

Practical Stats Newsletter for April 2018

Subscribe and unsubscribe: <http://practicalstats.com/news>

Archive of past newsletters <http://practicalstats.com/news/archive.html>

In this newsletter:

- A. Practical Stats Courses
- B. New Software for Nondetects
- C. The R Graph Gallery

A. Practical Stats Courses

Our self-paced Applied Environmental Statistics course is available in two parts on our online training site:

<http://practicalstats.teachable.com/>

The two courses separately are each \$650 USD for a 1-year access for one person. Or get both courses together (equivalent to our week-long course) in a bundle for \$1200 USD.

B. New Software for Nondetects

As part of placing our Nondetects And Data Analysis course online (we're working to finish that – expect it to be available this summer on our Training Center), I've been looking at alternatives to some of the routines currently in the NADA package for R. While the NADA package has been a great help to many, there are improvements and changes that need to be done. The new Nondetects And Data Analysis online course will also use methods for censored data available in the EnvStats package by Steven Millard (<https://cran.r-project.org/web/packages/EnvStats/index.html>) as well as routines I'm writing that make some analyses easier, and some graphs prettier, than what's gone before.

Examples of the routines that will be found in our online course that augment what is found in the NADA package are:

1. Bootstrap confidence intervals for Kaplan-Meier estimates of the mean using the EnvStats package. Below, a 95% two-sided bootstrap interval for the Kaplan-Meier mean pyrene concentration is computed:

Results of Distribution Parameter Estimation
Based on Type I Censored Data

```
-----  
Assumed Distribution:      None  
Censoring Side:           left  
Estimated Parameter(s):  mean      = 164.09450  
                          sd          = 389.41379  
                          se.mean    =  49.75292  
  
Estimation Method:       Kaplan-Meier  
Data:                    Pyrene  
Percent Censored:        19.64286%
```

```

Confidence Interval for:      mean
Confidence Interval Method:  Bootstrap
Number of Bootstraps:       5000
Confidence Interval Type:   two-sided
Confidence Level:           95%
Confidence Interval:        Pct.LCL = 99.91121
                             Pct.UCL = 264.31983

```

2. Bootstrap confidence intervals for Regression on Order Statistics (ROS) estimates of the mean using the EnvStats package. Below, a 95% two-sided bootstrap interval for the ROS mean pyrene concentration is computed:

```

Results of Distribution Parameter Estimation
Based on Type I Censored Data
-----

```

```

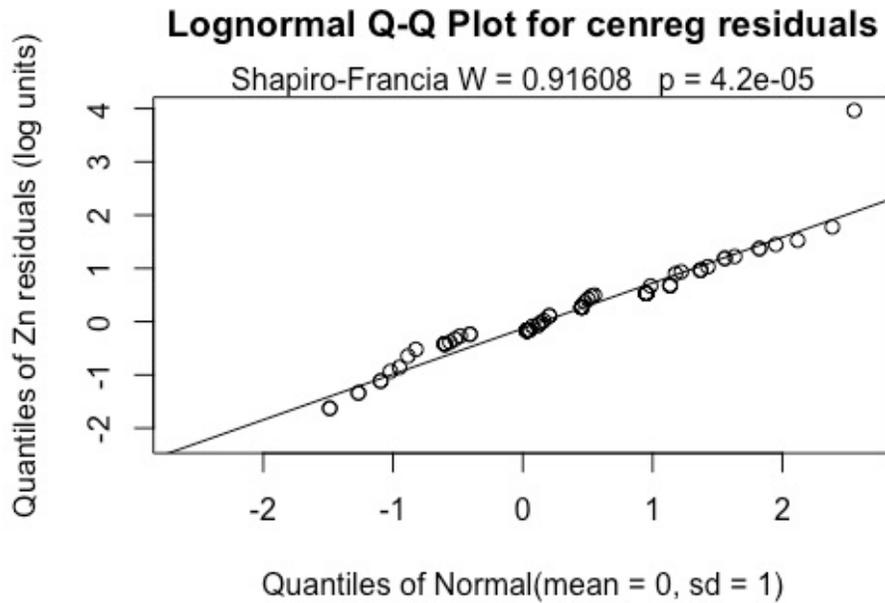
Assumed Distribution:        Lognormal
Estimated Parameter(s):     mean = 163.371129
                             cv   =  2.406266

Estimation Method:          Imputation with
Q-Q Regression (rROS)

Data:                        Pyrene
Confidence Interval for:    mean
Confidence Interval Method: Bootstrap
Number of Bootstraps:       5000
Confidence Interval Type:   two-sided
Confidence Level:           95%
Confidence Interval:        Pct.LCL = 99.77337
                             Pct.UCL = 262.12133

```

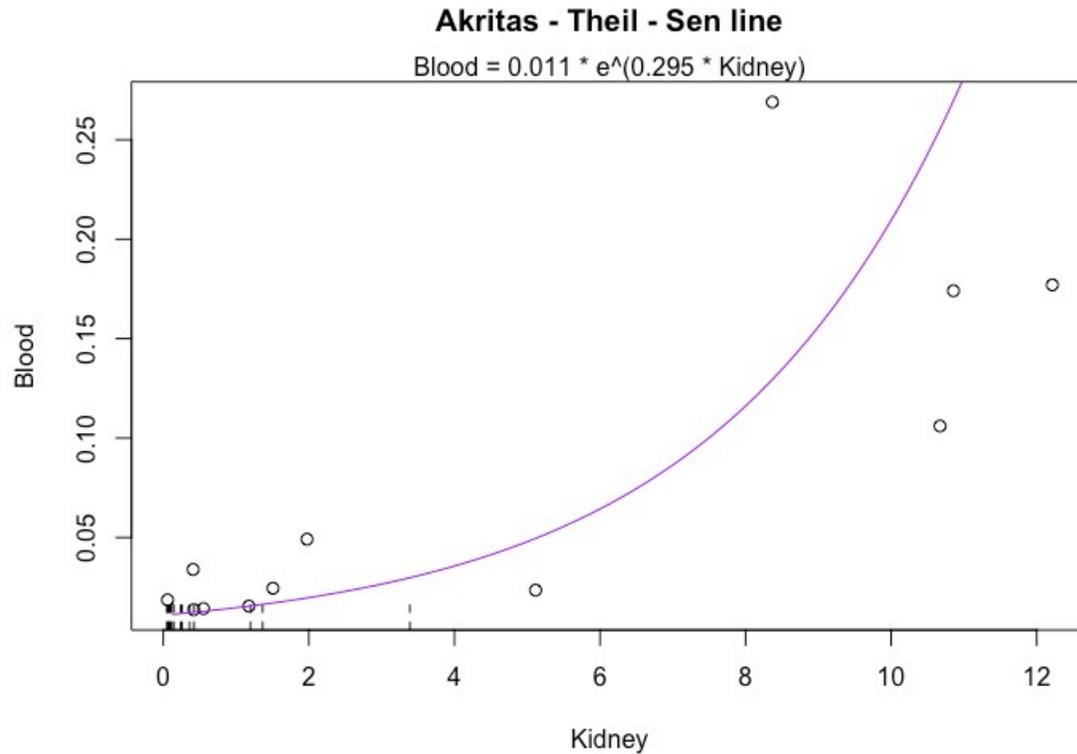
3. The new cenregQQ plot command (in the course script file) will compute a regression model for a Y variable with nondetects, and draw a QQ plot of the residuals along with performing a test for normality of censored data, to evaluate whether the assumed distribution for the residuals seems reasonable. This may also be done when using maximum likelihood to compute a parametric test (something like a t-test or ANOVA for censored data) for discerning differences in group means. Below is a QQ plot and test of the whether residuals of log of zinc concentrations follow a normal distribution when testing for a difference in mean log zinc concentrations at two locations:



Residuals appear to follow a normal distribution except for one point at the high end. That one large outlier causes the test to reject normality ($p = 0.000042$), which is sufficient to lower the test's ability to find differences that occur. As the test for differences in the mean of the two locations failed to find a difference, a nonparametric or permutation test of difference in means would be a better, more powerful procedure.

4. A new ATS command (in the course script file) performs the nonparametric Akritas-Theil-Sen test and line, analogous to the Theil-Sen method but for data with nondetects, and plots the resulting model along with the data. Here, lead in the blood of herons was related to lead concentrations in their kidneys. The equation was computed in log units of blood lead (Y variable), but results plotted on the original scale:

```
> Pbk <- ATS(Blood, BloodCen, Kidney, KidneyCen, retrans = TRUE)
```



There are additional capabilities we discuss in the course that augment what is available through the NADA package. Check out the Nondetects And Data Analysis course when it comes online this summer.

C. The R Graph Gallery

The R Graph Gallery provides colorful examples (with code to produce them) of familiar graph types that can be produced in R. Some examples for boxplots are found at:

<https://www.r-graph-gallery.com/boxplot/>

Take a look at the Gallery if you'd like to improve your R graphing capabilities.

Graphs for data with nondetects could use a bit more excitement themselves, and some colorful new examples will be found in the Nondetects And Data Analysis course coming this summer.

'Til next time,

Practical Stats

-- Make sense of your data