Practical Stats Newsletter for April 2011

Subscribe and Unsubscribe: http://practicalstats.com/news/
All of our past newsletters:
http://www.practicalstats.com/news/news/bydate.html

In this newsletter:
1. Registration deadlines for upcoming courses
2. Things done with nondetects that are just wrong!
3. Get a Helsel and Hirsch hardback book when you register for AES

1. Registration deadlines for upcoming courses

Registration for is open for our Applied Environmental Statistics course in Philadelphia the first week of May. As always, online registration is available through the Practical Stats "Upcoming Classes" page (URL below). Registering before April 11 will save your project or employer money. Also see section 3, below for a special offer.

Applied Environmental Statistics
Statistics, down to earth

May 2-6, 2011	$1395 registration before April 11
Temple Univ. City Center	$1495 on or after.
Philadelphia, PA 19102	

Time Series and Forecasting
for frequently-collected, "real-time" data

April 4-5, 2011	$995
Homewood Suites Littleton	registration still available
Littleton Colorado 80127	

Nondetects And Data Analysis
Correctly interpret data below detection limits

April 6-7, 2011	$995
Homewood Suites Littleton	registration still available
Littleton Colorado 80127	

You can always find our complete course listing on our "Upcoming Courses" page at http://www.practicalstats.com/new_classes/classes.html

2. Things done with nondetects that are just wrong!

Over the years I've seen some twisted things done with nondetect data. It hasn't seemed to stop – I just reviewed a report with one of the items below in it. So here's a rogue's gallery of things that are just plain wrong that people do with nondetects. These are from the upcoming 2nd edition of the NADA textbook, and in our upcoming webinar on April

11th -- "Why Subbing One-Half of the Detection Limit is Trouble and What You Can Do Instead". Registration for the webinar and its sequel in May is still available.

a) Deleting nondetects and just looking at detections.
If you had a dataset of entirely detected values, would you ever throw away the bottom 60% of the data and compute the mean of what's left? I didn't think so. But this is just what happens when someone computes the mean of detects only, and then lets the reader try to evaluate it. Or worse, does this for several groups. What does the mean of the top 40% of one group have to say in comparison to the mean of the top 70% of another group? This is throwing away the primary information in your nondetects, which is the proportion of values falling below each reporting limit. Use the information in your entire set of data, including nondetects. There are ways to compute a mean, median etc. for the entire group, detects plus nondetects.

b) Substituting one-half or one over the square-root of 2 times the RL for nondetects
If you are a regular reader of this newsletter, you've already heard many times that this is a terrible practice. Substitution produces an artificial pattern, which I call 'invasive data', into the dataset. For more on its dangers, see our Spring 2003 newsletter (yes, it has been a longtime topic for Practical Stats!) freely downloadable on our website, or the article "Fabricating Data" -- http://dx.doi.org/10.1016/j.chemosphere.2006.04.051
or hear it from a recent EPA guidance document: "Based upon the results of Singh, Maichle and Lee (2006) it is strongly recommended to avoid the use of the DL/2 method....even when the percentage of NDs is as low as 5%-10%."

c) Substituting for the standard deviation, CV or variance
Those unfamiliar with proper methods for determining estimates of the variability of censored data sometimes just fabricate a number that seems 'reasonable'. This has made its way into regulations, where 0.6 for the CV (the ratio of the standard deviation to the mean) is popular. That's no better than a guess. It could be very far off, with unwarranted consequences to health or to the cost of monitoring programs. The three methods in our NADA textbook for computing descriptive statistics, MLE, Kaplan-Meier and ROS, will each estimate the mean and standard deviation, and so the CV, for censored environmental data.

d) Comparing groups or testing trends of %detections when the RL is changing
Envision two sets of identical data, the first measured 10 years ago, the second measured this year. Same concentrations. No change over time. The early one was censored with a mix of two reporting limits, at 1 and 10 ug/L:
<1 <1 <1 3 5 7 9 <10 <10 <10 <10 <10
while this year's data was measured with better instruments. Now the only reporting limit is at 1 ug/L:
<1 <1 <1 3 5 7 9 <1 2 2 3 5

The analyst computes that there were only 33% detects 10 years ago, but now there are 67% detects of this dangerous chemical. There's been a huge change, and something must be done about it! Comparing %detections between groups or over time only makes sense when the mix of reporting limits is constant. In practice, this happens only when

there is one reporting limit. Instead of computing the %detections above a moving target, use methods that correctly account for differing reporting limits in your analysis. You'll find them in our NADA textbook and courses.

3. Get a Helsel and Hirsch hardback book when you register for AES

For the first time ever, when you register for our May Applied Environmental Statistics course in Philadelphia, you'll receive a hardback copy of the 2002 textbook *Statistical Methods in Water Resources* by D.R. Helsel and R. M. Hirsch. This book has become a classic. It was first published in 1992 as a hardback by Elsevier. They charged $200 for it (that's $313 in today's currency!). You can find copies of that hardback for around $400 in online used bookstores today. Eventually a paperback was released. In 2002 the book was republished by the US Geological Survey, but only as a pdf. Five errors in the original edition were corrected at that time. Now you can get a hardback copy of the corrected 2002 edition when you register for our AES class in Philly. In addition, we'll introduce you to R, the statistics software that is the world's standard, is incredibly comprehensive, and is free! Spring is a great time to be in Philadelphia – not too hot yet, flowers are blooming, and the Phillies are playing in town that week. The class is located in Center City, nearby to Independence Hall, shopping, and Philly cheese steaks. And you get a free book!

'Til next time,

Practical Stats (Dennis Helsel)
-- Make sense of your data