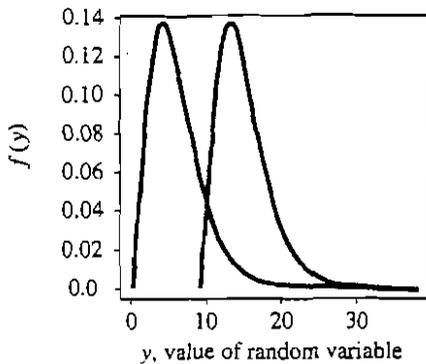


6.3 A Nonparametric Alternative: The Wilcoxon Rank Sum Test

The two-sample t test of the previous section was based on several conditions: independent samples, normality, and equal variances. When the conditions of normality and equal variances are not valid but the sample sizes are large, the

Wilcoxon rank sum test

FIGURE 6.7
Skewed population distributions identical in shape but shifted



results using a t (or t') test are approximately correct. There is, however, an alternative test procedure that requires less stringent conditions. This procedure, called the **Wilcoxon rank sum test**, is discussed here.

The assumptions for this test are that we have independent random samples taken from two populations whose distributions are identical except that one distribution may be shifted to the right of the other distribution, as shown in Figure 6.7. The Wilcoxon rank sum test does not require that populations have normal distributions. Thus, we have removed one of the three conditions that were required of the t -based procedures. The other conditions, equal variances and independence of the random samples, are still required for the Wilcoxon rank sum test. Because the two population distributions are assumed to be identical under the null hypothesis, independent random samples from the two populations should be similar if the null hypothesis is true. Because we are now allowing the population distributions to be nonnormal, the rank sum procedure must deal with the possibility of extreme observations in the data. One way to handle samples containing extreme values is to replace each data value with its rank (from lowest to highest) in the combined sample—that is, the sample consisting of the data from both populations. The smallest value in the combined sample is assigned the rank of 1 and the largest value is assigned the rank of $N = n_1 + n_2$. The ranks are not affected by how far the smallest (largest) data value is from next smallest (largest) data value. Thus, extreme values in data sets do not have a strong effect on the rank sum statistic as they did in the t -based procedures.

The calculation of the rank sum statistic consists of the following steps:

1. List the data values for both samples from smallest to largest.
2. In the next column, assign the numbers 1 to N to the data values with 1 to the smallest value and N to the largest value. These are the **ranks** of the observations.
3. If there are ties—that is, duplicated values—in the combined data set, the ranks for the observations in a tie are taken to be the average of the ranks for those observations.
4. Let T denote the sum of the ranks for the observations from population 1.

If the null hypothesis of identical population distributions is true, the n_1 ranks from population 1 are just a random sample from the N integers 1, . . . , N . Thus, under the null hypothesis, the distribution of the sum of the ranks T depends only on the sample sizes, n_1 and n_2 , and does not depend on the shape of the population distributions. Under the null hypothesis, the sampling distribution of T has mean and variance given by

$$\mu_T = \frac{n_1(n_1 + n_2 + 1)}{2} \quad \sigma_T^2 = \frac{n_1 n_2}{12} (n_1 + n_2 + 1)$$

Intuitively, if T is much smaller (or larger) than μ_T , we have evidence that the null hypothesis is false and in fact the population distributions are not equal. The rejection region for the rank sum test specifies the size of the difference between T and μ_T for the null hypothesis to be rejected. Because the distribution of T under the null hypothesis does not depend on the shape of the population distributions, Table 5 provides the critical values for the test regardless of the shape of the population distribution. The Wilcoxon rank sum test is summarized here.

Wilcoxon Rank Sum Test*

H_0 : The two populations are identical.

- H_a :
1. Population 1 is shifted to the right of population 2.
 2. Population 1 is shifted to the left of population 2.
 3. Populations 1 and 2 are shifted from each other.

$(n_1 \leq 10, n_2 \leq 10)$

T.S.: T , the sum of the ranks in sample 1

R.R.: For $\alpha = .05$, use Table 5 in the Appendix to find critical values for T_U and T_L ;

1. Reject H_0 if $T > T_U$.
2. Reject H_0 if $T < T_L$.
3. Reject H_0 if $T > T_U$ or $T < T_L$.

Wilcoxon Rank Sum Test: $n_1 > 10$ and $n_2 > 10$

T.S.: $z = \frac{T - \mu_T}{\sigma_T}$, where T denotes the sum of the ranks in sample 1.

R.R.: For a specified value of α ,

1. Reject H_0 if $z \geq z_{\alpha}$.
2. Reject H_0 if $z \leq -z_{\alpha}$.
3. Reject H_0 if $|z| \geq z_{\alpha/2}$.

Placebo	0.90	0.37	1.63	0.83	0.95	0.78	0.86	0.61	0.38	1.97
Alcohol	1.46	1.45	1.76	1.44	1.11	3.07	0.98	1.27	2.56	1.32

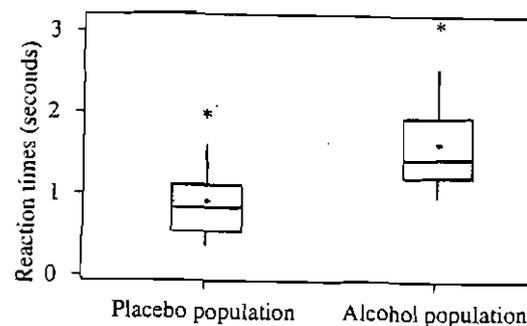
EXAMPLE 6.4

Many states are considering lowering the blood-alcohol level at which a driver is designated as driving under the influence (DUI) of alcohol. An investigator for a legislative committee designed the following test to study the effect of alcohol on reaction time. Ten participants consumed a specified amount of alcohol. Another group of ten participants consumed the same amount of a nonalcoholic drink, a placebo. The two groups did not know whether they were receiving alcohol or the placebo. The twenty participants' average reaction times (in seconds) to a series of simulated driving situations are reported in the following table. Does it appear that alcohol consumption increases reaction time?

- Why is the t test inappropriate for analyzing the data in this study?
- Use the Wilcoxon rank sum test to test the hypotheses:

H_0 : The distributions of reaction times for the placebo and alcohol populations are identical.

H_a : The distribution of reaction times for the placebo consumption population is shifted to the left of the distribution for the alcohol population. (Larger reaction times are associated with the consumption of alcohol.)



Solution

- A boxplot of the two samples is given here. The plots indicate that the population distributions are skewed to the right, because 10% of the data values are large outliers and the upper whiskers are longer than the lower whiskers. The sample sizes are both small, and hence the t test may be inappropriate for analyzing this study.
- The Wilcoxon rank sum test will be conducted to evaluate whether alcohol consumption increases reaction time. Table 6.6 contains the ordered data for the combined samples, along with their associated ranks. We will designate observations from the placebo group as 1 and from the alcohol group as 2.

For $\alpha = .05$, reject H_0 if $T < 83$, using Table 5 in the Appendix with $\alpha = .05$, one-tailed, and $n_1 = n_2 = 10$. The value of T is computed by summing the ranks from group 1: $T = 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 16 + 18 = 70$. Because 70 is less than 83, we reject H_0 and conclude there is significant evidence that the placebo population has smaller reaction times than the population of alcohol consumers.