Practical Stats Newsletter   for Winter, 2004

In this newsletter:
1. Nondetects And Data Analysis textbook
2. Scatterplots with Nondetects
3. Newsletter archive on the web
4. Upcoming Courses

1. Nondetects And Data Analysis textbook
Wondering why you haven't been getting your Practical Stats newsletter?
Well, it could be due to overactive spam filters (see #3 below), but part
of the fault is mine. This newsletter is a month late. All of my spare
time (what's that?) has been devoted to putting the finishing touches on my
upcoming book, Nondetects And Data Analysis, to be published by Wiley this
summer. NADA has been an idea of mine for over a decade, ever since my
"Less Than Obvious" article appeared in ES&T. To preview the book I've
enclosed an annotated outline below. The past two newsletters, today's
item #2, and the main topics for several upcoming newsletters, all contain
items from the text.

Nondetects And Data Analysis   by Dennis Helsel   Wiley, 2004   270
pages

Introduction
The introduction finds a parallel between the missteps in interpreting data
that led up to the Challenger space-shuttle accident in 1986, and today's
inability to interpret environmental data. Several quotes from the Rogers
Commission report on the Challenger accident are included.

Chapter 1. In Focus
A typical analysis comparing two groups of contaminant data, at an
upgradient 'control' site and at a downgradient site possibly contaminated
by an organic compound, illustrates why deleting less-thans, or
substituting values such as one-half the detection limit, lead to bad
decisions.

Chapter 2. The Three Approaches
The three approaches contrasted throughout the book are described in
detail. These are: substitution of some fraction of the detection limit,
maximum likelihood methods, and nonparametric methods. How each works, and
the assumptions behind each, are illustrated.

Chapter 3. Reporting Limits
Similarities and differences in the calculations of detection limits,
reporting limits, limits of quantitation, and other types of censoring
thresholds are given. How better data analyses are done by understanding

the differences.

Chapter 4. Storing Data with Nondetects
A short chapter describing the three most common methods to represent censored data in databases, and for use by statistical software.

Chapter 5. Plotting Data With Nondetects
Graphs, including scatterplots, boxplots and the survival function plot, for describing and illustrating one or more groups of data with nondetects.

Chapter 6. Computing Summary Statistics
The topic for censored data most discussed in the environmental literature. How can the mean, standard deviation, median, and other percentiles be computed when nondetects are present? What are the differences, strengths and weaknesses, of the variety of methods that have been proposed?

Chapter 7. Computing Interval Estimates
How to compute confidence, prediction and tolerance intervals for censored data.

Chapter 8. What Can Be Done When All Data Are Below the Reporting Limit?
There is actually a great deal of information in the knowledge that all measurements are below the limit of detection. What questions can be answered when this occurs?

Chapter 9. Comparing Two Groups
The most common question in a designed study is comparing background to a possibly contaminated site. Why substitution followed by a t-test does not work well. The two primary tests for this situation are unknown to almost all environmental scientists. A description is given of how they are computed.

Chapter 10. Comparing Three or More Groups
Extension of two-groups methods to three or more groups. How multiple comparison tests can be conducted for data with nondetects.

Chapter 11. Correlation
Correlation coefficients when one or both variables are censored. Computing parametric and nonparametric coefficients (including Kendall's tau) for data with nondetects.

Chapter 12. Regression and Trends
Several methods are available for fitting a predictive curve to censored data, including MLE regression. logistic regression, and the nonparametric Thiel-Sen estimator. Applications to trend analysis.
Appendices

A: Adding data with nondetects to get a sum (estimation of mass or load);
B: A listing of data sets and software used in the book and available on
the web;

References: articles and books on this topic from the fields of statistics,
astronomy, economics, and the variety of environmental disciplines
including soils, air and water quality. The most comprehensive listing
found in one location.

2.   Scatterplots With Nondetects
How should data be plotted on an X-Y plot when one or more values are below
reporting limits? Perhaps the most common method is to plot a symbol at
either the detection limits, or at one-half the detection limits. Both
suffer from the same problem = more is shown than is known. The locations
of nondetects are not known with assurance, only that values are somewhere
between zero and the detection limit. However by plotting data as a point
at one location, the eye is told something different -- the location is
known.

A method in agreement with what is actually known about the data is to plot
it as a bar representing the range of possible values for the observation.
Using a lighter shade than those for detected observations emphasizes the
uncertainty in value for nondetects. An example scatterplot for the small
data set below is below.

Year	Dissolved Iron	(data from Hughes and Millard, 1988)
1977	20	
1978	<10	
1979	<10	
1980	<10	
1981	<10	
1982	7	
1983	3	
1984	<3	
1985	<3	

In the book we also look at determining correlation, regression and a trend
test for these and similar data.

3. Newsletter archive on the web
Back issues of recent newsletters can be found and downloaded from the
Practical Stats web site. If you have subscribed but are not receiving
them (AOL routinely blocks them as spam) add our "ask" email address to
your "white list". Spam is a huge problem, not just for each of us
personally (I get tons) but for legitimate newsletters to make it through
to the readers who have subscribed. Newsletters currently on our web site

are
* Is Excel and adequate statistics package?
* Why substituting one-half the detection limit for nondetects is a bad idea.
* Cohen's MLE for estimating the mean of nondetects - commonly used but outdated.
* Correlation with censored data

You'll receive them two months earlier than when they appear on the web by signing up and making sure that they get through your spam filters.

4. Upcoming Courses
Less Than Obvious, the 2-day course on the analysis of data with nondetects, will be taught again on August 18-19, 2004 on the campus of the Colorado School of Mines in Golden, Colorado. Registration information is on the PracticalStats web site. Course content follows the NADA textbook outline listed above.

Applied Environmental Statistics will also be taught in Golden, on June 7-11, 2004. This is our survey course on practical statistics -- How to "make sense of your data". Course outline is on the web site. It covers topics from plotting data and hypothesis tests to regression, trend analysis, and comparing data to standards.

We're looking for other locations - if you know of at least 10 people ready to take either course we'll arrange to come to your area and fill the remaining slots through advertising. Or we'd be glad to teach it just for your group. Contact us at ask[at]practicalstats.com .

'Til next time,

Practical Stats
http://www.practicalstats.com
-- Make sense of your data

from Nondetects And Data Analysis,
D.R. Helsel, 2004