

Practical Stats Newsletter for Summer, 2003

In this newsletter:

1. New course: Less Than Obvious -- December 2003
2. Cohen's method - popular but outdated
3. Our privacy policy -- the address stops here

1. New Course on Handling Nondetects

Less Than Obvious, a 2-day course on the analysis of data with nondetects, will be offered Dec. 11-12 at the Univ. of California's Cooperative Extension Auditorium in Sacramento, CA. The course content of Less Than Obvious has undergone a complete remake, along with a completely new set of notes. Course content covers how to compute summary statistics, hypothesis tests, and regression models for censored data, without ever substituting a single number for data below detection limits. The excerpt from the Spring Practical Stats newsletter on the dangers of substituting one-half the detection limit, and the short summary of Cohen's method this month, are excerpted from the new course notes, where they are discussed in much more detail. For more information on course content and to register for the course, go to

<http://www.practicalstats.com/Pages/lto.html>

{Editor's note. The course is now titled ***Nondetects And Data Analysis***}

2. Cohen's Method - popular but outdated

In the late 1950s and early 60s, several papers in statistical journals by A.C. Cohen introduced maximum likelihood estimation (MLE) for determination of the mean and standard deviation of censored data. MLE uses both the un-censored (detected) observations, along with the proportion of data below one or more censoring thresholds (detection limits) to compute statistics for the entire data set. MLE requires that the distribution of the data be specified -- normal and lognormal distributions are commonly used in environmental work. These methods are computer intensive, and the necessary computing power at that time was not available to most people. So Cohen presented a version of the method which used a lookup table to calculate estimates of the mean and standard deviation by adjusting downwards the statistics of the un-censored observations as a function of the amount of censoring in the data set. The table lookup method was popularized for environmental sciences in Gilbert's 1987 widely-read book on environmental pollution monitoring.

The table lookup method has a serious drawback -- it was designed for use with one censoring threshold. Today environmental data usually exhibit multiple detection limits. With modern computer hardware and

software, Cohen's approximate method is now unnecessary. More accurate direct solutions of the likelihood equations are possible with commercially-available statistical software. Routines are available in Minitab, SAS, S-Plus and others for computing summary statistics for censored data. However, the table-adjustment method is still often recommended. Though its results are approximately correct, with the approximation better for data having one detection limit, its results are not as good as those produced by MLE routines in statistical software.

In order to use Cohen's method on multiply-censored data, all values below the highest limit must be set to a less-than. For example, below are 24 observations, 11 un-censored and 13 censored at 3 detection limits.

```
0.5 0.5 0.5 0.6 0.7 0.7 <0.9 0.9 <1.0 <1.0 <1.0 <1.0 1.5
1.7 <2.0 <2.0 <2.0 <2.0 <2.0 <2.0 <2.0 <2.0 <2.0 2.8 3.2
```

In order to compute the (one-dl) Cohen's method, all data below the highest detection limit must be considered censored. In other words, all values below 2 become <2, resulting in 22 <2s, plus the two un-censored values of 2.8 and 3.2 . Two coefficients, h and gamma, are computed in order to use the table Cohen developed, and which is reprinted in Gilbert's book. From this table is gotten another coefficient, lambda, which is used to adjust the mean and standard deviation of the detected observations down to levels which represent the entire data set. For the above data, the mean and standard deviation are estimated using Cohen's method as

```
mean      1.03
std dev   0.88
```

These were computed by assuming the data follow a lognormal distribution.

Cohen's method produces estimates more reasonable than those obtained by substituting values for less-thans (see the Spring Practical Stats newsletter), but they aren't that similar to results obtained using more exact procedures. For example, using the maximum likelihood routines in Minitab for lognormal distributions, the mean and standard deviation are estimated to be

```
mean      0.95
std dev   0.66
```

These MLE values are more consistent with other procedures for censored parameter estimation than are the table-lookup results [more details on these other procedures are given in the course]. The use of an 'approximate' or 'similar' method such as Cohen's is questionable when industry-standard and more accurate alternatives such as exact maximum likelihood are readily available. There is nothing that can be done

with Cohen's table lookup procedure that cannot be more directly computed with modern software. In addition there are fewer opportunities for error, and the software is capable of estimation for multiple detection limits, using the values of detected observations such as the 1.5 and 1.7 above, rather than calling these both <2 . Software is also capable of performing hypothesis tests and regression for censored data, in addition to computing summary statistics.

Therefore while Cohen's method is a vast improvement over substituting one-half the detection limit, it is far from being "best available technology" in the 21st century. There are better ways. We will present them in the December course.

3. Our privacy policy - the address stops here

Subscribers to this newsletter entrust their email addresses to us. We value that trust. Your address is not given to anyone else. Not for any reason. Not at any time. It cannot be seen by other subscribers. It is not released to other subscribers. Be sure that no other use is made of it other than to send you this newsletter. If you have any questions about this, just email us at [ask\[at\]practicalstats.com](mailto:ask@practicalstats.com).

Practical Stats

<http://www.practicalstats.com>

-- Make sense of your data