

Chapter 9

Nonparametric Tests

9.1 Sign Test

The reader should recall that the procedures discussed in Chapter 5 for testing the null hypothesis that $\mu = \mu_0$ are valid only if the population is approximately normal or if the sample is large. If $n < 30$ and the population is decidedly nonnormal, we must resort to a nonparametric test. The sign test is used to test hypotheses on a population median. In the case of many of the nonparametric procedures, the mean is replaced by the median as the pertinent location parameter under test. The population counterpart, denoted by $\tilde{\mu}$, has an analogous definition. Given a random variable X , $\tilde{\mu}$ is defined such that $P(X > \tilde{\mu}) \leq 0.5$ and $P(X < \tilde{\mu}) \leq 0.5$. In the continuous case, $P(X > \tilde{\mu}) = P(X < \tilde{\mu}) = 0.5$.

The appropriate test statistic for the sign test is the binomial random variable X , representing the number of plus signs in our random sample. If the null hypothesis that $\tilde{\mu} = \tilde{\mu}_0$ is true, the probability that a sample value results in either a plus or a minus sign is equal to $1/2$. Therefore, to test the null hypothesis that $\tilde{\mu} = \tilde{\mu}_0$, we actually test the null hypothesis that the number of plus signs is a value of a random variable having the binomial distribution with the parameter $p = 1/2$. P-values for both one-sided and two-sided alternatives can then be calculated using this binomial distribution. For example, in testing

$$H_0: \tilde{\mu} = \tilde{\mu}_0, \quad H_1: \tilde{\mu} < \tilde{\mu}_0,$$

we shall reject H_0 in favor of H_1 only if the proportion of plus signs is sufficiently less than $1/2$, that is, when the value x of our random variable is small. Hence, if the computed P-value

$$P = P(X \leq x \text{ when } p = 1/2)$$

is less than or equal to some preselected significance level α , we reject H_0 in favor of H_1 . For example, when $n = 15$ and $x = 3$, we find that

$$\begin{aligned} P &= P\left(X \leq x \text{ when } p = \frac{1}{2}\right) = \sum_{x=0}^3 b\left(x; 15, \frac{1}{2}\right) \\ &= \sum_{x=0}^3 \binom{15}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{15-x} = \sum_{x=0}^3 \binom{15}{x} \left(\frac{1}{2}\right)^{15} \\ &= \left(\frac{1}{2}\right)^{15} \left(\binom{15}{0} + \binom{15}{1} + \binom{15}{2} + \binom{15}{3} \right) \\ &= 0.0000305(1 + 15 + 105 + 455) = 0.0176 \end{aligned}$$

so the null hypothesis $\tilde{\mu} = \tilde{\mu}_0$ can certainly be rejected at the 0.05 level of significance but not at the 0.01 level. To test the hypothesis

$$H_0: \tilde{\mu} = \tilde{\mu}_0, \quad H_1: \tilde{\mu} > \tilde{\mu}_0,$$

we reject H_0 in favor of H_1 only if the proportion of plus signs is sufficiently greater than $1/2$, that is, when the value x of our random variable is small. Hence, if the computed P-value

$$P = P(X \geq x \text{ when } p = 1/2)$$

is less than α , we reject H_0 in favor of H_1 .

Finally, to test the hypothesis

$$H_0: \tilde{\mu} = \tilde{\mu}_0, \quad H_1: \tilde{\mu} \neq \tilde{\mu}_0,$$

we reject H_0 in favor of H_1 when the proportion of plus signs is significantly less than or greater than $1/2$. This, of course, is equivalent to x being sufficiently small or sufficiently large. Therefore, if $x < n/2$ and the computed P-value

$$P = 2P(X \leq x \text{ when } p = 1/2)$$

is less than or equal to α , or if $x > n/2$ and the computed P-value

$$P = 2P(X \geq x \text{ when } p = 1/2)$$

is less than or equal to α , we reject H_0 in favor of H_1 .

Example 9.1

The following data represent the number of hours that a rechargeable hedge trimmer operates before a recharge is required:

1.5, 2.2, 0.9, 1.3, 2.0, 1.6, 1.8, 1.5, 2.0, 1.2, 1.7.

Use the sign test to test the hypothesis, at the 0.05 level of significance, that this particular trimmer operates a median of 1.8 hours before requiring a recharge.

Solution

1. $H_0: \tilde{\mu} = 1.8$.
2. $H_1: \tilde{\mu} \neq 1.8$.
3. $\alpha = 0.05$.
4. Test statistic: Binomial variable X with $p = 1/2$.
5. Computations: Replacing each value by the symbol “+” if it exceeds 1.8 and by the symbol “-” if it is less than 1.8 and discarding the one measurement that equals 1.8, we obtain the sequence

1.5 2.2 0.9 1.3 2.0 1.6 1.8 1.5 2.0 1.2 1.7
- ' + ' - ' - ' + ' - ' ' - ' + ' - ' -

for which $n = 10$, $x = 3$, and $n/2 = 5$. Therefore, the computed P-value is

$$\begin{aligned}
P &= 2P(X \leq 3 \text{ when } p = 1/2) = 2 \sum_{x=0}^3 b\left(x; 10, \frac{1}{2}\right) \\
&= 2 \sum_{x=0}^3 \binom{10}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{10-x} = 2 \sum_{x=0}^3 \binom{10}{x} \left(\frac{1}{2}\right)^{10} \\
&= \left(\frac{1}{2}\right)^{10} \left(\binom{10}{0} + \binom{10}{1} + \binom{10}{2} + \binom{10}{3} \right) \\
&= 2 \times 0.0000977(1 + 10 + 45 + 120) = 0.3438 > 0.05
\end{aligned}$$

6. Decision: Do not reject the null hypothesis and conclude that the median operating time is not significantly different from 1.8 hours.

We can also use the sign test to test the null hypothesis $\tilde{\mu}_1 - \tilde{\mu}_2 = d_0$ for paired observations. Here we replace each difference, d_i , with a plus or minus sign depending on whether the adjusted difference, $d_i - d_0$, is positive or negative. Throughout this section, we have assumed that the populations are symmetric. However, even if populations are skewed, we can carry out the same test procedure, but the hypotheses refer to the population medians rather than the means.

9.2 Signed-Rank Test

The reader should note that the sign test utilizes only the plus and minus signs of the differences between the observations and $\tilde{\mu}_0$ in the one-sample case, or the plus and minus signs of the differences between the pairs of observations in the paired-sample case; it does not take into consideration the magnitudes of these differences. A test utilizing both direction and magnitude, proposed in 1945 by Frank Wilcoxon, is now commonly referred to as the **Wilcoxon signed-rank test**. The analyst can extract more information from the data in a nonparametric fashion if it is reasonable to invoke an additional restriction on the distribution from which the data were taken. The

Wilcoxon signed-rank test applies in the case of a **symmetric continuous distribution**. Under this condition, we can test the null hypothesis $\tilde{\mu} = \tilde{\mu}_0$. We first subtract $\tilde{\mu}_0$ from each sample value, discarding all differences equal to zero. The remaining differences are then ranked without regard to sign. A rank of 1 is assigned to the smallest absolute difference (i.e., without sign), a rank of 2 to the next smallest, and so on. When the absolute value of two or more differences is the same, assign to each the average of the ranks that would have been assigned if the differences were distinguishable. For example, if the fifth and sixth smallest differences are equal in absolute value, each is assigned a rank of 5.5. If the hypothesis $\tilde{\mu} = \tilde{\mu}_0$ is true, the total of the ranks corresponding to the positive differences should nearly equal the total of the ranks corresponding to the negative differences. Let us represent these totals by w_+ and w_- , respectively. We designate the smaller of w_+ and w_- by w . In selecting repeated samples, we would expect w_+ and w_- , and therefore w , to vary. Thus, we may think of w_+, w_- and w as values of the corresponding random variables W_+, W_- and W . The null hypothesis $\tilde{\mu} = \tilde{\mu}_0$ can be rejected in favor of the alternative $\tilde{\mu} < \tilde{\mu}_0$ only if w_+ is small and w_- is large. Likewise, the alternative $\tilde{\mu} > \tilde{\mu}_0$ can be accepted only if w_+ is large and w_- is small. For a two-sided alternative, we may reject H_0 in favor of H_1 if either w_+ or w_- , and hence w , is sufficiently small. Therefore, no matter what the alternative hypothesis may be, we reject the null hypothesis when the value of the appropriate statistic W_+, W_- , or W is sufficiently small.

Two Samples with Paired Observations

To test the null hypothesis that we are sampling two continuous symmetric populations with $\tilde{\mu}_1 - \tilde{\mu}_2$ for the paired-sample case, we rank the differences of the paired observations without regard to sign and proceed as in the single-sample case. The various test procedures

for both the single- and paired-sample cases are summarized in Table 9.2.

Table 9.2: Signed-Rank Test

H_0	H_1	Compute
$\tilde{\mu} = \tilde{\mu}_0$	$\begin{cases} \tilde{\mu} < \tilde{\mu}_0 \\ \tilde{\mu} > \tilde{\mu}_0 \\ \tilde{\mu} \neq \tilde{\mu}_0 \end{cases}$	$\begin{array}{l} W_+ \\ W_- \\ W \end{array}$
$\tilde{\mu}_1 = \tilde{\mu}_2$	$\begin{cases} \tilde{\mu}_1 < \tilde{\mu}_2 \\ \tilde{\mu}_1 > \tilde{\mu}_2 \\ \tilde{\mu}_1 \neq \tilde{\mu}_2 \end{cases}$	$\begin{array}{l} W_+ \\ W_- \\ W \end{array}$

It is not difficult to show that whenever $n < 5$ and the level of significance does not exceed 0.05 for a one-tailed test or 0.10 for a two-tailed test, all possible values of $w_+, w_-,$ or w will lead to the acceptance of the null hypothesis. However, when $5 \leq n \leq 30$, **Table Signed-Rank Test** shows approximate critical values of W_+ and W_- for levels of significance equal to 0.01, 0.025, and 0.05 for a one-tailed test and critical values of W for levels of significance equal to 0.02, 0.05, and 0.10 for a two-tailed test. The null hypothesis is rejected if the computed value $w_+, w_-,$ or w is less than or equal to the appropriate tabled value. For example, when $n = 12$, **Table Signed-Rank Test** shows that a value of $w_+ \leq 17$ is required for the one-sided alternative $\tilde{\mu} < \tilde{\mu}_0$ to be significant at the 0.05 level.

Example 9.2: Rework Example 9.1 by using the signed-rank test.

1. $H_0: \tilde{\mu} = 1.8.$
2. $H_1: \tilde{\mu} \neq 1.8.$
3. $\alpha = 0.05.$

4. Critical region: Since $n = 10$ after discarding the one measurement that equals 1.8, **Table Signed-Rank Test** shows the critical region to be $w \leq 8$.

5. Computations: Subtracting 1.8 from each measurement and then ranking the differences without regard to sign, we have

d_i	-0.3	0.4	-0.9	-0.5	0.2	-0.2	-0.3	0.2	-0.6	-0.1
Ranks	5.5	7	10	8	3	3	5.5	3	9	1

Now $w_+ = 13$ and $w_- = 42$, so $w = 13$, the smaller of w_+ and w_- .

6. Decision: As before, do not reject H_0 and conclude that the median operating time is not significantly different from 1.8 hours.

The signed-rank test can also be used to test the null hypothesis that $\tilde{\mu}_1 - \tilde{\mu}_2 = d_0$. In this case, the populations need not be symmetric. As with the sign test, we subtract d_0 from each difference, rank the adjusted differences without regard to sign, and apply the same procedure as above.

Example 9.3:

It is claimed that a college senior can increase his or her score in the major field area of the graduate record examination by at least 50 points if he or she is provided with sample problems in advance. To test this claim, 20 college seniors are divided into 10 pairs such that the students in each matched pair have almost the same overall grade-point averages for their first 3 years in college. Sample problems and answers are provided at random to one member of each pair 1 week prior to the examination. The examination scores are given in Table 9.3.

Table 9.3: Data for Example 9.3

	Pair									
	1	2	3	4	5	6	7	8	9	10
With Sample Problems	531	621	663	579	451	660	591	719	543	575
Without Sample Problems	509	540	688	502	424	683	568	748	530	524

Test the null hypothesis, at the 0.05 level of significance, that sample problems increase scores by 50 points against the alternative hypothesis that the increase is less than 50 points.

Solution:

Let $\tilde{\mu}_1$ and $\tilde{\mu}_2$ represent the median scores of all students taking the test in question with and without sample problems, respectively.

1. $H_0: \tilde{\mu}_1 - \tilde{\mu}_2 = 50$.
2. $H_1: \tilde{\mu}_1 - \tilde{\mu}_2 < 50$.
3. $\alpha = 0.05$.
4. Critical region: Since $n = 10$, **Table Signed-Rank Test** shows the critical region to be $w_+ \leq 11$.

5. Computations:

	Pair									
	1	2	3	4	5	6	7	8	9	10
d_i	22	81	-25	77	27	-23	23	-29	13	51
$d_i - d_0$	-28	31	-75	27	-23	-73	-27	-79	-37	1
Ranks	5	6	9	3.5	2	8	3.5	10	7	1

Now we find that $w_+ = 6 + 3.5 + 1 = 10.5$.

6. Decision: Reject H_0 and conclude that sample problems do not, on average, increase one's graduate record score by as much as 50 points.

9.3 Spearman Correlation Coefficient

In Section 7.3, we used the sample correlation coefficient r to measure the population correlation coefficient ρ , the linear relationship between two continuous variables X and Y . If ranks $1, 2, \dots, n$ are assigned to the x observations in order of magnitude and similarly to the y observations, and if these ranks are then substituted for the actual numerical values in the formula for the correlation coefficient in Section 7.3, we obtain the nonparametric counterpart of the conventional correlation coefficient. A correlation coefficient calculated in this manner is known as the Spearman rank correlation coefficient and is denoted by r_s .

Definition (Spearman rank correlation coefficient)

A nonparametric measure of association between two variables X and Y is given by the **rank correlation coefficient**

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)},$$

where d_i is the difference between the ranks assigned to x_i and y_i and n is the number of pairs of data.

Example 9.4: The figures listed in Table 9.4, released by the Federal Trade Commission, show the milligrams of tar and nicotine found in 10 brands of cigarettes. Calculate the rank correlation coefficient to measure the degree of relationship between tar and nicotine content in cigarettes.

Table 9.4: Tar and Nicotine Contents

Cigarette Brand	Tar Content	Nicotine Content
Viceroy	14	0.9
Marlboro	17	1.1
Chesterfield	28	1.6
Kool	17	1.3
Kent	16	1.0
Raleigh	13	0.8
Old Gold	24	1.5
Philip Morris	25	1.4
Oasis	18	1.2
Players	31	2.0

Solution: Let X and Y represent the tar and nicotine contents, respectively. First we assign ranks to each set of measurements, with the rank of 1 assigned to the lowest number in each set, the rank of 2 to the second lowest number in each set, and so forth, until the rank of 10 is assigned to the largest number. Table 9.10 shows the individual rankings of the measurements and the differences in ranks for the 10 pairs of observations.

Table 9.10: Rankings for Tar and Nicotine Content

Cigarette Brand	x_i	y_i	d_i
Viceroy	2.0	2.0	0.0
Marlboro	4.5	4.0	0.5
Chesterfield	9.0	9.0	0.0
Kool	4.5	6.0	-1.5
Kent	3.0	3.0	0.0
Raleigh	1.0	1.0	0.0
Old Gold	7.0	8.0	-1.0
Philip Morris	8.0	7.0	1.0
Oasis	6.0	5.0	1.0
Players	10.0	10.0	0.0

Substituting into the formula for r_s , we find that

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} = 1 - \frac{(6)(5.50)}{10(100 - 1)} = 0.967,$$

indicating a high positive correlation between the amounts of tar and nicotine found in cigarettes.