Nondetects And Data Analysis:
Matched Pair Tests with NDs

Dennis R. Helsel, Ph.D
Practical Stats

1

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

2

2

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.

3

3

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.

4

4

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

5

5

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

6

6

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
 ...
```

7

7

Interval-censored differences

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

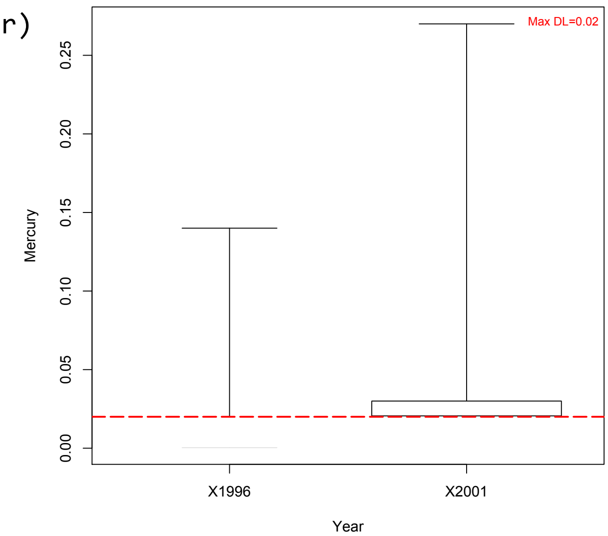
X2001	Cens01	X1996	Cens96	mindiff	maxdiff
<0.02	TRUE	<0.020	TRUE	-0.020	0.020
0.02	FALSE	0.021	FALSE	-0.001	-0.001
0.05	FALSE	<0.020	TRUE	0.030	0.050
0.02	FALSE	<0.020	TRUE	0.000	0.020
<0.02	TRUE	0.026	FALSE	-0.026	-0.006
0.10	FALSE	<0.020	TRUE	0.080	0.100

8

8

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

9

9

The Sign Test

1. Compute differences across each row = i $d_i = x_i - y_i$
2. Record sign of d_i: + or – (or 0 for x tied with y)
3. S^+ = # of positive differences n(x > y)
4. The probability of getting the n(x>y) out of i=N total rows, or a more extreme n, is the p-value.
5. A correction for ties, including pairs such as <0.05 vs 0.026, was given by Fong (2003).

10

10

The Sign Test

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

Alt. Hypothesis: Probability [2001 > 1996] does not = one-half
(median difference = 0) (two-sided)

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

11

11

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 = 9.7617e-10
Fong correction for ties:   p-value = 0.0013449
```

12

12

One-sided Sign Test

1-sided test is appropriate when we expected a specific direction in Hg concentrations (difference is greater than zero, for example) from first principles prior to seeing a plot or the data:

```
> cen_signtest(X2001, Cens01, X1996, Cens96, alt="greater")
Censored sign test for  (x:X2001 - y:X1996) equals 0
   alternative hypothesis: true difference X2001 - X1996 > 0
   n = 82    n+ = 55    n- = 8    ties: 19

  No correction for ties:   p-value = 4.8808e-10
Fong correction for ties:   p-value = 0.00067247
```

13

13

The Signed-Rank Test

Null Hypothesis: Median difference = 0

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

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

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

For environmental studies, this is usually true.

14

14

Computing the Signed-Rank Test

1. Compute differences $d_i = x_i - y_i$
2. Rank differences in order of absolute value
   Zero distances (x tied with y) get the smallest ranks.
3. Signed Rank -- put algebraic sign on the rank
4. For tied ranks, the differences = 0 are discarded after computing the signed ranks, not before (Pratt correction for ties).

Test statistic Z = sum of positive signed-ranks / standard deviation

The p-value is the probability of getting a Z at the observed value or something more extreme.

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
```

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

One-sided Signed-Rank Test

If we expected an increase in Hg prior to seeing the data:

```
> cen_signedrank.test(X2001, Cens01, X1996, Cens96, alt = "greater")

Censored signed-rank test for x:X2001 - y:X1996 equals 0
alternative hypothesis: true difference X2001 - X1996 is greater than 0

 Pratt correction for ties
 n = 82   Z= 5.673   p-value = 0.000000007034
```

With a p-value this small, we reject the null hypothesis and state that mercury concentrations are more frequently higher in 2001 than in 1996.

17

17

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.
- Yet different mechanics: it computes scores (similar to Kaplan-Meier percentiles) for both groups combined, then splits the scores back into their respective groups and determines whether the paired scores are significantly different.
- 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.

18

18

The PPW Test

Null Hypothesis: Both columns of paired data have similar cdfs

Alt. Hypothesis: The two columns have different cdfs (two-sided)

Alt. Hypothesis: One column is shifted higher than the other column (1-sided)

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

19

19

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.

20

20

One-sided PPW Test

If we expected an increase in Hg prior to seeing the data:

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

data:  X2001 and X1996
Paired Prentice Z = 6.044, n = 82, p-value = 7.518e-10
alternative hypothesis: true difference is greater than 0

Median difference is 0.015
```

With a p-value this small, we reject the null hypothesis and state that mercury concentrations are more frequently higher in 2001 than in 1996.

21

21

Censored Paired Test

(a parametric test)

Null Hypothesis: mean difference = 0

Alt. Hypothesis: mean difference is NOT = 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.

22

22

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

23

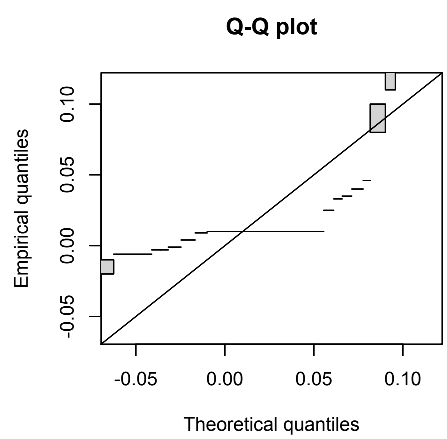
23

Q-Q plot for interval-censored data

These data do not follow a pattern similar to the straight line representing a normal distribution, but more like a bending "smile". Therefore these data do not follow 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.

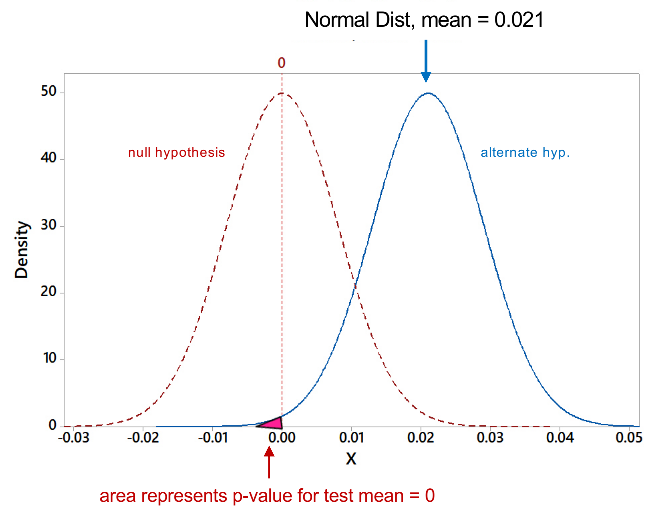


24

24

Mechanics of the Censored Paired Test

1. Computes min and max differences (slide 8) between the pairs of data.
2. Input these interval-censored data to the survreg command without a grouping X variable for the regression. Fits a normal distribution to the data by Maximum Likelihood.
3. Result is an estimate of the mean difference (0.021), a confidence interval on the mean, and a test for whether the mean equals zero assuming data follow a normal distribution.

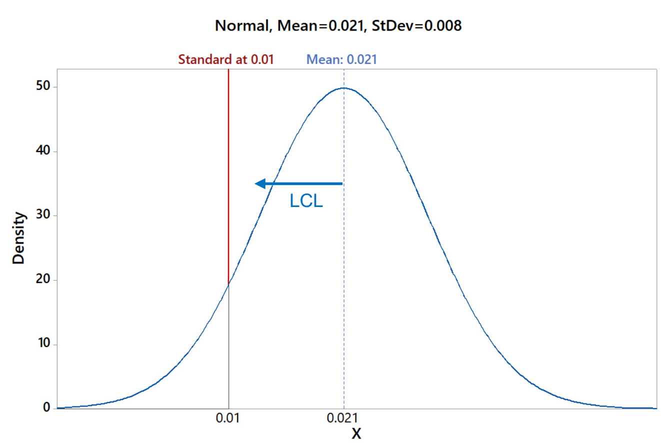


25

25

Mechanics of the Censored Paired Test

4. If using this test to compare one column of data to a standard, it is the same process as in Section 7b where a confidence interval is computed by MLE assuming a normal distribution.
5. 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.



26

26

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

27

27

Fixing Non-normality

If the p-value were in the range of 0.2 down to 0.05 so that you worry that non-normality may be obscuring what is an actual difference

1. Use a nonparametric test. Changes the test to whether the median difference equals zero (a frequency measure of difference).
2. Take the logs of the original x and y data and re-run (you can't take logs of the differences because many of them are negative). However, this means that the test is no longer a test for difference in the mean difference, but in the mean difference of the logs.
3. Run a permutation test of the mean difference.

28

28

One-sided Censored Paired test

One-sided test expecting increases in Hg conc:

```
> cen_paired(X2001, Cens01, X1996, Cens96, alt = "greater")

 Censored paired test for mean(X2001 - X1996) equals 0.
 alternative hypothesis: true mean difference is greater than 0.

 n = 82   Z= 4.4747   p-value = 0.000003826
 Mean difference = 0.02094
```

29

29

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.

30

30

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.

31

31

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)

32

32