Statistical analysis of bioassay data: Comparison of BioStat and EDAT

Prepared by Britta Voss (Ecology), Laura Inouye (Ecology), Brian Hester (USACE), Whitney Conard (EPA)

Introduction

Bioassay testing is a key component of the regulatory framework used to determine the toxicity of sediments at cleanup and dredging sites in the Pacific Northwest. In 1998, the U.S. Army Corps of Engineers (USACE) released the BioStat software program to provide the SMARM community with a consistent, defensible bioassay data analysis tool (Fox et al., 2007). The Department of Ecology's Toxics Cleanup Program also developed a web-based analysis tool called MyEIM (Ecology, 2015). Originally released in 2007, MyEIM provided analysis of both chemical and bioassay data stored in Ecology's Environmental Information Management (EIM) database. In 2024, Ecology released a redeveloped version of MyEIM called the EIM Data Analysis Tool (EDAT), which recreated the bioassay data analysis features of MyEIM as well as improved data presentation and mapping.

During the development of EDAT, subtle differences were identified between the calculations performed by BioStat and EDAT, which can in some cases result in different outcomes for the same dataset. The purpose of this paper is to document these differences and provide the SMARM community the opportunity to consider whether any revisions are needed for EDAT.

Problem Identification

Bioassay data consists of a test sample and a paired reference or negative control sample, each of which consists of multiple replicates. These paired samples are compared using both a numeric criteria and a statistical criteria to determine if the test sample exhibits greater toxic effects on the test organisms than the reference/control (DMMP, 2021; Ecology, 2021). BioStat does not evaluate numeric criteria and this paper does not address numeric criteria.

The statistical comparison is a one-sided test, meaning the null hypothesis[1] is expressed as $x_1 \leq x_2$ or $x_1 \geq x_2$ (with x_1 representing the test sample and x_2 the reference or negative control sample). Bioassay data can be in the form of proportion data (survival or mortality that has bounds between 0 and 100%), growth data (weight or lengths that have no upper limit), or absolute count (number of individuals).

The basic steps of bioassay statistical analysis, which are consistent between BioStat and EDAT, are:

1. Perform Shapiro-Wilk test on pooled test and reference/control replicates to determine if the data are normally distributed.
2. Perform Levene's test on pooled test and reference/control replicates to determine if the variances are homogeneous ("homoscedastic").
3. Based on the outcome of these tests, perform a statistical test to determine if the null hypothesis is supported.

---

[1] The null hypothesis is that the test sample does not produce a significant change in biological response (mortality, growth, etc.) compared to the reference/control sample.

a. Normal, Homoscedastic: Student's T-Test
b. Normal, Not Homoscedastic: Approximate T-Test
c. Not Normal, Homoscedastic: Mann-Whitney Test
d. Not Normal, Not Homoscedastic: Rankit transformation, followed by either Student's T-Test or Approximate T-Test

In general, parametric statistical tests (i.e. T-Tests, which are only appropriate for normally distributed data) are more powerful than non-parametric tests. Transformation formulas can be used to bring a non-parametric dataset closer to a normal distribution. It is the nuances of these transformations where BioStat and EDAT have subtle differences. In the discussion below, an example dataset is provided to demonstrate the impact of each difference.

While not necessary to examine the questions posed in this paper, it is worth noting the differences in how BioStat and EDAT approach transformation selections. In BioStat, the user chooses a transformation and must make other selections based on their analysis needs (null hypothesis, alpha level, minimum detectable difference, etc.). In EDAT, the user selects a criteria from a list of available SMS or DMMP criteria, and the selected criteria determines which transformations are applied to the data; users can also adjust these criteria properties if needed.

Difference #1: Radians versus degrees for arcsin(sqrt(x)) transformation.

When the arcsin(sqrt(x)) transformation is performed, the output can be conveyed in either radians or degrees. In Microsoft Excel, the result is conveyed in radians. In BioStat, the result of this transformation is converted to degrees while in EDAT the result is left in radians. The conversion from radians to degrees can result in small differences in results. If the calculated values are very close to the decision threshold (critical value) for the statistical test, it is possible to obtain a different final result. In EDAT, this transformation is used for mortality, Microtox luminosity, and fertilization criteria.

Example dataset

Replicate	Test Sample (x_1)	arcsin(sqrt(x_1)) (radians)	arcsin(sqrt(x_1)) (degrees)
1	0	0	0
2	0.25	0.5236	30
3	0.05	0.2255	12.92
4	0.05	0.2255	12.92
5	0.2	0.4636	26.56

Replicate	Reference Sample (x_2)	arcsin(sqrt(x_2)) (radians)	arcsin(sqrt(x_2)) (degrees)
1	0	0	0
2	0.05	0.2255	12.92
3	0.05	0.2255	12.92
4	0.05	0.2255	12.92
5	0	0	0

Null hypothesis: $x_1 \leq x_2$

Alpha: 0.05

Shapiro-Wilk Test: Normal

Levene's Test: Homoscedastic

Student's T-Test Critical Value: 1.86

- EDAT calculated value (radians): 1.3961
- BioStat calculated value (degrees): 1.3766

Result: The calculated values for the Student's T-Test differ slightly depending on whether the values in radians or degrees are used. However, both values are below the critical value, so the final result is the same.

Difference #2: Applying normality and homoscedasticity tests to untransformed versus transformed values.

In BioStat, the normality and homoscedasticity tests are performed on untransformed data, even if the user selects a transformation. In EDAT, the normality and homoscedasticity tests are performed on the data after the transformation has been applied.

Example dataset

Replicate	Test Sample (x_1)	arcsin(sqrt(x_1)) (radians)
1	0.15	0.3977
2	0.15	0.3977
3	0	0
4	0.15	0.3977
5	0	0

Replicate	Reference Sample (x_2)	arcsin(sqrt(x_2)) (radians)
1	0	0
2	0.05	0.2255
3	0.05	0.2255
4	0.05	0.2255
5	0	0

Null hypothesis: $x_1 \leq x_2$

Alpha: 0.05

Shapiro-Wilk Test:

- BioStat result (using untransformed values): Normal
- EDAT result (using transformed values): Not Normal

Levene's Test:

- BioStat result (using untransformed values): Not Homoscedastic
- EDAT result (using transformed values): Not Homoscedastic

Result: Using the untransformed values, the Shapiro-Wilk Test determined that the data are normally distributed; however, using the transformed values, the Shapiro-Wilk Test determined that the data are not normally distributed. This will result in a different statistical test being used to compare the samples.

Difference #3: Whether or not to apply a transformation when non-parametric statistical tests are used.

In BioStat, if the Shapiro-Wilk Test determines that the data are not normally distributed, a transformation selected by the user will not be applied to the data. In EDAT, the selected transformation is always applied, regardless of the outcome of the Shapiro-Wilk test.

Example dataset

Replicate	Test Sample (x_1)	arcsin(sqrt(x_1)) (radians)
1	0	0
2	0.1	0.3218
3	0.05	0.2255
4	0.1	0.3218
5	0.1	0.3218

Replicate	Reference Sample (x_2)	arcsin(sqrt(x_2)) (radians)
1	0.05	0.2255
2	0.05	0.2255
3	0.05	0.2255
4	0	0
5	0	0

Null hypothesis: $x_1 \leq x_2$

Alpha: 0.05

Shapiro-Wilk Test: Not Normal

Levene's Test: Homoscedastic

Mann-Whitney Test Critical Value: 21

- BioStat calculated value (using untransformed values): 19.5
- EDAT calculated value (using transformed values): 19.5

Result: EDAT applied the selected arcsin(sqrt(x)) transformation while BioStat overrode the user selection and did not apply a transformation before performing the Mann-Whitney Test. Both programs used the same critical value for the statistical test and calculated the same test statistic for the dataset. So although they differed in how the transformation was applied, the programs generated identical results.

Difference #4: Transforming data with $\log_{10}(x)$ versus $\log_{10}(x+1)$.

BioStat uses a logarithmic transformation of $\log_{10}(x+1)$. EDAT uses $\log_{10}(x)$. According to Zar (1984), the $\log_{10}(x+1)$ transformation used by BioStat "is preferred on theoretical grounds and is especially preferable when some of the observed values are small numbers (particularly zero)." In EDAT, this transformation is used for freshwater growth criteria. There are no DMMP criteria that use this transformation, but it is available in BioStat.

Example dataset

Replicate	Test Sample (x_1)	$\log_{10}(x_1)$	$\log_{10}(x_1+1)$
1	0	N/A	0
2	0	N/A	0
3	0.12	-0.9208	0.04922
4	2.04	0.3096	0.4829
5	0	N/A	0
6	0	N/A	0
7	0	N/A	0
8	0	N/A	0

Replicate	Control Sample (x_2)	$\log_{10}(x_2)$	$\log_{10}(x_2+1)$
1	1.61	0.2068	0.4166
2	1.84	0.2648	0.4533
3	1.23	0.08991	0.3483
4	1.42	0.1523	0.3838
5	1.31	0.1173	0.3636
6	1.50	0.1761	0.3979
7	1.35	0.1303	0.3711
8	1.89	0.2765	0.4609

Null hypothesis: $x_1 \geq x_2$

Alpha: 0.05

Shapiro-Wilk Test:

- BioStat result (using $\log_{10}(x)+1$ transformation): Not Normal
- EDAT result (using $\log_{10}(x)$ transformation): Not Normal

Levene's Test:

- BioStat result (using $\log_{10}(x+1)$ transformation): Homoscedastic
- EDAT result (using $\log_{10}(x)$ transformation): Not Homoscedastic

Statistical Test Result:

- BioStat result (using $\log_{10}(x+1)$ transformation): Hit
- EDAT result (using $\log_{10}(x)$ transformation): Pass

Result: This example dataset demonstrates how, for small numbers (i.e. final growth values < 1 mg), the $\log_{10}(x)$ transformation used by EDAT can result in invalid values, thus reducing the number of replicates available to evaluate the samples. This can result in a different outcome for the statistical tests, and potentially for the final criteria result. In a survey of freshwater growth data in EIM (1,147 total samples), using the BioStat version of this transformation resulted in selecting a different statistical test in about 14% of samples, and a different statistical test result in about 2% of samples.

Proposed Issue Solution and Rationale

Based on this evaluation, we do not propose making changes to either the EDAT or BioStat software programs at this time. The differences identified are subtle and do not affect the results of the bioassay statistical evaluation in most cases.

We recommend that the sediment cleanup and dredging community continue using the software program applicable to each site: for sites managed under SMS, use EDAT; for sites managed under DMMP, use BioStat.

We encourage interested users to analyze samples in both programs and share with the authors any differences they observe. If, in the future, there is credible evidence to support changing the bioassay statistical calculations in EDAT, we may consider making modifications.

If these programs are considered sufficiently equivalent, EDAT could provide an alternative to BioStat for DMMP users. Any changes to DMMP requirements would be announced through the SMARM process.

References

DMMP [Dredged Material Management Program]. 2021. Dredged Material Evaluation and Disposal Procedures User Manual. Seattle WA: U.S. Army Corps of Engineers, U.S. Environmental Protection Agency Region 10, Washington State Department of Natural Resources, Washington State Department of Ecology. https://www.nws.usace.army.mil/Missions/Civil-Works/Dredging/User-Manual/.

Ecology [Washington State Department of Ecology]. 2015. MyEIM User Manual. Olympia WA: Washington State Department of Ecology, Toxics Cleanup Program.

Ecology [Washington State Department of Ecology]. 2021. Sediment Cleanup User's Manual (SCUM). Publication Number 12-09-057. Olympia WA: Washington State Department of Ecology, Toxics Cleanup Program. https://apps.ecology.wa.gov/publications/SummaryPages/1209057.html.

Fox, D. F., D. A. Gustafson, and T. C. Shaw. 2007. BioStat 2.0 User's Guide. Seattle District Corps of Engineers.

Zar, J. H. 1984. *Biostatistical Analysis*. 2nd ed. Prentice Hall. Englewood Cliffs, NJ.