Chapter 12

NONPARAMETRIC HYPOTHESIS TESTS

12.1 INTRODUCTION

In this chapter, we shall develop some hypothesis tests in situations where the data come from a probability distribution whose underlying form is not specified. That is, it will not be assumed that the underlying distribution is normal, or exponential, or any other given type. Because no particular parametric form for the underlying distribution is assumed, such tests are called *nonparametric*.

The strength of a nonparametric test resides in the fact that it can be applied without any assumption on the form of the underlying distribution. Of course, if there is justification for assuming a particular parametric form, such as the normal, then the relevant parametric test should be employed.

In Section 12.2, we consider hypotheses concerning the median of a continuous distribution and show how the *sign test* can be used in their study. In Section 12.3, we consider the *signed rank test*, which is used to test the hypothesis that a continuous population distribution is symmetric about a specified value. In Section 12.4, we consider the two-sample problem, where one wants to use data from two separate continuous distributions to test the hypothesis that the distributions are equal, and present the *rank sum test*. Finally, in Section 12.5 we study the *runs test*, which can be used to test the hypothesis that a sequence of 0's and 1's constitutes a random sequence that does not follow any specified pattern.

12.2 THE SIGN TEST

Let $X_1, \ldots, X_n$ denote a sample from a continuous distribution F and suppose that we are interested in testing the hypothesis that the median of F, call it m, is equal to a specified value m_0. That is, consider a test of

$$H_0 : m = m_0 \qquad \text{versus} \qquad H_1 : m \neq m_0$$

where m is such that $F(m) = .5$.

This hypothesis can easily be tested by noting that each of the observations will, independently, be less than m_0 with probability $F(m_0)$. Hence, if we let

$$I_i = \begin{cases} 1 & \text{if } X_i < m_0 \\ 0 & \text{if } X_i \geq m_0 \end{cases}$$

then $I_1, \ldots, I_n$ are independent Bernoulli random variables with parameter $p = F(m_0)$; and so the null hypothesis is equivalent to stating that this Bernoulli parameter is equal to $\frac{1}{2}$. Now, if v is the observed value of $\sum_{i=1}^{n} I_i$ — that is, if v is the number of data values less than m_0 — then it follows from the results of Section 8.6 that the p-value of the test that this Bernoulli parameter is equal to $\frac{1}{2}$ is

$$p\text{-value} = 2\min(P\{\text{Bin}(n, 1/2) \leq v\}, P\{\text{Bin}(n, 1/2) \geq v\}) \tag{12.2.1}$$

where $\text{Bin}(n, p)$ is a binomial random variable with parameters n and p.
However,

$$\begin{aligned} P\{\text{Bin}(n,p) \geq v\} &= P\{n - \text{Bin}(n,p) \leq n - v\} \\ &= P\{\text{Bin}(n, 1-p) \leq n - v\} \quad \text{(why?)} \end{aligned}$$

and so we see from Equation 12.2.1 that the p-value is given by

$$\begin{aligned} p\text{-value} &= 2\min(P\{\text{Bin}(n, 1/2) \leq v\}, P\{\text{Bin}(n, 1/2) \leq n - v\}) \qquad (12.2.2) \\ &= \begin{cases} 2P\{\text{Bin}(n, 1/2) \leq v\} & \text{if } v \leq \dfrac{n}{2} \\ 2P\{\text{Bin}(n, 1/2) \leq n - v\} & \text{if } v \geq \dfrac{n}{2} \end{cases} \end{aligned}$$

Since the value of $v = \sum_{i=1}^{n} I_i$ depends on the signs of the terms $X_i - m_0$, the foregoing is called the *sign test*.

EXAMPLE 12.2a If a sample of size 200 contains 120 values that are less than m_0 and 80 values that are greater, what is the p-value of the test of the hypothesis that the median is equal to m_0?

SOLUTION From Equation 12.2.2, the p-value is equal to twice the probability that binomial random variable with parameters 200, $\frac{1}{2}$ is less than or equal to 80.

The text disk shows that

$$P\{\text{Bin}(200, .5) \leq 80\} = .00284$$

Therefore, the p-value is .00568, and so the null hypothesis would be rejected at even the 1 percent level of significance. ■

The sign test can also be used in situations analogous to ones in which the paired t-test was previously applied. For instance, let us reconsider Example 8.4c, which is interested in testing whether or not a recently instituted industrial safety program has had an effect on the number of man-hours lost to accidents. For each of 10 plants, the data consisted of the pair X_i, Y_i, which represented, respectively, the average weekly loss at plant i before and after the program. Letting $Z_i = X_i - Y_i, i = 1, \ldots, 10$, it follows that if the program had not had any effect, then $Z_i, i = 1, \ldots, 10$, would be a sample from a distribution whose median value is 0. Since the resulting values of Z_i, — namely, $7.5, -2.3, 2.6, 3.7, 1.5, -.5, -1, 4.9, 4.8, 1.6$ — contain three whose sign is negative and seven whose sign is positive, it follows that the hypothesis that the median of Z is 0 should be rejected at significance level α if

$$\sum_{i=0}^{3} \binom{10}{i} \left(\frac{1}{2}\right)^{10} \leq \frac{\alpha}{2}$$

Since

$$\sum_{i=0}^{3} \binom{10}{i} \left(\frac{1}{2}\right)^{10} = \frac{176}{1{,}024} = .172$$

it follows that the hypothesis would be accepted at the 5 percent significance level (indeed, it would be accepted at all significance levels less than the p-value equal to .344).

Thus, the sign test does not enable us to conclude that the safety program has had any statistically significant effect, which is in contradiction to the result obtained in Example 8.4c when it was assumed that the differences were normally distributed. The reason for this disparity is that the assumption of normality allows us to take into account not only the number of values greater than 0 (which is all the sign test considers) but also the magnitude of these values. (The next test to be considered, while still being nonparametric, improves on the sign test by taking into account whether those values that most differ from the hypothesized median value m_0 tend to lie on one side of m_0 — that is, whether they tend to be primarily bigger or smaller than m_0.)

We can also use the sign test to test one-sided hypotheses about a population median. For instance, suppose that we want to test

$$H_0 : m \leq m_0 \qquad \text{versus} \qquad H_1 : m > m_0$$

where m is the population median and m_0 is some specified value. Let p denote the probability that a population value is less than m_0, and note that if the null hypothesis is true then $p \geq 1/2$, and if the alternative is true then $p < 1/2$ (see Figure 12.1).

To use the sign test to test the preceding hypothesis, choose a random sample of n members of the population. If v of them have values that are less than m_0, then the

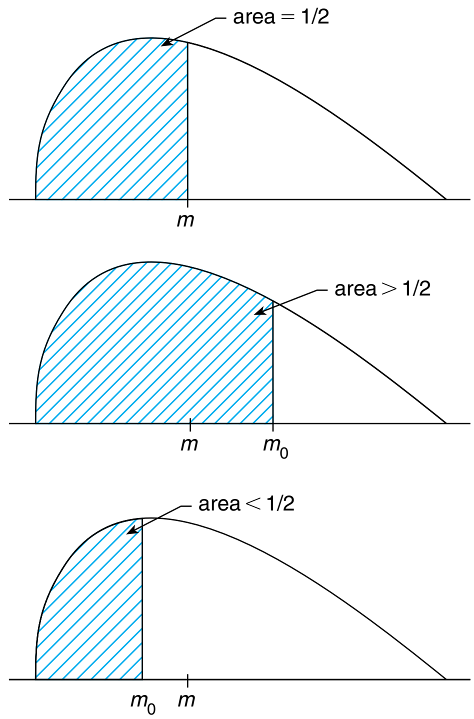

FIGURE 12.1

resulting p-value is the probability that a value of v or smaller would have occurred by chance if each element had probability 1/2 of being less than m_0. That is,

$$p\text{-value} = P\{\text{Bin}(n, 1/2) \le v\}$$

EXAMPLE 12.2b A financial institution has decided to open an office in a certain community if it can be established that the median annual income of families in the community is greater than \$90,000. To obtain information, a random sample of 80 families was chosen, and the family incomes determined. If 28 of these families had annual incomes below and 52 had annual incomes above \$90,000, is this significant enough to establish, say, at the 5 percent level of significance, that the median annual income in the community is greater than \$90,000?

SOLUTION We need to see if the data are sufficient to enable us to reject the null hypothesis when testing

$$H_0 : m \le 90 \qquad \text{versus} \qquad H_1 : m > 90$$

The preceding is equivalent to testing

$$H_0 : p \ge 1/2 \qquad \text{versus} \qquad H_1 : p < 1/2$$

where p is the probability that a randomly chosen member of the population has an annual income of less than \$90,000. Therefore, the p-value is

$$p\text{-value} = P\{\text{Bin}(80, 1/2) \leq 28\} = .0048$$

and so the null hypothesis that the median income is less than or equal to \$90,000 is rejected. ■

A test of the one-sided null hypothesis that the median is at least m_0 is obtained similarly. If a random sample of size n is chosen, and v of the resulting values are less than m_0, then the resulting p-value is

$$p\text{-value} = P\{\text{Bin}(n, 1/2) \geq v\}$$

12.3 THE SIGNED RANK TEST

The sign test can be employed to test the hypothesis that the median of a continuous distribution F is equal to a specified value m_0. However, in many applications one is really interested in testing not only that the median is equal to m_0 but that the distribution is symmetric about m_0. That is, if X has distribution function F, then one is often interested in testing the hypothesis $H_0 : P\{X < m_0 - a\} = P\{X > m_0 + a\}$ for all $a > 0$ (see Figure 12.2). Whereas the sign test could still be employed to test the foregoing hypothesis, it suffers in that it compares only the number of data values that are less than m_0 with the number that are greater than m_0 and does not take into account whether or not one of these sets tends to be farther away from m_0 than the other. A nonparametric test that does take this into account is the so-called *signed rank* test. It is described as follows.

Let $Y_i = X_i - m_0, i = 1, \ldots, n$ and rank (that is, order) the absolute values $|Y_1|, |Y_2|, \ldots, |Y_n|$. Set, for $j = 1, \ldots, n$,

$$I_j = \begin{cases} 1 & \text{if the } j\text{th smallest value comes from a data value that is smaller} \\ & \text{than } m_0 \\ 0 & \text{otherwise} \end{cases}$$

Now, whereas $\sum_{j=1}^{n} I_j$ represents the test statistic for the sign test, the signed rank test uses the statistic $T = \sum_{j=1}^{n} jI_j$. That is, like the sign test it considers those data values that are less than m_0, but rather than giving equal weight to each such value it gives larger weights to those data values that are farthest away from m_0.

EXAMPLE 12.3a If $n = 4$, $m_0 = 2$, and the data values are $X_1 = 4.2, X_2 = 1.8, X_3 = 5.3, X_4 = 1.7$, then the rankings of $|X_i - 2|$ are .2, .3, 2.2, 3.3. Since the first of these values — namely, .2 — comes from the data point X_2, which is less than 2, it follows that

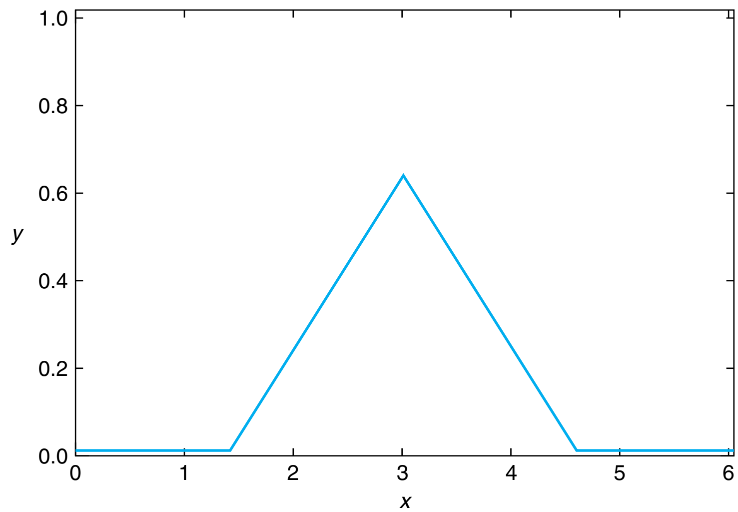

FIGURE 12.2 *A symmetric density: m = 3.*

$$f(x) = \begin{cases} \max\{0, .4(x-3) + \sqrt{.4}\} & x \le 3 \\ \max\{0, -.4(x-3) + \sqrt{.4}\} & x > 3 \end{cases}$$

$I_1 = 1$. Similarly, $I_2 = 1$, and I_3 and I_4 equal 0. Hence, the value of the test statistic is $T = 1 + 2 = 3$. ■

When H_0 is true, the mean and variance of the test statistic T are easily computed. This is accomplished by noting that, since the distribution of $Y_j = X_j - m_0$ is symmetric about 0, for any given value of $|Y_j|$ — say, $|Y_j| = y$ — it is equally likely that either $Y_j = y$ or $Y_j = -y$. From this fact it can be seen that under H_0, $I_1, \ldots, I_n$ will be independent random variables such that

$$P\{I_j = 1\} = \tfrac{1}{2} = P\{I_j = 0\}, \quad j = 1, \ldots, n$$

Hence, we can conclude that under H_0,

$$\begin{aligned} E[T] &= E\left[\sum_{j=1}^{n} jI_j\right] \\ &= \sum_{j=1}^{n} \frac{j}{2} = \frac{n(n+1)}{4} \end{aligned} \tag{12.3.1}$$

$$
\begin{aligned}
\operatorname{Var}(T) &= \operatorname{Var}\left(\sum_{j=1}^{n} jI_j\right) \\
&= \sum_{j=1}^{n} j^2 \operatorname{Var}(I_j) \\
&= \sum_{j=1}^{n} \frac{j^2}{4} = \frac{n(n+1)(2n+1)}{24}
\end{aligned} \tag{12.3.2}
$$

where the fact that the variance of the Bernoulli random variable I_j is $\frac{1}{2}(1-\frac{1}{2}) = \frac{1}{4}$ is used.

It can be shown that for moderately large values of n ($n > 25$ is often quoted as being sufficient) T will, when H_0 is true, have approximately a normal distribution with mean and variance as given by Equations 12.3.1 and 12.3.2. Although this approximation can be used to derive an approximate level α test of H_0 (which has been the usual approach until the recent advent of fast and cheap computational power), we shall not pursue this approach but rather will determine the p-value for given test data by an explicit computation of the relevant probabilities. This is accomplished as follows.

Suppose we desire a significance level α test of H_0. Since the alternative hypothesis is that the median is not equal to m_0, a two-sided test is called for. That is, if the observed value of T is equal to t, then H_0 should be rejected if either

$$
P_{H_0}\{T \le t\} \le \frac{\alpha}{2} \qquad \text{or} \qquad P_{H_0}\{T \ge t\} \le \frac{\alpha}{2} \tag{12.3.3}
$$

The p-value of the test data when $T = t$ is given by

$$
p\text{-value} = 2\min(P_{H_0}\{T \le t\}, P_{H_0}\{T \ge t\}) \tag{12.3.4}
$$

That is, if $T = t$, the signed rank test calls for rejection of the null hypothesis if the significance level α is at least as large as this p-value. The amount of computation necessary to compute the p-value can be reduced by utilizing the following equality (whose proof will be given at the end of the section).

$$
P_{H_0}\{T \ge t\} = P_{H_0}\left\{T \le \frac{n(n+1)}{2} - t\right\}
$$

Using Equation 12.3.4, the p-value is given by

$$
\begin{aligned}
p\text{-value} &= 2\min\left(P_{H_0}\{T \le t\}, P_{H_0}\left\{T \le \frac{n(n+1)}{2} - t\right\}\right) \\
&= 2P_{H_0}\{T \le t^*\}
\end{aligned}
$$

where

$$t^* = \min\left(t, \frac{n(n+1)}{2} - t\right)$$

It remains to compute $P_{H_0}\{T \le t^*\}$. To do so, let $P_k(i)$ denote the probability, under H_0, that the signed rank statistic T will be less than or equal to i when the sample size is k. We will determine a recursive formula for $P_k(i)$ starting with $k = 1$. When $k = 1$, since there is only a single data value, which, when H_0 is true, is equally likely to be either less than or greater than m_0, it follows that T is equally likely to be either 0 or 1. Thus

$$P_1(i) = \begin{cases} 0 & i < 0 \\ \frac{1}{2} & i = 0 \\ 1 & i \ge 1 \end{cases} \tag{12.3.5}$$

Now suppose the sample size is k. To compute $P_k(i)$, we condition on the value of I_k as follows:

$$\begin{aligned}
P_k(i) &= P_{H_0}\left\{\sum_{j=1}^{k} jI_j \le i\right\} \\
&= P_{H_0}\left\{\sum_{j=1}^{k} jI_j \le i|I_k = 1\right\} P_{H_0}\{I_k = 1\} \\
&\quad + P_{H_0}\left\{\sum_{j=1}^{k} jI_j \le i|I_k = 0\right\} P_{H_0}\{I_k = 0\} \\
&= P_{H_0}\left\{\sum_{j=1}^{k-1} jI_j \le i - k|I_k = 1\right\} P_{H_0}\{I_k = 1\} \\
&\quad + P_{H_0}\left\{\sum_{j=1}^{k-1} jI_j \le i|I_k = 0\right\} P_{H_0}\{I_k = 0\} \\
&= P_{H_0}\left\{\sum_{j=1}^{k-1} jI_j \le i - k\right\} P_{H_0}\{I_k = 1\} + P_{H_0}\left\{\sum_{j=1}^{k-1} jI_j \le i\right\} P_{H_0}\{I_k = 0\}
\end{aligned}$$

where the last equality utilized the independence of $I_1, \dots, I_{k-1}$ and I_k (when H_0 is true). Now $\sum_{j=1}^{k-1} jI_j$ has the same distribution as the signed rank statistic of a sample of size $k - 1$, and since

$$P_{H_0}\{I_k = 1\} = P_{H_0}\{I_k = 0\} = \tfrac{1}{2}$$

we see that

$$P_k(i) = \tfrac{1}{2}P_{k-1}(i-k) + \tfrac{1}{2}P_{k-1}(i) \tag{12.3.6}$$

Starting with Equation 12.3.5, the recursion given by Equation 12.3.6 can be successfully employed to compute $P_2(\cdot)$, then $P_3(\cdot)$, and so on, stopping when the desired value $P_n(t^*)$ has been obtained.

EXAMPLE 12.3b For the data of Example 12.3a,

$$t^* = \min\left(3, \frac{4 \cdot 5}{2} - 3\right) = 3$$

Hence the p-value is $2P_4(3)$, which is computed as follows:

$$
\begin{aligned}
P_2(0) &= \tfrac{1}{2}[P_1(-2) + P_1(0)] = \tfrac{1}{4} \\
P_2(1) &= \tfrac{1}{2}[P_1(-1) + P_1(1)] = \tfrac{1}{2} \\
P_2(2) &= \tfrac{1}{2}[P_1(0) + P_1(2)] = \tfrac{3}{4} \\
P_2(3) &= \tfrac{1}{2}[P_1(1) + P_1(3)] = 1 \\
P_3(0) &= \tfrac{1}{2}[P_2(-3) + P_2(0)] = \tfrac{1}{8} \qquad \text{since } P_2(-3) = 0 \\
P_3(1) &= \tfrac{1}{2}[P_2(-2) + P_2(1)] = \tfrac{1}{4} \\
P_3(2) &= \tfrac{1}{2}[P_2(-1) + P_2(2)] = \tfrac{3}{8} \\
P_3(3) &= \tfrac{1}{2}[P_2(0) + P_2(3)] = \tfrac{5}{8} \\
P_4(0) &= \tfrac{1}{2}[P_3(-4) + P_3(0)] = \tfrac{1}{16} \\
P_4(1) &= \tfrac{1}{2}[P_3(-3) + P_3(1)] = \tfrac{1}{8} \\
P_4(2) &= \tfrac{1}{2}[P_3(-2) + P_3(2)] = \tfrac{3}{16} \\
P_4(3) &= \tfrac{1}{2}[P_3(-1) + P_3(3)] = \tfrac{5}{16} \quad \blacksquare
\end{aligned}
$$

Program 12.3 will use the recursion in Equations 12.3.5 and 12.3.6 to compute the p-value of the signed rank test data. The input needed is the sample size n and the value of test statistic T.

EXAMPLE 12.3c Suppose we are interested in determining whether a certain population has an underlying probability distribution that is symmetric about 0. If a sample of size 20 from this population results in a signed rank test statistic of value 142, what conclusion can we draw at the 10 percent level of significance?

SOLUTION Running Program 12.3 yields that

$$p\text{-value} = .177$$

Thus the hypothesis that the population distribution is symmetric about 0 is accepted at the $\alpha = .10$ level of significance. ■

We end this section with a proof of the equality

$$P_{H_0}\{T \geq t\} = P_{H_0}\left\{T \leq \frac{n(n+1)}{2} - t\right\}$$

To verify the foregoing, note first that $1 - I_j$ will equal 1 if the jth smallest value of $|Y_1|, \ldots, |Y_n|$ comes from a data value larger than m_0, and it will equal 0 otherwise. Hence, if we let

$$T^1 = \sum_{j=1}^{n} j(1 - I_j)$$

then T^1 will represent the sum of the ranks of the $|Y_j|$ that correspond to data values larger than m_0. By symmetry, T^1 will have, under H_0, the same distribution as T. Now

$$T^1 = \sum_{j=1}^{n} j - \sum_{j=1}^{n} jI_j = \frac{n(n+1)}{2} - T$$

and so

$$\begin{aligned} P\{T \geq t\} &= P\{T^1 \geq t\} \qquad \text{since } T \text{ and } T^1 \text{ have the same distribution} \\ &= P\left\{\frac{n(n+1)}{2} - T \geq t\right\} \\ &= P\left\{T \leq \frac{n(n+1)}{2} - t\right\} \end{aligned}$$

REMARK ON TIES

Since we have assumed that the population distribution is continuous, there is no possibility of ties — that is, with probability 1, all observations will have different values. However, since in practice all measurements are quantized, ties are always a distinct possibility. If ties do occur, then the weights given to the values less than m_0 should be the average of the different weights they could have had if the values had differed slightly. For instance, if $m_0 = 0$ and the data values are 2, 4, 7, −5, −7, then the ordered absolute values are 2, 4, 5, 7, 7. Since 7 has rank both 4 and 5, the value of the test statistic

T is $T = 3 + 4.5 = 7.5$. The p-value should be computed as when we assumed that all values were distinct. (Although technically this is not correct, the discrepancy is usually minor.)

12.4 THE TWO-SAMPLE PROBLEM

Suppose that one is considering two different methods for producing items having measurable characteristics with an interest in determining whether the two methods result in statistically identical items.

To attack this problem let $X_1, \ldots, X_n$ denote a sample of the measurable values of n items produced by method 1, and, similarly, let $Y_1, \ldots, Y_m$ be the corresponding value of m items produced by method 2. If we let F and G, both assumed to be continuous, denote the distribution functions of the two samples, respectively, then the hypothesis we wish to test is $H_0 : F = G$.

One procedure for testing H_0 — which is known by such names as the rank sum test, the Mann-Whitney test, or the Wilcoxon test — calls initially for ranking, or ordering, the $n + m$ data values $X_1, \ldots, X_n, Y_1, \ldots, Y_m$. Since we are assuming that F and G are continuous, this ranking will be unique — that is, there will be no ties. Give the smallest data value rank 1, the second smallest rank 2, . . . , and the $(n + m)$th smallest rank $n + m$. Now, for $i = 1, \ldots, n$, let

$$R_i = \text{rank of the data value } X_i$$

The rank sum test utilizes the test statistic T equal to the sum of the ranks from the first sample — that is,

$$T = \sum_{i=1}^{n} R_i$$

EXAMPLE 12.4a An experiment designed to compare two treatments against corrosion yielded the following data in pieces of wire subjected to the two treatments.

Treatment 1	65.2, 67.1, 69.4, 78.2, 74, 80.3
Treatment 2	59.4, 72.1, 68, 66.2, 58.5

(The data represent the maximum depth of pits in units of one thousandth of an inch.) The ordered values are 58.5, 59.4, 65.2*, 66.2, 67.1*, 68, 69.4*, 72.1, 74*, 78.2*, 80.3* with an asterisk noting that the data value was from sample 1. Hence, the value of the test statistic is $T = 3 + 5 + 7 + 9 + 10 + 11 = 45$. ■

Suppose that we desire a significance level α test of H_0. If the observed value of T is $T = t$, then H_0 should be rejected if either

$$P_{H_0}\{T \le t\} \le \frac{\alpha}{2} \qquad \text{or} \qquad P_{H_0}\{T \ge t\} \le \frac{\alpha}{2} \tag{12.4.1}$$

That is, the hypothesis that the two samples are equivalent should be rejected if the sum of the ranks from the first sample is either too small or too large to be explained by chance.

Since for integral t,

$$\begin{aligned} P\{T \ge t\} &= 1 - P\{T < t\} \\ &= 1 - P\{T \le t - 1\} \end{aligned}$$

it follows from Equation 12.4.1 that H_0 should be rejected if either

$$P_{H_0}\{T \le t\} \le \frac{\alpha}{2} \qquad \text{or} \qquad P_{H_0}\{T \le t-1\} \ge 1 - \frac{\alpha}{2} \tag{12.4.2}$$

To compute the probabilities in Equation 12.4.2, let $P(N, M, K)$ denote the probability that the sum of the ranks of the first sample will be less than or equal to K when the sample sizes are N and M and H_0 is true. We will now determine a recursive formula for $P(N, M, K)$, which will then allow us to obtain the desired quantities $P(n, m, t) = P_{H_0}\{T \le t\}$ and $P(n, m, t - 1)$.

To compute the probability that the sum of the ranks of the first sample is less than or equal to K when N and M are the sample sizes and H_0 is true, let us condition on whether the largest of the $N + M$ data values belongs to the first or second sample. If it belongs to the first sample, then the sum of the ranks of this sample is equal to $N + M$ plus the sum of the ranks of the other $N - 1$ values from the first sample. Hence this sum will be less than or equal to K if the sum of the ranks of the other $N - 1$ values is less than or equal to $K - (N + M)$. But since the remaining $N - 1 + M$ — that is, all but the largest — values all come from the same distribution (when H_0 is true), it follows that the sum of the ranks of $N - 1$ of them will be less than or equal to $K - (N + M)$ with probability $P(N - 1, M, K - N - M)$. By a similar argument we can show that, given that the largest value is from the second sample, the sum of the ranks of the first sample will be less than or equal to K with probability $P(N, M - 1, K)$. Also, since the largest value is equally likely to be any of the $N + M$ values $X_1, \ldots, X_N, Y_1, \ldots, Y_M$, it follows that it will come from the first sample with probability $N/(N + M)$. Putting these together, we thus obtain that

$$\begin{aligned} P(N, M, K) = {} & \frac{N}{N+M} P(N-1, M, K-N-M) \\ & + \frac{M}{N+M} P(N, M-1, K) \end{aligned} \tag{12.4.3}$$

Starting with the boundary condition

$$P(1,0,K) = \begin{cases} 0 & K \le 0 \\ 1 & K > 0 \end{cases}, \qquad P(0,1,K) = \begin{cases} 0 & K < 0 \\ 1 & K \ge 0 \end{cases}$$

Equation 12.4.3 can be solved recursively to obtain $P(n, m, t-1)$ and $P(n, m, t)$.

EXAMPLE 12.4b Suppose we wanted to determine $P(2, 1, 3)$. We use Equation 12.4.3 as follows:

$$P(2,1,3) = \tfrac{2}{3}P(1,1,0) + \tfrac{1}{3}P(2,0,3)$$

and

$$\begin{aligned} P(1,1,0) &= \tfrac{1}{2}P(0,1,-2) + \tfrac{1}{2}P(1,0,0) = 0 \\ P(2,0,3) &= P(1,0,1) \\ &= P(0,0,0) = 1 \end{aligned}$$

Hence,

$$P(2,1,3) = \tfrac{1}{3}$$

which checks since in order for the sum of the ranks of the two X values to be less than or equal to 3, the largest of the values X_1, X_2, Y_1, must be Y_1, which, when H_0 is true, has probability $\frac{1}{3}$. ■

Since the rank sum test calls for rejection when either

$$2P(n,m,t) \le \alpha \qquad \text{or} \qquad \alpha \ge 2[1 - P(n,m,t-1)]$$

it follows that the p-value of the test statistic when $T = t$ is

$$p\text{-value} = 2\min\{P(n,m,t), 1 - P(n,m,t-1)\}$$

Program 12.4 uses the recursion in Equation 12.4.3 to compute the p-value for the rank sum test. The input needed is the sizes of the first and second samples and the sum of the ranks of the elements of the first sample. Whereas either sample can be designated as the first sample, the program will run fastest if the first sample is the one whose sum of ranks is smallest.

EXAMPLE 12.4c In Example 12.4a, the sizes of the two samples are 5 and 6, respectively, and the sum of the ranks of the first sample is 21. Running Program 12.4 yields the result:

$$p\text{-value} = .1255 \quad ■$$

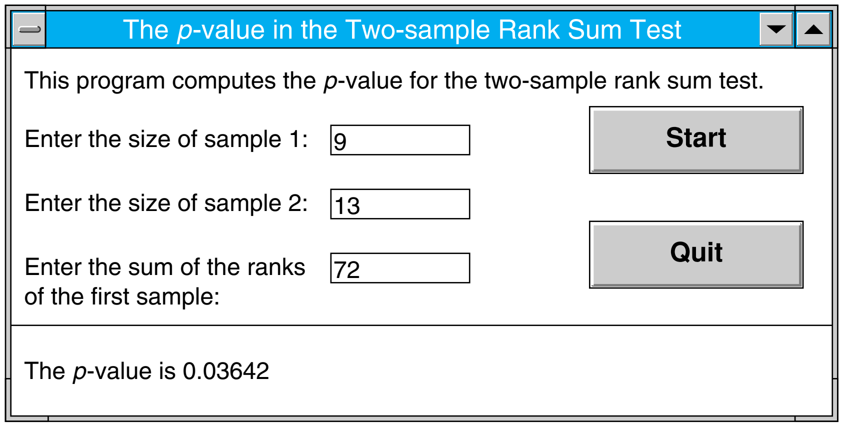

FIGURE 12.3

EXAMPLE 12.4d Suppose that in testing whether 2 production methods yield identical results, 9 items are produced using the first method and 13 using the second. If, among all 22 items, the sum of the ranks of the 9 items produced by method 1 is 72, what conclusions would you draw?

SOLUTION Run Program 12.4 to obtain the result shown in Figure 12.3. Thus, the hypothesis of identical results would be rejected at the 5 percent level of significance. ■

It remains to compute the value of the test statistic T. It is quite efficient to compute T directly by first using a standard computer science algorithm (such as quicksort) to sort, or order, the $n + m$ values. Another approach, easily programmed, although efficient for only small values of n and m, uses the following identity.

PROPOSITION 12.4.1 For $i = 1, \ldots, n, j = 1, \ldots, m$, let

$$W_{ij} = \begin{cases} 1 & \text{if } X_i > Y_j \\ 0 & \text{otherwise} \end{cases}$$

Then

$$T = \frac{n(n+1)}{2} + \sum_{i=1}^{n} \sum_{j=1}^{m} W_{ij}$$

Proof

Consider the values $X_1, \ldots, X_n$ of the first sample and order them. Let $X_{(i)}$ denote the ith smallest, $i = 1, \ldots, n$. Now consider the rank of $X_{(i)}$ among all $n + m$ data values.

This is given by

$$\text{rank of } X_{(i)} = i + \text{ number } j\colon Y_j < X_{(i)}$$

Summing over i gives

$$\sum_{i=1}^{n} \text{rank of } X_{(i)} = \sum_{i=1}^{n} i + \sum_{i=1}^{n} (\text{number } j\colon Y_j < X_{(i)}) \tag{12.4.4}$$

But since the order in which we add terms does not change the sum obtained, we see that

$$\sum_{i=1}^{n} \text{rank of } X_{(i)} = \sum_{i=1}^{n} \text{rank of } X_i = T \tag{12.4.5}$$

$$\sum_{i=1}^{n} (\text{number } j : Y_j < X_{(i)}) = \sum_{i=1}^{n} (\text{number } j : Y_j < X_i)$$

Hence, from Equations 12.4.4 and 12.4.5, we obtain that

$$\begin{aligned} T &= \sum_{i=1}^{n} i + \sum_{i=1}^{n} (\text{number } j : Y_j < X_i) \\ &= \frac{n(n+1)}{2} + \sum_{i=1}^{n} \sum_{j=1}^{m} W_{ij} \quad \blacksquare \end{aligned}$$

*12.4.1 The Classical Approximation and Simulation

The difficulty with employing the recursion in Equation 12.4.3 to compute the p-value of the two-sample sum of rank test statistic is that the amount of computation grows enormously as the sample sizes increase. For instance, if $n = m = 200$, then even if we choose the test statistic to be the smaller sum of ranks, since the sum of all the ranks is $1+2+\cdots+400 = 80{,}200$, it is possible that the test statistic could have a value as large as 40,100. Hence, there can be as many as 1.604×10^9 values of $P(N, M, K)$ that would have to be computed to determine the p-value. Thus, for large sample sizes the approach based on the recursion in Equation 12.4.3 is not viable. Two approximate methods that can be utilized in such cases are (a) a classical method based on approximating the distribution of the test statistic and (b) simulation.

(a) *The Classical Approximation* When the null hypothesis is true and so $F = G$, it follows that all $n + m$ data values come from the same distribution and thus all $(n + m)!$ possible rankings of the values $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ are equally likely.

* Simulation will be covered in Chapter 15.

From this it follows that choosing the n rankings of the first sample is probabilistically equivalent to randomly choosing n of the (possible rank) values $1, 2, \ldots,$ $n + m$. Using this, it can be shown that T has a mean and variance given by

$$E_{H_0}[T] = \frac{n(n+m+1)}{2}$$
$$\mathrm{Var}_{H_0}(T) = \frac{nm(n+m+1)}{12}$$

In addition, it can be shown that when both n and m are of moderate size (both being greater than 7 should suffice) T has, under H_0, approximately a normal distribution. Hence, when H_0 is true

$$\frac{T - \dfrac{n(n+m+1)}{2}}{\sqrt{\dfrac{nm(n+m+1)}{12}}} \mathrel{\dot\sim} \mathcal{N}(0,1) \tag{12.4.6}$$

If we let d denote the absolute value of the difference between the observed value of T and its mean value given above, then based on Equation 12.4.6 the approximate p-value is

$$\begin{aligned}
p\text{-value} &= P_{H_0}\{|T - E_{H_0}[T]| > d\} \\
&\approx P\left\{|Z| > d\Big/\sqrt{\frac{nm(n+m+1)}{12}}\right\} \qquad \text{where } Z \sim \mathcal{N}(0,1) \\
&= 2P\left\{Z > d\Big/\sqrt{\frac{nm(n+m+1)}{12}}\right\}
\end{aligned}$$

EXAMPLE 12.4e In Example 12.4a, $n = 5, m = 6$, and the test statistic's value is 21. Since

$$\frac{n(n+m+1)}{2} = 30$$
$$\frac{nm(n+m+1)}{12} = 30$$

we have that $d = 9$ and so

$$\begin{aligned}
p\text{-value} &\approx 2P\left\{Z > \frac{9}{\sqrt{30}}\right\} \\
&= 2P\{Z > 1.643108\}
\end{aligned}$$

$$= 2(1 - .9498)$$
$$= .1004$$

which can be compared with the exact value, as given in Example 12.4c, of .1225.

In Example 12.4d, $n = 9$, $m = 13$, and so

$$\frac{n(n+m+1)}{2} = 103.5$$
$$\frac{nm(n+m+1)}{12} = 224.25$$

Since $T = 72$, we have that

$$d = |72 - 103.5| = 31.5$$

Thus, the approximate p-value is

$$p\text{-value} \approx 2P\left\{Z > \frac{31.5}{\sqrt{224.25}}\right\}$$
$$= 2P\{Z > 2.103509\}$$
$$= 2(1 - .9823) = .0354$$

which is quite close to the exact p-value (as given in Example 12.4d) of .0364.

Thus, in the two examples considered, the normal approximation worked quite well in the second example — where the guideline that both sample sizes should exceed 7 held — and not so well in the first example — where the guideline did not hold. ■

(b) *Simulation* If the observed value of the test statistic is $T = t$, then the p-value is given by

$$p\text{-value} = 2 \min\left\{P_{H_0}\{T \geq t\}, P_{H_0}\{T \leq t\}\right\}$$

We can approximate this value by continually simulating a random selection of n of the values $1, 2, \ldots, n+m$ — noting on each occasion the sum of the n values. The value of $P_{H_0}\{T \geq t\}$ can be approximated by the proportion of time that the sum obtained is greater than or equal to t, and $P_{H_0}\{T \leq t\}$ by the proportion of time that it is less than or equal to t.

A Chapter 12 text disk program approximates the p-value by performing the preceding simulation. The program will run most efficiently when the sample of smallest size is designated as the first sample.

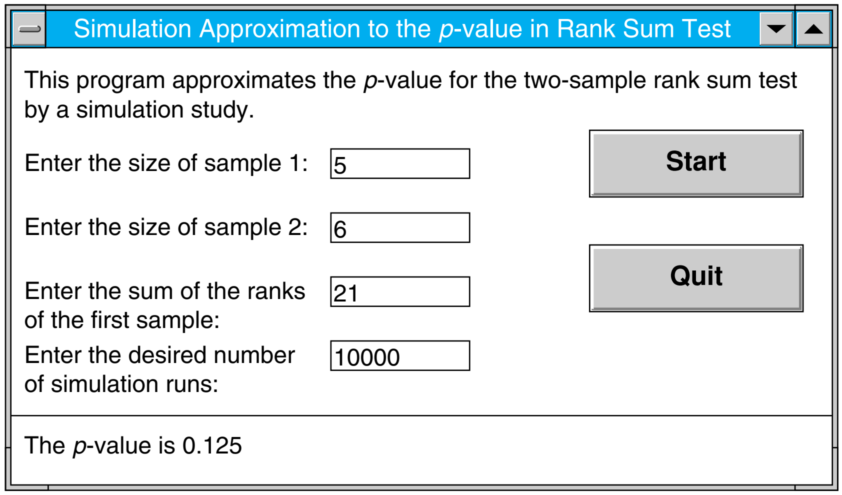

FIGURE 12.4

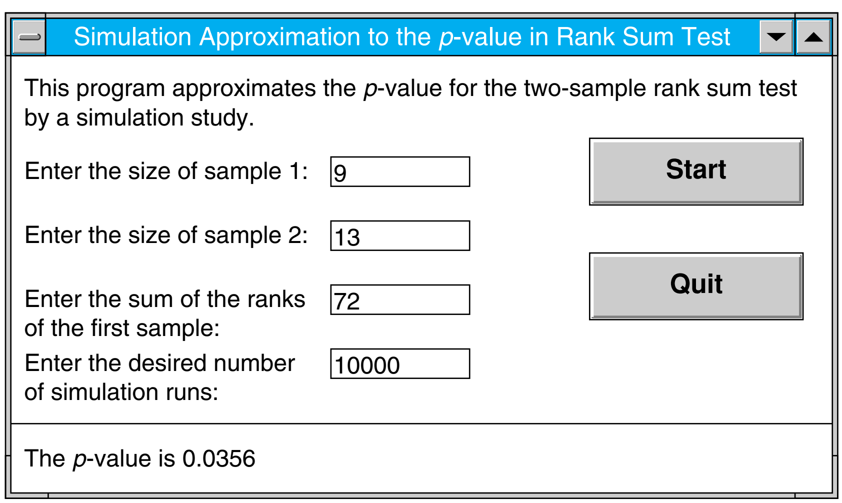

FIGURE 12.5

EXAMPLE 12.4f Running the text disk program on the data of Example 12.4c yields Figure 12.4, which is quite close to the exact value of .1225. Running the program using the data of Example 12.4d yields Figure 12.5, which is again quite close to the exact value of .0364. ■

Both of the approximation methods work quite well. The normal approximation, when n and m both exceed 7, is usually quite accurate and requires almost no computational time. The simulation approach, on the other hand, can require a great deal of computational time. However, if an immediate answer is not required and great accuracy is desired, then simulation, by running a large number of cases, can be made accurate to an arbitrarily prescribed precision.

12.4.2 Testing the Equality of Multiple Probability Distributions

Whereas the preceding sections showed how to test the hypothesis that two population distributions are identical, we are sometimes faced with the situation where there are more than two populations. So suppose there are k populations, that F_i is the distribution function of some measurable value of the elements of population i, and that we are interested in testing the null hypothesis

$$H_0: \quad F_1 = F_2 = \cdots = F_k$$

against the alternative

$$H_1: \quad \text{not all of the } F_i \text{ are equal}$$

To test the preceding null hypothesis, suppose that independent samples are drawn from each of the k populations. Let n_i denote the size of the sample chosen from population $i, i = 1, \ldots, k$. and let $N = \sum_{i=1}^{k} n_i$ denote the total number of data values obtained. Now, rank these N data values from the smallest to largest, and let R_i denote the sum of the ranks of the n_i data values from population i, $i = 1, \ldots, k$.

Now, when H_0 is true, the rank of any individual data value is equally likely to be any of the ranks $1, \ldots, N$, and thus the expected value of its rank is $\frac{1+2+\cdots+N}{N} = \frac{N+1}{2}$. Consequently, with $\bar{r} = \frac{N+1}{2}$, it follows when H_0 is true that the expected sum of the ranks of the n_i data values from population i is $n_i\bar{r}$. That is, when H_0 is true

$$E[R_i] = n_i\bar{r}.$$

Drawing inspiration from the goodness of fit test, let us consider the test statistic

$$T = \sum_{i=1}^{k} \frac{(R_i - n_i\bar{r})^2}{n_i\bar{r}}$$

and use a test that rejects the null hypothesis when T is large. Now,

$$\begin{aligned} T &= \frac{1}{\bar{r}} \sum_{i=1}^{k} \frac{R_i^2 - 2R_i n_i\bar{r} + n_i^2\bar{r}^2}{n_i} \\ &= \frac{1}{\bar{r}} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 2\sum_{i=1}^{k} R_i + \bar{r}\sum_{i=1}^{k} n_i \\ &= \frac{1}{\bar{r}} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - N\bar{r} \end{aligned}$$

where the final equality used that $\sum_{i=1}^{k} R_i$ is the sum of the ranks of all $N = \sum_i n_i$ data values and so

$$\sum_{i=1}^{k} R_i = 1 + 2 + \cdots + N = \frac{N(N+1)}{2} = N\bar{r}$$

Hence, rejecting H_0 when T is large is equivalent to rejecting H_0 when $\sum_{i=1}^{k} R_i^2/n_i$ is large. So we might as well let the test statistic be

$$TS = \sum_{i=1}^{k} \frac{R_i^2}{n_i}$$

To determine the appropriate α level significance test, we need the distribution of TS when H_0 is true. While its exact distribution is rather complicated, we can use the result that, when H_0 is true and all n_i are at least 5, the distribution of

$$\frac{12}{N(N+1)} TS - 3(N+1)$$

is approximately that of a chi-squared random variable with $k-1$ degrees of freedom. Using this, we see that an approximate significance level α test of the null hypothesis that all distributions are identical is to

$$\text{reject } H_0 \quad \text{if} \quad \frac{12}{N(N+1)} TS - 3(N+1) \geq \chi^2_{k-1,\alpha}$$

For even more accuracy, simulation can be used. To implement it, one should first compute the value of TS, say that it is equal to t. The resulting p-value is

$$p\text{-value} = P_{H_0}\{TS \geq t\}$$

To determine the preceding by a simulation, one should continually simulate $TS = \sum_i R_i^2/n_i$ under the assumption that H_0 is true and then use the fraction of the simulated values that are at least t as the estimate of the p-value. Because, under H_0, all possible orderings of the N data values are equal likely, one could simulate TS by generating a permutation of $1, \ldots, N$ that is equally likely to be any of the $N!$ permutations. (See Example 15.2b for an efficient way to simulate such a *random permutation*.) One can then let the first n_1 values of the permutation be the ranks of the first sample, the next n_2 values be the ranks of the second sample, and so on. That is, if $p_1, \ldots, p_N$ is the generated value of the permutation, then with $s_0 = 0$, $s_j = n_1 + \cdots + n_j$, $j \geq 1$, the simulated values of $R_1, \ldots, R_k$ would be

$$R_i = \sum_{j=s_{i-1}+1}^{s_i} p_j, \quad i = 1, \ldots, k$$

The preceding is known as the *Kruskal–Wallis* test.

EXAMPLE 12.4g The following data give the number of visitors to a medium size Los Angeles library on Tuesdays, Wednesdays, and Thursdays of 10 successive weeks.

Tuesday visitors	721, 660, 622, 738, 820, 707, 672, 589, 902, 688
Wednesday visitors	604, 626, 744, 802, 691, 665, 711, 715, 661, 729
Thursday visitors	642, 480, 705, 584, 713, 654, 704, 522, 683, 708

Are these data consistent with the hypothesis that the distributions of the number of visitors for the three days are identical?

SOLUTION Ordering the $N = 30$ data values, gives that the sum of the ranks of the three samples are

$$R_1 = 176, \quad R_2 = 175, \quad R_3 = 114$$

Therefore,

$$\frac{12}{N(N+1)} TS - 3(N+1) = \frac{12}{30 \cdot 31} \frac{176^2 + 175^2 + 114^2}{10} - 93 = 3.254$$

Because $\chi^2_{2,.05} = 5.99$, the null hypothesis that the distributions of the number of visitors for each of the three weekdays are identical cannot be rejected at the 5 percent level of significance. Indeed, the resulting p-value is

$$p\text{-value} = P\{\chi^2_2 \geq 3.254\} = .1965 \quad \blacksquare$$

12.5 THE RUNS TEST FOR RANDOMNESS

A basic assumption in much of statistics is that a set of data constitutes a random sample from some population. However, it is sometimes the case that the data are not generated by a truly random process but by one that may follow a trend or a type of cyclical pattern. In this section, we will consider a test — called the runs test — of the hypothesis H_0 that a given data set constitutes a random sample.

To begin, let us suppose that each of the data values is either a 0 or a 1. That is, we shall assume that each data value can be dichotomized as being either a success or a failure. Let $X_1, \ldots, X_N$ denote the set of data. Any consecutive sequence of either 1's or 0's is called a *run*. For instance, the data set

$$1\ 0\ 0\ 1\ 1\ 1\ 0\ 0\ 1\ 0\ 1\ 1\ 1\ 1\ 0\ 1\ 0\ 0\ 0\ 0\ 1\ 1$$

contains 11 runs — 6 runs of 1 and 5 runs of 0. Suppose that the data set $X_1, \ldots, X_N$ contains n 1's and m 0's, where $n + m = N$, and let R denote the number of runs. Now, if H_0 were true, then $X_1, \ldots, X_N$ would be equally likely to be any of the $N!/(n!m!)$ permutations of n 1's and m 0's, and therefore, given a total of n 1's and m 0's, it follows that, under H_0, the probability mass function of R, the number of runs is given by

$$P_{H_0}\{R = k\} = \frac{\text{number of permutations of } n \text{ 1's and } m \text{ 0's resulting in } k \text{ runs}}{\binom{n+m}{n}}$$

This number of permutations can be explicitly determined and it can be shown that

$$\begin{aligned} P_{H_0}\{R = 2k\} &= 2\frac{\binom{m-1}{k-1}\binom{n-1}{k-1}}{\binom{m+n}{n}} \\ P_{H_0}\{R = 2k+1\} &= \frac{\binom{m-1}{k-1}\binom{n-1}{k} + \binom{m-1}{k}\binom{n-1}{k-1}}{\binom{n+m}{n}} \end{aligned} \tag{12.5.1}$$

If the data contain n 1's and m 0's, then the runs test calls for rejection of the hypothesis that the data constitutes a random sample if the observed number of runs is either too large or too small to be explained by chance. Specifically, if the observed number of runs is r, then the p-value of the runs test is

$$p\text{-value} = 2\min(P_{H_0}\{R \geq r\}, P_{H_0}\{R \leq r\})$$

Program 12.5 uses Equation 12.5.1 to compute the p-value.

EXAMPLE 12.5a The following is the result of the last 30 games played by an athletic team, with W signifying a win and L a loss.

$$W\ W\ W\ L\ W\ W\ W\ L\ W\ W\ L\ W\ L\ W\ W\ L\ W\ W\ W\ W\ L\ W\ L\ W\ W\ W\ L\ W\ L\ W\ L$$

Are these data consistent with pure randomness?

SOLUTION To test the hypothesis of randomness, note that the data, which consist of 20 W's and 10 L's, contain 20 runs. To see whether this justifies rejection at, say, the 5 percent level of significance, we run Program 12.5 and observe the results in Figure 12.6. Therefore, the hypothesis of randomness would be rejected at the 5 percent level of significance. (The striking thing about these data is that the team always came back to win after losing a game, which would be quite unlikely if all outcomes containing 20 wins and 10 losses were equally likely.) ■

The above can also be used to test for randomness when the data values are not just 0's and 1's. To test whether the data $X_1, \ldots, X_N$ constitute a random sample, let s-med denote the sample median. Also let n denote the number of data values that are less than

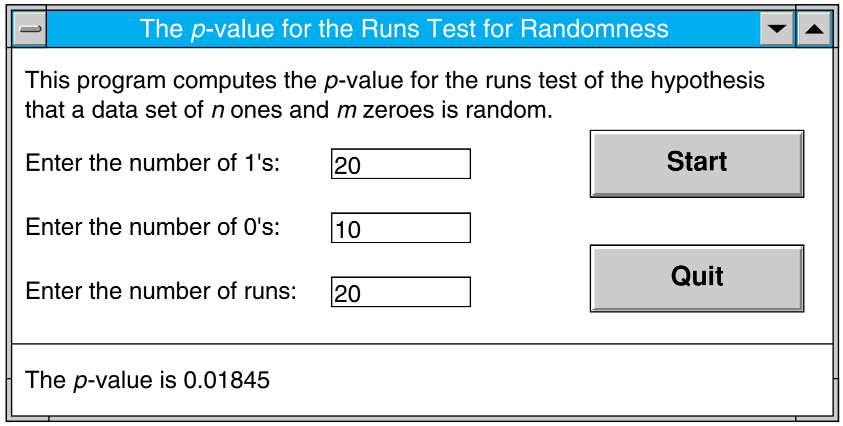

FIGURE 12.6

or equal to s-med and m the number that are greater. (Thus, if N is even and all data values are distinct, then $n = m = N/2$.) Define $I_1, \ldots, I_N$ by

$$I_j = \begin{cases} 1 & \text{if } X_j \leq s\text{-med} \\ 0 & \text{otherwise} \end{cases}$$

Now, if the original data constituted a random sample, then the number of runs in $I_1, \ldots, I_N$ would have a probability mass function given by Equation 12.5.1. Thus, it follows that we can use the preceding runs test on the data values $I_1, \ldots, I_N$ to test that the original data are random.

EXAMPLE 12.5b The lifetime of 19 successively produced storage batteries is as follows:

145 152 148 155 176 134 184 132 145 162 165
185 174 198 179 194 201 169 182

The sample median is the 10th smallest value — namely, 169. The data indicating whether the successive values are less than or equal to or greater than 169 are as follows:

1 1 1 1 0 1 0 1 1 1 1 0 0 0 0 0 0 1 0

Hence, the number of runs is 8. To determine if this value is statistically significant, we run Program 12.5 (with $n = 10, m = 9$) to obtain the result:

$$p\text{-value} = .357$$

Thus the hypothesis of randomness is accepted. ■

It can be shown that, when n and m are both large and H_0 is true, R will have approximately a normal distribution with mean and standard deviation given by

$$\mu = \frac{2nm}{n+m} + 1 \qquad \text{and} \qquad \sigma = \sqrt{\frac{2nm(2nm-n-m)}{(n+m)^2(n+m-1)}} \tag{12.5.2}$$

Therefore, when n and m are both large

$$\begin{aligned} P_{H_0}\{R \le r\} &= P_{H_0}\left\{\frac{R-\mu}{\sigma} \le \frac{r-\mu}{\sigma}\right\} \\ &\approx P\left\{Z \le \frac{r-\mu}{\sigma}\right\}, \qquad Z \sim \mathcal{N}(0,1) \\ &= \Phi\left(\frac{r-\mu}{\sigma}\right) \end{aligned}$$

and, similarly,

$$P_{H_0}\{R \ge r\} \approx 1 - \Phi\left(\frac{r-\mu}{\sigma}\right)$$

Hence, for large n and m, the p-value of the runs test for randomness is approximately given by

$$p\text{-value} \approx 2\min\left\{\Phi\left(\frac{r-\mu}{\sigma}\right), 1 - \Phi\left(\frac{r-\mu}{\sigma}\right)\right\}$$

where μ and σ are given by Equation 12.5.2 and r is the observed number of runs.

EXAMPLE 12.5c Suppose that a sequence of sixty 1's and sixty 0's resulted in 75 runs. Since

$$\mu = 61 \qquad \text{and} \qquad \sigma = \sqrt{\frac{3{,}540}{119}} = 5.454$$

we see that the approximate p-value is

$$\begin{aligned} p\text{-value} &\approx 2\min\{\Phi(2.567), 1 - \Phi(2.567)\} \\ &= 2 \times (1 - .9949) \\ &= .0102 \end{aligned}$$

On the other hand, by running Program 12.5 we obtain that the exact p-value is

$$p\text{-value} = .0130$$

If the number of runs was equal to 70 rather than 75, then the approximate p-value would be

$$p\text{-value} \approx 2[1 - \Phi(1.650)] = .0990$$

as opposed to the exact value of

$$p\text{-value} = .1189 \quad \blacksquare$$

Problems

1. A new medicine against hypertension was tested on 18 patients. After 40 days of treatment, the following changes of the diastolic blood pressure were observed.

$$\begin{array}{ccccccccc} -5, & -1, & +2, & +8, & -25, & +1, & +5, & -12, & -16 \\ -9, & -8, & -18, & -5, & -22, & +4, & -21, & -15, & -11 \end{array}$$

Use the sign test to determine if the medicine has an effect on blood pressure. What is the p-value?

2. An engineering firm is involved in selecting a computer system, and the choice has been narrowed to two manufacturers. The firm submits eight problems to the two computer manufacturers and has each manufacturer measure the number of seconds required to solve the design problem with the manufacturer's software. The times for the eight design problems are given below.

Design problem	1	2	3	4	5	6	7	8
Time with computer A	15	32	17	26	42	29	12	38
Time with computer B	22	29	1	23	46	25	19	47

Determine the p-value of the sign test when testing the hypothesis that there is no difference in the distribution of the time it takes the two types of software to solve problems.

3. The published figure for the median systolic blood pressure of middle-aged men is 128. To determine if there has been any change in this value, a random sample of 100 men has been selected. Test the hypothesis that the median is equal to 128 if

(a) 60 men have readings above 128;
(b) 70 men have readings above 128;
(c) 80 men have readings above 128.

In each case, determine the p-value.

4. To test the hypothesis that the median weight of 16-year-old females from Los Angeles is at least 110 pounds, a random sample of 200 such females was chosen. If 120 females weighed less than 110 pounds, does this discredit the hypothesis? Use the 5 percent level of significance. What is the *p*-value?

5. In 2004, the national median salary of all U.S. financial accountants was \$124,400. A recent random sample of 14 financial accountants showed 2007 incomes of (in units of \$1,000)

125.5, 130.3, 133.0, 102.6, 198.0, 232.5, 106.8,
114.5, 122.0, 100.0, 118.8, 108.6, 312.7, 125.5

Use these data to test the hypothesis that the median salary of financial accountants in 2007 was not greater than in 2004. What is the *p*-value?

6. An experiment was initiated to study the effect of a newly developed gasoline detergent on automobile mileage. The following data, representing mileage per gallon before and after the detergent was added for each of eight cars, resulted.

Car	Mileage without Additive	Mileage with Additive
1	24.2	23.5
2	30.4	29.6
3	32.7	32.3
4	19.8	17.6
5	25.0	25.3
6	24.9	25.4
7	22.2	20.6
8	21.5	20.7

Find the *p*-value of the test of the hypothesis that mileage is not affected by the additive when

(a) the sign test is used;
(b) the signed rank test is used.

7. Determine the *p*-value when using the signed rank statistic in Problems 1 and 2.

8. Twelve patients having high albumin content in their blood were treated with a medicine. Their blood content of albumin was measured before and after treatment. The measured values are shown in the table.

Blood Content of Albumin[a]

Patient	Before Treatment	After Treatment
1	5.02	4.66
2	5.08	5.15
3	4.75	4.30
4	5.25	5.07
5	4.80	5.38
6	5.77	5.10
7	4.85	4.80
8	5.09	4.91
9	6.05	5.22
10	4.77	4.50
11	4.85	4.85
12	5.24	4.56

[a] *Values given in grams per 100 ml.*

Is the effect of the medicine significant at the 5 percent level?

(a) Use the sign test.
(b) Use the signed rank test.

9. An engineer claims that painting the exterior of a particular aircraft affects its cruising speed. To check this, the next 10 aircraft off the assembly line were flown to determine cruising speed prior to painting, and were then painted and reflown. The following data resulted.

	Cruising Speed (knots)	
Aircraft	Not Painted	Painted
1	426.1	416.7
2	418.4	403.2
3	424.4	420.1
4	438.5	431.0
5	440.6	432.6
6	421.8	404.2
7	412.2	398.3
8	409.8	405.4
9	427.5	422.8
10	441.2	444.8

Do the data uphold the engineer's claim?

10. Ten pairs of duplicate spectrochemical determinations for nickel are presented below. The readings in column 2 were taken with one type of measuring instrument and those in column 3 were taken with another type.

Sample	Duplicates	
1	1.94	2.00
2	1.99	2.09
3	1.98	1.95
4	2.07	2.03
5	2.03	2.08
6	1.96	1.98
7	1.95	2.03
8	1.96	2.03
9	1.92	2.01
10	2.00	2.12

Test the hypothesis, at the 5 percent level of significance, that the two measuring instruments give equivalent results.

11. Let $X_1, \ldots, X_n$ be a sample from the continuous distribution F having median m; and suppose we are interested in testing the hypothesis $H_0 : m = m_0$ against the one-sided alternative $H_1 : m > m_0$. Present the one-sided analog of the signed rank test. Explain how the p-value would be computed.

12. In a study of bilingual coding, 12 bilingual (French and English) college students are divided into two groups. Each group reads an article written in French, and each answers a series of 25 multiple-choice questions covering the content of the article. For one group the questions are written in French; the other takes the examination in English. The score (total correct) for the two groups is:

Examination in French	11	12	16	22	25	25
Examination in English	10	13	17	19	21	24

Is this evidence at the 5 percent significance level that there is difficulty in transferring information from one language to another?

13. Fifteen cities, of roughly equal size, are chosen for a traffic safety study. Eight of them are randomly chosen, and in these cities a series of newspaper articles dealing with traffic safety is run over a 1-month period. The number of traffic accidents reported in the month following this campaign is as follows:

Treatment group	19	31	39	45	47	66	74	81
Control group	28	36	44	49	52	52	60	

Determine the exact p-value when testing the hypothesis that the articles have not had any effect.

14. Determine the p-value in Problem 13 by

(a) using the normal approximation;
(b) using a simulation study.

15. The following are the weights of random samples of adult males from different political affiliations.

Republicans: 204, 178, 195, 187, 240, 182, 152, 166
Democrats: 175, 200, 168, 192, 156, 164, 180, 138

We want to use these data to test the null hypothesis that the two distributions are identical.

(a) Find the exact p-value.
(b) Determine the p-value obtained when using the normal approximation.

16. In a 1943 experiment (Whitlock, H. V., and Bliss, D. H., "A bioassay technique for antihelminthics," *Journal of Parasitology*, **29**, pp. 48–58, 10), albino rats were used to study the effectiveness of carbon tetrachloride as a treatment for worms. Each rat received an injection of worm larvae. After 8 days, the rats were randomly divided into 2 groups of 5 each; each rat in the first group received a dose of .032 cc of carbon tetrachloride, whereas the dosage for each rat in the second group was .063 cc. Two days later the rats were killed, and the number of adult worms in each rat was determined. The numbers detected in the group receiving the .032 dosage were

$$421,\ 462,\ 400,\ 378,\ 413$$

whereas they were

$$207,\ 17,\ 412,\ 74,\ 116$$

for those receiving the .063 dosage. Do the data prove that the larger dosage is more effective than the smaller?

17. In a 10-year study of the dispersal patterns of beavers (Sun, L. and Muller-Schwarze, D., "Statistical resampling methods in biology: A case study of beaver dispersal patterns," *American Journal of Mathematical and Management Sciences*, **16**, pp. 463–502, 1996) a total of 332 beavers were trapped in Allegheny State Park in southwestern New York. The beavers were tagged (so as to be identifiable when later caught) and released. Over time a total of 32 of them, 9 female and

23 male, were discovered to have resettled in other sites. The following data give the dispersal distances (in kilometers) between these beavers' original and resettled sites for the females and for the males.

Females:	.660, .984, .984, 1.992, 4.368, 6.960, 10.656, 21.600, 31.680
Males:	.288, .312, .456, .528, .576, .720, .792, .984, 1.224,
	1.584, 2.304, 2.328, 2.496, 2.688, 3.096, 3.408, 4.296, 4.884,
	5.928, 6.192, 6.384, 13.224, 27.600

Do the data prove that the dispersal distances are gender related?

18. The following data give the numbers of people who visit a local health clinic in the day following

(1) a Saturday win by the local university football team;
(2) a Saturday loss by the team;
(3) a Saturday when the team does not play.

Number following a win	71, 66, 62, 79, 80, 70, 66, 59, 89, 68
Number following a loss	64, 62, 75, 81, 69, 67, 73, 71, 69, 74
Number when no game	49, 48, 70, 58, 73, 65, 55, 52, 68, 74

Do these data prove that the resulting number of clinic visits depends on what happens with the football team? Test at the 5 percent level.

19. A production run of 50 items resulted in 11 defectives, with the defectives occurring on the following items (where the items are numbered by their order of production): 8, 12, 13, 14, 31, 32, 37, 38, 40, 41, 42. Can we conclude that the successive items did not constitute a random sample?

20. The following data represent the successive quality levels of 25 articles: 100, 110, 122, 132, 99, 96, 88, 75, 45, 211, 154, 143, 161, 142, 99, 111, 105, 133, 142, 150, 153, 121, 126, 117, 155. Does it appear that these data are a random sample from some population?

21. Can we use the runs test if we consider whether each data value is less than or greater than some predetermined value rather than the value s-med?

22. The following table (taken from Quinn, W. H., Neal, T. V., and Antuñez de Mayolo, S. E., 1987, "El Niño occurrences over the past four-and-a-half centuries," *Journal of Geophysical Research*, **92** (C13), pp. 14,449–14,461) gives the years and magnitude (either moderate or strong) of major El Niño years between 1800 and 1987. Use it to test the hypothesis that the successive El Niño magnitudes constitute a random sample.

Year and Magnitude (0 = moderate, 1 = strong) of Major El Niño Events, 1800–1987

Year	Magnitude	Year	Magnitude	Year	Magnitude
1803	1	1866	0	1918	0
1806	0	1867	0	1923	0
1812	0	1871	1	1925	1
1814	1	1874	0	1930	0
1817	0	1877	1	1932	1
1819	0	1880	0	1939	0
1821	0	1884	1	1940	1
1824	0	1887	0	1943	0
1828	1	1891	1	1951	0
1832	0	1896	0	1953	0
1837	0	1899	1	1957	1
1844	1	1902	0	1965	0
1850	0	1905	0	1972	1
1854	0	1907	0	1976	0
1857	0	1911	1	1982	1
1860	0	1914	0	1987	0
1864	1	1917	1		