Practical Stats Newsletter for Summer, 2004

In this newsletter:
1. "Less Than Obvious" - Handling Nondetects. Aug 18-19 in Golden CO
2. Insider Censoring: A hidden problem
3. NADA coming in October

1. "Less Than Obvious" - Handling Nondetects. Aug 18-19 in Golden CO

Less Than Obvious, our 2-day course on the analysis of data with nondetects, will be taught August 18-19, 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 pre-release copy of the new textbook "Nondetects And Data Analysis: Statistics for censored environmental data", available October 2004. Get a jump on this topic by attending next month! Rocky Mountain Natn. Park and other amazing sites to visit and enjoy are just two hours away.

2. Insider Censoring: A hidden problem

Insider censoring (called "informative censoring" in the medical sciences) occurs when the choice of a reporting (detection, quantitation) limit is made based on - informed by - the measured value in the laboratory. Lower values get one limit, higher values another. For environmental sciences, values below a detection limit (the lower limit) are assigned the higher quantitation limit as a censoring level, and so are reported as <QL. Values measured between the two limits (so also measured as less than the quantitation limit) are assigned the (lower) detection limit, and so are reported as individual values rather than as less-thans. These in-between values are often reported as "Estimated" or "E-values". The choice of reporting limit is therefore a function of the measured concentration of the sample. The result is that all interpretations of data reported in this manner are biased. The two figures on the NADA web page illustrate the problem:
http://www.practicalstats.com/nada/insider.htm

Figure A shows the original measured concentrations as a histogram. Forty percent of observations are measured between 0 and the detection limit (the white bar). Twenty-five percent are measured between the detection and quantitation limit (light gray bar), and the remaining higher measurements reported as detected values (dark bars). The measurements between the

limits (in the gray bar) are reported along with a qualifier that these observations are 'estimated', but still reside between the two limits. This bar graph applies as long as the values measured below the detection limit are reported as "<DL". No bias exists at this stage.

Figure B shows the same data using insider censoring. The only difference is that values measured below the detection limit are now reported as being below the quantitation limit or "<QL", as if they might belong anywhere from zero up to the quantitation limit. To reflect how any interpretation process will interpret data reported this way, the probability (forty percent) that observations may fall below the detection limit is spread evenly along the entire range from zero to the quantitation limit. This is pictured in Figure B as two white bars totaling 40 percent evenly split between two categories, 20 percent of observations in each category. The result of insider censoring is that the probability that an observation might fall between the detection and quantitation limits is exaggerated, and the probability that it would fall below the detection limit is underestimated, in comparison to the proportions actually measured. The shape of the histogram has been changed, and so too will all interpretations that follow. This upward bias changes the proportions (the percentiles) of data, and is picked up by any subsequent procedure, from the simplest computation of means or percentiles to more complex methods such as maximum likelihood.

For example, suppose you were computing the mean by substituting one-half the reporting level for all less-thans (a simple, but not recommended, method). Here's the data as measured in the lab:
1.17 2.93 2.68 4.73 1.15 1.08 5.49 8.24 0.70 1.59 3.00 1.31
Mean = 2.84

After insider censoring with a DL of 2 and a QL of 4, the data are reported to the user as:
<4 2.93 2.68 4.73 <4 <4 5.49 8.24 <4 <4 3.00 <4

and after substituting one-half the reporting level, the data are considered to be:
2 2.93 2.68 4.73 2 2 5.49 8.24 2 2 3.00 2
Mean = 3.26

Insider censoring has produced a 15% upward bias in the mean for these data.

Perhaps you know of laboratories using this type of reporting method? There is an easy fix to this problem. Data users can re-censor their data, implementing one of three unbiased methods listed below:
1. censor all data below the quantitation limit as "<QL" (no 'estimated' or

"E" values),
2. censor data measured below the detection limit as "<DL" instead of "<QL" (avoid the upward shift), or
3. use interval-censoring methods for interpretation (see NADA. This may be the subject of a later newsletter).

3. NADA coming in October

Much more detail on handling censored data, data with nondetects, is coming in October.
Nondetects And Data Analysis. NADA. See
http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471671738.html
Look for it.

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