Nondetects And Data Analysis:
Two-Group Tests with NDs

Dennis R. Helsel, Ph.D
Practical Stats

1

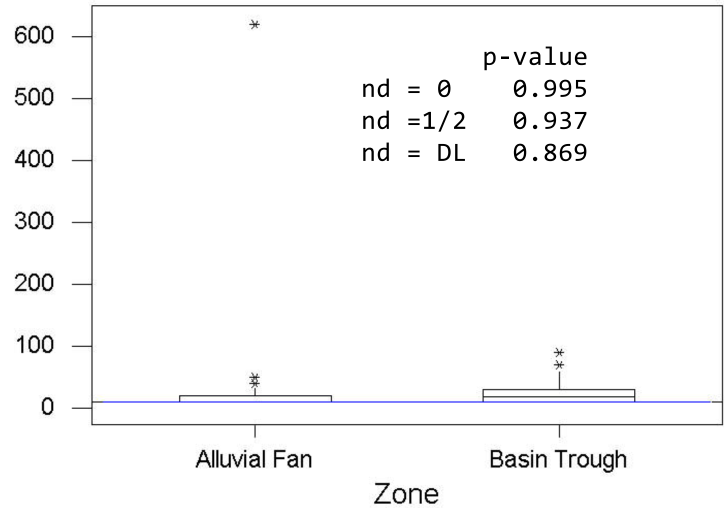
Methods for Testing Differences Between Groups

1. Parametric. Tests differences in group means - "Does one group have a higher mean than another group?" You must designate the assumed distribution that best matches the shape of your data.
2. Nonparametric. Tests differences in percentiles - "Is one group shifted higher than another?" No shape is assumed or necessary.
3. Simpler tests. Convert data to above or below a single, or the highest of multiple, DLs. Nonparametric -- no shape is assumed or necessary.

2

2

What NOT to do: t-Tests by substitution



No differences found, yet there are differences between these two groups.

Max DL= 10

3

3

Parallels between standard and censored data methods

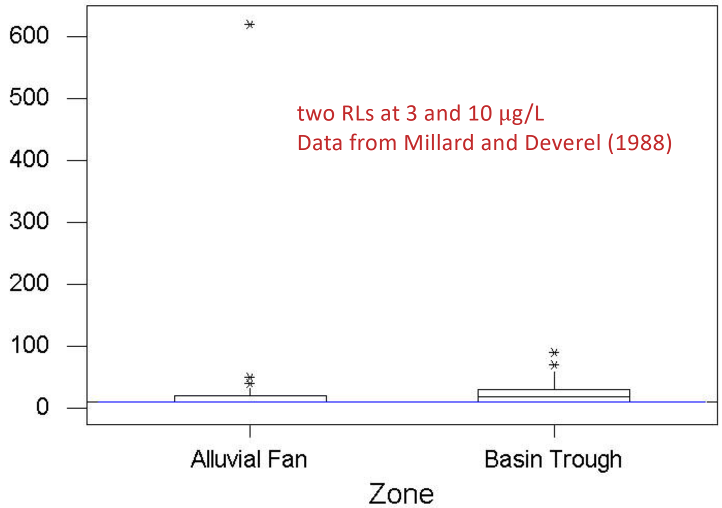
Comparing Two Groups

Standard Methods	Methods for Censored Data
t-test	Censored MLE regression with 0/1 group indicator
Wilcoxon rank-sum test	Peto-Peto (or similar) tests

4

4

Censored data methods use detected values between multiple RLs



Test while keeping detected 3, 4, 5, 6, 7, 8, 9 as detects.

Max DL= 10

5

5

1. Parametric: "t-test" by MLE regression

Regression of concentrations versus group id

Concentration = intercept + slope • Factor
where Factor = 0 for group A

Factor = 1 for group B

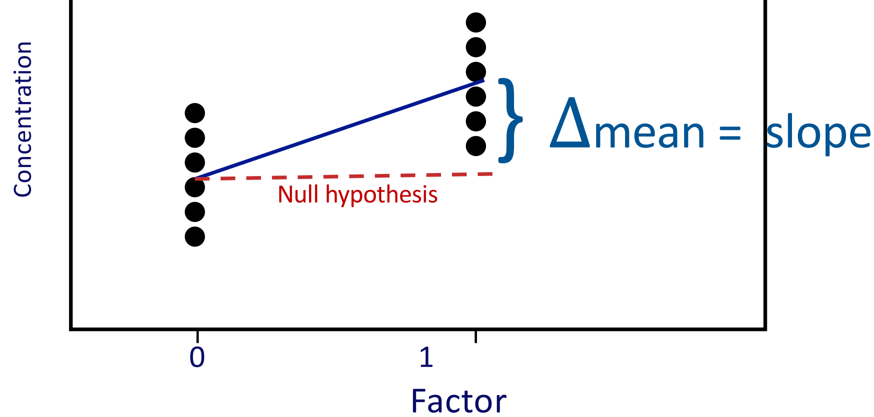
The slope is fit by Maximum Likelihood Estimation

Test for Factor slope = 0 is a test for difference between the means of the two groups. If the p-value is small, there is a significant difference between the two means

6

6

Parametric:
"t-test" by regression



7

7

"t-test" by MLE regression

- These tests are parametric (MLE), assuming a distributional shape. Residuals on the original scale or after taking logarithms are assumed to follow a normal distribution.
- The slope is computed incorporating both the detected values as well as the proportion of the dataset below each detection limit.
- If the distributional assumption is incorrect, the tests will have low power. Power is the ability to see differences when they exist. For parametric tests, violation of the assumptions increases the chance of not seeing differences that are present.

8

8

Data: Zinc concentrations in central California

- Zinc and other trace element concentrations were measured in groundwaters of two geochemical zones (based on geology) of central California.
- Two detection limits, at 3 and 10 ug/L were used by the laboratory.
- We'll perform a 2-group 't-test' using MLE regression, both on the original scale and on the logs of the concentrations.
- Based on the Q-Q plots produced for the residuals from the regression, which scale best fits the requirement of a normal distribution for this parametric test?
- Then follow up with a 2-group nonparametric test

Source of the dataset: *Millard, S.P. and S.J. Deverel, 1988, Water Resources Research 24, p. 2087-2098.*

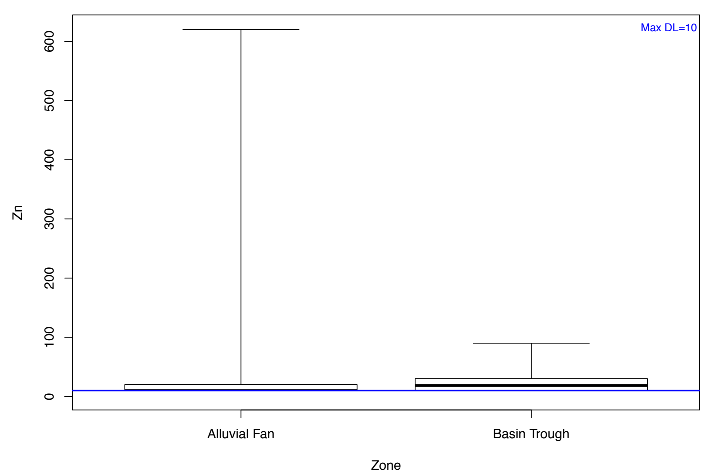
9

9

© PracticalStats.com

Boxplots for the Zinc data

- There is one large outlier in the Alluvial fan data.
- This will disrupt the "t-test"
- Logs would be a better scale to meet the test's requirements, but we'll do the test on both scales.
- Basin Trough box appears to be elevated in comparison to the Alluvial Fan (a percentile evaluation)




```
> cboxplot (Zn, ZnLT, Zone, minmax = TRUE)
```

10

10

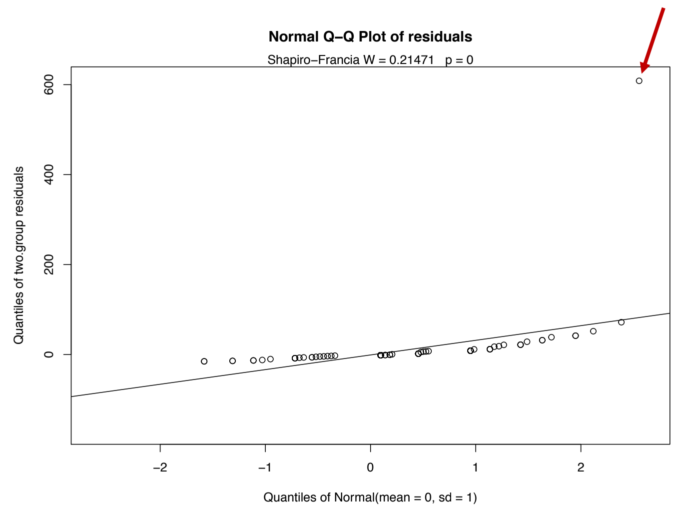
"t-test" by MLE regression

```
> cen2means (Zn, ZnLT, Zone, LOG=FALSE)

MLE 't-test' of mean CensData: Zn
by Factor: Zone

Assuming normal distribution of
residuals around group means
Chisq = 0.2928  on 1 degrees
 of freedom     p = 0.588
```

The assumption of a normal distribution is rejected.



11

11

Another issue with MLE tests assuming normality -- "data" go below zero

```
> cen2means(Zn, ZnLT, Zone, LOG=FALSE)
    MLE 't-test' of mean CensData: Zn   by Factor: Zone
    Assuming normal distribution of residuals around group means
    mean of Alluvial Fan = 11.49     mean of Basin Trough = 18.13
    Chisq = 0.2928  on 1 degrees of freedom     p = 0.588
```

Another problem of MLE tests assuming a normal distribution is that data (and sometimes even their means!) can be modeled as below 0 by MLE. This lowers the means from what they should be, and also makes the p-values unreliable, as the differences between groups can be exaggerated compared to what actually occurs -- data that are positive very close to 0 are modeled as being below 0. Standard deviations may also be exaggerated. MLE p-values may therefore be too high or too low with MLE assuming a normal distribution.

Instead,

1. use a permutation test (cenperm2 function)
2. work in log units, testing for a difference in geometric means

12

12

Use a permutation test instead of assuming normality with an MLE test

```
> cenperm2(Zn, ZnLT, Zone)
 Permutation test of mean CensData: Zn   by Factor: Zone
     9999 Permutations     alternative = two.sided
mean of Alluvial Fan = 21.22 to 23.51   mean of Basin Trough = 21.28 to 21.94
Mean(Alluvial Fan - Basin Trough) = -0.056 to 1.567   p = 1 to 0.998
```

These means are believable, as are the p-values.

MLE means were not:

```
 mean of Alluvial Fan = 11.49     mean of Basin Trough = 18.13
```

13

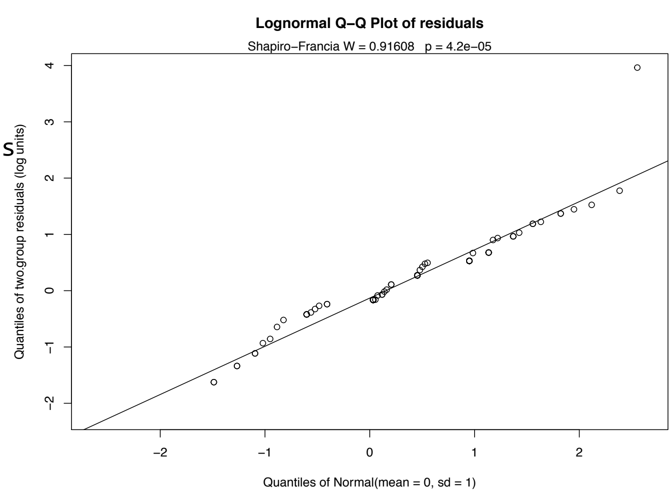
13

"t-test" on logs by MLE regression

```
> cen2means (Zn, ZnLT, Zone)

MLE 't-test' of mean natural logs of
CensData: Zn by Factor: Zone

Assuming lognormal distribution of
residuals around group geometric means
Chisq = 2.547  on 1 degrees of
freedom        p = 0.11
```



The Shapiro-Francia test for (log)normality shows this is a better set of units. After taking logs, differences in the group geometric means is being tested. Outlier is still inflating one mean.

14

14

2. Nonparametric tests for censored data

These nonparametric tests extend the Wilcoxon rank-sum test to censored data.

Scores are ranks adjusted for censoring - essentially are K-M percentiles

R version is called the "Peto-Peto" test.

Peto, R. and J. Peto, 1972, Asymptotically efficient rank invariant test procedures (with discussion): Journal of the Royal Statistical Society, Series A 135, 185–206.

Similar tests called Peto-Prentice, HF1, Generalized Wilcoxon, Tarone-Ware and Gehan.

Use the cen1way command in NADA2.

15

15

Nonparametric tests for censored data

Peto-Peto test is essentially a test for whether the survival curves (KM percentiles) are the same for all groups

```
> cen1way (Zn, ZnLT, Zone)

     Oneway Peto-Peto test of CensData: Zn   by Factor: Zone
     Chisq = 5.183   on 1 degrees of freedom     p = 0.0228
```

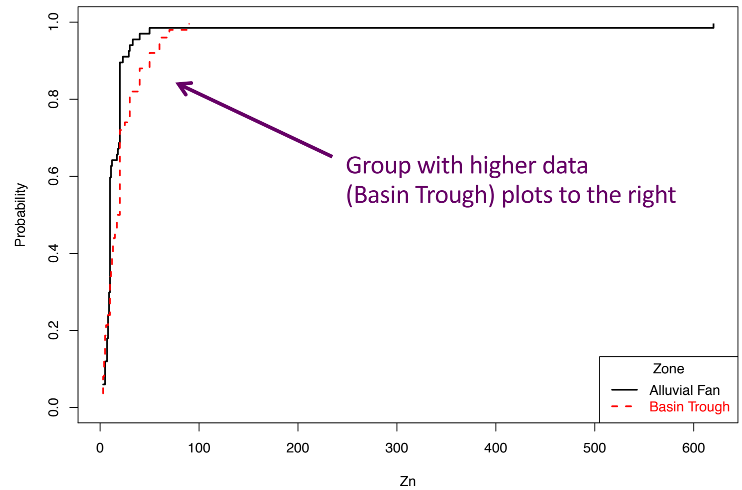
Non-normality is irrelevant for this nonparametric test.
Differences between cdfs (percentiles) of the two groups is significant.

16

16

Wilcoxon tests for censored data

```
> cen_ecdf(Zn, ZnLT, Zone)
```

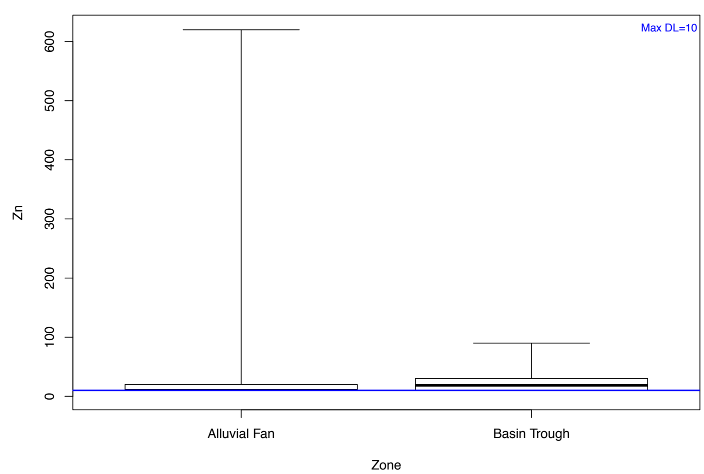


17

17

Saw this easily before in the boxplots

- The Peto-Peto test finds a difference in the percentiles of the two groups.
- Basin Trough box is significantly elevated in comparison to the Alluvial Fan (a percentile evaluation)




```
> cboxplot (Zn, ZnLT, Zone)
```

18

18

Some software also prints out the log-rank test for censored data. Why?

Wilcoxon-style tests such as the Peto-Peto test "weight the scores by a function of n, the number of detects at each value" -- It essentially uses a modified rank of the data to compute the test statistic. Larger modified ranks are assigned to larger numbers.

The log-rank test (which we don't use) uses a different approach more similar to a contingency table test, not "weighting the scores by a function of n" – using counts rather than ranks. It is more commonly used in medical statistics than the Wilcoxon-style tests, but is less powerful for skewed data and so less applicable to environmental sciences.

19

19

3. Simpler Methods

For data with 1 RL.

If more than 1 RL, set all observations below the highest RL as a <RL .

Then can use ordinary nonparametric tests.

For example:

```
  Data:  <1  <1   3  <5   7   8   8   8  12  15  22
Set to:  <5  <5  <5  <5   7   8   8   8  12  15  22
 Ranks: 2.5 2.5 2.5 2.5   5   7   7   7   9  10  11
```

Ties get average ranks: tie of ranks 1, 2, 3, 4 tie of ranks 6, 7, 8

20

20

3A. Contingency Tables

Do %detects differ between groups?

Categorize data to either "< DL" or "≥ DL". If more than 1 RL, use the highest detection limit (here, 5)

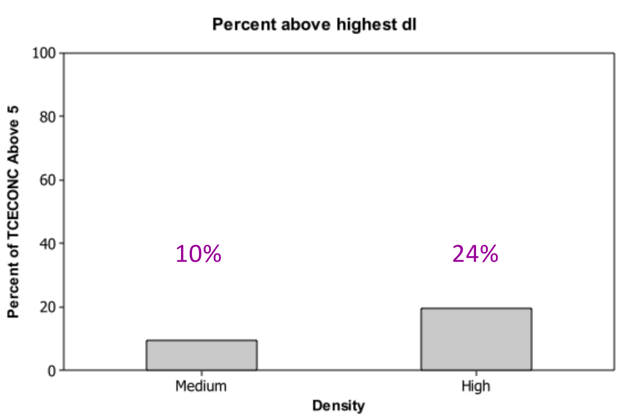
For example:

```
   Data: <1 <1  3 <5  7  8  8  8 12 15 22
Set to: <5 <5 <5 <5 ≥5 ≥5 ≥5 ≥5 ≥5 ≥5 ≥5
```

21

21

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




```
load("TCE2.RData")
> attach(TCE2)
> ftable(Density~Below5Cens)
           Density High Medium
Below5Cens
0                    18     12   >=5
1                    74    118   <5

> tab= xtabs(~Below5Cens+Density)
> chisq.test(tab)

Pearson's Chi-squared test with Yates'
continuity correction

data:  tab
X-squared = 4.0785, df = 1,
p-value = 0.04343
```

22

22

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

For 2 groups, use the Wilcoxon rank sum 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
- Results are unequivocal. No argument.
- No assumption of distribution

23

23

3B. Standard nonparametric tests

Nondetects are ranked lower than detected values, and tied with each other at rank $(n_T + 1)/2$

```
For 1 RL:  -1  -1  3  5  7  8  8  8 12  15  22
    rank: 1.5 1.5  3  4  5  7  7  7  9  10  11
```

If more than 1 RL, tie all observations below the highest RL at the lowest rank

```
original:  <1  <1   3   5   7   8 <10 <10  12  15  22
recode:    -1  -1  -1  -1  -1  -1  -1  -1  12  15  22
rank:     4.5 4.5 4.5 4.5 4.5 4.5 4.5 4.5   9  10  11
```

24

24

3B. Standard nonparametric tests

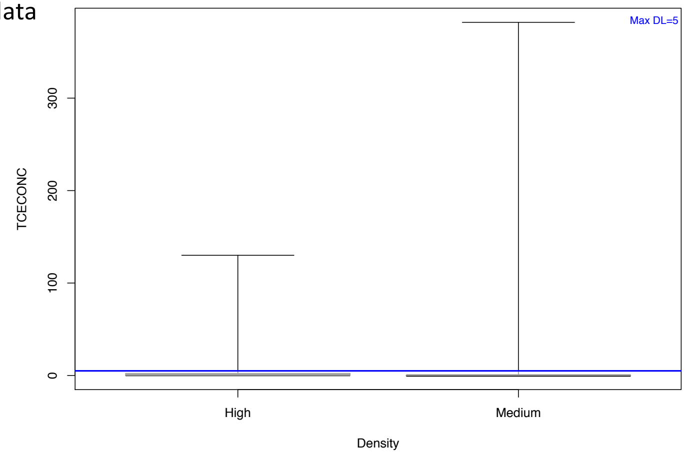
The TCE2 data

```
> cboxplot (TCECONC, TCECens, Density)

> wilcox.test (Below5~Density)
Wilcoxon rank sum test with continuity
correction
data:  Below5 by Density
W = 6599.5, p-value = 0.02713
```

compare to:

```
> t.test (Half.DL~Density)
          Welch Two Sample t-test
t = -0.065623, df = 201.88,
p-value = 0.9477
```



25

25

Conclusions
Two-sample 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 (Wilcoxon rank-sum test)

Using survival analysis methods:

1. use censored regression (parametric) for a "t-test"   cen2means
2. use nonparametric Peto-Peto test   cen1way

26

26

Methods for censored data

Method	Parametric	Nonparametric
Descriptive stats	MLE	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 cen2means	Peto-Peto cen1way
3+ Indep Groups	MLE Regression on 0/1 factor	Peto-Peto
Correlation	Likelihood R by MLE	Kendall’s tau
Regression	MLE Regression	ATS line

27

27