

Practical Stats Newsletter for July 2007

In this newsletter:

1. Excel worksheet for Kaplan-Meier method now on our website (free)
2. More software for handling nondetects
3. Upcoming courses

1. Excel worksheet for Kaplan-Meier method now on our website (free)

Kaplan-Meier is a standard procedure in the medical sciences for computing summary statistics of censored data. The NADA (Nondetects And Data Analysis, Wiley, 2005) textbook presents it as one of its recommended procedures for handling left-censored environmental data. Our Spring 04 newsletter, available in our newsletter archive,

<http://www.practicalstats.com/news/>

discussed the method in some detail. Here we'd like to mention software for computing it. One of the hurdles to using better statistical methods for nondetect data is the availability of software. We have provided Minitab macros and links to R routines for computing Kaplan-Meier and other nondetect-capable methods on our website for over two years. This free software has allowed people using those two software packages to employ modern analysis methods for handling nondetect data.

Kaplan-Meier is an easy to compute procedure, involving only counting and ranking data. It is so easy that it can be computed in Excel! So to make K-M more widely available to the masses, we have placed an Excel worksheet that performs it on our site. The worksheet estimates the mean, standard deviation, median, 25th, and 75th percentiles, and the (t-interval) UCL95 for censored data. It is not a macro, just a simple worksheet, to minimize the possibility of viruses, etc. It has been checked and found clean by Norton Antivirus. It is limited at present to data at 100 different values - there may be multiple observations at each value, however. It 'flips' the data and performs all processing, so you simply put in the values and get back the results. If you find it useful, cite the source as PracticalStats.com when presenting or publishing results. Give us credit -- after all, we're giving you the worksheet for free.

Download KMStats.xls at: <http://www.practicalstats.com/nada/nadafiles/downloads.html>

2. More software for handling nondetects

a) In our October 06 newsletter (Oct06_UCL) at

<http://www.practicalstats.com/news/>

(and click the NADA news link)

we discussed methods for computing the one-sided 95% upper confidence interval on the mean (UCL95) for censored data. The Kaplan-Meier method using a t-interval or bootstrapping performed best when studied by Singh et. al (2006). ROS procedures did not work as well, but performed far better than substitution of 1/2DL and similar values. Now software for computing the UCL95, developed by Singh and others for the USEPA, has incorporated these methods for handling nondetects, particularly the Kaplan-Meier procedures. ProUCL 4.0 is available for download at:

<http://www.epa.gov/esd/tsc/software.htm>

b) Minitab macros for computing the bootstrapped Kaplan-Meier estimates of the UCL95 and a variety of summary statistics has always been available on our web site. ROS methods are also there. New in version 2.3 is the macro KMStats that, like the Excel worksheet, computes Kaplan-Meier estimates while doing all the 'flipping' internally. We used to give this out only at the NADA courses, but now it is included in the online macros set. NADAforMTB23.zip (the numbers will change as newer releases occur) is available at:

<http://www.practicalstats.com/nada/nadafiles/downloads.html>

Also available there is a SAS macro for computing the ROS procedure. Download CROS.zip.

c) NCSS is the first commercial statistics software to explicitly incorporate the methods from the NADA textbook - no macros required. While we have not used the software, and so cannot at present vouch for accuracy or breadth of methodology, the methods listed on their web page certainly appear to incorporate much of what is found in the NADA text. If you are considering a statistics software purchase and will be dealing with nondetects, check them out at

<http://www.ncss.com>

3. Upcoming courses

Our newsletter has been more infrequent than usual -- we have been extremely busy teaching courses in 2007. Our pace has now slowed (by design), and so we will also publish an August newsletter in an attempt to catch up. The course keeping us the busiest is our new Untangling Multivariate Relationships (UMR) survey of multivariate methods applied to environmental (chemical and biological) data. The next offering with open enrollment is

Untangling Multivariate Relationships

November 1-2, 2007

Colorado School of Mines

Golden, Colorado.

For those of you attending the Geological Society of America meeting in Denver, it follows GSA. Stay for the course and make your travel dollars count! See the website for online registration and details, and join us there.

And for those of you in Europe, there is a NADA class in Helsinki, Finland in late August.

'Til next time,

Practical Stats

<http://www.practicalstats.com>

-- Make sense of your data