Sampling Strategies and Statistics Training Materials

for

Part 201 Cleanup Criteria

Michigan Department of Environmental Quality
Remediation and Redevelopment Division
www.michigan.gov/deq
John Engler, Governor
Russell J. Harding, Director

2002

AUTHORITY: PA451 of 1994, as amended
TOTAL COPIES: XXXX
TOTAL COST: $X,XXX.XX
COST PER COPY: $XX.XX
Michigan Department of Environmental Quality

TABLE OF CONTENTS

Tab 1.0

Introduction

INTRODUCTION

INTRODUCTION

Sampling Strategies and Statistics Training Materials for Part 201 Cleanup Criteria (S^3TM) was developed to help staff of the Michigan Department of Environmental Quality (MDEQ) by providing recommendations on:

1. sampling of environmental media for various sampling objectives under Part 201, Environmental Remediation, of the Natural Resources and Environmental Protection Act, 1994 PA 451, as amended (NREPA), and

2. determining when it is appropriate to use statistics and which statistical methods to use for comparing data to Part 201 cleanup criteria.

Appropriate sampling strategies differ based on the sampling objectives (e.g, *FACILITY* characterization, verification of remediation, comparison to criteria, or waste characterization), the variability of hazardous substances in the media to be sampled, knowledge about the distribution of hazardous substances on a property, and costs. The S^3TM provides recommendations on sampling strategies based on these considerations. Biased and statistical sampling strategies are presented and discussed.

After sampling has been completed, the degree to which statistics can be used and the selection of statistical method(s) will vary depending on the exposure pathway, land-use category, and type of determination being made (i.e., *FACILITY* determination, remedial action, and verification of remediation and closure). These training materials do not reflect an increased expectation by the department for the use of statistics, but rather are provided to guide decision-making when statistics are used or *PROPOSED* to help assure it is done properly (terms in capitalized italicized font are defined in the tabbed section titled, "Acronyms / Glossary").

Specifically, the S^3TM will help staff answer three basic questions for making cleanup determinations or know when to seek assistance if statistics are being used to assess compliance with applicable criteria:

1. Is a statistical analysis appropriate?
2. What is the appropriate data set to statistically derive a *REPRESENTATIVE CONCENTRATION* for comparison to cleanup criteria?
3. What is the appropriate statistical method to use for comparison to the cleanup criteria?

1. <u>Is a statistical analysis appropriate?</u>

Statistical Guidesheets have been developed to describe the extent to which statistical analysis of data may be relied upon to evaluate each exposure pathway and condition. At the top of each Statistical Guidesheet is an Applicability of Statistics Section which summarizes the primary factors to consider for that exposure pathway or condition. "**YES**," "Generally Not Practical (**GNP**)" or "**NO**" appear in a box to the right of the Applicability of Statistics heading to indicate the degree to which statistical analysis is appropriate. The Statistical Guidesheets are lettered and numbered to correspond with the Criteria Application Guidesheets presented in the Cleanup Criteria Training Material (CCTM).

Statistical Guidesheets categorized as "**YES**" indicate that use of statistics may be appropriate for the exposure pathway/condition and that sufficient data are likely to be available to calculate a *REPRESENTATIVE CONCENTRATION* for comparison to cleanup criteria. Statistical Guidesheets designated as "**GNP**" indicate that statistical applications may be appropriate but that data are not likely to be available and/or the complexities of the exposure pathway/condition make it difficult to derive a *REPRESENTATIVE CONCENTRATION* for comparison to cleanup criteria. Conditions for which no generic criteria have been developed (e.g., polluted soil runoff to surface water), are also designated as **GNP**. Finally, the exposure pathway categorized as "**NO**" means that statistical analysis is not allowed due to an administrative rule requirement. This is true only for the drinking water pathway for which Administrative Rule 709(3) requires that criteria be met at every point in the affected aquifer.

For quick reference, the CCTM general reference table titled, "Conditions to Evaluate in Assessing Compliance with Part 201 Cleanup Criteria," has been expanded to identify the applicability of statistics for each condition to evaluate. This table can be found in the tabbed section titled, "Applicability of Statistics." Remember that in cases where criteria are not applicable, it is not necessary to conduct a statistical analysis of *FACILITY* data.

2. What is the appropriate data set to statistically derive a *REPRESENTATIVE CONCENTRATION* for comparison to cleanup criteria?

Selecting the proper data set for a statistical analysis, if a statistical analysis is appropriate for the exposure pathway or condition, is an important step given the manner in which sampling data are typically obtained at sites. Section 2 of the Statistical Guidesheets addresses Selection of Data for Statistical Analysis.

FACILITY characterization is a necessary first step before an appropriate data set can be identified for statistical comparison to cleanup criteria. Adequate knowledge of contaminant distribution and the presence of *HOT SPOTS* are essential due to assumptions underlying the statistical methods recommended for comparing site data to cleanup criteria (i.e., 95% upper confidence limits (UCLs) for the mean concentration). Adherence to these assumptions is necessary if an accurate statistical conclusion is to be drawn. Once defined, *HOT SPOTS* should not be included in a statistical analysis for comparison to most criteria. *HOT SPOTS* must be addressed separately. These concepts are discussed in more detail in Section 2.4.1 of the tabbed section titled, "Sampling Strategies."

Once the nature and extent of contamination has been defined, it is necessary to identify and/or obtain data that will allow for appropriate comparison to criteria. The statistical methods described in this document require independence of the data (i.e., the data were obtained through *RANDOM* sampling). However, data gathered from *FACILITY* investigations may not be suitable for statistical comparison to cleanup criteria. Samples collected for the purpose of characterizing a *FACILITY* are typically biased, based on factors such as historical information, previous sampling, disposal practices, visual impacts, and aerial photos.

There are two primary considerations in determining if data sets are adequate. First, data sets must be obtained from locations that represent the exposure pathway or condition and the relevant land-use category. For many of the exposure pathways *EXPOSURE UNITS* are defined to describe the area over which a person may be exposed to hazardous substances and data required for each *EXPOSURE UNIT*. Second, if statistics are used, data sets must contain a sufficient number of *RANDOMLY* located sample results to adequately represent hazardous substance concentrations and allow for proper statistical analysis and development of *REPRESENTATIVE CONCENTRATIONS*. Therefore, additional sampling will often be required to

support statistical analyses after the nature and extent of contamination has been defined. Although *RANDOM* samples are preferred for deriving a *REPRESENTATIVE CONCENTRATION*, previous sample results may be used on a *FACILITY*-specific basis. See further discussion of this issue in Section 2.4.2 of the tabbed section titled, "Sampling Strategies."

The appropriate data set for statistical analysis also depends on the size and variability of hazardous substance concentrations in the *EXPOSURE UNIT*. The size of the *EXPOSURE UNIT* varies between different exposure pathways and the land-use category being considered. Generally, only data from one *EXPOSURE UNIT* may be used in each statistical analysis for comparison to cleanup criteria.

Section 2 of the Statistical Guidesheets also provides information related to unique aspects of the exposure pathway/condition that affect which data may be included in a statistical analysis for comparison to criteria. For example, only groundwater data from *GSI MONITORING WELLS* within the *AVERAGING AREA* may be used for statistical comparison to chronic mixing zone-based groundwater surface water interface (GSI) criteria.

3. What is the appropriate statistical method to use for comparison to the cleanup criteria?

Proper evaluation of data sets to assess compliance for an exposure pathway and/or condition is an important objective of the S^3TM. Once the applicability of statistics has been established and an appropriate data set identified, it is necessary to select the appropriate statistical method(s) for comparing those data to Part 201 criteria.

For characterizing human exposure potential to hazardous substances, the Environmental Protection Agency (EPA) recommends that a 95% UCL for the mean be used to estimate a reasonable maximum exposure (RME) concentration for Superfund risk assessments. The MDEQ also recommends use of a 95% UCL for the mean to compare *FACILITY* data to Part 201 criteria.

Use of a 95% UCL for the mean to compare *FACILITY* data to Part 201 criteria corresponds to a baseline assumption that the mean hazardous substance concentration is at or above its respective criterion unless the data provide sufficient evidence to conclude otherwise. This baseline assumption is consistent with EPA's recommendations in the context of federal cleanup programs (e.g., Superfund and Resource Conservation and Recovery Act (RCRA) Corrective Action).

Various methods are available for calculating UCLs for the mean concentration. Selection of the appropriate method requires an evaluation of the assumptions underlying each method. One of these assumptions is the statistical distribution of the data set (i.e., normal, lognormal, or neither). Consequently, each data set must be evaluated for the best-fitting statistical distribution. Chapter 1 of the tabbed section titled, "Statistical Methods" provides several techniques to accomplish this task. As described in Chapter 1, these techniques should be used in combination to best evaluate the statistical distribution.

Chapter 2 of the tabbed section titled, "Statistical Methods" provides techniques for identifying whether suspect data points are statistical outliers. Recommendations for treatment of outliers, once identified, are also provided in Chapter 2.

Methods for calculating UCLs for the mean concentration are provided in Chapter 3 of the tabbed section titled, "Statistical Methods."

Relationship to the Part 201 CCTM

This S^3TM builds on the framework of the CCTM of January 1998 by providing guidance for statistically analyzing sample data to assess compliance with Part 201 cleanup criteria. Part 201 Section 20a(14) states: "the department shall approve the use of probabilistic or statistical methods or other scientific methods of evaluating environmental data when determining compliance with a pertinent cleanup criterion if the methods are determined by the department to be reliable, scientifically valid, and best represent actual site conditions and exposure potential." Since many divisions of the MDEQ utilize Part 201 criteria, the S^3TM will be useful for this purpose across the MDEQ.

Statistical Considerations Related to *BACKGROUND*

Under Part 201, *BACKGROUND* becomes the Part 201 criterion when the *BACKGROUND* concentration for a hazardous substance is greater than its corresponding risk-based criterion. In this case, *FACILITY* data may be compared to *BACKGROUND* concentrations instead. Chapter 4 of the tabbed section titled, "Statistical Methods" provides recommended statistical methods for this purpose. Recommended methods vary depending on: 1) type of *BACKGROUND* being considered (i.e., *STATEWIDE DEFAULT BACKGROUND*, *REGIONAL BACKGROUND*, or *FACILITY-SPECIFIC BACKGROUND*) and 2) whether a statistical analysis of *FACILITY* data is appropriate for comparison to risk-based criteria. Because statistical distribution and presence of outliers remain important considerations for selection of appropriate method(s) to compare *FACILITY* data to *BACKGROUND* data, Chapter 4 refers back to Chapters 1 and 2 for these considerations.

Relationship to the "Verification of Soil Remediation" (VSR) Guidance Document (MDNR 1994; Revision 1)

Topics addressed in the VSR have been incorporated into the S^3TM and updated as necessary to reflect regulatory requirements under Part 201. The VSR was written in the context of the Michigan Environmental Response Act (MERA), 1982 PA 307, as amended, prior to the 1995 amendments. Consequently, recommendations in the VSR do not address concepts addressed by Part 201 such as evaluation of exposure pathways. Statistical methods presented in the VSR have also been updated to reflect more state-of-the-art recommendations in the statistical analysis of environmental data.

The VSR provided sampling recommendations for both verifying remediation and characterizing wastes. The VSR also presented some statistical methods for evaluating verification or characterization data. Sampling strategies for verifying remediation are described in Sections 1.3 and 2.3 of the tabbed section titled, "Sampling Strategies." Statistical methods for comparing data to criteria are provided the tabbed section titled, "Statistical Methods." Waste characterization is addressed in the tabbed section titled, "Waste Characterization."

Professional Judgment

The S^3TM supplements the tools available to aid in decision-making and do not replace or diminish the use of other appropriate tools such as professional judgment. For example, professional judgment may be used to determine that a *FACILITY* has been adequately characterized based primarily on biased sampling. Professional judgment may also be used to evaluate the significance of environmental data in a manner that does not require a statistical

analysis of *FACILITY* data. If it is determined that data are not representative of a quantity of hazardous substance that could result in an unacceptable risk, it may be appropriate to draw a conclusion without using statistics, even if one or more data points are greater than the cleanup criteria.

Cleanup criteria that are based on projections of the fate and transport of a hazardous substance from one media to another (e.g., groundwater volatilization to indoor air), and the screening levels (i.e., flammability and explosivity, acute inhalation) are good examples of pathways where the quantity of hazardous substance that is present can be considered in applying the screening levels or criteria. If a single data point exceeds an acute inhalation screening level, but that data point is representative of only a small area of groundwater, the sample is from considerable depth below ground surface, and the concentration is not substantially greater than the screening level, it may be concluded that there is no need for response activity to address this situation. This is a conclusion based on professional judgment, not on a statistical evaluation.

The S^3TM is aimed primarily at sampling (i.e., recommended approaches to data gathering) and the use of statistics in decision-making under Part 201. Some decisions, however, will be determined on a qualitative basis (e.g., source control), since cleanup criteria are not available for all conditions.

Self-Implementation

Part 201 Section 14(2) states: "A person may undertake response activity without prior approval by the department unless that response activity is being done pursuant to an administrative order or agreement or judicial decree which requires prior department approval. Any such action shall not relieve any person of liability for further response activity as may be required by the department." A self-implemented response activity using statistics to support determinations must be documented in a manner that fully and clearly addresses the three questions outlined in this Introduction.

Waste Characterization

Characterization of wastes for the purpose of disposal must often be addressed at Part 201 *FACILITIES*. Waste characterization is regulated under Part 111, Hazardous Waste Management and Part 115, Solid Waste Management, of the Natural Resources and Environmental Protection Act, 1994 PA 451, as amended (NREPA). Because there is a great deal of overlap in sampling strategies and statistical methods that can be used to compare data to criteria under Parts 201, 111 and 115, considerations related to waste characterization are also provided in this document. The tabbed section titled, "Waste Characterization" provides most of the recommendations related to waste characterization. Many of these considerations have also been incorporated in the tabbed sections titled, "Sampling Strategies" and "Statistical Methods."

APPLICABILITY OF STATISTICS TO EACH CONDITION TO EVALUATE

APPLICABILITY OF STATISTICS TO EACH CONDITION TO EVALUATE
General Reference Table

	CONDITION	REFERENCE GUIDESHEET(S)	APPLICABILITY OF STATISTICS
SOURCES:			
1	Abandoned substances not yet dispersed & free phase liquids	A 7, 20	GNP Pathway Dependent
RISKS DUE TO GROUNDWATER CONTAMINATION:			
2	Drinking water usage	**1**, 2, 7	NO
3	Dermal exposures such as by utility workers	**6**, 7, 8, 9	GNP
4	Indoor air hazards (chronic/systemic)	**4**, 5, 7, 8, 9	GNP
5	Hazards to surface waters	**3**, 7	YES / NO
RISKS DUE TO SOIL CONTAMINATION:			
6	Hazards due to direct contact (ingestion, dermal)	10, **19**, 20, 27, 28, 29	YES
7	Ambient air inhalation hazards	10, **15**, **16**, **17**, **18**, 20, 23, 24, 25, 26	YES
8	Indoor air inhalation hazards	10, **14**, 20, 22	GNP
9	Injury to drinking water use of aquifer	10, **11**, 20, 21	GNP
10	Risk from contact (utility work) with groundwater	10, **13**, 20	GNP
11	Causes groundwater to be hazardous to surface water	10, **12**, 20	GNP
12	Polluted soil runoff to surface water	B, 10, 20	GNP
RISKS DUE TO CONTAMINATION OF SURFACE WATER SEDIMENTS:			
13	Aquatic flora/fauna/food chain hazards/aesthetics	C	GNP
OTHER RISKS:			
14	Acute toxic impacts/physical hazards	D, 8, 9	GNP
15	Terrestrial flora/fauna/food chain hazards/aesthetics	E	GNP
16	Asbestos containing materials	F	Pathway Dependent

Note: **Bold** guidesheet references indicate generic residential cleanup criteria.

APPLICABILITY OF STATISTICS

YES: Use of statistics may be appropriate for this pathway/condition and sufficient data are likely to be available to calculate a *REPRESENTATIVE CONCENTRATION* for comparison to cleanup criteria.

NO: Use of statistics is not allowed for this pathway/condition due to an administrative rule requirement. This is true only for the drinking water pathway in which administrative rule 709(3) requires that criteria be met at every point in the affected aquifer.

GNP: Use of statistics may be appropriate for this pathway/condition but data are not likely to be available and/or the complexities of the exposure pathway/condition make it difficult to derive a *REPRESENTATIVE CONCENTRATION* for comparison to cleanup criteria. Conditions for which no generic criteria have been developed (e.g., polluted soil runoff to surface water) are also designated as “GNP.”

STATISTICAL ANALYSIS WORKSHEETS

Tab 3.0

Worksheets

STATISTICAL ANALYSIS REVIEW WORKSHEET

GENERAL INFORMATION

SITE/PROJECT: _____

☐ **RAP** ☐ **Final IR** ☐ **Due Care** ☐ **Other** _____

Submittal Being Evaluated: _____

Release Area: _____

Substance evaluated: _____

Pathway/Land use: _____

Current criterion value: _____

Applicability of Statistics: ☐ **Yes** ☐ **No*** ☐ **GNP***

*** DEQ Action Taken:** _____

Conclusion of Submittal For This Substance:

Representative Concentration: _____

Statistical Method Used: ☐ 95% UCL for the mean ☐ Other _____

ADEQUACY OF CHARACTERIZATION

1) Nature and extent of contaminant distribution determined? ☐ Yes ☐ No

Comments: _____

2) Is there evidence "Hot Spots" or potential "Hot Spots" remain present? ☐ Yes ☐ No ☐ Unknown

Comments: _____

Characterization adequate to proceed with statistical analysis? ☐ Yes ☐ No*

*DEQ Action Taken:

REVIEW OF DATA SET

CHARACTERISTICS	SUBMITTAL	GUIDESHEET RECOMMENDATION	RECOMMENDATION SATISFIED?
Size of Area Evaluated (Exposure Unit / Averaging Area)	_____ ☐ Acres ☐ Sq. Ft. ☐ _____	_____ Acres / sq.ft. guidesheet number _____ or, matches areas used for mixing zone determination.	☐ Yes ☐ No Comments:_____
Number of Observations		9 minimum per exposure unit per stats guidesheet	☐ Yes ☐ No Comments:_____
Basis For Sample Points		Random, per stats guidesheet	☐ Yes ☐ No ☐ Unknown Comments:_____
Detection Limit		_____ Per op memo #6	☐ Yes ☐ No Comments:_____
Percent Non-Detects		<50% for use of std. methods per statistical methods section (alt. methods to be proposed if >50%.)	☐ Yes ☐ No Comments:_____

REVIEW OF STATISTICAL ANALYSIS

1) STATISTICAL DISTRIBUTION OF DATA SET

Submittal

Distribution: ☐ Normal ☐ Lognormal ☐ Neither

Formal Test: ☐ Shapiro-Wilk Test ☐ Shapiro-Francia Test

Graphical Technique: ☐ Probability Plots ☐ Box Plots

Summary Statistics: ☐ Coefficient of Variation ☐ Coefficient of Skewness

DEQ Review

Distribution: ☐ Normal ☐ Lognormal ☐ Neither

Formal Test: ☐ Shapiro-Wilk Test ☐ Shapiro-Francia Test

Graphical Technique ☐ Probability Plots ☐ Box Plots

Summary Statistics: ☐ Coefficient of Variation ☐ Coefficient of Skewness

Attach notes / worksheets for above method(s) used. Analysis must include a formal test, a probability plot and summary statistics.

2) OUTLIER EVALUATION

Submittal

Was data set evaluated for outliers?

☐ No
☐ Yes

How were outliers identified?
(Check all that apply)

☐ Graphically

☐ Probability Plot ☐ Box Plot ☐ Other_____

☐ Formal tests (Assumed distribution: _____)

☐ Grubbs' Test ☐ Dixon's Test

☐ Rosner's Test ☐ Other_____

CONCLUSIONS

☐ Value(s) not outliers(s)
☐ Value(s) confirmed as outliers

Treatment of outlier value(s)

☐ Included in statistical analysis
☐ Excluded from statistical analysis

Justification (for either): _____

_____

_____

DEQ Review

☐ No potential outliers in data set based upon qualitative review and / or plots
(Proceed to #3)

☐ Potential outliers evident in dataset

Is data set either normal or lognormal?

☐ Yes, outlier(s) tested by:

☐ Grubbs' Test ☐ Dixon's Test
☐ Rosner's Test ☐ Other _____

☐ No, outlier(s) tested by:

☐ Iterative Approach
☐ Other _____

CONCLUSIONS

☐ Value(s) not outliers(s) (Proceed to #3)
☐ Value(s) confirmed as outlier(s)

Treatment of outlier(s)

☐ Outlier value investigated and found to be erroneous because:

_____

_____

Correct value =_____ **or** ☐ Not discernible
(Return to #1) (Exclude value and document, proceed to #3)

☐ Outlier value apparently accurate, but extreme for population and not from a Hot Spot (Proceed to #3)

☐ Outlier value apparently accurate and may represent a Hot Spot (Return to Adequacy of Characterization)

3) CALCULATION OF REPRESENTATIVE CONCENTRATION

Submittal

Statistical method used:

☐ Student's t (for normally distributed data sets)

☐ Land's Method (lognormally distributed data)

☐ Other _____

Representative Concentration:

DEQ Review

Statistical method used

☐ Student's t (for normally distributed data sets)

☐ Land's Method (lognormally distributed data)

☐ Other formulas as proposed and discussed with DEQ Statistician

Representative Concentration:

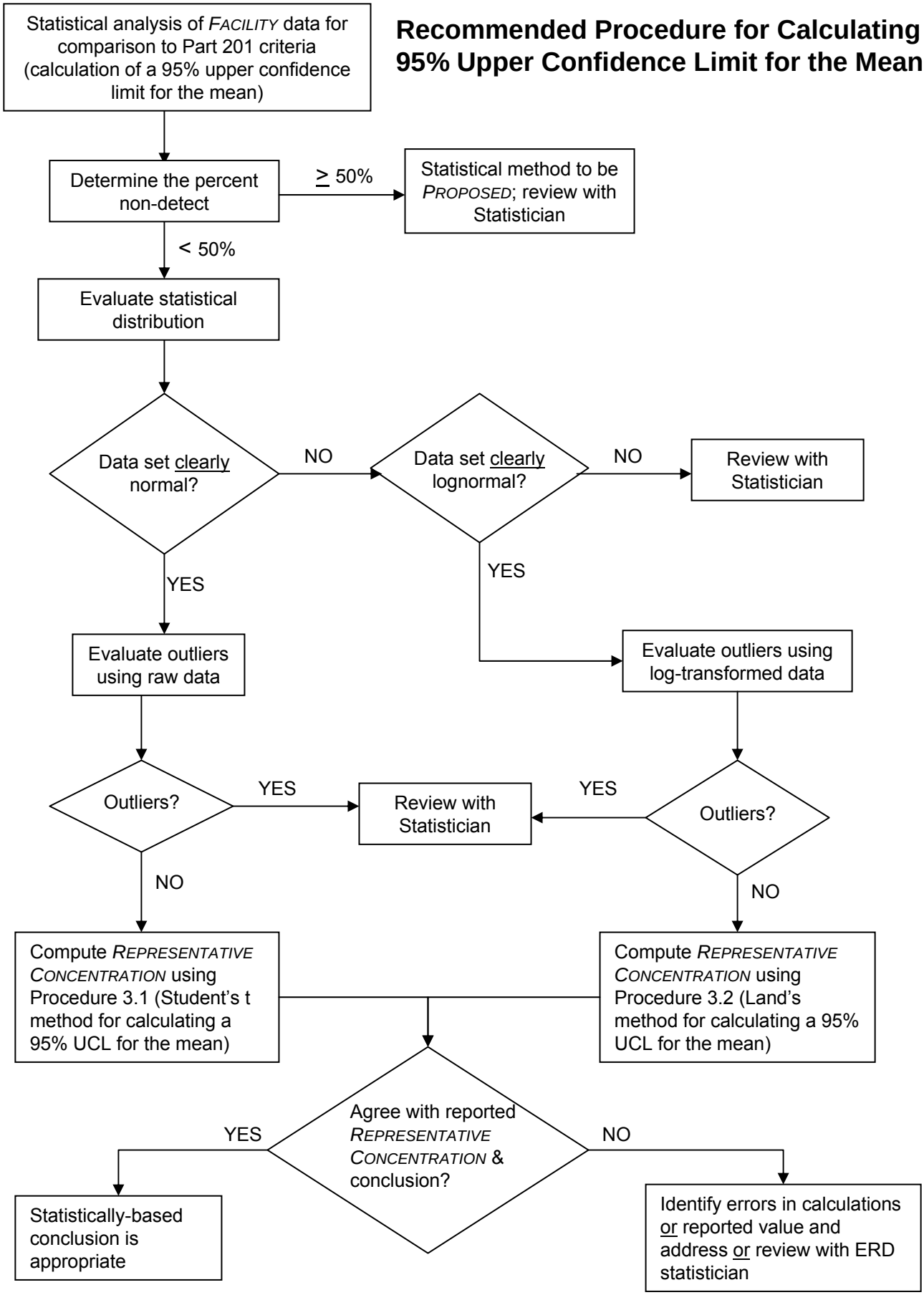
CONCLUSIONS OF REVIEW

☐ **Review supports conclusion of the submittal for the representative concentration of this substance at this location OR**
☐ **Review finds conclusion of the submittal for the representative concentration of this substance IS NOT appropriate because:**

AND Statistical Analysis ☐ DOES ☐ DOES NOT provide evidence of compliance with this criterion for this substance at this location.

Staff Reviewer: _____ Review Date: _____

Recommended Procedure for Calculating a 95% Upper Confidence Limit for the Mean



SAMPLING STRATEGIES

SAMPLING STRATEGIES TABLE OF CONTENTS

EXAMPLES

FIGURES

TABLES

INTRODUCTION

Environmental sampling is often conducted to support a variety of interrelated data objectives. When developing a sampling plan, these objectives require careful consideration if useful data are to be obtained. The objectives for sampling addressed in this tabbed section include:

- identifying and characterizing *RELEASE* areas;
- verifying remediation of *RELEASE* areas;
- comparing *FACILITY* data to Part 201 cleanup criteria, either on a point-by-point basis or using statistics to derive a *REPRESENTATIVE CONCENTRATION*;
- characterizing wastes; and
- establishing *BACKGROUND* concentrations.

Systematic selection of the appropriate sampling strategy is an important first step in satisfying the environmental sampling objectives listed above. Two basic sampling strategies or designs are commonly used for environmental sampling: biased (judgmental) or *RANDOMIZED* (statistical). Further discussion of these strategies occurs in Chapters 1 and 2, respectively.

The DEQ recognizes that sampling to meet these environmental management objectives is a challenging and complex undertaking. This is due in part to the dynamic nature of environmental media (air, water, soil, sediments). Data from environmental sampling are static, representing only a single location and point in time. The complexity and dynamic nature of environmental media is often overlooked in sample planning, collection and cleanup decisions based on static sample data. Thoughtful consideration of the following questions will result in a more effective and efficient sampling strategy.

- Why sample?
  - What is the goal or purpose (data objective) of sampling (e.g., characterization, release area identification, verification of remediation, demonstration of compliance using statistics, demonstration of due care)?
  - Will sample results be used to draw conclusions about human health risks through various exposure pathways, natural resource damage, biological impacts?
- What to sample?
  - Sample media: air, water, soil, sediments, waste materials
  - Hazardous substances: *FACILITY*-specific constituents of concern, organic constituents and their breakdown products, inorganic constituents, metals
- Where to sample?
  - Locations based on biased versus *RANDOM* sampling strategies
  - Vertical sampling components

- When to sample?
  - Are there seasonal variations?
  - Time or history of *RELEASE* – when did the *RELEASE* occur?
  - What are the dynamics of the system(s)?

- How to sample?
  - Discrete grab sample versus continuous sampling
  - Field methodologies

The location and number of samples to be collected depends on factors such as the sampling strategy, the spatial and temporal variability of hazardous substances in the media to be sampled, the level of confidence desired (either in locating a *HOT SPOT* or in drawing conclusions about a mean concentration), and the costs involved.

As described in the following chapters, a combination of sampling strategies is often recommended to best address the sampling objective(s). When characterizing a *FACILITY* (Sections 1.2 and 2.2), biased sampling should be used whenever information is available with which to reliably select sampling locations. However, there are limitations to using biased sampling strategies alone since unexpected areas of contamination will not be identified. To eliminate sampler bias, statistical or *RANDOMIZED* sampling strategies should often be used to supplement biased sampling. The number of samples can be estimated to locate a *HOT SPOT* of an assumed size and shape with a specified level of confidence (Section 2.2.1.1), to proportionally represent an area of a given size (Section 2.2.1.2) or by selecting a particular data objective to satisfy, such as estimating mean concentration with specified levels of precision and confidence.

For verifying remediation of soils, biased sampling is recommended in small areas (i.e., less than ¼ acre) and statistical sampling is recommended in medium- to large-sized areas (Sections 1.3 and 2.3, respectively). Because identification and consideration of *HOT SPOTS* is necessary for statistical comparison of verification data to criteria, biased sampling may also be necessary.

Hazardous substance concentrations in biased samples must generally be compared to criteria on a point-by-point basis (Section 1.4). If all concentrations meet criteria and the *FACILITY* is believed to be adequately characterized, sampling may be complete. If one or more concentrations are present above criteria, a statistical analysis may be considered for comparison to criteria. Further statistical sampling may be necessary with the sampling objective of statistical estimation of a *REPRESENTATIVE CONCENTRATION* for comparison to criteria (Section 2.4). The additional samples would be collected to obtain a sufficient number of *RANDOMLY* located samples to: 1) allow for an appropriate statistical analysis (see the tabbed section titled, “Statistical Methods”) and 2) adequately represent both exposures for a given exposure pathway and hazardous substance concentrations. However, considerations such as applicability of statistics and adequacy of characterization must first be addressed as described in Section 2.4.

Sampling for the purpose of waste characterization is discussed in Sections 1.5 and 2.5 and described in more detail in the tabbed section titled, “Waste Characterization.”

COMPOSITE SAMPLES

Demonstration of compliance with Part 201 cleanup criteria generally requires collection of discrete soil samples. Compositing of samples is not accepted without prior DEQ approval.

SAMPLE ANALYSIS

All test methods and associated target detection levels must be consistent with those specified in rules and procedures under Part 201. These include:

- analytical methodologies
- target detection levels
- quality control procedures

Generally, constituents in soil will be measured on a total, dry weight basis. Considerations for other media (i.e., groundwater, sediments, waste, leachate) must be addressed on a site-specific basis.

CHAPTER 1: BIASED SAMPLING STRATEGIES

1.1 INTRODUCTION TO BIASED SAMPLING

"Biased" sampling strategies generally involve use of professional judgment to collect soil samples from areas most likely to contain contamination. Often, biased sampling is utilized for smaller areas (e.g., less than a 1/4 acre). However, biased sampling also plays a role on large properties. Biased sampling should be used to focus on known or suspected areas of concern.

Use of biased sampling is premised on enough detailed property information on which to base selection of sample locations. The sample locations are purposefully chosen based on the goal of investigating known or suspected areas of concern. With sufficient knowledge of existing conditions, historic activities, or field indicators (e.g., visual, olfactory, or field screening instrumentation), these areas can be focused on reliably.

Any biased sampling plan requires use of professional judgment. A thorough justification must be documented for each sample location explaining the rationale used to select the location. Without this important detail, biased sampling alone will not be adequate. The reporting section of this document should be carefully followed to ensure adequate documentation of the selection of sample locations (see Section 1.4.2).

It is often necessary to use a combination of sampling strategies for both known or suspected areas of concern as well as areas believed to be unimpacted. Since unexpected areas of contamination will not be identified through biased sampling alone, statistical sampling should often be used to supplement biased sampling. This concept is addressed in the following sections for each of the sampling objectives.

Analytical results from biased sampling must generally be compared to Part 201 criteria on a point-by-point basis and individual exceedances noted. When point-by-point comparisons are made, professional judgment is required to interpret the significance of exceedances that are very close to criteria, or that may be associated with insignificant quantities of a hazardous substance.

A statistical analysis of data generated from biased sampling is generally not appropriate. This is due to the underlying assumptions of most statistical methods used to compare *FACILITY* data to cleanup criteria. One underlying assumption is that the data being evaluated were obtained through *RANDOM* sampling of a single, homogeneous population that can be described by a single statistical distribution (e.g., a normal distribution with a mean of 3.6 and a standard deviation of 0.78). Biased sampling can be used to help identify if this is the case or if differing populations (e.g., *RELEASE* areas) are present.

If statistical sampling is completed in addition to biased sampling, it may be appropriate to combine analytical results from the statistical sampling with some or even all of the biased sampling results in a statistical analysis. However, there are several key considerations which must first be addressed, as described in Section 2.4.1 of Sampling Strategies.

Biased sampling strategies require collection of discrete soil samples. Compositing of samples is not accepted without prior DEQ approval.

The remainder of this chapter provides considerations for selection of biased sample locations according to the following sampling objectives:

- *FACILITY* characterization
- Verification of remediation
- Comparison to criteria for demonstration of compliance
- Waste characterization

Chapter 2 of Sampling Strategies provides statistical sampling methods for each of these objectives.

1.2 *FACILITY* CHARACTERIZATION

FACILITY characterization often includes the collection of samples for the purpose of representing *FACILITY-SPECIFIC BACKGROUND* conditions in addition to the investigation of *RELEASE* areas. Sampling for the identification of *RELEASE* areas is described in Section 1.2.1. Sampling to characterize *FACILITY-SPECIFIC BACKGROUND* conditions is described in Section 1.2.2.

1.2.1 *RELEASE* Area(s)

Sampling strategies for investigation of *RELEASE* areas should incorporate information on known or suspected areas of contamination whenever this information is available. Existing information on areas of contamination is often incorporated in sampling plans through biased sampling of areas most likely to be impacted. Application of a statistical sampling approach such as simple *RANDOM* sampling alone would not be appropriate since it would not incorporate this site-specific knowledge. Known or suspected areas of contamination may not be sampled.

It is often necessary to use a combination of sampling strategies when characterizing soils on a property. For example, biased sampling should be used to focus on known or suspected areas of contamination. However, statistical sampling should also be considered to supplement biased sampling. The necessity of statistical sampling will depend on the accuracy and level of detail of site-specific information used to: 1) select biased sample locations and 2) rule out areas not sampled. For example, if known or suspected contamination is limited to a well defined area, statistical sampling of that area or surrounding areas may not be necessary. If well-defined locations do not exist, statistical sampling may be necessary to either locate the contamination or to adequately demonstrate that the area meets criteria. Statistical sampling should be considered for areas believed to be unimpacted in order to confirm this. Statistical sampling approaches such those described in Sections 2.1.1.1 and 2.2.1.2 can be used.

Use of biased sampling may result in reduced sampling costs when there is sufficient knowledge of known or suspected *RELEASE* areas and adequate documentation of the selected sample locations. However, analytical results from biased sampling must generally be compared to Part 201 criteria on a point-by-point basis. Statistical analysis of biased sampling data is generally not appropriate. If statistical sampling is completed to supplement biased sampling, it may be appropriate to combine analytical results from the statistical sampling with some or even all of the biased sampling results in a statistical analysis for comparison to Part 201 criteria (Section 2.4.2). However, the considerations described in Section 2.4.1 must first be addressed.

The number of samples to be collected will be based on professional judgment, considering the level of certainty and quality of information used to identify the location(s) and extent of known or suspected areas of contamination or to judge an area as unimpacted. The size of appropriate exposure units will also impact the number of samples necessary to adequately characterize a property.

Once sampling has been conducted for the purpose of characterization, a judgment must be made as to whether a site has been adequately characterized. This judgment is generally subjective based upon available information and knowledge of site conditions as described above. Furthermore, a *FACILITY* that was thought to be adequately characterized may need supplemental characterization based on data generated through subsequent *RANDOM* sampling if a statistical analysis is to be used to estimate a *REPRESENTATIVE CONCENTRATION* for comparison to criteria, as described in Section 2.4.2. This may be due to the discovery of unexpected results such as previously unidentified *HOT SPOTS*.

Considerations for Biased Sampling

The biased sampling approach specified in this document recommends sampling from areas most likely to exceed cleanup criteria. The location of soil samples relies on site-specific information on the *RELEASE* or contaminant distribution and specific conditions (e.g., soil type, *RELEASE* type) encountered. Sources of information about a property may include historic and/or current:

- aerial photos,
- property photos,
- detailed property maps,
- utility/activity maps or diagrams,
- historic documentation of activities associated with potential *RELEASES*,
- documentation of containment structures,
- documented field observations (such as stained or visible *RELEASE* areas, odors in an area), and/or
- previous remediation activities.

In addition to the sources of information listed above, other factors may be useful in biasing sample locations and depths towards areas most likely to be contaminated as well as in selection of appropriate analytical parameters. The following describes some of these factors.

a. Sample Locations

Using a biased sampling approach, samples must be collected where they will most likely encounter contamination which could exceed the cleanup criteria. This will minimize the number of samples needed to characterize a property. A sampling strategy that uses bias to choose sample locations is recommended. While it is inappropriate for this document to dictate exact locations for sample collection in this strategy, site specific information concerning the *RELEASE* (e.g., the location of leaks in an underground storage tank or its piping) and soil conditions should be used along with professional judgment and the general information provided here to select appropriate soil sampling locations.

Where to sample:

- Existing information on a property should be used to the degree possible when selecting biased sample locations. Field personnel present during the investigation and/or remediation activities should be sufficiently familiar with the conditions on-site to implement an appropriate biased sampling strategy. A soil sampling strategy should incorporate all pertinent biases of a site which may include, but are not limited to, those listed below:

    - source areas,
    - stained soils,
    - preferential pathways for contaminant migration,
    - other site specific "clues" (e.g., fractures in clays),
    - changes in soil characteristics (e.g., sand/clay interfaces),
    - soil types and characteristics , and/or
    - time lapse between *RELEASE* and investigation.

- For example, if a leak was confirmed on the south side of a tank, more extensive sampling should be conducted on the south side. It would be incorrect to sample the north side of the tank area as extensively as the south side when the leak was known and confirmed to be on the south side of the tank.

b. Depths and Soil Types

Medium sand or larger grains

Medium to larger grain size sand has from 20 to 40% porosity. Most sands in Michigan are composed of quartz, limestone, and small amounts of metamorphic rock fragments. These soils have a low capacity for adsorbing metals or hydrophilic (soluble) organic chemicals. Hydrophobic (insoluble) organic chemicals with low molecular weight will adsorb to this soil in small amounts. Hydrophobic chemicals with high molecular weight will adsorb in moderate amounts (Cline & Brown, 1989). These soils have a low capacity to hold contaminants in the grain interstices due to low capillary action. Contaminants that are held in these soils adhere to the grains themselves in dry soils and are forced into the smaller pore spaces in wet soils (Schwille, 1988).

Where to sample:

- In these soils, the capillary force is low enough to ignore its effects in transporting contaminants lateral to gravity. This is especially true for low surface tension products such as gasoline. Therefore, samples should generally be located below the source and/or *RELEASE* area at depths most likely to intercept contaminant migration.

- Limestone sand grains can act as a buffer to contaminants that cause pH changes (e.g., steel mill pickling acids). For these types of contaminants, the sampler should be on the lookout for intra-granular precipitates. These can appear as grain surface staining or make the soil appear clumpy or aggregated. Soils containing precipitates should be sampled.

Fine sand and silt

These soils have strong capillary action due to the small inter-granular distances. A determination of the fluid surface tension of the spilled product is helpful. High surface tension aids in the ability of a substance to overcome gravity by capillary action. As before, higher molecular weight products can be expected to adsorb to the grains to a greater degree. This allows a product to move lateral to gravity and, to a degree, upward from the leak location. Low surface tension products, such as TCE (trichloroethene), are more likely to go straight down than oils in these kinds of soils. However, the hydraulic head (i.e., the amount of product in the original spill) must be substantial to force a dense non-aqueous phase liquid through a medium with a hydraulic conductivity less than 1×10^{-3} cm/sec (Schwille, 1988).

Where to sample:

- Interfaces between fine sand layers with larger grains above should be sampled. When high surface tension contaminants are suspected, silt layers should be sampled.

Clays

Clay soils are very different from the sands and silts. Clays possess a net negative charge. This causes heavy metal cations (e.g., Cr^{+6}, Cd^{+2}, Pb^{+2}) to adsorb to the clay surface. In fact, this is true for any positive ionizable substance. Clays also have a much greater secondary porosity than primary (primary porosity is the space between the soil particles; secondary porosity is the space between fractures, bedding planes, and soil structures). As a result, spills in clay soils tend to follow preferred pathways. Clays will often show signs of shrinkage cracks or fractures that will allow contaminants to migrate in what would otherwise be considered a "tight" soil in a lab analysis of permeability. Signs of fracturing include "patterned" mottling. This is where the iron (and also manganese) will be oxidized to a red, yellow, or reddish brown color along the crack while the matrix remains the reduced blue/gray color (Lindsay, 1979). Additionally, studies have also indicated that dense non-aqueous phase liquids (DNAPL) tend to reside in clay layers where those conditions exist.

Where to sample:

- It is very important to take clay soil samples from fractures. The fractures are the avenue of travel for contaminants in clay soils. Clay soils may also have sand lenses which should always be sampled. Sand lenses in clays tend to collect fluids. As such, they may harbor contaminants.

Organic carbon content of soil

The organic carbon content of soils is a key factor in the ability of any soil to adsorb contaminants. For a variety of reasons (Lindsay, 1979), an increase in organic carbon content leads to an increase in the adsorption of several classes of chemicals.

Where to sample:

- Soils that appear to have excess organic carbon (e.g. peat, muck, darker soils) should be preferentially sampled.

Bedrock

RELEASES into bedrock present difficult problems. Unlike clay, some bedrock formations have substantial primary porosity as well as secondary porosity. In Michigan, these are sandstones, conglomerates, and brecciated/coarse grained limestone. Examples of bedrock in Michigan with low primary porosity are fine grained limestone, shale, and crystalline metamorphic rocks (e.g., gneiss). If the sampler is unaware of the type of bedrock that is being investigated in a *RELEASE* area, a geologist must be consulted.

Where to sample:

RELEASES into areas with shallow bedrock that have significant primary porosity must be sampled in both the fractures and the matrix. Bedrock without primary porosity should have sampling predominantly in the fractures as in the clay situation. Weathered zones in bedrock will hold contaminants better than unweathered zones. This is due to the increased number of adsorption sites available in weathered rock.

c. Changes Over Time and Chemical Transformations

Many organic chemicals may undergo aerobic and anaerobic degradation. A detailed description of these processes is beyond the scope of this document. The subject is approached here, however, to be sure that samplers are aware that the chemical(s) spilled may not be the only chemical(s) in the soil after a transformation has occurred. These transformations should be considered in the *FACILITY* investigation.

The professional literature contains many articles on this subject (Cline and Brown, 1989; Borden and Bedient, 1987; Wilson and Wilson, 1985). The interested reader is directed to these articles.

What to analyze:

- Analyses should be done for all chemicals that have been *RELEASED* as well as those identified as breakdown products of these chemicals.

- For example, soils surrounding a tank containing tetrachloroethylene that had valve leakage occurring over the last ten years should be sampled for the tetrachloroethylene as well as all breakdown products (TCE, dichloroethylene and vinyl chloride).

d. Exposure Pathways

The characteristics of the exposure pathway for each applicable cleanup criterion are critical to consider when locating samples, since sampling data is ultimately compared or interpreted relative to a cleanup criterion. Information on land use, receptor (exposed) population, exposure medium, exposure route (e.g., oral, inhalation and or dermal), hazardous substance and for some criteria pathways contaminant transport/migration, is all integrated into the development of each cleanup criterion. It follows, therefore, that these same characteristics should be considered when samples are collected for comparison to criteria.

Each cleanup criterion conveys an acceptable exposure concentration for a particular exposure pathway. An objective of biased sampling, therefore, may be to purposely collect samples from locations that will represent the unique characteristics of the exposure pathway of concern. The characteristics of each exposure pathway must be considered to assure that sample results are obtained from representative/appropriate locations. For some pathways, exposure to a

hazardous substance occurs in the medium being sampled (e.g., drinking water and soil direct contact criteria), while for other pathways exposure is to a different medium or location from that which is sampled (e.g., soil to groundwater protection criteria, GSI criteria). Biased sampling strategies may, therefore, be designed to achieve one of the following objectives depending on the criteria/pathway of interest: 1) represent the concentration in the medium to which a person may be exposed, or 2) represent the concentration in a medium that is protective of a different exposure medium.

The phrase "may be exposed" is underlined to emphasize that exposures under the generic context of unrestricted future use of property could occur to areas within the soil or groundwater contaminant profile that are currently not available for exposure, but may be in the future depending on property activities.

The exposure pathways considered under Part 201 are listed in the tabbed section titled, "Applicability of Statistics."

1.2.2 *BACKGROUND*

Under Part 201, the term *BACKGROUND* is defined as:

> *the concentration or level of a hazardous substance which exists in the environment at or regionally proximate to a site that is not attributable to any release at or regionally proximate to the site.*

Consequently, *BACKGROUND* samples must be collected from areas that are representative of *BACKGROUND* conditions and have not been impacted by a *RELEASE* at or regionally proximate to the site.

According to Section 20a(11), when *BACKGROUND* concentrations of a hazardous substance are greater than the corresponding Part 201 risk-based criterion, *BACKGROUND* becomes the Part 201 criterion. Consequently, *FACILITY* data will generally be compared to *BACKGROUND* concentrations only when *BACKGROUND* concentrations are greater than the applicable risk-based criterion.

Establishing *BACKGROUND* concentrations for soil can be accomplished by utilizing Operational Memorandum #15. The types of *BACKGROUND* that will generally be considered include: 1) *STATEWIDE DEFAULT BACKGROUND*, 2) *REGIONAL BACKGROUND*, and 3) *FACILITY-SPECIFIC BACKGROUND*. Additional information on each type of *BACKGROUND* is provided in Chapter 4 of the tabbed section titled, "Statistical Methods."

Sampling considerations for *FACILITY-SPECIFIC BACKGROUND* are described below. *FACILITY-SPECIFIC BACKGROUND* samples are not typically collected using statistical or probabilistic approaches. An effort is made to reflect the same natural conditions observed at a *FACILITY.* Consequently, *FACILITY-SPECIFIC BACKGROUND* data are somewhat biased. Although not described here, statistical or probabilistic methods could be incorporated into *FACILITY-SPECIFIC BACKGROUND* soil sampling. If multiple soil types or horizons are present, this could be accomplished through *RANDOM* sampling of each soil type or horizon independently.

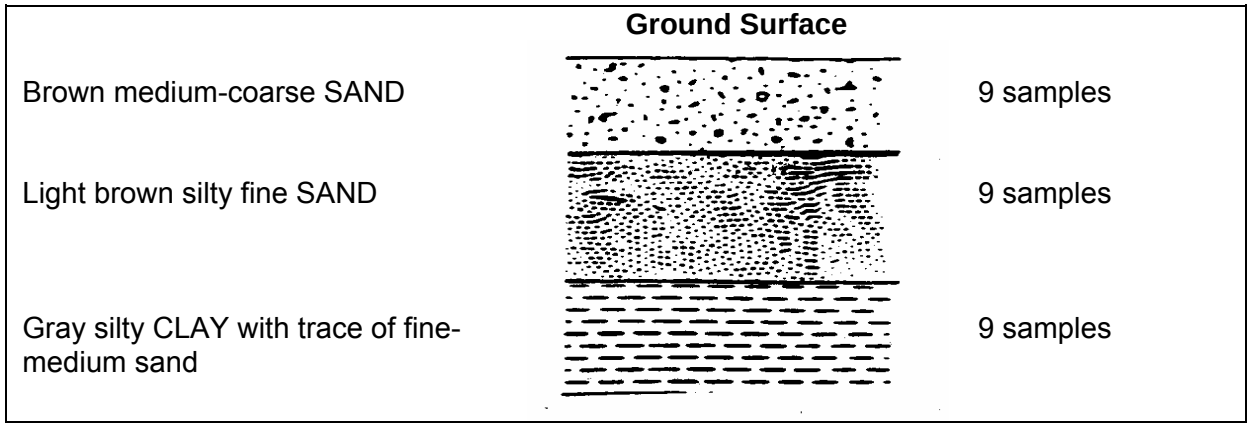
Number of *FACILITY-SPECIFIC BACKGROUND* Samples

Approximately nine samples must be used to establish *FACILITY-SPECIFIC BACKGROUND* concentrations in soils. This recommendation is based on statistical considerations only. It is necessary that an adequate number of samples is available to evaluate the underlying statistical distribution of the *FACILITY-SPECIFIC BACKGROUND* data (i.e., normal, lognormal, or neither).

The goal of collecting *FACILITY-SPECIFIC BACKGROUND* data is to adequately represent the magnitude and variability of naturally occurring concentrations in samples collected from the *FACILITY* of interest. Ideally, a *FACILITY-SPECIFIC BACKGROUND* data set should provide an equal representation of natural soil conditions identified at the *FACILITY*, the key difference being the potential for a *RELEASE*. If a *FACILITY* and the numbers of samples being collected from that *FACILITY* are large, it would be prudent to collect a sufficient number of *FACILITY-SPECIFIC BACKGROUND* samples to adequately represent the same naturally occurring conditions observed in *FACILITY* samples. For this reason, nine samples may not always be adequate to represent or characterize the magnitude and variability of *FACILITY-SPECIFIC BACKGROUND* concentrations.

Furthermore, if multiple soil horizons are present, approximately nine *FACILITY-SPECIFIC BACKGROUND* samples should be collected *from each soil horizon* and a *FACILITY-SPECIFIC BACKGROUND* concentration established for each horizon separately. It is generally not appropriate to combine data from multiple populations or statistical distributions for statistical analysis since this will inflate the variability of the data set, resulting in inflated *BACKGROUND* concentrations.

Example 1.1 Collection of *BACKGROUND* When Multiple Soil Horizons are Present



Additional Considerations for Selection of *FACILITY-SPECIFIC BACKGROUND* Sample Locations

Many factors can play a part in the *BACKGROUND* concentrations of a chemical in soil. Consideration of these factors is particularly important when establishing *FACILITY-SPECIFIC BACKGROUND* concentrations. For example, the geologic origin (e.g., the parent rock) of glacial drift may have been high in copper, lead, or other metals that may be potential contaminants. Additionally, the hydrogeologic situation can alter the quantity of these elements. Groundwater recharge areas (e.g., highlands) are frequently leached of metals while groundwater discharge areas (e.g., swamps, floodplain) are the recipients of leached metals. Thus, sites in low areas will usually have higher *BACKGROUND* concentrations than upland areas. Other conditions, such

as precipitation and atmospheric fallout from widely dispersed human and natural activities, also affect soil concentrations.

In addition, an estimate of contamination depth should be made and *BACKGROUND* samples taken at comparable depths for the particular soil type. This estimate should be made based on waste type, contaminant mobility, operation practices, and soil type (sand, silty sand, clay).

Selection of Analytical Parameters

BACKGROUND should be established as appropriate for site-specific waste constituents, specific chemicals used in various processes, *FACILITY* operations, or remedial investigation results. Sample analyses will generally include metals and other site-specific inorganic constituents of concern. Analytical parameters could possibly include organic constituents if consideration of *NON-RELEASE ANTHROPOGENIC BACKGROUND* conditions is justified.

1.3 VERIFICATION OF REMEDIATION

When verifying remediation of small areas (i.e., less than 10,890 ft^2 or 1/4 acre), biased sampling is recommended. "Biased" sampling involves collecting soil samples from areas most likely to still exceed cleanup criteria after remediation. When biased sampling is conducted to verify remediation of soils, analytical results must generally be compared to Part 201 cleanup criteria on a point-by-point basis. A statistical analysis of verification sample results may not be used to compare the data to Part 201 cleanup criteria.

Although it is unlikely, it is possible that biased sampling could be used to verify remediation in areas larger than 10,890 ft^2. However, statistical sampling methods are generally recommended for these larger areas. (See Section 2.3 of Sampling Strategies.) As noted in Section 2.3, when statistical sampling methods are used, it may be appropriate to conduct a statistical analysis of verification sample results for comparison to Part 201 criteria; however, the considerations summarized in Section 2.4 must first be addressed.

Compositing samples for verifying soil remediation is not acceptable without prior DEQ approval. When verifying a soil remediation is complete, contaminant concentrations will be low. Compositing may result in the contaminant concentrations not being representative of what remains in the soil. If concentrations are low, compositing may dilute the concentrations of a contaminant to below its threshold detection limit. Additionally, if contamination is indicated in a composite sample, the location of the contamination remains unknown.

Any biased sampling plan requires professional judgment. A thorough justification must be documented for each sample location explaining the rationale used to select the location. Without this important detail, it is often necessary to apply a broader sampling strategy to include unknown areas. (See Section 1.4.2)

1.3.1 Selecting Numbers and Locations of Verification Samples in Excavations

Verifying that contaminated soil is remediated by means of excavation requires samples from the excavation bottom and sidewalls. The following tables provide the minimum number of samples necessary to verify cleanup for various size excavations. Considerations for selection of biased sample locations are also discussed.

It should be noted that "excavation" as used here refers only to that area excavated for remediation purposes and being verified to meet Part 201 cleanup criteria.

When biased sampling is used for verifying remediation, a point-by-point comparison of verification sample results to cleanup criteria is specified. If the cleanup criteria are exceeded at any point, this verification methodology may require additional excavation at that point until the criteria are attained.

Number of Samples

The following tables are used to determine the minimum number of samples necessary from the floor and sidewalls of an excavation no greater than 10,890 ft^2 using a biased sampling approach. If the area of the excavation floor exceeds 10,890 ft^2, refer to Section 2.3 of Sampling Strategies.

Determine the minimum number of excavation floor samples from the table below.

Table 1.1 Number of Excavation Floor Samples

Area of Floor (ft^2)	*Number of Samples*
< 500	2
500 < 1,000	3
1,000 < 1,500	4
1,500 < 2,500	5
2,500 < 4,000	6
4,000 < 6,000	7
6,000 < 8,500	8
8,500 <10,890	9

Sidewall samples are required to verify that the horizontal extent of contamination has been remediated. Use Table 1.2 to determine the minimum number of required sidewall samples. In no case is less than one sample on each sidewall (i.e., four) acceptable. In the case of irregularly shaped excavations in which four walls are not readily discernible, divide the total wall area into four segments of approximately equal size. Sidewall samples should be located in accordance with "biases" outlined below.

Table 1.2 Number of Excavation Sidewall Samples

Total Area of Sidewalls (ft^2)	*Number of Samples*
< 500	4
500 < 1,000	5
1,000 < 1,500	6
1,500 < 2,000	7
2,000 < 3,000	8
3,000 < 4,000	9
> 4,000	1 sample per 45 lineal feet of sidewall

Considerations for Biased Sampling

"Biased" sampling involves collecting soil samples from areas most likely to still exceed cleanup criteria. Specific considerations for biasing sample locations are described in detail in Section 1.2.1 under *FACILITY* characterization. The fundamental approaches for biasing sample location are basically the same for verifying remediation; however, the biased sampling is now focused on post-remediation activities.

A site may have an appropriate number of samples collected for verification, but if the samples are not collected from the appropriate locations and adequately reported remediation may not be considered adequate. The location of the sample collection points relies on site-specific analysis of the *RELEASE* or contaminant distribution and the soil types encountered in the excavation. For example, when selecting verification sample locations in an excavation, more extensive verification sampling should be completed on the south side of the excavation if a leak was confirmed on the south side of a tank. It would be incorrect to sample the north side of an excavation pit as extensively as the south side when the leak was confirmed on the south side of the tank.

Sampling and analyzing the locations most likely to have contaminants can minimize the number of samples needed to verify remediation is complete. Professional judgment and site-specific knowledge are required for selection of biased sampling locations. The verification report must accurately locate and describe all sample locations. A thorough justification must be documented for each sample location explaining the rationale used to select the location.

1.3.2 Selecting Numbers and Locations of Soil Verification Samples for Ex Situ Remedies

Verification samples from ex situ remediation activities can also be collected in a biased manner. Again, sampling and analyzing the locations most likely to have contaminant concentrations above Part 201 cleanup criteria can minimize the number of samples needed to verify remediation is complete. However, biased verification soil sample results must generally be compared to cleanup criteria on a point-by-point basis. A statistical analysis of data generated from biased sampling is generally not appropriate.

Number of Samples

The number of verification soil samples to be collected from a soil pile should be based on the volume of the soil pile. The following table provides recommended numbers of verification samples for biased sampling.

Table 1.3 Number of Samples for Ex Situ Remedies

Volume (cubic yards)	0-25	26-100	101-500	501-1,000	1,001-2,000	> 2,000
Number of samples (depending on basis of bias)	3-4	6-8	8-10	10-12	13-15	15 + 3 for every additional 500 cubic yards

If it is demonstrated that concentrations of hazardous substances in the soil under consideration represent a single, homogeneous population, the effectiveness of ex situ soil remedies may be verified by three-dimensional *RANDOM* soil sampling. Refer to Section 2.2 of Sampling Strategies for recommended statistical sampling strategies. Because these strategies are two dimensional, a vertical component must be added. Certain ex situ remedies, such as bio-piles or aboveground vapor extraction, may be more amenable to statistical sampling strategies or batch sampling. Any *PROPOSED* sampling strategy for ex situ remedies should be pre-approved by the DEQ.

Considerations for Biased Sampling

Specific considerations for biasing sample locations are described in detail in Section 1.2.1 under *FACILITY* characterization. The fundamental approaches for biasing sample location are basically the same for verifying remediation; however, the biased sampling is now focused on post-remediation activities.

1.3.3 Selecting Numbers and Locations of Soil Verification Samples for In Situ Remedies

In situ verification soil sampling is required to evaluate the effectiveness of the remedy or to *PROPOSE* closure upon completion of an in situ soil corrective action (e.g., soil vapor extraction, bioventing, in situ bioremediation, and natural attenuation). The purpose of the in situ verification soil sampling is to demonstrate that the entire volume of contaminated soil is below Part 201 cleanup criteria. Because the in situ verification soil sampling is characterizing a volume of soil, additional samples will be required beyond the number of sidewall and floor samples recommended above for excavations, as described below.

Biased verification soil sample results must generally be compared to cleanup criteria on a point-by-point basis. A statistical analysis of data generated from biased sampling is generally not appropriate.

Number of Sample Locations

The number of biased sampling locations should be selected using Tables 1.1 and 1.2 shown in Section 1.3.1 (Excavation Floor Samples and Excavation Sidewall Samples, respectively). Table 1.2 is also considered since it is important to verify remediation around the lateral extent of the remediated area.

Statistical sampling methods are generally recommended for areas larger than 10,890 ft^2. These methods are described in Section 2.2 of Sampling Strategies. Because these strategies are two dimensional, a vertical component must be added. Any *PROPOSED* sampling strategy for in situ remedies should be pre-approved by the DEQ.

Considerations for Biased Sampling

Specific considerations for biasing sample locations are described in detail in Section 1.2.1 under *FACILITY* characterization. The fundamental approaches for biasing sample location are basically the same for verifying remediation; however, the biased sampling is now focused on post-remediation activities.

A biased sampling strategy should also be used to select the vertical sampling intervals that are the most likely to still contain contaminant concentrations above Part 201 cleanup criteria. The following areas should be considered for vertical sampling intervals: areas of probable high contaminant concentrations (based on previous sample results, field observations, and/or the field screening instrumentation) and areas where the flow of air, water and/or nutrients will be impeded (e.g., low permeability lenses and the capillary fringe zone). If air flow modeling indicates the possible presence of stagnation zones, these areas should be sampled.

At least one sample should be collected from each five feet of the verification sample boring, with the exception of borings that are advanced through uncontaminated backfill. Verification soil samples are not required from uncontaminated backfill material. If there is no basis for further biasing the sample from within a five-foot interval, then the five-foot interval may be subdivided into six-inch intervals, and an interval *RANDOMLY* selected for sampling.

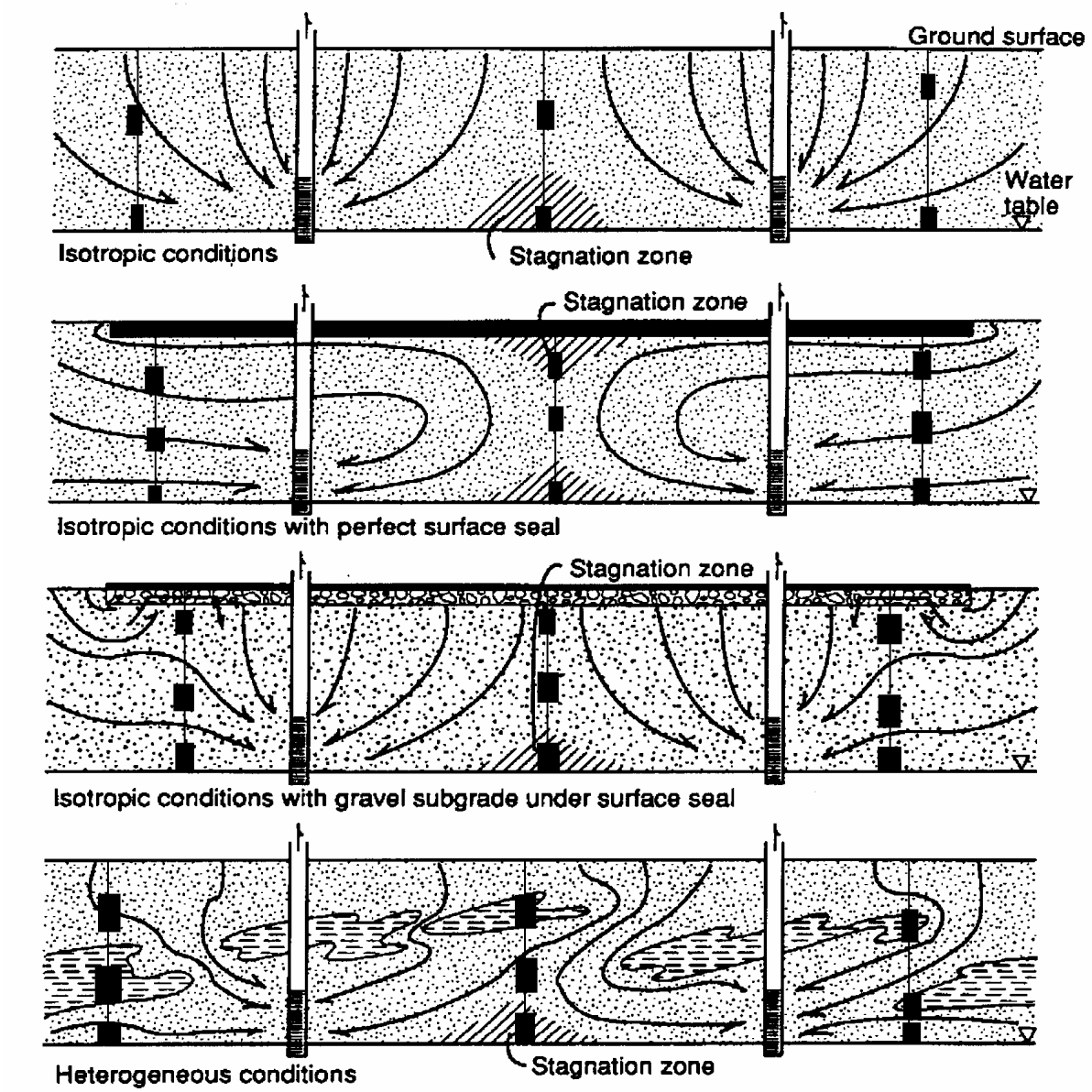
The verification soil borings must extend at least as far as the known depth of the soil contamination. If a confining layer is determined to be the lower boundary of the contamination (as should have already been determined during the investigation phase), at least one soil sample should be collected from the top of the lower confining layer to verify this. If the confining layer is not the lower boundary of the contamination, then a sampling of this layer and below must be conducted in relation to the remediation activity.

Other Considerations

Partial dewatering of an aquifer can allow soil vapor extraction and/or bioventing systems to remediate the residual soil contamination below the water table. If the groundwater potentiometric surface is being lowered during remediation, then the verification soil samples should be collected while the soils in question are still dewatered. The groundwater dewatering system should then be discontinued and verification groundwater samples collected.

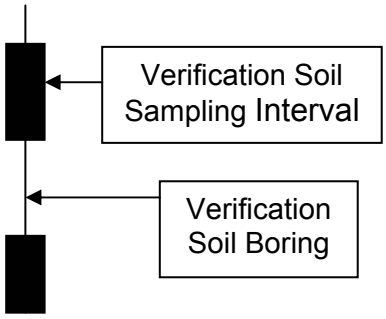
Examples of biased strategies for vertical sampling are shown on the following page.

Example 1.2 Examples of Biased Vertical Sampling Strategies



Note: Not to scale, for conceptual discussion purposes only
Flow is three dimensional. Recharge also occurs in the dimension perpendicular to the drawing.

(Revised from WDNR, 1993)



1.4 COMPARISON TO CRITERIA

1.4.1 Demonstrating Compliance on a Point-by-Point Basis

Use of biased sampling may require fewer samples to demonstrate compliance than statistical or probabilistic approaches. However, a limitation to biased sampling is that unexpected areas of contamination will not be identified. Consequently, statistical sampling should often be used to supplement biased sampling.

Analytical data generated using biased sampling strategies must generally be compared to Part 201 criteria on a point-by-point basis and individual exceedances noted. When point-by-point comparisons are made, professional judgment is required to interpret the significance of exceedances that are very close to criteria, or that may be associated with insignificant quantities of a hazardous substance.

Interpretation of Analytical Results

If all samples in an area are below cleanup criteria based on biased sampling and the area is believed to be adequately characterized, the investigation for this area may be complete. A thorough justification must be documented for each sample location explaining the rationale used to select the location. Adequate documentation of all sample locations must also be provided with respect to known or suspected *RELEASE* areas.

If one or more samples contain contaminant concentrations above cleanup criteria, this may indicate one or more of the following:

- additional site characterization is necessary to better understand the exceedance (e.g., if no exceedances were expected in an area/depth), and possibly a change in sampling strategy from biased to statistical (e.g., identification of an unexpected Hot Spot, unless documented to be a localized Release, may necessitate statistical sampling with the objective of identifying similar Hot Spots (see Section 2.2.1.1);
- vertical and horizontal delineation of elevated concentrations in the area of the biased sample location is necessary;
- alternate approaches should be considered to demonstrate appropriate pathway protection (e.g., leach testing, if a protection of groundwater criterion is exceeded); and/or
- remediation (or further remediation) is necessary.

Statistical analyses of data from biased sampling is generally not appropriate. This is due to the underlying assumptions of most statistical methods used to compare *FACILITY* data to cleanup criteria. One underlying assumption is that the data being evaluated were obtained through *RANDOM* sampling of a single population that can be described by a single statistical distribution (e.g., a normal distribution with a mean of 3.6 and a standard deviation of 0.78).

If statistical sampling is also completed in an area that was previously sampled using a biased approach, it may be appropriate to combine analytical results from the statistical sampling with some or even all of the biased sampling results in a statistical analysis. See Section 2.4.2 of Sampling Strategies for further detail. However, it is first necessary to address the considerations summarized in Section 2.4.1.

1.4.2 Comparison of *FACILITY* Data to *FACILITY-SPECIFIC BACKGROUND* Concentrations

When Part 201 criteria are established as *BACKGROUND* concentrations, the objective becomes to determine whether the *FACILITY* concentrations are significantly higher than *BACKGROUND* concentrations for a hazardous substance. To make this determination in most cases, *FACILITY* data will be compared to *FACILITY-SPECIFIC BACKGROUND* concentrations on a point-by-point basis. That is, concentrations of each hazardous substance in each *FACILITY* sample will be compared to directly to the *FACILITY-SPECIFIC BACKGROUND* concentration and individual exceedances will be noted. This will most often be the case when *FACILITY* samples are collected in a biased manner. However, it may also be necessary to compare *FACILITY* data collected using statistical sampling strategies to *FACILITY-SPECIFIC BACKGROUND* concentrations on a point-by-point basis.

For information on establishment of *BACKGROUND* concentrations, see Chapter 4 of the tabbed section titled, “Statistical Methods.”

1.4.3 Recommended Summary Report Format for Biased Sampling

Summary Reports for sites utilizing biased sampling strategies must identify the number and location of samples and provide the justification of the sample locations selected. The Summary Report should contain site maps and cross-sections, drawn to scale, which depict the area and volume of contaminated soil; any remediation area(s); the locations of any air/fluid injection and/or extraction wells with their estimated zones of influence; the sampling grids; and the soil boring/sample locations and vertical sampling intervals. The Summary Report should also include the calculations for determining the sampling grid intervals and a statement documenting the sampling strategy utilized for selection of the sampling locations.

The list below identifies items recommended to properly evaluate a closure certification. These items are not "absolutes." Other information or substitutions may be provided which technically justify and certify a "clean closure."

The report must include the following:

1. MAP(s) and CROSS SECTIONS
   Provide a scaled map of the investigated and remediated area (i.e. the estimated *RELEASE* area, or floor and walls of an excavation, or the vertical and horizontal area treated for in situ remediations, etc.) with sample locations identified. If a cross section is utilized to display the remediation activities and data, it should show the relation of key elevations, depict the stratigraphy, fractures, soil types, discolorations, unusual characteristics, possibly indicate the original *RELEASE* source location, sample locations/elevations, etc.

2. SAMPLE LOCATION RATIONALE
   - All sample locations including *BACKGROUND* samples, investigative samples, and/or verification samples
   - Sample depths
   - Sample collection procedures
   - Describe basis of sampling biases and the rationale used for collecting each sample (e.g., clay fractures, discolored soil, location of leak in tank)

3. DATA ANALYSES
   - Analytical parameters
   - Analytical methods used
   - Method detection limits
   - Laboratory Quality Assurance/Quality Control
   - Summary of decontamination procedures

4. STATISTICAL ANALYSIS OF BACKGROUND DATA, IF APPLICABLE
   - Lab results/data tabulation
   - Complete statistical calculations as described in Chapters 1, 2, and 4 of the tabbed section titled, "Statistical Methods"
   - Narrative explanation of *BACKGROUND* concentrations
   - Point-by-point comparisons of *FACILITY* data to cleanup criteria and/or *BACKGROUND* concentrations

5. CONCLUSIONS
   This portion of the report should include a summary of activities and final conclusions, accounting for the work and testing completed and how completed activity(ies) fit in with the site-wide remediation plan (i.e. land use issues, possible closure requirements, whether there is a Remedial Action Plan, resumption of operations, etc.)

1.5 WASTE CHARACTERIZATION

Characterization of wastes for the purpose of disposal must often be addressed at Part 201 *FACILITIES*. The regulatory context for waste characterization is different than the context described in previous sections of this chapter, however. Sections 1.1 through 1.4 describe sampling and analysis of data for the purpose of demonstrating compliance under Part 201. Waste characterization is regulated under Parts 111 and 115. Consequently, sampling strategies and statistical analysis of data for this purpose have been described in a separate tabbed section titled, "Waste Characterization."

CHAPTER 2: STATISTICAL SAMPLING STRATEGIES

2.1 INTRODUCTION TO STATISTICAL SAMPLING

Statistical sampling, also referred to as unbiased or probabilistic sampling, is based on the theory of *RANDOM* chance probabilities in order to choose samples which are representative of a given area. The probability of selecting any sampling location is equal or at least known as with stratified sampling designs. Because sampler bias is not of concern, the error in data accuracy of a *RANDOM* sampling scheme can be objectively measured. Furthermore, knowledge of the contaminant distribution is not always necessary depending on the sampling objective.

In some cases it is preferable to choose statistical sampling strategies over biased sampling strategies since they can be used to produce increased data accuracy while eliminating sampler bias. This will depend on the purpose of sampling and the amount of available information with which to bias sampling. In some cases, a combination of approaches will yield the most comprehensive information.

Several statistical sampling strategies can be used to produce an unbiased, representative sampling program. The principles behind the three basic types of *RANDOM* sampling and the situations for which they are best suited are described below. To achieve true *RANDOM* sampling, composite sampling is not acceptable.

1. ***Simple RANDOM sampling*** is a method that requires little or no prior knowledge of material distribution. It relies on *RANDOM* chance probability theory – where each sampling location has an equal and known probability of being selected. In this way, sampling error can be accurately estimated. Often, the area of interest is sectioned into a two- or three-dimensional grid pattern and *RANDOM* coordinates are chosen for sampling.

2. ***Systematic RANDOM sampling*** is an extension of simple *RANDOM* sampling that may produce a more efficient sampling survey. It can be more efficient by reducing the sampling error while maintaining the sample number, or by reducing the number of samples needed to achieve a specified sampling error, or by reducing the cost of collection. This method also requires little or no knowledge about the waste distribution, but bias and imprecision can be introduced if unseen trends or cycles exist. Two methods used to select sample locations under this method follow.

    A) *RANDOMLY* select a transect or transects and sample at pre-selected intervals.

    B) Pre-select both the transect or transects and the sampling interval and starting from a *RANDOMLY* selected point. This is the method used most throughout this chapter.

3. ***Stratified RANDOM sampling*** requires some knowledge about the waste distribution. When stratification is known or suspected, sampling efficiency can be improved by dividing the material into strata that are more homogeneous than the total area. Simple or systematic *RANDOM* sampling techniques can then be used to sample each stratum independently. Each stratum is divided into a grid pattern and the sampling points are selected *RANDOMLY*. If the area is vertically stratified, the sampling points in each stratum are selected *RANDOMLY* and then selected depths are sampled. If the area is horizontally stratified, the sampling

points within each stratum are selected *RANDOMLY*, but the total depth is sampled. An analysis of variance (ANOVA) may be used to determine if the analytical results differ significantly among strata. This can help evaluate whether use of stratified *RANDOM* sampling was necessary and statistically valid. When the volumes of the strata differ or the number of samples within each stratum differs, the results must be weighed appropriately to avoid bias if the data are to be combined in order to draw conclusions.

Of these methods, Systematic *RANDOM* sampling is generally recommended for each sampling purpose described in this chapter. As noted above, this approach is often the most efficient since it involves collection of samples at equal intervals, simplifying the location of samples in the field. Furthermore, systematic *RANDOM* sampling can serve many purposes. When samples are collected on a regular grid interval, conclusions can be drawn about the size of a *HOT SPOT* likely to be identified (or missed), as described in Section 2.1. Furthermore, because the sampling locations are *RANDOMIZED* through systematic *RANDOM* sampling (i.e., by *RANDOMLY* selecting transects and/or the initial sampling point), it may be appropriate to include the analytical results in a statistical analysis for the purpose of comparison to Part 201 cleanup criteria. However, the considerations described in Section 2.4.1 must first be addressed.

Three methods for establishing grids intervals for systematic *RANDOM* sampling are described. Selection of the appropriate method will depend on the purpose for sampling.

1) **Establish a grid interval based on the size of a *HOT SPOT* to be identified.** *HOT SPOT* shape and size and the level of confidence for finding the *HOT SPOT* are prespecified in this approach. This method is described in Section 2.2 for the purpose of *FACILITY* characterization. See Section 2.2.1.1 for details.
2) **Establish a grid interval based on the size of the area to be sampled.** This approach was originally presented in the medium- to large-site portion of the April 1994 MDEQ guidance document titled, "Verification of Soil Remediation (Revision 1)". This method is described in both Sections 2.2 and 2.3 since it may be useful for either *FACILITY* characterization or verification of remediation. See Section 2.2.1.2 and 2.3.1 for details.
3) **Collect *RANDOM* samples for the purpose of demonstrating compliance with Part 201 criteria using statistics.** This method involves collection of a prespecified number of samples (minimum of nine) for the purpose of estimating a *REPRESENTATIVE CONCENTRATION* for comparison to criteria. As described in Section 2.4, use of this method presumes that the area has already been adequately characterized and represents a single homogeneous population that can be described by a single statistical distribution.

It is often necessary to use a combination of sampling strategies. For example, biased sampling is commonly completed before statistical sampling strategies are employed. Statistical sampling strategies should be considered to supplement characterization. Once characterization is complete, it may be necessary to collect additional samples for the purpose of using statistics to estimate a *REPRESENTATIVE CONCENTRATION* for comparison to criteria.

When conducting the statistical analysis, it may be appropriate to combine some or all of the data from biased sampling with data collected using statistical sampling strategies. See Section 2.4 for further detail. See also the tabbed section titled, "Statistical Guidesheets" for key considerations which must be addressed before a statistical analysis is conducted to compare data to Part 201 criteria.

Statistical sampling strategies require collection of discrete soil samples. Compositing of samples is not accepted without prior DEQ approval.

The remainder of this chapter describes statistical sampling strategies for each of the following sampling objectives:

- *FACILITY* characterization
- Verification of remediation
- Comparison to criteria for demonstration of compliance using statistics
- Waste characterization

Chapter 1 of Sampling Strategies provides considerations for selecting biased sample locations for each of these objectives.

2.2 *FACILITY* CHARACTERIZATION

Statistical sampling strategies may be useful as part of an overall sampling plan for *FACILITY* characterization. This section provides several statistical sampling strategies for this purpose.

In addition to the investigation of *RELEASE* areas, *FACILITY* characterization often includes the collection of samples for the purpose of representing *FACILITY-SPECIFIC BACKGROUND* conditions. Sampling for the identification of *RELEASE* areas is described in Section 2.2.1. Sampling to characterize *FACILITY-SPECIFIC BACKGROUND* conditions is described in Section 2.2.2.

2.2.1 *RELEASE* Area(s)

An important goal of *FACILITY* characterization is to identify existing environmental conditions in areas potentially affected by *RELEASE*s of hazardous substances. This includes identifying the nature and extent of contamination and whether *HOT SPOTS* are present in these areas. The answers to these questions will play a role in subsequent site evaluations and/or actions.

HOT SPOTS

For the purpose of evaluating data from Part 201 *FACILITIES*, a *HOT SPOT* is defined as:

> *Two or more adjacent sample locations in reasonably close proximity at which concentrations are sufficiently above criteria and surrounding location (i.e., spatially correlated concentrations sufficiently above criteria) to indicate that they:*
>
> - *represent a different statistical population and*
> - *pose a potential risk that should not be masked by a statistical analysis.*

Professional judgment may be used to determine if the magnitude of concentrations and/or the number and proximity of spatially correlated samples above criteria are sufficient to classify an area as a *HOT SPOT*. That is, the data should support the conclusion that the samples reflect a second statistical population influenced by a localized *RELEASE*, such as a tank spill, resulting in concentrations above criteria. Spatially correlated concentrations below criteria may also exist; however, these will not be classified as *HOT SPOTS* for the purpose of statistical analysis. Particular consideration should be given to contaminants that are present above soil saturation

(Csat) screening levels and contaminants for which criteria are based on acute toxicological effects and/or physical hazards.

A single sample location may represent a potential *HOT SPOT*, or an area at which concentrations are sufficiently above criteria to necessitate additional sampling to determine if a *HOT SPOT* exists. Once again, professional judgment may be used by staff to determine if the magnitude of the concentration is sufficient to warrant additional sampling.

To identify *HOT SPOTS*, data should be qualitatively evaluated for spatial correlation, or patterns indicating that high or low concentrations generally occur in localized areas. Spatial correlation may exist horizontally and/or vertically. Spatial correlation may also be seen along geological bedding planes or between different lithologies. For example, contaminants could accumulate on top of a sloping clay layer, resulting in spatial correlation even though depth to contamination will vary.

HOT SPOT areas identified for one exposure pathway/condition may not represent *HOT SPOTS* for another exposure pathway/condition due to differences in criteria, *EXPOSURE UNITS* sizes, and significance of exceedances.

If *HOT SPOTS* are present, they must be addressed independent of other non-*HOT SPOT* areas, possibly through remediation or mitigation of exposures through institutional controls. A statistical analysis combining *HOT SPOT* data with non-*HOT SPOT* data is not appropriate.

Considerations for Statistical Sampling

As described in Chapter 1, sampling strategies for *FACILITY* characterization should incorporate information on known or suspected areas of contamination whenever this information is available. Often, biased sampling is conducted for this purpose. However, use of biased sampling to demonstrate compliance with Part 201 criteria requires sufficient knowledge of existing conditions, historic activities, or field indicators (e.g., visual, olfactory, or field screening instrumentation) and thorough documentation of the selection of sampling locations based on this information. Consequently, there is a limitation that must be recognized when using biased sampling for the purpose of *FACILITY* characterization: the resulting information is only as complete and accurate as the information used to select sample locations.

For example, a previously undocumented area of contamination may be present; however, this area may be missed using a biased sampling approach, particularly if there are no observable field indications that the contamination exists. Furthermore, if biased sampling is conducted in an area of known or suspected contamination, areas of contamination can still be missed if the information used to select sample locations is not accurate or sufficiently detailed. In these instances, statistical sampling should be used to supplement biased sampling.

Example 2.1 Example of When Statistical Sampling Strategies Should Be Considered

It is known that waste materials were placed in an area. However, there is limited documentation regarding the exact locations where the materials were placed. Some field observations can be made, but they cannot be fully relied upon due to the nature of the materials or the condition of the area (e.g., overgrown with dense vegetation). Some biased samples may be collected based on the limited knowledge of environmental conditions, but it would be necessary to supplement with statistical sampling of this area to evaluate in an unbiased manner whether *HOT SPOTS* are present.

General Recommendations

In general, the following recommendations are being made with regard to selection of sampling strategies for *FACILITY* characterization:

- A combination of sampling strategies should be considered for the purpose of *FACILITY* characterization.
- Biased sampling should be used whenever information is available with which to reliably bias sample locations.
- Systematic *RANDOM* sampling should be considered to supplement data obtained from most biased sampling programs. The number of samples and/or grid interval necessary to supplement biased sampling results will require use of professional judgment depending on the level of detail and accuracy of existing information used to bias sample locations.
- Systematic *RANDOM* sampling should also be conducted in areas believed to be unimpacted. Professional judgment will be necessary to determine the number of samples and/or the grid interval necessary to confirm that the area(s) meet cleanup criteria. For example, only a small number of samples may be needed if sufficiently detailed and accurate information exists to support that the area is unimpacted.

The *HOT SPOT* identification techniques described in Section 2.2.1.1 may be used to determine the required grid interval for finding a *HOT SPOT* of a given size and shape with a prespecified level of confidence. Section 2.2.1.2 provides formulas that can be used to determine a grid interval based on the size of the area to be sampled.

A benefit when using statistical sampling strategies, either alone or in conjunction with biased sampling, is that it may be appropriate to compare resulting data to Part 201 criteria using statistics. Furthermore, it may be appropriate to combine analytical results from statistical sampling with some or even all of the biased sampling results in a statistical analysis as described in Section 2.4.2 of this chapter. However, considerations summarized in Section 2.4.1 must first be addressed.

Once sampling has been conducted for the purpose of characterization, a judgment must be made as to whether a site has been adequately characterized. This judgment is generally subjective based upon the body of data collected and knowledge of site conditions. Furthermore, a *FACILITY* that was thought to be adequately characterized may need supplemental characterization based on data generated through subsequent sampling (e.g., when sampling for the purpose of estimating a *REPRESENTATIVE CONCENTRATION*, as described in Section 2.4.2.) This may be due to the discovery of unexpected results such as previously unidentified *HOT SPOTS*.

PROPOSALS for grid strategies other than those presented in this chapter may be submitted for DEQ review and approval on a case-by-case basis.

2.2.1.1 Sampling Grids for *HOT SPOT* Identification

Statistical tools are available that may be used to assist in the identification of *HOT SPOTS* during *FACILITY* characterization, particularly when there is limited knowledge about the location of *RELEASE* areas. These tools will be most useful to supplement biased sampling or for locating relatively large *HOT SPOTS.* In addition, these techniques may be useful for evaluating trade-offs between sampling costs and the probability of locating *HOT SPOTS* of a given size. For example, financial constraints may limit the number of samples to be collected from a property. The statistical methods for identifying *HOT SPOTS* described in this section can be used to evaluate the size of the *HOT SPOT* likely to be identified (or missed), given that a specified number of samples are collected along a sampling grid. Furthermore, when sampling is conducted using a grid system, these methods may give some perspective on the likelihood that *HOT SPOTS* exist which have not been found.

The methods described in this section may not be practical if the goal is to locate relatively small *HOT SPOTS* over large areas. Large numbers of samples may be necessary to locate small *HOT SPOTS* with an acceptable level of confidence. **It is not expected that these statistically-based tools will be used to identify small *HOT SPOTS* in this manner.**

Use of the statistical tools described in this section should not replace the use of professional judgment for characterization and/or locating *HOT SPOTS*.

Statistical Tools for Identifying *HOT SPOTS*

Statistical methods for identifying *HOT SPOTS* may be used to answer the following questions (Gilbert, 1987):

1. What size grid interval is necessary to locate a *HOT SPOT* with a specified level of confidence?
2. If a grid interval is pre-specified, what is the probability of locating a *HOT SPOT* of a specific size and shape?
3. What is the probability that a *HOT SPOT* exists when no *HOT SPOTS* were found by sampling on a grid?

Questions 2 and 3 may be of particular interest when sampling is conducted using the method described in Section 2.4.2, "*RANDOM* Sampling of *EXPOSURE UNITS*," since the number of samples (i.e., a minimum of nine) and the corresponding grid interval are predetermined. The size of a *HOT SPOT* likely to be identified (or missed) using this sampling method may be evaluated.

Procedures that can be used for *HOT SPOT* identification and detection are described in "Chapter 10: Locating *HOT SPOTS*" in Gilbert's book, "Statistical Methods for Environmental Pollution Monitoring" (1987). The methods outlined in Chapter 10 are based upon several important assumptions. These include:

- the *HOT SPOT* is circular or elliptical in shape
- samples are collected on a square, rectangular, or triangular grid
- contaminant concentration levels constituting a *HOT SPOT* are clearly defined (e.g., sufficiently above Part 201 criteria; see definition in Section 2.2.1)
- there are no measurement misclassification errors – that is, no errors are made in deciding when a *HOT SPOT* has been found

The method for identifying *HOT SPOTS* described by Gilbert does not incorporate a vertical component. For subsurface *HOT SPOTS*, the method targets the projection of the *HOT SPOT* to the surface. This is a limitation when applying this method for *FACILITY* characterization since many *RELEASE* areas may be present below the surface. Consequently, any sampling plan for identifying *HOT SPOTS* must also incorporate a vertical component.

A brief description of the method follows. For details, see Chapter 10 of Gilbert (1987). For the sake of simplicity, the square sampling grids described in Section 2.4.2, "*RANDOM* Sampling of *EXPOSURE UNITS*," are recommended. The equations and graphs in Gilbert (1987) provide the necessary tools to answer the three questions listed above.

The method for locating *HOT SPOTS* described by Gilbert generally involves specifying:

- the shape of the *HOT SPOT,*
- the allowable probability that a *HOT SPOT* will be missed,
- the spacing of the sampling grid,
- some combination of the above, depending on the objective of the analysis.

Then a graph and an equation are used to determine the quantity of interest. Overall, Gilbert's (1987) method is simple and easy to use, and it provides a sound, defensible technique for designing or evaluating a sampling program for *HOT SPOT* detection.

Example 2.2 Use of *HOT SPOT* Identification Techniques to Select a Grid Interval for Finding a Pre-Specified *HOT SPOT*

Suppose that we want to be able to detect a circular *HOT SPOT* with a radius of 10 ft (diameter of 20 ft) with 90% probability in a rectangular (80 x 125 ft) *EXPOSURE UNIT*. Using a square grid, what size of grid interval and approximately how many samples would be required to find this *HOT SPOT* with the specified probability?

Using Gilbert (1987), the grid interval should be about 18 ft. Consequently, approximately 31 samples would be required to cover this *EXPOSURE UNIT* using systematic sampling with a grid interval of 18 ft.

Example 2.3 Use of *HOT SPOT* Identification Techniques to Identify the Size of a *HOT SPOT* That Can Be Found Using a Pre-Specified Sampling Grid

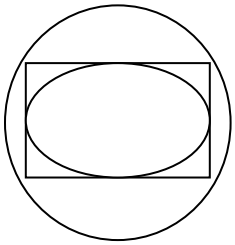
Now suppose that nine samples are to be collected from a square (300 x 300 ft) *EXPOSURE UNIT* using systematic sampling. The objective of the samples is to estimate a *REPRESENTATIVE CONCENTRATION* for comparison to criteria. Using the equation provided in the Section 2.4.2, "*RANDOM* Sampling of *EXPOSURE UNITS*," the approximate grid interval is 100 ft. using a square sampling grid. Assuming that a potential *HOT SPOT* is circular, what size *HOT SPOT* would be detected with 80% probability?

Given these specifications, the methods in Gilbert (1987) suggest that a circular *HOT SPOT* with a radius of 50 ft (diameter of 100 ft) would be detected 80% of the time using a 100 ft grid interval. A circular *HOT SPOT* with a radius of 70 ft (diameter of 140 ft) would be detected 100% of the time using a 100 ft grid interval.

It is not surprising that the grid intervals in the above examples are close in size to the intended target.

A limitation of the method described by Gilbert (1987) is that the *HOT SPOT* must be circular or elliptical in shape. But what about when the *HOT SPOT* is square or rectangular in shape? One option is to evaluate the objective considering a circle constructed around the hypothesized square or rectangular *HOT SPOT* and an ellipse fitting inside the *HOT SPOT* (Figure 2.1). Using the Gilbert (1987) method with the circle and the ellipse, bounds can be obtained for the quantity of interest.

Figure 2.1 Concentric Ellipse and Circle Bounding the Rectangular *HOT SPOT*



Example 2.4 Use of *HOT SPOT* Identification Techniques to Select a Grid Interval for Finding a Rectangular *HOT SPOT*

Suppose that you were designing a sampling plan for a 300 x 300 ft square site and wanted to find a 20 x 40 ft rectangular *HOT SPOT* with a probability of 80%. Using the methods in Gilbert (1987), a grid interval of 26.7 ft (approximately 126 systematic samples) is required to detect the small ellipse and an interval of 44.7 ft (approximately 45 systematic samples) is required to detect the larger circle with a probability of 80%. To find the rectangular *HOT SPOT*, the appropriate grid interval and number of samples lies somewhere between these bounds. The radius of the circle can be determined using the following equation:

$$Radius = \frac{\sqrt{w^2 + l^2}}{2}$$

where w is the width of the rectangle and l is the length of the rectangle.

Setting the Grid

After the grid interval is calculated, it is recommended that a scaled grid overlay be made to superimpose on a map of the area. A point (usually the southwest corner) should be designated as the (0,0) coordinate. The grid can then be adjusted to maximize sampling coverage. Some grid adjustment may be necessary for unusually shaped areas.

Variations on Basic Approach

When the goal is to identify *HOT SPOTS*, further *RANDOMIZATION* of sample locations as described in Section 2.2.1.2 is generally not appropriate. Grid stations must be equally spaced to be able to draw conclusions about the size of the *HOT SPOT* that can be identified.

Available Software for Identification of *HOT SPOTS*

Software tools are also available that can be used to assist in the design of sampling plans for *HOT SPOT* detection. Visual Sampling Plan (VSP) and Fully-Integrated Environmental Location Decision Support system (FIELDS) are two of these programs. VSP is a freeware downloadable program developed by Richard Gilbert and other staff of the Pacific Northwest National Laboratory and is available at http://vsp.pnnl.gov/. Once a site map and a sampling plan have been imported or generated, VSP has the capability of identifying the largest unsampled *HOT SPOT* that may be present. In addition, it can determine the number of samples needed to obtain a desired level of precision, the total cost of various sampling plans, and the locations that should be sampled based on *RANDOM* or systematic sampling.

The FIELDS program is also freeware, but the program disk must be requested. FIELDS offers many of the same sampling plan design capabilities (determining the number of samples, grid intervals, sample locations, and unsampled *HOT SPOT* sizes) with additional capabilities for spatial interpolation and modeling. However, ArcView (a Geographical Information System program) is required to run FIELDS. Because of this, FIELDS may be of limited utility for smaller-scale site assessments or for those without access to ArcView.

2.2.1.2 Sampling Grid Based on Size of Area to be Sampled

The equations and tables in this section provide a simple basis for establishing a grid system to facilitate unbiased selection of sampling points and sample coverage proportional to the area being sampled.

Basic Approach

This method for calculating grid intervals provides a grid point representation that is proportioned to the size of the area being sampled. The following equations provide grid interval estimates for small, medium and large sites, respectively. Small sites are defined as less than 1/4 acre (i.e., < 1/4 acre). Medium sized sites range from larger than 1/4 acre to three acres. Large sites are those which are more than three acres in size. Once a grid interval is determined, a grid system can established over the area to be sampled.

$$small\ \ site \qquad \frac{\sqrt{A/\pi}}{2} = GI$$

$$medium\ \ site \qquad \frac{\sqrt{A/\pi}}{4} = GI$$

$$large\ \ site \qquad \sqrt{\frac{A\pi}{SF}} = GI$$

where: A = area to be grid (ft^2)
GI = grid interval
SF = Site Factor, length of area to be grid (unitless)

To simplify this application, the following chart may be used based on an average size range of sites (one acre = 43,560 ft^2). The approximate grid ranges are provided as a quick check on numbers generated for specific sites using the above formulas.

Table 2.1 Approximate Grid Ranges Based on Size of Area to be Sampled

Acreage of Area to be Sampled	*Square Feet*	*≈Grid Interval Ranges*
up to 0.25 (small)	up to 10,890	0-29 ft
0.25-3.00 (medium)	10,890-130,680	15-50 ft
3.0 and over (large)	130,680 +	30 ft plus

Setting the Grid

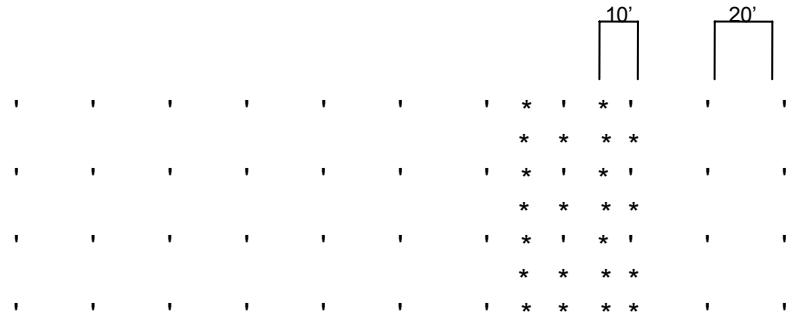
After the grid interval is calculated, it is recommended that a scaled grid overlay be made to superimpose on a map of the area. A point (usually the southwest corner) should be designated as the (0,0) coordinate. The grid can then be adjusted to maximize sampling coverage. Some grid adjustment may be necessary for unusually shaped areas.

Variations on Basic Approach

1. Subgridding

It may be warranted to apply grids with different intervals portions of the area to be sampled so that a proportional sampling can be focused on suspect areas (such as sumps, tank leak areas, etc.).

Example 2.5 Subgridding

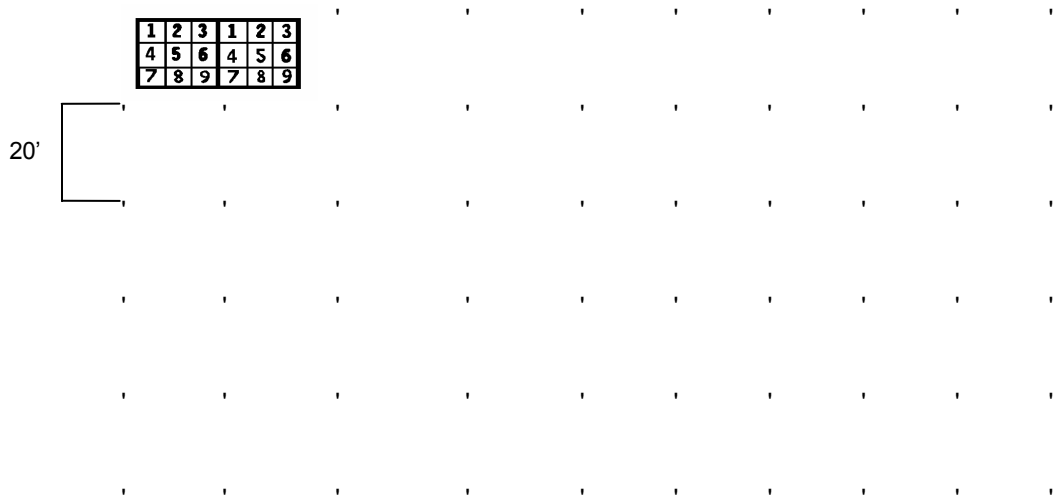


' = Area I Sample Station, 80' x 200', GI = 20'
* = Area II (subset of I) Sample Station, 30' x 50', GI = 10'

2. Further *RANDOMIZATION*

Sites that may have a patterned distribution of waste or contamination due to time sequence of filling, production sequences, or physical site conditions (i.e., furrows) may require a further *RANDOMIZATION* of sampling. In such cases, the following grid cell sampling format may be selected instead of using regular grid point stations. Each grid cell to be sampled may be divided into nine equal sized "subcells." Next, a *RANDOM* numbers table is used to select one of the nine subcells for sampling. The *RANDOM* numbers table can be used again to select a subcell for the next cell and so on.

Example 2.6 Further *RANDOMIZATION* Using Sampling Grids



Area = 120' x 200', GI = 20'

In the example above, a sampling grid was set up with grid point stations 20 ft apart using the appropriate formula. Two cells have been divided into nine subcells each to illustrate further *RANDOMIZATION*. Two *RANDOM* numbers were selected: 4 and 2. Samples should be collected from subcell #4 in the first cell and subcell #2 in the other cell. This process would be continued for all of the cells.

It would generally not be appropriate to apply this technique to the sampling grids described in Section 2.2.1.1. Because the goal of the sampling approach described in Section 2.2.1.1 is to identify *HOT SPOTS*, the grid stations must be equally spaced to be able to draw conclusions about the size of the *HOT SPOT* that can be identified.

2.2.2 *BACKGROUND*

Establishment of *FACILITY-SPECIFIC BACKGROUND* concentrations is described in Section 1.2.2. *FACILITY-SPECIFIC BACKGROUND* samples are not typically collected using statistical or probabilistic approaches. An effort is made to reflect the same natural conditions observed at a *FACILITY.* Consequently, *FACILITY-SPECIFIC BACKGROUND* data are somewhat biased. Although not described here, statistical or probabilistic methods could be incorporated into *FACILITY-SPECIFIC BACKGROUND* soil sampling. If multiple soil types or horizons are present, this could be accomplished through *RANDOM* sampling of each soil type or horizon independently.

2.3 VERIFICATION OF REMEDIATION

This section describes the use of statistical or probabilistic sampling strategies for the purpose of verifying remediation. When verifying remediation of medium- or large-sized areas, statistical sampling strategies are generally recommended. Medium-sized areas are generally defined as

areas ranging from 1/4 acre to three acres. Large areas are those which are larger than three acres.

For each type of remediation described below, statistical analyses for comparing verification data to Part 201 criteria may be appropriate if the data are generated using a statistical sampling approach. However, there are several key considerations which must first be addressed, as described in Section 2.4.1.

If a statistical analysis is used to compare verification data to Part 201 criteria, verification data from other areas should not generally be combined with data from remediated areas. In addition, verification of remediation should be demonstrated independently for each remediated area. If more than one area has been remediated (e.g., through excavation of soils), it is not appropriate to combine areas for the purpose of verifying remediation, regardless of the exposure pathway being evaluated. In the case of in situ or ex situ remediation, previous soil sample data should generally not be included with the new verification soil sample data for statistical analysis.

Compositing samples for verifying soil remediation is not acceptable without prior DEQ approval. When verifying a soil remediation is complete, contaminant concentrations will be low. Compositing may result in the contaminant concentrations not being representative of what remains in the soil. If concentrations are low, compositing may dilute the concentrations of a contaminant to below its threshold detection limit. Additionally, if contamination is indicated in a composited sample, the location of the contamination remains unknown.

2.3.1 Selecting Numbers and Locations of Verification Samples in Excavations

Verifying that contaminated soil is remediated by means of excavation requires collection of samples from the excavation bottom and sidewalls. It should be noted that "excavation" as used here refers only to that area excavated for remediation purposes and being verified to meet Part 201 cleanup criteria.

If sampling and statistical analysis indicate that Part 201 cleanup criteria have not been met, additional remediation will be required. If any portion of the soil in question appears to be causing the material to fail, the area above criteria may be identified through additional sampling and selectively removed. Subsequent sampling must be done to confirm that the remaining material meets Part 201 criteria.

Number of Samples

The number of samples to be collected will be determined based on application of a sampling grid to the area to be sampled. The method for establishing a grid interval described in Section 2.2.1.2 should generally be used. The “area” term in the formula used to determine the grid interval should reflect the total area of the excavation base and sidewalls.

A grid system should be established over the entire area of the excavation using the grid interval determined as above. The grid should extend over sidewalls and base. Grid placement may need to be adjusted to accommodate a minimum of at least one sample from each sidewall.

Sampling of Grid

If the use of statistics is appropriate for comparing soil verification data to Part 201 cleanup criteria, sampling of grids may include all of the grid stations or a phased subset of the total stations. A subset of grid stations may be created by assigning coordinates to all of the grid nodes and *RANDOMLY* selecting nodes for sampling using a *RANDOM* number generator or a *RANDOM* number table. A minimum of nine samples or 25% of the total number of grid points, whichever is larger, should be sampled to allow for a large enough data pool for statistical analysis. However, if statistics are to be used to compare verification data to Part 201 criteria, the considerations in Section 2.4.1 must also be addressed. If an excavation is divided into *EXPOSURE UNITS* to address a specific exposure pathway, a minimum of nine samples per *EXPOSURE UNIT* is necessary for statistical analysis. Consequently, it is advisable that extra samples also be collected and kept under proper chain of custody and storage procedures at the time of initial sampling to avoid an unnecessary return trip to the field.

Lambda Relationship

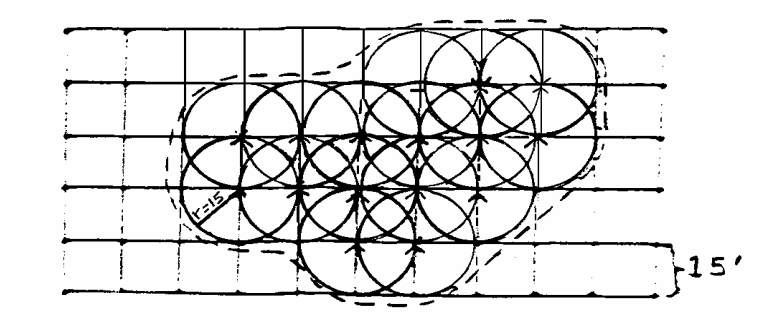
Methods for calculating the sample size requirements, including the Lambda Relationship, are presented in Section 3.2 of the tabbed section titled, “Statistical Methods.” These methods may be useful if an appropriately conducted statistical analysis results in a UCL for the mean that is above the cleanup criterion of interest, but the mean concentration is below. If this occurs, collection of additional samples may result in a lower UCL for the mean which is below the Part 201 criterion. The methods described in Section 3.2 can be used to estimate the number of additional samples that would be required, assuming that the data used in the original analysis were representative.

Grid Approach to Additional Remediation

One of the following two approaches may be used to help guide additional remediation through excavation. Selection of the appropriate method depends on whether verification samples were collected using a regular grid interval or if samples were further *RANDOMIZED* using subcells.

1. *Two-Dimensional Node Sampling Excavation Grid.* Verification sampling as described above will at times indicate that remediation is incomplete. Excavation of contaminated areas should be based on the established grid system interval. Where a subset of grid points has indicated that the entire area exceeds the cleanup, the nodes adjacent to the sampled nodes that are causing the exceedance should be sampled, and this process repeated until the "*HOT SPOTS*" requiring removal have been defined. The radius of excavation around the contaminated sample point(s) is equal to the grid interval (GI=r). Excavation depth is to the deepest point of contamination or to the depth where acceptable levels are anticipated. After excavation, the impacted point(s) must be resampled at their new elevations to verify that the area meets the selected cleanup criteria. If continued contamination is detected, the excavation format is repeated until a satisfactory result is obtained. Remediation of contaminated soil by excavation will be in accordance with the NREPA. The *PROPOSED* remedial action plan must be approved by the DEQ.

Example 2.7 Two-Dimensional Node Sampling Excavation Grid



GL = 150
A = 11,250
GI = 14.9

Sample
Station x
Contaminated Station
r = GI = 15 ft

2. *Two-Dimensional Subcell Sampling Excavation Grid*. The radius of excavation around a contaminated point may need to be adjusted to greater than the GI distance. This adjustment is due to the variable distances between sampling points.

2.3.2 Selecting Numbers and Locations of Soil Verification Samples for Ex Situ Remedies

The number of verification soil samples to be collected from a soil pile should be based on the volume of the soil pile. The table presented in Section 1.3.2 may be used for biased sampling.

If a statistical sampling strategy is to be used, a grid interval can be established for sampling of the soil pile. If the materials are heterogeneous and do not represent a single statistical distribution (e.g., a normal distribution with a mean of 3.6 and a standard deviation of 0.78), the sampling strategy described in Section 2.2.1.2 should be applied based on the size of a *HOT SPOT* to be identified. Since identification of *HOT SPOTS* is a two dimensional sampling technique, a vertical component must be added.

If it is demonstrated or reasonably concluded that concentrations of hazardous substances in the soil under consideration represent a single, homogeneous population, the effectiveness of ex situ soil remedies may be verified by three-dimensional *RANDOM* soil sampling. A minimum of nine samples should be collected from the waste pile if a statistical analysis is used for comparison to criteria. Additional samples may be necessary to adequately represent the variability of concentrations in the soil pile. Application of a sampling grid as described in Section 2.4 can be considered for placement of samples in the two-dimensional horizontal plane. A *RANDOM* numbers generator should be used to select the vertical sampling interval.

Fewer samples may be collected if hazardous substances concentrations are homogeneous; however, a point-by-point comparison to criteria must be made.

Certain ex situ remedies, such as bio-piles or aboveground vapor extraction, may be more amenable to statistical sampling strategies or batch sampling. Any *PROPOSED* sampling strategy for in situ or ex situ remedies should be pre-approved by the DEQ.

Statistical sampling strategies require discrete soil samples. Compositing of samples is not accepted without prior DEQ approval.

If ex situ treatment processes of contaminated soil or waste is used in the remediation, a sampling program for the process stream needs to be developed. The basis of this program is to get representative samples over time versus a spatial approach.

2.3.3 Selecting Numbers and Locations of Soil Verification Samples for In Situ Remedies

In situ verification soil sampling is required to evaluate the effectiveness of the remedy or to *PROPOSE* closure upon completion of an in situ soil corrective action (e.g., soil vapor extraction, bioventing, in situ bioremediation, and natural attenuation). The purpose of the in situ verification soil sampling is to demonstrate that the entire volume of contaminated soil is below Part 201 cleanup criteria. Because the in situ verification soil sampling is verifying remediation for a volume of soil, additional samples will be required beyond the number of sidewall and floor samples recommended above for excavations, as described below.

Number of Sample Locations

The number of samples to be collected will be determined based on application of a sampling grid over the area to be sampled. A grid interval can be established for sampling the volume of soil. If the materials are heterogeneous and do not represent a single statistical distribution (e.g., a normal distribution with a mean of 3.6 and a standard deviation of 0.78), the sampling strategies described in Sections 2.2.1.1 and 2.2.1.2 should be considered. Since these are two dimensional sampling techniques, a vertical component must be added.

However, if statistics are to be used to compare verification data to Part 201 criteria, the considerations in Section 2.4.1 must also be addressed. This may have an impact on the number of samples necessary to verify remediation. Fewer samples may be collected if hazardous substances concentrations are homogeneous; however, a point-by-point comparison to criteria must be made.

Grid placement may need to be adjusted to provide for a sufficient number of samples along the horizontal and vertical boundaries of the volume of soil that was remediated.

Determining Vertical Sampling Intervals for In Situ Verification Sampling

At least one sample should be collected from each five feet of the verification sample boring, with the exception of borings that are advanced through uncontaminated backfill or native soils. The five-foot interval may be subdivided into six-inch intervals, and an interval *RANDOMLY* selected for sampling.

The verification soil borings must extend at least as far as the known depth of the soil contamination. If a confining layer is determined to be the lower boundary of the contamination (as should have already been determined during the investigation phase), at least one soil sample should be collected from the top of the lower confining layer to verify this. If the confining layer is not the lower boundary of the contamination, then a sampling of this layer and below must be conducted in relation to the remediation activity.

Other Considerations

Partial dewatering of an aquifer can allow soil vapor extraction and/or bioventing systems to remediate the residual soil contamination below the water table. If the groundwater potentiometric surface is being lowered during remediation, then the verification soil samples should be collected while the soils in question are still dewatered. The groundwater dewatering system should then be discontinued and verification groundwater samples collected.

2.4 DEMONSTRATING COMPLIANCE WITH PART 201 CRITERIA USING STATISTICS

Sampling conducted for the purpose of identifying and characterizing *RELEASE* areas may not yield adequate data to estimate a *REPRESENTATIVE CONCENTRATION* (the 95% UCL for the mean) for comparison to Part 201 criteria. Therefore, once the nature and extent of contamination has been defined, it is necessary to identify and/or obtain data that will allow for appropriate comparison to cleanup criteria, if statistics are to be used. The goal of samples used to estimate a *REPRESENTATIVE CONCENTRATION* is to represent both exposures and hazardous substance concentrations in an appropriate *EXPOSURE UNIT*. Consequently, an *EXPOSURE UNIT* must first be identified that adequately represents exposures (e.g., a 1/4 acre *EXPOSURE UNIT* for evaluating compliance with generic residential soil direct contact criteria). Systematic *RANDOM* sampling may then be used to identify locations for collecting the minimum of nine samples necessary for statistical analysis.

Note that if all samples in an area are below cleanup criteria and the area is believed to be adequately characterized, the investigation for that area may be complete. A statistical analysis is not required. Adequate documentation of all sample locations must be provided based on sufficient knowledge of known or suspected *RELEASE* areas. A thorough justification must be documented for each sample location explaining the rationale used to select the location.

2.4.1 General Considerations When Demonstrating Compliance Using Statistics

Before assembling a data set and conducting a statistical analysis for comparison to criteria, it is important to give careful consideration to the following:

FACILITY Characterization

FACILITY characterization is a necessary first step before conducting an appropriate statistical analysis of *FACILITY* data for comparison to Part 201 cleanup criteria. Adequate knowledge of contaminant distribution and the presence of *HOT SPOTS* is essential.

Adequate *FACILITY* characterization is necessary due to assumptions underlying the statistical methods recommended by both the DEQ and the Environmental Protection Agency for comparing site data to cleanup criteria (i.e., 95% UCLs for the mean concentration). Adherence to these assumptions is necessary if an accurate statistical conclusion is to be drawn. One key assumption is that the data are independently and identically distributed (iid). For this assumption to be true, the following are generally necessary:

- Samples must be *independent* and representative of the area included in the analysis. In statistical terms, this means that the data were collected *RANDOMLY*.
- For the data to be identically distributed, each data point must have been drawn from the same *identical* statistical distribution (e.g., a normal distribution with a mean of 3.6 ppm and a standard deviation of 0.78 ppm). Data from a *HOT SPOT* area would be represented by a different statistical distribution than data from non-*HOT SPOT* areas. In other words, the mean concentration in a *HOT SPOT* area would be higher than in the non-*HOT SPOT* areas and the standard deviation would likely differ as well.

Consequently, adequate characterization and identification of *HOT SPOT* areas is necessary before an appropriate statistical analysis can be conducted.

Once defined, *HOT SPOTS* should not be included in a statistical analysis for comparison to most criteria. *HOT SPOTS* must be addressed separately. This is necessary to avoid combining data from different statistical distributions and violating the assumptions of the statistical methods. Furthermore, combining samples from *HOT SPOT* areas with samples from other areas of a property for statistical analysis will dilute the sample results that represent *HOT SPOTS*, potentially leaving unacceptable levels of hazardous substances in place. Specific recommendations for treatment of *HOT SPOTS* are provided in the Statistical Guidesheets for each of the pathways and closure categories (see tabbed section titled, "Statistical Guidesheets").

Applicability of Statistics

Before using statistics to compare *FACILITY* data to any Part 201 cleanup criterion, it is important to consider the applicability of statistics for demonstrating compliance with that criterion. Statistical Guidesheets have been developed to describe the extent to which statistical analysis of data may be relied upon to evaluate each exposure pathway and condition. At the top of each Statistical Guidesheet is an Applicability of Statistics Section which summarizes the primary factors to consider for that exposure pathway or condition. "YES," "Generally Not Practical (GNP)" or "NO" appear in a box to the right of the Applicability of Statistics heading to indicate the degree to which statistical analysis is appropriate.

Statistical Guidesheets categorized as "YES" indicate that use of statistics may be appropriate for the exposure pathway/condition and that sufficient data are likely to be available to calculate a *REPRESENTATIVE CONCENTRATION* for comparison to cleanup criteria. Statistical Guidesheets designated as "GNP" indicate that statistical applications may be appropriate but that data are not likely to be available and/or the complexities of the exposure pathway/condition make it difficult to derive a *REPRESENTATIVE CONCENTRATION* for comparison to cleanup criteria. Conditions for which no generic criteria have been developed (e.g., polluted soil runoff to surface water), are also designated as "GNP." Finally, the exposure pathway categorized as "NO" means that statistical analysis is not allowed due to an administrative rule requirement. This is true only for the drinking water pathway for which administrative rule 709(3) requires that criteria be met at every point in the affected aquifer.

Identification of an Appropriate Data Set for Statistical Analysis

Selecting the proper data set for a statistical analysis, if a statistical analysis is appropriate for the exposure pathway or condition, is an important step given the manner in which sampling data are typically obtained at sites. Section 2 of the Statistical Guidesheets addresses Selection of Data for Statistical Analysis.

All data gathered from *FACILITY* investigations may not be suitable for statistical comparison to cleanup criteria. Samples collected for the purpose of characterizing a *FACILITY* are typically biased, based on factors such as historical information, previous sampling, disposal practices, visual impacts, and aerial photos. Once the nature and extent of contamination has been defined, it is necessary to identify and/or obtain data that will allow for appropriate comparison to criteria. There are two primary considerations in determining if data sets are adequate. First, data sets must be obtained from locations that represent the exposure pathway or condition and the relevant land use category. For many of the exposure pathways *EXPOSURE UNITS* are defined to describe the area over which a person may be exposed to hazardous substances and data required for each *EXPOSURE UNIT*. Second, if statistics are used, data sets must contain a sufficient number of *RANDOMLY* located sample results to adequately represent hazardous substance concentrations and allow for proper statistical analysis and development

of *REPRESENTATIVE CONCENTRATIONS*. Therefore, additional sampling will often be required to support statistical analyses after the nature and extent of contamination has been defined. Although *RANDOM* samples are preferred for deriving a *REPRESENTATIVE CONCENTRATION*, previous sample results may be used on a *FACILITY*-specific basis. See further discussion of this issue in Section 2.4.2.

The appropriate data set for statistical analysis also depends on the size and variability of hazardous substance concentrations in the *EXPOSURE UNIT*. The size of the *EXPOSURE UNIT* varies between different exposure pathways and the land use category being considered. Generally, only data from one *EXPOSURE UNIT* may be used in each statistical analysis for comparison to cleanup criteria.

Section 2 of the Statistical Guidesheets also provides information related to unique aspects of the exposure pathway/condition that affect which data may be included in a statistical analysis for comparison to criteria. For example, only groundwater data from *GSI MONITORING WELLS* within the *AVERAGING AREA* may be used for statistical comparison to chronic mixing zone-based groundwater surface water interface (GSI) criteria.

Selecting the Appropriate Statistical Method for Comparison to Criteria

Once the applicability of statistics has been established and an appropriate data set identified, it is necessary to select the appropriate statistical method(s) for comparing those data to Part 201 criteria.

A 95% UCL for the mean should be utilized to compare *FACILITY* data to Part 201 criteria. Various methods are available for calculating UCLs for the mean concentration. Selection of the appropriate method requires an evaluation of the assumptions underlying each method. One of these assumptions is the statistical distribution of the data set (i.e., normal, lognormal, or neither). Consequently, each data set must be evaluated for the best-fitting statistical distribution. Chapter 1 of the tabbed section titled, “Statistical Methods” provides several techniques to accomplish this task. As described in Chapter 1, these techniques should be used in combination to best evaluate the statistical distribution.

Chapter 2 of the tabbed section titled, “Statistical Methods” provides techniques for identifying whether suspect data points are statistical outliers. Recommendations for treatment of outliers, once identified, are also provided in Chapter 2.

Methods for calculating UCLs for the mean concentration are provided in Chapter 3 of the tabbed section titled, “Statistical Methods.”

2.4.2 *RANDOM* Sampling of *EXPOSURE UNITS*

One of the most important factors that affect the quality of a statistical assessment of hazardous substances is sample design. Sample design simply refers to the methodology used to locate and collect samples. However, the conceptual simplicity of designing a sampling scheme belies its importance. Because most statistical methods require a *RANDOM* sample from the population under study, without a proper sampling design a *RANDOM* sample cannot be assured. Without a *RANDOM* sample, drawing accurate conclusions is difficult, if not impossible.

Sampling for the purpose of estimating a *REPRESENTATIVE CONCENTRATION* for comparison to Part 201 criteria should involve some form of *RANDOM* sampling within an *EXPOSURE UNIT.* Two methods for *RANDOM* sampling are described below. *RANDOM* sampling within an appropriate *EXPOSURE UNIT* will yield data which are representative of: 1) the exposure pathway or condition given the relevant land use category, and 2) concentrations of hazardous substances within the *EXPOSURE UNIT*.

Biased sampling strategies require discrete soil samples. Compositing samples is not accepted without prior DEQ approval.

2.4.2.1 Simple *RANDOM* Sampling

Simple *RANDOM* sampling is a common sampling design, particularly in applications other than environmental studies. This design consists of *RANDOMLY* selecting locations within a specified area (e.g., an *EXPOSURE UNIT*) and then collecting a sample at each selected location. Although simple *RANDOM* sampling does provide a *RANDOM* sample from the population, it can be somewhat inefficient and costly to implement.

Determining the *RANDOM* locations and subsequently locating them is a laborious task, possibly requiring a computer and a Global Positioning System (GPS). In addition, simple *RANDOM* sampling tends to unevenly sample the area under consideration. Some areas would likely contain several sample locations in close proximity, while other areas would remain unsampled. The uneven coverage that can result from simple *RANDOM* sampling is illustrated below.

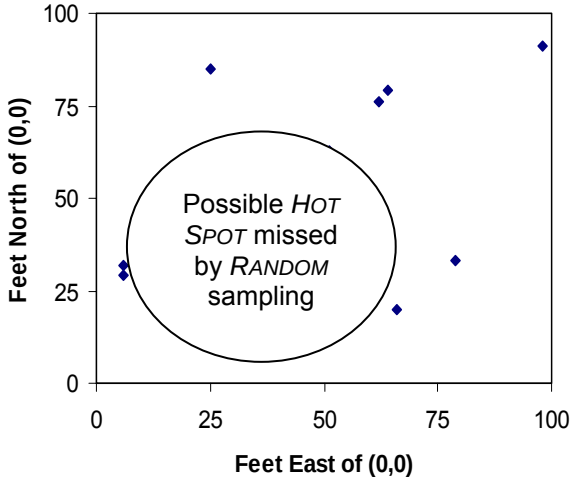
Example 2.8 Simple *RANDOM* Sampling

Figure 2.2 illustrates a sampling design where nine samples were collected *RANDOMLY* over a 100 ft x 100 ft area (approximately 1/4 acre). To locate each sample, a pair of *RANDOM* numbers between 0 and 100 was generated using the Microsoft Excel function:

=RANDBETWEEN(L,U)

where L is the lower number (set to 0 in this case) and U is the upper number (set to 100 in this case). The results for the first sample location were 5 and 27. Starting from the southwest corner of the *EXPOSURE UNIT*, this sample was located by moving 5 ft east and 27 ft north.

Figure 2.2 Simulated Simple *RANDOM* Sample of Nine Observations from a 100 ft x 100 ft (1/4 Acre) *EXPOSURE UNIT*.



Notice that several large areas are not sampled whereas other areas are sampled rather intensively. This is because in *RANDOM* sampling, each location in an area has an equal probability of being sampled, without regard to locations of other samples. Due to the uneven coverage associated with this *RANDOM* sampling design, the likelihood of detecting a *HOT SPOT*, if one exists, is relatively low. With increased sampling intensity (i.e., increasing the number of samples collected), more of the *EXPOSURE UNIT* would be sampled and hence coverage would be improved. However, because systematic *RANDOM* sampling (described below) is more efficient and results in better coverage with better *HOT SPOT* detection capabilities, it is generally superior to simple *RANDOM* sampling for a statistical assessment of hazardous substances.

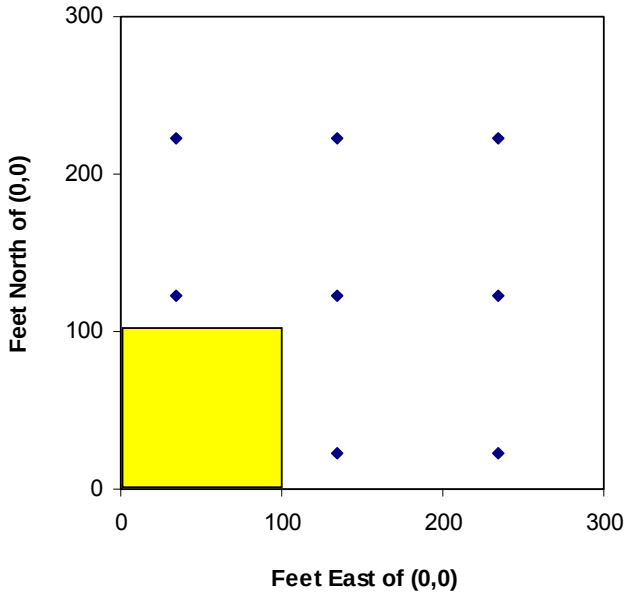
Note that we have made the assumption that the *FACILITY* has been adequately characterized prior to calculation of *REPRESENTATIVE CONCENTRATIONS*. Therefore, in most instances, *HOT SPOT* detection capabilities at this point should not be of major concern. However, because the approach described in the following section may serve both purposes of obtaining a *RANDOM* sample for estimation of a *REPRESENTATIVE CONCENTRATION* and *HOT SPOT* identification, this approach should be considered during the characterization stage in areas where no previous information is available to bias *FACILITY* characterization samples. Furthermore, to the extent that a *HOT SPOT* was missed in the characterization stage, this approach may help in the identification of remaining *HOT SPOTS*.

2.4.2.2 Systematic *RANDOM* Sampling

A preferable alternative to simple *RANDOM* sampling is systematic *RANDOM* sampling. This sampling design consists of dividing the total area to be sampled into subsections based on the number of samples to be collected (e.g., for nine samples, divide the total area into nine subsections) and *RANDOMLY* selecting a starting point within the first subsection. Subsequent sampling locations are then identified on a grid that is anchored at the starting point. The grid nodes represent locations to be sampled and all nodes are located based on the first *RANDOMLY*-selected location.

Figure 2.3 illustrates how systematic *RANDOM* sampling works. First, a 300 ft x 300 ft area was divided into nine subsections of equal area (100 ft x 100 ft). A point was *RANDOMLY* selected from the lower left cell of the *EXPOSURE UNIT*. Subsequent samples were identified systematically using a 100 ft grid extended from the first point.

Figure 2.3 Systematic *RANDOM* Sample of Nine Observations Collected from a 300 ft x 300 ft *EXPOSURE UNIT*.



The shaded area represents the cell from which the first sample was *RANDOMLY* selected.

The advantages of this design are numerous. First, it results in a *RANDOMIZED* sample from the population, thus satisfying the statistical requirements. Second, once the first location has been selected, locating the remaining sample locations is relatively straightforward and doesn't require a computer or GPS system. Third, because the coverage is fairly uniform, most of the *EXPOSURE UNIT* will be sampled and the likelihood of missing a large *HOT SPOT* is reduced. Furthermore, since sample locations are identified using a grid, statistical tools described in the tabbed section titled, "Identification and Consideration of *HOT SPOTS*," may be used to estimate the size of a *HOT SPOT* that might be identified (or missed) using this sampling approach. However, there is a danger that, if contamination occurs with some pattern, samples located on

a grid could systematically "miss" the contamination. If this is a concern, use of an unaligned grid (Gilbert, 1987, page 93) should be considered.

To determine the grid spacing, first determine the number of samples that are to be collected. The recommended minimum number is nine, based on statistical considerations only. Additional samples may be necessary to adequately represent spatial variability in the *EXPOSURE UNIT*. Next, use the following equation to determine an approximate grid interval:

$$\text{Grid Interval} = \sqrt{\frac{Area}{n}}$$

Where *Area* represents the total area of the *EXPOSURE UNIT* and *n* represents the number of samples that are to be collected.

The grid interval equation given above provides a rough approximation to a reasonable grid interval. The unique shape and size of the *EXPOSURE UNIT*, as well as the number of samples to be collected, will influence what the appropriate grid interval should be and where samples should be collected. The aim of systematic *RANDOM* sampling is to evenly cover the sampled area while collecting a *RANDOM* sample. Judgment must be used to decide on a sampling plan that is appropriate for individual *EXPOSURE UNITS*.

Example 2.9 Systematic *RANDOM* Sampling

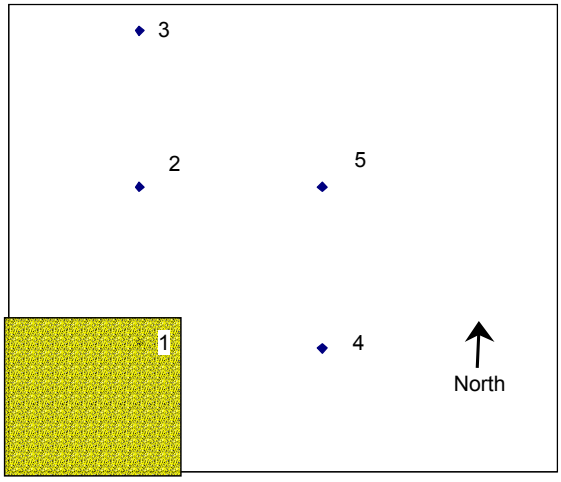
Figure 2.4 represents a square *EXPOSURE UNIT* of approximately two acres (i.e., 300 ft x 300 ft). Suppose that nine samples are to be collected. Using the above equation, the resulting grid interval is 100 ft. The *EXPOSURE UNIT* was divided into nine subsections of equal area (100 ft x 100 ft) and a sample location was *RANDOMLY* selected from the 100 ft x 100 ft cell in the lower left corner of the *EXPOSURE UNIT*. Based on the point selected, subsequent points are collected at the nodes of a grid with the grid interval equal to 100 ft.

The initial *RANDOM* sample location within the 100 x 100 ft cell in the southwest corner of the *EXPOSURE UNIT* was obtained as follows (which corner you start from is irrelevant, but for the sake of consistency, we recommend beginning in the southwest corner). First, generate two *RANDOM* numbers between 0 and 100 using the Microsoft Excel function:

=RANDBETWEEN(L,U)

where L is the lower number (set to 0 in this case) and U is the upper number (set to 100 in this case). The results were 80 and 94. Starting from the southwest corner of the *EXPOSURE UNIT*, move 80 ft east and 94 ft north to establish the *RANDOM* starting point within the southwest cell. The remaining eight sample locations are then positioned at the nodes of a grid with a 100 ft grid interval. For example, the second sample would be located 100 ft north of the first and the third would be located 200 ft north of the first. The fourth would be located 100 ft east of the first, and the fifth would be located 100 ft north of the fourth. This process would continue until all nine systematic *RANDOM* sample locations had been identified.

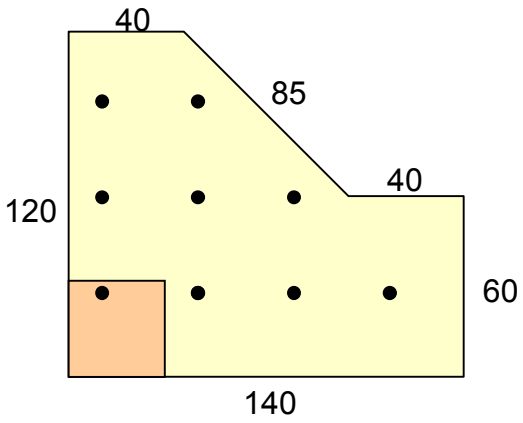
Figure 2.4 Systematic *RANDOM* Sampling of a Two Acre *EXPOSURE UNIT*



Example 2.10 Systematic *RANDOM* Sampling of an Odd-Shaped *EXPOSURE UNIT*

Figure 2.5 represents an odd-shaped *EXPOSURE UNIT* with the dimensions as listed (in feet). The first step is to determine the total area of the polygon, which is 12,600 ft^2. Suppose that nine samples need to be collected by systematic *RANDOM* sampling. Using the above equation, the approximate grid interval is 37 ft. To determine the *RANDOM* starting point, we generate two *RANDOM* numbers, 9 and 27. Starting in the lower-left corner of the polygon, we move nine ft east and 27 ft north to establish the first sample location. Subsequent samples are located on a grid anchored on the first point with a grid interval of 37 ft. If we denote the lower-left corner of the polygon as the (0,0) point on a (x, y) coordinate plane, the nine sample locations depicted are at (9,27), (9,64), (9,101), (46,27), (46,64), (46,101), (83,27), (83,64), and (120,27). Professional judgment may be used to increase the number of samples to be collected if nine samples do not appear to provide adequate coverage of an odd-shaped *EXPOSURE UNIT*.

Figure 2.5 Systematic *RANDOM* Sampling of an Odd-Shaped *EXPOSURE UNIT*



2.4.2.3 Three-Dimensional Sampling

There is often a need to sample a volume of soil rather than simply the surface soils. Sampling a volume of soil involves three-dimensional sampling. That is, rather than sampling in two dimensions only (e.g., North-South and East-West), a third dimension is incorporated (i.e., depth). In many cases, the vertical component is not sampled *RANDOMLY*, but rather based on knowledge of the underlying geology. For example, vertical sampling intervals may be selected so that samples are collected at an interface between two differing soil horizons.

The design of a three dimensional sampling plan (or any sampling plan for that matter) depends on how the population is defined. For example, if the population of concern is a fill layer from three-six feet in an *EXPOSURE UNIT*, then sampling surface soils in the *EXPOSURE UNIT* would be an improper sampling design. The sampling plan should be designed in such a way that representative samples from the intended population are collected.

If the goal is to use statistics to estimate a *REPRESENTATIVE CONCENTRATION* of hazardous substances in a defined layer, *RANDOM* sampling vertically as well as horizontally within the defined layer should be considered. That is, three-dimensional sampling should be considered in which systematic *RANDOM* sampling is completed in the horizontal dimension with an additional *RANDOM* component to sample depth. This approach will be most useful for evaluation of fill materials in which contamination is generally distributed throughout and differing soil horizons are not of a concern. **If initial *FACILITY* characterization shows that contamination is not homogeneous in the vertical dimension, this approach should not be used and sampling should be focused on depths of likely *RELEASES*.**

Using the methods outlined above, first define horizontal sampling locations. At each sample location, *RANDOMLY* select a depth for collecting the sample. The Microsoft Excel function shown below may be used to generate uniform *RANDOM* numbers between an upper and lower limit. This function is:

=RANDBETWEEN(L,U)

Where L is the lower limit and U is the upper limit for the *RANDOM* numbers. For example, if soils from 0-3 ft (i.e., 0-36 inches) are the intended population within an *EXPOSURE UNIT*, then this function can be used to generate a set of *RANDOM* depths from which to obtain samples. Since vertical sampling is often completed in six-inch intervals, the total depth of 36 inches can be divided into six intervals of six inches each. The above function can then be used to *RANDOMLY* select one of the six intervals (i.e., set L = 1 and U = 6) for sampling at each grid node.

An important caveat of three-dimensional sampling is the sample density (the number of samples per unit of volume or area). Sampling a volume (three dimensions) without increasing the total number of samples results in a lower sample density compared to sampling a two-dimensional area. If the sample density is too low, then the power to make statistical inferences is greatly reduced. Therefore, adequately sampling a volume will generally require more samples to be collected than would a two-dimensional sampling plan, particularly if the depth over which samples are to be collected is large.

For example, consider a two-dimensional systematic sampling plan where 10 samples from the top inch of soil are collected over a 25 x 30 ft area. This results in a sample density of approximately 0.16 samples/ft^3. Now contrast this with a three dimensional systematic sampling

plan where 10 samples from the top six inches of soil are collected over the same 25 x 30 ft area. This would result in a sample density of approximately 0.03 samples/ft^3. In fact, a total of 60 samples would need to be collected in the three-dimensional case to attain the same sample density as the two-dimensional case. This lower sample density in this three-dimensional sampling scenario reduces the relative power of the statistical inference about the sampled population.

2.4.2.4 Use of Site Characterization Data in Place of *RANDOM* Sample Locations

As previously stated, *RANDOM* sampling is important if accurate conclusions are to be drawn from a statistical analysis of data. *FACILITY* characterization data are typically biased, based on historical information such as previous sampling, past practices, visual impacts, and aerial photos. However, it may be impractical and cost prohibitive to sample every *EXPOSURE UNIT* from scratch using a *RANDOM* sampling design without use of existing data, where appropriate. Consequently, the following guidelines are recommended with regard to use of existing characterization data for a statistical analysis.

Divide the *EXPOSURE UNIT* into at least nine cells as described under Systematic *RANDOM* Sampling. (The number of cells is determined by the number of samples to be collected in the *EXPOSURE UNIT*.) If a single previous characterization sample location is present in a cell and the sample is representative of current conditions (e.g., the sample was collected from an area that has not been actively remediated), it may not be necessary to collect a new *RANDOM* sample from that cell. However, the sample must be comparable in terms of sampling and analytical methods and detection limits before making this determination and including the analytical result in the statistical analysis.

If more than one site characterization sample was collected from a cell, some judgment must be used to determine which data to include in the statistical analysis. In some cases, it will be appropriate to include all existing data from that cell in the statistical analysis; however, in some cases it will be appropriate to include only some of the data or to select one value to include in the statistical analysis. This determination must be made on a case-by-case basis.

The nature of the previous sampling is the main factor that needs to be considered when evaluating how much of the previous data can be included. For example, suppose that characterization samples present in a given cell are not clustered and are not associated with a *HOT SPOT*. Consider including each of the results in the statistical analysis because, presumably, the samples were collected independently and are representative of the overall variation in that area. However, if the characterization samples were highly clustered in a small area, and were collected for the purpose of investigating and/or confirming sample results in that limited area, the results may not represent the overall variability of concentrations in the *EXPOSURE UNIT.* Additional samples may be necessary. In this case, it would also be more appropriate to select a single value from the cluster of samples to avoid placing a disproportionately large weight on samples collected in a limited area.

It is also important to evaluate the extent to which the samples provide duplicative information. If sample analytical results are spatially correlated, the information provided by these samples is not independent, as required by the statistical methods presented in the S^3TM. Independence is necessary since, if sample results are correlated over space, the effective sample size is reduced. Each correlated sample does not provide as much “new” information about hazardous substance concentrations in an *EXPOSURE UNIT* because its value is partially determined by the value of adjacent observations.

Consequently, the following approach is recommended (alternate approaches may be acceptable on a case-by-case basis):

1. Divide each *EXPOSURE UNIT* into nine or more cells, as previously described.

2. Qualitatively evaluate the number of independent sample locations in each cell. Consider samples collected in a cluster to represent one sample location. At least one independent sample location is necessary in each cell if statistics are to be used to compare data from the *EXPOSURE UNIT* to criteria.

3. Select data for inclusion in the statistical analysis. Samples collected in a cluster: select the original sample result only for inclusion in the statistical analysis. That is, do not include any of the confirmatory sample results. Samples not collected in a cluster: consider including each sample in the statistical analysis. Note: Including more than one sample from a cell will result in more than nine samples from an *EXPOSURE UNIT* for statistical analysis, since at least one sample from each cell is necessary. Areas with spatially correlated concentrations above criteria (i.e., *HOT SPOTS*) must generally be excluded from the statistical analysis and addressed separately.

2.4.3 Comparison of *FACILITY* Data to *FACILITY-SPECIFIC BACKGROUND* Concentrations

When Part 201 criteria are established as *BACKGROUND* concentrations, the objective becomes to determine whether the *FACILITY* concentrations are significantly higher than *BACKGROUND* concentrations for a hazardous substance. To make this determination in most cases, *FACILITY* data will be compared to *FACILITY-SPECIFIC BACKGROUND* concentrations on a point-by-point basis. That is, concentrations of each hazardous substance in each *FACILITY* sample will be compared to directly to the *FACILITY-SPECIFIC BACKGROUND* concentration and individual exceedances will be noted. This will most often be the case when *FACILITY* samples are collected in a biased manner. However, it may also be necessary to compare *FACILITY* data collected using statistical sampling strategies to *FACILITY-SPECIFIC BACKGROUND* concentrations on a point-by-point basis.

If statistics are used to compare *FACILITY* data to *FACILITY-SPECIFIC BACKGROUND* data, a "two-sample test" may be used (i.e. a test that compares two data sets to each other). See Chapter 4 of the tabbed section titled, "Statistical Methods" for additional information. The appropriateness of using a statistical method to compare *FACILITY* data to *FACILITY-SPECIFIC BACKGROUND* data must be evaluated. This depends on: 1) the type of *BACKGROUND* being considered, 2) the manner in which *FACILITY* data were collected (i.e., using a biased or statistical approach), and 3) whether a statistical analysis of *FACILITY* data for comparison to *BACKGROUND* is appropriate. Statistical Guidesheets for the applicable pathways/conditions should be consulted to determine the applicability of statistics for comparing *FACILITY* data to Part 201 criteria and key considerations for selection of the appropriate data set(s). Recommendations provided in the Statistical Guidesheets apply to all Part 201 criteria, including *BACKGROUND*. A statistical analysis of *FACILITY* data for comparison to *BACKGROUND* will most likely be appropriate for pathways/conditions categorized as **YES**, or for those categorized as **GNP** for which a statistical analysis is demonstrated to be appropriate.

For additional information on statistical comparisons to *BACKGROUND*, see Section 2.2.2 of Sampling Strategies and Chapter 4 of the tabbed section titled, "Statistical Methods."

2.4.4 Recommended Summary Report Format

Soil Summary Reports for sites utilizing statistical sampling strategies must identify the number and location of samples and provide the justification of the sample locations selected (why and how). The Summary Report should contain site maps and cross-sections, drawn to scale, which depict the area and volume of contaminated soil; any remediation area(s); the locations of any air/fluid injection and/or extraction wells with their estimated zones of influence; the sampling grids; and the verification soil boring locations and vertical sampling intervals. The Summary Report should also include the calculations for determining the sampling grid intervals, a statement documenting the *RANDOM* sampling strategy utilized for selection of the sampling locations, the number of soil borings/samples and any statistical calculations used to evaluate the sample data.

The checklist below identifies items recommended to properly evaluate a closure certification. These items are not "absolutes." Other information or substitutions may be provided which technically justify and certify a "clean closure."

The verification report must include the following:

1. MAP(s) and CROSS SECTIONS
   Provide a scaled map of the investigated and remediated area (i.e. the estimated *RELEASE* area, or floor and walls of an excavation, or the vertical and horizontal area treated for in situ remediations, etc.) with sample locations identified. If a cross section is utilized to display the remediation activities and data, it should show the relation of key elevations, depict the stratigraphy, fractures, soil types, discolorations, unusual characteristics, possibly indicate the original *RELEASE* source location, sample locations/elevations, etc.

2. SAMPLE LOCATION RATIONALE
   - All sample locations (including *BACKGROUND* samples, *FACILITY* samples, and/or verification samples)
   - Sample depths
   - Sample collection procedures
   - Basis of sampling biases (where used) and the rationale used for collecting each sample (e.g., clay fractures, discolored soil, location of leak in tank)

3. DATA ANALYSES
   - Analytical parameters
   - Analytical methods used
   - Method detection limits
   - Laboratory Quality Assurance/Quality Control
   - Summary of decontamination procedures

4. STATISTICAL ANALYSES
   - Lab results/data tabulation
   - Complete statistical calculations as described in the tabbed section titled, “Statistical Methods”
   - If considered, development of *BACKGROUND* concentrations and/or statistical comparisons to *BACKGROUND* as described in Chapter 4 of the tabbed section titled, “Statistical Methods”

- Narrative explanation of all statistical calculations completed using *FACILITY* and/or *BACKGROUND* concentrations
- Comparisons of *FACILITY* data to cleanup criteria and/or *BACKGROUND* (either point-by-point or statistical)
- Completed worksheet for each statistical comparison (see the tabbed section titled, "Statistical Analysis Worksheets")

5. CONCLUSIONS
   This portion of the report should include a summary of activities and final conclusions, accounting for the work and testing completed and how completed activity(ies) fit in with the site-wide remediation plan (i.e. land use issues, possible closure requirements, whether there is a Remedial Action Plan, resumption of operations, etc.)

2.5 WASTE CHARACTERIZATION

Characterization of wastes for the purpose of disposal must often be addressed at Part 201 *FACILITIES*. The regulatory context for waste characterization is different than the context described in previous sections of this chapter, however. Sections 2.1 through 2.4 describe sampling and analysis of data for the purpose of demonstrating compliance under Part 201. Waste characterization is regulated under Parts 111 and 115. Consequently, sampling strategies and statistical analysis of data for this purpose have been described in a separate tabbed section titled, "Waste Characterization."

SUMMARY

Selection of a sampling plan to collect environmental samples is a complex, multifaceted, and multi-phased problem. As pointed out during the previous discussions of sampling strategies, the ***goal*** of any sampling plan is to collect samples that represent the environmental media and conditions being defined.

The beginning point is usually a conceptual model of the environmental conditions that exist at a property or *FACILITY* based on historical records, previous sampling, or site reviews. An evaluation of the completeness of the conceptual model is made to determine if existing information allows a reasonable model or if further characterization is needed to conceptualize the environmental conditions at the property or *FACILITY*.

If additional sampling is to be conducted, an objective should be set for the sampling plan (e.g., *FACILITY* characterization, identification of *RELEASE* areas, verification of remediation, comparison to regulatory criteria, waste characterization). This plan may involve a combination of sampling techniques (i.e., biased and/or statistical). The sampling plan should be guided by answering "why, what, when, where, and how" to sample. A continuous and sequential review and evaluation of the resulting data should be made to determine if the goal of the sampling plan has been met. Rarely is this sequence a straightforward, single-pathway, or simple analysis. Each sample as it is collected, analyzed, and evaluated may change the conceptual model and the overall sampling strategy. For example, identification of a previously unknown *HOT SPOT* may lead back to further characterization.

As new data become available, an iterative process should be conducted to continuously evaluate whether the data make sense in light of existing data and the conceptual model of the environmental conditions and whether characterization is complete.

The following example illustrates the iterative nature of this process and use of a combination of sampling techniques.

1) A site being investigated has a wealth of existing data (e.g., existing sampling locations and analytical results, documented site history, detailed and accurate information on operational procedures and equipment location). Therefore, additional site characterization will be completed using a biased sampling approach based on known conditions.
2) Sample results indicate that concentrations of one or more hazardous substances are highly elevated at a sample location. It must be decided whether to further characterize at this location by:
   a. continuing biased sampling, or
   b. initiating a statistical sampling plan.

The conceptual model of environmental conditions had been incomplete. To decide on a next step, it is necessary to evaluate why an unexpectedly high concentration was present at the sample location. Was it caused by a discrete release that could be documented as a result of further research? If so, biased sampling may be useful to focus on that location to identify the nature and extent of contamination. If not, it must be concluded that high concentrations may be present at more locations. If no reliable information is available with which to bias additional sampling, a statistical sampling strategy should be used to better characterize the area.

Summary

The decision was made to redirect the sampling plan to use a statistical sampling approach. Although some additional areas with elevated concentrations were identified as a result, none were confirmed to be *HOT SPOTS*. Biased samples were collected by stepping out around each elevated location to make this determination.

The next step is to make a determination on compliance with cleanup criteria. Since locations are present at which criteria are exceeded but none of these locations represented *HOT SPOTS*, it may be appropriate to consider a statistical analysis for comparison to criteria. However, before proceeding it is necessary to first evaluate several key considerations including the applicability of statistics for the exposure pathway(s) being evaluated, adequacy of characterization, and selection of the appropriate data set to represent exposures and hazardous substance concentrations in an appropriate exposure area. These concepts are addressed throughout Sampling Strategies (particularly Section 2.4.1) as well as in the tabbed section titled, "Statistical Guidesheets."

PART 201 STATISTICAL EVALUATION GUIDESHEETS

LIST OF PART 201 STATISTICAL EVALUATION GUIDESHEETS

A. Abandoned Substances Not Yet Dispersed and Free Phase Liquid
B. Soil: Contaminated Soil Runoff to Surface Waters
C. Surface Water Sediment: Aquatic Flora/Fauna/Food Chain Hazards and Aesthetics
D. Acute Toxicity and Physical Hazards: Acute Inhalation Toxicity and Flammability/ Explosivity, Corrosivity, Ignitability, and Reactivity
E. Ecological and Aesthetic Impacts: Terrestrial Flora, Fauna, Food Chain, Aesthetic or Other Impacts
F. Asbestos Containing Materials: Asbestos

GROUNDWATER

1. Generic Residential and Commercial I Drinking Water Criteria (DWC)
2. Generic Commercial II, III, IV and Industrial Drinking Water Criteria (DWC)
3. Generic and Mixing Zone-Based Groundwater Surface Water Interface (GSI) Criteria
4. Generic Residential and Commercial I Groundwater Volatilization to Indoor Air Inhalation Criteria (GVIIC)
5. Generic Commercial II, III, IV and Industrial Groundwater Volatilization to Indoor Air Inhalation Criteria (GVIIC)
6. Generic Groundwater Contact Criteria (GCC)
7. Water Solubility
8. Generic Screening Levels for Flammability and Explosivity
9. Generic Acute Inhalation Toxicity Screening Levels

SOIL

10. Soil Background
11. Generic Soil Criteria Protective of Residential and Commercial I Drinking Water
12. Soil Criteria Protective of the Groundwater Surface Water Interface (GSI)
13. Generic Soil Criteria Protective for Groundwater Contact
14. Generic Residential and Commercial I Soil Volatilization to Indoor Air Inhalation Criteria (SVIIC)

15.-17. Generic Residential and Commercial I Infinite and Finite Volatile Soil Inhalation Criteria (VSIC) for Ambient Air

18. Generic Residential and Commercial I Particulate Soil Inhalation Criteria (PSIC) for Ambient Air
19. Generic Residential and Commercial I Direct Contact Criteria (DCC)
20. Generic Soil Saturation (CSAT) Screening Levels
21. Generic Soil Criteria Protective of Commercial II, III, IV and Industrial Drinking Water
22. Generic Commercial II, III, IV and Industrial Soil Volatilization to Indoor Air Inhalation Criteria (SVIIC)

23.-25. Generic Commercial II, III, IV and Industrial Infinite Source Volatile Soil Inhalation Criteria (VSIC) for Ambient Air

26. Generic Commercial II, III, IV and Industrial Particulate Soil Inhalation Criteria (PSIC) for Ambient Air

27.-29. Generic Industrial and Commercial II, III, IV Direct Contact Criteria (DCC)

INTRODUCTION TO PART 201 STATISTICAL EVALUATION GUIDESHEETS

Statistical Guidesheets are lettered or numbered to correspond with the Criteria Application Guidesheets. Words that are in capital letters and italicized are defined terms that are presented in the tabbed section titled, "Acronyms / Glossary." Bullets presented as ★ are used to highlight considerations that are unique to the exposure pathway or condition.

The Statistical Guidesheets have been organized into sections to focus attention to the key considerations for each purpose for evaluating data. Section 1 of the Statistical Guidesheet addresses the use of statistical analysis for determining if a property is a *"FACILITY."* Section 2 identifies the key factors to consider when selecting data to include in a statistical analysis for the purpose of deriving a *REPRESENTATIVE CONCENTRATION* for comparison to cleanup criteria. Section 3 highlights the factors to consider when making comparison to the cleanup criteria, and Section 4 specifies the important considerations when using statistical analyses to demonstrate a verification of remediation or closure.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR **SOIL AND GROUNDWATER** DATA	Statistical Guidesheet A
ABANDONED SUBSTANCES NOT YET DISPERSED AND FREE PHASE LIQUID	
May assist in evaluation of condition(s):	1
Applicability of Statistics	**GNP**

Statistics are generally not practical as a tool for assessing this condition because source control evaluations will generally be conducted on a qualitative basis.

Key Considerations:		
	Section 1	**Facility Determination**
		★ The determination that property is a *FACILITY* is generally based on data which show that one or more generic residential cleanup criterion is exceeded. Information that documents the presence of abandoned hazardous substances which are not yet dispersed, or the presence of free phase liquid hazardous substances that have been *RELEASED* is also potentially sufficient to support a conclusion that property is a *FACILITY*. Consult your supervisor for guidance if necessary. Since *FACILITY* determinations that are based on the presence of abandoned sources or free phase liquids are generally not based on review of laboratory data, statistics are not expected to be relevant to those *FACILITY* determinations.
	Section 2	**Selection of Data For Statistical Analysis**
		• Not applicable - see Section 3.
	Section 3	**Risk Analysis – Comparison to Criteria**
		★ Since source control evaluations will generally be conducted on a qualitative basis (i.e., not based on rigorous comparison of *FACILITY* data to cleanup criteria or another value), statistics are not expected to be necessary.
	Section 4	**Verification of Remediation or Closure**
		• Not applicable.

Additional Information:	Criteria Application Guidesheet A and Appendix A (Part 201 Cleanup Criteria Training Materials: Source Control Obligations for Part 201 Facilities.)
Recommended Statistical Methods for Comparison to Criteria:	Not applicable.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR SOIL DATA	Statistical Guidesheet B
CONTAMINATED SOIL RUNOFF TO SURFACE WATERS	
May assist in evaluation of condition(s):	12
Applicability of Statistics	**GNP**
Statistics are generally not practical as a tool for assessing this condition because there are no generic cleanup criteria.	

Key Considerations:	**Section 1**	**Facility Determination**
		• The determination that property is a *FACILITY* is generally based on data which show that one or more generic residential cleanup criterion is exceeded. Since there are not generic cleanup criteria available for this condition, *FACILITY* determinations will not generally be based on this condition. If you are evaluating property where you suspect problems with this condition and no data are available to demonstrate that the property is a *FACILITY* for other reasons, consult your supervisor for guidance.
	Section 2	**Selection of Data For Statistical Analysis**
		• Not applicable - see Section 3.
	Section 3	**Risk Analysis – Comparison to Criteria**
		• Since there are no generic cleanup criteria available for this condition, statistical analysis of data is not generally expected to be used for risk analysis.
	Section 4	**Verification of Remediation or Closure**
		• Since there are no generic cleanup criteria available for this condition, statistical analysis of data is not generally expected to be used for verification of remediation or closure.

Additional Information:	Criteria Application Guidesheet B
Recommended Statistical Methods for Comparison to Criteria:	The statistical method for comparing *FACILITY* data to criteria will be determined on a case-by-case basis.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR SURFACE WATER SEDIMENT DATA	Statistical Guidesheet C
AQUATIC FLORA/FAUNA/FOOD CHAIN HAZARDS AND AESTHETICS	
May assist in evaluation of condition(s):	13
Applicability of Statistics	**GNP**

Statistics are generally not practical as a tool for initially assessing this condition because there are no generic sediment cleanup criteria. If the condition is determined to require response activity and site-specific criteria are developed, statistical evaluation of data may be possible. See Sections 2 and 3 below.

Key Considerations:

Section 1

Facility Determination

★ The determination that property is a *FACILITY* is generally based on data which show that one or more generic residential cleanup criterion is exceeded. Since there are no generic cleanup criteria available for this condition, *FACILITY* determinations using numerical data will not generally be based on this condition. However, site-specific sediment cleanup criteria may be developed for a *FACILITY* determination and Section 20a(17) allows for a qualitative *FACILITY* determination based on unacceptable risk to surface water and sediment. If you are evaluating property where you suspect problems with this condition and no data are available to demonstrate that the property is a *FACILITY* for other reasons, consult your supervisor for guidance.

Section 2

Selection of Data For Statistical Analysis

★ Generic cleanup criteria are not available for this condition; all sediment cleanup criteria are developed on a site-specific basis. This statistical guidesheet provides information on selecting data for comparison to site-specific criteria when available.

- Samples collected for the purpose of characterizing a *FACILITY* are typically biased, based on factors such as historical information, previous sampling, disposal practices, visual impacts, and aerial photos. Once the nature and extent of contamination has been defined, it is necessary to identify and/or obtain data that will allow for appropriate comparison to criteria. There are two primary considerations in determining if data sets are adequate. First, data sets must be obtained from locations that are consistent with the exposure assumptions for the relevant land use scenario. Second, if statistics are used, data sets must contain a sufficient number of *RANDOMLY* located sample results to allow for proper statistical analysis and development of *REPRESENTATIVE CONCENTRATIONS*. Therefore, additional sampling to develop *REPRESENTATIVE CONCENTRATIONS* will often be required after the nature and extent of contamination have been defined. Some characterization data may be used in the development of a *REPRESENTATIVE CONCENTRATION* as described in Section 2.4.2 of the tabbed section titled, "Sampling Strategies."

- A statistical analysis of soil data should be completed only if the nature and extent of contamination, including any *HOT SPOTS*, has been adequately defined.

	Section 2: (Continued)	• It is not appropriate to combine samples from *HOT SPOT* areas with samples from other areas of a property for statistical analysis. This is necessary to avoid averaging or diluting the samples that represent *HOT SPOTS*. *HOT SPOTS* must be addressed separately. See Section 2.2.2.1 of the tabbed section titled, "Sampling Strategies." ★ The selection of sediment characterization data that are appropriate for statistical analysis will depend on the exposures or other impacts that are the basis for the sediment cleanup criteria. See Guidesheet C in the Part 201 Cleanup Criteria Training Materials and Rule 717(5)(a) to (l) for information about the factors that must be considered in establishing sediment cleanup criteria.
	Section 3	**Risk Analysis – Comparison to Criteria** ★ Cleanup criteria for contaminated sediments will be established by the department, if necessary, based on the factors described in Rule 717(5)(a) to (l). The appropriateness of statistics as a tool for determining compliance will depend on which of the factors in that rule were used as the basis for each site-specific criterion (e.g., if adverse aesthetics is the basis for a sediment cleanup criterion, statistics may not be applicable). • If DEQ approval of a response activity is being sought, a *PROPOSAL* for a statistical analysis must be submitted to the DEQ for approval to assure that data needs and/or complexities of the pathway are addressed. *PROPOSALS* for the use of statistics must include a justification that relates the *PROPOSAL* to the basis for the cleanup criteria (e.g., mitigating degradation of benthos, or eliminating fish consumption restrictions).
	Section 4	**Verification of Remediation or Closure** ★ Cleanup criteria for contaminated sediments will be established by the Department, if necessary, based on the factors described in Rule 717(5)(a) to (l). The appropriateness of statistics as a tool for determining the adequacy of remedial actions will depend on which of the factors in that rule were used as the basis for each site-specific criterion (e.g., if adverse aesthetics is the basis for a sediment cleanup criterion, statistics may not be applicable). ★ If DEQ approval of a response activity is being sought, a *PROPOSAL* for a statistical analysis must be submitted to the DEQ for approval to assure that data needs and/or complexities of the pathway are addressed. *PROPOSALS* for use of statistics to evaluate sediment data for verifying remediation or closure must include a justification that relates the *PROPOSAL* to the basis for the cleanup criteria (e.g., mitigating degradation of benthos, or eliminating fish consumption restrictions).

Additional Information:	Criteria Application Guidesheet C
Recommended Statistical Methods for Comparison to Criteria:	The statistical method for comparing *FACILITY* data to criteria will be determined on a case-by-case basis.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR ACUTE TOXICITY AND PHYSICAL HAZARDS	**Statistical Guidesheet D**

ACUTE INHALATION TOXICITY AND FLAMABILITY/EXPOLSIVITY, CORROSIVITY, IGNITABILITY, AND REACTIVITY

May assist in evaluation of condition(s):	14
Applicability of Statistics	**GNP**

Statistics are generally not practical as a tool for assessing this condition because it deals with acute risks and any exceedance of screening levels requires further consideration.

Key Considerations:		
	Section 1	**Facility Determination** • The determination that property is a *FACILITY* is generally based on data which show that one or more generic residential cleanup criterion is exceeded. Since there are no generic cleanup criteria available for this condition, *FACILITY* determinations will not generally be based on this condition. If you are evaluating property where you suspect problems with this condition and no data are available to demonstrate that the property is a *FACILITY* for other reasons, consult your supervisor for guidance.
	Section 2	**Selection of Data For Statistical Analysis** • Not applicable – see Section 3.
	Section 3	**Risk Analysis – Comparison to Criteria** ★ Since this condition deals with acute risks, any exceedance of flammability/explosivity or acute inhalation screening levels requires further consideration. See Statistical Guidesheets 8 and 9 for additional discussion about application of screening values for acute effects. • Statistical treatment of data is not expected to be a practical tool for making decisions about the need for response activity related to this condition. • Professional judgment will be required to determine the significance of any data that relates to this condition. The quantity of the hazardous substance that is present above potential levels of concern is a factor that determines whether there may be an unacceptable risk.
	Section 4	**Verification of Remediation or Closure** • See Section 3.

Additional Information:	Statistical Guidesheets 8 and 9; Criteria Application Guidesheets D, 8 and 9.
Recommended Statistical Methods for Comparison to Criteria:	The statistical method for comparing *FACILITY* data to criteria will be determined on a case-by-case basis.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR **ECOLOGICAL AND AESTHETIC IMPACTS**	Statistical Guidesheet **E**

TERRESTIAL FLORA, FAUNA, FOOD CHAIN, AESTHETIC OR OTHER IMPACTS

May assist in evaluation of condition(s):	15
Applicability of Statistics	**GNP**

Statistics are generally not practical as a tool for assessing this condition because there are no generic cleanup criteria.

Key Considerations:		
	Section 1	**Facility Determination**
		• The determination that property is a *FACILITY* is generally based on data which show that one or more generic residential cleanup criterion is exceeded. Since there are no generic cleanup criteria available for this condition, *FACILITY* determinations will not generally be based on this condition. If you are evaluating property where you suspect problems with this condition and no data are available to demonstrate that the property is a *FACILITY* for other reasons, consult your supervisor for guidance.
	Section 2	**Selection of Data For Statistical Analysis**
		• Not applicable - see Section 3.
	Section 3	**Risk Analysis – Comparison to Criteria**
		• Since there are no generic cleanup criteria available for this condition, statistical analysis of data is not generally expected to be used for risk analysis.
	Section 4	**Verification of Remediation or Closure**
		• Since there are no generic cleanup criteria available for this condition, statistical analysis of data is not generally expected to be used for verification of remediation or closure.

Additional Information:	Criteria Application Guidesheet E
Recommended Statistical Methods for Comparison to Criteria:	The statistical method for comparing *FACILITY* data to criteria will be determined on a case-by-case basis.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR CONTROL OF **ASBESTOS CONTAINING MATERIALS**	Statistical Guidesheet **F**

ASBESTOS

May assist in evaluation of condition(s):	2, 6, 7, 12
Applicability of Statistics	**Pathway Dependent**

See Applicability of Statistics discussion on Statistical Guidesheet for relevant exposure pathway.

Key Considerations:		
	Section 1	**Facility Determination**
		• A *FACILITY* determination based on the presence of asbestos would generally depend on an exceedance of a generic residential cleanup criterion (e.g., drinking water or particulate soil inhalation). See relevant exposure pathway statistical guidesheets for further guidance on making *FACILITY* determinations. Also see Statistical Guidesheet A for discussion of the ways in which abandoned substances that are not yet dispersed may constitute a *FACILITY*.
	Section 2	**Selection of Data For Statistical Analysis**
		• See statistical guidesheets for relevant exposure pathways.
	Section 3	**Risk Analysis – Comparison to Criteria**
		• See statistical guidesheets for relevant exposure pathways.
	Section 4	**Verification of Remediation or Closure**
		• See statistical guidesheets for relevant exposure pathways.

Additional Information:	Statistical Guidesheets 1, 2, 18, and 26; Criteria Application Guidesheets 1, 2, 18 and 26.
Recommended Statistical Methods for Comparison to Criteria:	See statistical guidesheets for relevant exposure pathways.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR **GROUNDWATER** DATA	Statistical Guidesheet 1
GENERIC RESIDENTIAL AND COMMERCIAL I DRINKING WATER CRITERIA (DWC)	
May assist in evaluation of condition(s):	2
Applicability of Statistics	**NO**

Statistics are not applicable for assessing this exposure pathway because Rule 709(3) states that the point of exposure is presumed to be any point in the affected aquifer. As a result, cleanup criteria must be met at all points.

Key Considerations:		
	Section 1	**Facility Determination**
		• For the purposes of a BEA or initial *FACILITY* determination, property may be classified as a *FACILITY* if one or more samples contain concentrations above generic residential criteria or *BACKGROUND* concentrations, whichever is greater, provided that there is not a greater body of evidence that the property is not a *FACILITY*. For purposes other than *FACILITY* determination (e.g., remediation and/or closure), additional data will likely be required.
	Section 2	**Selection of Data For Statistical Analysis**
		• Not applicable – see Section 3.
	Section 3	**Risk Analysis – Comparison to Criteria**
		★ Statistical analysis of *FACILITY* groundwater concentrations for comparison to criteria (either across well locations or over time in an individual well) is not acceptable. Since the point of exposure is presumed to be any point in the affected aquifer [Rule 709(3)], groundwater drinking water criteria must be met at each point in an aquifer. Therefore, evaluation of groundwater concentrations must be completed on a point-by-point basis* (i.e., each concentration at a given time and location must be compared individually to criteria). • If *BACKGROUND* concentrations are greater than criteria for naturally occurring substances, groundwater concentrations may be compared to *BACKGROUND* concentrations. If comparison is made to *BACKGROUND* the comparison must be completed on a point-by-point basis.* See Statistical Guidesheet 10 and Chapter 4 of the tabbed section titled, "Statistical Methods." ★ Statistical analysis of groundwater concentrations over time in an individual well will not be allowed for comparison to criteria or *BACKGROUND*.
	Section 4	**Verification of Remediation or Closure**
		• **No Action Needed**: If the data evaluation completed under Section 3 demonstrates that there are no exceedances of cleanup criteria, no remedial action is required for this pathway and consequently no additional sampling or data analysis is required under Section 4.

	Section 4: (Continued)	• **In situ/Ex situ Treatment**: Plans for verifying remediation or closure must be made on a case-by-case basis and must demonstrate how treatment is effective in meeting cleanup criteria. To demonstrate verification of remediation or closure, concentrations in groundwater must meet criteria during a specified number of consecutive sampling events. The number of consecutive sampling events must be selected on a case-by-case basis to reflect seasonal variation in groundwater quality, flow rates, and initial distribution of contamination in groundwater. Ordinarily, this demonstration should be made using data from at least one year of quarterly sampling. • **Containment**: Generally groundwater sampling used to verify a groundwater containment remedy will be done at the perimeter of the containment system. Statistical analysis is not allowed for reasons described in Section 3. This verification will be an ongoing part of the remedy as long as hazardous substances are present above criteria. • Comparisons to criteria or *BACKGROUND* must be made on a point-by-point basis.* See Section 3.
Additional Information:		Criteria Application Guidesheet 1
Recommended Statistical Methods for Comparison to Criteria:		Not applicable.

*Professional judgment is required to interpret the significance of exceedances that are very close to criteria, or that may be associated with insignificant quantities of hazardous substances.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR **GROUNDWATER** DATA	Statistical Guidesheet 2
GENERIC COMMERCIAL II, III, IV AND INDUSTRIAL DRINKING WATER CRITERIA (DWC)	
May assist in evaluation of condition(s):	2
Applicability of Statistics	**NO**
Statistics are not applicable for assessing this exposure pathway because the point of exposure is presumed to be any point in the affected aquifer. As a result, cleanup criteria must be met at all points.	

Key Considerations:		
	Section 1	**Facility Determination**
		• Not applicable.
	Section 2	**Selection of Data For Statistical Analysis**
		• Not applicable – see Section 3.
	Section 3	**Risk Analysis – Comparison to Criteria**
		★ Statistical analysis of *FACILITY* groundwater concentrations for comparison to criteria (either across well locations or over time in an individual well) is not acceptable. Since the point of exposure is presumed to be any point in the affected aquifer, groundwater drinking water criteria must be met at each point in an aquifer. Therefore, evaluation of groundwater concentrations must be completed on a point-by-point basis* (i.e., each concentration at a given time and location must be compared individually to criteria). • If *BACKGROUND* concentrations are greater than criteria for naturally occurring substances, groundwater concentrations may be compared to *BACKGROUND* concentrations. If comparison is made to *BACKGROUND* the comparison must be completed on a point-by-point basis.* See Statistical Guidesheet 10 and Chapter 4 of the tabbed section titled, "Statistical Methods." ★ Statistical analysis of groundwater concentrations over time in an individual well will not be allowed for comparison to criteria or *BACKGROUND*.
	Section 4	**Verification of Remediation or Closure**
		• **No Action Needed**: If the data evaluation completed under Section 3 demonstrates that there are no exceedances of cleanup criteria, no remedial action is required for this pathway and consequently no additional sampling or data analysis is required under Section 4.

	Section 4: (Continued)	• **In situ/Ex situ Treatment**: Plans for verifying remediation or closure must be made on a case-by-case basis and must demonstrate how treatment is effective in meeting cleanup criteria. To demonstrate verification of remediation or closure, concentrations in groundwater must meet criteria during a specified number of consecutive sampling events. The number of consecutive sampling events must be selected on a case-by-case basis to reflect seasonal variation in groundwater quality, flow rates, and initial distribution of contamination in groundwater. Ordinarily, this demonstration should be made using data from at least one year of quarterly sampling. • **Containment**: Generally groundwater sampling used to verify a groundwater containment remedy will be done at the perimeter of the containment system. Statistical analysis is not allowed for reasons described in Section 3. This verification will be an ongoing part of the remedy as long as hazardous substances are present above criteria. • Comparisons to criteria or *BACKGROUND* must be made on a point-by-point basis.* See Section 3.
Additional Information:		Criteria Application Guidesheet 2
Recommended Statistical Methods for Comparison to Criteria:		Not applicable.

*Professional judgment is required to interpret the significance of exceedances that are very close to criteria, or that may be associated with insignificant quantities of hazardous substances.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR **GROUNDWATER** DATA	Statistical Guidesheet 3
GENERIC AND MIXING ZONE-BASED GROUNDWATER SURFACE WATER INTERFACE (GSI) CRITERIA	
May assist in evaluation of condition(s):	5
Applicability of Statistics	**YES - Chronic; NO - Generic and Acute**

Statistics are applicable for evaluating compliance with chronic mixing zone-based GSI criteria since the "average" impact of hazardous substances in groundwater is considered. Statistics are not applicable for acute mixing zone-based GSI criteria, since point-by-point exceedances of these criteria can result in unacceptable impacts. Statistics are also not applicable for comparison to generic GSI criteria, since these criteria do not take into account any dilution or mixing with receiving waters.

Key Considerations:

Section 1

Facility Determination

- For the purposes of a BEA or initial *FACILITY* determination, property may be classified as a *FACILITY* if one or more samples contain concentrations above generic residential criteria or *BACKGROUND* concentrations, whichever is greater, provided that there is not a greater body of evidence that the property is not a *FACILITY*. For purposes other than *FACILITY* determination (e.g., remediation and/or closure), additional data will likely be required.
- If a statistical analysis is relied upon for making a *FACILITY* determination see Sections 2 and 3.

★ In certain cases, professional judgment may dictate that this pathway is not relevant due to lack of proximity to the surface water and other factors described in Cleanup Criteria Application Guidesheet 3. In this case, these criteria are not applicable and are not the basis for a property being considered a *FACILITY*.

Section 2

Selection of Data For Statistical Analysis

★ Samples collected for the purpose of characterizing a *FACILITY* are typically biased, based on factors such as historical information, previous sampling, disposal practices, visual impacts, and aerial photos. Once the nature and extent of contamination has been defined, it is necessary to identify and/or obtain data that will allow for appropriate comparison to criteria. There are two primary considerations in determining if data sets are adequate. First, data sets must be obtained from appropriately placed wells at the GSI. Second, if statistics are used, data sets must be from the *AVERAGING AREA* and contain a sufficient number of sample results to allow for proper statistical analysis and development of *REPRESENTATIVE CONCENTRATIONS*. Therefore, additional sampling to develop *REPRESENTATIVE CONCENTRATIONS* is often required after the nature and extent of contamination have been defined.

★ A statistical analysis is appropriate only if the nature and extent of contamination, including any *HOT SPOTS*, has been adequately defined. This means that the horizontal and vertical extent of contamination exceeding generic GSI criteria in groundwater, and sources that may result in GSI criteria exceedances in the future, must be defined.

Section 2: (Continued)

The following bullets relate to evaluation of groundwater data from *GSI MONITORING WELLS* only:

★ Statistical analysis of groundwater data is not allowed for comparison to generic GSI criteria or acute mixing zone-based criteria.

★ Statistical analysis of groundwater data from *GSI MONITORING WELLS* is allowed for comparison to chronic mixing zone-based criteria only. Only groundwater data from *GSI MONITORING WELLS* within the *AVERAGING AREA* may be used for statistical comparison to chronic mixing zone-based GSI criteria. The *AVERAGING AREA* is the cross sectional area of the hazardous substance plume used to estimate the discharge rate of venting groundwater in the request for a mixing zone determination. This cross section represents the area in which hazardous substance concentrations exceed or are expected to exceed the generic GSI criteria.

★ It is appropriate to combine samples from *HOT SPOT* areas within the *AVERAGING AREA* with samples from other areas within the *AVERAGING AREA* for statistical purposes when comparing data to chronic mixing zone based criteria.

★ The horizontal limits of the cross sectional area typically extend to the nearest adjacent wells along the GSI in which groundwater concentrations are consistently below generic GSI criteria; it is assumed that groundwater concentrations exceed generic GSI criteria up to these boundary points. Therefore, the horizontal boundaries of the cross sectional area extend up to, but do not include, these adjacent wells. Consequently, the adjacent wells (below generic GSI) should not be included in the *AVERAGING AREA* and data from these wells should not be included in the statistical analysis.

★ Different *AVERAGING AREAS* may be required for different hazardous substances depending on whether more than one cross sectional area was used to estimate discharge rates of venting groundwater in the request for the mixing zone determination.

★ The groundwater data used in the statistical analysis must be representative of the *AVERAGING AREA* and include all monitoring points located within the *AVERAGING AREA*. Samples from a minimum of nine distinct GSI monitoring points must be used in the statistical analysis. This does not necessarily mean nine individual wells (e.g., three screened intervals in each of two wells would yield nine distinct monitoring points). This number is based on statistical considerations only, but may not be practical for groundwater plumes that have very narrow *AVERAGING AREAS*. Additional samples may be necessary to represent spatial variability in the *AVERAGING AREA* depending on such factors as soil type, and size of the *AVERAGING AREA*.

★ Statistical analysis of groundwater data over time is not allowed for comparison to chronic mixing zone-based criteria. The statistical analysis must generally be completed using data from a single sampling event. If resampling is conducted to confirm the presence of elevated hazardous substance concentrations in one or more individual wells, results from resampling should not be incorporated into the statistical analysis. However, all wells in the *AVERAGING AREA* may be resampled for the purpose of a separate statistical evaluation.

Section 3

Risk Analysis – Comparison to Criteria

★ In certain cases, professional judgment may dictate that these criteria are not applicable due to lack of proximity to surface water and other factors as described in Cleanup Criteria Application Guidesheet 3.

	Section 3: (Continued)	★ Comparison of groundwater data to generic GSI criteria and acute mixing zone-based criteria must be completed on a point-by-point basis.* That is, statistical analysis of groundwater concentrations across well locations or over time in an individual well will not be allowed for comparison to these criteria. ★ Whole effluent acute toxicity test results must be compared to the *ONE ACUTE TOXIC UNIT* criterion on a point-by-point basis.* The *ONE ACUTE TOXIC UNIT* has the same regulatory significance as an acute mixing-zone-based GSI criterion. ★ Statistical analysis of groundwater data is allowed for comparison to chronic mixing zone-based criteria only. Only data from *GSI MONITORING WELLS* within the *AVERAGING AREA* may be included in the statistical analysis. The statistical analysis must generally be completed using data from a single sampling event. See Section 2 for a description of the *AVERAGING AREA* and additional details on *FACILITY* characterization data that may be included in the statistical analysis. • Statistical analyses must include an evaluation of the underlying statistical distribution of the data (i.e., normal, lognormal, or neither) and the level of censoring (i.e., proportion of the data below the detection limit). See the tabbed section titled, "Statistical Methods." This evaluation must be completed for each hazardous substance. The statistical method used for comparison of groundwater data to chronic mixing zone-based criteria must be selected for each hazardous substance based on statistical distribution and level of censoring. ★ If *BACKGROUND* concentrations are greater than criteria for naturally occurring substances, groundwater data may be compared to *BACKGROUND* concentrations. This evaluation will generally be made on a point-by-point basis,* except for certain chronic mixing zone-based criteria. See Statistical Guidesheet 10 and Chapter 4 of the tabbed section titled, "Statistical Methods."
	Section 4	**Verification of Remediation or Closure** • For the purposes of verification of remediation or closure, compliance of groundwater data with GSI criteria at the GSI may be determined in the same manner described in Section 3. To demonstrate verification of remediation or closure, concentrations in groundwater must meet criteria during a specified number of consecutive sampling events. The number of consecutive sampling events must be selected on a case-by-case basis to reflect seasonal variation in groundwater quality, flow rates, and initial distribution of contamination in groundwater. ★ **No Action Needed**: If the data evaluation completed under Sections 2 and 3 demonstrated that there are no exceedances of generic GSI criteria, and there is no reason to believe there will be exceedances in the future at the GSI, no remedial action is required for this pathway and consequently no additional sampling or data analysis is required under Section 4. • **In situ/Ex situ Treatment**: For the purposes of verification of remediation or closure, compliance of groundwater data with these criteria may be determined in the same manner described in Section 3. Plans for verifying remediation or closure must be made on a case-by-case basis and must demonstrate how treatment is effective in meeting cleanup criteria. These plans may or may not depend on statistical analysis of data as allowed in Sections 2 and 3. To demonstrate verification of remediation or closure, concentrations in groundwater must meet criteria during a specified number of consecutive sampling events. The number of consecutive sampling events must be selected on a case-by-case basis to reflect seasonal variation in groundwater quality, flow rates, and initial distribution of contamination in groundwater.

	Section 4: (Continued)	• **Containment**: For the purposes of verification of remediation or closure, compliance of groundwater data with these criteria may be determined in the same manner described in Section 3. Generally groundwater sampling used to verify a groundwater containment remedy will be done at the perimeter of the containment system. These plans may or may not depend on statistical analysis of data as allowed in Sections 2 and 3. This verification will be ongoing part of the remedy as long as hazardous substances are present above criteria. ★ **For all the preceding cases**, in addition to demonstrating compliance with GSI criteria at the GSI, it is necessary to demonstrate that groundwater at the GSI will continue to meet GSI criteria for verification of remediation or closure. This demonstration must include an evaluation of upgradient groundwater and soil concentrations and sources. ★ If *BACKGROUND* concentrations are greater than criteria for naturally occurring substances, groundwater data may be compared to *BACKGROUND* concentrations. This evaluation will generally be made on a point-by-point basis,* except for certain chronic mixing zone-based criteria. See Statistical Guidesheet 10 and Chapter 4 of the tabbed section titled, "Statistical Methods."
Additional Information:		Criteria Application Guidesheet 3
Recommended Statistical Methods for Comparison to Criteria:		For chronic mixing zone-based criteria only, compare *FACILITY* data to criteria using a 95% upper confidence limit (UCL) for the mean concentration (see Chapter 3 of the tabbed section titled, "Statistical Methods"). Other statistical methods may be acceptable on a case-by-case basis.

*Professional judgment is required to interpret the significance of exceedances that are very close to criteria, or that may be associated with insignificant quantities of hazardous substances.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR GROUNDWATER DATA	**Statistical Guidesheet 4**

GENERIC RESIDENTIAL AND COMMERCIAL I GROUNDWATER VOLATILIZATION TO INDOOR AIR INHALATION CRITERIA (GVIIC)

May assist in evaluation of condition(s):	4
Applicability of Statistics	**GNP**

Statistics are generally not practical as a tool for assessing this exposure pathway because of the spatial and temporal variability of groundwater hazardous substance plumes and the resulting soil gas plumes that are generated from hazardous substances in groundwater.

Key Considerations:

Section 1

Facility Determination

- For the purposes of a BEA or initial *FACILITY* determination, property may be classified as a *FACILITY* if one or more samples contain concentrations above generic residential criteria, provided that there is not a greater body of evidence that the property is not a *FACILITY*. For purposes other than *FACILITY* determination (e.g., remediation and/or closure), additional data will likely be required.
- If a statistical analysis is relied upon for making a *FACILITY* determination see Sections 2 and 3.

Section 2

Selection of Data For Statistical Analysis

- Samples collected for the purpose of characterizing a *FACILITY* are typically biased, based on factors such as historical information, previous sampling, disposal practices, visual impacts, and aerial photos. Once the nature and extent of contamination has been defined, it is necessary to identify and/or obtain data that will allow for appropriate comparison to criteria. There are two primary considerations in determining if data sets are adequate. First, data sets must be obtained from locations that are consistent with the exposure assumptions for the relevant land use scenario. Second, if statistics are used, data sets must contain a sufficient number of *RANDOMLY* located sample results to allow for proper statistical analysis and development of *REPRESENTATIVE CONCENTRATIONS*. Therefore, additional sampling to develop *REPRESENTATIVE CONCENTRATIONS* will often be required after the nature and extent of contamination have been defined. Some characterization data may be used in the development of a *REPRESENTATIVE CONCENTRATION* as described in Section 2.4.2 of the tabbed section titled, "Sampling Strategies."
- A statistical analysis is appropriate only if the nature and extent of contamination, including any *HOT SPOTS*, has been adequately defined.
- It is not appropriate to combine samples from *HOT SPOT* areas with samples from other areas of a property for statistical analysis. This is necessary to avoid averaging or diluting the samples that represent *HOT SPOTS*. *HOT SPOTS* must be addressed separately. See Section 2.2.1.1 of the tabbed section titled "Sampling Strategies."
- ★ Sample locations with concentrations equal to or greater than the hazardous substance's water solubility must not be included in the data set for statistical evaluation. Areas where data exceed solubility must be addressed separately and cannot be ignored.

	Section 2 (Continued)	★ Generally only groundwater data from a 1,200 ft^2 *EXPOSURE UNIT* (i.e., the building footprint) may be used in a statistical calculation for remedial compliance.
	Section 3	**Risk Analysis – Comparison to Criteria** ★ Statistical analysis of *FACILITY* groundwater data for comparison to GVIIC is generally not practical given the spatial and temporal variability of groundwater plumes and the dynamics of the resulting soil gas plume that is generated by hazardous substances in groundwater. Determining the impact of these factors within a 1,200 ft^2 area (i.e., the building footprint) further complicates application of statistical approaches. As a result, comparison to these criteria will generally be completed on a point-by-point basis* (i.e., each concentration at each location must be compared individually to criteria). • If DEQ approval of a response activity is being sought, a *PROPOSAL* for a statistical analysis must be submitted to the DEQ for approval to assure that data needs and/or complexities of the pathway are addressed. • Statistical analysis of groundwater concentrations over time in an individual well will not be allowed for comparison to criteria.
	Section 4	**Verification of Remediation or Closure** • **No Action Needed**: If the data evaluation completed under Sections 2 and 3 demonstrated that there are no exceedances of cleanup criteria, no remedial action is required for this pathway and consequently no additional sampling or data analysis is required under Section 4. • **In situ/Ex situ Treatment**: Plans for verifying remediation or closure must be made on a case-by-case basis and must demonstrate how treatment is effective in meeting cleanup criteria. To demonstrate verification of remediation or closure, concentrations in groundwater must meet criteria during a specified number of consecutive sampling events. The number of consecutive sampling events must be selected on a case-by-case basis to reflect seasonal variation in groundwater quality, flow rates, and initial distribution of contamination in groundwater. Ordinarily, this demonstration should be made using data from at least one year of quarterly sampling. • **Cover/Containment**: In the context of this pathway, cover/containment will typically be a vapor barrier. Generally groundwater sampling will not be used to verify remediation or closure that relies on vapor barrier. Verification of remediation in areas outside the cover/containment structure will be done using the options above. • Statistical analysis of groundwater concentrations over time in an individual well will not be allowed for comparison to criteria.

Additional Information	Criteria Application Guidesheet 4
Recommended Statistical Methods for Comparison to Criteria:	If statistical analysis is documented to be practical, the recommended statistical method is comparison of *FACILITY* data to criteria using a 95% upper confidence limit (UCL) for the mean concentration (see Chapter 3 of the tabbed section titled, "Statistical Methods"). Other statistical methods may be acceptable on a case-by-case basis.

*Professional judgment is required to interpret the significance of exceedances that are very close to criteria, or that may be associated with insignificant quantities of hazardous substances.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR **GROUNDWATER** DATA	Statistical Guidesheet 5

GENERIC COMMERCIAL II, III, IV AND INDUSTRIAL GROUNDWATER VOLATILIZATION TO INDOOR AIR INHALATION CRITERIA (GVIIC)

May assist in evaluation of condition(s):	4
Applicability of Statistics	**GNP**

Statistics are generally not practical as a tool for assessing this exposure pathway because of the spatial and temporal variability of groundwater hazardous substance plumes and the resulting soil gas plumes that are generated from hazardous substances in groundwater.

Key Considerations:

Section 1

Facility Determination

- Not applicable.

Section 2

Selection of Data For Statistical Analysis

- Samples collected for the purpose of characterizing a *FACILITY* are typically biased, based on factors such as historical information, previous sampling, disposal practices, visual impacts, and aerial photos. Once the nature and extent of contamination has been defined, it is necessary to identify and/or obtain data that will allow for appropriate comparison to criteria. There are two primary considerations in determining if data sets are adequate. First, data sets must be obtained from locations that are consistent with the exposure assumptions for the relevant land use scenario. Second, if statistics are used, data sets must contain a sufficient number of *RANDOMLY* located sample results to allow for proper statistical analysis and development of *REPRESENTATIVE CONCENTRATIONS*. Therefore, additional sampling to develop *REPRESENTATIVE CONCENTRATIONS* will often be required after the nature and extent of contamination have been defined. Some characterization data may be used in the development of a *REPRESENTATIVE CONCENTRATION* as described in Section 2.4.2 of the tabbed section titled, "Sampling Strategies."

- A statistical analysis is appropriate only if the nature and extent of contamination, including any *HOT SPOTS*, has been adequately defined.

- It is not appropriate to combine samples from *HOT SPOT* areas with samples from other areas of a property for statistical analysis. This is necessary to avoid averaging or diluting the samples that represent *HOT SPOTS*. *HOT SPOTS* must be addressed separately. See Section 2.2.1.1 of the tabbed section titled, "Sampling Strategies."

★ Sample locations with concentrations equal to or greater than the hazardous substance's water solubility must not be included in the data set for statistical evaluation. Areas where data exceed solubility must be addressed separately and cannot be ignored.

★ Generally only groundwater data from a 4,000 ft^2 *EXPOSURE UNIT* (i.e., the building footprint) may be used in a statistical calculation for remedial compliance.

	Section 3	**Risk Analysis – Comparison to Criteria**
		★ Statistical analysis of *FACILITY* groundwater concentrations for comparison to GVIIC is generally not practical given the spatial and temporal variability of groundwater plumes and the dynamics of the resulting soil gas plume that is generated by hazardous substances in groundwater. Determining the impact of these factors within a 4,000 ft^2 area (i.e., the building footprint) adds further complications with application of statistical approaches. As a result, comparison to these criteria will generally be completed on a point-by-point basis* (i.e., each concentration at each location must be compared individually to criteria). • If DEQ approval of a response activity is being sought, a *PROPOSAL* for a statistical analysis must be submitted to the DEQ for approval to assure that data needs and/or complexities of the pathway are addressed. • Statistical analysis of groundwater concentrations over time in an individual well will not be allowed for comparison to criteria.
	Section 4	**Verification of Remediation or Closure**
		• **No Action Needed**: If the data evaluation completed under Sections 2 and 3 demonstrated that there are no exceedances of cleanup criteria, no remedial action is required for this pathway and consequently no additional sampling or data analysis is required under Section 4. • **In situ/Ex situ Treatment**: Plans for verifying remediation or closure must be made on a case-by-case basis and must demonstrate how treatment is effective in meeting cleanup criteria. To demonstrate verification of remediation or closure, concentrations in groundwater must meet criteria during a specified number of consecutive sampling events. The number of consecutive sampling events must be selected on a case-by-case basis to reflect seasonal variation in groundwater quality, flow rates, and initial distribution of contamination in groundwater. Ordinarily, this demonstration should be made using data from at least one year of quarterly sampling. • **Cover/Containment**: In the context of this pathway, cover/containment will typically be a vapor barrier. Generally groundwater sampling will not be used to verify remediation or closure that relies on a vapor barrier. Verification of remediation in areas outside the cover/containment structure will be done using the options above. • Statistical analysis of groundwater concentrations over time in an individual well will not be allowed for comparison to criteria.

Additional Information:	Criteria Application Guidesheet 5
Recommended Statistical Methods for Comparison to Criteria:	If statistical analysis is documented to be practical, the recommended statistical method is comparison of *FACILITY* data to criteria using a 95% upper confidence limit (UCL) for the mean concentration (see Chapter 3 of the tabbed section titled, "Statistical Methods"). Other statistical methods may be acceptable on a case-by-case basis.

*Professional judgment is required to interpret the significance of exceedances that are very close to criteria, or that may be associated with insignificant quantities of hazardous substances.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR **GROUNDWATER** DATA	Statistical Guidesheet 6
GENERIC GROUNDWATER CONTACT CRITERIA (GCC)	
May assist in evaluation of condition(s):	3
Applicability of Statistics	**GNP**

Statistics are generally not practical as a tool for assessing this exposure pathway because it is unlikely that a sufficient number of samples will be collected to allow for valid analysis in the very small areas where groundwater contact may occur.

Key Considerations:

Section 1

Facility Determination

- For the purposes of a BEA or initial *FACILITY* determination, property may be classified as a *FACILITY* if one or more samples contain concentrations above generic residential criteria or *BACKGROUND* concentrations, whichever is greater, provided that there is not a greater body of evidence that the property is not a *FACILITY*. For purposes other than *FACILITY* determination (e.g., remediation and/or closure), additional data will likely be required.
- If a statistical analysis is relied upon for making a *FACILITY* determination see Sections 2 and 3.
- ★ In certain cases, professional judgment may dictate that this pathway is not relevant if the depth to groundwater is greater than the depth to utilities and the depth at which other subsurface work may be performed. These factors and additional considerations are described in Criteria Application Guidesheet 6. If the pathway is not relevant, these criteria are not applicable and are not the basis for a property being considered a *FACILITY*.

Section 2

Selection of Data For Statistical Analysis

- Samples collected for the purpose of characterizing a *FACILITY* are typically biased, based on factors such as historical information, previous sampling, disposal practices, visual impacts, and aerial photos. Once the nature and extent of contamination has been defined, it is necessary to identify and/or obtain data that will allow for appropriate comparison to criteria. There are two primary considerations in determining if data sets are adequate. First, data sets must be obtained from locations that are consistent with the exposure assumptions for the relevant land use scenario. Second, if statistics are used, data sets must contain a sufficient number of *RANDOMLY* located sample results to allow for proper statistical analysis and development of *REPRESENTATIVE CONCENTRATIONS*. Therefore, additional sampling to develop *REPRESENTATIVE CONCENTRATIONS* will often be required after the nature and extent of contamination have been defined. Some characterization data may be used in the development of a *REPRESENTATIVE CONCENTRATION* as described in Section 2.4.2 of the tabbed section titled, "Sampling Strategies."
- A statistical analysis is appropriate only if the nature and extent of contamination, including any *HOT SPOTS*, has been adequately defined.
- It is not appropriate to combine samples from *HOT SPOT* areas with samples from other areas of a property for statistical analysis. This is necessary to avoid averaging or diluting the samples that represent *HOT SPOTS. HOT SPOTS* must be addressed separately. See Section 2.2.1.1 of the tabbed section titled, "Sampling Strategies."

	Section 3	**Risk Analysis – Comparison to Criteria**
		★ In certain cases, professional judgment may dictate that these criteria are not applicable due to factors such as depth to groundwater relative to the depth where subsurface activities may occur, and other factors as described in Criteria Application Guidesheet 6. ★ Statistical analysis of *FACILITY* groundwater concentrations for comparison to criteria is generally not practical, since exposure to groundwater may be limited to very small areas such as an area the size of a manhole. Therefore, comparisons to GCC will generally be completed on a point-by-point basis* (i.e., each concentration at each time and location must be compared individually to criteria). • If DEQ approval of a response activity is being sought, a *PROPOSAL* for a statistical analysis must be submitted to the DEQ for approval to assure that data needs and/or complexities of the pathway are addressed. • If *BACKGROUND* concentrations are greater than criteria for naturally occurring substances, groundwater concentrations may be compared to *BACKGROUND* concentrations. See Statistical Guidesheet 10 and Chapter 4 of the tabbed section titled, "Statistical Methods." This evaluation will generally be made on a point-by-point basis* because of the lack of data. • Statistical analysis of groundwater concentrations over time in an individual well will not be allowed for comparison to criteria or *BACKGROUND*.
	Section 4	**Verification of Remediation or Closure**
		• **No Action Needed**: If the data evaluation completed under Sections 2 and 3 demonstrated that there are no exceedances of cleanup criteria, no remedial action is required for this pathway and consequently no additional sampling or data analysis is required under Section 4. • **In situ/Ex situ Treatment**: For the purposes of verification of remediation or closure, compliance of groundwater data with these criteria may be determined in the same manner described in Section 3. Plans for verifying remediation or closure must be made on a case-by-case basis and must demonstrate how treatment is effective in meeting cleanup criteria. These plans may or may not depend on statistical analysis of data as allowed in Sections 2 and 3. To demonstrate verification of remediation or closure, concentrations in groundwater must meet criteria during a specified number of consecutive sampling events. The number of consecutive sampling events must be selected on a case-by-case basis to reflect seasonal variation in groundwater quality, flow rates, and initial distribution of contamination in groundwater. Ordinarily, this demonstration should be made using data from at least one year of quarterly sampling. • **Containment**: Generally groundwater sampling used to verify a groundwater containment remedy will be done at the perimeter of the containment system. Statistical analysis is generally not practical for reasons described in Sections 2 and 3. This verification will be an ongoing part of the remedy as long as hazardous substances are present above criteria. • If *BACKGROUND* concentrations are greater than criteria for naturally occurring substances, groundwater concentrations may be compared to *BACKGROUND* concentrations. See Statistical Guidesheet 10 and Chapter 4 of the tabbed section titled, "Statistical Methods." This evaluation will generally be made on a point-by-point basis* because of the lack of data.

Additional Information:	Criteria Application Guidesheet 6
Recommended Statistical Methods for Comparison to Criteria:	If statistical analysis is documented to be practical, the recommended statistical method is comparison of *FACILITY* data to criteria using a 95% upper confidence limit (UCL) for the mean concentration (see Chapter 3 of the tabbed section titled, "Statistical Methods"). Other statistical methods may be acceptable on a case-by-case basis.

*Professional judgment is required to interpret the significance of exceedances that are very close to criteria, or that may be associated with insignificant quantities of hazardous substances.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR **GROUNDWATER** DATA	Statistical Guidesheet 7
WATER SOLUBILITY	
May assist in evaluation of condition(s):	1-5
Applicability of Statistics	**Pathway Dependent**
See the Statistical Guidesheet for the exposure pathway where the cleanup criterion defaults to water solubility to determine the applicability of statistics.	

Key Considerations:	**Section 1**	**Facility Determination**
		• See the Statistical Guidesheet for the generic criterion that defaults to the water solubility value.
	Section 2	**Selection of Data For Statistical Analysis**
		• See the Statistical Guidesheet for the generic criterion that defaults to the water solubility value. • Data that are greater than water solubility should not be included in a statistical analysis.
	Section 3	**Risk Analysis – Comparison to Criteria**
		• See Section 2.
	Section 4	**Verification of Remediation or Closure**
		• See Section 2.

Additional Information:	Criteria Application Guidesheets 7 and A, and Appendix A.
Recommended Statistical Methods for Comparison to Criteria:	See Statistical Guidesheets for relevant exposure pathways.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR **GROUNDWATER** DATA	Statistical Guidesheet 8

GENERIC SCREENING LEVELS FOR FLAMMABILITY AND EXPLOSIVITY

May assist in evaluation of condition(s):	3, 4, 14
Applicability of Statistics	**GNP**

Statistics are generally not practical as a tool for assessing this condition because it deals with acute risks and any exceedance of screening levels requires further consideration.

Key Considerations:

Section 1

Facility Determination

- The determination that property is a *FACILITY* is generally based on data which show that one or more generic residential cleanup criterion is exceeded. Since the values developed for this condition are screening levels and not cleanup criteria, *FACILITY* determinations generally will not be based on this condition. If you are evaluating property where you suspect problems with this condition and no data are available to demonstrate that the property is a *FACILITY* for other reasons, consult your supervisor for guidance.

Section 2

Selection of Data For Statistical Analysis

- Not applicable - see Section 3.

Section 3

Risk Analysis – Comparison to Criteria

★ Since this condition deals with acute risks, any exceedance of screening levels requires further consideration. Statistical treatment of data is not expected to be a practical tool for making decisions about the need for response activity related to this condition. Professional judgment will be required to determine the significance of any data that relates to this condition. The quantity of the hazardous substance that is present above its screening level(s) is a factor that determines whether there may be an unacceptable risk.

Section 4

Verification of Remediation or Closure

- See Section 3.

Additional Information:	Criteria Application Guidesheet 8
Recommended Statistical Methods for Comparison to Criteria:	The statistical method for comparing *FACILITY* data to criteria will be determined on a case-by-case basis.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR **GROUNDWATER** DATA	Statistical Guidesheet 9

GENERIC ACUTE INHALATION TOXICITY SCREENING LEVELS

May assist in evaluation of condition(s):	3, 4, 14

Applicability of Statistics	GNP

Statistics are generally not practical as a tool for assessing this condition because it deals with acute risks and any exceedance of screening levels requires further consideration.

Key Considerations:

Section 1

Facility Determination

- The determination that property is a *FACILITY* is generally based on data which show that one or more generic residential cleanup criterion is exceeded. Since the values developed for this condition are screening levels and not cleanup criteria, *FACILITY* determinations generally will not be based on this condition. If you are evaluating property where you suspect problems with this condition and no data are available to demonstrate that the property is a *FACILITY* for other reasons, consult your supervisor for guidance.

Section 2

Selection of Data For Statistical Analysis

- Not applicable - see Section 3.

Section 3

Risk Analysis – Comparison to Criteria

★ Since this condition deals with acute risks, any exceedance of screening levels requires further consideration. Statistical treatment of data is not expected to be a practical tool for making decisions about the need for response activity related to this condition. Professional judgment will be required to determine the significance of any data that relates to this condition. The quantity of the hazardous substance that is present above its screening level(s) is a factor that determines whether there may be an unacceptable risk.

Section 4

Verification of Remediation or Closure

- See Section 3.

Additional Information:	Criteria Application Guidesheet 9
Recommended Statistical Methods for Comparison to Criteria:	The statistical method for comparing *FACILITY* data to criteria will be determined on a case-by-case basis.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR SOIL DATA	**Statistical Guidesheet 10**
SOIL BACKGROUND	
May assist in evaluation of Condition(s):	6-12
Applicability of Statistics	**YES - Determining *BACKGROUND*; Pathway Dependent - Comparison to *FACILITY* Data**

See the Statistical Guidesheet for the soil exposure pathway to determine if criteria may default to *BACKGROUND* and if statistical analysis of *FACILITY* data is appropriate for comparison to *BACKGROUND* (or if a point-by-point comparison of *FACILITY* data to *BACKGROUND* is necessary).

Key Considerations:	**Section 1**	**Facility Determination**
		• For the purposes of a BEA or initial *FACILITY* determination, property may be classified as a *FACILITY* if one or more samples contain concentrations above generic residential criteria or *BACKGROUND* concentrations, whichever is greater, provided that there is not a greater body of evidence that the property is not a *FACILITY*. However, if concentrations of a hazardous substance from the property are less than the *STATEWIDE DEFAULT BACKGROUND* concentration in Operational Memorandum #15 for that hazardous substance, the property is not a *FACILITY* with respect to that hazardous substance. For purposes other than *FACILITY* determination (e.g., remediation and/or closure), additional data will likely be required. • If a statistical analysis is relied upon for making a *FACILITY* determination see Sections 2 and 3.
	Section 2	**Selection of Data For Statistical Analysis**
		★ For all other Statistical Guidesheets, this section addresses selection of *FACILITY* data for statistical analysis. However, for this Statistical Guidesheet, the following comments address selection of *BACKGROUND* data for establishing *BACKGROUND* concentrations. ★ Establishing soil *BACKGROUND* can be accomplished by utilizing the *STATEWIDE DEFAULT BACKGROUND* criteria provided in Operational Memorandum #15 or by developing *FACILITY-SPECIFIC BACKGROUND* criteria. In addition, *REGIONAL BACKGROUND* may be *PROPOSED* on a case-by-case basis, as described in Operational Memorandum #15. ★ If multiple soil horizons are present at a *FACILITY*, *FACILITY-SPECIFIC BACKGROUND* should be established for each distinct soil horizon being evaluated at the *FACILITY*. ★ *FACILITY-SPECIFIC BACKGROUND* samples should be taken in each distinct soil type or horizon and at comparable depths to *FACILITY* soil samples being compared to *FACILITY-SPECIFIC BACKGROUND*. A minimum of nine samples for each distinct soil horizon should be used to establish *FACILITY-SPECIFIC BACKGROUND* concentrations. This is to help account for the natural variability inherent within each distinct soil horizon. Fewer than nine samples may be *PROPOSED* on a case-by-case basis.

<table>
<tr><td rowspan="3"></td><td>Section 2: (Continued)</td><td>★ FACILITY-SPECIFIC BACKGROUND samples must reflect naturally occurring concentrations except as described in the following bullet. This is necessary to comply with the definition of BACKGROUND under Part 201:

“The concentration or level of a hazardous substance which exists in the environment at or regionally proximate to a site that is not attributable to any release at or regionally proximate to the site.”

★ BACKGROUND soil samples are typically used to establish naturally occurring levels of metals. NON-RELEASE ANTHROPOGENIC BACKGROUND may be PROPOSED on a case-by-case basis for other classes of hazardous substances if the presence of these hazardous substances is not present due to a RELEASE (e.g., compounds present in the soil resulting from application of pesticides in accordance with label directions).

★ FACILITY-SPECIFIC BACKGROUND samples should be collected using the same methodology, analytical methods and detection limits as FACILITY samples.

★ BACKGROUND soil samples must be analyzed using methods and detection limits found in Operational Memorandum #6, Rev 5.</td></tr>
<tr><td rowspan="2">Section 3</td><td>Risk Analysis – Comparison to Criteria</td></tr>
<tr><td>★ According to Section 20a(11), BACKGROUND concentrations become the Part 201 cleanup criteria when BACKGROUND concentrations are greater than corresponding risk-based criteria.

★ The recommended statistical method for comparing FACILITY data to BACKGROUND depends on the type of BACKGROUND being used and whether a statistical analysis was appropriate for comparing FACILITY data to Part 201 risk-based criteria. See Chapter 4 of the tabbed section titled, “Statistical Methods.”

• When FACILITY-SPECIFIC BACKGROUND is utilized, the analysis OF FACILITY-SPECIFIC BACKGROUND data must include an evaluation of the underlying statistical distribution of the data (i.e., normal, lognormal, or neither) and level of censoring (i.e., proportion of data below the detection limit). See the tabbed section titled, “Statistical Methods.” This evaluation must be completed for each hazardous substance.

• The presence of outliers in the FACILITY-SPECIFIC BACKGROUND data set should be evaluated using the procedures outlined in Chapter 2 of the tabbed section titled, “Statistical Methods.”</td></tr>
<tr><td rowspan="2">Section 4</td><td>Verification of Remediation or Closure</td></tr>
<tr><td>★ Refer to the statistical guidesheets corresponding to the pathway/condition. When appropriate, verification samples may be compared to BACKGROUND concentrations as described in Sections 2 and 3 of this guidesheet.</td></tr>
<tr><td colspan="2">Additional Information:</td><td>Criteria Application Guidesheet 10</td></tr>
<tr><td colspan="2">Recommended Statistical Methods for Comparison to Criteria:</td><td>See Chapter 4 of the tabbed section titled, “Statistical Methods.” Other statistical methods may be acceptable on a case-by-case basis.</td></tr>
</table>

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR SOIL DATA	**Statistical Guidesheet 11**
GENERIC SOIL CRITERIA PROTECTIVE OF RESIDENTIAL AND COMMERCIAL I DRINKING WATER	
May assist in evaluation of condition(s):	9
Applicability of Statistics	**GNP**
Statistics are generally not practical as a tool for assessing this exposure pathway because of the difficulty in projecting the impact on groundwater of hazardous substances in soil and the need to assure that drinking water cleanup criteria are met at all points in the aquifer.	

Key Considerations:		
	Section 1	**Facility Determination** • For the purposes of a BEA or initial *FACILITY* determination, property may be classified as a *FACILITY* if one or more samples contain concentrations above generic residential criteria or *BACKGROUND* concentrations, whichever is greater, provided that there is not a greater body of evidence that the property is not a *FACILITY* or leachate sample results are less than the groundwater criterion. However, if concentrations of a hazardous substance from the property are less than the *STATEWIDE DEFAULT BACKGROUND* value in Operational Memorandum #15 for that hazardous substance, the property is not a *FACILITY* with respect to that hazardous substance. For purposes other than *FACILITY* determination (e.g., remediation and/or closure), additional data will likely be required. • If a statistical analysis is relied upon for making a *FACILITY* determination see Sections 2 and 3.
	Section 2	**Selection of Data For Statistical Analysis** • Samples collected for the purpose of characterizing a *FACILITY* are typically biased, based on factors such as historical information, previous sampling, disposal practices, visual impacts, and aerial photos. Once the nature and extent of contamination has been defined, it is necessary to identify and/or obtain data that will allow for appropriate comparison to criteria. There are two primary considerations in determining if data sets are adequate. First, data sets must be obtained from locations that are consistent with the exposure assumptions for the relevant land use scenario. Second, if statistics are used, data sets must contain a sufficient number of *RANDOMLY* located sample results to allow for proper statistical analysis and development of *REPRESENTATIVE CONCENTRATIONS*. Therefore, additional sampling to develop *REPRESENTATIVE CONCENTRATIONS* will often be required after the nature and extent of contamination have been defined. Some characterization data may be used in the development of a *REPRESENTATIVE CONCENTRATION* as described in Section 2.4.2 of the tabbed section titled, “Sampling Strategies.” • A statistical analysis is appropriate only if the nature and extent of contamination, including any *HOT SPOTS*, has been adequately defined. • It is not appropriate to combine samples from *HOT SPOT* areas with samples from other areas of a property for statistical analysis. This is necessary to avoid averaging or diluting the samples that represent *HOT SPOTS*. *HOT SPOTS* must be addressed separately. See Section 2.2.1.1 of the tabbed section titled, “Sampling Strategies.”

	Section 3	**Risk Analysis – Comparison to Criteria**
		★ Statistical analysis of *FACILITY* soil concentrations for comparison to soil criteria protective of the drinking water criteria may be acceptable in limited circumstances. Sufficient data must be available to demonstrate that areas of contaminated soil above criteria are not large enough to result in groundwater concentrations in an aquifer above criteria. This may require a fairly rigorous data set that is not often practical to obtain. Similarly, statistical analysis of *FACILITY* leachate concentrations for comparison to drinking water criteria may also be acceptable in limited circumstances. • If *BACKGROUND* concentrations in soil are greater than criteria, soil data may be compared to *BACKGROUND* instead. Statistical methods for comparing to *BACKGROUND* will also vary depending on which type of *BACKGROUND* is being considered. See Statistical Guidesheet 10 and Chapter 4 of the tabbed section titled, "Statistical Methods." • Statistical analysis of *FACILITY* data is not allowed for hazardous substances present above Csat screening levels without further evaluation of risk and the extent of the area exceeding Csat. • Statistical analysis of *FACILITY* data is not allowed for hazardous substances that have criteria based on acute toxicological effects and/or physical hazards.
	Section 4	**Verification of Remediation or Closure** • **No Action Needed**: If the data evaluation completed under Sections 2 and 3 demonstrated that there are no exceedances of cleanup criteria, no remedial action is required for this pathway and consequently no additional sampling or data analysis is required under Section 4. • **Excavation**: Analytical results from verification sampling will generally be compared to Part 201 criteria on a point-by-point basis* in unsaturated soil. Numbers and locations of samples collected for verifying remediation of soil by excavation can be selected in accordance with either the tabbed section titled, "Sampling Strategies" (Sections 1.3 or 2.3 as appropriate based on size of excavation), or a DEQ-approved sampling plan. • **In situ/Ex situ Treatment**: Plans for verifying remediation or closure must be made on a case-by-case basis and must demonstrate how treatment is effective in meeting cleanup criteria. These plans may or may not depend on statistical analysis of data as allowed in Sections 2 and 3. • **Cover/Containment**: Generally soil sampling will not be used to verify remediation or closure that relies on a cover or containment. Instead, verification of remediation will involve verifying the integrity of the cover or containment structure on an ongoing basis. Groundwater sampling may be used in some cases to verify the effectiveness of the cover/containment. Verification of remediation in areas outside the cover/containment structure will be done using the options above. • If *BACKGROUND* concentrations in soil are greater than criteria, soil data may be compared to *BACKGROUND* instead. Statistical methods for comparing to *BACKGROUND* will also vary depending on which type of *BACKGROUND* is being considered. See Statistical Guidesheet 10 and Chapter 4 of the tabbed section titled, "Statistical Methods."

Additional Information:	Criteria Application Guidesheet 11; Statistical Guidesheet 1
Recommended Statistical Methods for Comparison to Criteria:	If statistical analysis is documented to be practical, the recommended statistical method is comparison of *FACILITY* data to criteria using a 95% upper confidence limit (UCL) for the mean concentration (see Chapter 3 of the tabbed section titled, "Statistical Methods"). Other statistical methods may be acceptable on a case-by-case basis.

*Professional judgment is required to interpret the significance of exceedances that are very close to criteria, or that may be associated with insignificant quantities of hazardous substances.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR SOIL DATA	**Statistical Guidesheet 12**
SOIL CRITERIA PROTECTIVE OF THE GROUNDWATER SURFACE WATER INTERFACE (GSI)	
May assist in evaluation of condition(s):	11
Applicability of Statistics	**GNP**

Statistics are generally not practical as a tool for assessing this exposure pathway because of the difficulty in projecting the impact on groundwater of hazardous substances in soil and the need for analysis of transport of hazardous substance from areas that are remote from the GSI.

Key Considerations:

Section 1

Facility Determination

- For the purposes of a BEA or initial *FACILITY* determination, property may be classified as a *FACILITY* if one or more samples contain concentrations above generic residential criteria or *BACKGROUND* concentrations, whichever is greater, provided that there is not a greater body of evidence that the property is not a *FACILITY* or leachate sample results are less than the groundwater criterion. However, if concentrations of a hazardous substance from the property are less than the *STATEWIDE DEFAULT BACKGROUND* value in Operational Memorandum #15 for that hazardous substance, the property is not a *FACILITY* with respect to that hazardous substance. For purposes other than *FACILITY* determination (e.g., remediation and/or closure), additional data will likely be required.

- If a statistical analysis is relied upon for making a *FACILITY* determination see Sections 2 and 3.

★ In certain cases, professional judgment may dictate that this pathway is not relevant due to proximity to the surface water and other factors described in Cleanup Criteria Application Guidesheet 3. In this case, these criteria are not applicable and are not the basis for a property being considered a *FACILITY*.

Section 2

Selection of Data For Statistical Analysis

★ Samples collected for the purpose of characterizing a *FACILITY* are typically biased, based on factors such as historical information, previous sampling, disposal practices, visual impacts, and aerial photos. Once the nature and extent of contamination has been defined, it is necessary to identify and/or obtain data that will allow for appropriate comparison to criteria. There are two primary considerations in determining if data sets are adequate. First, data sets must be obtained from locations that are representative of potential impacts for this migration pathway. Second, if statistics are used, data sets must contain a sufficient number of *RANDOMLY* located sample results from an appropriate area to allow for proper statistical analysis and development of *REPRESENTATIVE CONCENTRATIONS*. Therefore, additional sampling to develop *REPRESENTATIVE CONCENTRATIONS* will often be required after the nature and extent of contamination have been defined. Some characterization data may be used in the development of a *REPRESENTATIVE CONCENTRATION* as described in section 2.4.2 of the tabbed section titled, “Sampling Strategies.”

- A statistical analysis is appropriate only if the nature and extent of contamination, including any *HOT SPOTS*, has been adequately defined.

	Section 2: (Continued)	• It is not appropriate to combine samples from *HOT SPOT* areas with samples from other areas of a property for statistical analysis. This is necessary to avoid averaging or diluting the samples that represent *HOT SPOTS*. *HOT SPOTS* must be addressed separately. See section 2.2.1.1 of the tabbed section titled, "Sampling Strategies."
	Section 3	**Risk Analysis – Comparison to Criteria** ★ Statistical analysis of *FACILITY* soil or leachate concentrations for comparison to soil criteria protective of the GSI based on generic critieria or mixing zone-based criteria may be acceptable in limited circumstances, but is generally not practical. Meaningful use of statistics in these cases would require identifying a contaminated soil volume that is remote from the GSI but related to points at the GSI with a degree of precision that is generally not practical to achieve. This is because of the uncertainties associated with distribution of hazardous substances and transport to the GSI at most sites. ★ Only soil or leachate concentrations from areas that impact groundwater concentrations at the *AVERAGING AREA* may be used for statistical comparison to a soil criteria protective of the GSI which is based on chronic mixing zone-based criteria. • If DEQ approval of a response activity is being sought, a *PROPOSAL* for a statistical analysis must be submitted to the DEQ for approval to assure that data needs and/or complexities of the pathway are addressed. • If *BACKGROUND* concentrations in soil are greater than criteria, soil data may be compared to *BACKGROUND* instead. Statistical methods for comparing to *BACKGROUND* will also vary depending on which type of *BACKGROUND* is being considered. See Statistical Guidesheet 10 and Chapter 4 of the tabbed section titled, "Statistical Methods." • Statistical analysis of *FACILITY* data is not allowed for hazardous substances present above Csat screening levels without further evaluation of risk and the extent of the area exceeding Csat.
	Section 4	**Verification of Remediation or Closure** • **No Action Needed**: If the data evaluation completed under Sections 2 and 3 demonstrated that there are no exceedances of cleanup criteria, no remedial action is required for this pathway and consequently no additional sampling or data analysis is required under Section 4. • **Excavation**: Statistical analysis of *FACILITY* soil or leachate data is generally not practical. See Section 3. Analytical results from verification sampling will therefore generally be compared to Part 201 criteria on a point-by-point basis* in unsaturated soil. Numbers and locations of samples collected for point-by-point comparison can be selected in accordance with either the tabbed section titled, "Sampling Strategies" (Sections 1.3 or 2.3 as appropriate based on size of excavation), or a DEQ-approved sampling plan. If a statistical analysis is relied upon for verifying remediation or closure follow the guidance presented in Sections 2 and 3 in this Guidesheet. • **In situ/Ex situ Treatment**: Plans for verifying remediation or closure must be made on a case-by-case basis and must demonstrate how treatment is effective in meeting cleanup criteria. These plans may or may not depend on statistical analysis of data.

	Section 4: (Continued)	• **Cover/Containment**: Generally soil sampling will not be used to verify remediation or closure that relies on a cover or containment. Instead, verification of remediation will involve verifying the integrity of the cover or containment structure on an ongoing basis. Groundwater sampling may be used in some cases to verify the effectiveness of the cover/containment. Verification of remediation in areas outside the cover/containment structure will be done using the options above. ★ For all the preceding cases, it is necessary to demonstrate that groundwater at the GSI will continue to meet GSI criteria. This conclusion must be based on a demonstration that soil conditions and sources will not result in a future exceedance of GSI criteria at the GSI. • If *BACKGROUND* concentrations in soil are greater than criteria, soil data may be compared to *BACKGROUND* instead. Statistical methods for comparing to *BACKGROUND* will also vary depending on which type of *BACKGROUND* is being considered. See Statistical Guidesheet 10 and Chapter 4 of the tabbed section titled, "Statistical Methods."

Additional Information:	Criteria Application Guidesheet 12; Statistical Guidesheet 3
Recommended Statistical Methods for Comparison to Criteria:	If statistical analysis is documented to be practical, the recommended statistical method is comparison of *FACILITY* data to criteria using a 95% upper confidence limit (UCL) for the mean concentration (see Chapter 3 of the tabbed section titled, "Statistical Methods"). Other statistical methods may be acceptable on a case-by-case basis.

*Professional judgment is required to interpret the significance of exceedances that are very close to criteria, or that may be associated with insignificant quantities of hazardous substances.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR SOIL DATA	**Statistical Guidesheet 13**
GENERIC SOIL CRITERIA PROTECTIVE FOR GROUNDWATER CONTACT	
May assist in evaluation of condition(s):	10
Applicability of Statistics	**GNP**

Statistics are generally not practical as a tool for assessing this exposure pathway because of the difficulty in projecting the impact on groundwater of hazardous substances in soil and the need to assure compliance with GCC in potentially small areas.

Key Considerations:

Section 1

Facility Determination

- For the purposes of a BEA or initial *FACILITY* determination, property may be classified as a *FACILITY* if one or more samples contain concentrations above generic residential criteria or *BACKGROUND* concentrations, whichever is greater, provided that there is not a greater body of evidence that the property is not a *FACILITY* or leachate sample results are less than the groundwater criterion. However, if concentrations of a hazardous substance from the property are less than the *STATEWIDE DEFAULT BACKGROUND* value in Operational Memorandum #15 for that hazardous substance, the property is not a *FACILITY* with respect to that hazardous substance. For purposes other than *FACILITY* determination (e.g., remediation and/or closure), additional data will likely be required.
- If a statistical analysis is relied upon for making a *FACILITY* determination see Sections 2 and 3.

★ In certain cases, professional judgment may dictate that this pathway is not relevant if the depth to groundwater is greater than the depth to utilities and the depth at which other subsurface work may be performed. These factors and additional considerations are described in Cleanup Criteria Application Guidesheet 6. If the pathway is not relevant, these criteria are not applicable and are not the basis for a property being considered a *FACILITY*.

Section 2

Selection of Data For Statistical Analysis

- Samples collected for the purpose of characterizing a *FACILITY* are typically biased, based on factors such as historical information, previous sampling, disposal practices, visual impacts, and aerial photos. Once the nature and extent of contamination has been defined, it is necessary to identify and/or obtain data that will allow for appropriate comparison to criteria. There are two primary considerations in determining if data sets are adequate. First, data sets must be obtained from locations that are consistent with the exposure assumptions for the relevant land use scenario. Second, if statistics are used, data sets must contain a sufficient number of *RANDOMLY* located sample results to allow for proper statistical analysis and development of *REPRESENTATIVE CONCENTRATIONS*. Therefore, additional sampling to develop *REPRESENTATIVE CONCENTRATIONS* will often be required after the nature and extent of contamination have been defined. Some characterization data may be used in the development of a *REPRESENTATIVE CONCENTRATION* as described in section 2.4.2 of the tabbed section titled, "Sampling Strategies."
- A statistical analysis is appropriate only if the nature and extent of contamination, including any *HOT SPOTS*, has been adequately defined.

	Section 2: (Continued)	• It is not appropriate to combine samples from *HOT SPOT* areas with samples from other areas of a property for statistical analysis. This is necessary to avoid averaging or diluting the samples that represent *HOT SPOTS*. *HOT SPOTS* must be addressed separately. See section 2.2.1.1 of the tabbed section titled, "Sampling Strategies."
	Section 3	**Risk Analysis – Comparison to Criteria** ★ Since sufficient data are generally not available to allow for a statistical evaluation within the limited area that would be necessary to assure compliance with GCC at all potential exposure points in groundwater, comparison of soil data to these criteria must generally be completed on a point-by-point basis* (i.e., each concentration at each location must be compared individually to criteria). Therefore, statistical analysis of *FACILITY* soil concentrations for comparison to soil criteria protective for groundwater contact is generally not practical. Statistical analysis of *FACILITY* leachate concentrations for comparison to criteria may be acceptable in limited circumstances, but is also generally not practical. ★ In certain cases, professional judgment may dictate that this pathway is not relevant if the depth to groundwater is greater than the depth to utilities and the depth at which other subsurface work may be performed. These factors and additional considerations are described in Cleanup Criteria Application Guidesheet 6. If the pathway is not relevant, these criteria are not applicable and are not the basis for a property being considered a *FACILITY*. • If DEQ approval of a response activity is being sought, a *PROPOSAL* for a statistical analysis must be submitted to the DEQ for approval to assure that data needs and/or complexities of the pathway are addressed. • If *BACKGROUND* concentrations in soil are greater than criteria, soil data may be compared to *BACKGROUND* instead. Statistical methods for determining *BACKGROUND* concentrations will vary depending on which type of *BACKGROUND* is being considered. See Statistical Guidesheet 10 and Chapter 4 of the tabbed section titled, "Statistical Methods." • Statistical analysis of *FACILITY* data is not allowed for hazardous substances present above Csat screening levels without further evaluation of risk and the extent of the area exceeding Csat.
	Section 4	**Verification of Remediation or Closure** • **No Action Needed**: If the data evaluation completed under Sections 2 and 3 demonstrated that there are no exceedances of cleanup criteria, no remedial action is required for this pathway and consequently no additional sampling or data analysis is required under Section 4. • **Excavation**: Statistical analysis of *FACILITY* soil or leachate data is generally not practical. See Section 3. Analytical results from verification sampling will therefore generally be compared to Part 201 criteria on a point-by-point basis* in unsaturated soil. Numbers and locations of samples collected for point-by-point comparison can be selected in accordance with either the tabbed section titled, "Sampling Strategies" (Sections 1.3 or 2.3 as appropriate based on size of excavation), or a DEQ-approved sampling plan. If a statistical analysis is relied upon for verifying remediation or closure follow the guidance presented in Sections 2 and 3 of this Guidesheet.

	Section 4: (Continued)	• **In situ/Ex situ Treatment**: Plans for verifying remediation or closure must be made on a case-by-case basis and must demonstrate how treatment is effective in meeting cleanup criteria. These plans may or may not depend on statistical analysis of data. • **Cover/Containment**: Generally soil sampling will not be used to verify remediation or closure that relies on a cover or containment. Instead, verification of remediation will involve verifying the integrity of the cover or containment structure on an ongoing basis. Groundwater sampling may be used in some cases to verify the effectiveness of the cover/containment in which case groundwater data should be evaluated as described in Section 4 of Statistical Guidesheet 6. Verification of remediation in areas outside the cover/containment structure will be done using the options above. • If *BACKGROUND* concentrations in soil are greater than criteria, soil data may be compared to *BACKGROUND* instead. Statistical methods for determining *BACKGROUND* concentrations will vary depending on which type of *BACKGROUND* is being considered. See Statistical Guidesheet 10 and Chapter 4 of the tabbed section titled, "Statistical Methods."
Additional Information:		Criteria Application Guidesheet 13; Statistical Guidesheet 6
Recommended Statistical Methods for Comparison to Criteria:		If statistical analysis is documented to be practical, the recommended statistical method is comparison of *FACILITY* data to criteria using a 95% upper confidence limit (UCL) for the mean concentration (see Chapter 3 of the tabbed section titled, "Statistical Methods"). Other statistical methods may be acceptable on a case-by-case basis.

*Professional judgment is required to interpret the significance of exceedances that are very close to criteria, or that may be associated with insignificant quantities of hazardous substances.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR SOIL DATA	**Statistical Guidesheet 14**
GENERIC RESIDENTIAL AND COMMERCIAL I SOIL VOLATILIZATION TO INDOOR AIR INHALATION CRITERIA (SVIIC)	
May assist in evaluation of condition(s):	8
Applicability of Statistics	**GNP**
Statistics are generally not practical as a tool for assessing this exposure pathway because it is unlikely that there will be a sufficient number of samples available from the generic building footprint size.	

Key Considerations:

Section 1

Facility Determination

- For the purposes of a BEA or initial *FACILITY* determination, property may be classified as a *FACILITY* if one or more samples contain concentrations above generic residential criteria, provided that there is not a greater body of evidence that the property is not a *FACILITY*. For purposes other than *FACILITY* determination (e.g., remediation and/or closure), additional data will likely be required.
- If a statistical analysis is relied upon for making a *FACILITY* determination see Sections 2 and 3.

Section 2

Selection of Data For Statistical Analysis

★ Generally, only data from a 1,200 ft^2 *EXPOSURE UNIT* may be used in any particular statistical calculation.

- Samples collected for the purpose of characterizing a *FACILITY* are typically biased, based on factors such as historical information, previous sampling, disposal practices, visual impacts, and aerial photos. Once the nature and extent of contamination has been defined, it is necessary to identify and/or obtain data that will allow for appropriate comparison to criteria. There are two primary considerations in determining if data sets are adequate. First, data sets must be obtained from locations that are consistent with the exposure assumptions for the relevant land use scenario. Second, if statistics are used, data sets must contain a sufficient number of *RANDOMLY* located sample results to allow for proper statistical analysis and development of *REPRESENTATIVE CONCENTRATIONS*. Therefore, additional sampling to develop *REPRESENTATIVE CONCENTRATIONS* will often be required after the nature and extent of contamination have been defined. Some characterization data may be used in the development of a *REPRESENTATIVE CONCENTRATION* as described in Section 2.4.2 of the tabbed section titled, "Sampling Strategies."
- A statistical analysis is appropriate only if the nature and extent of contamination, including any *HOT SPOTS*, has been adequately defined.
- It is not appropriate to combine data from *HOT SPOT* areas with data from other areas of a property for statistical analysis. This is necessary to avoid averaging out or diluting the samples that represent *HOT SPOTS*. *HOT SPOTS* must be addressed separately. See Section 2.2.1.1 of the tabbed section titled, "Sampling Strategies."

★ Generally only soil data from a 1,200 ft^2 *EXPOSURE UNIT* (i.e., the building footprint) may be used in a statistical calculation for remedial compliance.

	Section 3	**Risk Analysis – Comparison to Criteria**
		★ Statistical analysis of *FACILITY* soil concentrations for comparison to residential SVIIC is generally not practical since sufficient data (i.e., a minimum of nine *RANDOMLY* located samples) are generally not available to allow for a statistical evaluation within the generic building footprint area of 1,200 ft^2. Therefore, comparisons to SVIIC must generally be completed on a point-by-point basis* (i.e., each concentration at each location must be compared individually to criteria). ★ Any statistical analysis must consider the potential for spatial variability of soil types to influence vapor migration. • If DEQ approval of a response activity is being sought, a *PROPOSAL* for a statistical analysis must be submitted to the DEQ for approval to assure that data needs and/or complexities of the pathway are addressed. • Statistical analysis of *FACILITY* data is not allowed for hazardous substances present above Csat screening levels without further evaluation of risk and the extent of the area exceeding Csat.
	Section 4	**Verification of Remediation or Closure**
		• **No Action Needed**: If the data evaluation completed under Sections 2 and 3 demonstrated that there are no exceedances of cleanup criteria, no remedial action is required for this pathway and consequently no additional sampling or data analysis is required under Section 4. • **Excavation**: Statistical analysis of *FACILITY* soil data is generally not practical. See Section 3. Analytical results from verification sampling will therefore generally be compared to Part 201 criteria on a point-by-point basis* in unsaturated soil. Numbers and locations of samples collected for point-by-point comparison can be selected in accordance with either the tabbed section titled, "Sampling Strategies" (Sections 1.3 or 2.3 as appropriate based on size of excavation), or a DEQ-approved sampling plan. If a statistical analysis is relied upon for verifying remediation or closure follow the guidance presented in Sections 2 and 3 of this Guidesheet. • **In situ/Ex situ Treatment**: Plans for verifying remediation or closure must be made on a case-by-case basis and must demonstrate how treatment is effective in meeting cleanup criteria. These plans may or may not depend on statistical analysis of data. • **Cover/Containment**: In the context of this pathway, cover/containment will typically be a vapor barrier. Generally soil sampling will not be used to verify remediation or closure that relies on a cover or containment. Verification of remediation in areas outside the cover/containment structure will be done using the other options in this section. • If *BACKGROUND* concentrations in soil are greater than criteria, soil data may be compared to *BACKGROUND* instead. Statistical methods for determining *BACKGROUND* concentrations will vary depending on which type of *BACKGROUND* is being considered. See Chapter 4 of the tabbed section titled, "Statistical Methods."

Additional Information:	Criteria Application Guidesheet 14
Recommended Statistical Methods for Comparison to Criteria:	If statistical analysis is documented to be practical, the recommended statistical method is comparison of *FACILITY* data to criteria using a 95% upper confidence limit (UCL) for the mean concentration (see Chapter 3 of the tabbed section titled, "Statistical Methods"). Other statistical methods may be acceptable on a case-by-case basis.

*Professional judgment is required to interpret the significance of exceedances that are very close to criteria, or that may be associated with insignificant quantities of hazardous substances.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR SOIL DATA	**Statistical Guidesheets 15, 16 and 17**

GENERIC RESIDENTIAL AND COMMERCIAL I INFINITE AND FINITE VOLATILE SOIL INHALATION CRITERIA (VSIC) FOR AMBIENT AIR

May assist in evaluation of condition(s):	7
Applicability of Statistics	**YES**

Statistics are applicable for evaluating this exposure pathway. Use of statistics is practical when there are adequate data sets available in the *EXPOSURE UNITS* or *EMISSION SOURCE AREAS*, as appropriate to the *FACILITY*.

Key Considerations:

Section 1

Facility Determination

- For the purposes of a BEA, or initial *FACILITY* determination, property may be classified as a *FACILITY* if one or more samples contain concentrations above generic residential, provided that there is not a greater body of evidence that the property is not a *FACILITY*. For purposes other than *FACILITY* determination (e.g., remediation and/or closure), additional data will likely be required.
- If a statistical analysis is relied upon for making a *FACILITY* determination see Sections 2 and 3.

Section 2

Selection of Data For Statistical Analysis

- Samples collected for the purpose of characterizing a *FACILITY* are typically biased, based on factors such as historical information, previous sampling, disposal practices, visual impacts, and aerial photos. Once the nature and extent of contamination has been defined, it is necessary to identify and/or obtain data that will allow for appropriate comparison to criteria. There are two primary considerations in determining if data sets are adequate. First, data sets must be obtained from locations that are consistent with the exposure assumptions for the relevant land use scenario. Second, if statistics are used, data sets must contain a sufficient number of *RANDOMLY* located sample results to allow for proper statistical analysis and development of *REPRESENTATIVE CONCENTRATIONS*. Therefore, additional sampling to develop *REPRESENTATIVE CONCENTRATIONS* will often be required after the nature and extent of contamination have been defined. Some characterization data may be used in the development of a *REPRESENTATIVE CONCENTRATION* as described in Section 2.4.2 of the tabbed section titled, "Sampling Strategies."
- A statistical analysis of soil data should be completed only if the nature and extent of contamination, including any *HOT SPOTS*, has been adequately defined.
- It is not appropriate to combine samples from *HOT SPOT* areas with samples from other areas of a property for statistical analysis. This is necessary to avoid averaging or diluting the samples that represent *HOT SPOTS*. *HOT SPOTS* must be addressed separately. See Section 2.2.1.1 of the tabbed section titled, "Sampling Strategies."
- ★ The horizontal and vertical extent of the *EMISSION SOURCE AREA* must be estimated to allow for selection of the *SOURCE SIZE MODIFIER* and infinite or finite source generic VSIC. Only data from the estimated *EMISSION SOURCE AREA* may be used in statistical analysis for comparison to criteria. See the "Part 201 Generic Soil Inhalation Criteria for Ambient Air: Technical Support Document" for information about source size characterization.

	Section 2: (Continued)	★ Characterization of the horizontal extent of contamination is necessary for both finite and infinite generic VSIC to estimate the *EMISSION SOURCE AREA* contributing to volatile emissions from soil. ★ Infinite generic VSIC are applicable to the entire contaminated vertical soil column since both surface and subsurface concentrations of hazardous substances in soil may contribute to volatile emissions. ★ Finite generic VSIC are applicable only when contamination is demonstrated to be limited to a two or five-foot vertical interval. Consequently, only data from the contaminated interval can be included in a statistical analysis. If contamination is not limited to a two or five-foot interval, infinite source VSIC are applicable. ★ A *SOURCE SIZE MODIFIER* must be selected for *EMISSION SOURCE SIZES* that are different than the assumed 1/2 acre *EMISSION SOURCE SIZE*. The *SOURCE SIZE MODIFIER* must correspond to an *EMISSION SOURCE SIZE* that is at least as large as the *EMISSION SOURCE SIZE* of the *FACILITY*. For example, if the *EMISSION SOURCE SIZE* at a *FACILITY* is at least eight acres, the generic soil inhalation criteria for a 1/2 acre *EMISSION SOURCE SIZE* is multiplied by the *SOURCE SIZE MODIFIER* for 10 acres to provide generic criteria for a 10 acre *FACILITY*. *SOURCE SIZE MODIFIERS* to adjust generic criteria for *EMISSION SOURCE SIZES* other than a 1/2 acre are provided in a table at the end of this guidesheet.
	Section 3	**Risk Analysis – Comparison to Criteria** ★ Criteria for this pathway depend on the *EMISSION SOURCE SIZE*, which must be estimated in both the horizontal and vertical dimension. Criteria shown in the cleanup criteria tables are based on a 1/2 acre *EMISSION SOURCE SIZE*. *SOURCE SIZE MODIFIERS* to adjust generic criteria for *EMISSION SOURCE SIZES* other than a 1/2 acre are provided in a table at the end of this guidesheet. For convenience, this table has been updated to include 1/4 acre and two acre *EMISSION SOURCE SIZES* for use at residential/commercial I and commercial II, III, IV and industrial land uses, respectively. Soil contamination for the VSIC pathway may occur in two general patterns that will affect the *EMISSION SOURCE SIZE* used to adjust the criteria for comparison to the *FACILITY* data: ★ **Properties or *FACILITIES* with a single *EMISSION SOURCE AREA*** If hazardous substances are detected in only a limited area within a 1/4 acre *EXPOSURE UNIT* and are not detected in other areas of the *EXPOSURE UNIT*, the *EMISSION SOURCE AREA* is equal to only the horizontal extent of the area with detectable concentrations. Only data within the *EMISSION SOURCE AREA* may be included in a statistical analysis. For an *EMISSION SOURCE AREA* that is larger than a 1/4 acre (i.e., *EXPOSURE UNIT*), soil concentrations within each 1/4 acre *EXPOSURE UNIT* must meet the VSIC adjusted for the *EMISSION SOURCE SIZE*.

Section 3: (Continued)

★ **Properties with multiple *EMISSION SOURCE AREAS***

Properties larger than a 1/4 acre may contain several small *EMISSION SOURCE AREAS*. Where hazardous substances are not detected in the areas of the property between *EMISSION SOURCE AREAS*, the final *EMISSION SOURCE SIZE* is the sum of the horizontal extent of all individual *EMISSION SOURCE AREAS*. Soil concentrations within each *EMISSION SOURCE AREA* are compared to the generic criteria adjusted for the summed area of the individual *EMISSION SOURCE AREA* sizes.

★ Statistical analysis is possible for either the *EXPOSURE UNIT* or the *EMISSION SOURCE AREA* provided that sufficient data are available as described in Section 2.

★ Nine *RANDOMLY* located samples per *EXPOSURE UNIT* or *EMISSION SOURCE AREA*, whichever is smaller, should generally be used to conduct the statistical analysis. The actual number of samples to be collected may vary based on size of the *EMISSION SOURCE AREA*. Data collected for other purposes (as for identification of nature and extent) may be used where *RANDOM* sample locations fall on or reasonably close to existing sample locations and where data will be consistent in terms of sampling and analytical methods.

- Statistical analyses must include an evaluation of the underlying statistical distribution of the data (i.e., normal, lognormal, or neither) and the level of censoring (i.e., proportion of the data below the detection limit). See the tabbed section titled, Statistical Methods." This evaluation must be completed for each hazardous substance. Additionally, the statistical methods used to compare *FACILITY* data to criteria must be selected based on statistical distribution and level of censoring.

★ Compliance with infinite generic VSIC must be demonstrated for the entire vertical soil column since both surface and subsurface soil concentrations of hazardous substances may contribute to volatile emissions.

- Since a hazardous substance will <u>not</u> volatilize more when present at concentrations greater than Csat, sample data that exceeds Csat can be included in a statistical analysis to determine compliance with the VSIC as long as the Csat exceedance is not a *HOT SPOT*.

- Statistical analysis of *FACILITY* data is not allowed for hazardous substances that have criteria based on acute toxicological effects and/or physical hazards.

Section 4

Verification of Remediation or Closure

- **No Action Needed**: If the data evaluation completed under Sections 2 and 3 demonstrated that there are no exceedances of cleanup criteria, no remedial action is required for this pathway and consequently no additional sampling or data analysis is required under Section 4.

	Section 4: (Continued)	• **Excavation**: Numbers and locations of samples collected for verifying remediation of soil can be selected in accordance with either the tabbed section titled, "Sampling Strategies" (Sections 1.3 or 2.3 as appropriate based on size of excavation), or a DEQ-approved sampling plan when a person is seeking DEQ approval of the response activity. Analytical results from verification sampling must be compared to Part 201 criteria on a point-by-point basis* unless a minimum of nine *RANDOMLY* located samples are available within the *EXPOSURE UNIT(S)* or *EMISSION SOURCE AREA*, whichever is smaller. If a statistical analysis is used, analytical results from verification sampling must be compared to Part 201 criteria as described in Sections 2 and 3. • **In situ/Ex situ Treatment**: Plans for verifying remediation or closure must be made on a case-by-case basis and must demonstrate how treatment is effective in meeting cleanup criteria. These plans may or may not depend on statistical analysis of data as allowed in Sections 2 and 3. • **Cover/Containment**: Generally soil sampling will not be used to verify Limited or Site-specific remediation or closure that relies on a cover or containment. Instead, verification of remediation will involve verifying the integrity of the cover or containment structure on an ongoing basis. Verification of remediation in areas outside the cover/containment structure will be done using the options above.
Additional Information:		Criteria Application Guidesheet 15; "Part 201 Generic Soil Inhalation Criteria for Ambient Air: Technical Support Document"
Recommended Statistical Methods for Comparison to Criteria:		Compare *FACILITY* data to criteria using a 95% upper confidence limit (UCL) for the mean concentration (see Chapter 3 of the tabbed section titled, "Statistical Methods"). Other statistical methods may be acceptable on a case-by-case basis.

*Professional judgment is required to interpret the significance of exceedances that are very close to criteria, or that may be associated with insignificant quantities of hazardous substances.

***SOURCE SIZE MODIFIERS* for Ambient Air Soil Inhalation Criteria**

Modifiers		
Source Size (ft^2 or acres)	Q/C (g/m^2-s per kg/m^3)	Modifier
400 ft^2	261.26	3.17
1000 ft^2	180.76	2.2
2000 ft^2	144.91	1.76
1/4 acre	94.56	1.15
1/2 acre	82.33	1
1 acre	71.74	0.87
2 acre	63.51	0.77
5 acre	54.62	0.66
10 acre	49.13	0.6
32 acre	41.55	0.5
100 acre	35.66	0.43

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR SOIL DATA	**Statistical Guidesheet 18**
GENERIC RESIDENTIAL AND COMMERCIAL I PARTICULATE SOIL INHALATION CRITERIA (PSIC) FOR AMBIENT AIR	
May assist in evaluation of condition(s):	7
Applicability of Statistics	**YES**
Statistics are applicable for evaluating this exposure pathway. Use of statistics is practical when there are adequate data sets available in the *EXPOSURE UNITS* or *EMISSION SOURCE AREAS*, as appropriate to the *FACILITY*.	

Key Considerations:	**Section 1**	**Facility Determination**
		• For the purposes of a BEA, or initial *FACILITY* determination, property may be classified as a *FACILITY* if one or more samples contain concentrations above generic residential criteria or *BACKGROUND* concentrations, whichever is greater, provided that there is not a greater body of evidence that the property is not a *FACILITY*. However, if concentrations of a hazardous substance from the property are less than the *STATEWIDE DEFAULT BACKGROUND* value in Operational Memorandum #15 for that hazardous substance, the property is not a *FACILITY* with respect to that hazardous substance. For purposes other than *FACILITY* determination (e.g., remediation and/or closure), additional data will likely be required. • If a statistical analysis is relied upon for making a *FACILITY* determination see Sections 2 and 3.
	Section 2	**Selection of Data For Statistical Analysis**
		• Samples collected for the purpose of characterizing a *FACILITY* are typically biased, based on factors such as historical information, previous sampling, disposal practices, visual impacts, and aerial photos. Once the nature and extent of contamination has been defined, it is necessary to identify and/or obtain data that will allow for appropriate comparison to criteria. There are two primary considerations in determining if data sets are adequate. First, data sets must be obtained from locations that are consistent with the exposure assumptions for the relevant land use scenario. Second, if statistics are used, data sets must contain a sufficient number of *RANDOMLY* located sample results to allow for proper statistical analysis and development of *REPRESENTATIVE CONCENTRATIONS*. Therefore, additional sampling to develop *REPRESENTATIVE CONCENTRATIONS* will often be required after the nature and extent of contamination have been defined. Some characterization data may be used in the development of a *REPRESENTATIVE CONCENTRATION* as described in section 2.4.2 of the tabbed section titled, “Sampling Strategies.” • A statistical analysis of soil data should be completed only if the nature and extent of contamination, including any *HOT SPOTS*, has been adequately defined. • It is not appropriate to combine samples from *HOT SPOT* areas with samples from other areas of a property for statistical analysis. This is necessary to avoid averaging or diluting the samples that represent *HOT SPOTS*. *HOT SPOTS* must be addressed separately. See section 2.2.1.1 of the tabbed section titled, “Sampling Strategies.”

	Section 2: (Continued)	★ A *SOURCE SIZE MODIFIER* must be selected for *EMISSION SOURCE SIZES* that are different than the assumed 1/2 acre *EMISSION SOURCE SIZE*. The *SOURCE SIZE MODIFIER* must correspond to an *EMISSION SOURCE SIZE* that is at least as large as the *EMISSION SOURCE SIZE* of the *FACILITY*. For example, if the *EMISSION SOURCE SIZE* at a *FACILITY* is at least eight acres, the generic soil inhalation criteria for a 1/2 acre *EMISSION SOURCE SIZE* is multiplied by the *SOURCE SIZE MODIFIER* for 10 acres to provide generic criteria for a 10 acre *FACILITY*. *SOURCE SIZE MODIFIERS* to adjust generic criteria for *EMISSION SOURCE SIZES* other than a 1/2 acre are provided in a table at the end of this guidesheet.
	Section 3	**Risk Analysis – Comparison to Criteria** ★ Criteria for this pathway depend on the *EMISSION SOURCE SIZE*, which must be estimated in both the horizontal and vertical dimension. Criteria shown in the cleanup criteria tables are based on a 1/2 acre *EMISSION SOURCE SIZE*. *SOURCE SIZE MODIFIERS* to adjust generic criteria for *EMISSION SOURCE SIZES* other than a 1/2 acre are provided in a table at the end of this guidesheet. For convenience, this table has been updated to include 1/4 acre and two acre *EMISSION SOURCE SIZES* for use at residential/commercial I and commercial II, III, IV and industrial land uses, respectively. Soil contamination for the PSIC pathway may occur in two general patterns that will affect the *EMISSION SOURCE SIZE* used to adjust the criteria for comparison to the *FACILITY* data: ★ **Properties or *FACILITIES* with a single *EMISSION SOURCE AREA*** If hazardous substances are detected in only a limited area within a 1/4 acre *EXPOSURE UNIT* and are not detected in other areas of the *EXPOSURE UNIT*, the *EMISSION SOURCE AREA* is equal to only the horizontal extent of the area with detectable concentrations. Only data within the *EMISSION SOURCE AREA* may be included in a statistical analysis. For an *EMISSION SOURCE AREA* that is larger than a 1/4 acre (i.e., *EXPOSURE UNIT*), soil concentrations within each 1/4 acre *EXPOSURE UNIT* must meet the PSIC adjusted for the *EMISSION SOURCE SIZE.* ★ **Properties with multiple *EMISSION SOURCE AREAS*** Properties larger than a 1/4 acre may contain several small *EMISSION SOURCE AREAS*. Where hazardous substances are not detected in the areas of the property between *EMISSION SOURCE AREAS*, the final *EMISSION SOURCE SIZE* is the sum of the horizontal extent of all individual *EMISSION SOURCE AREAS*. Soil concentrations within each *EMISSION SOURCE AREA* are compared to the generic criteria adjusted for the summed area of the individual *EMISSION SOURCE AREA* sizes. ★ Statistical analysis is possible for either the *EXPOSURE UNIT* or the *EMISSION SOURCE AREA* provided that sufficient data are available as described in Section 2.

	Section 3: (Continued)	★ Nine *RANDOMLY* located samples per *EXPOSURE UNIT* or *EMISSION SOURCE AREA*, whichever is smaller, should generally be used to conduct the statistical analysis. The actual number of samples to be collected may vary based on size of the *EMISSION SOURCE AREA*. Data collected for other purposes (as for identification of nature and extent) may be used where *RANDOM* sample locations fall on or reasonably close to existing sample locations and where data will be consistent in terms of sampling and analytical methods. • Statistical analyses must include an evaluation of the underlying statistical distribution of the data (i.e., normal, lognormal, or neither) and the level of censoring (i.e., proportion of the data below the detection limit). See the tabbed section titled, Statistical Methods." This evaluation must be completed for each hazardous substance. Additionally, the statistical methods used to compare *FACILITY* data to criteria must be selected based on statistical distribution and level of censoring. ★ Compliance with infinite generic PSIC must be demonstrated for the entire vertical soil column since hazardous substances in subsurface soil may contribute to emissions if moved to the surface in the future. • Statistical analysis of *FACILITY* data is not allowed for hazardous substances that have criteria based on acute toxicological effects and/or physical hazards. • If *BACKGROUND* concentrations in soil are greater than criteria, soil data may be compared to *BACKGROUND* instead. Statistical methods for determining *BACKGROUND* concentrations will vary depending on which type of *BACKGROUND* is being considered. See Statistical Guidesheet 10 and Chapter 4 of the tabbed section titled, "Statistical Methods."
	Section 4	**Verification of Remediation or Closure** • **No Action Needed**: If the data evaluation completed under Sections 2 and 3 demonstrated that there are no exceedances of cleanup criteria, no remedial action is required for this pathway and consequently no additional sampling or data analysis is required under Section 4. • **Excavation**: Numbers and locations of samples collected for verifying remediation of soil can be selected in accordance with either the tabbed section titled, "Sampling Strategies" (Sections 1.3 or 2.3 as appropriate based on size of excavation), or a DEQ-approved sampling plan when a person is seeking DEQ approval of the response activity. Analytical results from verification sampling must be compared to Part 201 criteria on a point-by-point basis* unless a minimum of nine *RANDOMLY* located samples are available within the *EXPOSURE UNIT(S)* or *EMISSION SOURCE AREA*, whichever is smaller. If a statistical analysis is used, analytical results from verification sampling must be compared to Part 201 criteria as described in Sections 2 and 3. • **In situ/Ex situ Treatment**: Plans for verifying remediation or closure must be made on a case-by-case basis and must demonstrate how treatment is effective in meeting cleanup criteria. These plans may or may not depend on statistical analysis of data as allowed in Sections 2 and 3.

	Section 4: (Continued)	• **Cover/Containment**: Generally soil sampling will not be used to verify Limited or Site-specific remediation or closure that relies on a cover or containment. Instead, verification of remediation will involve verifying the integrity of the cover or containment structure on an ongoing basis. Verification of remediation in areas outside the cover/containment structure will be done using the options above.
Additional Information:		Criteria Application Guidesheet 18; "Part 201 Generic Soil Inhalation Criteria for Ambient Air: Technical Support Document"
Recommended Statistical Methods for Comparison to Criteria:		Compare *FACILITY* data to criteria using a 95% upper confidence limit (UCL) for the mean concentration (see Chapter 3 of the tabbed section titled, "Statistical Methods"). Other statistical methods may be acceptable on a case-by-case basis.

*Professional judgment is required to interpret the significance of exceedances that are very close to criteria, or that may be associated with insignificant quantities of hazardous substances.

***SOURCE SIZE MODIFIERS* for Ambient Air Soil Inhalation Criteria**

Modifiers		
Source Size (ft^2 or acres)	Q/C (g/m^2-s per kg/m^3)	Modifier
400 ft^2	261.26	3.17
1000 ft^2	180.76	2.2
2000 ft^2	144.91	1.76
1/4 acre	94.56	1.15
1/2 acre	82.33	1
1 acre	71.74	0.87
2 acre	63.51	0.77
5 acre	54.62	0.66
10 acre	49.13	0.6
32 acre	41.55	0.5
100 acre	35.66	0.43

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR SOIL DATA	**Statistical Guidesheet 19**
GENERIC RESIDENTIAL AND COMMERCIAL I DIRECT CONTACT CRITERIA (DCC)	
May assist in evaluation of condition(s):	6
Applicability of Statistics	**YES**

Statistics are applicable for evaluating this exposure pathway. Use of statistics is practical when there are adequate data sets available for the *EXPOSURE UNITS*.

Key Considerations:

Section 1

Facility Determination

- For the purposes of a BEA, or initial *FACILITY* determination, property may be classified as a *FACILITY* if one or more samples contain concentrations above generic residential criteria or *BACKGROUND* concentrations, whichever is greater, provided that there is not a greater body of evidence that the property is not a *FACILITY*. However, if concentrations of a hazardous substance from the property are less than the *STATEWIDE DEFAULT BACKGROUND* value in Operational Memorandum #15 for that hazardous substance, the property is not a *FACILITY* with respect to that hazardous substance. For purposes other than *FACILITY* determination (e.g., remediation and/or closure), additional data will likely be required.
- If a statistical analysis is relied upon for making a *FACILITY* determination see Sections 2 and 3.

Section 2

Selection of Data For Statistical Analysis

- Samples collected for the purpose of characterizing a *FACILITY* are typically biased, based on factors such as historical information, previous sampling, disposal practices, visual impacts, and aerial photos. Once the nature and extent of contamination has been defined, it is necessary to identify and/or obtain data that will allow for appropriate comparison to criteria. There are two primary considerations in determining if data sets are adequate. First, data sets must be obtained from locations that are consistent with the exposure assumptions for the relevant land use scenario. Second, if statistics are used, data sets must contain a sufficient number of *RANDOMLY* located sample results to allow for proper statistical analysis and development of *REPRESENTATIVE CONCENTRATIONS*. Therefore, additional sampling to develop *REPRESENTATIVE CONCENTRATIONS* will often be required after the nature and extent of contamination have been defined. Some characterization data may be used in the development of a *REPRESENTATIVE CONCENTRATION* as described in Section 2.4.2 of the tabbed section titled, “Sampling Strategies.”
- A statistical analysis of soil data should be completed only if the nature and extent of contamination, including any *HOT SPOTS*, has been adequately defined.
- It is not appropriate to combine samples from *HOT SPOT* areas with samples from other areas of a property for statistical analysis. This is necessary to avoid averaging or diluting the samples that represent *HOT SPOTS*. *HOT SPOTS* must be addressed separately. See Section 2.2.1.1 of the tabbed section titled, “Sampling Strategies.”

★ Generally only soil data from a 1/4 acre *EXPOSURE UNIT* may be used in a statistical calculation for remedial compliance.

	Section 2: (Continued)	★ Soil data from an *EXPOSURE UNIT* that reflects current human activity patterns may be used in a statistical calculation for Due Care compliance. ★ Compliance with these criteria must be demonstrated separately for surface and subsurface soils unless there is little variability between surface and subsurface data. Surface soils are typically defined as the top six inches of the soil column. However, if contamination is predominantly located at the immediate surface (such as through air deposition of hazardous substances), surface soil samples should represent the immediate surface (e.g., top one inch). ★ Subsurface soils should be evaluated in the same manner as surface soils; however, larger *EXPOSURE UNITS* and/or fewer samples may be acceptable if characterization of the property indicates that contamination is predominantly at the surface. ★ *EXPOSURE UNITS* larger than the standard 1/4 acre may be used for both Due Care and Remedial Compliance if it has been demonstrated that there is little variability among concentrations of hazardous substances in the *EXPOSURE UNIT* to be evaluated.
	Section 3	**Risk Analysis – Comparison to Criteria** ★ For remedial compliance, the *FACILITY* must generally be divided into 1/4 acre *EXPOSURE UNITS* for comparison to generic residential criteria. *EXPOSURE UNIT* size can be different for surface and subsurface soil. Areas classified as *HOT SPOTS* must be evaluated separately. A minimum of nine *RANDOMLY* located samples per *EXPOSURE UNIT* (considering surface soil and subsurface soil separately) are required to conduct the statistical analysis. This number is based on statistical considerations only. Additional samples may be necessary to adequately characterize and represent spatial variability in the *EXPOSURE UNIT* depending on such factors as soil type and size of the *EXPOSURE UNIT*. ★ *EXPOSURE UNITS* for Due Care should be based on exposures currently occurring and reasonably likely to occur based on human activity patterns at the *FACILITY*. If activities are concentrated in an area smaller than the standard *EXPOSURE UNIT*, *EXPOSURE UNITS* smaller than 1/4 acre should be used. • Statistical analyses must include an evaluation of the underlying statistical distribution of the data (i.e., normal, lognormal, or neither) and the level of censoring (i.e., proportion of the data below the detection limit). See the tabbed section titled, “Statistical Methods.” This evaluation must be completed for each hazardous substance. The statistical methods used to compare *FACILITY* data to criteria must be selected based on statistical distribution and level of censoring. • If *BACKGROUND* concentrations in soil are greater than criteria, soil data may be compared to *BACKGROUND* instead. Statistical methods for determining *BACKGROUND* concentrations will vary depending on which type of *BACKGROUND* is being considered. See Statistical Guidesheet 10 and Chapter 4 of the tabbed section titled, “Statistical Methods.” • Statistical analysis of *FACILITY* data is not allowed for hazardous substances present above Csat screening levels without further evaluation of risk and the extent of the area exceeding Csat. • Statistical analysis of *FACILITY* data is not allowed for hazardous substances that have criteria based on acute toxicological effects and/or physical hazards.

	Section 4	**Verification of Remediation or Closure**
		• **No Action Needed**: If the data evaluation completed under Sections 2 and 3 demonstrated that there are no exceedances of cleanup criteria, no remedial action is required for this pathway and consequently no additional sampling or data analysis is required under Section 4. • **Excavation**: Numbers and locations of samples collected for verifying remediation of soil can be selected in accordance with either the tabbed section titled, "Sampling Strategies" (Sections 1.3 or 2.3 as appropriate based on size of excavation), or a DEQ-approved sampling plan when a person is seeking DEQ approval of the response activity. Analytical results from verification sampling must be compared to Part 201 criteria on a point-by-point basis* unless a minimum of nine *RANDOMLY* located samples are available within the *EXPOSURE UNIT(S).* If a statistical analysis is used, analytical results from verification sampling must be compared to Part 201 criteria as described in Sections 2 and 3. • **In situ/Ex situ Treatment**: Plans for verifying remediation or closure must be made on a case-by-case basis and must demonstrate how treatment is effective in meeting cleanup criteria. These plans may or may not depend on statistical analysis of data as allowed in Sections 2 and 3. • **Cover/Containment**: Generally soil sampling will not be used to verify Limited or Site-specific remediation or closure that relies on a cover or containment. Instead, verification of remediation will involve verifying the integrity of the cover or containment structure on an ongoing basis. Verification of remediation in areas outside the cover/containment structure will be done using the options above. • If *BACKGROUND* concentrations in soil are greater than criteria, soil data may be compared to *BACKGROUND* instead. Statistical methods for determining *BACKGROUND* concentrations will vary depending on which type of *BACKGROUND* is being considered. See Statistical Guidesheet 10 and Chapter 4 of the tabbed section titled, "Statistical Methods."
Additional Information:		Criteria Application Guidesheet 19
Recommended Statistical Methods for Comparison to Criteria:		Compare *FACILITY* data to criteria using a 95% upper confidence limit (UCL) for the mean concentration (see Chapter 3 of the tabbed section titled, "Statistical Methods"). Other statistical methods may be acceptable on a case-by-case basis.

*Professional judgment is required to interpret the significance of exceedances that are very close to criteria, or that may be associated with insignificant quantities of hazardous substances.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR SOIL DATA	Statistical Guidesheet 20
GENERIC SOIL SATURATION (CSAT) SCREENING LEVELS	
May assist in evaluation of condition(s):	1 and 6-12
Applicability of Statistics	**Pathway Dependent**
See the Statistical Guidesheet for the exposure pathway where the cleanup criterion defaults to Csat to determine the applicability of statistics.	

Key Considerations:	**Section 1**	**Facility Determination**
		• See the Statistical Guidesheet for the generic criteria that defaults to the Csat screening level.
	Section 2	**Selection of Data For Statistical Analysis**
		• See the Statistical Guidesheet for the generic criteria that defaults to the Csat screening level. • Data that are greater than Csat should not be included in a statistical analysis unless a site-specific evaluation has shown that there is no free-phase hazardous substance. This does not apply to the generic volatile soil inhalation criteria (VSIC). See Section 3 of Statistical Guidesheets 15, 16 and 17 and 23, 24 and 25 for guidance in conducting statistical analysis of data with concentrations greater than Csat.
	Section 3	**Risk Analysis – Comparison to Criteria**
		• See Section 2.
	Section 4	**Verification of Remediation or Closure**
		• See Section 2.

Additional Information:	Criteria Application Guidesheet 20 and A, and Appendix A; Part 201 Generic Soil Saturation Screening Concentrations: Technical Support Document.
Recommended Statistical Methods for Comparison to Criteria:	See Statistical Guidesheet 7 for relevant exposure pathways. Pathway dependent.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR **SOIL** DATA	Statistical Guidesheet 21
GENERIC SOIL CRITERIA PROTECTIVE OF COMMERCIAL II, III, IV AND INDUSTRIAL DRINKING WATER	
May assist in evaluation of condition(s):	9
Applicability of Statistics	**GNP**
Statistics are generally not practical as a tool for assessing this exposure pathway because of the difficulty in projecting the impact on groundwater of hazardous substances in soil and the need to assure that drinking water cleanup criteria are met at all points in the aquifer.	

Key Considerations:

Section 1

Facility Determination

- Not applicable.

Section 2

Selection of Data For Statistical Analysis

- Samples collected for the purpose of characterizing a *FACILITY* are typically biased, based on factors such as historical information, previous sampling, disposal practices, visual impacts, and aerial photos. Once the nature and extent of contamination has been defined, it is necessary to identify and/or obtain data that will allow for appropriate comparison to criteria. There are two primary considerations in determining if data sets are adequate. First, data sets must be obtained from locations that are consistent with the exposure assumptions for the relevant land use scenario. Second, if statistics are used, data sets must contain a sufficient number of *RANDOMLY* located sample results to allow for proper statistical analysis and development of *REPRESENTATIVE CONCENTRATIONS*. Therefore, additional sampling to develop *REPRESENTATIVE CONCENTRATIONS* will often be required after the nature and extent of contamination have been defined. Some characterization data may be used in the development of a *REPRESENTATIVE CONCENTRATION* as described in Section 2.4.2 of the tabbed section titled, "Sampling Strategies."

- A statistical analysis is appropriate only if the nature and extent of contamination, including any *HOT SPOTS*, has been adequately defined.

- It is not appropriate to combine samples from *HOT SPOT* areas with samples from other areas of a property for statistical analysis. This is necessary to avoid averaging or diluting the samples that represent *HOT SPOTS*. *HOT SPOTS* must be addressed separately. See Section 2.2.1.1 of the tabbed section titled, "Sampling Strategies."

Section 3

Risk Analysis – Comparison to Criteria

★ Statistical analysis of *FACILITY* soil concentrations for comparison to soil criteria protective of the drinking water criteria may be acceptable in limited circumstances. Sufficient data must be available to demonstrate that areas of contaminated soil above criteria are not large enough to result in groundwater concentrations in an aquifer above criteria. This may require a fairly rigorous data set that is not often practical to obtain. Similarly, statistical analysis of *FACILITY* leachate concentrations for comparison to drinking water criteria may also be acceptable in limited circumstances.

	Section 3: (Continued)	• If *BACKGROUND* concentrations in soil are greater than criteria, soil data may be compared to *BACKGROUND* instead. Statistical methods for comparing to *BACKGROUND* will also vary depending on which type of *BACKGROUND* is being considered. See Statistical Guidesheet 10 and Chapter 4 of the tabbed section titled, "Statistical Methods." • Statistical analysis of *FACILITY* data is not allowed for hazardous substances present above Csat screening levels without further evaluation of risk and the extent of the area exceeding Csat. • Statistical analysis of *FACILITY* data is not allowed for hazardous substances that have criteria based on acute toxicological effects and/or physical hazards.
	Section 4	**Verification of Remediation or Closure** • **No Action Needed**: If the data evaluation completed under Sections 2 and 3 demonstrated that there are no exceedances of cleanup criteria, no remedial action is required for this pathway and consequently no additional sampling or data analysis is required under Section 4. • **Excavation**: Analytical results from verification sampling will generally be compared to Part 201 criteria on a point-by-point basis* in unsaturated soil. Numbers and locations of samples collected for verifying remediation of soil by excavation can be selected in accordance with either the tabbed section titled, "Sampling Strategies" (Sections 1.3 or 2.3 as appropriate based on size of excavation), or a DEQ-approved sampling plan. • **In situ/Ex situ Treatment**: Plans for verifying remediation or closure must be made on a case-by-case basis and must demonstrate how treatment is effective in meeting cleanup criteria. These plans may or may not depend on statistical analysis of data as allowed in Sections 2 and 3. • **Cover/Containment**: Generally soil sampling will not be used to verify remediation or closure that relies on a cover or containment. Instead, verification of remediation will involve verifying the integrity of the cover or containment structure on an ongoing basis. Groundwater sampling may be used in some cases to verify the effectiveness of the cover/containment. Verification of remediation in areas outside the cover/containment structure will be done using the options above. • If *BACKGROUND* concentrations in soil are greater than criteria, soil data may be compared to *BACKGROUND* instead. Statistical methods for comparing to *BACKGROUND* will also vary depending on which type of *BACKGROUND* is being considered. See Statistical Guidesheet 10 and Chapter 4 of the tabbed section titled, "Statistical Methods."

Additional Information:	Criteria Application Guidesheet 21; Statistical Guidesheet 2
Recommended Statistical Methods for Comparison to Criteria:	If statistical analysis is documented to be practical, the recommended statistical method is comparison of *FACILITY* data to criteria using a 95% upper confidence limit (UCL) for the mean concentration (see Chapter 3 of the tabbed section titled, "Statistical Methods"). Other statistical methods may be acceptable on a case-by-case basis.

*Professional judgment is required to interpret the significance of exceedances that are very close to criteria, or that may be associated with insignificant quantities of hazardous substances.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR SOIL DATA	Statistical Guidesheet 22

GENERIC COMMERCIAL II, III, IV AND INDUSTRIAL SOIL VOLATILIZATION TO INDOOR AIR INHALATION CRITERIA (SVIIC)

May assist in evaluation of condition(s):	8
Applicability of Statistics	**GNP**

Statistics are generally not practical as a tool for assessing this exposure pathway because it is unlikely that there will be a sufficient number of samples available from the generic building footprint size.

Key Considerations:

Section 1

Facility Determination

- Not applicable.

Section 2

Selection of Data For Statistical Analysis

- Samples collected for the purpose of characterizing a *FACILITY* are typically biased, based on factors such as historical information, previous sampling, disposal practices, visual impacts, and aerial photos. Once the nature and extent of contamination has been defined, it is necessary to identify and/or obtain data that will allow for appropriate comparison to criteria. There are two primary considerations in determining if data sets are adequate. First, data sets must be obtained from locations that are consistent with the exposure assumptions for the relevant land use scenario. Second, if statistics are used, data sets must contain a sufficient number of *RANDOMLY* located sample results to allow for proper statistical analysis and development of *REPRESENTATIVE CONCENTRATIONS*. Therefore, additional sampling to develop *REPRESENTATIVE CONCENTRATIONS* will often be required after the nature and extent of contamination have been defined. Some characterization data may be used in the development of a *REPRESENTATIVE CONCENTRATION* as described in Section 2.4.2 of the tabbed section titled, "Sampling Strategies."

- A statistical analysis is appropriate only if the nature and extent of contamination, including any *HOT SPOTS*, has been adequately defined.

- It is not appropriate to combine data from *HOT SPOT* areas with data from other areas of a property for statistical analysis. This is necessary to avoid averaging out or diluting the samples that represent *HOT SPOTS*. *HOT SPOTS* must be addressed separately. See Section 2.2.1.1 of the tabbed section titled, "Sampling Strategies."

★ Generally only soil data from a 4,000 ft^2 *EXPOSURE UNIT* (i.e., the building footprint) may be used in a statistical calculation for remedial compliance.

Section 3

Risk Analysis – Comparison to Criteria

★ Statistical analysis of *FACILITY* soil concentrations for comparison to residential SVIIC is generally not practical since sufficient data (i.e., a minimum of nine *RANDOMLY* located samples) are generally not available to allow for a statistical evaluation within the generic building footprint area of 4,000 ft^2. Therefore, comparisons to SVIIC must generally be completed on a point-by-point basis* (i.e., each concentration at each location must be compared individually to criteria).

	Section 3: (Continued)	★ Any statistical analysis must consider the potential for spatial variability of soil types to influence vapor migration. • If DEQ approval of a response activity is being sought, a *PROPOSAL* for a statistical analysis must be submitted to the DEQ for approval to assure that data needs and/or complexities of the pathway are addressed. • Statistical analysis of *FACILITY* data is not allowed for hazardous substances present above Csat screening levels without further evaluation of risk and the extent of the area exceeding Csat.
	Section 4	**Verification of Remediation or Closure** • **No Action Needed**: If the data evaluation completed under Sections 2 and 3 demonstrated that there are no exceedances of cleanup criteria, no remedial action is required for this pathway and consequently no additional sampling or data analysis is required under Section 4. • **Excavation**: Statistical analysis of *FACILITY* soil data is generally not practical. See Section 3. Analytical results from verification sampling will therefore generally be compared to Part 201 criteria on a point-by-point basis* in unsaturated soil. Numbers and locations of samples collected for point-by-point comparison can be selected in accordance with either the tabbed section titled, "Sampling Strategies" (Sections 1.3 or 2.3 as appropriate based on size of excavation), or a DEQ-approved sampling plan. If a statistical analysis is relied upon for verifying remediation or closure follow the guidance presented in Sections 2 and 3 of this Guidesheet. • **In situ/Ex situ Treatment**: Plans for verifying remediation or closure must be made on a case-by-case basis and must demonstrate how treatment is effective in meeting cleanup criteria. These plans may or may not depend on statistical analysis of data. • **Cover/Containment**: In the context of this pathway, cover/containment will typically be a vapor barrier. Generally soil sampling will not be used to verify remediation or closure that relies on a cover or containment. Verification of remediation in areas outside the cover/containment structure will be done using the options above.
Additional Information:		Criteria Application Guidesheet 22
Recommended Statistical Methods for Comparison to Criteria:		If statistical analysis is documented to be practical, the recommended statistical method is comparison of *FACILITY* data to criteria using a 95% upper confidence limit (UCL) for the mean concentration (see Chapter 3 of the tabbed section titled, "Statistical Methods"). Other statistical methods may be acceptable on a case-by-case basis.

*Professional judgment is required to interpret the significance of exceedances that are very close to criteria, or that may be associated with insignificant quantities of hazardous substances.

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR SOIL DATA	**Statistical Guidesheets 23, 24 and 25**

GENERIC COMMERCIAL II, III, IV AND INDUSTRIAL INFINITE SOURCE VOLATILE SOIL INHALATION CRITERIA (VSIC) FOR AMBIENT AIR

May assist in evaluation of condition(s):	7
Applicability of Statistics	**YES**

Statistics are applicable for evaluating this exposure pathway. Use of statistics is practical when there are adequate data sets available in the *EXPOSURE UNITS* or *EMISSION SOURCE AREAS*, as appropriate to the *FACILITY*.

Key Considerations:

Section 1

Facility Determination

- Not applicable.

Section 2

Selection of Data For Statistical Analysis

- Samples collected for the purpose of characterizing a *FACILITY* are typically biased, based on factors such as historical information, previous sampling, disposal practices, visual impacts, and aerial photos. Once the nature and extent of contamination has been defined, it is necessary to identify and/or obtain data that will allow for appropriate comparison to criteria. There are two primary considerations in determining if data sets are adequate. First, data sets must be obtained from locations that are consistent with the exposure assumptions for the relevant land use scenario. Second, if statistics are used, data sets must contain a sufficient number of *RANDOMLY* located sample results to allow for proper statistical analysis and development of *REPRESENTATIVE CONCENTRATIONS*. Therefore, additional sampling to develop *REPRESENTATIVE CONCENTRATIONS* will often be required after the nature and extent of contamination have been defined. Some characterization data may be used in the development of a *REPRESENTATIVE CONCENTRATION* as described in Section 2.4.2 of the tabbed section titled, "Sampling Strategies."

- A statistical analysis of soil data should be completed only if the nature and extent of contamination, including any *HOT SPOTS*, has been adequately defined.

- It is not appropriate to combine samples from *HOT SPOT* areas with samples from other areas of a property for statistical analysis. This is necessary to avoid averaging or diluting the samples that represent *HOT SPOTS*. *HOT SPOTS* must be addressed separately. See Section 2.2.1.1 of the tabbed section titled, "Sampling Strategies."

★ The horizontal and vertical extent of the *EMISSION SOURCE AREA* must be estimated to allow for selection of the *SOURCE SIZE MODIFIER* and infinite or finite source generic VSIC. Only data from the estimated *EMISSION SOURCE AREA* may be used in statistical analysis for comparison to criteria. See the "Part 201 Generic Soil Inhalation Criteria for Ambient Air: Technical Support Document" for information about source size characterization.

★ Characterization of the horizontal extent of contamination is necessary for both finite and infinite generic VSIC to estimate the *EMISSION SOURCE AREA* contributing to volatile emissions from soil.

Section 2: (Continued)

★ Infinite generic VSIC are applicable to the entire contaminated vertical soil column since both surface and subsurface soil concentrations of hazardous substances may contribute to volatile emissions.

★ For an *EMISSION SOURCE AREA* that is larger than two acres (i.e., *EXPOSURE UNIT*), soil concentrations within each two acre *EXPOSURE UNIT* must meet the VSIC adjusted for the *EMISSION SOURCE SIZE*.

★ A *SOURCE SIZE MODIFIER* must be selected for *EMISSION SOURCE SIZES* that are different than the assumed 1/2 acre *EMISSION SOURCE SIZE*. The *SOURCE SIZE MODIFIER* must correspond to an *EMISSION SOURCE SIZE* that is at least as large as the *EMISSION SOURCE SIZE* of the *FACILITY*. For example, if the *EMISSION SOURCE SIZE* at a *FACILITY* is at least eight acres, the generic soil inhalation criteria for a 1/2 acre *EMISSION SOURCE SIZE* is multiplied by the *SOURCE SIZE MODIFIER* for 10 acres to provide generic criteria for a 10 acre *FACILITY*. *SOURCE SIZE MODIFIERS* to adjust generic criteria for *EMISSION SOURCE SIZES* other than a 1/2 acre are provided in a table at the end of this guidesheet.

Section 3

Risk Analysis – Comparison to Criteria

★ Criteria for this pathway depend on the *EMISSION SOURCE SIZE*, which must be estimated in both the horizontal and vertical dimension. Criteria shown in the cleanup criteria tables are based on a 1/2 acre *EMISSION SOURCE SIZE*. *SOURCE SIZE MODIFIERS* to adjust generic criteria for *EMISSION SOURCE SIZES* other than a 1/2 acre are provided in a table at the end of this guidesheet. For convenience, this table has been updated to include 1/4 acre and two acre *EMISSION SOURCE SIZES* for use at residential/commercial I and commercial II, III, IV and industrial land uses, respectively.

Soil contamination for the VSIC pathway may occur in two general patterns that will affect the *EMISSION SOURCE SIZE* used to adjust the criteria for comparison to the *FACILITY* data:

★ **Properties or *FACILITIES* with a single *EMISSION SOURCE AREA***

If hazardous substances are detected in only a limited area within a two acre *EXPOSURE UNIT* and are not detected in other areas of the *EXPOSURE UNIT*, the *EMISSION SOURCE AREA* is equal to only the horizontal extent of the area with detectable concentrations. Only data within the *EMISSION SOURCE AREA* may be included in a statistical analysis.

For an *EMISSION SOURCE AREA* that is larger than two acres (i.e., *EXPOSURE UNIT*), soil concentrations within each two acre *EXPOSURE UNIT* must meet the VSIC adjusted for the *EMISSION SOURCE SIZE*.

★ **Properties with multiple *EMISSION SOURCE AREAS***

Properties larger than a two acre may contain several small *EMISSION SOURCE AREAS*. Where hazardous substances are not detected in the areas of the property between *EMISSION SOURCE AREAS*, the final *EMISSION SOURCE SIZE* is the sum of the horizontal extent of all individual *EMISSION SOURCE AREAS*. Soil concentrations within each *EMISSION SOURCE AREA* are compared to the generic criteria adjusted for the summed area of the individual *EMISSION SOURCE AREA* sizes.

	Section 3: (Continued)	★ Generally only soil data from a two acre *EXPOSURE UNIT* may be used in a statistical calculation for remedial compliance unless the property is smaller than two acres and the *FACILITY* is confined to the property. In that case, the *EXPOSURE UNIT* size will be same as the property size. <u>However, in cases where the *EMISSION SOURCE AREA* is smaller than two acres only soil data from the *EMISSION SOURCE AREA* may be used in a statistical calculation</u>. ★ Statistical analysis is possible for either the *EXPOSURE UNIT* or the *EMISSION SOURCE AREA* provided that sufficient data are available as described in Section 2. ★ Nine *RANDOMLY* located samples per *EXPOSURE UNIT* or *EMISSION SOURCE AREA*, whichever is smaller, should generally be used to conduct the statistical analysis. The actual number of samples to be collected may vary based on size of the *EMISSION SOURCE AREA*. Data collected for other purposes (as for identification of nature and extent) may be used where *RANDOM* sample locations fall on or reasonably close to existing sample locations and where data will be consistent in terms of sampling and analytical methods. • Statistical analyses must include an evaluation of the underlying statistical distribution of the data (i.e., normal, lognormal, or neither) and the level of censoring (i.e., proportion of the data below the detection limit). See the tabbed section titled, "Statistical Methods." This evaluation must be completed for each hazardous substance. Additionally, the statistical methods used to compare *FACILITY* data to criteria must be selected based on statistical distribution and level of censoring. ★ Compliance with infinite generic VSIC must be demonstrated for the entire vertical soil column since both surface and subsurface soil concentrations of hazardous substances may contribute to volatile emissions. • Since a hazardous substance will <u>not</u> volatilize more when present at concentrations greater than Csat, sample data that exceeds Csat can be included in a statistical analysis to determine compliance with the VSIC as long as the Csat exceedance is not a *HOT SPOT*. • Statistical analysis of *FACILITY* data is not allowed for hazardous substances that have criteria based on acute toxicological effects and/or physical hazards.
	Section 4	**Verification of Remediation or Closure** • **No Action Needed**: If the data evaluation completed under Sections 2 and 3 demonstrated that there are no exceedances of cleanup criteria, no remedial action is required for this pathway and consequently no additional sampling or data analysis is required under Section 4. • **Excavation**: Numbers and locations of samples collected for verifying remediation of soil can be selected in accordance with either the tabbed section titled, "Sampling Strategies" (Sections 1.3 or 2.3 as appropriate based on size of excavation), or a DEQ-approved sampling plan when a person is seeking DEQ approval of the response activity. Analytical results from verification sampling must be compared to Part 201 criteria on a point-by-point basis* unless a minimum of nine *RANDOMLY* located samples are available within the *EXPOSURE UNIT(S)* or *EMISSION SOURCE AREA*, whichever is smaller. If a statistical analysis is used, analytical results from verification sampling must be compared to Part 201 criteria as described in Sections 2 and 3.

	Section 4: (Continued)	• **In situ/Ex situ Treatment**: Plans for verifying remediation or closure must be made on a case-by-case basis and must demonstrate how treatment is effective in meeting cleanup criteria. These plans may or may not depend on statistical analysis of data as allowed in Sections 2 and 3. • **Cover/Containment**: Generally soil sampling will not be used to verify Limited or Site-specific remediation or closure that relies on a cover or containment. Instead, verification of remediation will involve verifying the integrity of the cover or containment structure on an ongoing basis. Verification of remediation in areas outside the cover/containment structure will be done using the options above.
Additional Information:		Criteria Application Guidesheets 23, 24 and 25; "Part 201 Generic Soil Inhalation Criteria for Ambient Air: Technical Support Document".
Recommended Statistical Methods for Comparison to Criteria:		Compare *FACILITY* data to criteria using a 95% upper confidence limit (UCL) for the mean concentration (see Chapter 3 of the tabbed section titled, "Statistical Methods"). Other statistical methods may be acceptable on a case-by-case basis.

*Professional judgment is required to interpret the significance of exceedances that are very close to criteria, or that may be associated with insignificant quantities of hazardous substances.

***SOURCE SIZE MODIFIERS* for Ambient Air Soil Inhalation Criteria**

Modifiers		
Source Size (ft^2 or acres)	Q/C (g/m^2-s per kg/m^3)	Modifier
400 ft^2	261.26	3.17
1000 ft^2	180.76	2.2
2000 ft^2	144.91	1.76
1/4 acre	94.56	1.15
1/2 acre	82.33	1
1 acre	71.74	0.87
2 acre	63.51	0.77
5 acre	54.62	0.66
10 acre	49.13	0.6
32 acre	41.55	0.5
100 acre	35.66	0.43

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR **SOIL** DATA	Statistical Guidesheet 26

GENERIC COMMERCIAL II, III, IV AND INDUSTRIAL PARTICULATE SOIL INHALATION CRITERIA (PSIC) FOR AMBIENT AIR

May assist in evaluation of condition(s):	7
Applicability of Statistics	**YES**

Statistics are applicable for evaluating this exposure pathway. Use of statistics is practical when there are adequate data sets available in the *EXPOSURE UNITS* or *EMISSION SOURCE AREAS*, as appropriate to the *FACILITY*.

Key Considerations:

Section 1

Facility Determination

- Not applicable.

Section 2

Selection of Data For Statistical Analysis

- Samples collected for the purpose of characterizing a *FACILITY* are typically biased, based on factors such as historical information, previous sampling, disposal practices, visual impacts, and aerial photos. Once the nature and extent of contamination has been defined, it is necessary to identify and/or obtain data that will allow for appropriate comparison to criteria. There are two primary considerations in determining if data sets are adequate. First, data sets must be obtained from locations that are consistent with the exposure assumptions for the relevant land use scenario. Second, if statistics are used, data sets must contain a sufficient number of *RANDOMLY* located sample results to allow for proper statistical analysis and development of *REPRESENTATIVE CONCENTRATIONS*. Therefore, additional sampling to develop *REPRESENTATIVE CONCENTRATIONS* will often be required after the nature and extent of contamination have been defined. Some characterization data may be used in the development of a *REPRESENTATIVE CONCENTRATION* as described in Section 2.4.2 of the tabbed section titled, "Sampling Strategies."

- A statistical analysis of soil data should be completed only if the nature and extent of contamination, including any *HOT SPOTS*, has been adequately defined.

- It is not appropriate to combine samples from *HOT SPOT* areas with samples from other areas of a property for statistical analysis. This is necessary to avoid averaging or diluting the samples that represent *HOT SPOTS*. *HOT SPOTS* must be addressed separately. See Section 2.2.1.1 of the tabbed section titled, "Sampling Strategies."

★ A *SOURCE SIZE MODIFIER* must be selected for *EMISSION SOURCE SIZES* that are different than the assumed 1/2 acre *EMISSION SOURCE SIZE*. The *SOURCE SIZE MODIFIER* must correspond to an *EMISSION SOURCE SIZE* that is at least as large as the *EMISSION SOURCE SIZE* of the *FACILITY*. For example, if the *EMISSION SOURCE SIZE* at a *FACILITY* is at least eight acres, the generic soil inhalation criteria for a 1/2 acre *EMISSION SOURCE SIZE* is multiplied by the *SOURCE SIZE MODIFIER* for 10 acres to provide generic criteria for a 10 acre *FACILITY*. *SOURCE SIZE MODIFIERS* to adjust generic criteria for *EMISSION SOURCE SIZES* other than a 1/2 acre are provided in a table at the end of this guidesheet.

Section 3

Risk Analysis – Comparison to Criteria

★ Criteria for this pathway depend on the *EMISSION SOURCE SIZE*, which must be estimated in both the horizontal and vertical dimension. Criteria shown in the cleanup criteria tables are based on a 1/2 acre *EMISSION SOURCE SIZE*. *SOURCE SIZE MODIFIERS* to adjust generic criteria for *EMISSION SOURCE SIZES* other than a 1/2 acre are provided in a table at the end of this guidesheet. For convenience, this table has been updated to include 1/4 acre and two acre *EMISSION SOURCE SIZES* for use at residential/commercial I and commercial II, III, IV and industrial land uses, respectively.

Soil contamination for the PSIC pathway may occur in two general patterns that will affect the *EMISSION SOURCE SIZE* used to adjust the criteria for comparison to the *FACILITY* data:

★ **Properties or *FACILITIES* with a single *EMISSION SOURCE AREA***

If hazardous substances are detected in only a limited area within a two acre *EXPOSURE UNIT* and are not detected in other areas of the *EXPOSURE UNIT*, the *EMISSION SOURCE AREA* is equal to only the horizontal extent of the area with detectable concentrations. Only data within the *EMISSION SOURCE AREA* may be included in a statistical analysis.

For an *EMISSION SOURCE AREA* that is larger than two acres (i.e., *EXPOSURE UNIT*), soil concentrations within each two acre *EXPOSURE UNIT* must meet the PSIC adjusted for the *EMISSION SOURCE SIZE*.

★ **Properties with multiple *EMISSION SOURCE AREAS***

Properties larger than two acres may contain several small *EMISSION SOURCE AREAS*. Where hazardous substances are not detected in the areas of the property between *EMISSION SOURCE AREAS*, the final *EMISSION SOURCE SIZE* is the sum of the horizontal extent of all individual *EMISSION SOURCE AREAS*. Soil concentrations within each *EMISSION SOURCE AREA* are compared to the generic criteria adjusted for the summed area of the individual *EMISSION SOURCE AREA* sizes.

★ Statistical analysis is possible for either the *EXPOSURE UNIT* or the *EMISSION SOURCE AREA* provided that sufficient data are available as described in Section 2.

★ Nine *RANDOMLY* located samples per *EXPOSURE UNIT* or *EMISSION SOURCE AREA*, whichever is smaller, should generally be used to conduct the statistical analysis. The actual number of samples to be collected may vary based on size of the *EMISSION SOURCE AREA*. Data collected for other purposes (as for identification of nature and extent) may be used where *RANDOM* sample locations fall on or reasonably close to existing sample locations and where data will be consistent in terms of sampling and analytical methods.

- Statistical analyses must include an evaluation of the underlying statistical distribution of the data (i.e., normal, lognormal, or neither) and the level of censoring (i.e., proportion of the data below the detection limit). See the tabbed section titled, Statistical Methods." This evaluation must be completed for each hazardous substance. Additionally, the statistical methods used to compare *FACILITY* data to criteria must be selected based on statistical distribution and level of censoring.

	Section 3: (Continued)	★ Compliance with infinite generic PSIC must be demonstrated for the entire vertical soil column since hazardous substances in subsurface soil may contribute to emissions if moved to the surface in the future. • Statistical analysis of *FACILITY* data is not allowed for hazardous substances that have criteria based on acute toxicological effects and/or physical hazards. • If *BACKGROUND* concentrations in soil are greater than criteria, soil data may be compared to *BACKGROUND* instead. Statistical methods for determining *BACKGROUND* concentrations will vary depending on which type of *BACKGROUND* is being considered. See Statistical Guidesheet 10 and Chapter 4 of the tabbed section titled, "Statistical Methods."
	Section 4	**Verification of Remediation or Closure** • **No Action Needed**: If the data evaluation completed under Sections 2 and 3 demonstrated that there are no exceedances of cleanup criteria, no remedial action is required for this pathway and consequently no additional sampling or data analysis is required under Section 4. • **Excavation**: Numbers and locations of samples collected for verifying remediation of soil can be selected in accordance with either the tabbed section titled, "Sampling Strategies" (Sections 1.3 or 2.3 as appropriate based on size of excavation), or a DEQ-approved sampling plan when a person is seeking DEQ approval of the response activity. Analytical results from verification sampling must be compared to Part 201 criteria on a point-by-point basis* unless a minimum of nine *RANDOMLY* located samples are available within the *EXPOSURE UNIT(S)* or *EMISSION SOURCE AREA*, whichever is smaller. If a statistical analysis is used, analytical results from verification sampling must be compared to Part 201 criteria as described in Sections 2 and 3. • **In situ/Ex situ Treatment**: Plans for verifying remediation or closure must be made on a case-by-case basis and must demonstrate how treatment is effective in meeting cleanup criteria. These plans may or may not depend on statistical analysis of data as allowed in Sections 2 and 3. • **Cover/Containment**: Generally soil sampling will not be used to verify Limited or Site-specific remediation or closure that relies on a cover or containment. Instead, verification of remediation will involve verifying the integrity of the cover or containment structure on an ongoing basis. Verification of remediation in areas outside the cover/containment structure will be done using the options above.

Additional Information:	Criteria Application Guidesheet 22; "Part 201 Generic Soil Inhalation Criteria for Ambient Air: Technical Support Document"
Recommended Statistical Methods for Comparison to Criteria:	Compare *FACILITY* data to criteria using a 95% upper confidence limit (UCL) for the mean concentration (see Chapter 3 of the tabbed section titled, "Statistical Methods"). Other statistical methods may be acceptable on a case-by-case basis.

*Professional judgment is required to interpret the significance of exceedances that are very close to criteria, or that may be associated with insignificant quantities of hazardous substances.

SOURCE SIZE MODIFIERS for Ambient Air Soil Inhalation Criteria

Modifiers		
Source Size (ft^2 or acres)	Q/C (g/m^2-s per kg/m^3)	Modifier
400 ft^2	261.26	3.17
1000 ft^2	180.76	2.2
2000 ft^2	144.91	1.76
1/4 acre	94.56	1.15
1/2 acre	82.33	1
1 acre	71.74	0.87
2 acre	63.51	0.77
5 acre	54.62	0.66
10 acre	49.13	0.6
32 acre	41.55	0.5
100 acre	35.66	0.43

PART 201 STATISTICAL EVALUATION GUIDESHEET FOR SOIL DATA	Statistical Guidesheets 27, 28 and 29
GENERIC INDUSTRIAL AND COMMERCIAL II, III, IV DIRECT CONTACT CRITERIA (DCC)	
May assist in evaluation of condition(s):	6
Applicability of Statistics	**YES**
Statistics are applicable for evaluating this exposure pathway. Use of statistics is practical when there are adequate data sets available for the *EXPOSURE UNITS*.	

Key Considerations:	**Section 1**	**Facility Determination**
		• Not applicable.
	Section 2	**Selection of Data For Statistical Analysis**
		• Samples collected for the purpose of characterizing a *FACILITY* are typically biased, based on factors such as historical information, previous sampling, disposal practices, visual impacts, and aerial photos. Once the nature and extent of contamination has been defined, it is necessary to identify and/or obtain data that will allow for appropriate comparison to criteria. There are two primary considerations in determining if data sets are adequate. First, data sets must be obtained from locations that are consistent with the exposure assumptions for the relevant land use scenario. Second, if statistics are used, data sets must contain a sufficient number of *RANDOMLY* located sample results to allow for proper statistical analysis and development of *REPRESENTATIVE CONCENTRATIONS*. Therefore, additional sampling to develop *REPRESENTATIVE CONCENTRATIONS* will often be required after the nature and extent of contamination have been defined. Some characterization data may be used in the development of a *REPRESENTATIVE CONCENTRATION* as described in Section 2.4.2 of the tabbed section titled, "Sampling Strategies."
		• A statistical analysis of soil data should be completed only if the nature and extent of contamination, including any *HOT SPOTS*, has been adequately defined.
		• It is not appropriate to combine samples from *HOT SPOT* areas with samples from other areas of a property for statistical analysis. This is necessary to avoid averaging or diluting the samples that represent *HOT SPOTS*. *HOT SPOTS* must be addressed separately. See Section 2.2.1.1 of the tabbed section titled, "Sampling Strategies."
		★ Generally only soil data from a two acre *EXPOSURE UNIT* may be used in a statistical calculation for remedial compliance unless the property is smaller than 2 acres and the *FACILITY* is confined to the property. In that case, the *EXPOSURE UNIT* size will be same as the property size.
		★ Soil data from an *EXPOSURE UNIT* that reflects current human activity patterns may be used in a statistical calculation for Due Care compliance.

	Section 2: (Continued)	★ Compliance with these criteria must be demonstrated separately for surface and subsurface soils unless there is little variability between surface and subsurface data. Surface soils are typically defined as the top six inches of the soil column. However, if contamination is predominantly located at the immediate surface (such as through air deposition of hazardous substances), surface soil samples should represent the immediate surface (e.g., top one inch). ★ Subsurface soils should be evaluated in the same manner as surface soils; however, larger *EXPOSURE UNITS* and/or fewer samples may be acceptable if characterization of the property indicates that hazardous substances are predominantly at the surface. ★ *EXPOSURE UNITS* larger than the standard two acres may be used for both Due Care and Remedial Compliance if it has been demonstrated that there is little variability among concentrations of hazardous substances in the *EXPOSURE UNIT* to be evaluated.
	Section 3	**Risk Analysis – Comparison to Criteria** ★ For remedial compliance, the *FACILITY* must generally be divided into two acre *EXPOSURE UNITS* for comparison to these criteria. *EXPOSURE UNIT* size can be different for surface and subsurface soil. Areas classified as *HOT SPOTS* must be evaluated separately. A minimum of nine *RANDOMLY* located samples per *EXPOSURE UNIT* (considering surface soil and subsurface soil separately) are required to conduct the statistical analysis. This number is based on statistical considerations only. Additional samples may be necessary to adequately characterize and represent spatial variability in the *EXPOSURE UNIT* depending on such factors as soil type, and size of the *EXPOSURE UNIT*. ★ *EXPOSURE UNITS* for Due Care should be based on exposures currently occurring and reasonably likely to occur based on human activity patterns at the *FACILITY*. If activities are concentrated in an area smaller than the standard *EXPOSURE UNIT*, *EXPOSURE UNITS* smaller than two acres should be used. • Statistical analyses must include an evaluation of the underlying statistical distribution of the data (i.e., normal, lognormal, or neither) and the level of censoring (i.e., proportion of the data below the detection limit). See the tabbed section titled, “Statistical Methods.” This evaluation must be completed for each hazardous substance. The statistical methods used to compare *FACILITY* data to criteria must be selected based on statistical distribution and level of censoring. • If *BACKGROUND* concentrations in soil are greater than criteria, soil data may be compared to *BACKGROUND* instead. Statistical methods for determining *BACKGROUND* concentrations will vary depending on which type of *BACKGROUND* is being considered. See Statistical Guidesheet 10 and Chapter 4 of the tabbed section titled, “Statistical Methods.” • Statistical analysis of *FACILITY* data is not allowed for hazardous substances present above Csat screening levels without further evaluation of risk and the extent of the area exceeding Csat. • Statistical analysis of *FACILITY* data is not allowed for hazardous substances that have criteria based on acute toxicological effects and/or physical hazards.

	Section 4	**Verification of Remediation or Closure**
		• **No Action Needed**: If the data evaluation completed under Sections 2 and 3 demonstrated that there are no exceedances of cleanup criteria, no remedial action is required for this pathway and consequently no additional sampling or data analysis is required under Section 4. • **Excavation**: Numbers and locations of samples collected for verifying remediation of soil can be selected in accordance with either the tabbed section titled, "Sampling Strategies" (Sections 1.3 or 2.3 as appropriate based on size of excavation), or a DEQ-approved sampling plan when a person is seeking DEQ approval of the response activity. Analytical results from verification sampling must be compared to Part 201 criteria on a point-by-point basis* unless a minimum of nine *RANDOMLY* located samples are available within the *EXPOSURE UNIT(S).* If a statistical analysis is used analytical results from verification sampling must be compared to Part 201 criteria as described in Sections 2 and 3. • **In situ/Ex situ Treatment**: Plans for verifying remediation or closure must be made on a case-by-case basis and must demonstrate how treatment is effective in meeting cleanup criteria. These plans may or may not depend on statistical analysis of data. • **Cover/Containment**: Generally soil sampling will not be used to verify Limited or Site-specific remediation or closure that relies on a cover or containment. Instead, verification of remediation will involve verifying the integrity of the cover or containment structure on an ongoing basis. Verification of remediation in areas outside the cover/containment structure will be done using the options above. • If *BACKGROUND* concentrations in soil are greater than criteria, soil data may be compared to *BACKGROUND* instead. Statistical methods for determining *BACKGROUND* concentrations will vary depending on which type of *BACKGROUND* is being considered. See Statistical Guidesheet 10 and Chapter 4 of the tabbed section titled, "Statistical Methods."
Additional Information:		Criteria Application Guidesheets 27, 28 & 29
Recommended Statistical Methods for Comparison to Criteria:		Compare *FACILITY* data to criteria using a 95% upper confidence limit (UCL) for the mean concentration (see Chapter 3 of the tabbed section titled, "Statistical Methods"). Other statistical methods may be acceptable on case-by-case basis.

*Professional judgment is required to interpret the significance of exceedances that are very close to criteria, or that may be associated with insignificant quantities of hazardous substances.

USE OF STATISTICS IN ASSESSING "DUE CARE" COMPLIANCE

USE OF STATISTICS IN ASSESSING "DUE CARE" COMPLIANCE

Staff and *FACILITY* owner/operators may need to evaluate data to assess compliance with Section 7a ("Due Care") and the Part 10 rules. Generally, the applicability of statistics will be the same in the Due Care context for the other purposes described in the Statistical Guidesheets (e.g., verification of remediation). Therefore, if the Applicability of Statistics box on the Statistical Guidesheet for an exposure pathway or condition denotes YES, it is appropriate to use statistics in a Due Care evaluation, if the data are sufficient. If Applicability is denoted as GNP, it is important to determine whether an adequate data set is available for evaluation of Due Care. The minimum data requirements are the same for Due Care as for other applications (e.g., number of samples per *EXPOSURE UNIT*, need for data to reflect *REPRESENTATIVE CONCENTRATIONS*).

The goal of Section 7a(1)(b) is protection of public health and safety. Compliance with this Section requires that response activity be undertaken to mitigate unacceptable exposure to hazardous substances and fire and explosion hazards. Recall that under Rule 1003(4)(a)(i), Due Care compliance is evaluated for the exposure pathways that are complete or likely to be complete in light of the intended use of the property. Not all exposure pathways and conditions to consider are relevant to Due Care. For example, the drinking water ingestion pathway is not relevant when drinking water is provided from a municipal supply that is unaffected by contamination from the *FACILITY*. The GSI pathway is generally not relevant for Due Care. Also, the intended use at a *FACILITY* may limit exposure to contaminated media, such that only a portion of the contaminated media needs to evaluated. For example, if planned activities at a *FACILITY* would not result in excavation below the top 12 inches of soil, the direct contact pathway would not need to be evaluated for soil more than 12 inches deep.

The following table summarizes how Due Care compliance may vary from evaluation in the remedial action context.

Statistical Guidesheet	Pathway	How Due Care and Use of Statistics to Evaluate Compliance May Be Different than Remedial Compliance
A	Abandoned substances not yet dispersed & free phase liquids	Source control may contribute to preventing or mitigating unacceptable exposures. Statistics generally not practical in defining or assessing effectiveness of source control activities because this is typically a qualitative evaluation.
B	Polluted soil runoff to surface water	Generally not relevant to Due Care compliance.
C	Aquatic flora/fauna/food chain hazards/ aesthetics	Generally not relevant to Due Care compliance.
D	Acute toxic impacts/physical hazards	This condition may need to be addressed as part of Due Care compliance to mitigate unacceptable exposures or fire and explosion hazards. Statistics generally not practical for evaluating compliance because of the acute nature of the hazards.
E	Terrestrial flora/fauna/food chain hazards/ aesthetics	Generally not relevant to Due Care compliance.

F	Asbestos	Asbestos may need to be addressed to mitigate unacceptable exposure. Relevant to Due Care in areas and through exposure pathways where asbestos exposures may occur. Use of statistics in those areas would be the same as described in the Statistical Guidesheet for the relevant exposure pathways.
1	Groundwater: Drinking water usage (Res)	Relevant to Due Care compliance if groundwater is being used for drinking water. Statistics not applicable for evaluating Due Care compliance.
2	Groundwater: Drinking water usage (C/I)	Relevant to Due Care compliance if groundwater is being used for drinking water. Statistics not applicable for evaluating Due Care compliance.
3	Groundwater: Hazards to surface waters (All)	Generally not relevant to Due Care compliance except for conditions in venting groundwater that could result in unacceptable human exposure at GSI (e.g., very high or low pH) or if there is a drinking water intake close to point of venting.
4	Groundwater: Indoor air hazards (chronic /systemic) (Res)	Relevant to Due Care compliance if volatile contaminants are present in groundwater under existing or planned structures. Use of statistics is the same as described in the Statistical Guidesheet. Due Care compliance can be based on size of existing or planned structure, and can take into account building characteristics (e.g., ventilation rates) that are different from generic assumptions. Data needs for *REPRESENTATIVE CONCENTRATION* and minimum number of data points per building footprint/*EXPOSURE UNIT* are the same as for remedial compliance. *HOT SPOTS* must be evaluated separately to determine if they represent an unacceptable risk.
5	Groundwater: Indoor air hazards (chronic /systemic) (C/I)	Relevant to Due Care compliance if volatile contaminants are present in groundwater under existing or planned structures. Use of statistics is the same as described in the Statistical Guidesheet. Due Care compliance can be based on size of existing or planned structure, and can take into account building characteristics (e.g., ventilation rates) that are different from generic assumptions. Data needs for *REPRESENTATIVE CONCENTRATION* and minimum number of data points per building footprint/*EXPOSURE UNIT* are the same as for remedial compliance. *HOT SPOTS* must be evaluated separately to determine if they represent an unacceptable risk.
6	Groundwater: Dermal exposures such as by utility workers (All)	Relevant to Due Care compliance if contact with groundwater is reasonably likely to occur given intended use of property. Area evaluated can be limited to location of likely groundwater contact, in light of existing features. Use of statistics is the same as described in the Statistical Guidesheet. Data needs for *REPRESENTATIVE CONCENTRATION* and minimum number of data points per building footprint/*EXPOSURE UNIT* are the same as for remedial compliance. *HOT SPOTS* must be evaluated separately to determine if they represent an unacceptable risk.

7	Groundwater: Water Solubility (All)	Relevant to Due Care if a generic cleanup criterion for an exposure pathway or condition that is relevant defaults to water solubility. Use of statistics is the same as described in the Statistical Guidesheet for the relevant pathway.
8	Groundwater: Flammability / Explosivity (All)	This condition may need to be addressed as part of Due Care compliance to mitigate unacceptable fire and explosion hazards. Statistics generally not practical for evaluating compliance because of the acute nature of the hazards.
9	Groundwater: Acute Inhalation Risks (All)	This condition may need to be addressed as part of Due Care compliance to mitigate unacceptable exposures. Statistics generally not practical for evaluating compliance because of the acute nature of the hazards.
10	Background (All)	*BACKGROUND* may be relevant in evaluating Due Care compliance if an applicable cleanup criterion defaults to *BACKGROUND*. Use of statistics is the same as described in the Statistical Guidesheet for the relevant exposure pathway.
11	Soil: Injury to drinking water use of aquifer (Res)	Generally not relevant to Due Care compliance.
12	Soil: Causes groundwater to be hazardous to surface water (All)	Generally not relevant to Due Care compliance.
13	Soil: Risk from contact (utility work) with groundwater (All)	Generally not relevant to Due Care compliance.
14	Soil: Indoor air inhalation hazards (Res)	Relevant to Due Care compliance if volatile contaminants are present in soil under existing or planned structures. Use of statistics is the same as described in the Statistical Guidesheet. Due Care compliance can be based on size of an existing or planned structure, and can take into account building characteristics (e.g., ventilation rates) that are different from generic assumptions. Data needs for *REPRESENTATIVE CONCENTRATION* and minimum number of data points per building footprint/*EXPOSURE UNIT* are the same as for remedial compliance. *HOT SPOTS* must be evaluated separately to determine if they represent an unacceptable risk.
15, 16, 17	Soil: Ambient air inhalation hazards (Volatile) (All)	Relevant to Due Care compliance if volatile contaminants are present in soil. Use of statistics is the same as described in the Statistical Guidesheet. *EXPOSURE UNIT* size may be modified from generic size if appropriate to the activity patterns on the property. *EMISSION SOURCE AREA* the same as for remedial compliance. Data needs for *REPRESENTATIVE CONCENTRATION* and minimum number of data points per *EXPOSURE UNIT* or *EMISSION SOURCE AREA* are the same as for remedial compliance. *HOT SPOTS* must be evaluated separately to determine if they represent an unacceptable risk.

18	Soil: Ambient air inhalation hazards (Particulate) (C/I)	Relevant to Due Care compliance. Use of statistics is the same as described in the Statistical Guidesheet. Depth of soil evaluated can be limited if activity at property will not result in disturbance of existing surface soils. Data needs for *REPRESENTATIVE CONCENTRATION* and minimum number of data points per *EXPOSURE UNIT* or *EMISSION SOURCE AREA* are the same as for remedial compliance. *HOT SPOTS* must be evaluated separately to determine if they represent an unacceptable risk.
19	Hazards due to direct contact (Res)	Relevant to Due Care compliance. Use of statistics is the same as described in the Statistical Guidesheet. Depth of soil evaluated can be limited if activity at property will not result in disturbance of existing surface soils. *EXPOSURE UNIT* size can be based on existing and planned activity patterns, rather than standard size. Data needs for *REPRESENTATIVE CONCENTRATION* and minimum number of data points per *EXPOSURE UNIT* are the same as for remedial compliance. *HOT SPOTS* must be evaluated separately to determine if they represent an unacceptable risk.
20	Soil Saturation	Relevant to Due Care if a generic cleanup criterion for an exposure pathway or condition that is relevant defaults to Csat. Use of statistics is the same as described in the Statistical Guidesheet. Data needs for *REPRESENTATIVE CONCENTRATION* and minimum number of data points per building footprint/*EXPOSURE UNIT* are the same as for remedial compliance. *HOT SPOTS* must be evaluated separately to determine if they represent an unacceptable risk.
21	Soil: Injury to drinking water use of aquifer (C/I)	Generally not relevant to Due Care compliance.
22	Soil: Indoor air inhalation hazards (C/I)	Relevant to Due Care compliance if volatile contaminants are present in soil under existing or planned structures. Use of statistics is the same as described in the Statistical Guidesheet. Due Care compliance can be based on size of existing or planned structure, and can take into account building characteristics (e.g., ventilation rates) that are different from generic assumptions. Data needs for *REPRESENTATIVE CONCENTRATION* and minimum number of data points per building footprint/*EXPOSURE UNIT* are the same as for remedial compliance. *HOT SPOTS* must be evaluated separately to determine if they represent an unacceptable risk.

23, 24, 25	Soil: Ambient air inhalation hazards (Volatile) (C/I)	Relevant to Due Care compliance if volatile contaminants are present in soil. Use of statistics is the same as described in the Statistical Guidesheet. *EXPOSURE UNIT* size may be modified from generic size if appropriate to the activity patterns on the property. *EMISSION SOURCE AREA* the same as for remedial compliance. Data needs for *REPRESENTATIVE CONCENTRATION* and minimum number of data points per *EXPOSURE UNIT* or *EMISSION SOURCE AREA* are the same as for remedial compliance. *HOT SPOTS* must be evaluated separately to determine if they represent an unacceptable risk.
26	Soil: Ambient air inhalation hazards (Particulate) (C/I)	Relevant to Due Care compliance. Use of statistics is the same as described in the Statistical Guidesheet. Depth of soil evaluated can be limited if activity at property will not result in disturbance of existing surface soils. Data needs for *REPRESENTATIVE CONCENTRATION* and minimum number of data points per *EXPOSURE UNIT* or *EMISSION SOURCE AREA* are the same as for remedial compliance. *HOT SPOTS* must be evaluated separately to determine if they represent an unacceptable risk.
27, 28, 29	Soil: Hazards due to direct contact (C/I)	Relevant to Due Care compliance. Use of statistics is the same as described in the Statistical Guidesheet. Depth of soil evaluated can be limited if activity at property will not result in disturbance of existing surface soils. *EXPOSURE UNIT* size can be based on existing and planned activity patterns, rather than standard size. Data needs for *REPRESENTATIVE CONCENTRATION* and minimum number of data points per *EXPOSURE UNIT* are the same as for remedial compliance. *HOT SPOTS* must be evaluated separately to determine if they represent an unacceptable risk.

Key: Res – Residential and Commercial I Land Use Categories
C/I – Commercial II, III, IV and Industrial Land Use Categories
All – All Land Use Categories

STATISTICAL METHODS FOR COMPARING FACILITY DATA TO PART 201 CRITERIA

STATISTICAL METHODS TABLE OF CONTENTS

EXAMPLES

FIGURES

PROCEDURES

TABLES

INTRODUCTION

"[Statistics are] the only tools by which an opening can be cut through the formidable thicket of difficulties that bars the path of those who pursue the science of man."
-Sir Francis Galton

"By a small sample, we may judge of the whole piece."
-Miguel de Cervantes from Don Quixote

“Data! Data! Data! I can't make bricks without clay!”
-Sherlock Holmes

Use of this Document

Statistics provide important tools for describing and understanding the characteristics of data. Without statistics, quantifying the properties of a data set and drawing conclusions about the population from which the data were sampled is impossible. The following document has grown out of a need for a simple, yet comprehensive, approach to analyzing environmental data sets. It was composed with the intent of providing a clear, easy-to-use summary of the statistical methods available for various uses and how to implement these methods. Rather than providing a cookbook of statistical techniques where you just look up what you need and then “plug-and-chug” with the appropriate numbers and formulas, this document was meant to be used as a whole. Analyzing any data set is a process with several steps, not just one. This document is intended to provide a simple description and rationale behind those steps.

Common uses of statistics in Michigan’s Part 201 program include:

1) comparison of *FACILITY* data to the risk-based Part 201 criteria, and
2) comparison of *FACILITY* data to *FACILITY-SPECIFIC BACKGROUND*, *STATEWIDE DEFAULT BACKGROUND*, or *REGIONAL BACKGROUND* concentrations.

Statistical methods needed for these tasks are presented in the order that they are typically completed in conducting these statistical analyses. Chapter 1 provides recommended methods for evaluating the statistical distribution of a data set. Data sets will typically include either *FACILITY* data within a specified area (e.g., an *EXPOSURE UNIT*) or *FACILITY-SPECIFIC BACKGROUND* data. Chapter 2 describes methods for evaluating and recommendations for handling statistical outliers. Chapter 3 presents acceptable methods for comparing *FACILITY* data to Part 201 criteria. Chapter 4 provides statistical methods for comparing *FACILITY* data to the various types of *BACKGROUND* data that may be used. Recommendations for handling data below the detection limit or nondetects (generally substituting 1/2 of the detection limit) are provided throughout. Handling nondetects, including alternative methods, is discussed in more detail in Chapter 5.

For each method described herein, the following are provided: background information, assumptions underlying the method and their verification, a description including procedures for implementing the method, and an example analysis illustrating the steps that are required.

The approaches described in this document include use of summary statistics, graphical techniques, and formal statistical tests. Summary statistics help the user to understand characteristics of the data. Graphical techniques allow the user to view the data and greatly enhance the user’s understanding of the data. Formal statistical tests provide an objective

framework for making decisions about the data. Used together, these methods provide a strong framework for making decisions about and understanding the data.

Use of professional judgment is also an integral part of any statistical analysis. Statistics is a tool that may be used to help make decisions about Part 201 *FACILITIES*. However, used incorrectly, the decisions that are made may not reflect reality. Consequently, professional judgment must be used throughout in evaluating the data, selecting the appropriate statistical test, and drawing conclusions about a *FACILITY* based on a statistical analysis.

The statistical methods described herein are generally acceptable without approval when conducted within the framework described throughout the S^{3}TM. Alternative statistical methods are acceptable on a case-by-case basis. If departmental approval for a response action is being sought, alternative methods must be *PROPOSED* for approval by the DEQ statistician. The terms "*PROPOSE,*" "*PROPOSED,*" and "*PROPOSAL*" are used throughout the S^{3}TM to describe the process through which a statistical analysis of data is submitted to and reviewed by the department for approval, when necessary. Self-implemented response activities using statistics to support determinations must be documented in a manner that fully and clearly addresses the three questions outlined in the tabbed section titled, "Introduction."

Use of Computers in Statistical Analyses

Use of computer programs such as statistical software or spreadsheets is recommended to the extent possible when conducting statistical analyses. Calculation of statistical quantities by hand can not only be tedious, but is prone to error. Although formulas for calculating these quantities are provided herein, use of statistical software and/or spreadsheets is described and recommended throughout.

To highlight when a computer can be used to assist in an analysis, a computer icon is shown wherever computer "short-cuts" are provided. Microsoft Excel formulas are provided, where applicable, in the form:

=function(data range)

where "function" refers to the operation applied to the data (e.g., "average" or "stdev") and "data range" refers to the range over which the data to be evaluated appear on the spread sheet (e.g., A1:A20).

Occasionally a statistical method is a bit too complicated to conduct using a spreadsheet and a more powerful statistical software package is required. The DEQ has developed a customized statistical software package called "Statistical Interface for Part 201 Evaluations." This package was developed using a statistical software package called "S-PLUS StatServer 6" (StatServer) and many of the specialized statistical functions necessary for environmental data analysis provided by EnvironmentalStats for S-PLUS. This software is accessible to all staff in the department as well as the public through the Internet.

Interpretation of Statistical Results and Use of Professional Judgment

To aid understanding and comprehension, this document includes examples for each method presented. At the end of each example, interpretation of the results is discussed. The icon to the left is shown to highlight where information on interpretation of the statistical results can be found.

As stated above, use of professional judgment is an integral part of any statistical analysis and must be used when interpreting statistical results. The results of any statistical analysis should be carefully examined. This may be particularly important if the results of a statistical analysis do not conform with expectations. For example, if hazardous substance concentrations in an area of a property are clearly above criteria, yet a statistical analysis concludes that concentrations are below criteria, the test should be reevaluated. It is also possible that a statistical test will indicate that an area is above criteria when there are only one or two marginal exceedances of criteria. Several aspects of the analysis should be checked, including:

- Characterization – Was the property adequately characterized? (See the tabbed section titled, "Sampling Strategies.")

- Applicability of statistics – For exposure pathways/conditions in which use of statistics is "Generally Not Practical" (GNP), was it appropriate to conduct a statistical analysis of the data given the above? (See the appropriate Statistical Guidesheet.)

- Data set – Was the data set selected to represent a single population within an appropriate area (e.g., an *EXPOSURE UNIT*)? (See Sections 2 and 3 of the appropriate Statistical Guidesheet and the tabbed section titled, "Sampling Strategies.")

- Statistical test – Was the appropriate statistical test selected? Were the assumptions of the statistical test adequately evaluated? Were there any questions or uncertainties in whether the assumptions were met? Should an alternate statistical test be *PROPOSED*?

It is important that the above issues are addressed in any statistical analysis. If the entire S^3TM is followed through the course of an analysis, these issues will be addressed up front. However, unexpected or questionable conclusions may still be reached. It is possible that the original expectations about a data set or a *FACILITY* were incorrect and that the statistical analysis shed new insight on conditions at the *FACILITY*. However, the likelihood that a correct conclusion is drawn as well as a user's level of comfort with a statistical analysis can only be improved by scrutinizing statistical results.

CHAPTER 1: STATISTICAL DISTRIBUTIONS

Common uses of statistics in Michigan's Part 201 program include:

1) comparison of *FACILITY* data to Part 201 cleanup criteria, and
2) comparison of *FACILITY* data to *FACILITY-SPECIFIC BACKGROUND*, *STATEWIDE DEFAULT BACKGROUND*, or *REGIONAL BACKGROUND* concentrations.

Chapters 3 and 4 of Statistical Methods describe recommended statistical methods for these uses, respectively. However, before proceeding to these chapters, it is important to evaluate *FACILITY* data and, when available, *FACILITY-SPECIFIC BACKGROUND* data as described in Chapters 1 and 2 in order to select an appropriate statistical method for comparison.

1.1 IDENTIFICATION OF APPROPRIATE DATA SETS

Before utilizing statistics to compare *FACILITY* data to Part 201 criteria or *BACKGROUND* concentrations, it is important to first select the data set(s) you want to evaluate and the associated population(s). It is a relatively straightforward matter to enter numbers into formulas, but meaningful results and accurate conclusions are not likely unless careful consideration is given to the data being used in the analysis and the manner in which they were collected.

Proper identification of data sets for statistical analysis relies on adequate site characterization information. (See Sections 1.3 and 2.3 of the tabbed section titled, "Sampling Strategies.") Once a *FACILITY* has been adequately characterized, appropriate data sets must be selected as described in Parts 2 and 3 of the appropriate Statistical Guidesheets. Information on *HOT SPOTS* is also provided in Section 2.2.1.1 of the tabbed section titled, "Sampling Strategies." In addition, instructions for obtaining *RANDOM* samples is described in Section 2.4.2 of the tabbed section titled, "Sampling Strategies." This section also describes the importance of *RANDOM* sampling when comparing *FACILITY* data to Part 201 criteria, as well as recommendations on how to incorporate *FACILITY* characterization data into the data set for statistical analysis.

1.2 DISTRIBUTION TYPES

Before selecting between the various statistical methods provided in Chapters 3 and 4 for comparing *FACILITY* data to Part 201 risk-based criteria or *BACKGROUND* concentrations, respectively, it is important to understand the underlying assumptions for each method. Many statistical methods are based on an assumption that the data being evaluated come from a normal probability distribution. (The concepts of probability distributions and the normal distribution are described below.) If this assumption is not accurate, alternative methods may be required. Consequently, once an appropriate data set has been selected according to the considerations described in Section 1.1, the next step is to evaluate the data to determine its underlying statistical distribution.

In statistics, a probability distribution is mathematic rule or formula that gives the probability associated with obtaining various observations from a population. Values that have a high probability will be observed in *RANDOM* samples more frequently than values with low probability. Similarly, values with equal probability will show up in *RANDOM* samples with roughly the same frequency. As an example, for a fair coin toss there is a 50% chance of obtaining a "head" and a 50% chance of obtaining a "tail." This 50/50 chance of obtaining a head or a tail is an example of a simple probability distribution.

The most commonly known probability distribution is the normal distribution. The normal distribution can be illustrated by the familiar bell-shaped curve (Figure 1.1). The mean and variance determine the location and shape, respectively, of different normal curves. For example, Figure 1.2a illustrates the difference between two normal distributions with differing means, but the same variance. The shapes of these curves are the same, but the locations on the x-axis differ. Figure 1.2b shows two normal distributions with the same mean value, but differing variances. Although these curves are centered at the same location, the shapes of the curves are different. The curve with the larger variance is wider than the other curve, illustrating that there is a larger amount of variability in the distribution of the data. Many statistical methods are based on an assumption of normality (i.e., the data being evaluated come from a normal distribution). Alternate statistical methods may be needed if the data under evaluation are not normal.

The normal distribution can be used to describe positive-valued data, negative-valued data, or both. Because negative values are frequently not plausible for environmental data (e.g., negative concentrations), the normal distribution has some limitations in use.

Figure 1.1 Standard Normal Distribution

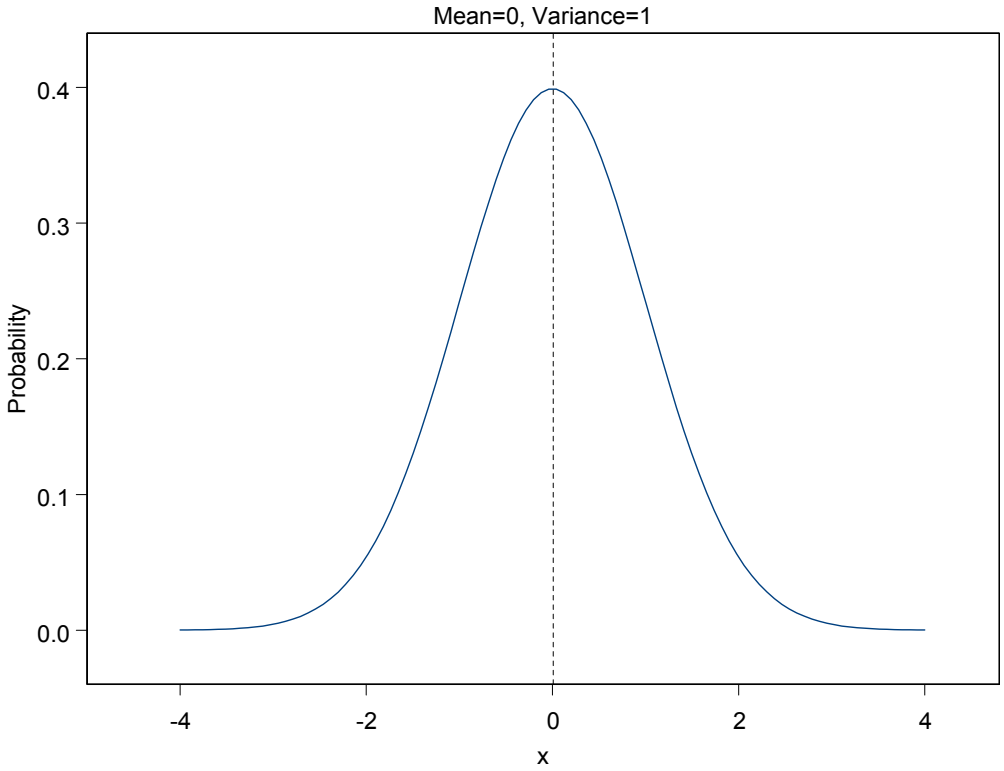


Figure 1.2a Two Normal Distributions (Different Means, Equal Variances)

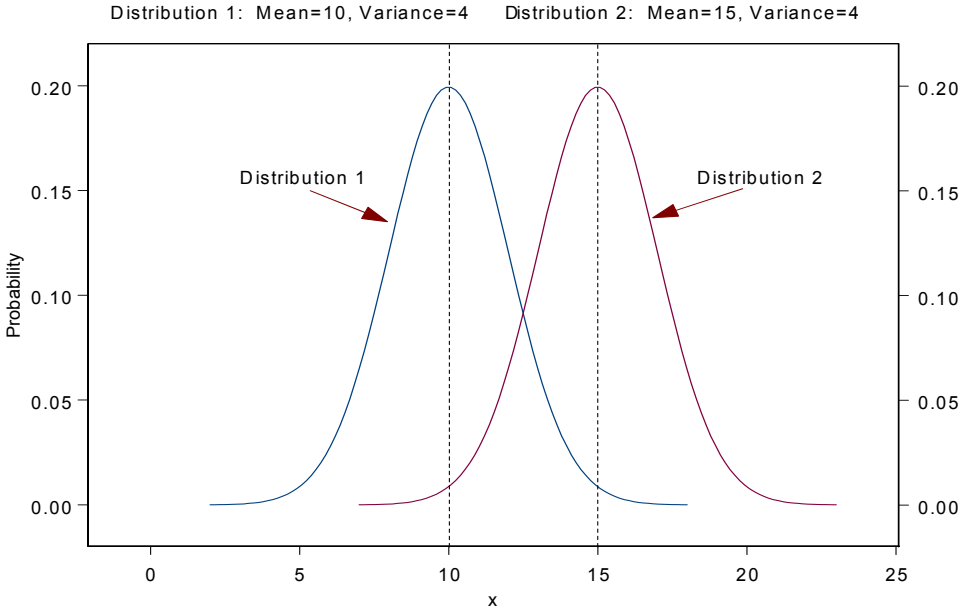
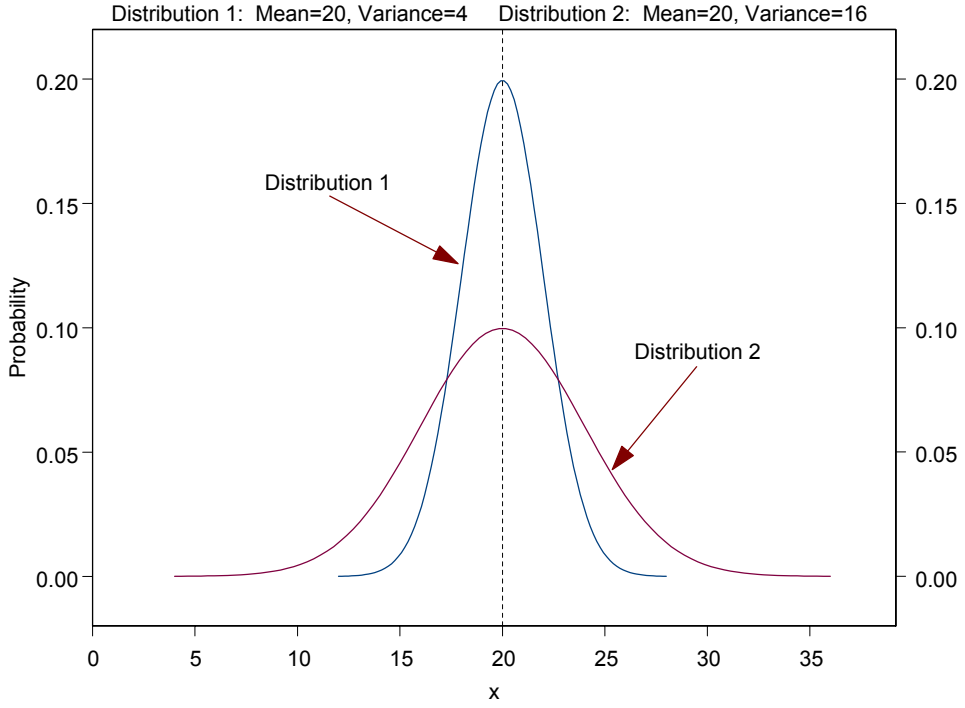
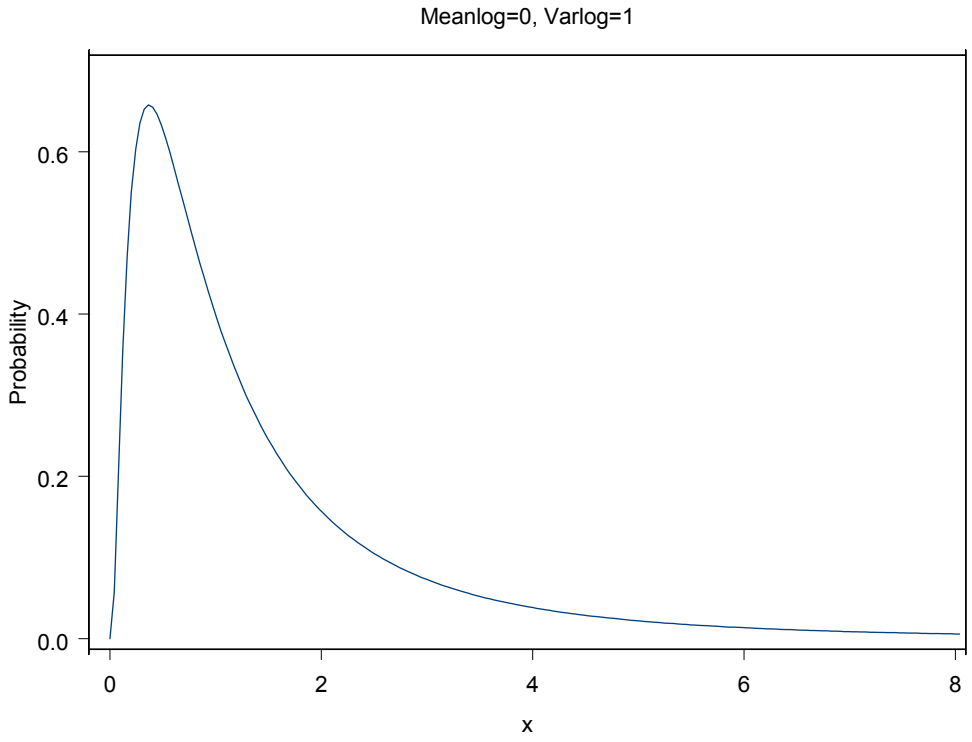


Figure 1.2b Two Normal Distributions (Equal Means, Different Variances)



Another common distribution used to model environmental data is the lognormal distribution (Figure 1.3). The lognormal distribution is characterized by a lower bound of zero and a right-skewed density function (i.e., the shape of the distribution is asymmetric with a long right tail). Because of the zero lower bound, the lognormal distribution can only be used to describe positive-valued data. These distributional characteristics often make the lognormal model a better candidate for describing environmental data sets. This idea is supported by Ott (1990) who demonstrates that there is a theoretical basis for the common occurrence of the lognormal distribution in environmental data.

Figure 1.3 Lognormal Distribution



As their names might suggest, there is a close relationship between the normal and lognormal distributions. Simply put, if a set of data $x_1, x_2, ..., x_n$ are lognormally distributed, then the set $y_1, y_2, ..., y_n$ representing the natural logs of the original observations (i.e., $y_i = \ln(x_i)$) will be normally distributed. In this case, log-transforming the original data set results in a normally-distributed data set.

Calculating the natural logs of the data is an example of a transformation of the original data set to achieve a normally-distributed data set. Transformations are used for changing the scale and range of data to obtain better statistical properties. Other methods for transforming data to normality are available (e.g., the Box-Cox family of transformations *PROPOSED* by Box and Cox (1964)); however, these methods are outside of the scope of this document. Transforming data to other scales often results in additional complexities when interpreting statistical results. Because of this and the common occurrence of the lognormal distribution in environmental data, the EPA (1992c) generally recommends testing for normality and lognormality only. Alternate methods for transforming data may be *PROPOSED* for review by the DEQ statistician.

Various statistical guidance documents suggest using a default assumption of normality or lognormality (i.e., test first for normality or lognormality; if the data pass a test for the default distribution, no further testing is recommended). **However, it is recommended here that all data be evaluated for *both* the normal and lognormal distributions. The results of these evaluations should be compared to determine which distribution provides a better fit to the data and, as a result, which methods are most appropriate.**

Experience has shown that some data sets do not conform to either distribution type. This may be due to many reasons, including:

- the presence of one or more outliers (see Chapter 2),
- the combination of data from multiple populations (e.g., *HOT SPOT* and non- *HOT SPOT* data) into a single data set, or
- the proportion of concentrations below the detection limit is too large (> 50%) to adequately evaluate the underlying distribution.

Appropriate selection of data sets as described in the Statistical Guidesheets should lessen problems associated with the first two points above. When the percent of concentrations below the detection limit is > 50%, it is generally not possible to evaluate the statistical distribution as described in Section 1.3. In some cases, alternate statistical methods are described for evaluating data sets with 50% or more of values below the detection limit. Consultation with a professional statistician is advised. In any case, alternate statistical methods may be *PROPOSED* for approval by the DEQ statistician.

1.3 METHODS FOR EVALUATING STATISTICAL DISTRIBUTIONS

Several methods are available for evaluating the statistical distribution of a data set. These methods include summary statistics, graphical techniques and formal tests. Most of these methods can be used to evaluate data for both normality and lognormality.

For each of these methods, it is recommended that a minimum of nine samples be used to evaluate for statistical distribution. A minimum of nine samples was selected in an effort to balance costs of sampling and analysis with statistical rigor. Tests for normality can be completed using as few as three samples (EPA 2000); however, the statistical power to conclude that the data are not normal (or lognormal) is low. Other sources recommend that sample sizes of 20 or more are required to attain reasonable power in tests of normality (e.g., EPA 1992a, EPA 1992c). Recognizing that this minimum number of samples must be collected from each *EXPOSURE UNIT* or area for which a statistical analysis is completed, the recommended minimum of samples was selected to represent the lower end of the range cited in literature. Because use of a relatively small sample size results in reduced power in tests of normality, a variety of techniques for evaluating distribution are recommended.

If it can be reasonably concluded that concentrations from multiple *EXPOSURE UNITS* or areas can be modeled by the same statistical distribution (e.g., due to placement of homogeneous fill materials over a large area), it may be appropriate to pool data from these areas when testing for distribution in order to increase sample size. If this approach is considered, a *PROPOSAL* should be submitted for review by the DEQ statistician. Consultation with a professional statistician is advised when developing this approach.

In some cases, it may be difficult to determine which distribution is most appropriate. **A final choice of statistical distribution should be made based on evaluation of the collective results of the different methods described below.**

A number of methods are provided in this section to evaluate data for a normal or lognormal distribution. These methods are summarized below:

Type of Method	Method	Section
Summary Statistic	Coefficient of Variation	1.3.1
Summary Statistic	Coefficient of Skewness	1.3.2
Graphical Technique	Probability Plots	1.3.3
Graphical Technique	Box Plots	1.3.4
Formal Test	Shapiro-Wilk Test ($n \leq 50$)	1.3.5
Formal Test	Shapiro-Francia Test ($n > 50$)	1.3.6
Formal Test	D'Agostino's Test ($n > 50$)	1.3.7

Summary statistics provide a simple way to evaluate the distribution of data. Summary statistics include descriptive values like the mean and standard deviation. These values are called summary statistics, or descriptive statistics, because they summarize the information contained in the data set and describe certain properties of the data set, such as central tendency and variability. The coefficient of variation and the coefficient of skewness are two more examples of summary statistics. The coefficient of variation measures the relative variability of the data set and provides a rough indication of the likelihood that the data are normal. The coefficient of skewness provides a measure of symmetry or asymmetry (i.e., skewness) of the data.

Graphical techniques provide a visual depiction of the data and are an excellent tool for inspecting data and their resemblance to a particular distribution. For example, a data set may fail a formal test of normality due to the presence of an outlier (i.e., without the outlier, the data would be normally distributed). This may not be apparent based on inspection of summary statistics or formal test results. Upon graphing the data, the cause for non-normality (an outlier, in this case) may be evident, allowing for possible modifications to the data (e.g., correction of an erroneous analytical result) and/or statistical approach. Note that identification and treatment of outliers is discussed in detail in Chapter 2. Because interpretation of graphical results may be somewhat subjective, decisions regarding statistical distribution should not be made using graphical methods alone.

Formal testing provides an objective framework for making decisions about the statistical distribution since a data set will either pass or fail a test for normality or lognormality. Subjectivity is largely absent when conducting a formal test and interpreting the result. Thus, formal tests provide the benefit of an objective framework for making decisions about data. However, as noted above, if a data set fails a formal test for normality or lognormality, the formal test results alone may not provide insight as to why a particular data set failed.

Used together, formal tests, graphical techniques, and summary statistics provide a strong framework for making decisions about and understanding the data. Consequently, a final choice of statistical distribution should be made based on evaluation of the collective results of the methods described below. **At a minimum, all statistical reports or *PROPOSALS* should include calculation of summary statistics, normal and lognormal probability plots, and**

the Shapiro-Wilk or Shapiro-Francia Tests for each contaminant in order to evaluate both normality and lognormality. Alternative methods for evaluating statistical distributions may be *PROPOSED* for review by the DEQ statistician.

1.3.1 Coefficient of Variation (CV)

The CV has been used historically by the DEQ (formerly the Department of Natural Resources [DNR]) as well as the Environmental Protection Agency (EPA) to evaluate the assumption of normality (DNR, 1994; EPA, 1989c). However, questions have been raised about the usefulness of using the CV for this purpose (EPA 1992c). The CV is sensitive to biased estimates of the mean should be used cautiously for censored data. As a measure of symmetry or asymmetry, the coefficient of skewness (Section 1.3.2) provides a more reliable tool. Furthermore, the Shapiro-Wilk and Shapiro-Francia tests (Sections 1.3.5 and 1.3.6, respectively) provide a better assessment of normality. **Consequently, it is recommended that the CV be calculated and interpreted together with the other methods presented below. Decisions regarding statistical distribution should not be made based solely on the CV.**

Procedure 1.1 Calculation of the Coefficient of Variation

To evaluate for normality, calculate the CV using the following equations. To evaluate for lognormality, first define y = ln(x) and use these y (log-transformed) values in place of the x values in the equations below. For nondetects, substitute 1/2 of the detection limit up to 50% nondetects.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \text{Equation 1.1}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \qquad \text{Equation 1.2}$$

$$CV = \frac{s}{\bar{x}} \qquad \text{Equation 1.3}$$

Where n represents the sample size, x_i represents the i th observed value, s represents the sample standard deviation, and $\bar{x}$ represents the sample mean. Note: The CV is often reported as a percentage (i.e., 100 times the CV calculated above).

It should not be necessary to calculate the above equations by hand considering the availability of basic statistical computations in statistical software packages and most spreadsheet packages. Most statistical software packages provide the CV as well as the sample standard deviation (s) and the sample mean ($\bar{x}$) together with other common summary statistics. Alternatively, Microsoft Excel can be used to obtain s and $\bar{x}$ for use in Equation 1.3. The Excel functions that should be used to obtain s and $\bar{x}$, respectively, are:

=STDEV(data range)

=AVERAGE(data range)

Note: The Excel function =STDEVP(data range) provides the population standard deviation (denominator = n) rather than the sample standard deviation (denominator = n-1). The population standard deviation is only appropriate for use with census data, which represent every unit of a population. In practice, census data are almost never available and the true standard deviation must be estimated. The sample standard deviation is an unbiased estimate of the true standard deviation (i.e., it does not consistently underestimate or overestimate the true standard deviation) and should be used throughout this document in all calculations calling for an estimate of the standard deviation.

Interpretation: If the data come from a normal distribution, then the CV will generally be less than one (EPA, 1989c). However, a CV < 1 does not automatically imply that the data are normal. The other methods described in this chapter must be considered to determine if the normal distribution is appropriate for the data. If the CV > 1, the normal distribution may not provide an adequate model for the data. If the data come from a lognormal or some other distribution, then the CV will likely be greater than one. A log-transformation of the data set will decrease the CV if the data come from a lognormal distribution.

Example 1.1 Sample Calculation of the Coefficient of Variation

At a *FACILITY*, 20 soil samples were collected for analysis of nickel concentrations. The results are shown below in parts per billion (ppb).

	A	B	C	D
1	Observation	Nickel concentration (ppb)	Ln (Nickel conc.)	
2	1	58.8	4.074	=LN(B2)
3	2	1.0	0.000	=LN(B3)
4	3	262.0	5.568	=LN(B4)
5	4	56.0	4.025	⋮
6	5	8.7	2.163	
7	6	19.0	2.944	
8	7	81.5	4.401	
9	8	331.0	5.802	
10	9	14.0	2.639	
11	10	64.4	4.165	
12	11	39.0	3.664	
13	12	151.0	5.017	
14	13	27.0	3.296	
15	14	21.4	3.063	
16	15	578.0	6.360	
17	16	3.1	1.131	
18	17	942.0	6.848	
19	18	85.6	4.450	
20	19	10.0	2.303	
21	20	637.0	6.457	
22				
23	n:	20	20	
24	Mean:	169.5	3.9	
25	Std. Dev:	259.72	1.80	
26	CV:	1.53	0.46	
27	Skewness:	2.00	-0.27	
28				
29				
30		=count(B2:B21)	=count(C2:C21)	
31		=average(B2:B21)	=average(C2:C21)	
32		=stdev(B2:B21)	=stdev(C2:C21)	
33		=B25/B24	=C25/C24	
34		=skew(B2:B21)	=skew(C2:c21)	

Coefficients of variation were calculated for both the raw data and the log-transformed data. First, the averages and standard deviations were calculated using the Microsoft Excel functions shown above. The CV was then calculated by dividing the standard deviations by their respective mean concentrations, which may be denoted CV_{Raw} and CV_{Ln}, respectively.

Since the CV for the raw nickel concentrations is greater than 1 (CV_{Raw}=1.53), this indicates that the data set may not be normally distributed. The CV for the log-transformed data set is less than 1 (CV_{Ln}=0.46), suggesting that it is more appropriate to assume that the data come from a lognormally distributed population. A final decision regarding the distribution of the data set will be made after considering the results of other methods presented throughout this section.

1.3.2 Coefficient of Skewness

The coefficient of skewness (Skew; also denoted γ_1) measures the skewness or asymmetry of a data set with respect to the mean. It is relatively easy to compute and it can be useful in evaluating symmetry.

The following coefficients of skewness may be expected for the various types of distributions:

Distribution Type	Coefficient of Skewness
Symmetric (e.g., normal distribution)	Zero (i.e., Skew = 0)
Right-skewed (with a long right tail)	Positive (i.e., Skew > 0)
Left-skewed (with a long left tail)	Negative (i.e., Skew < 0)

Right-skewed distributions are common in environmental data analysis (i.e., Skew > 0).

Procedure 1.2 (following page) describes the method for calculating the coefficient of skewness and recommendations for interpretation.

Example 1.2 Sample Calculation of the Coefficient of Skewness

Nickel concentrations from 20 soil samples collected at a *FACILITY* were presented in Example 1.1.

The coefficient of skewness was calculated for both the raw and log-transformed data using the Excel function shown in Example 1.1. The results obtained were as follows:

$$Skew_{Raw} = 2.00$$

$$Skew_{Ln} = -0.27$$

$Skew_{Raw}$ is a positive number, indicating that the data set is positively or right skewed, as is common with environmental data sets. Because $Skew_{Raw}$ is greater than 1, the data may not be normally distributed.

Since $-1 < Skew_{Ln} < 1$, the data may be lognormally distributed.

Procedure 1.2 Calculation of the Coefficient of Skewness

To evaluate for normality, the coefficient of skewness may be calculated using the following equation. To evaluate for lognormality, first define y = ln(x) and use these y (log-transformed) values in place of the x values in the equation below. For nondetects, substitute 1/2 of the detection limit up to 50% nondetects.

$$Skew = \frac{\frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3} \qquad \text{Equation 1.4}$$

Where n represents the sample size, x_i represents the i th observation, $\bar{x}$ represents the arithmetic mean (Equation 1.1), and s represents the sample standard deviation (Equation 1.2).

Most statistical software packages calculate the coefficient of skewness. Alternatively, Microsoft Excel can be used to obtain the coefficient of skewness, as follows:

=SKEW(data range)

Interpretation: Small degrees of skewness may not affect the results of an analysis conducted using an assumption of normality; however, a large coefficient of skewness may lead to inaccurate results. In general, if -1 < Skew < 1, the normal distribution may provide a reasonable approximation of the data. If Skew > 1 or Skew < -1, the data may not be normally distributed. When calculating the coefficient of skewness for both the raw data set and the log-transformed (Ln) data set, the results (which may be denoted $Skew_{Raw}$ and $Skew_{Ln}$, repsectively) may be interpreted as follows:

- If $Skew_{Raw} > 1$ and $Skew_{Ln} > 1$, the data set is highly skewed and may not represent either a normal or lognormal distribution.
- If $Skew_{Raw} > 1$ but $-1 < Skew_{Ln} < 1$, the raw data are right-skewed, but the log-transformed data are approximately symmetric. This is evidence for the assumption of lognormality for the raw data set.
- If $-1 < Skew_{Raw} < 1$ and $-1 < Skew_{Ln} < 1$, but $|Skew_{Raw}| > |Skew_{Ln}|$, the log-transformed data set is more symmetric. Consequently, an assumption of lognormality may be more appropriate for the data. (Note: $|Skew_{Raw}|$ denotes the absolute value of the skewness of the raw, non-transformed data.)

It is also possible to obtain a Skew < -1. This may be due to the presence of one or more low outliers or simply sampling variability. If a data set is negatively skewed, a log-transformation of the data will not improve the symmetry. If Skew < -1 and the raw data fail a formal test of normality, a conclusion regarding the distribution cannot be drawn. It may be necessary to consult a professional statistician and/or *PROPOSE* an alternate statistical method to compare *FACILITY* data (see Chapter 3).

1.3.3 Normal Probability Plots

One of the best tools for evaluating a data set for normality (or lognormality) is the normal probability plot. This tool allows for visual inspection of the data as well as an assessment of the fit to a specific probability distribution. Irregularities in the data when compared to a known probability distribution are easy to identify.

Normal probability plots are typically constructed by plotting concentration measurements sorted in increasing order along the y-axis versus corresponding quantiles or "z-scores" of a standard normal distribution (i.e., a normal distribution with a mean of zero and a standard deviation of one, also denoted N(0,1)) on the x-axis. If the data are normal, the plotted data points will approximate a straight line. If the data are not normal, departures from normality may be evident as bends or curves in the plotted points. In addition, unusual values, such as outliers, are often identifiable on the resulting plot.

Normal probability plots may also be used to evaluate a data set for lognormality by constructing a probability plot using a natural log transformation of the analytical results instead of the raw (untransformed) analytical results. **It is recommended that probability plots be constructed for both raw (untransformed) and log-transformed (Ln) data and compared to identify the best-fitting distribution.**

Because interpretation of probability plots may be somewhat subjective, conclusions should not be drawn regarding statistical distributions based on probability plots alone. Probability plots should be used in combination with the other methods described in this chapter.

Many statistical packages can create normal probability plots at the click of a button. (The axes may be reversed or may differ slightly from those described below; however, they may be interpreted in the same manner.) Different statistical software packages may refer to these plots by alternate names, such as "normal probability plots," "normal quantile-quantile plots," or "QQ normal plots." For those without access to a statistical package, spreadsheets provide an alternative means to construct probability plots. The instructions are straightforward and are provided in Procedure 1.3.

Procedure 1.3 Construction of a Normal Probability Plot

To evaluate for normality, complete the steps below using the raw data. To evaluate for lognormality, first define y = ln(x) and use these y (log-transformed) values in place of the x values below. For nondetects, substitute 1/2 of the detection limit. Probability plots may be constructed using data with more than 50% nondetects; however, as the percent nondetects increases, the amount of information provided by the plot decreases.

1. Order the data from the smallest to the largest value ($x_{(i)}$, i = 1, …, n).

2. Calculate the cumulative probabilities corresponding to each $x_{(i)}$ (representing the proportion of values less than or equal to $x_{(i)}$) as follows:

3.
$$p_i = \frac{i}{n+1} \qquad \text{Equation 1.5}$$

Where n represents the sample size and i represents the rank of the i th ordered concentration.

4. Determine the quantiles or z-scores from the standard normal distribution corresponding to the cumulative probabilities in Step 2.

5.
$$z_i = \Phi^{-1}(p_i) \qquad \text{Equation 1.6}$$

where Φ^{-1} denotes the inverse of the cumulative standard normal distribution. These values are easily calculated using Excel (see below).

6. Plot z_i (z-scores) versus $x_{(i)}$ (the ordered concentrations for each sample).

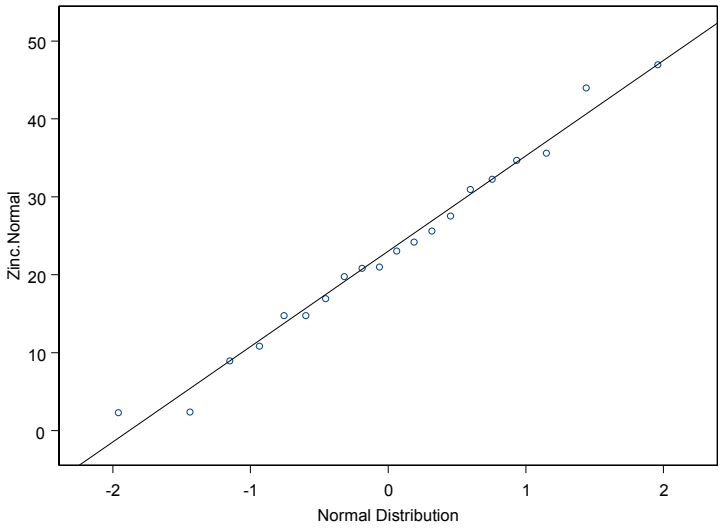
Microsoft Excel can be used to plot the above results as well as obtain z-scores (Equation 1.6) for each probability p_i using the following function:

=NORMSINV(p_i)

Interpretation: If a data set is approximately normal, the plotted points should fall on or near a straight line. Curves or bends in the line indicate that the data are not normally distributed. Lognormality of the data set can be evaluated by plotting the log-transformed y-values in place of the x-values above. If the data set is approximately lognormal and the probability plot is constructed using log-transformed data, the plotted points should fall on or near a straight line.

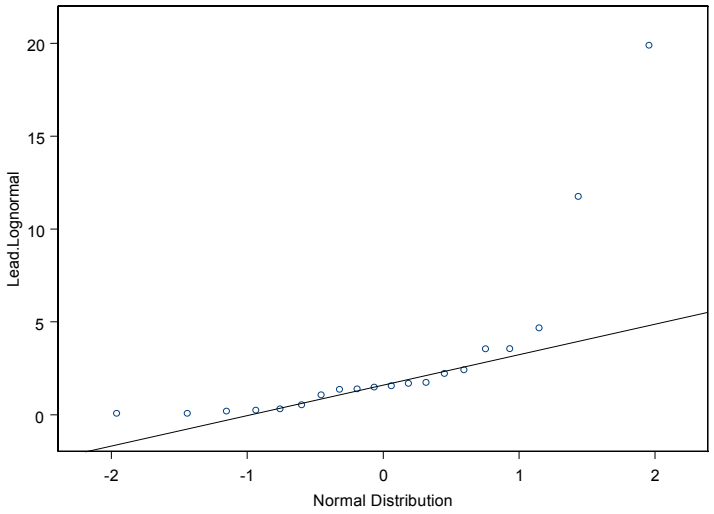
Sample plots are shown on Figures 1.4 through 1.8 to illustrate typical patterns that may be observed on probability plots. Interpretations are provided to the right of each plot. Figures 1.4 through 1.8 were created using the statistical software package S-PLUS. Figures 1.9 and 1.10 (Example 1.3) were constructed using Microsoft Excel.

Figure 1.4 Normal Probability Plot of Zinc Data



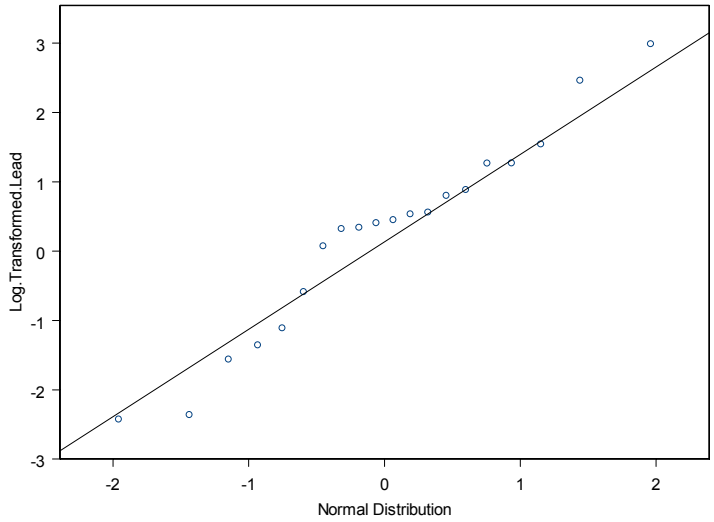
Using raw (untransformed) data, the points fall approximately on a straight line. Therefore, the data set appears to be approximately normally distributed.

Figure 1.5 Normal Probability Plot of Lead Data



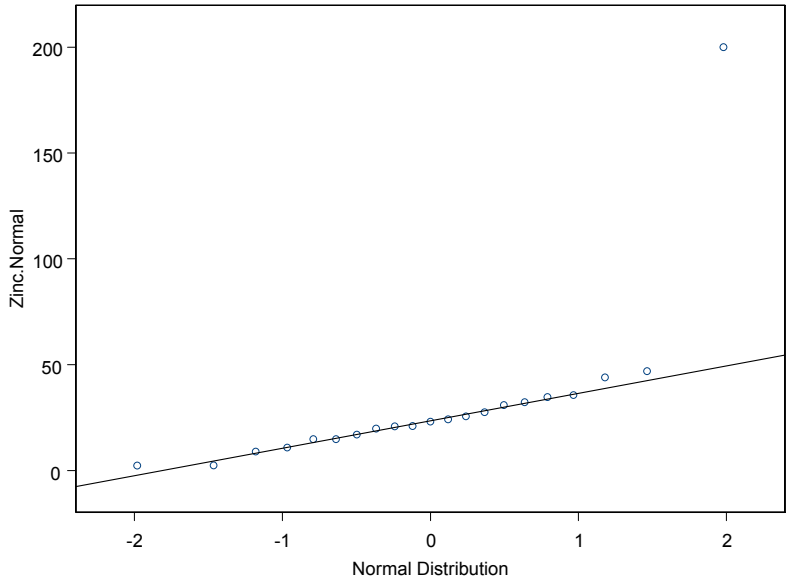
Using raw lead data, the points curve about the straight line. This may indicate that the data are approximately lognormal instead. Compare to the plot constructed using log-transformed data below.

Figure 1.6 Normal Probability Plot of Log-Transformed (Ln) Lead Data



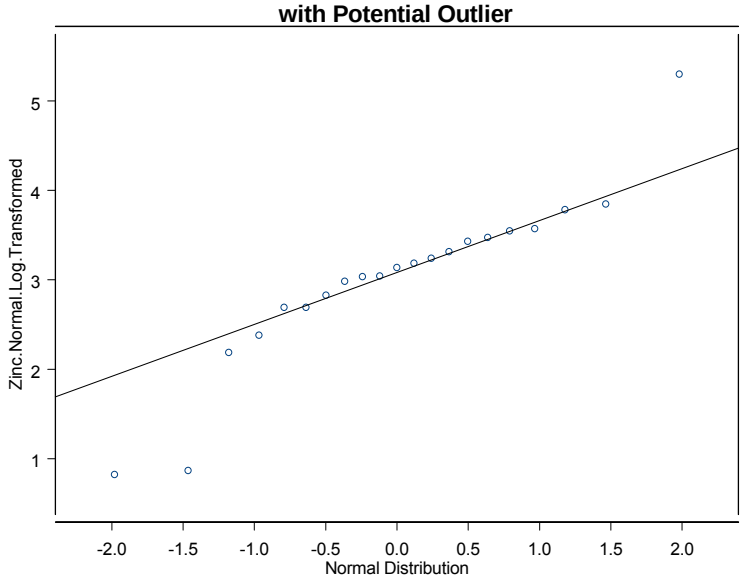
Using log-transformed lead data, the points fall approximately on a straight line. Because the log-transformed data are approximately normal, it can be concluded that the raw (untransformed) data set is approximately lognormal.

Figure 1.7 Normal Probability Plot of Zinc Data with Potential Outlier



Although one might initially guess this data set is lognormally distributed, this plot does not illustrate the characteristic curve of a data set that is lognormally distributed. Most of the data fall on or near a straight line; however, one outlying value is evident. This indicates that the data may be normally distributed with the exception of a single outlier.

Figure 1.8 Normal Probability Plot of Log-Transformed Zinc Data with Potential Outlier



The data shown in Figure 1.7 were log-transformed to illustrate that log-transformations do not always solve problems with high outliers. The high value still stands out as a potential outlier in this plot. Furthermore, note that two low values now appear to be outlying results as well.

The data set shown in Figures 1.7 and 1.8 are evaluated further in Chapter 2 on identification of outliers.

Example 1.3 Sample Construction of a Probability Plot

The nickel data set from Example 1.1 was used to create probability plots for both the raw nickel concentrations and the log-transformed concentrations. Microsoft Excel was used to complete the calculations described in Procedure 1.3 (calculations shown below) and to plot the results. Column D was plotted versus Column C to create the probability plot for the raw (untransformed) data set (Figure 1.9). Column E was plotted versus Column C to create the normal probability plot for log-transformed (Ln) data (Figure 1.10).

	A	B	C	D	E
1			Normal		
2	Order (i)	i / (n+1)	quantile	Nickel conc. (ppb)	Ln (Nickel conc.)
3	1	0.05	-1.67	1.0	0.000
4	2	0.10	-1.31	3.1	1.131
5	3	0.14	-1.07	8.7	2.163
6	4	0.19	-0.88	10.0	2.303
7	5	0.24	-0.71	14.0	2.639
8	6	0.29	-0.57	19.0	2.944
9	7	0.33	-0.43	21.4	3.063
10	8	0.38	-0.30	27.0	3.296
11	9	0.43	-0.18	39.0	3.664
12	10	0.48	-0.06	56.0	4.025
13	11	0.52	0.06	58.8	4.074
14	12	0.57	0.18	64.4	4.165
15	13	0.62	0.30	81.5	4.401
16	14	0.67	0.43	85.6	4.450
17	15	0.71	0.57	151.0	5.017
18	16	0.76	0.71	262.0	5.568
19	17	0.81	0.88	331.0	5.802
20	18	0.86	1.07	578.0	6.360
21	19	0.90	1.31	637.0	6.457
22	20	0.95	1.67	942.0	6.848
23					
24					
25		=A3/21	=NORMSINV(B3)		=LN(D3)
26		=A4/21	=NORMSINV(B4)		=LN(D4)
27		⋮	⋮		⋮

Figure 1.9 Normal Probability Plot of Raw Nickel Data (Plotted Using Microsoft Excel)

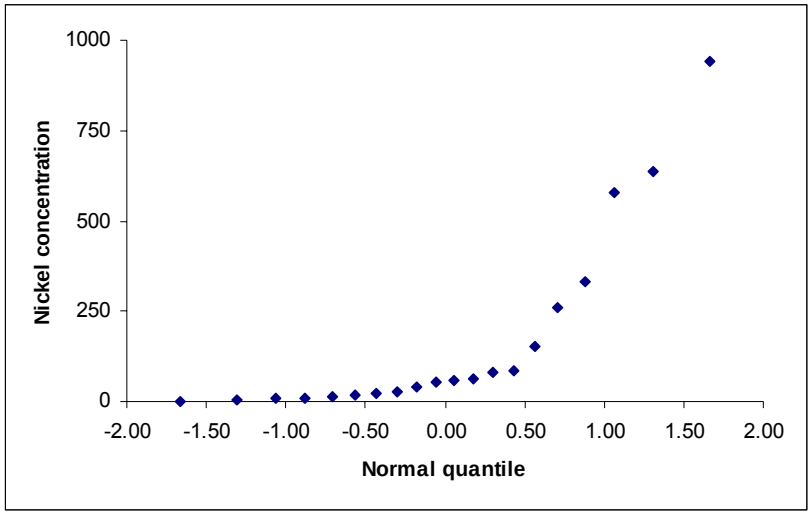
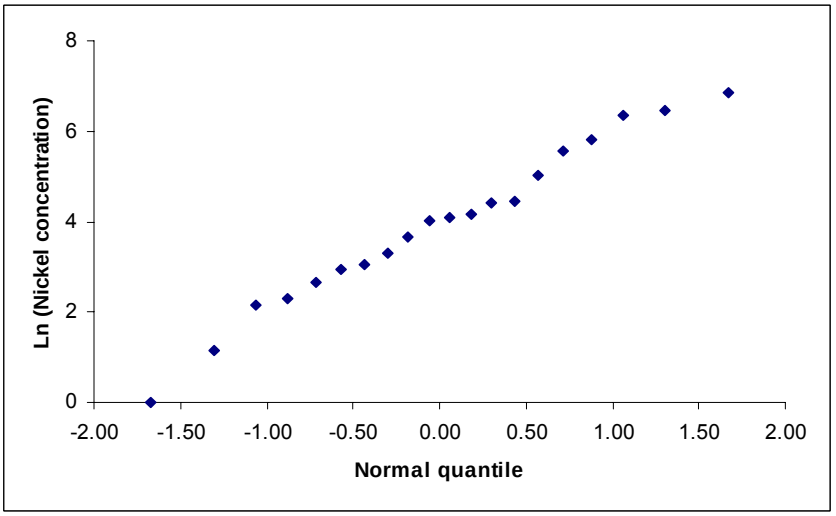


Figure 1.10 Normal Probability Plot of Log-Transformed Nickel Concentrations (Plotted Using Microsoft Excel)



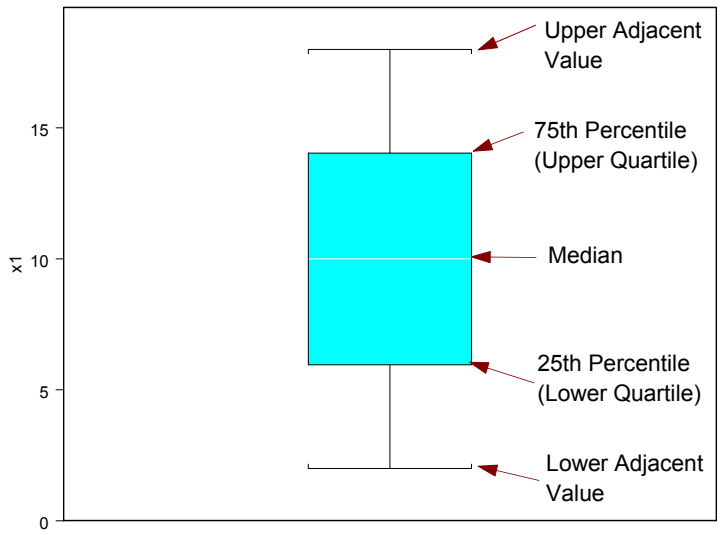
Review of the probability plots indicates that the lognormal distribution provides a better fit to the nickel concentrations than the normal distribution. The plot of the raw data in Figure 1.9 illustrates the distinct curve often seen when the data are better approximated by the lognormal distribution. However, the plot of the log-transformed data closely approximates a straight line. This indicates that the log-transformed nickel concentrations are approximately normal, thus the raw data are approximately lognormal. These findings are consistent with the conclusions drawn in Examples 1.1 and 1.2.

1.3.4 Box Plots

Box plots are another way to graphically examine the characteristics of a data set (see Figure 1.11). The 25th and 75th percentiles of the data set define the lower and upper ends of the box, respectively. A line across the center of the box denotes the median (50th percentile, or middle value). The length of the box, or distance from the 25th to the 75th percentile, is called the interquartile range (IQR).

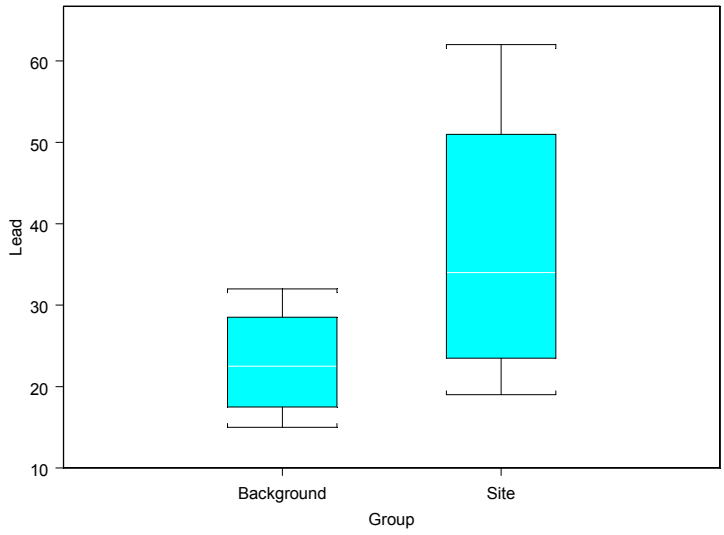
Vertical lines extending from the ends of the box (whiskers) may be constructed in various ways. Two common methods for drawing whiskers are described here. The simplest method involves drawing vertical lines from the upper edge of the box to the maximum concentration and from the lower edge of the box to the minimum concentration.

Figure 1.11 Box and Whisker Plot



A second method involves drawing vertical lines from the upper edge of the box to the next larger value within one step (i.e., the upper adjacent value) and from the 25th percentile down to the next smaller value within one step (i.e., the lower adjacent value). A step is traditionally defined as 1.5 times the IQR. This step size is useful in that the outside values shown as asterisks or horizontal lines beyond the whiskers are potential outliers, though formal verification as described in Chapter 2 is required. This screening criteria can be made more or less strict by changing the step size.

Figure 1.12 Side-by-Side Box Plots



Box plots are particularly useful when comparing two or more data sets. Side-by-side box plots can be inspected for differences between the data sets (Figure 1.12).

Since interpretation of box plots can be somewhat subjective, box plots should only be used in combination with other methods for determining the distribution of a data set, not as a sole determining factor.

Although Microsoft Excel does not currently have capabilities to produce box plots, most statistical software packages (e.g., S-Plus, Minitab, Statistica, Systat, SPSS, and SAS) can easily produce box plots with the click of a button or a few command lines of code.

Interpretation: Box plots provide another useful tool for examining the statistical distributions of data sets. The horizontal line dividing the box graphically shows the median or middle value of a data set. The length of the box provides information on the variability or spread in the data set (similar to the CV). Skewness (i.e., symmetry versus asymmetry) can be inferred by examining the relative lengths of the box halves. For example, if the portion of the box and/or whiskers above the median are longer than the portion of the box and/or whiskers below the median, the data set is right-skewed (i.e., asymmetric, with a long right tail).

Box plots also provide an easy way to identify any unusual values (e.g., potential outliers) in the data set since outside values are often shown as asterisks or horizontal lines beyond the whiskers.

1.3.5 Shapiro-Wilk Test for Normality ($n \leq 50$)

The Shapiro-Wilk (SW) Test for normality (Shapiro and Wilk, 1965) is recommended as a formal test of normality when the data set contains 50 or fewer results. As noted by the EPA (1992c), the SW Test is considered to be one of the very best tests of normality available (Miller, 1986; Madansky, 1988). Essentially, the SW Test for normality provides a measure of the degree to which a probability plot approximates a straight line. When the probability plot shows a nearly straight line, the SW Test statistic will be large. Conversely, when the probability plot shows substantial bends or curves, the SW Test statistic will be small.

As previously discussed, various statistical guidance documents provide recommendations to consider default assumptions of normality or lognormality (i.e., test first for normality or lognormality; if the data pass a test for the default distribution, no further testing is recommended). **However, it is recommended here that the SW Test be completed both on the raw data and on the log-transformed (Ln) data. The results of these two tests should be compared to determine which distribution provides a better fit to the data.**

Most statistical software packages include tests of statistical distributions; although not all include the SW Test. Of those that include this test, not all include the same version of the test. Consequently, care should be exercised. EnvironmentalStats for S-PLUS provides this test along with several supplemental graphics that are available with a few clicks of a button. For those without access to a statistical package, spreadsheets provide an alternative means for completing the SW Test. Instructions for Microsoft Excel are provided in Procedure 1.4.

Procedure 1.4 Shapiro-Wilk Test for Normality (n ≤ 50)

To test for normality, follow the steps below. To test for lognormality, first define y = ln(x) and use these y (transformed) values in place of the x values below. For nondetects, substitute 1/2 of the detection limit up to 50% nondetects.

1. Order the data from the smallest to the largest value, where $x_{(1)}$ is the smallest value (also called the first order statistic), $x_{(2)}$ is the next largest value, …, and $x_{(n)}$ is the largest value.

2. Set *k* equal to the largest integer less than or equal to (*n*/2), where n represents the sample size.

3. Calculate the differences $[x_{(n-i+1)} - x_{(i)}]$ for each $i = 1, 2, \ldots, k$.

4. Find the coefficients a_i for i = 1, 2, …, *k* using Table 1.1. Coefficients are provided in this table for samples sizes from *n* = 3 to *n* = 50.

5. Calculate *b* as follows:

$$b = \sum_{i=1}^{k} b_i = \sum_{i=1}^{k} a_{n-i+1}(x_{(n-i+1)} - x_{(i)}) \qquad \text{Equation 1.7}$$

6. Calculate the sample standard deviation (*s*) of the data set (Equation 1.2).

Microsoft Excel can be used to obtain an estimate of the sample standard deviation (*s*) using the following function:

=STDEV(data range)

7. Calculate the Shapiro-Wilk test statistic as follows:

$$SW = \left[\frac{b}{s\sqrt{n-1}}\right]^2 \qquad \text{Equation 1.8}$$

8. Using Table 1.2 and a significance level of $\alpha = 0.05$, determine the critical value (sw_α) based on *n* observations and compare to *SW*.

Interpretation: The value SW will tend to be large when a probability plot of the data approximates a straight line. SW will tend to be small when a probability plot shows substantial departures from a straight line (e.g., bends, curves, or outliers). Specifically:

- If $SW \geq sw_\alpha$ conclude that the data set is approximately normal (lognormal, if calculated using y values).
- If $SW < sw_\alpha$ conclude that the data set is **not** normally (lognormally) distributed.

***p*-Values:** Some statistical software packages provide *p*-values instead of the *SW* and sw_α values described above. The *p*-value may be interpreted as the probability that the given data set and the corresponding *SW* value would be obtained if the sampled population (i.e., the population that was sampled to obtain the data set) were truly normally distributed (or lognormally distributed, if testing for lognormality). For example, if a raw or untransformed data set was tested for normality and the resulting *p*-value was small, it is unlikely that the data set was obtained from a normal distribution. When the significance level is set at $\alpha = 0.05$ (the recommended level of significance for most situations), the p-value should be compared directly to this value. That is:

If $p \geq 0.05$, conclude that the data set is approximately normal (lognormal, if obtained using y values).

If $p < 0.05$, conclude that the data set is **not** normally (lognormally) distributed.

As previously noted, the SW Test should be completed for both the raw data set and the log-transformed (Ln) data set to test for normality and lognormality, respectively. If the results indicate that the data pass both tests, the test resulting in the higher *SW* value (or *p*-value) should be relied upon to draw conclusions about the distribution of the data.

A review of plots and summary statistics should also be completed to confirm the results of the SW Test.

In some cases, data sets fail both tests for normality and lognormality. When the SW Test results in a conclusion that a data set is not normal or lognormal, it does not indicate which characteristic(s) of the data, say skewness to a heavy-tailed distribution (or both) was responsible for the lack of normality or lognormality. Inspection of the histogram, box plot, and particularly the probability plot, may provide some insight as to why this occurred. It may be clear that the data set is highly skewed, even when log-transformed. If there is a potential outlier in the data set, see Chapter 2 of this section for specific recommendations. The presence of a true outlier(s) may be an indication of an unsuspected *HOT SPOT*. *HOT SPOTS* generally represent a separate population and must be addressed separately. If there is no clear reason for the data to fail tests of normality and lognormality, it may be necessary to proceed to an alternate statistical method for comparing *FACILITY* data to Part 201 criteria or *BACKGROUND*, as described in Chapters 3 and 4. Consultation with a professional statistician is advised.

Example 1.4 Sample Shapiro-Wilk Test for Normality

Based on the evaluations presented in Examples 1.1, 1.2, and 1.3, the nickel concentrations in soil appear to fit a lognormal distribution better than a normal distribution pattern. To formally test this, the SW Test was conducted on both the raw data and the log-transformed data.

Calculations for the Shapiro-Wilk Test of Normality (Raw Data)

	A	B	C	D	E	F	G
1	i	$x_{(i)}$	$x_{(n-i+1)}$	$x_{(n-i+1)} - x_{(i)}$	$a_{(n-i+1)}$	$b_{(i)}$	
2	1	1.0	942.0	941.0	0.4734	445.4694	← =D2*E2
3	2	3.1	637.0	633.9	0.3211	203.5453	← =D3*E3
4	3	8.7	578.0	569.3	0.2565	146.0255	← =D4*E4
5	4	10.0	331.0	321.0	0.2085	66.9285	⋮
6	5	14.0	262.0	248.0	0.1686	41.8128	
7	6	19.0	151.0	132.0	0.1334	17.6088	
8	7	21.4	85.6	64.2	0.1013	6.5035	
9	8	27.0	81.5	54.5	0.0711	3.8750	
10	9	39.0	64.4	25.4	0.0422	1.0719	
11	10	56.0	58.8	2.8	0.0140	0.0392	
12	11	58.8	56.0		b:	932.880	
13	12	64.4	39.0			=SUM(F2:F11)	
14	13	81.5	27.0				
15	14	85.6	21.4				
16	15	151.0	19.0				
17	16	262.0	14.0				
18	17	331.0	10.0				
19	18	578.0	8.7				
20	19	637.0	3.1				
21	20	942.0	1.0				
22							
23	Std. Dev:	259.7	← =STDEV(B2:B21)				
24	n:	20	← =COUNT(B2:B21)				
25	k:	10					

$$W = \left[\frac{932.9}{259.7\sqrt{19}}\right]^2 = 0.679$$

Calculations for the Shapiro-Wilk Test of Normality (Log-Transformed Data)

	A	B	C	D	E	F	G
1	i	$x_{(i)}$	$x_{(n-i+1)}$	$x_{(n-i+1)} - x_{(i)}$	$a_{(n-i+1)}$	$b_{(i)}$	
2	1	0.0	6.8	6.8	0.4734	3.2418	← =D2*E2
3	2	1.1	6.5	5.3	0.3211	1.7100	← =D3*E3
4	3	2.2	6.4	4.2	0.2565	1.0763	← =D4*E4
5	4	2.3	5.8	3.5	0.2085	0.7297	⋮
6	5	2.6	5.6	2.9	0.1686	0.4939	
7	6	2.9	5.0	2.1	0.1334	0.2765	
8	7	3.1	4.4	1.4	0.1013	0.1404	
9	8	3.3	4.4	1.1	0.0711	0.0785	
10	9	3.7	4.2	0.5	0.0422	0.0212	
11	10	4.0	4.1	0.0	0.0140	0.0007	
12	11	4.1	4.0		b:	7.769	
13	12	4.2	3.7			=SUM(F2:F11)	
14	13	4.4	3.3				
15	14	4.4	3.1				
16	15	5.0	2.9				
17	16	5.6	2.6				
18	17	5.8	2.3				
19	18	6.4	2.2				
20	19	6.5	1.1				
21	20	6.8	0.0				
22							
23	Std. Dev:	1.8	← =STDEV(B2:B21)				
24	n:	20	← =COUNT(B2:B21)				
25	k:	10					

$$W = \left[\frac{7.769}{1.8\sqrt{19}}\right]^2 = 0.980$$

From Table 1.2, it can be seen that the critical value sw_α, given a sample size of 20, is 0.905 at the α = 0.05 level. The SW statistic calculated from the raw nickel data (SW_{Raw}) is equal to 0.679. The SW statistic calculated using the log-transformed data (SW_{Ln}) is equal to 0.980. Because $SW_{Raw} < sw_\alpha$ (0.679 < 0.905), it can be concluded that the data are not normally distributed. However, because $SW_{Ln} > sw_\alpha$ (0.980 > 0.905), it can be concluded that the data are approximately lognormally distributed.

The results of the SW Test confirm our initial conclusions based on use of summary statistics and plots. Consequently, we can be more confident in our conclusion that the lognormal model provides a better approximation to the data. Inference made under the assumption of lognormality is expected to be more reliable than under the assumption of normality.

1.3.6 Shapiro-Francia Test for Normality (*n* > 50)

When a data set is large (i.e., data sets with more than 50 observations), the Shapiro-Francia (SF) Test for normality is appropriate (Shapiro and Francia, 1972). Similar to the SW Test, the SF Test statistic will be a large when a probability plot of the data shows a nearly straight line.

EnvironmentalStats for S-PLUS provides the SF Test for normality (or lognormality) along with several supplemental graphics that are available with a few clicks of a button. For those without access to a statistical package, spreadsheets provide an alternative means for completing the SF Test. Instructions for Microsoft Excel are provided in Procedure 1.5.

The SF Test should be completed for both the raw data set and the log-transformed (Ln) data set to test for normality and lognormality, respectively. If the results indicate that the data pass both tests, the test resulting in the higher SF value should be relied upon to draw conclusions about the distribution of the data.

A review of plots and summary statistics should also be completed to confirm the results of the SF Test.

In some cases, data sets fail both tests for normality and lognormality. When the SF Test results in a conclusion that a data set is not normal or lognormal, it does not indicate which characteristic(s) of the data, say skewness to a heavy-tailed distribution (or both) was responsible for the lack of normality or lognormality. Inspection of the histogram, box plot, and particularly the probability plot, may provide some insight as to why this occurred. It may be clear that the data set is highly skewed, even when log-transformed. If there is a potential outlier in the data set, see Chapter 2 of this section for specific recommendations. The presence of a true outlier(s) may be an indication of an unsuspected *HOT SPOT*. *HOT SPOTS* generally represent a separate population and must be addressed separately. If there is no clear reason for the data to fail tests of normality and lognormality, it may be necessary to proceed to an alternate statistical method for comparing *FACILITY* data to Part 201 criteria or *BACKGROUND*, as described in Chapters 3 and 4. Consultation with a professional statistician is advised.

Procedure 1.5 Shapiro-Francia Test for Normality (n > 50)

To test for normality, follow the steps below. To test for lognormality, first define y = ln(x) and use these y (transformed) values in place of the x values below. Substitute 1/2 of the detection limit for nondetects for up to 50% nondetects.

1. Order the data from the smallest to the largest value, where $x_{(1)}$ is the smallest value (also called the first order statistic), $x_{(2)}$ is the next largest value, …, and $x_{(n)}$ is the largest value.

2. Calculate the sample standard deviation (s^2) using Equation 1.2 and squaring the result.

Microsoft Excel can be used to obtain an estimate of the sample variance (s^2) using the following function:

=VAR(data range)

3. Calculate m_i, or i th ordered normal quantiles (z-scores) using:

$$m_i = \Phi^{-1}\left(\frac{i}{n+1}\right) \qquad \text{Equation 1.9}$$

Where Φ^{-1} denotes the inverse of the cumulative standard normal distribution.

Microsoft Excel can be used to obtain z-scores (m_i) for each value using the following function:

m_i =NORMSINV$\left(\frac{i}{n+1}\right)$

4. The SF Test statistic is calculated using the following formula:

$$SF = \left[\sum_{i=1}^{n} m_i x_{(i)}\right]^2 \Bigg/ \left[(n-1)s^2 \sum_{i=1}^{n} m_i^{\,2}\right] \qquad \text{Equation 1.10}$$

5. Using Table 1.3 and a significance level of α = 0.05, determine the critical value (sf_α) based on n observations and compare to SF.

Interpretation: Similar to the SW statistic, the value of the SF statistic will tend to be large when a probability plot of the data approximates a straight line. SF will tend to be small when a probability plot shows substantial departures from a straight line (e.g., bends, curves, or outliers). Specifically:

- If $SF \geq sf_\alpha$ conclude that the data set is approximately normal (lognormal, if calculated using y values).
- If $SF < sf_\alpha$ conclude that the data set is **not** normally (lognormally) distributed.

***p*-Values:** Some statistical software packages provide *p*-values instead of the *SF* and sf_α values described above. The *p*-value may be interpreted as the probability that the given data set and the corresponding *SF* value would be obtained if the sampled population (i.e., the population that was sampled to obtain the data set) were truly normally distributed (or lognormally distributed, if testing for lognormality). For example, if a raw or untransformed data set was tested for normality and the resulting *p*-value was small, it is unlikely that the data set was obtained from a normal distribution. When the significance level is set at $\alpha = 0.05$ (the recommended level of significance for most situations), the p-value should be compared directly to this value. That is:

If $p \geq 0.05$, conclude that the data set is approximately normal (lognormal, if obtained using y values).

If $p < 0.05$, conclude that the data set is **not** normally (lognormally) distributed.

As previously noted, the SF Test should be completed for both the raw data set and the log-transformed (Ln) data set to test for normality and lognormality, respectively. If the results indicate that the data pass both tests, the test resulting in the higher *SF* value (or *p*-value) should be relied upon to draw conclusions about the distribution of the data.

Example 1.5 Sample Shapiro-Francia Test for Normality

Fifty-two samples of soil dioxin (in ppb) were collected at a hypothetical *FACILITY*. Probability plots, the coefficient of skewness, and the CV all suggested that the data were lognormally distributed. A SF Test was then conducted to formally test whether the data were lognormally distributed.

The results were SF_{Ln} = 0.9763 and SF_{Raw} = 0.7682 (only the calculations for the log-transformed data set are shown in the figure). The sf_α for n = 52 and α = 0.05 is 0.955. Since $SF_{Ln} > sf_\alpha$, we conclude that the dioxin data set is lognormally distributed.

Calculations for Shapiro-Francia Test on Example Data Set

	A	B	C	D	E	F	G	H	I	J
1	Observation	Ln (Dioxin)	i / (n + 1)	Quantiles (m_i)	$m_i * x_{(i)}$	m_i^2				
2	1	0.07	0.02	-2.08	-0.14	4.32		n:	52	
3	2	0.36	0.04	-1.78	-0.64	3.16			=COUNT(B2:B53)	
4	3	0.41	0.06	-1.58	-0.65	2.51				
5	4	0.61	0.08	-1.44	-0.88	2.06		s^2:	0.83	
6	5	0.69	0.09	-1.31	-0.91	1.73			=VAR(B2:B53)	
7	6	0.73	0.11	-1.21	-0.89	1.46				
8	7	0.79	0.13	-1.12	-0.88	1.25		SF:	0.9763	
9	8	0.86	0.15	-1.03	-0.89	1.07		= E54^2 / (51 * I5 * F54)		
10	9	0.92	0.17	-0.95	-0.88	0.91				
11	10	1.04	0.19	-0.88	-0.92	0.78				
12	11	1.18	0.21	-0.81	-0.97	0.66	$SF = \left[\sum_{i=1}^{n} m_i x_{(i)}\right]^2 \Big/ \left[(n-1)s^2 \sum_{i=1}^{n} m_i^2\right]$			
13	12	1.27	0.23	-0.75	-0.95	0.56				
14	13	1.36	0.25	-0.69	-0.94	0.48				
15	14	1.42	0.26	-0.63	-0.90	0.40				
16	15	1.45	0.28	-0.57	-0.83	0.33				
17	16	1.48	0.30	-0.52	-0.77	0.27				
18	17	1.51	0.32	-0.47	-0.70	0.22				
19	18	1.57	0.34	-0.41	-0.65	0.17				
20	19	1.61	0.36	-0.36	-0.58	0.13				
21	20	1.62	0.38	-0.31	-0.50	0.10				
22	21	1.70	0.40	-0.26	-0.45	0.07				
23	22	1.71	0.42	-0.21	-0.37	0.05				
24	23	1.72	0.43	-0.17	-0.29	0.03				
25	24	1.73	0.45	-0.12	-0.20	0.01				
26	25	1.73	0.47	-0.07	-0.12	0.01				
27	26	1.75	0.49	-0.02	-0.04	0.00				
28	27	1.76	0.51	0.02	0.04	0.00				
29	28	1.77	0.53	0.07	0.13	0.01				
30	29	1.82	0.55	0.12	0.22	0.01				
31	30	1.86	0.57	0.17	0.31	0.03				
32	31	1.94	0.58	0.21	0.42	0.05				
33	32	1.94	0.60	0.26	0.51	0.07				
34	33	1.98	0.62	0.31	0.62	0.10				
35	34	2.02	0.64	0.36	0.73	0.13				
36	35	2.10	0.66	0.41	0.87	0.17				
37	36	2.18	0.68	0.47	1.02	0.22				
38	37	2.40	0.70	0.52	1.25	0.27				
39	38	2.51	0.72	0.57	1.44	0.33				
40	39	2.57	0.74	0.63	1.62	0.40				
41	40	2.65	0.75	0.69	1.83	0.48				
42	41	2.68	0.77	0.75	2.01	0.56				
43	42	2.82	0.79	0.81	2.30	0.66				
44	43	2.90	0.81	0.88	2.56	0.78				
45	44	2.96	0.83	0.95	2.82	0.91				
46	45	3.01	0.85	1.03	3.10	1.07				
47	46	3.09	0.87	1.12	3.45	1.25				
48	47	3.25	0.89	1.21	3.93	1.46				
49	48	3.30	0.91	1.31	4.34	1.73				
50	49	3.44	0.92	1.44	4.94	2.06				
51	50	3.48	0.94	1.58	5.51	2.51				
52	51	3.65	0.96	1.78	6.48	3.16				
53	52	3.78	0.98	2.08	7.86	4.32				
54					43.34	45.44				
55					=SUM(E2:E53)	=SUM(F2:F53)				
56										
57				=NORMSINV(C2)						
58				=NORMSINV(C3)						
59				=NORMSINV(C4)						
60				⋮						
61										

1.3.7 D'Agostino Test for Normality (*n* > 50)

Some EPA statistical guidance documents recommend the D'Agostino Test (D'Agostino, 1971) to test for normality when the sample size is between 50 and 1000. The D'Agostino's test and the Shapiro-Francia test tend to provide similar results and therefore can be considered roughly equivalent. EnvironmentalStats for S-PLUS does not provide a function that performs the D'Agostino Test for normality, but the necessary calculations can easily be performed using a spreadsheet. Both Gibbons (1994) and Gilbert (1987) describe the method and provide the necessary tables. Although the SF Test is preferred, analyses using the D'Agostino Test are acceptable.

1.4 EVALUATING STATISTICAL DISTRIBUTIONS WHEN DATA CONTAIN > 50% NONDETECTS

When the proportion of concentrations below the detection limit is > 50%, evaluation of the probability distribution of the data is not recommended. Therefore, the analyses described in this chapter are generally not needed. However, these methods may still provide some insight, in particular the graphical methods, to the extent that the data contain some detectable concentrations (e.g., a data set with only slightly more than 50% nondetects).

Table 1.1 Coefficients (a_i) for the Shapiro-Wilk Test for Normality for Various Sample Sizes (n). From Shapiro and Wilk, 1965.

i \ n	2	3	4	5	6	7	8	9	10
1	0.7071	0.7071	0.6872	0.6646	0.6431	0.6233	0.6052	0.5888	0.5739
2	-	0.0000	0.1677	0.2413	0.2806	0.3031	0.3164	0.3244	0.3291
3	-	-	-	0.0000	0.0875	0.1401	0.1743	0.1976	0.2141
4	-	-	-	-	-	0.0000	0.0561	0.0947	0.1224
5	-	-	-	-	-	-	-	0.0000	0.0399

i \ n	11	12	13	14	15	16	17	18	19	20
1	0.5601	0.5475	0.5359	0.5251	0.5150	0.5056	0.4968	0.4886	0.4808	0.4734
2	0.3315	0.3325	0.3325	0.3318	0.3306	0.3290	0.3273	0.3253	0.3232	0.3211
3	0.2260	0.2347	0.2412	0.2460	0.2495	0.2521	0.2540	0.2553	0.2561	0.2565
4	0.1429	0.1586	0.1707	0.1802	0.1878	0.1939	0.1988	0.2027	0.2059	0.2085
5	0.0695	0.0922	0.1099	0.1240	0.1353	0.1447	0.1524	0.1587	0.1641	0.1686
6	0.0000	0.0303	0.0539	0.0727	0.0880	0.1005	0.1109	0.1197	0.1271	0.1334
7	-	-	0.0000	0.0240	0.0433	0.0593	0.0725	0.0837	0.0932	0.1013
8	-	-	-	-	0.0000	0.0196	0.0359	0.0496	0.0612	0.0711
9	-	-	-	-	-	-	0.0000	0.0163	0.0303	0.0422
10	-	-	-	-	-	-	-	-	0.0000	0.0140

i \ n	21	22	23	24	25	26	27	28	29	30
1	0.4643	0.4590	0.4542	0.4493	0.4450	0.4407	0.4366	0.4328	0.4291	0.4254
2	0.3185	0.3156	0.3126	0.3098	0.3069	0.3043	0.3018	0.2992	0.2968	0.2944
3	0.2578	0.2571	0.2563	0.2554	0.2543	0.2533	0.2522	0.2510	0.2499	0.2487
4	0.2119	0.2131	0.2139	0.2145	0.2148	0.2151	0.2152	0.2151	0.2150	0.2148
5	0.1736	0.1764	0.1787	0.1807	0.1822	0.1836	0.1848	0.1857	0.1864	0.1870
6	0.1399	0.1443	0.1480	0.1512	0.1539	0.1563	0.1584	0.1601	0.1616	0.1630
7	0.1092	0.1150	0.1201	0.1245	0.1283	0.1316	0.1346	0.1372	0.1395	0.1415
8	0.0804	0.0878	0.0941	0.0997	0.1046	0.1089	0.1128	0.1162	0.1192	0.1219
9	0.0530	0.0618	0.0696	0.0764	0.0823	0.0876	0.0923	0.0965	0.1002	0.1036
10	0.0263	0.0368	0.0459	0.0539	0.0610	0.0672	0.0728	0.0778	0.0822	0.0862
11	0.0000	0.0122	0.0228	0.0321	0.0403	0.0476	0.0540	0.0598	0.6500	0.0697
12	-	-	0.0000	0.0107	0.0200	0.0284	0.0358	0.0424	0.0483	0.0537
13	-	-	-	-	0.0000	0.0094	0.0178	0.0253	0.0320	0.0381
14	-	-	-	-	-	-	0.0000	0.0084	0.0159	0.0227
15	-	-	-	-	-	-	-	-	0.0000	0.0076

Table 1.1 (*continued*) Coefficients (a_i) for the Shapiro-Wilk Test for Normality for Various Sample Sizes (n). From Shapiro and Wilk, 1965.

i \ n	31	32	33	34	35	36	37	38	39	40
1	0.4220	0.4188	0.4156	0.4127	0.4096	0.4068	0.4040	0.4015	0.3989	0.3964
2	0.2921	0.2898	0.2876	0.2854	0.2834	0.2813	0.2794	0.2774	0.2755	0.2737
3	0.2475	0.2462	0.2451	0.2439	0.2427	0.2415	0.2403	0.2391	0.2380	0.2368
4	0.2145	0.2141	0.2137	0.2132	0.2127	0.2121	0.2116	0.2110	0.2104	0.2098
5	0.1874	0.1878	0.1880	0.1882	0.1883	0.1883	0.1883	0.1881	0.1880	0.1878
6	0.1641	0.1651	0.1660	0.1667	0.1673	0.1678	0.1683	0.1686	0.1689	0.1691
7	0.1433	0.1449	0.1463	0.1475	0.1487	0.1496	0.1505	0.1513	0.1520	0.1526
8	0.1243	0.1265	0.1284	0.1301	0.1317	0.1331	0.1344	0.1356	0.1366	0.1376
9	0.1066	0.1093	0.1118	0.1140	0.1160	0.1179	0.1196	0.1211	0.1225	0.1237
10	0.0899	0.0931	0.0961	0.0988	0.1013	0.1036	0.1056	0.1075	0.1092	0.1108
11	0.0739	0.0777	0.0812	0.0844	0.0873	0.0900	0.0924	0.0947	0.0967	0.0986
12	0.0585	0.0629	0.0669	0.0706	0.0739	0.0770	0.0798	0.0824	0.0848	0.0870
13	0.0435	0.0485	0.0530	0.0572	0.0610	0.0645	0.0677	0.0706	0.0733	0.0759
14	0.0289	0.0344	0.0395	0.0441	0.0484	0.0523	0.0559	0.0592	0.0622	0.0651
15	0.0144	0.0206	0.0262	0.0314	0.0361	0.0404	0.0444	0.0481	0.0515	0.0546
16	0.0000	0.0068	0.0131	0.0187	0.0239	0.0287	0.0331	0.0372	0.0409	0.0444
17	-	-	0.0000	0.0062	0.0119	0.0172	0.0220	0.0264	0.0305	0.0343
18	-	-	-	-	0.0000	0.0057	0.0110	0.0158	0.0203	0.0244
19	-	-	-	-	-	-	0.0000	0.0053	0.0101	0.0146
20	-	-	-	-	-	-	-	-	0.0000	0.0049

i \ n	41	42	43	44	45	46	47	48	49	50
1	0.3940	0.3917	0.3894	0.3872	0.3850	0.3830	0.3808	0.3789	0.3770	0.3751
2	0.2719	0.2701	0.2684	0.2667	0.2651	0.2635	0.2620	0.2604	0.2589	0.2574
3	0.2357	0.2345	0.2334	0.2323	0.2313	0.2302	0.2291	0.2281	0.2271	0.2260
4	0.2091	0.2085	0.2078	0.2072	0.2065	0.2058	0.2052	0.2045	0.2038	0.2032
5	0.1876	0.1874	0.1871	0.1868	0.1865	0.1862	0.1859	0.1855	0.1851	0.1847
6	0.1693	0.1694	0.1695	0.1695	0.1695	0.1695	0.1695	0.1693	0.1692	0.1691
7	0.1531	0.1535	0.1539	0.1542	0.1545	0.1548	0.1550	0.1551	0.1553	0.1554
8	0.1384	0.1392	0.1398	0.1405	0.1410	0.1415	0.1420	0.1423	0.1427	0.1430
9	0.1249	0.1259	0.1269	0.1278	0.1286	0.1293	0.1300	0.1306	0.1312	0.1317
10	0.1123	0.1136	0.1149	0.1160	0.1170	0.1180	0.1189	0.1197	0.1205	0.1212
11	0.1004	0.1020	0.1035	0.1049	0.1062	0.1073	0.1085	0.1095	0.1105	0.1113
12	0.0891	0.0909	0.0927	0.0943	0.0959	0.0972	0.0986	0.9980	0.1010	0.1020
13	0.0782	0.0804	0.0824	0.0842	0.0860	0.0876	0.0892	0.0906	0.0919	0.0932
14	0.0677	0.0701	0.0724	0.0745	0.0765	0.0783	0.0801	0.0817	0.0832	0.0846
15	0.0575	0.0602	0.0628	0.0651	0.0673	0.0694	0.0713	0.0731	0.0748	0.0764
16	0.0476	0.0506	0.0534	0.0560	0.0584	0.0607	0.0628	0.0648	0.0667	0.0685
17	0.0379	0.0411	0.0442	0.0471	0.0497	0.0522	0.0546	0.0568	0.0588	0.0608
18	0.0283	0.0318	0.0352	0.0383	0.0412	0.0439	0.0465	0.0489	0.0511	0.0532
19	0.0188	0.0227	0.0263	0.0296	0.0328	0.0357	0.0385	0.0411	0.0436	0.0459
20	0.0094	0.0136	0.0175	0.0211	0.0245	0.0277	0.0307	0.0335	0.0361	0.0386
21	0.0000	0.0045	0.0087	0.0126	0.0163	0.0197	0.0229	0.0259	0.0288	0.0314
22	-	-	0.0000	0.0042	0.0081	0.0118	0.0153	0.0185	0.0215	0.0244
23	-	-	-	-	0.0000	0.0039	0.0076	0.0111	0.0143	0.0174
24	-	-	-	-	-	-	0.0000	0.0037	0.0071	0.0104
25	-	-	-	-	-	-	-	-	0.0000	0.0035

Table 1.2 Critical Values (sw_α, α = 0.05 and α = 0.10) for the Shapiro-Wilk Test for Normality. From Shapiro and Wilk, 1965.

n	$sw_{0.05}$	$sw_{0.10}$
3	0.767	0.789
4	0.748	0.792
5	0.762	0.806
6	0.788	0.826
7	0.803	0.838
8	0.818	0.851
9	0.829	0.859
10	0.842	0.869
11	0.850	0.876
12	0.859	0.883
13	0.866	0.889
14	0.874	0.895
15	0.881	0.901
16	0.887	0.906
17	0.892	0.910
18	0.897	0.914
19	0.901	0.917
20	0.905	0.920
21	0.908	0.923
22	0.911	0.926
23	0.914	0.928
24	0.916	0.930
25	0.918	0.931
26	0.920	0.933
27	0.923	0.935
28	0.924	0.936
29	0.926	0.937
30	0.927	0.939
31	0.929	0.940
32	0.930	0.941
33	0.931	0.942
34	0.933	0.943
35	0.934	0.944
36	0.935	0.945
37	0.936	0.946
38	0.938	0.947
39	0.939	0.948
40	0.940	0.949
41	0.941	0.950
42	0.942	0.951
43	0.943	0.951
44	0.944	0.952
45	0.945	0.953
46	0.945	0.953
47	0.946	0.954
48	0.947	0.954
49	0.947	0.955
50	0.947	0.955

Table 1.3 Critical Values (sf_α, $\alpha = 0.05$) for the Shapiro-Francia Test for Normality. From Shapiro and Francia, 1972.

Sample Size	$sf_{0.05}$
50	0.953
51	0.954
53	0.957
55	0.958
57	0.961
59	0.962
61	0.963
63	0.964
65	0.965
67	0.966
69	0.966
71	0.967
73	0.968
75	0.969
77	0.969
79	0.970
81	0.970
83	0.971
85	0.972
87	0.972
89	0.972
91	0.973
93	0.973
95	0.974
97	0.975
99	0.976

CHAPTER 2: IDENTIFICATION AND TREATMENT OF OUTLIERS

Prior to statistically comparing *FACILITY* data to Part 201 risk-based criteria (Chapter 3) or *BACKGROUND* concentrations (Chapter 4), it is first necessary to evaluate the underlying statistical distribution (Chapter 1) and screen for outliers. In addition, if *FACILITY-SPECIFIC BACKGROUND* data are obtained, these data must also be evaluated as described in Chapter 1 prior to screening for outliers.

The importance of outlier testing is noted by Gilbert (1987), who states that, "statistical tests for outliers are one part of the data validation process wherein data are screened and examined in various ways before being placed in a data bank and used for estimating population parameters or making decisions."

Outliers are typically thought of as values that are extreme with respect to other data values. As stated by Millard and Neerchal (2001), "an outlier can be defined as an observation that is 'far away' from the rest of the observations."

An outlier can occur for various reasons:

- It may be an incorrect value due to errors in sampling, laboratory analysis, data entry or transcription.

- It may be an accurate result that was sampled from a different population than the one previously identified for investigation. For example, a population may be identified as soil concentrations in a 1/4 acre *EXPOSURE UNIT* used to evaluate direct contact exposures. These soil concentrations must representative of a single statistical distribution. If an outlier is identified in a sample collected from this *EXPOSURE UNIT*, it may reflect a different statistical distribution and possibly a *HOT SPOT* that was previously unidentified. Additional characterization may be necessary.

- It may be an accurate but extreme value sampled from the originally identified population.

- It may be an accurate value that appears to be extreme with respect to the remaining values in the data set due to failure to obtain a representative sample. This may occur because an insufficient number of samples was collected to reflect the true variability in the population or possibly because biased sampling was conducted rather than *RANDOM* sampling of the population.

In practice, true outliers can be difficult to identify because rarely do we know the parameter values of the population from which our data were sampled. Generally, the parameters must be estimated from the data, thus introducing uncertainty to outlier analysis.

Statistical methods are available for identifying outliers, as presented in Section 2.1. However**, it is important to note that classification of an observation as an outlier does not automatically imply that the observation should be removed from the data set**. This is supported by Gilbert (1987) and EPA (1992c), among others. Outlier testing simply provides methods for quantitatively identifying observations that need to be investigated further and checked for possible errors. Often, outliers provide important information that should not be casually dismissed. As noted by Millard (1998), the hole in the ozone layer over Antarctica

could have been discovered earlier; however it was delayed as a result of "flagging" low-level outliers in the data (Stolarski et al., 1986).

Recommended treatment of outliers depends on the possible cause of the outlier and the context of the evaluation. Specific recommendations are provided in Section 2.2.

HOT SPOT Identification

Outlier testing may serve as a way to quantitatively assess for the presence of *HOT SPOTS*. *HOT SPOTS,* discussed in Section 2.2.1.1 of the tabbed section titled, "Sampling Strategies," are defined as "two or more adjacent sample locations in reasonably close proximity at which concentrations are sufficiently above criteria and surrounding locations (i.e., spatially correlated concentrations sufficiently above criteria) to indicate that they: 1) represent a different statistical population, or 2) pose a potential risk that should not be masked by a statistical analysis." Professional judgment may generally be used to determine whether the magnitude of concentrations and/or the number and proximity of spatially correlated samples above criteria are sufficient to classify an area as a *HOT SPOT*. Spatially correlated concentrations below criteria may also exist; however, these will not be classified as *HOT SPOTS* for the purpose of statistical analysis.

Because outlier testing provides a quantitative approach for evaluating whether elevated measurements are significantly different from a sampled population, it may be considered as a tool for identifying potential *HOT SPOTS* within an area such as an *EXPOSURE UNIT*. However, outlier testing does not take into consideration the spatial distribution of the data or the proximity of concentrations to criteria. For example, a result that is below criteria may be classified as an outlier simply because it is significantly higher than the remaining concentrations in the data set. This value would not represent a *HOT SPOT* as defined above since it is below criteria. Therefore, outlier testing may be *PROPOSED* as a tool, but it may not be used solely in identification of *HOT SPOTS*.

2.1 STATISTICAL METHODS FOR IDENTIFICATION OF OUTLIERS

Formal testing for outliers should be completed only if the presence of one or more outliers is suspected, as recommended by EPA (1992c). Many of the methods described in Chapter 1 of Statistical Methods serve as screening tools for identification of outliers. For example, outliers may be suspected if either a high coefficient of variation (Section 1.2.1) or a coefficient of skewness (Section 1.2.2) far from zero is observed, particularly if these conditions exist when calculated using both the raw and log-transformed data. Furthermore, potential outliers can be easily identified on graphs such as probability plots (Section 1.2.3) and box plots (Section 1.2.4). Since plots are often of greater utility than summary statistics for identifying potential outliers, graphical techniques are always recommended as outlier screening tools.

Once a potential outlier is identified (i.e., through visual inspection of tabulated data or use of other screening tools), formal testing should be completed before classifying the observation as an outlier. Three formal tests are provided for this purpose. In addition, an informal, iterative approach is described.

Type of Method	Method	Section
Graphical Technique	Probability Plots	2.1.1
Graphical Technique	Box Plots	2.1.2
Formal Test	Grubbs' Test (single outlier)	2.1.3
Formal Test	Dixon's Test (multiple outliers, n $\leq$ 25)	2.1.4
Formal Test	Rosner's Test (multiple outliers, n > 25)	2.1.5
Iterative Approach	Retesting with methods described in Chapter 1	2.1.6

It should be mentioned that each of the formal tests for outliers shown above assumes that the data under consideration are normally distributed. Formal outlier testing should be completed only in conjunction with tests for normality to ensure that this assumption is met. This is important since values that appear to be anomalous on the original scale may no longer appear inconsistent when transformed to the log scale. Consequently, the following recommendations are made:

- If a data set is concluded to be **normally distributed**, formal outlier testing should be completed on the raw (untransformed) data.

- If a data set is concluded to be **lognormally distributed**, formal outlier testing should be completed on the log-transformed (natural log) data.

The presence of outliers will sometimes cause a data set to fail tests for normality and lognormality. Therefore, if a data set is found to be **neither normal nor lognormal** and graphical techniques indicate the presence of a potential outlier, an iterative approach using methods described in Chapter 1 may be taken, as described in Section 2.1.6.

For data sets with 50% or more of the values below the detection limit, it is generally not possible to identify the statistical distribution of the data. Consequently, neither formal testing nor the iterative approach can be applied to the data. Therefore, it is necessary to identify outliers qualitatively using the graphical techniques shown above.

2.1.1 Probability Plots

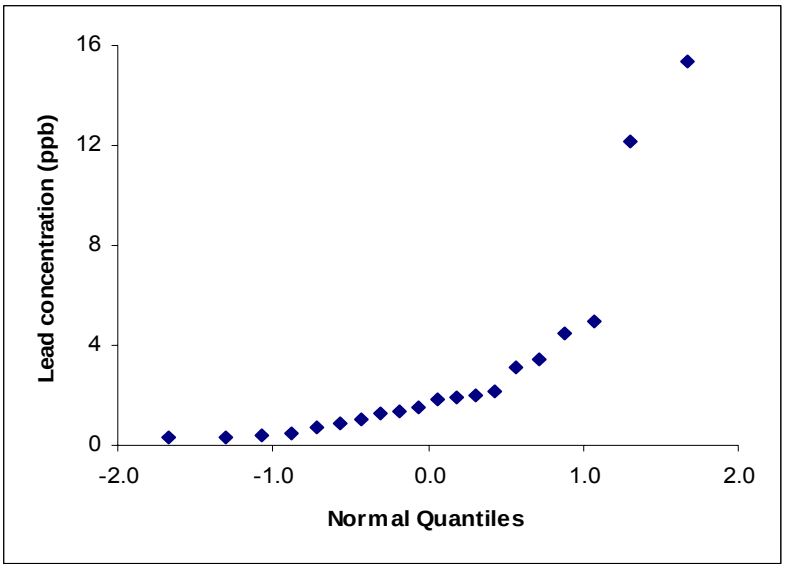
As previously mentioned, one of the first steps in analyzing data is to determine the underlying statistical distribution. Chapter 1 outlines several methods for evaluating and identifying the underlying distribution of a data set. Probability plots provide an excellent tool not only for graphically assessing the distribution of a data set, but also for identifying potential outliers. See Procedure 1.3 for a description of how to construct a probability plot.

Interpretation: On a probability plot, potential outliers will appear as isolated points away from the other points. The other points may form a pattern, such as a line, with potential outliers deviating from this pattern. A probability plot can indicate whether and how many potential outliers there may be, but further testing should be

conducted before classifying the observations as outliers. In addition, some of the formal tests for outliers presented in this chapter (e.g., Rosner's Test) require an initial estimate of the number of outliers in the data set. A probability plot may be used to select this number.

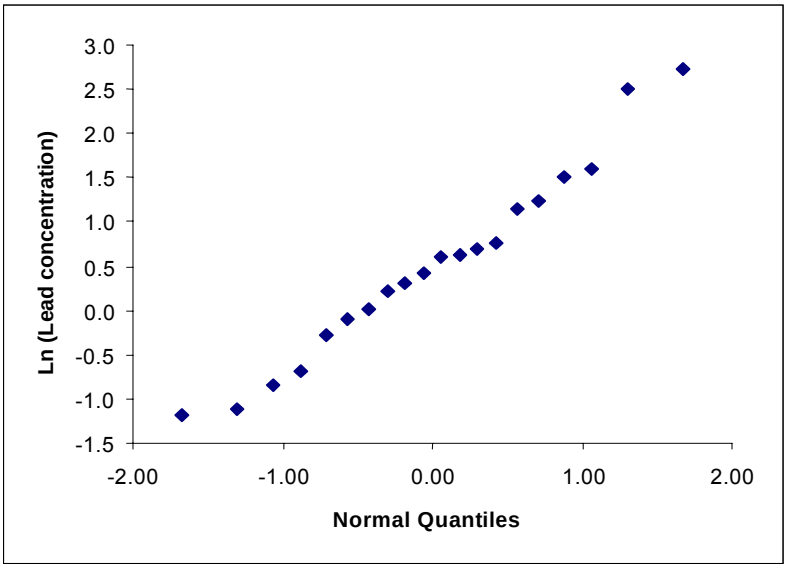
Example 2.1 Sample Probability Plots

Figure 2.1 Probability Plot of Lead Data



This probability plot, constructed using raw (untransformed) lead data, indicates the presence of two potential outliers. A probability plot of log-transformed (Ln) values should be reviewed before identifying these as potential

Figure 2.2 Probability Plot of Log-Transformed (Ln) Lead Data



This probability plot of log-transformed (Ln) lead concentrations does not clearly indicate the presence of potential outliers. Although the two highest values still stand out somewhat, they now fall relatively close to the straight line. Consequently, formal testing for outliers is not clearly necessary. Testing may be completed, however, based on professional judgment.

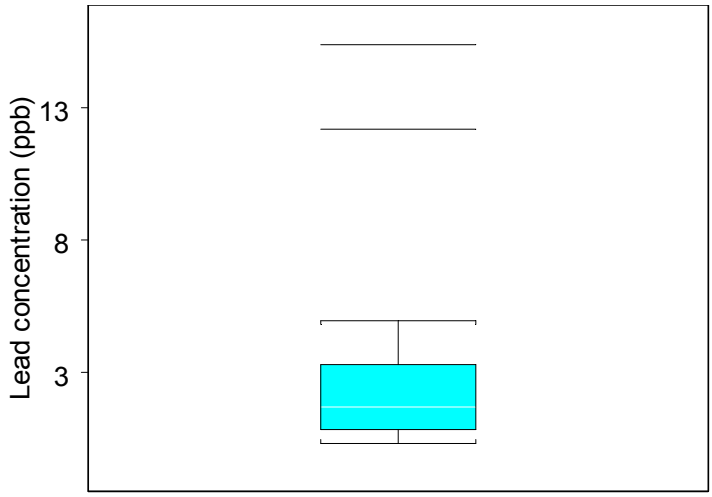
2.1.2 Box Plots

Box plots are another way to represent data graphically. Box plots were presented and described in Section 1.2.5.

Interpretation: On a box plot, potential outliers will appear as points (typically represented as asterisks or horizontal lines) beyond the whiskers of the plot. As with probability plots, box plots may be used to identify if potential outliers are present and, if so, how many. Formal testing should be completed before classifying these values as outliers.

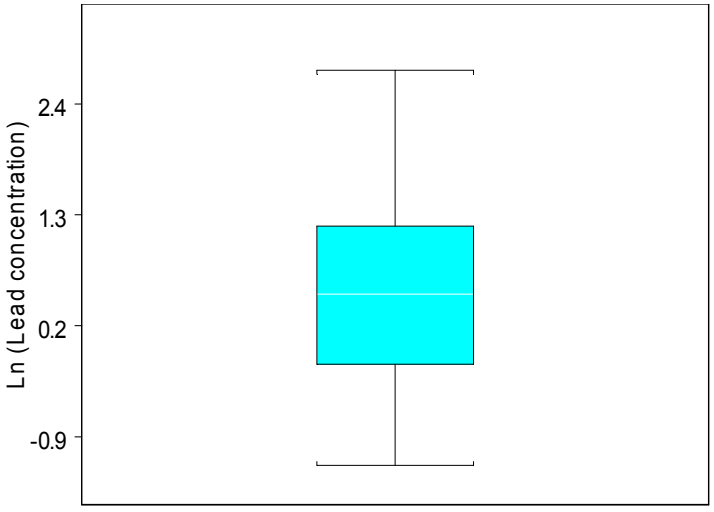
Example 2.2 Sample Box and Whisker Plots

Figure 2.3 Box Plot of Lead Data



This box plot, constructed using raw (untransformed) lead data, indicates the presence of two potential outliers. A box plot of log-transformed (Ln) values should be reviewed before identifying these as potential outliers.

Figure 2.4 Box Plot of Log-Transformed (Ln) Lead Data



This box plot of log-transformed (Ln) lead concentrations does not indicate the presence of any potential outliers. Consequently, formal testing for outliers may not be necessary.

2.1.3 Grubbs' Test

Grubbs' Test (Grubbs 1950, 1969) can be used to identify single outliers in most data sets with sample sizes ranging from small to relatively large ($3 < n < 100$). The test makes an assumption of normality, so the data set should first be evaluated for statistical distribution (Chapter 1). If the data are lognormally distributed, log-transform the data and conduct Grubbs' Test on the transformed values. Although Grubbs' Test can be used in an iterative fashion to evaluate whether there are multiple outliers, the procedure below describes only how to determine whether a single, large observation is an outlier. Other tests (e.g., Dixon's Test and Rosner's Test) should be used in cases where multiple outliers are suspected.

Procedure 2.1 Grubbs' Test for Single Outliers (3 < n < 100)

To test for outliers when the underlying distribution is normal, complete this procedure using the raw (untransformed) data, substituting 1/2 of the detection limit for nondetects up to 50% nondetect. To test for outliers when the underlying distribution is lognormal, first define y = ln(x) and use the y (log-transformed) values in place of the x values in the procedure below.

1. Calculate the sample mean ($\overline{x}$) and sample standard deviation (s) using all values, including the suspected outlier. (Equations 1.1 and 1.2 are reproduced below for convenience.)

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$$

2. Denote the maximum value (the suspected outlier) by x_n and calculate T_G using the following equation:

$$T_G = (x_n - \overline{x}) / s \qquad \text{Equation 2.1}$$

3. Compare the value of T_G to the critical value T_α in Table 2.1 based on the sample size (n) and a 95% level of confidence (i.e., α=0.05).

4. If $T_G \geq T_\alpha$, conclude that the observation x_n is an outlier. Follow the guidelines presented in Section 2.2 regarding the treatment of outliers.

Statistical software packages may be used to obtain some of the calculated values shown above. Most of these packages provide the sample standard deviation (s), the sample mean ($\bar{x}$), and the maximum value together with other common summary statistics. Alternatively, Microsoft Excel can be used to obtain *s*, $\bar{x}$, and the maximum for use in Equation 1.3. The Excel functions that should be used to obtain these values, respectively, are:

=STDEV(data range)

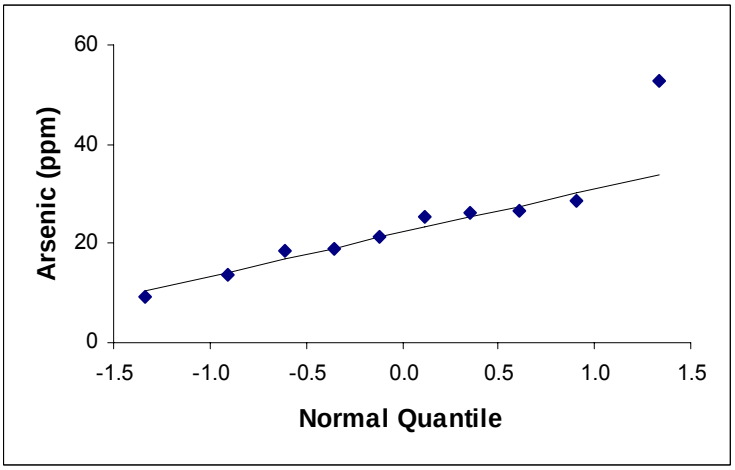
=AVERAGE(data range)

=MAX(data range)

Example 2.3 Sample Calculation of Grubbs' Test

At a hypothetical *FACILITY*, 10 samples for arsenic were collected. The measurements were 21.2, 26.0, 9.1, 28.7, 13.6, 52.6, 18.8, 25.5, 18.5, and 26.4 ppm. Suppose for this example that an evaluation as described in Chapter 1 indicated that the data were normally distributed, except for one potential outlier. Based on a normal probability plot (Figure 2.5), the largest observation (52.6 ppm) was suspected to be an outlier.

Figure 2.5 Probability Plot of Arsenic Data Set



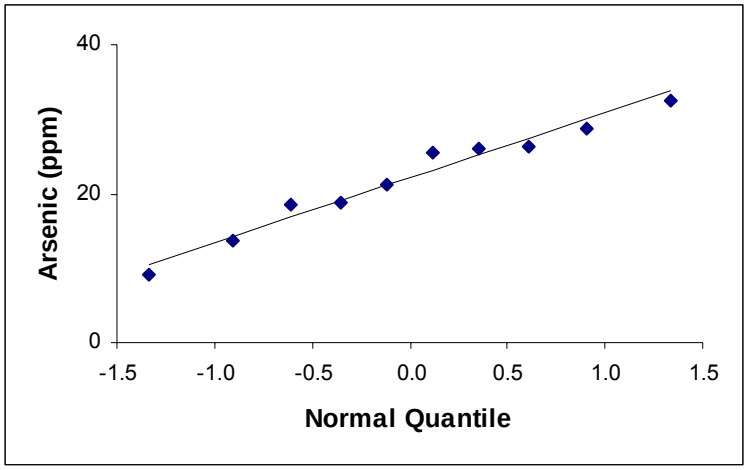
Spreadsheet Calculations for Grubbs' Test on the Arsenic Data Set

	A	B	C	D	E
1	Observation	Arsenic (ppm)			
2	1	9.1		**n:**	10
3	2	13.6			=COUNT(B2:B11)
4	3	18.5			
5	4	18.8		**Average:**	24.0
6	5	21.2			=AVERAGE(B2:B11)
7	6	25.5			
8	7	26.0		**Max:**	52.6
9	8	26.4			=MAX(B2:B11)
10	9	28.7			
11	10	52.6		**Std. Dev:**	11.8
12					=STDEV(B2:B11)
13					
14				**Grubb's T**	2.42
15					=(E8-E5)/E11

As shown above, the full data set was used to compute Grubbs' Test. The calculated value for T_G was 2.42 and the critical value T_α for a sample size of 10 at the α = 0.05 level is 2.176 (Table 2.1). Because $T_G > T_\alpha$ (2.42 > 2.176), we conclude that the 52.6 ppm observation is a statistical outlier.

Suppose that upon review of the data sheets, a data-entry error was found and the 52.6 ppm observation was actually measured as 32.6 ppm. After making this correction, we see that the probability plot of the raw data (Figure 2.6) now appears linear. Therefore, we would conclude that there are no outliers in the data set.

Figure 2.6 Probability Plot of Arsenic Data Set After Correcting Data-Entry Error



2.1.4 Dixon's Test

Dixon's Test (Dixon 1953) is used for identifying outliers in relatively small data sets ($n \leq 25$). Essentially, the test statistic is formed by a ratio, with the numerator representing the relative distance from the outlier to the next highest value and the denominator representing the spread of the data set (i.e., the range).

Dixon's Test is based on an assumption that the data are normally distributed. Therefore, the data set should first be evaluated for statistical distribution to make sure it conforms to this assumption (Chapter 1). If the data are lognormally distributed, log-transform the data and conduct Dixon's Test on the transformed values.

The test was initially developed for testing whether an individual observation is an outlier, but can be modified to accommodate testing for multiple outliers (Gibbons 1994). Consequently, if the data set is smaller than 25 and more than one outlier is suspected, Dixon's test is appropriate and can be used as follows:

Evaluate the *least extreme* observation first, temporarily excluding the more extreme observations from the data set. If the least extreme observation is identified as an outlier, then the more extreme observations can be classified as outliers as well. If the smallest potential outlier is not classified as such, then the next largest observation may be tested using the same procedure. This procedure, which is described in more detail in Procedure 2.2, may be continued until a set of outliers has been identified or until the test finds no outliers in the data set.

If more than one outlier is suspected and the sample size is greater than 25, Rosner's Test (Section 2.1.5) should be considered as an alternative to Dixon's Test. For more than 25 samples, Rosner's Test has the advantage of testing for both high and low outliers simultaneously. Dixon's Test can be used to test for high or low outliers, but no procedure has been established for testing for both high and low outliers simultaneously. DEQ recommends that Dixon's Test be used to test for high outliers with sample sizes less than 25. If sample sizes are less than 25 and low outliers or both high and low outliers are suspected, consultation with a professional statistician is advised.

Procedure 2.2 Dixon's Test for Single or Multiple Outliers (n < 25)

To test for outliers when the underlying distribution is normal, complete this procedure using the raw (untransformed) data, substituting 1/2 of the detection limit for nondetects up to 50% nondetects. To test for outliers when the underlying distribution is lognormal, first define y = ln(x) and use the y (log-transformed) values in place of the x values in the procedure below.

1. Order the data set from least to greatest and label the observations $x_{(1)}, x_{(2)}, ..., x_{(n)}$ where $x_{(1)}$ is the smallest observation and $x_{(n)}$ is the largest.

2. Based on the sample size (*n*), use the appropriate equation below to calculate the test statistic T_D:

n	Dixon's Test Statistic (T_D)	
3 - 7	$(x_{(n)} - x_{(n-1)}) / (x_{(n)} - x_{(1)})$	Equation 2.2
8 - 10	$(x_{(n)} - x_{(n-1)}) / (x_{(n)} - x_{(2)})$	Equation 2.3
11 - 13	$(x_{(n)} - x_{(n-2)}) / (x_{(n)} - x_{(2)})$	Equation 2.4
14 - 25	$(x_{(n)} - x_{(n-2)}) / (x_{(n)} - x_{(3)})$	Equation 2.5

3. Obtain the critical value for Dixon's test (T_α) based on the sample size (*n*) and a 95% level of confidence (i.e., α = 0.05) in Table 2.2.

4. Compare T_D to T_α. If $T_D > T_\alpha$, the value (and all potential outliers higher than this value) may be classified as outliers. Follow the recommendations for dealing with outliers (Section 2.2).

Microsoft Excel can be used to obtain each of the values for use in the above equations. The Excel functions that should be used to obtain the *k* th largest or smallest values, respectively, are:

=LARGE(data range,k)

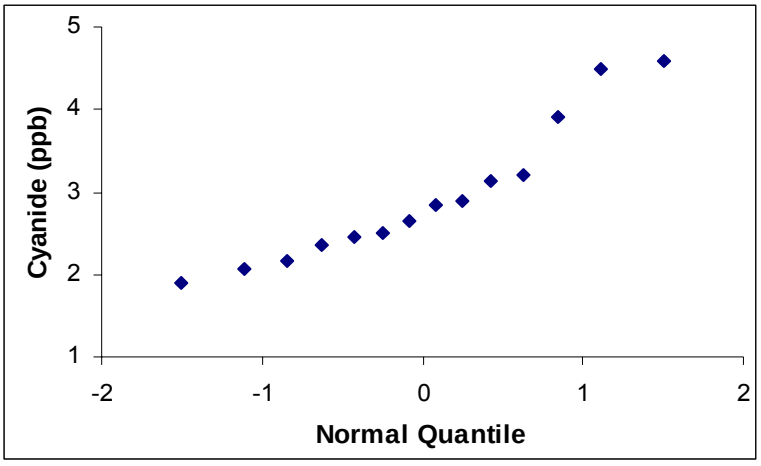
=SMALL(data range,k)

Where k denotes the *k* th largest or *k* th smallest value, respectively, in the data set.

Example 2.4 Sample Calculation of Dixon's Test

At a hypothetical *FACILITY*, 14 measurements for cyanide were collected. The values were 2.5, 4.5, 2.9, 1.9, 2.1, 3.9, 2.2, 2.4, 2.5, 4.6, 2.7, 2.8, 3.1, and 3.2 ppb. Suppose for this example that an evaluation as described in Chapter 1 indicated that the data were normally distributed, except for three potential outliers. Construction of a normal probability plot reveals that the three largest observations (3.9, 4.5, and 4.6 ppb) may be outliers (Figure 2.7).

Figure 2.7 Normal Probability Plot for Cyanide Data Set



Dixon's Test Calculations

	A	B	C	D	E	F
1	Observation	Cyanide				
2	1	1.9		Evaluating **3.9** :		
3	2	2.1			$x_{(12)}$:	3.9
4	3	2.2			$x_{(10)}$:	3.1
5	4	2.4			$x_{(2)}$:	2.1
6	5	2.5				
7	6	2.5			T_D:	0.444
8	7	2.7			= (3.9-3.1) / (3.9-2.1)	
9	8	2.8				
10	9	2.9			T (n=12):	0.546
11	10	3.1				
12	11	3.2		Evaluating **4.5**:		
13	12	**3.9**			$x_{(13)}$:	4.5
14	13	**4.5**			$x_{(11)}$:	3.2
15	14	**4.6**			$x_{(2)}$:	2.1
16						
17					T_D:	0.542
18					= (4.5-3.2) / (4.5-2.1)	
19						
20					T (n=13):	0.521

Because multiple outliers are suspected and the data set has less than 25 observations, Dixon's Test is appropriate. The least extreme potential outlier is evaluated first (3.9). The larger observations (4.5 and 4.6) are temporarily excluded from the data set, thus changing the sample size to 12 for the purposes of this test. With a sample size of 12, Equation 2.4 is used.

For n=12, the calculated value for T_D is 0.444 and the critical value for T based on a sample size of 12 is 0.546 (Table 2.2). Because 0.444 < 0.546, we conclude that 3.9 **is not an outlier**.

Next we move on to the 4.5 ppb observation. The 4.6 observation is temporarily excluded from the data set, changing the sample size to 13 for the purposes of this test. The calculated value for T_D is 0.542 (Figure 2.9) and the critical value for T based on a sample size of 13 is 0.521 (Table 2.2). Because 0.542 > 0.521, we conclude that the 4.5 ppb observation **is an outlier**. And because the 4.6 ppb observation is greater than the 4.5 ppb observation (which was just determined to be an outlier), we **consider both observations to be outliers**. Next we would follow the procedures outlined in Section 2.2 for dealing with these two outliers.

2.1.5 Rosner's Test

Rosner's Test (Rosner 1983) is an effective method for identifying outliers in moderate to large-sized data sets (i.e., n > 25). Potential outliers are tested in groups.

Similar to Grubbs' and Dixon's Tests, Rosner's Test assumes that the sampled population is normally distributed. Therefore, the data set should first be evaluated for statistical distribution to make sure it conforms to this assumption (Chapter 1). If the data are lognormally distributed, log-transform the data and conduct Rosner's Test on the transformed values.

Before completing Rosner's Test, the total number of potential outliers (k) must be identified. Probability plots, box plots, or a visual inspection of the tabulated data should be completed to screen first for potential outliers.

Rosner's Test procedure is iterative. As described in Procedure 2.3, the mean, standard deviation and maximum values are calculated first with the entire data set and again excluding potential outliers one by one, from largest to smallest, until all potential outliers have been removed. Then the first test statistic is calculated to test whether all k values are outliers. If the result is significant, all k observations are classified as outliers. If not significant, the group of possible outliers under evaluation is reduced by one (i.e., the smallest potential outlier is placed back in the data set) and the test statistic is recalculated considering the remaining k - 1 possible outliers. This process is repeated until a group of outliers is identified or until the test finds no outliers in the data set.

Procedure 2.3 Rosner's Test for Multiple Outliers (n > 25)

To test for outliers when the underlying distribution is normal, complete this procedure using the raw (untransformed) data, substituting 1/2 of the detection for nondetects up to 50% nondetect. To test for outliers when the underlying distribution is lognormal, first define y = ln(x) and use the y (log-transformed) values in place of the x values in the procedure below.

1. Order the data set from smallest to largest and denote the values as $x_{(1)}, x_{(2)}, \ldots x_{(n)}$. From plots or examination of the tabulated data, identify the number (*k*) of possible outliers.

2. Set *i* = 0 and use the following formulas

$$\bar{x}^{(i)} = (x_1 + x_2 + \ldots x_{n-i})/(n-i) \qquad \text{Equation 2.6}$$

$$s^{(i)} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_{n-i} - \bar{x})^2}{n-i}} \qquad \text{Equation 2.7}$$

Calculate the sample mean $(\bar{x})$ and sample standard deviation (s) for the full data set (i.e., when *i* = 0, the full data set is included in the above equations). Denote these values as $\bar{x}^{(0)}$ and $s^{(0)}$. Determine the value of the measurement furthest from $\bar{x}^{(0)}$ and denote it as $y^{(0)}$.

3. Remove the observation $y^{(0)}$ from the data set and recalculate the mean and standard deviation, denoting them as $\bar{x}^{(1)}$ and $s^{(1)}$ (i.e., set *i* = 1). Determine the value of the measurement furthest from $\bar{x}^{(1)}$ and denote it as $y^{(1)}$.

4. Remove the observation $y^{(1)}$ from the data set and recalculate the mean and standard deviation, denoting them as $\bar{x}^{(2)}$ and $s^{(2)}$ (i.e., set *i* = 2). Note that *i* refers to the number of observations that have been removed from the data set.

5. Repeat steps 3 and 4 until k potential outliers have been removed. This should provide a set of results similar to:

$$\left[\bar{x}^{(0)}, s^{(0)}, y^{(0)}\right], \left[\bar{x}^{(1)}, s^{(1)}, y^{(1)}\right], \ldots, \left[\bar{x}^{(k-1)}, s^{(k-1)}, y^{(k-1)}\right]$$

6. Evaluate the test for all k possible outliers first. To test for k outliers, compute the test statistic:

$$R_k = \left| y^{(k-1)} - \bar{x}^{(k-1)} \right| \Big/ s^{(k-1)} \qquad \text{Equation 2.8}$$

(Continued on next page)

Procedure 2.3 Rosner's Test (continued)

7. Obtain the critical value (R_α) from Table 2.3 given the sample size (n) and a 95% level of confidence (α = 0.05). If $R_k > R_\alpha$, conclude that there are k outliers. If not, repeat the procedure testing for $k-1$ outliers. Continue in this fashion until a group of outliers has been identified or until the test finds no outliers in the data set.

8. If one or more outliers are identified, see Section 2.2 regarding treatment of outliers.

In most cases, it should not be necessary to calculate Equations 2.6 through 2.8 by hand. Most statistical software packages provide the sample standard deviation (s) and the sample mean ($\bar{x}$) together with other common summary statistics. Alternatively, Microsoft Excel can be used to obtain s and $\bar{x}$ for use in Equation 2.8. The Excel functions that should be used to obtain these values, respectively, are:

=STDEV(data range)

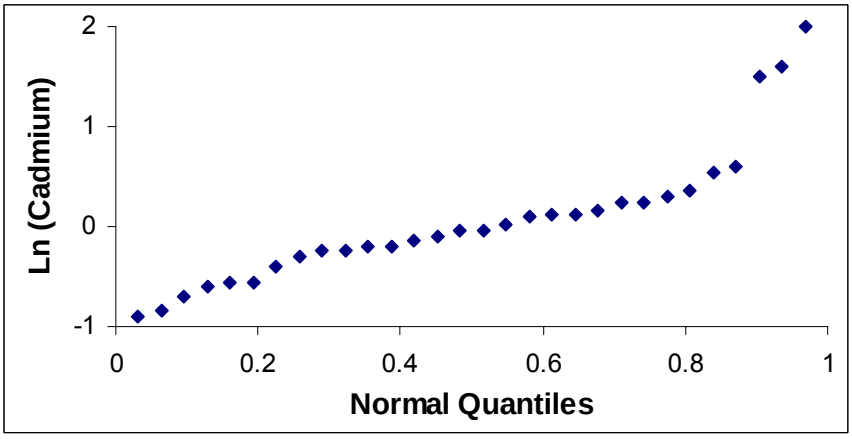
=AVERAGE(data range)

Where the data range should be modified as necessary for each iteration described above.

Example 2.5 Sample Calculation of Rosner's Test

Suppose that 30 observations for cadmium are collected at a site. For the purpose of this example, further suppose that the data were concluded to be lognromally distributed using the methods described in Chapter 1. Based on a probability plot of the log-transformed (Ln) data, the data appeared to contain three possible outliers (Figure 2.10).

Figure 2.8 Probability Plot of Log-Transformed (Ln) Cadmium



Calculations for Rosner's Test With the Cadmium Data Set

	A	B	C	D	E	F	G	H	I	J
1	Observation	Ln (Cadmium)								
2	1	-0.90		i	n-i	$\bar{y}^{(i)}$	$s_y^{(i)}$	$y^{(i)}$	R	$R_{crit.}$
3	2	-0.85		0	30	0.06	0.68	2.00	2.87	2.91
4	3	-0.70		1	29	0.00	0.58	1.60	2.77	2.89
5	4	-0.59		2	28	-0.06	0.50	1.50	3.13	2.88
6	5	-0.57								
7	6	-0.57				=AVERAGE(B2:B31)	=STDEV(B2:B31)	=MAX(B2:B31)	=ABS(H3-F3)/G3	
8	7	-0.40				=AVERAGE(B2:B30)	=STDEV(B2:B30)	=MAX(B2:B30)	=ABS(H4-F4)/G4	
9	8	-0.29				=AVERAGE(B2:B29)	=STDEV(B2:B29)	=MAX(B2:B29)	=ABS(H5-F5)/G5	
10	9	-0.23								
11	10	-0.23								
12	11	-0.20								
13	12	-0.19								
14	13	-0.13								
15	14	-0.10								
16	15	-0.03								
17	16	-0.03								
18	17	0.03								
19	18	0.10								
20	19	0.11								
21	20	0.13								
22	21	0.16								
23	22	0.25								
24	23	0.25								
25	24	0.30								
26	25	0.36								
27	26	0.55								
28	27	0.60								
29	28	1.50								
30	29	1.60								
31	30	2.00								

The three highest observations were evaluated using Rosner's Test. By equation 2.8, to test for three outliers, we must calculate R_3. The result of this calculation was 3.13 and the critical value for R (with n = 30 and i = 3) was 2.88 (Table 2.3). Because 3.13 > 2.88, we conclude that there are three outliers in the cadmium data set (1.5, 1.6, and 2.0).

2.1.6 Informal Iterative Approach to Outlier Testing When Data Set is Not Normal or Lognormal

As previously noted, the presence of outliers will sometimes cause a data set to fail tests for normality and lognormality. If a data set is found to be neither normal nor lognormal and graphical techniques indicate the presence of a potential outlier, an iterative approach using methods described in Chapter 1 may be taken, as described below.

- If one outlier is suspected, remove the potential outlier and reevaluate for normality and lognormality using the remaining observations and the methods described in Section 1. If the data pass a test for normality or lognormality without the suspected outlier, classify the anomalous value as an outlier.

- If multiple outliers are suspected, exclude them one by one and repeat the above step until the data pass a test of normality or lognormality or all of the potential outliers have been removed.

2.1.7 Walsh's Test

The formal outlier tests presented in this chapter have relied on an assumption of normality. Walsh's Test (Walsh 1958) provides a nonparametric alternative for evaluating potential outliers. Unfortunately, rather large sample sizes are required. For example, $n > 60$ is necessary to obtain a significance level of $\alpha = 0.10$ and $n > 220$ to obtain a significance level of $\alpha = 0.05$. Because of the large sample sizes necessary, it is doubtful that Walsh's Test can be applied to most environmental data sets and therefore a detailed procedure for conducting the test is not provided. However, if enough samples have been collected and the distribution appears to be neither normal nor lognormal, then Walsh's Test may be appropriate and can be *PROPOSED* for review by the DEQ statistician.

2.1.8 Evaluating for Outliers when Data Contain >50% Nondetects

As previously stated, for data sets with > 50% nondetects, it is generally not possible to identify the statistical distribution of the data (Helsel, 1990). Consequently, neither formal testing nor the iterative approach can be applied to the data. In this case, it will be necessary to evaluate for and identify outliers qualitatively using the graphical techniques described in Sections 2.1.1 and 2.1.2.

2.2 TREATMENT OF OUTLIERS

All values classified as outliers through formal testing (Section 2.1) should be thoroughly investigated before deciding how to treat them. The "chain of custody" records for the outlier should be reviewed as a check of the steps of the sampling and analysis up to this point. Each of the possible causes for outliers described in the introductory portion of this chapter should be considered.

Once all outliers have been investigated, one of the following actions should be taken:

1. If a transcription error is found and the correct value can be determined, replace the outlier with the correct value and conduct statistical analyses with the corrected value. The procedures described in Chapters 1 and 2 should be completed again with the revised data set.

2. If the observation can be proven erroneous, but the correct value cannot be determined, the outlier can be deleted and subsequent analyses conducted on the reduced data set. If a value is deleted from the data set, this fact must be reported with the statistical results.

3. If no error in the value can be found, the outlier should be regarded as a true, but extreme, observation. If this is the case, one of the following courses of action should be taken, depending on the general location of the sample containing the outlier.

   - ***FACILITY-SPECIFIC BACKGROUND*** **samples collected off of the property of interest:** If samples from the data set of interest were collected from off-site locations for the purpose of determining *FACILITY-SPECIFIC BACKGROUND* concentrations *and* the

outlier is not believed to represent naturally occurring *BACKGROUND* concentrations, the outlying value should be removed from the data set and documented as such. If the value is believed to be representative of *BACKGROUND* conditions, the value should be retained in the data set. Alternate statistical methods and/or additional sampling may be necessary.

- ***FACILITY-SPECIFIC BACKGROUND* samples collected on the property of interest:** If *FACILITY-SPECIFIC BACKGROUND* samples were collected from locations on the property of interest and the outlier is not believed to represent naturally occurring *BACKGROUND* concentrations, the outlying value may represent a previously unidentified area of contamination. The outlying result should be compared to the appropriate Part 201 criteria to determine if it represents a *FACILITY*. Additional characterization may be necessary in this area.

- ***FACILITY* samples:** If samples from the data set of interest were collected from locations within a *FACILITY*, the outlying value may represent a previously unidentified *HOT SPOT* if the concentration is sufficiently above criteria. Additional characterization may be necessary to determine if a *HOT SPOT* exists.

Table 2.1 Grubbs' Test Critical Values (T_α, α = 0.05). From Grubbs and Beck, 1972.

n	T_α	n	T_α	n	T_α
3	1.153	51	2.964	101	3.210
4	1.463	52	2.971	102	3.214
5	1.672	53	2.978	103	3.217
6	1.822	54	2.986	104	3.220
7	1.938	55	2.992	105	3.224
8	2.032	56	3.000	106	3.227
9	2.110	57	3.006	107	3.230
10	2.176	58	3.013	108	3.233
11	2.234	59	3.019	109	3.236
12	2.285	60	3.025	110	3.239
13	2.331	61	3.032	111	3.242
14	2.371	62	3.037	112	3.245
15	2.409	63	3.044	113	3.248
16	2.443	64	3.049	114	3.251
17	2.475	65	3.055	115	3.254
18	2.504	66	3.061	116	3.257
19	2.532	67	3.066	117	3.259
20	2.557	68	3.071	118	3.262
21	2.580	69	3.076	119	3.265
22	2.603	70	3.082	120	3.267
23	2.624	71	3.087	121	3.270
24	2.644	72	3.092	122	3.274
25	2.663	73	3.098	123	3.276
26	2.681	74	3.102	124	3.279
27	2.698	75	3.107	125	3.281
28	2.714	76	3.111	126	3.284
29	2.730	77	3.117	127	3.286
30	2.745	78	3.121	128	3.289
31	2.759	79	3.125	129	3.291
32	2.773	80	3.130	130	3.294
33	2.786	81	3.134	131	3.296
34	2.799	82	3.139	132	3.298
35	2.811	83	3.143	133	3.302
36	2.823	84	3.147	134	3.304
37	2.835	85	3.151	135	3.306
38	2.846	86	3.155	136	3.309
39	2.857	87	3.160	137	3.311
40	2.866	88	3.163	138	3.313
41	2.877	89	3.167	139	3.315
42	2.887	90	3.171	140	3.318
43	2.896	91	3.174	141	3.320
44	2.905	92	3.179	142	3.322
45	2.914	93	3.182	143	3.324
46	2.923	94	3.186	144	3.326
47	2.931	95	3.189	145	3.328
48	2.940	96	3.193	146	3.331
49	2.948	97	3.196	147	3.334
50	2.956	98	3.201		
		99	3.204		
		100	3.207		

Table 2.2 Critical Values for Dixon's Test (T_α, α = 0.05). From Dixon, 1953.

n	T
3	0.941
4	0.765
5	0.642
6	0.560
7	0.507
8	0.554
9	0.512
10	0.477
11	0.576
12	0.546
13	0.521
14	0.546
15	0.525
16	0.507
17	0.490
18	0.475
19	0.462
20	0.450
21	0.440
22	0.430
23	0.421
24	0.413
25	0.406

Table 2.3 Rosner's Test Critical Values (R_α, α = 0.05 for Various Levels of *n* and *k*). From Rosner, 1983.

n	k	Rα	n	k	Rα	n	k	Rα
25	1	2.82	35	1	2.98	45	1	3.09
	2	2.80		2	2.97		2	3.08
	3	2.78		3	2.95		3	3.07
	4	2.76		4	2.94		4	3.06
	5	2.73		5	2.92		5	3.05
	10	2.59		10	2.84		10	2.99
26	1	2.84	36	1	2.99	46	1	3.09
	2	2.82		2	2.98		2	3.09
	3	2.80		3	2.97		3	3.08
	4	2.78		4	2.95		4	3.07
	5	2.76		5	2.94		5	3.06
	10	2.62		10	2.86		10	3.00
27	1	2.86	37	1	3.00	47	1	3.10
	2	2.84		2	2.99		2	3.09
	3	2.82		3	2.98		3	3.09
	4	2.80		4	2.97		4	3.08
	5	2.78		5	2.95		5	3.07
	10	2.65		10	2.88		10	3.01
28	1	2.88	38	1	3.01	48	1	3.11
	2	2.86		2	3.00		2	3.10
	3	2.84		3	2.99		3	3.09
	4	2.82		4	2.98		4	3.09
	5	2.80		5	2.97		5	3.08
	10	2.68		10	2.91		10	3.03
29	1	2.89	39	1	3.03	49	1	3.12
	2	2.88		2	3.01		2	3.11
	3	2.86		3	3.00		3	3.10
	4	2.84		4	2.99		4	3.09
	5	2.82		5	2.98		5	3.09
	10	2.71		10	2.91		10	3.04
30	1	2.91	40	1	3.04	50	1	3.13
	2	2.89		2	3.03		2	3.12
	3	2.88		3	3.01		3	3.11
	4	2.86		4	3.00		4	3.10
	5	2.84		5	2.99		5	3.09
	10	2.73		10	2.92		10	3.05
31	1	2.92	41	1	3.05	60	1	3.20
	2	2.91		2	3.04		2	3.19
	3	2.89		3	3.03		3	3.19
	4	2.88		4	3.01		4	3.18
	5	2.86		5	3.00		5	3.17
	10	2.76		10	2.94		10	3.14
32	1	2.94	42	1	3.06	70	1	3.26
	2	2.92		2	3.05		2	3.25
	3	2.91		3	3.04		3	3.25
	4	2.89		4	3.03		4	3.24
	5	2.88		5	3.01		5	3.24
	10	2.78		10	2.95		10	3.21
33	1	2.95	43	1	3.07	80	1	3.31
	2	2.94		2	3.06		2	3.30
	3	2.92		3	3.05		3	3.30
	4	2.91		4	3.04		4	3.29
	5	2.89		5	3.03		5	3.29
	10	2.80		10	2.97		10	3.26
34	1	2.97	44	1	3.08	100	1	3.38
	2	2.95		2	3.07		2	3.38
	3	2.94		3	3.06		3	3.38
	4	2.92		4	3.05		4	3.37
	5	2.91		5	3.04		5	3.37
	10	2.82		10	2.98		10	3.35

CHAPTER 3: CALCULATION OF A 95% UPPER CONFIDENCE LIMIT (UCL) FOR THE MEAN CONCENTRATION

A goal of sampling environmental media is often to identify characteristics and/or draw conclusions about a defined population. An example of a typical population of interest under the Part 201 program would be concentrations of a hazardous substance in surface soil within a 1/4 acre *EXPOSURE UNIT*. One characteristic of the population that is useful to identify is the mean concentration. The mean, which is a measure of centrality, represents an average value. Other measures of centrality include the median (i.e., the middle value) and the mode (i.e., the most frequently occurring value). The mean and the median are more commonly used than the mode in environmental applications.

It is the MDEQ's policy that a mean rather than a median or other measure of centrality be used to estimate concentrations of hazardous substances for the purpose of comparison to Part 201 criteria. In terms of exposure to hazardous substances, the mean provides the best representation of average exposure levels at a *FACILITY* because it incorporates the magnitude of all observations. This is consistent with EPA guidance, which recommends for purposes of risk assessment that the mean concentration be used to estimate risks through exposure to a hazardous substance (EPA, 1992a). **Due to uncertainty in estimating the true mean concentration based on sample data, a UCL for the mean must generally be used to compare concentration data to Part 201 criteria, as described below.**

The true mean, also referred to as the population mean, is typically estimated by collecting data and using these data to calculate a sample mean. A sample mean is an example of a point estimate. It provides a single value to estimate the population mean; it does not represent the variability or uncertainty associated with the estimate. Confidence intervals around the sample mean are used to represent the range of uncertainty or variability associated with this estimate of the mean.

Confidence intervals can also be used to conduct a statistical test of the mean. For example, a UCL for the mean can be compared to a fixed value, such as a Part 201 criterion, to test whether a sample mean concentration is below the Part 201 criterion. This corresponds directly with EPA's recommendation that a UCL for the mean be used to estimate a reasonable maximum exposure (RME) concentration for Superfund risk assessments (EPA, 1992a). The RME "is intended to account for both uncertainty in the hazardous substance concentration and variability in exposure parameters (e.g., exposure frequency, averaging time)."

Use of a UCL for the mean to compare *FACILITY* data to Part 201 criteria corresponds to the following null and alternative hypotheses:

Hypothesis 1

H_o: The mean hazardous substance concentration in a given *EXPOSURE UNIT* is greater than or equal to the Part 201 criterion

H_a: The mean hazardous substance concentration in a given *EXPOSURE UNIT* is less than the Part 201 criterion

The null hypothesis (H_o) represents the condition that is assumed to be true. The alternative hypothesis (H_a), also known as the research hypothesis (H_R), is the converse of the null hypothesis. The alternative hypothesis will be concluded only if the sample data provide sufficient evidence that the null hypothesis is incorrect.

Practically speaking, the baseline assumption stated above in Hypothesis 1 is that the mean concentration is at or above its respective criterion unless the sample data provide sufficient evidence to conclude otherwise. Use of this baseline assumption is consistent with EPA's recommendations in the context of their cleanup programs (e.g., the Superfund program and RCRA Corrective Action) as described in many EPA statistical guidance documents (1989a, 1989b, 1992a, 1992b, 1993, 1996a, 1996b, 1999).

Further, EPA (1992b) recommends:

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

By setting the null hypothesis as something one wants to disprove, the motivation is to conduct a proper and rigorous statistical analysis in order to disprove this condition. In other words, be skeptical that concentrations meet criteria until evidence proves otherwise. Conversely, if the null hypothesis is set as something one wants to prove (e.g., hazardous substance concentrations are below Part 201 criteria), there is no motivation to obtain sufficient data or to utilize statistical rigor to disprove this assumption.

When calculating a UCL for the mean for the purpose of comparing *FACILITY* data to Part 201 criteria, a significance level of $\alpha = 0.05$ should be used. A significance level of $\alpha = 0.10$ may be used for the purpose of waste characterization (SW-846 Chapter 9; EPA 1986). See the tabbed section titled, "Waste Characterization" for further detail.

3.1 STEPS FOR CALCULATING AN UPPER CONFIDENCE LIMIT FOR THE MEAN

A minimum of nine *RANDOMLY* located samples per *EXPOSURE UNIT* is required if statistics are to be used to compare *FACILITY* data to Part 201 criteria. (For waste characterization purposes, a minimum of nine *RANDOM* samples is required if statistics are to be used to compare characterization data to regulatory thresholds.) This minimum number is necessary to evaluate the underlying statistical distribution of the data set, as described in Chapter 1. The necessity of evaluating the distribution of the data set is discussed below and described in detail in Chapter 1.

The following methods for calculating UCLs for the mean are presented or discussed in this chapter:

Method	Section
Student's t (recommended for normally distributed data)	3.1.1
Land's Method (recommended for lognormally distributed data)	3.1.2

Alternate methods for data sets which are neither normal nor lognormal (may be *PROPOSED* on a case-by-case basis)	3.1.3

Assumptions

Each of the statistical methods described in this chapter was developed based on certain underlying assumptions. For example, all of the methods described in this chapter require an assumption that the data are statistically independent (i.e., there are no trends in the data and obtained through *RANDOM* sampling) and representative of a single statistical distribution. Data should therefore be plotted on a map to identify spatial trends and or *HOT SPOTS* before conducting a statistical analysis. Furthermore, Statistical Guidesheets must be referred to for key considerations on the selection of an appropriate data set for development of a *REPRESENTATIVE CONCENTRATION*. Additional considerations and recommendations are presented in the tabbed section titled, "Sampling Strategies." (See Sections 1.2 and 2.2 on *FACILITY* Characterization and Section 2.4 on demonstrating compliance with Part 201 criteria using statistics.)

A second assumption required by many statistical methods is that the data were obtained from a specified underlying statistical distribution. Statistical methods requiring knowledge of the statistical distribution are called parametric methods. The method for constructing a UCL for the mean presented in Section 3.1.1 is based on an assumption that the data follow a normal distribution. Section 3.1.2 presents a method that was designed for data assumed to be lognormally distributed (Gilbert, 1987; EPA, 1992a).

Steps for Constructing a UCL for the Mean

The following procedure describes the selection of an appropriate method for calculating a UCL for the mean:

1) Determine the percent of data below the detection limit (e.g., a data set with nine samples, three of which are below the detection limit, contains 33% nondetects).

2) For data sets with < 50% nondetects, evaluate and identify the underlying statistical distribution of the data using the methods presented in Chapter 1. This step is not required for data sets with ≥ 50% nondetects, although the methods presented in Chapter 1 may still provide some insight into the data to the extent that detectable concentrations are contained in the data set.

3) All data sets should be evaluated for outliers as described in Chapter 2 whenever *FACILITY* data are being statistically compared to Part 201 criteria. Many of the methods for calculating UCLs for the mean concentration presented in this chapter are sensitive to outliers. As noted in Chapter 2, formal testing for outliers is recommended only if initial screening of the data (i.e., a review of tabulated data and/or plots) indicates the presence of one or more potential outliers. For data sets with ≥ 50% nondetects, the data should be qualitatively evaluated for outliers using the graphical techniques presented in Chapter 2.

4) Select the appropriate formula for calculation of a 95% UCL for the mean:
   - Data sets with < 50% nondetects that are approximately normal: see Section 3.1.1.
   - Data sets with < 50% nondetects that are approximately lognormal: see Section 3.1.2.

- Data sets with ≥ 50% nondetects and/or are neither normal nor lognormal: see Section 3.1.3.

Unfortunately, widely accepted methods are not available to calculate UCLs for the mean of data sets that contain ≥ 50% nondetects and/or are neither normal nor lognormal. Consequently, alternative methods for statistically comparing *FACILITY* data to Part 201 criteria must be *PROPOSED* if departmental approval of a response activity is being sought. Section 3.1.3 describes some alternative methods that may be considered, including nonparametric (i.e., distribution-free methods) and large sample approximations.

The importance of evaluating the underlying assumptions for each method can not be over-stated if accurate conclusions are to be drawn. Therefore, it is necessary to identify the underlying assumptions of each method up front and evaluate these assumptions as described above before calculating UCLs for the mean concentration for comparison to Part 201 criteria.

3.1.1 UCL for the Mean of a Normally Distributed *RANDOM* Variable (Student's t)

When a data set is normally distributed, a UCL for the mean may be calculated based on the Student's t distribution for comparison to Part 201 criteria. Procedure 3.1 describes the calculation of a UCL for the mean of a normal distribution using the Student's t distribution.

Assumptions

The assumption underlying this method is that the data set is approximately normal in its distribution. Although this method is somewhat robust to slight deviations from normality, this assumption should be evaluated using the methods described in Chapter 1.

Example 3.1 Sample Calculation of a Student's t 95% Upper Confidence Limit for the Mean of a Normally Distributed *RANDOM* Variable

Suppose that we have 10 observations of lead concentrations and they are 4.4, 2.4, 5.5, 7.6, 7.4, 8.5, 0.6, 4.5, 7.2, and 2.8 ppb. An evaluation of the data as described in Chapter 1 indicates that the data set is approximately normal. The calculations necessary to calculate a 95% UCL for the mean are presented in Figure 3.1 using Microsoft Excel. For this data set, the 95% UCL for the mean is 6.59 ppb.

Figure 3.1 Spreadsheet Calculations for a 95% UCL for the Mean

	A	B	C	D	E
1	Observation	Lead			
2	1	4.4		n:	10
3	2	2.4			=COUNT(B2:B11)
4	3	5.5			
5	4	7.6		$\bar{x}$:	5.09
6	5	7.4			=AVERAGE(B2:B11)
7	6	8.5			
8	7	0.6		s:	2.58
9	8	4.5			=STDEV(B2:B11)
10	9	7.2			
11	10	2.8		t (0.95, 9):	1.83
12					=TINV(0.1,9)
13					
14				95% UCL:	6.59
15				= E5 + E8 * E11 / SQRT(10)	

Procedure 3.1 Student's t Upper (1-α)% Confidence Limit for the Mean of a Normally Distributed *RANDOM* Variable

The Student's t method for calculating a UCL for the mean relies on having normal or nearly normal data. First evaluate the statistical distribution of the data using the methods described in Chapter 1. Outlier testing (Chapter 2) should be completed if suspect values are identified based on review of the data and/or graphs described in Chapter 1.

1. Calculate the sample mean ($\bar{x}$) and the sample standard deviation (*s*). (Equations 1.1 and 1.2 are reproduced below for convenience.)

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

where x_i represents the *i* th observed value, *s* represents the sample standard deviation, and $\bar{x}$ represents the sample mean and *n* is the number of samples in the data set. Substitute ½ of the detection limit for nondetects up to 50% nondetects. Additional options for handling nondetects are described in Chapter 5.

2. Using Table 3.1, look up the Student's t value for a (1-α)% level of confidence (α = 0.05 for Part 201 applications; α = 0.10 may be considered for waste characterization) and *n-1* degrees of freedom. This value is denoted $t_{1-\alpha,n-1}$.

3. Calculate the one-sided (1-α)% UCL for the mean as follows:

$$UCL_{1-\alpha} = \bar{x} + t_{1-\alpha,n-1}\frac{s}{\sqrt{n}} \qquad \text{Equation 3.1}$$

In most cases, it should not be necessary to calculate the above equations by hand. Most statistical software packages provide the sample standard deviation (s) and the sample mean ($\bar{x}$) with other common summary statistics. Alternatively, Microsoft Excel can be used to obtain *s*, $\bar{x}$, and the tabled value of $t_{1-\alpha,n-1}$. The Excel functions that can be used to obtain these values, respectively, are:

=STDEV(data range)
=AVERAGE(data range)
=TINV(2α,*n* -1)

Note that a value of 2 times α (e.g., 0.10 for α = 0.05 or 0.20 for α = 0.10 is used in the =TINV() function rather than α. This is because Excel automatically provides t-values for two-sided intervals, rather than one-sided intervals. Using α = 0.1 in this function will yield the correct t-value for a one-sided 95% UCL for the mean. This may be necessary for many of the statistical software packages as well.

Interpretation: If $UCL_{1-\alpha}$ > criterion, conclude that the mean concentration is above the criterion. If $UCL_{1-\alpha}$ < criterion, conclude that the mean concentration is below the criterion.

Note that a more correct statement of the interpretation would be to conclude that the mean is at or above the criterion if the UCL for the mean is greater than or equal to the criterion. However, because the data being evaluated are measured on a continuous scale, the probability of obtaining a UCL for the mean that is exactly equal to a given criterion is essentially zero. Therefore, the interpretation stated above will be relied upon for these comparisons.

3.1.2 UCL on the Mean of a Lognormally Distributed *RANDOM* Variable (Land's Method)

In cases where the data are lognormally distributed, Land's method can be used to calculate a UCL for the mean (Land, 1971; Gilbert, 1987; EPA, 1992a). Procedure 3.2 describes the calculation of a (1-α)% UCL for the mean using Land's method.

Land's method is sensitive to departures from lognormality. For example, a right-skewed data set (i.e., asymmetric with a long right tail) may be concluded to be lognormal based on an analysis as described in Chapter 1. However, the data set may be more highly skewed than a true lognormal distribution and/or contain one or more outliers. The resulting UCL for the mean may be inappropriately high. This is more likely to be the case when the sample size is small and/or highly variable. If this occurs, consider increasing the sample size until a more reasonable value is obtained. Alternative statistical methods may also be *PROPOSED*. Suggested methods for consideration include a method described by Parkin, et. al. (1990) or the methods described in Section 3.1.3.

Assumptions

Land's method is based on an assumption that the sampled population is lognormally distributed. Even slight deviations from lognormality may result in unreasonably high UCLs for the mean. The methods outlined in Section 1 should be used to evaluate this assumption.

Procedure 3.2 Land's Method for Calculating a (1-α)% UCL for the Mean of a Lognormally Distributed *RANDOM* Variable

Land's method is intended for lognormally distributed data only. Use Chapter 1 to ensure that this is the case before using Land's method. Outlier testing (Chapter 2) should be completed if suspect values are identified based on review of the data and/or graphs described in Chapter 1.

1. For nondetects, substitute 1/2 of the detection limit up to 50% nondetects. (See Chapter 5 for additional options for handling nondetects.) Then calculate the natural logarithm (Ln) of each observation in the data set (i.e., log-transform the data set). Define y = ln(x).

2. Calculate the mean $(\bar{y})$ and standard deviation (s_y) of the log-transformed data set using the y values in place of the x values in Equations 1.1 and 1.2. (Equations 1.1 and 1.2 are reproduced below for convenience.)

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

3. Look up the value for $H_{1-\alpha}$ in Table 3.2 for α = 0.05 (Part 201 applications) or Table 3.3 for α = 0.10 (for some waste characterization applications). This value is dependent upon the sample size (*n*) and the standard deviation of the log-transformed data (s_y). If the correct number of samples and standard deviation are not represented on the table, it will be necessary to interpolate between adjacent points. It may be necessary to interpolate in both directions. The details of a double linear interpolation are described in Box 4-31 of EPA (2000), provided in Appendix A of the tabbed section titled, Appendices." The interpolation is illustrated using Table A-10 of EPA (2000). Double linear interpolation can be done in the same manner on Tables 3.2 and 3.3.

4. Calculate the (1-α)% UCL of the mean using the following equation:

$$UCL_{1-\alpha} = \exp\left[\bar{y} + \frac{s_y^2}{2} + \frac{s_y H_{1-\alpha}}{\sqrt{n-1}}\right] \qquad \text{Equation 3.2}$$

Microsoft Excel can be used to take natural logarithms and exponentiate, as well as calculate s_y, s_y^2, and $\bar{y}$. The Excel functions that should be used to obtain these values, respectively, are:

=LN(cell) =EXP(cell)

=STDEV(data range) =VAR(data range)

=AVERAGE(data range)

Interpretation: If the resulting *UCL* > criterion, conclude that the mean concentration is at or above the criterion. If *UCL* < criterion, conclude that the mean concentration is below the criterion.

Note that a more correct statement of the interpretation would be to conclude that the mean is at or above the Part 201 criterion if the UCL for the mean is greater than or equal to the criterion. However, because the data being evaluated are measured on a continuous scale, the probability of obtaining a UCL for the mean that is exactly equal to a given criterion is essentially zero. Therefore, the interpretation stated above will be relied upon for these comparisons.

Example 3.2 Sample Calculation of 95% UCL for the Mean of a Lognormally Distributed *RANDOM* Variable Using Land's Method

Fifteen soil samples were collected for analysis of arsenic concentrations. Using the methods described in Chapter 1, the data were concluded to be lognormally distributed. Figure 3.2 shows the data set, the log-transformed data set, and the calculations involved in determining a 95% UCL for the mean using Land's method. For this data set, the calculated values were $\bar{y}$ = 1.92, s_y^2 = 0.76, s_y = 0.87, and $H_{0.95}$ = 2.545. Using Equation 3.2, the 95% UCL for the mean arsenic concentration was 18.2 ppb. Note that this value is below three of the individual concentrations in the data set.

Figure 3.2 Spreadsheet Calculations of a 95% UCL Using Land's Method

	A	B	C	D	E	F	G	H
1	Observation	Arsenic	Ln (Arsenic)					
2	1	3.20	1.16		**n:**	15		
3	2	27.70	3.32			=COUNT(C2:C16)		
4	3	2.57	0.94					
5	4	9.98	2.30		**$\bar{y}$:**	1.92		
6	5	6.38	1.85			=AVERAGE(C2:C16)		
7	6	2.12	0.75					
8	7	2.80	1.03		**s_y^2:**	0.76		
9	8	12.11	2.49			=VAR(C2:C16)		
10	9	6.92	1.93					
11	10	5.32	1.67		**s_y:**	0.87		
12	11	23.19	3.14			=STDEV(C2:C16)		
13	12	7.02	1.95					
14	13	7.32	1.99		**$H_{0.95}$:**	2.545		
15	14	28.62	3.35			Interpolated value from Table 3.1		
16	15	2.59	0.95					
17			⌒		**Square bracket:**	2.90		
18			=LN(B2)			= F5 + F8 / 2 + F11 * F14 / SQRT(14)		
19			=LN(B3)					
20			⋮		**95% UCL:**	18.2		
21						=EXP(F17)		

3.1.3 Alternate Methods for Calculating UCLs for the Mean

As previously noted, the Student-t method and Land's method rely on assumptions of normality and lognormality, respectively. For data sets which are neither normal nor lognormal, alternative methods must be *PROPOSED* if departmental approval of a response activity is being sought.

3.1.3.1 Large Sample Methods

The Central Limit Theorem states that, for large sample sizes, the mean will tend to be normally distributed regardless of the distribution of the sampled population. A large sample method for computing a UCL for the mean is presented by Gilbert (1987, pg. 139). Large sample methods for calculating UCLs for the mean perform better as the number of samples increases. For smaller data sets with significant skewness, the method does not tend to perform well (Singh, et al. 1997). As an alternative, Chen (1995) describes a UCL that uses the Central Limit Theorem and incorporates an adjustment factor based on the skewness of the data set. This method is also described by Singh, et al. (1997). Because this method accounts for possible skewness in the data, it may provide more reasonable and accurate results than the standard large sample method presented by Gilbert for data sets that are right skewed.

However, for data sets that are highly skewed, Singh, et al. (1999) have found that these large sample methods do not provide adequate coverage for the mean (i.e., they may underestimate the mean concentration). Therefore, they should not be considered in this case unless the number of samples is sufficiently large. Gilbert (1987) suggests that, for highly skewed distributions, a sample size of 50 or more may be required. As always, it is important that the data set be identified as described in the appropriate Statistical Guidesheet (i.e., the data should be from samples within an appropriate *EXPOSURE UNIT* and should not include data representing *HOT SPOTS*).

Since these methods involve calculation of a mean and standard deviation, they should not be used for data sets with ≥ 50% nondetects.

Large sample methods for calculating UCLs for the mean concentration must be *PROPOSED* for review and approval if departmental approval of a response activity is being sought.

3.1.3.2 Alternative Methods for Small Data Sets and Data Sets With ≥ 50% Nondetects

For most *FACILITIES*, the number of samples in a given area (e.g., *EXPOSURE UNIT*) will generally be too small to consider the large sample methods described above. Statistical methods for calculating UCLs for the mean are generally not available for small data sets that are neither normally nor lognormally distributed. Consequently, additional samples may be collected according to a DEQ-approved sampling plan or alternate statistical methods may be *PROPOSED* to compare *FACILITY* data to Part 201 criteria.

For small data sets and for data sets with ≥ 50% nondetects, UCLs for a percentile may be considered in place of UCLs for a mean. Two methods for selecting a percentile are described below:

1) **Select a percentile that will provide an estimate of the mean of the distribution.** Since the goal is to estimate the mean concentration in a given area, a specific percentile (generally higher than the 50th percentile or median) may be selected on a case-by-case basis that will provide an estimate of the mean concentration. Parkin, et. al. (1990) describe a method for identifying a percentile to estimate the mean of a lognormal distribution. This method may be *PROPOSED* for right-skewed data sets that are approximately lognormal. Parkin's method may provide a useful alternative to Land's method for evaluating data sets that are concluded to be lognormal, but result in unreasonably high UCLs for the mean using the Land's method.

2) **Select a percentile that is higher than the percent of data below the detection limit.** For highly censored data sets, EPA (2000) suggests considering a percentile higher than the percent of data below the detection limit. For example, if 67% of the data are below the detection limit, EPA recommends consideration of a UCL for the 70th or 75th percentile to statistically compare data to a given criterion. This method, and a specified percentile for comparing data to criteria, may be *PROPOSED* if department approval of a response action is being sought.

Once a percentile has been selected, a nonparametric UCL for the percentile can be calculated to compare *FACILITY* data to Part 201 criteria. The method described by Gilbert (1987, pg. 141) may be *PROPOSED* for data sets with greater than 20 samples. Alternatively, for data sets with 20 or fewer samples, the procedure described by Conover (1980, pg. 112) may be *PROPOSED*.

Note: Past EPA guidance (1989c, 1992c) recommended use of a UCL for the median (i.e., the 50th percentile) when comparing data sets that are neither normal nor lognormal to fixed criteria. For populations with symmetric distributions, such as the normal distribution, the median and the mean are equivalent. However, the lognormal distribution is more commonly used to describe environmental data. This is because many environmental data sets have asymmetric distributions with a long right tail representing high concentration levels (i.e., right-skewed distributions).

For data sets which are lognormally distributed, the mean is always greater than the median. The median will underestimate the mean when the distribution of the data is right-skewed. Consequently, the selected percentile must generally be higher than the median, as described above. This recommendation is supported by more recent EPA guidance (2000), which recommends use of an upper percentile higher than the median for data sets with a CV > 0.5, particularly if the proportion of data below the detection limit is high (i.e., greater than 30 %).

3.2 NUMBER OF SAMPLES

When obtaining samples to characterize soils or waste materials, it is important to assure that the analytical results obtained will provide an accurate estimation of the nature of the entire area/volume under consideration. The location and number of discrete samples to be collected at a particular site depend on many factors: the degree of accuracy desired, the spatial and temporal variability of the media being sampled (e.g., soils, treated media, waste, etc.) to be sampled, and the costs involved. An important objective in any sampling program is to obtain the most accurate and representative data possible while minimizing the associated costs. One method to

accomplish this goal is to use statistically valid sampling strategies. The appropriate sample size can be estimated and the sampling locations can be chosen without bias.

Sample size, that is the number of samples to be collected, is a critical issue for determining compliance using UCLs for the mean. Larger numbers of samples result in lower UCLs for the mean. Although sample size analysis is not strictly necessary, it may be useful for demonstrating compliance.

The DEQ recommends a minimum of nine samples per *EXPOSURE UNIT*. However, when the sample mean concentration is close to the cleanup criterion, it may be beneficial to collect additional samples with the goal of lowering the UCL for the mean to demonstrate compliance. Following are three methods that can be considered during planning for sample collection.

3.2.1 Normal Distributions

When the underlying statistical distribution of the data set has been determined to be normal the appropriate number of samples required for waste characterization can be calculated by either of the following methods. Additional methods may also be *PROPOSED*.

3.2.1.1 Lambda Test

If the preliminary data indicate that more samples are needed to make a statistical comparison to Michigan's cleanup or waste classification criteria (e.g., the UCL for the mean is greater than the criterion, but the mean concentration is below), the Lambda (λ) relationship may be used to identify the number of samples necessary to demonstrate compliance. The total number of samples necessary to demonstrate compliance (assuming that the initial data were representative) can be estimated by use of the Lambda (λ) relationship and then consulting a table of values. A step by step approach to calculating the appropriate sample size follows:

1. Using data from the *n* initial samples, calculate λ

$$\lambda = \frac{C - \overline{X}}{s} \qquad \text{Equation 3.3}$$

Where:

C = the Part 201 criterion (or regulatory threshold for waste characterization),
$\overline{X}$ = the arithmetic mean of the data, and
s = the sample standard deviation.

The lower the resulting value for λ, the more samples are required to maintain a certain level of confidence. Also, as $\overline{X}$ approaches C, λ becomes smaller, and therefore a greater sample size is indicated for a certain level of confidence.

2. Refer to Table 3.4 to obtain the appropriate total sample size (n_{total}) from the table of values based on the resulting value for λ and a one-sided α at the desired significance level (α = 0.05 for comparison to Part 201 criteria). The resulting number of samples (n_{total}) reflects the total number of samples necessary to demonstrate compliance.

3. Determine the number of additional samples necessary (n_{total}-n). Collect the additional samples using the same sampling and analytical procedures as the first n samples. All field and laboratory procedures should be kept as consistent as possible to lower the amount of variability in the data.

4. Evaluate the spatial distribution of the combined data to determine if any new *HOT SPOTS* have been identified. If *HOT SPOTS* are present, these must be addressed separately.

5. Reevaluate the combined data (excluding *HOT SPOTS*) for statistical distribution as described in Chapter 1 and outliers (if apparent) as described in Chapter 2. Use all data values to calculate a new $\overline{X}$ and s.

6. If the new $\overline{X} \geq$ C, then a 95% UCL for the mean will clearly be above C and collection of additional samples is not likely to result in a UCL for the mean that is below C. Therefore, it can be concluded that the contaminant is present at an unacceptable concentration and the study would be complete.

7. If the new $\overline{X}$ < C, recalculate the 95% UCL for the mean as described in Section 3.1. If the new 95% UCL for the mean is below C, it can be concluded that the mean concentration is below the criterion (C) with 95% confidence. However, if the new 95% UCL for the mean is above C, either conclude that contaminant concentrations are unacceptable or start again at step 1.

3.2.1.2 SW-846 Method

An appropriate number of samples can also be calculated for a data set displaying the characteristics of a normal distribution by using the formula obtained from SW-846. This formula uses the t-statistic at the α level of significance with n-1 degrees of freedom.

$$n_{total} = \frac{t^2_{n-1,\alpha} s^2}{\Delta^2} \qquad \text{Equation 3.4}$$

Where:

Δ = C - $\overline{x}$
C = the Part 201 criterion or the regulatory threshold for waste characterization,
$\overline{x}$ = the arithmetic mean of the data set,
s^2 = the sample variance of the data set,
$t_{n-1, \alpha}$ = the appropriate t-statistic derived using degrees of freedom of n-1, and
n = the number of samples previously collected, and
n_{total} = the total number of samples to be collected.

For Part 201 comparisons table 3.1 is used with α = 0.05. When characterizing waste, the t-statistic is found by entering table 3.1 with α = 0.10. This method directly calculates the number of samples required (n_{total}) at the selected confidence level.

3.2.2 Lognormal Distributions

For a lognormal distribution, no simple established sample size formula, such as those presented above for normal distributions, is available which can be used to establish the number of samples necessary to achieve a specified error limit (EPA 1999). The use of the

Land's Method is recommended to define the confidence limit. Other methods may be *PROPOSED* to determine the appropriate sample size for lognormal distributions.

Sample Size Determination Using Land's Procedure for Lognormal Data

When data can be shown to be lognormally distributed, Land's procedure can be applied to calculate UCLs for the mean (Land, 1971 and 1975). For lognormally distributed populations, a UCL for the mean is (Gilbert 1987):

$$UCL = e^{\left(\bar{y}+\frac{s^2}{2}+\frac{sH}{\sqrt{n-1}}\right)} = \hat{X}e^{\left(\frac{sH}{\sqrt{n-1}}\right)}$$

To attain a UCL for the mean with 100*d*% relative error for the mean we require that:

$$\frac{UCL-\hat{X}}{\hat{X}} = e^{\left(\frac{sH}{\sqrt{n-1}}\right)} - 1 < d$$

Solving for *n*, a sample size formula to insure that Land's procedure results in confidence limits with 100xd% precision is given by

$$n_{total} \cong \frac{s^2 H_{1-\alpha}{}^2}{\left(\ln(1+d)\right)^2} + 1 \qquad \text{Equation 3.5}$$

Where:

s_y^2 = the variance of the log transformed data,
α = the significance level (α = 0.05 or 0.10), and
$H_{1-\alpha}$ = tabled value that can be obtained from Table 3.2 for α = 0.05 or Table 3.3 for α = 0.10,
n = the number of samples previously collected, and
n_{total} = the total number of samples to be collected.

For example, with s=0.2 and n ranging from 10 to 51, H ≈ 1.8. For d=0.1,

$$n = (0.2)^2(1.8)^2 / \log\left((1.1)\right)^2 + 1 = 15.3$$

Rounding up, the total sample n_{total} would be 16.

It should be noted that relative error term (d) must be specified in the original untransformed scale.

3.3 LOWER CONFIDENCE LIMITS FOR THE MEAN

Lower confidence limits (LCLs) for the mean provide a lower bound for the true mean concentration. When used to statistically compare *FACILITY* data to Part 201 criteria, the LCL for the mean corresponds to the following null and alternative hypotheses:

Hypothesis 2

H_o: The mean hazardous substance concentration in a given *EXPOSURE UNIT* is less than or equal to the Part 201 criterion

H_a: The mean hazardous substance concentration in a given *EXPOSURE UNIT* is greater than the Part 201 criterion

The baseline assumption in Hypothesis 2 is that the mean hazardous substance concentration is at or below its respective criterion unless the sample data provide sufficient evidence to conclude that the mean concentration is significantly above the criterion. Because the Part 201 cleanup program pertains to sites of environmental contamination, this set of assumptions generally does not apply. Furthermore, a review EPA statistical guidance documents indicates that EPA's recommendations regarding the use of LCLs for the mean are limited to the context of RCRA compliance monitoring (1989c, 1992c). As stated in the introduction to this chapter, UCLs for the mean are generally recommended by the EPA in the context of their cleanup programs (e.g., the Superfund program and RCRA Corrective Action).

Consequently, the baseline assumption under Hypothesis 2 will be justifiable only in limited circumstances under Michigan's Part 201 program. For example, an LCL for the mean may be justified for comparing hazardous substance concentrations to Part 201 criteria for the purpose of a *FACILITY* determination if there is no evidence to suggest that there has been a *RELEASE* anywhere on the property. Because of the limited utility of LCLs for the mean in demonstrating compliance with Part 201 criteria, this statistical method is not described in detail.

Table 3.1 Values of $t_{1-\alpha, n-1}$, α = 0.05, 0.10

Cumulative t Distribution

	α	0.10	0.05
	1	3.078	6.314
	2	1.886	2.920
	3	1.638	2.353
	4	1.533	2.132
	5	1.476	2.015
	6	1.440	1.943
	7	1.415	1.895
	8	1.397	1.860
	9	1.383	1.833
	10	1.372	1.812
	11	1.363	1.796
	12	1.356	1.782
	13	1.350	1.771
	14	1.345	1.761
	15	1.341	1.753
	16	1.337	1.746
	17	1.333	1.740
	18	1.330	1.734
df	19	1.328	1.729
(n-1)	20	1.325	1.725
	21	1.323	1.721
	22	1.321	1.717
	23	1.319	1.714
	24	1.318	1.711
	25	1.316	1.708
	26	1.315	1.706
	27	1.314	1.703
	28	1.313	1.701
	29	1.311	1.699
	30	1.310	1.697
	40	1.303	1.684
	60	1.296	1.671
	120	1.289	1.658
		1.282	1.645

Table 3.2 Land's Method Values for $H_{0.95}$. From Land (1975).

	n									
sy	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	104.60	37.06	26.20	20.55	18.71	17.01	15.18	13.74	12.51	11.45
9.00	117.70	41.68	29.46	23.10	21.03	19.11	17.05	15.43	14.05	12.85
10.00	130.800	46.31	32.73	25.66	23.35	21.22	18.93	17.13	15.59	14.26

Table 3.3 Land's Method Values for $H_{0.90.}$ From Land (1975).

	n									
sy	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 3.4 Number of Observations for t-Test of the Mean

Level for *t* Test (α)

One-sided / Two-sided	α = 0.005 / = 0.01					α = 0.01 / = 0.02					α = 0.025 / = 0.05					α = 0.05 / = 0.1				
λ	β= 0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5	0.01	0.05	0.1	0.2	0.5
0.05																				
0.10																				
0.15																				122.0
0.20										139.0					99.0					70.0
0.25					110.0					90.0				128.0	64.0			139.0	101.0	45.0
0.30				134.0	78.0				115.0	63.0			119.0	90.0	45.0		122.0	97.0	71.0	32.0
0.35			125.0	99.0	58.0			109.0	85.0	47.0		109.0	88.0	67.0	34.0		90.0	72.0	52.0	24.0
0.40		115.0	97.0	77.0	45.0		101.0	85.0	66.0	37.0	117.0	84.0	68.0	51.0	26.0	101.0	70.0	55.0	40.0	19.0
0.45		92.0	77.0	62.0	37.0	110.0	81.0	68.0	53.0	30.0	93.0	67.0	54.0	41.0	21.0	80.0	55.0	44.0	33.0	15.0
0.50	100.0	75.0	63.0	51.0	30.0	90.0	66.0	55.0	43.0	25.0	76.0	54.0	44.0	34.0	18.0	65.0	45.0	36.0	27.0	13.0
0.55	83.0	63.0	53.0	42.0	26.0	75.0	55.0	46.0	36.0	21.0	63.0	45.0	37.0	28.0	15.0	54.0	38.0	30.0	22.0	11.0
0.60	71.0	53.0	45.0	36.0	22.0	63.0	47.0	39.0	31.0	18.0	53.0	38.0	32.0	24.0	13.0	46.0	32.0	26.0	19.0	9.0
0.65	61.0	46.0	39.0	31.0	20.0	55.0	41.0	34.0	27.0	16.0	46.0	33.0	27.0	21.0	12.0	39.0	28.0	22.0	17.0	8.0
0.70	53.0	40.0	34.0	28.0	17.0	47.0	35.0	30.0	24.0	14.0	40.0	29.0	24.0	19.0	10.0	34.0	24.0	19.0	15.0	8.0
0.75	47.0	36.0	30.0	25.0	16.0	42.0	31.0	27.0	21.0	13.0	35.0	26.0	21.0	16.0	9.0	30.0	21.0	17.0	13.0	7.0
0.80	41.0	32.0	27.0	22.0	14.0	37.0	28.0	24.0	19.0	12.0	31.0	22.0	19.0	15.0	9.0	27.0	19.0	15.0	12.0	6.0
0.85	37.0	29.0	24.0	20.0	13.0	33.0	25.0	21.0	17.0	11.0	28.0	21.0	17.0	13.0	8.0	24.0	17.0	14.0	11.0	6.0
0.90	34.0	26.0	22.0	18.0	12.0	29.0	23.0	19.0	16.0	10.0	25.0	19.0	16.0	12.0	7.0	21.0	15.0	13.0	10.0	5.0
0.95	31.0	24.0	20.0	17.0	11.0	27.0	21.0	18.0	14.0	9.0	23.0	17.0	14.0	11.0	7.0	19.0	14.0	11.0	9.0	5.0
1.00	28.0	22.0	19.0	16.0	10.0	25.0	19.0	16.0	13.0	9.0	21.0	16.0	13.0	10.0	6.0	18.0	13.0	11.0	8.0	5.0
1.1	24.0	19.0	16.0	14.0	9.0	21.0	16.0	14.0	12.0	8.0	18.0	13.0	11.0	9.0	6.0	15.0	11.0	9.0	7.0	
1.2	21.0	16.0	14.0	12.0	8.0	18.0	14.0	12.0	10.0	7.0	15.0	12.0	10.0	8.0	5.0	13.0	10.0	8.0	6.0	
1.3	18.0	15.0	13.0	11.0	8.0	16.0	13.0	11.0	9.0	6.0	14.0	10.0	9.0	7.0		11.0	8.0	7.0	6.0	
1.4	16.0	13.0	12.0	10.0	8.0	14.0	11.0	10.0	9.0	6.0	12.0	9.0	8.0	7.0		10.0	8.0	7.0	5.0	
1.5	15.0	12.0	11.0	9.0	7.0	13.0	10.0	9.0	8.0	6.0	11.0	8.0	7.0	6.0		9.0	7.0	6.0		
1.6	13.0	11.0	10.0	8.0	6.0	12.0	10.0	9.0	7.0	5.0	10.0	8.0	7.0	6.0		8.0	6.0	6.0		
1.7	12.0	10.0	9.0	8.0	6.0	11.0	9.0	8.0	7.0		9.0	7.0	6.0	5.0		8.0	6.0	5.0		
1.8	12.0	10.0	9.0	8.0	6.0	10.0	8.0	7.0	7.0		8.0	7.0	6.0			7.0	6.0			
1.9	11.0	9.0	8.0	7.0	6.0	10.0	8.0	7.0	6.0		8.0	6.0	6.0			7.0	5.0			
2.0	10.0	8.0	8.0	7.0	5.0	9.0	7.0	7.0	6.0		7.0	6.0	5.0			6.0				
2.1	10.0	8.0	7.0	7.0		8.0	7.0	6.0	6.0		7.0	6.0				6.0				
2.2	9.0	8.0	7.0	6.0		8.0	7.0	6.0	5.0		7.0	6.0				6.0				
2.3	9.0	7.0	7.0	6.0		8.0	6.0	6.0			6.0	5.0				5.0				
2.4	8.0	7.0	7.0	6.0		7.0	6.0	6.0			6.0									
2.5	8.0	7.0	6.0	6.0		7.0	6.0	6.0			6.0									
3.0	7.0	6.0	6.0	5.0		6.0	5.0	5.0			5.0									
3.5	6.0	5.0	5.0			5.0														
4.0	6.0																			
						99% confidence										95% confidence				

CHAPTER 4: COMPARISON OF *FACILITY* DATA TO *BACKGROUND*

As stated in Rule 701(c), the term *BACKGROUND* is defined as:

> *the concentration or level of a hazardous substance which exists in the environment at or regionally proximate to a site that is not attributable to any release at or regionally proximate to the site.*

According to Section 20a(11), when *BACKGROUND* concentrations of a hazardous substance are greater than the corresponding Part 201 risk-based criterion, *BACKGROUND* becomes the Part 201 criterion. Consequently, consideration of applicable Part 201 criteria is necessary before comparing *FACILITY* data to *BACKGROUND*. In general, *FACILITY* data will be compared to *BACKGROUND* concentrations only when *BACKGROUND* concentrations are greater than the applicable risk-based criterion.

When *BACKGROUND* concentrations are being considered, the objective becomes to determine whether the *FACILITY* concentrations are significantly higher than *BACKGROUND* concentrations for a hazardous substance. This determination may or may not involve a statistical comparison to *BACKGROUND*, depending on: 1) the type of *BACKGROUND* being considered, and 2) whether a statistical analysis of *FACILITY* data for comparison to *BACKGROUND* is appropriate.

The types of *BACKGROUND* that will generally be considered include: 1) *STATEWIDE DEFAULT BACKGROUND*, 2) *REGIONAL BACKGROUND*, and 3) *FACILITY-SPECIFIC BACKGROUND*. Additional information on each type of *BACKGROUND* is provided below.

In most cases, *FACILITY* data will be compared to *BACKGROUND* on a point-by-point basis. That is, concentrations of each hazardous substance in each *FACILITY* sample will be compared to directly to the *BACKGROUND* concentration and individual exceedances will be noted. When point-by-point comparisons are made, professional judgment is required to interpret the significance of exceedances that are very close to criteria or *BACKGROUND*, or that may be associated with insignificant quantities of a hazardous substance.

Statistical analysis of *FACILITY* data for comparison to *BACKGROUND* may be appropriate. Statistical Guidesheets for the applicable pathways/conditions should be consulted to determine the applicability of statistics for comparing *FACILITY* data to Part 201 criteria and key considerations for selection of the appropriate data set(s). Recommendations provided in the Statistical Guidesheets apply to all Part 201 criteria, including *BACKGROUND*. A statistical analysis of *FACILITY* data for comparison to *BACKGROUND* will most likely be appropriate for pathways/conditions categorized as **YES**, or for those categorized as **GNP** for which a statistical analysis is demonstrated to be appropriate.

Recommended methods for comparing *FACILITY* data to *BACKGROUND* are provided in the following sections:

Type of *BACKGROUND*	Statistical Analysis of *FACILITY* Data Appropriate?	Method for Comparing *FACILITY* Data to *BACKGROUND*
STATEWIDE DEFAULT BACKGROUND	No	Section 4.1.1
	Yes	Section 4.1.2
REGIONAL BACKGROUND	No	Section 4.2.1
	Yes	Section 4.2.2
FACILITY-SPECIFIC BACKGROUND	No	Section 4.3.1
	Yes	Section 4.3.2

Where the column "Statistical Analysis of *FACILITY* Data Appropriate?" is "No," a point-by-point comparison of *FACILITY* data to *BACKGROUND* concentrations is necessary.

If departmental approval of a response action is being sought, alternate statistical methods for comparing *FACILITY* data to *BACKGROUND* data may be *PROPOSED* on a case-by-case basis. Self-implemented response activities using statistics to support determinations must be documented in a manner that fully and clearly addresses the three questions outlined in the tabbed section titled, "Introduction."

4.1 *STATEWIDE DEFAULT BACKGROUND* CONCENTRATIONS

STATEWIDE DEFAULT BACKGROUND concentrations are provided in Operational Memorandum #15 for naturally occurring metals. (See Appendix B of the tabbed section titled, "Appendices.") For the purpose of statistical comparisons with *FACILITY* concentrations, the MDEQ considers these criteria to be fixed values, rather than statistically-derived numbers.

4.1.1 Point-by-Point Comparison of *FACILITY* Data to *STATEWIDE DEFAULT BACKGROUND* Concentrations

When the *BACKGROUND* concentration being used is a *STATEWIDE DEFAULT BACKGROUND* concentration, *FACILITY* concentrations will generally be compared to this value on a point-by-point basis. Therefore, in general, simply compare each individual *FACILITY* concentration to the *STATEWIDE DEFAULT BACKGROUND* value and note individual exceedances.

4.1.2 Statistical Comparison of *FACILITY* Data to *STATEWIDE DEFAULT BACKGROUND* Concentrations, When Appropriate

When comparing *FACILITY* data to a *STATEWIDE DEFAULT BACKGROUND* concentration and a statistical analysis of *FACILITY* data is documented to be appropriate, a 95% UCL for the mean of the *FACILITY* data set may be calculated for comparison to the *STATEWIDE DEFAULT BACKGROUND* concentration. This is the same approach taken for statistical comparison of *FACILITY* data to risk-based criteria since these criteria are also treated as fixed values.

Procedures for calculating a UCL for the mean are presented in Chapter 3. The first step is to determine the distribution of the *FACILITY* data set (Chapter 1) and whether outliers are present (Chapter 2). If the data are normally distributed, use Procedure 3.1 to calculate a 95% UCL for the mean ($UCL_{0.95}$) using the Student's t method and compare this value to the *STATEWIDE DEFAULT BACKGROUND* concentration. If the data are lognormally distributed, use Procedure 3.2 to calculate $UCL_{0.95}$ based on Land's method and compare this value to the *STATEWIDE DEFAULT BACKGROUND* concentration. Alternate methods may be *PROPOSED* on a case-by-case basis, if departmental approval of a response action is being sought.

Interpretation: If $UCL_{0.95}$ > *STATEWIDE DEFAULT BACKGROUND* value, conclude that the mean *FACILITY* concentration is above this value. If $UCL_{0.95}$ < *STATEWIDE DEFAULT BACKGROUND* value, conclude that the mean *FACILITY* concentration is below the *STATEWIDE DEFAULT BACKGROUND* concentration.

Note that a more correct statement of the interpretation would be to conclude that the mean is at or above the *STATEWIDE DEFAULT BACKGROUND* value if the 95% UCL for the mean is greater than or equal to the *BACKGROUND* value. However, because the data being evaluated are measured on a continuous scale, the probability of obtaining a 95% UCL for the mean that is exactly equal to the *STATEWIDE DEFAULT BACKGROUND* value is essentially zero. Therefore, the interpretation stated above will be relied upon for these comparisons.

4.2 *REGIONAL BACKGROUND* CONCENTRATIONS

REGIONAL BACKGROUND concentrations may be *PROPOSED* on a case-by-case basis for comparison to *FACILITY* data. *REGIONAL BACKGROUND* data are typically considered in one of two manners:

1) **Development of a *REGIONAL BACKGROUND* concentration:** Data provided in sources such as the 1991 Michigan *BACKGROUND* Soil Survey (MBSS) can be *PROPOSED* to develop a *REGIONAL BACKGROUND* concentration. The MBSS should not generally be used as the only source of information for this purpose because the data in this survey do not equally represent all areas in Michigan (i.e., large numbers of *BACKGROUND* samples were collected at some locations, but small numbers or no samples in others). This yields a disproportionate weight to the locations with large numbers of samples. Other potential sources of data include approved *FACILITY-SPECIFIC BACKGROUND* data for *FACILITIES* in the nearby region, data published by the United States Geological Survey or other approved sources. Once a regional data set has been compiled, the statistical method for calculating a *REGIONAL BACKGROUND* concentration must also be *PROPOSED* on a case-by-case basis if departmental approval of a response action is being sought.

2) **Use of Professional Judgment:** In some cases, professional judgment based on general knowledge of *BACKGROUND* conditions in the region is used rather than the more rigorous approach described above. For example, if *FACILITY* concentrations are generally within the range of *BACKGROUND* concentrations that have been approved at other *FACILITIES* or are known to be present nearby within the region, it could be concluded that the *FACILITY* concentrations comply with *REGIONAL BACKGROUND* concentrations.

4.2.1 Point-by-Point Comparison of *FACILITY* Data to *REGIONAL BACKGROUND* Data

The method for establishing a *REGIONAL BACKGROUND* concentration and the *REGIONAL BACKGROUND* data used for this purpose must generally be *PROPOSED* on a case-by-case basis if departmental approval of a response action is being sought.

When a *BACKGROUND* concentration has been established as a *REGIONAL BACKGROUND* concentration, *FACILITY* concentrations will generally be compared to the *REGIONAL BACKGROUND* concentration on a point-by-point basis. Therefore, in general, simply compare each *FACILITY* concentration to the *REGIONAL BACKGROUND* concentration and note individual exceedances.

4.2.2 Statistical Comparison of *FACILITY* Data to *REGIONAL BACKGROUND* Data, When Appropriate

When the *BACKGROUND* concentration has been established as a *REGIONAL BACKGROUND* concentration and a statistical analysis of *FACILITY* data is appropriate, a statistical analysis of *FACILITY* data may be conducted for comparison to *REGIONAL BACKGROUND.* The method for comparing *FACILITY* data to *REGIONAL BACKGROUND* data must be *PROPOSED* on a case-by-case basis if departmental approval of a response action is being sought.

4.3 *FACILITY-SPECIFIC BACKGROUND* CONCENTRATIONS

Comparison of *FACILITY* data to *FACILITY-SPECIFIC BACKGROUND* concentrations has historically been completed using the mean plus three standard deviations calculated using *FACILITY-SPECIFIC BACKGROUND* data. This method of deriving *BACKGROUND* concentrations is still allowable in certain circumstances, as described in Section 4.3.1. However, it should be noted that this upper limit statistically represents the expected variability of sample results in an uncontaminated area, not a mean concentration; consequently, **only individual *FACILITY* concentrations may be compared to the mean plus three standard deviations**. **It is not appropriate to compare a mean concentration or a UCL for the mean concentration to an upper limit for *BACKGROUND* calculated as the mean plus three standard deviations.**

If a statistical analysis of *FACILITY* data is demonstrated to be appropriate for comparison to *BACKGROUND*, an alternative to the mean plus three standard deviation approach may be taken to compare *FACILITY* concentrations to *FACILITY-SPECIFIC BACKGROUND* concentrations, as described in Section 4.3.2. The recommended methods described in this section may be used to statistically compare the two populations represented by *FACILITY* concentrations and *FACILITY-SPECIFIC BACKGROUND* concentrations.

4.3.1 Point-by-Point Comparison of *FACILITY* Data to *FACILITY-SPECIFIC BACKGROUND* Data

When a *BACKGROUND* concentration has been established as a *FACILITY-SPECIFIC BACKGROUND* concentration, *FACILITY* concentrations will generally be compared to the *FACILITY-SPECIFIC BACKGROUND* concentration on a point-by-point basis. Therefore, in general, simply compare each individual *FACILITY* concentration to the *FACILITY-SPECIFIC BACKGROUND* concentration and note individual exceedances.

4.3.1.1 Normal or Lognormal Distributions

A *FACILITY-SPECIFIC BACKGROUND* concentration may be established as described in Procedure 4.1. This procedure is recommended for *FACILITY-SPECIFIC BACKGROUND* data sets that are either normally or lognormally distributed. If departmental approval of a response action is being sought, *PROPOSALS* must be made to calculate *FACILITY-SPECIFIC BACKGROUND* concentrations using alternative approaches.

Before calculating the *FACILITY-SPECIFIC BACKGROUND* concentration, it is still necessary to evaluate the data set for underlying statistical distribution (i.e., normal, lognormal, or neither) and the presence of outliers using the methods described in Chapters 1 and 2, respectively. If potential outliers are identified based on a review of the tabulated data and/or graphs of the data, formal outlier testing should be completed as described in Chapter 2, taking into account the underlying statistical distribution of the data.

Figure 4.1 Data and Calculations for Determining the *FACILITY-SPECIFIC BACKGROUND* Concentration and Evaluating for Exceedances

	A	B	C	D	E	F	G
1							
2		Background	Facility				
3	Observation	Conc.	Conc.	Exceeds?			
4	1	24.4	25.0	No		BG average:	25.1
5	2	22.4	29.7	No		=AVERAGE(B4:B13)	
6	3	25.5	31.1	No			
7	4	27.6	28.4	No		BG stdev:	2.6
8	5	27.4	23.8	No		=STDEV(B4:B13)	
9	6	28.5	34.2	Yes			
10	7	20.6	24.3	No	Facility-specific BG conc:		32.8
11	8	24.5	26.4	No		= G4 + 3 * G7	
12	9	27.2	42.6	Yes			
13	10	22.8	30.5	No			
14	11		31.0	No			
15	12		30.9	No			

Example 4.1 Sample Calculation of a *FACILITY-SPECIFIC BACKGROUND* Concentration

Suppose that 10 *FACILITY-SPECIFIC BACKGROUND* observations and 12 *FACILITY* observations of lead concentrations were collected. The data are provided in the spreadsheet below (Figure 4.1). No apparent outliers were present in the *FACILITY-SPECIFIC BACKGROUND* data. Using the methods outlined in Procedure 4.1, the average *BACKGROUND* concentration was estimated as 25.1 ppm and the standard deviation was 2.6 ppm. The resulting *FACILITY-SPECIFIC BACKGROUND* concentration was 32.8 ppm. Two exceedances were noted (34.2 and 42.6 ppm) as shown below.

Procedure 4.1 Determining a *FACILITY-SPECIFIC BACKGROUND* Concentration

1. For less than 50% nondetects, calculate the sample mean ($\bar{x}$) and sample standard deviation (*s*) using the following equation and substituting 1/2 of the detection limit for nondetects. (Equations 1.1 and 1.2 are reproduced below for convenience.)

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

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

2. For less than 50% nondetects, use the above values for $\bar{x}$ and s to calculate the *FACILITY-SPECIFIC BACKGROUND* concentration as:

$$\bar{x} + 3s$$

   For 50% nondetects or more, calculate the *FACILITY-SPECIFIC BACKGROUND* concentration using a nonparametric upper tolerance limit (EPA 1992c). This value should be determined as the maximum detected concentration in the *FACILITY-SPECIFIC BACKGROUND* data set.

3. Compare the *FACILITY* concentrations to the *FACILITY-SPECIFIC BACKGROUND* concentration on a point-by-point basis* and note exceedances.

In most cases, it should not be necessary to calculate the above equations by hand considering the availability of basic statistical computations in statistical software packages and most spreadsheet packages. Most statistical software packages provide the sample mean ($\bar{x}$) and the sample standard deviation (s). Alternatively, Microsoft Excel can be used to obtain $\bar{x}$ and *s*. The Excel functions that should be used to obtain these values, respectively, are:

=AVERAGE(data range)
=STDEV(data range)

* When point-by-point comparisons are made, professional judgment is required to interpret the significance of exceedances that are very close to criteria or *BACKGROUND*, or that may be associated with insignificant quantities of a hazardous substance.

4.3.1.2 Alternate Methods for Calculating *FACILITY-SPECIFIC BACKGROUND* Concentrations

Alternate methods are available for calculating *FACILITY-SPECIFIC BACKGROUND* concentrations. These methods include upper tolerance limits and upper prediction limits, among others. Upper tolerance limits and upper prediction limits are described in many documents and texts, including those by EPA (1992c), Gibbons (1994), Gibbons and Coleman (2001), Helsel and Hirsch (1992), and Millard and Neerchal (2001).

Use of alternate methods must be *PROPOSED* if departmental approval of a response action is being sought.

4.3.2 Statistical Comparison of *FACILITY* Data to *FACILITY-SPECIFIC BACKGROUND* Data, When Appropriate

If a statistical analysis of the *FACILITY* data is appropriate, statistical testing may be conducted to evaluate whether the population represented *FACILITY* data is signficantly greater than the population represented by *FACILITY-SPECIFIC BACKGROUND* data. A statistical comparison of two populations is referred to as a "two-sample test." Recommended two-sample tests (EPA 2000) include: 1) Student's t-test, 2) Satterthwaite's t-test, or 3) Wilcoxon Rank Sum test followed by the Quantile test. Selection of the appropriate test depends upon an evaluation of the underlying assumptions of these tests, as described below.

Two-sample testing is an alternative to the approach described in Section 4.3.1 because the focus is no longer on individual exceedances of a *BACKGROUND* value. Rather, two-sample tests typically compare some parameter of the population distributions, such as the mean. The Student t-test and Satterthwaite's t-test both provide methods for identifying differences in mean concentrations. The Wilcoxon Rank Sum test can be conducted to test for differences in the median concentration. The EPA (2000) recommends that the Wilcoxon Rank Sum test be conducted and interpreted together with the Quantile test. The Quantile test allows for detection of instances where only parts of the data set are different, rather than a complete shift in the data set. The EPA (2000) recommends this "tandem testing" approach since the combined tests are most powerful for detecting true differences between two population distributions.

Because EPA's *Guidance for Data Quality Assessment* (EPA, 2000) contains detailed explanations of the procedures for conducting the two-sample tests recommended in this section, this EPA document has been referenced as a source for descriptions of these tests. Excerpts from EPA (2000) are provided in Appendix A of the tabbed section titled, "Appendices."

Regardless of whether a statistical analysis is appropriate to compare *FACILITY* data to risk-based criteria, it is necessary to compute a *FACILITY-SPECIFIC BACKGROUND* concentration as described in Procedure 4.1 to determine if *BACKGROUND* is higher than the corresponding risk-based criteria. If the resulting *BACKGROUND* mean plus three standard deviations is greater than the risk-based criterion for a given hazardous substance, then this value becomes the Part 201 criterion for that hazardous substance. However, if the resulting *BACKGROUND* mean plus three standard deviations is not greater than the risk-based criterion, *BACKGROUND* does not become the Part 201 criterion and no further comparisons to *FACILITY-SPECIFIC BACKGROUND* are relevant.

4.3.2.1 Parametric Methods for Comparing *FACILITY-SPECIFIC BACKGROUND* and *FACILITY* Data Sets

First, determine the percent of nondetects in the combined data set. That is, combine the *FACILITY-SPECIFIC BACKGROUND* and *FACILITY* data sets determine the percent nondetect. If the percent nondetect in the combined data sets is ≥ 15%, proceed to Section 4.3.2.2.

If the combined data set is < 15% nondetect, determine the distribution of the *FACILITY-SPECIFIC BACKGROUND* and *FACILITY* data sets separately (Chapter 1). That is, complete the methods described in Chapter 1 for both data sets individually.

Next, evaluate each data set for potential outliers. If outliers are suspected, formal testing should be completed as described in Chapter 2. If sufficient data remain to allow for a valid statistical analysis, proceed as follows:

If both data sets are normally distributed, or both data sets are lognormally distributed, conduct Levene's test for equality of variances between the two data sets. The procedure for conducting Levene's test can be found in Box 4-26 on page 4-37 and an example can be found in Box 4-27 on page 4-38 of EPA (2000). These pages are provided in Appendix A of the tabbed section titled, "Appendices."

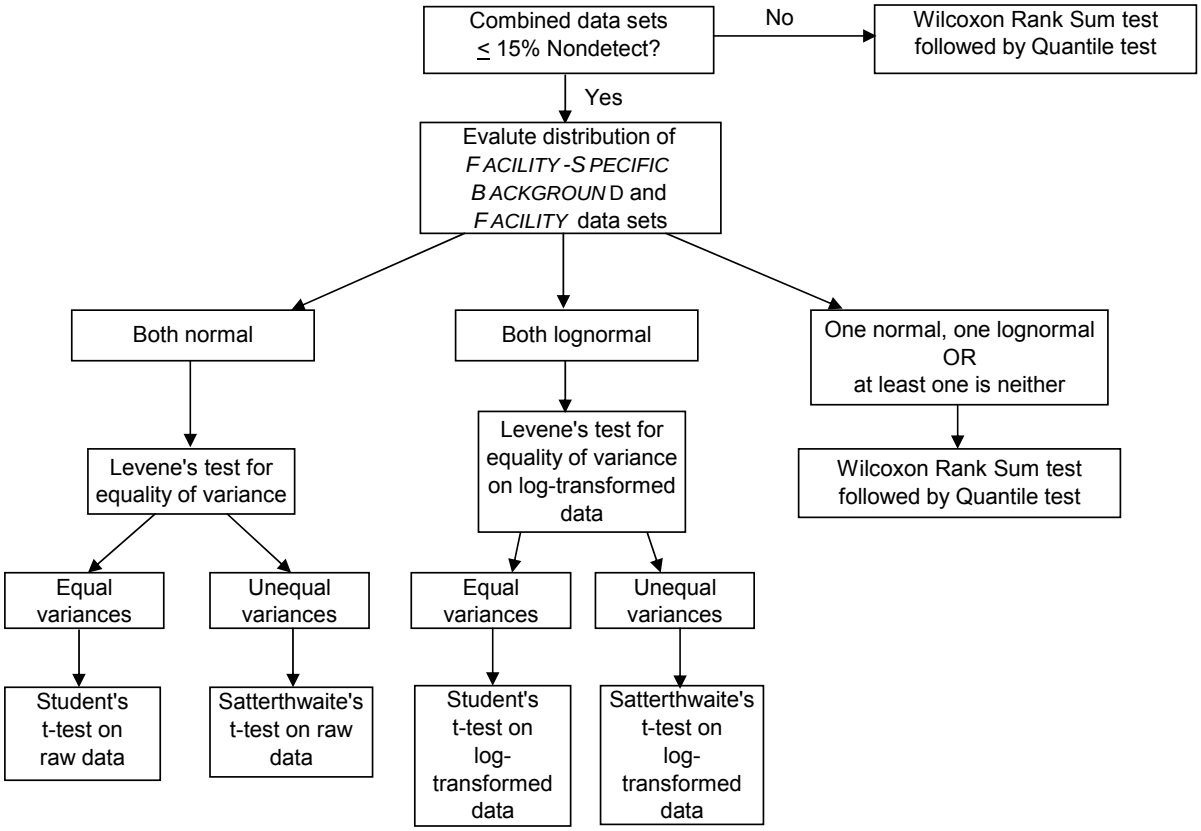
- If both data sets are normal and Levene's test indicates that variances of the data sets are equal, conduct the Student's t-test using the raw, untransformed data. Equality of variance is necessary since the Student's t-test assumes that both populations have equal variance. If both data sets are lognormal and the variances of the data sets are concluded to be equal, conduct the Student's t-test using log-transformed data. The procedure for conducting the Student's t-test and an example can be found in Box 3-14 on page 3-24 and Box 3-15 on page 3-25 of EPA (2000), respectively. These pages are provided in Appendix A of the tabbed section titled, Appendices."

- If both data sets are found to be normal and Levene's tests indicates that data sets do not have equal variances, conduct Satterthwaite's t-test using the raw, untransformed data. Satterthwaite's t-test provides an alternative to the Student's t-test when the populations being compared have unequal variances. If both data sets are lognormal and the variances of the data sets are not equal, conduct Satterthwaite's t-test using log-transformed data. Details for conducting Satterthwaite's t-test and an example can be found in Box 3-16 on page 3-26 and Box 3-17 on page 3-27 in EPA (2000), respectively. These pages are provided in Appendix A of the tabbed section titled, "Appendices." As noted by Millard and Neerchal (2001), it is important to decide whether it makes sense to focus on a difference in mean concentrations if you already know there is a difference in the variances. If this is the case, consider evaluating the data using the methods described in Section 4.3.2.2.

- If one data set is found to be normal and the other is found to be lognormal or if either data set is found to be neither normal nor lognormal, compare the data sets using the Wilcoxon Rank Sum Test followed by the Quantile Test, as described in Section 4.3.2.2.

See Figure 4.2 for a flowchart describing selection of the appropriate two-sample test.

Figure 4.2 Selection of Appropriate Two-Sample Test for Comparison of *FACILITY* Data to *FACILITY-SPECIFIC BACKGROUND* Data, When Appropriate



4.3.2.2 Nonparametric Methods for Comparing *FACILITY-SPECIFIC BACKGROUND* and *FACILITY* Data Sets

Nonparametric (i.e., distribution-free) methods for comparing *FACILITY-SPECIFIC BACKGROUND* data to *FACILITY* data are recommended when:

- One data set is found to be normal and the other is found to be lognormal, or
- Either data set is found to be neither normal nor lognormal, or
- The percent of nondetects in the combined data set is ≥ 15% (EPA 1992c).

The recommended nonparametric method involves comparing the data sets using the Wilcoxon Rank Sum Test followed by the Quantile Test. Directions for conducting a Wilcoxon Rank Sum Test can be found in Box 3-20 on page 3-32 and an example can be found in Box 3-21 on page 3-33 of EPA (2000). A large sample approximation to the Wilcoxon Rank Sum Test is provided in Box 3-22 on page 3-34 of EPA (2000) for comparing data sets each containing 20 or more samples. The Quantile Test is discussed on page 3-35 of EPA (2000). Directions for conducting a modified Quantile test can be found in Box 3-23 on page 3-36 and an example can be found in Box 3-24 on page 3-37 of EPA (2000). Each of these pages is provided in Appendix A of the tabbed section titled, "Appendices."

Although the Wilcoxon Rank Sum test can be done using a spreadsheet, the Quantile test is rather difficult to compute. S-PLUS contains functions that can perform both of these tests.

4.4 *BACKGROUND* CONCENTRATIONS FOR GROUNDWATER

Some issues particular to calculating *BACKGROUND* groundwater concentrations deserve special consideration. Careful consideration should be given to the selection of an appropriate *BACKGROUND* groundwater data set. First, *BACKGROUND* groundwater data should adequately represent *BACKGROUND* conditions. A sufficient number of groundwater samples should be collected over a time frame that will reflect seasonal variation (e.g., quarterly samples for two years). When *BACKGROUND* groundwater concentrations are established using upgradient wells, the number of wells should be sufficient to represent natural spatial variability.

BACKGROUND groundwater data should be closely evaluated for natural spatial variability and between-well variability before pooling data from upgradient wells for calculation of *BACKGROUND* concentrations. A *BACKGROUND* data set should represent a single population that can be described by a single statistical distribution (e.g., a normal distribution with a mean of 3.6 and a standard deviation of 0.78). *BACKGROUND* groundwater data should, therefore, be evaluated to determine if more than one population or distribution is present (e.g., due to the presence of multiple aquifers or significant spatial variability across an aquifer). This evaluation can be completed through a general review of the data or by using graphical techniques such as probability plots, side-by-side box plots, and/or stiff diagrams. It is generally not appropriate to combine data from different aquifers or from significantly differing areas, if present within an aquifer. Inappropriately combining data across wells can result in inflated *BACKGROUND* concentrations due to the large overall variability.

BACKGROUND groundwater data should be independent. In general, this means that: 1) the data should be *RANDOMLY* obtained, and 2) there should be no trends in the data. For groundwater monitoring purposes, samples are typically collected at fixed time intervals from wells which have been located based on professional judgment. Consequently, groundwater samples are not generally *RANDOM* with respect to space or time. However, it is important that

time intervals between samples be spaced far enough apart, considering groundwater flow rates, that groundwater samples are independent (i.e., the same groundwater is not sampled each time). Furthermore, data used to establish *BACKGROUND* groundwater concentrations should be non-trending. Use of trending data to establish *BACKGROUND* groundwater concentrations will result in inflated *BACKGROUND* concentrations.

Once *BACKGROUND* groundwater concentrations have been established, it may be necessary to reevaluate and/or update the *BACKGROUND* concentrations on a periodic basis. Revisions to *BACKGROUND* groundwater concentrations must be *PROPOSED* for DEQ approval.

The statistical methods described in Chapter 1 (evaluating statistical distributions) and Chapter 2 (evaluating for outliers) of this tabbed section are also appropriate for evaluating *BACKGROUND* groundwater data. In addition, the methods presented in Section 4.3.1 of Chapter 4 for calculating upper limits based on *FACILITY-SPECIFIC BACKGROUND* data are generally appropriate and can be used for calculating *BACKGROUND* groundwater concentrations.

Additional information and detail will be forthcoming in future statistics training materials. Any questions or reviews regarding statistic analyses of groundwater data for Part 201 applications should be submitted to the Department Statistician.

CHAPTER 5: CENSORED DATA

Contaminant concentrations often lie below the detection limits of the equipment used to estimate the concentrations of the contaminants. For these samples, instead of reporting an actual concentration, the detection limit of the equipment is typically reported (e.g., less than 7 ppb). The true concentration at the sampled location is likely below this detection limit, but how far below the limit is left to speculation. These types of data are termed **censored**.

Censored data can be rather difficult to deal with statistically. As one person put it, "censored data is like holding on to a tail and being unsure whether the other end is connected to an elephant or a mouse." This analogy gets at the heart of the difficulty that censored data presents. As you might imagine, if we have a five quantified observations above and five censored observations below some detection limit, then determining a "good" estimate of the mean, variance, or even the type of distribution (e.g., normal or lognormal) can be difficult.

Censored data can either be singly or multiply censored. Singly censored means that the data set contains only one censoring level or detection limit (e.g., < 7 ppb). A multiply-censored data set contains multiple censoring levels or detection limits (e.g., data may be reported at < 7 ppb, < 5 ppb, and < 3 ppb). A singly-censored data set is much easier to deal with than a multiply censored data set.

Recommendation

A simple approach for dealing with censored data has been to simply replace the censored observations with half of their respective detection limits. This method was found to perform adequately when calculating upper confidence limits (UCLs) for the mean concentration based on simulation studies conducted during the preparation of the draft document, Statistical Analysis of Ground-Water Monitoring Data at RCRA *FACILITIES* – Unified Guidance.

After substituting 1/2 of the detection limit for concentrations below the limit of detection, proceed with the analysis as usual by evaluating the distribution of the resulting data set (Chapter 1), examining for outliers (Chapter 2), and calculating UCLs for the mean when appropriate (Chapter 3).

This approach is recommended only for data sets with less than 50% nondetects.

When calculating a *FACILITY-SPECIFIC BACKGROUND* concentration as the mean plus three standard deviations for comparison to *FACILITY* data (Chapter 4), 1/2 of the detection limit may also be substituted for *FACILITY-SPECIFIC BACKGROUND* concentrations below the detection limit up to 50% nondetect. For greater than 50% nondetects, nonparametric upper tolerance limits may be used to obtain a *FACILITY-SPECIFIC BACKGROUND* concentration. As described by EPA (1992c), the upper limit may be established as the maximum concentration in the *FACILITY-SPECIFIC BACKGROUND* data set.

When conducting two-sample tests, such as those used to compare two populations (i.e., *FACILITY* concentrations and *FACILITY-SPECIFIC BACKGROUND* concentrations), nonparametric procedures should be used in place of parametric procedures when the percent of concentrations below the detection limit is greater than 15% in either data set, as recommended by EPA (1992c). The nonparametric procedure recommended in Chapter 4 is the Wilcoxon Rank Sum Test followed by the Quantile Test.

Alternate Methods

Several alternate methods are available for estimating the mean and standard deviation of censored data sets. Two of these are Cohen's method and Aitchison's method (EPA, 1992c). Cohen's method is based on the assumption that the censored observations belong to the same continuous distribution as the detected observations, except that they are censored at the detection limit (Cohen, 1991). Aitchison's method is based on the assumption that censored observations actually represent zero concentrations and detected values arise from some other distribution (Aitchison, 1955). These methods may be *PROPOSED*; however, it is necessary to evaluate the underlying assumptions and justify the choice between the two methods.

Censored and detects-only probability plots may be used to determine if Cohen's method or Aitchison's method is more appropriate for a given data set. These plots are described in detail in EPA (1992c). If the censoring appears to be of the Aitchison type, the data set should be evaluated further to determine whether two populations were being sampled (i.e., one represented by zero values and another represented by detected values) before proceeding with Aitchinson's method for handling nondetects. This could be achieved by examining the spatial distribution of the samples and their associated values.

Additional statistical methods for dealing with censored data are available and may be *PROPOSED* on a case-by-case basis, where departmental approval of a response action is being sought. These methods are generally too complex to implement using spreadsheet software. EnvironmentalStats for S-PLUS has several built-in functions that can be used to estimate a mean, variance, or even a UCL for the mean based on a singly or multiply censored data set.

Unfortunately, there are only a few studies that examine the performance of the various methods for calculating UCLs for the mean with censored data. Schmee et al. (1985) provides exact confidence limits, using maximum likelihood, for the parameters of a normal or lognormal distribution for various sample sizes and levels of single censoring. Millard and Neerchal (2001) provide the results of simulations examining confidence interval coverage probabilities using various sample sizes, levels of censoring, types of distributions, and methods of estimation using the functions within EnvironmentalStats for S-PLUS. As mentioned earlier, this software package contains several methods for calculating a UCL for the mean with censored data. Consult with a statistician to choose among the available methods and determine which is most appropriate for the situation at hand.

SOLID AND HAZARDOUS WASTE CHARACTERIZATION

SOLID AND HAZARDOUS WASTE CHARACTERIZATION

1.0 WASTE CHARACTERIZATION

For the purpose of waste characterization, it is assumed that there is no knowledge of the origin, process of generation or history of the material to be characterized and that the characteristics and waste class need to be determined through sampling and analysis. If sufficient knowledge of the material exists, sampling and/or statistical analysis may be reduced or eliminated.

The considerations presented here apply to any waste characterization required under the NREPA (Part 111, for Hazardous Waste Management or Part 115, Solid Waste Management). The appropriate DEQ staff must be consulted for an inertness designation for solid wastes (Part 115) and de-list petitions for listed hazardous wastes (Part 111). Other methods of determining waste classifications are available, but must be reviewed and approved by the appropriate DEQ staff.

Before statistical methods can be considered for comparing data to the appropriate criteria, the nature (waste process origin) and extent (the horizontal and vertical extent of the material) of the waste media in question, must first be defined. This information will aid in designing the sampling strategy to adequately characterize the material (i.e., identify if there is a single, homogeneous population that can be described by a single statistical distribution or whether there are *HOT SPOTS* present). Refer to the tabbed section titled, "Sampling Strategies" for more details on sampling options.

Each unique waste type must be characterized independently due to the heterogeneous nature of the waste material and the potential for stratification.

In general, compositing of samples is not recommended under any circumstances, but for volatile organic sampling it is never acceptable. Use of composite samples should be discussed with the appropriate DEQ staff.

2.0 IDENTIFICATION OF SAMPLING STRATEGIES

When obtaining samples to characterize a material, it is important to insure that the analytical results obtained will provide an accurate estimation of the nature of the entire area/volume under consideration. The location and number of samples to be taken of a particular material depend on many factors: the degree of accuracy desired, the spatial and temporal variability of the material to be sampled, and the costs involved. An important objective in any sampling program is to obtain the most accurate data possible while minimizing the associated costs. One method to accomplish this goal is to use statistically valid sampling strategies. The appropriate sample number can be estimated and the sampling intervals (which may be based on volume, location depth or time) can be chosen without bias.

A combination of sampling strategies may be necessary when characterizing wastes. The nature and extent of the material (boundaries) must be adequately defined. Biased sampling should generally be used to determine whether the material is **homogeneous** (i.e. origin with little variability), **heterogeneous** (i.e. origin with a great deal of variability) and/or **stratified**. Each unique waste type must be characterized independently. Biased sampling results may be sufficient to characterize a waste material depending on: 1) whether the biased sampling results adequately represent the waste materials, and 2) whether any individual sample result

exceeds the regulatory threshold of interest. If one or more samples exceed the regulatory threshold, a more extensive sampling may be pursued to apply a statistical analysis to compare the new characterization data to the regulatory threshold.

If a statistical analysis is to be used to compare characterization data to a regulatory threshold, it will be necessary to reevaluate data needs. Additional data will often be necessary for the purpose of calculating an upper confidence limit (UCL) for the mean concentration. These data should be obtained through statistical or *RANDOM* sampling.

If several materials on a site are under investigation, it is advisable to evaluate them separately. This is especially true if information does not exist to indicate that the materials contain similar constituents or that they were placed at the same time period.

2.1 Biased Sampling

"Biased" sampling strategies generally involve use of judgment to collect samples from areas most likely to contain contamination. Often, biased sampling is utilized for smaller areas (e.g., less than a 1/4 acre) and/or smaller volumes of materials. However, biased sampling also plays a role when characterizing large volumes of materials. In this case, biased sampling may be used to initially determine the homogeneity or heterogeneity of waste materials.

Use of biased sampling is premised on enough detailed information on which to base selection of sample locations. The sample locations are purposefully chosen based on the goal of identifying localized areas or volumes in which contaminant concentrations are elevated (i.e., *HOT SPOTS*). With sufficient knowledge of existing conditions, historic activities, or field indicators (e.g., visual, olfactory, or field screening instrumentation), these areas can be focused on reliably.

Any biased sampling plan requires use of professional judgment. A thorough justification must be documented for each sample location explaining the rationale used to select the location. Without this important detail, biased sampling alone will not be adequate.

Analytical results from biased sampling must generally be compared to regulatory thresholds on a point-by-point basis. A statistical analysis of data generated from biased sampling is generally not appropriate. This is due to the underlying assumptions of most statistical methods used to compare characterization data to regulatory thresholds. One underlying assumption is that the data being evaluated were obtained through *RANDOM* sampling of a single, homogeneous population that can be described by a single statistical distribution (e.g., a normal distribution with a mean of 3.6 and a standard deviation of 0.78). Biased sampling can be used to help identify if this is the case or if differing populations (e.g., *HOT SPOTS*) are present.

If statistical sampling is completed in addition to biased sampling, it may be appropriate to combine analytical results from the statistical sampling with some or even all of the biased sampling results in a statistical analysis. However, there are several key considerations which must first be addressed, as described in Section 3.2.

Biased sampling strategies require collection of discrete soil samples. Compositing of samples is not accepted without prior DEQ approval.

2.1.1 Sampling Process Streams

Although sampling is generally thought to occur on a pile of material or over an area of treated soil, other schemes are possible. Process wastes are commonly sampled at the point of generation. This is the preferred method, since it is most representative of the material under study. The lack of exposure to elements that might cause chemical degradation and/or leaching will result in material most indicative of actual conditions. A sampling point along the material conveyor that can be fairly easily and safely reached should be chosen. It should be in an area where the entire belt can be accessed for sampling. Sampling intervals can be assigned by specific process intervals, by generated volume (see Table 2.1 below) or by a timed interval that would represent key process time/volume.

When biased sampling is conducted for the purpose of waste characterization, it should account for process variability. Ideally, the entire active time of the waste process stream would be represented in the sampling scheme, but selected times and volumes should be identified for specific sampling (i.e., to provide a basis for biased sampling).

2.1.2 Sampling Ex Situ/Waste Piles

The number of samples to be collected from a waste pile should be based on the volume of the waste pile. The following table provides recommended numbers of samples for waste characterization using biased sampling.

Table 2.1 Number of Biased Samples for Waste Piles

Volume (cubic yards)	0-25	26-100	101-500	501-1,000	1,001-2,000	> 2,000
Number of samples (depending on basis of bias)	3-4	6-8	8-10	10-12	13-15	15 + 3 for every additional 500 cubic yards

Specific considerations for biasing sample locations are described in detail in Section 1.2.1 of the tabbed section titled, “Sampling Strategies” under *FACILITY* characterization. The fundamental approaches for biasing sample location are basically the same; however, the biased sampling is now focused on waste characterization.

2.2 Statistical Sampling

Statistical sampling, also referred to as unbiased or probabilistic sampling, is based on the theory of *RANDOM* chance probabilities in order to choose samples which are representative of a given area or volume. The probability of selecting any sampling location is equal. Because sampler bias is not of concern, the error in data accuracy of a *RANDOM* sampling scheme can be objectively measured. Furthermore, knowledge of the waste distribution is not always necessary depending on the purpose for collecting samples.

In some cases it is preferable to choose statistical sampling strategies over biased sampling strategies since they can be used to produce increased data accuracy while eliminating sampler bias. This will depend on the amount of available information with which to bias sampling. In some cases, a combination of approaches will yield the most comprehensive information.

Several statistical sampling strategies can be used to produce an unbiased, representative sampling program. The principles behind the three basic types of *RANDOM* sampling and the situations for which they are best suited are described below. To achieve true *RANDOM* sampling, composite sampling is not acceptable.

1. ***Simple RANDOM sampling*** is a method that requires little or no prior knowledge of material distribution. It relies on *RANDOM* chance probability theory – where each sampling location has an equal and known probability of being selected. In this way, sampling error can be accurately estimated. Often, the area of interest is sectioned into a two- or three-dimensional grid pattern and *RANDOM* coordinates are chosen for sampling.

2. ***Systematic RANDOM sampling*** is an extension of simple *RANDOM* sampling that may produce a more efficient sampling survey. It can be more efficient by reducing the sampling error while maintaining the sample number, or by reducing the number of samples needed to achieve a specified sampling error, or by reducing the cost of collection. This method also requires little or no knowledge about the waste distribution, but bias and imprecision can be introduced if unseen trends or cycles exist. Two methods used to select sample locations under this method follow.

    A) *RANDOMLY* select a transect or transects and sample at pre-selected intervals.

    B) Pre-select both the transect or transects and the sampling interval and starting from a *RANDOMLY* selected point. This is the method used most throughout this chapter.

3. ***Stratified RANDOM sampling*** requires some knowledge about the waste distribution. When stratification is known or suspected, sampling efficiency can be improved by dividing the material into strata that are more homogeneous than the total area. Simple or systematic *RANDOM* sampling techniques can then be used to sample each stratum independently. Each stratum is divided into a grid pattern and the sampling points are selected *RANDOMLY*. If the area is vertically stratified, the sampling points in each stratum are selected *RANDOMLY* and then selected depths are sampled. If the area is horizontally stratified, the sampling points within each stratum are selected *RANDOMLY*, but the total depth is sampled. An analysis of variance (ANOVA) may be done on the analytical results to determine if the strata differ significantly. This can help evaluate whether use of stratified *RANDOM* sampling was necessary and statistically valid. When the volumes of the strata differ or the number of samples within each stratum differs, the results must be weighed appropriately to avoid bias if the data are to be combined in order to draw conclusions.

Of these methods, Systematic *RANDOM* sampling is generally recommended. This approach is often the most efficient since it involves collection of samples at equal intervals, simplifying the collection of samples. Furthermore, systematic *RANDOM* sampling can serve many purposes. When samples are collected on a regular grid interval, conclusions can be drawn about the size of a *HOT SPOT* likely to be identified (or missed), as described in Section 2.1 of the tabbed section titled, “Sampling Strategies.” Furthermore, because the sampling locations are *RANDOMIZED* through systematic *RANDOM* sampling (i.e., by *RANDOMLY* selecting transects and/or the initial

sampling point), it may be appropriate to include the analytical results in a statistical analysis for the purpose of comparison to regulatory thresholds. However, the considerations described in Section 3.2 of this tabbed section must first be addressed.

The number of samples needed to conduct a statistical evaluation will depend on the volume of the material being characterized and the homogeneity of the material. In all cases, a population sufficiency test must be completed and passed to satisfy the number of samples required for the statistical test used. See Section 4.0 of this tabbed section and Section 3.2 of the tabbed section titled, “Statistical Methods.”

For further discussion on sampling strategies and sample collection methods, see "Test Methods for Evaluating Solid Waste," SW846 Volume II: Field Methods, November 1986, Third Edition, USEPA.

2.2.1 Sampling a Process Stream

As noted in Section 2.1.1, process wastes are commonly sampled at the point of generation. This is the preferred method, since it is most representative of the material under study. The lack of exposure to elements that might cause chemical degradation and/or leaching will result in material most indicative of actual conditions. A sampling point along the material conveyor that can be fairly easily and safely reached should be chosen. It should be in an area where the entire belt can be accessed for sampling.

When statistical sampling is conducted to reflect temporal rather than spatial variability, time strata should be established over the course of the process day. The time strata should be established to account for process variability. Ideally, the entire active time of the line should be included in the sampling scheme. Once time strata are chosen, *RANDOM* numbers can be selected to establish sampling times. When the appropriate sampling time arrives, the identified material would be collected.

For example, in a four hour period, a point somewhere on the table would be chosen and numbers greater than 0 but less than 240 would be selected until the number of samples for that strata were obtained. The *RANDOM* numbers would represent time in minutes. These numbers would be added to the starting time for each stratum to determine the time of sampling.

The Microsoft Excel function =RANDBETWEEN(L,U) can be used to select a *RANDOM* number between a specified range of numbers where L is the lower bound of the range and U is the upper bound. In the example above, the function =RANDBETWEEN(0,240) would be used.

If the time strata chosen are of unequal lengths, the number of samples chosen from any one stratum should reflect the percentage contribution that stratum makes to the time frame as a whole. If, for example, a 24 hour operating time is divided such that stratum one is four hours and stratum two is eight hours, stratum two should have twice as many samples as stratum one.

When the appropriate sampling time arrives, the material from the conveyor belt point that had been identified would be sampled.

2.2.2 Sampling Ex Situ/Waste Piles

This section assumes that an adequate investigation has been completed and the volume of the material to be characterized has been confirmed to represent a single, homogeneous population

that can be described by a single statistical distribution (e.g., a normal distribution with a mean of 3.6 and a standard deviation of 0.78).

Samples collected from a waste pile or some ex situ soils may need to be located using a three dimensional gridding method. Characterizations involving soils, piles and/or wastes with a significant vertical component should be evaluated in **three dimensions** (volume evaluation). Examples of such characterizations would be ex situ soil characterization or waste pile characterization involving several cubic feet of soil and/or waste. A grid would be superimposed on the area and a vertical component added at each node.

Establishing Grid Intervals

Statistical sampling strategies often employ the use of gridding to facilitate the unbiased selection of sampling points and the use of accepted statistical tools for evaluating the resultant data.

The equations and tables provided in Section 2.2.1.2 of the tabbed section titled, "Sampling Strategies" can be used to identify grid intervals. Use of these methods facilitate unbiased selection of sampling points and sample coverage proportional to the area of the pile or material being characterized. Other methods can be *PROPOSED* for DEQ approval.

Sampling of the Grid

A minimum of nine samples or 25%, whichever is greater, of the total grid stations should be sampled and analyzed initially to allow a large enough data pool for the statistical analysis. Extra samples should be taken and kept under proper chain of custody and handling procedures at the time of initial sampling. If the statistical analysis indicates that two or three more samples are needed, an additional trip to the field may not be necessary. This may also avoid the need to reestablish the grid pattern at a later date.

Stratified *RANDOM* Sampling

Stratified *RANDOM* Sampling requires some knowledge about the waste distribution. When stratification is known or suspected, sampling efficiency can be improved by dividing the material into strata that are more homogeneous than the total area. Simple or systematic *RANDOM* sampling techniques can then be used to sample each stratum independently.

If the area is vertically stratified, the sampling points in each stratum are selected *RANDOMLY* and then selected depths are sampled. Typically, layers in a stratified material will be between two and five feet thick. Any material over five feet thick must be vertically (as well as horizontally) sampled.

If the area is horizontally stratified, the sampling points within each stratum should be selected *RANDOMLY*, but the total depth sampled. An analysis of variance (ANOVA) may be done on the analytical results to determine if the strata differ significantly. This would assure that the use of stratified *RANDOM* sampling was statistically valid. When the volume of the strata differs or the number of samples within each stratum differs, the results must be weighed appropriately to avoid bias.

3.0 COMPARISON TO REGULATORY THRESHOLDS

Regulatory thresholds for solid waste characterization are found in R 299.4115. Regulatory thresholds for the hazardous waste toxicity characteristic characterization are found in

Part 111 R 299.9217. To determine if environmental media contain a listed hazardous waste, characterization data should be compared to the Type B criteria under the NREPA as found in MERA Memo 8: Revision 3.

3.1 Comparing Characterization Data to Regulatory Thresholds on a Point-by-Point Basis

Analytical data generated using biased sampling strategies must generally be compared to regulatory thresholds on a point-by-point basis. If all samples are below the regulatory threshold and sampling of the waste materials is believed to be adequately representative, no additional samples are warranted. Adequate documentation of all sample locations must be provided based on sufficient knowledge of the waste materials.

If one or more samples contain contaminant concentrations above regulatory thresholds, additional sampling may be necessary to better understand the exceedance (e.g., if the waste materials were expected to be homogeneous and no exceedances were expected), and possibly a change in sampling strategy from biased to statistical. Vertical and horizontal delineation of elevated concentrations in the area of the biased sample location may be necessary.

Statistical analyses of data from biased sampling is generally not appropriate. This is due to the underlying assumptions of most statistical methods used to compare *FACILITY* data to cleanup criteria. One underlying assumption is that the data being evaluated were obtained through *RANDOM* sampling of a single population that can be described by a single statistical distribution (e.g., a normal distribution with a mean of 3.6 and a standard deviation of 0.78).

However, if statistical sampling is completed, use of a statistical analysis to compare characterization data to regulatory thresholds may be appropriate. It is first necessary to evaluate the considerations described below in Section 3.2.

3.2 Comparing Characterization Data to Regulatory Thresholds Using Statistics

Before assembling a data set and conducting a statistical analysis, it is important to give careful consideration to the following:

Population to be sampled

Adequate knowledge of contaminant distribution and the presence of *HOT SPOTS* is essential due to assumptions underlying the statistical methods used to compare characterization data to regulatory thresholds (i.e., 95% UCLs for the mean concentration). Adherence to these assumptions is necessary if an accurate statistical conclusion is to be drawn. One key assumption is that the data are independently and identically distributed (iid). For this assumption to be true, the following are generally necessary:

- Samples must be *independent* and representative of the area included in the analysis. In statistical terms, this means that the data were collected *RANDOMLY*.

- For the data to be identically distributed, each data point must have been drawn from the same *identical* statistical distribution (e.g., a normal distribution with a mean of 3.6 ppm and a standard deviation of 0.78 ppm). Data from a *HOT SPOT* area would be represented by a different statistical distribution than data from non-*HOT SPOT* areas. In

other words, the mean concentration in a *HOT SPOT* area would be higher than in the non-*HOT SPOT* areas and the standard deviation would likely differ as well.

Consequently, identification of *HOT SPOTS* is necessary before an appropriate statistical analysis can be conducted.

Once defined, *HOT SPOTS* should not be included in a statistical analysis for comparison to regulatory thresholds. *HOT SPOTS* must be characterized separately. This is necessary to avoid combining data from different statistical distributions and violating the assumptions of the statistical methods.

Selecting the Appropriate Statistical Method for Comparison to Criteria

A UCL for the mean should be utilized to compare characterization data to regulatory thresholds. Various methods are available for calculating UCLs for the mean concentration. These methods are presented in Chapter 3 of the tabbed section titled, "Statistical Methods."

Selection of the appropriate method for calculating a UCL for the mean requires an evaluation of the assumptions underlying each method. One of these assumptions is the statistical distribution of the data set (i.e., normal, lognormal, or neither). Consequently, each data set must be evaluated for the best-fitting statistical distribution. Chapter 1 of the tabbed section titled, "Statistical Methods" provides several techniques to accomplish this task. As described in Chapter 1, these techniques should be used in combination to best evaluate the statistical distribution.

Chapter 2 of the tabbed section titled, "Statistical Methods" provides techniques for identifying whether suspect data points are statistical outliers. Recommendations for treatment of outliers, once identified, are also provided in Chapter 2.

For the purpose of hazardous waste determinations, a 10% level of significance ($\alpha = 0.10$) may be used. For solid waste determinations, a significance level of 5% ($\alpha = 0.05$) must be used.

4.0 STATISTICAL VALIDATION

It is suggested that a minimum of one sample be collected for any container (55 gallon drum or smaller). A minimum of nine *RANDOMLY* located, discrete samples should be collected per layer/stratification for waste piles or ex situ material if statistics are to be used to compare characterization data to regulatory thresholds. The rationale for this minimum number of samples is described in Section 1.3 of the tabbed section titled, "Statistical Methods."

Further evaluation of the sample size used for waste characterization must be conducted on the data to show that a sufficient number of samples had been collected to use the chosen statistical method. Section 3.2 of the tabbed section titled, "Statistical Methods" provides statistical methods to address this issue, including the Lambda Test.

ACRONYMS / GLOSSARY

Tab 9.0 Acronyms / Glossary

ACRONYMS

ANOVA	Analysis of Variance
BCCs	Bioaccumulative Contaminants of Concern
BEA	Baseline Environmental Assessment
CCTM	Cleanup Criteria Training Material
CFR	Code of Federal Regulations
Csat	Soil Saturation Concentration
CV	Coefficient of Variation
DCC	Direct Contact Criteria
DNAPL	Dense Non-Aqueous Phase Liquids
DWC	Drinking Water Criteria
EU	Exposure Units
FIELDS	Fully-Integrated Environmental Location Decision Support
GCC	Groundwater Contact Criteria
GI	Grid Interval
GNP	Generally Not Practical
GPS	Global Positioning System
GSI	Groundwater Surface Water Interface
GVIIC	Groundwater Volatilization to Indoor Air Inhalation Criteria
IQR	Interquartile Range
LCL	Lower Confidence Limit
Ln	Log-transformed
MBSS	Michigan Background Soil Survey
MERA	Michigan Environmental Response Act
MDEQ (DEQ)	Michigan Department of Environmental Quality
MDNR (DNR)	Michigan Department of Natural Resources
MZB	Mixing Zone-Based
NDs	Non-Detects
NREPA	Natural Resources and Environmental Protection Act
ppb	parts per billion
PSIC	Particulate Soil Inhalation Criteria
RME	Reasonable Maximum Exposure
RCRA	Resource Conservation and Recovery Act

SF	Shapiro-Francia
S^3TM	Sampling Strategies and Statistics Training Materials
SVIIC	Soil Volatilization to Indoor Air Inhalation Criteria
SW	Shapiro-Wilk
TCE	Trichloroethene
U.S.EPA (EPA)	United States Environmental Protection Agency
UCL	Upper Confidence Limit
UTL	Upper Tolerance Limit
VSIC	Volatile Soil Inhalation Criteria
VSP	Visual Sampling Plan
VSR	Verification of Soil Remediation
WET	Whole Effluent Toxicity

GLOSSARY

AVERAGING AREA: The cross sectional area of the contaminated plume used to estimate the discharge rate of venting groundwater in the request for a mixing zone determination, generally the cross sectional area with concentrations greater than the generic GSI criterion.

***BACKGROUND*:** The concentration or level of a hazardous substance which exists in the environment at or regionally proximate to a site that is not attributable to any release at or regionally proximate to the site. [See Statistical Guidesheet 10 for information on each type of *BACKGROUND* described below.

> ***STATEWIDE DEFAULT BACKGROUND*:** The concentrations provided in the September 30, 1993, Michigan Environmental Response Act (MERA) Operational Memorandum #15: Default Type A Cleanup Criteria (Op Memo #15) which represent acceptable *BACKGROUND* concentrations at all *FACILITIES*. [See Appendix A and Chapter 4 of the tabbed section titled, "Statistical Methods."]
>
> ***FACILITY-SPECIFIC BACKGROUND:*** The *BACKGROUND* concentrations in soil at or adjacent to a *FACILITY*. According to Op Memo #15, it is acceptable to establish *BACKGROUND* concentrations higher than the *STATEWIDE DEFAULT BACKGROUND* concentrations for a *FACILITY*. [See Chapter 4 of the tabbed section titled, "Statistical Methods."]
>
> ***REGIONAL BACKGROUND:*** The *BACKGROUND* concentrations in soil that are found regionally proximate to a *FACILITY*. According to Op Memo #15, use of regionally proximate background values higher than the *STATEWIDE DEFAULT BACKGROUND* concentrations may be established for a *FACILITY*. *REGIONAL BACKGROUND* values must be *PROPOSED* on a case-by-case basis.
>
> ***NON-RELEASE ANTHROPOGENIC BACKGROUND:*** *BACKGROUND* concentrations that are affected by anthropogenic compounds which are excluded from the definition of a *RELEASE (*e.g., compounds present in the soil resulting from historic widespread application of pesticides for insect control if applied in accordance with label directions).

***EMISSION SOURCE SIZE/EMISSION SOURCE AREA*:** A term specific to the generic soil inhalation criteria for ambient air that is defined as the horizontal and vertical extent of soil contamination with detectable concentrations.

***EXPOSURE UNIT*:** The area over which an individual is expected to move *RANDOMLY*, such that equivalent amounts of time are assumed to be spent at each location. This area should be logical and generally regular in shape.

FACILITY: Any area, place, or property where a hazardous substance is in excess of the concentrations which satisfy the requirements of Section 20120a(1)(a) or (17) of the cleanup criteria for unrestricted residential use under Part 213, Leaking Underground Storage Tanks, of the Natural Resources and Environmental Protection Act, 1994 PA 451, as amended (NREPA), has been *RELEASED*, deposited, disposed of, or otherwise comes to be located. *FACILITY* does not include any area, place, or property at which response activities have been completed which satisfy the cleanup criteria for the residential category provided for in Section 20120a(1)(a) and (17) or at which corrective action has been completed under Part 213 which satisfies the cleanup criteria for unrestricted residential use.

***FACILITY-SPECIFIC BACKGROUND*:** See *BACKGROUND* definition.

GSI MONITORING WELLS: Vertical wells placed along and near the surface water body, no closer than the ordinary high water mark, that are used to monitor groundwater venting to surface water.

***HOT SPOT*:** Two or more adjacent sample locations in reasonably close proximity at which concentrations are sufficiently above cleanup criteria and sample concentrations from the surrounding area (i.e., spatially correlated concentrations sufficiently above criteria) to indicate that they: 1) represent a different statistical population, and 2) pose a potential risk that should not be masked by a statistical analysis. Judgment must be used to determine whether concentrations are sufficiently above cleanup criteria and surrounding location. [See Section 2.2.1.1 of the tabbed section titled, "Sampling Strategies."]

***INTERIM RESPONSES*:** Cleanup or removal of hazardous substances from the environment or performing other actions prior to the selection of a remedial action that is necessary to prevent, minimize, or mitigate injury to the public health, safety, or welfare, the environment, or natural resources, which injury might otherwise result from a release of a hazardous substance.

NON-RELEASE ANTHROPOGENIC BACKGROUND: See *BACKGROUND* definition.

***ONE ACUTE TOXIC UNIT*:** A value calculated as $100/LC_{50}$, where the LC_{50} is determined from a whole effluent toxicity (WET) test which produces a result that is statistically or graphically estimated to be lethal to 50% of the test organisms.

PROPOSAL/PROPOSE/ PROPOSED: A plan describing the use and, in some cases, justifying the applicability of statistics under the Part 201 program. *PROPOSALS* will typically be submitted for the following:

1. comparison of *FACILITY* data to Part 201 cleanup criteria,
2. development of *REGIONAL BACKGROUND* concentrations, or
3. use of statistical methods not described in the tabbed section titled, "Statistical Methods."

PROPOSALS are necessary only if departmental approval is being sought for a response activity. Self-implemented approaches using statistics to support determinations must be documented in a manner that address the three objectives described in the tabbed section titled, "Introduction."

PROPOSALS for comparison of *FACILITY* data to Part 201 cleanup criteria must include a justification for the use of statistics for this purpose that addresses the issues described in the Statistical Guidesheet corresponding to the pathway/condition. These issues generally relate to the practicality of obtaining enough data within limited areas and/or the complexity of the pathway/condition considering contaminant distribution and transport to the point of exposure. These proposals must also include a description of the statistical methods to be used in the analysis and the basis for their selection (e.g., consideration of the appropriate statistical distribution and proportion of concentrations below the detection limit).

PROPOSALS for development of *REGIONAL BACKGROUND* concentrations must include a description of the *REGIONAL BACKGROUND* data to be used as well as the statistical method to be used to develop the *REGIONAL BACKGROUND* concentrations.

PROPOSALS for the use of statistical methods not described in the tabbed section titled, "Statistical Methods" must provide a basis for use of the alternative methods (e.g., consideration of the appropriate statistical distribution and proportion of concentrations below the detection limit) as well as a reference for the alternative method and/or the actual document. Consultation with a professional statistician is advised.

RANDOM/RANDOMIZATION/RANDOMIZED/RANDOMIZING/RANDOMLY: A method utilized to sample *EXPOSURE UNITS*, excluding areas identified as *HOT SPOTS*, such that each location in the *EXPOSURE UNIT* has an equal likelihood of being sampled. [See Section 2.4.2 of the tabbed section titled, "Sampling Strategies."]

REGIONAL BACKGROUND: See *BACKGROUND* definition.

RELEASE: Part 201 Section 1(bb) states in part: "*RELEASE*" includes, but is not limited to, any spilling, leaking, pumping, pouring, emitting, emptying, discharging, injecting, escaping, leaching, dumping, or disposing of a hazardous substance into the environment, or the abandonment or discarding of barrels, containers, and other closed receptacles containing a hazardous substance. For exceptions to this definition, see Part 201 Section 1(bb).

REPRESENTATIVE CONCENTRATION: The concentration of a hazardous substance derived from a statistical analysis of *FACILITY* data obtained from sample locations representing: 1) the exposure pathway and land use category in question and 2) contaminant concentrations within a specified *EXPOSURE UNIT*. *RANDOM* sampling within the *EXPOSURE UNIT* (or other area for pathway conditions without default *EXPOSURE UNITS*) is necessary to obtain unbiased, representative samples; however, biased samples collected for the purpose of site characterization may be used in some circumstances, as described in Section 2.4.2 of the tabbed section titled, "Sampling Strategies."

SOURCE SIZE MODIFIER: A multiplier for adjusting the generic soil inhalation criteria for ambient air to correspond to the *EMISSION SOURCE SIZE/EMISSION SOURCE AREA* determined for the *FACILITY*.

***STATEWIDE DEFAULT BACKGROUND*:** See *BACKGROUND* definition.

GLOSSARY OF COMMON STATISTICAL TERMS

μ – Lowercase Greek letter pronounced "mu." Commonly used to represent the population mean of a distribution which, in practice, is usually unknown. The arithmetic mean is an estimate of μ.

σ^2 – Lowercase Greek letter sigma, squared. Commonly used to represent the population variance. The variance describes the width or spread of a distribution and is usually unknown. The sample variance, s^2, is an estimate of σ^2.

σ – Lowercase Greek letter sigma. Commonly used to representthe population standard deviation, the square root of σ^2. The standard deviation is also a measure of the width or spread of a distribution and is usually unknown. The sample standard deviation, *s*, is an estimate of σ.

Bias – the amount by which an estimate differs from the "true" parameter value.

Censored – an observation without a definitive quantity associated with it. In contaminant monitoring, typically reported as an observation below some threshold detection limit (e.g., <7 ppb).

Confidence interval – an interval used to describe the uncertainty associated with a parameter of interest. It is calculated from the sample observations and believed to contain the true parameter value a particular percentage of the time (commonly 95%).

Distribution – a function describing the probability associated with obtaining a measurable quantity of interest.

Interval estimate – two estimated values used to construct an interval, which is intended to enclose a parameter of interest. Values within the interval represent possible values for the parameter of interest.

Lognormal – a distribution characterized by a left-skewed peak and a long right tail. The shape of the distribution depends on the parameters for μ and σ^2. The natural logarithms of observations from a lognormal distribution are normally distributed. The mean and median are generally not the same for a lognormal distribution. The lognormal distribution ranges in value from 0 to $+\infty$.

Mean – the average value of a population.

Median – the "middle" value of a population. Half the population is above the median and half is below.

Normal – a symmetric, bell-shaped, probability distribution. The shape and position of the distribution depends on the parameters for μ and σ^2. Many statistical methods rely upon having normally-distributed data. The mean and median for a normal distribution are the same. The normal distribution ranges in value from $-\infty$ to $+\infty$.

Outlier – a value that has been determined to be inconsistent with an assumed distribution as determined based on the remaining data using statistical methods described in Section 2.1 of Statistical Methods. Values identified as outliers must be evaluated before determining a course of action regarding treatment of the outlier.

Point estimate – a single value, or point, that estimates a parameter of interest. For example, $\bar{x}$ is a point estimate of μ.

Potential outlier – a value that has been identified through the screening measures described in Section 2.1 of Statistical Methods. A values identified as a potential outlier must be further evaluated through formal testing before classifying it as an outlier.

Prediction interval – an interval used to describe the uncertainty associated with obtaining a single new observation.

Skewness – the amount by which a distribution is shifted away from symmetry. For example, a lognormal distribution is typically a positively-skewed (or right-skewed with a long right tail) distribution in which the median is less than the mean.

Transform – a technique for changing the scale and range of data by the application of some function to obtain better statistical properties. For example, to satisfy a method's assumption of normality, the natural log transform is often applied to lognormal data to obtain a normally distributed data set.

REFERENCES

Tab 10.0

References

REFERENCES

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

Borden, R.C. and P.B. Bedient. 1987. In-Situ Measurement of Adsorption and Biotransformation at a Hazardous Waste Site. *Water Resources Bulletin*, 23(4), 629-636.

Box, G.E.P. and D.R. Cox 1964. An Analysis of Transformations (with Discussion). *Journal of the Royal Statistical Society*, Series B 26(2), 211-252.

Chen, L. 1995. Testing the mean of skewed distributions. *Journal of the American Statistical Association*, 90:767-772. (Skewness-Adjusted t)

Cline, P.V. and S. Brown. 1989. Site Specific Evaluation of Contaminant Fate: Test Applications and Limitations. *Superfund '89*, Hazardous Materials Control Research Institute, Silver Springs, MD.

Cohen, A.C. 1991. Truncated and Censored Samples: Theory and Applications. M. Dekker, New York. 312 pp.

Conover, W.J. 1980. *Practical Nonparametric Statistics*, 2nd ed. Wiley, New York.

D'Agostino, R.B. 1971. An omnibus test of normality for moderate and large size samples. *Biometrika*, 58:341-348.

Dixon, W.J. 1953. Processing data for outliers. *Biometrics*, 9:74-89.

EPA 1986. Test Methods for Evaluating Solid Waste, SW846 Volume II: Field Methods, Third Edition.

EPA 1989a. Risk Assessment Guidance for Superfund, Volume 1 Human Health Evaluation Manual (Part A) – Interim Final.

EPA 1989b. Methods for Evaluating the Attainment of Cleanup Standards. Volume 1: Soils and Solid Media.

EPA 1989c. Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities – Interim Final Guidance.

EPA 1992a. Supplemental Guidance to RAGS: Calculating the Concentration Term.

EPA 1992b. Methods for Evaluating the Attainment of Cleanup Standards. Volume 2: Ground Water.

EPA 1992c. Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities – Addendum to Interim Final Guidance.

EPA 1993. Supplemental Guidance to RAGS: Estimating Risk from Groundwater Contamination.

EPA 1994. Statistical Training Course for Groundwater Monitoring Data Analysis. EPA/530/R-93/003. Solid Waste and Emergency Response, U.S. Environmental Protection Agency, Washington, D.C.

EPA 1996a. Soil Screening Guidance: User's Guide.

EPA 1996b. Soil Screening Guidance: Technical Background Document.

EPA 1999. Some Practical Aspects of Sample Size and Power Computations for Estimating the Mean of Positively Skewed Distributions in Environmental Applications. EPA/600/s-99/006. Office of Research and Development. U.S. Environmental Protection Agency, Washington, D.C.

EPA 2000. Guidance for Data Quality Assessment: Practical Methods for Data Analysis. EPA/600/R-96/084. Office of Environmental Information, U.S. Environmental Protection Agency, Washington, D.C.

Gibbons, R.D. 1994. *Statistical Methods for Groundwater Monitoring.* John Wiley & Sons, New York.

Gibbons, R.D. and D.E. Coleman. 2001. *Statistical Methods for Detection and Quantification of Environmental Contamination.* John Wiley & Sons, New York.

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

Grubbs, F.E. 1950. Sample Criteria for Testing Outlying Observations. *Annals of Mathematical Statistics*, 21(1):27-58.

Grubbs, F.E. 1969. Procedures for Detecting Outlying Observations in Samples. *Technometrics*, 11(1):1-21.

Grubbs, F.E. and G. Beck. 1972. Extension of sample sizes and percentage points for significance tests of outlying observations. *Technometrics*, 14(4):847-854.

Helsel, D.R. and R.M. Hirsch. 1992. *Statistical Methods in Water Resources.* Elsevier Science Publishers, B.V., Amsterdam, The Netherlands.

Land, C.E. 1971. Confidence intervals for linear functions of the normal mean and variance. *Annals of Mathematical Statistics*, 42(4):1187-1205.

Lindsay, W.L. 1979. Chemical Equilibria In Soils. John Wiley and Sons, New York.

Madansky, A. 1988. Prescriptions for Working Statisticians. Springer-Verlag, New York.

Michigan Department of Natural Resources. 1994. Verification of Soil Remediation (Revision 1).

Millard, S.P. 1998. EnrironmentalStats for S-Plus User's Manual for Widows and UNIX. Springer-Verlag, New York.

Millard, S.P. and N.K.Neerchal 2001. Environmental Statistics with S-PLUS. CRC Press LLC (CRC Applied Environmental Statistics Series), Boca Raton, Florida.

Miller, R.G. 1986. Beyond ANOVA: Basics of Applied Statistics. John Wiley & Sons, New York.

Ott, W.R. 1990. A physical explanation of the lognormality of pollutant concentrations. *Journal of the Air and Waste Management Association*, 40:1378-1383.

Parkin, T.B., S.T. Chester, and J.A. Ribinson. 1990. Calculating confidence intervals for the mean of a lognormally distributed variable. *Journal of the Soil Science Society of America*, 54:321-326.

Rosner, B. 1983. Percentage points for a generalized ESD many-outlier procedure. *Technometrics*, 25:165-172.

Schwille, F. 1988. Dense Non-Aqueous Phase Liquids In Porous and Fractured Media-Model Experiments. Lewis Publishers, Boca Raton, LA.

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

Shapiro, S.S. and R.S. Francia. 1972. An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, 63:1343-1372.

Singh, A.K., A. Singh, and M. Engelhardt. 1997. The lognormal distribution in environmental applications. EPA/600/R-97/006. Technology Support Center for Monitoring and Site Characterization, Technology Innovation Office, Office of Research and Development, Office of Solid Waste and Emergency Response, U.S. Enviornmental Protection Agency, Washington, D.C., December 1997.

Singh, A.K., A. Singh, and M. Engelhardt. 1999. Some practical aspects of sample size and power computations for estimating the mean of positively skewed distribution in environmental applications. EPA/600/s-99/006. Technology Support Center for Monitoring and Site Characterization, Technology Innovation Office, Office of Research and Development, Office of Solid Waste and Emergency Response, U.S. Environmental Protection Agency, Washington, D.C., November 1999.

Stolarski, R.S., A.J. Krueger, M.R. Schoeberl, R.D. McPeters, P.A. Newman, and J.C. Alpert. (1986). Nimbus 7 Satellite Measurements of the Spring-time Antarctic Ozone Decrease. *Nature*, 322, 808-811.

Walsh, J.E. 1958. Large sample nonparametric rejection of outlying observations. *Annals of the Institute of Statistical Mathematics*, 10:223-232.

Wilson, J.T. and B.H. Wilson. 1985. Biotransformation of Trichloroethylene in Soil. *Applied Environmental Microbiology*, 49:242-243.

Wisconsin Department of Natural Resources, Emergency and Remedial Response Section. 1993. Guidance for Design, Installation and Operation of Soil Venting Systems, July, 1993, PUBL-SW 185-93.

Tab A.0

Appendices

APPENDICES – REFERENCE MATERIALS

A. EPA GUIDANCE FOR DATA QUALITY ASSESSMENT, PRACTICAL METHODS FOR DATA ANALYSIS, EPA QA/G-9, QA00 UPDATE, JULY 2000

METHOD FOR DOUBLE LINEAR INTERPOLATION

SECTION 3.3: TESTS FOR COMPARING TWO POPULATIONS

SECTION 3.3.1: COMPARING TWO MEANS

SECTION 3.3.3: NONPARAMETRIC COMPARISONS OF TWO POPULATIONS

SECTION 4.5.4: LEVENE'S TEST FOR THE EQUALITY OF TWO OR MORE VARIANCES

B. MERA OPERATIONAL MEMORANDUM #15: DEFAULT TYPE A CLEANUP CRITERIA, SEPTEMBER 30, 1993

United States
Environmental Protection
Agency

Office of Environmental
Information
Washington, DC 20460

EPA/600/R-96/084
July, 2000

Guidance for Data Quality Assessment

Practical Methods for Data Analysis

EPA QA/G-9

QA00 UPDATE

FOREWORD

This document is the 2000 (QA00) version of the *Guidance for Data Quality Assessment* which provides general guidance to organizations on assessing data quality criteria and performance specifications for decision making. The Environmental Protection Agency (EPA) has developed a process for performing Data Quality Assessment (DQA) Process for project managers and planners to determine whether the type, quantity, and quality of data needed to support Agency decisions 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 distinctly different from other guidance documents; it is not intended to be read in a linear or continuous fashion. The intent of the document is for it 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 different problems using a systematic methodology.

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 for the Data Quality Objectives Process*
EPA QA/G-4D	*DEFT Software for the Data Quality Objectives Process*
EPA QA/G-4HW	*Guidance for the Data Quality Objectives Process for Hazardous Waste Site Investigations*
EPA QA/G-9D	*Data Quality Evaluation Statistical Toolbox (DataQUEST)*

This document is intended to be a "living document" that will be updated periodically to incorporate new topics and revisions or refinements to existing procedures. Comments received on this 2000 version will be considered for inclusion in subsequent versions. Please send your written comments on *Guidance for Data Quality Assessment* to:

Quality Staff (2811R)
Office of Environmental Information
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

Box 4-31: Double Linear Interpolation

The details of the double linear interpolation are provided to assist in the use of Table A-10 of Appendix A. The desired value for $\hat{\lambda}$ corresponds to $\gamma = 0.083$ and, $h = 0.125$ from Box 4-30, Step 3. The values from Table A-10 for interpolatation are:

γ	$h = 0.10$	$h = 0.15$
0.05	0.11431	0.17925
0.10	0.11804	0.18479

There are 0.05 units between 0.10 and 0.15 on the h-scale and 0.025 units between 0.10 and 0.125. Therefore, the value of interest lies (0.025/0.05)100% = 50% of the distance along the interval between 0.10 and 0.15. To linearly interpolate between tabulated values on the h axis for $\gamma = 0.05$, the range between the values must be calculated, 0.17925 - 0.11431 = 0.06494; the value that is 50% of the distance along the range must be computed, 0.06494 x 0.50 = 0.03247; and then that value must be added to the lower point on the tabulated values, 0.11431 + 0.03247 = 0.14678. Similarly for $\gamma = 0.10$, 0.18479 - 0.11804 = 0.06675, 0.06675 x 0.50 = 0.033375, and 0.11804 + 0.033375 = 0.151415.

On the γ-axis there are 0.033 units between 0.05 and 0.083 and there are 0.05 units between 0.05 and 0.10. The value of interest (0.083) lies (0.033/0.05 x 100) = 66% of the distance along the interval between 0.05 and 0.10, so 0.151415 - 0.14678 = 0.004635, 0.004635 * 0.66 = 0.003059. Therefore,

$$\hat{\lambda} = 0.14678 + 0.003059 = 0.149839.$$

4.7.2.2 Trimmed Mean

Trimming discards the data in the tails of a data set in order to develop an unbiased estimate of the population mean. For environmental data, nondetects usually occur in the left tail of the data so trimming the data can be used to adjust the data set to account for nondetects when estimating a mean. Developing a 100p% trimmed mean involves trimming p% of the data in both the lower and the upper tail. Note that p must be between 0 and .5 since p represents the portion deleted in both the upper and the lower tail. After np of the largest values and np of the smallest values are trimmed, there are n(1-2p) data values remaining. Therefore, the proportion trimmed is dependent on the total sample size (n) since a reasonable amount of samples must remain for analysis. For approximately symmetric distributions, a 25% trimmed mean (the midmean) is a good estimator of the population mean. However, environmental data are often skewed (non-symmetric) and in these cases a 15% trimmed mean performance may be a good estimator of the population mean. It is also possible to trim the data only to replace the nondetects. For example, if 3% of the data are below the detection limit, a 3% trimmed mean could be used to estimate the population mean. Directions for developing a trimmed mean are contained in Box 4-32 and an example is given in Box 4-33. A trimmed variance is rarely calculated and is of limited use.

4.7.2.3 Winsorized Mean and Standard Deviation

Winsorizing replaces data in the tails of a data set with the next most extreme data value. For environmental data, nondetects usually occur in the left tail of the data. Therefore, winsorizing can be used to adjust the data set to account for nondetects. The mean and standard deviation can then be computed on the new data set. Directions for winsorizing data (and revising the sample size) are contained in Box 4-34 and an example is given in Box 4-35

TABLE A-10: VALUES OF THE PARAMETER $\hat{\lambda}$ FOR COHEN'S ESTIMATES ADJUSTING FOR NONDETECTED VALUES

γ	h .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

γ	h .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

The concept of a confidence interval can be shown by a simple example. Suppose a stable situation producing data without any anomalies was sampled many times. Each time the sample was taken, the mean and standard deviation was calculated from the sample and a confidence interval constructed using the method of Box 3-10.

Box 3-12: Directions for a Confidence Interval for a Mean for Simple and Systematic Random Samples

Let $X_1, X_2, ..., X_n$ represent a sample of size n from a population of normally distributed values.

Step 1: Use the directions in Box 2-2 to calculate the sample mean, $\overline{X}$. Use the directions in Box 2-3 to calculate the sample standard deviation, s.

Step 2: Use Table A-1 of Appendix A to find the critical value $t_{1-\alpha/2}$ such that 100(1-α/2)% of the t distribution with n - 1 degrees of freedom is below $t_{1-\alpha/2}$. For example, if α = 0.10 and n = 16, then n-1 = 15 and $t_{1-\alpha/2}$ = 1.753.

Step 3: The (1-α)100% confidence interval is: $\overline{X} - \frac{t_{1-\alpha/2}s}{\sqrt{n}}$ *to* $\overline{X} + \frac{t_{1-\alpha/2}s}{\sqrt{n}}$

Box 3-13: An Example of a Confidence Interval for a Mean for a Random or Systematic Random Samples

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: 8.1, 7.9, 7.9. 8.2, 8.2, 8.0,7.9. The directions in Box 3-12 will be used to develop a 95% confidence interval for the mean.

Step 1: Using Box 2-2, $\overline{X}$ =8.03. Use Box 2-3, s=0.138.

Step 2: Using Table A-1 of Appendix A and 6 *degrees of freedom,* $t_{1-\alpha/2}$ = 2.447.

Step 3: The (1-α)100% confidence interval is:

$$8.03 - \frac{2.447x0.138}{\sqrt{7}} \text{ to } 8.03 + \frac{2.447x0.138}{\sqrt{7}} \text{ or 7.902 to 8.158.}$$

3.3 TESTS FOR COMPARING TWO POPULATIONS

A two-sample test involves the comparison of two populations or a "before and after" comparison. 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. The comparison of the two populations may be based on a statistical parameter that characterizes the relative location (e.g., a mean or median), or it may be based on a distribution-free comparison of the two population distributions. Tests that do not assume an underlying distributions (e.g., normal or lognormal) are called distribution-free or nonparametric tests. These tests are often more useful for comparing two populations than those that assume a specific distribution because they make less stringent assumptions. Section 3.3.1 covers tests for differences in the means of two populations. Section 3.3.2 covers tests for differences in the

proportion or percentiles of two populations. Section 3.3.3 describes distribution-free comparisons of two populations. Section 3.3.4 describes tests for comparing two medians.

Often, a two-sample test involves the comparison of the difference of two population parameters to a threshold value. For environmental applications, the threshold value is often zero, representing the case where the data are used to determine which of the two population parameters is greater than the other. For example, concentration levels from a Superfund site may be compared to a background site. Then, if the Superfund site levels exceed the background levels, the site requires further investigation. A two-sample test may also be used to compare readings from two instruments or two separate populations of people.

If the exact same sampling locations are used for both populations, then the two samples are not independent. This case should be converted to a one-sample problem by applying the methods described in Section 3.2 to the differences between the two populations at the same location. For example, one could compare contaminant levels from several wells after treatment to contaminant levels from the same wells before treatment. The methods described in Section 3.2 would then be applied to the differences between the before and after treatment contaminant levels for each well.

3.3.1 Comparing Two Means

Let μ_1 represent the mean of population 1 and μ_2 represent the mean of population 2. The hypotheses considered in this section are:

Case 1: H_0: $\mu_1 - \mu_2 \leq \delta_0$ vs. H_A: $\mu_1 - \mu_2 > \delta_0$; and

Case 2: H_0: $\mu_1 - \mu_2 \geq \delta_0$ vs. H_A: $\mu_1 - \mu_2 < \delta_0$.

An example of a two-sample test for population means is comparing the mean contaminant level at a remediated Superfund site to a background site; in this case, δ_0 would be zero. Another example is a Record of Decision for a Superfund site which specifies that the remediation technique must reduce the mean contaminant level by 50 ppm each year. Here, each year would be considered a separate population and δ_0 would be 50 ppm.

The information required for these tests includes the null and alternative hypotheses (either Case 1 or Case 2); the gray region (i.e., a value $\delta_1 > \delta_0$ for Case 1 or a value $\delta_1 < \delta_0$ for Case 2 representing the bound of the gray region); the false rejection error rate α at δ_0; the false acceptance error rate β at δ_1; and any additional limits on decision errors. It may be helpful to label additional false rejection error limits as α_2 at $\delta_{\alpha 2}$, α_3 at $\delta_{\alpha 3}$, etc., and to label additional false acceptance error limits as β_2 at $\delta_{\beta 2}$, β_3 at $\delta_{\beta 3}$, etc.

3.3.1.1 Student's Two-Sample t-Test (Equal Variances)

PURPOSE

Student's two-sample t-test can be used to compare two population means based on the independent random samples $X_1, X_2, \ldots, X_m$ from the first population, and $Y_1, Y_2, \ldots, Y_n$ from the second population. This test assumes the variabilities (as expressed by the variance) of the two populations are approximately equal. If the two variances are not equal (a test is described in Section 4.5), use Satterthwaite's t test (Section 3.3.1.2).

ASSUMPTIONS AND THEIR VERIFICATION

The principal assumption required for the two-sample t-test is that a random sample of size m ($X_1, X_2, \ldots, X_m$) is drawn from population 1, and an independent random sample of size n ($Y_1, Y_2, \ldots, Y_n$) is drawn from population 2. Validity of the random sampling and independence assumptions should be confirmed by reviewing the procedures used to select the sampling points.

The second assumption required for the two-sample t-tests are that the sample means $\overline{X}$ (sample 1) and $\overline{Y}$ (sample 2) are approximately normally distributed. If both m and n are large, one may make this assumption without further verification. For small sample sizes, approximate normality of the sample means can be checked by testing the normality of each of the two samples.

LIMITATIONS AND ROBUSTNESS

The two-sample t-test with equal variances is robust to violations of the assumptions of normality and equality of variances. However, if the investigator has tested and rejected normality or equality of variances, then nonparametric procedures may be applied. The t-test is not robust to outliers because sample means and standard deviations are sensitive to outliers.

SEQUENCE OF STEPS

Directions for the two-sample t-test for a simple random sample and a systematic simple random sample are given in Box 3-14 and an example in Box 3-15.

3.3.1.2 Satterthwaite's Two-Sample t-Test (Unequal Variances)

Satterthwaite's t-test should be used to compare two population means when the variances of the two populations are not equal. It requires the same assumptions as the two-sample t-test (Section 3.3.1.1) except the assumption of equal variances.

Directions for Satterthwaite's t-test for a simple random sample and a systematic simple random sample are given in Box 3-16 and an example in Box 3-17.

Box 3-14: Directions for the Student's Two-Sample t-Test (Equal Variances) for Simple and Systematic Random Samples

This describes the steps for applying the two-sample t-tests for differences between the population means when the two population variances are equal for Case 1 (H_0: $\mu_1 - \mu_2 \leq \delta_0$). Modifications for Case 2
(H_0: $\mu_1 - \mu_2 \geq \delta_0$) are given in parentheses { }.

STEP 1: Calculate the sample mean $\overline{X}$ and the sample variance s_X^2 for sample 1 and compute the sample mean $\overline{Y}$ and the sample variance s_Y^2 for sample 2.

STEP 2: Use Section 4.5 to determine if the variances of the two populations are equal. If the variances of the two populations are not equal, use Satterthwaite's t test (Section 3.3.1.2). Otherwise, compute the pooled standard deviation

$$s_E = \sqrt{\frac{(m-1)s_X^2 + (n-1)s_Y^2}{(m-1)+(n-1)}}.$$

STEP 3: Calculate $t = \dfrac{\overline{X} - \overline{Y} - \delta_0}{s_E\sqrt{1/n + 1/m}}$.

Use Table A-1 of Appendix A to find the critical value $t_{1-\alpha}$ such that 100(1-α)% of the t-distribution with (m+n-2) degrees of freedom is below $t_{1-\alpha}$.

If $t > t_{1-\alpha}$ {$t < -t_{1-\alpha}$}, the null hypothesis may be rejected. Go to Step 5.

If $t \not> t_{1-\alpha}$ {$t \not< -t_{1-\alpha}$}, there is not enough evidence to reject the null hypothesis. Therefore, the false acceptance error rate will need to be verified. Go to Step 4.

STEP 4: To calculate the power of the test, assume that the true values for the mean and standard deviation are those obtained in the sample and use a statistical software package like the DEFT software (EPA, 1994) or the DataQUEST software (EPA, 1996) to generate the power curve of the two-sample t-test. If only one false acceptance error rate (β) has been specified (at δ_1), it is possible to calculate the sample size which achieves the DQOs, assuming the true mean and standard deviation are equal to the values estimated from the sample, instead of calculating the power of the test. Calculate

$$m^* = n^* = \frac{2s^2(z_{1-\alpha} + z_{1-\beta})^2}{(\delta_1 - \delta_0)^2} + (0.25)z_{1-\alpha}^2$$

If $m^* \leq m$ and $n^* \leq n$, the false acceptance error rate has been satisfied. Otherwise, the false acceptance error rate has not been satisfied.

STEP 5: The results of the test could be:

1) the null hypothesis was rejected, and it seems $\mu_1 - \mu_2 > \delta_0$ {$\mu_1 - \mu_2 < \delta_0$};

2) the null hypothesis was not rejected, the false acceptance error rate was satisfied, and it seems $\mu_1 - \mu_2 \leq \delta_0$ {$\mu_1 - \mu_2 \geq \delta_0$}; or

3) the null hypothesis was not rejected, the false acceptance error rate was not satisfied, and it seems $\mu_1 - \mu_2 \leq \delta_0$ {$\mu_1 - \mu_2 \geq \delta_0$}, but this conclusion is uncertain because the sample size

Box 3-15: An Example of a Student's Two-Sample t-Test (Equal Variances) for Simple and Systematic Random Samples

At a hazardous waste site, area 1 (cleaned using an in-situ methodology) was compared with a similar (but relatively uncontaminated) reference area, area 2. If the in-situ 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. Seven random samples were taken from area 1, and eight were taken from area 2. Because the contaminant concentrations in the two areas are supposedly equal, the null hypothesis is H_0: $\mu_1 - \mu_2 \leq 0$ (Case 1). 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 ppb.

STEP 1:

	Sample Mean	Sample Variance
Area 1	7.8 ppm	2.1 ppm^2
Area 2	6.6 ppm	2.2 ppm^2

STEP 2: Methods described in Section 4.5 were used to determine that the variances were essentially equal. Therefore,

$$s_E = \sqrt{\frac{(7-1)2.1 + (8-1)2.2}{(7-1)+(8-1)}} = 1.4676$$

STEP 3:

$$t = \frac{7.8 - 6.6 - 0}{1.4676\sqrt{1/7+1/8}} = 1.5798$$

Table A-1 of Appendix A was used to find that the critical value $t_{0.95}$ with (7 + 8 - 2) = 13 degrees of freedom is 1.771.

Because $t \not> t_{1-\alpha}$ (i.e., 1.5798 $\not>$ 1.771), there is not enough evidence to reject the null hypothesis. The false acceptance error rate will need to be verified.

STEP 4: Assuming the true values for the mean and standard deviation are those obtained in the sample:

$$m^* = n^* = \frac{2(1.4676^2)(1.645+0.842)^2}{(2.5-0)^2} + (0.25)1.645^2 = 4.938, \text{ i.e., } 5.$$

Because $m^* \leq$ m (7) and $n^* \leq$ n (8), the false acceptance error rate has been satisfied.

STEP 5: The null hypothesis was not rejected and the false acceptance error rate was satisfied. Therefore, it seems there is no difference between the two areas and that the in-situ methodology worked as expected.

Box 3-16: Directions for Satterthwaite's t-Test (Unequal Variances) for Simple and Systematic Random Samples

This describes the steps for applying the two-sample t-test for differences between the population means for Case 1 (H_0: $\mu_1 - \mu_2 \leq \delta_0$). Modifications for Case 2 (H_0: $\mu_1 - \mu_2 \geq \delta_0$) are given in parentheses { }.

STEP 1: Calculate the sample mean $\overline{X}$ and the sample variance s_X^2 for sample 1 and compute the sample mean $\overline{Y}$ and the sample variance s_Y^2 for sample 2.

STEP 2: Using Section 4.5, test whether the variances of the two populations are equal. If the variances of the two populations are not equal, compute:

$$s_{NE} = \sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}}$$

If the variances of the two populations appear approximately equal, use Student's two-sample t-test (Section 3.3.1.1, Box 3-14).

STEP 3: Calculate $t = \dfrac{\overline{X} - \overline{Y} - \delta_0}{s_{NE}}$

Use Table A-1 of Appendix A to find the critical value $t_{1-\alpha}$ such that 100(1-α)% of the t-distribution with f degrees of freedom is below $t_{1-\alpha}$, where

$$f = \frac{\left[\dfrac{s_X^2}{m} + \dfrac{s_Y^2}{n}\right]^2}{\dfrac{s_X^4}{m^2(m-1)} + \dfrac{s_Y^4}{n^2(n-1)}}$$

(Round f down to the nearest integer.)

If $t > t_{1-\alpha}$ {$t < -t_{1-\alpha}$}, the null hypothesis may be rejected. Go to Step 5.

If $t \not> t_{1-\alpha}$ {$t \not< -t_{1-\alpha}$}, there is not enough evidence to reject the null hypothesis and therefore, the false acceptance error rate will need to be verified. Go to Step 4.

STEP 4: If the null hypothesis (H_0) was not rejected, calculate either the power of the test or the sample size necessary to achieve the false rejection and false acceptance error rates. To calculate the power of the test, assume that the true values for the mean and standard deviation are those obtained in the sample and use a statistical software package to generate the power curve of the two-sample t-test. A simple method to check on statistical power does not exist.

STEP 5: The results of the test could be:

1) the null hypothesis was rejected, and it seems $\mu_1 - \mu_2 > \delta_0$ {$\mu_1 - \mu_2 < \delta_0$};

2) the null hypothesis was not rejected, the false acceptance error rate was satisfied, and it seems $\mu_1 - \mu_2 \leq \delta_0$ {$\mu_1 - \mu_2 \geq \delta_0$}; or

3) the null hypothesis was not rejected, the false acceptance error rate was not satisfied, and it seems $\mu_1 - \mu_2 \leq \delta_0$ {$\mu_1 - \mu_2 \geq \delta_0$}, but this conclusion is uncertain because the sample

Box 3-17: An Example of Satterthwaite's t-Test (Unequal Variances) for Simple and Systematic Random Samples

At a hazardous waste site, area 1 (cleaned using an in-situ methodology) was compared with a similar (but relatively uncontaminated) reference area, area 2. If the in-situ 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. Seven random samples were taken from area 1, and eight were taken from area 2. Because the contaminant concentrations in the two areas are supposedly equal, the null hypothesis is H_0: $\mu_1 - \mu_2 \leq 0$ (Case 1). 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 ppb.

STEP 1:

	Sample Mean	Sample Variance
Area 1	9.2 ppm	1.3 ppm^2
Area 2	6.1 ppm	5.7 ppm^2

STEP 2: Using Section 4.5, it was determined that the variances of the two populations were not equal, and therefore using Satterthwaite's method is appropriate:

$$s_{NE} = \sqrt{1.3/7 + 5.7/8} = 0.9477$$

STEP 3:

$$t = \frac{9.2 - 6.1 - 0}{0.9477} = 3.271$$

Table A-1 was used with f degrees of freedom, where

$$f = \frac{[1.3/7 + 5.7/8]^2}{\frac{1.3^2}{7^2(7-1)} + \frac{5.7^2}{8^2(8-1)}} = 10.307 \ (i.e., \ 10 \ degrees \ of \ freedom)$$

(recall that f is rounded down to the nearest integer), to find $t_{1-\alpha} = 1.812$.

Because $t > t_{0.95}$ (3.271 > 1.812), the null hypothesis may be rejected.

STEP 5: Because the null hypothesis was rejected, it would appear there is a difference between the two areas (area 1 being more contaminated than area 2, the reference area) and that the in-situ methodology has not worked as intended.

3.3.2 Comparing Two Proportions or Percentiles

This section considers hypotheses concerning two population proportions (or two population percentiles); for example, one might use these tests to compare the proportion of children with elevated blood lead in one urban area compared with the proportion of children with elevated blood lead in another area. The population proportion is the ratio of the number of elements in a subset of the total population to the total number of elements, where the subset has some specific characteristic that the rest of the elements do not. A population percentile represents the percentage of elements of a population having values less than some threshold value C.

3.3.3 Nonparametric Comparisons of Two Populations

In many cases, assumptions on distributional characteristics are difficult to verify or difficult to satisfy for both populations. In this case, several distribution-free test procedures are available that compare the shape and location of the two distributions instead of a statistical parameter (such as a mean or median). The statistical tests described below test the null hypothesis "H_0: the distributions of population 1 and population 2 are identical (or, the site is not more contaminated than background)" versus the alternative hypothesis "H_A: part of the distribution of population 1 is located to the right of the distribution of population 2 (or the site is more contaminated than background)." Because of the structure of the hypothesis tests, the labeling of populations 1 and 2 is of importance. For most environmental applications, population 1 is the area of interest (i.e., the potentially contaminated area) and population 2 is the reference area.

There is no formal statistical parameter of interest in the hypotheses stated above. However, the concept of false rejection and false acceptance error rates still applies.

3.3.3.1 The Wilcoxon Rank Sum Test

PURPOSE

The Wilcoxon rank sum test can be used to compare two population distributions based on m independent random samples $X_1, X_2, \ldots, X_m$ from the first population, and n independent random samples $Y_1, Y_2, \ldots, Y_n$ from the second population. When applied with the Quantile test (Section 3.3.3.2), the combined tests are most powerful for detecting true differences between two population distributions.

ASSUMPTIONS AND THEIR VERIFICATION

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 the same shape and dispersion, so that one distribution differs by some fixed amount (or is increased by a constant) when compared to the other distribution. For large samples, to test whether both site distributions have approximately the same shape, one can create and compare histograms for the samples.

LIMITATIONS AND ROBUSTNESS

The Wilcoxon rank sum test may produce misleading results if many data values are the same. When values are the same, their relative ranks are the same, and this has the effect of diluting the statistical power of the Wilcoxon rank sum test. Estimated concentrations should be reported for data below the detection limit, even if these estimates are negative, because their relative magnitude to the rest of the data is of importance. An important advantage of the Wilcoxon rank sum test is its partial robustness to outliers, because the analysis is conducted in

terms of rankings of the observations. This limits the influence of outliers because a given data point can be no more extreme than the first or last rank.

SEQUENCE OF STEPS

Directions and an example for the Wilcoxon rank sum test are given in Box 3-20 and Box 3-21. However, if a relatively large number of samples have been taken, it is more efficient in terms of statistical power to use a large sample approximation to the Wilcoxon rank sum test (Box 3-22) to obtain the critical values of W.

Box 3-20: Directions for the Wilcoxon Rank Sum Test for Simple and Systematic Random Samples

Let $X_1, X_2, \ldots, X_n$ represent the n data points from population 1 and $Y_1, Y_2, \ldots, Y_m$ represent the m data points from population 2 where both n and m are less than or equal to 20. For Case 1, the null hypothesis will be that population 1 is shifted to the left of population 2 with the alternative that population 1 is either the same as or shifted to the right of population 2; Case 2 will be that population 1 is shifted to the right of population 2 with the alternative that population 1 is the same as or shifted to the left of population 2; for Case 3, the null hypothesis will be that there is no difference between the two populations and the alternative hypothesis will be that population 1 is shifted either to the right or left of population 2. If either m or n are larger than 20, use Box 3-22.

STEP 1: List and rank the measurements from both populations from smallest to largest, keeping track of which population contributed each measurement. The rank of 1 is assigned to the smallest value, the rank of 2 to the second smallest value, and so forth. If there are ties, assign the average of the ranks that would otherwise have been assigned to the tied observations.

STEP 2: Calculate R as the sum of the ranks of the data from population 1, then calculate

$$W = R - \frac{n(n+1)}{2}.$$

STEP 3: Use Table A-7 of Appendix A to find the critical value w_α (or $w_{\alpha/2}$ for Case 3). For Case 1, reject the null hypothesis if $W > nm - w_\alpha$. For Case 2, reject the null hypothesis if $W < w_\alpha$. For Case 3, reject the null hypothesis if $W > nm - w_{\alpha/2}$ or $W < w_{\alpha/2}$. If the null hypothesis is rejected, go to Step 5. Otherwise, go to Step 4.

STEP 4: If the null hypothesis (H_0) was not rejected, the power of the test or the sample size necessary to achieve the false rejection and false acceptance error rates should be calculated. For small samples sizes, these calculations are too complex for this document.

STEP 5: The results of the test could be:

1) the null hypothesis was rejected and it seems that population 1 is shifted to the right (Case 1), to the left (Case 2) or to the left or right (Case 3) of population 2.

2) the null hypothesis was not rejected and it seems that population 1 is shifted to the left (Case 1) or to the right (Case 2) of population 2, or there is no difference between the two populations (Case 3).

Box 3-21: An Example of the Wilcoxon Rank Sum Test for Simple and Systematic Random Samples

At a hazardous waste site, area 1 (cleaned using an in-situ methodology) was compared with a similar (but relatively uncontaminated) reference area, area 2. If the in-situ 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 null hypothesis will be that area 1 is shifted to the right of area 2 and the alternative hypothesis will be that there is no difference between the two areas or that area 1 is shifted to the left of area 2 (Case 2). The false rejection error rate was set at 10% 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 eight samples were taken from area 2:

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

STEP 1: The data listed and ranked by size are (Area 1 denoted by *):

Data (ppb): 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*

STEP 2: R = 3.5 + 8 + 9.5 + 9.5 + 12 + 14 + 15 = 71..5. W = 71.5 - 7(7 + 1)/2 = 43.5

STEP 3: Using Table A-7 of Appendix A, $\alpha = 0.10$ and $W_{\alpha} = 17$. Since 43.5 > 17, do not reject the null hypothesis.

STEP 4: The null hypothesis was not rejected and it would be appropriate to calculate the probable power of the test. However, because the number of samples is small, extensive computer simulations are required in order to estimate the power of this test which is beyond the scope of this guidance.

STEP 5: The null hypothesis was not rejected. Therefore, it is likely that there is no difference between the investigated area and the reference area, although the statistical power is low due to the small sample sizes involved.

Box 3-22: Directions for the Large Sample Approximation to the Wilcoxon Rank Sum Test for Simple and Systematic Random Samples

Let $X_1, X_2, \ldots, X_n$ represent the n data points from population 1 and $Y_1, Y_2, \ldots, Y_m$ represent the m data points from population 2 where both n and m are greater than 20. For Case 1, the null hypothesis will be that population 1 is shifted to the left of population 2 with the alternative that population 1 is the same as or shifted to the right of population 2; for Case 2, the null hypothesis will be that population 1 is shifted to the right of population 2 with the alternative that population 1 is the same as or shifted to the left of population 2; for Case 3, the null hypothesis will be that there is no difference between the populations and the alternative hypothesis will be that population 1 is shifted either to the right or left of population 2.

STEP 1: List and rank the measurements from both populations from smallest to largest, keeping track of which population contributed each measurement. The rank of 1 is assigned to the smallest value, the rank of 2 to the second smallest value, and so forth. If there are ties, assign the average of the ranks that would otherwise have been assigned to the tied observations.

STEP 2: Calculate W as the sum of the ranks of the data from population 1.

STEP 3: Calculate $w_p = \frac{mn}{2} + Zp\sqrt{mn(n + m + 1)/12}$ where p = 1 - α for Case 1, p = α for Case 2, and z_p is the p^{th} percentile of the standard normal distribution (Table A-1 of Appendix A). For Case 3, calculate both $w_{\alpha/2}$ (p = α/2) and $w_{1-\alpha/2}$ (p = 1 - α/2).

STEP 4: For Case 1, reject the null hypothesis if $W > w_{1-\alpha}$. For Case 2, reject the null hypothesis if $W < w_{\alpha}$. For Case 3, reject the null hypothesis if $W > w_{1-\alpha/2}$ or $W < w_{\alpha/2}$. If the null hypothesis is rejected, go to Step 6. Otherwise, go to Step 5.

STEP 5: If the null hypothesis (H_0) was not rejected, calculate either the power of the test or the sample size necessary to achieve the false rejection and negative error rates. If only one false acceptance error rate (β) has been specified (at δ_1), it is possible to calculate the sample size that achieves the DQOs, assuming the true mean and standard deviation are equal to the values estimated from the sample, instead of calculating the power of the test. If m and n are large, calculate:

$$m^* = n^* = \frac{2s^2(z_{1-\alpha} + z_{1-\beta})^2}{(\delta_1 - \delta_0)^2} + (0.25)z_{1-\alpha}^2$$

where z_p is the p^{th} percentile of the standard normal distribution (Table A-1 of Appendix A). If $1.16m^* \leq m$ and $1.16n^* \leq n$, the false acceptance error rate has been satisfied.

STEP 6: The results of the test could be:

1) the null hypothesis was rejected, and it seems that population 1 is shifted to the right (Case 1), to the left (Case 2) or to the left or right (Case 3) of population 2.

2) the null hypothesis was not rejected, the false acceptance error rate was satisfied, and it seems that population 1 is shifted to the left (Case 1) or to the right (Case 2) of population 2, or there is no difference between the two populations (Case 3).

3) the null hypothesis was not rejected, the false acceptance error rate was not satisfied, and it seems that population 1 is shifted to the left (Case 1) or to the right (Case 2) of population

3.3.3.2 The Quantile Test

PURPOSE

The Quantile test can be used to compare two populations based on the independent random samples $X_1, X_2, \ldots, X_m$ from the first population and $Y_1, Y_2, \ldots, Y_n$ from the second population. When the Quantile test and the Wilcoxon rank sum test (Section 3.3.3.1) are applied together, the combined tests are the most powerful at detecting true differences between two populations. The Quantile test is useful in detecting instances where only parts of the data are different rather than a complete shift in the data. It essentially looks at a certain number of the largest data values to determine if too many data values from one population are present to be accounted for by pure chance.

ASSUMPTIONS AND THEIR VERIFICATION

The Quantile test assumes that the data $X_1, X_2, \ldots, X_m$ are a random sample from population 1, and the data $Y_1, Y_2, \ldots, Y_n$ are a random sample from population 2, and the two random samples are independent of one another. The validity of the random sampling and independence assumptions is assured by using proper randomization procedures, either random number generators or tables of random numbers. The primary verification required is to review the procedures used to select the sampling points. The two underlying distributions are assumed to have the same underlying dispersion (variance).

LIMITATIONS AND ROBUSTNESS

The Quantile test is not robust to outliers. In addition, the test assumes either a systematic (e.g., a triangular grid) or simple random sampling was employed. The Quantile test may not be used for stratified designs. In addition, exact false rejection error rates are not available, only approximate rates.

SEQUENCE OF STEPS

The Quantile test is difficult to implement by hand. Therefore, directions are not included in this guidance but the DataQUEST software (EPA, 1996) can be used to conduct this test. However, directions for a modified Quantile test that can be implemented by hand are contained in Box 3-23 and an example is given in Box 3-24.

Box 3-23: Directions for a Modified Quantile Test for Simple and Systematic Random Samples

Let there be 'm' measurements from population 1 (the reference area or group) and 'n' measurement from population 2 (the test area or group). The Modified Quantile test can be used to detect differences in shape and location of the two distributions. For this test, the significance level (α) can either be approximately 0.10 or approximately 0.05. The null hypothesis for this test is that the two population are the same (i.e., the test group is the same as the reference group) and the alternative is that population 2 has larger measurements than population 1 (i.e., the test group has larger values than the reference group).

STEP 1:	Combine the two samples and order them from smallest to largest keeping track of which sample a value came from.
STEP 2:	Using Table A-13 of Appendix A, determine the critical number (C) for a sample size n from the reference area, sample size m from the test area using the significance level α. If the C^{th} largest measurement of the combined population is the same as others, increase C to include all of these tied values.
STEP 3:	If the largest C measurements from the combined samples are all from population 2 (the test group), then reject the null hypothesis and conclude that there are differences between the two populations. Otherwise, the null hypothesis is not rejected and it appears that there is no difference between the two populations.

3.3.4 Comparing Two Medians

Let $\tilde{\mu}_1$ represent the median of population 1 and $\tilde{\mu}_2$ represent the median of population 2. The hypothesis considered in this section are:

Case 1: H_0: $\tilde{\mu}_1 - \tilde{\mu}_2 \leq \delta_0$ vs. H_A: $\tilde{\mu}_1 - \tilde{\mu}_2 > \delta_0$; and

Case 2: H_0: $\tilde{\mu}_1 - \tilde{\mu}_2 \geq \delta_0$ vs. H_A: $\tilde{\mu}_1 - \tilde{\mu}_2 < \delta_0$.

An example of a two-sample test for the difference between two population medians is comparing the median contaminant level at a Superfund site to the median of a background site. In this case, δ_0 would be zero.

The median is also the 50^{th} percentile, and, therefore, the methods described in Section 3.3.2 for percentiles and proportions may be used to test hypotheses concerning the difference between two medians by letting $P_1 = P_0 = 0.50$. The Wilcoxon rank sum test (Section 3.3.3.1) is also recommended for comparing two medians. This test is more powerful than those for proportions for symmetric distributions.

to departures from normality. With long-tailed distributions, the test too often rejects equality (homogeneity) of the variances.

Bartlett's test requires the calculation of the variance for each sample, then calculation of a statistic associated with the logarithm of these variances. This statistic is compared to tables and if it exceeds the tabulated value, the conclusion is that the variances differ as a complete set. It does *not* mean that one is significantly different from the others, nor that one or more are larger (smaller) than the rest. 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 homogeneity of variance (testing for differences among the dispersions of several groups). 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. However, Levene's test is more difficult to apply than Bartlett's test since it involves applying an analysis of variance (ANOVA) to the absolute deviations from the group means. Directions and an example of Levene's test are contained in Box 4-26 and Box 4-27, respectively.

Box 4-24: Directions for Bartlett's Test

Consider k groups with a sample size of n_i for each group. Let N represent the total number of samples, i.e., let $N = n_1 + n_2 + \ldots + n_k$. For example, consider two wells where 4 samples have been taken from well 1 and 3 samples have been taken from well 2. In this case, k = 2, $n_1 = 4$, $n_2 = 3$, and N = 4 + 3 = 7.

STEP 1: For each of the k groups, calculate the sample variances, s_i^2 (Section 2.2.3).

STEP 2: Compute the pooled variance across groups: $$s_p^2 = \frac{1}{(N-k)} \sum_{i=1}^{k} (n_i - 1) s_i^2$$

STEP 3: Compute the test statistic: $$TS = (N - k) \ln(s_p^2) - \sum_{i=1}^{k} (n_i - 1) \ln(s_i^2)$$

where "ln" stands for natural logarithms.

STEP 4: Using a chi-squared table (Table A-8 of Appendix A), find the critical value for χ^2 with (k-1) degrees of freedom at a predetermined significance level. For example, for a significance level of 5% and 5 degrees of freedom, $\chi^2 = 11.1$. If the calculated value (*TS*) is greater than the tabulated value, conclude that the variances are not equal at that significance level.

Box 4-26: Directions for Levene's Test

Consider k groups with a sample size of n_i for the ith group. Let N represent the total number of samples, i.e., let $N = n_1 + n_2 + \ldots + n_k$. For example, consider two wells where 4 samples have been taken from well 1 and 3 samples have been taken from well 2. In this case, k = 2, $n_1 = 4$, $n_2 = 3$, and N = 4 + 3 = 7.

STEP 1: For each of the k groups, calculate the group mean, $\bar{x}_i$ (Section 2.2.2), i.e., calculate:

$$\bar{X}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} x_{1j}, \quad \bar{X}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} x_{2j}, \quad \ldots, \quad \bar{X}_k = \frac{1}{n_k}\sum_{j=1}^{n_k} x_{kj}.$$

STEP 2: Compute the absolute residuals $z_{ij} = |X_{ij} - \bar{X}_i|$ where X_{ij} represents the j^{th} value of the i^{th} group. For each of the k groups, calculate the means, $\bar{z}_i$, of these residuals, i.e., calculate:

$$\bar{z}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} z_{1j}, \quad \bar{z}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} z_{2j}, \quad \ldots, \quad \bar{z}_k = \frac{1}{n_k}\sum_{j=1}^{n_k} z_{kj}.$$

Also calculate the overall mean residual as $\bar{z} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} z_{ij} = \frac{1}{N}\sum_{i=1}^{k} n_i \bar{z}_i$.

STEP 3: Compute the following sums of squares for the absolute residuals:

$$SS_{TOTAL} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} z_{ij}^2 - \frac{\bar{z}}{N}, \quad SS_{GROUPS} = \sum_{i=1}^{k}\frac{\bar{z}_i^2}{n_i} - \frac{\bar{z}}{N}, \text{ and } SS_{ERROR} = SS_{TOTAL} - SS_{GROUPS}.$$

STEP 4: Compute $f = \dfrac{SS_{GROUPS}/(k-1)}{SS_{ERROR}/(N-k)}$

STEP 5: Using Table A-9 of Appendix A, find the critical value of the F-distribution with (k-1) numerator degrees of freedom, (N-k) denominator degrees of freedom, and a desired level of significance (α). For example, if $\alpha = 0.05$, the numerator degrees of freedom is 5, and the denominator degrees of freedom is 18, then using Table A-9, F = 2.77. If f is greater than F, reject the assumptions of equal variances.

Box 4-27: An Example of Levene's Test

Four months of data on arsenic concentration 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 will be used to make this determination.

STEP 1: The group mean for each well ($\bar{x}_i$) is shown in the last row of the table below.

Arsenic Concentration (ppm)						
Month	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6
1	22.90	2.00	2.0	7.84	24.90	0.34
2	3.09	1.25	109.4	9.30	1.30	4.78
3	35.70	7.80	4.5	25.90	0.75	2.85
4	4.18	52.00	2.5	2.00	27.00	1.20
Group Means	$\bar{x}_1$=16.47	$\bar{x}_2$=15.76	$\bar{x}_3$=29.6	$\bar{x}_4$=11.26	$\bar{x}_5$=13.49	$\bar{x}_6$=2.29

STEP 2: To compute the absolute residuals z_{ij} in each well, the value 16.47 will be subtracted from Well 1 data, 15.76 from Well 2 data, 29.6 from Well 3 data, 11.26 from Well 4 data, 13.49 from Well 5 data, and 2.29 from Well 6 data. The resulting values are shown in the following table with the new well means ($\bar{z}_i$) and the total mean $\bar{z}$.

Residual Arsenic Concentration (ppm)						
Month	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6
1	6.43	13.76	27.6	3.42	11.41	1.95
2	13.38	14.51	79.8	1.96	12.19	2.49
3	19.23	7.96	25.1	14.64	12.74	0.56
4	12.29	36.24	27.1	9.26	13.51	1.09
Residual Means	$\bar{z}_1$=12.83	$\bar{z}_2$=18.12	$\bar{z}_3$=39.9	$\bar{z}_4$=7.32	$\bar{z}_5$=12.46	$\bar{z}_6$=1.52
Total Residual Mean $\bar{z}$ = (1/6)(12.83 + 18.12 + 39.9 + 7.32 + 12.46 + 1.52) = 15.36						

STEP 3: The sum of squares are: SS_{TOTAL} = 6300.89, SS_{WELLS} = 3522.90, and SS_{ERROR} = 2777.99.

STEP 4: $$f = \frac{SS_{WELLS}/(k-1)}{SS_{ERROR}/(N-k)} = \frac{3522.9/(6-1)}{2777.99/(24-6)} = 4.56$$

STEP 5: Using Table A-9 of Appendix A, the F statistic for 5 and 18 degrees of freedom with α = 0.05 is 2.77. Since f=4.56 exceeds $F_{.05}$=2.77, the assumption of equal variances should be rejected.

MICHIGAN DEPARTMENT OF NATURAL RESOURCES

INTEROFFICE COMMUNICATION

September 30, 1993

TO: Environmental Response Division Staff

FROM: Alan J. Howard, Chief, Environmental Response Division

SUBJECT: MERA Operational Memorandum #15: Default Type A Cleanup Criteria

In order to facilitate cleanup decisions at sites at which naturally occurring metals may be of concern, the following acceptable default Type A soil cleanup criteria have been established. These values are based on analysis of the database for the Michigan Background Soil Survey (April 1991) which is maintained by Waste Management Division (WMD). They represent the mean plus one standard deviation for WMD data from combined clay, topsoil, and sand categories. The values are presented in two significant figures. Data should be rounded to two significant figures for comparison.

Table 1: ACCEPTABLE DEFAULT VALUES TYPE A SOIL CLEANUP CRITERIA

Substance	Acceptable Concentration(mg/Kg)	Substance	Acceptable Concentration(mg/Kg)
Aluminum	6900	Iron	12000
Arsenic	5.8	Mercury	0.13
Barium	75	Lithium	9.8
Cadmium	1.2	Manganese	440
Cobalt	6.8	Nickel	20
Chromium	18	Lead	21
(total)	32	Selenium	0.41
Copper	0.39	Silver	1.0
Cyanide		Zinc	47

The default values apply as follows:

1. If measured concentrations at a site do not exceed the values listed in Table 1, site specific samples to establish background are not required.

2. The values apply to all soil types, statewide.

3. It is acceptable to establish site-specific background concentration higher than the default values. Such sampling should be conducted according to requirements in existence before the issuance of this memorandum. Comparison of site values is made against the mean plus three standard deviations calculated from background samples as

provided for in existing ERD guidance regarding verification of soil remediation.

4. Staff also may approve Type A cleanups based on a regionally proximate background value higher than the default values. Comparison should be made as in #3, above.

This memorandum is intended to provide guidance to Division staff to foster consistent application of the Michigan Environmental Response Act (1982 PA 307, as amended) and the Administrative Rules promulgated thereunder. This document is not intended to convey any rights to any parties or create any duties or responsibilities under law. This document and matters addressed herein are subject to revision.

Any questions about this memorandum should be directed to Bill Iversen at 517-373-0907.

rev. 0
cc: Dennis Drake, Air Quality Division
Bob Miller, Surface Water Quality Division
Tom Segall, Geological Survey Division
Jim Sygo, Waste Management Division