# 9

# Tests of Hypotheses
# for a Single Sample

## CHAPTER OUTLINE

## LEARNING OBJECTIVES

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

1. Structure engineering decision-making problems as hypothesis tests
2. Test hypotheses on the mean of a normal distribution using either a Z-test or a t-test procedure
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 hypotheses tests
6. Compute power, type II error probability, and make sample size selection decisions for tests on means, variances, and proportions
7. Explain and use the relationship between confidence intervals and 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

Statistical hypothesis testing and confidence interval estimation of parameters are the fundamental methods used at the data analysis stage of a **comparative experiment,** in which the engineer is interested, for example, in comparing the mean of a population to a specified value.

## Definition

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

# 9-1 Hypothesis Testing

## 9-1.1 Statistical Hypotheses

For example, suppose that we are interested in the burning rate of a solid propellant used to power aircrew escape systems.

• Now burning rate is a random variable that can be described by a probability distribution.

• Suppose that our interest focuses on the **mean** burning rate (a parameter of this distribution).

• Specifically, we are interested in deciding whether or not the mean burning rate is 50 centimeters per second.

Two-sided Alternative Hypothesis

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

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

One-sided Alternative Hypotheses

$H_0$: $\mu = 50$ 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

## 9-1.1 Statistical Hypotheses

**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 Hypothesis Testing

## 9-1.2 Tests of Statistical Hypotheses

$$H_0: \mu = 50 \text{ centimeters per second}$$
$$H_1: \mu \neq 50 \text{ centimeters per second}$$

| Reject $H_0$ | Fail to Reject $H_0$ | Reject $H_0$ |
|---|---|---|
| $\mu \neq 50$ cm/s | $\mu = 50$ cm/s | $\mu \neq 50$ cm/s |

48.5          50          51.5          $\bar{x}$

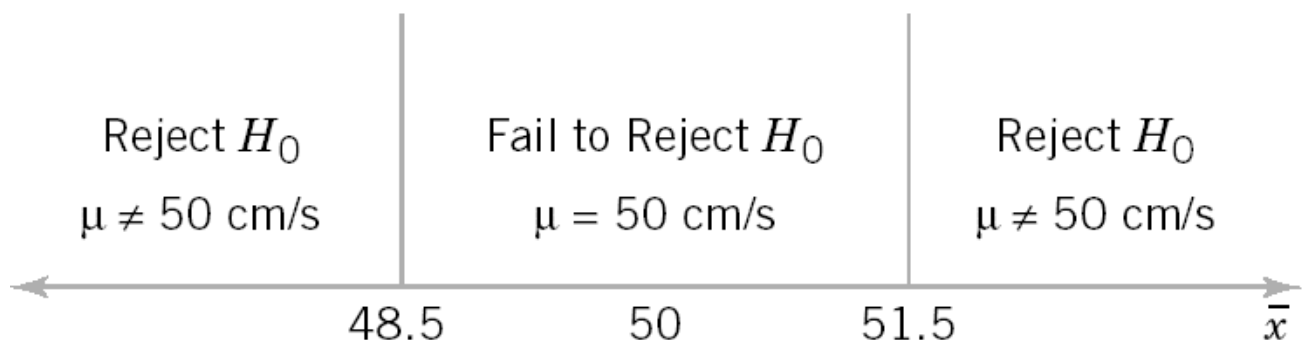**Figure 9-1** Decision criteria for testing $H_0: \mu = 50$ centimeters per second versus $H_1: \mu \neq 50$ centimeters per second.

# 9-1 Hypothesis Testing

## 9-1.2 Tests of Statistical Hypotheses

### Definitions

> Rejecting the null hypothesis $H_0$ when it is true is defined as a **type I error.**

> Failing to reject the null hypothesis when it is false is defined as a **type II error.**

# 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 |

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true})$$

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

## 9-1.2 Tests of Statistical Hypotheses

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

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

$$\alpha = P(Z < -1.90) + P(Z > 1.90) = 0.028717 + 0.028717 = 0.057434$$
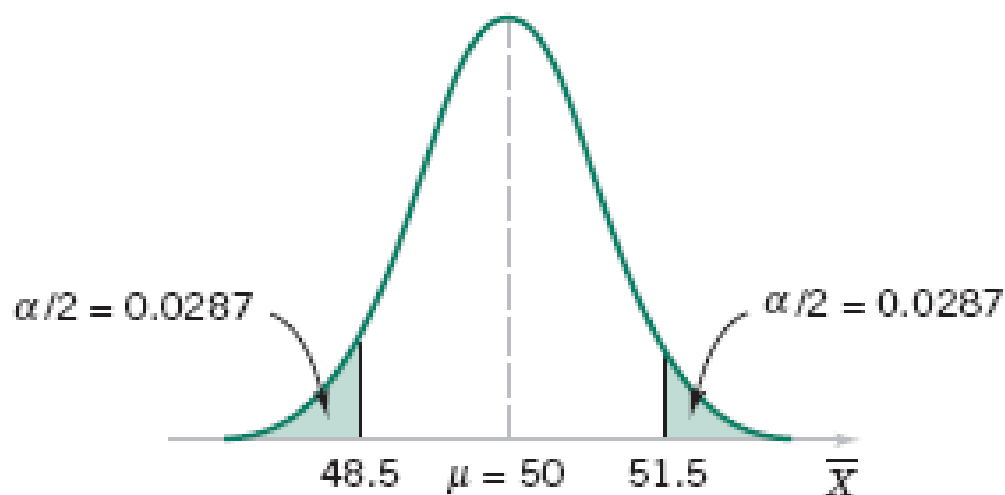
# 9-1 Hypothesis Testing



Figure 9-2   The critical region for $H_0$: $\mu = 50$ versus $H_1$: $\mu \neq 50$ and $n = 10$.

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true}) \qquad (9\text{-}3)$$

# 9-1 Hypothesis Testing

$$\beta = P(\text{type II error}) = P(\text{fail to reject } H_0 \text{ when } H_0 \text{ is false}) \qquad (9\text{-}4)$$



**Figure 9-3** The probability of type II error when $\mu = 52$ and $n = 10$.

# 9-1 Hypothesis Testing

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

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

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

Therefore

$$\beta = P(-4.43 \le Z \le -0.63) = P(Z \le -0.63) - P(Z \le -4.43)$$
$$= 0.2643 - 0.0000 = 0.2643$$

# 9-1 Hypothesis Testing

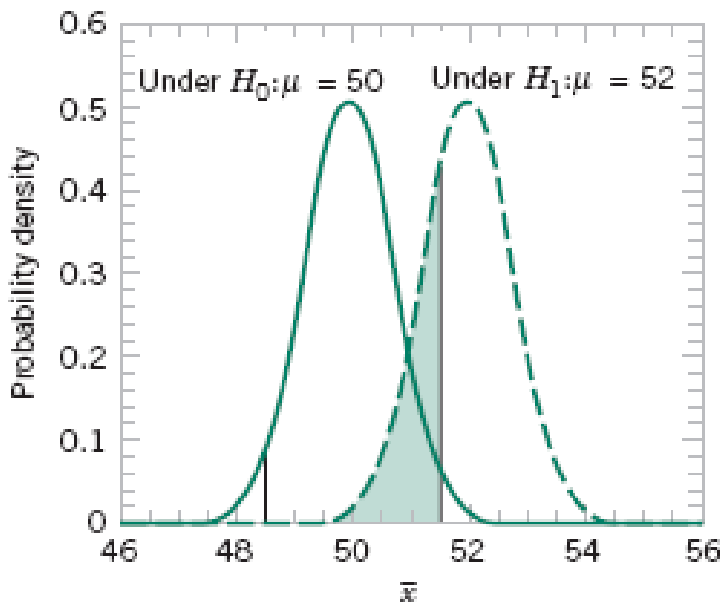$$\beta = P(48.5 \leq \overline{X} \leq 51.5 \text{ when } \mu = 50.5)$$



**Figure 9-4** The probability of type II error when $\mu = 50.5$ and $n = 10$.

# 9-1 Hypothesis Testing

$$\beta = P(48.5 \leq \overline{X} \leq 51.5 \text{ when } \mu = 50.5)$$

As shown in Fig. 9-4, the $z$-values corresponding to 48.5 and 51.5 when $\mu = 50.5$ are

$$z_1 = \frac{48.5 - 50.5}{0.79} = -2.53 \quad \text{and} \quad z_2 = \frac{51.5 - 50.5}{0.79} = 1.27$$

Therefore

$$\beta = P(-2.53 \leq Z \leq 1.27) = P(Z \leq 1.27) - P(Z \leq -2.53)$$
$$= 0.8980 - 0.0057 = 0.8923$$

# 9-1 Hypothesis Testing

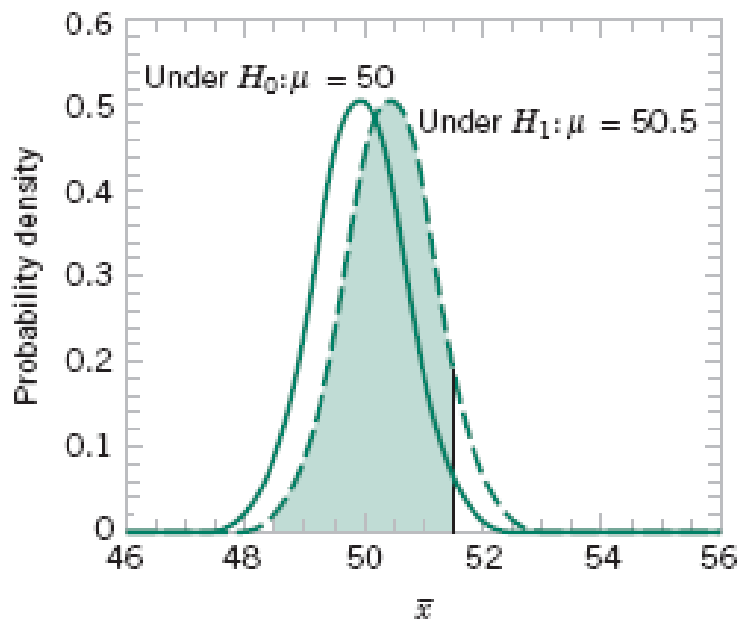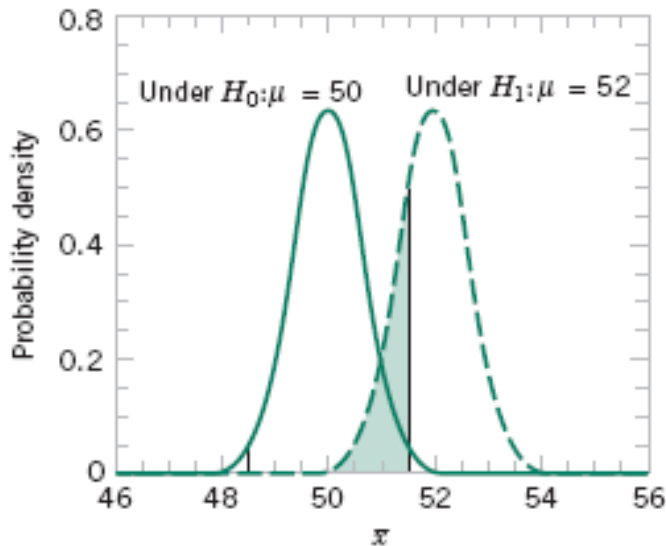$$\beta = P(48.5 \leq \overline{X} \leq 51.5 \text{ when } \mu = 52)$$



**Figure 9-5** The probability of type II error when $\mu = 2$ and $n = 16$.

# 9-1 Hypothesis Testing

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

When $n = 16$, the standard deviation of $\overline{X}$ is $\sigma/\sqrt{n} = 2.5/\sqrt{16} = 0.625$, and the z-values corresponding to 48.5 and 51.5 when $\mu = 52$ are

$$z_1 = \frac{48.5 - 52}{0.625} = -5.60 \quad \text{and} \quad z_2 = \frac{51.5 - 52}{0.625} = -0.80$$

Therefore

$$\beta = P(-5.60 \leq Z \leq -0.80) = P(Z \leq -0.80) - P(Z \leq -5.60)$$
$$= 0.2119 - 0.0000 = 0.2119$$

# 9-1 Hypothesis Testing

| Acceptance Region | Sample Size | $\alpha$ | $\beta$ at $\mu = 52$ | $\beta$ at $\mu = 50.5$ |
|---|---|---|---|---|
| $48.5 < \bar{x} < 51.5$ | 10 | 0.0576 | 0.2643 | 0.8923 |
| $48 < \bar{x} < 52$ | 10 | 0.0114 | 0.5000 | 0.9705 |
| $48.5 < \bar{x} < 51.5$ | 16 | 0.0164 | 0.2119 | 0.9445 |
| $48 < \bar{x} < 52$ | 16 | 0.0014 | 0.5000 | 0.9918 |

# 9-1 Hypothesis Testing

## Definition

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 $1 - \beta$, and power can be interpreted as *the probability of correctly rejecting a false null hypothesis*. We often compare statistical tests by comparing their power properties.
• For example, consider the propellant burning rate problem when we are testing $H_0 : \mu = 50$ centimeters per second against $H_1 : \mu$ not equal 50 centimeters per second . Suppose that the true value of the mean is $\mu = 52$. When $n = 10$, we found that $\beta = 0.2643$, so the power of this test is $1 - \beta = 1 - 0.2643 = 0.7357$ when $\mu = 52$.

# 9-1 Hypothesis Testing

## 9-1.3 One-Sided and Two-Sided Hypotheses

### Two-Sided Test:

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

### One-Sided Tests:

$$H_0: \mu = \mu_0 \qquad \text{or} \qquad H_0: \mu = \mu_0$$
$$H_1: \mu > \mu_0 \qquad \qquad H_1: \mu < \mu_0$$

# 9-1 Hypothesis Testing

## Example 9-1

Consider the propellant burning rate problem. Suppose that if the burning rate is less than 50 centimeters per second, we wish to show this with a strong conclusion. The hypotheses should be stated as

$$H_0: \mu = 50 \text{ centimeters per second}$$
$$H_1: \mu < 50 \text{ centimeters per second}$$

Here the critical region lies in the lower tail of the distribution of $\overline{X}$. Since the rejection of $H_0$ is always a strong conclusion, this statement of the hypotheses will produce the desired outcome if $H_0$ is rejected. Notice that, although the null hypothesis is stated with an equal sign, it is understood to include any value of $\mu$ not specified by the alternative hypothesis. Therefore, failing to reject $H_0$ does not mean that $\mu = 50$ centimeters per second exactly, but only that we do not have strong evidence in support of $H_1$.

# 9-1 Hypothesis Testing

The bottler wants to be sure that the bottles meet the specification on mean internal pressure or bursting strength, which for 10-ounce bottles is a minimum strength of 200 psi. The bottler has decided to formulate the decision procedure for a specific lot of bottles as a hypothesis testing problem. There are two possible formulations for this problem, either

$$H_0: \mu = 200 \text{ psi}$$
$$H_1: \mu > 200 \text{ psi}$$

or

$$H_0: \mu = 200 \text{ psi}$$
$$H_1: \mu < 200 \text{ psi}$$

# 9-1 Hypothesis Testing

## 9-1.4 P-Values in Hypothesis Tests

### Definition

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

# 9-1 Hypothesis Testing

## 9-1.4 P-Values in Hypothesis Tests

Consider the two-sided hypothesis test for burning rate

$$H_0 : \mu = 50 \quad H_1 : \mu \neq 50$$

with $n = 16$ and $\sigma = 2.5$. Suppose that the observed sample mean is $\bar{x} = 51.3$ 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 $\alpha$ associated with this critical region. Any smaller value for $\alpha$ 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. In this example

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

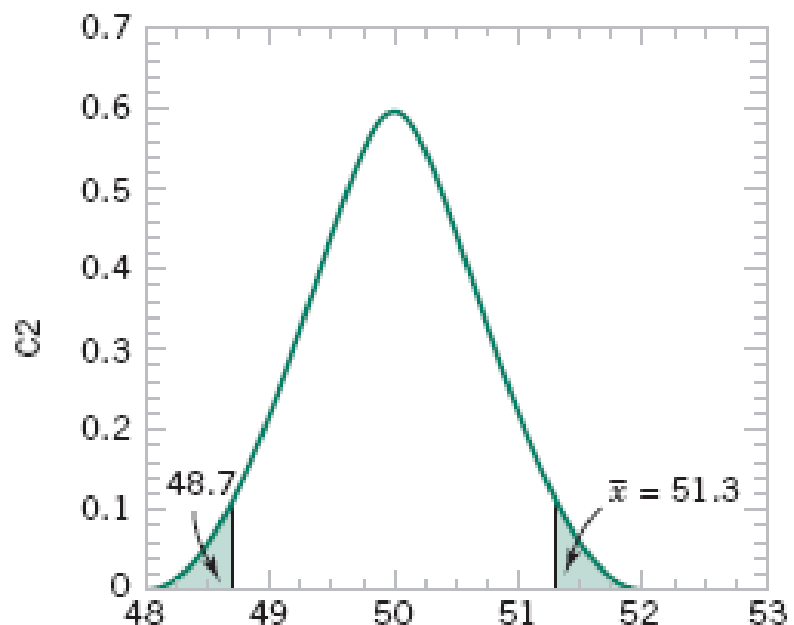# 9-1 Hypothesis Testing

## 9-1.4 P-Values in Hypothesis Tests



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

# 9-1 Hypothesis Testing

## 9-1.5 Connection between Hypothesis Tests and Confidence Intervals

There is a close relationship between the test of a hypothesis about any parameter, say $\theta$, and the confidence interval for $\theta$. If $[l, u]$ is a $100(1 - \alpha)\%$ confidence interval for the parameter $\theta$, the test of size $\alpha$ 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 $\theta_0$ is **not** in the $100(1 - \alpha)\%$ CI $[l, u]$. As an illustration, consider the escape system propellant problem with $\bar{x} = 51.3$, $\sigma = 2.5$, and $n = 16$. The null hypothesis $H_0: \mu = 50$ was rejected, using $\alpha = 0.05$. The 95% two-sided CI on $\mu$ can be calculated using Equation 8-7. This CI is $51.3 \pm 1.96(2.5/\sqrt{16})$ and this is $50.075 \leq \mu \leq 52.525$. Because the value $\mu_0 = 50$ is not included in this interval, the null hypothesis $H_0: \mu = 50$ is rejected.

# 9-1 Hypothesis Testing

## 9-1.6 General Procedure for Hypothesis Tests

1. From the problem context, identify the parameter of interest.

2. State the null hypothesis, $H_0$ .

**3.** Specify an appropriate alternative hypothesis, $H_1$.

4. Choose a significance level, $\alpha$.

5. Determine an appropriate tst statistic.

6. State the rejection region for the statistic.

7. Compute any necessary sample quantities, substitute these into the equation for the test statistic, and compute that value.

8. Decide whether or not $H_0$ should be rejected and report that in the problem context.

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

## 9-2.1 Hypothesis Tests on the Mean

We wish to test:

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

The **test statistic** is:

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \qquad (9\text{-}8)$$

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

## 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

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

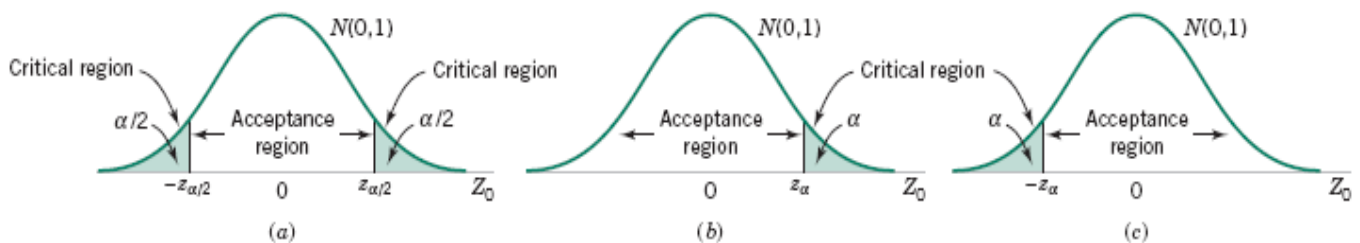# 9-2 Tests on the Mean of a Normal Distribution, Variance Known



**Figure 9-7** The distribution of $Z_0$ when $H_0: \mu = \mu_0$ is true, with critical region for (a) the two-sided alternative $H_1: \mu \neq \mu_0$, (b) the one-sided alternative $H_1: \mu > \mu_0$, and (c) the one-sided alternative $H_1: \mu < \mu_0$.

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

## Example 9-2

Aircrew 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. We know that the standard deviation of burning rate is $\sigma = 2$ centimeters per second. The experimenter decides to specify a type I error probability or significance level of $\alpha = 0.05$ and selects a random sample of $n = 25$ and obtains a sample average burning rate of $\bar{x} = 51.3$ centimeters per second. What conclusions should be drawn?

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

## Example 9-2

We may solve this problem by following the eight-step procedure outlined in Section 9-1.4. This results in

1. The parameter of interest is $\mu$, the mean burning rate.
2. $H_0$: $\mu = 50$ centimeters per second
3. $H_1$: $\mu \neq 50$ centimeters per second
4. $\alpha = 0.05$
5. The test statistic is

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

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

## Example 9-2

6. Reject $H_0$ if $z_0 > 1.96$ or if $z_0 < -1.96$. Note that this results from step 4, where we specified $\alpha = 0.05$, and so the boundaries of the critical region are at $z_{0.025} = 1.96$ and $-z_{0.025} = -1.96$.

7. Computations: Since $\bar{x} = 51.3$ and $\sigma = 2$,

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

8. Conclusion: Since $z_0 = 3.25 > 1.96$, we reject $H_0$: $\mu = 50$ at the 0.05 level of significance. Stated more completely, we conclude that the mean burning rate differs from 50 centimeters per second, based on a sample of 25 measurements. In fact, there is strong evidence that the mean burning rate exceeds 50 centimeters per second.

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

## 9-2.1 Hypothesis Tests on the Mean

We may also develop procedures for testing hypotheses on the mean $\mu$ where the alternative hypothesis is one-sided. Suppose that we specify the hypotheses as

$$H_0: \mu = \mu_0$$
$$H_1: \mu > \mu_0 \tag{9-11}$$

In defining the critical region for this test, we observe that a negative value of the test statistic $Z_0$ would never lead us to conclude that $H_0: \mu = \mu_0$ is false. Therefore, we would place the critical region in the **upper tail** of the standard normal distribution and reject $H_0$ if the computed value of $z_0$ is too large. That is, we would reject $H_0$ if

$$z_0 > z_\alpha \tag{9-12}$$

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

## 9-2.1 Hypothesis Tests on the Mean (Continued)

as shown in Figure 9-7(b). Similarly, to test

$$H_0: \mu = \mu_0$$
$$H_1: \mu < \mu_0 \tag{9-13}$$

we would calculate the test statistic $Z_0$ and reject $H_0$ if the value of $z_0$ is too small. That is, the critical region is in the **lower tail** of the standard normal distribution as shown in Figure 9-7(c), and we reject $H_0$ if

$$z_0 < -z_\alpha \tag{9-14}$$

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

## 9-2.1 Hypothesis Tests on the Mean (Continued)

Null hypothesis:     $H_0: \mu = \mu_0$

Test statistic:     $Z_0 = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$

| Alternative hypothesis | Rejection criteria |
|---|---|
| $H_1: \mu \neq \mu_0$ | $z_0 > z_{\alpha/2,n-1}$   or   $z_0 < -z_{\alpha/2,n-1}$ |
| $H_1: \mu > \mu_0$ | $z_0 > z_{\alpha,n-1}$ |
| $H_1: \mu < \mu_0$ | $z_0 < -z_{\alpha,n-1}$ |

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

## *P*-Values in Hypothesis Tests

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

$$P = \begin{cases} 2[1 - \Phi(|z_0|)] & \text{for a two-tailed test: } H_0: \mu = \mu_0 \quad H_1: \mu \neq \mu_0 \\ 1 - \Phi(z_0) & \text{for a upper-tailed test: } H_0: \mu = \mu_0 \quad H_1: \mu > \mu_0 \\ \Phi(z_0) & \text{for a lower-tailed test: } H_0: \mu = \mu_0 \quad H_1: \mu < \mu_0 \end{cases} \quad (9\text{-}15)$$

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

## 9-2.2 Type II Error and Choice of Sample Size

### Finding the Probability of Type II Error β

Consider the two-sided hypothesis

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

Suppose that the null hypothesis is false and that the true value of the mean is $\mu = \mu_0 + \delta$, say, where $\delta > 0$. The test statistic $Z_0$ is

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\overline{X} - (\mu_0 + \delta)}{\sigma/\sqrt{n}} + \frac{\delta\sqrt{n}}{\sigma}$$

Therefore, the distribution of $Z_0$ when $H_1$ is true is

$$Z_0 \sim N\left(\frac{\delta\sqrt{n}}{\sigma}, 1\right) \tag{9-16}$$

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

## 9-2.2 Type II Error and Choice of Sample Size

### Finding the Probability of Type II Error β

$$\beta = \Phi\left(z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) \tag{9-17}$$

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

## 9-2.2 Type II Error and Choice of Sample Size

**Finding the Probability of Type II Error β**



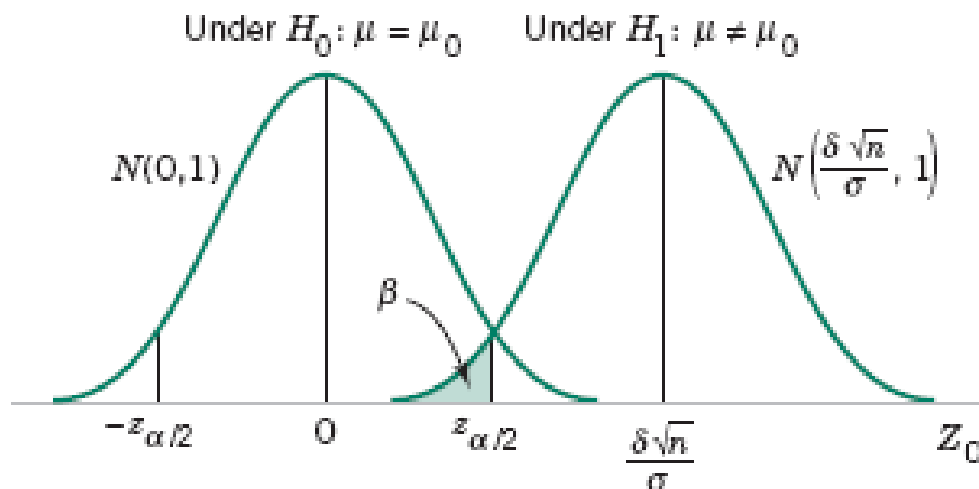**Figure 9-7** The distribution of $Z_0$ under $H_0$ and $H_1$

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

## 9-2.2 Type II Error and Choice of Sample Size

**Sample Size Formulas**

For a two-sided alternative hypothesis:

$$n \simeq \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{\delta^2} \qquad \text{where} \qquad \delta = \mu - \mu_0 \qquad (9\text{-}19)$$

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

## 9-2.2 Type II Error and Choice of Sample Size

### Sample Size Formulas

For a one-sided alternative hypothesis:

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{\delta^2} \quad \text{where} \quad \delta = \mu - \mu_0 \qquad (9\text{-}20)$$

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

### Example 9-3

Consider the rocket propellant problem of Example 9-2. 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 9-19 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 $\Phi(-z_{\alpha/2} - \delta\sqrt{n}/\sigma) = \Phi(-1.96 - (1)\sqrt{42}/2) = \Phi(-5.20) \simeq 0$, which is small relative to $\beta$.

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

## 9-2.2 Type II Error and Choice of Sample Size

### Using Operating Characteristic Curves

When performing sample size or type II error calculations, it is sometimes more convenient to use the **operating characteristic (OC) curves** in Appendix Charts VI*a* and VI*b*. These curves plot $\beta$ as calculated from Equation 9-17 against a parameter $d$ for various sample sizes $n$. Curves are provided for both $\alpha = 0.05$ and $\alpha = 0.01$. The parameter $d$ is defined as

$$d = \frac{|\mu - \mu_0|}{\sigma} = \frac{|\delta|}{\sigma} \tag{9-21}$$

so one set of operating characteristic curves can be used for all problems regardless of the values of $\mu_0$ and $\sigma$. From examining the operating characteristic curves or Equation 9-17 and Fig. 9-7, we note that

1. The further the true value of the mean $\mu$ is from $\mu_0$, the smaller the probability of type II error $\beta$ for a given $n$ and $\alpha$. That is, we see that for a specified sample size and $\alpha$, large differences in the mean are easier to detect than small ones.

2. For a given $\delta$ and $\alpha$, the probability of type II error $\beta$ decreases as $n$ increases. That is, to detect a specified difference $\delta$ in the mean, we may make the test more powerful by increasing the sample size.

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

## Example 9-4

Consider the propellant problem in Example 9-2. Suppose that the analyst is concerned about the probability of type II error if the true mean burning rate is $\mu = 51$ centimeters per second. We may use the operating characteristic curves to find $\beta$. Note that $\delta = 51 - 50 = 1$, $n = 25$, $\sigma = 2$, and $\alpha = 0.05$. Then using Equation 9-21 gives

$$d = \frac{|\mu - \mu_0|}{\sigma} = \frac{|\delta|}{\sigma} = \frac{1}{2}$$

and from Appendix Chart VII$a$, with $n = 25$, we find that $\beta = 0.30$. That is, if the true mean burning rate is $\mu = 51$ centimeters per second, there is approximately a 30% chance that this will not be detected by the test with $n = 25$.

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

## 9-2.3 Large Sample Test

We have developed the test procedure for the null hypothesis $H_0$: $\mu = \mu_0$ assuming that the population is normally distributed and that $\sigma^2$ is known. In many if not most practical situations $\sigma^2$ will be unknown. Furthermore, we may not be certain that the population is well modeled by a normal distribution. In these situations if $n$ is large (say $n > 40$) the sample standard deviation $s$ can be substituted for $\sigma$ in the test procedures with little effect. Thus, while we have given a test for the mean of a normal distribution with known $\sigma^2$, it can be easily converted into a **large-sample test procedure for unknown $\sigma^2$** that is valid regardless of the form of the distribution of the population. This large-sample test relies on the central limit theorem just as the large-sample confidence interval on $\mu$ that was presented in the previous chapter did. Exact treatment of the case where the population is normal, $\sigma^2$ is unknown, and $n$ is small involves use of the $t$ distribution and will be deferred until Section 9-3.

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

## 9-3.1 Hypothesis Tests on the Mean

**One-Sample $t$-Test**

Null hypothesis:    $H_0: \mu = \mu_0$

Test statistic:    $T_0 = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$

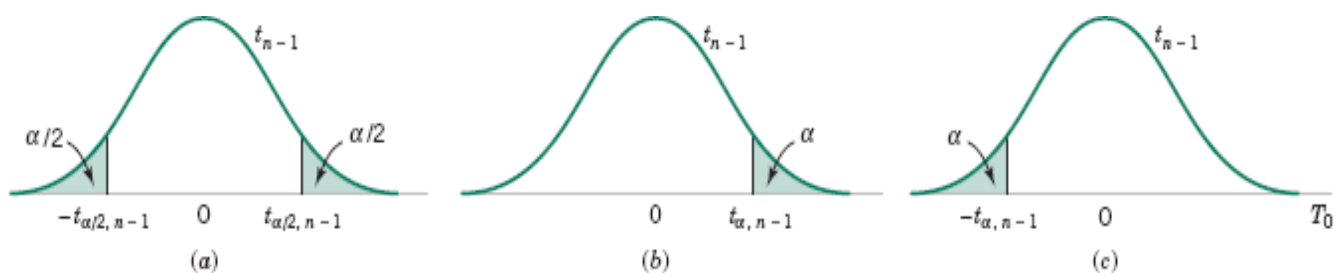| Alternative hypothesis | Rejection criteria |
|---|---|
| $H_1: \mu \neq \mu_0$ | $t_0 > t_{\alpha/2,n-1}$   or   $t_0 < -t_{\alpha/2,n-1}$ |
| $H_1: \mu > \mu_0$ | $t_0 > t_{\alpha,n-1}$ |
| $H_1: \mu < \mu_0$ | $t_0 < -t_{\alpha,n-1}$ |

**Figure 9-9** The reference distribution for $H_0: \mu = \mu_0$ with critical region for (a) $H_1: \mu \neq \mu_0$ , (b) $H_1: \mu > \mu_0$, and (c) $H_1: \mu < \mu_0$.

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

## Example 9-6

The increased availability of light materials with high strength has revolutionized the design and manufacture of golf clubs, particularly drivers. Clubs with hollow heads and very thin faces can result in much longer tee shots, especially for players of modest skills. This is due partly to the "spring-like effect" that the thin face imparts to the ball. Firing a golf ball at the head of the club and measuring the ratio of the outgoing velocity of the ball to the incoming velocity can quantify this spring-like effect. The ratio of velocities is called the coefficient of restitution of the club. An experiment was performed in which 15 drivers produced by a particular club maker were selected at random and their coefficients of restitution measured. In the experiment the golf balls were fired from an air cannon so that the incoming velocity and spin rate of the ball could be precisely controlled. 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 follow:

| | | | | |
|---|---|---|---|---|
| 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 |

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

## Example 9-6

The sample mean and sample standard deviation are $\bar{x} = 0.83725$ and $s = 0.02456$. The normal probability plot of the data in Fig. 9-9 supports the assumption that the coefficient of restitution is normally distributed. Since the objective of the experimenter is to demonstrate that the mean coefficient of restitution exceeds 0.82, a one-sided alternative hypothesis is appropriate.

The solution using the eight-step procedure for hypothesis testing is as follows:

1. The parameter of interest is the mean coefficient of restitution, $\mu$.

2. $H_0: \mu = 0.82$

3. $H_1: \mu > 0.82$. We want to reject $H_0$ if the mean coefficient of restitution exceeds 0.82.

4. $\alpha = 0.05$

5. The test statistic is

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$
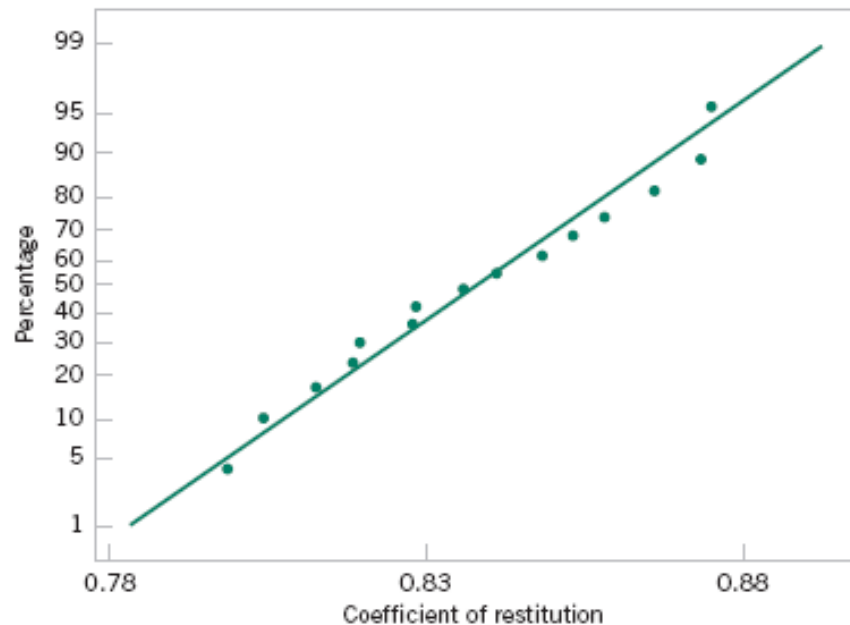
6. Reject $H_0$ if $t_0 > t_{0.05,14} = 1.761$

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

**Example 9-6**

**Figure 9-10**
Normal probability plot of the coefficient of restitution data from Example 9-6.

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

**Example 9-6**

7. Computations: Since $\bar{x} = 0.83725$, $s = 0.02456$, $\mu_0 = 0.82$, and $n = 15$, we have

$$t_0 = \frac{0.83725 - 0.82}{0.02456/\sqrt{15}} = 2.72$$

8. Conclusions: Since $t_0 = 2.72 > 1.761$, we reject $H_0$ and conclude at the 0.05 level of significance that the mean coefficient of restitution exceeds 0.82.

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

## 9-3.2 *P*-value for a *t*-Test

The *P*-value for a *t*-test is just the smallest level of significance at which the null hypothesis would be rejected.

To illustrate, consider the *t*-test based on 14 degrees of freedom in Example 9-6. The relevant critical values from Appendix Table IV are as follows:

| Critical Value: | 0.258 | 0.692 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 3.326 | 3.787 | 4.140 |
|---|---|---|---|---|---|---|---|---|---|---|
| Tail Area: | | 0.40 | 0.25 | 0.10 | 0.05 | 0.025 | 0.01 | 0.005 | 0.0025 | 0.001 | 0.0005 |

Notice that $t_0 = 2.72$ in Example 9-6, and that this is between two tabulated values, 2.624 and 2.977. Therefore, the *P*-value must be between 0.01 and 0.005. These are effectively the upper and lower bounds on the *P*-value.

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

## 9-3.3 Type II Error and Choice of Sample Size

The type II error of the two-sided alternative (for example) would be

$$\beta = P\{-t_{\alpha/2,n-1} \leq T_0 \leq t_{\alpha/2,n-1} | \delta \neq 0\}$$
$$= P\{-t_{\alpha/2,n-1} \leq T_0' \leq t_{\alpha/2,n-1}\}$$

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

## Example 9-7

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 $\sigma$. Then $d = |\delta|/\sigma = 0.02/0.02456 = 0.81$. By referring to the operating characteristic curves in Appendix Chart VIIg (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 $\sigma^2$ equals a specified value, say $\sigma_0^2$, or equivalently, that the standard deviation $\sigma$ is equal to $\sigma_0$. Let $X_1$, $X_2, \ldots, X_n$ be a random sample of $n$ observations from this population. To test

$$H_0: \sigma^2 = \sigma_0^2$$
$$H_1: \sigma^2 \neq \sigma_0^2 \qquad (9\text{-}26)$$

we will use the test statistic:

$$X_0^2 = \frac{(n-1)S^2}{\sigma_0^2} \qquad (9\text{-}27)$$

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

## 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 9-27 follows the chi-square distribution with $n-1$ degrees of freedom. This is the reference distribution for this test procedure. Therefore, we calculate $\chi_0^2$, the value of the test statistic $X_0^2$, and the null hypothesis $H_0: \sigma^2 = \sigma_0^2$ would be rejected if

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

where $\chi_{\alpha/2,n-1}^2$ and $\chi_{1-\alpha/2,n-1}^2$ are the upper and lower $100\alpha/2$ percentage points of the chi-square distribution with $n-1$ degrees of freedom, respectively. Figure 9-10(a) shows the critical region.

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

$$H_0: \sigma^2 = \sigma_0^2$$
$$H_1: \sigma^2 > \sigma_0^2 \tag{9-28}$$

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

$$H_0: \sigma^2 = \sigma_0^2$$
$$H_1: \sigma^2 < \sigma_0^2 \tag{9-29}$$

we would reject $H_0$ if $\chi_0^2 < \chi_{1-\alpha,n-1}^2$. The one-sided critical regions are shown in Figure 9-10(b) and (c).

# 9-4 Hypothesis Tests on the Variance and Standard Deviation of a Normal Distribution
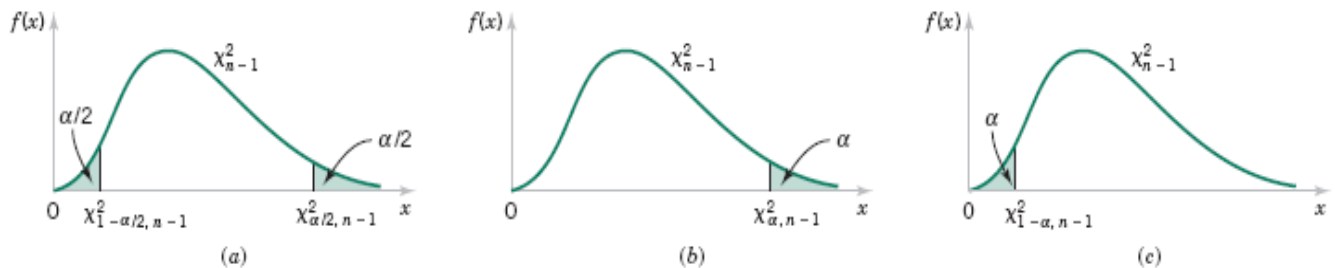
## 9-4.1 Hypothesis Test on the Variance



**Figure 9-11** 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) $H_1$: $\sigma^2 > \sigma_0^2$, and (c) $H_1$: $\sigma^2 < \sigma_0^2$.

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

## Example 9-8

An automatic 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 eight-step procedure results in the following:

1. The parameter of interest is the population variance $\sigma^2$.
2. $H_0$: $\sigma^2 = 0.01$
3. $H_1$: $\sigma^2 > 0.01$
4. $\alpha = 0.05$
5. The test statistic is

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

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

**Example 9-8**

6.  Reject $H_0$ if $\chi_0^2 > \chi_{0.05,19}^2 = 30.14$.
7.  Computations:

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

8.  Conclusions: Since $\chi_0^2 = 29.07 < \chi_{0.05,19}^2 = 30.14$, we conclude that there is no strong evidence that the variance of fill volume exceeds 0.01 (fluid ounces)$^2$.

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

## 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*:

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

## Example 9-9

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 $\beta \approx 0.6$. 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 $\beta$-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

$$Z_0 = \frac{X - np_0}{\sqrt{np_0(1 - p_0)}} \tag{9-32}$$

and reject $H_0: p = p_0$ if

# 9-5 Tests on a Population Proportion

## Example 9-10

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 eight-step hypothesis-testing procedure as follows:

1. The parameter of interest is the process fraction defective $p$.
2. $H_0$: $p = 0.05$
3. $H_1$: $p < 0.05$
   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.
4. $\alpha = 0.05$

# 9-5 Tests on a Population Proportion

## Example 9-10

5. The test statistic is (from Equation 9-32)

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

   where $x = 4$, $n = 200$, and $p_0 = 0.05$.
6. Reject $H_0$: $p = 0.05$ if $z_0 < -z_{0.05} = -1.645$
7. Computations: The test statistic is

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

8. Conclusions: Since $z_0 = -1.95 < -z_{0.05} = -1.645$, we reject $H_0$ and conclude that the process fraction defective $p$ is less than 0.05. The $P$-value for this value of the test statistic is $P = 0.0256$, which is less than $\alpha = 0.05$. 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

$$\beta = \Phi\left(\frac{p_0 - p + z_{\alpha/2}\sqrt{p_0(1 - p_0)/n}}{\sqrt{p(1 - p)/n}}\right) - \Phi\left(\frac{p_0 - p - z_{\alpha/2}\sqrt{p_0(1 - p_0)/n}}{\sqrt{p(1 - p)/n}}\right) \quad (9\text{-}34)$$

If the alternative is $p < p_0$

$$\beta = 1 - \Phi\left(\frac{p_0 - p - z_\alpha\sqrt{p_0(1 - p_0)/n}}{\sqrt{p(1 - p)/n}}\right) \quad (9\text{-}35)$$

If the alternative is $p > p_0$

$$\beta = \Phi\left(\frac{p_0 - p + z_\alpha\sqrt{p_0(1 - p_0)/n}}{\sqrt{p(1 - p)/n}}\right) \quad (9\text{-}36)$$

# 9-5 Tests on a Population Proportion

## 9-5.3 Type II Error and Choice of Sample Size

For a two-sided alternative

$$n = \left[ \frac{z_{\alpha/2}\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p(1-p)}}{p - p_0} \right]^2 \qquad (9\text{-}37)$$

For a one-sided alternative

$$n = \left[ \frac{z_\alpha\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p(1-p)}}{p - p_0} \right]^2 \qquad (9\text{-}38)$$

# 9-5 Tests on a Population Proportion

## Example 9-11

Consider the semiconductor manufacturer from Example 9-10. Suppose that its process fall-out is really $p = 0.03$. What is the $\beta$-error for a test of process capability that uses $n = 200$ and $\alpha = 0.05$?

The $\beta$-error can be computed using Equation 9-35 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 $\beta$-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 $\beta$-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 $\alpha = 0.05$, what sample size would be required?

The required sample size can be computed from Equation 9-38 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 9-38. Note that $n = 832$ is a very large sample size. However, we are trying to detect a fairly small deviation from the null value $p_0 = 0.05$.

## IMPORTANT TERMS AND CONCEPTS

$\alpha$ and $\beta$

Connection between hypothesis tests and confidence intervals

Critical region for a test statistic

Hypothesis test

Inference

Null distribution

Null hypothesis

One- and two-sided alternative hypotheses

Operating characteristic (OC) curves

Power of a test

P-value

Reference distribution for a test statistic

Sample size determination for hypothesis tests

Significance level of a test

Statistical hypotheses

Statistical versus practical significance

Test for goodness of fit

Test for homogeneity

Test for independence

Test statistic

Type I and type II errors