E-Learning Module on Wilcoxon Signed Rank **Test**

Learning Objectives

At the end of this session, you will be able to know

- About Wilcoxon signed rank test
- Principle and logic behind the test
- Assumptions and Procedure of the test
- Difference between sign test and Wilcoxon signed Rank test

Introduction

- The **Wilcoxon signed-rank test** is a non parametric statistical hypothesis test used when comparing two related samples, matched samples, or repeated measurements on a single sample to assess whether their population mean ranks differ (that is it is a paired difference test). It can be used as an alternative to the paired Students t test, *t*-test for matched pairs, or the *t*-test for dependent samples when the population cannot be assumed to be Normally distributed.

The sign test is designed to test a hypothesis about the location of a population distribution. It is most often used to test the hypothesis about a population median, and often involves the use of matched pairs, for example, before and after data, in which case it tests for a median difference of zero.

We now turn to consider a somewhat analogous alternative to the t-test for correlated samples. The correlated-samples t-test makes certain assumptions and can be meaningfully applied only in so far as these assumptions are met. Namely,

- That the scale of measurement for X_1 and X_2 has the properties of an equal-interval scale;
- that the differences between the paired values of X_1 and X_2 have been randomly drawn from the source population; and
- that the source population from which these differences have been drawn can be reasonably supposed to have a normal distribution

Here again, it is not simply a question of good manners or good taste. If there is one or more of these assumptions that we cannot reasonably suppose to be satisfied, then the t-test for correlated samples cannot be legitimately applied.

Of all the correlated-samples situations that run afoul of these assumptions, we expect the most common are those in which the scale of measurement for X_1 and X_2 cannot be assumed to have the properties of an equal-interval scale.

The most obvious example would be the case in which the measures for X_1 and X_2 derive from some sort of rating scale. In any event, when the data within two correlated samples fail to meet one or another of the assumptions of the **t**-test, an appropriate non-parametric alternative can often be found in the Wilcoxon Signed Rank test.

The test is named for Frank Wilcoxon (1892 – 1965) who, in a single paper, proposed both it and the rank sum test for two independent samples (Wilcoxon, 1945). The test was popularized by Siegal (1956) in his influential text book on non-parametric statistics. Siegel used the symbol *T* for the value defined below as W

In consequence, the test is sometimes referred to as the **Wilcoxon *T* test**, and the test statistic is reported as a value of *T*. Other names may include the "*t*-test for matched pairs" or the "*t*-test for dependent samples".

Assumptions

- Data are paired and come from the same population.
- Each pair is chosen randomly and independent.
- The data are measured on an interval scale (ordinal is not sufficient because we take differences), but need not be normal.

When to use it

You use the sign test when there are two nominal variables and one ,measurement variable ranked variable. One of the nominal variables has only two values, such as "before" and "after" or "left" and "right," and the other nominal variable identifies the pairs of observations. The data could be analyzed using a paired t-test or a Wilcoxon Signed Rank test if the null hypothesis is that the mean or median difference between pairs of observations is zero. The sign test is used to test the null hypothesis that there are equal numbers of differences in each direction.

For example:

One situation in which a sign test is appropriate is when the biological null hypothesis is that there may be large differences between pairs of observations, but they are random in direction. For example, let's say you want to know whether copper pennies will reduce the number of mosquito larvae in backyard ponds.

You measure the abundance of larvae in your pond, add some pennies, then measure the abundance of larvae again a month later. You do this for several other backyard ponds, with the before- and after-pennies measurements at different times in the summer. Based on prior research, you know that mosquito larvae abundance varies a lot throughout the summer, due to variation in the weather and random fluctuations in the number of adult mosquitoes that happen to find a particular pond; even if the pennies have no effect, you expect big differences in the abundance of larvae between the before and after samples.

Why a paired t-test would be inappropriate?

To see why a paired t-test would be inappropriate for the mosquito experiment, imagine that you've done the experiment in a neighborhood with 100 backyard ponds. Due to changes in the weather, etc., the abundance of mosquito larvae increases in half the ponds and decreases in half the ponds; in other words, the probability that a random pond will decrease in mosquito larvae abundance is 0.5.

If you do the experiment on four ponds picked at random, and all four happen show the same direction of difference (all four increase or all four decrease) even though the pennies really have no effect, you'll probably get a significant paired t-test. However, the probability that all four ponds will show the same direction of change is 0.125. Thus you would get a "significant" P-value from the paired t-test 12.5% of the time, which is much higher than the P less than 0.05 you want.

The other time you'd use a sign test is when you don't know the size of the difference, only its direction; in other words, you have a ranked variable with only two values, "greater" and "smaller." For example, let's say you're comparing the abundance of adult mosquitoes between your front yard and your back yard. You stand in your front yard for 5 minutes, swatting at every mosquito that lands on you, and then you stand in your back yard for 5 minutes. You intend to count every mosquito that lands on you, but they are so abundant that soon you're dancing around, swatting yourself wildly, with no hope of getting an accurate count.

You then repeat this in your back yard and rate the mosquito abundance in your back yard as either "more annoying" or "less annoying" than your front yard. You repeat this on several subsequent days. You don't have any numbers for mosquito abundance, but you can do a sign test and see whether there are significantly more times where your front yard has more mosquitoes than your back yard, or vice versa.

Null hypothesis

The H0 is that an equal number of pairs of observations have a change in each direction. If the pairs are "before" and "after," the H0 would be that the number of pairs showing an increase equals the number showing a decrease.

Note that this is different from the null hypothesis tested by a paired t-test, which is that the mean difference between pairs is 0. The difference would be illustrated by an example in which 19 pairs have an increase of 1 unit, while one pair have a decrease of 19 units. The 19:1 ratio of increases to decreases would be highly significant under a sign test, but the mean change would be 0.

General Test procedure

Let n be the sample size, the number of pairs. Thus, there are a total of *2n* data points. For i=1,2, ..., n, let x_{1i} and x_{2i} denote the measurements.

H0: Median difference between the pairs is zero

H1: Median difference is not zero

- For , i=1,2,...,n calculate $|x_{2i} - x_{1i}|$ and sgn of $|x_{2i} - x_{1i}|$ Exclude pairs with $|x_{2i} - x_{1i}| = 0$

- Let n_r be the reduced sample size.

- Rank the pairs, starting with the smallest as 1. Ties receive a rank equal to the average of the ranks they span. Let R_i denote the rank.

.

- Calculate the test statistic

$$W = \left| \sum_{i=1}^{m} [\operatorname{sgn}(x_{2i} - x_{1i}).R_i] \right|$$

the absolute value of the sum of the signed ranks.

- As n_r increases, the sampling distribution of W converges to a normal distribution.

Logic behind the test

Both paired sample sign test and paired sample Wilcoxon signed rank test are alternative non-parametric methods of paired sample t-test. When the normality assumption is not satisfied or the sample size is too small, t-test is not valid. In this case, paired sample sign test or paired sample Wilcoxon signed rank test should be used to test their means.

While paired sample sign test only uses the sign of paired difference ranks, paired sample Wilcoxon signed rank test uses magnitude and sign of the paired difference ranks.

As an example, suppose researchers want to test whether a new antihypertension drug is more effective than a traditional one. They collected the reduced value of blood pressure by these two antihypertension drugs in 10 hypertensive patients.

Since the normality is not satisfied, paired sample t-test is not valid. Now we use nonparametric methods to test whether their means are equal.

Uses of Wilcoxon signed rank test

As for the sign test, the Wilcoxon signed rank sum test is used to test the null hypothesis that the median of a distribution is equal to some value. It can be used

a) in place of a one-sample t-test
b) in place of a paired t-test or
c) for ordered categorical data where a numerical scale is inappropriate but where it is possible to rank the observations.

In many applications, this test is used in place of the one sample t-test when the normality assumption is questionable. It is a more powerful alternative to the sign test, but does assume that the population probability distribution is symmetric.

This test can also be applied when the observations in a sample of data are ranks, that is, ordinal data rather than direct measurements.

Practical Procedure for different situations

Case 1: Paired data

1. State the null hypothesis - in this case it is that the median deference, M, is equal to zero.

2. Calculate each paired deference, di= xi - yi,, where xi , yi are the pairs of observations.

3. Rank the differences, ignoring the signs (i.e. assign rank 1 to the smallest modulus of di, rank 2 to the next etc.)

4. Label each rank with its sign, according to the sign of di

5. Calculate W+, the sum of the ranks of the positive differences, and W−, the sum of the ranks of the negative differences. (As a check the total, (W+)+ (W−) , should be equal to n(n+1)/2 , where n is the number of pairs of observations in the sample).

Case 2: Single set of observations

1. State the null hypothesis - the median value is equal to some value M.
2. Calculate the deterrence between each observation and the hypothesized median, $d_i = x_i - M$.
3. Apply Steps 3-5 as above.

Under the null hypothesis, we would expect the distribution of the deference's to be approximately symmetric around zero and the distribution of positives and negatives to be distributed at random among the ranks.

Under this assumption, it is possible to work out the exact probability of every possible outcome for W. To carry out the test, we therefore proceed as follows:

6. Choose W = min(W−., W+).

7. Use tables of critical values for the Wilcoxon signed rank sum test to find the probability of observing a value of W or more extreme. Most tables give both one-sided and two-sided p-values. If not, double the one-sided p-value to obtain the two-sided p-value. This is an exact test.

Normal approximation

If the number of observations/pairs is such that n(n+1)/2 is large enough (> 20), a normal approximation can be used with

Mean of W = n(n + 1)/4 ,

$$\sigma_W = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

$$Z = \frac{W - Mean}{\sigma_W}$$

If $|Z| \geq Z_{\alpha/2}$ we reject H0

Dealing with ties:

There are two types of tied observations that may arise when using the Wilcoxon signed rank test:

- Observations in the sample may be exactly equal to M (i.e. 0 in the case of paired deference's). Ignore such observations and adjust n accordingly.
- Two or more observations/differences may be equal. If so, average the ranks across
- the tied observations and reduce the variance by $t^3 - t/48$ for each group of t

Difference between sign test and Wilcoxon Sign test

The sign test could be used to test the null hypothesis that two populations have the same distribution of values, when the data consisted of paired samples. If the null hypothesis were true then the median of these differences would be zero.

The only information needed for the sign test of the equality of two population distributions, when paired samples are used, is the number of times the first data value in a pair is larger than the second. That is the sign test does not require the actual values of the data pairs, only the knowledge of which is larger. However although it is easy to use the sign test is not particularly efficient test of the hypothesis that the population distributions are the same.

For if this null hypothesis is indeed true, then not only will the distribution of the paired differences have median zero but it will also have the stronger property of being symmetric about zero. That is for any number x it will be just as likely for the first value in the pair to be larger than the second by the amount x as for the second value to be larger than the first by this amount. The sign test however does not check for symmetry of the distribution of the differences, only that its median value is equal to zero

For instance suppose that the data consists of 12 paired values whose differences are as follows:

2 , 5, -0.1, -0.4, -0.3, 9,7,8, 12,-0.5,-1,-0.6

Since 6 of the differences are positive and the 6 are negative this data is perfectly consistent with the hypothesis that the median of the differences is 0.

On the other hand since all the large values are on the positive side the data do not appear to be consistent with the hypothesis that the two population distributions are equal.

A test which is more sensitive than the sign test is called the Wilcoxon signed rank test. And it proceeds by testing whether the distribution of the differences of the paired values is symmetric about zero.

Suppose that paired samples of size n are chosen from two populations. Di denote xi-yi. For i=1,2,..,n. Now order these n differences according to their absolute values. The test statistic for the signed rank test is the sum of the ranks (or positions) of the negative numbers in the resulting sequence

Example:

Suppose the data consists of the following four paired sample values

i	Xi	Yi	Xi - Yi	Ranks according to absolute value
1	4.6	6.2	-1.6	1
2	3.8	1.5	2.3	2
3	6.6	11.7	-5.1	4
4	6	2.1	3.9	3

Since the ranks of the negative values are 1 and 4 , the value of the signed rank test statistic is

TS= 1+4=5

The signed rank test is like a sign test, in that it considers those data pairs in which the first population value is less than the second. But whereas the sign test gives equal weight to each such pair, the signed rank test gives larger weights to the pairs whose differences are farthest from zero

The signed rank test calls for the rejection of the null hypothesis when we are testing

H0: The two population distributions are equal

H1: The two population distributions are not equal

If the test statistic TS is either sufficiently large or sufficiently small. A large value of TS indicates that the majority of the larger values of the differences have negative signs whereas the small value indicates that the majority have positive signs.