

*November 1998*  
*Evaluation Framework*

# **Appendix 9A**

## **BIOASTAT SOFTWARE**

**DMMP CLARIFICATION PAPER  
SMS TECHNICAL INFORMATION MEMORANDUM**

**BIOSTAT SOFTWARE FOR THE ANALYSIS OF DMMP/SMS  
BIOASSAY DATA**

Prepared by David F. Fox (U.S. Army Corps of Engineers), David A. Gustafson (U.S. Army Corps of Engineers) and Travis C. Shaw (U.S. Army Corps of Engineers) for the DMMP/SMS agencies.

**INTRODUCTION**

Biological testing can be used to determine both the toxicity of sediments to a suite of organisms and the bioavailability of chemicals for uptake and storage. In both cases, the experimental results are statistically analyzed to determine whether there is a significant difference between test and reference samples. Thus, statistical analysis plays a critical role in the interpretation of bioassay results and in regulatory determinations made with regard to the sediment. It is important that the statistical procedures used be technically sound, consistently applied and provide reproducible results by regulators, bioassay practitioners and consultants alike.

**PROBLEM IDENTIFICATION**

Statistical procedures for the DMMP were first established in the PSDDA Management Plan Report Phase II (PSDDA 1989, page 5-25). Modifications of these procedures have been made twice by the regulatory agencies via the annual review process. In 1994, the experimental significance level for the larval test was increased to 0.10 and the use of power analysis was established for that bioassay (Fox & Littleton, 1994). In 1996, use of the Shapiro-Wilk test for normality was incorporated into the statistical procedures (Michelsen & Shaw, 1996). An additional modification, replacement of Cochran's test with Levene's test for equality of variances, was proposed but not formally adopted during the 1997 annual review process (Shaw & Fox, 1997).

The modifications, adopted and proposed, provide a statistically more rigorous treatment of bioassay data but also increase the complexity of the analysis. When Levene's test was proposed for adoption as the standard test for equality of variances, concern was expressed that the statistics were becoming more complicated than agency staff and most consultants could readily handle. It was therefore proposed (Shaw & Fox, 1997) that statistical software be developed to facilitate bioassay data analysis. This software would incorporate Levene's test, as well as the other modifications made earlier.

The primary purpose of this paper is to introduce the software that was promised in 1997 and to describe the statistical procedures used by the program. Secondly, where modifications have

been made to statistical procedures currently used in the DMMP and SMS programs, technical justification is provided.

## TECHNICAL DISCUSSION

*Student's t-test and underlying assumptions.* Interpretation of DMMP and SMS bioassays includes a statistical comparison between test and reference sediment data. The basic statistic used in this analysis is the student's t-test. However, use of the t-test is based on the assumption that test and reference samples have been taken from a normally distributed population and have equal variances. The consequences of violating these assumptions include loss of confidence in the type I error rate and a decrease in statistical power. Violations of the assumptions can be addressed through the use of data transformation or the application of alternate statistical procedures.

*Development of the BioStat Software.* Seattle District developed BioStat to automate the testing of statistical assumptions and to perform the comparison test between experimental treatments that best matches the outcome of the assumptions tests. BioStat also provides the ability to do data transformations prior to the statistical analysis. Figure 1 is a flow diagram which depicts the basic statistical logic and procedures incorporated into the software. Details are provided in the BioStat Users Guide (Fox *et al.*, 1998) and the following sections of this paper.

*Test for Normality.* The test and reference data must be evaluated to determine whether or not they have been taken from a normally distributed population. As indicated in Michelsen and Shaw (1996) and EPA/USACE (1994), the recommended test for normality is the Shapiro-Wilk W-statistic (Shapiro and Wilk, 1965).

*Test for Equality of Variance.* The statistical clarification paper presented at the 1996 Sediment Management Annual Review Meeting recommended the use of Cochran's test to evaluate equality of variance (Michelsen and Shaw, 1996). Subsequent to presentation of that paper, simulations conducted at the Corps of Engineers Waterways Experiment Station (WES) revealed that Cochran's test may have very high Type I error rates when the data set has an asymmetric non-normal distribution (Clarke and Brandon, 1995).

In its work, WES determined that Levene's test outperforms all the commonly used tests for equality of variance. Levene's test is performed by conducting an analysis of variance on the absolute deviations of treatment observations from the treatment means (Levene, 1960). The analysis of variance simplifies to a t-test when a single test treatment is being compared to a single reference treatment, which is the case in the interpretation of DMMP and SMS bioassays.

The first step in conducting Levene's test is to transform the data set into absolute deviations from the mean in each of the two treatment groups. The transformed scores are then tested using a two-tailed t-test. If the results are significant, then the conclusion is that heterogeneity of variances exists and a key assumption of the Student's t-test (for the comparison of the bioassay

endpoint data) is violated by the data set. The data set must then be transformed (e.g. arcsine-square root) or the approximate t-test used.

User-selected Data Transformations. In cases where at least one of the distribution assumptions is violated, a simple transformation may allow both assumptions to be met and the t-test employed. BioStat includes three common data transformations that are user-selectable:

- 1) arcsine square root =  $\sin^{-1} \sqrt{x}$
- 2) square root =  $\sqrt{x+.375}$
- 3) log =  $\log_{10}(x + 1)$

The arcsine square root transformation is used with percentage data and is the most commonly used transformation for DMMP and SMS bioassays. The square root transformation is used when the variances are proportional to the means (Zar 1984, p. 241). The logarithmic transformation is sometimes useful in the analysis of growth data (Sokal and Rohlf, 1969).

Rank Transformation. In the event that none of these transformations can establish normality or homoscedasticity, BioStat automatically transforms the data to rankits (Sokal and Rohlf, 1969). Rank transformation normalizes the distribution, permitting the transformed data to be evaluated using a t-test (Conover and Iman, 1981).

Statistical Comparison of Treatment Means. Depending on the outcome of the tests for normality and equality of variance, BioStat uses the following statistical tests to compare treatment means:

Outcome of W-test	Outcome of Levene's test	Statistic used to compare treatment means	References
normal distribution	equal variances	student's t-test	Sokal & Rohlf 1969, p. 220
normal distribution	unequal variances	approximate t-test	Zar 1984, p.131
non-normal distribution	equal variances	Mann-Whitney	Sokal & Rohlf 1969, p. 393 Zar 1984, p. 139 Potvin & Roff, 1993
non-normal distribution	unequal variances	t-test on rankits	Sokal & Rohlf 1969, p. 121 Conover & Iman, 1981

One-sample t-test. There are two cases where a one-sample t-test would be used. The evaluation of bioaccumulation data sometimes includes a statistical comparison of replicate test data to a numerical standard, such as a Food and Drug Administration Action Level. The standard is not an experimental treatment and does not have replicate data, therefore a one-sample t-test must be run (EPA/USACE, 1994, page D-43 and Zar, 1984, page 102).

A second case in which BioStat uses the one-sample t-test is one in which there is no variance in the reference treatment replicates. This is an uncommon occurrence but is possible if, for example, the amphipod test is run and there is zero mortality in all of the reference treatment replicates. In this case, BioStat automatically applies the one-sample t-test.

Power Analysis. Power analysis procedures have been incorporated into BioStat for all three forms of the t-test. Technical guidance for this portion of the software came from Dixon and Massey (1957, Chapter 14), supplemented by Cohen (1988, Chapter 12) and is fully documented in the BioStat Users Guide (Fox *et al.*, 1998).

## **PROPOSED MODIFICATION.**

The BioStat software provides for a statistically rigorous treatment of bioassay data and will be used in the future by the DMMP and SMS agencies to compare test and reference treatment data. Concomitant to the implementation of BioStat, the agencies officially adopt Levene's test to assess equality of variance rather than Cochran's test.

BioStat can be downloaded from Seattle District's FTP (file transfer protocol) server. For instructions, contact David Fox at *david.f.fox@usace.army.mil*.

## **Cited Literature**

Clarke, J.U. and D.L. Brandon, 1995, *Applications Guide for Statistical Analyses in Dredged Sediment Evaluations*, U.S. Army Corps of Engineers, Waterways Experimental Station, Vicksburg, MS.

Cohen, J., 1988, *Statistical Power Analysis for the Behavioral Sciences, 2nd Ed.*, Lawrence Erlbaum Associates, Publishers, Hillsdale, NJ.

Conover, W.J. and R.L. Iman, 1981, Rank transformations as a bridge between parametric and nonparametric statistics, *The American Statistician*, **35**, 124-133.

Dixon, W. F. and F.J. Massey, 1957, *Introduction to Statistical Analysis, 2nd Ed.*, McGraw-Hill, New York.

EPA/USACE, 1994, *Evaluation of Dredged Material Proposed for Discharge in Waters of the U.S. - Testing Manual (Draft)*, prepared by the Environmental Protection Agency Office of Science and Technology and the Department of the Army Corps of Engineers, Washington, D.C.

Public Review Draft  
April 1998

Fox, D.F. and T.M. Littleton, 1994, *Interim Revised Performance Standards for the Sediment Larval Bioassay*, In: Puget Sound Dredged Disposal Analysis Sixth Annual Review Meeting Minutes.

Fox, D.F., D.A. Gustafson and T.C. Shaw, 1998, *BioStat Users Guide*, U.S. Army Corps of Engineers, Seattle District.

Levene, H., 1960, Robust tests for equality of variances, In: I. Olkin (Ed.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, 278-292, Stanford University Press, Stanford, CA.

Michelsen, T. and T.C. Shaw, 1996, *Statistical Evaluation of Bioassay Results*, In: Puget Sound Dredged Disposal Analysis Eighth Annual Review Meeting Minutes,

PSDDA, 1989, Management Plan Report - Unconfined Open-Water Disposal of Dredged Material, Phase II, U.S. Army Corps of Engineers, Seattle District; U.S. Environmental Protection Agency, Region X; Washington State Department of Natural Resources; Washington State Department of Ecology.

Potvin, C. and D.A. Roff, 1993, Distribution-free and robust statistical methods: viable alternatives to parametric statistics?, *Ecology*, **74**, 1617-1628.

Shaw, T.C. and D.F. Fox, 1997, *Refinements to DMMP/SMS Bioassay Statistics (DMMP/SMS Status Report)*, In: mailout preceding the Sediment Management Annual Review Meeting, May 7, 1997.

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

Sokal R.R. and F.J. Rohlf, 1969, *Biometry*, W.H. Freeman and Company, San Francisco.

Zar, J.H., 1984, *Biostatistical Analysis, 2nd Ed.*, Simon & Schuster Co., Englewood Cliffs, NJ.