10.3 HYPOTHESIS TESTING

Example 10.37
The times between two successive crashes are recorded for two competing computer systems as follows (time in weeks):

System $X$ : 1.8 0.4 2.7 3.0  
System $Y$ : 2.0 5.4 1.3 4.5 0.8

In order to test the hypothesis that both systems have the same mean time between crashes against the alternative that system $X$ has a shorter mean time between crashes, we first arrange the combined data in ascending order and assign ranks ($y$ ranks are underlined for easy identification):

<table>
<thead>
<tr>
<th>Original data</th>
<th>0.4</th>
<th>0.8</th>
<th>1.3</th>
<th>1.8</th>
<th>2.0</th>
<th>2.7</th>
<th>3.0</th>
<th>4.5</th>
<th>5.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ranks</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
</tbody>
</table>

The rank sum $W = 2 + 3 + 5 + 8 + 9 = 27$. Looking up the table of rank-sum critical values with $n_1 = 4$ and $n_2 = 5$, we find that

$$P(W \geq 27 \mid H_0) = 0.056.$$  

Therefore, we reject the null hypothesis at 0.056 level of significance.

Noether [NOET 1967] points out several reasons why in general the rank-sum test is preferable to the $t$ test. Since the rank-sum test is distribution free, whatever the true population distribution, as long as both samples come from the same population (i.e., $f_X = f_Y$), the significance level of the test is known. On the other hand, for nonnormal populations, the significance level of the $t$ test may differ considerably from the calculated value. In contrast to the $t$ test, the rank-sum test is not overly affected by large deviations (so called outliers). On the other hand, when there is sufficient justification for assuming that the population distribution is normal, it would be a mistake not to use that information [HOEL 1971].

Problems

1. Returning to the server-selection problem considered in problem 4 of Section 10.3.1, the sample means of the response times for the first two servers are 2.28 and 2.52 s, respectively. The sample size in both cases is 64, and the variances can be assumed to be 0.6 and 0.8 s$^2$, respectively. Test the hypothesis that the mean response times of the two servers are the same against the alternative that server 1 has a smaller mean response time.

2. In this section we assumed that $X$ and $Y$ samples were chosen independently. In practice, the observations often occur in pairs, $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. On the basis of pairwise differences $d_i = x_i - y_i$, construct a $t$ test using the statistic

$$T = \frac{D - d_0}{S_D / \sqrt{n}}$$