© PracticalStats.com

Nondetects And Data Analysis:
Three+ Group Tests with NDs

Dennis R. Helsel, Ph.D

Practical Stats

1

© PracticalStats.com

Methods for Testing Differences Between Groups

1. **Parametric.** Tests differences in group means - "Does at least one group mean differ from another group mean?" You must designate the assumed distribution that best matches the shape of your data.

2. **Nonparametric.** Tests differences in percentiles - "Is at least one group shifted away from others?" No shape is assumed or required.

3. **Simpler tests.** Convert data to above or below a single, or the highest of multiple, detection limits. Nonparametric -- no shape is assumed or required.

2

2

Parallels between standard methods and censored data methods

Standard Methods   Methods for Censored Data

Comparing Three or More Groups

ANOVA   Censored MLE regression with
                                                0/1 group indicators

Kruskal–Wallis test   Peto-Peto test
                                                (or variations)

3

3

Data: Ontario Pollen Monitoring Network

dataset: PollenThia.RData

- Pesticide concentrations are measured in pollen at beehives located across the province.
- Neonicotinoids are neurotoxins that kill insects through attacking receptors in nerve synapses.
- Nearly 100% of corn seed and roughly 60% of soybean seed are treated with neonicotinoids.
- Thiamethoxam is a neonicotinoid pesticide; the concern is its affect on honeybees.
- Do thiamethoxam concentrations differ in pollen between 4 stages of plant growth (pre-plant, post-plant, corn tassel, goldenrod)?
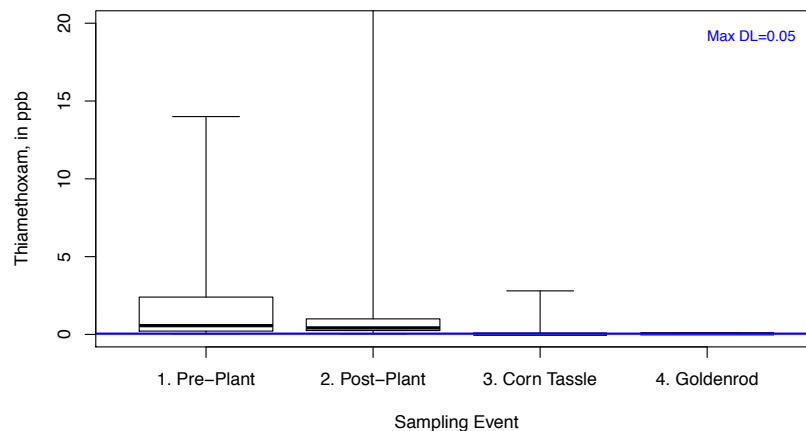
Source: Ontario Ministry of the Environment, Conservation and Parks

4

4

© PracticalStats.com

Concentrations are skewed

```
> cboxplot(Thiamethoxam, ThiaCens, SamplingEvent, Ylab =
  "Thiamethoxam, in ppb", Xlab = "Sampling Event", show = TRUE,
  Ymax = 20)
```


5

© PracticalStats.com

1. Parametric:
"ANOVA" using censored regression

Set up regression versus group ids (0/1 variables) :

$Y = \text{intercept} + \text{slope}_1 \cdot \text{Pre} + \text{slope}_2 \cdot \text{Post} + \text{slope}_3 * \text{Tassle}$

Where a 1 in either Pre, Post or Tassle defines membership in that group. A 0 in all three defines membership in the 4th group (Goldenrod).

For k groups, k-1 group 0/1 identifiers are needed. The slopes are fit by MLE

The test for "do group means differ?" is the test for whether all slopes are equal to 0. The test compares this three variable model against an equation that just has the intercept.

6

Parametric Method -- Maximum Likelihood Estimation (MLE)

- Starts with the observed data (including nondetects), and your choice of which distribution should be used

- Given the observed data, what values for parameters (intercept, slopes) are most likely to have given rise to these data?

- Optimization is performed to maximize the fit between observed data and the parameters

- Error in the fit is computed as a 'log likelihood', which is minimized as the parameters are changed in value until the best set of parameters are chosen

- No values are substituted for nondetects! Instead…….

- For censored data, the fit is to both observed detected observations and how the observed proportions of data below each detection limit (detects and nondetects) change with changing X_i values.

7

Parametric Method -- Maximum Likelihood Estimation (MLE)

- No values are substituted for nondetects. Information provided by nondetects is the observed proportion below each detection limit.

For example, if there were two DLs, 0.05 and 0.10,
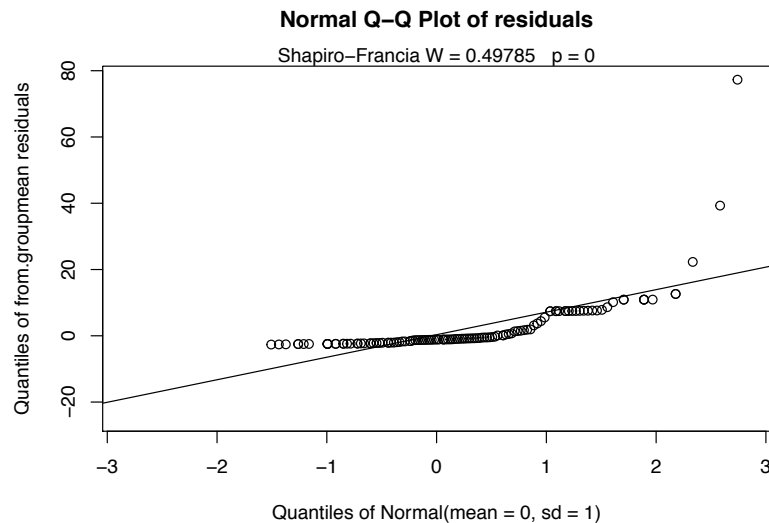
proportion <0.05 = (the number of observed <0.05) / n

proportion <0.10 = sum [# <0.05 + # <0.10 + # detects 0.05-0.099] / n

This is how the information in nondetects is used, without substituting values that you don't know for them.

8

Use Q-Q plot of residuals to evaluate whether the Y variable is on correct scale
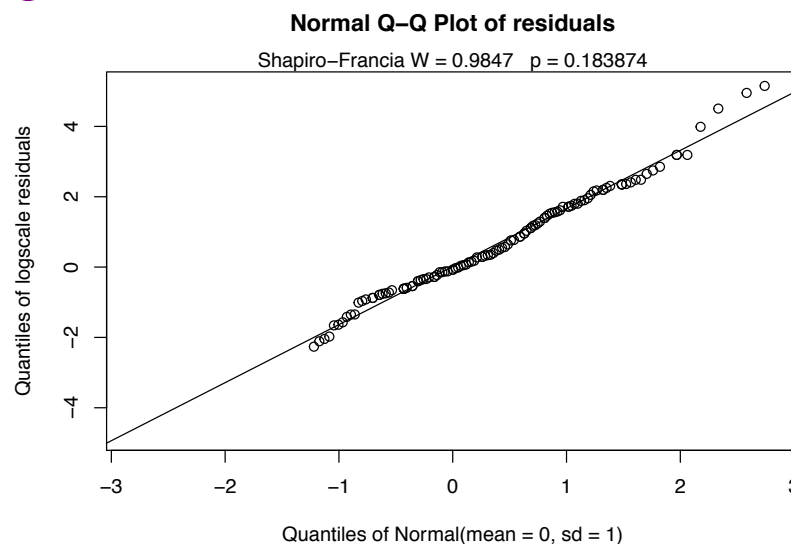
© PracticalStats.com

- Compares residuals of detected observations from their group mean to a normal distribution.
- Nondetects not plotted, but space for them left so that residuals of detected observations are plotted correctly.
- Straight line represents the fitted normal distribution.
- These data are curved -- not a good fit to a normal distribution. Shapiro-Francia test rejects normality.

Normal Q–Q Plot of residuals

Shapiro–Francia W = 0.49785   p = 0

Quantiles of from.groupmean residuals

Quantiles of Normal(mean = 0, sd = 1)

9

Thiamethoxam Residuals Are Fit Well by a Lognormal Distribution

© PracticalStats.com

- Log(Concentration) is used as the Y variable.
- Straight line pattern of residuals -- a good fit to a lognormal distribution.
- Will test the logs of Thiamethoxam using (parametric MLE) ANOVA

Normal Q–Q Plot of residuals

Shapiro–Francia W = 0.9847   p = 0.183874

Quantiles of logscale residuals

Quantiles of Normal(mean = 0, sd = 1)

10

Logs of Concentrations very symmetric

```
> cboxplot(Thiamethoxam, ThiaCens, SamplingEvent, Ylab =
   "Thiamethoxam, in ppb", Xlab = "Sampling Event", show = TRUE,
   LOG=TRUE)
```


11
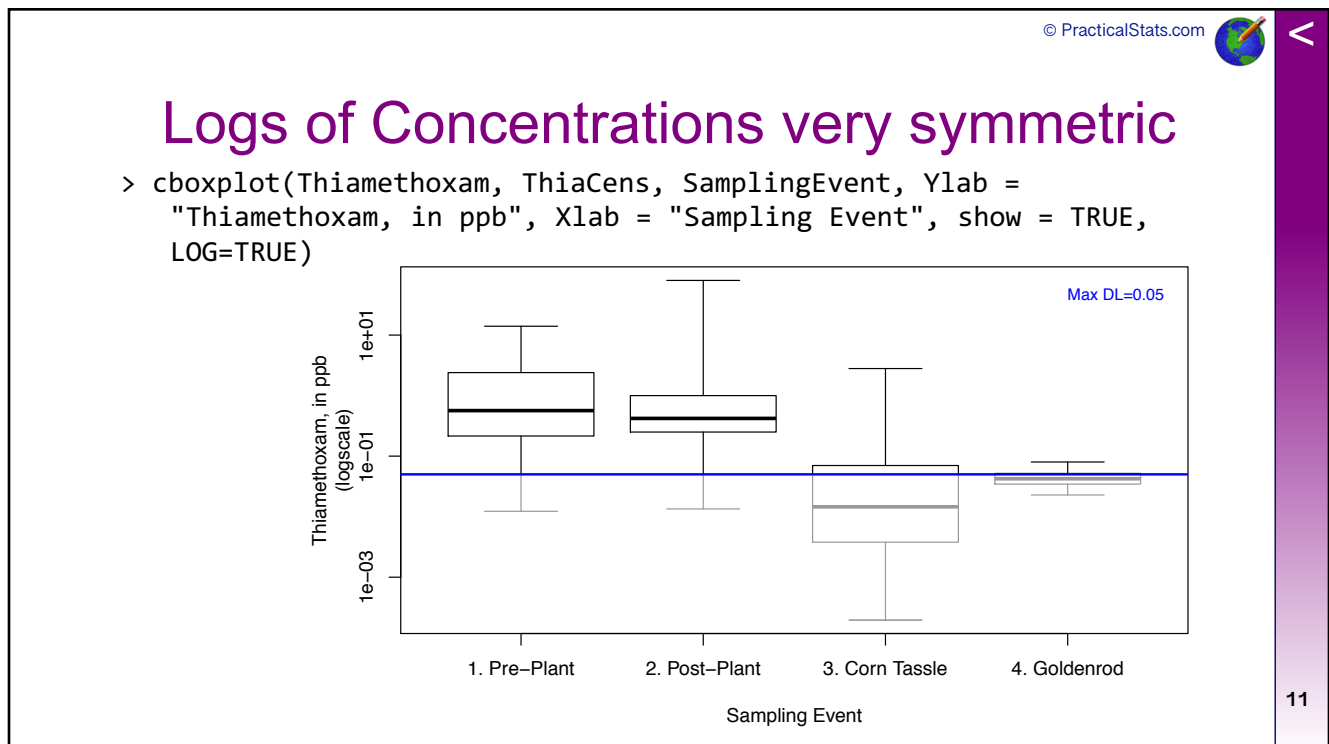
How the MLE test is computed

- The log-likelihood of a "no model" fit (all group means are similar, so slopes in the MLE regression are all zero) is compared to the log-likelihood for the "model" fit using the observed group means (at least one slope is non-zero).

- For k = 4 groups there are k-1=3 degrees of freedom used by the model, just as in ANOVA.

- A Chisquare statistic is computed as -2*(difference in the two log likelihoods).

- The Chisquare statistic (the signal strength for the difference in means) is compared to a chi-square distribution with k-1 degrees of freedom to determine the p-value

- If the Chisquare statistic is small (not much evidence for a difference in group means), the p-value is large (>0.05) and the group means are not significantly different from one another.

- If the Chisquare statistic is large, the p-value is small (<0.05) and at least one group mean differs from the others.

- This is an overall test -- no statement of which group means differ from the others is made.

12

Censored "ANOVA" on logs of Concentration

```
> cenanova (Thiamethoxam, ThiaCens, SamplingEvent)

MLE test of mean natural logs of CensData: Thiamethoxam by
Factor: SamplingEvent

Assuming lognormal distribution of CensData

Chisq = 146.3  on 3 degrees of freedom

 p = 1.64e-31
```
-- the mean logs (geometric means) differ

```
     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|)
2. Post-Plant - 1. Pre-Plant == 0    -0.2197     0.3348  -0.656    0.912
3. Corn Tassle - 1. Pre-Plant == 0   -3.3733     0.3823  -8.823   <0.001
     ***
4. Goldenrod - 1. Pre-Plant == 0     -4.4589     0.4932  -9.040   <0.001
     ***
3. Corn Tassle - 2. Post-Plant == 0  -3.1536     0.3795  -8.310   <0.001
     ***
4. Goldenrod - 2. Post-Plant == 0    -4.2392     0.4909  -8.636   <0.001
     ***
4. Goldenrod - 3. Corn Tassle == 0   -1.0857     0.5004  -2.169    0.128
```

Pre- A	Post- A	Corn Tassle B	Goldenrod B

13

13

MLE Multiple Comparisons

For each each comparison of two groups' mean logs, an estimate of difference (meanlog1 - meanlog2) is computed along with its standard error.

If zero is outside the confidence interval on that estimate, the two groups differ and p-values are below 0.05.

p-values are adjusted for the multiple k(k-1)/2 comparisons of k group mean logs that are made using Tukey's method

Here 4 of the 6 differences in mean logs are significant.

```
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|)
2. Post-Plant - 1. Pre-Plant == 0    -0.2197     0.3348  -0.656    0.912
3. Corn Tassle - 1. Pre-Plant == 0   -3.3733     0.3823  -8.823   <0.001
     ***
4. Goldenrod - 1. Pre-Plant == 0     -4.4589     0.4932  -9.040   <0.001
     ***
3. Corn Tassle - 2. Post-Plant == 0  -3.1536     0.3795  -8.310   <0.001
     ***
4. Goldenrod - 2. Post-Plant == 0    -4.2392     0.4909  -8.636   <0.001
     ***
4. Goldenrod - 3. Corn Tassle == 0   -1.0857     0.5004  -2.169    0.128
```

Pre- A	Post- A	Corn Tassle B	Goldenrod B

14

14

© PracticalStats.com

What if we wanted to run the ANOVA in original units?

```
> cenanova(Thiamethoxam, ThiaCens, SamplingEvent, LOG=FALSE)
 MLE test of mean CensData: Thiamethoxam   by Factor: SamplingEvent
       Assuming normal distribution of CensData
       Chisq = 55.02  on 3 degrees of freedom    p = 6.81e-12   unreliable

  NOTE: Data with nondetects may be projected below 0 with MLE normal distribution.
  If so, p-values will be unreliable (often too small).  Use perm test instead.

    mean(1. Pre-Plant)   mean(2. Post-Plant)   mean(3. Corn Tassle)   mean(4. Goldenrod)
            1.37561              2.708575             -7.348195             -10.83883
```

Use a permutation test (cenpermanova) instead:

15

© PracticalStats.com

cenpermanova -- permutation test with nondetects

```
> cenpermanova(Thiamethoxam, ThiaCens, SamplingEvent)
 Permutation test of mean CensData: Thiamethoxam   by Factor: SamplingEvent
      9999 Permutations
Test Statistic = 826.5 to 830.1        p = 0.0064 to 0.0073  (not 6.81e-12!)

mean(1.Pre-Plant)  mean(2.Post-Plant)   mean(3.Corn Tassle)   mean(4.Goldenrod)
      2.00300             3.38700             0.09991             0.04426
```

These are realistic !

16

© PracticalStats.com

2. Nonparametric Peto-Peto test (cen1way)

- Extends the Kruskal-Wallis test to censored data
- Scores (ordered values like percentiles or ranks) are computed for the detected data and separately, the censored nondetect data. No distribution assumed.
- From the scores are computed a log-likelihood, a joint (detects and nondetects) measure of error
- The log-likelihood of a model with group differences is compared to a null situation of no group differences
- If the group difference model has significantly lower error than the null model, the p-value is small, and the group assignment is explaining part of the variation of the data
- No values are substituted for nondetects!
- There are several similar tests which differ in the details of their score function. These include the Peto-Peto and Tarone-Ware tests. For skewed data which look something like a lognormal distribution these two versions have more power than a third version, named the logrank test

17

17

© PracticalStats.com

Computing the Peto-Peto test statistic

- Peto-Peto test is essentially a test for whether the cdfs are the same for all groups
- A Chisquare statistic is computed as -2*(difference in the two log likelihoods), $-2*(loglikelihood_{null} - loglikelihood_{model})$.
- The computed Chisquare statistic is compared to a chi-square distribution with k-1 degrees of freedom to determine the p-value
- If the Chisquare statistic (the signal of group differences) is small, the p-value is large (>0.05) and the group cdfs are not significantly different from one another.
- If the Chisquare statistic is large, the p-value is small (<0.05) and at least one group cdf differs from the others.
- This is an overall test -- no statement of which group cdfs differ from the others is made. That is done with multiple comparisons.

18

18

Peto-Peto test of Difference in Group Concentration Percentiles

```
> cen1way (Thiamethoxam, ThiaCens, SamplingEvent)
    Oneway Peto-Peto test of CensData: Thiamethoxam   by Factor: SamplingEvent
    Chisq = 127   on 3 degrees of freedom      p = 2.35e-27

  Pairwise comparisons using Peto & Peto test
data:  CensData and Factor
              1. Pre-Plant 2. Post-Plant 3. Corn Tassle
2. Post-Plant  0.416         -             -
3. Corn Tassle 6.5e-15       6.5e-15       -
4. Goldenrod   6.5e-15       7.1e-15       0.055
```

Pre-A	Post-A	Corn Tassle B	Goldenrod B

19

19

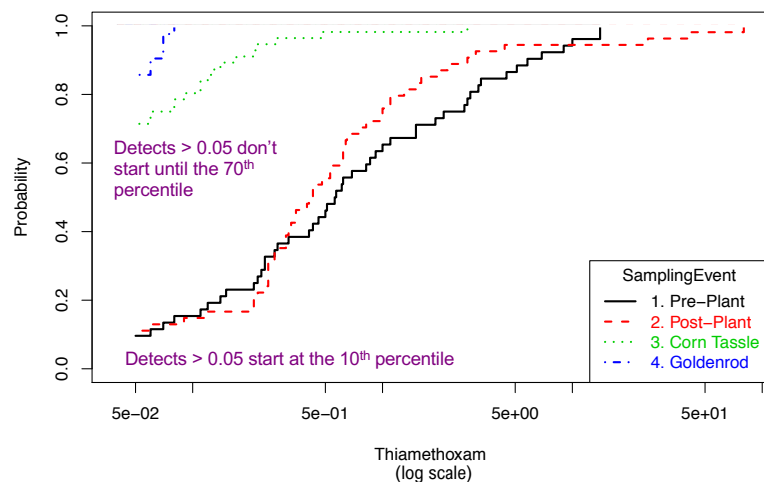# Graph the Data: Sample CDFs Incorporating Nondetects

```
> cen_ecdf (log(Thiamethoxam), ThiaCens, SamplingEvent)
```

Data shown as step function.

CDF is a plot of quantiles.
Probability = 0.5 is a median, etc.

Higher quantiles for a given probability → cdf plots to the right.

Pre- and Post- Plant are similar. All other comparisons appear different.

Detects > 0.05 don't start until the 70th percentile

Detects > 0.05 start at the 10th percentile

SamplingEvent
1. Pre–Plant
2. Post–Plant
3. Corn Tassle
4. Goldenrod

Thiamethoxam (log scale)

20

20

© PracticalStats.com

Saw this before in the boxplots of logs

```
> cboxplot(Thiamethoxam, ThiaCens, SamplingEvent, Ylab =
  "Thiamethoxam, in ppb", Xlab = "Sampling Event", show = TRUE,
  LOG=TRUE)
```


21

© PracticalStats.com
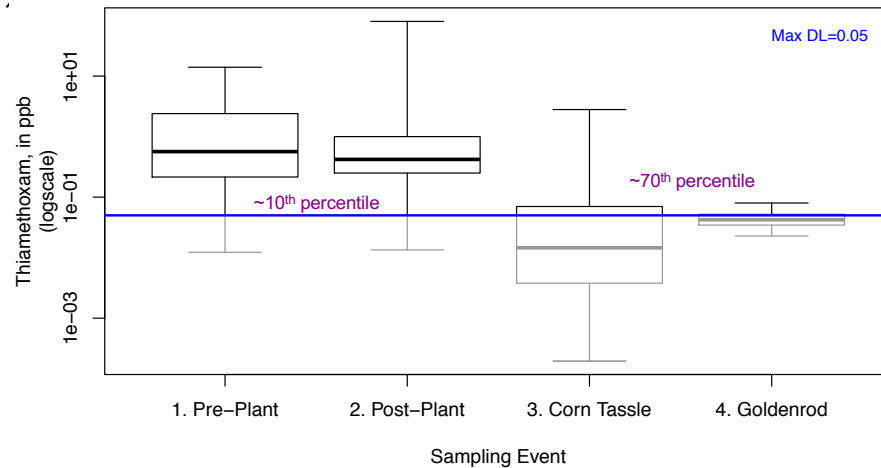
3. Using Simpler Nonparametric Tests

Re-censor to highest DL and use standard nonparametric tests

For DLs of 0.05 and <0.10, all data below 0.10 (<0.10s, <0.05s and detected 0.051 to 0.099) are set to <0.10

A. Categorize the data to either "< highest DL" or "≥ highest DL" and compute a contingency table test

B. Re-censor to highest DL and test differences using the Kruskal-Wallis test

22

© PracticalStats.com

3A. Contingency Tables

Does the % ≥ highest (or single) DL differ between groups?

Categorize data to either "< DL" or "≥ DL". If more than 1 RL, use the highest detection limit (here, 0.05)

23

© PracticalStats.com

3A. Test for difference in % detects with a contingency table test

```
> head(Pollen_Thiamethoxam)
  Thiamethoxam ThiaCens  SamplingEvent ThiaAbvBelow
1        0.05        1 3. Corn Tassle        Below
2        0.05        1 3. Corn Tassle        Below
3        0.86        0  1. Pre-Plant        Above
4        0.05        1 3. Corn Tassle        Below
5        0.05        1   4. Goldenrod        Below
6        1.00        0  1. Pre-Plant        Above
> ThiaAbvBelow <- ThiaCens
> ThiaAbvBelow[Thiamethoxam < 0.05] <- 1
> ftable(SamplingEvent~ThiaAbvBelow)
             1. Pre-Plant 2. Post-Plant 3. Corn Tassle 4. Goldenrod
ThiaAbvBelow
    0              47            48            16            6
    1               5             6            40           36

> tab.Thia= xtabs(~ThiaAbvBelow+SamplingEvent)
> chisq.test(tab.Thia)
        Pearson's Chi-squared test
data:  tab.Thia
X-squared = 95.968, df = 3, p-value < 2.2e-16
```
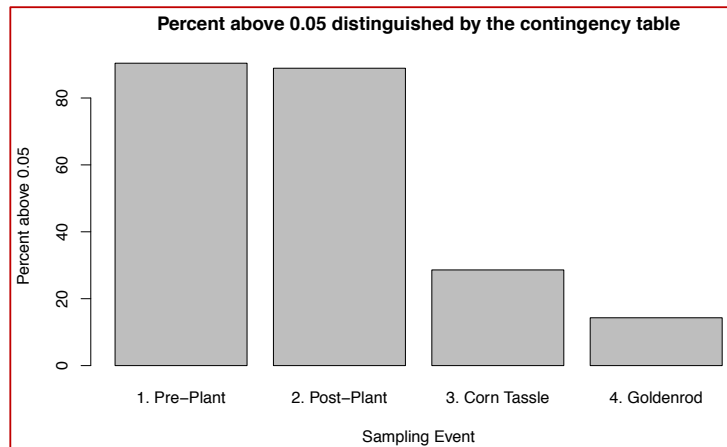
≥ 0.05

< 0.05

24

3A. Bars of % above highest DL to illustrate the contingency table test

```
> Gps <- levels(SamplingEvent)
> Pct.Abv <- c( 90.4, 88.9, 28.6, 14.3)
> barplot(Pct.Abv, names.arg = Gps, ylab = "Percent above 0.05", xlab = "Sampling Event",
  main = "Percent above 0.05 distinguished by the contingency table")
```


25

3B. Re-censor at the highest DL & run a rank-based nonparametric test

- For 3+ groups this is the Kruskal-Wallis test.
- All nondetects are tied, will be represented by a tied rank lower than the ranks of detected data.
- If multiple DLs, re-code data below highest RL with the same number. Number must be below the highest DL and all the same.
- I use -1 for all values below the (highest or single) detection limit.
- Test for differences in ranks of data.

Advantages:
- – No fabrication
- – No assumption of distribution

26

© PracticalStats.com

3B. Standard nonparametric tests

```
> Below05 <- Thiamethoxam
> Below05 [ThiaCens == 1] <- -1
> Below05 [Thiamethoxam < 0.05] <- -1
> kruskal.test(Below05~SamplingEvent)

        Kruskal-Wallis rank sum test

data:  Below05 by SamplingEvent
Kruskal-Wallis chi-squared = 114.04,
df = 3, p-value < 2.2e-16
```

Below 05: Concentrations + DLs to start with

Sets all censored values to -1

Sets all detects below highest DL (0.05) to -1

27

© PracticalStats.com

Conclusions
Three+ group tests for censored data

Substitution often gives wrong results! Whether wrong or right, you'll never be sure.

Instead, for simple yet effective methods:

1. re-censor at highest RL and run the binary contingency table test, or

2. re-censor at highest RL and run standard nonparametric methods (Kruskal-Wallis test)

Using survival analysis methods:

1. use censored regression (parametric) for an MLE "ANOVA"   cenanova

2. use nonparametric Peto-Peto test   cen1way

28

Conclusions

1. Hypothesis test methods are available for censored data for both 1 and multiple DLs <u>without</u> substituting fabricated values for nondetects
2. Use parametric methods if you want to test for differences in means, and you feel confident that a particular distribution fits the data well (Note: transformations change the meaning of a mean)
3. Use nonparametric methods if you want to test for whether some groups have generally higher values than others (different percentiles). No distribution needs to be assumed.
4. Multiple comparison tests are available for both types of methods.
5. For simpler methods, re-censor to the highest DL (or use the 1 DL if that's all you have) and run either contingency tables or the KW test.

29

29

Methods for censored data

Method	Parametric	Nonparametric
Descriptive stats	MLE ROS	Kaplan-Meier
Intervals	Bootstrapping MLE	Bootstrapping K-M
Paired Data	CI on differences by MLE	PPW
2 Indep Groups	MLE Regression on 0/1 Factor	Peto-Peto
3+ Indep Groups	MLE Regression on 0/1 Factor cenanova	Peto-Peto cen1way
Correlation	Likelihood R by MLE	Kendall's tau
Regression	MLE Regression	ATS line

30

30