WILEY

Applied Statistics and Probability for Engineers

Sixth Edition

Douglas C. Montgomery   George C. Runger

Chapter 9

Tests of Hypotheses for a Single Sample

9 Tests of Hypotheses for a Single Sample

CHAPTER OUTLINE

WILEY

Learning Objectives for Chapter 9

After careful study of this chapter, you should be able to do the following:

1. Structure engineering decision-making as hypothesis tests.
2. Test hypotheses on the mean of a normal distribution using a *Z*-test or a *t*-test.
3. Test hypotheses on the variance or standard deviation of a normal distribution.
4. Test hypotheses on a population proportion.
5. Use the *P*-value approach for making decisions in hypothesis tests.
6. Compute power & Type II error probability. Make sample size selection decisions for tests on means, variances & proportions.
7. Explain & use the relationship between confidence intervals & hypothesis tests.
8. Use the chi-square goodness-of-fit test to check distributional assumptions.
9. Use contingency table tests.

9-1 Hypothesis Testing

9-1.1 Statistical Hypotheses

A statistical hypothesis is a statement about the parameters of one or more populations.

Let $H_0 : \mu = 50$ centimeters per second and $H_1 : \mu \neq 50$ centimeters per second

The statement $H_0 : \mu = 50$ is called the **null hypothesis.**

The statement $H_1 : \mu \neq 50$ is called the **alternative hypothesis.**

<u>One-sided Alternative Hypotheses</u>

$H_0 : \mu = 50$ centimeters per second $H_0 : \mu = 50$ centimeters per second

or

$H_1 : \mu < 50$ centimeters per second $H_1 : \mu > 50$ centimeters per second

9-1 Hypothesis Testing

Test of a Hypothesis

- A procedure leading to a decision about a particular hypothesis
- Hypothesis-testing procedures rely on using the information in a. **random sample from the population of interest**
- If this information is *consistent* with the hypothesis, then we will conclude that the hypothesis is **true**; if this information is *inconsistent* with the hypothesis, we will conclude that the hypothesis is **false**.

9-1.2 Tests of Statistical Hypotheses

H_0 : $\mu = 50$ centimeters per second

H_1 : $\mu \neq 50$ centimeters per second

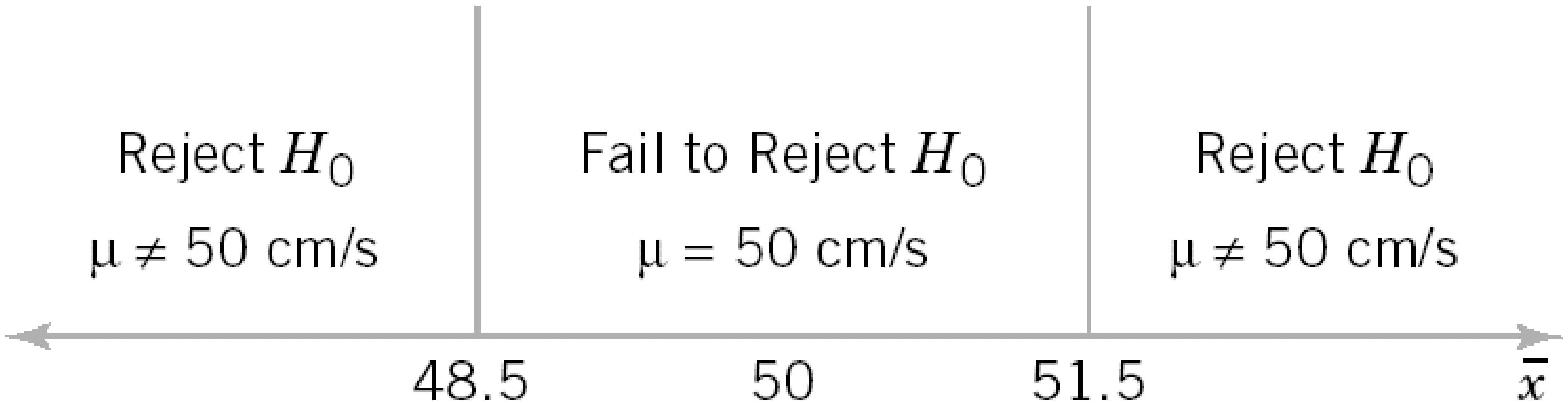

Figure 9-1 Decision criteria for testing H_0:$\mu = 50$ centimeters per second versus H_1:$\mu \neq 50$ centimeters per second.

9-1 Hypothesis Testing

9-1.2 Tests of Statistical Hypotheses

Table 9-1 Decisions in Hypothesis Testing

Decision	H_0 is True	H_0 is False
Fail to reject H_0	No error	Type II error
Reject H_0	Type I error	No error

Probability of Type I and Type II Error

$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true})$
$\beta = P(\text{type II error}) = P(\text{fail to reject } H_0 \text{ when } H_0 \text{ is false})$

Sometimes the type I error probability is called the **significance level**, or the **α-error**, or the **size** of the test.

9-1 Hypothesis Testing

Computing the Probability of Type I Error

$$\alpha = P(\overline{X} < 48.5 \text{ when } \mu = 50) + P(\overline{X} > 51.5 \text{ when } \mu = 50)$$

The z-values that correspond to the critical values 48.5 and 51.5 are

$$z_1 = \frac{48.5 - 50}{0.79} = -1.90 \qquad \text{and} \qquad z_2 = \frac{51.5 - 50}{0.79} = 1.90$$

Therefore

$\alpha = P(Z < -1.90) + P(Z > 1.90) = 0.028717 + 0.028717 = 0.057434$

which implies 5.74% of all random samples would lead to rejection of the hypothesis H_0: $\mu = 50$.

9-1 Hypothesis Testing

Computing the Probability of Type II Error

$$\beta = P(48.5 \le \overline{X} \le 51.5 \text{ when } \mu = 52)$$

The *z*-values corresponding to 48.5 and 51.5 when $\mu = 52$ are

$$z_1 = \frac{48.5 - 52}{0.79} = -4.43 \qquad \text{and} \qquad z_2 = \frac{51.5 - 52}{0.79} = -0.63$$

Hence,

$$\begin{aligned}\beta = P(-4.43 \le Z \le -0.63) &= P(Z \le -0.63) - P(Z \le -4.43) \\ &= 0.2643 - 0.0000 \\ &= 0.2643\end{aligned}$$

which means that the probability that we will fail to reject the false null hypothesis is 0.2643.

9-1 Hypothesis Testing

The **power** of a statistical test

The **power** of a statistical test is the probability of rejecting the null hypothesis H_0 when the alternative hypothesis is true.

- The power is computed as 1 - β, and power can be interpreted as *the probability of correctly rejecting a false null hypothesis.*

- For example, consider the propellant burning rate problem when we are testing H_0 : μ = 50 centimeters per second against H_1 : μ not equal 50 centimeters per second . Suppose that the true value of the mean is μ = 52.

  When *n* = 10, we found that β = 0.2643, so the power of this test is

$$1 - \beta = 1 - 0.2643$$
$$= 0.7357$$

9-1 Hypothesis Testing

9-1.3 One-Sided and Two-Sided Hypotheses

Two-Sided Test:

$$H_0\colon \mu = \mu_0$$
$$H_1\colon \mu \neq \mu_0$$

One-Sided Tests:

$$H_0\colon \mu = \mu_0 \qquad H_1\colon \mu > \mu_0$$

or

$$H_0\colon \mu = \mu_0 \qquad H_1\colon \mu < \mu_0$$

In formulating one-sided alternative hypotheses, one should remember that rejecting H_0 is always a strong conclusion. Consequently, we should put the statement about which it is important to make a strong conclusion in the alternative hypothesis.

9-1 Hypothesis Testing

9-1.4 *P*-Value

The ***P*-value** is the smallest level of significance that would lead to rejection of the null hypothesis H_0 with the given data.

P-value is the **observed significance level.**

9-1 Hypothesis Testing

9-1.4 *P*-Values in Hypothesis Tests

Consider the two-sided hypothesis test H_0:$\mu = 50$ against H_1:$\mu \neq 50$ with $n = 16$ and $\sigma = 2.5$. Suppose that the observed sample mean is centimeters per second. Figure 9-6 shows a critical region for this test with critical values at 51.3 and the symmetric value 48.7.

The *P*-value of the test is the α associated with the critical region. Any smaller value for α expands the critical region and the test fails to reject the null hypothesis when $\bar{x} = 51.3$.

The *P*-value is easy to compute after the test statistic is observed.

$$\bar{x} = 51.3$$

$$\begin{aligned}
P\text{-value} &= 1 - P(48.7 < \bar{X} < 51.3) \\
&= 1 - P\left(\frac{48.7 - 50}{2.5/\sqrt{16}} < Z < \frac{51.3 - 50}{2.5/\sqrt{16}}\right) \\
&= 1 - P(-2.08 < Z < 2.08) \\
&= 1 - 0.962 = 0.038
\end{aligned}$$

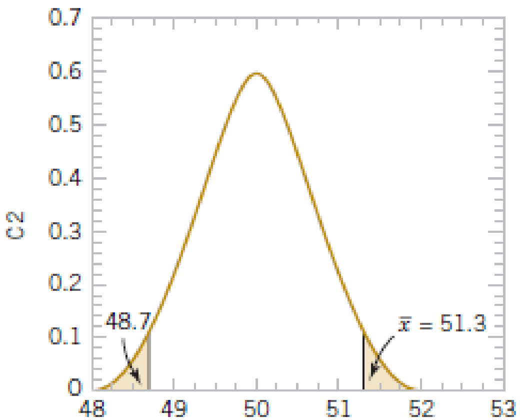

Figure 9-6 *P*-value is area of shaded region when $\bar{x} = 51.3$

WILEY

9-1 Hypothesis Testing

9-1.5 Connection between Hypothesis Tests and Confidence Intervals

A close relationship exists between the test of a hypothesis for θ, and the confidence interval for θ.

If [l, u] is a $100(1-\alpha)\%$ confidence interval for the parameter θ, the test of size α of the hypothesis

$$H_0: \theta = \theta_0$$
$$H_1: \theta \neq \theta_0$$

will lead to rejection of H_0 if and only if θ_0 is **not** in the $100(1-\alpha)\%$ CI [l, u].

9-1 Hypothesis Testing

9-1.6 General Procedure for Hypothesis Tests

1. Identify the parameter of interest.
2. Formulate the null hypothesis, H_0 .
3. Specify an appropriate alternative hypothesis, H_1.
4. Choose a significance level, α.
5. Determine an appropriate test statistic.
6. State the rejection criteria for the statistic.
7. Compute necessary sample quantities for calculating the test statistic.
8. Draw appropriate conclusions.

9-2.1 Hypothesis Tests on the Mean

Consider the two-sided hypothesis test

$$H_0\text{: } \mu = \mu_0$$
$$H_1\text{: } \mu \neq \mu_0$$

The **test statistic** is:

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \qquad (9\text{-}1)$$

9-2.1 Hypothesis Tests on the Mean

Reject H_0 if the observed value of the test statistic z_0 is either:

$$z_0 > z_{\alpha/2} \text{ or } z_0 < -z_{\alpha/2}$$

Fail to reject H_0 if the observed value of the test statistic z_0 is

$$-z_{\alpha/2} < z_0 < z_{\alpha/2}$$

EXAMPLE 9-2 Propellant Burning Rate

Air crew escape systems are powered by a solid propellant. The burning rate of this propellant is an important product characteristic. Specifications require that the mean burning rate must be 50 centimeters per second and the standard deviation is $\sigma = 2$ centimeters per second. The significance level of $\alpha = 0.05$ and a random sample of $n = 25$ has a sample average burning rate of $\bar{x} = 51.3$ centimeters per second. Draw conclusions.

The seven-step procedure is

1. **Parameter of interest:** The parameter of interest is μ, the mean burning rate.
2. **Null hypothesis:** H_0: $\mu = 50$ centimeters per second
3. **Alternative hypothesis:** H_1: $\mu \neq 50$ centimeters per second

EXAMPLE 9-2 Propellant Burning Rate

4. Test statistic: The test statistic is

$$z_0 = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$

5. Reject H_0 if: Reject H_0 if the P-value is less than 0.05. The boundaries of the critical region would be $z_{0.025} = 1.96$ and $-z_{0.025} = -1.96$.

6. Computations: Since $\overline{x} = 51.3$ and $\sigma = 2$,

$$z_0 = \frac{51.3 - 50}{2/\sqrt{25}} = 3.25$$

7. Conclusion: Since $z_0 = 3.25$ and the p-value is $= 2[1 - \Phi(3.25)] = 0.0012$, we reject H_0: $\mu = 50$ at the 0.05 level of significance.

Interpretation: The mean burning rate differs from 50 centimeters per second, based on a sample of 25 measurements.

9-2.2 Type II Error and Choice of Sample Size

Finding the Probability of Type II Error β

Consider the two-sided hypotheses test H_0: $\mu = \mu_0$ and H_1: $\mu \neq \mu_0$

Suppose the null hypothesis is false and the true value of the mean is $\mu = \mu_0 + \delta$, where $\delta > 0$.

The test statistic Z_0 is $$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{X} - (\mu_0 + \delta)}{\sigma/\sqrt{n}} + \frac{\delta\sqrt{n}}{\sigma}$$

Hence, the distribution of Z_0 when H_1 is true is

$$Z_0 \sim N\left(\frac{\delta\sqrt{n}}{\sigma}, 1\right) \quad (9\text{-}2)$$

$$\beta = \Phi\left(z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) \quad (9\text{-}3)$$

9-2 Tests on the Mean of a Normal Distribution, Variance Known

Sample Size for a Two-Sided Test

For a two-sided alternative hypothesis:

$$n \simeq \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{\delta^2} \quad \text{where} \quad \delta = \mu - \mu_0 \qquad (9\text{-}4)$$

Sample Size for a One-Sided Test

For a one-sided alternative hypothesis:

$$n = \frac{(z_{\alpha} + z_{\beta})^2 \sigma^2}{\delta^2} \quad \text{where} \quad \delta = \mu - \mu_0 \qquad (9\text{-}5)$$

EXAMPLE 9-3 Propellant Burning Rate Type II Error

Consider the rocket propellant problem of Example 9-2. The true burning rate is 49 centimeters per second. Find β for the two-sided test with $\alpha = 0.05$, $\sigma = 2$, and $n = 25$?

Here $\delta = 1$ and $z_{\alpha/2} = 1.96$. From Equation 9-3,

$$\beta = \Phi\left(1.96 - \frac{\sqrt{25}}{\sigma}\right) - \Phi\left(-1.96 - \frac{\sqrt{25}}{\sigma}\right)$$
$$= \Phi(-0.54) - \Phi(-4.46) = 0.295$$

The probability is about 0.3 that the test will fail to reject the null hypothesis when the true burning rate is 49 centimeters per second.

Interpretation: A sample size of $n = 25$ results in reasonable, but not great power $= 1 - \beta = 1 - 0.3 = 0.70$.

WILEY

EXAMPLE 9-3 Propellant Burning Rate Type II Error - Continuation

Suppose that the analyst wishes to design the test so that if the true mean burning rate differs from 50 centimeters per second by as much as 1 centimeter per second, the test will detect this (i.e., reject H_0: $\mu = 50$) with a high probability, say, 0.90. Now, we note that $\sigma = 2$, $\delta = 51 - 50 = 1$, $\alpha = 0.05$, and $\beta = 0.10$.

Since $z_{\alpha/2} = z_{0.025} = 1.96$ and $z_\beta = z_{0.10} = 1.28$, the sample size required to detect this departure from H_0: $\mu = 50$ is found by Equation 9-4 as

$$n \simeq \frac{(z_{\alpha/2} + z_\beta)^2 \sigma^2}{\delta^2} = \frac{(1.96 + 1.28)^2 2^2}{(1)^2} \simeq 42$$

The approximation is good here, since $\Phi\left(-z_{\alpha/2} - \delta\sqrt{n}/\sigma\right) = \Phi\left(-1.96 - (1)\sqrt{42}/2\right) = \Phi(-5.20) \simeq 0$, which is small relative to β.

<u>Interpretation:</u> To achieve a much higher power of 0.90 we need a considerably large sample size, $n = 42$ instead of $n = 25$.

9-2 Tests on the Mean of a Normal Distribution, Variance Known

9-2.3 Large Sample Test

A test procedure for the null hypothesis H_0: $\mu = \mu_0$ assuming that the population is normally distributed and that σ^2 known is developed. In most practical situations, σ^2 will be unknown. Even, we may not be certain that the population is normally distributed.

In such cases, if n is large (say, $n > 40$) the sample standard deviation s can be substituted for σ in the test procedures. Thus, while we have given a test for the mean of a normal distribution with known σ^2, it can be easily converted into a **large-sample test procedure for unknown** σ^2 regardless of the form of the distribution of the population.

Exact treatment of the case where the population is normal, σ^2 is unknown, and n is small involves use of the t distribution.

9-3 Tests on the Mean of a Normal Distribution, Variance Unknown

9-3.1 Hypothesis Tests on the Mean

One-Sample *t*-Test

Consider the two-sided hypothesis test

$$H_0\text{: } \mu = \mu_0 \text{ and } H_1\text{: } \mu \neq \mu_0$$

Test statistic:

$$T_0 = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

Alternative hypothesis	Rejection criteria
H_1: $\mu \neq \mu_0$	$t_0 > t_{\alpha/2,n-1}$ or $t_0 < -t_{\alpha/2,n-1}$
H_1: $\mu > \mu_0$	$t_0 > t_{\alpha,n-1}$
H_1: $\mu < \mu_0$	$t_0 < -t_{\alpha,n-1}$

EXAMPLE 9-6 Golf Club Design

An experiment was performed in which 15 drivers produced by a particular club maker were selected at random and their coefficients of restitution measured. It is of interest to determine if there is evidence (with $\alpha = 0.05$) to support a claim that the mean coefficient of restitution exceeds 0.82.

The observations are:

0.8411	0.8191	0.8182	0.8125	0.8750
0.8580	0.8532	0.8483	0.8276	0.7983
0.8042	0.8730	0.8282	0.8359	0.8660

The sample mean and sample standard deviation are $\bar{x} = 0.83725$ and $s = 0.02456$. The objective of the experimenter is to demonstrate that the mean coefficient of restitution exceeds 0.82, hence a one-sided alternative hypothesis is appropriate.

The seven-step procedure for hypothesis testing is as follows:

1. **Parameter of interest:** The parameter of interest is the mean coefficient of restitution, μ.

2. **Null hypothesis:** H_0: $\mu = 0.82$

3. **Alternative hypothesis:** H_1: $\mu > 0.82$

EXAMPLE 9-6 Golf Club Design - Continued

4. Test Statistic: The test statistic is

5. Reject H_0 if: Reject H_0 if the *P*-value is less than 0.05.

6. Computations: Since $\overline{x} = 0.83725$, $s = 0.02456$, $\mu = 0.82$, and $n = 15$, we have

$$t_0 = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} \qquad t_0 = \frac{0.83725 - 0.82}{0.02456/\sqrt{15}} = 2.72$$

7. Conclusions: From Appendix A Table II, for a *t* distribution with 14 degrees of freedom, $t_0 = 2.72$ falls between two values: 2.624, for which $\alpha = 0.01$, and 2.977, for which $\alpha = 0.005$. Since, this is a one-tailed test the *P*-value is between those two values, that is, $0.005 < P < 0.01$. Therefore, since $P < 0.05$, we reject H_0 and conclude that the mean coefficient of restitution exceeds 0.82.

Interpretation: There is strong evidence to conclude that the mean coefficient of restitution exceeds 0.82.

9-3 Tests on the Mean of a Normal Distribution, Variance Unknown

9-3.2 Type II Error and Choice of Sample Size

The type II error of the two-sided alternative would be

$$\beta = P(-t_{\alpha/2,n-1} \le T_0 \le t_{\alpha/2,n-1} \mid \delta \ne 0)$$
$$= P(-t_{\alpha/2,n-1} \le T_0' \le t_{\alpha/2,n-1})$$

Curves are provided for two-sided alternatives on Charts VII*e and VIIf .* The abscissa scale factor *d* on these charts is defined as

$$d = \frac{|\mu - \mu_0|}{\sigma} = \frac{|\delta|}{\sigma}$$

For the one-sided alternative $\mu > \mu_0$ or $\mu < \mu_0$, use charts VII*g and VIIh.* The abscissa scale factor *d* on these charts is defined as

$$d = \frac{|\mu - \mu_0|}{\sigma} = \frac{|\delta|}{\sigma}$$

9-3 Tests on the Mean of a Normal Distribution, Variance Unknown

EXAMPLE 9-7 Golf Club Design Sample Size

Consider the golf club testing problem from Example 9-6. If the mean coefficient of restitution exceeds 0.82 by as much as 0.02, is the sample size $n = 15$ adequate to ensure that H_0: $\mu = 0.82$ will be rejected with probability at least 0.8?

To solve this problem, we will use the sample standard deviation $s = 0.02456$ to estimate σ. Then $d = |\delta|/\sigma = 0.02/0.02456 = 0.81$. By referring to the operating characteristic curves in Appendix Chart VII*g* (for $\alpha = 0.05$) with $d = 0.81$ and $n = 15$, we find that $\beta = 0.10$, approximately. Thus, the probability of rejecting H_0: $\mu = 0.82$ if the true mean exceeds this by 0.02 is approximately $1 - \beta = 1 - 0.10 = 0.90$, and we conclude that a sample size of $n = 15$ is adequate to provide the desired sensitivity.

9-4 Hypothesis Tests on the Variance and Standard Deviation of a Normal Distribution

9-4.1 Hypothesis Test on the Variance

Suppose that we wish to test the hypothesis that the variance of a normal population σ^2 equals a specified value, say , or equivalently, that the standard deviation σ is equal to σ_0. Let $X_1, X_2, \ldots, X_n$ be a random sample of n observations from this population. To test

$$H_0: \sigma^2 = \sigma_0^2$$
$$H_1: \sigma^2 \neq \sigma_0^2 \qquad (9\text{-}6)$$

we will use the test statistic:

$$X_0^2 = \frac{(n-1)S^2}{\sigma_0^2} \qquad (9\text{-}7)$$

9-4.1 Hypothesis Test on the Variance

If the null hypothesis $H_0: \sigma^2 = \sigma_0^2$ is true, the test statistic X_0^2 defined in Equation 9-7 follows the chi-square distribution with $n - 1$ degrees of freedom. This is the reference distribution for this test procedure. Therefore, we calculate χ_0^2, the value of the test statistic X_0^2, and the null hypothesis $H_0: \sigma^2 = \sigma_0^2$ would be rejected if

$$\chi_0^2 > \chi_{\alpha/2,n-1}^2 \qquad \text{or if} \qquad \chi_0^2 < \chi_{1-\alpha/2,n-1}^2$$

where $\chi_{\alpha/2,n-1}^2$ and $\chi_{1-\alpha/2,n-1}^2$ are the upper and lower 100 α /2 percentage points of the chi-square distribution with $n - 1$ degrees of freedom, respectively. Figure 9-17(a) shows the critical region.

9-4 Hypothesis Tests on the Variance and Standard Deviation of a Normal Distribution

9-4.1 Hypothesis Test on the Variance

The same test statistic is used for one-sided alternative hypotheses. For the one-sided hypotheses.

$$H_0: \sigma^2 = \sigma_0^2 \qquad (9\text{-}8)$$
$$H_1: \sigma^2 > \sigma_0^2$$

we would reject H_0 if $\chi_0^2 > \chi_{\alpha,n-1}^2$, whereas for the other one-sided hypotheses

$$H_0: \sigma^2 = \sigma_0^2 \qquad (9\text{-}9)$$
$$H_1: \sigma^2 < \sigma_0^2$$

we would reject H_0 if $\chi_0^2 < \chi_{\alpha-1,n-1}^2$. The one-sided critical regions are shown in Fig. 9-17(b) and (c).

9-4.1 Hypothesis Tests on the Variance

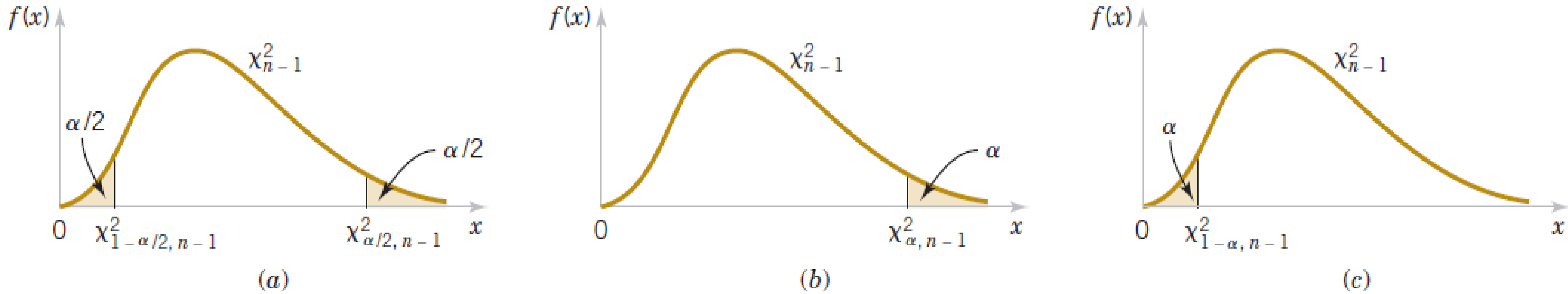

FIGURE 9-17 Reference distribution for the test of $H_0: \sigma^2 = \sigma_0^2$ with critical region values for (a) $H_1: \sigma^2 \neq \sigma_0^2$. (b) $H_1: \sigma^2 > \sigma_0^2$. (c) $H_1: \sigma^2 < \sigma_0^2$.

EXAMPLE 9-8 Automated Filling

An automated filling machine is used to fill bottles with liquid detergent. A random sample of 20 bottles results in a sample variance of fill volume of $s^2 = 0.0153$ (fluid ounces)^2. If the variance of fill volume exceeds 0.01 (fluid ounces)^2, an unacceptable proportion of bottles will be underfilled or overfilled. Is there evidence in the sample data to suggest that the manufacturer has a problem with underfilled or overfilled bottles? Use $\alpha = 0.05$, and assume that fill volume has a normal distribution.

Using the seven-step procedure results in the following:

1. **Parameter of Interest:** The parameter of interest is the population variance σ^2.

2. **Null hypothesis:** H_0: $\sigma^2 = 0.01$

3. **Alternative hypothesis:** H_1: $\sigma^2 > 0.01$

9-4 Hypothesis Tests on the Variance and Standard Deviation of a Normal Distribution

Example 9-8

4. **Test statistic:** The test statistic is

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

5. **Reject H_0:** Use α **=** 0.05**,** and reject H_0 if $\chi_0^2 > \chi^2_{0.05,19} = 30.14$.

6. **Computations:**

$$\chi_0^2 = \frac{19(0.0153)}{0.01} = 29.07$$

7. **Conclusions:** Since $\chi_0^2 = 29.07 < \chi^2_{0.05,19} = 30.14$, we conclude that there is no strong evidence that the variance of fill volume exceeds 0.01 (fluid ounces)^2. So there is no strong evidence of a problem with incorrectly filled bottles.

9-4.2 Type II Error and Choice of Sample Size

For the two-sided alternative hypothesis:

$$\lambda = \frac{\sigma}{\sigma_0}$$

Operating characteristic curves are provided in Charts VI*i* and VI*j*:

EXAMPLE 9-9 Automated Filling Sample Size

Consider the bottle-filling problem from Example 9-8. If the variance of the filling process exceeds 0.01 (fluid ounces)^2, too many bottles will be underfilled. Thus, the hypothesized value of the standard deviation is $\sigma_0 = 0.10$. Suppose that if the true standard deviation of the filling process exceeds this value by 25%, we would like to detect this with probability at least 0.8. Is the sample size of $n = 20$ adequate?

To solve this problem, note that we require

$$\lambda = \frac{\sigma}{\sigma_0} = \frac{0.125}{0.10} = 1.25$$

This is the abscissa parameter for Chart VII*k.* From this chart, with $n = 20$ and $\lambda = 1.25$, we find that $\beta \simeq 0.6$. Therefore, there is only about a 40% chance that the null hypothesis will be rejected if the true standard deviation is really as large as $\sigma = 0.125$ fluid ounce.

To reduce the β-error, a larger sample size must be used. From the operating characteristic curve with $\beta = 0.20$ and $\lambda = 1.25$, we find that $n = 75$, approximately. Thus, if we want the test to perform as required above, the sample size must be at least 75 bottles.

9-5 Tests on a Population Proportion

9-5.1 Large-Sample Tests on a Proportion

Many engineering decision problems include hypothesis testing about *p*.

$$H_0 : p = p_0$$
$$H_1 : p \neq p_0$$

An appropriate **test statistic** is

$$Z_0 = \frac{X - np_0}{\sqrt{np_0(1 - p_0)}} \qquad (9\text{-}10)$$

9-5 Tests on a Population Proportion

EXAMPLE 9-10 Automobile Engine Controller A semiconductor manufacturer produces controllers used in automobile engine applications. The customer requires that the process fallout or fraction defective at a critical manufacturing step not exceed 0.05 and that the manufacturer demonstrate process capability at this level of quality using $\alpha = 0.05$. The semiconductor manufacturer takes a random sample of 200 devices and finds that four of them are defective. Can the manufacturer demonstrate process capability for the customer?

We may solve this problem using the seven-step hypothesis-testing procedure as follows:

1. **Parameter of Interest:** The parameter of interest is the process fraction defective *p.*

2. **Null hypothesis:** $H_0{:}p = 0.05$

3. **Alternative hypothesis:** H_1: $p < 0.05$

This formulation of the problem will allow the manufacturer to make a strong claim about process capability if the null hypothesis H_0: $p = 0.05$ is rejected.

9-5 Tests on a Population Proportion

Example 9-10

4. The test statistic is (from Equation 9-10)

$$z_0 = \frac{x - np_0}{\sqrt{np_0(1 - p_0)}}$$

where $x = 4$, $n = 200$, and $p_0 = 0.05$.

5. **Reject H_0 if:** Reject H_0: $p = 0.05$ if the p-value is less than 0.05.

6. **Computations:** The test statistic is

$$z_0 = \frac{4 - 200(0.05)}{\sqrt{200(0.05)(0.95)}} = -1.95$$

7. **Conclusions:** Since $z_0 = -1.95$, the *P*-value is $\Phi(-1.95) = 0.0256$, so we reject H_0 and conclude that the process fraction defective p is less than 0.05. Practical Interpretation: We conclude that the process is capable.

9-5 Tests on a Population Proportion

Another form of the test statistic Z_0 is

$$Z_0 = \frac{X/n - p_0}{\sqrt{p_0(1 - p_0)/n}} \quad \text{or} \quad Z_0 = \frac{\hat{P} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

9-5 Tests on a Population Proportion

9-5.2 Type II Error and Choice of Sample Size

For a two-sided alternative

$$\beta = \Phi\left(\frac{p_0 - p + z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right) - \Phi\left(\frac{p_0 - p - z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right) \quad (9\text{-}11)$$

If the alternative is $p < p_0$

$$\beta = 1 - \Phi\left(\frac{p_0 - p - z_{\alpha}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right) \quad (9\text{-}12)$$

If the alternative is $p > p_0$

$$\beta = \Phi\left(\frac{p_0 - p + z_{\alpha}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right) \quad (9\text{-}13)$$

9-5 Tests on a Population Proportion

9-5.3 Type II Error and Choice of Sample Size

For a two-sided alternative

$$n = \left[\frac{z_{\alpha/2}\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p(1-p)}}{p - p_0} \right]^2 \qquad (9\text{-}14)$$

For a one-sided alternative

$$n = \left[\frac{z_{\alpha}\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p(1-p)}}{p - p_0} \right]^2 \qquad (9\text{-}15)$$

9-5 Tests on a Population Proportion

Example 9-11 Automobile Engine Controller Type II Error
Consider the semiconductor manufacturer from Example 9-10. Suppose that its process fallout is really p = 0.03. What is the β-error for a test of process capability that uses n = 200 and a = 0.05?

The β-error can be computed using Equation 9-12 as follows:

$$\beta = 1 - \Phi\left[\frac{0.05 - 0.03 - (1.645)\sqrt{0.05(0.95)/200}}{\sqrt{0.03(1-0.03)/200}}\right]$$
$$= 1 - \Phi(-0.44) = 0.67$$

Thus, the probability is about 0.7 that the semiconductor manufacturer will fail to conclude that the process is capable if the true process fraction defective is p = 0.03 (3%). That is, the power of the test against this particular alternative is only about 0.3. This appears to be a large β-error (or small power), but the difference between p = 0.05 and p = 0.03 is fairly small, and the sample size n = 200 is not particularly large.

9-5 Tests on a Population Proportion

Example 9-11

Suppose that the semiconductor manufacturer was willing to accept a β-error as large as 0.10 if the true value of the process fraction defective was $p = 0.03$. If the manufacturer continues to use a = 0.05, what sample size would be required?

The required sample size can be computed from Equation 9-15 as follows:

$$n = \left[\frac{1.645\sqrt{0.05(0.95)} + 1.28\sqrt{0.03(0.97)}}{0.03 - 0.05} \right]^2$$

$$\simeq 832$$

where we have used $p = 0.03$ in Equation 9-15.

Conclusion: Note that $n = 832$ is a very large sample size. However, we are trying to detect a fairly small deviation from the null value $p_0 = 0.05$.

WILEY

9-7 Testing for Goodness of Fit

- The test is based on the chi-square distribution.
- Assume there is a sample of size *n* from a population whose probability distribution is unknown.
- Let O_i be the observed frequency in the *i*th class interval.
- Let E_i be the expected frequency in the *i*th class interval.

The test statistic is

$$X_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad (9\text{-}16)$$

9-7 Testing for Goodness of Fit

EXAMPLE 9-12 **Printed Circuit Board Defects**

Poisson Distribution

The number of defects in printed circuit boards is hypothesized to follow a Poisson distribution. A random sample of n = 60 printed boards has been collected, and the following number of defects observed.

Number of Defects	Observed Frequency
0	32
1	15
2	9
3	4

9-7 Testing for Goodness of Fit

Example 9-12

The mean of the assumed Poisson distribution in this example is unknown and must be estimated from the sample data. The estimate of the mean number of defects per board is the sample average, that is, (32·0 + 15·1 + 9·2 + 4·3)/60 = 0.75. From the Poisson distribution with parameter 0.75, we may compute p_i, the theoretical, hypothesized probability associated with the *i*th class interval. Since each class interval corresponds to a particular number of defects, we may find the p_i as follows:

$$p_1 = P(X = 0) = \frac{e^{-0.75}(0.75)^0}{0!} = 0.472$$

$$p_2 = P(X = 1) = \frac{e^{-0.75}(0.75)^1}{1!} = 0.354$$

$$p_3 = P(X = 2) = \frac{e^{-0.75}(0.75)^2}{2!} = 0.133$$

$$p_4 = P(X \geq 3) = 1 - \left(p_1 + p_2 + p_3\right) = 0.041$$

WILEY

9-7 Testing for Goodness of Fit

Example 9-12

The expected frequencies are computed by multiplying the sample size n = 60 times the probabilities p_i. That is, $E_i = np_i$. The expected frequencies follow:

Number of Defects	Probability	Expected Frequency
0	0.472	28.32
1	0.354	21.24
2	0.133	7.98
3 (or more)	0.041	2.46

WILEY

9-7 Testing for Goodness of Fit

Example 9-12

Since the expected frequency in the last cell is less than 3, we combine the last two cells:

Number of Defects	Observed Frequency	Expected Frequency
0	32	28.32
1	15	21.24
2 (or more)	13	10.44

The chi-square test statistic in Equation 9-16 will have $k - p - 1 = 3 - 1 - 1 = 1$ degree of freedom, because the mean of the Poisson distribution was estimated from the data.

9-7 Testing for Goodness of Fit

Example 9-12

The seven-step hypothesis-testing procedure may now be applied, using $\alpha = 0.05$, as follows:

1. **Parameter of interest:** The variable of interest is the form of the distribution of defects in printed circuit boards.

2. **Null hypothesis:** H_0: The form of the distribution of defects is Poisson.

3. **Alternative hypothesis:** H_1: The form of the distribution of defects is not Poisson.

4. **Test statistic:** The test statistic is

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(o_i - E_i)^2}{E_i}$$

9-7 Testing for Goodness of Fit

Example 9-12

5. **Reject H_0 if:** Reject H_0 if the *P*-value is less than 0.05.

6. **Computations:**

$$\chi_0^2 = \frac{(32-28.32)^2}{28.32} + \frac{(15-21.24)^2}{21.24} + \frac{(13-10.44)^2}{10.44} = 2.94$$

7. **Conclusions:** We find from Appendix Table III that $\chi^2_{0.10,1} = 2.71$ and $\chi^2_{0.05,1} = 3.84$. Because $\chi_0^2 = 2.94$ lies between these values, we conclude that the *P*-value is between 0.05 and 0.10. Therefore, since the *P*-value exceeds 0.05 we are unable to reject the null hypothesis that the distribution of defects in printed circuit boards is Poisson. The exact *P*-value computed from Minitab is 0.0864.

9-8 Contingency Table Tests

Many times, the n elements of a sample from a population may be classified according to two different criteria. It is then of interest to know whether the two methods of classification are statistically independent;

Table 9-2 An $r \times c$ Contingency Table

		Columns			
		1	2	…	c
Rows	1	O_{11}	O_{12}	…	O_{1c}
	2	O_{21}	O_{22}	…	O_{2c}
	⋮	⋮	⋮	⋮	⋮
	r	O_{r1}	O_{r2}	…	O_{rc}

9-8 Contingency Table Tests

We are interested in testing the hypothesis that the row-and-column methods of classification are independent. If we reject this hypothesis, we conclude there is some interaction between the two criteria of classification. The exact test procedures are difficult to obtain, but an approximate test statistic is valid for large *n*. Let p_{ij} be the probability that a randomly selected element falls in the *ij*th cell, given that the two classifications are independent. Then $p_{ij}=u_i v_j$, where u_i is the probability that a randomly selected element falls in row class *i* and *vj* is the probability that a randomly selected element falls in column class ***j***. Now. assuming independence, the estimators of u_i and v_j are

$$\hat{u}_i = \frac{1}{n}\sum_{j=1}^{c} O_{ij}$$

$$\hat{v}_j = \frac{1}{n}\sum_{i=1}^{r} O_{ij} \qquad (9\text{-}17)$$

9-8 Contingency Table Tests

Therefore, the expected frequency of each cell is

$$E_{ij} = n\hat{u}_i\hat{v}_j = \frac{1}{n}\sum_{j=1}^{c} O_{ij} \sum_{i=1}^{r} O_{ij} \quad (9\text{-}18)$$

Then, for large *n*, the statistic

$$\chi_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (9\text{-}19)$$

has an approximate chi-square distribution with (*r* –1)(*c* – 1) degrees of freedom if the null hypothesis is true. We should reject the null hypothesis if the value of the test statistic χ_0^2 is too large. The *P*-value would be calculated as the probability beyond χ_0^2 on the $\chi^2_{(r-1)(c-1)}$ distribution, or $P = p\left(\chi^2_{(r-1)(c-1)} > \chi_0^2\right)$. For a fixed-level test, we would reject the hypothesis of independence if the observed value of the test statistic χ_0^2 exceeded $\chi^2_{\alpha,(r-1)(c-1)}$.

9-8 Contingency Table Tests

A company has to choose among three health insurance plans. Management wishes to know whether the preference for plans is independent of job classification and wants to use $\alpha = 0.05$.

The opinions of a random sample of 500 employees are shown in Table 9-3.

Table 9-3 Observed Data for Example 9-14

	Health Insurance Plan			
Job Classification	1	2	3	Totals
Salaried workers	160	140	40	340
Hourly workers	40	60	60	160
Totals	200	200	100	500

To find the expected frequencies, we must first compute $\hat{u}_1 = (340/500) = 0.68$, $\hat{u}_2 = (160/500) = 0.32$, $\hat{v}_1 = (200/500) = 0.40$, $\hat{v}_2 = (200/500) = 0.40$, and $\hat{v}_3 = (100/500) = 0.20$

WILEY

9-8 Contingency Table Tests

The expected frequencies may now be computed from Equation 9-18.

For example, the expected number of salaried workers favoring health insurance plan 1 is

$$E_{11} = n\hat{u}_1\hat{v}_1 = 500(0.68)(0.40) = 136$$

The expected frequencies are shown in below table

	Health Insurance Plan			
Job Classification	1	2	3	Totals
Salaried workers	160	140	40	340
Hourly workers	40	60	60	160
Totals	200	200	100	500

9-8 Contingency Table Tests

Example 9-14

The seven-step hypothesis-testing procedure may now be applied to this problem.

1. Parameter of Interest: The variable of interest is employee preference among health insurance plans.

2. Null hypothesis: H_0: Preference is independent of salaried versus hourly job classification.

3. Alternative hypothesis: H_1: Preference is not independent of salaried versus hourly job classification.

4. Test statistic: The test statistic is

$$\chi_0^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(o_{ij} - E_{ij})^2}{E_{ij}}$$

5. Reject H_0 if: We will use a fixed-significance level test with a = 0.05. Therefore, since r = 2 and c = 3, the degrees of freedom for chi-square are (r – 1)(c – 1) = (1)(2) = 2, and we would reject H_0 if $\chi_0^2 > \chi_{0.05,2}^2 = 5.99$.

9-8 Contingency Table Tests

Example 9-14

6. **Computations:**

$$\chi_0^2 = \sum_{i=1}^{2} \sum_{j=1}^{3} \frac{(o_{ij} - E_{ij})^2}{E_{ij}}$$

$$= \frac{(160-136)^2}{136} + \frac{(140-136)^2}{136} + \frac{(40-68)^2}{68} + \frac{(40-64)^2}{64} + \frac{(60-64)^2}{64} + \frac{(60-32)^2}{32}$$

$$= 49.63$$

7. **Conclusions:** Since $\chi_0^2 = 49.63 > \chi_{0.05,2}^2 = 5.99$, we reject the hypothesis of independence and conclude that the preference for health insurance plans is not independent of job classification. The *P*-value for $\chi_0^2 = 49.63$ is $P = 1.671 \times 10^{-11}$. (This value was computed from computer software.)

WILEY

9-9 Nonparametric Procedures

9-9.1 The Sign Test

- The sign test is used to test hypotheses about the median $\overline{\mu}$ of a continuous distribution.
- Suppose that the hypotheses are $H_0: \tilde{\mu} = \tilde{\mu}_0 \qquad H_1: \tilde{\mu} < \tilde{\mu}_0$
- Test procedure: Let $X_1, X_2, \ldots, X_n$ be a random sample from the population of interest. Form the differences $X_i - \tilde{\mu}_0$, i =1,2,…,n.
- An appropriate test statistic is the number of these differences that are positive, say R^+.
- *P*-value for the observed number of plus signs r^+ can be calculated directly from the binomial distribution.
- If the computed *P*-value is less than or equal to the significance level α, we will reject H_0.
- For two-sided alternative $H_0: \tilde{\mu} = \tilde{\mu}_0 \qquad H_1: \tilde{\mu} \neq \tilde{\mu}_0$
- If the computed *P*-value is less than the significance level α, we will reject H_0.

EXAMPLE 9-15 Propellant Shear Strength Sign Test

Montgomery, Peck, and Vining (2012) reported 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 9-5. Test the hypothesis that the median shear strength is 2000 psi, using $\alpha = 0.05$.

Table 9-5 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	–

EXAMPLE 9-15 Propellant Shear Strength Sign Test - Continued

The seven-step hypothesis-testing procedure is:

1. **Parameter of Interest:** The variable of interest is the median of the distribution of propellant shear strength.
2. **Null hypothesis:** $H_0 : \tilde{\mu} = 2000 \text{ psi}$
3. **Alternative hypothesis:** $H_1 : \tilde{\mu} \neq 2000 \text{ psi}$
4. **Test statistic:** The test statistic is the observed number of plus differences in Table 9-5, i.e., $r^+ = 14$.
5. **Reject H_0:** If the *P*-value corresponding to $r^+ = 14$ *is* less than or equal to $\alpha = 0.05$
6. **Computations :** $r^+ = 14$ is greater than $n/2 = 20/2 = 10$.

   P-value :

$$P = 2P\left(R^+ \geq 14 \text{ when } p = \frac{1}{2} \right)$$

$$= 2\sum_{r=14}^{20} \binom{20}{r} (0.5)^r (0.5)^{20-r}$$

$$= 0.1153$$

7. **Conclusions:** Since the P-value is greater than $\alpha = 0.05$ we cannot reject the null hypotheses that the median shear strength is 2000 psi.

9-9 Nonparametric Procedures

9-9.2 The Wilcoxon Signed-Rank Test

- A test procedure that uses both direction (sign) and magnitude.

- Suppose that the hypotheses are $H_0: \mu = \mu_0 \qquad H_1: \mu \neq \mu_0$

  Test procedure : Let $X_1, X_2, \ldots, X_n$ be a random sample from continuous and symmetric distribution with mean (and Median) μ. Form the differences $X_i - \mu_0$.

- Rank the absolute differences $|X_i - \mu_0|$ in ascending order, and give the ranks to 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, and let $W = min(W^+, W^-)$.
  Critical values of *W, can be found in* Appendix Table IX.

- If the computed value is less than the critical value, we will reject H_0 .

- For one-sided alternatives $H_1: \mu > \mu_0$ reject H_0 if $W^- \leq$ critical value
  $H_1: \mu < \mu_0$ reject H_0 if $W^+ \leq$ critical value

WILEY

EXAMPLE 9-16 Propellant Shear Strength-Wilcoxon Signed-Rank Test

Let's illustrate the Wilcoxon signed rank test by applying it to the propellant shear strength data from Table 9-5. Assume that the underlying distribution is a continuous symmetric distribution. Test the hypothesis that the median shear strength is 2000 psi, using $\alpha = 0.05$.

The seven-step hypothesis-testing procedure is:

1. **Parameter of Interest:** The variable of interest is the mean or median of the distribution of propellant shear strength.
2. **Null hypothesis:** $H_0: \mu = 2000 \text{ psi}$
3. **Alternative hypothesis:** $H_1: \mu \neq 2000 \text{ psi}$
4. **Test statistic:** The test statistic is $W = min(W^+, W^-)$
5. **Reject H_0 if:** $W \leq 52$ (from Appendix Table IX).
6. **Computations :** 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 $w^- = (7 + 8 + 9 + 10 + 12 + 14) = 60$.

EXAMPLE 9-16 Propellant Shear Strength-Wilcoxon Signed-Rank Test

Observation *i*	Differences 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

$W = \min(W^+, W^-) = \min(150 , 60) = 60$

7. **Conclusions:** Since W = 60 is not ≤ 52 we fail to reject the null hypotheses that the mean or median shear strength is 2000 psi.

Important Terms & Concepts of Chapter 9

- α and β
- Connection between hypothesis tests & confidence intervals
- Critical region for a test statistic
- Goodness-of-fit test
- Homogeneity test
- Inference
- Independence test
- Non-parametric or distribution-free methods
- Normal approximation to non-parametric tests
- Null distribution
- Null hypothesis
- 1 & 2-sided alternative hypotheses
- Operating Characteristic (OC) curves
- Power of a test
- P-value
- Ranks
- Reference distribution for a test statistic
- Sample size determination for hypothesis tests
- Significance level of a test
- Sign test
- Statistical hypotheses
- Statistical vs. practical significance
- Test statistic
- Type I & Type II errors
- Wilcoxen signed-rank test