

12.10: Wilcoxon Signed-Rank Test

Learning Objectives

- To use the Wilcoxon signed-rank test when you'd like to use the paired t -test, but the differences are severely non-normally distributed.

When to use it

Use the Wilcoxon signed-rank test when there are two nominal variables and one measurement variable. One of the nominal variables has only two values, such as "before" and "after," and the other nominal variable often represents individuals. This is the non-parametric analogue to the paired t -test, and you should use it if the distribution of differences between pairs is severely non-normally distributed.

For example, Laureysens et al. (2004) measured metal content in the wood of 13 poplar clones growing in a polluted area, once in August and once in November. Concentrations of aluminum (in micrograms of Al per gram of wood) are shown below.

Clone	August	November	August–November
Columbia River	18.3	12.7	-5.6
Fritzi Pauley	13.3	11.1	-2.2
Hazendans	16.5	15.3	-1.2
Primo	12.6	12.7	0.1
Raspalje	9.5	10.5	1.0
Hoogvorst	13.6	15.6	2.0
Balsam Spire	8.1	11.2	3.1
Gibecq	8.9	14.2	5.3
Beaupre	10.0	16.3	6.3
Unal	8.3	15.5	7.2
Trichobel	7.9	19.9	12.0
Gaver	8.1	20.4	12.3
Woltersen	13.4	36.8	23.4

There are two nominal variables: time of year (August or November) and poplar clone (Columbia River, Fritzi Pauley, etc.), and one measurement variable (micrograms of aluminum per gram of wood). The differences are somewhat skewed; the Woltersen clone, in particular, has a much larger difference than any other clone. To be safe, the authors analyzed the data using a Wilcoxon signed-rank test, and I'll use it as the example.

Null hypothesis

The null hypothesis is that the median difference between pairs of observations is zero. Note that this is different from the null hypothesis of the paired t -test, which is that the *mean* difference between pairs is zero, or the null hypothesis of the [sign test](#), which is that the numbers of differences in each direction are equal.

How the test works

Rank the absolute value of the differences between observations from smallest to largest, with the smallest difference getting a rank of 1, then next larger difference getting a rank of 2, etc. Give average ranks to ties. Add the ranks of all differences in one direction, then add the ranks of all differences in the other direction. The smaller of these two sums is the test statistic, W (sometimes symbolized T_s). Unlike most test statistics, *smaller* values of W are less likely under the null hypothesis. For the aluminum in

wood example, the median change from August to November (3.1 micrograms Al/g wood) is significantly different from zero ($W = 16$, $P = 0.040$).

Example

Buchwalder and Huber-Eicher (2004) wanted to know whether turkeys would be less aggressive towards unfamiliar individuals if they were housed in larger pens. They tested 10 groups of three turkeys that had been reared together, introducing an unfamiliar turkey and then counting the number of times it was pecked during the test period. Each group of turkeys was tested in a small pen and in a large pen. There are two nominal variables, size of pen (small or large) and the group of turkeys, and one measurement variable (number of pecks per test). The median difference between the number of pecks per test in the small pen vs. the large pen was significantly greater than zero ($W = 10$, $P = 0.04$).



Fig. 4.12.1 A turkey's head.

Ho et al. (2004) inserted a plastic implant into the soft palate of 12 chronic snorers to see if it would reduce the volume of snoring. Snoring loudness was judged by the sleeping partner of the snorer on a subjective 10-point scale. There are two nominal variables, time (before the operations or after the operation) and individual snorer, and one measurement variable (loudness of snoring). One person left the study, and the implant fell out of the palate in two people; in the remaining nine people, the median change in snoring volume was significantly different from zero ($W = 0$, $P = 0.008$).

Graphing the results

You should graph the data for a Wilcoxon signed rank test the same way you would graph the data for a paired t -test, a bar graph with either the values side-by-side for each pair, or the differences at each pair.

Similar tests

You can analyze paired observations of a measurement variable using a paired t -test, if the null hypothesis is that the mean difference between pairs of observations is zero and the differences are normally distributed. If you have a large number of paired observations, you can plot a histogram of the differences to see if they look normally distributed. The paired t -test isn't very sensitive to non-normal data, so the deviation from normality has to be pretty dramatic to make the paired t -test inappropriate.

Use the sign test when the null hypothesis is that there are equal number of differences in each direction, and you don't care about the size of the differences.

How to do the test

Spreadsheet

I have prepared a spreadsheet to do the Wilcoxon signed-rank test [signedrank.xls](#). It will handle up to 1000 pairs of observations.

Web pages

There is a web page that will perform the Wilcoxon signed-rank test. You may enter your paired numbers directly onto the web page; it will be easier if you enter them into a spreadsheet first, then copy them and paste them into the web page.

R

Salvatore Mangiafico's *R Companion* has a sample [R program for the Wilcoxon signed rank test](#).

SAS

To do Wilcoxon signed-rank test in SAS, you first create a new variable that is the difference between the two observations. You then run PROC UNIVARIATE on the difference, which automatically does the Wilcoxon signed-rank test along with several

others. Here's an example using the poplar data from above:

```
DATA POPLARS;  
INPUT clone $ augal noval;  
diff=augal - noval;  
ATALINES;  
Balsam_Spire 8.1 11.2  
Beaupre 10.0 16.3  
Hazendans 16.5 15.3  
Hoogvorst 13.6 15.6  
Raspalje 9.5 10.5  
Unal 8.3 15.5  
Columbia_River 18.3 12.7  
Fritzi_Pauley 13.3 11.1  
Trichobel 7.9 19.9  
Gaver 8.1 20.4  
Gibecq 8.9 14.2  
Primo 12.6 12.7  
Wolterson 13.4 36.8  
;
```

```
PROC UNIVARIATE DATA=poplars;  
VAR diff;  
RUN;
```

PROC UNIVARIATE returns a bunch of descriptive statistics that you don't need; the result of the Wilcoxon signed-rank test is shown in the row labeled "Signed rank":

Tests for Location: $\mu_0=0$

Test -Statistic- -----p Value-----

Student's t t -2.3089 Pr > |t| 0.0396
Sign M -3.5 Pr >= |M| 0.0923
Signed Rank S -29.5 Pr >= |S| 0.0398

References

1. Picture of a turkey's head from Ohio State University 4-H Poultry.
2. Buchwalder, T., and B. Huber-Eicher. 2004. Effect of increased floor space on aggressive behaviour in male turkeys (*Melagris gallopavo*). Applied Animal Behavior Science 89: 207-214.
3. Ho, W.K., W.I. Wei, and K.F. Chung. 2004. Managing disturbing snoring with palatal implants: a pilot study. Archives of Otolaryngology Head and Neck Surgery 130: 753-758.
4. Laureysens, I., R. Blust, L. De Temmerman, C. Lemmens and R. Ceulemans. 2004. Clonal variation in heavy metal accumulation and biomass production in a poplar coppice culture. I. Seasonal variation in leaf, wood and bark concentrations. Environmental Pollution 131: 485-494.

This page titled [12.10: Wilcoxon Signed-Rank Test](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [John H. McDonald](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- **4.12: Wilcoxon Signed-Rank Test** by [John H. McDonald](#) has no license indicated. Original source: <http://www.biostat handbook.com>.