Practical Stats Newsletter for September 2005

In this newsletter:

1. Courses for 2006

Two of our courses are scheduled on the west coast for Spring 2006. Applied Environmental Statistics, our week-long survey of all the basics, will be taught in Portland OR at the Mark Spencer Hotel on March 27-31. Registration is $1495, which covers all course materials, digital textbook, and data. Links for registration can be found at
http://www.practicalstats.com/Pages/aes.html

Nondetects And Data Analysis, the 2-day course on handling nondetect data using modern methods of survival analysis, is scheduled to be taught May 4-5 in San Jose, CA. The exact location has still not been resolved, so check the Practical Stats web page for more details soon.

Course content for each course is outlined on our web page. For any other information, email us at ask[atsign]practicalstats.com.

2. Hypothesis Testing With Nondetects

Hypothesis tests when data include nondetect values can appear problematic, especially when multiple detection limits occur. A traditional solution has been to substitute some value for nondetects, such as one-half the detection limit. This practice works poorly, often placing a signal into the data that was not there initially. The result - wrong test results. There are better ways.

Three general approaches can be taken to test data with nondetects. All three will give valid answers, unlike substitution. The choice of which approach to take depends on whether you must stick to tests more traditional within the environmental community, and whether there are multiple detection limits.

Approach A. Tests for categorical data

Data can be re-coded into two categories, below and above the detection limit, and tests performed for dichotomous (2-category) data. The tests discern whether the proportion of detects is identical for each group, or whether some differences occur. With multiple detection limits the data must be coded as below or above

the highest limit. These are standard tests familiar to many scientists. To compare among groups analogous to ANOVA and t-tests, use Contingency Tables (also called the Test of Proportions). For regression, use logistic regression.

Approach B. Familiar nonparametric tests

Nonparametric tests are performed by first ranking the data from lowest to highest. The lowest value receives a rank of 1 and the highest n, where n is the number of data or data pairs. Tests are then computed using these ranks, rather than the original values. When comparing a <1 to a 6, for example, the nondetect receives a lower rank. Therefore the information in nondetects can be efficiently captured by using rank-based tests. These tests are familiar to most environmental scientists: the rank-sum (Mann-Whitney) test to compare two groups, the Kruskal-Wallis nonparametric ANOVA, and Kendall's tau or Spearman's rho correlation coefficients. All of these tests can be computed without substitution of fabricated values, and therefore (unlike substitution) their results are believable. These tests are most applicable for data with one detection limit, where they can be applied with no special adjustments or considerations.

These familiar tests can also be applied to data with multiple detection limits, though there likely will be some loss of power. To run them, the data must first be re-coded so that all values below the highest limit (DLmax) are coded as <DLmax, and then the test is performed. If the highest limit were 10, for example, all <1s are coded as <10, as well as any <5s and all detected values of 2, 3, etc. up to just below 10. Then the test is computed. Though this approach will find differences in fewer cases than with Approach 3, it is a simple procedure requiring little explanation, and if differences are found, is sufficiently powerful to find them!

Approach C. Methods for survival analysis

Survival analysis methods are much less familiar to environmental scientists, and their application to environmental data is the central theme of the book Nondetects And Data Analysis (Helsel, 2005, published by Wiley). Their primary use is to test data with multiple detection limits. These tests handle multiple limits without re-coding values, or substituting fabricated numbers. Plus, there are both parametric and nonparametric tests within this approach.

The parametric tests use maximum likelihood estimation (MLE) to solve for parameters such as regression slopes or differences between means. ANOVA, t-tests and regression all have their analogues with MLE methods. As with those more familiar tests, MLE methods require first

designating a distribution such as the normal or lognormal to approximately fit the shape of the data. With a good fit, p-values and coefficients of these tests are accurate. In addition, MLE tests are the only way to validly perform parametric tests on data with one detection limit.

Nonparametric tests with multiple detection limits go by the name of "score tests". Scores are modified ranks or percentiles, usually computed only for detected values but adjusted for the proportion of nondetects above and below each detect. Tests between groups determine if the cumulative distribution functions, the sets of percentiles or scores, are the same or different for each group. Score tests extend the rank-sum and Kruskal-Wallis tests to data with multiple detection limits. Kendall's tau correlation coefficient can be computed for data with multiple detection limits, as can adaptations of the Kendall-Theil-Sen-Mann median line and test for trend.

While methods under this approach are more complex and less familiar to environmental scientists than the other two approaches, their increased power makes them an important tool for the ever more frequent situation of data with multiple detection limits. We'll discuss tests using this third approach in more detail in upcoming newsletters. Or get a copy of the textbook Nondetects And Data Analysis
http://www.practicalstats.com/nada
and get all the details now.

3. 15 years later, its now More Than Obvious
In 1990, Environmental Science and Technology published a review article on handling nondetects titled "Less Than Obvious". Cited in numerous publications, it was the precursor to the 2005 textbook "Nondetects And Data Analysis". Now in mid-October, ES&T will publish an updated review on this issue entitled "More Than Obvious". Read it to see how and why better methods like those above should be a part of your data analysis toolkit.

'Til next time,

Practical Stats
http://www.practicalstats.com

-- Make sense of your data