
APPENDIX D
STATISTICAL METHODS

TABLE OF CONTENTS

		Page No.
	Table of Contents	i
	List of Tables	ii
	List of Figures	iii
D1.0	INTRODUCTION	D-1
D1.1	Basic Statistics	D-2
D1.2	Hypothesis Testing	D-6
D1.3	Experimental Design	D-7
D2.0	BIOLOGICAL EFFECTS	D-9
D2.1	Tier III Water Column Toxicity Tests	D-9
D2.1.1	Comparison of 100% Elutriate and Dilution Water	D-9
D2.1.1.1	Methods	D-9
D2.1.1.2	Analysis of Example Data	D-16
D2.1.2	Calculating Median Lethal Concentration	D-20
D2.1.2.1	Methods For Calculating LC₅₀	D-22
D2.1.2.2	Analysis of Example Data	D-24
D2.2	Tier III Benthic Toxicity Tests	D-25
D2.2.1	Methods	D-27
D2.2.2	Analyses of Example Data	D-29
D3.0	BIOACCUMULATION	D-36
D3.1	Tier III Single-Time Point Laboratory Bioaccumulation Study	D-36
D3.1.1	Comparisons with a Reference Sediment	D-37
D3.1.1.1	Less Than Detection Limit Data	D-43
D3.1.2	Comparisons with an Action Level	D-46
D3.1.2.1	Less Than Detection Limit Data	D-49
D3.2	Tier IV Time-Sequenced Laboratory Bioaccumulation Study	D-50
D3.2.1	Calculating Steady-State Concentrations	D-51
D3.2.2	Comparison with Reference Sediments and Action Levels	D-56
D3.3	Steady-State Bioaccumulation from Field Data	D-59
D4.0	SAS PROGRAMS AND OUTPUT FOR EXAMPLE DATA	D-59
D4.1	Program WATTOX.SAS for Water Column Toxicity Test Data Analysis	D-60
D4.1.1	WATTOX.SAS Program Statements	D-60
D4.1.2	WATTOX.SAS Program Output	D-62
D4.2	Program BENTOX.SAS for Benthic Toxicity Test Data Analysis	D-64
D4.2.1	BENTOX.SAS Program Statements	D-65
D4.2.2	BENTOX.SAS Program Output	D-68
D4.3	Program BIOACC.SAS for Single-Time Point Bioaccumulation Test Data Analysis ..	D-73
D4.3.1	BIOACC.SAS Program Statements	D-73
D4.3.2	BIOACC.SAS Program Output	D-78
D4.4	Program BIOACCSS.SAS for Time-Sequenced Bioaccumulation Test Data Analysis ..	D-85
D4.4.1	BIOACCSS.SAS Program Statements	D-86
D4.4.2	BIOACCSS.SAS Program Output	D-91
D4.5	SAS Program Statements for Censored Data Methods	D-106
D4.5.1	SAS Statements for DL, DL/2 and ZERO	D-107
D4.5.1.1	SAS Program Output for DL, DL/2, or ZERO	D-107
D4.5.2	SAS Statements for UNIF	D-107
D4.5.2.1	SAS Program Output for UNIF	D-108

D4.5.3	SAS Statements for LR	D-109
D4.5.3.1	SAS Program Output for LR	D-110
D5.0	REFERENCES	D-111

LIST OF TABLES

	Page No.
Table D-1.	Types of Errors in Hypothesis Testing and Associated Probabilities. D-7
Table D-2.	Suggested α Levels to Use for Tests of Assumptions. D-12
Table D-3.	Number of Survivors in a Hypothetical Water Column Toxicity Test After 96 h. D-18
Table D-4.	Tests of Assumptions and Hypothesis Tests on Arcsine-Transformed Water Column Toxicity Test Example Data. D-18
Table D-5.	Calculated LC_{50} Values for Example Water Column Toxicity Test Data. D-25
Table D-6.	Number of Survivors in a Hypothetical Benthic Toxicity Test. D-30
Table D-7.	Tests of Assumptions and Parametric Tests of Hypotheses on Arcsine-Transformed Benthic Toxicity Test Example Data. D-33
Table D-8.	Tests of Assumptions and Nonparametric Hypothesis Tests on Benthic Toxicity Test Example Data Converted to Rankits and Ranks. D-34
Table D-9.	Results from a Hypothetical Single-Time Point Bioaccumulation Test, Showing Contaminant Concentrations ($\mu\text{g/g}$) in Tissues of Animals Exposed to Different Treatments. D-37
Table D-10.	Tests of Assumptions and Parametric Hypothesis Tests on Untransformed and Log_{10} -Transformed Bioaccumulation Example Data. D-41
Table D-11.	Tests of Assumptions and Nonparametric Hypothesis Tests on Bioaccumulation Example Data Converted to Rankits and Ranks. D-42
Table D-12.	Recommended Censored Data Methods for Small Samples to be Used in Statistical Comparisons. D-44
Table D-13.	Results from a Hypothetical Time-Sequenced Bioaccumulation Test, Showing Contaminant Concentrations ($\mu\text{g/g}$) in Tissues of Animals Exposed to Different Treatments. D-52
Table D-14.	Regression Parameters Estimated from Example Time-Sequenced Bioaccumulation Data. D-54
Table D-15.	Tests of Assumptions and Parametric Hypothesis Tests on Untransformed Steady-State Bioaccumulation Example Data. D-57

LIST OF FIGURES

	Page No.
Figure D-1. Water Column Toxicity Test Decision Tree.	D-17
Figure D-2. LC ₅₀ Decision Tree.	D-21
Figure D-3. Probit Plot of Water Column Toxicity Test Example Data.	D-26
Figure D-4A. Benthic Toxicity Test Decision Tree (Parametric Tests).	D-31
Figure D-4B. Benthic Toxicity Test Decision Tree (Nonparametric Tests).	D-32
Figure D-5A. Bioaccumulation Test Decision Tree (Parametric Tests).	D-38
Figure D-5B. Bioaccumulation Test Decision Tree (Nonparametric Tests).	D-39
Figure D-6. Comparison of Mean Dredged Sediment Contaminant Tissue Levels (mean) and 95% Upper Confidence Level (<i>UCL</i>) with Hypothetical Action Level.	D-48
Figure D-7. Comparison of Mean Dredged Sediment Contaminant Steady-State Tissue Levels (<i>C_{ss}</i>) (mean) and 95% Upper Confidence Levels (<i>UCL</i>) with Hypothetical Action Level.	D-58
Figure D-8. Plot of Time-Sequenced Bioaccumulation Reference Sediment Example Data by Replicate.	D-94
Figure D-9. Plot of Time-Sequenced Bioaccumulation Dredged Sediment 1 Example Data by Replicate.	D-95
Figure D-10. Plot of Time-Sequenced Bioaccumulation Dredged Sediment 2 Example Data by Replicate.	D-96
Figure D-11. Plot of Time-Sequenced Bioaccumulation Dredged Sediment 3 Example Data by Replicate	D-97

APPENDIX D - STATISTICAL METHODS

D1.0 INTRODUCTION

This Appendix presents the appropriate statistical methods for analyzing data from toxicity and bioaccumulation tests. The methodology is not intended to be exhaustive, nor is it intended to be a "cook-book" approach to data analysis. Statistical analyses are routine only under ideal experimental conditions. The methods presented here will usually be adequate for the tests conducted under the conditions specified in this document. An experienced applied statistician should be consulted whenever there are questions.

The following are examples of departures from ideal experimental conditions that may require additions to or modifications of the statistical methods presented in this chapter:

- Unequal numbers of experimental animals assigned to each treatment container, or loss of animals during the experiment
- Unequal numbers of replications (i.e., containers or aquaria) of the treatments
- Measurements scheduled at selected time intervals actually performed at other times
- Different conditions of salinity, pH, dissolved oxygen, temperature, etc., among exposure chambers
- Differences in placement conditions of the testing containers, or in the animals assigned to different treatments
- Contaminant concentration data reported as less than detection limit.

Problems such as these, which result in non-ideal data, will be examined and illustrated in detail in an Applications Guide to be published by the USACE as a supplement to this Appendix (Clarke and Brandon, in press).

The following statistical methods will be presented as each applies to a specific test procedure:

- Tests of assumptions (normality and equality of variances)
 - Data-scale transformations
 - Two-sample *t*-test
 - Nonparametric two-sample test
-

- Power and sample size calculations
- LC_{50} calculations
- Parametric multiple comparisons among treatments
- Nonparametric multiple comparisons among treatments
- Confidence interval calculations
- Comparisons to action levels

Decision trees are included to provide a general overview of each biological test. These trees illustrate which of the above statistical methods are appropriate for analyzing the results of each biological test, and the order in which the statistical procedures should be conducted. The trees include three general levels of decisions in the biological testing evaluation process: (1) decisions made by evaluating the experimental QA/QC and examining dredged material and reference means, (2) decisions concerning which statistical comparison procedure to use based on tests of assumptions, and (3) decisions concerning the significance of statistical comparisons.

The statistical methods (with the exception of LC_{50} procedures) are illustrated in this Appendix with example data analyzed by SAS IBM-compatible PC programs (SAS Institute, Inc., 1988a-c). This manual does not constitute official endorsement or approval of these or any other commercial hardware or software products. Other equally acceptable hardware and software products are commercially available and may be used to perform the necessary analyses. For example, all analyses required for this Appendix can be conducted using SYSTAT (Steinberg, 1988; Wilkinson, 1990; Steinberg and Colla, 1991), with different tests for normality and equality of variances. If it is necessary to write original programs to perform statistical analysis, the appropriateness of the techniques and accuracy of the calculations must be very carefully verified and documented.

Each example data set included in this Appendix is analyzed using several different statistical methods (usually, all of the possible tests in the appropriate decision tree) for illustrative purposes only. *Note that the results of different statistical tests will occasionally disagree, and it is never appropriate to conduct several tests in order to choose the result one likes best.* Decisions concerning the proper statistical tests to use should be made *a priori*, based on such considerations as experimental design, hypotheses of interest, relative importance of Type I and Type II error rates (Section D1.2), and tests of assumptions (Section D2.1.1.1).

D1.1 Basic Statistics

Statistical methods are used to make inferences about *populations*, based on *samples* from those populations. In most toxicity and bioaccumulation tests, samples of exposed organisms are used to estimate the response of the population of laboratory organisms. The response from the samples is usually compared with the response to a reference¹, or with some fixed standard such as an FDA action level. In any toxicity or bioaccumulation test, summary statistics such as means and standard errors for response variables (e.g., survival, contaminant levels in tissue) should be provided for each treatment (e.g., elutriate concentration, sediment).

In the tests described herein, samples or observations refer to *replicates* of treatments. Sample size n is the number of replicates (i.e., experimental units, test containers) in an individual treatment, not the number of organisms in a test container. Overall sample size N is the total number of replicates in all treatments combined, i.e.,

$$N = n_1 + n_2 + n_3 + \dots + n_k$$

where k is the total number of treatments in the experiment.

The statistical methods discussed in this Appendix are described in general statistics texts such as Steel and Torrie (1980), Sokal and Rohlf (1981), Dixon and Massey (1983), Zar (1984), and Snedecor and Cochran (1989). We recommend that investigators using this Appendix have at least one of these texts on hand. A nonparametric statistics text such as Conover (1980) can also be helpful.

Mean

The sample mean (\bar{x}) is the average value, or $\Sigma x_i / n$, where

$$\begin{aligned} n &= \text{number of observations (replicates)} \\ x_i &= \textit{ith} \text{ observation, e.g., } x_2 \text{ is the second observation} \\ \Sigma x_i &= \text{every } x \text{ summed} = x_1 + x_2 + x_3 + \dots + x_n ; \textit{ usually written } \Sigma x \end{aligned}$$

Most calculators and statistical software packages will provide means.

Standard deviation

The sample standard deviation (s) is a measure of the variation of the data around the mean. The sample variance, s^2 , is given by:

¹ Reference is used generically to refer either to a reference sediment (as in benthic toxicity and bioaccumulation testing), or to dilution water or control water (used in water column toxicity testing).

$$s^2 = \frac{\Sigma x^2 - (\Sigma x)^2/n}{n - 1} \quad (\text{Eq. 1})$$

Standard error of the mean

The standard error of the mean (SE, or s/\sqrt{n}) estimates variation among sample means rather than among individual values. The SE is an estimate of the SD among means that would be obtained from several samples of n observations each. Most of the statistical tests in this manual compare means with other means (e.g., dredged sediment mean with reference mean) or with a fixed standard (e.g., FDA action level). Therefore, the "natural" or "random" variation of sample means (estimated by SE), rather than the variation among individual observations (estimated by s), is required for the tests.

In addition to the summary statistics above, two other statistics derived from the normal (bell-shaped) frequency distribution are central to statistical testing and to the tests described in this Appendix. These two statistics are normal deviates (z -scores) and Student's t .

Normal deviates (z)

Z -scores or normal deviates measure distance from the mean in standard deviation units in a normal distribution. For example, a point 1 standard deviation greater than the mean has a z -score of 1; the mean has a z -score of 0. Z -scores are usually associated with a cumulative probability or proportion. For example, suppose an investigator wants to know the proportion of values in a normal distribution less than or equal to the mean plus 1 standard deviation. In this situation $z=0.84$, i.e., in a normal distribution 84% of values will be less than or equal to the mean plus 1 standard deviation. Alternatively, an investigator may want to determine the z -score associated with a specific proportion or probability. For example, he or she may want to know the range in which 95% of the values in a normal distribution should fall. That range is the mean ± 1.96 standard deviation (z -scores from -1.96 to +1.96).

Tables of z -scores can be found in most statistical texts, and bear titles such as "Standard Normal Cumulative Probabilities," "Ordinates of the Normal Curve," or "Normal Curve Areas." Typically the z -scores are listed in the column (top) and row (left) margins, with the column marginal value being added to the row marginal value to obtain the z -score. The body of the table contains the probability associated with each z -score. However, depending on the table, that probability may refer to the proportion of all values less than the z -score, the proportion of values falling between 0 and the z -score, or the proportion of values greater than the z -score. For example, if the z -score is 1.96, 97.5% of the values in a normal distribution fall below the z -score (Kleinbaum and Kupper, 1978, Table A-1), 47.5% fall between 0 and the z -score (Rohlf and Sokal, 1969, Table P), and 2.5% fall above the z -score (Steel and Torrie, 1980, Table A.4). It is important to distinguish which probability is of interest.

Z-scores can also be obtained from functions in statistical software packages. For example, in SAS the PROBIT function will return a z -score for a specified probability, and the PROBNORM function will compute the proportion of values less than a given z -score.

Student's t

Normal deviates can only be used to make inferences when the standard deviation is known, rather than estimated. The true population mean (μ) and standard deviation (σ) are only known if the entire population is sampled, which is rare. In most cases samples are taken randomly from the population, and the s calculated from those samples is only an estimate of σ . Student's t -values account for this uncertainty, but are otherwise similar to normal deviates. For example, an investigator may want to determine the range in which 95% of the values in a population should fall, based on a sample of 20 observations from that population. If the sample consisted of the entire population, μ and σ would be known with certainty, and normal deviates would be used to estimate the desired range (as in the above paragraph). However, if the sample represented only a small proportion of the population, t -values would be used to estimate the desired range. The degrees of freedom for the test, which is defined as the sample size minus one ($n-1$), must be used to obtain the correct t -value. Student t -values decrease with increasing sample size, because larger samples provide a more precise estimate of μ and σ . For a probability of 95%, the appropriate range of t -values is -2.09 to $+2.09$. In other words, 95% of the values in the population should lie within the range: sample mean $\pm 2.09 s$. Note that this is wider than the corresponding range calculated using normal deviates. As sample size increases, t -values converge on the z -scores for the same probability.

Tables of t -values typically give the degrees of freedom (df or ν) in the row (left) margin and probabilities or percentiles in the column (top) margin. Percentiles refer to the cumulative proportion of values less than t , whereas probabilities (also known as α in this case) refer to the proportion of values less than $-t$ and/or greater than $+t$. A two-tailed probability refers to both "tails" of the t -distribution curve, i.e., the probability of a value either $>+t$ or $<-t$. A one-tailed probability refers to only one of the tails of the curve, e.g., the probability of a value $>+t$.

When using a t table, it is crucial to determine whether the table is based on one-tailed probabilities (such as Table V in McClave and Dietrich, 1979, and Table A-2 in Kleinbaum and Kupper, 1978), or two-tailed probabilities (such as Table A.3 of Steel and Torrie, 1980). Some tables give both (such as Table B.3 of Zar, 1984). For most applications involving t -values in this Appendix, one-tailed probabilities are desired. The body of the table contains the t -value for each df and percentile (or α). The t -value for a one-tailed probability may be found in a two-tailed table by looking up t under the column for twice the desired one-tailed probability. For example, the one-tailed t -value for $\alpha = 0.05$ and $df = 20$ is 1.725, and is found in a two-tailed table using the column for $\alpha = 0.10$.

Statistical software packages may also provide functions to determine t -values or their associated probabilities. In SAS, these functions are TINV and PROBT.

D1.2 Hypothesis Testing

The goal in analyzing toxicity and bioaccumulation test data is to determine whether the mean effect of exposure to a dredged sediment is significantly greater than the mean effect of exposure to a reference. Two formal hypotheses underlie the statistical analysis of data in the two-sample situation. Let μ_r denote the mean effect of exposure to the reference R and let μ_d denote the mean effect of exposure to the dredged sediment D. Then, these two hypotheses are defined as follows:

Null hypothesis

Case 0: $H_0: \mu_d = \mu_r$

There is no difference in mean effect between the treatment (dredged sediment) and reference.

Alternative hypotheses

Case 1: $H_1: \mu_d < \mu_r$

The mean effect of the dredged sediment is less than the mean effect of the reference (e.g., survival).

OR

Case 2: $H_1: \mu_d > \mu_r$

The mean effect of the dredged sediment is greater than the mean effect of the reference (e.g., bioaccumulation).

Our hypothesis test will either reject H_0 for H_1 (Case 1 or Case 2), or will be unable to reject H_0 (Case 0). A one-tailed test is used because there is little concern about identifying a lesser negative effect from the dredged sediment than from the reference.

In performing the hypothesis test, and in determining the sample size to use in the test, the investigator must be aware of the probabilities for two types of errors that can occur in the conclusion. Type I errors occur if, after analysis of the data, H_0 is rejected when it was actually true. In Case 1 for example, a Type I error occurs when it is concluded that the mean effect (e.g., survival) of the dredged sediment is less than the mean effect of the reference when, in fact, the true mean effect of the dredged sediment is not less than that for the reference. Type II errors occur when H_0 is not rejected when it actually should have been rejected (e.g., in Case 2, it is concluded that there is no difference in mean effects of the dredged sediment and reference when, in fact, the true mean effect of the dredged sediment is greater than that of the reference).

To be environmentally protective in dredged sediment disposal evaluations, it is more important to guard against Type II errors. A Type II error could result in inappropriate placement of dredged sediment in the aquatic environment, while a Type I error could result in more costly alternatives to aquatic disposal. The probability of a Type I error is often represented by the letter α ; the probability of a Type II error is often written as β . The

significance level or confidence level of a statistical test is $1 - \alpha$. The power of a test is $1 - \beta$, which is the probability of rejecting H_0 when it should be rejected, or in other words, the power to detect true significant differences. For example, in Case 2 above, the power is the probability of concluding that the mean effect is greater in the dredged-sediment group when, in fact, this is true. The types of errors and their associated probabilities are summarized in Table D-1.

Table D-1. Types of Errors in Hypothesis Testing and Associated Probabilities.

Hypothesis Test Conclusion	True State of Nature	
	H_0 True	H_0 False
H_0 True (do not reject)	Correct (probability = $1 - \alpha$)	Type II Error (probability = β)
H_0 False (reject)	Type I Error (probability = α)	Correct (probability = $1 - \beta$)

In hypothesis testing, the Type I error rate is usually prespecified (biological tests, by convention, generally set $\alpha = 0.05$, although there is nothing magical about this probability). An ideal statistical procedure for hypothesis testing seeks to maintain the predetermined α , while minimizing the Type II error rate (i.e., maximizing power). It may not be possible to do both, particularly if the sample data depart from a normal distribution. A test that does well in maintaining the predetermined α , regardless of the characteristics of the sample data, is considered "robust." Tests included in this Appendix were chosen primarily on the basis of power rather than robustness, as the consequences of Type II error were considered more severe than those of Type I error.

Simple formulae for calculating the power of the statistical tests used in this Appendix are presented along with the descriptions of the tests in Sections D2.1.1.1, D2.2.1, D2.2.2, D3.1.2, and D3.2.2. The formulae may be used to calculate the sample size required to ensure a specific power of detecting an effect of a given magnitude (effect size), assuming that effect exists. The formulae can also be used to calculate the power of a specific sample size to detect a specified difference. This latter approach is often more relevant than calculating required sample sizes because budget or logistical constraints usually limit the number of replicates that can be used in biological tests. This is especially true if the tests include expensive chemical analyses (e.g., Tiers III and IV bioaccumulation tests).

D1.3 Experimental Design

Once the investigator has formulated the null hypotheses to be tested, decided upon significance (α) and power ($1 - \beta$) levels for hypothesis testing, and determined the sample size necessary to achieve the desired power, the next step is to design an experiment to test the hypotheses. Instructions for setting up and conducting sediment toxicity and bioaccumulation experiments are outlined in Chapters 11 and 12, but it is important at this point

to review the basic principles of *experimental design*. These principles include replication, randomization, interspersions, and controls (Hurlbert, 1984).

Replication refers to the assignment of a treatment to more than one experimental unit. The number of replicates, as stated earlier, is the sample size for that treatment. Recall that an experimental unit or replicate is the test container (e.g., a beaker or an aquarium), *not* an individual organism in the test container. The number of organisms in the test container is important only in terms of constituting an adequate measure of the endpoint being tested (e.g., providing sufficient tissue to measure contaminant bioaccumulation). Replication of treatments is necessary to control for random error in the conduct of the experiment. Appendix E includes guidelines for minimum number of replicates for various Tier III and IV bioassays. However, we strongly recommend determining sample size *a priori* using the power formulae in Sections D2.1.1.1, D2.2.1, D2.2.2, and D3.2.2. In many cases, the number of replicates necessary for a powerful statistical test will be greater than the minimum guidelines.

Randomization and interspersions refer to the actual placement of experimental units in the laboratory setup. A random numbers table, available in most statistical texts, may be used to randomly assign treatments to the experimental units. If the randomization does not achieve a reasonable interspersions of treatments, e.g. if several experimental units of the same treatment are clumped together, then a new randomization should be tried. Randomization and interspersions are necessary to control for investigator bias, for initial or inherent variability among experimental units, and for variability in environmental conditions such as lighting, water flow, etc.

Replication, randomization, and interspersions all function to control extraneous sources of variability in an experiment. In addition, *control treatment(s)* are needed to control temporal or procedural variability. In the broadest sense, the control treatment is simply the treatment against which the other treatments are compared. This is the dilution water (or control water) in water column toxicity testing, and the reference sediment in benthic toxicity and bioaccumulation testing. Laboratory controls, such as a clean sand exposure in bioaccumulation testing, may also be included. In Tiers III and IV testing, laboratory controls are used for quality assurance, and are not included in the statistical analyses.

Testing in Tiers III and IV can in most cases be best accomplished using simple experimental designs, either a completely randomized design or a randomized complete blocks design. These designs are discussed in most general statistics texts. In a completely randomized design, treatments are assigned to experimental units randomly over the entire experimental setup. A randomized complete blocks design should be used when the experimental units are placed on or in several different tables, benches or water baths (i.e., "blocks"). Each block holds a certain proportion of the experimental units. Treatments are assigned to experimental units randomly within each block, and each block contains an equal number of replicates of each treatment. Either of these designs is acceptable, providing the principles of replication, randomization, interspersions, and controls are followed. Adherence to the principles of experimental design ensures that the most basic assumption of statistical hypothesis testing, the assumption that treatments are sampled independently, is met.

D2.0 BIOLOGICAL EFFECTS

D2.1 Tier III Water Column Toxicity Tests

The objective of the analysis of Tier III water column toxicity test data is to assess the evidence for reduced survival due to toxicity of suspended plus dissolved dredged sediment constituents. If reduced survival is evident, then the median lethal concentration (LC_{50}) or effective sublethal concentration (EC_{50}) of the dredged sediment is calculated from the serial dilution experiment described in Section 11.1.4. Figures D-1 and D-2 provide an overview of water column toxicity test data analysis. Control survival must be $\geq 90\%$ or some other appropriate value, otherwise the test must be repeated (Section 13.3.17.3). At the end of the exposure period, the effects, if any, on the survival of the test organisms should be clearly manifest in the 100% elutriate concentration. When the dilutions are prepared with other than control water, the dilution water treatment is preferred over the control water for the data analysis. If the elutriate survival exceeds the control survival, then the toxicity test indicates no adverse impact from the dredged sediment (Section 11.1.5).

D2.1.1 Comparison of 100% Elutriate and Dilution Water

D2.1.1.1 Methods

Two-sample t -test

The usual statistical test for comparing two independent samples such as the 100% elutriate and the dilution water is the two-sample t -test (Snedecor and Cochran, 1989). The t -test will also be used in some circumstances in benthic toxicity and bioaccumulation tests, to compare individual dredged sediments with a reference (see Figures D-1, D-4A, D-5A).

The t -statistic for testing the equality of means \bar{x}_1 and \bar{x}_2 from two independent samples with n_1 and n_2 replicates is:

$$t = (\bar{x}_1 - \bar{x}_2) / \sqrt{s_{pooled}^2 (1/n_1 + 1/n_2)}, \quad (\text{Eq. 2})$$

where s_{pooled}^2 , the pooled variance, is calculated as:

$$s_{pooled}^2 = [s_1^2(n_1 - 1) + s_2^2(n_2 - 1)] / (n_1 + n_2 - 2), \quad (\text{Eq. 3})$$

and where s_1^2 and s_2^2 are the sample variances of the two groups. If the sample sizes are equal ($n_1 = n_2$), then:

$$s_{pooled}^2 (1/n_1 + 1/n_2) = 2s_{pooled}^2/n \quad . \quad (\text{Eq. 4})$$

The calculated t is compared with the Student t distribution with $n_1 + n_2 - 2$ degrees of freedom.

The use of Eq.2 to calculate t assumes that the variances of the two groups are equal. If the variances are unequal (see Tests for Equality of Variances below), t is computed as:

$$t = (\bar{x}_1 - \bar{x}_2) / \sqrt{s_1^2/n_1 + s_2^2/n_2} \quad . \quad (\text{Eq. 5})$$

This statistic is compared with the Student t distribution with degrees of freedom given by Satterthwaite's (1946) approximation:

$$df = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2 / (n_1 - 1) + (s_2^2/n_2)^2 / (n_2 - 1)} \quad . \quad (\text{Eq. 6})$$

This formula can result in fractional degrees of freedom, in which case one should round df down to the nearest integer in order to use a t table. The degrees of freedom for the t -test for unequal variances will usually be less than the degrees of freedom for the t -test for equal variances.

Tests of Assumptions

The two-sample t -test for equal variances (and other parametric tests such as analysis of variance) is only appropriate if:

- there are independent, replicate experimental units for each treatment,
- each treatment is sampled from a normally distributed population, and
- variances for both treatments are equal or similar.

The first assumption is an essential component of experimental design (Section D1.3). The second and third assumptions can be tested using the data obtained from the experiment. Therefore, prior to conducting the t -test, tests for normality and equality of variances should be performed. In some statistical software packages, these

tests of assumptions are done in conjunction with *t*-tests or as part of data summary or screening routines that also provide means, *s*, SE and various diagnostic statistics.

Outliers (extreme values) and systematic departures from a normal distribution (e.g., a log-normal distribution) are the most common causes of departures from normality and/or equality of variances. An appropriate transformation will normalize many distributions. In fact, the arcsine transformation (arcsine, in radians, of \sqrt{p} , where *p* is the survival expressed as a proportion) is so effective, and so frequently necessary, that this Appendix recommends applying it automatically to all survival data in the analysis of toxicity tests. Problems with outliers can usually be solved only by using nonparametric tests, but careful laboratory practices can reduce the frequency of outliers.

Tests for Normality

The most commonly used test for normality for small sample sizes (<50 observations total) is the Shapiro-Wilk's Test. This test determines if residuals are normally distributed. Residuals are the differences between individual observations and the treatment mean. Residuals, rather than raw observations, are tested because subtracting the treatment mean removes any differences among treatments. This scales the observations so that the mean of residuals for each treatment and over all treatments is zero. The Shapiro-Wilk's Test provides a test statistic *W*, which is compared to values of *W* expected from a normal distribution. *W* will generally vary between 0.3 and 1.0, with lower values indicating greater departure from normality. Because normality is desired, one looks for a high value of *W* with an associated probability greater than the prespecified α level.

Table D-2 provides α levels to determine whether departures from normality are significant. Normality should be rejected when the probability associated with *W* (or other normality test statistic) is less than α for the appropriate total number of replicates (*N*) and design. A balanced design means that all treatments have an equal (or nearly equal) number of replicate experimental units. For applications in this Appendix, a design may be considered unbalanced when the treatment with the largest number of replicates (n_{\max}) has at least twice as many replicates as the treatment with the fewest replicates (n_{\min}). Note that higher α levels are used when number of replicates is small, or when the design is unbalanced, because these are the cases in which departures from normality have the greatest effects on *t*-tests and other parametric comparisons. If data fail the test for normality, even after transformation, nonparametric tests should be used (see Nonparametric Tests below).

Table D-2. Suggested α Levels to Use for Tests of Assumptions.

Test	Number of Observations ^a	α When Design Is	
		Balanced	Unbalanced ^b
Normality	$N = 3$ to 9	0.10	0.25
	$N = 10$ to 19	0.05	0.10
	$N = 20$ or more	0.01	0.05
Equality of Variances	$n = 2$ to 9	0.10	0.25
	$n = 10$ or more	0.05	0.10

^a N = total number of observations (replicates) in all treatments combined; n = number of observations (replicates) in an individual treatment

^b $n_{\max} \geq 2n_{\min}$

Tables of quantiles of W can be found in Shapiro and Wilk (1965), Gill (1978), Conover (1980), USEPA (1989) and other statistical texts. These references also provide methods of calculating W , although the calculations can be tedious. For that reason, computer programs are preferred for the calculation of W . SAS can calculate W using the NORMAL option in PROC UNIVARIATE (see Program WATTOX.SAS in Section D4.1).

The Kolmogorov-Smirnov (K-S) Test is also an acceptable test for normality for small sample sizes, provided that the probabilities developed by Lilliefors (1967) are used (Sokal and Rohlf, 1981). The SYSTAT NPAR module provides the appropriate test, and specifically identifies the test as Lilliefors Test (Wilkinson, 1990). Other statistical packages providing K-S Tests may not use the Lilliefors probabilities, and the package documentation should always be checked to determine if the appropriate probabilities are provided. The chi-square (χ^2) test for normality can be used for larger sample sizes (e.g., $N > 50$) (Sokal and Rohlf, 1981).

Tests for Equality of Variances

There are a number of tests for equality of variances. Some of these tests are sensitive to departures from normality, which is why a test for normality should be performed first. Bartlett's Test, Levene's Test, and Cochran's Test (Winer, 1971; Snedecor and Cochran, 1989) all have similar power for small, equal sample sizes ($n=5$) (Conover et al., 1981), and any one of these tests is adequate for the analyses in this Appendix. Many software packages for t -tests and analysis of variance (ANOVA) provide at least one of the tests. Levene's Test can easily be performed by comparing the absolute values of residuals between treatments using t -tests or ANOVA. SAS statements for conducting Levene's Test are provided in BENTOX.SAS, BIOACC.SAS and BIOACCSS.SAS programs (Sections D4.2.1, D4.3.1 and D4.4.1).

If no tests for equality of variances are included in the available statistical software, Hartley's F_{\max} can easily be calculated:

$$F_{\max} = (\text{larger of } s_1^2, s_2^2) / (\text{smaller of } s_1^2, s_2^2)$$

When F_{\max} is large, the hypothesis of equal variances is more likely to be rejected. F_{\max} is a two-tailed test because it does not matter which variance is expected to be larger. Some statistical texts provide critical values of F_{\max} (Winer, 1971; Gill, 1978 [includes a table for unequal replication, but only for $\alpha = 0.05$]; Rohlf and Sokal, 1969). In the two-sample case, Hartley's F_{\max} is the same as the Folded- F or F' test. The F' test is conducted automatically in the SAS TTEST procedure.

Cochran's Test, where C = the largest variance divided by the sum of the variances, is also simple to calculate by hand, and is somewhat more powerful than Hartley's F_{\max} for small, equal sample sizes (Conover et al., 1981). However, tables of critical values of Cochran's C are not available in most statistical texts. Winer (1971) and Dixon and Massey (1983) include a table for Cochran's Test, but the tables are limited to tests with equal sample sizes. Tables of critical values for tests such as Cochran's C and Hartley's F_{\max} may also be restricted to one or two α levels (usually 0.05 and 0.01). Because of the limitations of these tables, computer programs are preferred for tests of equality of variances.

Levels of α for tests of equality of variances are provided in Table D-2; these depend upon number of replicates in a treatment (n) and allotment of replicates among treatments (design). Relatively high α 's are recommended because the power of the above tests for equality of variances is rather low when n is small. Equality of variances is rejected if the probability associated with the test statistic is less than the appropriate α . If the test for equality of variances is significant even after transformation, the t -test for unequal (separate) variances should be selected rather than the t -test for equal (pooled) variances.

Nonparametric Tests

Tests such as the t -test, which analyze the original or transformed data, and which rely on the properties of the normal distribution, are referred to as parametric tests. Nonparametric tests, which do not require that data be normally distributed, generally analyze the ranks of data, comparing medians rather than means. The median of a sample is the middle or 50th percentile observation when the data are ordered from smallest to largest. In many cases, nonparametric tests can be performed simply by converting the data to ranks or normalized ranks, and then conducting the usual parametric test procedures on the ranks.

Nonparametric tests are useful because of their generality, but may have less statistical power than corresponding parametric tests when the parametric test assumptions are met.

When parametric tests are not appropriate for comparisons because the normality assumption is not met, we recommend converting the data to normalized ranks (rankits). Rankits are simply the z -scores expected for the rank in a normal distribution. Thus, using rankits imposes a normal distribution over all the data, although not

necessarily within each treatment. Rankits can be obtained by ranking the data, then converting the ranks to rankits using the following formula:

$$\text{rankit} = z_{[(\text{rank} - 0.375) / (N + 0.25)]} \quad (\text{Eq. 7})$$

where z is the normal deviate and N is the total number of observations. For example, the approximate rankit for the sixth lowest value (rank=6) of 20 would be $z_{[(6 - 0.375)/(20 + 0.25)]}$, which is $z_{0.278}$ or -0.59 .

In SAS, normalized ranks or rankits can be provided in PROC RANK with the NORMAL=BLOM option. In SYSTAT and other packages, the ranks must be converted to rankits using the formula above (the conversion is a one-line command). In some programs the conversion may be more difficult to make, especially if functions to provide z -scores for any probability are not available. When rankits cannot easily be calculated, the original data may be converted to ranks.

In comparisons involving only two treatments, there is no real need to test assumptions on the rankits or ranks; simply proceed with a one-tailed t -test for unequal variances using the rankits or ranks.

Statistical Power

For a t -test, the basic formula for calculating the sample size (number of replicate experimental units, n) per treatment necessary to provide a specified power ($1-\beta$) to detect a given effect size (d) is:

$$n = 2 (t_{1-\alpha, v} + t_{1-\beta, v})^2 (s^2/d^2) \quad (\text{Eq. 8})$$

where v = degrees of freedom (df) or $(n_1 + n_2 - 2)$

$t_{1-\alpha, v}$ = Student t -value for probability $1-\alpha$ and v df

$t_{1-\beta, v}$ = Student t -value for probability $1-\beta$ and v df

d = the effect size or difference to be detected.

Recall that β is the probability of committing a Type II error. This formula for n must be solved iteratively, because an initial value of n must be used to determine v . A new n is then calculated using the initial value, and the process is repeated until n and v are consistent. The iterative process can be tedious if computer programs are not used. It is easier to use the following approximate formula (from Alldredge, 1987):

$$n = 2 (z_{1-\alpha} + z_{1-\beta})^2 (s^2/d^2) + 0.25(z_{1-\alpha}^2) \quad (\text{Eq. 9})$$

where $z_{1-\alpha}$ = normal deviate for $1-\alpha$
 $z_{1-\beta}$ = normal deviate for $1-\beta$
 $0.25(z_{1-\alpha}^2)$ = correction term to increase sample size when n is small

Calculated n derived from this formula should be regarded as approximate for $n < 5$. Regardless of which formula is used, a fractional n is always rounded up to the next integer.

A useful exercise when sample sizes are fixed because of budget or logistic constraints is to calculate the power of the test to detect a specific effect size (d). In a test comparing 100% elutriate survival with dilution water survival, d is some selected reduction in mean 100% elutriate survival from mean dilution water survival. Eq. 8 can be rearranged and solved for $t_{1-\beta}$ to determine the power:

$$t_{1-\beta, v} = \frac{\sqrt{nd}}{\sqrt{2s}} - t_{1-\alpha, v} \quad (\text{Eq. 10})$$

We then enter a t table at v df and find the column closest to the value of $t_{1-\beta}$; power $\approx 1-P$, where P is the probability for that column. SAS can calculate power more exactly using the PROBT function for $t_{1-\beta}$ and v df. Note that t -values can be used because both n and v are known. One can also calculate the difference that can be detected for any given power and sample size:

$$d = (t_{1-\alpha, v} + t_{1-\beta, v})\sqrt{2s^2/n} \quad (\text{Eq. 11})$$

The simplest power to use is 0.50, because then $t_{1-\beta}=0$. Many computer programs will provide this difference, usually referred to as the "minimum significant difference", "least significant difference" or some similar term. The term "average detectable difference" would also be applicable, as this is the difference we expect to be able to detect 50% of the time. In this Appendix, we recommend reporting the minimum significant difference or some other indication of power along with the results of statistical analyses. If power is consistently and regularly reported, investigators will gain an appreciation of the strengths and limitations of various toxicity tests and analyses.

If values are transformed prior to analyses, all power calculations should be done on the transformed scale. In the case of arcsine-transformed survival, a constant effect size d on the percentage or proportion scale will not be constant on the arcsine scale, because the latter scale spreads out high and low values. Therefore, a reference survival must be specified and arcsine-transformed, and the effect size also transformed to a difference on the arcsine scale. For example, suppose we wanted to calculate the power of a t -test to detect a 25% reduction in survival from the reference. A reasonable reference survival (e.g., 90%) would be specified and arcsine-transformed (=1.249). We would also arcsine-transform a 25% reduction (=65% survival or 0.938 after

transformation). The difference d would then be $1.249 - 0.938$ or 0.311 , and that value would be used in power calculations. Experimentation with arcsine-transformed data will rapidly reveal that toxicity tests are more powerful, in terms of the size of differences that can be detected on the original (untransformed) scale, when reference survival is higher. In other words, we are more likely to detect a 25% reduction in survival if reference survival is 90% than if reference survival is 75%. This is precisely what happens in real toxicity tests, which is why the arcsine transformation is used for survival data.

Simple formulae for calculation of sample size or power are not available for the tests of assumptions recommended in this Appendix.

D2.1.1.2 Analysis of Example Data

Table D-3 contains example data from a 96-h water column toxicity test using a dilution water and a dredged-sediment elutriate at four serial dilutions. In this example, control (laboratory) water was also used for dilutions, and no separate control was necessary. In other cases, the dilution water may be receiving water and a separate laboratory control would be required. Analysis of this example data will be conducted using the decision tree in Figure D-1. Numbers in parentheses in the text refer to numbered nodes of the decision tree. The SAS program WATTOX and complete results for water column toxicity test data analyses are provided in Section D4.1; some additional analyses were conducted using SYSTAT programs.

Means (\bar{I}) and SE for the survival data are provided in Table D-3. Overall mean survival in the control (= dilution) water was 98%, indicating that the test was acceptable (2). The statistical comparison of 100% elutriate survival and dilution water survival was then conducted because the 100% elutriate survival was at least 10% lower than the dilution water survival (3). The next step was to arcsine-transform the survival proportions for the dilution water and 100% elutriate treatments (4).

Tests of Assumptions

Following arcsine-transformation, the data were tested for normality (5) to determine whether parametric or nonparametric procedures should be used. Table D-4 provides the results of tests for normality and equality of variances for the example data. The value of Shapiro-Wilk's W for the arcsine-transformed data was 0.846, with associated probability (P) = 0.051. Because this value of P exceeds 0.05 (α level from Table D-2, $N=10$, balanced design), we conclude that the data do not depart significantly from the normal distribution (5), and we now examine the results of the tests for equality of variances (6).

Bartlett's Test (from SYSTAT) and F' both indicated that the variances of arcsine-transformed data were not significantly different for the two treatments, with $P>0.10$ (α level from Table D-2, $n=5$, balanced design). Thus, on the basis of these tests, we would proceed with a t -test for equal variances (7).

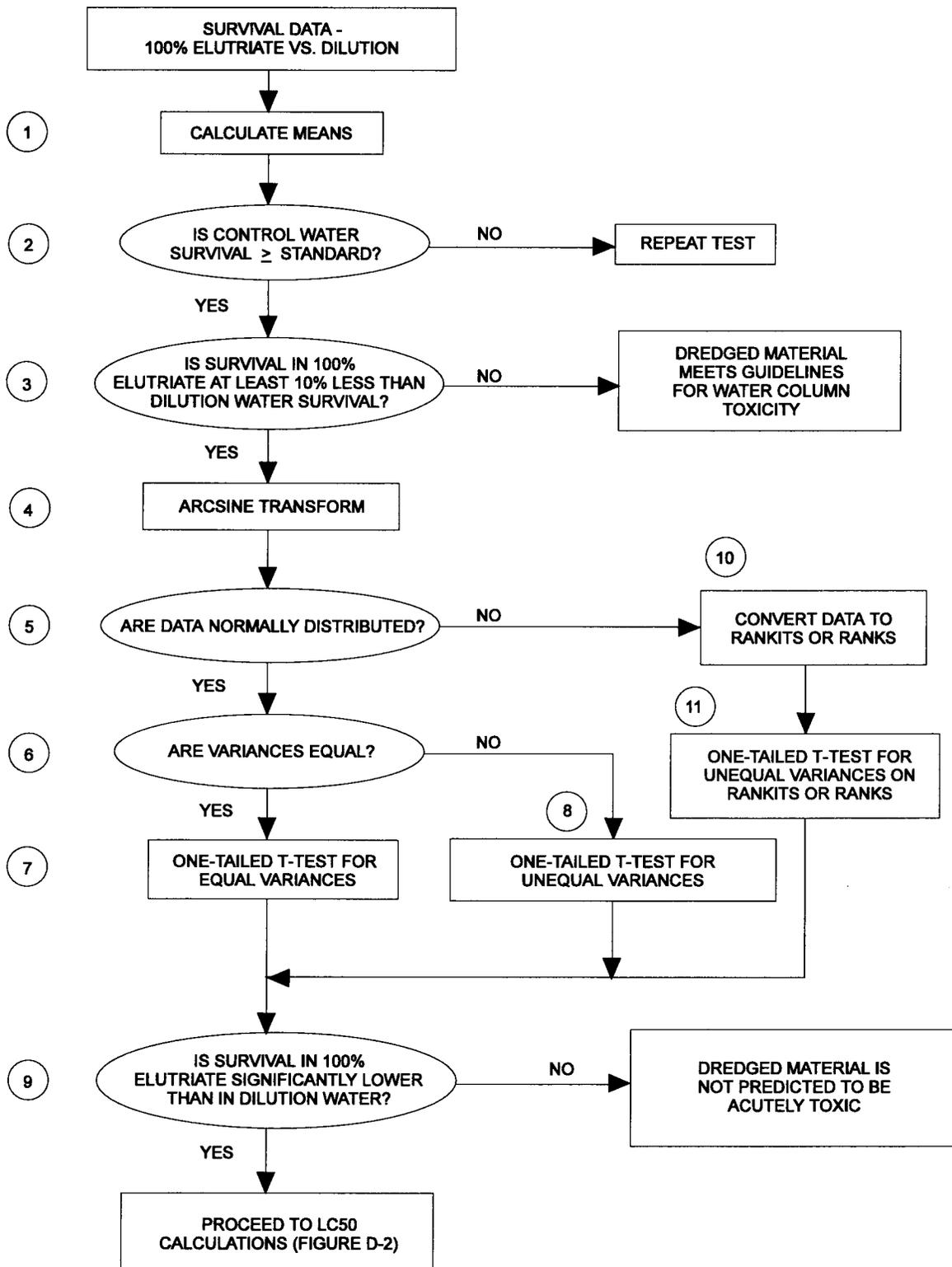


Figure D-1. Water Column Toxicity Test Decision Tree.

Table D-3. Number of Survivors in a Hypothetical Water Column Toxicity Test After 96 h.

Replicate ^b	Treatment ^a				
	Dilution Water ^c	100%	50%	25%	12.5%
1	20	6	8	12	17
2	19	7	8	18	17
3	20	9	9	15	18
4	20	5	10	14	16
5	19	8	11	13	18
Total	98	35	46	72	86
Mean	19.6 (98%)	7.0 (35%)	9.2 (46%)	14.4 (72%)	17.2 (86%)
SE	0.24	0.71	0.58	1.03	0.37

^a Percent concentrations of dredged-material elutriate:
 100% = 1 part elutriate plus 0 part dilution water
 50% = 1 part elutriate plus 1 part dilution water
 25% = 1 part elutriate plus 3 parts dilution water
 12.5% = 1 part elutriate plus 7 parts dilution water

^b 20 organisms per replicate at initiation of test

^c In this example, the dilution water was control (laboratory) water

Table D-4. Tests of Assumptions and Hypothesis Tests on Arcsine-Transformed Water Column Toxicity Test Example Data.

Null Hypothesis: Mean 100% Elutriate Survival Equals Mean Dilution Water Survival ^a				
Test	Test Statistic	Probability <i>P</i>	α	Conclusion
Normality Assumption: Shapiro-Wilk's Test	<i>W</i> =0.846	0.051	0.05	do not reject
Equality of Variances Assumption: Bartlett's Test <i>F'</i> Test	<i>F</i> =0.5 <i>F'</i> =2.18	0.47 0.468	0.25 0.25	do not reject do not reject
Null Hypothesis: <i>t</i> -Test (equal variances) <i>t</i> -Test (unequal variances) <i>t</i> -test on rankits (unequal variances)	<u><i>t</i>=12.734</u> <i>t</i> =12.734 <i>t</i> = 4.631	<u><0.0001</u> <0.0001 0.0010	<u>0.05</u> 0.05 0.05	<u>reject</u> reject reject

^a Based on tests of assumptions, appropriate statistical test of null hypothesis is underlined. Other test results are included for illustration only.

Two-sample t-tests

Table D-4 provides the results of t -tests for equal (7) and unequal variances (8). The t -test for equal variances indicated that survival in the 100% elutriate was significantly ($P < 0.05$) less than in the dilution water (9). If the data had been normally distributed with unequal variances, the t -test for unequal variances would have been used. With the example data, both test results are the same, but this will not always be the case.

Nonparametric Test

Nonparametric tests would generally not be performed on these data because the sample data did not depart significantly from a normal distribution. However, the data were converted to rankits (10), and a t -test for unequal variances (11) was conducted on the rankits (SAS Program WATTOX) for illustrative purposes. The t -test indicated that median survival in the 100% elutriate was significantly lower than in the dilution water (Table D-4).

Statistical Power

The difference in survival between the 100% elutriate and the dilution water was so large (63%) that it was easily detected (declared significant) even though there were only five replicates per treatment. The power of a t -test to detect such a large decrease in survival ($d = 0.848$ on the arcsine scale) when $n = 5$ and $s = 0.1055$ (also on the arcsine scale) is > 0.99 . However, it is reasonable to ask if $n = 5$ is adequate for detecting smaller differences. For example, what sample size would be required to provide a ≥ 0.95 chance ($1 - \beta = 0.95$; $z_{1-\beta} = 1.645$) of detecting a reduction of survival to $\leq 80\%$, with $\alpha = 0.05$ ($z_{1-\alpha} = 1.645$)? In the example data, mean arcsine-transformed dilution water survival was 1.4806 ($\approx 99\%$ survival; back-transformation of means of transformed values will not be the same as means based on original data, although the difference is trivial in this case); the arcsine-transformed value for 80% survival is 1.1071, giving a reduction (d) of 0.3736 on the arcsine scale; and the pooled s was 0.1055. Using Eq. 9:

$$n = 2(1.645 + 1.645)^2 (0.1055^2/0.3736^2) + 0.25(1.645^2) = 2.40$$

Rounding up gives $n = 3$. A more exact iterative computer program (SYSTAT DESIGN) based on t -values (Eq. 8) also yields $n = 3$. The sample size required for a 0.95 probability of detecting a reduction in survival to 90% is $n = 6$, again calculated with the iterative program. The minimum significant difference (i.e., the difference we have a 0.50 probability of detecting) when $n = 5$ is $t_{0.95,8}(2s^2/n)^{1/2}$ or $1.86[2(0.1055^2/5)]^{1/2} = 0.1241$. Subtracting that from the mean transformed dilution water survival, and back-transforming gives 95.5% survival. In other words, given the example data, the test can be expected to detect a reduction in survival from $\approx 99\%$ to $\approx 95\text{--}96\%$ approximately half the time.

When dilution water survival is near 100% and variation among replicates is low, as with the example data, a test with $n = 5$ replicates may be too powerful. In many cases, we would declare survival of $\geq 90\%$ in the 100% elutriate significantly lower than in the dilution water, yet that $\geq 90\%$ survival would be acceptable for the dilution water. For this reason, if survival in the 100% elutriate is not at least 10% lower than in the dilution

water, the difference should not be considered significant and no statistical tests need be performed. *It is important to remember that a statistically significant difference is not necessarily biologically significant (and vice versa).* If dilution water survival were lower, say 90% instead of 98%, and s remained the same, the t -test would have less power. For example, $n=13$ would be required to provide a 0.95 probability of detecting a reduction in survival in the 100% elutriate to 80%. Much higher standard deviations can also be expected in many toxicity tests.

The SAS program WATTOX (Section D4.1) provides minimum significant difference and power of a t -test. Power is determined for 10, 20, 30, 40 and 50 percent reductions in true population survival from the mean dilution water survival.

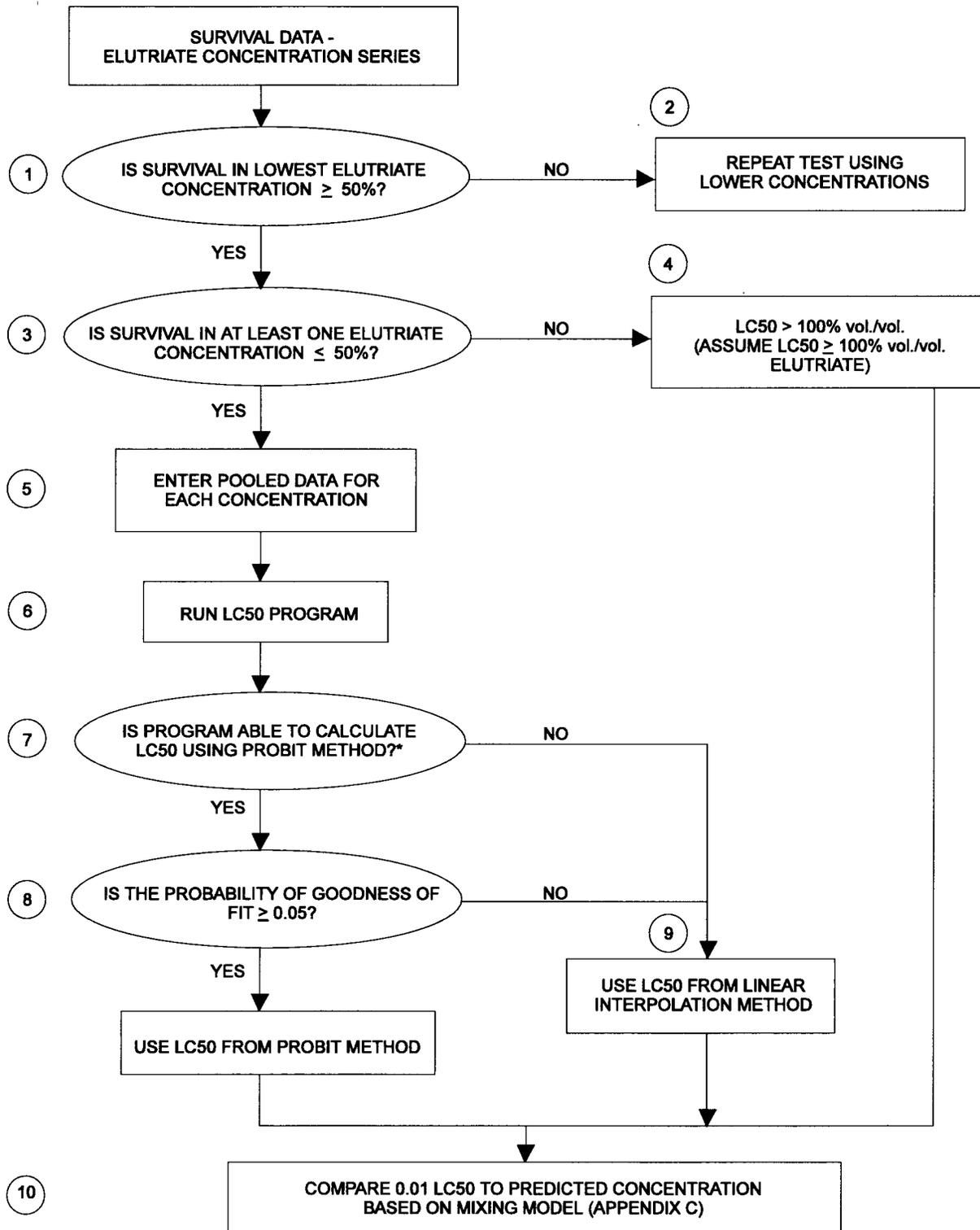
D2.1.2 Calculating Median Lethal Concentration

In Tier III water column toxicity tests, the median lethal concentration (LC_{50}) or median effective concentration (EC_{50}) are calculated when 100% elutriate survival is significantly lower than dilution water survival. The LC_{50} is the concentration lethal to 50% of the test organisms; the EC_{50} is the concentration causing some sublethal effect (e.g., abnormality, immobility) in 50% of the test organisms. The remainder of this section will discuss the LC_{50} but all comments apply equally to EC_{50} . Steps and decisions in the LC_{50} determination are shown in the decision tree in Figure D-2. Numbers in parentheses in the text refer to numbered nodes of the decision tree.

Ideally, data for at least five elutriate concentrations should be available to calculate an LC_{50} , although most methods described below can be used for fewer concentrations. The control or dilution water survival is not included. Survival in the lowest elutriate concentration must be at least 50% (1); otherwise the test must be repeated using lower concentrations (2). An LC_{50} should not be calculated unless at least 50% of the test organisms die in at least one of the serial dilutions (3). If there are no mortalities greater than 50%, then the LC_{50} is assumed to be $\geq 100\%$ elutriate (4).

If the conditions in (1) and (3) are met, then replicate mortality data for each concentration are pooled (5) for calculation of LC_{50} (6). The Probit method (7) can be used if the data meet the requirements of the Probit method listed below and fit the probit model (8). The Trimmed Spearman-Kärber (TSK) and Logistic methods (described below) are acceptable substitutes for the Probit method, provided that the data meet the requirements of these alternative methods. If the data do not meet the requirements of the Probit method or alternatives, then the Linear Interpolation method should be used (9). When an LC_{50} value has been determined, 1% of that value is entered into the mixing model (10) provided in Appendix C for mixing zone evaluation.

Calculation of LC_{50} values is also recommended for reference toxicant tests to determine the relative health of the organisms used in toxicity and bioaccumulation testing (Section 13.3.17.2).



* Trimmed Spearman-Kärber and logistic methods are acceptable substitutes for Probit method.

Figure D-2. LC₅₀ Decision Tree.

D2.1.2.1 Methods For Calculating LC₅₀

Stephan (1977) and Gelber et al. (1985) provide careful reviews of LC₅₀ estimation procedures. In addition, USEPA (1985) discusses in detail the mechanics of calculating LC₅₀ using current methods and contains, as an appendix, computer programs for each statistical method. The most commonly used methods are the Probit, Trimmed Spearman-Kärber (TSK) and Linear Interpolation. This Appendix recommends use of the Probit, TSK or Logistic methods if the data are appropriate; otherwise the Linear Interpolation method may be used (Figure D-2). In general, results from different methods should be similar. Programs commonly used to calculate LC₅₀ are PROBIT, developed for and available from the USEPA (Environmental Monitoring and Support Laboratory, Cincinnati, OH), and several programs developed by Dr. C.E. Stephan of the USEPA Environmental Research Laboratory in Duluth, Minnesota. Procedures in statistical packages such as SAS or SYSTAT may not be easily adaptable for routine calculations of LC₅₀, and specialized packages are generally preferred. This Appendix does not include SAS programs for LC₅₀.

Probit

The Probit method is based on regression of the probit of mortality on the log of concentration. A probit is the same as a z-score; for example, the Probit corresponding to 70% mortality is $z_{0.70}$ or ≈ 0.52 . The LC₅₀ is calculated from the regression, and is the concentration associated with $z=0$ (mortality = 50%). The Probit method can be used whenever the following conditions are met:

- there are at least two concentrations with partial mortality (i.e., >0 and <100%)
- the data points fit the probit regression line reasonably well.

The first condition is necessary because the regression line is estimated from the partial mortalities. The second condition, called goodness-of-fit, can be tested by the χ^2 statistic, which is a measure of the distance of the data points from the regression line. A low χ^2 indicates a good fit. By convention, the fit is considered adequate if the P -value for χ^2 is >0.05 (in other words, goodness-of-fit is rejected if $P \leq 0.05$). Programs such as PROBIT will only provide χ^2 , in which case χ^2 should be compared against tabled values with $k - 2$ df, where k is the number of partial mortalities. If there are only two partial mortalities ($k=2$), then there are 0 df, and the goodness-of-fit cannot be tested (i.e., a line between two points is always a perfect fit). When there are only two partial mortalities, the LC₅₀ is identical to the LC₅₀ which would be calculated by Linear Interpolation (see below) with mortality expressed on a probit scale. Goodness-of-fit can also be assessed by eye, if the data are plotted on log-probit paper, or if the computer program provides a plot.

Linear Interpolation Method

The Linear Interpolation method should be used when:

- there are 0 or 1 partial mortalities
-

- the data do not fit the Probit (or Logistic) models

The Linear Interpolation method should also be used when LC_{50} s are calculated and compared over an extended time series (i.e., for tracking reference toxicant results), because inevitably, one or more data sets will fail to meet the requirements for the Probit, TSK or Logistic methods. Linear Interpolation may also be used if programs for the other methods are unavailable, but we strongly recommend that investigators have programs available for one or more of the other methods.

The Linear Interpolation method calculates an LC_{50} by interpolation between the two concentrations with mortality nearest to, and on either side of 50%. The interpolation is made on a log concentration scale, using the following formula:

$$LC_{50} = \text{antilog} \frac{(50 - M_L) (\log C_U) + (M_U - 50) (\log C_L)}{M_U - M_L} , \quad (\text{Eq. 11})$$

where C_L = concentration with mortality nearest to and below 50%

C_U = concentration with mortality nearest to and above 50%

M_L = % mortality at C_L

M_U = % mortality at C_U .

If there are no partial mortalities, the formula simplifies to:

$$LC_{50} = \sqrt{(C_U)(C_L)} .$$

For the example data given in Table D-3, $C_L=25\%$ elutriate ($\log=1.398$); $M_L=28\%$ mortality; $C_U=50\%$ elutriate ($\log=1.699$); and $M_U=54\%$ mortality. Therefore:

$$LC_{50} = \text{antilog} \frac{(50 - 28) (1.699) + (54 - 50) (1.398)}{54 - 28} ,$$

or 44.9%.

The formula and example given above express mortality on an arithmetic (untransformed) scale. Some computer programs or investigators may use arcsine-transformed mortalities (Stephan, 1977; see Section D.2.1.1.1 Tests of Assumptions). One could also express mortality on a probit or logit scale, if there were one partial mortality on each side of 50%. In those cases, the Linear Interpolation should produce the same LC_{50} estimate as the Probit or Logistic methods. In this manual, we recommend the use of untransformed mortality for simplicity and consistency. However, LC_{50} estimates using other scales can easily be calculated for comparison.

Trimmed Spearman-Kärber (TSK) Method

The TSK method is a nonparametric method that can be calculated by hand using the procedure in Gelber et al. (1985). The calculations can be tedious, especially for processing large numbers of tests, and computer programs are usually used. The method is labelled "trimmed" because extreme values (mortality much higher or lower than 50%) are "trimmed" or removed prior to calculation of the LC_{50} . Thus, the LC_{50} is calculated using points near 50% mortality, which may produce a more robust estimate. The TSK method can be used in many cases where the Probit method is unsuitable. Access to appropriate computer programs, and difficulties in deciding what values to trim are probably the major factors limiting widespread use of the TSK method. Investigators with access to reliable programs should not hesitate to use the TSK method whenever there are two or more partial mortalities. Information concerning TSK computer programs may be obtained from the USEPA Environmental Research Laboratories in Athens, GA, or Duluth, MN, or CSC/USEPA, Cincinnati, OH.

Logistic Method

The Logistic method is similar to the Probit method except that mortalities are converted to logits rather than probits. A logit is $\log [M/(100 - M)]$, where M is % mortality. The LC_{50} is derived from a regression of logits on log concentration. As with the Probit method, the Logistic method can be used whenever there are two or more partial mortalities, and the data fit the regression line. Logistic regression is not commonly used in aquatic toxicology only because Probit programs are more available, but the two methods are equally acceptable. Logistic regression programs in SAS and SYSTAT are designed for complex analyses and comparisons of logistic regressions, and may be inconvenient to use for simple and routine calculations of LC_{50} for single tests.

D2.1.2.2 Analysis of Example Data

Table D-5 provides LC_{50} estimates calculated by several different methods using the example data in Table D-3. In all cases, the data from the five replicates for each concentration were pooled, and entered as the number responding (dying) out of 100. *Because pooling over replicates ignores any additional variance in survival among replicates (i.e., beyond the expected error from sampling the binomial distribution), the confidence limits provided by the programs may not be accurate and should not be reported or used.* Because the LC_{50} is required only for use in the mixing model (Appendix C), confidence limits are not needed.

Table D-5. Calculated LC₅₀ Values for Example Water Column Toxicity Test Data.

Method	LC ₅₀ Estimate (% v/v)
Probit	52.6
Linear Interpolation - untransformed mortality - arcsine-transformed mortality	44.9 45.1
Trimmed Spearman-Kärber	48.4
Logistic	52.6

The Probit LC₅₀ was calculated with the EPA PROBIT program, and was almost identical to the Logistic LC₅₀ calculated using the SYSTAT LOGISTIC program. The χ^2 goodness-of-fit for the Probit line was 1.756, indicating a good fit ($P > 0.05$ with $4 - 2 = 2$ df), which could be verified by examining the plot provided (Figure D-3). The LC₅₀ estimated by Linear Interpolation, with untransformed mortality, was almost identical to the LC₅₀ calculated using arcsine-transformed mortality. The TSK LC₅₀ was calculated using a program modified from an original program described in Hamilton et al. (1977), and was intermediate between the Linear Interpolation and regression (Probit and Logistic) estimates.

The various estimates in Table D-5 differed by up to 7.7% elutriate, which is not unusual or alarming. The Probit or Logistic LC₅₀ would be the preferred estimate, because the regression lines fit the data well, and the regression methods use more of the data in such cases. However, any of the estimates would be adequate for use in the mixing model in Appendix C, because the imprecision and uncertainty involved in the model calculations and estimates are undoubtedly far greater than the differences among the LC₅₀ estimates.

D2.2 Tier III Benthic Toxicity Tests

The objective of Tier III benthic toxicity tests is to determine if sediments taken from a potential dredge site are significantly more toxic than a reference sediment. The test procedure is described in Section 11.2. The statistical analysis recommended below assumes that individual dredge sites are relatively large, and that a decision about potential sediment toxicity, and subsequently about disposal options, will be made independently for each site. If only one dredge site is tested, and compared to a reference sediment, statistical analysis is the same as that given in Section D2.1.1 for comparison of 100% elutriate and dilution water (Figure D-1 and SAS program WATTOX in Section D4.1). However, in many cases, more than one dredge site is tested simultaneously with one reference sediment. In those cases, recommended statistical methods will differ from the two-sample case. Methods for comparison of more than one dredged sediment with a reference sediment are described below, and computer procedures are given in SAS program BENTOX (Section D4.2).

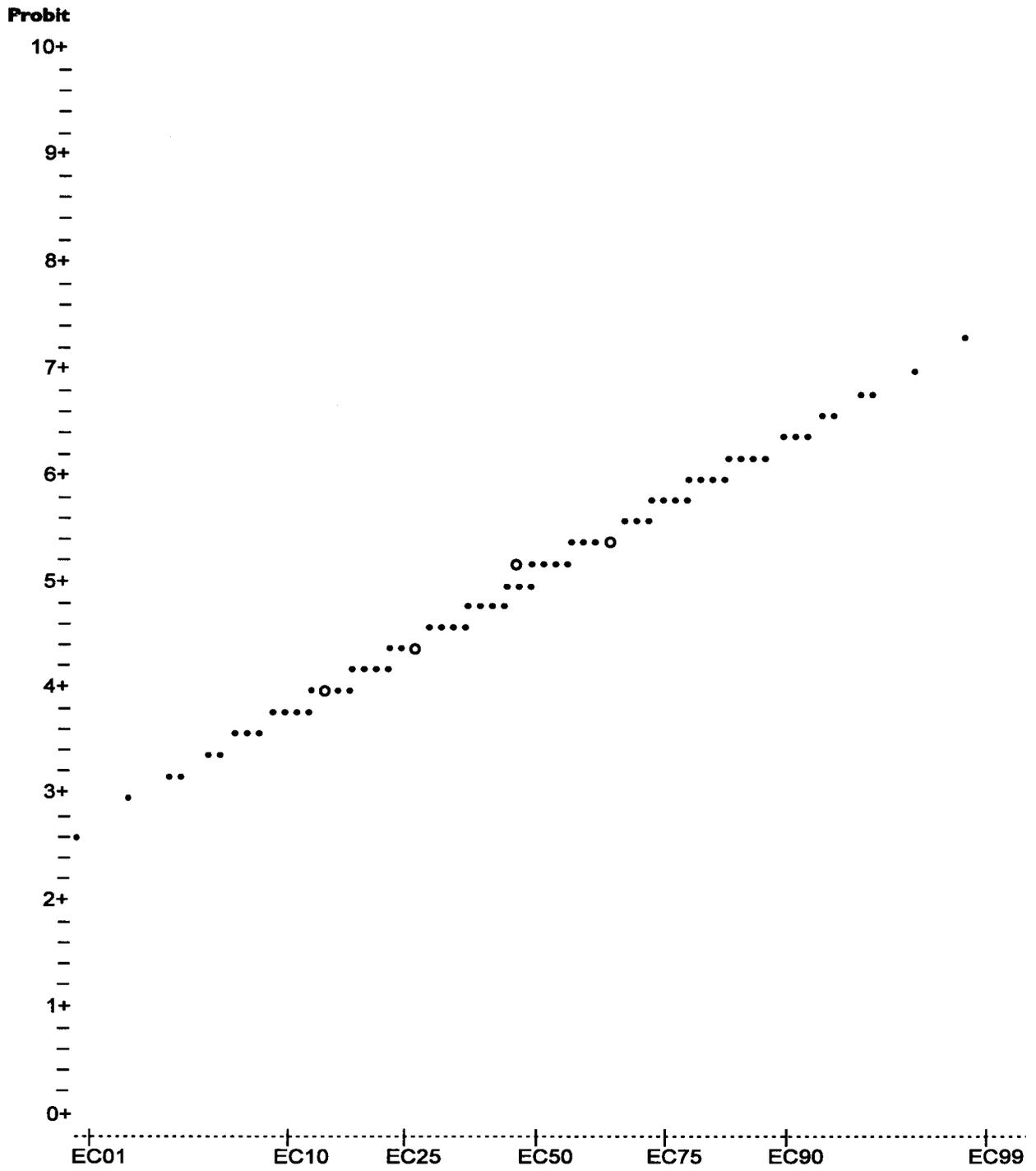


Figure D-3. Probit Plot of Water Column Toxicity Test Example Data.

D2.2.1 Methods
Fisher's Least Significant Difference (LSD)

Fisher's Least Significant Difference (LSD) is the appropriate parametric statistical test for assessing differences in survival or other response when more than two means are being compared. This *a posteriori* multiple comparison technique is discussed in many statistical texts, e.g., Steel and Torrie (1980); SAS Institute, Inc. (1988b); Snedecor and Cochran (1989); and Wilkinson (1990). The LSD controls the pairwise Type I error rate rather than the experimentwise Type I error rate. This means that when the test assumptions are met, the Type I error rate for each comparison is held to the preset α even though the overall Type I error rate for all comparisons (i.e., experimentwise error rate) may be higher. A test that controls the pairwise error rate is appropriate because disposal decisions are to be made independently for each dredge site regardless of how many sites are compared to the same reference. The LSD replaces the previously recommended Dunnett's test, which is not appropriate because it controls experimentwise error rate.

The LSD is usually performed in conjunction with analysis of variance (ANOVA), and only if the data meet the assumptions of normality and equal variances. The ANOVA is conducted primarily to provide the mean square error (*MSE*), which is an estimate of the pooled variance across all treatments. The ANOVA *F*-statistic and its associated probability are ignored in this application.

The test statistic for the LSD is *t*, calculated in much the same way as for a *t*-test:

$$t = (\bar{x}_1 - \bar{x}_2) / \sqrt{MSE (1/n_1 + 1/n_2)} \quad (\text{Eq. 13})$$

This *t*-statistic is compared against the distribution of Student's *t* with $N - k$ degrees of freedom, where N is the total number of observations (Σn) and k is the number of treatments including the reference. A *t*-statistic is computed for each possible pair of treatments in the analysis.

The *MSE* can be calculated as:

$$MSE = \Sigma[s_i^2 (n_i - 1)] / \Sigma(n_i - 1) \quad , \quad (\text{Eq. 14})$$

where s_i^2 and n_i are the variance and number of replicates for the *i*th treatment. The term $\Sigma(n_i - 1)$ is equivalent to $N - k$.

If sample sizes are equal, then:

$$MSE (1/n_1 + 1/n_2) = 2MSE/n \quad . \quad (\text{Eq. 15})$$

The major advantage of using the LSD as opposed to conducting individual two-sample t -tests comparing each dredged sediment to the reference is that the MSE is a better estimate of the true population variance than the pooled variance calculated from only two samples. Consequently, the LSD test is more powerful, as reflected in the greater df for the calculated t . It also follows that a pooled variance should only be calculated, and the LSD test conducted, if the variances for the treatments are not significantly different.

Tests of Assumptions

The Shapiro-Wilk's Test described in Section D2.1.1.1 can also be used to test for normality when more than two treatments are compared. If the data are not normally distributed, even after an appropriate transformation, then nonparametric tests should be used (see Nonparametric Tests below).

Bartlett's Test, Levene's Test, F_{\max} , or Cochran's Test can be used to test for equality of variances. If there are more than two samples, then F_{\max} is equal to the largest variance divided by the smallest variance. If variances are significantly unequal, even after transformation, then each dredged sediment should be compared with the reference using two-sample t -tests.

Nonparametric Tests

When parametric tests are not appropriate for multiple comparisons because the normality assumption is violated, the data should be converted to rankits, and the rankits should be tested for normality and equality of variances. If these assumptions are not violated, an LSD is then performed on the rankits (Conover, 1980, refers to this as van der Waerden's Test). Tests performed on rankits are robust to departures from normality, and can still be used when the normality assumption is violated. Rankits will rarely fail tests for normality, partly because a normal distribution is imposed over the entire data set. The rankit data may fail the test for equality of variances, but then t -tests can be conducted for each pair of treatments to be compared. If rankit-transformed data fail normality tests, it is probably safest to use the t -tests for unequal variances, as some tests for equality of variance are not robust when data are non-normal.

When rankits cannot be easily calculated, the original data may be converted to ranks (using SAS PROC RANK, for example). Equality of variances should be tested after the data are ranked. There is a common misconception that nonparametric tests can be used when variances are not equal as well as when data are not normally distributed. However, nonparametric tests are not very robust if the variances of the ranks are not similar among treatments. Bartlett's Test should not be used to test equality of variances of ranks, as ranks will follow a uniform, rather than a normal distribution, and Bartlett's Test is unduly sensitive to non-normality. Other tests discussed in Section D2.1.1.1 Tests for Equality of Variances may be used on ranks; there are also nonparametric tests for equality of variances provided in Conover (1980).

If the variances of the ranks are not significantly different, the Conover *T*-Test (Conover, 1980) should be performed. This test can most easily be conducted by performing an LSD on the ranks. If the variances of ranks are significantly unequal, a one-tailed *t*-test for unequal variances should be performed (using ranks) for each pair of treatments to be compared.

Statistical Power

Power calculations for the LSD are the same as for the *t*-test (see Eq. 8), except that the degrees of freedom for $t_{1-\alpha}$ and $t_{1-\beta}$ are $N - k$, and *MSE* replaces s^2 :

$$n = 2 (t_{1-\alpha, \nu} + t_{1-\beta, \nu})^2 (MSE/d^2) , \quad (\text{Eq. 16})$$

If the *z*-approximation (Eq. 9 with *MSE* replacing s^2) is used to calculate sample size, the result will be a slight overestimate, although the overestimation is rarely of practical importance. Finally, the minimum significant difference should be reported for LSD tests. Note that the test is named the Least Significant Difference because another way to conduct the test is to compare the observed differences to the minimum significant difference.

If an increase in power ($1-\beta$) is desired, because variance is high or sample size low, one effective method of increasing power is to increase the number of reference replicates rather than increase the sample size for each treatment. It is even possible to increase power without increasing overall sample size by increasing sample size for the reference, and decreasing sample size for the dredged sediments. The optimal apportionment of replicates is to make the sample size for the reference \sqrt{k} times the sample size for the other sediments (Dunnett, 1955). Increasing sample size for the reference sediment is effective because the reference is involved in every comparison, whereas the dredged sediments are involved in only one comparison each.

D2.2.2 Analyses of Example Data

Table D-6 presents survival data from a hypothetical benthic toxicity test comparing survival from three dredged sediments with reference sediment survival. The example data are used to illustrate the steps in benthic toxicity data analysis, with numbers in parentheses in the text referring to numbered nodes in the decision tree (Figures D-4A,B). In this example, survival in the control (data not shown) was $\geq 90\%$, indicating the acceptability of the test (Figure D-4A,1). Mean survival in all dredged sediments was more than 10% below mean survival in the reference sediment, indicating that the significance of the reductions should be tested statistically (2). All data were arcsine-transformed prior to analyses (3). Data were analyzed using SAS program BENTOX (Section D4.2), and results for the analyses are given in Section D4.2.2.

Table D-6. Number of Survivors in a Hypothetical Benthic Toxicity Test.

Replicate ^a	Treatment			
	Reference	Sediment 1	Sediment 2	Sediment 3
1	20	17	15	17
2	20	16	16	12
3	19	18	13	10
4	19	17	17	16
5	20	15	11	13
Total	98	83	72	68
Mean	19.6 (98%)	16.6 (83%)	14.4 (72%)	13.6 (68%)
SE	0.24	0.51	1.08	1.29

^a 20 organisms per replicate at initiation of test

Tests of Assumptions

Following arcsine-transformation, the data were tested for normality (4) to determine whether parametric (Figure D-4A) or nonparametric (Figure D-4B) procedures should be used. Results of tests for normality (4) and equality of variances (5) are provided in Table D-7. The *P*-value for the Shapiro-Wilk's Test was 0.32, indicating no significant departure from normality because *P* exceeds 0.01 (α level in Table D-2 for $N=20$, balanced data). Bartlett's Test, Levene's Test, and F_{\max} all indicated that variances were not significantly different among groups, as all *P*-values were >0.10 (α level in Table D-2 for $n=5$, balanced data). Note that these three tests were included for the sake of comparison, but generally only one of them would be conducted. Because the data are normally distributed and variances are not significantly different, the LSD is the most appropriate test for comparing each dredged sediment to the reference (6).

Parametric Tests

Relevant results from the LSD test are provided in Table D-7 (note that LSD results are given separately for each dredged sediment-reference sediment comparison, but only one LSD test is actually performed, comparing each pair of sediments simultaneously). The *P*-values for the LSD comparisons of each sediment with the reference were all much less than 0.05; thus, we conclude that survival in each of the dredged sediments was significantly less than reference sediment survival (7). SAS output for the LSD test (Section D4.2.2) does not provide *t*-values and probabilities for the individual comparisons, and it is not necessary to calculate these.

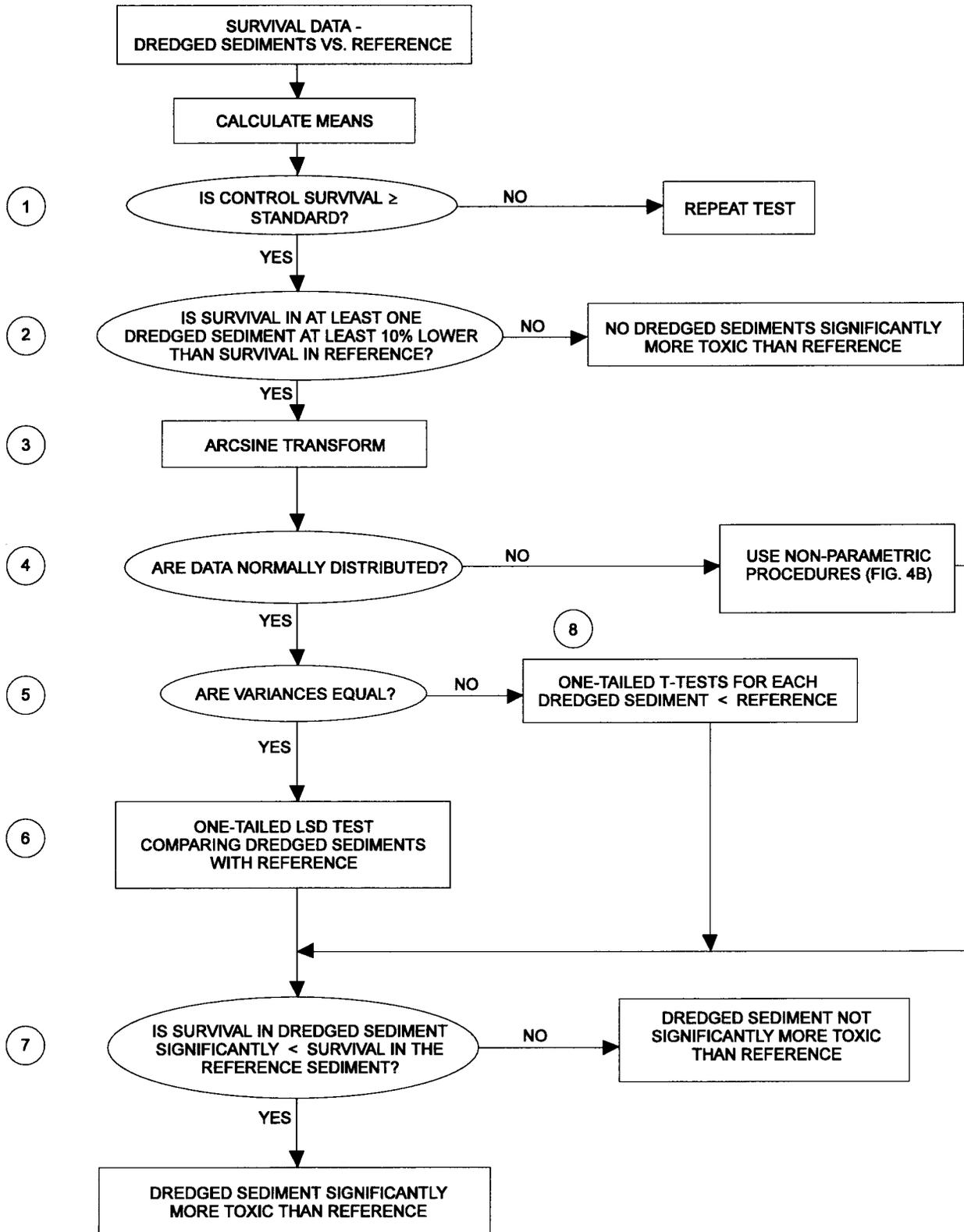


Figure D-4A. Benthic Toxicity Test Decision Tree (Parametric Tests).

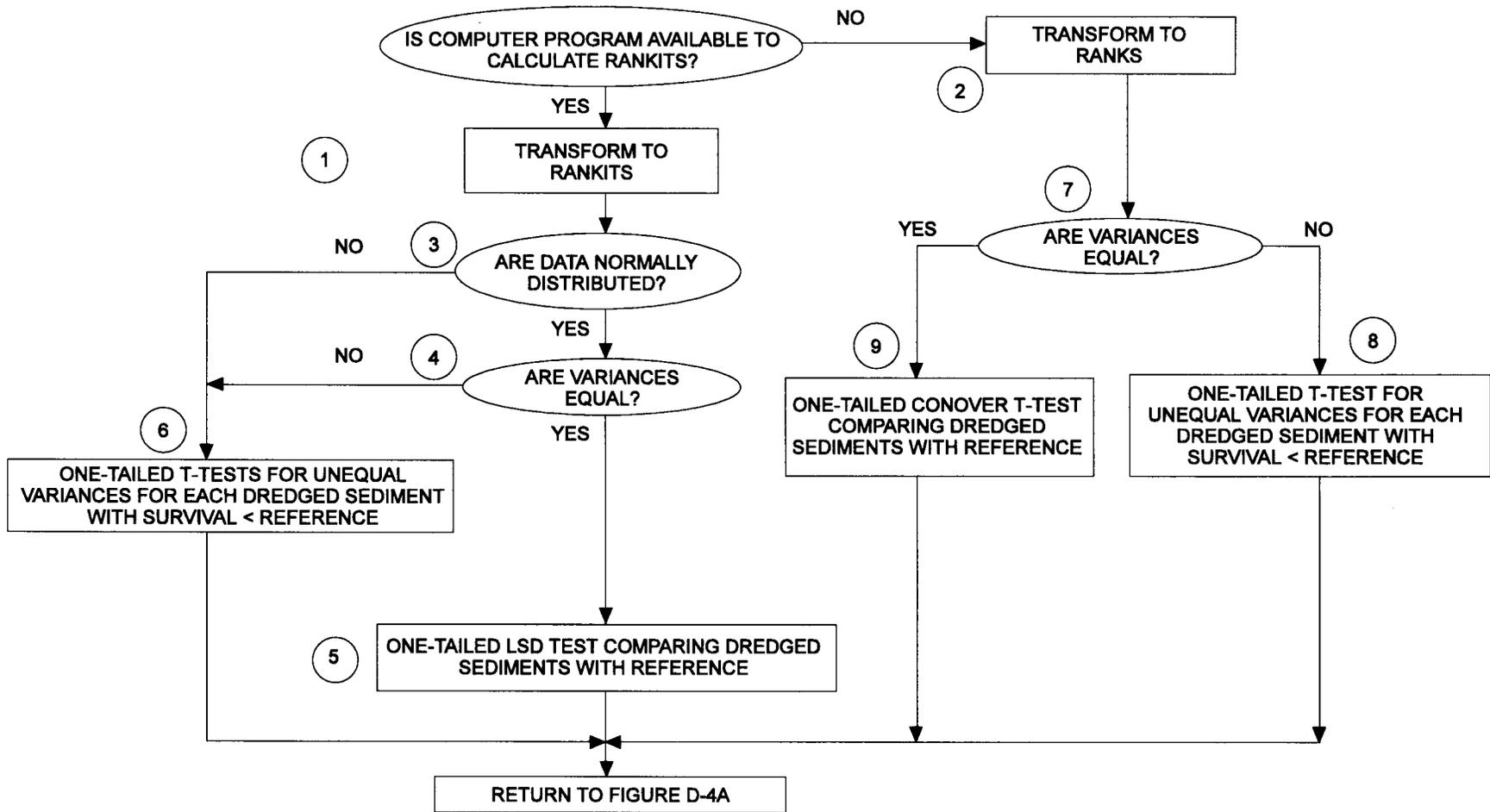


Figure D-4B. Benthic Toxicity Test Decision Tree (Nonparametric Tests).

Table D-7. Tests of Assumptions and Parametric Tests of Hypotheses on Arcsine-Transformed Benthic Toxicity Test Example Data.

Null Hypothesis: Mean Dredged Sediment Survival Equals Mean Reference Sediment Survival ^a				
Test	Test Statistic	Probability <i>P</i>	α	Conclusion
Normality Assumption: Shapiro-Wilk's Test	$W=0.946$	0.322	0.01	do not reject
Equality of Variances Assumption: Bartlett's Test	$F=0.6$	0.61	0.10	do not reject
Levene's Test	$F=1.74$	0.199	0.10	do not reject
F_{\max} Test	$F_{\max}=4.4$	>0.25	0.10	do not reject
Null Hypotheses: Sediment 1 = Reference <u>LSD Test</u>	<u>$t=4.11$</u>	<u>0.0017</u>	<u>0.05</u>	<u>reject</u>
t -Test (unequal variances)	$t=5.09$	0.0009	0.05	reject
Sediment 2 = Reference <u>LSD Test</u>	<u>$t=5.73$</u>	<u>0.0002</u>	<u>0.05</u>	<u>reject</u>
t -Test (unequal variances)	$t=5.63$	0.0003	0.05	reject
Sediment 3 = Reference <u>LSD Test</u>	<u>$t=6.25$</u>	<u>0.0001</u>	<u>0.05</u>	<u>reject</u>
t -Test (unequal variances)	$t=5.57$	0.0004	0.05	reject

^a Based on tests of assumptions, appropriate statistical tests of null hypotheses are underlined. Other test results are included for illustration only.

SAS indicates significant differences by using different letters under the "T Grouping" column. Mean reference survival was highest (A); mean survivals for sediments 1 (B) and 2 (BC) were significantly less than reference but not different from each other, and sediment 3 mean survival (C) was significantly lower than reference and sediment 1 but not sediment 2.

If the variances had been unequal, survival data would have been compared using t -tests (8). These results are included in Table D-7 for illustration. Again, the P -values indicate that all dredged sediment survivals were significantly less than reference sediment survival. Note that these P -values are one-half those given in the output from SAS program BENTOX in Section D4.2.2, because the SAS TTEST procedure returns two-tailed, rather than one-tailed probabilities.

Nonparametric Tests

Although the arcsine-transformed example data did not violate parametric hypothesis testing assumptions, nonparametric tests were performed to illustrate the steps in the nonparametric decision tree (Figure D-4B). The example data were converted using both rankits (1) and ranks (2), and the appropriate tests of assumptions were conducted (Table D-8). The rankits passed both the normality (3) and equality of variances (4) tests, so the next

step would be the LSD on rankits (5). Had either of these assumptions been violated, t -tests for unequal variances would have been performed on the rankits (6). If the ranks had failed the Levene's Test for equality of variances (7), t -tests for unequal variances would have been performed on the ranks (8), rather than the Conover T -Test (9). Results for all of these nonparametric hypothesis tests are shown in Table D-8. SAS Program BENTOX does not perform Levene's Test on ranks, the Conover T -Test, or 2-sample t -tests on ranks, as SAS can easily calculate rankits, and ranks-based tests would not be needed. The P -values for the nonparametric hypothesis tests in Table D-8 were in most cases slightly greater than those for the parametric tests, suggesting slightly lower power for the nonparametric tests. Nevertheless, all tests indicated that survival was significantly reduced in the dredged sediments compared to reference sediment survival. These results could easily have been predicted prior to analyses, because survival in the dredged sediment samples did not overlap with survival in the reference sediment samples (Table D-6).

Table D-8. Tests of Assumptions and Nonparametric Hypothesis Tests on Benthic Toxicity Test Example Data Converted to Rankits and Ranks.

Null Hypothesis: Median Dredged Sediment Survival Equals Median Reference Sediment Survival				
Test	Test Statistic	Probability P	α	Conclusion
Normality Assumption: Shapiro-Wilk's Test (rankits)	$W=0.982$	0.940	0.01	do not reject
Equality of Variances Assumption: Levene's Test (rankits)	$F=1.18$	0.349	0.10	do not reject
Levene's Test (ranks)	$F=2.25$	0.122	0.10	do not reject
Null Hypotheses:				
Sediment 1 = Reference				
LSD Test (rankits)	$t=3.05$	0.0079	0.05	reject
t -Test (rankits, unequal variances)	$t=4.57$	0.0011	0.05	reject
Conover T -Test (ranks)	$t=3.04$	0.0080	0.05	reject
t -Test (ranks, unequal variances)	$t=4.27$	0.0036	0.05	reject
Sediment 2 = Reference				
LSD Test (rankits)	$t=4.71$	0.0008	0.05	reject
t -Test (rankits, unequal variances)	$t=5.44$	0.0007	0.05	reject
Conover T -Test (ranks)	$t=4.90$	0.0006	0.05	reject
t -Test (ranks, unequal variances)	$t=5.80$	0.0012	0.05	reject
Sediment 3 = Reference				
LSD Test (rankits)	$t=5.28$	0.0004	0.05	reject
t -Test (rankits, unequal variances)	$t=4.91$	0.0019	0.05	reject
Conover T -Test (ranks)	$t=5.30$	0.0004	0.05	reject
t -Test (ranks, unequal variances)	$t=5.51$	0.0018	0.05	reject

Statistical Power

From Eq. 11, the minimum significant difference (d_{\min} , when $t_{1-\beta}=0$) for the parametric LSD test was:

$$d_{\min} = (t_{1-\alpha, \nu}) \sqrt{2MSE/n} \quad (\text{Eq. 17})$$

= $1.746[2(0.01618)/5]^{1/2} = 0.1405$, where $\nu=16$ df. Subtracting 0.1405 from the mean arcsine-transformed survival in the reference (1.481), and back-transforming gives 95%. That is, any survival less than 95% measured in a sample would be significantly lower than in the reference, and we would have a 0.50 probability of detecting a reduction in survival in any case where true population survival was 95%. Modifying Eq. 10, the probability (power or $1-\beta$) of detecting a difference if true population survival in a dredged sediment is <90% can be determined by:

$$t_{1-\beta, \nu} = d\sqrt{n/2MSE} - t_{1-\alpha, \nu} \quad (\text{Eq. 18})$$

= $(1.481 - 1.249) [5/2(0.01618)]^{1/2} - 1.746 = 1.138$. Using the SAS PROBT function to determine $1-\beta$ for $t=1.138$ with 16 df, power = 0.86. As with the water column toxicity test example data, the level of replication for the benthic toxicity example data is adequate to detect any reductions in survival that would be considered biologically significant. Investigators can expect lower reference survival and/or greater variance, and consequently less power, in real toxicity tests.

Suppose that we required an increase in power, but could not afford to add any more replicates. The optimal solution, assuming that variance could not be reduced by improving laboratory practices, would be to use 8 replicates for the reference, and 4 for each of the dredged sediments. The overall sample size remains 20. Note that a ratio of reference:dredged sediment replicates of 8:4 (2:1) is approximately equal to the optimal ratio of $\sqrt{k}:1$ or 1.73:1 ($k=3$ with 3 dredged sediments). Assuming that $MSE=0.01618$, as above, the minimum significant difference for an LSD test, again with 16 df, would be:

$$d_{\min} = (t_{1-\alpha, \nu}) \sqrt{MSE(1/n_1 + 1/n_2)} \quad (\text{Eq. 19})$$

= $1.746[0.01618(1/4 + 1/8)]^{1/2} = 0.1360$. This value is lower, although by <5%, than the minimum significant difference of 0.1405 for equal sample sizes of 5. The increase in power using the optimal ratio of reference:dredged sediment replicates will be greater when k is greater (more sediments tested).

SAS program BENTOX (Section D4.2) provides power calculations for the LSD test when true population survival from a dredged sediment is 10, 20, 30, 40 and 50 percent lower than mean reference sediment survival.

D3.0 BIOACCUMULATION

Bioaccumulation tests described in Section 12 are applied to determine whether an organism's exposure to the dredged material is likely to cause an elevation of contaminants in its tissues, i.e., bioaccumulation. Bioaccumulation tests may be conducted in the laboratory or in the field. Data analysis for these tests uses statistical procedures that have already been described for benthic toxicity test data analysis. These procedures are illustrated with example data in the following sections.

Statistical procedures for bioaccumulation data analysis in this Appendix cannot be applied directly in the common situation where some contaminant concentrations are reported only as less than some numerical detection limit (DL). The actual concentrations of these "censored" data (hereafter referred to as nondetects) are unknown and are presumed to fall between zero and the DL. Whenever possible, laboratories that analyze contaminant residues should be encouraged to report observed concentrations below DL (Porter et al., 1988), even though the precision of these measurements is less than that of above-DL measurements. When below-DL concentrations (sometimes called "J-values") are reported, they should be used as legitimate data in statistical comparisons. On the other hand, when bioaccumulation samples include nondetects, the unknown values must be replaced using a censored data method prior to statistical analysis. Recommended censored data methods are discussed in Sections D3.1.1.1 and D3.1.2.1.

D3.1 Tier III Single-Time Point Laboratory Bioaccumulation Study

The Tier III single-time point laboratory bioaccumulation test produces tissue concentration measurements for each contaminant of concern. Table D-9 presents example results for one contaminant from a hypothetical laboratory test. Chemical analysis of the tissue samples from each replicate shows that concentrations of the example contaminant varied among and within sediments. Two types of analyses may be performed on these data:

- comparisons between each dredged sediment and the reference, and
- comparisons with an action level when applicable.

Although Section 6.3 stipulates that applicable comparisons with an action level be conducted first, the statistical analysis can be performed more efficiently if comparisons with the reference are done first. Computer procedures for statistical analysis of single-time point bioaccumulation data are given in SAS program BIOACC (Section D4.3).

Table D-9. Results from a Hypothetical Single-Time Point Bioaccumulation Test, Showing Contaminant Concentrations ($\mu\text{g/g}$) in Tissues of Animals Exposed to Different Treatments.

Replicate	Treatment			
	Reference	Sediment 1	Sediment 2	Sediment 3
1	0.06	0.16	0.24	0.13
2	0.05	0.19	0.10	0.05
3	0.05	0.18	0.13	0.17
4	0.08	0.22	0.18	0.08
5	0.09	0.31	0.30	0.22
Mean	0.066	0.212	0.190	0.130
SE	0.008	0.026	0.036	0.030

D3.1.1 Comparisons with a Reference Sediment

Analysis of the example data follows the decision tree steps in Figures D-5A and 5B, with numbers in parentheses in the text referring to numbered nodes of the decision trees. The objective of this type of analysis is to determine whether organisms exposed to the dredged sediments accumulate greater tissue contaminant levels than organisms exposed to the reference sediment. One-sided tests are appropriate because there is little concern if bioaccumulation from a dredged sediment is less than bioaccumulation from the reference sediment. If mean tissue concentrations of contaminants of concern in organisms exposed to a dredged sediment are less than or equal to those of organisms exposed to the reference sediment (I), the dredged sediment meets the guidelines (Section 6.3), and no statistical analysis is required.

If only one dredged sediment is compared to the reference, then the procedures described in Section D2.1.1.1 (tests of assumptions followed by a t -test using a transformation or rankits if necessary) for comparing two samples are used. If more than one sediment is compared to the reference, then the procedures described in Section D2.2.1 (tests of assumptions followed by LSD, t -tests, or nonparametric equivalents) are used. Because contaminant concentration data are not easily expressed as proportions, the arcsine transformation is not appropriate. The raw data are analyzed first and, if necessary, a logarithmic (either natural or base 10) transformation may be employed. Although other transformations (such as square root) are possible, we recommend the log transformation because contaminant concentration data often follow a lognormal distribution. As always, tests of assumptions must be rerun on the data following transformation. If the transformed data violate the normality assumption, then data are converted to rankits (or ranks) and the assumptions are retested.

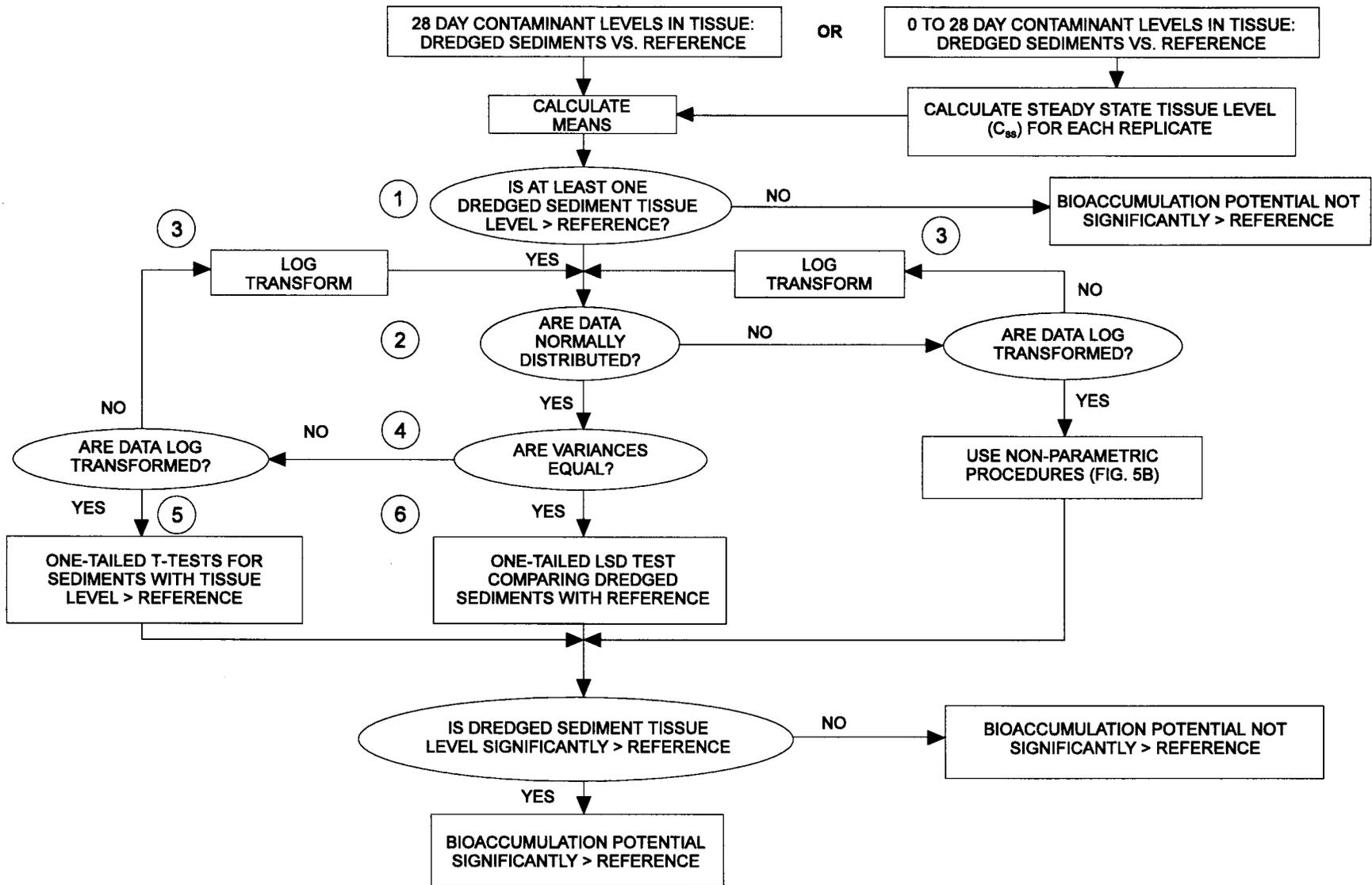


Figure D-5A. Bioaccumulation Test Decision Tree (Parametric Tests).

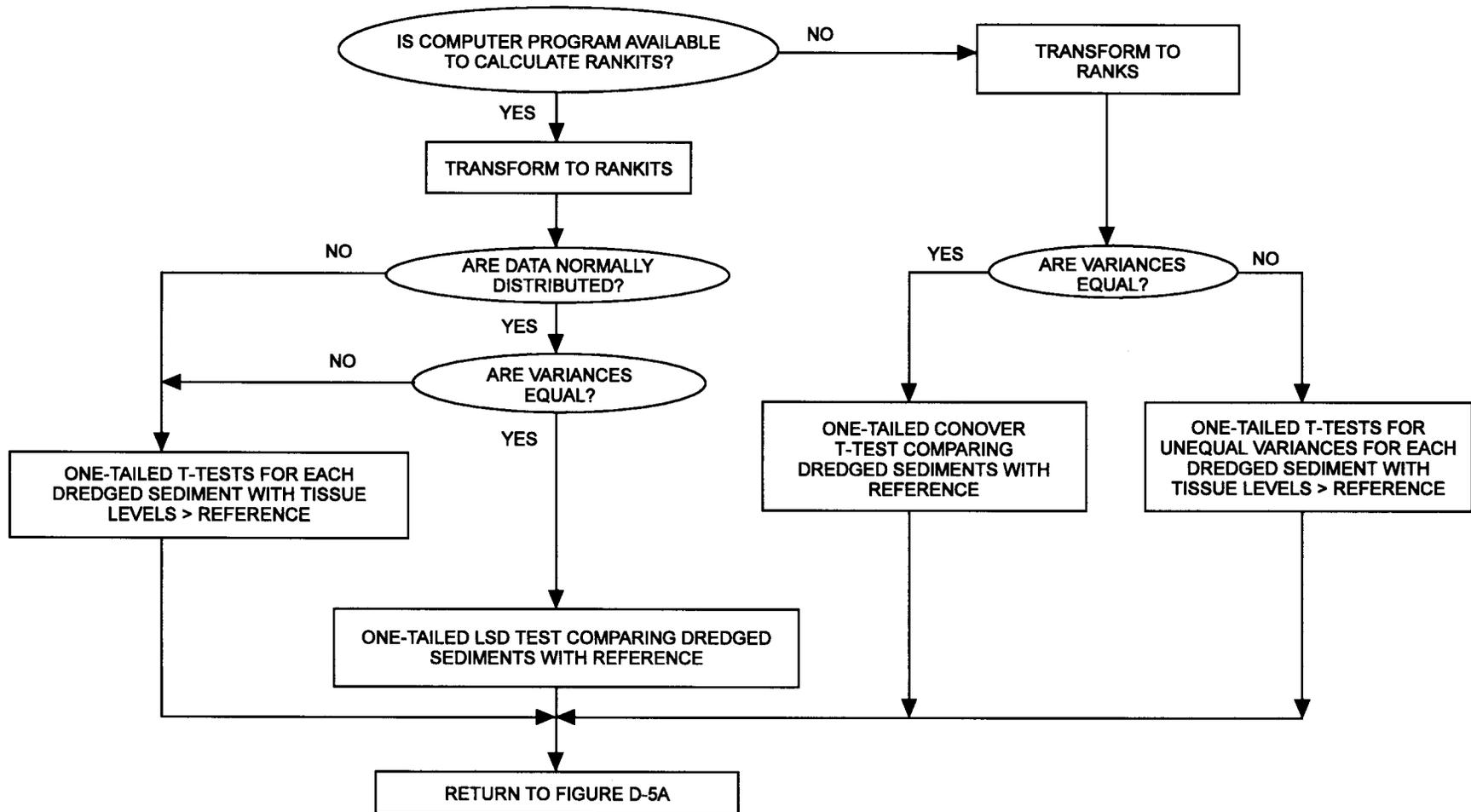


Figure D-5B. Bioaccumulation Test Decision Tree (Nonparametric Tests).

The data in Table D-9 were analyzed using SAS program BIOACC (Section D4.3), and the results are reported in Tables D-10 and D-11. The probability value for Shapiro-Wilk's Test (2) was >0.01 (α level in Table D-2 for $N=20$, balanced data), indicating no significant departure from normality. If the raw data had failed the normality test, then a log transformation (3) would be applied and the Shapiro-Wilk's Test rerun (2). If the log-transformed data still departed significantly from normality, then nonparametric hypothesis testing procedures would be performed (Figure D-5B); these procedures are described in Section D2.2.1.

The P -value for Levene's Test (4) was >0.10 (α level in Table D-2, $n=5$, balanced data), indicating that assumption of equality of variances need not be rejected for the raw data. If the variances had been significantly unequal, a log transformation would have been applied (3) and the tests of assumptions (2,4) rerun. Data that passed the normality test but failed the test for equality of variances would be analyzed using a t -test for each dredged sediment-reference sediment comparison (5).

Because the example data passed both tests of assumptions, the LSD (6) was conducted on the untransformed data to compare bioaccumulation from each dredged sediment with bioaccumulation from the reference sediment. LSD results indicated that mean tissue levels for organisms exposed to dredged sediments 1 and 2 (but not 3) were significantly greater than mean tissue levels for organisms exposed to the reference sediment (Table D-10).

For the sake of illustration, Table D-10 also includes results for log-transformed example data and for t -tests. Table D-11 gives nonparametric test results for the example data. Note that the different statistical tests give conflicting hypothesis test conclusions for the sediment 3-reference sediment comparison, because the P -values of the tests are close to α . This situation will often arise in the analysis of actual bioaccumulation data. Once again, *it is not acceptable to conduct several different statistical tests in order to choose the results one prefers*. For dredged sediment disposal evaluations, the decision trees in this Appendix should be followed to determine the appropriate statistical procedures in any given situation. In the case of the example data, the tests of assumptions indicate that the appropriate hypothesis testing procedure is the LSD test using untransformed data, and the results of this test should be accepted. However, in making decisions concerning disposal, it is entirely appropriate to consider that the significance of the sediment 3-reference sediment comparison is marginal. The power of the LSD test (calculated below) should also be taken into consideration.

Power calculations for the example data are performed on the untransformed data. Using Eq. 17, the minimum significant difference for the parametric LSD test was:

$$d_{\min} = 1.746[2(0.003763)/5]^{1/2} = 0.0677 \text{ } \mu\text{g/g.}$$

SAS conveniently provides this value as the "Least Significant Difference" in the GLM or ANOVA procedures when the LSD test is requested (and sample sizes are equal).

Table D-10. Tests of Assumptions and Parametric Hypothesis Tests on Untransformed and Log₁₀-Transformed Bioaccumulation Example Data.

Null Hypothesis: Mean Dredged Sediment Bioaccumulation Equals Mean Reference Sediment Bioaccumulation ^a				
Test	Test Statistic	Probability <i>P</i>	α	Conclusion
Normality Assumption:				
Shapiro-Wilk's Test				
Untransformed data	W=0.958	0.511	0.01	do not reject
Log-transformed data	W=0.980	0.921	0.01	do not reject
Equality of Variances Assumption:				
Levene's Test				
Untransformed data	F=2.15	0.134	0.10	do not reject
Log-transformed data	F=2.19	0.129	0.10	do not reject
Null Hypotheses:				
Sediment 1 = Reference				
<u>LSD Test</u>				
<u>Untransformed data</u>	<u>t=3.76</u>	<u>0.0028</u>	<u>0.05</u>	<u>reject</u>
Log-transformed data	t=4.45	0.0011	0.05	reject
<i>t</i> -Test (unequal variances)				
Untransformed data	t=5.30	0.0020	0.05	reject
Log-transformed data	t=7.04	<0.0001	0.05	reject
Sediment 2 = Reference				
<u>LSD Test</u>				
<u>Untransformed data</u>	<u>t=3.20</u>	<u>0.0063</u>	<u>0.05</u>	<u>reject</u>
Log-transformed data	t=3.84	0.0025	0.05	reject
<i>t</i> -Test (unequal variances)				
Untransformed data	t=3.33	0.0129	0.05	reject
Log-transformed data	t=4.34	0.0020	0.05	reject
Sediment 3 = Reference				
<u>LSD Test</u>				
<u>Untransformed data</u>	<u>t=1.65</u>	<u>0.0688</u>	<u>0.05</u>	<u>do not reject</u>
Log-transformed data	t=2.20	0.0295	0.05	reject
<i>t</i> -Test (unequal variances)				
Untransformed data	t=2.03	0.0523	0.05	do not reject
Log-transformed data	t=1.98	0.0495	0.05	reject

^a Based on tests of assumptions, appropriate statistical tests of null hypotheses are underlined. Other test results are included for illustration only.

Table D-11. Tests of Assumptions and Nonparametric Hypothesis Tests on Bioaccumulation Example Data Converted to Rankits and Ranks.

Null Hypothesis: Median Dredged Sediment Bioaccumulation Equals Median Reference Sediment Bioaccumulation				
Test	Test Statistic	Probability <i>P</i>	α	Conclusion
Normality Assumption: Shapiro-Wilk's Test (rankits)	<i>W</i> =0.972	0.791	0.01	do not reject
Equality of Variances Assumption: Levene's Test (rankits)	<i>F</i> =0.61	0.621	0.10	do not reject
(ranks)	<i>F</i> =1.57	0.236	0.10	do not reject
Null Hypotheses:				
Sediment 1 = Reference				
LSD Test (rankits)	<i>t</i> =3.87	0.0024	0.05	reject
<i>t</i> -Test (rankits, unequal variances)	<i>t</i> =4.69	0.0011	0.05	reject
Conover <i>T</i> -Test	<i>t</i> =4.14	0.0016	0.05	reject
<i>t</i> -Test (ranks, unequal variances)	<i>t</i> =6.18	0.0003	0.05	reject
Sediment 2 = Reference				
LSD Test (rankits)	<i>t</i> =3.32	0.0053	0.05	reject
<i>t</i> -Test (rankits, unequal variances)	<i>t</i> =3.76	0.0040	0.05	reject
Conover <i>T</i> -Test	<i>t</i> =3.54	0.0038	0.05	reject
<i>t</i> -Test (ranks, unequal variances)	<i>t</i> =3.95	0.0046	0.05	reject
Sediment 3 = Reference				
LSD Test (rankits)	<i>t</i> =1.66	0.0677	0.05	do not reject
<i>t</i> -Test (rankits, unequal variances)	<i>t</i> =1.69	0.0706	0.05	do not reject
Conover <i>T</i> -Test	<i>t</i> =1.86	0.0497	0.05	reject
<i>t</i> -Test (ranks, unequal variances)	<i>t</i> =1.85	0.1215	0.05	do not reject

Using Eq. 18, the power of the LSD test for detecting a 100% increase in dredged sediment bioaccumulation over the mean reference bioaccumulation (i.e., $d=0.066 \mu\text{g/g}$) can be determined by:

$$t_{1-\beta} = (0.066) [5/2(0.003763)]^{1/2} - 1.746 = -0.045$$

and $1-\beta$ for $t=-0.045$ with 16 df is 0.48. Power values for 10, 25, 50, 100, 200 and 300% increases over mean reference bioaccumulation are given in the output for SAS program BIOACC (Section D4.3.2).

The sample size (n) required to provide a 0.95 probability ($1-\beta=0.95$) of detecting a 25% increase ($0.0165 \mu\text{g/g}$) over the mean reference bioaccumulation, calculated using the z -approximation (Eq. 9) with MSE replacing s^2 , is:

$$n = 2(1.645 + 1.645)^2 [0.003763 / (0.0165)^2] + 0.25(1.645)^2 = 300 !$$

Using the same equation, to detect a 100% increase ($0.066 \mu\text{g/g}$) over the mean reference bioaccumulation with a power of 0.95, $n = 20$. Assuming we are limited to 5 replicates, there is a 0.95 probability of detecting a

difference (d) of 0.135 $\mu\text{g/g}$, which is a 205% increase over the mean reference bioaccumulation. Other values of d when power = 0.5, 0.6, 0.7, 0.8, 0.9, and 0.99 are given in the output for SAS program BIOACC (Section D4.3.2).

D3.1.1.1 Less Than Detection Limit Data

A number of methods can be used to permit statistical comparisons of censored data. A simulation study was conducted to identify which of 10 censored data methods work best to maintain power and minimize α in LSD comparisons when n is small, for various situations depending on type of frequency distribution, equality or inequality of variances, coefficient of variation (CV), and amount of censoring (Clarke, 1995a). The 10 censored data methods include three simple substitution methods, two uniform distribution substitution methods, three maximum likelihood methods, and two regression methods. General results from all simulations combined indicate that the simple substitution methods perform as well as or better than the more complicated censored data techniques in most situations (Clarke, 1994). In particular, substitution of the detection limit when up to 40 percent of the data are nondetects, or one-half the detection limit when more than 40 percent of the data are nondetects, are methods that work reasonably well for small sample sizes in most cases.² These methods are not limited to untransformed data, but may also be used when data will subsequently be log-transformed or converted to rankits.

Nevertheless, the simulations have shown that substitution of the detection limit or half the detection limit are not the most advantageous censored data methods for all combinations of frequency distribution and variance characteristics. Detailed guidelines for statistical treatment of less than detection limit data developed from the simulation study are described in Clarke (1995b); investigators wishing to maximize the effectiveness of statistical comparisons that include nondetects are encouraged to read this publication carefully. The guidelines are summarized below; the recommendations table from Clarke (1995b) is condensed as Table D-12 and includes the following methods:

- **DL.** Substitution of the detection limit for all nondetects.
- **DL/2.** Substitution of one-half the detection limit for all nondetects.
- **ZERO.** Substitution of zero for all nondetects.

When data are subsequently transformed to rankits, the above three methods produce the exact same results (assuming all uncensored observations in the sample are $> \text{DL}$), and are called **CONST** for substitution of any constant between 0 and DL.

²Power will generally decline as censoring increases; when the data are more than 60 to 80 percent nondetects, it is unlikely that any method will perform acceptably.

Table D-12. Recommended Censored Data Methods for Small Samples to be Used in Statistical Comparisons.

Amount of Censoring	Variances	Coefficient of Variation	Data Transformation (Distribution)		
			Log (Lognormal)	None (Normal)	Rankit (Nonnormal)
≤20 %	Equal	≤0.25	<i>DL</i> ^a	DL	CONST, UNIF
		0.26 - 1	<i>DL/2, DL</i>	<i>DL/2, ZERO</i>	CONST, UNIF
		0.51 - 1	<i>DL/2, DL</i>	<i>ZERO, DL/2</i>	CONST, UNIF
		>1	<i>DL/2, DL</i>	-- ^b	CONST, UNIF
	Increase as Means Increase		<i>DL, DL/2</i>	<i>LR, DL/2</i>	CONST, UNIF
	Mixed		<i>DL</i> ^c	<i>DL, DL/2</i>	CONST, UNIF ^c
21 - 40 %	Equal	≤0.25	DL	DL	CONST, UNIF
		0.26 - 1	<i>DL/2</i>	<i>DL/2, ZERO</i>	CONST, UNIF
		>1	<i>DL/2, DL</i>	-- ^b	CONST, UNIF
	Increase as Means Increase		<i>DL/2, DL</i>	<i>DL, DL/2</i>	CONST, UNIF
	Mixed		<i>DL</i>	<i>ZERO, DL/2</i> ^c	CONST, UNIF
41 - 60 %	Equal	≤0.25	<i>DL/2, DL</i>	<i>DL/2, ZERO</i>	CONST
		>0.25	<i>DL/2</i>	<i>DL/2, ZERO</i>	CONST
	Increase as Means Increase		<i>DL/2</i>	<i>DL/2, ZERO</i>	CONST
	Mixed		<i>DL/2</i>	-- ^d	CONST
61 - 80 %	Equal	≤0.25	<i>DL/2, DL</i>	<i>DL/2</i>	CONST
		0.26 - 1	<i>DL/2</i>	<i>DL/2, ZERO</i>	CONST
		>1	<i>DL/2</i> ^c	-- ^b	-- ^d
	Increase as Means Increase		<i>DL/2</i>	<i>DL/2, ZERO</i>	CONST
	Mixed		<i>DL/2</i> ^c	-- ^d	CONST ^c

^a Non-italicized methods have been $\alpha < 0.06$; italicized methods have been α between 0.06 and 0.10

^b When coefficient of variation > 1 , normal distribution is unlikely for chemical concentration data due to increasing proportion of negative values

^c All methods with acceptable power have $\alpha \geq 0.06$

^d All methods have unacceptably low power and/or high α

- **UNIF.** Nondetects are replaced by ordered observations x_i ($i = 1, 2 \dots nc$, where nc is the number of censored observations in the sample) between 0 and DL, where

$$x_i = DL(i - 1)/(nc - 1)$$

and $x_i = DL/2$ when $nc = 1$.

- **LR.** Substitution of estimated values from a lognormal distribution using linear regression of logarithms of above-DL concentrations vs their rankits. The regression equation is used to extrapolate values for which antilogs are taken to replace the nondetects. This method (called Helsel's Robust Method) is available in a software package called UNCENSOR³ (Newman and Dixon, 1990).

SAS program statements for DL, DL/2, ZERO, UNIF, and LR are given in Section D4.5.

Deletion of nondetects is not recommended as it results in excessive loss of information and power as amount of censoring increases.

The following steps should be used to select the best censored data method in a given situation. For each contaminant reporting nondetects:

- Determine proportion of data that are censored (all samples combined).
- Determine characteristics of the variances and statistical data distribution for the contaminant of interest. This can be done if the data are not severely censored by applying two or more censored data methods to obtain a range of possible variances and CVs. Alternatively, one might use uncensored sample data for the same or similar contaminants, or historical data for the same contaminant from the same area.
 - Determine whether variances are equal or unequal among samples (Section D2.1.1.1). If unequal, do the variances increase as means increase, or are the variances seemingly random (mixed)?
 - Calculate CV of combined samples, where $CV = s / \bar{x}$
 - Determine whether combined sample residuals are distributed normally, lognormally, or nonnormally (Section D2.1.1.1). If $CV \geq 1$, data are unlikely to be distributed normally as such a population would include a fair proportion of negative concentrations; therefore, assume lognormal or nonnormal distribution

³A public domain program that can be obtained from Dr. Michael C. Newman, University of Georgia Savannah River Ecology Laboratory, P.O. Drawer E, Aiken, SC 29801.

- Refer to Table D-12 to determine most appropriate method given approximate amount of data censoring, properties of variances and type of statistical distribution
- If it is crucial to maintain α at approximately 0.05 or less, choose non-italicized methods where available in Table D-12
- Apply selected method to censored data, then continue with tests of assumptions and statistical comparison procedures as outlined in Section D3.1.1. Avoid a data transformation for which no method is given in Table D-12 due to low power or excessively high α
- Do not attempt statistical comparisons of severely censored samples in situations where no censored data methods are considered appropriate. In such cases, the probability of an erroneous outcome is high.

It is quite possible that an evaluation including a number of sediments and contaminants would produce comparisons involving several different combinations of censoring proportions, variance characteristics and data frequency distributions. Following the guidelines herein would result in the application of more than one censored data method to the project data. This is entirely acceptable when the censored data methods are selected for the purpose of maximizing power and minimizing type I error. *What is not acceptable is to try several censored data methods for the purpose of finding one that will produce a desired statistical comparison outcome.*

The simulation study did not address the performance of censored data methods in the common situation of multiple detection limits within a set of replicate observations. Until guidelines are developed for analysis of multiple detection limits, the same procedures should be followed as for single detection limits. SAS programs for the censored data methods can be applied without modification to multiple detection limit censored samples.

D3.1.2 Comparisons with an Action Level

In this comparison, the objective is to determine whether the mean bioaccumulation of contaminants in animals exposed to a dredged sediment is significantly less than a specified action level or standard. If the mean tissue concentration of one or more contaminants of concern is greater than or equal to the applicable action level, then no statistical testing is required. The conclusion would be that the dredged sediment does not meet the guidelines associated with the action level (Section 6.3). If the mean tissue concentrations of a contaminant of concern are less than the applicable action level, then a confidence-interval approach is used to determine if these means are *significantly* less than the action level. One-sided tests are appropriate since there is concern only if bioaccumulation from the dredged sediment is not significantly less than the action level. There are two different approaches to conducting these tests, and both are acceptable.

The first approach is to calculate a value of t , much as in a t -test (this approach is often called a one-sample t -test):

$$t = \frac{\bar{x} - \text{action level}}{\sqrt{s^2/n}} , \quad (\text{Eq. 20})$$

where \bar{x} , s^2 and n refer to mean, variance, and number of replicates for contaminant bioaccumulation from the dredged sediment.

If tests of equality of variances in the comparison of dredged sediments with the reference indicate that variances are equal for all sediments, then MSE from the ANOVA is used as s^2 , and calculated t is compared to $t_{0.95}$, with $N - k$ degrees of freedom. If the variances are not equal, then s^2 for the individual sediment is used, and calculated t compared with $t_{0.95}$, with $n - 1$ degrees of freedom. If the data were log-transformed to normalize the distributions or equalize variances, then all calculations should be carried out on log-transformed values.

Another approach is to calculate the upper one-sided 95% confidence limit (UCL), and compare it to the action level:

$$UCL = \bar{x} + (t_{0.95,v})(\sqrt{s^2/n}) . \quad (\text{Eq. 21})$$

As in the first approach, the MSE is used in place of s^2 if variances are not significantly different, and the degrees of freedom (v) are $N - k$. If variances are significantly different, s^2 for the individual sediment is used, and v for each sediment i is $n_i - 1$. There is a 0.95 probability that the true population mean tissue level is below the UCL . If the UCL is below the action level, there is a ≥ 0.95 probability that the population mean tissue level for the dredged sediment is below the action level, and we conclude that the action level is not exceeded. If the UCL is above the action level, we cannot be sure that the mean population tissue level does not exceed the action level.

Either of the above procedures may be used with data that have failed the normality test, but the results should be considered approximate.

The choice of which approach to use depends on the computer software and the presentation method to be used. In SAS, it is more convenient to calculate the UCL and compare with the action level, as in program BIOACC (Section D4.3). In SYSTAT, it is simpler to conduct a one-sample t -test. Both approaches can easily be performed by hand. If the data are presented graphically, as in Figure D-6, the confidence-level approach is used. If the investigator wants to provide the exact probability that the mean tissue level is less than the action level, then the one-sample t -test is used.

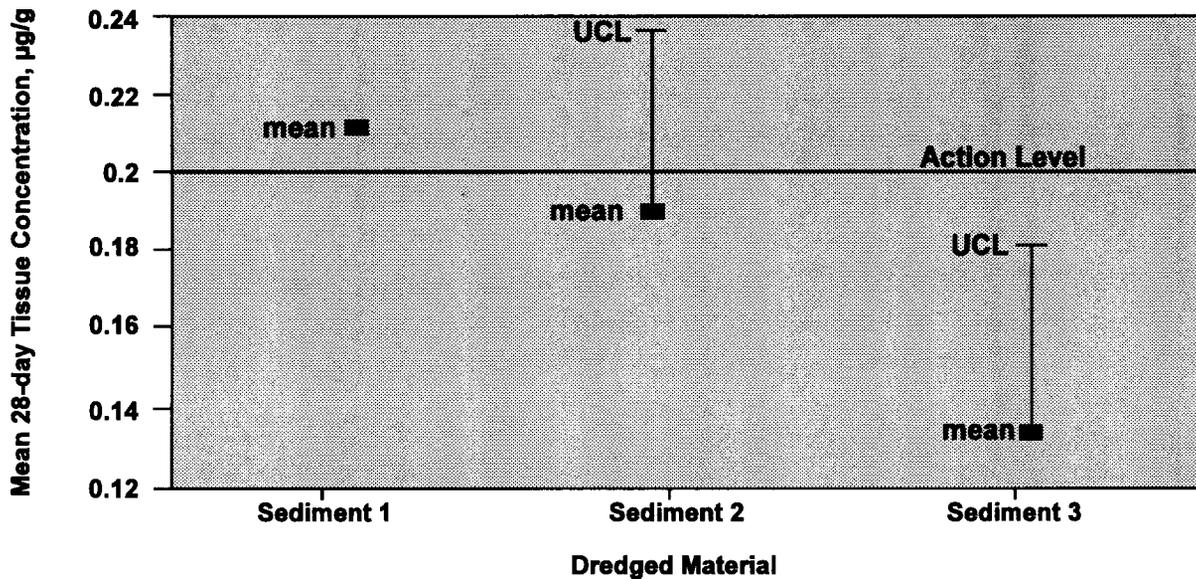


Figure D-6 Comparison of Mean Dredged Sediment Contaminant Tissue Levels (mean) and 95% Upper Confidence Level (UCL) with Hypothetical Action Level.

Figure D-6 presents a comparison of mean bioaccumulation from the three dredged sediments (see Table D-9) with a hypothetical action level of 0.2 µg/g. There is no need to calculate the UCL for sediment 1 as the mean exceeds the action level. Because variances were not significantly different for the untransformed data (Table D-10), we use $MSE=0.003763$ and $t_{0.95,16}=1.746$ in Eq. 21 to obtain:

$$UCL = 0.190 + 1.746(0.003763/5)^{1/2} = 0.238$$

for sediment 2, and $UCL = 0.178$ for sediment 3. SAS program BIOACC (Section D4.3) calculates UCL for both equal and unequal variances.

If the UCL lies below the action level, there is a >0.95 probability that the true population mean tissue level for that sediment is less than the action level. Thus, we would conclude that mean bioaccumulation for dredged sediment 3 is less than the action level. Because the UCL for sediment 2 exceeds the action level even though the sample mean does not, we cannot be sure that the true population mean tissue level for this sediment is less than the action level.

Formulae for calculating statistical power for comparisons to a fixed standard such as an action level are very similar to Eqs. 8 and 9:

$$n = (t_{1-\alpha, v} + t_{1-\beta, v})^2 (s^2/d^2), \quad (\text{Eq. 22})$$

where s^2 and v (degrees of freedom) are MSE and $N - k$ if variances are equal (or expected to be equal, if the calculation is made prior to testing), and s^2 for the individual sediment and $n_i - 1$ if variances are unequal. It is usually easier to use the z -approximation (from Alldredge, 1987) to avoid solving for n iteratively:

$$n = (z_{1-\alpha} + z_{1-\beta})^2 (s^2/d^2) + 0.5(z_{1-\alpha}^2) \quad . \quad (\text{Eq. 23})$$

The formulae indicate that the sample size required to detect any given difference d will be approximately one-half that required for a comparison of two treatments. The sample size required is lower because the comparison is made to a fixed value, rather than to a reference which can also vary. Thus, there is no sampling uncertainty or error for the fixed standard and we know the true value of one of the two things being compared.

Using the z -approximation and $s^2=MSE$, the sample size required to provide a 0.95 probability ($1-\beta=0.95$) of detecting a tissue level 25% ($0.05 \mu\text{g/g}$) below the action level is:

$$n = (1.645 + 1.645)^2(0.003763)/0.0025 + 0.5(1.645)^2 = 18.$$

The minimum significant difference is:

$$d_{\min} = t_{0.95,16}(MSE/n)^{1/2} = 1.746(0.003763/5)^{1/2} = 0.048 \mu\text{g/g}.$$

The power of a comparison can be determined by:

$$t_{1-\beta} = \frac{d\sqrt{n}}{s} - t_{1-\alpha,v} \quad , \quad (\text{Eq. 24})$$

When variances are not significantly different, s is replaced by $(MSE)^{1/2}$ and $v = N - k$ df. Using $MSE=0.003763$ as above, the power to detect a 10% decrease in mean bioaccumulation below the action level is 0.16, and power to detect a 50% decrease is 0.96. Power for 10, 20, 30, 40 and 50% decreases are given in the output for SAS program BIOACC (Section D4.3.2).

D3.1.2.1 Less Than Detection Limit Data

Recommendations for censored data methods in Table D-12 were developed to facilitate comparisons of two or more samples. When a sample of contaminant bioaccumulation concentrations must be compared with an action level or standard, accurate estimation of the sample mean and standard deviation is important. In general, this may require different censored data methods than does the comparison of samples in the previous section. Most recommendations for censored data methods in estimation

problems have been based on relatively large sample sizes ($n=10$ or more). Gleit (1985) identified certain methods that perform better than others for estimating the mean and variance of normal populations based on samples of $n=5$. The best methods, depending on mean, CV, and amount of censoring, included DL, DL/2, ZERO, and an iterative method using expected values of order statistics. The latter method (which Gleit recommended), along with several others including LR and some MLE techniques, are available in UNCENSOR (Newman and Dixon, 1990).

Recommendations for censored data methods for estimating mean and standard deviation when n is small are provided in the Applications Guide as a supplement to this Appendix (Clarke and Brandon, in press). If zero is substituted for all nondetects and the sample mean is greater than or equal to the applicable action level, then clearly no statistical testing is required as the mean contaminant concentration cannot be less than the action level.

D3.2 Tier IV Time-Sequenced Laboratory Bioaccumulation Study

The time-sequenced laboratory bioaccumulation test in Tier IV is designed to detect differences, if any, between steady-state bioaccumulation in organisms exposed to the dredged sediments and steady-state bioaccumulation in organisms exposed to the reference sediment. If organisms are exposed to biologically available contaminants under constant conditions for a sufficient period of time, bioaccumulation will eventually reach a steady state in which maximum bioaccumulation has occurred, and the net exchange of contaminant between sediment and organism is zero.

A simple kinetic model (McFarland and Clarke, 1987; Clarke and McFarland, 1991) can be used with data collected over a relatively short period of constant exposure to project tissue concentrations at steady state. This model integrated for constant exposure is:

$$C_t = \frac{k_1 C_w}{k_2} (1 - e^{-k_2 t}) \quad , \quad (\text{Eq. 25})$$

where C_t = concentration of a compound in tissues of an organism at time t ,

k_1 = uptake rate constant,

C_w = exposure (water) concentration of the compound,

k_2 = elimination rate constant, and

t = time in days.

Using this model, contaminant uptake occurs rapidly at first, and then the rate of uptake gradually diminishes as uptake begins to level off and approach an asymptote (steady state).

As duration of exposure increases, the exponential term in the model approaches zero, and the tissue concentration at steady state (i.e., infinite exposure) is calculated as:

$$C_t = \frac{k_1 C_w}{k_2} = C_{ss} \quad , \quad (\text{Eq. 26})$$

where C_{ss} is an estimate of the whole-body concentration of the compound at steady state.

Steady-state concentration estimates from organisms exposed to dredged sediments are compared to applicable action levels and to steady-state estimates from organisms exposed to the reference sediment. The data analysis involves several steps:

1. Calculate a separate nonlinear regression for each replicate using Eq. 25.
2. Use the regression coefficients (k_1 and k_2) to calculate the steady-state concentrations (C_{ss}) from Eq. 26, or set up the regression analysis to estimate/output C_{ss} directly (see below).
3. Use the estimates of C_{ss} as data in a statistical test comparing each dredged sediment to the reference (as in Section D3.1.1). Conclusions possible from these comparisons and evaluative factors that should be assessed are detailed in Section 6.3.
4. Use confidence intervals or one-sample t -tests to compare the steady-state estimates with applicable action levels (as in Section D3.1.2).

If nondetects occur in the early days of uptake, values may be assigned to them using a censored data method such as DL/2. If nondetects occur in the later portion of uptake, or if all of the bioaccumulation data for a replicate are near the detection limit, then the data probably do not fit the simple kinetic model and that replicate should be dropped from the analysis.

D3.2.1 Calculating Steady-State Concentrations

Table D-13 presents example data resulting from a hypothetical 28-day time-sequenced laboratory bioaccumulation test using three dredged sediments and a reference sediment. There are five replicates of each treatment, and tissue samples were analyzed on days 2, 4, 7, 10, 18, and 28 of the test. More sampling days are scheduled in the early part of the test to enable more accurate characterization of the early, rapidly changing portion of the uptake curve.

Table D-13. Results from a Hypothetical Time-Sequenced Bioaccumulation Test, Showing Contaminant Concentrations ($\mu\text{g/g}$) in Tissues of Animals Exposed to Different Treatments.

Replicate	Day	Treatment			
		Reference	Sediment 1	Sediment 2	Sediment 3
1	2	0.054	0.159	0.869	0.745
	4	0.441	0.516	0.838	1.316
	7	0.687	0.881	1.246	1.583
	10	0.037	0.278	1.767	1.578
	18	0.856	0.904	1.631	2.822
	28	0.514	0.172	1.178	1.295
2	2	0.163	0.292	0.726	1.703
	4	0.797	0.158	0.633	0.930
	7	0.177	0.317	0.816	2.715
	10	0.549	0.485	1.272	2.268
	18	0.598	1.300	1.877	2.607
	28	0.839	1.049	1.721	2.964
3	2	0.391	0.428	0.394	2.045
	4	0.203	0.743	0.452	2.141
	7	0.862	0.270	0.897	1.016
	10	0.884	0.051	1.003	1.756
	18	0.016	0.671	1.487	3.414
	28	0.793	0.476	1.366	2.109
4	2	0.234	0.558	1.232	1.855
	4	0.564	0.324	0.728	1.150
	7	0.413	0.562	1.639	2.221
	10	0.787	0.909	1.158	2.899
	18	0.806	0.934	1.216	1.319
	28	0.899	0.712	1.513	2.820
5	2	0.034	0.256	0.977	1.135
	4	0.018	0.126	1.314	1.621
	7	0.029	0.603	0.688	2.134
	10	0.294	0.718	1.415	0.890
	18	0.119	1.173	1.280	1.866
	28	0.226	1.245	1.843	3.325
Mean Sediment Concentration		0.45	4.0	33.0	44.0

These data can be used with iterative nonlinear regression methods such as those in the SAS NLIN or SYSTAT NONLIN procedures to solve for the parameters (k_1 and k_2) in the model above. Then C_{ss} , the steady-state concentration, is simply $k_1 C_w / k_2$. In this iterative calculation method, the contaminant concentration in the sediment (C_s), or even a constant such as 1, may be used instead of C_w . This is because the values of the rate constants and the exposure concentration are not of interest in this application, only their ratio (i.e., C_{ss}). Thus, the equation could be written as:

$$C_t = C_{ss} (1 - e^{-k_2 t}) \quad , \quad (\text{Eq. 27})$$

and C_{ss} estimated directly by the regression software. The estimate of C_{ss} should be the same regardless of which approach is used. SAS program BIOACCSS (Section D4.4) performs the steady-state calculations using C_s , and outputs the regression parameters and C_{ss} for each replicate to a new data set. These are displayed in Table D-14.

Nonlinear regressions for the example data were calculated using the SAS NLIN procedure with the DUD method. This method does not require derivatives. Other methods may be used but the derivatives for the parameters (k_1 and k_2 , or C_{ss} and k_2 if C_{ss} is estimated directly) must be specified. The Marquardt method and the Gauss-Newton method produced results similar to DUD for the example data.

Iterative curve-fitting techniques will provide better fits to some data than to others, and the asymptotic relationship will not always be the best fit to the data. Thus, investigators should be aware of the following problems:

1. Failure to converge on a solution within the allowed number of iterations. Always have the regression software print out the results, even though the regressions are only used to create a new data set of C_{ss} values. SAS will output the parameter estimates from the final iteration, regardless of whether convergence occurred. If the last few iterations approach convergence (i.e., there is little change in parameter estimates and residual error mean square), then the parameter estimates from the last iteration may be used. If convergence was not approached, then the program should be run again for that replicate, using the parameter values from the last iteration as starting values.
 2. No relationship between concentration and time. This can occur in sediments with low or non-detectable contaminant levels. The model-derived estimate of C_{ss} will usually converge on the mean tissue concentration over all days.
 3. A non-asymptotic relationship. If the relationship between tissue levels and time is linear, rather than asymptotic, the estimated asymptote (C_{ss}) will approach infinity. A linear relationship will occur if the experiment was not conducted for a long enough time for the tissue levels to approach the asymptote, or because of anomalously high tissue levels later in the experiment. Always plot the data prior to calculating the regressions to make sure the relationships are asymptotic. SAS program
-

Table D-14. Regression Parameters Estimated from Example Time-Sequenced Bioaccumulation Data.

Treatment	C_s $\mu\text{g/g}$	Replicate	k_1	k_2	C_{ss}
Reference	0.45	1	0.237	0.176	0.608
		2	0.306	0.201	0.687
		3	0.540	0.407	0.597
		4	0.318	0.162	0.883
		5	0.045	0.087	0.234
Sediment 1	4.0	1	0.059	0.427	0.554
		2	0.019	0.047	1.644
		3	0.243	2.206	0.441
		4	0.051	0.243	0.833
		5	0.024	0.060	1.600
Sediment 2	33.0	1	0.014	0.319	1.488
		2	0.007	0.113	1.907
		3	0.006	0.120	1.511
		4	0.034	0.878	1.290
		5	0.023	0.568	1.350
Sediment 3	44.0	1	0.011	0.250	1.964
		2	0.015	0.236	2.776
		3	0.094	1.977	2.087
		4	0.024	0.458	2.259
		5	0.008	0.139	2.648

BIOACCSS (Section D4.4) provides separate plots for each treatment, with the replicates identified. Anomalies/outliers and non-asymptotic relationships for any replicate can easily be spotted using plots such as these.

If relationships for only one or a few replicates are non-asymptotic, then those replicates can be dropped from the analysis, or the maximum measured tissue concentrations used as an estimate of C_{ss} . If relationships for several replicates (i.e., >5 total, or >1 for any individual sediment) are non-asymptotic, then there is little justification for assuming that a steady state has been approached. The test should be repeated, but over a longer time interval. Measuring concentrations in field-collected organisms is also an alternative, if steady state is not reached in laboratory experiments (see Section D3.3).

4. Estimates of C_{ss} that are negative. This can happen if tissue concentrations decrease over time and k_2 is negative. If there are only one or a few replicates with negative C_{ss} values, then these replicates can be dropped from the analysis. Alternatively, the minimum or mean measured concentration could be used as an estimate of C_{ss} . If there are several (i.e., >5 total, or >1 for any individual sediment), then the test should be repeated. High initial contaminant levels in the test organisms are the most probable cause of negative C_{ss} values. Prior to repeating the test, these initial contaminant levels should be measured, and the source of test organisms should be changed if these levels are greater than bioaccumulation of the contaminant at the end of the previous test.

If difficulties are encountered, approaches such as those discussed by Draper and Smith (1981) and SCI (1989) should be considered. Investigators with limited experience should always consult an applied statistician familiar with nonlinear regression prior to analyzing time-sequenced bioaccumulation data. It is important to remember that these data are usually very expensive to obtain, because of the extensive number of chemical analyses required, and the data should be carefully and correctly analyzed.

In the example data analysis, the DUD method failed to converge within the default number of iterations (50) for sediment 3, replicate 5. However, the procedure was close enough to convergence that the regression coefficients output at the final iteration produced a reasonable estimate of C_{ss} .

The approach recommended in this Appendix for comparison of Tier IV dredged sediment and reference sediment bioaccumulation data differs from that described in the Ocean Disposal Manual (the "Green Book"). The approach of comparing 95% confidence intervals for C_{ss} is not recommended because:

- The 95% confidence intervals apply to the estimate of C_{ss} rather than to the *difference* between estimates
- The 95% confidence intervals are based on regressions through points from all replicates for a treatment, ignoring variation among replicates within a treatment
- Different programs or methods will provide different confidence intervals for the same data
- Measurements of tissue levels taken at different times may not be independent.

If the objective of the Tier IV investigation is only to compare bioaccumulation from reference and dredged sediments over the duration of the experiment, and estimates of C_{ss} are not required, there are other alternatives to analyze the data:

- Repeated measures analysis of variance, testing for linear and quadratic components of the time trend
-

- Multivariate analysis of variance (MANOVA), with tissue levels for each day considered separate variables (linear and quadratic trends can also be tested in MANOVA).

These alternatives are equivalent with respect to testing for linear and quadratic trends over time, and some repeated measures programs (e.g., SYSTAT MGLH) will provide MANOVA results as well. These alternatives should only be used by experienced investigators who are familiar with them. Both alternatives would be most useful in testing for an overall quadratic trend, as the absence of such a trend over time would indicate that tissue levels did not approach an asymptote within the duration of the experiment.

D3.2.2 Comparison with Reference Sediments and Action Levels

The difficult part of analyzing time-sequenced bioaccumulation tests is obtaining sound estimates of C_{ss} . Once these estimates are obtained, they are analyzed using the same procedures as for single time-point bioaccumulation tests (Section D3.1). Steady-state concentration estimates for the dredged sediments are compared to steady-state concentration estimates for the reference sediment using the appropriate methods from the decision trees in Figures D-5A or 5B.

The values of C_{ss} in Table D-14 were analyzed using the decision tree steps in Figure D-5A. Although SAS Program BIOACCSS (Section D4.4) conducts all parametric and rankit analyses from the decision trees, only the appropriate results are reported in Table D-15. The untransformed C_{ss} values were normally distributed (Shapiro-Wilk's Test, $P > 0.01$, the α level from Table D-2 for $N=20$, balanced data). However, neither the untransformed nor log-transformed C_{ss} passed Levene's Test for equality of variances ($P < 0.10$, the α level in Table D-2 for $n=5$, balanced data). Therefore, t -tests were conducted, comparing each dredged sediment C_{ss} with reference sediment C_{ss} , using the untransformed C_{ss} estimates. Note that t -tests for equal variances could be used because the F' tests for each dredged sediment-reference comparison did not reject equal variances, even though the overall test of equality of variances indicated unequal variances within the data set as a whole. Mean estimated concentrations at steady state for dredged sediments 2 and 3 (but not sediment 1) were significantly greater than that of the reference sediment (Table D-15).

Table D-15. Tests of Assumptions and Parametric Hypothesis Tests on Untransformed Steady-State Bioaccumulation Example Data.

Null Hypothesis: Mean Dredged Sediment Steady-State Bioaccumulation Equals Mean Reference Sediment Steady-State Bioaccumulation				
Test	Test Statistic	Probability <i>P</i>	α	Conclusion
Normality Assumption: Shapiro-Wilk's Test				
Untransformed data	W=0.963	0.613	0.01	do not reject
Log-transformed data	W=0.943	0.280	0.01	do not reject
Equality of Variances Assumption: Levene's Test	F=4.74	0.015	0.10	reject
Untransformed data	F=3.68	0.034	0.10	reject
Log-transformed data				
Null Hypotheses: Sediment 1 = Reference				
<i>t</i> -Test (equal variances)	<i>t</i> =1.49	0.0873	0.05	do not reject
Sediment 2 = Reference				
<i>t</i> -Test (equal variances)	<i>t</i> =6.03	0.0002	0.05	reject
Sediment 3 = Reference				
<i>t</i> -Test (equal variances)	<i>t</i> =9.21	<0.0001	0.05	reject

Power calculations for an LSD test using untransformed data are performed in SAS program BIOACCSS (Section D4.4). From Eq. 18, a 50% increase over the mean reference C_{ss} (0.602 $\mu\text{g/g}$) can be detected with a probability of 0.32, and a 100% increase with a probability of 0.78. Likewise, there is a 0.95 probability of detecting a 138% increase in C_{ss} over the mean reference C_{ss} . The least significant difference from the LSD is 0.415 $\mu\text{g/g}$, which is a 69% increase over the mean reference C_{ss} . Sample size (n) required to provide a 0.95 probability of detecting a 25% increase over the mean reference C_{ss} is 136 (Eq. 9, using MSE in place of s^2).

The C_{ss} values for the dredged sediment can also be compared to an action level, if available, using the one-sample t -test or one-sided upper confidence limits (UCL) as in Section D3.1.2. UCL for both equal variances and unequal variances may be calculated using SAS program BIOACCSS (Section D4.4). Figure D-7 provides the mean C_{ss} and UCL for each example dredged sediment, along with a hypothetical action level of 2 $\mu\text{g/g}$. The UCL for sediments 1 and 2 were below the action level, indicating that the C_{ss} for these sediments were significantly lower than the action level. The mean C_{ss} for sediment 3 was above the action level, so there was no need to calculate a UCL to conclude that the C_{ss} was not significantly lower than the action level.

Power to detect a true population steady-state concentration 10, 20, 30, 40 and 50% below an action level is calculated in SAS program BIOACCSS (Section D4.4).

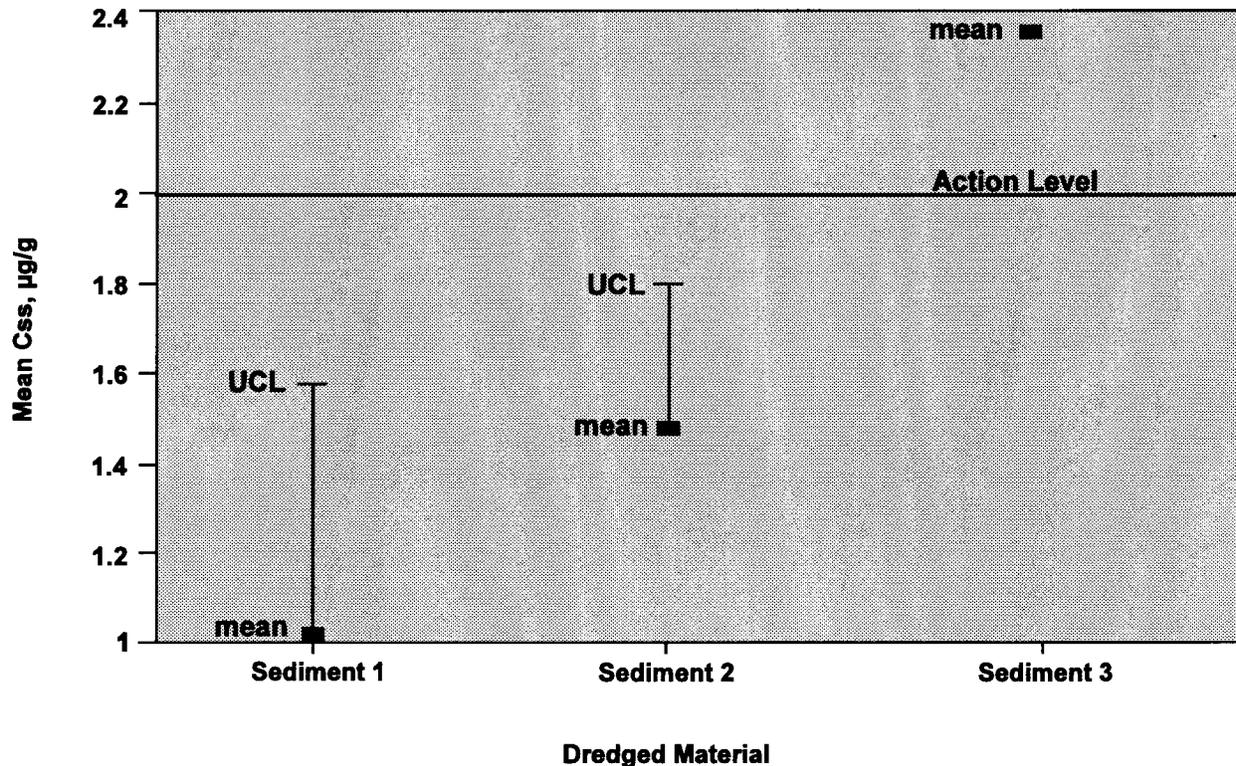


Figure D-7. Comparison of Mean Dredged Sediment Contaminant Steady-State Tissue Levels (C_{ss}) (mean) and 95% Upper Confidence Levels (UCL) with Hypothetical Action Level.

D3.3 Steady-State Bioaccumulation from Field Data

The field bioaccumulation test is designed to show differences, if any, between organisms living at the proposed disposal site and organisms living in the sediments in the reference area. This approach is valid only under the conditions described in Section 12.2.2.

Replicate tissue concentrations in organisms collected at the disposal site(s) are compared with replicate tissue concentrations in organisms collected from the reference area using the decision tree steps in Figures D-5A and 5B. If comparisons involve organisms from only one disposal site, then the appropriate statistical comparison procedures, depending on the results of the tests of assumptions, are the two-sample *t*-test for equal or unequal variances, or the *t*-test for unequal variances using rankits or ranks (Section D2.1.1.1).

D4.0 SAS PROGRAMS AND OUTPUT FOR EXAMPLE DATA

This Section provides SAS programs to analyze the example data sets given in Appendix D. Each program includes all analyses from the corresponding decision tree that would be performed using SAS. While it is certainly possible to conduct the statistical analysis of a data set in a stepwise fashion, we find it much more efficient to perform all analyses at once, and then select the appropriate results based on the steps in the decision tree. Power calculations are provided in addition to the decision tree analyses.

SAS statements in the sections that follow are given in uppercase letters (although this is not required for SAS). Comments within the body of the programs are in upper and lowercase letters in the following format: /* Comment line. */. Every SAS statement must end with a semicolon, but several statements may be included on the same line. The programs are designed for the analysis of Appendix D example data, but can be used with other data sets after minor modifications. Investigators wishing to use these programs should have some familiarity with SAS. SAS output follows each program; the output has been edited to remove much of the nonessential information.

We recommend that data analysis reports include at least the following:

- Number of replicates, mean and SE for each treatment
 - Treatment of less-than detection limit data, if any
 - Results of tests of assumptions
 - Data transformation used, if any
 - Name of statistical hypothesis testing procedure, its calculated test statistic and associated probability, and conclusion reached regarding the null hypothesis
-

- Minimum significant difference or some other indication of power for a parametric LSD test or *t*-test.

SAS programs and output are also provided for censored data methods used when bioaccumulation data include nondetects.

D4.1 Program WATTOX.SAS for Water Column Toxicity Test Data Analysis

WATTOX.SAS is a program to compare dilution water survival vs. 100% elutriate survival, using an arcsine-square root transformation on the data. The program performs all statistical analyses in Figure D-1. Included in these analyses are: mean survival for dilution water and elutriates, Shapiro-Wilk's Test for normality, *t*-test for equal or unequal variances, and a *t*-test for unequal variances on data converted to rankits. Refer to the decision tree in Figure D-1 to determine which test results should be used. Minimum significant difference and some other power calculations for the parametric *t*-test are also provided.

D4.1.1 WATTOX.SAS Program Statements

```
LIBNAME Q 'C:\SAS';
OPTIONS LINESIZE=79 PAGESIZE=59 NODATE NONUMBER;

/* Identify the treatment codes. */

PROC FORMAT;
  VALUE TRTFMT
    0='DILUTION WATER '
    1='100% ELUTRIATE '
    2='50% ELUTRIATE '
    3='25% ELUTRIATE '
    4='12.5% ELUTRIATE';

/* Input the toxicity test data after the CARDS statement, listing the      */
/* treatment code, replicate, and number of survivors. A permanent SAS      */
/* data set is created in the directory specified in the LIBNAME statement. */

DATA Q.WATCOL;
  INPUT TRT REP SURV @@;
  CARDS;
0 1 20 0 2 19 0 3 20 0 4 20 0 5 19
1 1 6 1 2 7 1 3 9 1 4 5 1 5 8
2 1 8 2 2 8 2 3 9 2 4 10 2 5 11
3 1 12 3 2 18 3 3 15 3 4 14 3 5 13
4 1 17 4 2 17 4 3 18 4 4 16 4 5 18
;
/* Input no. of organisms (M) per test container at start of test. */
/* Calculate proportion of survivors (SURV/M) and take the SQRT. */
/* Arcsine transform SQRT(SURV/M). */
/* Format, print, sort the data. Print no. of observations, mean, and */
/* standard error for survival in each treatment. */

DATA A0;
  SET Q.WATCOL;
  M=20;
  ARCSURV=AR SIN(SQRT(SURV/M));
  LABEL TRT='TREATMENT GROUP'
```

```
REP='REPLICATE'
M='NO. OF ORGANISMS PER REPLICATE'
SURV='NUMBER OF SURVIVORS'
ARCSURV='ARCSINE TRANSFORMATION';
FORMAT TRT TRTFMT.;
TITLE 'WATER COLUMN TOXICITY DATA';
PROC PRINT LABEL; VAR TRT REP M SURV ARCSURV;
PROC SORT; BY TRT;
PROC MEANS NOPRINT; BY TRT; VAR SURV;
OUTPUT OUT=Y N=N SUM=TOTAL MEAN=MEANSURV STDERR=SE;
PROC PRINT LABEL; VAR TRT N MEANSURV SE;
LABEL MEANSURV='MEAN SURVIVAL';

/* Delete data not needed for the dilution water-100% elutriate comparison. */
/* Print descriptive statistics. */

DATA A;
SET A0;
IF TRT>1 THEN DELETE;
TITLE2 'ARCSINE-SQUARE ROOT TRANSFORMATION';
PROC MEANS NOPRINT; VAR ARCSURV; BY TRT; ID M;
OUTPUT OUT=X N=N MEAN=MEAN VAR=VARIANCE STD=S STDERR=SE;
PROC PRINT LABEL; VAR TRT N MEAN VARIANCE S SE;

/* Test normality of residuals using Shapiro-Wilk's Test. */

PROC GLM DATA=A NOPRINT;
CLASS TRT;
MODEL ARCSURV=TRT;
OUTPUT OUT=Z R=RESID;
PROC UNIVARIATE NORMAL DATA=Z;
VAR RESID;
TITLE3 'SHAPIRO-WILKS TEST';

/* Conduct t-test, which includes F' test for equality of variances. */

PROC TTEST DATA=A;
CLASS TRT;
VAR ARCSURV;
TITLE3 'T-TEST';

/* Convert data to rankits and conduct t-test. */

PROC RANK DATA=A NORMAL=BLOM OUT=A1;
VAR SURV; RANKS RANKIT;
PROC TTEST DATA=A1;
CLASS TRT;
VAR RANKIT;
TITLE2 'DATA CONVERTED TO RANKITS';

/* Calculate minimum significant difference and power of a t-test to detect */
/* true population differences of 10, 20, 30, 40 and 50% below mean */
/* dilution water survival. */

DATA B0;
MERGE X Y;
IF TRT^=0 THEN DELETE;
MEAN0=MEAN; N0=N; S20=VARIANCE;
MEANPCT=MEANSURV/M;
DATA B1;
SET X;
IF TRT^=1 THEN DELETE;
N1=N; S21=VARIANCE;
```

```

DATA B2;
MERGE B0 B1;
DF=N0+N1-2;
N=(N0+N1)/2;
S2POOL=(S20*(N0-1)+S21*(N1-1))/DF;
TALPHA=TINV(.95,DF);
DMIN=TALPHA*SQRT(2*S2POOL/N);
LABEL N='NO. OF REPLICATES'
      MEANPCT='MEAN DILUTION WATER SURVIVAL'
      S2POOL='POOLED VARIANCE'
      DF='DEGREES OF FREEDOM, DF'
      TALPHA='T VALUE FOR (1-ALPHA=0.95,DF)'
      DMIN='MINIMUM SIGNIFICANT DIFFERENCE';
TITLE2 'POWER OF T-TEST TO DETECT A TRUE POPULATION DIFFERENCE (D)';
TITLE3 'FROM MEAN DILUTION WATER SURVIVAL USING ARCSINE TRANSFORMATION';
PROC PRINT LABEL NOOBS; VAR M N MEANPCT S2POOL DF TALPHA DMIN;
DATA B3;
SET B2;
DO PCTDIFF=10 TO 50 BY 10;
  SEDSURV=MEANPCT-PCTDIFF/100;
  ARCSURV=ARCSIN(SQRT(SED SURV));
  ARCDIFF=MEAN0-ARCSURV;
  TBETA=(SQRT(N)*ARCDIFF)/SQRT(2*S2POOL)-TALPHA;
  POWER=PROBT(TBETA,DF);
  OUTPUT;
END;
LABEL PCTDIFF='% REDUCTION IN SURVIVAL FROM DIL. WATER'
      SEDSURV='100% ELUTRIATE SURVIVAL'
      ARCSURV='ARCSINE 100% ELUTRIATE SURVIVAL'
      ARCDIFF='D'
      TBETA='T VALUE FOR (1-BETA,DF)';
PROC PRINT LABEL;
VAR PCTDIFF SEDSURV ARCSURV ARCDIFF TBETA POWER;
TITLE;

```

D4.1.2 WATTOX.SAS Program Output

WATER COLUMN TOXICITY DATA

OBS	TREATMENT GROUP	REPLICATE	NO. OF ORGANISMS PER REPLICATE	NUMBER OF SURVIVORS	ARCSINE TRANSFORMATION
1	DILUTION WATER	1	20	20	1.57080
2	DILUTION WATER	2	20	19	1.34528
3	DILUTION WATER	3	20	20	1.57080
4	DILUTION WATER	4	20	20	1.57080
5	DILUTION WATER	5	20	19	1.34528
6	100% ELUTRIATE	1	20	6	0.57964
7	100% ELUTRIATE	2	20	7	0.63305
8	100% ELUTRIATE	3	20	9	0.73531
9	100% ELUTRIATE	4	20	5	0.52360
10	100% ELUTRIATE	5	20	8	0.68472
11	50% ELUTRIATE	1	20	8	0.68472
12	50% ELUTRIATE	2	20	8	0.68472
13	50% ELUTRIATE	3	20	9	0.73531
14	50% ELUTRIATE	4	20	10	0.78540
15	50% ELUTRIATE	5	20	11	0.83548
16	25% ELUTRIATE	1	20	12	0.88608
17	25% ELUTRIATE	2	20	18	1.24905
18	25% ELUTRIATE	3	20	15	1.04720

19	25% ELUTRIATE	4	20	14	0.99116
20	25% ELUTRIATE	5	20	13	0.93774
21	12.5% ELUTRIATE	1	20	17	1.17310
22	12.5% ELUTRIATE	2	20	17	1.17310
23	12.5% ELUTRIATE	3	20	18	1.24905
24	12.5% ELUTRIATE	4	20	16	1.10715
25	12.5% ELUTRIATE	5	20	18	1.24905

OBS	TREATMENT GROUP	N	MEAN SURVIVAL	SE
1	DILUTION WATER	5	19.6	0.24495
2	100% ELUTRIATE	5	7.0	0.70711
3	50% ELUTRIATE	5	9.2	0.58310
4	25% ELUTRIATE	5	14.4	1.02956
5	12.5% ELUTRIATE	5	17.2	0.37417

WATER COLUMN TOXICITY DATA
ARCSINE-SQUARE ROOT TRANSFORMATION

OBS	TREATMENT GROUP	N	MEAN	VARIANCE	S	SE
1	DILUTION WATER	5	1.48059	0.015257	0.12352	0.055239
2	100% ELUTRIATE	5	0.63126	0.006986	0.08358	0.037379

WATER COLUMN TOXICITY DATA
ARCSINE-SQUARE ROOT TRANSFORMATION
SHAPIRO-WILKS TEST

UNIVARIATE PROCEDURE

Variable=RESID

N	10
W:Normal	0.846238 Prob<W 0.0507

WATER COLUMN TOXICITY DATA
ARCSINE-SQUARE ROOT TRANSFORMATION
T-TEST

TTEST PROCEDURE

Variable: ARCSURV ARCSINE TRANSFORMATION

TRT	N	Mean	Std Dev	Std Error
DILUTION WATER	5	1.48059096	0.12351878	0.05523928
100% ELUTRIATE	5	0.63126480	0.08358232	0.03737915

Variances	T	DF	Prob> T
Unequal	12.7340	7.0	0.0001
Equal	12.7340	8.0	0.0000

For H0: Variances are equal, $F' = 2.18$ DF = (4,4) Prob>F' = 0.4679

WATER COLUMN TOXICITY DATA
DATA CONVERTED TO RANKITS

TTEST PROCEDURE

Variable: RANKIT RANK FOR VARIABLE SURV

TRT	N	Mean	Std Dev	Std Error
DILUTION WATER	5	0.74011839	0.44830825	0.20048954
100% ELUTRIATE	5	-0.74011839	0.55672332	0.24897424

Variances	T	DF	Prob> T
Unequal	4.6306	7.7	0.0019
Equal	4.6306	8.0	0.0017

For HO: Variances are equal, $F' = 1.54$ $DF = (4,4)$ $Prob>F' = 0.6850$

WATER COLUMN TOXICITY DATA
POWER OF T-TEST TO DETECT A TRUE POPULATION DIFFERENCE (D)
FROM MEAN DILUTION WATER SURVIVAL USING ARCSINE TRANSFORMATION

NO. OF ORGANISMS PER REPLICATE	N	MEAN DILUTION WATER SURVIVAL	POOLED VARIANCE	DEGREES OF FREEDOM, DF	T VALUE FOR (1-ALPHA=0.95, DF)	MINIMUM SIGNIFICANT DIFFERENCE
20	5	0.98	0.011121	8	1.85955	0.12403

OBS	% REDUCTION IN SURVIVAL FROM DIL. WATER	100% ELUTRIATE SURVIVAL	ARCSINE 100% ELUTRIATE SURVIVAL	D	T VALUE FOR (1-BETA, DF)	POWER
1	10	0.88	1.21705	0.26354	2.09166	0.96508
2	20	0.78	1.08259	0.39800	4.10768	0.99830
3	30	0.68	0.96953	0.51106	5.80277	0.99980
4	40	0.58	0.86574	0.61485	7.35888	0.99996
5	50	0.48	0.76539	0.71520	8.86344	0.99999

D4.2 Program BENTOX.SAS for Benthic Toxicity Test Data Analysis

BENTOX.SAS is a program to compare benthic toxicity data from dredged sediments vs. reference sediment, using an arcsine-square root transformation on the data. Included in these analyses are: mean survival from each sediment exposure, Shapiro-Wilk's Test for normality, Levene's test for equality of variances, *t*-tests for equal or unequal variances, LSD test, and tests on rankits (normalized ranks for survival). Refer to the decision tree in Figures D-4A and 4B to determine which test results should be used. The program includes power calculations (on an arcsine-transformed scale) for an LSD test.

D4.2.1 BENTOX.SAS Program Statements

```

LIBNAME Q 'C:\SAS';
OPTIONS LINESIZE=79 PAGESIZE=59 NODATE NONUMBER;

/* Identify the treatment codes. */

PROC FORMAT;
VALUE TRTFMT
  1='REFERENCE '
  2='SEDIMENT 1'
  3='SEDIMENT 2'
  4='SEDIMENT 3';

/* Input the toxicity test data after the CARDS statement, listing the */
/* treatment code, replicate, and number of survivors. A permanent SAS */
/* data set is created in the directory specified in the LIBNAME statement. */

DATA Q.BENTHIC;
  INPUT TRT REP SURV @@;
  CARDS;
1 1 20 1 2 20 1 3 19 1 4 19 1 5 20
2 1 17 2 2 16 2 3 18 2 4 17 2 5 15
3 1 15 3 2 16 3 3 13 3 4 17 3 5 11
4 1 17 4 2 12 4 3 10 4 4 16 4 5 13
;
/* Input no. of organisms (M) per test container at start of test. */
/* Calculate proportion of survivors (SURV/M) and take the SQRT.*/
/* Arcsine transform SQRT(SURV/M). */
/* Format, print, sort the data. Print no. of observations, mean, and */
/* standard error for survival in each treatment. */

DATA A0;
SET Q.BENTHIC;
M=20;
ARCSURV=ARSIN(SQRT(SURV/M));
LABEL TRT='TREATMENT GROUP'
      REP='REPLICATE'
      M='NO. OF ORGANISMS PER REPLICATE'
      SURV='NUMBER OF SURVIVORS'
      ARCSURV='ARCSINE TRANSFORMATION';
FORMAT TRT TRTFMT.;
TITLE 'BENTHIC TOXICITY DATA';
PROC RANK NORMAL=BLOM OUT=A;
VAR SURV; RANKS RANKIT;
PROC PRINT LABEL; VAR TRT REP M SURV ARCSURV RANKIT;
LABEL RANKIT='NORMALIZED RANK FOR SURVIVAL';
PROC SORT; BY TRT;
PROC MEANS NOPRINT; BY TRT; VAR SURV; ID M;
OUTPUT OUT=Y N=N SUM=TOTAL MEAN=MEANSURV STDERR=SE;
PROC PRINT LABEL; VAR TRT N TOTAL MEANSURV SE;
LABEL MEANSURV='MEAN SURVIVAL';

/* Print descriptive statistics for the arcsine-transformed survival data. */

PROC MEANS NOPRINT DATA=A; VAR ARCSURV; BY TRT;
OUTPUT OUT=X N=N MEAN=MEAN VAR=VARIANCE STD=S STDERR=SE;
TITLE2 'ARCSINE-SQUARE ROOT TRANSFORMATION';
PROC PRINT LABEL; VAR TRT N MEAN VARIANCE S SE;

/* Test normality of residuals using Shapiro-Wilk's Test. */

```

```
PROC GLM DATA=A NOPRINT;
  CLASS TRT;
  MODEL ARCSURV=TRT;
  OUTPUT OUT=Z R=RESID;
PROC UNIVARIATE NORMAL DATA=Z;
  VAR RESID;
  TITLE3 'SHAPIRO-WILKS TEST FOR NORMALITY';

/* Conduct Levene's Test for equality of variances. */

DATA AX;
  MERGE A X; BY TRT;
  ABSDEV=ABS(ARCSURV-MEAN);
  LABEL ABSDEV='ABSOLUTE DEVIATIONS FROM MEAN';
PROC GLM;
  CLASS TRT;
  MODEL ABSDEV=TRT;
  TITLE3 'LEVENE'S TEST FOR EQUALITY OF VARIANCES';

/* Perform LSD Test. */

PROC GLM DATA=A OUTSTAT=W;
  CLASS TRT;
  MODEL ARCSURV=TRT;
  MEANS TRT/LSD ALPHA=.1;
  TITLE3 'LSD TEST';

/* Perform t-tests for each dredged sediment-reference sediment comparison. */

DATA T1;
  SET A;
  IF TRT>2 THEN DELETE;
PROC TTEST;
  CLASS TRT;
  VAR ARCSURV;
  TITLE3 'T-TEST';
DATA T2;
  SET A;
  IF TRT=2 OR TRT=4 THEN DELETE;
PROC TTEST;
  CLASS TRT;
  VAR ARCSURV;
DATA T3;
  SET A;
  IF TRT=2 OR TRT=3 THEN DELETE;
PROC TTEST;
  CLASS TRT;
  VAR ARCSURV;

/* Test normality and equality of variances of rankits. */

PROC GLM NOPRINT DATA=A;
  CLASS TRT;
  MODEL RANKIT=TRT;
  OUTPUT OUT=Z1 R=RESID;
  TITLE2 'SURVIVAL DATA CONVERTED TO RANKITS';
PROC UNIVARIATE NORMAL;
  VAR RESID;
  TITLE3 'SHAPIRO-WILKS TEST FOR NORMALITY';
PROC MEANS DATA=A NOPRINT;
  BY TRT; VAR RANKIT;
  OUTPUT OUT=X1 MEAN=MEAN;
DATA AX1;
```

```
MERGE A X1; BY TRT;
ABSDEV=ABS(RANKIT-MEAN);
LABEL ABSDEV='ABSOLUTE DEVIATIONS FROM MEAN';
PROC GLM;
  CLASS TRT;
  MODEL ABSDEV=TRT;
  TITLE3 'LEVENE'S TEST';

/* Perform LSD test on rankits. */

PROC GLM DATA=A;
  CLASS TRT;
  MODEL RANKIT=TRT;
  MEANS TRT/LSD ALPHA=.1;
  TITLE3 'LSD TEST ON RANKITS';

/* Perform t-tests comparing each dredged sediment with the reference */
/* using rankits. */

PROC TTEST DATA=T1;
  CLASS TRT;
  VAR RANKIT;
  TITLE3 'T-TEST ON RANKITS';
PROC TTEST DATA=T2;
  CLASS TRT;
  VAR RANKIT;
PROC TTEST DATA=T3;
  CLASS TRT;
  VAR RANKIT;

/* Calculate power of an LSD test to detect true population differences */
/* of 10, 20, 30, 40 and 50% below mean (arcsine-transformed) reference */
/* sediment survival. */

DATA C1;
  SET W;
  IF _TYPE_ ^= 'ERROR' THEN DELETE;
  MSE=SS/DF;
  KEEP MSE DF;
DATA C2;
  MERGE Y X;
  IF TRT ^= 1 THEN DELETE;
  MEANPCT=MEANSURV/M;
DATA C3;
  MERGE C1 C2;
  TALPHA=TINV(.95,DF);
  LABEL M='NO. OF ORGANISMS AT START OF TEST'
        N='NO. OF REPLICATES'
        MEANPCT='MEAN REFERENCE SURVIVAL'
        MSE='MEAN SQUARE ERROR'
        DF='DEGREES OF FREEDOM, DF'
        TALPHA='T VALUE FOR (1-ALPHA=0.95,DF)';
  TITLE2 'POWER OF LSD TO DETECT A TRUE POPULATION DIFFERENCE (D)';
  TITLE3 'FROM MEAN REFERENCE SURVIVAL USING ARCSINE TRANSFORMATION';
PROC PRINT LABEL NOOBS; VAR M N MEANPCT MSE DF TALPHA;
DATA C;
  SET C3;
  DO PCTDIFF=10 TO 50 BY 10;
    SEDSURV=MEANPCT-PCTDIFF/100;
    ARCSURV=ARSIN(SQRT(SEDSURV));
    ARCDIFF=MEAN-ARCSURV;
    TBETA=ARCDIFF*SQRT(N/(2*MSE))-TALPHA;
```

```

POWER=PROBT(TBETA,DF);
OUTPUT;
END;
LABEL PCTDIFF='% REDUCTION IN SURVIVAL FROM REFERENCE'

      SEDSURV='DREDGED SEDIMENT SURVIVAL'
      ARCSURV='ARCSINE DREDGED SEDIMENT SURVIVAL'
      ARCDIFF='D'
      TBETA='T VALUE FOR (1-BETA,DF)';
PROC PRINT LABEL;
VAR PCTDIFF SEDSURV ARCSURV ARCDIFF TBETA POWER;
TITLE;

```

D4.2.2 BENTOX.SAS Program Output

BENTHIC TOXICITY DATA

OBS	TREATMENT GROUP	REPLICATE	NO. OF ORGANISMS PER REPLICATE	NUMBER OF SURVIVORS	ARCSINE TRANSFORMATION	NORMALIZED RANK FOR SURVIVAL
1	REFERENCE	1	20	20	1.57080	1.46660
2	REFERENCE	2	20	20	1.57080	1.46660
3	REFERENCE	3	20	19	1.34528	0.83164
4	REFERENCE	4	20	19	1.34528	0.83164
5	REFERENCE	5	20	20	1.57080	1.46660
6	SEDIMENT 1	1	20	17	1.17310	0.25276
7	SEDIMENT 1	2	20	16	1.10715	-0.18775
8	SEDIMENT 1	3	20	18	1.24905	0.58946
9	SEDIMENT 1	4	20	17	1.17310	0.25276
10	SEDIMENT 1	5	20	15	1.04720	-0.51861
11	SEDIMENT 2	1	20	15	1.04720	-0.51861
12	SEDIMENT 2	2	20	16	1.10715	-0.18775
13	SEDIMENT 2	3	20	13	0.93774	-0.83164
14	SEDIMENT 2	4	20	17	1.17310	0.25276
15	SEDIMENT 2	5	20	11	0.83548	-1.40341
16	SEDIMENT 3	1	20	17	1.17310	0.25276
17	SEDIMENT 3	2	20	12	0.88608	-1.12814
18	SEDIMENT 3	3	20	10	0.78540	-1.86824
19	SEDIMENT 3	4	20	16	1.10715	-0.18775
20	SEDIMENT 3	5	20	13	0.93774	-0.83164

BENTHIC TOXICITY DATA

OBS	TREATMENT GROUP	N	TOTAL	MEAN SURVIVAL	SE
1	REFERENCE	5	98	19.6	0.24495
2	SEDIMENT 1	5	83	16.6	0.50990
3	SEDIMENT 2	5	72	14.4	1.07703
4	SEDIMENT 3	5	68	13.6	1.28841

BENTHIC TOXICITY DATA
ARCSINE-SQUARE ROOT TRANSFORMATION

OBS	TREATMENT GROUP	N	MEAN	VARIANCE	S	SE
1	REFERENCE	5	1.48059	0.015257	0.12352	0.055239
2	SEDIMENT 1	5	1.14992	0.005820	0.07629	0.034119
3	SEDIMENT 2	5	1.02013	0.018147	0.13471	0.060244
4	SEDIMENT 3	5	0.97789	0.025477	0.15962	0.071382

BENTHIC TOXICITY DATA
ARCSINE-SQUARE ROOT TRANSFORMATION
SHAPIRO-WILKS TEST FOR NORMALITY

UNIVARIATE PROCEDURE

Variable=RESID

N	20
W:Normal	0.945932 Prob<W 0.3217

BENTHIC TOXICITY DATA
ARCSINE-SQUARE ROOT TRANSFORMATION
LEVENE'S TEST FOR EQUALITY OF VARIANCES

General Linear Models Procedure

Source	DF	Sum of Squares	Absolute Deviations from Mean Square	F Value	Pr > F
Model	3	0.01373434	0.00457811	1.74	0.1985
Error	16	0.04201517	0.00262595		
Corrected Total	19	0.05574951			

BENTHIC TOXICITY DATA
ARCSINE-SQUARE ROOT TRANSFORMATION
LSD TEST

General Linear Models Procedure

T tests (LSD) for variable: ARCSURV

NOTE: This test controls the type I comparisonwise error rate not the experimentwise error rate.

Alpha= 0.1 df= 16 MSE= 0.016175
Critical Value of T= 1.75
Least Significant Difference= 0.1404

Means with the same letter are not significantly different.

T Grouping	Mean	N	TRT
A	1.4806	5	REFERENCE
B	1.1499	5	SEDIMENT 1
B			
C	1.0201	5	SEDIMENT 2
C			
C	0.9779	5	SEDIMENT 3

BENTHIC TOXICITY DATA
ARCSINE-SQUARE ROOT TRANSFORMATION
T-TEST

TTEST PROCEDURE

Variable: ARCSURV ARCSINE TRANSFORMATION

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	1.48059096	0.12351878	0.05523928
SEDIMENT 1	5	1.14991717	0.07629145	0.03411857

Variances	T	DF	Prob> T
Unequal	5.0930	6.7	0.0017
Equal	5.0930	8.0	0.0009

For H0: Variances are equal, $F' = 2.62$ DF = (4,4) Prob>F' = 0.3733

BENTHIC TOXICITY DATA
ARCSINE-SQUARE ROOT TRANSFORMATION
T-TEST

TTEST PROCEDURE

Variable: ARCSURV ARCSINE TRANSFORMATION

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	1.48059096	0.12351878	0.05523928
SEDIMENT 2	5	1.02013391	0.13470903	0.06024371

Variances	T	DF	Prob> T
Unequal	5.6335	7.9	0.0005
Equal	5.6335	8.0	0.0005

For H0: Variances are equal, $F' = 1.19$ DF = (4,4) Prob>F' = 0.8706

BENTHIC TOXICITY DATA
ARCSINE-SQUARE ROOT TRANSFORMATION
T-TEST

TTEST PROCEDURE

Variable: ARCSURV ARCSINE TRANSFORMATION

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	1.48059096	0.12351878	0.05523928
SEDIMENT 3	5	0.97789308	0.15961511	0.07138205

Variances	T	DF	Prob> T
Unequal	5.5695	7.5	0.0007
Equal	5.5695	8.0	0.0005

For H0: Variances are equal, $F' = 1.67$ DF = (4,4) Prob>F' = 0.6315

BENTHIC TOXICITY DATA
 SURVIVAL DATA CONVERTED TO RANKITS
 SHAPIRO-WILKS TEST FOR NORMALITY

UNIVARIATE PROCEDURE

Variable=RESID

N 20
 W:Normal 0.981773 Prob<W 0.9399

BENTHIC TOXICITY DATA
 SURVIVAL DATA CONVERTED TO RANKITS
 LEVENE'S TEST

General Linear Models Procedure

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.31609842	0.10536614	1.18	0.3493
Error	16	1.43149144	0.08946821		
Corrected Total	19	1.74758986			

BENTHIC TOXICITY DATA
 SURVIVAL DATA CONVERTED TO RANKITS
 LSD TEST ON RANKITS

General Linear Models Procedure

T tests (LSD) for variable: RANKIT

NOTE: This test controls the type I comparisonwise error rate not the experimentwise error rate.

Alpha= 0.1 df= 16 MSE= 0.346143
 Critical Value of T= 1.75
 Least Significant Difference= 0.6496

Means with the same letter are not significantly different.

T Grouping	Mean	N	TRT
A	1.213	5	REFERENCE
B	0.078	5	SEDIMENT 1
B			
C	-0.538	5	SEDIMENT 2
C			
C	-0.753	5	SEDIMENT 3

BENTHIC TOXICITY DATA
SURVIVAL DATA CONVERTED TO RANKITS
T-TEST ON RANKITS

TTEST PROCEDURE

Variable: RANKIT RANK FOR VARIABLE SURV

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	1.21261524	0.34778201	0.15553284
SEDIMENT 1	5	0.07772091	0.43279236	0.19355063

Variances	T	DF	Prob> T
Unequal	4.5707	7.6	0.0021
Equal	4.5707	8.0	0.0018

For H0: Variances are equal, $F' = 1.55$ DF = (4,4) Prob>F' = 0.6821

BENTHIC TOXICITY DATA
SURVIVAL DATA CONVERTED TO RANKITS
T-TEST ON RANKITS

TTEST PROCEDURE

Variable: RANKIT RANK FOR VARIABLE SURV

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	1.21261524	0.34778201	0.15553284
SEDIMENT 2	5	-0.53773198	0.62918751	0.28138121

Variances	T	DF	Prob> T
Unequal	5.4442	6.2	0.0014
Equal	5.4442	8.0	0.0006

For H0: Variances are equal, $F' = 3.27$ DF = (4,4) Prob>F' = 0.2773

BENTHIC TOXICITY DATA
SURVIVAL DATA CONVERTED TO RANKITS
T-TEST ON RANKITS

TTEST PROCEDURE

Variable: RANKIT RANK FOR VARIABLE SURV

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	1.21261524	0.34778201	0.15553284
SEDIMENT 3	5	-0.75260418	0.82488344	0.36889909

Variances	T	DF	Prob> T
Unequal	4.9088	5.4	0.0038
Equal	4.9088	8.0	0.0012

For H0: Variances are equal, $F' = 5.63$ DF = (4,4) Prob>F' = 0.1229

BENTHIC TOXICITY DATA
POWER OF LSD TO DETECT A TRUE POPULATION DIFFERENCE (D)
FROM MEAN REFERENCE SURVIVAL USING ARCSINE TRANSFORMATION

NO. OF ORGANISMS AT START OF TEST	NO. OF REPLICATES	MEAN REFERENCE SURVIVAL	MEAN SQUARE ERROR	DEGREES OF FREEDOM, DF	T VALUE FOR (1-ALPHA=0.95, DF)
20	5	0.98	0.016175	16	1.74588

OBS	% REDUCTION IN SURVIVAL FROM REFERENCE	DREDGED SEDIMENT SURVIVAL	ARCSINE DREDGED SEDIMENT SURVIVAL	D	T VALUE FOR (1-BETA, DF)	POWER
1	10	0.88	1.21705	0.26354	1.53043	0.92728
2	20	0.78	1.08259	0.39800	3.20210	0.99722
3	30	0.68	0.96953	0.51106	4.60766	0.99985
4	40	0.58	0.86574	0.61485	5.89797	0.99999
5	50	0.48	0.76539	0.71520	7.14555	1.00000

D4.3 Program BIOACC.SAS for Single-Time Point Bioaccumulation Test Data Analysis

BIOACC.SAS is a program to compare Tier III bioaccumulation data from dredged sediments vs. reference sediment, using raw data and \log_{10} transformation. Included in these analyses are: mean bioaccumulation from each sediment exposure, Shapiro-Wilk's Test for normality, Levene's Test for equality of variances, t -tests for equal or unequal variances, LSD test, and tests on rankits (normalized ranks for contaminant concentration). Refer to the decision tree in Figures D-5A and 5B to determine which test results should be used. The program includes power calculations for an LSD test on untransformed bioaccumulation data.

D4.3.1 BIOACC.SAS Program Statements

```
LIBNAME Q 'C:\SAS';
OPTIONS LINESIZE=79 PAGESIZE=59 NODATE NONUMBER;

/* Identify the treatment codes. */

PROC FORMAT;
VALUE TRTFMT
  1='REFERENCE '
  2='SEDIMENT 1'
  3='SEDIMENT 2'
  4='SEDIMENT 3';

/* Input the bioaccumulation data after the CARDS statement, listing the */
/* treatment code, replicate, and contaminant concentration. A permanent */
/* SAS data set is created in the directory specified in the LIBNAME */
/* statement. */
```

```
DATA Q.BIOACC;
  INPUT TRT REP CONC @@;
  CARDS;
1 1 .06 1 2 .05 1 3 .05 1 4 .08 1 5 .09
2 1 .16 2 2 .19 2 3 .18 2 4 .22 2 5 .31
3 1 .24 3 2 .10 3 3 .13 3 4 .18 3 5 .30
4 1 .13 4 2 .05 4 3 .17 4 4 .08 4 5 .22
;

/* Format, print, sort the data. Print no. of observations, mean, and */
/* standard error for concentration in each treatment for both */
/* untransformed and log10-transformed data. Calculate rankits. */

DATA A0;
  SET Q.BIOACC;
  LOGCONC=LOG10(CONC);
  MERGEVAR=1;
  LABEL TRT='TREATMENT GROUP'
        REP='REPLICATE'
        CONC='CONTAMINANT CONCENTRATION, ug/g'
        LOGCONC='LOG10 CONCENTRATION';
  FORMAT TRT TRTFMT.;
  TITLE 'SINGLE-TIME POINT CONTAMINANT BIOACCUMULATION DATA';
  PROC RANK NORMAL=BLOM OUT=A;
  VAR CONC; RANKS RANKIT;
  PROC PRINT LABEL; VAR TRT REP CONC LOGCONC RANKIT;
  LABEL RANKIT='NORMALIZED RANK FOR CONCENTRATION';
  PROC SORT; BY TRT;
  PROC MEANS NOPRINT; BY TRT; VAR CONC LOGCONC; ID MERGEVAR;
  OUTPUT OUT=Y N=N NLOG MEAN=MEANCONC MEANLOG VAR=S2 S2LOG STDERR=SE SELOG;
  PROC PRINT LABEL; VAR TRT N MEANCONC S2 SE MEANLOG S2LOG SELOG;
  LABEL MEANCONC='MEAN CONTAMINANT CONC.'
        S2='VARIANCE'
        SE='STANDARD ERROR'
        MEANLOG='MEAN LOG10 CONC.'
        S2LOG='VARIANCE OF LOGS'
        SELOG='STANDARD ERROR OF LOGS';

/* Test normality of residuals of untransformed and log-transformed data */
/* using Shapiro-Wilk's Test. */

PROC GLM NOPRINT DATA=A;
  CLASS TRT;
  MODEL CONC LOGCONC=TRT;
  OUTPUT OUT=Z R=RESID RESIDLOG;
  PROC UNIVARIATE NORMAL;
  VAR RESID RESIDLOG;
  TITLE2 'SHAPIRO-WILKS TEST FOR NORMALITY';

/* Conduct Levene's Test for equality of variances of untransformed and */
/* log-transformed data. */

DATA AY;
  MERGE A Y; BY TRT;
  ABSDEV=ABS(CONC-MEANCONC);
  ABSLOG=ABS(LOGCONC-MEANLOG);
  LABEL ABSDEV='ABSOLUTE DEVIATIONS FROM MEAN CONC.'
        ABSLOG='ABSOLUTE DEVIATIONS FROM MEAN LOGCONC.';
  PROC GLM;
  CLASS TRT;
  MODEL ABSDEV ABSLOG=TRT;
  TITLE2 'LEVENE'S TEST';
  /* Perform LSD on untransformed and log-transformed data. */
```

```
PROC GLM DATA=A OUTSTAT=W1;
  CLASS TRT;
  MODEL CONC=TRT;
  MEANS TRT/LSD ALPHA=.1;
  TITLE2 'LSD TEST (UNTRANSFORMED DATA)';
PROC GLM DATA=A OUTSTAT=W2;
  CLASS TRT;
  MODEL LOGCONC=TRT;
  MEANS TRT/LSD ALPHA=.1;
  TITLE2 'LSD TEST (LOG-TRANSFORMED DATA)';

/* Perform t-tests for each dredged sediment-reference sediment comparison */
/* using untransformed and log-transformed data. */

DATA T1;
  SET A;
  IF TRT>2 THEN DELETE;
PROC TTEST;
  CLASS TRT;
  VAR CONC LOGCONC;
  TITLE2 'T-TEST';
DATA T2;
  SET A;
  IF TRT=2 OR TRT=4 THEN DELETE;
PROC TTEST;
  CLASS TRT;
  VAR CONC LOGCONC;
DATA T3;
  SET A;
  IF TRT=2 OR TRT=3 THEN DELETE;
PROC TTEST;
  CLASS TRT;
  VAR CONC LOGCONC;

/* Test normality and equality of variances of rankits. */

PROC GLM NOPRINT DATA=A;
  CLASS TRT;
  MODEL RANKIT=TRT;
  OUTPUT OUT=Z2 R=RESID;
  TITLE2 'BIOACCUMULATION DATA CONVERTED TO RANKITS';
PROC UNIVARIATE NORMAL;
  VAR RESID;
  TITLE3 'SHAPIRO-WILKS TEST FOR NORMALITY';
PROC MEANS DATA=A NOPRINT;
  BY TRT; VAR RANKIT;
  OUTPUT OUT=X MEAN=MEAN;
DATA AX;
  MERGE A X; BY TRT;
  ABSDEV=ABS(RANKIT-MEAN);
PROC GLM;
  CLASS TRT;
  MODEL ABSDEV=TRT;
  TITLE3 'LEVENE'S TEST';

/* Perform LSD on rankits. */

PROC GLM DATA=A;
  CLASS TRT;
  MODEL RANKIT=TRT;
  MEANS TRT/LSD ALPHA=.1;
```

```
TITLE3 'LSD TEST';

/* Perform t-tests for each dredged sediment-reference sediment comparison */
/* using rankits. */

PROC TTEST DATA=T1;
  CLASS TRT;
  VAR RANKIT;
  TITLE3 'T-TEST';
PROC TTEST DATA=T2;
  CLASS TRT;
  VAR RANKIT;
PROC TTEST DATA=T3;
  CLASS TRT;
  VAR RANKIT;

/* Calculate power of an LSD test to detect true population differences */
/* 10, 25, 50, and 100% above the reference mean contaminant concentration. */

DATA C1;
  SET W1;
  IF _TYPE_ ^= 'ERROR' THEN DELETE;
  MSE=SS/DF;
  MERGEVAR=1;
  KEEP MSE DF MERGEVAR;
DATA C2;
  SET Y;
  IF TRT ^= 1 THEN DELETE;
DATA C3;
  MERGE C1 C2;
  TALPHA=TINV(.95,DF);
  LABEL      N='NO. OF REPLICATES, N'
             MEANCONC='REFERENCE MEAN CONTAMINANT CONCENTRATION'
             MSE='MEAN SQUARE ERROR, MSE'
             DF='DEGREES OF FREEDOM, DF'
             TALPHA='T VALUE FOR (1-ALPHA=0.95,DF)';
  TITLE2 'POWER OF LSD TO DETECT A TRUE POPULATION DIFFERENCE (D)';
  TITLE3 'ABOVE REFERENCE MEAN CONTAMINANT CONCENTRATION';
PROC PRINT LABEL NOOBS; VAR N MEANCONC MSE DF TALPHA;
DATA C4;
  SET C3;
  DO PCTDIFF=10,25,50,100,200,300;
    SEDCONC=MEANCONC+((PCTDIFF/100)*MEANCONC);
    D=SEDCONC-MEANCONC;
    TBETA=D*SQRT(N/(2*MSE))-TALPHA;
    POWER=PROBT(TBETA,DF);
    OUTPUT;
  END;
  LABEL      PCTDIFF='% INCREASE IN CONC. ABOVE REFERENCE'
             SEDCONC='DREDGED SEDIMENT BIOACCUMULATION'
             TBETA='T VALUE FOR (1-BETA,DF)'
             POWER='POWER (1-BETA)';
PROC PRINT LABEL NOOBS; VAR PCTDIFF SEDCONC D TBETA POWER;
  TITLE 'POWER OF LSD TO DETECT % INCREASE IN CONCENTRATION ABOVE REFERENCE';
  TITLE2 'MEAN CONTAMINANT CONCENTRATION GIVEN N, MSE AND DF SHOWN ABOVE';
DATA C5;
  SET C3;
  DO POWER=.5,.6,.7,.8,.9,.95,.99;
    TBETA=TINV(POWER,DF);
    D=((TBETA+TALPHA)*SQRT(2*MSE))/SQRT(N);
    SEDCONC=MEANCONC+D;
    PCTDIFF=(D*100)/MEANCONC;
    OUTPUT;
```

```
END;
LABEL SEDCONC='DREDGED SEDIMENT BIOACCUMULATION'
PCTDIFF='% INCREASE IN CONC. ABOVE REFERENCE'
TBETA='T VALUE FOR (1-BETA,DF)'
POWER='POWER (1-BETA)';
PROC PRINT LABEL NOOBS; VAR POWER D SEDCONC PCTDIFF TBETA;
TITLE 'MINIMUM DREDGED SEDIMENT BIOACCUMULATION THAT CAN BE DETECTED BY LSD';
TITLE2 'AS SIGNIFICANT GIVEN SPECIFIED POWER AND N, MSE, AND DF SHOWN ABOVE';

/* Calculation of upper confidence limits (UCL) for comparison of mean */
/* dredged sediment bioaccumulation with an action level. */

DATA D;
MERGE C1 Y; BY MERGEVAR;
IF TRT=1 THEN DELETE;
TALPHA1=TINV(.95,DF);
TALPHA2=TINV(.95,N-1);
UCL1=MEANCONC+TALPHA1*(SQRT(MSE/N));
UCL2=MEANCONC+TALPHA2*(SQRT(S2/N));
DMIN1=TALPHA1*SQRT(MSE/N);
DMIN2=TALPHA2*SQRT(S2/N);
LABEL UCL1='UCL (EQUAL VARIANCES)'
UCL2='UCL (UNEQUAL VARIANCES)'
TALPHA1='T VALUE FOR (1-ALPHA=.95,DF)'
TALPHA2='T VALUE FOR (1-ALPHA=.95,N-1)'
DMIN1='MINIMUM SIGNIFICANT DIFFERENCE'
DMIN2='MINIMUM SIGNIFICANT DIFFERENCE'
MSE='MEAN SQUARE ERROR'
S2='VARIANCE'
MEANCONC='MEAN BIOACCUMULATION';
TITLE 'COMPARISON OF MEAN DREDGED SEDIMENT BIOACCUMULATION WITH ACTION
LEVEL: ';
PROC PRINT LABEL NOOBS; VAR TRT MEANCONC UCL1 MSE TALPHA1 DF DMIN1;
TITLE2 'UPPER CONFIDENCE LIMITS (UCL) WHEN VARIANCES ARE EQUAL';
PROC PRINT LABEL NOOBS; VAR TRT MEANCONC UCL2 S2 TALPHA2 N DMIN2;
TITLE2 'UPPER CONFIDENCE LIMITS (UCL) WHEN VARIANCES ARE UNEQUAL';

/* Calculate power of dredged sediment-action level comparisons using */
/* MSE given 10, 20, 30, 40, and 50% decreases in mean concentration */
/* below action level. */

DATA D1;
SET C3;
ACTION=.2;
DO PCTDIFF=10 TO 50 BY 10;
D=PCTDIFF*ACTION/100;
SEDCONC=ACTION-D;
TBETA=D*SQRT(N/MSE)-TALPHA;
POWER=PROBT(TBETA,DF);
OUTPUT;
END;
LABEL PCTDIFF='% DECREASE BELOW ACTION LEVEL'
SEDCONC='MEAN DREDGED SEDIMENT BIOACCUMULATION'
TBETA='T VALUE FOR (1-BETA,DF)'
POWER='POWER (1-BETA)';
PROC PRINT NOOBS LABEL; VAR PCTDIFF SEDCONC D TBETA POWER;
TITLE 'POWER TO DETECT % DECREASE IN CONCENTRATION BELOW';
TITLE2 'ACTION LEVEL OF 0.2 ug/g GIVEN N, MSE AND DF SHOWN ABOVE';
```

D4.3.2 BIOACC.SAS Program Output

SINGLE-TIME POINT CONTAMINANT BIOACCUMULATION DATA

OBS	TREATMENT GROUP	REPLICATE	CONTAMINANT CONCENTRATION, ug/g	LOG10 CONCENTRATION	NORMALIZED RANK FOR CONCENTRATION
1	REFERENCE	1	0.06	-1.22185	-0.91914
2	REFERENCE	2	0.05	-1.30103	-1.46660
3	REFERENCE	3	0.05	-1.30103	-1.46660
4	REFERENCE	4	0.08	-1.09691	-0.66680
5	REFERENCE	5	0.09	-1.04576	-0.44777
6	SEDIMENT 1	1	0.16	-0.79588	0.06193
7	SEDIMENT 1	2	0.19	-0.72125	0.58946
8	SEDIMENT 1	3	0.18	-0.74473	0.38117
9	SEDIMENT 1	4	0.22	-0.65758	0.83164
10	SEDIMENT 1	5	0.31	-0.50864	1.86824
11	SEDIMENT 2	1	0.24	-0.61979	1.12814
12	SEDIMENT 2	2	0.10	-1.00000	-0.31457
13	SEDIMENT 2	3	0.13	-0.88606	-0.12434
14	SEDIMENT 2	4	0.18	-0.74473	0.38117
15	SEDIMENT 2	5	0.30	-0.52288	1.40341
16	SEDIMENT 3	1	0.13	-0.88606	-0.12434
17	SEDIMENT 3	2	0.05	-1.30103	-1.46660
18	SEDIMENT 3	3	0.17	-0.76955	0.18676
19	SEDIMENT 3	4	0.08	-1.09691	-0.66680
20	SEDIMENT 3	5	0.22	-0.65758	0.83164

SINGLE-TIME POINT CONTAMINANT BIOACCUMULATION DATA

OBS	TREATMENT GROUP	N	MEAN CONTAMINANT CONC.	VARIANCE	STANDARD ERROR	MEAN LOG10 CONC.	VARIANCE OF LOGS	STANDARD ERROR OF LOGS
1	REFERENCE	5	0.066	.00033	0.008124	-1.19332	0.013772	0.05248
2	SEDIMENT 1	5	0.212	.00347	0.026344	-0.68561	0.012257	0.04951
3	SEDIMENT 2	5	0.190	.00660	0.036332	-0.75469	0.037367	0.08645
4	SEDIMENT 3	5	0.130	.00465	0.030496	-0.94223	0.066666	0.11547

SINGLE-TIME POINT CONTAMINANT BIOACCUMULATION DATA
SHAPIRO-WILKS TEST FOR NORMALITY

UNIVARIATE PROCEDURE

Variable=RESID

N	20
W:Normal	0.957973 Prob<W 0.5111

Variable=RESIDLOG

N	20
W:Normal	0.980207 Prob<W 0.9208

SINGLE-TIME POINT CONTAMINANT BIOACCUMULATION DATA
LEVENE'S TEST
General Linear Models Procedure

Dependent Variable: ABSDEV ABSOLUTE DEVIATIONS FROM MEAN CONC.					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.00647280	0.00215760	2.15	0.1339
Error	16	0.01605600	0.00100350		
Corrected Total	19	0.02252880			

Dependent Variable: ABSLOG ABSOLUTE DEVIATIONS FROM MEAN LOGCONC.					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.04702396	0.01567465	2.19	0.1291
Error	16	0.11456390	0.00716024		
Corrected Total	19	0.16158786			

SINGLE-TIME POINT CONTAMINANT BIOACCUMULATION DATA
LSD TEST (UNTRANSFORMED DATA)

General Linear Models Procedure

T tests (LSD) for variable: CONC

NOTE: This test controls the type I comparisonwise error rate not the experimentwise error rate.

Alpha= 0.1 df= 16 MSE= 0.003763
Critical Value of T= 1.75
Least Significant Difference= 0.0677

Means with the same letter are not significantly different.

T Grouping	Mean	N	TRT
A	0.2120	5	SEDIMENT 1
A			
B	0.1900	5	SEDIMENT 2
B			
B	0.1300	5	SEDIMENT 3
C			
C	0.0660	5	REFERENCE

LSD TEST (LOG-TRANSFORMED DATA)

Alpha= 0.1 df= 16 MSE= 0.032515
 Critical Value of T= 1.75
 Least Significant Difference= 0.1991

Means with the same letter are not significantly different.

T Grouping	Mean	N	TRT
A	-0.686	5	SEDIMENT 1
A			
B	-0.755	5	SEDIMENT 2
B			
B	-0.942	5	SEDIMENT 3
C	-1.193	5	REFERENCE

SINGLE-TIME POINT CONTAMINANT BIOACCUMULATION DATA
 T-TEST

TTEST PROCEDURE

Variable: CONC CONTAMINANT CONCENTRATION, ug/g

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	0.06600000	0.01816590	0.00812404
SEDIMENT 1	5	0.21200000	0.05890671	0.02634388

Variances	T	DF	Prob> T
Unequal	-5.2960	4.8	0.0039
Equal	-5.2960	8.0	0.0007

For H0: Variances are equal, $F' = 10.52$ DF = (4,4) Prob>F' = 0.0426

Variable: LOGCONC LOG10 CONCENTRATION

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	-1.19331525	0.11735241	0.05248159
SEDIMENT 1	5	-0.68561391	0.11071260	0.04951218

Variances	T	DF	Prob> T
Unequal	-7.0366	8.0	0.0001
Equal	-7.0366	8.0	0.0001

For H0: Variances are equal, $F' = 1.12$ DF = (4,4) Prob>F' = 0.9128

SINGLE-TIME POINT CONTAMINANT BIOACCUMULATION DATA
T-TEST

TTEST PROCEDURE

Variable: CONC CONTAMINANT CONCENTRATION, ug/g

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	0.06600000	0.01816590	0.00812404
SEDIMENT 2	5	0.19000000	0.08124038	0.03633180

VariANCES	T	DF	Prob> T	
Unequal	-3.3307	4.4	0.0258	
Equal	-3.3307	8.0	0.0104	

For H0: Variances are equal, F' = 20.00 DF = (4,4) Prob>F' = 0.0132

Variable: LOGCONC LOG10 CONCENTRATION

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	-1.19331525	0.11735241	0.05248159
SEDIMENT 2	5	-0.75469033	0.19330562	0.08644890

VariANCES	T	DF	Prob> T	
Unequal	-4.3371	6.6	0.0040	
Equal	-4.3371	8.0	0.0025	

For H0: Variances are equal, F' = 2.71 DF = (4,4) Prob>F' = 0.3570

SINGLE-TIME POINT CONTAMINANT BIOACCUMULATION DATA
T-TEST

TTEST PROCEDURE

Variable: CONC CONTAMINANT CONCENTRATION, ug/g

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	0.06600000	0.01816590	0.00812404
SEDIMENT 3	5	0.13000000	0.06819091	0.03049590

VariANCES	T	DF	Prob> T	
Unequal	-2.0279	4.6	0.1045	
Equal	-2.0279	8.0	0.0771	

For H0: Variances are equal, F' = 14.09 DF = (4,4) Prob>F' = 0.0252

Variable: LOGCONC LOG10 CONCENTRATION

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	-1.19331525	0.11735241	0.05248159
SEDIMENT 3	5	-0.94222501	0.25819757	0.11546947

Variances	T	DF	Prob> T
Unequal	-1.9796	5.6	0.0990
Equal	-1.9796	8.0	0.0831

For H0: Variances are equal, $F' = 4.84$ DF = (4,4) Prob>F' = 0.1558

SINGLE-TIME POINT CONTAMINANT BIOACCUMULATION DATA
BIOACCUMULATION DATA CONVERTED TO RANKITS
SHAPIRO-WILKS TEST FOR NORMALITY

UNIVARIATE PROCEDURE

Variable=RESID

N 20
W:Normal 0.972308 Prob<W 0.7907

SINGLE-TIME POINT CONTAMINANT BIOACCUMULATION DATA
BIOACCUMULATION DATA CONVERTED TO RANKITS
LEVENE'S TEST

General Linear Models Procedure

Dependent Variable: ABSDEV

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.24147324	0.08049108	0.61	0.6212
Error	16	2.12865866	0.13304117		
Corrected Total	19	2.37013190			

SINGLE-TIME POINT CONTAMINANT BIOACCUMULATION DATA
BIOACCUMULATION DATA CONVERTED TO RANKITS
LSD TEST

General Linear Models Procedure

T tests (LSD) for variable: RANKIT

NOTE: This test controls the type I comparisonwise error rate not the experimentwise error rate.

Alpha= 0.1 df= 16 MSE= 0.503649
 Critical Value of T= 1.75
 Least Significant Difference= 0.7836

Means with the same letter are not significantly different.

T Grouping	Mean	N	TRT
A	0.746	5	SEDIMENT 1
A			
B A	0.495	5	SEDIMENT 2
B			
B C	-0.248	5	SEDIMENT 3
C			
C	-0.993	5	REFERENCE

SINGLE-TIME POINT CONTAMINANT BIOACCUMULATION DATA
 BIOACCUMULATION DATA CONVERTED TO RANKITS

TTEST PROCEDURE

Variable: RANKIT

RANK FOR VARIABLE CONC

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	-0.99338019	0.46306944	0.20709095
SEDIMENT 1	5	0.74648762	0.68780736	0.30759680

Variances	T	DF	Prob> T
Unequal	-4.6920	7.0	0.0022
Equal	-4.6920	8.0	0.0016

For H0: Variances are equal, $F' = 2.21$ DF = (4,4) Prob>F' = 0.4623

SINGLE-TIME POINT CONTAMINANT BIOACCUMULATION DATA
 BIOACCUMULATION DATA CONVERTED TO RANKITS

TTEST PROCEDURE

Variable: RANKIT

RANK FOR VARIABLE CONC

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	-0.99338019	0.46306944	0.20709095
SEDIMENT 2	5	0.49476200	0.75465812	0.33749337

Variances	T	DF	Prob> T
Unequal	-3.7583	6.6	0.0079
Equal	-3.7583	8.0	0.0056

For H0: Variances are equal, $F' = 2.66$ DF = (4,4) Prob>F' = 0.3671

SINGLE-TIME POINT CONTAMINANT BIOACCUMULATION DATA
BIOACCUMULATION DATA CONVERTED TO RANKITS

TTEST PROCEDURE

Variable: RANKIT		RANK FOR VARIABLE CONC		
TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	-0.99338019	0.46306944	0.20709095
SEDIMENT 3	5	-0.24786944	0.87038805	0.38924937

Variances	T	DF	Prob> T
Unequal	-1.6908	6.1	0.1411
Equal	-1.6908	8.0	0.1293

For H0: Variances are equal, $F' = 3.53$ $DF = (4,4)$ $Prob>F' = 0.2491$

SINGLE-TIME POINT CONTAMINANT BIOACCUMULATION DATA
POWER OF LSD TO DETECT A TRUE POPULATION DIFFERENCE (D)
ABOVE REFERENCE MEAN CONTAMINANT CONCENTRATION

NO. OF REPLICATES, N	REFERENCE MEAN CONTAMINANT CONCENTRATION	MEAN SQUARE ERROR, MSE	DEGREES OF FREEDOM, DF	T VALUE FOR (1-ALPHA=0.95, DF)
5	0.066	.0037625	16	1.74588

POWER OF LSD TO DETECT % INCREASE IN CONCENTRATION ABOVE REFERENCE
MEAN CONTAMINANT CONCENTRATION GIVEN N, MSE AND DF SHOWN ABOVE

% INCREASE IN CONC. ABOVE REFERENCE	DREDGED SEDIMENT BIOACCUMULATION	D	T VALUE FOR (1-BETA, DF)	POWER (1-BETA)
10	0.0726	0.0066	-1.57576	0.06732
25	0.0825	0.0165	-1.32056	0.10261
50	0.0990	0.0330	-0.89524	0.19196
100	0.1320	0.0660	-0.04460	0.48249
200	0.1980	0.1320	1.65668	0.94147
300	0.2640	0.1980	3.35796	0.99800

MINIMUM DREDGED SEDIMENT BIOACCUMULATION THAT CAN BE DETECTED BY LSD
AS SIGNIFICANT GIVEN SPECIFIED POWER AND N, MSE, AND DF SHOWN ABOVE

POWER (1-BETA)	D	DREDGED SEDIMENT BIOACCUMULATION	% INCREASE IN CONC. ABOVE REFERENCE	T VALUE FOR (1-BETA, DF)
0.50	0.06773	0.13373	102.622	0.00000
0.60	0.07772	0.14372	117.763	0.25760
0.70	0.08849	0.15449	134.069	0.53501
0.80	0.10127	0.16727	153.446	0.86467
0.90	0.11959	0.18559	181.195	1.33676
0.95	0.13546	0.20146	205.244	1.74588
0.99	0.16796	0.23396	254.477	2.58349

COMPARISON OF MEAN DREDGED SEDIMENT BIOACCUMULATION WITH ACTION LEVEL:
UPPER CONFIDENCE LIMITS (UCL) WHEN VARIANCES ARE EQUAL

TREATMENT GROUP	MEAN BIOACCUMULATION	UCL (EQUAL VARIANCES)	MEAN SQUARE ERROR	T VALUE FOR (1-ALPHA=.95, DF)	DF	MINIMUM SIGNIFICANT DIFFERENCE
SEDIMENT 1	0.212	0.25989	.0037625	1.74588	16	0.047893
SEDIMENT 2	0.190	0.23789	.0037625	1.74588	16	0.047893
SEDIMENT 3	0.130	0.17789	.0037625	1.74588	16	0.047893

COMPARISON OF MEAN DREDGED SEDIMENT BIOACCUMULATION WITH ACTION LEVEL:
UPPER CONFIDENCE LIMITS (UCL) WHEN VARIANCES ARE UNEQUAL

TREATMENT GROUP	MEAN BIOACCUMULATION	UCL (UNEQUAL VARIANCES)	VARIANCE	T VALUE FOR (1-ALPHA=.95, N-1)	N	MINIMUM SIGNIFICANT DIFFERENCE
SEDIMENT 1	0.212	0.26816	.00347	2.13185	5	0.056161
SEDIMENT 2	0.190	0.26745	.00660	2.13185	5	0.077454
SEDIMENT 3	0.130	0.19501	.00465	2.13185	5	0.065013

POWER TO DETECT % DECREASE IN CONCENTRATION BELOW
ACTION LEVEL OF 0.2 ug/g GIVEN N, MSE AND DF SHOWN ABOVE

% DECREASE BELOW ACTION LEVEL	MEAN DREDGED SEDIMENT BIOACCUMULATION	D	T VALUE FOR (1-BETA, DF)	POWER (1-BETA)
10	0.18	0.02	-1.01680	0.16219
20	0.16	0.04	-0.28772	0.38863
30	0.14	0.06	0.44136	0.66757
40	0.12	0.08	1.17045	0.87052
50	0.10	0.10	1.89953	0.96216

D4.4 Program BIOACCSS.SAS for Time-Sequenced Bioaccumulation Test Data Analysis

BIOACCSS.SAS is a program to compare Tier IV estimated steady-state bioaccumulation (C_{ss}) from dredged sediments vs. reference sediment, using untransformed data and \log_{10} transformation. Included are: data plots, estimation of C_{ss} , mean C_{ss} from each sediment exposure, Shapiro-Wilk's test for normality, Levene's test for equality of variances, LSD test, t -tests for equal or unequal variances, and tests on rankits (normalized ranks for C_{ss}). Refer to the decision tree in Figures D-5A and 5B to determine which test results should be used. The program includes power calculations for an LSD test on untransformed C_{ss} estimates.

D4.4.1 BIOACCSS.SAS Program Statements

```
LIBNAME Q 'C:\SAS';
OPTIONS LINESIZE=79 PAGESIZE=59 NONUMBER NODATE;

/* Identify the treatment codes. */

PROC FORMAT;
  VALUE TRTFMT
    1='REFERENCE '
    2='SEDIMENT 1'
    3='SEDIMENT 2'
    4='SEDIMENT 3';

/* Input the bioaccumulation data after the CARDS statement, listing the */
/* day, replicate, treatment code, and contaminant concentration. A */
/* permanent SAS data set is created in the directory specified in the */
/* LIBNAME statement. */

DATA Q.BIOACCSS;
  INPUT DAY REP TRT CONC @@;
  CARDS;
2 1 1 .054 2 2 1 .163 2 3 1 .391 2 4 1 .234 2 5 1 .034
2 1 2 .159 2 2 2 .292 2 3 2 .428 2 4 2 .558 2 5 2 .256
2 1 3 .869 2 2 3 .726 2 3 3 .394 2 4 3 1.232 2 5 3 .977
2 1 4 .745 2 2 4 1.703 2 3 4 2.045 2 4 4 1.855 2 5 4 1.135
4 1 1 .441 4 2 1 .797 4 3 1 .203 4 4 1 .564 4 5 1 .018
4 1 2 .516 4 2 2 .158 4 3 2 .743 4 4 2 .324 4 5 2 .126
4 1 3 .838 4 2 3 .633 4 3 3 .452 4 4 3 .728 4 5 3 1.314
4 1 4 1.316 4 2 4 .930 4 3 4 2.141 4 4 4 1.150 4 5 4 1.621
7 1 1 .687 7 2 1 .177 7 3 1 .862 7 4 1 .413 7 5 1 .029
7 1 2 .881 7 2 2 .317 7 3 2 .270 7 4 2 .562 7 5 2 .603
7 1 3 1.246 7 2 3 .816 7 3 3 .897 7 4 3 1.639 7 5 3 .688
7 1 4 1.583 7 2 4 2.715 7 3 4 1.016 7 4 4 2.221 7 5 4 2.134
10 1 1 .037 10 2 1 .549 10 3 1 .884 10 4 1 .787 10 5 1 .294
10 1 2 .278 10 2 2 .485 10 3 2 .051 10 4 2 .909 10 5 2 .718
10 1 3 1.767 10 2 3 1.272 10 3 3 1.003 10 4 3 1.158 10 5 3 1.415
10 1 4 1.578 10 2 4 2.268 10 3 4 1.756 10 4 4 2.899 10 5 4 .890
18 1 1 .856 18 2 1 .598 18 3 1 .016 18 4 1 .806 18 5 1 .119
18 1 2 .904 18 2 2 1.300 18 3 2 .671 18 4 2 .934 18 5 2 1.173
18 1 3 1.631 18 2 3 1.877 18 3 3 1.487 18 4 3 1.216 18 5 3 1.280
18 1 4 2.822 18 2 4 2.607 18 3 4 3.414 18 4 4 1.319 18 5 4 1.866
28 1 1 .514 28 2 1 .839 28 3 1 .793 28 4 1 .899 28 5 1 .226
28 1 2 .172 28 2 2 1.049 28 3 2 .476 28 4 2 .712 28 5 2 1.245
28 1 3 1.178 28 2 3 1.721 28 3 3 1.366 28 4 3 1.513 28 5 3 1.843
28 1 4 1.295 28 2 4 2.964 28 3 4 2.109 28 4 4 2.820 28 5 4 3.325
;

/* Specify contaminant concentrations in the sediments. Format, sort, */
/* and print the data. */

DATA AA;
  SET Q.BIOACCSS;
  SELECT (TRT);
    WHEN (1) CS=.45;
    WHEN (2) CS=4;
    WHEN (3) CS=33;
    WHEN (4) CS=44;
  OTHERWISE;
  END;
  LABEL TRT='TREATMENT GROUP'
```

```
REP='REPLICATE'
CONC='CONC. IN TISSUE'
CS='CONC. IN SEDIMENT';
FORMAT TRT TRTFMT.;
TITLE 'TIME-SEQUENCED BIOACCUMULATION';
PROC SORT; BY TRT REP;
PROC PRINT LABEL; BY TRT; VAR REP DAY CONC CS;

/* Plot the data by treatment group, identifying the replicates. Plots */
/* may be sent to the screen using the first GOPTIONS statement, or to a */
/* printer using the second GOPTIONS statement. Consult the SAS/GRAPH */
/* User's Guide (SAS Institute, Inc., 1988c) for appropriate device names */
/* and instructions for GACCESS=. */

*GOPTIONS DEVICE=VGA;
GOPTIONS DEVICE=HPLJ3P GACCESS='SASGASTD>LPT2:' VSIZE=6 IN HSIZE=6.5 IN
  VORIGIN=3 IN HORIGIN=0.3 IN;
PROC GPLOT UNIFORM; BY TRT;
  PLOT CONC*DAY=REP;

/* Perform nonlinear regressions on each treatment and replicate. */
/* If you wish to use a method other than DUD, include the following */
/* derivative statements after the MODEL statement: DER.K1=CS/K2*(1-EX); */
/* and DER.K2=CS*(K1/K2)*(DAY*EX-(1-EX)/K2);. Save regression parameters */
/* in a permanent SAS data set. */

PROC NLIN BEST=10 METHOD=DUD;
  BY TRT REP;
  PARS K1=0 TO 3 BY .1 K2=.01 TO 2 BY .1;
  EX=EXP(-K2*DAY);
  MODEL CONC=CS*(K1/K2)*(1-EX);
  OUTPUT OUT=Q.REGPARMS PARS=K1 K2;

/* Calculate and print Css and regression parameters. Log-transform Css. */
/* Calculate rankits. Save these variables in a permanent SAS data set. */

DATA A;
  SET Q.REGPARMS;
  IF DAY<28 THEN DELETE;
  CSS=CS*K1/K2;
  LOGCSS=LOG10(CSS);
  DROP DAY CONC;
  LABEL    CSS='STEADY STATE CONC., Css'
          LOGCSS='Log10 Css'
          K1='UPTAKE RATE CONSTANT, k1'
          K2='DEPURATION RATE CONSTANT, k2';
  MERGEVAR=1;

PROC RANK NORMAL=BLOM OUT=Q.CSS;
  VAR CSS; RANKS RANKIT;
PROC PRINT LABEL DATA=Q.CSS; VAR TRT REP K1 K2 CSS LOGCSS RANKIT;
  LABEL    RANKIT='NORMALIZED RANK FOR Css';

/* Calculate and print descriptive statistics for Css and logCss. */

PROC MEANS NOPRINT DATA=Q.CSS; BY TRT; VAR CSS LOGCSS; ID MERGEVAR;
  OUTPUT OUT=Y N=N NLOG MEAN=MEANCSS MEANLOG VAR=S2 S2LOG STDERR=SE SELOG;
PROC PRINT LABEL; VAR TRT N MEANCSS S2 SE MEANLOG S2LOG SELOG;
  LABEL    MEANCSS='MEAN Css'
          S2='VARIANCE'
          SE='STANDARD ERROR'
          MEANLOG='MEAN Log10 Css'
          S2LOG='VARIANCE OF LOGS'
```

```
SELOG='STANDARD ERROR OF LOGS';

/* Test normality of residuals of untransformed and log-transformed Css */
/* using Shapiro-Wilk's Test. */

PROC GLM NOPRINT DATA=Q.CSS;
  CLASS TRT;
  MODEL CSS LOGCSS=TRT;
  OUTPUT OUT=Z R=RESID RESIDLOG;
PROC UNIVARIATE NORMAL;
  VAR RESID RESIDLOG;
  TITLE2 'SHAPIRO-WILKS TEST FOR NORMALITY';
/* Conduct Levene's Test for equality of variances of untransformed and */
/* log-transformed Css. */

DATA AX;
  MERGE Q.CSS Y; BY TRT;
  ABSDEV=ABS(CSS-MEANCSS);
  ABSLOG=ABS(LOGCSS-MEANLOG);
  LABEL   ABSDEV='ABSOLUTE DEVIATIONS FROM Css MEAN'
          ABSLOG='ABSOLUTE DEVIATIONS FROM logCss MEAN';
PROC GLM;
  CLASS TRT;
  MODEL ABSDEV ABSLOG=TRT;
  TITLE2 'LEVENE'S TEST';

/* Perform LSD on untransformed and log-transformed Css. */

PROC GLM DATA=Q.CSS OUTSTAT=W1;
  CLASS TRT;
  MODEL CSS=TRT;
  MEANS TRT/LSD ALPHA=.1;
  TITLE2 'LSD TEST (UNTRANSFORMED DATA)';
PROC GLM DATA=Q.CSS OUTSTAT=W2;
  CLASS TRT;
  MODEL LOGCSS=TRT;
  MEANS TRT/LSD ALPHA=.1;
  TITLE2 'LSD TEST (LOG-TRANSFORMED DATA)';

/* Perform t-tests for each dredged sediment-reference sediment comparison */
/* using untransformed and log-transformed Css. */

DATA T1;
  SET Q.CSS;
  IF TRT>2 THEN DELETE;
PROC TTEST;
  CLASS TRT;
  VAR CSS LOGCSS;
  TITLE2 'T-TEST';
DATA T2;
  SET Q.CSS;
  IF TRT=2 OR TRT=4 THEN DELETE;
PROC TTEST;
  CLASS TRT;
  VAR CSS LOGCSS;
DATA T3;
  SET Q.CSS;
  IF TRT=2 OR TRT=3 THEN DELETE;
PROC TTEST;
  CLASS TRT;
  VAR CSS LOGCSS;

/* Test normality and equality of variances of rankits. */
```

```
PROC GLM NOPRINT DATA=Q.CSS;
  CLASS TRT;
  MODEL RANKIT=TRT;
  OUTPUT OUT=Z1 R=RESID;
  TITLE2 'Css CONVERTED TO RANKITS';
PROC UNIVARIATE NORMAL;
  VAR RESID;
  TITLE3 'SHAPIRO-WILKS TEST FOR NORMALITY';
PROC MEANS DATA=Q.CSS NOPRINT;
  BY TRT; VAR RANKIT;
  OUTPUT OUT=X2 MEAN=MEAN;
DATA AX2;
  MERGE Q.CSS X2; BY TRT;
  ABSDEV=ABS(RANKIT-MEAN);
PROC GLM;
  CLASS TRT;
  MODEL ABSDEV=TRT;
  TITLE3 'LEVENE'S TEST';

/* Perform LSD on rankits. */

PROC GLM DATA=Q.CSS;
  CLASS TRT;
  MODEL RANKIT=TRT;
  MEANS TRT/LSD ALPHA=.1;
  TITLE3 'LSD TEST';

/* Perform t-tests for each dredged sediment-reference sediment comparison */
/* using rankits. */

PROC TTEST DATA=T1;
  CLASS TRT; VAR RANKIT;
  TITLE3 'T-TEST';
PROC TTEST DATA=T2;
  CLASS TRT; VAR RANKIT;
PROC TTEST DATA=T3;
  CLASS TRT; VAR RANKIT;

/* Calculate power of an LSD test to detect true population differences */
/* 10, 25, 50, and 100% above the reference mean Css. */

DATA C1;
  SET W1;
  IF _TYPE_ ^= 'ERROR' THEN DELETE;
  MSE=SS/DF;
  MERGEVAR=1;
  KEEP MSE DF MERGEVAR;
DATA C2;
  SET Y;
  IF TRT ^= 1 THEN DELETE;
DATA C3;
  MERGE C1 C2;
  TALPHA=TINV(.95,DF);
  LABEL      N='NO. OF REPLICATES, N'
            MEANCSS='REFERENCE MEAN Css'
            MSE='MEAN SQUARE ERROR, MSE'
            DF='DEGREES OF FREEDOM, DF'
            TALPHA='T VALUE FOR (1-ALPHA=0.95,DF)';
  TITLE2 'POWER OF LSD TO DETECT A TRUE POPULATION DIFFERENCE (D)';
  TITLE3 'ABOVE REFERENCE MEAN Css';
PROC PRINT LABEL NOOBS; VAR N MEANCSS MSE DF TALPHA;
DATA C4;
  SET C3;
```

```

DO PCTDIFF=10,25,50,100,200,300;
  SEDCSS=MEANCSS+((PCTDIFF/100)*MEANCSS);
  D=SEDCSS-MEANCSS;
  TBETA=D*SQRT(N/(2*MSE))-TALPHA;
  POWER=PROBT(TBETA,DF);
  OUTPUT;
END;
LABEL      PCTDIFF='% INCREASE IN Css ABOVE REFERENCE'
          SEDCSS='DREDGED SEDIMENT Css'
          TBETA='T VALUE FOR (1-BETA,DF)'
          POWER='POWER (1-BETA)';
PROC PRINT LABEL NOOBS; VAR PCTDIFF SEDCSS D TBETA POWER;
  TITLE 'POWER OF LSD TO DETECT % INCREASE IN Css ABOVE REFERENCE';
  TITLE2 'MEAN Css GIVEN N, MSE AND DF SHOWN ABOVE';
DATA C5;
  SET C3;
DO POWER=.5,.6,.7,.8,.9,.95,.99;
  TBETA=TINV(POWER,DF);
  D=((TBETA+TALPHA)*SQRT(2*MSE))/SQRT(N);
  SEDCSS=MEANCSS+D;
  PCTDIFF=(D*100)/MEANCSS;
  OUTPUT;
END;
LABEL      SEDCSS='DREDGED SEDIMENT Css'
          PCTDIFF='% INCREASE IN Css ABOVE REFERENCE'
          TBETA='T VALUE FOR (1-BETA,DF)'
          POWER='POWER (1-BETA)';
PROC PRINT LABEL NOOBS; VAR POWER D SEDCSS PCTDIFF TBETA;
  TITLE 'MINIMUM DREDGED SEDIMENT Css THAT CAN BE DETECTED BY LSD';
  TITLE2 'AS SIGNIFICANT GIVEN SPECIFIED POWER AND N, MSE, AND DF SHOWN ABOVE';

/* Calculation of upper confidence limits (UCL) for comparison of mean */
/* dredged sediment Css with an action level. */

DATA D;
MERGE C1 Y; BY MERGEVAR;
IF TRT=1 THEN DELETE;
TALPHA1=TINV(.95,DF);
TALPHA2=TINV(.95,N-1);
UCL1=MEANCSS+TALPHA1*(SQRT(MSE/N));
UCL2=MEANCSS+TALPHA2*(SQRT(S2/N));
DMIN1=TALPHA1*SQRT(MSE/N);
DMIN2=TALPHA2*SQRT(S2/N);
LABEL      UCL1='UCL (EQUAL VARIANCES)'
          UCL2='UCL (UNEQUAL VARIANCES)'
          TALPHA1='T VALUE FOR (1-ALPHA=.95,DF)'
          TALPHA2='T VALUE FOR (1-ALPHA=.95,N-1)'
          DMIN1='MINIMUM SIGNIFICANT DIFFERENCE'
          DMIN2='MINIMUM SIGNIFICANT DIFFERENCE'
          MSE='MEAN SQUARE ERROR'
          S2='VARIANCE'
          MEANCSS='MEAN DREDGED SEDIMENT Css';
  TITLE 'COMPARISON OF MEAN DREDGED SEDIMENT Css WITH ACTION LEVEL: ';
PROC PRINT LABEL NOOBS; VAR TRT MEANCSS UCL1 MSE TALPHA1 DF DMIN1;
  TITLE2 'UPPER CONFIDENCE LIMITS (UCL) WHEN VARIANCES ARE EQUAL';
PROC PRINT LABEL NOOBS; VAR TRT MEANCSS UCL2 S2 TALPHA2 N DMIN2;
  TITLE2 'UPPER CONFIDENCE LIMITS (UCL) WHEN VARIANCES ARE UNEQUAL';

/* Calculate power of dredged sediment-action level comparisons using */
/* MSE given 10, 20, 30, 40, and 50% decreases in mean Css below */
/* action level. */

```

```

DATA D1;
SET C3;
ACTION=2;
DO PCTDIFF=10 TO 50 BY 10;
D=PCTDIFF*ACTION/100;
SEDCSS=ACTION-D;
TBETA=D*SQRT(N/MSE)-TALPHA;
POWER=PROBT(TBETA,DF);
OUTPUT;
END;
LABEL    PCTDIFF='% DECREASE BELOW ACTION LEVEL'
        SEDCSS='DREDGED SEDIMENT Css'
        TBETA='T VALUE FOR (1-BETA,DF)'
        POWER='POWER (1-BETA)';
PROC PRINT NOOBS LABEL; VAR PCTDIFF SEDCSS D TBETA POWER;
TITLE 'POWER TO DETECT % DECREASE IN Css BELOW';
TITLE2 'ACTION LEVEL OF 2 ug/g GIVEN N, MSE AND DF SHOWN ABOVE';

```

D4.4.2 BIOACCSS.SAS Program Output

TIME-SEQUENCED BIOACCUMULATION

----- TREATMENT GROUP=REFERENCE -----

OBS	REPLICATE	DAY	CONC. IN TISSUE	CONC. IN SEDIMENT
1	1	2	0.054	0.45
2	1	4	0.441	0.45
3	1	7	0.687	0.45
4	1	10	0.037	0.45
5	1	18	0.856	0.45
6	1	28	0.514	0.45
7	2	2	0.163	0.45
8	2	4	0.797	0.45
9	2	7	0.177	0.45
10	2	10	0.549	0.45
11	2	18	0.598	0.45
12	2	28	0.839	0.45
13	3	2	0.391	0.45
14	3	4	0.203	0.45
15	3	7	0.862	0.45
16	3	10	0.884	0.45
17	3	18	0.016	0.45
18	3	28	0.793	0.45
19	4	2	0.234	0.45
20	4	4	0.564	0.45
21	4	7	0.413	0.45
22	4	10	0.787	0.45
23	4	18	0.806	0.45
24	4	28	0.899	0.45
25	5	2	0.034	0.45
26	5	4	0.018	0.45
27	5	7	0.029	0.45
28	5	10	0.294	0.45
29	5	18	0.119	0.45
30	5	28	0.226	0.45

----- TREATMENT GROUP=SEDIMENT 1 -----

31	1	2	0.159	4
32	1	4	0.516	4
33	1	7	0.881	4
34	1	10	0.278	4
35	1	18	0.904	4
36	1	28	0.172	4
37	2	2	0.292	4
38	2	4	0.158	4
39	2	7	0.317	4
40	2	10	0.485	4
41	2	18	1.300	4
42	2	28	1.049	4
43	3	2	0.428	4
44	3	4	0.743	4
45	3	7	0.270	4
46	3	10	0.051	4
47	3	18	0.671	4
48	3	28	0.476	4
49	4	2	0.558	4
50	4	4	0.324	4
51	4	7	0.562	4
52	4	10	0.909	4
53	4	18	0.934	4
54	4	28	0.712	4
55	5	2	0.256	4
56	5	4	0.126	4
57	5	7	0.603	4
58	5	10	0.718	4
59	5	18	1.173	4
60	5	28	1.245	4

----- TREATMENT GROUP=SEDIMENT 2 -----

61	1	2	0.869	33
62	1	4	0.838	33
63	1	7	1.246	33
64	1	10	1.767	33
65	1	18	1.631	33
66	1	28	1.178	33
67	2	2	0.726	33
68	2	4	0.633	33
69	2	7	0.816	33
70	2	10	1.272	33
71	2	18	1.877	33
72	2	28	1.721	33
73	3	2	0.394	33
74	3	4	0.452	33
75	3	7	0.897	33
76	3	10	1.003	33
77	3	18	1.487	33
78	3	28	1.366	33
79	4	2	1.232	33
80	4	4	0.728	33
81	4	7	1.639	33

TIME-SEQUENCED BIOACCUMULATION

82	4	10	1.158	33
83	4	18	1.216	33
84	4	28	1.513	33
85	5	2	0.977	33
86	5	4	1.314	33
87	5	7	0.688	33
88	5	10	1.415	33
89	5	18	1.280	33
90	5	28	1.843	33

----- TREATMENT GROUP=SEDIMENT 3 -----

91	1	2	0.745	44
92	1	4	1.316	44
93	1	7	1.583	44
94	1	10	1.578	44
95	1	18	2.822	44
96	1	28	1.295	44
97	2	2	1.703	44
98	2	4	0.930	44
99	2	7	2.715	44
100	2	10	2.268	44
101	2	18	2.607	44
102	2	28	2.964	44
103	3	2	2.045	44
104	3	4	2.141	44
105	3	7	1.016	44
106	3	10	1.756	44
107	3	18	3.414	44
108	3	28	2.109	44
109	4	2	1.855	44
110	4	4	1.150	44
111	4	7	2.221	44
112	4	10	2.899	44
113	4	18	1.319	44
114	4	28	2.820	44
115	5	2	1.135	44
116	5	4	1.621	44
117	5	7	2.134	44
118	5	10	0.890	44
119	5	18	1.866	44

TIME-SEQUENCED BIOACCUMULATION

TREATMENT GROUP=REFERENCE

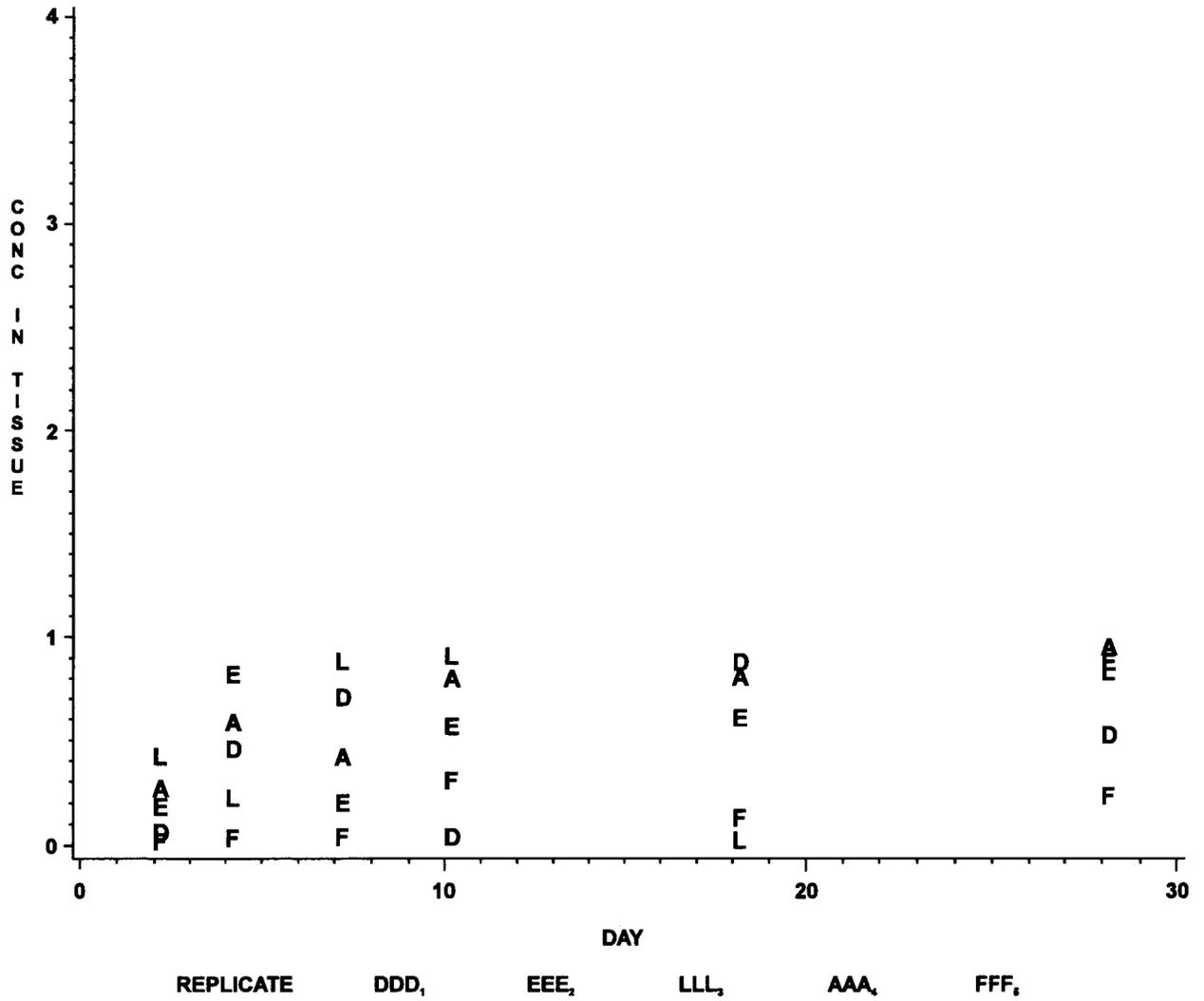


Figure D-8. Plot of Time-Sequenced Bioaccumulation Reference Sediment Example Data by Replicate.

TIME-SEQUENCED BIOACCUMULATION

TREATMENT GROUP=SEDIMENT 1

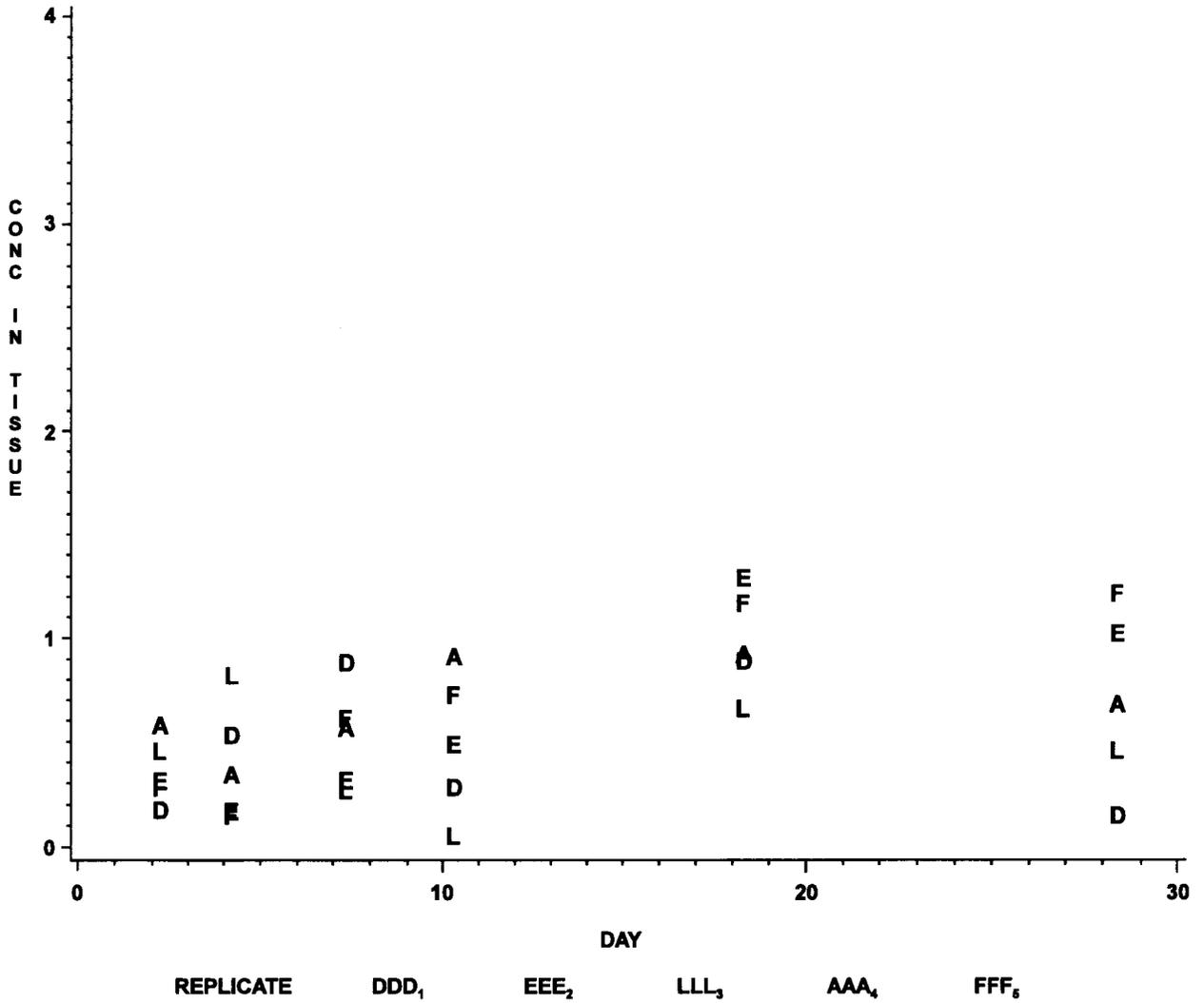


Figure D-9. Plot of Time-Sequenced Bioaccumulation Dredged Sediment 1 Example Data by Replicate.

TIME-SEQUENCED BIOACCUMULATION

TREATMENT GROUP=SEDIMENT 2

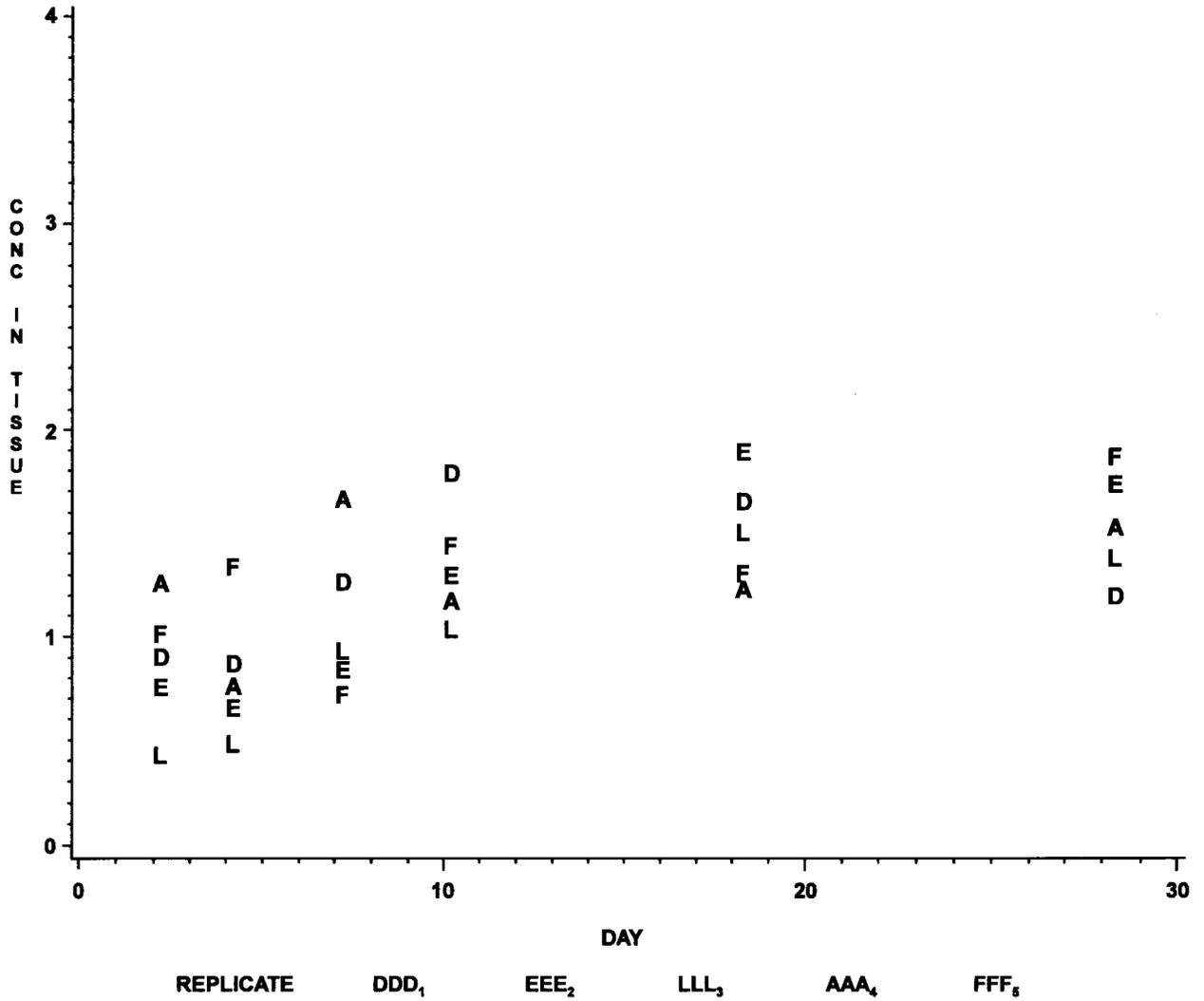


Figure D-10. Plot of Time-Sequenced Bioaccumulation Dredged Sediment 2 Example Data by Replicate.

TIME-SEQUENCED BIOACCUMULATION

TREATMENT GROUP=SEDIMENT 3

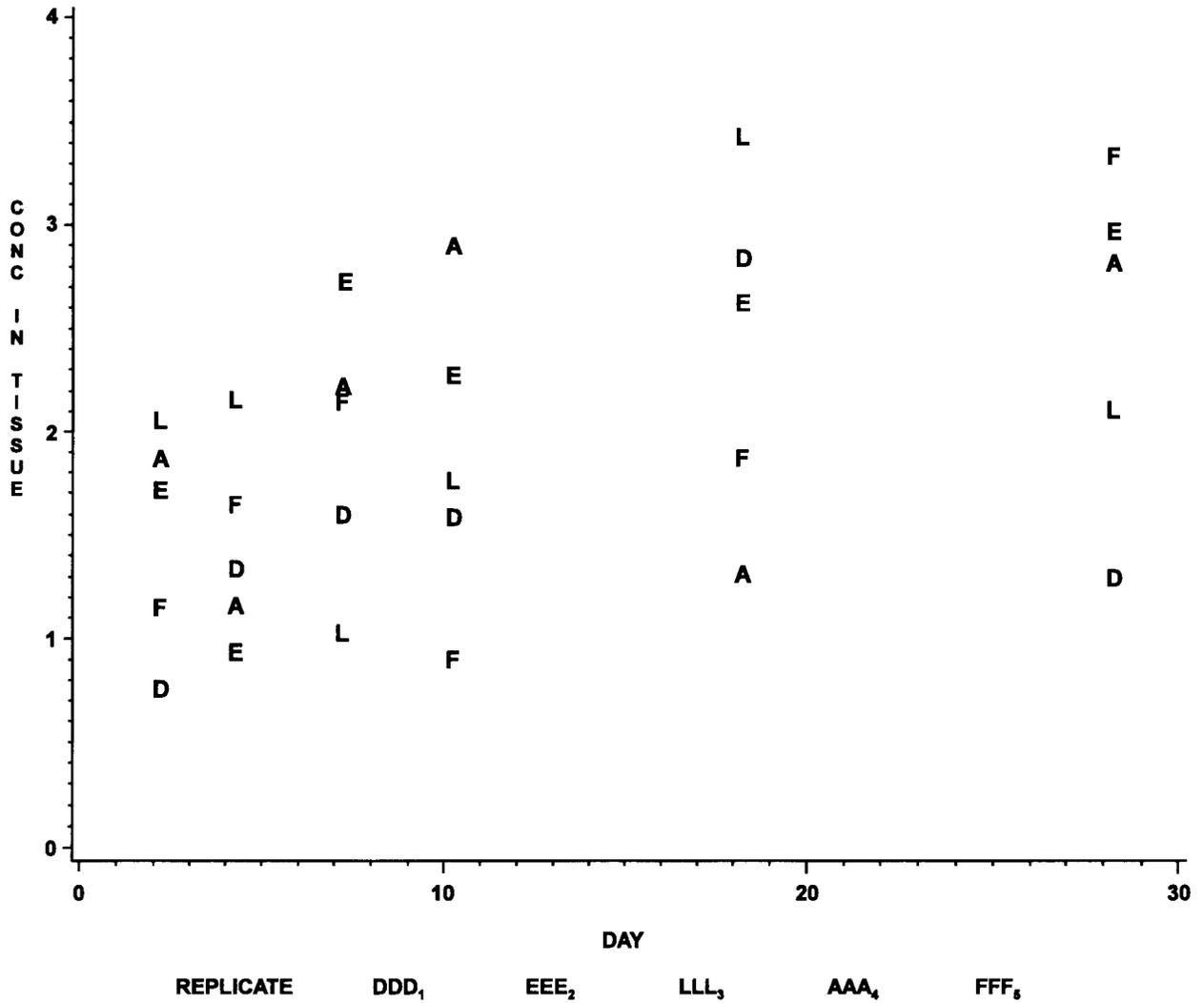


Figure D-11. Plot of Time-Sequenced Bioaccumulation Dredged Sediment 3 Example Data by Replicate.

(Note: the following PROC NLIN output is given as an example only for the reference sediment replicate 1. NLIN output for the other replicates and sediments has been deleted.)

TIME-SEQUENCED BIOACCUMULATION

----- TREATMENT GROUP=REFERENCE REPLICATE=1 -----

Non-Linear Least Squares Grid Search		Dependent Variable CONC	
	K1	K2	Sum of Squares
	0.300000	0.210000	0.416199
	0.400000	0.310000	0.425788
	0.500000	0.410000	0.441222
	0.200000	0.110000	0.448040
	0.400000	0.410000	0.454330
	0.600000	0.510000	0.457317
	0.300000	0.310000	0.457654
	0.500000	0.510000	0.460598
	0.600000	0.610000	0.470393
	0.700000	0.610000	0.472661

Non-Linear Least Squares DUD Initialization		Dependent Variable CONC	
DUD	K1	K2	Sum of Squares
-3	0.300000	0.210000	0.416199
-2	0.330000	0.210000	0.461659
-1	0.300000	0.231000	0.405093

Non-Linear Least Squares Iterative Phase		Dependent Variable CONC		Method: DUD
Iter	K1	K2	Sum of Squares	
0	0.300000	0.231000	0.405093	
1	0.239451	0.178897	0.400026	
2	0.241348	0.179839	0.400014	
3	0.241312	0.179738	0.400013	
4	0.237752	0.176113	0.399983	
5	0.237547	0.175943	0.399983	
6	0.237563	0.175943	0.399983	
7	0.237360	0.175718	0.399983	
8	0.237337	0.175695	0.399983	

NOTE: Convergence criterion met.

Non-Linear Least Squares Summary Statistics			Dependent Variable CONC	
Source	DF	Sum of Squares	Mean Square	
Regression	2	1.2676841229	0.6338420614	
Residual	4	0.3999828771	0.0999957193	
Uncorrected Total	6	1.6676670000		
(Corrected Total)	5	0.5505135000		

Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95 % Confidence Interval	
			Lower	Upper
K1	0.2373370301	0.22487054331	-.38699524147	0.86166930175
K2	0.1756952550	0.21727444929	-.42754716392	0.77893767392

TIME-SEQUENCED BIOACCUMULATION

OBS	TREATMENT GROUP	REPLICATE	UPTAKE RATE CONSTANT, k1	DEPURATION RATE CONSTANT, k2	STEADY STATE CONC., C _{ss}	Log10 C _{ss}	NORMALIZED RANK FOR C _{ss}
1	REFERENCE	1	0.23734	0.17570	0.60788	-0.21618	-0.74414
2	REFERENCE	2	0.30596	0.20060	0.68636	-0.16345	-0.58946
3	REFERENCE	3	0.53975	0.40677	0.59712	-0.22394	-0.91914
4	REFERENCE	4	0.31799	0.16208	0.88285	-0.05411	-0.31457
5	REFERENCE	5	0.04515	0.08670	0.23434	-0.63015	-1.86824
6	SEDIMENT 1	1	0.05916	0.42709	0.55411	-0.25641	-1.12814
7	SEDIMENT 1	2	0.01924	0.04682	1.64392	0.21588	0.44777
8	SEDIMENT 1	3	0.24301	2.20563	0.44071	-0.35584	-1.40341
9	SEDIMENT 1	4	0.05059	0.24290	0.83305	-0.07933	-0.44777
10	SEDIMENT 1	5	0.02419	0.06046	1.60020	0.20418	0.31457
11	SEDIMENT 2	1	0.01439	0.31909	1.48791	0.17258	0.06193
12	SEDIMENT 2	2	0.00653	0.11306	1.90667	0.28028	0.58946
13	SEDIMENT 2	3	0.00548	0.11964	1.51129	0.17935	0.18676
14	SEDIMENT 2	4	0.03430	0.87782	1.28959	0.11045	-0.18676
15	SEDIMENT 2	5	0.02323	0.56773	1.35040	0.13046	-0.06193
16	SEDIMENT 3	1	0.01117	0.25025	1.96371	0.29308	0.74414
17	SEDIMENT 3	2	0.01490	0.23622	2.77595	0.44341	1.86824
18	SEDIMENT 3	3	0.09375	1.97656	2.08697	0.31952	0.91914
19	SEDIMENT 3	4	0.02351	0.45781	2.25943	0.35400	1.12814
20	SEDIMENT 3	5	0.00838	0.13921	2.64810	0.42293	1.40341

TIME-SEQUENCED BIOACCUMULATION

OBS	TREATMENT GROUP	N	MEAN C _{ss}	VARIANCE	STANDARD ERROR	MEAN Log10 C _{ss}	VARIANCE OF LOGS	STANDARD ERROR OF LOGS
1	REFERENCE	5	0.60171	0.05531	0.10517	-0.25757	0.047978	0.09796
2	SEDIMENT 1	5	1.01440	0.32833	0.25625	-0.05430	0.068052	0.11666
3	SEDIMENT 2	5	1.50917	0.05797	0.10768	0.17462	0.004314	0.02937
4	SEDIMENT 3	5	2.34683	0.12421	0.15761	0.36659	0.004214	0.02903

TIME-SEQUENCED BIOACCUMULATION
SHAPIRO-WILKS TEST FOR NORMALITY

UNIVARIATE PROCEDURE

Variable=RESID

N 20
W:Normal 0.963283 Prob<W 0.6122

Variable=RESIDLOG

N 20
W:Normal 0.942525 Prob<W 0.2796

TIME-SEQUENCED BIOACCUMULATION
LEVENE'S TEST
General Linear Models Procedure

Dependent Variable: ABSDEV ABSOLUTE DEVIATIONS FROM C _{SS} MEAN					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.37008913	0.12336304	4.74	0.0150
Error	16	0.41648071	0.02603004		
Corrected Total	19	0.78656984			

Dependent Variable: ABSLOG ABSOLUTE DEVIATIONS FROM logC _{SS} MEAN					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.09646576	0.03215525	3.68	0.0344
Error	16	0.13965602	0.00872850		
Corrected Total	19	0.23612178			

TIME-SEQUENCED BIOACCUMULATION
LSD TEST (UNTRANSFORMED DATA)
General Linear Models Procedure

T tests (LSD) for variable: CSS

NOTE: This test controls the type I comparisonwise error rate not the experimentwise error rate.

Alpha= 0.1 df= 16 MSE= 0.141456
Critical Value of T= 1.75
Least Significant Difference= 0.4153

Means with the same letter are not significantly different.

T Grouping	Mean	N	TRT
A	2.347	5	SEDIMENT 3
B	1.509	5	SEDIMENT 2
C	1.014	5	SEDIMENT 1
C	0.602	5	REFERENCE

TIME-SEQUENCED BIOACCUMULATION
LSD TEST (LOG-TRANSFORMED DATA)
General Linear Models Procedure

T tests (LSD) for variable: LOGCSS

NOTE: This test controls the type I comparisonwise error rate not the experimentwise error rate.

Alpha= 0.1 df= 16 MSE= 0.03114
Critical Value of T= 1.75
Least Significant Difference= 0.1949

Means with the same letter are not significantly different.

T Grouping	Mean	N	TRT
A	0.367	5	SEDIMENT 3
A			
A	0.175	5	SEDIMENT 2
B	-0.054	5	SEDIMENT 1
C	-0.258	5	REFERENCE

TIME-SEQUENCED BIOACCUMULATION
T-TEST

TTEST PROCEDURE

Variable: CSS STEADY STATE CONC., C_{ss}

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	0.60171086	0.23517166	0.10517196
SEDIMENT 1	5	1.01440008	0.57300347	0.25625494

Variances	T	DF	Prob> T
Unequal	-1.4899	5.3	0.1935
Equal	-1.4899	8.0	0.1746

For H₀: Variances are equal, F' = 5.94 DF = (4,4) Prob>F' = 0.1127

Variable: LOGCSS Log₁₀ C_{ss}

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	-0.25756572	0.21903881	0.09795713
SEDIMENT 1	5	-0.05430384	0.26086789	0.11666367

Variances	T	DF	Prob> T
Unequal	-1.3343	7.8	0.2200
Equal	-1.3343	8.0	0.2188

For H₀: Variances are equal, F' = 1.42 DF = (4,4) Prob>F' = 0.7431

TTEST PROCEDURE
STEADY STATE CONC., C_{ss}

Variable: CSS

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	0.60171086	0.23517166	0.10517196
SEDIMENT 2	5	1.50916957	0.24077410	0.10767745

Variances	T	DF	Prob> T
Unequal	-6.0289	8.0	0.0003
Equal	-6.0289	8.0	0.0003

For H0: Variances are equal, F' = 1.05 DF = (4,4) Prob>F' = 0.9647

TTEST PROCEDURE
Log10 C_{ss}

Variable: LOGCSS

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	-0.25756572	0.21903881	0.09795713
SEDIMENT 2	5	0.17462207	0.06568351	0.02937456

Variances	T	DF	Prob> T
Unequal	-4.2261	4.7	0.0097
Equal	-4.2261	8.0	0.0029

For H0: Variances are equal, F' = 11.12 DF = (4,4) Prob>F' = 0.0386

TTEST PROCEDURE
STEADY STATE CONC., C_{ss}

Variable: CSS

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	0.60171086	0.23517166	0.10517196
SEDIMENT 3	5	2.34683295	0.35243662	0.15761445

Variances	T	DF	Prob> T
Unequal	-9.2100	7.0	0.0001
Equal	-9.2100	8.0	0.0000

For H0: Variances are equal, F' = 2.25 DF = (4,4) Prob>F' = 0.4525

TTEST PROCEDURE
Log10 C_{ss}

Variable: LOGCSS

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	-0.25756572	0.21903881	0.09795713
SEDIMENT 3	5	0.36658794	0.06491256	0.02902978

Variances	T	DF	Prob> T
Unequal	-6.1091	4.7	0.0023
Equal	-6.1091	8.0	0.0003

For H0: Variances are equal, F' = 11.39 DF = (4,4) Prob>F' = 0.0370

TIME-SEQUENCED BIOACCUMULATION
 Css CONVERTED TO RANKITS
 SHAPIRO-WILKS TEST FOR NORMALITY

UNIVARIATE PROCEDURE

Variable=RESID

N 20
 W:Normal 0.970187 Prob<W 0.7497

TIME-SEQUENCED BIOACCUMULATION
 Css CONVERTED TO RANKITS
 LEVENE'S TEST

General Linear Models Procedure

Dependent Variable: ABSDEV

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.52458729	0.17486243	1.88	0.1741
Error	16	1.49037397	0.09314837		
Corrected Total	19	2.01496126			

TIME-SEQUENCED BIOACCUMULATION
 Css CONVERTED TO RANKITS
 LSD TEST

General Linear Models Procedure
 T tests (LSD) for variable: RANKIT

NOTE: This test controls the type I comparisonwise error rate not the experimentwise error rate.

Alpha= 0.1 df= 16 MSE= 0.33088
 Critical Value of T= 1.75
 Least Significant Difference= 0.6352

Means with the same letter are not significantly different.

T Grouping	Mean	N	TRT
A	1.213	5	SEDIMENT 3
B	0.118	5	SEDIMENT 2
B			
C B	-0.443	5	SEDIMENT 1
C			
C	-0.887	5	REFERENCE

TIME-SEQUENCED BIOACCUMULATION
 CSS CONVERTED TO RANKITS
 T-TEST

TTEST PROCEDURE

Variable: RANKIT RANK FOR VARIABLE CSS

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	-0.88710960	0.59170982	0.26462068
SEDIMENT 1	5	-0.44339680	0.83054481	0.37143093

Variances	T	DF	Prob> T
Unequal	-0.9729	7.2	0.3621
Equal	-0.9729	8.0	0.3591

For H0: Variances are equal, $F' = 1.97$ $DF = (4,4)$ $Prob>F' = 0.5275$

TTEST PROCEDURE

Variable: RANKIT RANK FOR VARIABLE CSS

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	-0.88710960	0.59170982	0.26462068
SEDIMENT 2	5	0.11789116	0.29807434	0.13330290

Variances	T	DF	Prob> T
Unequal	-3.3918	5.9	0.0151
Equal	-3.3918	8.0	0.0095

For H0: Variances are equal, $F' = 3.94$ $DF = (4,4)$ $Prob>F' = 0.2126$

TTEST PROCEDURE

Variable: RANKIT RANK FOR VARIABLE CSS

TRT	N	Mean	Std Dev	Std Error
REFERENCE	5	-0.88710960	0.59170982	0.26462068
SEDIMENT 3	5	1.21261524	0.44129976	0.19735525

Variances	T	DF	Prob> T
Unequal	-6.3607	7.4	0.0003
Equal	-6.3607	8.0	0.0002

For H0: Variances are equal, $F' = 1.80$ $DF = (4,4)$ $Prob>F' = 0.5839$

TIME-SEQUENCED BIOACCUMULATION
POWER OF LSD TO DETECT A TRUE POPULATION DIFFERENCE (D)
ABOVE REFERENCE MEAN C_{ss}

NO. OF REPLICATES, N	REFERENCE MEAN C_{ss}	MEAN SQUARE ERROR, MSE	DEGREES OF FREEDOM, DF	T VALUE FOR (1-ALPHA=0.95, DF)
5	0.60171	0.14146	16	1.74588

POWER OF LSD TO DETECT % INCREASE IN C_{ss} ABOVE REFERENCE
MEAN C_{ss} GIVEN N, MSE AND DF SHOWN ABOVE

% INCREASE IN C_{ss} ABOVE REFERENCE	DREDGED SEDIMENT C_{ss}	D	T VALUE FOR (1-BETA, DF)	POWER (1-BETA)
10	0.66188	0.06017	-1.49293	0.07746
25	0.75214	0.15043	-1.11349	0.14097
50	0.90257	0.30086	-0.48110	0.31848
100	1.20342	0.60171	0.78369	0.77767
200	1.80513	1.20342	3.31327	0.99780
300	2.40684	1.80513	5.84285	0.99999

MINIMUM DREDGED SEDIMENT C_{ss} THAT CAN BE DETECTED BY LSD
AS SIGNIFICANT GIVEN SPECIFIED POWER AND N, MSE, AND DF SHOWN ABOVE

POWER (1-BETA)	D	DREDGED SEDIMENT C_{ss}	% INCREASE IN C_{ss} ABOVE REFERENCE	T VALUE FOR (1-BETA, DF)
0.50	0.41529	1.01700	69.019	0.00000
0.60	0.47657	1.07828	79.202	0.25760
0.70	0.54256	1.14427	90.169	0.53501
0.80	0.62097	1.22268	103.201	0.86467
0.90	0.73327	1.33498	121.864	1.33676
0.95	0.83059	1.43230	138.038	1.74588
0.99	1.02983	1.63154	171.150	2.58349

COMPARISON OF MEAN DREDGED SEDIMENT C_{ss} WITH ACTION LEVEL:
UPPER CONFIDENCE LIMITS (UCL) WHEN VARIANCES ARE EQUAL

TREATMENT GROUP	MEAN DREDGED SEDIMENT C_{ss}	UCL (EQUAL VARIANCES)	MEAN SQUARE ERROR	T VALUE FOR (1-ALPHA=.95, DF)	DF	MINIMUM SIGNIFICANT DIFFERENCE
SEDIMENT 1	1.01440	1.30806	0.14146	1.74588	16	0.29366
SEDIMENT 2	1.50917	1.80283	0.14146	1.74588	16	0.29366
SEDIMENT 3	2.34683	2.64049	0.14146	1.74588	16	0.29366

COMPARISON OF MEAN DREDGED SEDIMENT C_{ss} WITH ACTION LEVEL:
UPPER CONFIDENCE LIMITS (UCL) WHEN VARIANCES ARE UNEQUAL

TREATMENT GROUP	MEAN DREDGED SEDIMENT C_{ss}	UCL (UNEQUAL VARIANCES)	VARIANCE	T VALUE FOR (1-ALPHA=.95,N-1)	N	MINIMUM SIGNIFICANT DIFFERENCE
SEDIMENT 1	1.01440	1.56070	0.32833	2.13185	5	0.54630
SEDIMENT 2	1.50917	1.73872	0.05797	2.13185	5	0.22955
SEDIMENT 3	2.34683	2.68284	0.12421	2.13185	5	0.33601

POWER TO DETECT % DECREASE IN C_{ss} BELOW
ACTION LEVEL OF 2 ug/g GIVEN N, MSE AND DF SHOWN ABOVE

% DECREASE BELOW ACTION LEVEL	DREDGED SEDIMENT C_{ss}	D	T VALUE FOR (1-BETA, DF)	POWER (1-BETA)
10	1.8	0.2	-0.55682	0.29268
20	1.6	0.4	0.63224	0.73192
30	1.4	0.6	1.82131	0.95634
40	1.2	0.8	3.01037	0.99585
50	1.0	1.0	4.19943	0.99966

D4.5 SAS Program Statements for Censored Data Methods

SAS statements are given for the censored data methods DL, DL/2, ZERO, UNIF, and LR. Appropriate censored data methods from Table D-12 should be applied to bioaccumulation data sets that contain nondetects, prior to running BIOACC.SAS or BIOACCSS.SAS. The revised concentration data set obtained from the selected censored data method may then be used as the input data set for BIOACC.SAS or BIOACCSS.SAS.

First, create a contaminant concentration data set as in BIOACC.SAS (note that some of the concentrations have been changed from BIOACC.SAS in order to illustrate the censored data methods):

```
LIBNAME Q 'C:\SAS';
DATA BIOACC;
  INPUT TRT REP CONC @@;
  CARDS;
1 1 -.06 1 2 -.06 1 3 -.06 1 4 -.06 1 5 .09
2 1 -.06 2 2 .19 2 3 .18 2 4 .33 2 5 .31
3 1 .24 3 2 .10 3 3 .13 3 4 .18 3 5 .30
4 1 .13 4 2 -.06 4 3 .17 4 4 -.06 4 5 2.2
;
```

The minus signs are a convenient way of indicating nondetects and do not imply negative concentrations. All SAS programs that follow assume that nondetects have been coded as negatives. In the data above, DL = 0.06. This example data set is 35% censored, has unequal variances that increase as the means increase, CV equal to 2.0, and is lognormally or nonnormally distributed. The variance and distribution characteristics were determined by applying several of the methods that follow, and then testing the data for equality of variances using Levene's Test, and for normality and lognormality of residuals using Shapiro-Wilk's Test. From Table D-12, one would select and apply either DL or DL/2, and then proceed with BIOACC.SAS, using log-transformed data or rankits as appropriate. If rankits are needed, the method UNIF could also be used.

D4.5.1 SAS Statements for DL, DL/2 and ZERO

Read the data set created above into a new set, assign the DL, and use the statement corresponding to the selected simple substitution method:

```
DATA Q.BIOACC;
  SET BIOACC;
  IF CONC<0 THEN DL=ABS(CONC);
  OCONC=CONC;
  IF CONC<0 THEN CONC=DL;      /* Include this statement if using DL */
  IF CONC<0 THEN CONC=DL/2;    /* Include this statement if using DL/2 */
  IF CONC<0 THEN CONC=0; /* Include this statement if using ZERO */
PROC PRINT LABEL;              /* Print the revised data set */
VAR TRT REP OCONC CONC DL;
LABEL TRT='TREATMENT GROUP'
      REP='REPLICATE'
      OCONC='ORIGINAL CONCENTRATION'
      CONC='REVISED CONCENTRATION'
      DL='DETECTION LIMIT';
TITLE 'Uncensoring Using Simple Substitution Methods';
```

D4.5.1.1 SAS Program Output for DL, DL/2, or ZERO

Uncensoring Using Simple Substitution Methods

OBS	TREATMENT GROUP	REPLICATE	ORIGINAL CONCEN- TRATION	REVISED CONCEN- TRATION (DL)	REVISED CONCEN- TRATION (DL/2)	REVISED CONCEN- TRATION (ZERO)	DETECTION LIMIT
1	1	1	-0.06	0.06	0.03	0.00	0.06
2	1	2	-0.06	0.06	0.03	0.00	0.06
3	1	3	-0.06	0.06	0.03	0.00	0.06
4	1	4	-0.06	0.06	0.03	0.00	0.06
5	1	5	0.09	0.09	0.09	0.09	.
6	2	1	-0.06	0.06	0.03	0.00	0.06
7	2	2	0.19	0.19	0.19	0.19	.
8	2	3	0.18	0.18	0.18	0.18	.
9	2	4	0.33	0.33	0.33	0.33	.
10	2	5	0.31	0.31	0.31	0.31	.
11	3	1	0.24	0.24	0.24	0.24	.
12	3	2	0.10	0.10	0.10	0.10	.
13	3	3	0.13	0.13	0.13	0.13	.
14	3	4	0.18	0.18	0.18	0.18	.
15	3	5	0.30	0.30	0.30	0.30	.
16	4	1	0.13	0.13	0.13	0.13	.
17	4	2	-0.06	0.06	0.03	0.00	0.06
18	4	3	0.17	0.17	0.17	0.17	.
19	4	4	-0.06	0.06	0.03	0.00	0.06
20	4	5	2.20	2.20	2.20	2.20	.

D4.5.2 SAS Statements for UNIF

Create a contaminant concentration data set as in the first step above. Now, define DL and count number of reps (NREP) and censored (NC) and uncensored observations (NUC) in each treatment.

```

DATA A;
  SET BIOACC;
  IF CONC<0 THEN DL=ABS(CONC);
  OCONC=CONC;
  IF CONC<0 THEN COUNT=1; ELSE COUNT=0;
PROC MEANS NOPRINT;
  BY TRT;
  VAR COUNT;
  OUTPUT OUT=B0 SUM=NC N=NREP;
DATA B; SET B0;
  NUC=NREP-NC;
  DROP _TYPE_ _FREQ_;

```

/* The following statements initialize a counter at the first observation of each treatment, and then implement the UNIF formula. If there is only one nondetect in a treatment, it is set equal to DL/2. */

```

DATA Q.BIOACC;
  MERGE A B; BY TRT;
  IF FIRST.TRT THEN I=1;
  IF CONC<0 THEN DO;
    CONC=DL*(I-1)/(NC-1);
    IF NC=1 THEN CONC=DL/2;
    I+1;
  END;
PROC PRINT LABEL; /* Print the revised data set */
  VAR TRT REP OCONC CONC DL;
  LABEL TRT='TREATMENT GROUP'
        REP='REPLICATE'
        OCONC='ORIGINAL CONCENTRATION'
        CONC='REVISED CONCENTRATION'
        DL='DETECTION LIMIT';
  TITLE 'Uncensoring Using UNIF';

```

D4.5.2.1 SAS Program Output for UNIF

Uncensoring Using UNIF

OBS	TREATMENT GROUP	REPLICATE	ORIGINAL CONCENTRATION	REVISED CONCENTRATION	DETECTION LIMIT
1	1	1	-0.06	0.00	0.06
2	1	2	-0.06	0.02	0.06
3	1	3	-0.06	0.04	0.06
4	1	4	-0.06	0.06	0.06
5	1	5	0.09	0.09	.
6	2	1	-0.06	0.03	0.06
7	2	2	0.19	0.19	.
8	2	3	0.18	0.18	.
9	2	4	0.33	0.33	.
10	2	5	0.31	0.31	.
11	3	1	0.24	0.24	.
12	3	2	0.10	0.10	.
13	3	3	0.13	0.13	.
14	3	4	0.18	0.18	.
15	3	5	0.30	0.30	.
16	4	1	0.13	0.13	.
17	4	2	-0.06	0.00	0.06
18	4	3	0.17	0.17	.
19	4	4	-0.06	0.06	0.06
20	4	5	2.20	2.20	.

D4.5.3 SAS Statements for LR

Create a contaminant concentration data set as in the methods above. Now, define DL and count number of reps (NREP) and censored (NC) and uncensored observations (NUC) in each treatment, same as for UNIF above.

```
DATA A;
  SET BIOACC;
  IF CONC<0 THEN DL=ABS(CONC);
  OCONC=CONC;
  IF CONC<0 THEN COUNT=1; ELSE COUNT=0;
PROC MEANS NOPRINT;
  BY TRT;
  VAR COUNT;
  OUTPUT OUT=B0 SUM=NC N=NREP;
DATA B; SET B0;
  NUC=NREP-NC;
  DROP _TYPE_ _FREQ_;

/* LR should not be used unless there are at least 3 uncensored observations in a
treatment. If a treatment has more than one nondetect, each nondetect must be assigned
a different value below the DL. When nondetects have been originally scored as
negative concentrations, this can be done easily by multiplying each negative concen-
tration by its rep number. */

DATA C;
  MERGE A B; BY TRT;
  IF NUC<3 THEN DELETE;
  IF CONC<0 THEN CONC=CONC*REP;

/* Assign normal scores (rankits) to all concentrations and store in variable RANKIT
*/

PROC RANK NORMAL=BLOM OUT=C1;
  BY TRT; VAR CONC; RANKS RANKIT;

/* Make a new data set including only above-DL observations. These will be used with
their rankits in the REG procedure to calculate regression parameters. */

DATA C2; SET C1;
  IF CONC<0 THEN DELETE;
  SLOPE=RANKIT;
  LOGCONC=LOG10(CONC);          /* Take logs of above-DL concentrations */

/* Regress logs of above-DL concentrations against their rankits and output the
regression parameters */

PROC REG NOPRINT OUTEST=D;
  BY TRT;
  MODEL LOGCONC=SLOPE;
```

```

/* Make a new data set of just the nondetects. Merge it with the set of regression
parameters. Then estimate log concentrations for the nondetects using the slope and
intercept from the regression model, and the previously calculated rankits of the
nondetects. Take the antilogs to obtain estimated concentrations for the nondetects.
One problem with the LR method is that regression estimates of concentrations for
nondetects may exceed the DL. In such cases the concentration should be set equal to
the DL. */

```

```

DATA C3; SET C1;
  IF CONC<0;
DATA D1;
  MERGE D C3; BY TRT;
  LOGCONC=INTERCEP+SLOPE*RANKIT;
  CONC=10**LOGCONC;
  IF CONC=. THEN DELETE;
  IF CONC>DL THEN CONC=DL;

```

```

/* Combine with above-DL observations. Sort the data and print. Note that the new
data set will not include any treatments having fewer than 3 above-DL observations. */

```

```

DATA Q.BIOACC;
  SET C2 D1;
PROC SORT; BY TRT REP;
PROC PRINT LABEL;
  VAR TRT REP OCONC LRCONC CONC DL;
  LABEL TRT='TREATMENT GROUP'
        REP='REPLICATE'
        OCONC='ORIGINAL CONCENTRATION'
        LRCONC='CONCENTRATION ESTIMATED BY LR'
        CONC='REVISED CONCENTRATION'
        DL='DETECTION LIMIT';
  TITLE 'Uncensoring with LR';

```

D4.5.3.1 SAS Program Output for LR

Uncensoring with LR

OBS	TREATMENT		ORIGINAL CONCENTRATION	CONCENTRATION		DETECTION LIMIT
	GROUP	REPLICATE		ESTIMATED BY LR	REVISED CONCENTRATION	
1	2	1	-0.06	0.13291	0.06000	0.06
2	2	2	0.19	.	0.19000	.
3	2	3	0.18	.	0.18000	.
4	2	4	0.33	.	0.33000	.
5	2	5	0.31	.	0.31000	.
6	3	1	0.24	.	0.24000	.
7	3	2	0.10	.	0.10000	.
8	3	3	0.13	.	0.13000	.
9	3	4	0.18	.	0.18000	.
10	3	5	0.30	.	0.30000	.
11	4	1	0.13	.	0.13000	.
12	4	2	-0.06	0.02662	0.02662	0.06
13	4	3	0.17	.	0.17000	.
14	4	4	-0.06	0.00490	0.00490	0.06
15	4	5	2.20	.	2.20000	.

D5.0 REFERENCES

- Allredge, J.R. 1987. Sample size for monitoring of toxic chemical sites. *Environ. Monit. Assess.* 9:143-154.
- Clarke, J.U. 1994. Evaluating methods for statistical analysis of less than detection limit data using simulated small samples. 2. General results. *Dredging'94, Proceedings of the Second International Conference on Dredging and Dredged Material Placement, Vol. 1.* E.C. McNair, Jr. (ed.), American Society of Civil Engineers, New York, pp. 747-755.
- Clarke, J.U. 1995a. Evaluating methods for statistical analysis of less than detection limit data using simulated small samples. 1. Study description. *Water Quality '94 Proceedings of the 10th Seminar, 15-18 February 1994, Savannah, Georgia.* Miscellaneous Paper W-95-1, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS, pp. 330-335
- Clarke, J.U. 1995b. Guidelines for statistical treatment of less than detection limit data in dredged sediment evaluations. *Environmental Effects of Dredging Technical Note EEDP-04-23*, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.
- Clarke, J.V. and D.L. Brandon. In Press. Applications guide for statistical analyses in dredged sediment evaluations. Miscellaneous Paper D-96-2, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.
- Clarke, J.U. and V.A. McFarland. 1991. Assessing bioaccumulation in aquatic organisms exposed to contaminated sediments. Miscellaneous Paper D-91-2, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS. 77 pp.
- Conover, W.J. 1980. *Practical Nonparametric Statistics.* 2nd Ed. John Wiley & Sons, New York, NY. 493 pp.
- Conover, W.J., M.E. Johnson and M.M. Johnson. 1981. A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics* 23:351-361.
- Dixon, W.J. and F.J. Massey, Jr. 1983. *Introduction to Statistical Analysis.* 4th Ed. McGraw-Hill Book Company, New York, NY. 678 pp.
- Draper, N.R. and H. Smith. 1981. *Applied Regression Analysis.* 2nd Ed. John Wiley and Sons, Inc., New York, NY. 709 pp.
- Dunnett, C.W. 1955. Multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Assoc.* 50:1096-1121.
- Gelber, R.D., P.T. Lavin, C.R. Mehta and D.A. Schoenfeld. 1985. Chapter 5: Statistical analysis. In G.M. Rand and S.R. Petrocelli (Eds.), *Fundamentals of Aquatic Toxicology: Methods and Applications.* Hemisphere Publishing Corp., Washington, DC, pp. 110-123.
- Gill, J.L. 1978. *Design and Analysis of Experiments in the Animal and Medical Sciences.* Vol. 3. Appendices. The Iowa State University Press, Ames, IA. 173 pp.
- Gleit, A. 1985. Estimation for small normal data sets with detection limits. *Environ. Sci. Technol.* 19:1201-1206.
- Hamilton, M.A., R.C. Russo and R.V. Thurston. 1977. Trimmed Spearman-Kärber method for estimating median lethal concentrations in toxicity bioassays. *Environ. Sci. Technol.* 11:714-719.
-

-
- Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* 54:187-211.
- Kleinbaum, D.G. and L.L. Kupper. 1978. *Applied Regression Analysis and Other Multivariable Methods*. Duxbury Press, North Scituate, MA. 556 pp.
- Lilliefors, H.W. 1967. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *J. Amer. Statist. Assoc.* 62:399-402.
- McClave, J.T. and F.H. Dietrich, II. 1979. *Statistics*. Dellen Publishing Company, San Francisco, CA. 681 pp.
- McFarland, V.A. and J.U. Clarke. 1987. Simplified approach for evaluating bioavailability of neutral organic chemicals in sediment. *Environmental Effects of Dredging Technical Note EEDP-01-8*, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.
- Newman, M.C. and P.M. Dixon. 1990. UNCENSOR: A program to estimate means and standard deviations for data sets with below detection limit observations. *Amer. Environ. Lab.* April 1990:26-30.
- Porter, P.S., R.C. Ward and H.F. Bell. 1988. The detection limit. *Environ. Sci. Technol.* 22:856-861.
- Rohlf, F.J. and R.R. Sokal. 1969. *Statistical Tables*. W.H. Freeman and Company, New York, NY. 253 pp.
- SAS Institute, Inc. 1988a. *SAS® Language Guide for Personal Computers*, Release 6.03 Edition. SAS Institute, Inc., Cary, NC. 558 pp.
- SAS Institute, Inc. 1988b. *SAS/STAT™ User's Guide*, Release 6.03 Edition. SAS Institute, Inc., Cary, NC. 1028 pp.
- SAS Institute, Inc. 1988c. *SAS/GRAPH® User's Guide*, Release 6.03 Edition. SAS Institute, Inc., Cary, NC. 549 pp.
- Satterthwaite, F.W. 1946. An approximate distribution of estimates of variance components. *Biom. Bull.* 2:110-114.
- SCI. 1989. *PCNONLIN Version 3.0*. Statistical Consultants, Inc., 300 E. Main Street, Quality Place, Suite 400, Lexington, KY 40507-1539.
- Shapiro, S.S. and M.B. Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52:591-611.
- Snedecor, G.W. and G.C. Cochran. 1989. *Statistical Methods*. 8th Ed. The Iowa State University Press, Ames, IA. 507 pp.
- Sokal, R.R. and F.J. Rohlf. 1981. *Biometry*, 2nd Edition. W.H. Freeman and Company, San Francisco, CA. 859 pp.
- Steel, R.G.D. and J.H. Torrie. 1980. *Principles and Procedures of Statistics*, 2nd Edition. McGraw-Hill Book Company, New York, NY. 633 pp.
- Steinberg, D. 1988. *PROBIT: A Supplementary Module for SYSTAT and SYGRAPH*. SYSTAT Inc., Evanston, IL. 16 pp.
-

- Steinberg, D. and P. Colla. 1991. LOGIT: A Supplementary Module for SYSTAT. SYSTAT Inc., Evanston, IL. 225 pp.
- Stephan, C.E. 1977. Methods for Calculating an LC_{50} . In F.L. Mayer and J.A. Hamelink (Eds.), Aquatic Toxicology and Hazard Evaluation. ASTM Spec. Tech. Publ. 634. American Society for Testing and Materials, Philadelphia, PA, pp. 65-84.
- USEPA. 1985. Methods for measuring the acute toxicity of effluents to freshwater and marine organisms. 3rd Ed. Environmental Protection Agency, Environmental Monitoring and Support Laboratory, Cincinnati, OH. EPA-600/4-85/013.
- USEPA. 1989. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to aquatic organisms. 2nd Ed. U.S. Environmental Protection Agency Rept. No. EPA/600/4-89/001. 262 pp.
- Wilkinson, L. 1990. SYSTAT: The System for Statistics. SYSTAT Inc., Evanston, IL. 677 pp.
- Winer, B.J. 1971. Statistical Principles in Experimental Design. 2nd Ed. McGraw-Hill Book Company, New York, NY. 907 pp.
- Zar, J.H. 1984. Biostatistical Analysis. 2nd Ed. Prentice-Hall, Inc., Englewood Cliffs, NJ. 717 pp.
-