STAT 621 Lecture Notes
Wilcoxon Signed Rank Test

Like the sign test, the signed rank test is a procedure for drawing an inference about the location parameter of a population, and can be used for both paried and one-sample data. No normality assumptions are made, but the assumptions here are slightly more stringent than those for the sign test. Again, we will define the procedure for paired data, then generalize it to the one-sample case. First a little set-up and notation.

Data: Two measurements on each subject for $2n$ observations: $(X_1, Y_1), (X_2, Y_2), \ldots (X_n, Y_n)$. Define the differences between paired measurements as

$$Z_i = Y_i - X_i, \qquad i = 1, \ldots, n.$$

Assumptions:

1. $Z_1, \ldots, Z_n$ are mutually independent.

2. $Z_i \sim f_i$ where f_i is a continuous distribution, symmetric about its median, θ.

- Briefly discuss the assumptions.

- What does θ represent?

The Wilcoxon Signed Rank Test

Here interest is in testing the null hypothesis of no treatment effect,

$$H_0 : \theta = 0.$$

Below we outline the Wilcoxon procedure for testing this null by defining the test statistic and rejection rules, and considering an example. After that we discuss the logic behind the procedure, as well as a few special considerations. Finally we compare this test to the large-sample (Normal-theory) test for location.

<u>Test Statistic:</u> We will denote the test statistic for the Wilcoxon signed rank test by W, and the procedure for its computation is outlined below. Note, there are different versions of how to compute this statistic, all of which are equivalent in terms of the ultimate conclusion. The procedure below is consistent with what the R software will report.

1. Compute the absolute differences $|Z_1|, |Z_2|, \ldots, |Z_n|$, where $|Z_i| = |Y_i - X_i|$.

2. Order the absolute differences from smallest to largest and assign ranks, $R_i = \text{Rank}(|Z_i|)$.

3. The Wilcoxon signed rank test statistic is the sum of the ranks of those pairs whose difference is positive,
$$W = \sum_{i=1}^{n} R_i I(Z_i > 0)$$

A little discussion... What is the smallest value W can take? The largest value? What is $E(W)$ if H_0 is true?

What would you expect if in reality $\theta < 0$? If $\theta > 0$? If $\theta \neq 0$?

Wilcoxon derived the null distribution of W. You can find critical values and p-values for the test in R with the `qsignrank` and `psignrank` functions respectively.

Example (Lehman 1975): Two methods for closing wounds were compared by making incisions on each of $n = 10$ rats. One incision was closed with medical tape and the other with sutures. Tensile strength was measured on each wound 40 days after the procedure.

Rat	Tape	Suture	Z
1	659	452	207
2	984	587	397
3	397	460	-63
4	574	787	-213
5	447	351	96
6	479	277	202
7	676	234	442
8	761	516	245
9	647	577	70
10	577	513	64

Consider testing if tape-closed wounds tend to be stronger than sutured wounds. State the hypotheses.

Compute the test statistic W_{obs}.

Critical values can be found with the function `qsignrank`. You need to be a little careful because this test statistic is discrete, and R reports upper-tail areas in terms of a strict inequality. Here, for our right-tailed test, we want the value w so that $P(W \geq w) = \alpha$ when H_0 is true. That's the same thing as saying $P(W > w - 1) = \alpha$. So

```
#  Critical Value
>  qsignrank(.05, 10, lower.tail=F)

 [1] 44
```

So our rule is to reject the null when $W_{obs} > 44$, or equivalently $W_{obs} \geq 45$.

P-values can be found with the function `psignrank`. We want to find $P(W \geq 47) = P(W > 46)$ under the null. This is,

```
# P-value
> psignrank(46,10,lower.tail=F)

  [1] 0.02441406
```

So what is the indicated conclusion?

A final note. It turns out that the null distribuiton of the signed rank statistic is actually symmetric. So, for example, a p-value for a two-tailed test can be found by multiplying the *appropriate* tail probabilty by two. A picture:

The R function `wilcox.test` performs the Wilcoxon signed rank test, and also computes a confidence interval for the location parameter θ. Since we are doing a one-sided test, R reports a one-sided interval as well.

```
> Tape=c(659,984,397,574,447,479,676, 761,647,577)
> Suture=c(452,587,460,787,351,277,234,516,577,513)
>
>
> wilcox.test(Tape, Suture, alternative="greater", paired=TRUE,
               conf.int=TRUE, exact=TRUE, conf.level=.95)


Wilcoxon signed rank test

data:  Tape and Suture
V = 47, p-value = 0.02441
alternative hypothesis: true location shift is greater than 0
95 percent confidence interval:
  16 Inf
sample estimates:
(pseudo)median
           149
```

Sampling Distribution of W

That's how the signed rank test works. Now why does it work? To answer this we need to understand the sampling distribution of the test statistic under the null hypothesis. Once we find this distribution, we can evaluate observed values of W, i.e., compute p-values, etc.

For this test, if the null is true then the distribution of Z, the difference between paired measurements, has median zero. What is the probability that any Z_i is positive, negative, or zero?

Now imagine taking our n absolute differences $|Z_i|$ and randomly assigning them a $+$ or a $-$. How many such sign combinations are there?

The same holds if we consider assigning the $\pm$ to the rank instead of to the absloute difference. Moreover, each sequence of these signed ranks is equally likely and occurs with probability $(1/2)^n$. Why are these equally likely?

Consider the following simple example. An agricultural study divided each of $n = 3$ strawberry fields into two parts. A new fertilizer was applied to one part, and a standard fertilizer to the other part. Is there a difference in yield between the two fertilizers?

The following (x, y) pairs of strawberry yields were obtained:

$$(76, 78), (80, 86), (82, 91)$$

Compute W.

Is this outcome unusual if the yields for the two fertilizers have the same median? The table below enumerates all possible outcomes under this null hypothesis.

Z_i	$R_i I(Z_i > 0)$	W	Prob of W
-2, -6, -9	0, 0, 0	0	1/8
+2, -6, -9	1, 0, 0	1	1/8
-2, +6, -9	0, 2, 0	2	1/8
-2, -6, +9	0, 0, 3	3	1/8
+2, +6, -9	1, 2, 0	3	1/8
+2, -6, +9	1, 0, 3	4	1/8
-2, +6, +9	0, 2, 3	5	1/8
+2, +6, +9	1, 2, 3	6	1/8

Discuss. What if we were to test the hypothesis

$$H_0 : \theta = 0 \qquad H_A : \theta > 0$$

What if we used the two-sided alternative, $H_A : \theta \neq 0$?

<u>What To Do About Ties:</u> The assumption that data come from continuous distributions means that the probability of two Z's taking the same value, or a Z equal to 0, is negligible. In reality, however, ties are not uncommon. Next we will discuss some strategies for dealing with ties.

- Approaches for 0-differences

- Approaches for non-zero ties

Large Sample Approximation

For large samples, computing the exact sampling distribution of W can be time consuming. However W has an asymptotically normal distribution under the null hypothesis that $\theta = 0$. That is,

$$W^* = \frac{W - E(W)}{\sqrt{V(W)}} \sim N(0, 1).$$

approximately for large n. Let's discuss this idea for a second. Aren't we trying to avoid things like appealing to asymptotic results? What might be some other good options for large samples?

For a given data set, one can show that

$$E(W) = \frac{n(n+1)}{4} \qquad \text{and} \qquad V(W) = \frac{n(n+1)(2n+1)}{24}$$

These results can be used to define a large sample test based on the standard normal distribution. For example, define and apply this test to the suture data set to test $H_0 : \theta = 0$ vs. $H_A : \theta \neq 0$.

The Signed Rank Test for Unpaired Data

Suppose we have a single random sample, $X_1, \ldots, X_n$. We'd like to test whether θ, the median of population, is equal to a specific value. That is,

$$H_0 : \theta = \theta_0$$

Sketch how we can adapt the procedure for paired data to this one-sample problem. What is the test statistic? How can we find the p-value?

Example: Use the following random sample of size $n = 7$ to test the hypothesis

$$H_0 : \theta = 4 \qquad \text{vs.} \qquad H_A : \theta \neq 4$$

```
6.74   1.59   0.57   9.75   9.00   3.27  0.76
```

Here is the procedure in R. A good exercise would be to verify the value of the test statistic.

```
> wilcox.test(x, alternative="two.sided", mu=4)

Wilcoxon signed rank test

data:  x
V = 16, p-value = 0.8125
alternative hypothesis: true location is not equal to 4
```