X_1, X_2, \ldots, X_n$ represent the data values collected in sequence over equally spaced periods of time.

COMPUTATIONS: Order the data measurements from smallest to largest and assign rank r_i to measurement X_i. If measurements are tied, then assign the average rank.

STEP 1. Null Hypothesis: H_0: The data are independent.

STEP 2. Alternative Hypothesis: H_A: The data are not independent.

STEP 3. Test Statistic: $v_0 = \frac{12}{n\left(n^2-1\right)}\sum_{i=2}^{n}\left(r_i - r_{i-1}\right)^2$

STEP 4. a) Critical Value: Use Table A-16 to find $v_{n,\alpha}$.

STEP 5. a) Conclusion: If $v_0 < v_{n,\alpha}$, then reject the null hypothesis that the data are independent.

NOTE: If the Rank von Neumann ratio test indicates significant evidence of dependence in the data, then a statistician should be consulted before further analysis is performed. If ranks are tied, the power of the statistical test is diminished

Box 4-37: An Example of the Rank von Neumann Test

The following are hourly readings from a discharge monitor: hourly readings from a discharge monitor.

COMPUTATIONS:

Time	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00	24:00
Reading	6.5	6.6	6.7	6.4	6.3	6.4	6.2	6.2	6.3	6.6	6.8	6.9	7.0
Rank	7	8.5	10	5.5	3.5	5.5	1.5	1.5	3.5	8.5	11	12	13

STEP 1. Null Hypothesis: H_0: The data are independent.

STEP 2. Alternative Hypothesis: H_A: The data are not independent.

STEP 3. Test Statistic: $$v_0 = \frac{12}{13 \cdot \left(13^2 - 1\right)} \left\{ (8.5 - 7)^2 + (10 - 8.5)^2 + \cdots + (13 - 12)^2 \right\} = 0.473$$

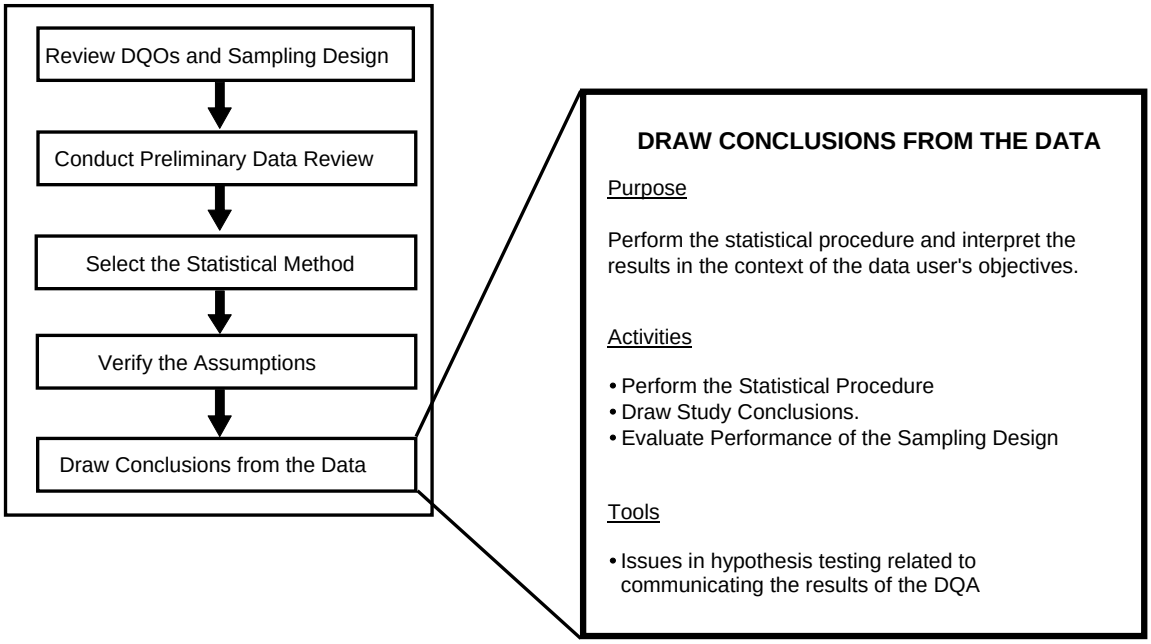
STEP 4. a) Critical Value: Using Table A-16, $v_{13,\,0.05} = 1.14$.

STEP 5. a) Conclusion: Since $v_0 = 0.473 < 1.14 = v_{13,\,0.05}$, we reject the null hypothesis that the data are independent.

CHAPTER 5

STEP 5: DRAW CONCLUSIONS FROM THE DATA

THE DATA QUALITY ASSESSMENT PROCESS



Step 5: Draw Conclusions from the Data

- Perform the calculations for the statistical method.
  - Perform the calculations and document them clearly.
  - If anomalies or outliers are present in the data set, perform the calculations with and without the questionable data.
- Evaluate the results and draw conclusions.
  - If the null hypothesis is rejected, then draw the conclusions and document the analysis.
  - If the null hypothesis is not rejected, verify whether the tolerable limits on false acceptance decision errors have been satisfied. If so, draw conclusions and document the analysis; if not, determine corrective actions, if any.
  - Interpret the results of the test or confidence interval.
- Evaluate the performance of the sampling design if the design is to be used again.
  - Evaluate the statistical power of the design over the full range of parameter values; consult a statistician as necessary.

List of Boxes

CHAPTER 5

STEP 5: DRAW CONCLUSIONS FROM THE DATA

5.1 OVERVIEW AND ACTIVITIES

In this final step of the DQA, the analyst performs the statistical hypothesis test or computes the confidence interval and draws conclusions that address the data user's objectives.

5.2 PERFORM THE STATISTICAL METHOD

The goal of this activity is to conduct the statistical hypothesis test or compute the confidence interval procedure chosen in Chapter 3. The calculations for the procedure should be clearly documented and easily verifiable. In addition, documentation of the results should be understandable so they can be communicated effectively to those who may hold a stake in the resulting decision. If computer software is used to perform the calculations, then the procedures should be adequately documented.

5.3 DRAW STUDY CONCLUSIONS

5.3.1 Hypothesis Tests

The goal of this activity is to translate the results of the statistical hypothesis test so that the data user may draw a conclusion from the data; the results being either:

(a) *reject the null hypothesis*, in which case there is significant evidence in favor of the alternative hypothesis. The decision can be made with sufficient confidence and without further analysis. This is because the statistical tests described in this document inherently control the false rejection error rate within the data user's tolerable limits when the underlying assumptions are valid.

(b) *fail to reject the null hypothesis*, in which case there is not significant evidence for the alternative hypothesis. The analyst is concerned about a possible false acceptance error. The most thorough procedure for verifying whether the false acceptance error limits have been satisfied is to compute the estimated power of the statistical test.

Alternatively the sample size required to satisfy the data user's objectives can be calculated retrospectively using an estimate of the variance or an upper confidence limit on variance obtained from the actual data. If this theoretical sample size is less than or equal to the number of samples actually taken, then the test is probably sufficiently powerful. The equations required to perform these calculations have been provided in the instructions for many of the hypothesis test procedures in Chapter 3. An example of this method is contained in Box 5-1, but it is emphasized that this only gives an estimate of power, not an absolute determination.

Box 5-1: Checking Adequacy of Sample Size for a One-Sample *t*-Test for Simple Random Sampling

In Box 3-3, the one-sample *t*-test was used to test the hypothesis H_0: $\mu \leq 95$ ppm vs. H_A: $\mu > 95$ ppm. DQOs specified that the test should limit the false rejection error rate to 5% and the false acceptance error rate to 20% if the true mean was 105 ppm.

A random sample of size $n = 9$ had sample mean $\overline{X} = 99.38$ ppm and sample standard deviation $s = 10.41$ ppm. The null hypothesis was not rejected. Assuming that the true value of the standard deviation was equal to the sample estimate of 10.41 ppm, it was found that a sample size of 9 would be required. This validated the sample size of 9 which had actually been used.

In such a case it makes sense to build in some conservatism, for example, by using an upper 90% confidence limit for σ in the sample size calculation of Box 3-3. Using Box 4-22, it is found that an upper 90% confidence limit for the true standard deviation is

$$s\sqrt{\frac{n-1}{U}} = 10.41\sqrt{\frac{8}{3.49}} = 15.76 .$$

Using this value for *s* in the sample size calculation of Box 3-3 leads to the sample size estimate of 17. Hence, a sample size of at least 17 should be used to be 90% sure of achieving the DQOs.

Since it is generally desirable to avoid the need for additional sampling, it is advisable to conservatively estimate sample size in the first place. In cases where DQOs depend on a variance estimate, this conservatism is achieved by intentionally overestimating the variance.

5.3.2 Confidence Intervals or Limits

A confidence interval is simply an interval estimate for the population parameter of interest. The interval's width is dependent upon the variance of the point estimate, the sample size, and the confidence level. More specifically, the width is large if the variance is large, the sample size is small, or the confidence level is large.

The interpretation of a confidence interval makes use of probability in an intuitive sense. When a confidence interval has been constructed using the data, there is still a chance that the interval does not include the true value of the parameter estimated. For example, consider this confidence interval statement: "the 95% confidence interval for the unknown population mean is 43.5 to 48.9". It is interpreted as, "I can be 95% certain that the interval 43.5 to 48.9 captures the unknown mean." Notice how there is a 5% chance that the interval does not capture the mean.

The confidence level is the 'confidence' we have that the population parameter lies within the interval. This concept is analogous to the false rejection error rate. The width of the interval is related to statistical power, or the false acceptance error rate. Rather than specifying a desired false acceptance error rate, the desired interval width can be specified.

A confidence interval can be used to make to decisions and in some situations a test of hypothesis is set up as a confidence interval. Confidence intervals are analogous to two-sided hypothesis tests. If the threshold value lies outside of the interval, then there is evidence that the population parameter differs from the threshold value. In a similar manner, confidence limits can also be related to one-sided hypothesis tests. If the threshold value lies above (below) an

upper (lower) confidence bound, then there is evidence that the population parameter is less (greater) than the threshold.

5.3.3 Tolerance Intervals or Limits

A tolerance interval is an interval estimate for a certain proportion of the population. The interval's width is dependent upon the variance of the population, the sample size, the desired proportion of the population, and the confidence level. More specifically, the width is large if the variance is large, the sample size is small, the proportion is large, or the confidence level is large.

When a tolerance interval has been constructed using the data, there is still a chance that the interval does not include the desired proportion of the population. For example, consider this tolerance interval statement: "the 99% tolerance interval for 90% of the population is 7.5 to 9.9". It is interpreted as, "I can be 99% certain that the interval 7.5 to 9.9 captures 90% of the population." Notice how there is a 1% chance that the interval does not capture at least the desired proportion.

The confidence level is the 'confidence' we have that the desired proportion of the population lies within the interval. This concept is analogous to the false rejection error rate. The width of the interval is related to statistical power, or the false acceptance error rate. Rather than specifying a desired false acceptance error rate, the desired interval width can be specified.

A tolerance interval can be used to make to decisions, and in some situations a test of hypothesis can be set up as a tolerance interval. Tolerance intervals are analogous to two-sided hypothesis tests. If the threshold value lies outside of the interval, then there is evidence that the desired proportion of the population differs from the threshold value. In a similar manner, tolerance limits can also be related to one-sided hypothesis tests. If the threshold value lies above (below) an upper (lower) tolerance limit, then there is evidence that the desired proportion of the population is less (greater) than the threshold.

5.4 EVALUATE PERFORMANCE OF THE SAMPLING DESIGN

If the sampling design is to be used again, either in a later phase of the current study or in a similar study, the analyst will be interested in evaluating the overall performance of the design. To evaluate the sampling design, the analyst performs a statistical power analysis that describes the estimated power of the statistical test over the range of possible parameter values. The estimated power is computed for all parameter values under the alternative hypothesis to create a power curve. A power analysis helps the analyst evaluate the adequacy of the sampling design when the true parameter value lies in the vicinity of the action level. In this manner, the analyst may determine how well a statistical test performed and compare this performance with that of other tests. Box 5-2 illustrates power calculations for a test of a single proportion.

Box 5-2: Example of Power Calculations for the One-Sample Test of a Single Proportion

This box illustrates power calculations for the test of H_0: $P \geq 0.20$ vs. H_A: $P < 0.20$, with a false rejection error rate of 5% when $P = 0.20$ presented in Box 3-13. The power of the test will be calculated assuming $P_1 = 0.15$ and before data is available. Since nP_0 and $n(1\text{-}P_0)$ both exceed 5, the normal approximation may be used.

STEP 1: Determine the general conditions for rejection of the null hypothesis. In this case, the null hypothesis is rejected if the sample proportion is sufficiently smaller than P_0. Using Box 3-11, H_0 is rejected if

$$\frac{p + 0.5/n - P_0}{\sqrt{P_0(1-P_0)/n}} < -z_{1-\alpha} \text{ or } p + 0.5/n < P_0 - z_{1-\alpha}\sqrt{P_0 Q_0/n},$$

where p is the sample proportion and $-z_{1-\alpha}$ is the standard normal critical value.

STEP 2: Determine the specific conditions for rejection of the null hypothesis if P_1 is the true value of the proportion P. Using the equations above, rejection occurs if

$$\frac{p + 0.5/n - P_1}{\sqrt{P_1(1-P_1)/n}} < \frac{P_0 - P_1 - z_{1-\alpha}\sqrt{P_0(1-P_0)/n}}{\sqrt{P_1(1-P_1)/n}} = \frac{0.20 - 0.15 - 1645\sqrt{0.2 \cdot 0.8/85}}{\sqrt{0.15 \cdot 0.85/85}} = -0.55$$

STEP 3: Find the probability of rejection if P_1 is the true proportion. The quantity on the left-hand side of the above inequality is a standard normal variable. Hence the power at $P_1 = 0.15$ is the probability that a standard normal variable is less than -0.55. Using Table A-1, this probability is approximately 0.3, which is fairly small.

5.5 INTERPRET AND COMMUNICATE THE RESULTS

At this point, the analyst has performed the applicable statistical procedure and has drawn conclusions. In many cases, the conclusions are straightforward and convincing so they lead to an unambiguous path forward for the project. In other cases, however, it is advantageous to consider these conclusions in a broader context in order to determine a course of action, see *Data Quality Assessment: A Reviewer's Guide* (EPA QA/G-9R) (U.S. EPA 2004).

APPENDIX A:

STATISTICAL TABLES

List of Tables

TABLE A-1. STANDARD NORMAL DISTRIBUTION

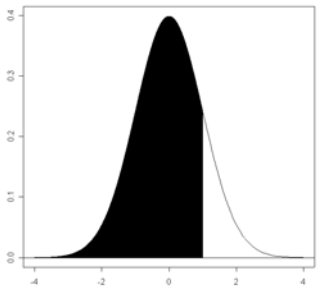


Table values are $P(Z \leq z_p) = p$.

z_p	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
-0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002

TABLE A-1. STANDARD NORMAL DISTRIBUTION (CONT.)

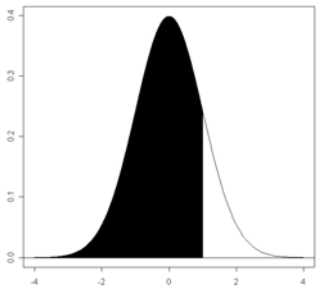
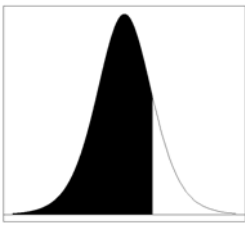


Table values are $P(Z \leq z_p) = p$.

z_p	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6247	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8897	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9889	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9137	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

TABLE A-2. CRITICAL VALUES OF STUDENT'S-*t* DISTRIBUTION



$t_{1-\alpha}$

Degrees of Freedom	1 - α								
	0.70	**0.75**	**0.80**	**0.85**	**0.90**	**0.95**	**0.975**	**0.99**	**0.995**
1	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.536	0.691	0.866	1.074	1.34	1.753	2.131	2.602	2.947
16	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.531	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.531	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.530	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750
40	0.529	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704
60	0.527	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660
120	0.526	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617
∞	0.524	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576

Note: The last row of the table (∞ degrees of freedom) gives the critical values for a standard normal distribution (*Z*), e.g., $t_{\infty, 0.95} = z_{0.95} = 1.645$.

TABLE A-3. CRITICAL VALUES FOR THE STUDENTIZED RANGE TEST

	Level of Significance α					
	0.10		0.05		0.01	
n	***a***	***b***	***a***	***b***	***a***	***b***
3	1.78	2.00	1.76	2.00	1.74	2.00
4	2.04	2.41	1.98	2.43	1.87	2.45
5	2.22	2.71	2.15	2.75	2.02	2.80
6	2.37	2.95	2.28	3.01	2.15	3.10
7	2.49	3.14	2.40	3.22	2.26	3.34
8	2.59	3.31	2.50	3.40	2.35	3.54
9	2.68	3.45	2.59	3.55	2.44	3.72
10	2.76	3.57	2.67	3.69	2.51	3.88
11	2.84	3.68	2.74	3.80	2.58	4.01
12	2.90	3.78	2.80	3.91	2.64	4.13
13	2.96	3.87	2.86	4.00	2.70	4.24
14	3.02	3.95	2.92	4.09	2.75	4.34
15	3.07	4.02	2.97	4.17	2.80	4.44
16	3.12	4.09	3.01	4.24	2.84	4.52
17	3.17	4.15	3.06	4.31	2.88	4.60
18	3.21	4.21	3.10	4.37	2.92	4.67
19	3.25	4.27	3.14	4.43	2.96	4.74
20	3.29	4.32	3.18	4.49	2.99	4.80
25	3.45	4.53	3.34	4.71	3.15	5.06
30	3.59	4.70	3.47	4.89	3.27	5.26
35	3.70	4.84	3.58	5.04	3.38	5.42
40	3.79	4.96	3.67	5.16	3.47	5.56
45	3.88	5.06	3.75	5.26	3.55	5.67
50	3.95	5.14	3.83	5.35	3.62	5.77
55	4.02	5.22	3.90	5.43	3.69	5.86
60	4.08	5.29	3.96	5.51	3.75	5.94
65	4.14	5.35	4.01	5.57	3.80	6.01
70	4.19	5.41	4.06	5.63	3.85	6.07
75	4.24	5.46	4.11	5.68	3.90	6.13
80	4.28	5.51	4.16	5.73	3.94	6.18
85	4.33	5.56	4.20	5.78	3.99	6.23
90	4.36	5.60	4.24	5.82	4.02	6.27
95	4.40	5.64	4.27	5.86	4.06	6.32
100	4.44	5.68	4.31	5.90	4.10	6.36
150	4.72	5.96	4.59	6.18	4.38	6.64
200	4.90	6.15	4.78	6.39	4.59	6.84
500	5.49	6.72	5.47	6.94	5.13	7.42
1000	5.92	7.11	5.79	7.33	5.57	7.80

TABLE A-4. CRITICAL VALUES FOR THE EXTREME VALUE TEST (DIXON'S TEST)

	Level of Significance α		
n	**0.10**	**0.05**	**0.01**
3	0.886	0.941	0.988
4	0.679	0.765	0.889
5	0.557	0.642	0.780
6	0.482	0.560	0.698
7	0.434	0.507	0.637
8	0.479	0.554	0.683
9	0.441	0.512	0.635
10	0.409	0.477	0.597
11	0.517	0.576	0.679
12	0.490	0.546	0.642
13	0.467	0.521	0.615
14	0.492	0.546	0.641
15	0.472	0.525	0.616
16	0.454	0.507	0.595
17	0.438	0.490	0.577
18	0.424	0.475	0.561
19	0.412	0.462	0.547
20	0.401	0.450	0.535
21	0.391	0.440	0.524
22	0.382	0.430	0.514
23	0.374	0.421	0.505
24	0.367	0.413	0.497
25	0.360	0.406	0.489

TABLE A-5. CRITICAL VALUES FOR DISCORDANCE TEST

	Level of Significance α	
N	**0.01**	**0.05**
3	1.155	1.153
4	1.492	1.463
5	1.749	1.672
6	1.944	1.822
7	2.097	1.938
8	2.221	2.032
9	2.323	2.110
10	2.410	2.176
11	2.485	2.234
12	2.550	2.285
13	2.607	2.331
14	2.659	2.371
15	2.705	2.409
16	2.747	2.443
17	2.785	2.475
18	2.821	2.504
19	2.854	2.532
20	2.884	2.557
21	2.912	2.580
22	2.939	2.603
23	2.963	2.624
24	2.987	2.644
25	3.009	2.663
26	3.029	2.681
27	3.049	2.698
28	3.068	2.714
29	3.085	2.730
30	3.103	2.745

	Level of Significance α	
n	**0.01**	**0.05**
31	3.119	2.759
32	3.135	2.773
33	3.150	2.786
34	3.164	2.799
35	3.178	2.811
36	3.191	2.823
37	3.204	2.835
38	3.216	2.846
39	3.228	2.857
40	3.240	2.866
41	3.251	2.877
42	3.261	2.887
43	3.271	2.896
44	3.282	2.905
45	3.292	2.914
46	3.302	2.923
47	3.310	2.931
48	3.319	2.940
49	3.329	2.948
50	3.336	2.956

TABLE A-6. APPROXIMATE CRITICAL VALUES λ_r FOR ROSNER'S TEST

		α	
n	***r***	**0.05**	**0.01**
25	**1**	2.82	3.14
	2	2.80	3.11
	3	2.78	3.09
	4	2.76	3.06
	5	2.73	3.03
	10	2.59	2.85
26	**1**	2.84	3.16
	2	2.82	3.14
	3	2.80	3.11
	4	2.78	3.09
	5	2.76	3.06
	10	2.62	2.89
27	**1**	2.86	3.18
	2	2.84	3.16
	3	2.82	3.14
	4	2.80	3.11
	5	2.78	3.09
	10	2.65	2.93
28	**1**	2.88	3.20
	2	2.86	3.18
	3	2.84	3.16
	4	2.82	3.14
	5	2.80	3.11
	10	2.68	2.97
29	**1**	2.89	3.22
	2	2.88	3.20
	3	2.86	3.18
	4	2.84	3.16
	5	2.82	3.14
	10	2.71	3.00
30	**1**	2.91	3.24
	2	2.89	3.22
	3	2.88	3.20
	4	2.86	3.18
	5	2.84	3.16
	10	2.73	3.03
31	**1**	2.92	3.25
	2	2.91	3.24
	3	2.89	3.22
	4	2.88	3.20
	5	2.86	3.18
	10	2.76	3.06

		α	
n	***r***	**0.05**	**0.01**
32	**1**	2.94	3.27
	2	2.92	3.25
	3	2.91	3.24
	4	2.89	3.22
	5	2.88	3.20
	10	2.78	3.09
33	**1**	2.95	3.29
	2	2.94	3.27
	3	2.92	3.25
	4	2.91	3.24
	5	2.89	3.22
	10	2.80	3.11
34	**1**	2.97	3.30
	2	2.95	3.29
	3	2.94	3.27
	4	2.92	3.25
	5	2.91	3.24
	10	2.82	3.14
35	**1**	2.98	3.32
	2	2.97	3.30
	3	2.95	3.29
	4	2.94	3.27
	5	2.92	3.25
	10	2.84	3.16
36	**1**	2.99	3.33
	2	2.98	3.32
	3	2.97	3.30
	4	2.95	3.29
	5	2.94	3.27
	10	2.86	3.18
37	**1**	3.00	3.34
	2	2.99	3.33
	3	2.98	3.32
	4	2.97	3.30
	5	2.95	3.29
	10	2.88	3.20
38	**1**	3.01	3.36
	2	3.00	3.34
	3	2.99	3.33
	4	2.98	3.32
	5	2.97	3.30
	10	2.91	3.22

		α	
n	***r***	**0.05**	**0.01**
39	**1**	3.03	3.37
	2	3.01	3.36
	3	3.00	3.34
	4	2.99	3.33
	5	2.98	3.32
	10	2.91	3.24
40	**1**	3.04	3.38
	2	3.03	3.37
	3	3.01	3.36
	4	3.00	3.34
	5	2.99	3.33
	10	2.92	3.25
41	**1**	3.05	3.39
	2	3.04	3.38
	3	3.03	3.37
	4	3.01	3.36
	5	3.00	3.34
	10	2.94	3.27
42	**1**	3.06	3.40
	2	3.05	3.39
	3	3.04	3.38
	4	3.03	3.37
	5	3.01	3.36
	10	2.95	3.29
43	**1**	3.07	3.41
	2	3.06	3.40
	3	3.05	3.39
	4	3.04	3.38
	5	3.03	3.37
	10	2.97	3.30
44	**1**	3.08	3.43
	2	3.07	3.41
	3	3.06	3.40
	4	3.05	3.39
	5	3.04	3.38
	10	2.98	3.32
45	**1**	3.09	3.44
	2	3.08	3.43
	3	3.07	3.41
	4	3.06	3.40
	5	3.05	3.39
	10	2.99	3.33

TABLE A-6. APPROXIMATE CRITICAL VALUES λ_r FOR ROSNER'S TEST (CONT.)

		α	
n	***r***	**0.05**	**0.01**
46	**1**	3.09	3.45
	2	3.09	3.44
	3	3.08	3.43
	4	3.07	3.41
	5	3.06	3.40
	10	3.00	3.34
47	**1**	3.10	3.46
	2	3.09	3.45
	3	3.09	3.44
	4	3.08	3.43
	5	3.07	3.41
	10	3.01	3.36
48	**1**	3.11	3.46
	2	3.10	3.46
	3	3.09	3.45
	4	3.09	3.44
	5	3.08	3.43
	10	3.03	3.37
49	**1**	3.12	3.47
	2	3.11	3.46
	3	3.10	3.46
	4	3.09	3.45
	5	3.09	3.44
	10	3.04	3.38
50	**1**	3.13	3.48
	2	3.12	3.47
	3	3.11	3.46
	4	3.10	3.46
	5	3.09	3.45
	10	3.05	3.39
60	**1**	3.20	3.56
	2	3.19	3.55
	3	3.19	3.55
	4	3.18	3.54
	5	3.17	3.53
	10	3.14	3.49
70	**1**	3.26	3.62
	2	3.25	3.62
	3	3.25	3.61
	4	3.24	3.60
	5	3.24	3.60
	10	3.21	3.57
80	**1**	3.31	3.67
	2	3.30	3.67
	3	3.30	3.66
	4	3.29	3.66
	5	3.29	3.65
	10	3.26	3.63
90	**1**	3.35	3.72
	2	3.34	3.71
	3	3.34	3.71
	4	3.34	3.70
	5	3.33	3.70
	10	3.31	3.68
100	**1**	3.38	3.75
	2	3.38	3.75
	3	3.38	3.75
	4	3.37	3.74
	5	3.37	3.74
	10	3.35	3.72
150	**1**	3.52	3.89
	2	3.51	3.89
	3	3.51	3.89
	4	3.51	3.88
	5	3.51	3.88
	10	3.50	3.87
200	**1**	3.61	3.98
	2	3.60	3.98
	3	3.60	3.97
	4	3.60	3.97
	5	3.60	3.97
	10	3.59	3.96
250	**1**	3.67	4.04
	5	3.67	4.04
	10	3.66	4.03
300	**1**	3.72	4.09
	5	3.72	4.09
	10	3.71	4.09
350	**1**	3.77	4.14
	5	3.76	4.13
	10	3.76	4.13
400	**1**	3.80	4.17
	5	3.80	4.17
	10	3.80	4.16
450	**1**	3.84	4.20
	5	3.83	4.20
	10	3.83	4.20
500	**1**	3.86	4.23
	5	3.86	4.23
	10	3.86	4.22

TABLE A-7. QUANTILES OF THE WILCOXON SIGNED RANKS TEST

Values in the table are such that $P(T^+ \leq w_\alpha)$ is approximately equal to, but less than α. For example, if $n = 12$, then $P(T^+ \leq 17) = 0.0461$, which is slightly less than 0.05. Note the exact probability was computed using the statistical software package R.

N	$w_{0.005}$	$w_{0.01}$	$w_{0.025}$	$w_{0.05}$	$w_{0.075}$	$w_{0.10}$	$w_{0.15}$	$w_{0.20}$
4	-	-	-	-	0	0	1	2
5	-	-	-	0	1	2	2	3
6	-	-	0	2	2	3	4	5
7	-	0	2	3	4	5	7	8
8	0	1	3	5	7	8	9	11
9	1	3	5	8	9	10	12	14
10	3	5	8	10	12	14	16	18
11	5	7	10	13	16	17	20	22
12	7	9	13	17	19	21	24	27
13	9	12	17	21	24	26	29	32
14	12	15	21	25	28	31	35	38
15	15	19	25	30	33	36	40	44
16	19	23	29	35	39	42	47	50
17	23	27	34	41	45	48	53	57
18	27	32	40	47	51	55	60	65
19	32	37	46	53	58	62	68	73
20	37	43	52	60	65	69	76	81

TABLE A-8. CRITICAL VALUES FOR THE WILCOXON RANK-SUM TEST

Table values are the largest x values such that $P(W_{rs} \le x) \le \alpha$. Therefore, significance levels, α, are approximate. If there are ties, then the test is approximate.

min(m, n)	α	max(m, n)								
		2	3	4	5	6	7	8	9	10
2	0.010	-	-	-	-	-	-	-	-	-
	0.025	-	-	-	-	-	-	0	0	0
	0.050	-	-	-	0	0	0	1	1	1
	0.100	-	0	0	1	1	1	2	2	3
3	0.010		-	-	-	0	0	0	1	1
	0.025		-	-	0	1	1	2	2	3
	0.050		0	0	1	2	2	3	4	4
	0.100		1	1	2	3	4	5	5	6
4	0.010			-	0	1	1	2	3	3
	0.025			0	0	2	3	4	4	5
	0.050			1	2	3	4	5	6	7
	0.100			3	4	5	6	7	9	10
5	0.010				1	2	3	4	5	6
	0.025				2	3	5	6	7	8
	0.050				4	5	6	8	9	11
	0.100				5	7	8	10	12	13
6	0.010					3	4	6	7	8
	0.025					5	6	8	10	11
	0.050					7	8	10	12	14
	0.100					9	11	13	15	17
7	0.010						6	7	9	11
	0.025						8	10	12	14
	0.050						11	13	15	17
	0.100						13	16	18	21
8	0.010							9	11	13
	0.025							13	15	17
	0.050							15	18	20
	0.100							19	22	25
9	0.010								14	16
	0.025								17	20
	0.050								21	24
	0.100								25	28
10	0.010									13
	0.025									23
	0.050									27
	0.100									32

TABLE A-8. CRITICAL VALUES FOR THE WILCOXON RANK-SUM TEST (CONT.)

Table values are the largest x values such that $P(W_{rs} \leq x) \leq \alpha$. Therefore, significance levels, α, are approximate. If there are ties, then the test is approximate.

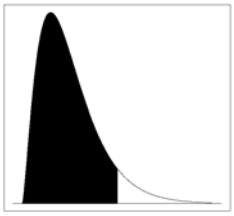
min(m, n)	α	max(m, n)									
		11	12	13	14	15	16	17	18	19	20
2	0.010	-	-	0	0	0	0	0	0	1	1
	0.025	0	1	1	1	1	1	2	2	2	2
	0.050	1	2	2	3	3	3	3	4	4	4
	0.100	3	4	4	4	5	5	6	6	7	7
3	0.010	1	2	2	2	3	3	4	4	4	5
	0.025	3	4	4	5	5	6	6	7	7	8
	0.050	5	5	6	7	7	8	9	9	10	11
	0.100	7	8	9	10	10	11	12	13	14	15
4	0.010	4	5	5	6	7	7	8	9	9	10
	0.025	6	7	8	9	10	11	11	12	13	14
	0.050	8	9	10	11	12	14	15	16	17	18
	0.100	11	12	13	15	16	17	18	20	21	22
5	0.010	7	8	9	10	11	12	13	14	15	16
	0.025	9	11	12	13	14	15	16	18	19	20
	0.050	12	13	15	16	18	19	20	22	23	25
	0.100	15	17	18	20	22	23	25	27	28	30
6	0.010	9	11	12	13	15	16	18	19	20	22
	0.025	13	14	16	17	19	21	22	24	25	27
	0.050	16	17	19	21	23	25	26	28	30	32
	0.100	19	21	23	25	27	29	31	34	36	38
7	0.010	12	14	16	17	19	21	23	24	26	28
	0.025	16	18	20	22	24	26	28	30	32	33
	0.050	19	21	24	26	28	30	33	35	37	39
	0.100	23	26	28	31	33	36	38	41	43	46
8	0.010	15	17	20	22	24	26	28	30	32	34
	0.025	19	22	24	26	28	31	34	36	38	41
	0.050	23	26	28	31	33	36	39	41	44	47
	0.100	27	30	33	36	39	42	45	48	51	54
9	0.010	18	21	23	26	28	31	33	36	38	40
	0.025	23	26	28	31	24	37	39	42	45	48
	0.050	27	30	33	36	39	42	45	48	51	54
	0.100	31	35	38	41	45	48	52	55	58	62
10	0.010	22	24	27	30	33	36	38	41	44	47
	0.025	26	29	33	36	39	42	45	48	52	55
	0.050	31	34	37	41	44	48	51	55	58	62
	0.100	36	39	43	47	51	54	58	62	66	70

TABLE A-8. CRITICAL VALUES FOR THE WILCOXON RANK-SUM TEST (CONT.)

Table values are the largest x values such that $P(W_{rs} \le x) \le \alpha$. Therefore, significance levels, α, are approximate. If there are ties, then the test is approximate.

min(m, n)	α	max(m, n)									
		11	12	13	14	15	16	17	18	19	20
11	0.010	25	28	31	34	37	41	44	47	50	53
	0.025	30	33	37	40	44	47	51	55	58	62
	0.050	34	38	42	46	50	54	57	61	65	69
	0.100	40	44	48	52	57	61	65	69	73	78
12	0.010		31	35	38	52	46	48	53	56	60
	0.025		37	41	45	49	53	57	61	65	69
	0.050		42	47	51	55	60	64	68	72	77
	0.100		49	53	58	63	68	72	77	81	86
13	0.010			39	43	47	51	55	59	63	67
	0.025			45	50	54	59	63	67	72	76
	0.050			51	56	61	65	70	75	80	84
	0.100			58	63	68	74	79	84	89	94
14	0.010				47	51	56	60	65	69	73
	0.025				55	59	64	69	74	78	83
	0.050				61	66	71	77	82	87	92
	0.100				69	74	80	85	91	97	102
15	0.010					56	61	66	70	75	80
	0.025					64	70	75	80	85	90
	0.050					72	77	83	88	94	100
	0.100					80	86	92	98	104	110
16	0.010						66	71	76	82	87
	0.025						75	81	86	92	98
	0.050						83	89	95	101	107
	0.100						93	99	106	112	119
17	0.010							77	82	88	93
	0.025							87	93	99	105
	0.050							96	102	109	115
	0.100							106	113	120	127
18	0.010								88	94	100
	0.025								99	106	112
	0.050								109	116	123
	0.100								120	128	135
19	0.010									101	107
	0.025									113	119
	0.050									123	130
	0.100									135	143
20	0.010										114
	0.025										127
	0.050										138
	0.100										151

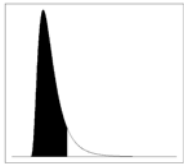
TABLE A-9. PERCENTILES OF THE CHI-SQUARE DISTRIBUTION



$\chi^2_{1-\alpha}$

	1 - α									
df	**0.005**	**0.010**	**0.025**	**0.050**	**0.100**	**0.900**	**0.950**	**0.975**	**0.990**	**0.995**
1	0.0^4393	0.0^3157	0.0^3982	0.0^2393	0.0158	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	4.61	5.99	7.38	9.21	10.60
3	0.072	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	7.78	9.49	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	1.61	9.24	11.07	12.83	15.09	16.75
6	0.676	0.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	0.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	3.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	53.33	85.53	90.53	95.02	100.4	104.2
80	51.17	53.54	57.15	60.39	64.28	96.58	101.9	106.6	112.3	116.3
90	59.20	61.75	65.65	69.13	73.29	107.6	113.1	118.1	124.1	128.3
100	67.33	70.06	74.22	77.93	82.36	118.5	124.3	129.6	135.8	140.2

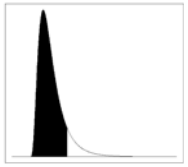
TABLE A-10. PERCENTILES OF THE *F*-DISTRIBUTION



$F_{1-\alpha}$

Degrees Freedom for Denominator		Degrees of Freedom for Numerator																	
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	60	120	∞
1	.90	39.9	49.5	53.6	55.8	57.2	58.2	58.9	59.4	59.9	60.2	60.7	61.2	61.7	62.0	62.3	62.8	63.1	63.3
	.95	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	252	253	254
	.975	648	800	864	900	922	937	948	957	963	969	977	985	993	997	1001	1010	1014	1018
	.99	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157	6209	6235	6261	6313	6339	6366
2	.90	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.48	9.49
	.95	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5
	.975	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	39.4	39.4	39.5	39.5	39.5	39.5	39.5
	.99	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5
3	.90	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.15	5.14	5.13
	.95	10.1	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	8.57	8.55	8.53
	.975	17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4	14.3	14.3	14.2	14.1	14.1	14.0	13.9	13.9
	.99	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.6	26.5	26.3	26.2	26.1
4	.90	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.79	3.78	3.76
	.95	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	5.69	5.66	5.63
	.975	12.2	10.6	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.36	8.31	8.26
	.99	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.9	13.8	13.7	13.6	13.5
	.999	74.1	61.2	56.2	53.4	51.7	50.5	49.7	49.0	48.5	48.1	47.4	46.8	46.1	45.8	45.4	44.7	44.4	44.1
5	.90	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.39	3.27	3.24	3.21	3.19	3.17	3.14	3.12	3.11
	.95	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	4.43	4.40	4.37
	.975	10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.12	6.07	6.02
	.99	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.47	9.38	9.20	9.11	9.02
	.999	47.2	37.1	33.2	31.1	29.8	28.8	28.2	27.6	27.2	26.9	26.4	25.9	25.4	25.1	24.9	24.3	24.1	23.8
6	.90	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.76	2.74	2.72
	.95	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	3.74	3.70	3.67
	.975	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	4.96	4.90	4.85
	.99	22.8	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.06	6.97	6.88
	.999	35.5	27.0	23.7	21.9	20.8	20.0	19.5	19.0	18.7	18.4	18.0	17.6	17.1	16.9	16.7	16.2	16.0	15.7

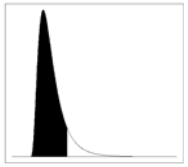
TABLE A-10. PERCENTILES OF THE *F*-DISTRIBUTION (CONT.)



$F_{1-\alpha}$

Degrees Freedom for Denominator		Degrees of Freedom for Numerator																	
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	60	120	∞
7	.90	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.51	2.49	2.47
	.95	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	3.30	3.27	3.23
	.975	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.25	4.20	4.14
	.99	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.82	5.74	5.65
	.999	29.2	21.7	18.8	17.2	16.2	15.5	15.0	14.6	14.5	14.1	13.7	13.3	12.9	12.7	12.5	12.1	11.9	11.7
8	.90	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.34	2.32	2.29
	.95	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	3.01	2.97	2.93
	.975	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.78	3.73	3.67
	.99	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.03	4.95	4.86
	.999	25.4	18.5	15.8	14.4	13.5	12.9	12.4	12.0	11.8	11.5	11.2	10.8	10.5	10.3	10.1	9.73	9.53	9.33
9	.90	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.21	2.18	2.16
	.95	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	2.79	2.75	2.71
	.975	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.45	3.39	3.33
	.99	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.48	4.40	4.31
	.999	22.9	16.4	13.9	12.6	11.7	11.1	10.7	10.4	10.1	9.89	9.57	9.24	8.90	8.72	8.55	8.19	8.00	7.81
10	.90	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.11	2.08	2.06
	.95	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.84	2.77	2.74	2.70	2.62	2.58	2.54
	.975	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.20	3.14	3.08
	.99	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.08	4.00	3.91
	.999	21.0	14.9	12.6	11.3	10.5	9.93	9.52	9.20	8.96	8.75	8.45	8.13	7.80	7.64	7.47	7.12	6.94	6.76
12	.90	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.96	1.93	1.90
	.95	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	2.38	2.34	2.30
	.975	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.85	2.79	2.72
	.99	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.54	3.45	3.36
	.999	18.6	13.0	10.8	9.63	8.89	8.38	8.00	7.71	7.48	7.29	7.00	6.71	6.40	6.25	6.09	5.76	5.59	5.42

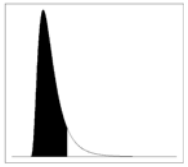
TABLE A-10. PERCENTILES OF THE *F*-DISTRIBUTION (CONT.)



$F_{1-\alpha}$

Degrees Freedom for Denominator		Degrees of Freedom for Numerator																	
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	60	120	∞
15	.90	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.82	1.79	1.76
	.95	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	2.16	2.11	2.07
	.975	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.52	2.46	2.40
	.99	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.05	2.96	2.87
	.999	16.6	11.3	9.34	8.25	7.57	7.09	6.74	6.47	6.26	6.08	5.81	5.54	5.25	5.10	4.95	4.64	4.48	4.31
20	.90	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.68	1.64	1.61
	.95	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	1.95	1.90	1.84
	.975	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.22	2.16	2.09
	.99	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.61	2.52	2.42
	.999	14.8	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24	5.08	4.82	4.56	4.29	4.15	4.00	3.70	3.54	3.38
24	.90	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.61	1.57	1.53
	.95	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	1.84	1.79	1.73
	.975	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.08	2.01	1.94
	.99	7.82	6.66	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.40	2.31	2.21
	.999	14.0	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.80	4.64	4.39	4.14	3.87	3.74	3.59	3.29	3.14	2.97
30	.90	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.62	1.64	1.61	1.54	1.50	1.46
	.95	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	1.74	1.68	1.62
	.975	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	1.94	1.87	1.79
	.99	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.21	2.11	2.01
	.999	13.3	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39	4.24	4.00	3.75	3.49	3.36	3.22	2.92	2.76	2.59
60	.90	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.40	1.35	1.29
	.95	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	1.53	1.47	1.39
	.975	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.67	1.58	1.48
	.99	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.84	1.73	1.60
	.999	12.0	7.77	6.17	5.31	4.76	4.37	4.09	3.86	3.69	3.54	3.32	3.08	2.83	2.69	2.55	2.25	2.08	1.89

TABLE A-10. PERCENTILES OF THE *F*-DISTRIBUTION (CONT.)



$F_{1-\alpha}$

Degrees Freedom for Denominator		Degrees of Freedom for Numerator																
		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	60	120	∞
120	.90	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.32	1.26	1.19
	.95	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.43	1.35	1.25
	.975	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.95	1.82	1.76	1.69	1.53	1.43	1.31
	.99	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.66	1.53	1.38
	.999	11.4	7.32	5.78	4.95	4.42	4.04	3.77	3.55	3.38	3.24	3.02	2.78	2.53	2.40	2.26	1.95	1.77	1.54
∞	.90	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.24	1.17	1.00
	.95	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	1.32	1.22	1.00
	.975	5.02	3.69	3.12	2.79	2.57	22.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.39	1.27	1.00
	.99	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.47	1.32	1.00
	.999	10.8	6.91	5.42	4.62	4.10	3.74	3.47	3.27	3.10	2.96	2.74	2.51	2.27	2.13	1.99	1.66	1.45	1.00

TABLE A-11. 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**
.00	.010100	.020400	.030902	.041583	.052507	.063625	.074953	.08649	.09824	.11020	.17342	.24268
.05	.010551	.021294	.032225	.043350	.054670	.066159	.077909	.08983	.10197	.11431	.17925	.25033
.10	.010950	.022082	.033398	.044902	.056596	.068483	.080563	.09285	.10534	.11804	.18479	.25741
.15	.011310	.022798	.034466	.046318	.058356	.070586	.083009	.09563	.10845	.12148	.18985	.26405
.20	.011642	.023459	.035453	.047829	.059990	.072539	.085280	.09822	.11135	.12469	.19460	.27031
.25	.011952	.024076	.036377	.048858	.061522	.074372	.087413	.10065	.11408	.12772	.19910	.27626
.30	.012243	.024658	.037249	.050018	.062969	.076106	.089433	.10295	.11667	.13059	.20338	.28193
.35	.012520	.025211	.038077	.051120	.064345	.077736	.091355	.10515	.11914	.13333	.20747	.28737
.40	.012784	.025738	.038866	.052173	.065660	.079332	.093193	.10725	.12150	.13595	.21129	.29250
.45	.013036	.026243	.039624	.053182	.066921	.080845	.094958	.10926	.12377	.13847	.21517	.29765
.50	.013279	.026728	.040352	.054153	.068135	.082301	.096657	.11121	.12595	.14090	.21882	.30253
.55	.013513	.027196	.041054	.055089	.069306	.083708	.098298	.11208	.12806	.14325	.22225	.30725
.60	.013739	.027849	.041733	.055995	.070439	.085068	.099887	.11490	.13011	.14552	.22578	.31184
.65	.013958	.028087	.042391	.056874	.071538	.086388	.10143	.11666	.13209	.14773	.22910	.31630
.70	.014171	.028513	.043030	.057726	.072505	.087670	.10292	.11837	.13402	.14987	.23234	.32065
.75	.014378	.029927	.043652	.058556	.073643	.088917	.10438	.12004	.13590	.15196	.23550	.32489
.80	.014579	.029330	.044258	.059364	.074655	.090133	.10580	.12167	.13775	.15400	.23858	.32903
.85	.014773	.029723	.044848	.060153	.075642	.091319	.10719	.12225	.13952	.15599	.24158	.33307
.90	.014967	.030107	.045425	.060923	.075606	.092477	.10854	.12480	.14126	.15793	.24452	.33703
.95	.015154	.030483	.045989	.061676	.077549	.093611	.10987	.12632	.14297	.15983	.24740	.34091
1.00	.015338	.030850	.046540	.062413	.078471	.094720	.11116	.12780	.14465	.16170	.25022	.34471

	h											
γ	**.25**	**.30**	**.35**	**.40**	**.45**	**.50**	**.55**	**.60**	**.65**	**.70**	**.80**	**.90**
.00	.31862	.4021	.4941	.5961	.7096	.8388	.9808	1.145	1.336	1.561	2.176	3.283
.05	.32793	.4130	.5066	.6101	.7252	.8540	.9994	1.166	1.358	1.585	2.203	3.314
.10	.33662	.4233	.5184	.6234	.7400	.8703	1.017	1.185	1.379	1.608	2.229	3.345
.15	.34480	.4330	.5296	.6361	.7542	.8860	1.035	1.204	1.400	1.630	2.255	3.376
.20	.35255	.4422	.5403	.6483	.7673	.9012	1.051	1.222	1.419	1.651	2.280	3.405
.25	.35993	.4510	.5506	.6600	.7810	.9158	1.067	1.240	1.439	1.672	2.305	3.435
.30	.36700	.4595	.5604	.6713	.7937	.9300	1.083	1.257	1.457	1.693	2.329	3.464
.35	.37379	.4676	.5699	.6821	.8060	.9437	1.098	1.274	1.475	1.713	2.353	3.492
.40	.38033	.4735	.5791	.6927	.8179	.9570	1.113	1.290	1.494	1.732	2.376	3.520
.45	.38665	.4831	.5880	.7029	.8295	.9700	1.127	1.306	1.511	1.751	2.399	3.547
.50	.39276	.4904	.5967	.7129	.8408	.9826	1.141	1.321	1.528	1.770	2.421	3.575
.55	.39679	.4976	.6061	.7225	.8517	.9950	1.155	1.337	1.545	1.788	2.443	3.601
.60	.40447	.5045	.6133	.7320	.8625	1.007	1.169	1.351	1.561	1.806	2.465	3.628
.65	.41008	.5114	.6213	.7412	.8729	1.019	1.182	1.368	1.577	1.824	2.486	3.654
.70	.41555	.5180	.6291	.7502	.8832	1.030	1.195	1.380	1.593	1.841	2.507	3.679
.75	.42090	.5245	.6367	.7590	.8932	1.042	1.207	1.394	1.608	1.851	2.528	3.705
.80	.42612	.5308	.6441	.7676	.9031	1.053	1.220	1.408	1.624	1.875	2.548	3.730
.85	.43122	.5370	.6515	.7781	.9127	1.064	1.232	1.422	1.639	1.892	2.568	3.754
.90	.43622	.5430	.6586	.7844	.9222	1.074	1.244	1.435	1.653	1.908	2.588	3.779
.95	.44112	.5490	.6656	.7925	.9314	1.085	1.255	1.448	1.668	1.924	2.607	3.803
1.00	.44592	.5548	.6724	.8005	.9406	1.095	1.287	1.461	1.882	1.940	2.626	3.827

TABLE A-12a: CRITICAL VALUES FOR THE MANN-KENDALL TEST FOR TREND

Significance Level α	Sample size n																
	4	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
0.20	4	6	7	7	8	10	11	13	14	16	17	19	20	22	25	27	28
0.10	6	8	9	11	12	14	17	19	20	24	25	29	30	34	35	39	42
0.05	6	8	11	13	16	18	21	23	26	28	31	35	38	42	45	49	52
0.01	-	10	13	17	20	24	27	31	34	40	43	47	52	58	63	67	72

TABLE A-12b: PROBABILITIES FOR THE SMALL-SAMPLE MANN-KENDALL TEST FOR TREND

S	n				S	n		
	4	**5**	**8**	**9**		**6**	**7**	**10**
0	0.625	0.592	0.548	0.540	**1**	0.500	0.500	0.500
2	0.375	0.408	0.452	0.460	**3**	0.360	0.386	0.431
4	0.167	0.242	0.360	0.381	**5**	0.235	0.281	0.364
6	0.042	0.117	0.274	0.306	**7**	0.136	0.191	0.300
8		0.042	0.199	0.238	**9**	0.068	0.119	0.242
10		0.0083	0.138	0.179	**11**	0.028	0.068	0.190
12			0.089	0.130	**13**	0.0083	0.035	0.146
14			0.054	0.090	**15**	0.0014	0.015	0.108
16			0.031	0.060	**17**		0.0054	0.078
18			0.016	0.038	**19**		0.0014	0.054
20			0.0071	0.022	**21**		0.00020	0.036
22			0.0028	0.012	**23**			0.023
24			0.00087	0.0063	**25**			0.014
26			0.00019	0.0029	**27**			0.0083
28			0.000025	0.0012	**29**			0.0046
30				0.00043	**31**			0.0023
32				0.00012	**33**			0.0011
34				0.000025	**35**			0.00047
36				0.0000028	**37**			0.00018
					39			0.000058
					41			0.000015
					43			0.0000028
					45			0.00000028

TABLE A-13. QUANTILES FOR THE WALD-WOLFOWITZ TEST FOR RUNS

n	*m*	$W_{0.01}$	$W_{0.05}$	$W_{0.10}$
4	**4**	-	-	3
	5	-	-	3
	6	-	3	4
	7	-	3	4
	8	-	3	4
	9	-	3	4
	10	-	4	5
	11	3	4	5
	12	3	4	5
	13	3	4	5
	14	3	5	6
	15	3	5	6
	16	4	5	6
	17	4	5	6
	18	4	5	6
	19	4	5	6

n	*m*	$W_{0.01}$	$W_{0.05}$	$W_{0.10}$
5	**5**	3	4	4
	6	3	4	4
	7	3	4	5
	8	3	4	5
	9	3	4	5
	10	4	4	5
	11	4	5	6
	12	4	5	6
	13	4	5	6
	14	5	6	6
	15	5	6	7
	16	5	6	7
	17	5	6	7
	18	5	6	7
	19	5	6	7
	20	5	6	7

n	*m*	$W_{0.01}$	$W_{0.05}$	$W_{0.10}$
6	**6**	3	4	5
	7	4	5	6
	8	4	5	6
	9	4	5	6
	10	4	6	7
	11	5	6	7
	12	5	6	7
	13	5	7	8
	14	5	7	8
	15	5	7	8
	16	6	7	8
	17	6	7	8
	18	6	7	8
	19	6	7	8
	20	6	7	8

n	*m*	$W_{0.01}$	$W_{0.05}$	$W_{0.10}$
7	**7**			
	8	4	5	6
	9	4	5	6
	10	4	6	7
	11	5	6	7
	12	5	6	7
	13	5	7	7
	14	5	7	8
	15	6	7	8
	16	6	7	8
	17	6	8	8
	18	7	8	9
	19	7	8	9
	20	7	8	9

TABLE A-13. QUANTILES FOR THE WALD-WOLFOWITZ TEST FOR RUNS (CONT.)

n	*m*	$W_{0.01}$	$W_{0.05}$	$W_{0.10}$
	20	4	5	6
8	**8**	5	6	6
	9	5	6	7
	10	5	7	7
	11	6	7	8
	12	6	7	8
	13	6	7	8
	14	6	8	8
	15	6	8	9
	16	7	8	9
	17	7	8	9
	18	7	9	10
	19	7	9	10
	20	7	9	10

n	*m*	$W_{0.01}$	$W_{0.05}$	$W_{0.10}$
9	**9**	5	6	7
	10	6	7	8
	11	6	7	8
	12	6	7	8
	13	7	8	9
	14	7	8	9
	15	7	9	9
	16	7	9	10
	17	8	9	10
	18	8	9	10
	19	8	10	11
	20	8	10	11

n	*m*	$W_{0.01}$	$W_{0.05}$	$W_{0.10}$
10	**10**	7	8	9
	11	7	8	9
	12	7	8	9
	13	7	9	10
	14	8	9	10
	15	8	9	10
	16	8	9	11
	17	8	10	11
	18	9	10	11
	19	9	11	12
	20	9	11	12

n	*m*	$W_{0.01}$	$W_{0.05}$	$W_{0.10}$
11	**11**	7	8	9
	12	7	8	9
	13	7	9	10
	14	8	9	10
	15	8	9	10
	16	8	10	11
	17	9	10	11
	18	9	10	11
	19	9	11	12
	20	9	11	12

TABLE A-13. QUANTILES FOR THE WALD-WOLFOWITZ TEST FOR RUNS (CONT.)

n	*m*	$W_{0.01}$	$W_{0.05}$	$W_{0.10}$
12	**12**	7	9	10
	13	8	10	10
	14	8	10	10
	15	8	10	11
	16	9	11	12
	17	9	11	12
	18	9	11	12
	19	10	12	13
	20	10	12	13
13	**13**	9	10	11
	14	9	11	12
	15	10	11	12
	16	10	11	12
	17	10	11	12
	18	10	11	13
	19	10	12	13
	20	11	12	13
14	**14**	9	11	12
	15	9	11	12
	16	10	11	13
	17	10	12	13
	18	10	12	13
	19	11	13	14
	20	11	13	14
15	15	8	10	11
	16	9	11	12
	17	10	12	13
	18	10	12	14
	19	11	13	15
	20	12	14	16
16	**16**	10	12	13
	17	11	13	14
17	**17**	11	13	14
	18	11	13	15
18	**18**	13	14	16
	19	13	14	16
19	**19**	14	16	17
	20	14	16	17

TABLE A-13. QUANTILES FOR THE WALD-WOLFOWITZ TEST FOR RUNS (CONT.)

n	*m*	$W_{0.01}$	$W_{0.05}$	$W_{0.10}$
	18	11	13	14
	19	12	14	15
	20	13	14	16

n	*m*	$W_{0.01}$	$W_{0.05}$	$W_{0.10}$
	19	12	14	16
	20	12	14	16

n	*m*	$W_{0.01}$	$W_{0.05}$	$W_{0.10}$
	20	14	15	16

n	*m*	$W_{0.01}$	$W_{0.05}$	$W_{0.10}$
20	**20**	14	16	17

When n or m is greater than 20 the W_p quantile is given by:

$$W_p = 1 + \frac{2mn}{m+n} + z_p\sqrt{\frac{2mn(2mn-m-n)}{(m+n)^2(m+n-1)}}$$ where z_p is the appropriate quantile from the standard normal (see Table A-1).

TABLE A-14. CRITICAL VALUES FOR THE SLIPPAGE TEST LEVEL OF SIGNIFICANCE (α) APPROXIMATELY 0.01

		***n* = size of population 1 (site)**															
		5	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
***m* = size of population 2 (background)**	**5**	5	6	6	7	8	8	9	9	10	11	11	12	12	13	14	14
	6	5	5	6	6	7	8	8	9	9	10	10	11	11	12	12	13
	7	5	5	6	6	7	7	8	8	9	9	10	10	10	11	11	12
	8	4	5	5	6	6	7	7	8	8	8	9	9	10	10	11	11
	9	4	5	5	5	6	6	7	7	8	8	8	9	9	10	10	10
	10	4	4	5	5	6	6	6	7	7	7	8	8	9	9	9	10
	11	4	4	5	5	5	6	6	6	7	7	7	8	8	9	9	9
	12	4	4	4	5	5	5	6	6	6	7	7	7	8	8	8	9
	13	4	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8
	14	4	4	4	4	5	5	5	6	6	6	7	7	7	7	8	8
	15	3	4	4	4	5	5	5	5	6	6	6	7	7	7	7	8
	16	3	4	4	4	4	5	5	5	6	6	6	6	7	7	7	7
	17	3	4	4	4	4	5	5	5	5	6	6	6	6	7	7	7
	18	3	3	4	4	4	5	5	5	5	5	6	6	6	6	7	7
	19	3	3	4	4	4	4	5	5	5	5	6	6	6	6	6	7
	20	3	3	4	4	4	4	5	5	5	5	5	6	6	6	6	7

		***n* = size of population 1 (site)**															
		25	**30**	**35**	**40**	**45**	**50**	**55**	**60**	**65**	**70**	**75**	**80**	**85**	**90**	**95**	**100**
***m* = size of population 2 (background)**	**25**	7	8	8	9	10	11	12	13	13	14	15	16	17	18	19	19
	30	6	7	7	8	9	10	10	11	12	13	13	14	15	14	16	17
	35	5	6	7	7	8	9	9	10	11	11	12	12	13	14	14	15
	40	5	6	6	7	7	8	9	9	10	10	11	11	12	12	13	13
	45	5	5	6	6	7	7	8	8	9	9	10	10	11	11	12	12
	50	5	5	5	6	6	7	7	8	8	9	9	10	10	10	11	11
	55	4	5	5	6	6	6	7	7	8	8	9	9	9	10	10	11
	60	4	5	5	5	6	6	7	7	7	8	8	8	9	9	10	10
	65	4	4	5	5	5	6	6	7	7	7	8	8	8	9	9	9
	70	4	4	5	5	5	6	6	6	7	7	7	8	8	8	9	9
	75	4	4	4	5	5	5	6	6	6	7	7	7	8	8	8	9
	80	4	4	4	5	5	5	6	6	6	6	7	7	7	8	8	8
	85	4	4	4	4	5	5	5	6	6	6	6	7	7	7	8	8
	90	3	4	4	4	5	5	5	5	6	6	6	6	7	7	7	8
	95	3	4	4	4	4	4	5	5	6	6	6	6	7	7	7	7
	100	3	4	4	4	4	5	5	5	5	6	6	6	6	7	7	7

TABLE A-14. CRITICAL VALUES FOR THE SLIPPAGE TEST (CONT.) LEVEL OF SIGNIFICANCE (α) APPROXIMATELY 0.05

		***n* = size of population 1 (site)**															
***m* = size of population 2 (background)**		**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
	5	4	4	5	5	6	6	7	7	8	8	9	9	9	10	10	11
	6	4	4	4	5	5	6	6	6	7	7	8	8	8	9	9	10
	7	3	4	4	5	5	5	6	6	6	7	7	7	8	8	8	9
	8	3	4	4	4	5	5	5	6	6	6	6	7	7	7	8	8
	9	3	3	4	4	4	5	5	5	5	6	6	6	7	7	7	7
	10	3	3	4	4	4	4	5	5	5	5	6	6	6	6	7	7
	11	3	3	3	4	4	4	4	5	5	5	5	6	6	6	6	7
	12	3	3	3	3	4	4	4	4	5	5	5	5	6	6	6	6
	13	3	3	3	3	4	4	4	4	4	5	5	5	5	6	6	6
	14	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5	6
	15	3	3	3	3	3	4	4	4	4	4	4	5	5	5	5	5
	16	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
	17	2	3	3	3	3	3	4	4	4	4	4	4	5	5	5	5
	18	2	3	3	3	3	3	3	4	4	4	4	4	4	5	5	5
	19	2	3	3	3	3	3	3	3	4	4	4	4	4	4	5	5
	20	2	2	3	3	3	3	3	3	4	4	4	4	4	4	4	5

		***n* = size of population 1 (site)**															
***m* = size of population 2 (background)**		**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**	**65**	**70**	**75**	**80**	**85**	**90**	**95**	**100**
	25	5	5	6	6	7	8	8	9	9	10	10	11	12	12	13	13
	30	4	5	5	6	6	7	7	8	8	9	9	9	10	10	11	11
	35	4	4	5	5	6	6	6	7	7	8	8	8	9	9	10	10
	40	4	4	4	5	5	5	6	6	7	7	7	8	8	8	9	9
	45	3	4	4	4	5	5	5	6	6	6	7	7	7	8	8	8
	50	3	3	4	4	4	5	5	5	6	6	5	7	7	7	7	8
	55	3	3	4	4	4	4	5	5	5	6	6	6	6	7	7	7
	60	3	3	3	4	4	4	4	5	5	5	5	6	6	6	6	7
	65	3	3	3	4	4	4	4	5	5	5	5	5	6	6	6	6
	70	3	3	3	3	4	4	4	4	5	5	5	5	5	6	6	6
	75	3	3	3	3	4	4	4	4	4	5	5	5	5	5	6	6
	80	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5	6
	85	3	3	3	3	3	3	3	4	4	4	4	5	5	5	5	5
	90	2	3	3	3	3	3	4	4	4	4	4	4	5	5	5	5
	95	2	3	3	3	3	3	3	4	4	4	4	4	4	5	5	5
	100	2	3	3	3	3	3	3	4	4	4	4	4	4	4	5	5

TABLE A-14. CRITICAL VALUES FOR THE SLIPPAGE TEST (CONT.) LEVEL OF SIGNIFICANCE (α) APPROXIMATELY 0.10

		***n* = size of population 1 (site)**															
		5	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
***m* = size of population 2 (background)**	**5**	3	4	4	4	5	5	6	6	6	7	7	7	8	8	9	9
	6	3	3	4	4	4	5	5	5	6	6	6	7	7	7	8	8
	7	3	3	3	4	4	4	5	5	5	5	6	6	6	7	7	7
	8	3	3	3	4	4	4	4	5	5	5	5	6	6	6	6	7
	9	3	3	3	3	4	4	4	4	4	5	5	5	5	6	6	6
	10	2	3	3	3	3	4	4	4	4	4	5	5	5	5	5	6
	11	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
	12	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5
	13	2	2	3	3	3	3	3	3	4	4	4	4	4	4	5	5
	14	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4	5
	15	2	2	2	3	3	3	3	3	3	3	4	4	4	4	4	4
	16	2	2	2	3	3	3	3	3	3	3	4	4	4	4	4	4
	17	2	2	2	2	3	3	3	3	3	3	3	4	4	4	4	4
	18	2	2	2	2	3	3	3	3	3	3	3	3	4	4	4	4
	19	2	2	2	2	2	3	3	3	3	3	3	3	3	4	4	4
	20	2	2	2	2	2	3	3	3	3	3	3	3	3	3	4	4

		***n* = size of population 1 (site)**															
		25	**30**	**35**	**40**	**45**	**50**	**55**	**60**	**65**	**70**	**75**	**80**	**85**	**90**	**95**	**100**
***m* = size of population 2 (background)**	**25**	4	4	5	5	6	6	6	7	7	8	8	9	9	9	10	10
	30	3	4	4	4	5	5	6	6	6	7	7	7	8	8	9	9
	35	3	3	4	4	4	5	5	5	6	6	6	7	7	7	8	8
	40	3	3	3	4	4	4	5	5	5	5	6	6	6	7	7	7
	45	3	3	3	4	4	4	4	5	5	5	5	6	6	6	6	7
	50	3	3	3	3	4	4	4	4	4	5	5	5	5	6	6	6
	55	2	3	3	3	3	4	4	4	4	4	5	5	5	5	5	6
	60	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
	65	2	2	3	3	3	3	3	4	4	4	4	4	4	5	5	5
	70	2	2	3	3	3	3	3	3	4	4	4	4	4	4	5	5
	75	2	2	2	3	3	3	3	2	3	4	4	4	4	4	4	5
	80	2	2	2	3	3	3	3	3	3	3	4	4	4	4	4	4
	85	2	2	2	3	3	3	3	3	3	3	4	4	4	4	4	4
	90	2	2	2	2	3	3	3	3	3	3	3	4	4	4	4	4
	95	2	2	2	2	3	3	3	3	3	3	3	3	4	4	4	4
	100	2	2	2	2	2	3	3	3	3	3	3	3	3	4	4	4

TABLE A-15. DUNNETT'S TEST (ONE TAILED)

Degrees of Freedom	α	Total Number of Investigated Groups (k – 1)											
		2	3	4	5	6	7	8	9	10	12	14	16
2	.05	3.80	4.34	4.71	5.08	5.24	5.43	5.60	5.75	5.88	6.11	6.29	6.45
	.10	2.54	2.92	3.20	3.40	3.57	3.71	3.83	3.94	4.03	4.19	4.32	4.44
3	.05	2.94	3.28	3.52	3.70	3.85	3.97	4.08	4.17	4.25	4.39	4.51	4.61
	.10	2.13	2.41	2.61	2.76	2.87	2.97	3.06	3.13	3.20	3.31	3.41	3.49
4	.05	2.61	2.88	3.08	3.22	3.34	3.44	3.52	3.59	3.66	3.77	3.86	3.94
	.10	1.96	2.20	2.37	2.50	2.60	2.68	2.75	2.82	2.87	2.97	3.05	3.11
5	.05	2.44	2.68	2.85	2.98	3.08	3.16	3.24	3.30	3.36	3.45	3.53	3.60
	.10	1.87	2.09	2.24	2.36	2.45	2.53	2.59	2.65	2.70	2.78	2.86	2.92
6	.05	2.34	2.56	2.71	2.83	2.92	3.00	3.06	3.12	3.17	3.26	3.33	3.48
	.10	1.82	2.02	2.17	2.27	2.36	2.43	2.49	2.54	2.59	2.67	2.74	2.79
7	.05	2.27	2.48	2.82	2.73	2.81	2.89	2.95	3.00	3.05	3.13	3.20	3.26
	.10	1.78	1.98	2.11	2.22	2.30	2.37	2.42	2.47	2.52	2.59	2.66	2.71
8	.05	2.22	2.42	2.55	2.66	2.74	2.81	2.87	2.92	2.96	3.04	3.11	3.16
	.10	1.75	1.94	2.08	2.17	2.25	2.32	2.38	2.42	2.47	2.54	2.60	2.65
9	.05	2.18	2.37	2.50	2.60	2.68	2.75	2.81	2.86	2.90	2.97	3.04	3.09
	.10	1.73	1.92	2.05	2.14	2.22	2.28	2.34	2.39	2.43	2.50	2.56	2.61
10	.05	2.15	2.34	2.47	2.56	2.64	2.70	2.76	2.81	2.85	2.92	2.98	3.03
	.10	1.71	1.90	2.02	2.12	2.19	2.26	2.31	2.35	2.40	2.46	2.52	2.57
12	.05	2.11	2.29	2.41	2.50	2.58	2.64	2.69	2.74	2.78	2.84	2.90	2.95
	.10	1.69	1.87	1.99	2.08	2.16	2.22	2.27	2.31	2.35	2.42	2.47	2.52
16	.05	2.06	2.23	2.34	2.43	2.50	2.56	2.61	2.65	2.69	2.75	2.81	2.85
	.10	1.66	1.83	1.95	2.04	2.11	2.17	2.22	2.26	2.30	2.36	2.41	2.46

TABLE A-15. DUNNETT'S TEST (ONE TAILED) (CONT.)

Degrees of Freedom	α	Total Number of Investigated Groups (k – 1)											
		2	3	4	5	6	7	8	9	10	12	14	16
20	.05	2.03	2.19	2.30	2.39	2.46	2.51	2.56	2.60	2.64	2.70	2.75	2.80
	.10	1.64	1.81	1.93	2.01	2.08	2.14	2.19	2.23	2.26	2.33	2.38	2.42
24	.05	2.01	2.17	2.28	2.36	2.43	2.48	2.53	2.57	2.60	2.66	2.72	2.76
	.10	1.63	2.80	1.91	2.00	2.06	2.12	2.17	2.21	2.24	2.30	2.35	2.40
30	.05	1.99	2.15	2.25	2.34	2.40	2.45	2.50	2.54	2.57	2.63	2.68	2.72
	.10	1.62	1.79	1.90	1.98	2.05	2.10	2.15	2.19	2.22	2.28	2.33	2.37
40	.05	1.97	2.13	2.23	2.31	2.37	2.42	2.47	2.51	2.54	2.60	2.65	2.69
	.10	1.61	1.77	1.88	1.96	2.03	2.08	2.13	2.17	2.20	2.26	2.31	2.35
50	.05	1.96	2.11	2.22	2.29	2.32	2.41	2.45	2.49	2.52	2.58	2.63	2.67
	.10	1.61	1.77	1.88	1.96	2.02	2.07	2.12	2.16	2.19	2.25	2.30	2.34
60	.05	1.95	2.10	2.21	2.28	2.34	2.40	2.44	2.48	2.51	2.57	2.61	2.65
	.10	1.60	1.76	1.87	1.95	2.01	2.06	2.11	2.15	2.18	2.24	2.29	2.33
70	.05	1.95	2.10	2.21	2.28	2.34	2.40	2.44	2.48	2.51	2.56	2.61	2.65
	.10	1.60	1.76	1.87	1.95	2.01	2.06	2.11	2.15	2.18	2.24	2.29	2.33
80	.05	1.94	2.10	2.20	2.28	2.34	2.39	2.43	2.47	2.50	2.55	2.60	2.64
	.10	1.60	1.76	1.87	1.95	2.01	2.06	2.10	2.15	2.18	2.23	2.28	2.32
90	.05	1.94	2.09	2.20	2.27	2.33	2.39	2.43	2.47	2.50	2.55	2.60	2.63
	.10	1.60	1.76	1.86	1.94	2.00	2.06	2.10	2.14	2.17	2.23	2.28	2.31
100	.05	1.93	2.08	2.18	2.27	2.33	2.38	2.42	2.46	2.49	2.54	2.59	2.63
	.10	1.59	1.75	1.85	1.93	1.99	2.05	2.09	2.14	2.17	2.22	2.27	2.31
120	.05	1.93	2.08	2.18	2.26	2.32	2.37	2.41	2.45	2.48	2.53	2.58	2.62
	.10	1.59	1.75	1.85	1.93	1.99	2.05	2.09	2.13	2.16	2.22	2.27	2.31
∞	.05	1.92	2.06	2.16	2.23	2.29	2.34	2.38	2.42	2.45	2.50	2.55	2.58
	.10	1.58	1.73	1.84	1.92	1.98	2.03	2.07	2.11	2.14	2.20	2.24	2.28

TABLE A-16. APPROXIMATE α-LEVEL CRITICAL POINTS FOR RANK VON NEUMANN RATIO TEST

	α	
n	**0.05**	**0.10**
4		
5	0.70	0.60
6	0.80	0.97
7	0.86	1.11
8	0.93	1.14
9	0.98	1.18
10	1.04	1.23
11	1.08	1.26
12	1.11	1.29
13	1.14	1.32
14	1.17	1.34
15	1.19	1.36
16	1.21	1.38
17	1.24	1.40
18	1.26	1.41
19	1.27	1.43
20	1.29	1.44
21	1.31	1.45
22	1.32	1.46
23	1.33	1.48
24	1.35	1.49
25	1.36	1.50
26	1.37	1.51
27	1.38	1.51
28	1.39	1.52
29	1.40	1.53
30	1.41	1.54
32	1.43	1.55
34	1.45	1.57
36	1.46	1.58
38	1.48	1.59
40	1.49	1.60
42	1.50	1.61
44	1.51	1.62
46	1.52	1.63
48	1.53	1.63
50	1.54	1.64
55	1.56	1.66
60	1.58	1.67
65	1.60	1.68
70	1.61	1.70
75	1.62	1.71
80	1.64	1.71
85	1.65	1.72
90	1.66	1.73
95	1.66	1.74
100	1.67	1.74

TABLE A-17. VALUES OF $H_{1-\alpha} = H_{0.90}$ FOR COMPUTING A ONE-SIDED UPPER 90% CONFIDENCE LIMIT ON A LOGNORMAL MEAN

s_y	n									
	3	**5**	**7**	**10**	**12**	**15**	**21**	**31**	**51**	**101**
0.10	1.686	1.438	1.381	1.349	1.338	1.328	1.317	1.308	1.301	1.295
0.20	1.885	1.522	1.442	1.396	1.380	1.365	1.348	1.335	1.324	1.314
0.30	2.156	1.627	1.517	1.453	1.432	1.411	1.388	1.370	1.354	1.339
0.40	2.521	1.755	1.607	1.523	1.494	1.467	1.437	1.412	1.390	1.371
0.50	2.990	1.907	1.712	1.604	1.567	1.532	1.494	1.462	1.434	1.409
0.60	3.542	2.084	1.834	1.696	1.650	1.606	1.558	1.519	1.485	1.454
0.70	4.136	2.284	1.970	1.800	1.743	1.690	1.631	1.583	1.541	1.504
0.80	4.742	2.503	2.119	1.914	1.845	1.781	1.710	1.654	1.604	1.560
0.90	5.349	2.736	2.280	2.036	1.955	1.880	1.797	1.731	1.672	1.621
1.00	5.955	2.980	2.450	2.167	2.073	1.985	1.889	1.812	1.745	1.686
1.25	7.466	3.617	2.904	2.518	2.391	2.271	2.141	2.036	1.946	1.866
1.50	8.973	4.276	3.383	2.896	2.733	2.581	2.415	2.282	2.166	2.066
1.75	10.48	4.944	3.877	3.289	3.092	2.907	2.705	2.543	2.402	2.279
2.00	11.98	5.619	4.380	3.693	3.461	3.244	3.005	2.814	2.648	2.503
2.50	14.99	6.979	5.401	4.518	4.220	3.938	3.629	3.380	3.163	2.974
3.00	18.00	8.346	6.434	5.359	4.994	4.650	4.270	3.964	3.697	3.463
3.50	21.00	9.717	7.473	6.208	5.778	5.370	4.921	4.559	4.242	3.965
4.00	24.00	11.09	8.516	7.062	6.566	6.097	5.580	5.161	4.796	4.474
4.50	27.01	12.47	9.562	7.919	7.360	6.829	6.243	5.769	5.354	4.989
5.00	30.01	13.84	10.61	8.779	8.155	7.563	6.909	6.379	5.916	5.508
6.00	36.02	16.60	12.71	10.50	9.751	9.037	8.248	7.607	7.048	6.555
7.00	42.02	19.35	14.81	12.23	11.35	10.52	9.592	8.842	8.186	7.607
8.00	48.03	22.11	16.91	13.96	12.96	12.00	10.94	10.08	9.329	8.665
9.00	54.03	24.87	19.02	15.70	14.56	13.48	12.29	11.32	10.48	9.725
10.00	60.04	27.63	21.12	17.43	16.17	14.97	13.64	12.56	11.62	10.79

TABLE A-17. VALUES OF $H_{\alpha} = H_{0.10}$ FOR COMPUTING A ONE-SIDED LOWER 10% CONFIDENCE LIMIT ON A LOGNORMAL MEAN

s_y	n									
	3	**5**	**7**	**10**	**12**	**15**	**21**	**31**	**51**	**101**
0.10	-1.431	-1.320	-1.296	-1.285	-1.281	-1.279	-1.277	-1.277	-1.278	-1.279
0.20	-1.350	-1.281	-1.268	-1.266	-1.266	-1.266	-1.268	-1.272	-1.275	-1.280
0.30	-1.289	-1.252	-1.250	-1.254	-1.257	-1.260	-1.266	-1.272	-1.280	-1.287
0.40	-1.245	-1.233	-1.239	-1.249	-1.254	-1.261	-1.270	-1.279	-1.289	-1.301
0.50	-1.213	-1.221	-1.234	-1.250	-1.257	-1.266	-1.279	-1.291	-1.304	-1.319
0.60	-1.190	-1.215	-1.235	-1.256	-1.266	-1.277	-1.292	-1.307	-1.324	-1.342
0.70	-1.176	-1.215	-1.241	-1.266	-1.278	-1.292	-1.310	-1.329	-1.349	-1.370
0.80	-1.168	-1.219	-1.251	-1.280	-1.294	-1.311	-1.332	-1.354	-1.377	-1.403
0.90	-1.165	-1.227	-1.264	-1.298	-1.314	-1.333	-1.358	-1.383	-1.409	-1.439
1.00	-1.166	-1.239	-1.281	-1.320	-1.337	-1.358	-1.387	-1.414	-1.445	-1.478
1.25	-1.184	-1.280	-1.334	-1.384	-1.407	-1.434	-1.470	-1.507	-1.547	-1.589
1.50	-1.217	-1.334	-1.400	-1.462	-1.491	-1.523	-1.568	-1.613	-1.663	-1.716
1.75	-1.260	-1.398	-1.477	-1.551	-1.585	-1.624	-1.677	-1.732	-1.790	-1.855
2.00	-1.310	-1.470	-1.562	-1.647	-1.688	-1.733	-1.795	-1.859	-1.928	-2.003
2.50	-1.426	-1.634	-1.751	-1.862	-1.913	-1.971	-2.051	-2.133	-2.223	-2.321
3.00	-1.560	-1.817	-1.960	-2.095	-2.157	-2.229	-2.326	-2.427	-2.536	-2.657
3.50	-1.710	-2.014	-2.183	-2.341	-2.415	-2.499	-2.615	-2.733	-2.864	-3.007
4.00	-1.871	-2.221	-2.415	-2.596	-2.681	-2.778	-2.913	-3.050	-3.200	-3.366
4.50	-2.041	-2.435	-2.653	-2.858	-2.955	-3.064	-3.217	-3.372	-3.542	-3.731
5.00	-2.217	-2.654	-2.897	-3.126	-3.233	-3.356	-3.525	-3.698	-3.889	-4.100
6.00	-2.581	-3.104	-3.396	-3.671	-3.800	-3.949	-4.153	-4.363	-4.594	-4.849
7.00	-2.955	-3.564	-3.904	-4.226	-4.377	-4.549	-4.790	-5.037	-5.307	-5.607
8.00	-3.336	-4.030	-4.418	-4.787	-4.960	-5.159	-5.433	-5.715	-6.026	-6.370
9.00	-3.721	-4.500	-4.937	-5.352	-5.547	-5.771	-6.080	-6.399	-6.748	-7.136
10.00	-4.109	-4.973	-5.459	-5.920	-6.137	-6.386	-6.730	-7.085	-7.474	-7.906

TABLE A-17. VALUES OF $H_{1-\alpha} = H_{0.95}$ FOR COMPUTING A ONE-SIDED UPPER 95% CONFIDENCE LIMIT ON A LOGNORMAL MEAN

s_y	n									
	3	**5**	**7**	**10**	**12**	**15**	**21**	**31**	**51**	**101**
0.10	2.750	2.035	1.886	1.802	1.775	1.749	1.722	1.701	1.684	1.670
0.20	3.295	2.198	1.992	1.881	1.843	1.809	1.771	1.742	1.718	1.697
0.30	4.109	2.402	2.125	1.977	1.927	1.882	1.833	1.793	1.761	1.733
0.40	5.220	2.651	2.282	2.089	2.026	1.968	1.905	1.856	1.813	1.777
0.50	6.495	2.947	2.465	2.220	2.141	2.068	1.989	1.928	1.876	1.830
0.60	7.807	3.287	2.673	2.368	2.271	2.181	2.085	2.010	1.946	1.891
0.70	9.120	3.662	2.904	2.532	2.414	2.306	2.191	2.102	2.025	1.960
0.80	10.43	4.062	3.155	2.710	2.570	2.443	2.307	2.202	2.112	2.035
0.90	11.74	4.478	3.420	2.902	2.738	2.589	2.432	2.310	2.206	2.117
1.00	13.05	4.905	3.698	3.103	2.915	2.744	2.564	2.423	2.306	2.205
1.25	16.33	6.001	4.426	3.639	3.389	3.163	2.923	2.737	2.580	2.447
1.50	19.60	7.120	5.184	4.207	3.896	3.612	3.311	3.077	2.881	2.713
1.75	22.87	8.250	5.960	4.795	4.422	4.081	3.719	3.437	3.200	2.997
2.00	26.14	9.387	6.747	5.396	4.962	4.564	4.141	3.812	3.533	3.295
2.50	32.69	11.67	8.339	6.621	6.067	5.557	5.013	4.588	4.228	3.920
3.00	39.23	13.97	9.945	7.864	7.191	6.570	5.907	5.388	4.947	4.569
3.50	45.77	16.27	11.56	9.118	8.326	7.596	6.815	6.201	5.681	5.233
4.00	52.31	18.58	13.18	10.38	9.469	8.630	7.731	7.024	6.424	5.908
4.50	58.85	20.88	14.80	11.64	10.62	9.669	8.652	7.854	7.174	6.590
5.00	65.39	23.19	16.43	12.91	11.77	10.71	9.579	8.688	7.929	7.277
6.00	78.47	27.81	19.68	15.45	14.08	12.81	11.44	10.36	9.449	8.661
7.00	91.55	32.43	22.94	18.00	16.39	14.90	13.31	12.05	10.98	10.05
8.00	104.6	37.06	26.20	20.55	18.71	17.01	15.18	13.74	12.51	11.45
9.00	117.7	41.68	29.64	23.10	21.03	19.11	17.05	15.43	14.05	12.85
10.00	130.8	46.31	32.73	25.66	23.35	21.22	18.93	17.13	15.59	14.26

TABLE A-17. VALUES OF $H_{\alpha} = H_{0.05}$ FOR COMPUTING A ONE-SIDED LOWER 5% CONFIDENCE LIMIT ON A LOGNORMAL MEAN

s_y	n									
	3	**5**	**7**	**10**	**12**	**15**	**21**	**31**	**51**	**101**
0.10	-2.130	-1.806	-1.731	-1.690	-1.677	-1.666	-1.655	-1.648	-1.644	-1.642
0.20	-1.949	-1.729	-1.678	-1.653	-1.646	-1.640	-1.636	-1.636	-1.637	-1.641
0.30	-1.816	-1.669	-1.639	-1.627	-1.625	-1.625	-1.627	-1.632	-1.638	-1.648
0.40	-1.717	-1.625	-1.611	-1.611	-1.613	-1.617	-1.625	-1.635	-1.647	-1.662
0.50	-1.644	-1.594	-1.594	-1.603	-1.609	-1.618	-1.631	-1.646	-1.663	-1.683
0.60	-1.589	-1.573	-1.584	-1.602	-1.612	-1.625	-1.643	-1.662	-1.685	-1.711
0.70	-1.549	-1.560	-1.582	-1.608	-1.622	-1.638	-1.661	-1.686	-1.713	-1.744
0.80	-1.521	-1.555	-1.586	-1.620	-1.636	-1.656	-1.685	-1.714	-1.747	-1.783
0.90	-1.502	-1.556	-1.595	-1.637	-1.656	-1.680	-1.713	-1.747	-1.785	-1.826
1.00	-1.490	-1.562	-1.610	-1.658	-1.681	-1.707	-1.745	-1.784	-1.827	-1.874
1.25	-1.486	-1.596	-1.662	-1.727	-1.758	-1.793	-1.842	-1.893	-1.949	-2.012
1.50	-1.508	-1.650	-1.733	-1.814	-1.853	-1.896	-1.958	-2.020	-2.091	-2.169
1.75	-1.547	-1.719	-1.819	-1.916	-1.962	-2.015	-2.088	-2.164	-2.247	-2.341
2.00	-1.598	-1.799	-1.917	-2.029	-2.083	-2.144	-2.230	-2.318	-2.416	-2.526
2.50	-1.727	-1.986	-2.138	-2.283	-2.351	-2.430	-2.540	-2.654	-2.780	-2.921
3.00	-1.880	-2.199	-2.384	-2.560	-2.644	-2.740	-2.874	-3.014	-3.169	-3.342
3.50	-2.051	-2.429	-2.647	-2.855	-2.953	-3.067	-3.226	-3.391	-3.574	-3.780
4.00	-2.237	-2.672	-2.922	-3.161	-3.275	-3.406	-3.589	-3.779	-3.990	-4.228
4.50	-2.434	-2.924	-3.206	-3.476	-3.605	-3.753	-3.960	-4.176	-4.416	-4.685
5.00	-2.638	-3.183	-3.497	-3.798	-3.941	-4.107	-4.338	-4.579	-4.847	-5.148
6.00	-3.062	-3.715	-4.092	-4.455	-4.627	-4.827	-5.106	-5.397	-5.721	-6.086
7.00	-3.499	-4.260	-4.699	-5.123	-5.325	-5.559	-5.886	-6.227	-6.608	-7.036
8.00	-3.945	-4.812	-5.315	-5.800	-6.031	-6.300	-6.674	-7.066	-7.502	-7.992
9.00	-4.397	-5.371	-5.936	-6.482	-6.742	-7.045	-7.468	-7.909	-8.401	-8.953
10.00	-4.852	-5.933	-6.560	-7.168	-7.458	-7.794	-8.264	-8.755	-9.302	-9.918

TABLE A-18. CRITICAL VALUES FOR THE SIGN TEST

The table values are such that for order statistics $x_{[lower]}$ and $x_{[upper]}$, $P\left(x_{[lower]} < \text{true median} < x_{[upper]}\right) \geq 1 - \alpha$. Note that the significance levels are for two-sided tests. For one-sided test, divide the significance level in half.

n	α									
	0.20		0.10		0.05		0.02		0.01	
	lower	upper	lower	upper	lower	upper	lower	upper	lower	upper
5	1	5	1	5	-	-	-	-	-	-
6	1	6	1	6	1	6	-	-	-	-
7	2	6	1	7	1	7	1	7	-	-
8	2	7	2	7	1	8	1	8	1	8
9	3	7	2	8	2	8	1	9	1	9
10	3	8	2	9	2	9	1	10	1	10
11	3	9	3	9	2	10	2	10	1	11
12	4	9	3	10	3	10	2	11	2	11
13	4	10	4	10	3	11	2	12	2	12
14	5	10	4	11	3	12	3	12	2	13
15	5	11	4	12	4	12	3	13	3	13
16	5	12	5	12	4	13	3	14	3	14
17	6	12	5	13	5	13	4	14	3	15
18	6	13	6	13	5	14	4	15	4	15
19	7	13	6	14	5	15	5	15	4	16
20	7	14	6	15	6	15	5	16	4	17
21	8	14	7	15	6	16	5	17	5	17
22	8	15	7	16	6	17	6	17	5	18
23	8	16	8	16	7	17	6	18	5	19
24	9	16	8	17	7	18	6	19	6	19
25	9	17	8	18	8	18	7	19	6	20
26	10	17	9	18	8	19	7	20	7	20
27	10	18	9	19	8	20	8	20	7	21
28	11	18	10	19	9	20	8	21	7	22
29	11	19	10	20	9	21	8	22	8	22
30	11	20	11	20	10	21	9	22	8	23
31	12	20	11	21	10	22	9	23	8	24
32	12	21	11	22	10	23	9	24	9	24
33	13	21	12	22	11	23	10	24	9	25
34	13	22	12	23	11	24	10	25	10	25
35	14	22	13	23	12	24	11	25	10	26
36	14	23	13	24	12	25	11	26	10	27
37	15	23	14	24	13	25	11	27	11	27
38	15	24	14	25	13	26	12	27	11	28
39	16	24	14	26	13	27	12	28	12	28
40	16	25	15	26	14	27	13	28	12	29

TABLE A-19. CRITICAL VALUES FOR THE QUANTILE TEST

m is the number of background samples, *n* is the number of site samples, and *c* is the number values larger than the b_1^{th} quantile.

If *s* is greater than or equal to the table value, q_α, then reject the null hypothesis of no difference at the given significance level.

			N																														
			4			**5**			**6**			**7**			**8**			**9**			**10**			**11**			**12**		
			0.50	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**
m	**4**	***c***	*4*	*2*	*1*	*4*	*2*	*1*	*5*	*3*	*1*	*5*	*3*	*1*	*6*	*3*	*2*	*6*	*3*	*2*	*7*	*4*	*2*	*7*	*4*	*2*	*8*	*4*	*2*
		0.10																						7					
		0.05	4			4			5			5			6			6			7						8		
		0.01																											
	5	***c***	*4*	*2*	*1*	*5*	*3*	*1*	*5*	*3*	*1*	*6*	*3*	*2*	*6*	*3*	*2*	*7*	*4*	*2*	*7*	*4*	*2*	*8*	*4*	*2*	*8*	*4*	*2*
		0.10					3																						
		0.05							5						6			7			7			8			8		
		0.01	4			5						6																	
	6	***c***	*5*	*3*	*1*	*5*	*3*	*1*	*6*	*3*	*2*	*6*	*3*	*2*	*7*	*4*	*2*	*7*	*4*	*2*	*8*	*4*	*2*	*8*	*4*	*2*	*9*	*5*	*2*
		0.10				4	3			3		5			6	4		6	4		7			7			8	5	
		0.05	4	3					5																				
		0.01				5			6			6			7			7			8			8			9		
	7	***c***	*5*	*3*	*1*	*6*	*3*	*2*	*6*	*3*	*2*	*7*	*4*	*2*	*7*	*4*	*2*	*8*	*4*	*2*	*8*	*4*	*2*	*9*	*5*	*2*	*9*	*5*	*2*
		0.10								3						4			4			4			5			5	
		0.05	4	3			3		5			6	4		6			7			7			8			8		
		0.01				5			6			7			7			8			8			9			9		
	8	***c***	*6*	*3*	*2*	*6*	*3*	*2*	*7*	*4*	*2*	*7*	*4*	*2*	*8*	*4*	*2*	*8*	*4*	*2*	*9*	*5*	*2*	*9*	*5*	*2*	*10*	*5*	*2*
		0.10			2	4			5						6				4		7						8	5	
		0.05	4	3			3			4			4			4		7				5		8	5				
		0.01				5			6			6			7			8			8			9			9		
	9	***c***	*6*	*3*	*2*	*7*	*4*	*2*	*7*	*4*	*2*	*8*	*4*	*2*	*8*	*4*	*2*	*9*	*5*	*2*	*9*	*5*	*2*	*10*	*5*	*2*	*10*	*5*	*2*
		0.10			2		3														7						8		
		0.05	4	3		5			5	4		6	4		6	4		7	5			5		8	5			5	
		0.01					4		6			7			7			8			8			9			9		
	10	***c***	*7*	*4*	*2*	*7*	*4*	*2*	*8*	*4*	*2*	*8*	*4*	*2*	*9*	*5*	*2*	*9*	*5*	*2*	*10*	*5*	*2*	*10*	*5*	*2*	*11*	*6*	*3*
		0.10			2		3	2	5						6	4					7						8		
		0.05	4	3								6	4					7	5		8	5		8	5		9	6	
		0.01		4		5	4		6	4		7			7	5		8			9			9			10		
	11	***c***	*7*	*4*	*2*	*8*	*4*	*2*	*8*	*4*	*2*	*9*	*5*	*2*	*9*	*5*	*2*	*10*	*5*	*2*	*10*	*5*	*2*	*11*	*6*	*3*	*11*	*6*	*3*
		0.10			2		3	2		3					6	4			4		7				5		8	5	
		0.05	4	3		5			5			6	4					7				5		8					
		0.01		4			4		6	4		7	5		7	5		8	5		8			9	6		9	6	
	12	***c***	*8*	*4*	*2*	*8*	*4*	*2*	*9*	*5*	*2*	*9*	*5*	*2*	*10*	*5*	*2*	*10*	*5*	*2*	*11*	*6*	*3*	*11*	*6*	*3*	*12*	*6*	*3*
		0.10					3	2	5		2				6	4			4		7		3		5	3		5	
		0.05	4	3	2					4		6	4					7			8	5		8			9		
		0.01		4		5	4		6	5		7	5		7	5		8	5		9	6		9	6		10	6	

TABLE A-19. CRITICAL VALUES FOR THE QUANTILE TEST (CONT.)

m is the number of background samples, *n* is the number of site samples, and *c* is the number values larger than the b_1^{th} quantile.
If *s* is greater than or equal to the table value, q_α, then reject the null hypothesis of no difference at the given significance level.

			n																														
			4			**5**			**6**			**7**			**8**			**9**			**10**			**11**			**12**		
			0.50	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**
m	**13**	***c***	*8*	*4*	*2*	*9*	*5*	*2*	*9*	*5*	*2*	*10*	*5*	*2*	*10*	*5*	*2*	*11*	*6*	*3*	*11*	*6*	*3*	*12*	*6*	*3*	*12*	*6*	*3*
		0.10					3	2			2				6					3	7		3			3	8		3
		0.05	4	3	2	5			5	4		6	4			4		7	5		8	5		8	5		9	5	
		0.01		4			4		6	5		7	5		7	5		8	6		9	6		9	6		10	6	
	14	***c***	*9*	*5*	*2*	*9*	*5*	*2*	*10*	*5*	*2*	*10*	*5*	*2*	*11*	*6*	*3*	*11*	*6*	*3*	*12*	*6*	*3*	*12*	*6*	*3*	*13*	*7*	*3*
		0.10					3	2	5		2			2	6	4							3			3			3
		0.05	4	3	2	5				4		6	4		7	5	3	7	5	3	8	5		8	5		9	6	
		0.01		4			4		6	5		7	5		8	6		8	6		9	6		9	6		10	7	
	15	***c***	*9*	*5*	*2*	*10*	*5*	*2*	*10*	*5*	*2*	*11*	*6*	*3*	*11*	*6*	*3*	*12*	*6*	*3*	*12*	*6*	*3*	*13*	*7*	*3*	*13*	*7*	*3*
		0.10					3	2	5		2		4		6	4					7		3	8	5	3	8		3
		0.05	4	3	2	5				4		6		3			3	7	5	3	8	5			6		9	6	
		0.01		4			4		6	5		7	5		7	5		8	6		9	6		9	7		10	7	
	16	***c***	*10*	*5*	*2*	*10*	*5*	*2*	*11*	*6*	*3*	*11*	*6*	*3*	*12*	*6*	*3*	*12*	*6*	*3*	*13*	*7*	*3*	*13*	*7*	*3*	*14*	*7*	*3*
		0.10					3		5						6	4			4			5			5	3		5	3
		0.05	4	3	2	5		2		4	3	6	4	3	7		3	7	5	3	8		3	8			9	6	
		0.01		4			4		6	5		7	5		8	5		8	6		9	6		9	6		10	7	
	17	***c***	*10*	*5*	*2*	*11*	*6*	*3*	*11*	*6*	*3*	*12*	*6*	*3*	*12*	*6*	*3*	*13*	*7*	*3*	*13*	*7*	*3*	*14*	*7*	*3*	*14*	*7*	*3*
		0.10							5						6	4								8	5		8	5	3
		0.05	4	3	2	5				4	3	6	4	3			3	7	5	3	8	5	3			3	9	6	
		0.01		4			4	3	6	5		7	5		7	5		8	6		9	6		9	6		10	7	
	18	***c***	*11*	*6*	*3*	*11*	*6*	*3*	*12*	*6*	*3*	*12*	*6*	*3*	*13*	*7*	*3*	*13*	*7*	*3*	*14*	*7*	*3*	*14*	*7*	*3*	*15*	*8*	*3*
		0.10			2		3	3	5																5				3
		0.05	4	3		5				4		6	4	3	7	5	3	7	5	3	8	5	3	8		3	9	6	
		0.01		4	3		4		6	5	3	7	5		8	6		8	6		9	6		9	6		10	7	
	19	***c***	*11*	*6*	*3*	*12*	*6*	*3*	*12*	*6*	*3*	*13*	*7*	*3*	*13*	*7*	*3*	*14*	*7*	*3*	*14*	*7*	*3*	*15*	*8*	*3*	*15*	*8*	*3*
		0.10			2		3	2	5				4		6	4		7			7			8	5				
		0.05	4	3		5				4		6		3	7	5	3		5	3	8	5	3	9	6	3	9	6	3
		0.01		4	3		4	3	6	5	3	7	5		8	6		8	6		9	6		10	7		10	7	
	20	***c***	*12*	*6*	*3*	*12*	*6*	*3*	*13*	*7*	*3*	*13*	*7*	*3*	*14*	*7*	*3*	*14*	*7*	*3*	*15*	*8*	*3*	*15*	*8*	*3*	*16*	*8*	*4*
		0.10			2		3	2	5							4								8	5				
		0.05	4	3		5				4		6	4	3	7		3	7	5	3	8	5	3		6	3	9	6	4
		0.01		4	3		4	3	6	5	3	7	5		8	5		8	6		9	6		9	7		10	7	

TABLE A-19. CRITICAL VALUES FOR THE QUANTILE TEST (CONT.)

m is the number of background samples, *n* is the number of site samples, and *c* is the number values larger than the b_1^{th} quantile.
If *s* is greater than or equal to the table value, q_α, then reject the null hypothesis of no difference at the given significance level.

			n																							
			13			**14**			**15**			**16**			**17**			**18**			**19**			**20**		
			0.50	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**
m	**4**	***c***	*8*	*4*	*2*	*9*	*5*	*2*	*9*	*5*	*2*	*10*	*5*	*2*	*10*	*5*	*2*	*11*	*6*	*3*	*11*	*6*	*3*	*12*	*6*	*3*
		0.10	8						9						10						11					
		0.05				9						10						11						12		
		0.01																								
	5	***c***	*9*	*5*	*2*	*9*	*5*	*2*	*10*	*5*	*2*	*10*	*5*	*2*	*11*	*6*	*3*	*11*	*6*	*3*	*12*	*6*	*3*	*12*	*6*	*3*
		0.10																								
		0.05	9			9			10			10			11			11			12			12		
		0.01																								
	6	***c***	*9*	*5*	*2*	*10*	*5*	*2*	*10*	*5*	*2*	*11*	*6*	*3*	*11*	*6*	*3*	*12*	*6*	*3*	*12*	*6*	*3*	*13*	*7*	*3*
		0.10	8			9			9			10			10			11			11			12		
		0.05																								
		0.01	9			10			10			11			11			12			12			13		
	7	***c***	*10*	*5*	*2*	*10*	*5*	*2*	*11*	*6*	*3*	*11*	*6*	*3*	*12*	*6*	*3*	*12*	*6*	*3*	*13*	*7*	*3*	*13*	*7*	*3*
		0.10		5			5			6			6			6						7			7	
		0.05	9			9			10			10			11			11			12			12		
		0.01	10			10			11			11			12			12			13			13		
	8	***c***	*10*	*5*	*2*	*11*	*6*	*3*	*11*	*6*	*3*	*12*	*6*	*3*	*12*	*6*	*3*	*13*	*7*	*3*	*13*	*7*	*3*	*14*	*7*	*3*
		0.10		5		9						10	6			6						7			7	
		0.05	9			10	6		10	6		11			11			12	7		12			13		
		0.01	10			11			11			12			12			13			13			14		
	9	***c***	*11*	*6*	*3*	*11*	*6*	*3*	*12*	*6*	*3*	*12*	*6*	*3*	*13*	*7*	*3*	*13*	*7*	*3*	*14*	*7*	*3*	*14*	*7*	*3*
		0.10				9						10						11			12			12		
		0.05	9	6			6		10	6			6		11	7			7			7			7	
		0.01	10			10			11			11			12			12			13			13		
	10	***c***	*11*	*6*	*3*	*12*	*6*	*3*	*12*	*6*	*3*	*13*	*7*	*3*	*13*	*7*	*3*	*14*	*7*	*3*	*14*	*7*	*3*	*15*	*8*	*3*
		0.10																								
		0.05	9	6		10	6		10	6		11	7		11	7		12	7		12	7		13	8	
		0.01	10			11			11			12			12			13			13			14		
	11	***c***	*12*	*6*	*3*	*12*	*6*	*3*	*13*	*7*	*3*	*13*	*7*	*3*	*14*	*7*	*3*	*14*	*7*	*3*	*15*	*8*	*3*	*15*	*8*	*3*
		0.10				9			10	6		10			11			11			12			12		
		0.05	9	6		10	6					11	7			7		12	7		13	8		13	8	
		0.01	10			11			11	7		12			12			13			14			14		
	12	***c***	*12*	*6*	*3*	*13*	*7*	*3*	*13*	*7*	*3*	*14*	*7*	*3*	*14*	*7*	*3*	*15*	*8*	*3*	*15*	*8*	*3*	*16*	*8*	*4*
		0.10		5			6			6			6						7			7				
		0.05	9			10			10			11			11	7		12			12			13	8	
		0.01	10	6		11	7		11	7		12	7		12			13	8		13	8		14		

TABLE A-19. CRITICAL VALUES FOR THE QUANTILE TEST (CONT.)

m is the number of background samples, *n* is the number of site samples, and *c* is the number values larger than the b_1^{th} quantile.
If *s* is greater than or equal to the table value, q_α, then reject the null hypothesis of no difference at the given significance level.

			n																					
			13			**14**			**15**			**16**			**17**			**18**			**19**			**20**		
			0.50	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**	**0.50**	**0.75**	**0.90**
m	**13**	***c***	*13*	*7*	*3*	*13*	*7*	*3*	*14*	*7*	*3*	*14*	*7*	*3*	*15*	*8*	*3*	*15*	*8*	*3*	*16*	*8*	*4*	*16*	*8*	*4*
		0.10	9			9			10	6		10	6		11			11	7		12	7		12	7	
		0.05		6		10	6		11			11			12	7		12			13			13		
		0.01	10	7		11	7		12	7		12	7		13	8		13	8		14	8		14	8	
	14	***c***	*13*	*7*	*3*	*14*	*7*	*3*	*14*	*7*	*3*	*15*	*8*	*3*	*15*	*8*	*3*	*16*	*8*	*4*	*16*	*8*	*4*	*17*	*9*	*4*
		0.10			3															4	12	7	4			
		0.05	9	6		10	6		10	6		11	7		11	7		12	7					13	8	
		0.01	10	7		11	7		11	7		12	8		12	8		13	8		13	8		14	9	
	15	***c***	*14*	*7*	*3*	*14*	*7*	*3*	*15*	*8*	*3*	*15*	*8*	*3*	*16*	*8*	*4*	*16*	*8*	*4*	*17*	*9*	*4*	*17*	*9*	*4*
		0.10	9		3	9		3	10						11		4			4	12		4			4
		0.05	10	6		10	6		11	7		11	7		12	7		12	7		13	8		13	8	
		0.01	11	7		11	7		12	8		12	8		13	8		13	8		14	9		14	9	
	16	***c***	*14*	*7*	*3*	*15*	*8*	*3*	*15*	*8*	*3*	*16*	*8*	*4*	*16*	*8*	*4*	*17*	*9*	*4*	*17*	*9*	*4*	*18*	*9*	*4*
		0.10			3		6	3	10	6				4	11		4		7	4	12		4			4
		0.05	9	6		10				7		11	7		12	7		12	8		13	8		13	8	
		0.01	10	7		11	7		11	8		12	8		13	8		13	9		14	9		14	9	
	17	***c***	*15*	*8*	*3*	*15*	*8*	*3*	*16*	*8*	*4*	*16*	*8*	*4*	*17*	*9*	*4*	*17*	*9*	*4*	*18*	*9*	*4*	*18*	*9*	*4*
		0.10	9		3		6	3	10	6			6		11	7	4		7	4	12	7	4			4
		0.05	10	6		10			11	7	4	11	7	4	12			12	8		13	8		13	8	
		0.01	11	7		11	7		12	8		12	8		13	8		13	9		14	9		14	9	
	18	***c***	*15*	*8*	*3*	*16*	*8*	*4*	*16*	*8*	*4*	*17*	*9*	*4*	*17*	*9*	*4*	*18*	*9*	*4*	*18*	*9*	*4*	*19*	*10*	*4*
		0.10	9		3				10	6					11				7	4	12	7	4	13		4
		0.05		6		10	6	4	11		4	11	7	4	12	7	4	12			13	8		14	8	
		0.01	10	7		11	7		12	7		12	8		13	8		13	8		14	9		15	9	
	19	***c***	*16*	*8*	*4*	*16*	*8*	*4*	*17*	*9*	*4*	*17*	*9*	*4*	*18*	*9*	*4*	*18*	*9*	*4*	*19*	*10*	*4*	*19*	*10*	*4*
		0.10	9						10						11				7		12		4			4
		0.05	10	6	4	10	6	4	11	7	4	11	7	4	12	7	4	12		4	13	8		13	8	
		0.01	11	7		11	7		12	8		12	8		13	8		13	8		14	9		14	9	
	20	***c***	*16*	*8*	*4*	*17*	*9*	*4*	*17*	*9*	*4*	*18*	*9*	*4*	*18*	*9*	*4*	*19*	*10*	*4*	*19*	*10*	*4*	*20*	*10*	*4*
		0.10	9				6		10						11			12	7		12			13		4
		0.05	10	6	4	10	7	4	11	7	4	11	7	4	12	7	4	13	8	4	13	8	4	14	8	
		0.01	11	7		11	8		12	8		12	8		13	8		14	9		14	9		15	9	

APPENDIX B:

REFERENCES

APPENDIX B: REFERENCES

This appendix provides references for the topics and procedures described in this document. The references are broken into three groups: Primary, Basic Statistics Textbooks, and Secondary. This classification does not refer in any way to the subject matter content but to the relevance to the intended audience for this document, ease in understanding statistical concepts and methodologies, and accessibility to the non-statistical community. Primary references are those thought to be of particular benefit as hands-on material, where the degree of sophistication demanded by the writer seldom requires extensive training in statistics; most of these references should be on an environmental statistician's bookshelf. Users of this document are encouraged to send recommendations on additional references to the address listed in the Foreword.

Some sections within the chapters reference materials found in most introductory statistics books. This document uses Walpole and Myers (1985), Freedman, Pisani, Purves, and Adhakari (1991), Mendenhall (1987), and Dixon and Massey (1983). Table B-1 (at the end of this appendix) lists specific chapters in these books where topics contained in this guidance may be found. This list could be extended much further by use of other basic textbooks; this is acknowledged by the simple statement that further information is available from introductory text books.

Some important books specific to the analysis of environmental data include: Gilbert (1987), an excellent all-round handbook having strength in sampling, estimation, and hot-spot detection; Gibbons (1994), a book specifically concentrating on the application of statistics to groundwater problems with emphasis on method detection limits, censored data, and the detection of outliers; and Madansky (1988), a slightly more theoretical volume with important chapters on the testing for Normality, transformations, and testing for independence. In addition, Ott (1995) describes modeling, probabilistic processes, and the Lognormal distribution of contaminants, and Berthouex and Brown (1994) provide an engineering approach to problems including estimation, experimental design and the fitting of models. Millard and Neerchal (2001) has excellent discussions of applied statistical methods using the software package S-Plus. Ginevan and Splitstone (2004) contains applied examples of many of the statistical tests discussed in this guidance and includes some more sophisticated tests for the more statistically minded. Gibbons and Coleman (2001) apply sophisticated statistical theory to environmental issues, but the text is not intended for the casual reader.

B.1 CHAPTER 1

Chapter 1 establishes the framework of qualitative and quantitative criteria against which the data that has been collected will be assessed. The most important feature of this chapter is the concept of the test of hypotheses framework which is described in any introductory textbook. A non-technical exposition of hypothesis testing is also to be found in U.S. EPA (2000, 1994b) which provides guidance on planning for environmental data collection. An application of the DQO Process to geostatistical error management may be found in Myers (1997).

A full discussion of sampling methods with the attendant theory are to be found in Gilbert (1987) and a shorter discussion may be found in U.S. EPA (1989). Cochran (1966) and Kish (1965) also provide more advanced theoretical concepts but may require the assistance of a statistician for full comprehension. More sophisticated sampling designs such as composite sampling, adaptive sampling, and ranked set sampling, will be discussed in future Agency guidance.

B.2 CHAPTER 2

Standard statistical quantities and graphical representations are discussed in most introductory statistics books. In addition, Berthouex & Brown (1994) and Madansky (1988) both contain thorough discussions on the subject. There are also several textbooks devoted exclusively to graphical representations, including Cleveland (1993), which may contain the most applicable methods for environmental data, Tufte (1983), and Chambers, Cleveland, Kleiner and Tukey (1983).

Two EPA sources for temporal data that keep theoretical discussions to a minimum are U.S. EPA (1992a) and U.S. EPA (1992b). For a more complete discussion on temporal data, specifically time series analysis, see Box and Jenkins (1970), Wei (1990), or Ostrum (1978). These more complete references provide both theory and practice; however, the assistance of a statistician may be needed to adapt the methodologies for immediate use. Theoretical discussions of spatial data may be found in Journel and Huijbregts (1978), Cressie (1993), and Ripley (1981).

B.3 CHAPTER 3

The hypothesis tests covered in this edition of the guidance are well known and straight-forward; basic statistics texts cover these subjects. Besides basic statistical text books, Berthouex & Brown (1994), Hardin and Gilbert (1993), and U.S. EPA (1989, 1994) may be useful to the reader. In addition, there are some statistics books devoted specifically to hypothesis testing, for example, see Lehmann (1991). These books may be too theoretical for most practitioners, and their application to environmental situations may not be obvious.

The statement in this document that the sign test requires approximately 1.225 times as many observations as the Wilcoxon rank sum test to achieve a given power at a given significance level is attributable to Lehmann (1975).

B.4 CHAPTER 4

This chapter is essentially a compendium of statistical tests drawn mostly from the primary references and basic statistics textbooks. Gilbert (1987) and Madansky (1988) have an excellent collection of techniques and U.S. EPA (1992a) contains techniques specific to water problems.

For Normality (Section 4.2), Madansky (1988) has an excellent discussion on tests as does Shapiro (1986). For trend testing (Section 4.3), Gilbert (1987) has an excellent discussion on statistical tests and U.S. EPA (1992) provides adjustments for trends and seasonality in the calculation of descriptive statistics.

There are several very good textbooks devoted to the treatment of outliers (Section 4.4). Two authoritative texts are Barnett and Lewis (1978) and Hawkins (1980). Additional information is also to be found in Beckman and Cook (1983) and Tietjen and Moore (1972).

Tests for dispersion (Section 4.5) are described in the basic textbooks and examples are to be found in U.S. EPA (1992a). Transformation of data (Section 4.6) is a sensitive topic and thorough discussions may be found in Gilbert (1987), and Dixon and Massey (1983). Equally sensitive is the analysis of data where some values are recorded as non-detected (Section 4.7); Gibbons (1994) and U.S. EPA (1992a) have relevant discussions and examples.

B.5 CHAPTER 5

Chapter 5 discusses some of the philosophical issues related to hypothesis testing which may help in understanding and communicating the test results. Although there are no specific references for this chapter, many topics (e.g., the use of p-values) are discussed in introductory textbooks. Future editions of this guidance will be expanded by incorporating practical experiences from the environmental community into this chapter.

B.6 LIST OF REFERENCES

B.6.1 Primary References

Berthouex, P.M., and L.C. Brown, 1994. *Statistics for Environmental Engineers*. Lewis, Boca Raton, FL.

Gibbons, R.D. and D. E. Coleman, 2001. *Statistical Methods for Detection and Quantification of Environmental Contamination*. John Wiley, New York, NY.

Gilbert, R.O., 1987. *Statistical Methods for Environmental Pollution Monitoring.* John Wiley, New York, NY.

Ginevan, M.E. and D.E. Splitstone, 2004. *Statistical Tools for Environmental Quality Measurement*. Chapman and Hall, Boca Raton, FL.

Myers, J.C., 1997. *Geostatistical Error Management*, John Wiley, New York, NY.

Ott, W.R., 1995. *Environmental Statistics and Data Analysis.* Lewis, Boca Raton, FL.

U.S. Environmental Protection Agency , 1992a. *Addendum to Interim Final Guidance Document on the Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities.* EPA/530/R-93/003. Office of Solid Waste. (NTIS: PB89-151026)

U.S. Environmental Protection Agency, 2000. *Guidance for the Data Quality Objectives Process* (EPA QA/G4). EPA/600/R-96/055. Office of Research and Development.

U.S. Environmental Protection Agency, 2002. *Choosing a Sampling Design for Environmental Data Collection* (EPA QA/G5S). Office of Environmental Information.

U.S. Environmental Protection Agency, 2004. *Data Quality Assessment: A Reviewer's Guide (Final Draft)* (EPA QA/G9R). Office of Environmental Information.

B.6.2 Basic Statistics Textbooks

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

Freedman, D. R. Pisani, R. Purves, and A. Adhikari, 1991. *Statistics.* W.W. Norton & Co., New York, NY.

Mendenhall, W., 1987. *Introduction to Probability and Statistics* (Seventh Edition). PWS-Kent, Boston, MA.

Walpole, R., and R. Myers, 1985. *Probability and Statistics for Engineers and Scientists* (Third Ed.). MacMillan, New York, NY.

B.6.3 Secondary 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 T. Lewis, 1978. *Outliers in Statistical Data.* John Wiley, New York, NY.

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

Box, G.E.P., and G.M. Jenkins, 1970. *Time Series Analysis, Forecasting, and Control.* Holden-Day, San Francisco, CA.

Chambers, J.M, W.S. Cleveland, B. Kleiner, and P.A. Tukey, 1983. *Graphical Methods for Data Analysis.* Wadsworth & Brooks/Cole Publishing Co., Pacific Grove, CA.

Chen, L., 1995. Testing the mean of skewed distributions, *Journal of the American Statistical Association* 90:767-772.

Cleveland, W.S., 1993. *Visualizing Data.* Hobart Press, Summit, NJ.
Cochran, W.G., 1966. *Sampling Techniques* (Third Edition). John Wiley, New York, NY.

Cohen, A.C., Jr. 1959. Simplified estimators for the normal distribution when samples are singly censored or truncated, *Technometrics* 1:217-237.

Conover, W.J., 1980. *Practical Nonparametric Statistics* (Second Edition). John Wiley, New York, NY.

Cressie, N., 1993. *Statistics for Spatial Data.* John Wiley, New York, NY.

D'Agostino, R.B., 1971. An omnibus test of normality for moderate and large size samples, *Biometrika* 58:341-348.

David, H.A., H.O. Hartley, and E.S. Pearson, 1954. The distribution of the ratio, in a single normal sample, of range to standard deviation, *Biometrika* 48:41-55.

Dixon, W.J., 1953. Processing data for outliers, *Biometrika* 9:74-79.

Filliben, J.J., 1975. The probability plot correlation coefficient test for normality, *Technometrics* 17:111-117.

Geary, R.C., 1947. Testing for normality, *Biometrika* 34:209-242.

Geary, R.C., 1935. The ratio of the mean deviation to the standard deviation as a test of normality, *Biometrika* 27:310-32.

Gibbons, R. D., 1994. *Statistical Methods for Groundwater Monitoring*. John Wiley, New York, NY.

Grubbs, F.E., 1969. Procedures for detecting outlying observations in samples, *Technometrics* 11:1-21.

Hardin, J.W., and R.O. Gilbert, 1993. *Comparing Statistical Tests for Detecting Soil Contamination Greater than Background, Report to U.S. Department of Energy,* PNL-8989, UC-630, Pacific Northwest Laboratory, Richland, WA.
Hawkins, D.M., 1980. *Identification of Outliers.* Chapman and Hall, New York, NY.

Hochberg, Y., and A. Tamhane, 1987. *Multiple Comparison Procedures*. John Wiley, New York, NY.

Journel, A.G., and C.J. Huijbregts, 1978. *Mining Geostatistics.* Academic Press, London.

Kish, L., 1965. *Survey Sampling*. John Wiley, New York, NY.

Kleiner, B., and J.A. Hartigan, 1981. Representing points in many dimensions by trees and castles, *Journal of the American Statistical Association* 76:260.

Lehmann, E.L., 1991. *Testing Statistical Hypotheses*. Wadsworth & Brooks/Cole Publishing Co., Pacific Grove, CA.

Lehmann, E.L., 1975. *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, Inc., San Francisco, CA.

Lilliefors, H.W., 1969. Correction to the paper "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," *Journal of the American Statistical Association* 64:1702.

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.

Madansky, A., 1988. *Prescriptions for Working Statisticians.* Springer-Verlag, New York, NY.

Millard, S.P. and N.K. Neerchal, 2001. Environmental Statistics with S-Plus. CRC Press, Boca Raton, FL.

Ostrum, C.W., 1978. *Time Series Analysis* (Second Edition). Sage University Papers Series, Vol 9. Beverly Hills and London.

Ripley, B.D., 1981. *Spatial Statistics.* John Wiley and Sons, Somerset, NJ.

Rosner, B., 1975. On the detection of many outliers, *Technometrics* 17:221-227.

Royston, J.P., 1982. An extension of Shapiro and Wilk's W test for normality to large samples, *Applied Statistics* 31:161-165.

Sen, P.K., 1968a. Estimates of the regression coefficient based on Kendall's tau, *Journal of the American Statistical Association* 63:1379-1389.

Sen, P.K., 1968b. On a class of aligned rank order tests in two-way layouts, *Annals of Mathematical Statistics* 39:1115-1124.

Shapiro, S., 1986. *Volume 3: How to Test Normality and Other Distributional Assumptions.* American Society for Quality Control, Milwaukee, WI.

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

Stefansky, W., 1972. Rejecting outliers in factorial designs, *Technometrics* 14:469-478.

Stephens, M.A., 1974. EDF statistics for goodness-of-fit and some comparisons, *Journal of the American Statistical Association* 90:730-737.

Tietjen, G.L., and R.M. Moore, 1972. Some Grubbs-type statistics for the detection of several outliers, *Technometrics* 14:583-597.

Tufte, E.R., 1983. *The Visual Display of Quantitative Information.* Graphics Press, Cheshire, CN.

Tufte, E.R., 1990. *Envisioning Information.* Graphics Press, Cheshire, CN.

U. S. Environmental Protection Agency, 1989. *Methods for Evaluating the Attainments of Cleanup Standards: Volume 1: Soils and Solid Media.* EPA/230/02-89-042. Office of Policy, Planning, and Evaluation. (NTIS: PB89-234959*)*

U. S. Environmental Protection Agency, 1992b. *Methods for Evaluating the Attainments of Cleanup Standards: Volume 2: Ground Water.* EPA/230/R-92/014. Office of Policy, Planning, and Evaluation. (NTIS: PB94-138815)

U. S. Environmental Protection Agency, 1994. *Methods for Evaluating the Attainments of Cleanup Standards: Volume 3: Reference-Based Standards*. EPA/230/R-94-004. Office of Policy, Planning, and Evaluation. (NTIS: PB94-176831)

Walsh, J.E., 1958. Large sample nonparametric rejection of outlying observations, *Annals of the Institute of Statistical Mathematics* 10:223-232.

Walsh, J.E., 1953. Correction to "Some nonparametric tests of whether the largest observations of a set are too large or too small," *Annals of Mathematical Statistics* 24:134-135.

Walsh, J.E., 1950. Some nonparametric tests of whether the largest observations of a set are too large or too small, *Annals of Mathematical Statistics* 21:583-592.

Wang, Peter C.C., 1978. *Graphical Representation of Multivariate Data.* Academic Press, New York, NY.

Wegman, Edward J., 1990. Hyperdimensional data analysis using parallel coordinates, *Journal of the American Statistical Association* 85: 664.

Wei, W.S., 1990. *Time Series Analysis (Second Edition)*. Addison Wesley, Menlo Park, CA.