

Practical Stats Newsletter for March 2019

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

Archive of past newsletters

<http://practicalstats.com/news/archive.html>

In this newsletter:

- A. Practical Stats Courses and Webinars
- B. Testing Group Differences with Nondetects
- C. Upcoming Free Webinar on March 19th

A. Practical Stats Courses and Webinars

Our online training site: <http://practicalstats.teachable.com/>

Our self-paced Applied Environmental Statistics course is available in two parts, each \$650 USD for a 1-year access for one person. Or get both courses together (equivalent to our week-long in-person course) in a bundle for \$1200 USD. See our online training site at the link above.

The recording of our February webinar “Fitting Distributions to Data with Nondetects: How and Why?” is available for free viewing on our training site. It joins four other free webinars from the past that are available there for viewing.

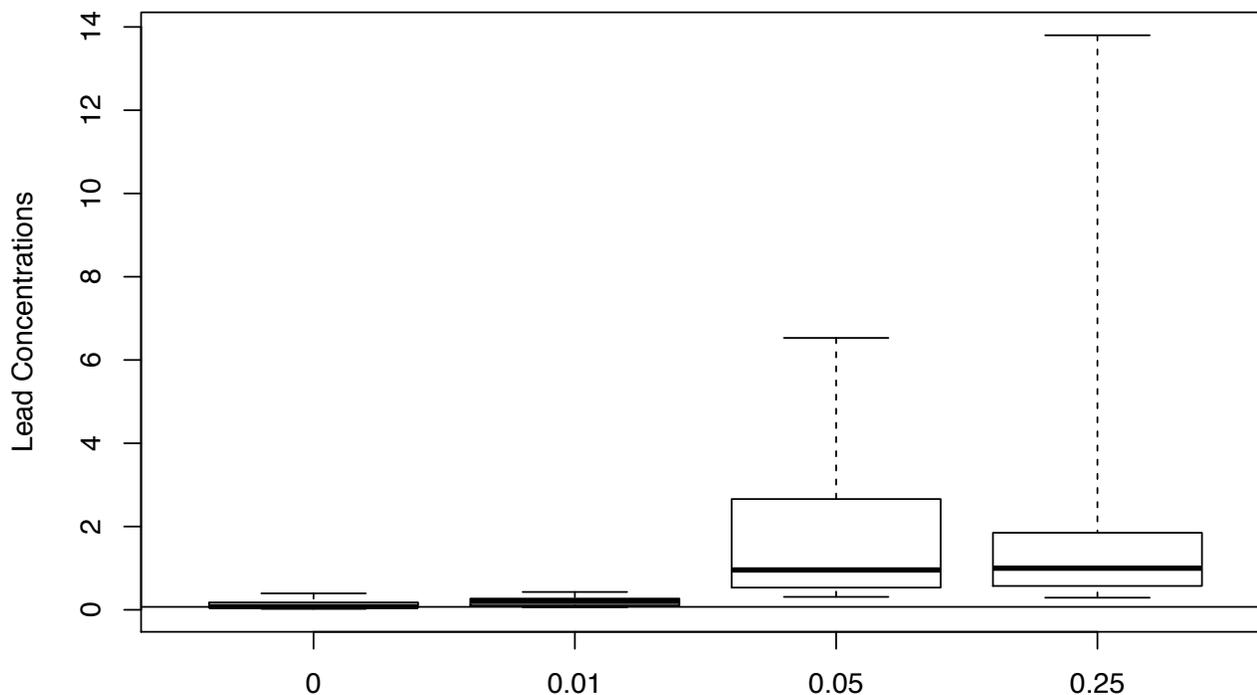
B. Testing Group Differences with Nondetects

Testing differences between groups of data is familiar to scientists. Analysis of variance (ANOVA), an expansion of the t-test to more than two groups, tests differences between group means. The nonparametric Kruskal-Wallis test determines if differences occur between cumulative distribution functions (cdfs, the set of percentiles) between the groups. However, what can be done when some observations are nondetects?

In my book *Statistics for Censored Environmental Data using Minitab and R* I covered how to use standard nonparametric methods like Kruskal-Wallis when there is only one detection limit. Our May 2010 and September 2005 (!) newsletters provide some additional background. Here I’ll demonstrate a parametric and nonparametric test for group differences that are designed for use with multiple detection limits.

The dataset is from my textbook and available through the NADA package in R. Named “Golden” after the scientist who created it, it measures lead concentrations in blood and several organs of herons. There are four dosage groups, birds who were exposed to from zero to 0.25 ug/L lead. Here I’ll look at lead levels in their feathers.

```
> cenboxplot(Feather, FeatherCen, as.factor(Dosage), log=FALSE, ylab="Lead Concentrations")
```



It certainly appears that differences occur between the four groups, with the two highest-dose groups not having any nondetects (bottom whisker of the boxplots ends above the horizontal detection limit line).

The parametric analogue to ANOVA for nondetects is a maximum likelihood procedure. No values are substituted for nondetects. Instead, the information content of nondetects is provided through the percent of data falling below each detection limit. A distribution such as the normal is assumed, and its mathematical function is fit to the data to best match both the values of detected observations and the percentages of data (detects and nondetects) falling below each detection limit. As these data are skewed, a lognormal distribution will be assumed, which is typical for trace contaminants on a scale where observations cannot go below zero. The `cenanova` command below is a script that comes with our online Nondetects And Data Analysis course. It uses the `survreg` command of the survival package of R, as does the `cenreg` command available in the NADA package. As you see, the `cenanova` command adds Tukey's multiple comparisons to the output as well. But the ANOVA-like overall test, answering the question "Are there differences in the group means or not?", can be performed with the NADA or survival packages.

```
> cenanova(Feather, FeatherCen, Dosage)
(Intercept) Factor0.01 Factor0.05 Factor0.25
 5.9884088 -0.5865279 -2.5244721 -2.5635843

ANOVA on natural logs of CensData: Feather by Factor: Dosage
Assuming lognormal distribution of CensData
Chisq = 24.17 on 3 degrees of freedom    p = 2.3e-05

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts
Fit: survreg(formula = logCensData ~ Factor, dist = "gaussian")
```

Linear Hypotheses:

	Estimate	Std. Error	z value	Pr(> z)	
0.01 - 0 == 0	0.58653	0.51951	1.129	0.6715	
0.05 - 0 == 0	2.52447	0.53976	4.677	<0.001	***
0.25 - 0 == 0	2.56358	0.51951	4.935	<0.001	***
0.05 - 0.01 == 0	1.93794	0.52797	3.671	0.0013	**
0.25 - 0.01 == 0	1.97706	0.50726	3.898	<0.001	***
0.25 - 0.05 == 0	0.03911	0.52797	0.074	0.9999	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

The overall test has a p-value of 0.000023 (2.3e-05), indicating a strong difference between group means. The Tukey's multiple comparison tests show that the two lowest groups (0 and 0.01) exhibit no difference, and the two highest groups (0.05 and 0.25) exhibit no difference, but that differences across all four low/high boundaries (0.05 vs. 0.25, etc.) are strongly significant.

The nonparametric analogue to the Kruskal-Wallis test to handle nondetects with multiple as well as one detection limit is a survival analysis procedure. No values are substituted for nondetects. Instead, the information content of nondetects is again provided through the percent of data falling below each detection limit. Scores, similar to Kaplan-Meier percentiles, are computed for detected observations and the test determines if these percentiles differ among the groups. The cen1way command below is a script that comes with our online Nondetects And Data Analysis course. It uses the survdiff command of the survival package of R, as does the cendiff command available in the NADA package. As you see, the cen1way command adds nonparametric multiple comparisons to the output as well. But the overall test, answering the question "Are there differences in the group percentiles or not?", can be performed with the NADA or survival packages.

```
> cen1way (Feather, FeatherCen, Dosage)
```

```
Oneway Peto-Peto test of CensData: Feather by Factor: Dosage  
Chisq = 22.45 on 3 degrees of freedom p = 5.25e-05
```

```
Pairwise comparisons using Peto & Peto test
```

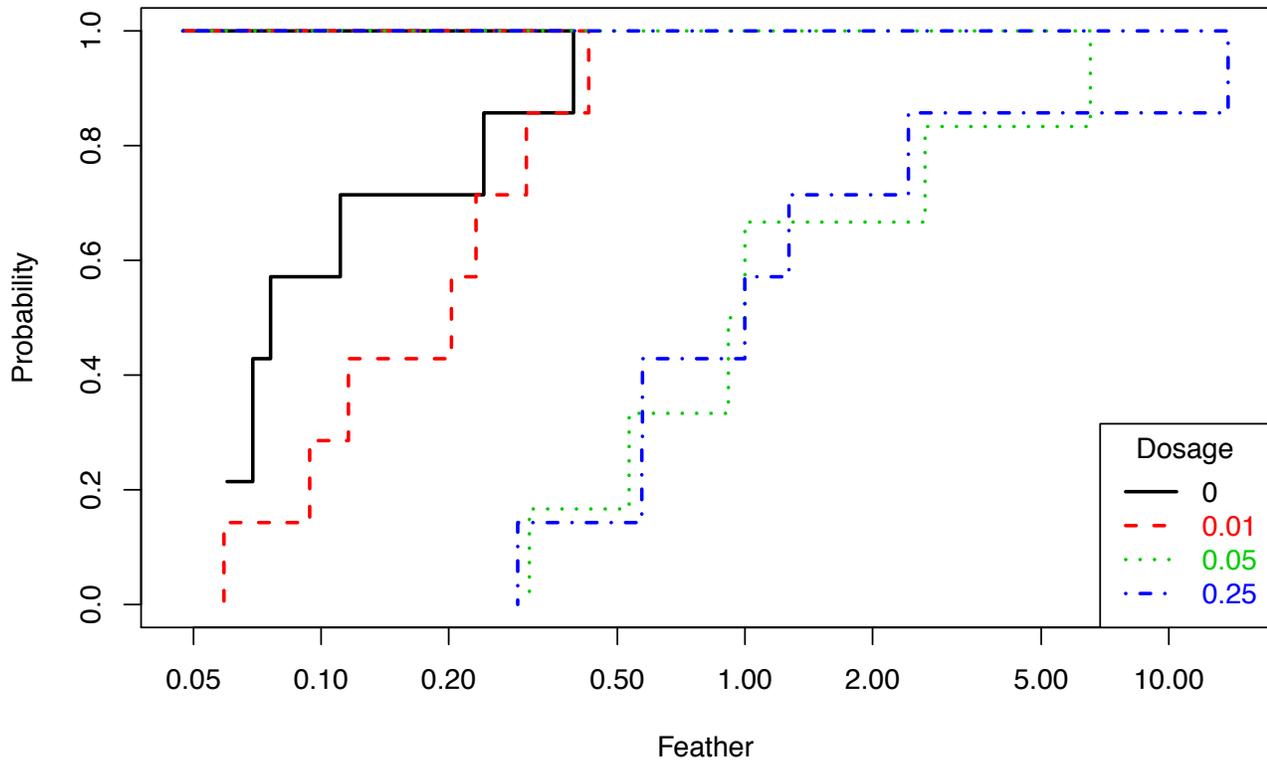
```
data: CensData and Factor
```

	0	0.01	0.05
0.01	0.3415	-	-
0.05	0.0024	0.0024	-
0.25	0.0024	0.0026	1.0000

```
P value adjustment method: BH
```

The overall and multiple comparison tests indicate the same results as did cenanova. A graph showing the cdfs (percentiles) of all four groups is an alternate to boxplots for illustrating the group patterns.

```
> plotcdf (Feather, FeatherCen, Dosage)
```



The similarity and differences among the four groups is clearly seen. This function takes elements from the NADA package and makes them easier to perform, and more colorful.

For more information about these tests, attend our free webinar on March 19th (see below).

C. Upcoming Free Webinar on March 19th

Testing Group Differences for Data with Nondetects

Tuesday Mar 19, 2019 at 1:00 PM Eastern, 10 am Pacific.

Testing group differences is familiar to scientists -- ANOVA for differences in means and Kruskal-Wallis for cdfs (percentiles). If differences occur, multiple comparison methods determine which groups differ from others. Tests with the same objectives exist that are designed for censored data, data subject to one or multiple detection limits. In this webinar you will learn how these tests work and how you can compute them.

To register:

<https://attendee.gotowebinar.com/register/1074923646089662989>

'Til next time,

Dennis Helsel

Practical Stats LLC

-- Make sense of your data