STAT 464 - Applied Nonparametric Statistics

4.3 - Wilcoxon Rank Sum Test

Printer-friendly version

Let $X_1, \ldots, X_N$ be a set of N observations. The method presented here does not use the observations but rather the ranks of the observations. The **rank** of X_i among the N observations, denoted $R(X_i)$, equals the number of $X'_j\, s \le x_i$. There are many cases where ranks are the observed values. For example, consider a contest where the judge has to rank the contests. It is also helpful to consider the ranks of the data where there are extreme outliers present.

Example

Suppose we have the following five observations: 9, 19, 11, 16, 20. Find the ranks of the observations.

Solution:

First, we order the observations: 9, 11, 16, 19, 20. The minimum (or smallest) observation will have the rank of 1 and the largest, 20, will have a rank of 5. Similarly, we can find the ranks for all the observations.

x	9	11	16	19	20
Rank	1	2	3	4	5

Therefore, the rank of $x = 19$ denoted $R(19) = 4$.

Now, consider the example we presented in the last section.

Recall that a company wants to see if the new method for teaching their new employees is better than the traditional method they use now.

New Method: 37, 49, 55, 57
Traditional Method: 23, 31, 46

Now, lets combine all of the data and rank them. The table below shows the combined data as well as the ranks. The numbers (and ranks) for the New Method are shown in red and the numbers (and ranks) for the Traditional Method are shown in blue below.

Combined Data	23	31	37	46	49	55	57
Rank	1	2	3	4	5	6	7
New Method	37	49	55	57			
Rank	3	5	6	7			
Traditional Method	23	31	46				
Rank	1	2	4				

The test statistic would be

$$W = \text{ sum of the ranks of the observations for Treatment 1 (or Treatment 2).}$$

Here we are actually testing:

$$H_0 : F_1(x) = F_2(x)$$
$$H_1 : F_1(x) \geq F_2(x)$$

or

$$H_0 : F_1(x) = F_2(x)$$
$$H_1 : F_1(x) \leq F_2(x)$$

where $F_j(x)$ is the distribution function for sample $j = 1, 2$. Conducting this test is similar to conducting a Permutation test. The step are detailed below.

Steps in Conducting the Wilcoxon Rank-Sum Test:

Assume that no two observations have the same value so that the ranks are distinct. We will discuss shortly how to deal with "ties" in the data. Also assume that treatment 1 has *m* observations and treatment 2 has *n* observations.

1. Combine the $m + n$ observations into one group and rank the observations from smallest to largest. Find the observed rank sum, W, of treatment 1.
2. Find all the possible permutation of the ranks into which *m* ranks are assigned to treatment 1 and *n* ranks are assigned into treatment 2.
3. For each permutation of the ranks, find the sum of the ranks for treatment 1.
4. Determine the *p*-value:

$$P_{\text{upper}} = \frac{\#\text{of rank sums } \leq \text{observed rank sum } W}{\binom{m+n}{n}}$$

Note: The rank sum of either treatment can be used; the choice of treatment 1 is arbitrary.

Instead of using the sum of the ranks, the test could also be based on the difference of mean ranks. Let W_1 be the Wilcoxon sum rank for treatment 1. Also, is $N = n + m$, then $T = 1 + 2 + \ldots + N = \sum_{i=1}^{N} i = \frac{N(N+1)}{2}$. Then

$$\text{Difference of mean ranks} = W_1\left(\frac{1}{m} + \frac{1}{n}\right) - \frac{N(N+1)}{2n}$$

$$= \frac{W_1}{n} + \frac{W_1}{m} = \frac{W_1}{n} + \frac{1}{m}\left(W_1 - \frac{N(N+1)}{2}\right)\frac{W_1}{n} - \frac{W_2}{m}$$

$$= \frac{W_1}{n} - \frac{1}{n}W_2 = \frac{W_1}{n} - \frac{1}{m}\left(\frac{N(N+1)}{2} - W_1\right)$$

This implies that it will have the same *p*-value as just using the first method discussed above.

To summarize, we have the following data:

New Method: 37, 49, 55, 57
Traditional Method: 23, 31, 46

Research Question: Is the new method "better"? In other words, does the new method tend to give higher scores?

Model: The shape of the distributions are the same but unspecified.

$$H_0 : F_1(x) = F_2(x) \qquad \Rightarrow \mu_1 = \mu_2 \qquad \Rightarrow \eta_1 = \eta_2$$

$$H_1 : F_1(x) > F_2(x) \qquad \Rightarrow \mu_1 > \mu_2 \qquad \Rightarrow \eta_1 > \eta_2$$

Here are some drawbacks from using the traditional $\bar{x}_1 - \bar{x}_2$:

1. $\bar{x}_1 - \bar{x}_2$ is strongly influenced by outliers and gross errors.
2. Simulation must be done for each dataset. The *p*-value is conditional on given data.
3. The population mean is hard to interpret in asymmetric distributions.

Notes:

Using the medians rather than the means helps in (1) and (2) but not (3). But this may be one reason to use the difference of mean ranks, $\bar{R}_1 - \bar{R}_2$.

Recall that W_1 = sum of the ranks of treatment 1 is related to $\bar{R}_1 - \bar{R}_2$ by

$$\bar{R}_1 - \bar{R}_2 = W_1 \left(\frac{1}{n} + \frac{1}{m}\right) - \frac{N(N+1)}{2}$$

Tables for upper and lower tail critical values can be found in the text, Table A3, on page 340. For small datasets, using the table is fine but its only valid if *n* and *m* are less than or equal to 10. Thankfully, we have technology to help.

Example

A researcher is interested in seeing if men download more movies onto their computers than women. They randomly sampled four men and four women and recorded how many gigabytes of movies they had on their primary computer.

Men	305	16	122	68
Women	25	63	84	103

Solution:

If we use the table in the text, we look up the value under $n = 4$ and $m = 4$. Since we want to know if men have more videos on their computers, we want an upper-tail critical value. From the table, we would reject the null hypothesis if the observed W is greater than 81. Let's find the observed value. First, combine the data (the observations for men are in red) and rank then. Finally, sum the ranks for the men.

Combined Data	16	25	63	68	84	103	122	305
Rank	1	2	3	4	5	6	7	8

The sum of the ranks is $W = 20$.

$$W = 1 + 4 + 7 + 8 = 20$$

Since 20 is less than 81, we fail to reject the null hypothesis.

Using R | Using Minitab

When in R, type the following to enter the data and fi

```
> m=c(305, 16, 122, 68)
> w=c(25, 63, 84, 103)
> wilcox.test(m, w, alternative="greater", 

   Wilcoxon rank sum test

   data: m and w
   W = 10, p-value = 0.3429
   alternative hypothesis: true location shi
```

In the above the *p*-value (exact in this case) is 0.3419. Since the p
level of 0.05, we fail to reject the null hypothesis and conclude th
to suggest that men store more movies on their computers than w

The first step is to type the data into the columns. For

women be in column 2 (C2).

Then, click `Stat > Nonparametric > Mann-Whitney.`
Wilcoxon Rank Sum test is the same as the Mann-Whitney test. I
Whitney-Wilcoxon (MWW) test.

The space labeled **First Sample** enter the column for men and in
the column for women.

For **Confidence**, enter the level we are testing (usually 0.05 unle
should be "greater".

Finally, click **OK**.

How to Handle Ties

So far, we have been assuming that the data come from a continuous distribution and that there are no ties observed. But what if there are ties?

We group all the tied observations and assign the **average rank** to tied values in that group. Call these ranks **adjusted ranks**. Consider the following example.

Example

We have systolic blood pressures from 5 men and 5 women. The observations in red are the ties.

Ordered Data	118	121	121	121	122	1
Rank w/out ties	1	2	3	4	5	6
Average Ranks	1	3	3	3	5	6

There are three values of 121 in the combined observed data. Instead of using the ranks 2, 3, and 4, we would average the three values (2 + 3 + 4) / 3 = 3 and use the average rank instead. You can see in the above table that we used the average value of 3. Similarly, we used the rank of (8 + 9) / 2 = 8.5 for the two values of 136. We then use these average ranks to nd the Wilcoxon rank-sum statistic and denote it, W_{ties}.

We can perform the permutation test using the average ranks and calculating W_{ties}. If the number of ties is small, we can use Table A3, if not we can use the large sample approximation which we will discuss a little later.

There is a test that can be carried out for ties and also a formula that adjusts for ties but we will not spend time on it. Computer software will perform these calculations for us.

‹ 4.2 - Permutation Principle up

Printer-friendly version Login to post comments