# Chapter 8
# Hypothesis Testing
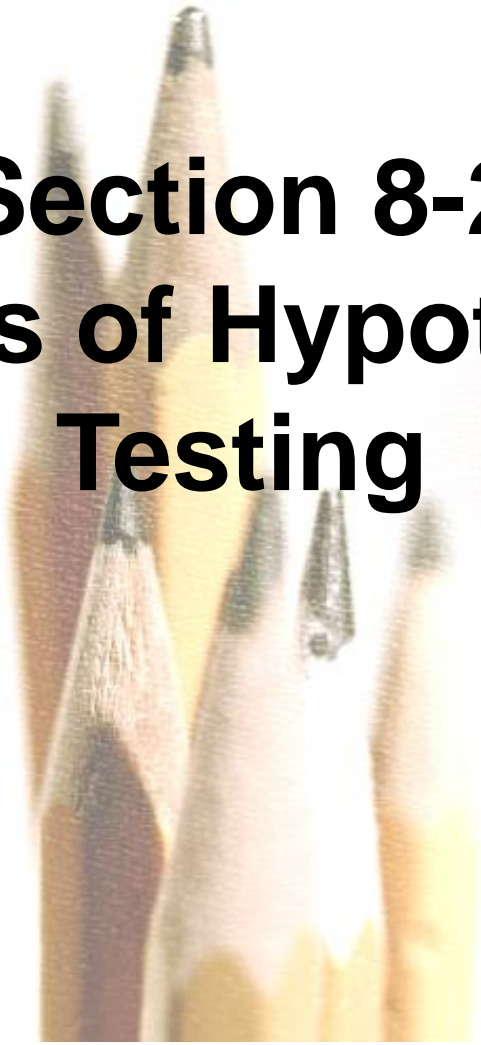
# Section 8-1
# Review and Preview

# Definitions

In statistics, a **hypothesis** is a claim or statement about a property of a population.

A **hypothesis test** (or **test of significance**) is a standard procedure for testing a claim about a property of a population.

# Section 8-2
# Basics of Hypothesis Testing

# Part 1:

# The Basics of Hypothesis Testing

# Components of a Formal Hypothesis Test

# Null Hypothesis: $H_0$

- **The null hypothesis (denoted by $H_0$) is a statement that the value of a population parameter (such as proportion, mean, or standard deviation) is equal to some claimed value.**

- **We test the null hypothesis directly.**

- **Either reject $H_0$ or fail to reject $H_0$.**

# Alternative Hypothesis: $H_1$

- **The alternative hypothesis (denoted by $H_1$ or $H_a$ or $H_A$) is the statement that the parameter has a value that somehow differs from the null hypothesis.**

- **The symbolic form of the alternative hypothesis must use one of these symbols: ≠, <, >.**

# Example:

**Consider the claim that the mean weight of airline passengers (including carry-on baggage) is at most 195 lb (the current value used by the Federal Aviation Administration). Follow the three-step procedure outlined in Figure 8-2 to identify the null hypothesis and the alternative hypothesis.**

# Example:

**Step 1:** **Express the given claim in symbolic form. The claim that the mean is at most 195 lb is expressed in symbolic form as $\mu \leq 195$ lb.**

**Step 2:** **If $\mu \leq 195$ lb is false, then $\mu > 195$ lb must be true.**

# Example:

**Step 3:** **Of the two symbolic expressions**
$\mu \le 195$ **lb and** $\mu > 195$ **lb, we see that**
$\mu > 195$ **lb does not contain equality, so we let the alternative hypothesis** $H_1$ **be** $\mu > 195$ **lb. Also, the null hypothesis must be a statement that the mean equals 195 lb, so we let** $H_0$ **be** $\mu = 195$ **lb.**

**Note that the original claim that the mean is at most 195 lb is neither the alternative hypothesis nor the null hypothesis. (However, we would be able to address the original claim upon completion of a hypothesis test.)**

# Test Statistic

The **test statistic** is a value used in making a decision about the null hypothesis, and is found by converting the sample statistic to a score with the assumption that the null hypothesis is true.

# Test Statistic - Formulas

**Test statistic for proportion**

$$z = \frac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}}$$

**Test statistic for mean**

$$z = \frac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}} \quad \text{or} \quad t = \frac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n}}}$$

**Test statistic for standard deviation**

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

# Example:

Let's again consider the claim that the XSORT method of gender selection increases the likelihood of having a baby girl. Preliminary results from a test of the XSORT method of gender selection involved 14 couples who gave birth to 13 girls and 1 boy. Use the given claim and the preliminary results to calculate the value of the test statistic. Use the format of the test statistic given above, so that a normal distribution is used to approximate a binomial distribution. (There are other exact methods that do not use the normal approximation.)

# Example:

**The claim that the XSORT method of gender selection increases the likelihood of having a baby girl results in the following null and alternative hypotheses $H_0$: $p = 0.5$ and $H_1$: $p > 0.5$. We work under the assumption that the null hypothesis is true with $p = 0.5$. The sample proportion of 13 girls in 14 births results in $\hat{p} = 13/14 = 0.929$. Using $p = 0.5$, $\hat{p} = 0.929$ and $n = 14$, we find the value of the test statistic as follows:**

# Example:

$$z = \frac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}} = \frac{0.929 - 0.5}{\sqrt{\dfrac{(0.5)(0.5)}{14}}} = 3.21$$

**We know from previous chapters that a z score of 3.21 is "unusual" (because it is greater than 2). It appears that in addition to being greater than 0.5, the sample proportion of 13/14 or 0.929 is significantly greater than 0.5. The figure on the next slide shows that the sample proportion of 0.929 does fall within the range of values considered to be significant because**

# Example:

**they are so far above 0.5 that they are not likely to occur by chance (assuming that the population proportion is $p = 0.5$).**



Unusually high
sample proportions

Critical region:
Area of $\alpha = 0.05$ used
as criterion for
identifying unusually
high sample proportions

$p = 0.5$
or
$z = 0$

$z = 1.645$

Critical
value

Sample proportion of: $\hat{p} = 0.929$
or
Test Statistic $z = 3.21$
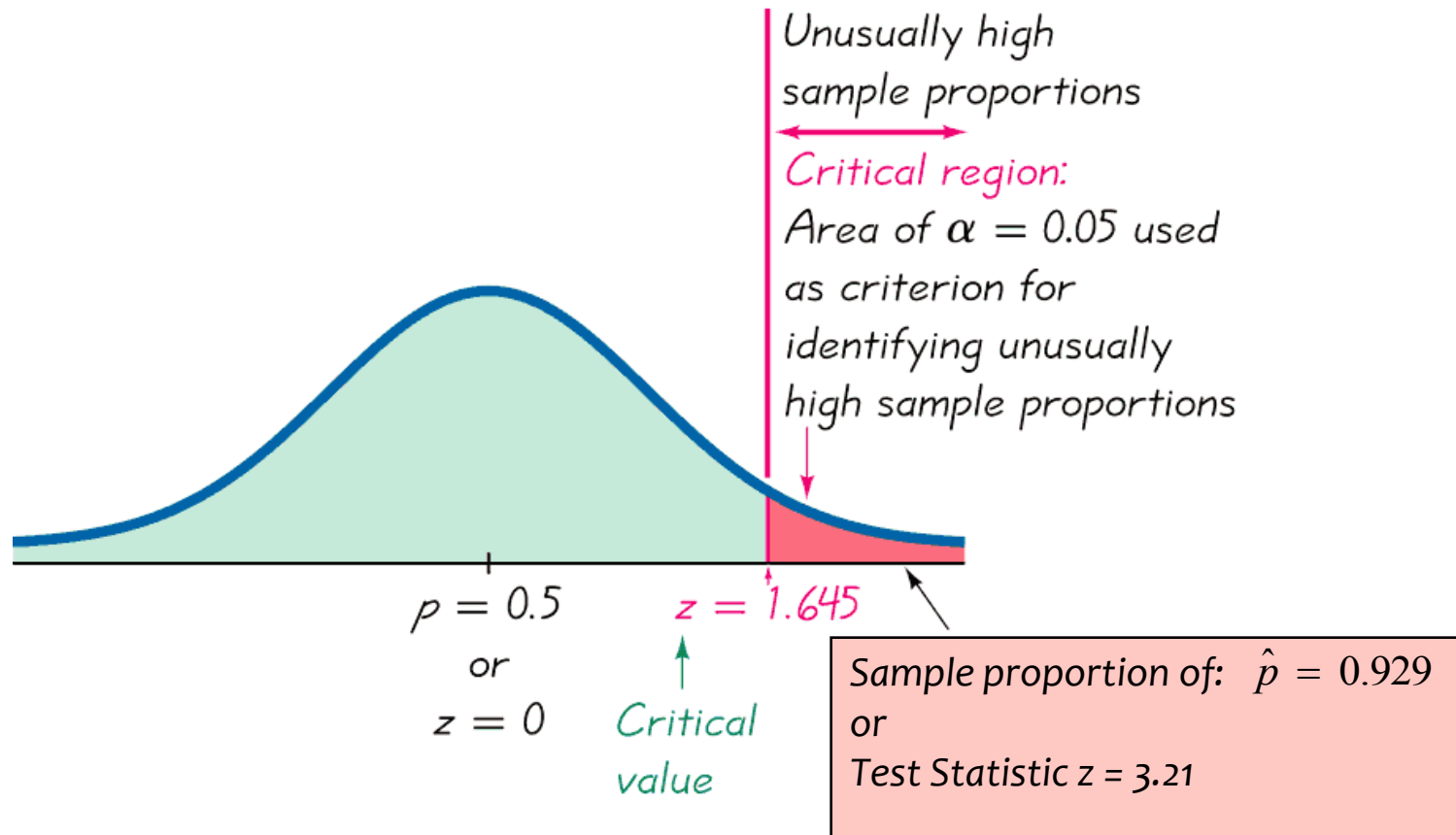
# Critical Region

**The critical region (or rejection region) is the set of all values of the test statistic that cause us to reject the null hypothesis. For example, see the red-shaded region in the previous figure.**

# Significance Level

The **significance level** (denoted by $\alpha$) is the probability that the test statistic will fall in the critical region when the null hypothesis is actually true. This is the same $\alpha$ introduced in Section 7-2. Common choices for $\alpha$ are 0.05, 0.01, and 0.10.

# Critical Value

A **critical value** is any value that separates the critical region (where we reject the null hypothesis) from the values of the test statistic that do not lead to rejection of the null hypothesis. The critical values depend on the nature of the null hypothesis, the sampling distribution that applies, and the significance level $\alpha$. See the previous figure where the critical value of $z = 1.645$ corresponds to a significance level of $\alpha = 0.05$.

# *P*-Value

The *P*-value (or *p*-value or probability value) is the probability of getting a value of the test statistic that is at least as extreme as the one representing the sample data, assuming that the null hypothesis is true.

Critical region in the left tail:

*P*-value = area to the left of the test statistic

Critical region in the right tail:

*P*-value = area to the right of the test statistic

Critical region in two tails:

*P*-value = twice the area in the tail beyond the test statistic

# *P*-Value

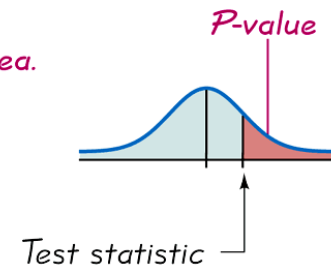The null hypothesis is rejected if the *P*-value is very small, such as 0.05 or less.

Here is a memory tool useful for interpreting the P-value:

If the P is low, the null must go.
If the P is high, the null will fly.

# Procedure for Finding *P*-Values

Figure 8-5
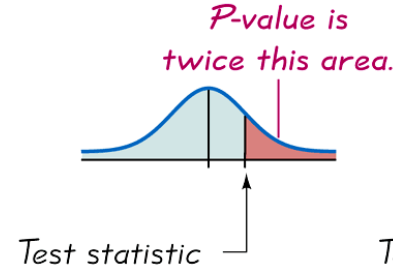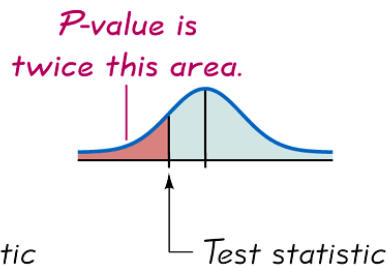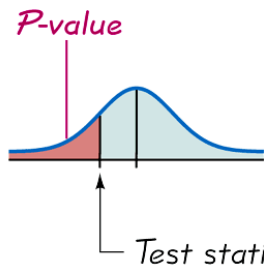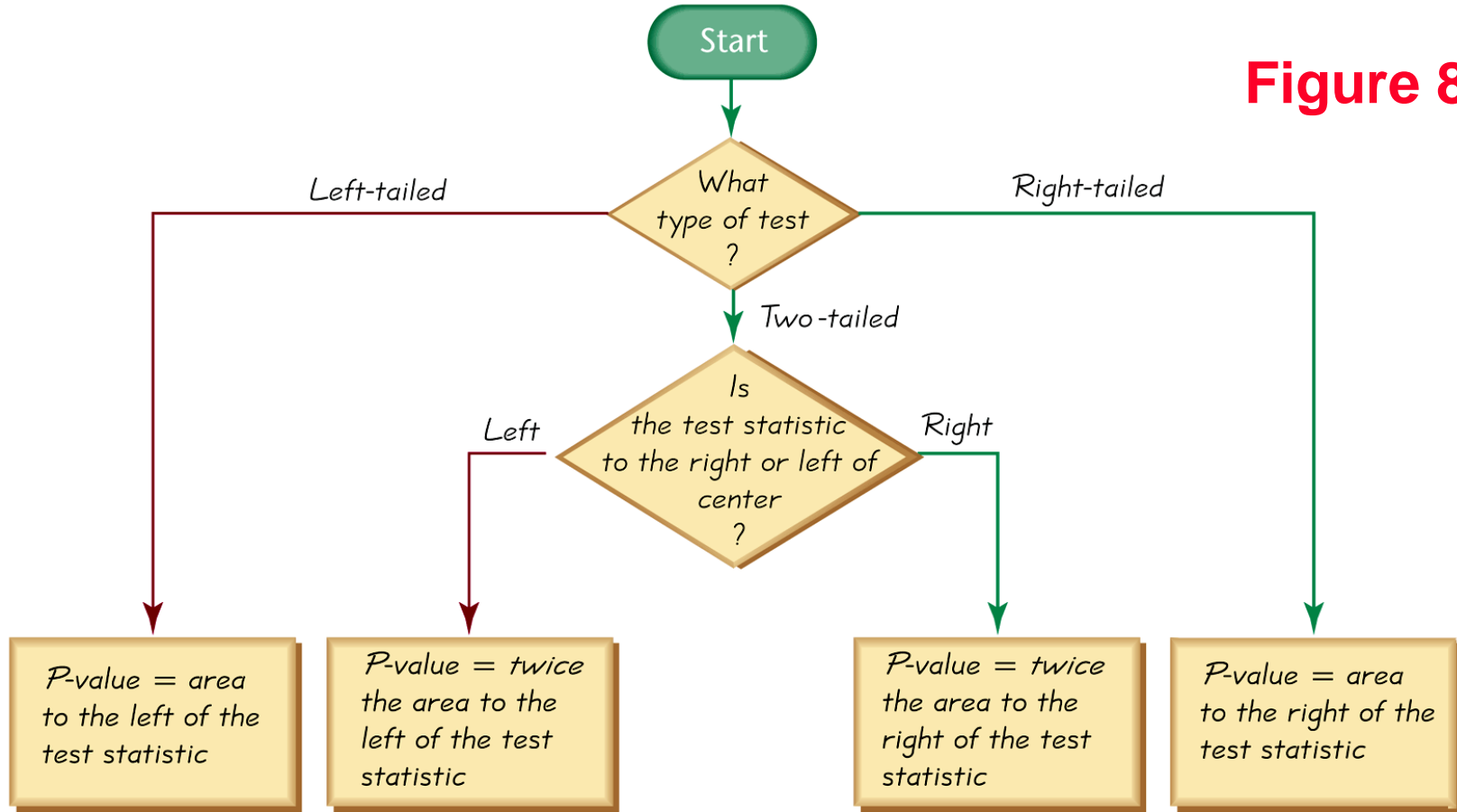
# Types of Hypothesis Tests: Two-tailed, Left-tailed, Right-tailed

The **tails** in a distribution are the extreme regions bounded by critical values.

Determinations of *P*-values and critical values are affected by whether a critical region is in two tails, the left tail, or the right tail. It therefore becomes important to correctly characterize a hypothesis test as two-tailed, left-tailed, or right-tailed.

# Two-tailed Test

$H_0$:  =

$\alpha$ is divided equally between the two tails of the critical region

$H_1$:  $\neq$

Means less than or greater than



Reject $H_0$

Fail to reject $H_0$

Reject $H_0$

Sign used in $H_1$: $\neq$
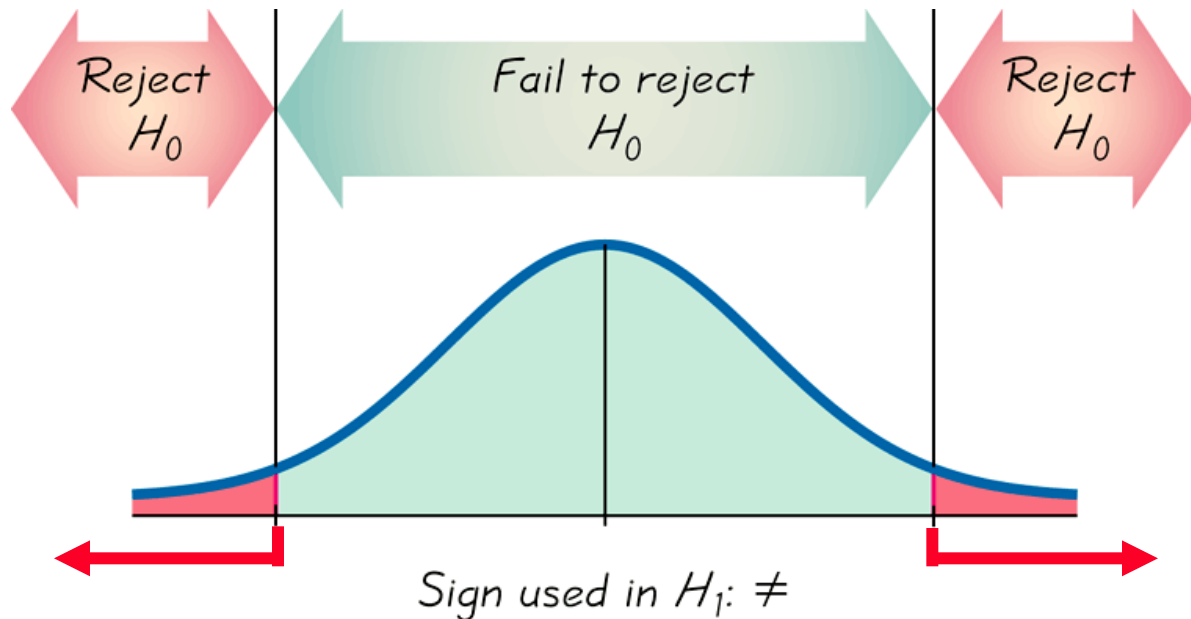
# Left-tailed Test

$H_0$: =

$H_1$: <

$\alpha$ the left tail

**Points Left**
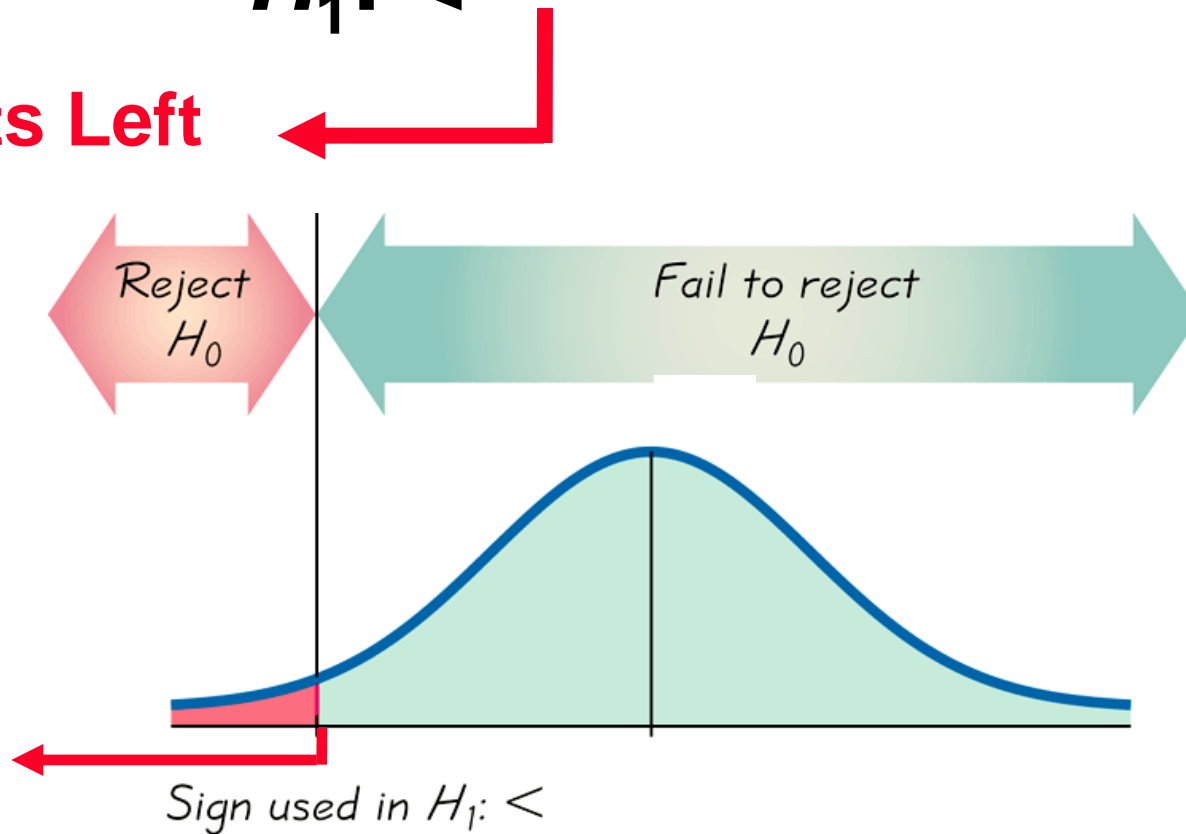
# Right-tailed Test

$H_0$:  =

$H_1$:  >

**Points Right**



Fail to reject $H_0$

Reject $H_0$

Sign used in $H_1$: >
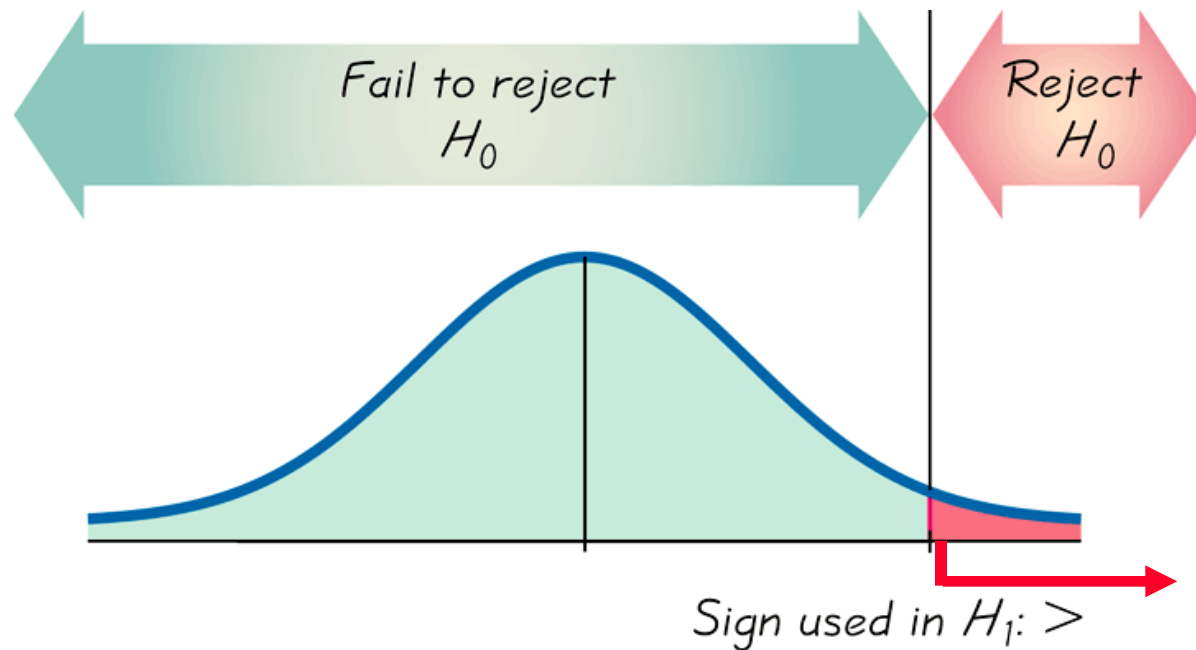
# Conclusions
# in Hypothesis Testing

**We always test the null hypothesis. The initial conclusion will always be one of the following:**

**1. Reject the null hypothesis.**

**2. Fail to reject the null hypothesis.**

# Decision Criterion

**P-value method**:
**Using the significance level $\alpha$:**

**If *P*-value $\leq \alpha$ ,   reject *$H_0$*.**

**If *P*-value $> \alpha$ ,   fail to reject *$H_0$*.**

# Decision Criterion

**Traditional method:**

**If the test statistic falls within the critical region, <span style="color:red">reject $H_0$</span>.**

**If the test statistic does not fall within the critical region, <span style="color:red">fail to reject $H_0$</span>.**

# Decision Criterion

**Another option:**

**Instead of using a significance level such as 0.05, simply identify the *P*-value and leave the decision to the reader.**

# Wording of Final Conclusion
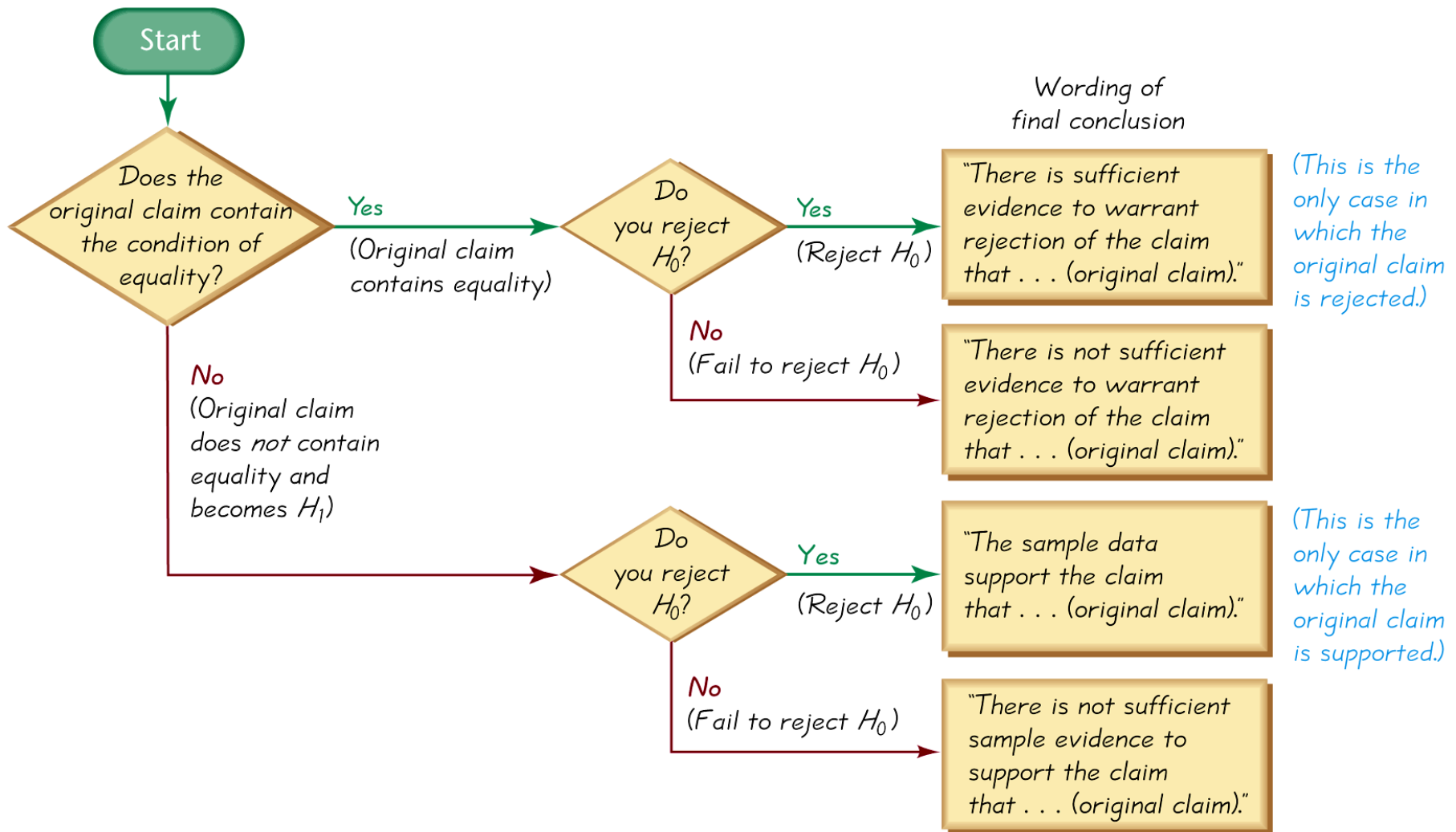


Figure 8-7

# Caution

**Never conclude a hypothesis test with a statement of "reject the null hypothesis" or "fail to reject the null hypothesis." Always make sense of the conclusion with a statement that uses simple nontechnical wording that addresses the original claim.**

# Accept Versus Fail to Reject

- **Some texts use "accept the null hypothesis."**

- **We are not proving the null hypothesis.**

- **Fail to reject says more correctly**

- **The available evidence is not strong enough to warrant rejection of the null hypothesis (such as not enough evidence to convict a suspect).**

# Comprehensive Hypothesis Test – *P*-Value Method

**Start**

1. Identify the specific claim or hypothesis to be tested, and put it in symbolic form.

2. Give the symbolic form that must be true when the original claim is false.

3. Of the two symbolic expressions obtained so far, let the alternative hypothesis $H_1$ be the one not containing equality, so that $H_1$ uses the symbol $>$ or $<$ or $\neq$. Let the null hypothesis $H_0$ be the symbolic expression that the parameter equals the fixed value being considered.

4. Select the significance level $\alpha$ based on the seriousness of a type 1 error. Make $\alpha$ small if the consequences of rejecting a true $H_0$ are severe. The values of 0.05 and 0.01 are very common.

5. Identify the statistic that is relevant to this test and determine its sampling distribution (such as normal, t, chi-square).

6. Find the test statistic and find the P-value (see Figure 8-6). Draw a graph and show the test statistic and P-value.

7. Reject $H_0$ if the P-value is less than or equal to the significance level $\alpha$. Fail to reject $H_0$ if the P-value is greater than $\alpha$.

8. Restate this previous decision in simple, nontechnical terms, and address the original claim.

**Stop**

# Comprehensive Hypothesis Test – Traditional Method

**Start**

1. Identify the specific claim or hypothesis to be tested, and put it in symbolic form.

2. Give the symbolic form that must be true when the original claim is false.

3. Of the two symbolic expressions obtained so far, let the alternative hypothesis $H_1$ be the one not containing equality, so that $H_1$ uses the symbol $>$ or $<$ or $\neq$. Let the null hypothesis $H_0$ be the symbolic expression that the parameter equals the fixed value being considered.

4. Select the significance level $\alpha$ based on the seriousness of a type 1 error. Make $\alpha$ small if the consequences of rejecting a true $H_0$ are severe. The values of 0.05 and 0.01 are very common.

5. Identify the statistic that is relevant to this test and determine its sampling distribution (such as normal, t, chi-square).

6. Find the test statistic, the critical values, and the critical region. Draw a graph and include the test statistic, critical value(s), and critical region.

7. Reject $H_0$ if the test statistic is in the critical region. Fail to reject $H_0$ if the test statistic is not in the critical region.

8. Restate this previous decision in simple, nontechnical terms, and address the original claim.

**Stop**

# Recap

In this section we have discussed:

❖ **Null and alternative hypotheses.**

❖ **Test statistics.**

❖ **Significance levels.**

❖ *P*-**values.**

❖ **Decision criteria.**

# Section 8-3
# Testing a Claim About a Proportion

# Notation

$n$ = number of trials

$\hat{p} = \dfrac{x}{n}$ (**sample** proportion)

$p$ = population proportion (used in the null hypothesis)

$q = 1 - p$

# Test Statistic for Testing a Claim About a Proportion

$$Z = \frac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}}$$

**P-values:** Use the standard normal distribution (Table A-2) and refer to Figure 8-5

**Critical Values:** Use the standard normal distribution (Table A-2).

# Example:

**The text refers to a study in which 57 out of 104 pregnant women correctly guessed the sex of their babies. Use these sample data to test the claim that the success rate of such guesses is no different from the 50% success rate expected with random chance guesses. Use a 0.05 significance level.**

# Example:

**Requirements are satisfied: simple random sample; fixed number of trials (104) with two categories (guess correctly or do not); $np = (104)(0.5) = 52 \geq 5$ and $nq = (104)(0.5) = 52 \geq 5$**

**Step 1:** original claim is that the success rate is no different from 50%: $p = 0.50$

**Step 2:** opposite of original claim is $p \neq 0.50$

**Step 3:** $p \neq 0.50$ does not contain equality so it is $H_1$.

$H_0$: $p = 0.50$ null hypothesis and original claim

$H_1$: $p \neq 0.50$ alternative hypothesis

# Example:

**Step 4:** significance level is $\alpha = 0.50$

**Step 5:** sample involves proportion so the relevant statistic is the sample proportion, $\hat{p}$

**Step 6:** calculate *z*:

$$z = \frac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}} = \frac{\dfrac{57}{104} - 0.50}{\sqrt{\dfrac{(0.50)(0.50)}{104}}} = 0.98$$

**two-tailed test, *P*-value is twice the area to the right of test statistic**

# Example:

**Table A-2: $z = 0.98$ has an area of 0.8365 to its left, so area to the right is 1 − 0.8365 = 0.1635, doubles yields 0.3270 (technology provides a more accurate *P*-value of 0.3268**

**Step 7:** the *P*-value of 0.3270 is greater than the significance level of 0.50, so fail to reject the null hypothesis

**Here is the correct conclusion: There is not sufficient evidence to warrant rejection of the claim that women who guess the sex of their babies have a success rate equal to 50%.**