Matched Pair Tests with Nondetects

Dennis R. Helsel

PracticalStats.com

1

Objectives for this webinar

1. Inform you of the challenges of computing statistical tests for paired data with nondetects
2. Inform you of what methods are available for paired data with nondetects.

 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

 3

3

Matched - Pairs Tests

Group X		Group Y
Observation 1	↔	Observation 1
Observation 2	↔	Observation 2
Observation 3	↔	Observation 3
Observation 4	↔	Observation 4

There is a direct relationship between observations in the same row

 4

4

Pairing in Environmental Studies

Most commonly by time or location

"Blocks"

	Urban	Ag
Jan		
Feb		
March		
April		

Tests are run by computing the differences between the two values in the same row. No differences computed between values in different rows.

5

5

Pairing in Environmental Studies

Most commonly by time or location

"Blocks"

	Old Method	New Method
Well 1		
Well 2		
Well 3		
Well 4		

Upgradient/Downgradient, Pre-event/Post-event are other common paired columns.

6

6

Matched Pair Tests for censored data

Test	Distributional Assumption?	Tests for _____ of Paired Differences
Censored Paired	Parametric	mean
Sign Test	Nonparametric	median
Signed - Rank	Nonparametric	percentiles
PPW	Nonparametric	percentiles

Example: Matched Pairs

- Mercury in soils was measured at the same sites in 1996 before a major fire, and after the fire in 2001 (Eppinger et al. , 2003).
- Measurements were 'blocked' by location. This minimizes causes of change other than the difference between the two years, which is attributed to the effect of the fire.

Q1: Is the mass of mercury over the site before the fire different than after the fire (a two-sided test on mean of paired differences)? Parametric

Q2: Are concentrations consistently higher or lower after the fire (a two-sided test on the percentiles of differences)? Nonparametric

The Data

Hg in Soils before (X1996) and after (X2001) a major fire.

"Stacked" format

```
> head(SedHg)
 Mercury  Year CenHg
   0.020 X1996     1
   0.021 X1996     0
   0.020 X1996     1
   0.020 X1996     1
   0.026 X1996     0
   0.020 X1996     1
   ...
```

"Unstacked" format

```
> head(EppsedHg)
 X2001 X1996 Cens01 Cens96
  0.02 0.020      1      1
  0.02 0.021      0      0
  0.05 0.020      0      1
  0.02 0.020      0      1
  0.02 0.026      1      0
  0.02 0.020      0      1
  ...
```

 9

9

Interval-censored differences

(one of the challenges)

Pattern of censored concentrations ("<" added to view) results in interval-censored paired differences (mindiff, maxdiff):

X2001	Cens01	X1996	Cens96	mindiff	maxdiff
<0.02	TRUE	<0.020	TRUE	-0.020	0.020
0.02	FALSE	0.021	FALSE	-0.001	-0.001
0.05	FALSE	<0.020	TRUE	0.030	0.050
0.02	FALSE	<0.020	TRUE	0.000	0.020
<0.02	TRUE	0.026	FALSE	-0.026	-0.006
0.10	FALSE	<0.020	TRUE	0.080	0.100

 10

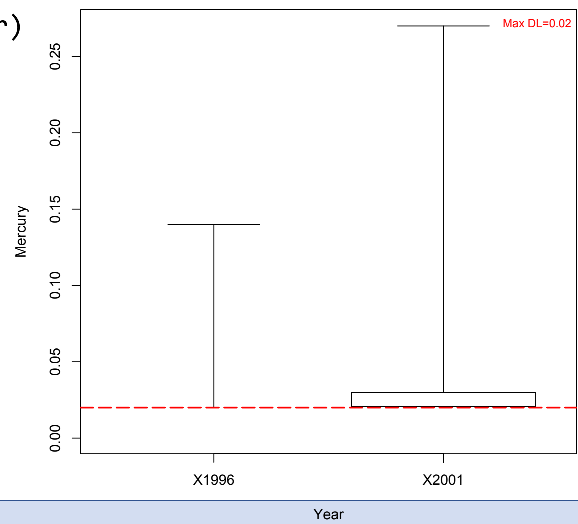
10

Individual Boxplots of The Two Columns of Data

```
> cboxplot(Mercury, CenHg, Year)
```

Group boxplots DO NOT show the effect of pairing of samples at the same site.

The 2001 data appear to be higher than 1996 in the upper percentiles



11

11

The Sign Test

Null Hypothesis: Probability [2001 Hg > 1996 Hg] = one-half

(median difference = 0)

Alt. Hypothesis: Probability [2001 > 1996] does not ≠ one-half
(two-sided)

Alt. Hypothesis: Probability [2001 > 1996] > one-half
(median difference > 0) (one-sided)

The sign test looks only at the +/-, not the magnitude, of the differences.

12

12

Sign Test Computation

(censoring -> lots of ties)

Pattern of censored concentrations (“<“ added to view) results in several ties:

X	CensX	Y	CensY	Sign(X-Y)
<0.05	TRUE	<0.020	TRUE	0 (tie)
0.05	FALSE	0.021	FALSE	+
0.05	FALSE	<0.020	TRUE	+
0.03	FALSE	<0.050	TRUE	0 (tie)
<0.02	TRUE	0.026	FALSE	-
0.10	FALSE	<0.020	TRUE	+

Two-sided Sign Test

using the cen_signtest script in NADAscripts.R

```
> cen_signtest(X2001, Cens01, X1996, Cens96)
Censored sign test for  (x:X2001 - y:X1996) equals 0
   alternative hypothesis: true difference X2001 - X1996 not = 0
   n = 82    n+ = 55    n- = 8     ties: 19

  No correction for ties:    p-value = 9.76e-10
Fong correction for ties:    p-value = 0.0013
```

(many ties: is too small)

The Signed-Rank Test

The test ranks the differences x - y = d in absolute value, then places the algebraic sign +/- back onto the ranks.

Null Hypothesis: Median difference = 0; the two groups have similar cdfs

Alt. Hypothesis: Median difference is ≠ 0 (two-sided); the cdfs are not the same

Alt. Hypothesis: Median difference is > 0 (one-sided); frequently X > Y

The signed-rank test generally has more power than the sign test because it uses the magnitudes of differences – a larger difference is stronger evidence than a smaller difference.

The test also requires a correction for ties -- these two tests assumed that ties were not present and their p-values are only correct if few ties are present. Censoring results in many ties.

15

15

Censored Signed-Rank Test in R

Two-sided test:

```
> cen_signedrank.test (X2001, Cens01, X1996, Cens96)
Censored signed-rank test for x:X2001 - y:X1996 equals 0
alternative hypothesis: true difference X2001 - X1996 does not equal 0

 Pratt correction for ties
 n = 82   Z= 5.673   p-value = 0.00000001407
```

Other corrections have been proposed. Is this one the best one? It is the only one available in R (in the coin package). With a p-value this small, we reject the null hypothesis and state that mercury concentrations for the two years are not the same.

16

16

Paired Prentice Wilcoxon Test

- A nonparametric test (O'Brien and Fleming, *Biometrics 43*, p. 169-180, 1987)
- Similar goal to the signed-rank test, to determine if one of the paired groups has "a shift in location" -- that y has consistently larger or smaller values than x.
- Has more power than the signed-rank test and sign tests when the differences are asymmetric (which environmental data often are).
- Has a good 'pedigree' of theory as part of a class of tests validated for censored data (linear rank tests on Prentice-Wilcoxon scores).
- Is built to handle ties. Doesn't need the retro-fit corrections of the sign and signed-rank tests for frequent ties.
- Can be thought of as something like a paired t-test on the scores.
- In terms of power and theoretical justification, I'd rank the usefulness of the three tests for testing censored environmental data as
  PPW > signed-rank > sign test

 17

17

The PPW Test in R

Two-sided test:

```
> ppw.test(X2001, Cens01, X1996, Cens96)
        Paired Prentice-Wilcoxon test

data:  X2001 and X1996
Paired Prentice Z = 6.044, n = 82, p-value = 1.504e-09
alternative hypothesis: true difference is not equal to 0

Median difference is 0.015
```

With a p-value this small, we reject the null hypothesis and state that mercury concentrations for the two years are not the same.

 18

18

Censored Paired Test
(a parametric test)

Null Hypothesis: mean difference = 0

Alt. Hypothesis: mean difference ≠ 0 (two-sided)

Alt. Hypothesis: mean difference > 0 (one-sided)

If the mean difference is not 0, then the means of the two separate columns are not equal.

19

Assumptions of the Censored Paired Test

- Primary assumption is that the paired differences have a normal distribution.
- Check this with a Q-Q plot for interval-censored data
- If non-normal, p-values may be too high (will not be too low)
- If p-values are low (0.01 etc.) then the violation of the normal assumption hasn't obscured the nonzero mean difference

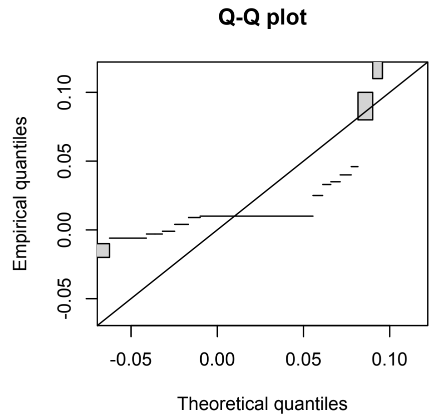
20

Q-Q plot for interval-censored data

If the data do not follow a pattern similar to the straight line representing a normal distribution, the test's p-value may be too high.

Shaded rectangles show an area where the quantile (percentile) can be anywhere within the range of the box, due to the interval-censored nature for some differences.

If you need to transform the data, transform the X and Y original variables. Cannot transform the differences -- cannot take a log of a negative difference, for example.



21

21

Censored Paired test in R: 2-sided

```
> cen_paired(X2001, Cens01, X1996, Cens96)

Censored paired test for mean(X2001 - X1996) equals 0.
alternative hypothesis: true mean difference does not
equal 0.

 n = 82   Z= 4.4747   p-value = 0.000007653
 Mean difference = 0.02093835
```

the p-value is very low, so non-normality is not causing a problem here

22

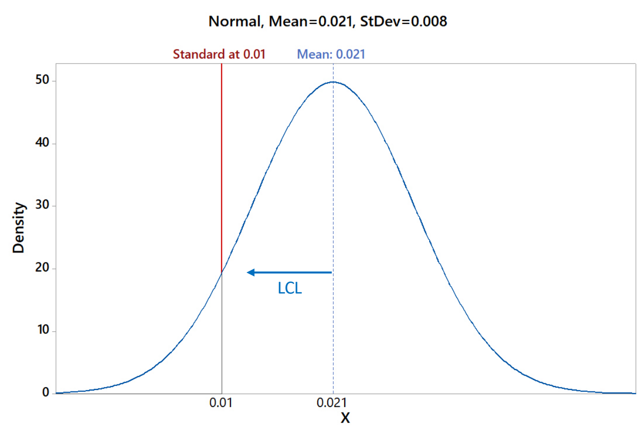
22

Comparing the mean of censored data to a standard

- Instead of a second column of data, insert a single number (the standard value) in the place of the y variable in cen_paired.
- To determine whether the mean exceeds a standard, set alternative = “greater”.
- Is the same process used by computing a one-sided lower confidence interval on the mean in section 7b. If the LCL95 is above the standard then the mean is shown to exceed the standard with 95% confidence, and correspondingly the p-value for the test will be less than (1-0.95) = 0.05.

The Censored Paired Test can be used to compare data to a standard

- This test is computed by MLE assuming a normal distribution.
- The methods in Section 7b allow you to test whether the mean exceeds a standard while assuming a lognormal or gamma distribution, which are often a better fit than is the normal distribution.



Comparing the mean of censored data to a standard

Does the mean of 2001 Hg concentrations exceed a 'standard' of 0.01?

```
> cen_paired(X2001, Cens01, 0.01, alt = "greater")
 Censored paired test for mean(X2001) equals 0.01
 alternative hypothesis: true mean X2001 exceeds 0.01.

 n = 82   Z= 5.2293   p-value = 8.508e-08
 Mean X2001 = 0.03608
```

The p-value is small, so reject the null hypothesis of equality. The mean exceeds the standard.

Summary: Tests for Difference Between Matched Pairs

Parametric and nonparametric tests answer two different questions. Which did you want to answer?

Nonparametric: The Hg concentrations were more often higher in 2001 at a site than in 1996. (Signed-rank, PPW or sign tests)

Parametric: The mass of Hg over the entire site is higher in 2001 than in 1996. (Censored Paired Test)

Next Video
Tuesday June 16th 2020 at anytime

How Many Observations Are Censored Data Worth?

Your data set has 30 observations, 12 of which are nondetects. How many equivalent observations do you have for the information content possessed in these data? Is it more than just 30 - 12 = 18 detected observations? How much more?

- No need to register. Free to view in any time zone. Will be posted on June 16th to our Videos page https://practicalstats.com/videos/
- Sign up for our newsletter/announcement list to get the announcement and description of content emailed to you. http://practicalstats.com/news/
- More videos related to our Training Courses are located on our Training Site: https://practicalstats.teachable.com

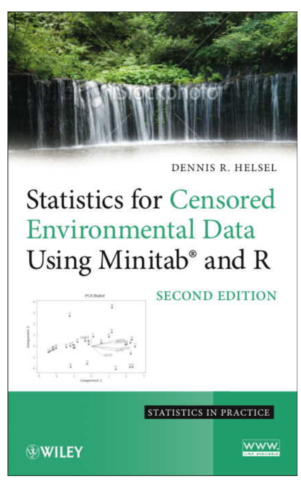
 27

27

For more on stats for data with NDs:

Statistics for Censored Environmental Data (the second edition)

by Dennis R. Helsel (2012)



 28

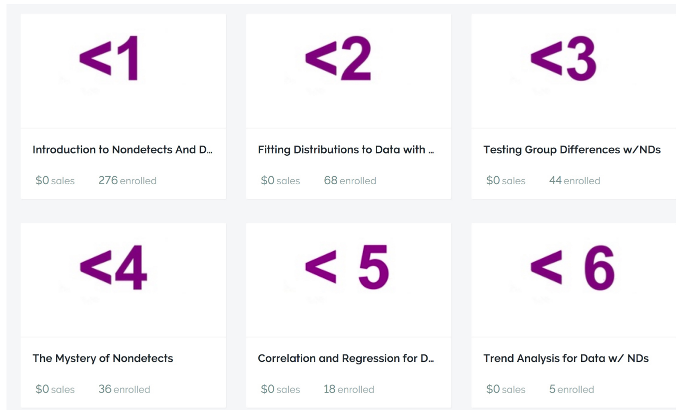
28

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 who missed it know about it.

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



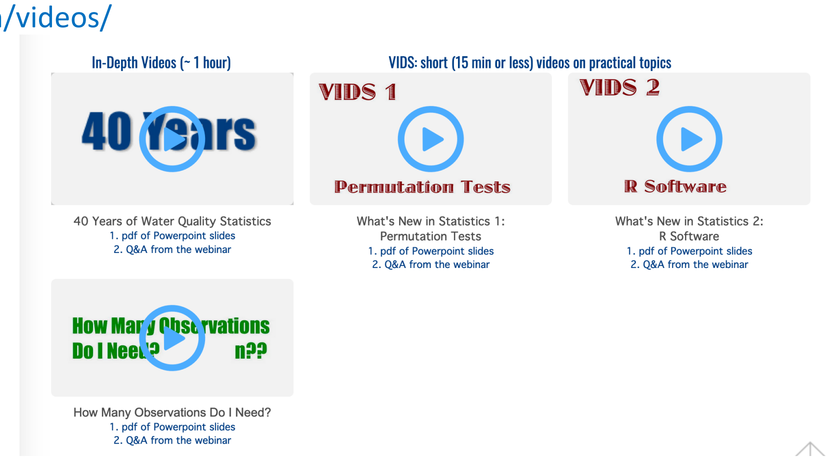
 29

29

Webinars on general statistics on our Video page

at no charge
http://www.practicalstats.com/videos/

Current free videos on our video page:



 30

30

Thank you for viewing

- Some of the material is based on 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