

How Many Observations Are Censored Data Worth?

Dennis R. Helsel

PracticalStats.com



© 2020 PracticalStats.com

1

Topics of this webinar

1. Don't give up when there is a high percent of nondetects
2. Don't make up rules for data with high detection limits. Let the data decide.
3. Learn how to compute the 'worth' of censored observations.



© 2020 PracticalStats.com

2

2

For More Information

on this topic and other methods for data analysis with nondetects, see our online course



Nondetects And Data Analysis
at <https://practicalstats.teachable.com>

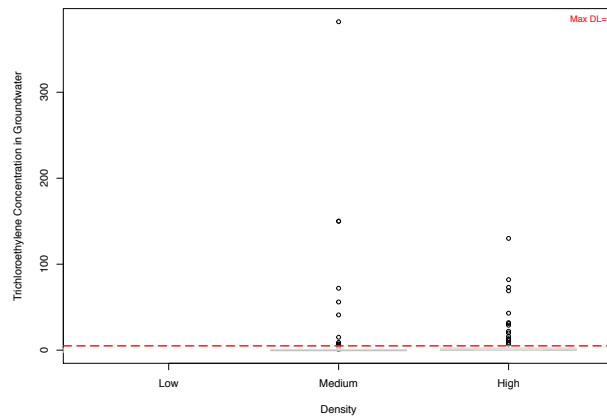


© 2020 PracticalStats.com

3

3

1. Don't Give Up With a High% of Nondetects



Example: TCE concentrations in GW

- Reporting limits at 1, 2, 3, 4, and 5 ug/L. ~ 80% censoring
- Do TCE distributions differ among the three land-use groups?



© 2020 PracticalStats.com

4

4

There is still much confusion over the value of nondetects

USEPA (RCRA Guidance, 1989) said to substitute DL/2 when there were fewer than 15% NDs.

USEPA (Guidance for Data Quality Assessment: Practical Methods for Data Analysis, 1998) said to substitute DL/2 when there were fewer than 15% NDs.

USACE (Environmental Statistics Manual, 2008) said to use the test of proportions (a contingency table test) for data with 50-90% nondetects.

USEPA (Unified Guidance, 2009) said to use nonparametric methods when there were greater than 50% NDs.

© 2020 PracticalStats.com
5

5

Substitution ruins the possibility of interpreting your data correctly

ANOVA sub <5 = 0
p = 0.57

ANOVA sub <5 = 2.5
p = 0.56

ANOVA sub <5 = 5
p = 0.50

No differences found!

Max DL=5

See our

VIDS 4

Why Not Sub DL/2
For Nondetects?

Why Not Just Substitute DL/2 For Nondetects? 1. part of PowerPoint slides

for more information

Substitution is Fabrication

© 2020 PracticalStats.com
6

6

Use Better Tests For Censored Data

```
> cen1way (TCECONC, BDL_1, Density)
      n n.cen median  mean  sd
Factor=High  92   58   <5  7.778019 19.7826151
Factor=Low   25   23   <5  2.083333  0.5768489
Factor=Medium 130  113  <5  7.867264 39.7664396
```

Oneway Peto-Peto test of CensData: TCECONC by Factor:
Density

Chisq = 16.25 on 2 degrees of freedom p = 0.000295

Distribution of TCE differs
between land-use groups

Pairwise comparisons using Peto & Peto test

	High	Low
Low	0.0098	-
Medium	0.0024	0.2796

Shows which groups differ from others

Nonparametric Test allowing Multiple Detection Limits
No Substitution!
from our Nondetects And Data Analysis course

© 2020 PracticalStats.com

7

7

2. Don't make up rules for data with high detection limits

- Some organizations have realized that data with high detection limits distort results if $DL/2$ is substituted.
- For detected values in the range of 0.5 - 6, for example, a detection limit of <50 for one observation (due to 10x dilution of the sample?) would result in $DL/2 = 25$, distorting the mean and standard deviation.
- However people have defined arbitrary thresholds above which they throw out the 'high nondetects'.
- Typical thresholds are when 'high nondetects' are at detection limits 2 - 3 times the highest detected value. I have also seen 80% of the highest detected value used. Data above those thresholds are deleted.

© 2020 PracticalStats.com

8

8

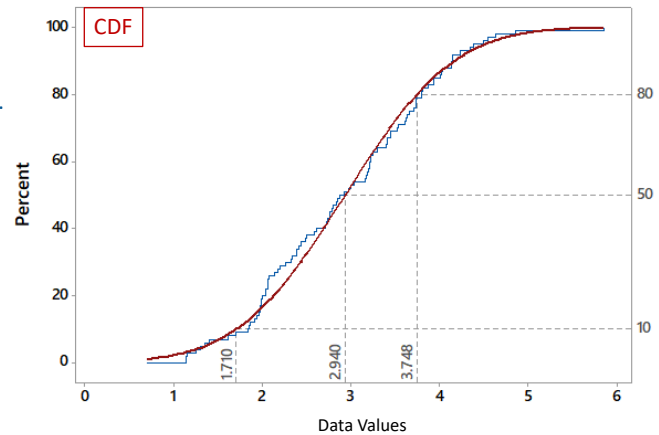
What if a nondetect is higher than all detects? <10

A nondetect higher than all detects (<50) has a DL at the 100th percentile

All that is known is that the value is somewhere between 0th and 100th percentile. Cannot be added to the step function.

This provides no information of whether that <50 is higher or lower than any of the data. Therefore, this 'high nondetect' contains no information for this dataset.

This would also be true for a <8, <10, <30 etc.



9

Information in high nondetects

- The information in nondetects is contained in the percentiles the detection limits are at in the dataset
- If a detection limit is higher than all the detected values, there is no way to assign a percentile to the detection limit. It is anywhere between the 0th and 100th percentile. Noninformative.
- Therefore data with a detection limit higher than the highest detected value has 0 information content and can be discarded.
- Censored data methods such as Kaplan-Meier and Maximum Likelihood correctly evaluate this, so their results will be identical whether those data are included in the dataset or not. They don't 'throw them away', just give them their appropriate weight, which is 0.
- No need for arbitrary rules when censored data methods are used.



10

3. What is the 'worth' of censored observations?

This topic is based on a presentation given by Dr. Brenda Gillespie of the University of Michigan at the 2019 American Statistical Association conference in Denver, CO.

<https://ww2.amstat.org/meetings/jsm/2019/onlineprogram/AbstractDetails.cfm?abstractid=301776>



11

The 'worth' of censored observations depends on the question asked

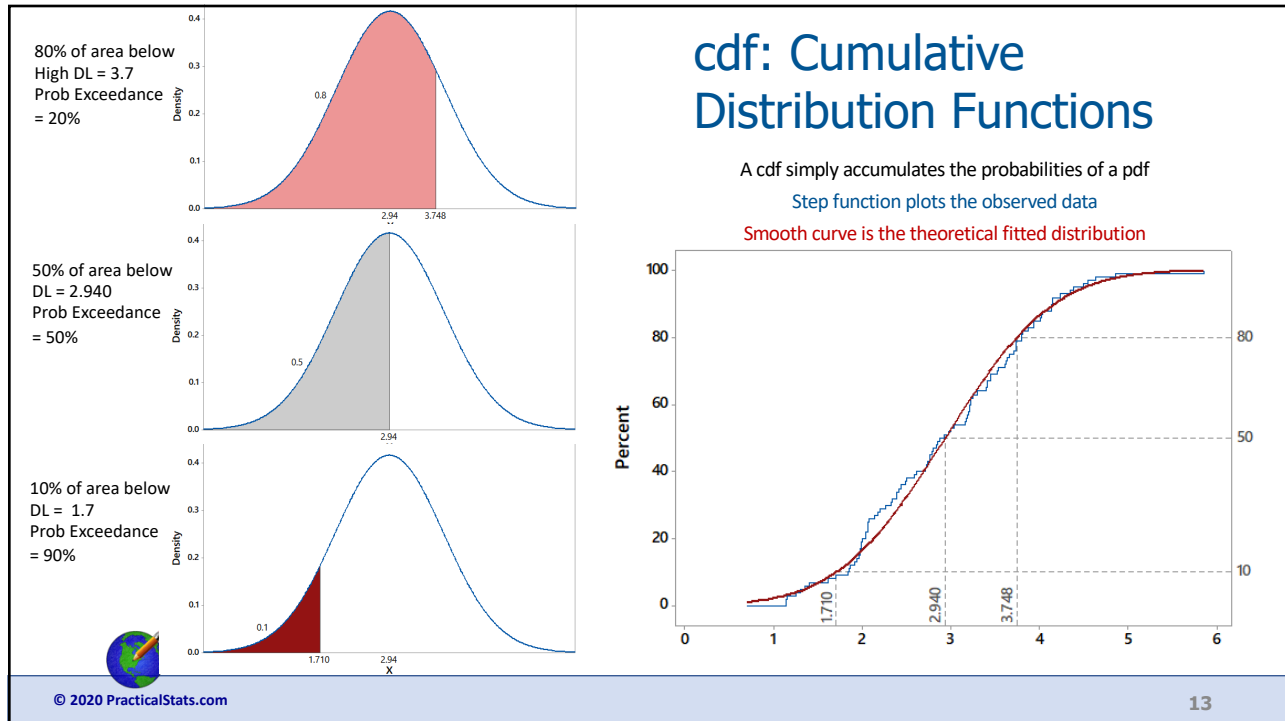
When the question is regression, maximum likelihood methods use the percent of observations below each detection limit (DL) to determine slope and intercept. For a smaller percent of observations below a DL, the larger the information available.

When the question is 'do groups differ in value?' using nonparametric methods, a smaller percent of observations below a DL produces more information in comparison to other observations.

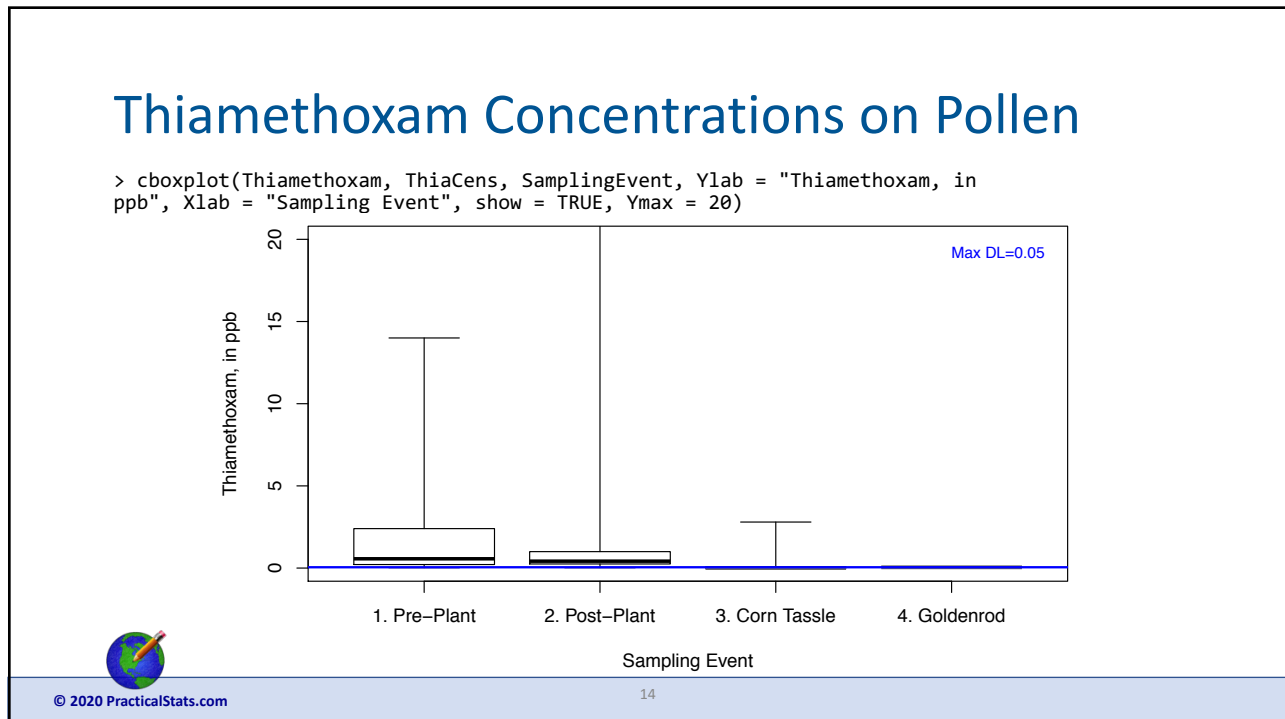
Therefore the amount of information in a nondetect is measured by the probability of exceedance of the detection limit. The lower the detection limit the smaller the percent of observations below, and the larger the probability of exceedance above. (An increase can be seen going from a <3 to a detected 10 but not from a <20 to the same detected 10.)



12



13



14

The 'worth' of censored observations in MLE and nonparametric procedures


```

> equivalent_n(Thiamethoxam, ThiaCens)
Thiamethoxam
all:
  n      n.cen  pct.cen  min  max
204.00000 87.00000 42.64706 0.05000 80.00000

limits:
  limit n uncen  pexceed
1 0.05 87 117 0.5735294

equivalent sample size:
  n.equiv  n.cen.equiv  n.detected
166.8971  49.89706    117
    
```

The "equivalent n" will increase as the detection limit gets smaller




© 2020 PracticalStats.com 15

15

Objective: Comparing Data to a Standard

- Different 'worth' so a different and binary equivalent n.
- Any NDs less than the standard have an equivalent n (eq_n) = 1:

<0.05	Standard = 1	obs is below standard	eq_n = 1
detected 0.18	Standard = 1	obs is below standard	eq_n = 1
< 5	Standard = 1	cannot say obs <, > std	eq_n = 0



© 2020 PracticalStats.com 16

16

Summary: How Many Observations Are Censored Data Worth?

1. Don't give up when there is a high percent of nondetects.
Use censored data methods instead of arbitrary cutoffs for "usefulness"
2. Don't make up rules for data with high detection limits. Let the data decide.
Censored methods recognize that 'high nondetects' have no information. No need to delete them. No need for arbitrary rules for deletion.
3. Learn how to compute the 'worth' of censored observations.
Censored data have partial n for continuous scale questions such as group tests. They have $0/1 n$ for comparison to standards.



17

Next Video

Tuesday August 18th 2020 at 11 am Mountain Time

State of The Stats Address:

What Can Be Done With Nondetects?

An overview of everything that you can now do for data with nondetects, including estimating summary statistics, comparing to standards, regression, trend analysis and more, all without substituting values like $DL/2$. Invite the people you know that have no idea of what can be done without fabricating numbers for nondetects to join us.

A link to the livestream will be on our website and sent in the August newsletter.

- Sign up for our newsletter/announcement list to get the announcement and description of content emailed to you. <https://practicalstats.com/news/>



18

view at:


© PracticalStats.com 2020

Our other videos on methods for nondetects available now for streaming

at no charge on our Online Training Center
<https://practicalstats.teachable.com/>

Let colleagues know about them.

Current free videos on Stats with NDs at our Training Center:



© 2020 PracticalStats.com

<p style="font-size: 2em; color: purple;"><1</p> <p style="font-size: 0.8em;">Introduction to Nondetects And D...</p> <p style="font-size: 0.7em;">\$0 sales 276 enrolled</p>	<p style="font-size: 2em; color: purple;"><2</p> <p style="font-size: 0.8em;">Fitting Distributions to Data with ...</p> <p style="font-size: 0.7em;">\$0 sales 68 enrolled</p>	<p style="font-size: 2em; color: purple;"><3</p> <p style="font-size: 0.8em;">Testing Group Differences w/NDs</p> <p style="font-size: 0.7em;">\$0 sales 44 enrolled</p>
<p style="font-size: 2em; color: purple;"><4</p> <p style="font-size: 0.8em;">The Mystery of Nondetects</p> <p style="font-size: 0.7em;">\$0 sales 36 enrolled</p>	<p style="font-size: 2em; color: purple;"><5</p> <p style="font-size: 0.8em;">Correlation and Regression for D...</p> <p style="font-size: 0.7em;">\$0 sales 18 enrolled</p>	<p style="font-size: 2em; color: purple;"><6</p> <p style="font-size: 0.8em;">Trend Analysis for Data w/ NDs</p> <p style="font-size: 0.7em;">\$0 sales 5 enrolled</p>

Thank you for viewing


- More on methods for nondetects can be found in my book [Statistics For Censored Environmental Data](#) by Dennis Helsel (2012).
- All opinions are my own and do not represent those of anyone else.

Questions about this material or our courses?

Get in touch!

Dennis Helsel ask@practicalstats.com

Courses & free recordings at: <http://practicalstats.teachable.com>



© 2020 PracticalStats.com