12.8 Wilcoxon Signed Ranks Test: Nonparametric Analysis for Two Related Populations

In Section 10.2, you used the paired t test to compare the means of two populations with repeated measures or matched samples. The paired t test assumes that the data are measured on an interval or a ratio scale and are normally distributed. If you cannot make these assumptions, you can use the nonparametric Wilcoxon signed ranks test to test for the median difference. The Wilcoxon signed ranks test requires that the differences are approximately symmetric and that the data are measured on an ordinal, interval, or ratio scale. When the assumptions for the Wilcoxon signed ranks test are met but the assumptions of the t test are violated, the Wilcoxon signed ranks test is usually more powerful in detecting a difference between the two populations. Even under conditions appropriate to the paired t test, the Wilcoxon signed ranks test is almost as powerful.

The Wilcoxon signed ranks test uses the test statistic \( W \). Exhibit 12.1 lists the steps required to compute \( W \).

**EXHIBIT 12.1 STEPS IN COMPUTING THE WILCOXON SIGNED RANKS TEST STATISTIC \( W \)**

1. For each item in a sample of \( n \) items, compute a difference score, \( D_i \), between the two paired values.
2. Neglect the + and − signs and list the set of \( n \) absolute differences, \( |D_i| \).
3. Omit any absolute difference score of zero from further analysis, thereby yielding a set of \( n' \) nonzero absolute difference scores, where \( n' \leq n \). After you remove values with absolute difference scores of zero, \( n' \) becomes the actual sample size.
4. Assign ranks \( R_i \) from 1 to \( n' \) to each of the \( |D_i| \) such that the smallest absolute difference score gets rank 1 and the largest gets rank \( n' \). If two or more \( |D_i| \) are equal, assign each of them the mean of the ranks they would have been assigned individually had ties in the data not occurred.
5. Reassign the symbol + or − to each of the \( n' \) ranks, \( R_i \), depending on whether \( D_i \) was originally positive or negative.
6. Compute the Wilcoxon test statistic, \( W \), as the sum of the positive ranks [see Equation (12.11)].

**WILCOXON SIGNED RANKS TEST**

The Wilcoxon test statistic \( W \) is computed as the sum of the positive ranks.

\[
W = \sum_{i=1}^{n'} R_i^{+} \tag{12.11}
\]

The null and alternative hypotheses for the Wilcoxon signed rank test are

<table>
<thead>
<tr>
<th>Two-Tail Test</th>
<th>One-Tail Test</th>
<th>One-Tail Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0: M_D = 0 )</td>
<td>( H_0: M_D \geq 0 )</td>
<td>( H_0: M_D \leq 0 )</td>
</tr>
<tr>
<td>( H_1: M_D \neq 0 )</td>
<td>( H_1: M_D &lt; 0 )</td>
<td>( H_1: M_D &gt; 0 )</td>
</tr>
</tbody>
</table>

Because the sum of the first \( n' \) integers (1, 2, ..., \( n' \)) equals \( n'(n' + 1)/2 \), the Wilcoxon test statistic \( W \) ranges from a minimum of 0 (where all the difference scores are negative) to a maximum of \( n'(n' + 1)/2 \) (where all the difference scores are positive). If the null hypothesis is true, the test statistic \( W \) is expected to be close to its mean, \( \mu_W = n'(n' + 1)/4 \). If the null hypothesis is false, the value of the test statistic is expected to be close to one of the extremes.
You use Table 12.19 to find the critical values of the test statistic \( W \) for both one- and two-tail tests for samples of the test statistic \( W \) is approximately normally distributed with mean \( \mu_W \) and standard deviation \( \sigma_W \). The mean of the test statistic \( W \) is

\[
\mu_W = \frac{n'(n' + 1)}{4}
\]

and the standard deviation of the test statistic \( W \) is

\[
\sigma_W = \sqrt{\frac{n'(n' + 1)(2n' + 1)}{24}}
\]
Therefore, Equation (12.12) defines the $Z_{STAT}$ test statistic.

**LARGE-SAMPLE WILCOXON SIGNED RANKS TEST**

\[
Z_{STAT} = \frac{W - \frac{n(n' + 1)}{4}}{\sqrt{\frac{n(n' + 1)(2n' + 1)}{24}}} \quad (12.12)
\]

You use this large-sample approximation formula when sample sizes are outside the range of Table 12.19. You reject the null hypothesis if the computed $Z_{STAT}$ test statistic falls in the rejection region. The region of rejection used depends on the level of significance and whether the test is one-tail or two-tail (see Figure 12.20).

To demonstrate how to use the Wilcoxon signed ranks test, return to the example concerning textbook prices discussed in Section 10.2 on page 379. If you cannot assume that the differences are from normally distributed populations, you can use the Wilcoxon signed ranks test to determine whether there is a difference in median prices at the local bookstore and at the online retailer.

The null and alternative hypotheses for this two-tail test are

\[
H_0: M_D = 0
\]

\[
H_1: M_D \neq 0
\]

To perform the test, follow the six steps listed in Exhibit 12.1. First, compute a set of difference scores, $D_i$, between each of the $n$ paired values:

\[
D_i = X_{1i} - X_{2i}
\]

where

\[
i = 1, 2, \ldots, n
\]

In this example, you compute a set of $n$ difference scores, using $D_i = X_{\text{local bookstore}} - X_{\text{online retailer}}$. If there is no difference in the price of textbooks between the two retailers, the difference scores will cluster near zero (i.e., $\equiv 0$) and you will not reject $H_0$. The remaining steps of the six-step procedure are developed in Table 12.20 on the next page.

Fourteen of the 19 difference scores have a positive sign. You now compute the test statistic $W$ as the sum of the positive ranks:

\[
W = \sum_{i=1}^{n'} R_i^{(+) = 9 + 7 + 14 + 8 + 12 + 15 + 16 + 5 + 2 + 4 + 17 + 11 + 18 + 6 = 144}
\]

Because $n = 19 \leq 20$, you can use Table 12.19 to determine the upper-tail critical value. Using $\alpha = 0.05$ for this two-tail test, the lower-tail critical value is 46 and the upper-tail critical value is 144 (see Table 12.21, which is a portion of Table 12.19).

Because $W = 144 \geq 144$, you reject the null hypothesis. There is evidence of a difference in the median prices of textbooks obtained from the local bookstore and the online retailer. However, the Minitab results shown in Figure 12.21 show that the $p$-value is 0.051. Because this $p$-value is greater than 0.05, you would not reject the null hypothesis! There is no evidence of a difference in the median prices of the textbooks from the two sources.

These seemingly contradictory results arise because Minitab computes the exact $p$-value, whereas Table 12.19 shows integer values that approximate the lower and upper critical values. The Minitab results are more precise and, in this example, that precision alters the outcome of the statistical test. As a general rule, if you have access to statistical applications such as Minitab that compute exact $p$-values, you should choose to use the $p$-value approach. That choice will ensure that you get the most accurate results.
### TABLE 12.20
Setting Up the Wilcoxon Signed Ranks Test for the Median Difference

| Author    | Title                        | Bookstore | Online | $D_i$ | $|D_i|_i$ | $R_i$ | Sign of $D_i$ |
|-----------|------------------------------|-----------|--------|-------|---------|-------|-------------|
| Pride     | *Business* 10/e               | 132.75    | 136.91 | -4.16 | 4.16    | 3     | -           |
| Carroll   | *Business* and Society        | 201.50    | 178.58 | 22.92 | 22.92   | 9     | +           |
| Quinn     | *Ethics for the Information* | 80.00     | 65.00  | 15.00 | 15.00   | 7     | +           |
| Bade      | *Foundations of Microeconomics* 5/e | 153.50 | 120.43 | 33.07 | 33.07   | 14    | +           |
| Case      | *Principles of Macroeconomics* 9/e | 153.50 | 217.99 | -64.49 | 64.49 | 19 | -           |
| Brigham   | *Financial Management* 13/e   | 216.00    | 197.10 | 18.90 | 18.90   | 8     | +           |
| Griffin   | *Organizational Behavior* 9/e | 199.75    | 168.71 | 31.04 | 31.04   | 12    | +           |
| George    | *Understanding and Managing*  Organizational Behavior* 5/e | 147.00 | 178.63 | -31.63 | 31.63 | 13 | -           |
| Grewal    | *Marketing* 2/e               | 132.00    | 95.89  | 36.11 | 36.11   | 15    | +           |
| Barlow    | *Abnormal Psychology*        | 182.25    | 145.49 | 36.76 | 36.76   | 16    | +           |
| Foner     | *Give Me Liberty: Seagull Ed.* (V2) 2/e | 45.50 | 37.60  | 7.90 | 7.90    | 5     | +           |
| Federer   | *Mathematical Interest Theory* 2/e | 89.95 | 91.69  | -1.74 | 1.74    | 1     | -           |
| Hoyle     | *Advanced Accounting* 9/e    | 123.02    | 148.41 | -25.39 | 25.39 | 10 | -           |
| Haviland  | *Talking About People* 4/e   | 57.50     | 53.93  | 3.57  | 3.57    | 2     | +           |
| Fuller    | *Information Systems Project Management* | 88.25 | 83.69  | 4.56  | 4.56    | 4     | +           |
| Pindyck   | *Microeconomics* 7/e         | 189.25    | 133.32 | 55.93 | 55.93   | 17    | +           |
| Mankiw    | *Microeconomics* 7/e         | 179.25    | 151.48 | 27.77 | 27.77   | 11    | +           |
| Shapiro   | *Multinational Financial Management* 9/e | 210.25 | 147.30 | 62.95 | 62.95 | 18 | +           |

### TABLE 12.21
Finding the Lower and Upper-Tail Critical Value for the Wilcoxon Signed Ranks Test Statistic $W$ Where $n = 19$ and $\alpha = 0.05$

<table>
<thead>
<tr>
<th>n</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>W</strong> (Lower, Upper)</td>
<td>41,112, 34,119, 27,126, 23,130</td>
<td>47,124, 40,131, 32,139, 27,144</td>
<td>53,137, 46,144, 37,153, 32,158</td>
<td>60,150, 52,158, 43,167, 37,173</td>
<td>41,112, 34,119, 27,126, 23,130</td>
</tr>
</tbody>
</table>

Source: Extracted from Table E.9.

### FIGURE 12.21
Minitab Wilcoxon signed ranks test results for the textbook prices example

**Wilcoxon Signed Rank Test: Difference**

Test of median = 0.000000 versus median not = 0.000000

<table>
<thead>
<tr>
<th>N for Wilcoxon</th>
<th>Test statistic</th>
<th>p</th>
<th>Estimated Median Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>19</td>
<td>144.0</td>
<td>0.051</td>
</tr>
</tbody>
</table>

These results are somewhat different from those of Section 10.2. When you used the paired $t$ test, the $p$-value was 0.0865 as compared to 0.051 for the Wilcoxon signed ranks test.

Table 12.19 provides critical values only for situations involving small samples (where $n' \leq 20$). If the sample size $n'$ is greater than 20, you must use the large-sample $Z$ approximation formula [Equation (12.12)]. For small samples, you can use either Table 12.19 or the $Z$ approximation.
To demonstrate the large-sample Z approximation, consider again the textbook prices data. Using Equation (12.12),

\[
Z_{\text{STAT}} = \frac{W - \frac{n'(n' + 1)}{4}}{\sqrt{\frac{n'(n' + 1)(2n' + 1)}{24}}}
\]

\[
144 - \frac{(19)(20)}{4} = \frac{(19)(20)(39)}{24} = 144 - 95 = \frac{1.971}{24.85} = 1.971
\]

The decision rule is

Reject \( H_0 \) if \( Z_{\text{STAT}} > +1.96 \) or if \( Z_{\text{STAT}} < -1.96 \);

otherwise, do not reject \( H_0 \).

Because \( Z_{\text{STAT}} = 1.971 > +1.96 \), reject \( H_0 \). The \( p \)-value using the normal approximation is 0.0488, which is less than 0.05. Because the \( p \)-value is less than \( \alpha = 0.05 \), you reject the null hypothesis. The fact that this result is slightly different from the exact \( p \)-value reported by Minitab is due to the use of the normal approximation, which only approximates the exact \( p \)-value based on the binomial distribution. Again, if you have access to statistical applications such as Minitab that provides the exact \( p \)-value, you should use the \( p \)-value approach.

The Wilcoxon signed ranks test makes fewer and less stringent assumptions than does the paired \( t \) test. These are the assumptions:

- The data are a random sample of \( n \) independent difference scores. The difference scores result from repeated measures or matched pairs.
- The underlying variable is continuous.
- The data are measured on an ordinal, interval, or ratio scale.
- The distribution of the population of difference scores is approximately symmetric.

### Problems for Section 12.8

#### LEARNING THE BASICS

12.91 Using Table 12.19, determine the lower- and upper-tail critical values for the Wilcoxon signed ranks test statistic \( W \) in each of the following two-tail tests:

- **a.** \( \alpha = 0.10, n' = 11 \)
- **b.** \( \alpha = 0.05, n' = 11 \)
- **c.** \( \alpha = 0.02, n' = 11 \)
- **d.** \( \alpha = 0.01, n' = 11 \)

12.92 Using Table 12.19, determine the upper-tail critical value for the Wilcoxon signed ranks test statistic \( W \) in each of the following one-tail tests:

- **a.** \( \alpha = 0.05, n' = 11 \)
- **b.** \( \alpha = 0.025, n' = 11 \)
- **c.** \( \alpha = 0.01, n' = 11 \)
- **d.** \( \alpha = 0.005, n' = 11 \)

12.93 Using Table 12.19, determine the lower-tail critical value for the Wilcoxon signed ranks test statistic \( W \) in each of the following one-tail tests:

- **a.** \( \alpha = 0.05, n' = 11 \)
- **b.** \( \alpha = 0.025, n' = 11 \)
- **c.** \( \alpha = 0.01, n' = 11 \)
- **d.** \( \alpha = 0.005, n' = 11 \)

12.94 Consider the following \( n = 12 \) difference scores \((D_i)\) from two related samples:

\[+3.2, +1.7, +4.5, 0.0, +11.1, -0.8\]
\[+2.3, -2.0, 0.0, +14.8, +5.6, +1.7\]

What is the value of the test statistic \( W \) if you are testing \( H_0: M_D = 0 \)?
12.95 In Problem 12.94, what are the lower- and upper-tail critical values for the test statistic $W$ from Table 12.19 if the level of significance is 0.05 and the alternative hypothesis is $H_1: M_D \neq 0$?

12.96 In Problems 12.94 and 12.95, what is your statistical decision?

12.97 Consider the following $n' = 12$ signed ranks ($R_i$) computed from the difference scores ($D_i$) from two related samples:

$$+5, +6.5, +4, +11, -8, +2.5, -2.5$$
$$+1, +12, +6.5, +10, +9$$

What is the value of the test statistic $W$ if you are testing $H_0: M_D = 0$ against $H_1: M_D > 0$?

12.98 For Problem 12.97, at a level of significance of 0.05, determine the upper-tail critical value for the Wilcoxon signed ranks test statistic $W$ if you want to test $H_0: M_D \leq 0$ against $H_1: M_D > 0$.

12.99 For Problems 12.97 and 12.98, what is your statistical decision?

### APPLYING THE CONCEPTS

12.100 Nine experts rated two brands of Colombian coffee in a taste-testing experiment. A rating on a 7-point scale ($1 = $extremely unpleasing, $7 = $extremely pleasing) is given for each of four characteristics: taste, aroma, richness, and acidity. The following table (data stored in the file Coffee) displays the summated ratings—accumulated over all four characteristics.

<table>
<thead>
<tr>
<th>Expert</th>
<th>Brand A</th>
<th>Brand B</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.C.</td>
<td>24</td>
<td>26</td>
</tr>
<tr>
<td>S.E.</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>E.G.</td>
<td>19</td>
<td>22</td>
</tr>
<tr>
<td>B.L.</td>
<td>24</td>
<td>27</td>
</tr>
<tr>
<td>C.M.</td>
<td>22</td>
<td>25</td>
</tr>
<tr>
<td>C.N.</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>G.N.</td>
<td>27</td>
<td>26</td>
</tr>
<tr>
<td>R.M.</td>
<td>25</td>
<td>27</td>
</tr>
<tr>
<td>P.V.</td>
<td>22</td>
<td>23</td>
</tr>
</tbody>
</table>

a. At the 0.05 level of significance, is there evidence of a difference in the median summated rating between brand A and brand B?
b. Compare the results of (a) with those of Problem 10.20 on page 384.

12.101 In industrial settings, alternative methods often exist for measuring variables of interest. The data in the file Measurement (coded to maintain confidentiality) represent measurements in-line (i.e., collected from an analyzer during the production process) and from an analytical lab. (Data extracted from M. Leitnaker, “Comparing Measurement Processes: In-Line Versus Analytical Measurements,” *Quality Engineering*, 13, 2000–2001, pp. 293–298.)

a. At the 0.05 level of significance, is there evidence of a difference in the median measurements in-line and from an analytical lab?
b. Compare the results of (a) with those of Problem 10.21 on page 384.

12.102 Is there a difference between the prices at a warehouse club such as Costco and store brands? To investigate this, a random sample of 10 purchases was selected, and the prices were compared. (Data extracted from “Shop Smart and Save Big,” *Consumer Reports*, May 2009, p. 17.) The prices for the products are stored in Shopping.

a. At the 0.05 level of significance, is there evidence of a difference between the median price of Costco purchases and store brand purchases?
b. Compare the results of (a) with those of Problem 10.22 on page 384.

12.103 Over the past year, the vice president for human resources at a large medical center conducted a series of three-month workshops aimed at increasing worker motivation and performance. As a check on the effectiveness of the workshops, she selected a random sample of 35 employees from the personnel files and recorded their most recent annual performance ratings along with the ratings attained prior to attending the workshops. The data are stored in the file Perform.

a. At the 0.05 level of significance, is there evidence of a difference between the median performance ratings?
b. Compare the results of (a) with those of Problem 10.25 on page 385.

12.104 The data in the file Concrete represent the compressive strength in thousands of pounds per square inch (psi) of 40 samples of concrete taken two and seven days after pouring.


a. At the 0.01 level of significance, is there evidence that the median strength is less at two days than at seven days?
b. Compare the results of (a) with those of Problem 10.26 on page 385.
EG12.8 EXCEL GUIDE FOR THE WILCOXON SIGNED RANKS TEST

There are no Excel Guide instructions for this section.

MG12.8 MINITAB GUIDE FOR THE WILCOXON SIGNED RANKS TEST

Use Calculator to compute differences and then use 1-Sample Wilcoxon to perform the Wilcoxon signed ranks test. For example, to perform this test for the Table 12.20 textbook price data, open the BookPrices worksheet. To compute the price differences between the local bookstore and the online retailer, select Calc ➔ Calculator ➔ In the Calculator dialog box:

1. Enter C5 in the Store result in variable box.
2. Enter C3 - C4 in the Expression box. (Column C3 contains the Bookstore column and C4 contains the Online column, so C3 - C4 is a shorthand way of entering Bookstore - Online.)
3. Click OK.

Next, enter Difference as the variable name for column C5. Then select Stat ➔ Nonparametrics ➔ 1-Sample Wilcoxon. In the 1-Sample Wilcoxon dialog box:

1. Double-click C5 Difference in the variables list to add Difference to the Variables box.
2. Click Test median and enter 0.0 in its box.
3. Select not equal from the Alternative drop-down list (to perform the two-tail test).
4. Click OK.