15 Nonparametric Statistics

CHAPTER OUTLINE

Bootstrap Method

15-1 Introduction

- Most of the hypothesis-testing and confidence interval procedures discussed in previous chapters are based on the assumption that we are working with random samples from normal populations.
- These procedures are often called **parametric methods**
- In this chapter, **nonparametric** and **distribution free** methods will be discussed.
- We usually make no assumptions about the distribution of the underlying population.

15-2 Sign Test

15-2.1 Description of the Test

- The **sign test** is used to test hypotheses about the **median** of a continuous distribution.

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

- Let R^+ represent the number of differences

$$X_i - \tilde{\mu}_0$$

that are positive.

- What is the sampling distribution of R^+ under H_0?

15-2 Sign Test

15-2.1 Description of the Test

If the following hypotheses are being tested:

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

The appropriate P-value is

$$P = P\left(R^+ \leq r^+ \text{ when } p = \frac{1}{2}\right)$$

15-2 Sign Test

15-2.1 Description of the Test

If the following hypotheses are being tested:

$$H_0\colon \tilde{\mu} = \tilde{\mu}_0$$

$$H_1\colon \tilde{\mu} > \tilde{\mu}_0$$

The appropriate P-value is

$$P = P\left(R^+ \geq r^+ \text{ when } p = \frac{1}{2}\right)$$

15-2 Sign Test

15-2.1 Description of the Test

If the following hypotheses are being tested:

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

If $r^+ < n/2$, then the appropriate P-value is

$$P = 2P\left(R^+ \leq r^+ \text{ when } p = \frac{1}{2}\right)$$

If $r^+ > n/2$, then the appropriate P-value is

$$P = 2P\left(R^+ \geq r^+ \text{ when } p = \frac{1}{2}\right)$$

15-2 Sign Test

Example 15-1

Montgomery, Peck, and Vining (2001) report on a study in which a rocket motor is formed by binding an igniter propellant and a sustainer propellant together inside a metal housing. The shear strength of the bond between the two propellant types is an important characteristic. The results of testing 20 randomly selected motors are shown in Table 15-1. We would like to test the hypothesis that the median shear strength is 2000 psi, using $\alpha = 0.05$.

This problem can be solved using the eight-step hypothesis-testing procedure introduced in Chapter 9:

1. The parameter of interest is the median of the distribution of propellant shear strength.
2. H_0: $\tilde{\mu} = 2000$ psi
3. H_1: $\tilde{\mu} \neq 2000$ psi
4. $\alpha = 0.05$
5. The test statistic is the observed number of plus differences in Table 15-1, or $r^+ = 14$.
6. We will reject H_0 if the P-value corresponding to $r^+ = 14$ is less than or equal to $\alpha = 0.05$.

Example 15-1

Table 15-1 Propellant Shear Strength Data

Observation i	Shear Strength x_i	Differences $x_i - 2000$	Sign
1	2158.70	+158.70	+
2	1678.15	−321.85	−
3	2316.00	+316.00	+
4	2061.30	+61.30	+
5	2207.50	+207.50	+
6	1708.30	−291.70	−
7	1784.70	−215.30	−
8	2575.10	+575.10	+
9	2357.90	+357.90	+
10	2256.70	+256.70	+
11	2165.20	+165.20	+
12	2399.55	+399.55	+
13	1779.80	−220.20	−
14	2336.75	+336.75	+
15	1765.30	−234.70	−
16	2053.50	+53.50	+
17	2414.40	+414.40	+
18	2200.50	+200.50	+
19	2654.20	+654.20	+
20	1753.70	−246.30	−

15-2 Sign Test

Example 15-1

7. Computations: Since $r^+ = 14$ is greater than $n/2 = 20/2 = 10$, we calculate the P-value from

$$P = 2P\left(R^+ \geq 14 \text{ when } p = \frac{1}{2}\right)$$
$$= 2\sum_{r=14}^{20}\binom{20}{r}(0.5)^r(0.5)^{20-r}$$
$$= 0.1153$$

8. Conclusions: Since $P = 0.1153$ is not less than $\alpha = 0.05$, we cannot reject the null hypothesis that the median shear strength is 2000 psi. Another way to say this is that the observed number of plus signs $r^+ = 14$ was not large or small enough to indicate that median shear strength is different from 2000 psi at the $\alpha = 0.05$ level of significance.

15-2 Sign Test

15-2.2 Sign Test for Paired Samples

The sign test can also be applied to paired observations drawn from continuous populations. Let (X_{1j}, X_{2j}), $j = 1, 2, \ldots, n$ be a collection of paired observations from two continuous populations, and let

$$D_j = X_{1j} - X_{2j} \qquad j = 1, 2, \ldots, n$$

be the paired differences. We wish to test the hypothesis that the two populations have a common median, that is, that $\tilde{\mu}_1 = \tilde{\mu}_2$. This is equivalent to testing that the median of the differences $\tilde{\mu}_D = 0$. This can be done by applying the sign test to the n observed differences d_j, as illustrated in the following example.

See Example 15-3.

15-2 Sign Test

15-2.3 Type II Error for the Sign Test

- **Depends on both the true population distribution and alternative value!**

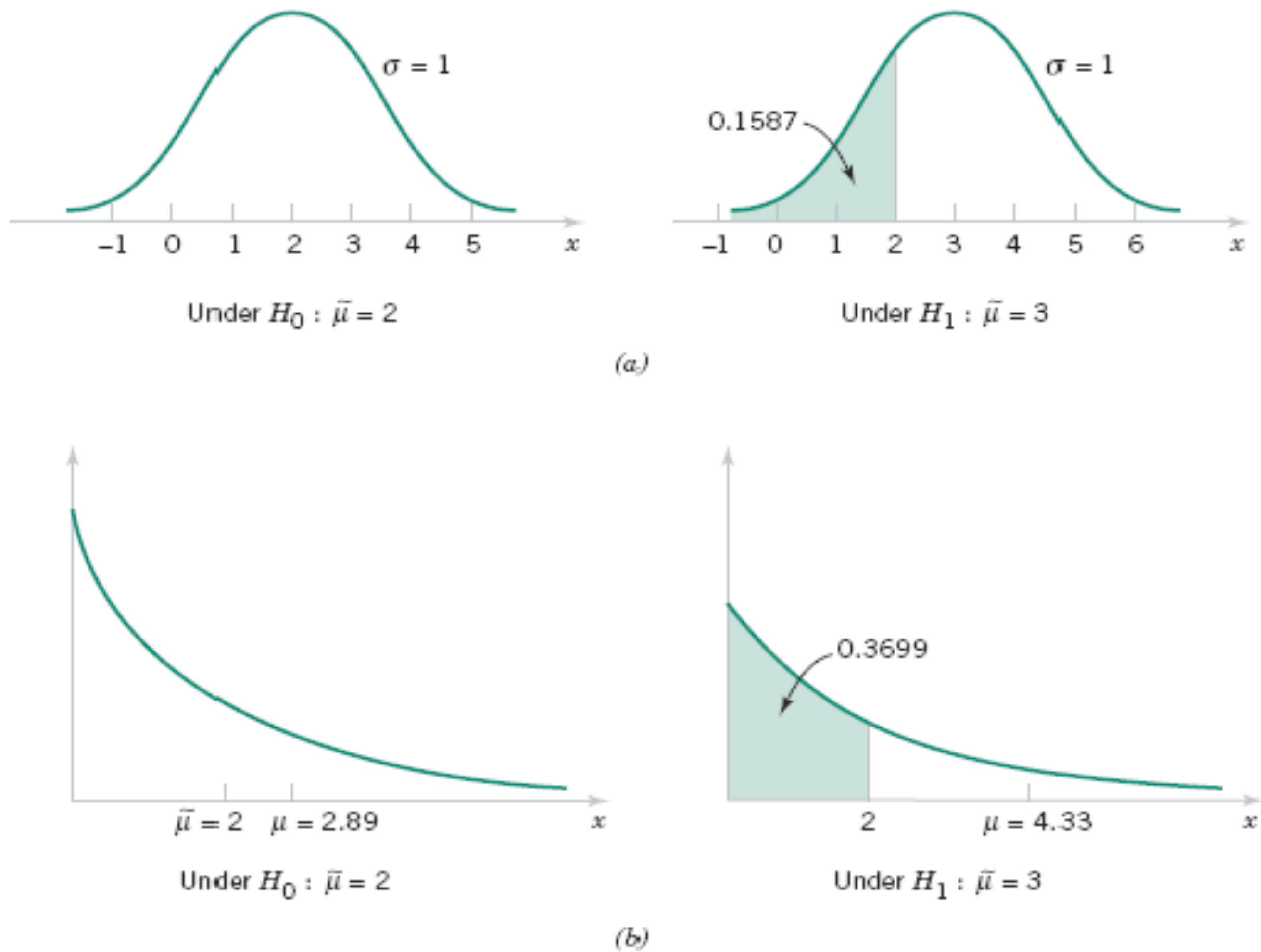

Figure 15-1 Calculation of β for the sign test. (a) Normal distributions. (b) Exponential distributions

15-3 Wilcoxon Signed-Rank Test

- The **Wilcoxon signed-rank test** applies to the case of symmetric continuous distributions.
- Under this assumption, the mean equals the median.
- The null hypothesis is H_0: $\mu = \mu_0$

15-3 Wilcoxon Signed-Rank Test

- 15-3.1 Description of the Test

- Assume that $X_1, X_2, \ldots, X_n$ is a random sample from a continuous and symmetric distribution with mean (and median) μ.

Procedure:

- Compute the differences $X_i - \mu_0$, $i = 1, 2, \ldots, n$.
- Rank the absolute differences $|X_i - \mu_0|$, $i = 1, 2, \ldots, n$ in ascending order.
- Give the ranks the signs of their corresponding differences.
- Let W^+ be the sum of the positive ranks and W^- be the absolute value of the sum of the negative ranks.
- Let $W = \min(W^+, W^-)$.

15-3 Wilcoxon Signed-Rank Test

Decision rules:

Appendix Table IX contains critical values of W, say w_α^*

If the alternative is $H_1\colon \mu \neq \mu_0$, reject $H_0\colon \mu = \mu_0$ if $w \leq w_\alpha^*$

If the alternative is $H_1\colon \mu > \mu_0$, reject $H_0\colon \mu = \mu_0$ if $w^- \leq w_\alpha^*$

If the alternative is $H_1\colon \mu < \mu_0$, reject $H_0\colon \mu = \mu_0$ if $w^+ \leq w_\alpha^*$

Appendix Table IX provides significance levels of $\alpha = 0.10$, $\alpha = 0.05$, $\alpha = 0.02$, $\alpha = 0.01$ for the two-sided test.

The significance levels for one-sided tests provided in Appendix Table IX are $\alpha = 0.05, 0.025, 0.01$, and 0.005.

15-3 Wilcoxon Signed-Rank Test

Example 15-4

We will illustrate the Wilcoxon signed-rank test by applying it to the propellant shear strength data from Table 15-1. Assume that the underlying distribution is a continuous symmetric distribution. The eight-step procedure is applied as follows:

1. The parameter of interest is the mean (or median) of the distribution of propellant shear strength.
2. H_0: $\mu = 2000$ psi
3. H_1: $\mu \neq 2000$ psi
4. $\alpha = 0.05$
5. The test statistic is

$$w = \min(w^+, w^-)$$

6. We will reject H_0 if $w \leq w^*_{0.05} = 52$ from Appendix Table VIII.

Example 15-4

7. Computations: The signed ranks from Table 15-1 are shown in the following table:

Observation	Difference $x_i - 2000$	Signed Rank
16	+53.50	+1
4	+61.30	+2
1	+158.70	+3
11	+165.20	+4
18	+200.50	+5
5	+207.50	+6
7	−215.30	−7
13	−220.20	−8
15	−234.70	−9
20	−246.30	−10
10	+256.70	+11
6	−291.70	−12
3	+316.00	+13
2	−321.85	−14
14	+336.75	+15
9	+357.90	+16
12	+399.55	+17
17	+414.40	+18
8	+575.10	+19
19	+654.20	+20

15-3 Wilcoxon Signed-Rank Test

Example 15-4

The sum of the positive ranks is $w^+ = (1 + 2 + 3 + 4 + 5 + 6 + 11 + 13 + 15 + 16 + 17 + 18 + 19 + 20) = 150$, and the sum of the absolute values of the negative ranks is $w^- = (7 + 8 + 9 + 10 + 12 + 14) = 60$. Therefore,

$$w = \min(150, 60) = 60$$

8. Conclusions: Since $w = 60$ is not less than or equal to the critical value $w_{0.05} = 52$, we cannot reject the null hypothesis that the mean (or median, since the population is assumed to be symmetric) shear strength is 2000 psi.

15-3 Wilcoxon Signed-Rank Test

15-3.2 Large-Sample Approximation

Therefore, a test of H_0: $\mu = \mu_0$ can be based on the statistic

$$Z_0 = \frac{W^+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \qquad (15\text{-}6)$$

Z_0 is approximately standard normal when n is large.

15-4 Wilcoxon Rank-Sum Test

Suppose that we have two independent continuous populations X_1 and X_2 with means μ_1 and μ_2. Assume that the distributions of X_1 and X_2 have the same shape and spread and differ only (possibly) in their locations. The Wilcoxon rank-sum test can be used to test the hypothesis H_0: $\mu_1 = \mu_2$. This procedure is sometimes called the Mann-Whitney test, although the Mann-Whitney test statistic is usually expressed in a different form.

15-4.1 Description of the Test

Let $X_{11}, X_{12}, \ldots, X_{1n_1}$ and $X_{21}, X_{22}, \ldots, X_{2n_2}$ be two independent random samples of sizes $n_1 \leq n_2$ from the continuous populations X_1 and X_2 described earlier. We wish to test the hypotheses

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

15-4 Wilcoxon Rank-Sum Test

15-4.1 Description of the Test

The test procedure is as follows. Arrange all $n_1 + n_2$ observations in ascending order of magnitude and assign ranks to them. If two or more observations are tied (identical), use the mean of the ranks that would have been assigned if the observations differed.

Let W_1 be the sum of the ranks in the smaller sample (1), and define W_2 to be the sum of the ranks in the other sample. Then,

$$W_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - W_1 \tag{15-7}$$

Now if the sample means do not differ, we will expect the sum of the ranks to be nearly equal for both samples after adjusting for the difference in sample size. Consequently, if the sums of the ranks differ greatly, we will conclude that the means are not equal.

15-4 Wilcoxon Rank-Sum Test

Example 15-6

The mean axial stress in tensile members used in an aircraft structure is being studied. Two alloys are being investigated. Alloy 1 is a traditional material, and alloy 2 is a new aluminum-lithium alloy that is much lighter than the standard material. Ten specimens of each alloy type are tested, and the axial stress is measured. The sample data are assembled in Table 15-3. Using $\alpha = 0.05$, we wish to test the hypothesis that the means of the two stress distributions are identical.

TABLE 15-3

Axial Stress for Two Aluminum-Lithium Alloys

Alloy 1		Alloy 2	
3238 psi	3254 psi	3261 psi	3248 psi
3195	3229	3187	3215
3246	3225	3209	3226
3190	3217	3212	3240
3204	3241	3258	3234

15-4 Wilcoxon Rank-Sum Test

Example 15-6

We will apply the eight-step hypothesis-testing procedure to this problem:

1. The parameters of interest are the means of the two distributions of axial stress.
2. H_0: $\mu_1 = \mu_2$
3. H_1: $\mu_1 \neq \mu_2$
4. $\alpha = 0.05$
5. We will use the Wilcoxon rank-sum test statistic in Equation 15-7,

$$w_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - w_1$$

6. Since $\alpha = 0.05$ and $n_1 = n_2 = 10$, Appendix Table IX gives the critical value as $w_{0.05} = 78$. If either w_1 or w_2 is less than or equal to $w_{0.05} = 78$, we will reject H_0: $\mu_1 = \mu_2$.

Example 15-6

7. Computations: The data from Table 15-3 are analyzed in ascending order and ranked as follows:

Alloy Number	Axial Stress	Rank
2	3187 psi	1
1	3190	2
1	3195	3
1	3204	4
2	3209	5
2	3212	6
2	3215	7
1	3217	8
1	3225	9
2	3226	10
1	3229	11
2	3234	12
1	3238	13
2	3240	14
1	3241	15
1	3246	16
2	3248	17
1	3254	18
2	3258	19
2	3261	20

15-4 Wilcoxon Rank-Sum Test

Example 15-6

The sum of the ranks for alloy 1 is

$$w_1 = 2 + 3 + 4 + 8 + 9 + 11 + 13 + 15 + 16 + 18 = 99$$

and for alloy 2

$$w_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - w_1 = \frac{(10 + 10)(10 + 10 + 1)}{2} - 99 = 111$$

8. Conclusions: Since neither w_1 nor w_2 is less than or equal to $w_{0.05} = 78$, we cannot reject the null hypothesis that both alloys exhibit the same mean axial stress.

15-5 Nonparametric Methods in the Analysis of Variance

The single-factor analysis of variance model for comparing *a* population means is

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n_i \end{cases}$$

The hypothesis of interest is

$$H_0\colon \mu_1 = \mu_2 = \cdots = \mu_a$$

The Kruskal-Wallis test (w/o assumption of normality)

- Basic idea: Use **ranks** instead of actual numbers

Parametric vs. Nonparametric Tests

- When the normality assumption is correct, t-test or F-test is more powerful.
  - Wilcoxon signed-rank or rank-sum test is approximately 95% as efficient as the t-test in large samples.
- On the other hand, regardless of the form of the distributions, nonparametric tests may be more powerful.
  - Wilcoxon signed-rank or rank-sum test will always be at least 86% as efficient.
- The efficiency of the Wilcoxon test relative to the t-test is usually high if the underlying distribution has heavier tails than the normal
  - because the behavior of the t-test is very dependent on the sample mean, which is quite unstable in heavy-tailed distributions.