Practical Stats Newsletter   for Autumn, 2004

In this newsletter:
1. "Less Than Obvious" - Handling Nondetects. October 28-29 in Golden CO
2. Storing nondetects in databases
3. New textbook "Nondetects And Data Analysis" now available

1. "Less Than Obvious" - Handling Nondetects. October 28-29 in Golden CO
Less Than Obvious, our 2-day course on the analysis of data with
nondetects, will be taught October 28-29, 2004 on the campus of the
Colorado
School of Mines in Golden, Colorado. Registration information is on the
PracticalStats web site. The course demonstrates methods for interpreting
data with nondetects, without substituting arbitrary values such as
one-half the detection limit. It covers methods for hypothesis testing,
regression and correlation, and computing stats like the UCL95. The course
also includes a copy of the just-published new textbook "Nondetects And
Data Analysis: Statistics for censored environmental data". There are still
openings as of today.

Just Scheduled!
Applied Environmental Statistics
March 21-25, 2005   Colorado School of Mines
                Golden, Colorado
Our 1-week survey of applied statistics. How to "make sense of your data".
Registration info will be at PracticalStats.com soon.

2. Storing nondetects in databases
There are two conventions for storing data with nondetect values in a
database for subsequent statistical analysis. Both require two columns to
represent each variable. The first method, utilizing an indicator
variable, is used by nonparametric methods for censored data. The second
method stores interval endpoints bracketing the possible values each
measurement might take. This format is generally used by parametric
methods for censored data.

Indicator variable
In this method the first variable stores the measured values for detected
observations, as well as values for the detection limits of all nondetects.
 The second variable contains a code that indicates whether the value in
the first column is a detected concentration or the detection limit for a
censored observation. Typically a 0 is used to indicate one state, and a 1
for the other. Some software also allows text phrases such as "censored"
and "uncensored" or "above" and "below" to be stored as the indicator

variable. If the remark variable is ignored, summary statistics computed on the first variable will be biased high, because the detection limit will be erroneously considered as the measured value for all nondetects.

The method's primary disadvantage is that if text is not used it can sometimes be difficult to remember which state each indicator value refers to.

Example
The three values of <1, <5 and 10 are represented by two columns below. A 1 indicates a censored observation.

Col 1. Concentration	Col 2. Indicator
1	1
5	1
10	0

Interval endpoints
In this format, all values in the data set are represented by the interval those data fall within. Endpoints for the interval, high and low, are stored as two variables. For detected observations the values of the two variables are identical. For censored data the values differ.

Interval endpoint storage is a flexible system that can represent left-censored data (less-thans) as well as right-censored (greater-thans) data, and both may be stored within the same data set. The format is illustrated below.

Example, cont.
For a <1, <5 and a 10, two variables represent each observation, one at the lower end (Start) of an interval within which the measured values lie, and a second at the upper end (End) of the interval. For left-censored data a value of 0 is entered in the Start variable to denote that the nondetect is no lower than zero, and the detection limit is entered as the largest possible value in the End variable. The detected value of 10 has the same number for both variables.

Start	End
0	1
0	5
10	10

If it is possible for data to extend in value out to infinity or minus infinity, an asterisk (*) is placed in the appropriate column to represent infinity. The asterisk states that there is no known boundary, unlike the 0 lower boundary in the Start variable above. A right-censored value of >100 would be designated as:

```
   Start   End
    100    *
```

Maximum likelihood methods within commercial statistics software generally use this format for their computations. Reading a direct printout of the data set is clear -- it is perhaps the least confusing of the three methods. However if the data are not recognized as being censored, summary statistics of either variable alone will give erroneous answers; the Start variable will be biased low and the End variable biased high.

3. New textbook "Nondetects And Data Analysis" now available
The new textbook Nondetects And Data Analysis: Statistics for censored environmental data" is now available for ordering from Wiley publishers. It provides solutions for environmental scientists and professionals who need to interpret and analyze data that fall below the laboratory detection limit. Adapting survival analysis methods that have been successfully used in medical and industrial research, the author demonstrates, for the first time, their practical applications for studies of trace chemicals in air, water, soils, and biota. Readers quickly become proficient in these methods through the use of real-world examples that are solved using MINITAB® Release 14, a popular statistical software package, as well as other commonly used software packages.

http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471671738.html

We'd be glad to hear your comments and reactions. Email us at ask[at]practicalstats.com .

'Til next time,

Practical Stats
http://www.practicalstats.com

-- Make sense of your data