Introductory Statistics

Chapter 9

Hypothesis Tests for One Population Mean

Tests of Significance

- The scheme of reasoning
- Stating hypotheses
- Test statistics
- P-values
- Statistical significance
- Test for population mean
- Two-sided test and confidence intervals

Statistical Inference

- The field of statistical inference consists of those methods used to make decisions or draw conclusions about a **population**.

- These methods utilize the information contained in a **sample** from the population in drawing conclusions.

Tests of Significance-Hypothesis Testing

This common type of inference is used to assess the evidence provided by the data in favor of or against some claim (hypothesis) about the population…

…rather than to estimate unknown population parameter, for which we would use confidence intervals.

Point Estimation

A **point estimate** of some population parameter θ is a single numerical value $\hat{\theta}$ of a statistic $\hat{\Theta}$.

Unknown Parameter θ	Statistic $\hat{\Theta}$	Point Estimate $\hat{\theta}$
μ	$\overline{X} = \frac{\Sigma X_i}{n}$	$\overline{x}$
σ^2	$S^2 = \frac{\Sigma(X_i - \overline{X})^2}{n-1}$	s^2
p	$\hat{P} = \frac{X}{n}$	$\hat{p}$
$\mu_1 - \mu_2$	$\overline{X}_1 - \overline{X}_2 = \frac{\Sigma X_{1i}}{n_1} - \frac{\Sigma X_{2i}}{n_2}$	$\overline{x}_1 - \overline{x}_2$
$p_1 - p_2$	$\hat{P}_1 - \hat{P}_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2}$	$\hat{p}_1 - \hat{p}_2$

Section 9.1

The Nature of Hypothesis Testing

Stating Hypotheses

- The hypothesis is a statement about the **parameters in a population** or model. Not about the data at hand.
- The results of a test are expressed in terms of a **probability** that measures how well the **data and the hypothesis agree**.
- In hypothesis testing, we need to state two hypotheses:
  - The **null** hypothesis H_0
  - The **alternative** hypothesis H_a

Statistical Hypotheses

- Two Parts
  - a null hypothesis
  - an alternative hypothesis
- Null Hypothesis – nothing new is happening
- Alternative Hypothesis – something new is happening
- Notation
  - null: H_0
  - alternative: H_a or H_1.

Definition 9.1

Null and Alternative Hypotheses; Hypothesis Test

Null hypothesis: A hypothesis to be tested. We use the symbol H_0 to represent the null hypothesis.

Alternative hypothesis: A hypothesis to be considered as an alternative to the null hypothesis. We use the symbol H_a to represent the alternative hypothesis.

Hypothesis test: The problem in a hypothesis test is to decide whether the null hypothesis should be rejected in favor of the alternative hypothesis.

What is a Hypothesis?

A Hypothesis is a statement about the value of a population parameter developed for the purpose of testing.

Examples of hypotheses made about a population parameter are:

- The mean monthly income for systems analysts is $3,625.
- Twenty percent of all customers at Bovine's Chop House return for another meal within a month.

Null and Alternative Hypotheses

- The Null and Alternative Hypotheses are mutually exclusive. Only one of them can be true.
- The Null and Alternative Hypotheses are collectively exhaustive. They are stated to include all possibilities. (An abbreviated form of the null hypothesis is often used.)
- The Null Hypothesis is assumed to be true.
- The burden of proof falls on the Alternative Hypothesis.

Null hypothesis:

- The null hypothesis is the claim which is initially favored or believed to be true. Often **default** or uninteresting **situation** of “no effect” or “no difference”.

We usually need to determine if there is a strong enough evidence **against it**.

- The test of significance is designed to assess this strength of the evidence against the null hypothesis.

Alternative hypothesis:

- The alternative hypothesis is the claim that we “hope” or “suspect” something else is true instead of H_0.
- We often begin with the alternative hypothesis H_a and then set up H_0 as the statement that the hoped-for effect is not present.

Example

- ✓ **Example:** Suppose you are investigating the effects of a new pain reliever. You hope the new drug relieves minor muscle aches and pains longer than the leading pain reliever. State the null and alternative hypotheses.

Solutions:

- H_o: The new pain reliever is no better than the leading pain reliever
- H_a: The new pain reliever lasts longer than the leading pain reliever

Example

- ✓ **Example:** You are investigating the presence of radon in homes being built in a new development. If the mean level of radon is greater than 4 then send a warning to all home owners in the development. State the null and alternative hypotheses.

Solutions:

- H_o: The mean level of radon for homes in the development is 4 (or less)

- H_a: The mean level of radon for homes in the development is greater than 4

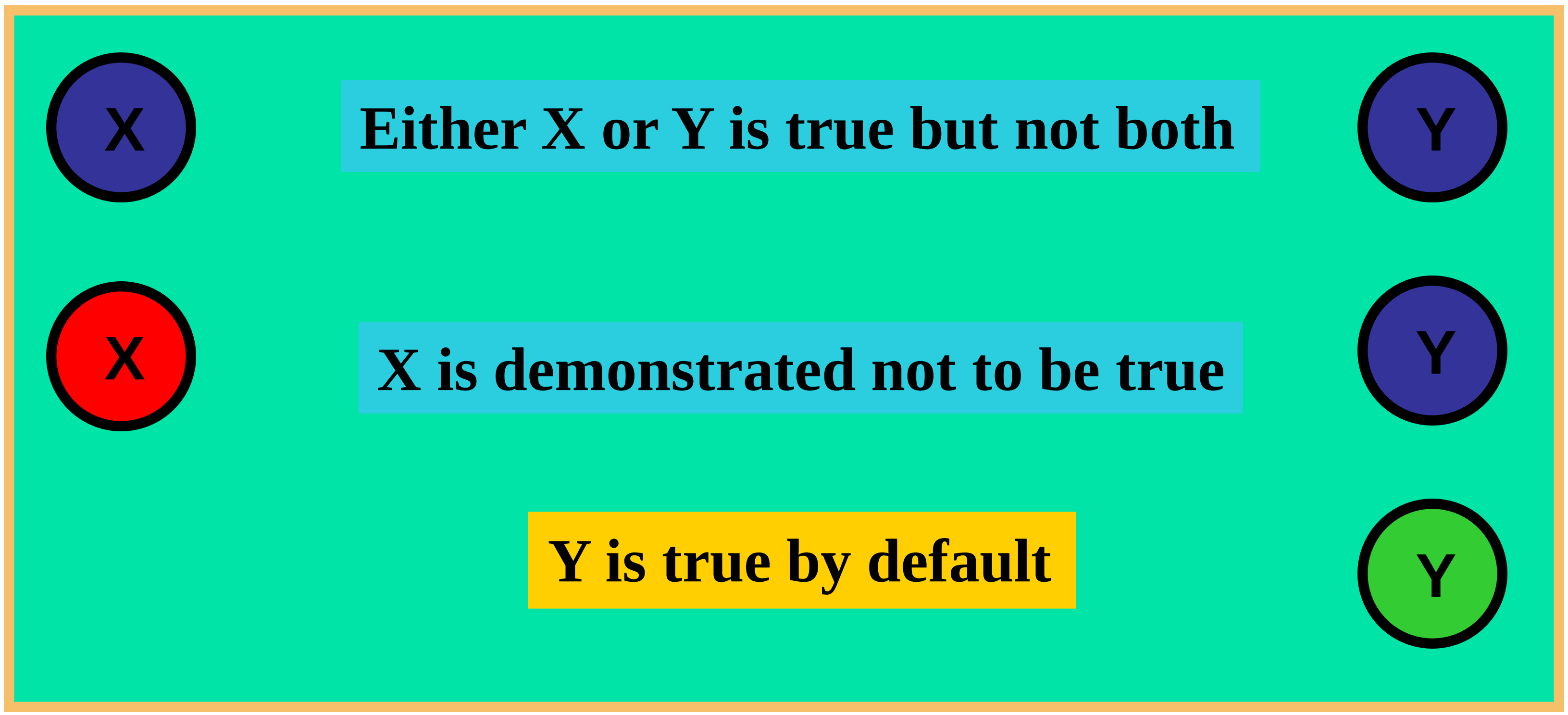
What is Hypothesis Testing?

Hypothesis testing is a procedure, based on sample evidence and probability theory, used to determine whether the hypothesis is a reasonable statement and should not be rejected, or is unreasonable and should be rejected:

- 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**.

Note: it does not tell us if a statement or the hypothesis is correct or not.

Method of Indirect Proof



Definition 9.3

Significance Level

The probability of making a Type I error, that is, of rejecting a true null hypothesis, is called the **significance level,** α **,** of a hypothesis test.

Procedure to Test Hypotheses

Task 1:
HYPOTHESIZE

Task 2:
TEST

Task 3:
TAKE STATISTICAL ACTION

Task 4:
DETERMINING THE BUSINESS IMPLICATIONS

Hypothesis Test Outcomes

Decision	Null Hypothesis: True	Null Hypothesis: False
Fail to reject H_o	Type A correct decision	Type II error
Reject H_o	Type I error	Type B correct decision

Type A correct decision: Null hypothesis true, decide in its favor

Type B correct decision: Null hypothesis false, decide in favor of alternative hypothesis

Type I error: Null hypothesis true, decide in favor of alternative hypothesis

Type II error: Null hypothesis false, decide in favor of null hypothesis

Definition 9.3

Type I and Type II Errors

Type I error: Rejecting the null hypothesis when it is in fact true.

Type II error: Not rejecting the null hypothesis when it is in fact false.

Example

Example: A calculator company has just received a large shipment of parts used to make the screens on graphing calculators. They consider the shipment acceptable if the proportion of defective parts is 0.01 (or less). If the proportion of defective parts is greater than 0.01 the shipment is unacceptable and returned to the manufacturer. State the null and alternative hypotheses, and describe the four possible outcomes and the resulting actions that would occur for this test.

Solutions:

- H_o: The proportion of defective parts is 0.01 (or less)
- H_a: The proportion of defective parts is greater than 0.01

Fail To Reject H_o

Null Hypothesis Is True:

Type A correct decision

Truth of situation: The proportion of defective parts is 0.01 (or less)

Conclusion: It was determined that the proportion of defective parts is 0.01 (or less)

Action: The calculator company received parts with an acceptable proportion of defectives

Null Hypothesis Is False:

Type II error

Truth of situation: The proportion of defective parts is greater than 0.01

Conclusion: It was determined that the proportion of defective parts is 0.01 (or less)

Action: The calculator company received parts with an unacceptable proportion of defectives

Reject H_o

Null hypothesis is true:

Type I error

Truth of situation: The proportion of defectives is 0.01 (or less)

Conclusion: It was determined that the proportion of defectives is greater than 0.01

Action: Send the shipment back to the manufacturer. The proportion of defectives is acceptable

Null hypothesis is false:

Type B correct decision

Truth of situation: The proportion of defectives is greater than 0.01

Conclusion: It was determined that the proportion of defectives is greater than 0.01

Action: Send the shipment back to the manufacturer. The proportion of defectives is unacceptable

Errors

Notes:

1. The type II error sometimes results in what represents a *lost opportunity*
2. Since we make a decision based on a *sample*, there is always the chance of making an error

Probability of a type I error = α

Probability of a type II error = β

Error in Decision	**Type**	**Probability**
Rejection of a true H_o	I	α
Failure to reject a false H_o	II	β

Correct Decision	**Type**	**Probability**
Failure to reject a true H_o	A	$1 - \alpha$
Rejection of a false H_o	B	$1 - \beta$

Notes

1. Would like α and β to be as small as possible

2. α and β are inversely related

3. Usually set α (and don't worry too much about β. Why?)

4. Most common values for α and β are 0.01 and 0.05

5. $1 - \beta$: the power of the statistical test

   A measure of the ability of a hypothesis test to reject a false null hypothesis

6. Regardless of the outcome of a hypothesis test, we never really know for sure if we have made the correct decision

Critical Region & Critical Value(s)

Critical Region: The set of values for the test statistic that will cause us to reject the null hypothesis. The set of values that are not in the critical region is called the noncritical region (sometimes called the acceptance region).

Critical Value(s): The first or boundary value(s) of the critical region(s)

Example

Example: A company that produces snack foods uses a machine to package 454 g. bags of pretzels. Assume that the net weights are normally distributed and the population standard deviation is 7.8 g. A simple random sample of 25 bags of pretzels has the net weights in grams, display in the table below. At the 0.0456 significance level, do the data provide sufficient evidence to conclude that the packaging machine is not working properly?

465	456	438	454	447
449	442	449	446	447
468	433	454	463	450
446	447	456	452	444
447	456	456	435	450

Level of Significance & Test Statistic

Level of Significance, α : The probability of committing the type I error

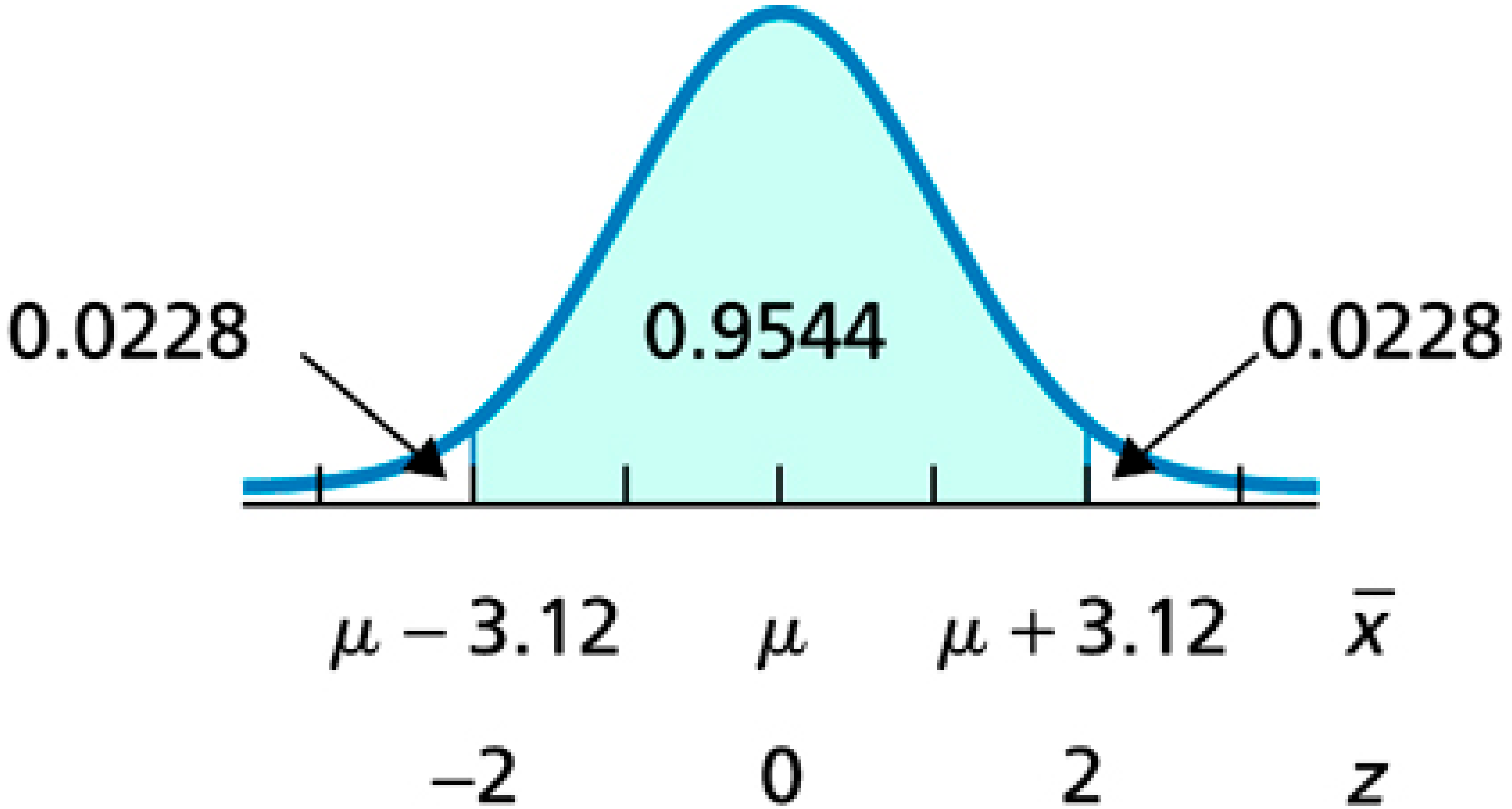
Test Statistic: A random variable whose value is calculated from the sample data and is used in making the decision *fail to reject H_o* or *reject H_o*

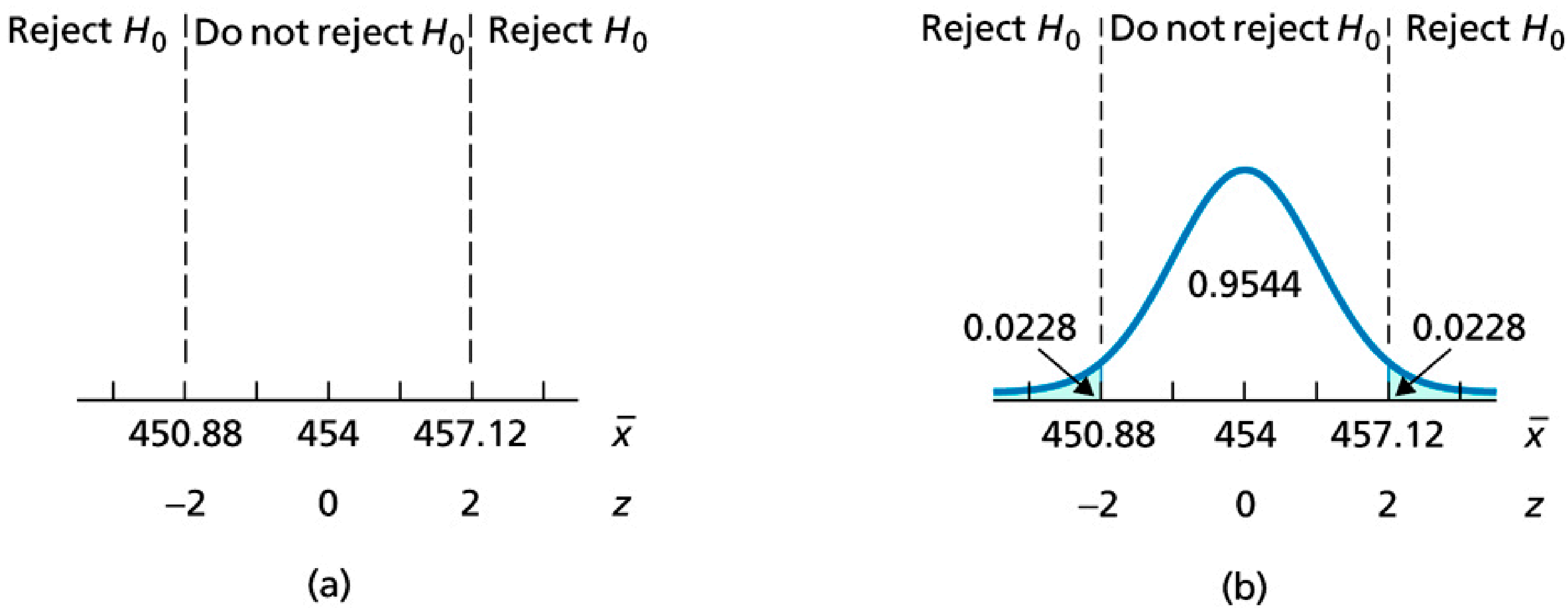
Notes:

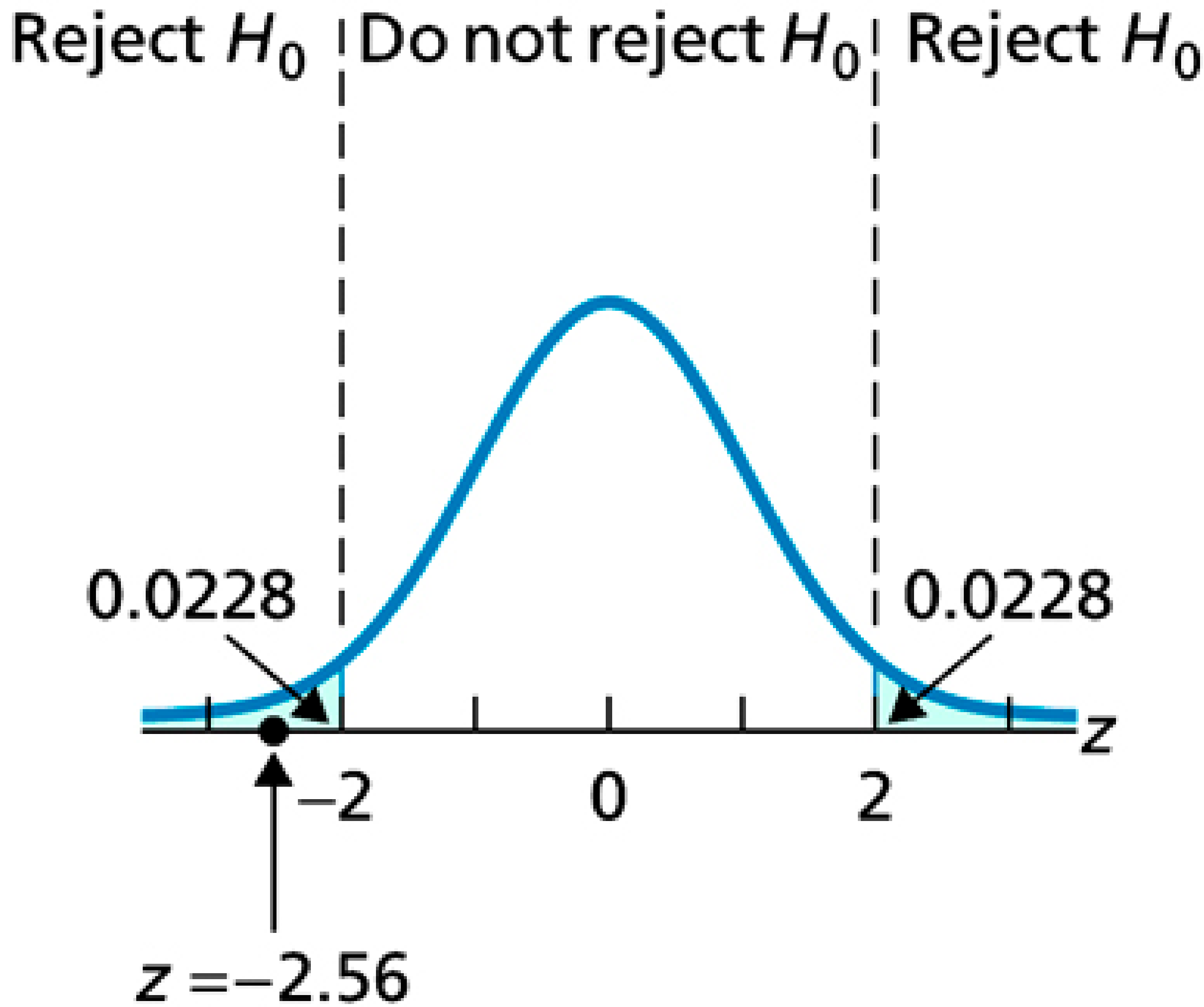
- The value of the test statistic is used in conjunction with a **decision rule** to determine *fail to reject H_o* or *reject H_o*
- The decision rule is established prior to collecting the data and specifies how you will reach the decision

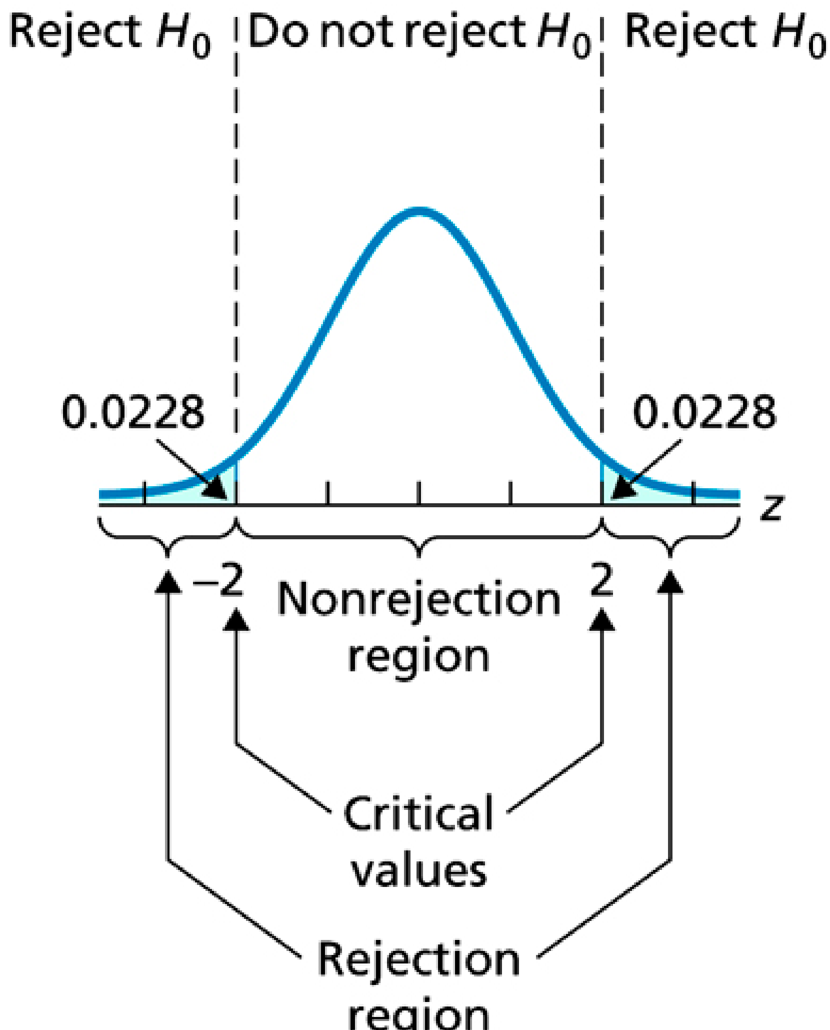
Solutions:

- H_o: $\mu = 454$ g.
- H_a: $\mu \neq 454$ g.




Reject H_0
Do not reject H_0
Reject H_0
450.88
454
457.12
$\bar{x}$
–2
0
2
z
(a)
Reject H_0
Do not reject H_0
Reject H_0
0.9544
0.0228
0.0228
450.88
454
457.12
$\bar{x}$
–2
0
2
z
(b)


Reject H_0
Do not reject H_0
Reject H_0
0.0228
0.0228
z
−2
0
2
z =−2.56


Reject H_0
Do not reject H_0
Reject H_0
0.0228
0.0228
z
−2
2
Nonrejection region
Critical values
Rejection region

Section 9.2

Critical-Value Approach to Hypothesis Testing

Example on Page 366

Example: Jack tells Jean that his average drive of a golf ball is 275 yards. A sample of 25 randomly selected drives were recorded and yielded an average of 264.4 yards. At 5% significance level, do the data provide sufficient evidence to conclude that Jack's mean driving distance is less than 275 yards? Assume that population is normal and population standard deviation is 20 yards.

Figure 9.1

Criterion for deciding whether to reject the null hypothesis

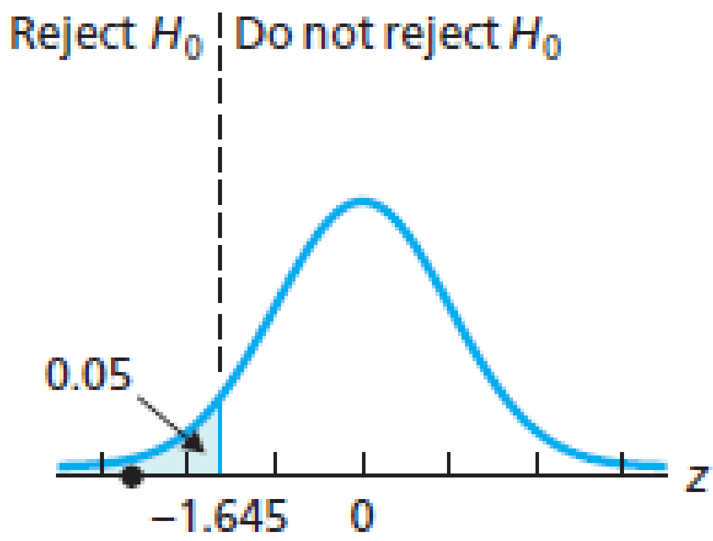


Figure 9.2

Rejection region, nonrejection region, and critical value for the golf-driving-distances hypothesis test

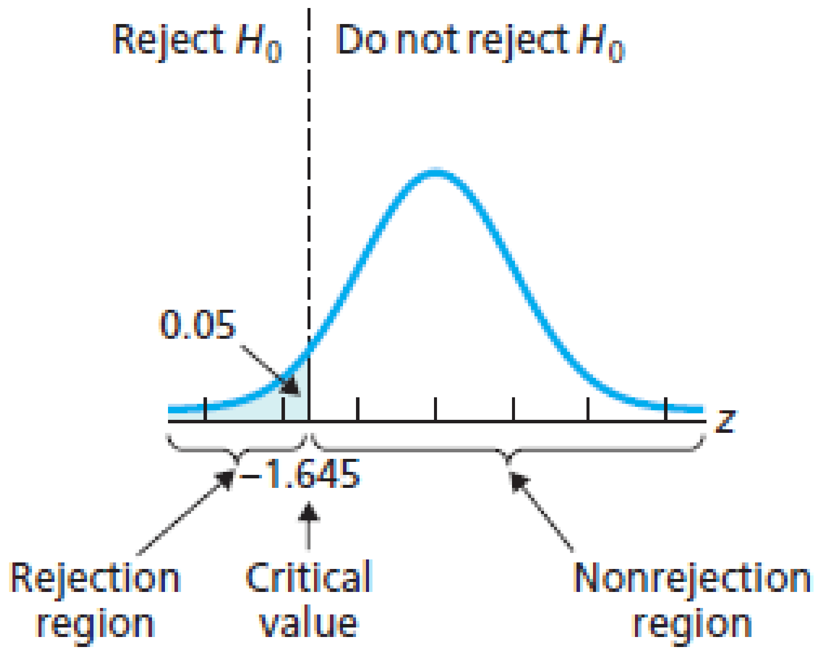
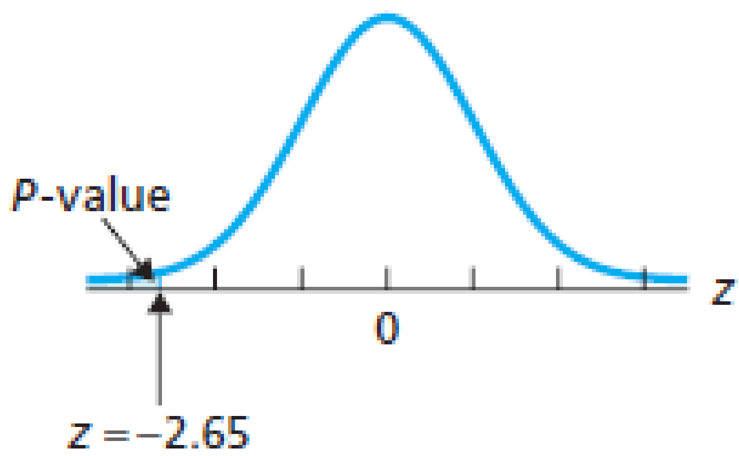


Figure 9.6

P-value for golf-driving-distances hypothesis test



More on the null hypothesis

Failing to reject the null hypothesis does not prove that H_0 is true, it means we have failed to disprove H_0 with some degree of confidents; i.e. we are 95% sure that H_0 is true.

Hence there is space for errors, since we are not 100% sure, so we will define various types of errors in later stage.

Examples for hypothesis testing:

1. Are Tim Kelly's weight measurements compatible with the claim that his true mean weight is 187 pounds?

Example: cont. (interpretation):

H_0: $\mu = 187$

In words: true weight is 187 pounds.

H_a: $\mu > 187$

In words: He weighs more than 187 pounds.

A so-called **one-sided** alternative H_a.

(Looking for a departure in one direction.)

Example: cont. (other possible settings):

- H_0: $\mu = 187$ vs. H_a: $\mu < 187$

Suspect the weight too low. **One-sided** H_a.

- H_0: $\mu = 187$ vs. H_a: $\mu > 187$

Suspect the weight too high. **One-sided** H_a.

- H_0: $\mu = 187$ vs. H_a: $\mu \neq 187$

Suspect weight is different. Two-sided H_a.

Note: decide on the setting **before** you see the data based on general knowledge or **other** measurements.

One-tailed and Two-tailed Tests

<u>Two-Sided Test</u>:

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0$$

<u>One-Sided Tests</u>:

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

One-Tailed Tests of Significance

- A test is one-tailed when the alternate hypothesis, H_1 , states a direction, such as:
  - H_1: The mean yearly commissions earned by full-time realtors is more than \$35,000. (μ>\$35,000)
  - H_1: The mean speed of trucks traveling on I-95 in Georgia is less than 60 miles per hour. (μ<60)
  - H_1: Less than 20 percent of the customers pay cash for their gasoline purchase. (P <.20)

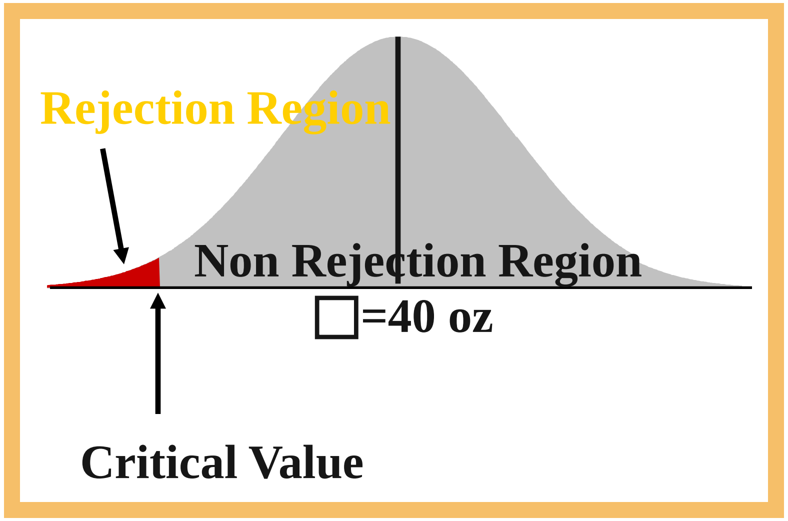
One-tailed Tests

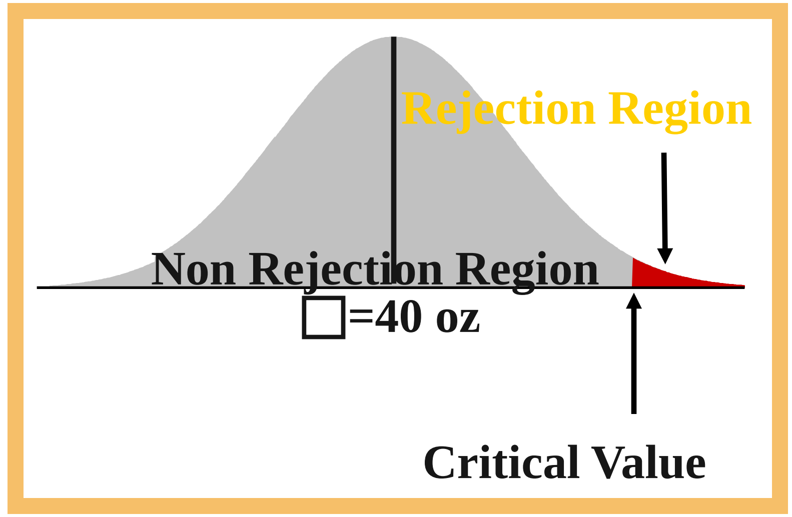
$$H_o : \mu \geq 40$$

$$H_a : \mu < 40$$

$$H_o : \mu \leq 40$$

$$H_a : \mu > 40$$





Two-Tailed Tests of Significance

- A test is two-tailed when no direction is specified in the alternate hypothesis H_1 , such as:
  - H_1: The mean amount spent by customers at the Wal-Mart in Georgetown is not equal to $25. ($\mu \neq \$25$).
  - H_1: The mean price for a gallon of gasoline is not equal to $1.54. ($\mu \neq \$1.54$).

Two-tailed Tests

$$H_o : \mu = 40$$

$$H_a : \mu \neq 40$$

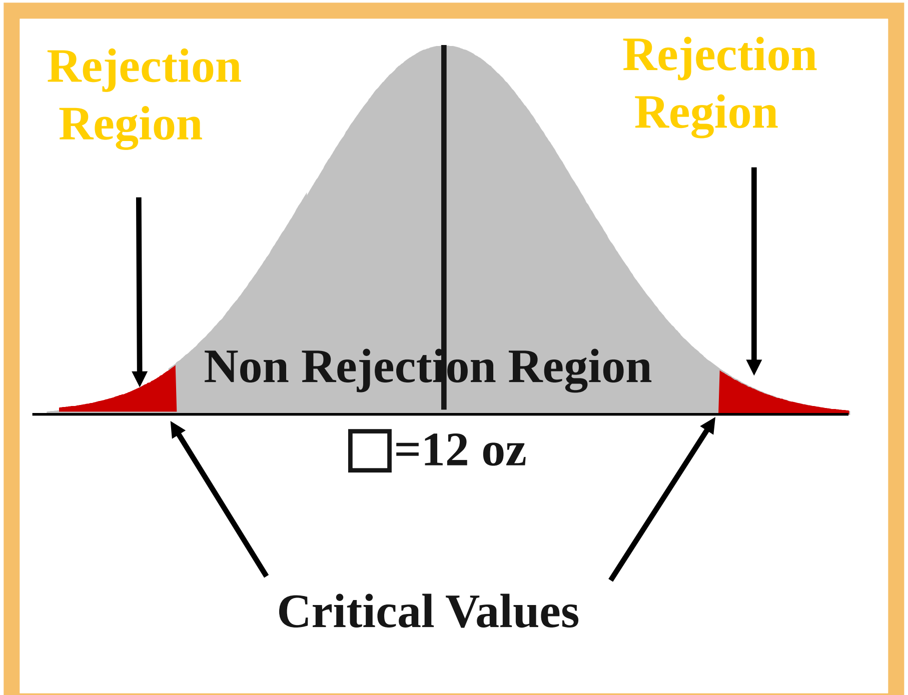


Figure 9.3

Graphical display of rejection regions for two-tailed, left-tailed, and right-tailed tests

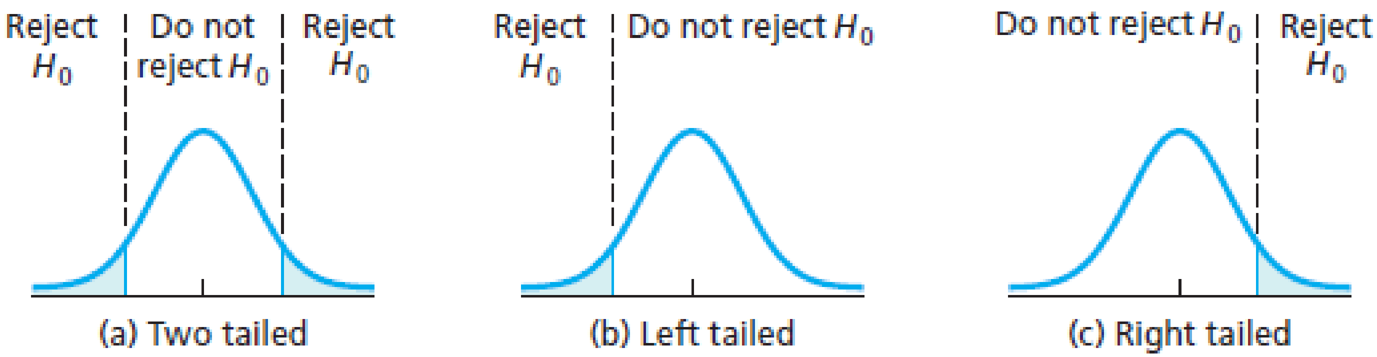
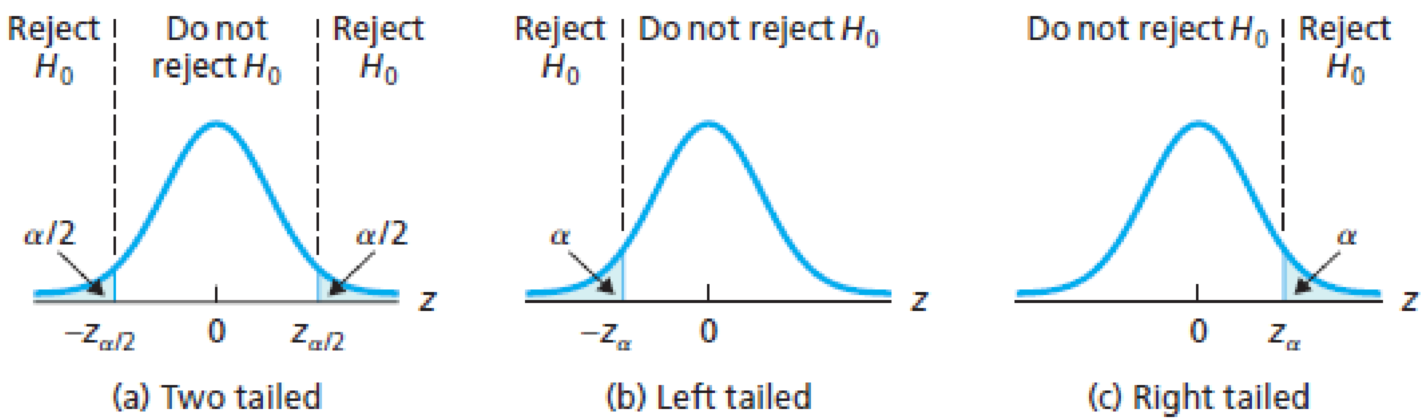


Figure 9.5

Critical value(s) for a hypothesis test at the significance level α if the test is (a) two tailed, (b) left tailed, or (c) right tailed



Decision Rule

Decision Rule:

a. If the test statistic falls within the critical region, we will reject H_o (the critical value is part of the critical region)

b. If the test statistic is in the noncritical region, we will fail to reject H_o

Notes

1. The null hypothesis specifies a particular value of a population parameter

2. The alternative hypothesis can take three forms. Each form dictates a specific location of the critical region(s)

3. For many hypothesis tests, the sign in the alternative hypothesis points in the direction in which the critical region is located

Sign in the Alternative Hypothesis	$<$	$\neq$	$>$
Critical Region	One region Left side **One-tailed test**	Two regions Half on each side **Two-tailed test**	One region Right side **One-tailed test**

4. Significance level: α

Section 9.3

P-Value Approach to Hypothesis Testing

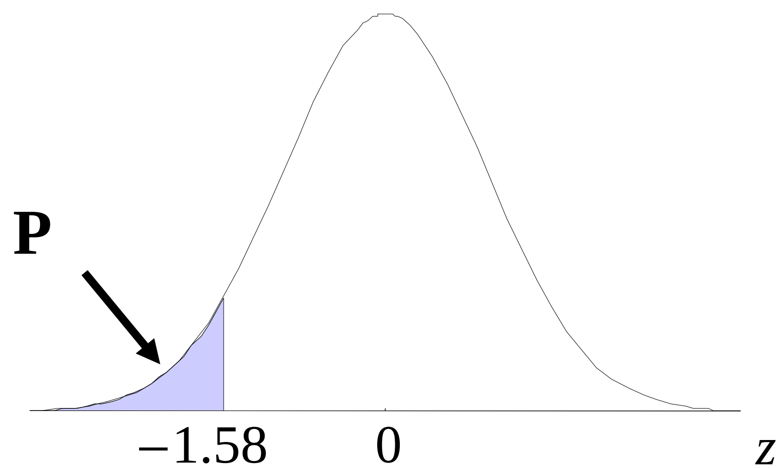
Definition 9.5

P-Value

The ***P*-value** of a hypothesis test is the probability of getting sample data at least as inconsistent with the null hypothesis (and supportive of the alternative hypothesis) as the sample data actually obtained. We use the letter ***P*** to denote the *P*-value.

Probability-Value or *p*-Value

Probability-Value, or *p*-Value: The probability that the test statistic could be the value it is or a more extreme value (in the direction of the alternative hypothesis) when the null hypothesis is true (*Note*: the symbol **P** will be used to represent the *p*-value, especially in algebraic situations)



$$\mathbf{P} = P(z < z^*) = P(z < -1.58)$$
$$= 0.0571$$

Solution Continued

b. Determine whether or not the *p*-value is smaller than α

The *p*-value (0.0571) is greater than α (0.05)

5. *The Results*

Decision Rule:

a. If the *p*-value is **less than or equal** to the level of significance α, then the decision must be to **reject** H_o

b. If the *p*-value is **greater than** the level of significance α, then the decision must be to **fail to reject** H_o

a. State the decision about H_o

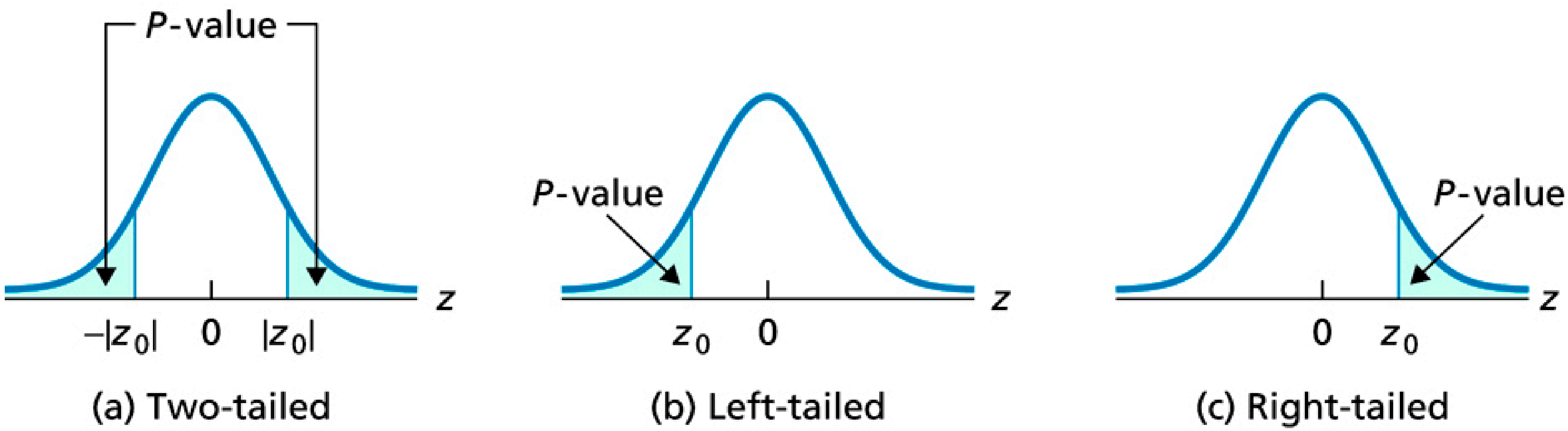
Decision about H_o : Fail to reject H_o

b. Write a conclusion about H_a

There is not sufficient evidence at the 0.05 level of significance to show that the mean weight of cereal boxes is less than 24 ounces

Figure 9.7

P-value for a one-mean z-test if the test is (a) two tailed, (b) left tailed, or (c) right tailed



p-Value in Hypothesis Testing

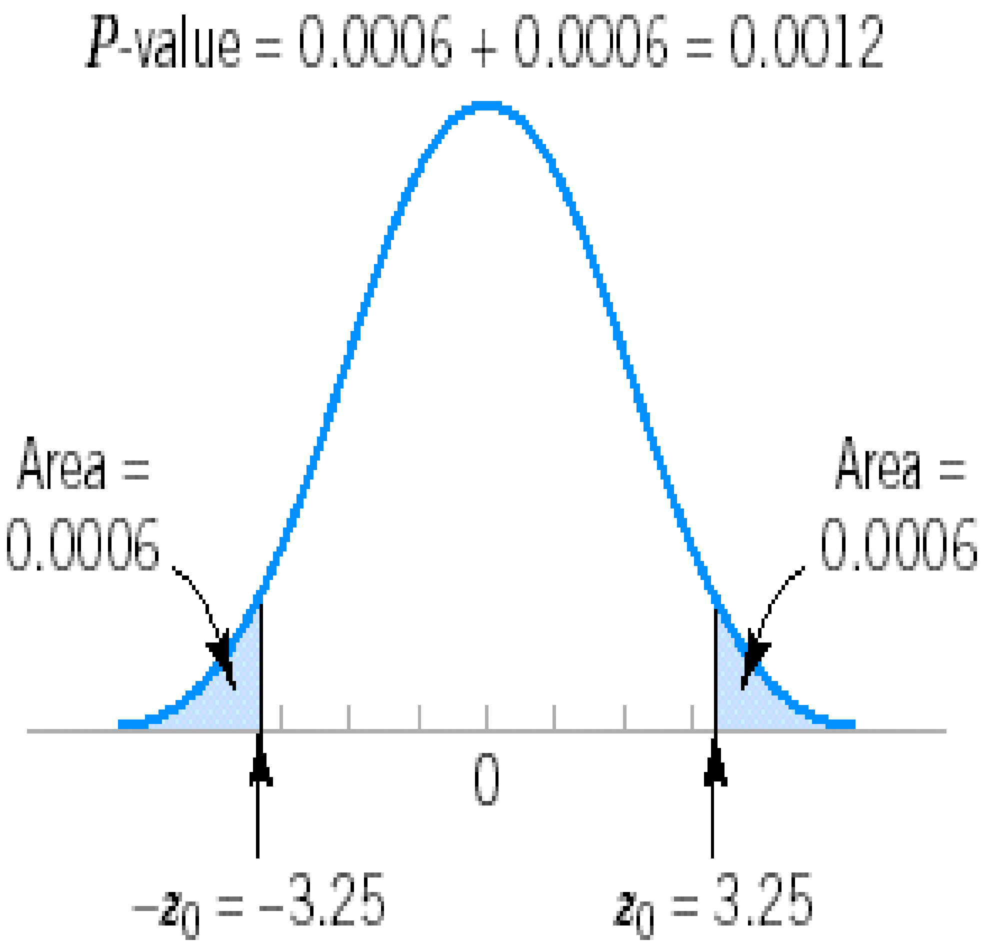
- A *p*-Value is the probability, assuming that the null hypothesis is true, of finding a value of the test statistic at least as extreme as the computed value for the test.
  - If the *p*-Value is smaller than the significance level, H_0 is rejected.
  - If the *p*-Value is larger than the significance level, H_0 is not rejected.

p-Value in Hypothesis Testing

In other words, in the *p*-Value approach, instead of fixing the significance level beforehand, we report the highest level at which the test statistic is observed to be significant. i.e. Do we have evidence that H_0 is not true.

p-Values in Hypothesis Testing

Illustration of the P-value for the two-tailed test



Computation of the *p*-Value

- One-Tailed Test:

  p-Value = $P\{z \geq$ absolute value of the computed test statistic value$\}$

- Two-Tailed Test:

  p-Value = $2P\{z \geq$ absolute value of the computed test statistic value$\}$

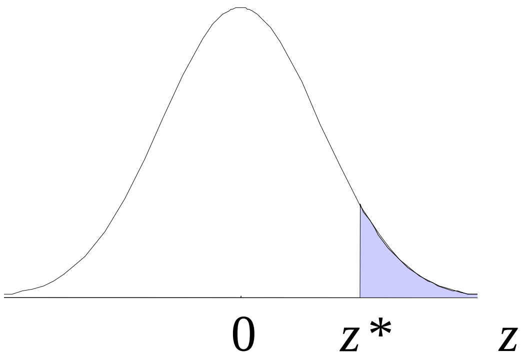
Notes

- If we fail to reject H_o, there is no evidence to suggest the null hypothesis is false. This does not mean H_o is true.
- The *p*-value is the area, under the curve of the probability distribution for the test statistic, that is more extreme than the calculated value of the test statistic.
- There are 3 separate cases for *p*-values. The direction (or sign) of the alternative hypothesis is the key.

Finding *p*-Values

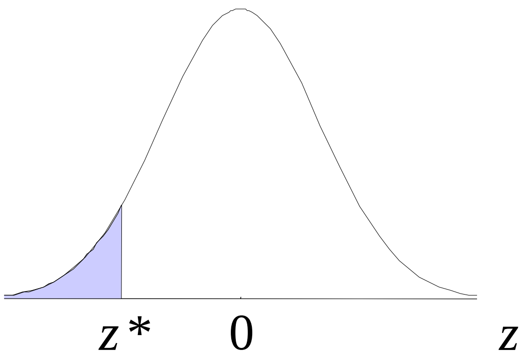
1. H_a contains $>$ (Right tail)

   p-value = $P(z > z^*)$



2. H_a contains $<$ (Left tail)

   p-value = $P(z < z^*)$



3. H_a contains $\neq$ (Two-tailed)

   p-value = $P(z < -|z^*|) + P(z > |z^*|)$

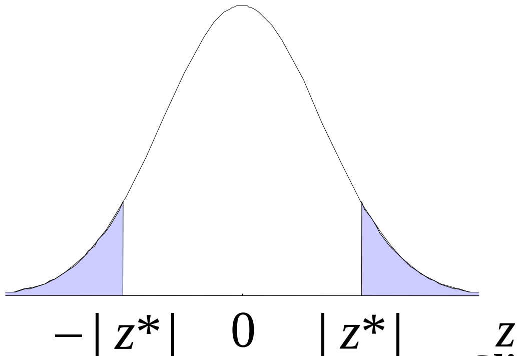
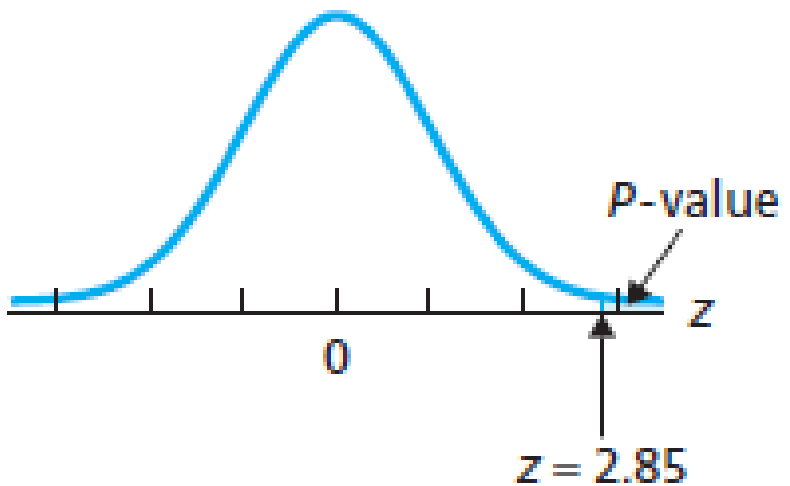


Figure 9.9

Value of the test statistic and the P-value



How to interpret P-values

Significance level	Reported probability	z-value one-sided test	z-value two-sided test	description
> 20%	P > 0.2	Z < 0.842	Z < 1.282	not significant
< 20%	P < 0.2	Z > 0.842	Z > 1.282	possibly significant
< 10%	P < 0.1	Z > 1.282	Z > 1.645	nearly significant
< 5%	P < 0.05	Z > 1,645	Z > 1.960	significant
< 1%	P < 0.01	Z > 2.326	Z > 2.576	very significant
< 0.5%	P < 0.005	Z > 2.576	Z > 2.807	highly significant
< 0.1%	P < 0.001	Z > 3.090	Z > 3.291	very highly significant

Table 9.8

Guidelines for using the P-value to assess the evidence against the null hypothesis

***P*-value**	**Evidence against H_0**
$P > 0.10$	Weak or none
$0.05 < P \leq 0.10$	Moderate
$0.01 < P \leq 0.05$	Strong
$P \leq 0.01$	Very strong

Section 9.4

Hypothesis Tests for One Population Mean When Sigma Is Known

Procedure 9.1

One-Mean z-Test

Purpose To perform a hypothesis test for a population mean, μ

Assumptions

1. Simple random sample
2. Normal population or large sample
3. σ known

Step 1 **The null hypothesis is H_0: $\mu = \mu_0$, and the alternative hypothesis is**

$$H_a: \mu \neq \mu_0 \quad \text{or} \quad H_a: \mu < \mu_0 \quad \text{or} \quad H_a: \mu > \mu_0$$

(Two tailed) or (Left tailed) or (Right tailed)

Step 2 **Decide on the significance level, α.**

Step 3 **Compute the value of the test statistic**

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

and denote that value z_0.

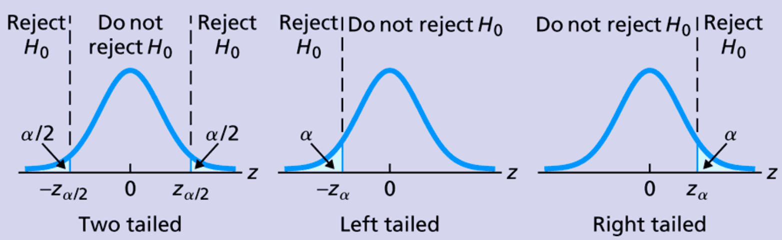
Procedure 9.1 (cont.)

CRITICAL-VALUE APPROACH	OR	P-VALUE APPROACH

Step 4 The critical value(s) are

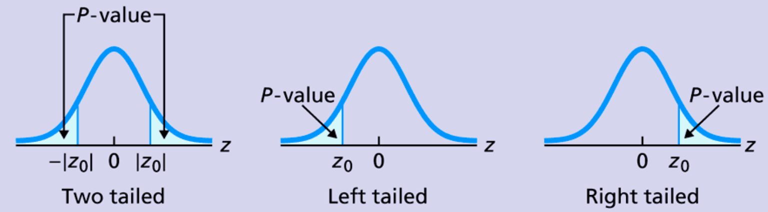
$\pm z_{\alpha/2}$ (Two tailed) or $-z_{\alpha}$ (Left tailed) or z_{α} (Right tailed)

Use Table II to find the critical value(s).



Step 5 If the value of the test statistic falls in the rejection region, reject H_0; otherwise, do not reject H_0.

Step 4 Use Table II to obtain the P-value.



Step 5 If $P \leq \alpha$, reject H_0; otherwise, do not reject H_0.

Step 6 Interpret the results of the hypothesis test.

Note: The hypothesis test is exact for normal populations and is approximately correct for large samples from nonnormal populations.

Critical Value Approach Procedure

The One-Sample *z*-Test for a Population Mean (Critical-Value Approach)

Assumptions

1. Normal population or large sample
2. σ known

Step 1 The null hypothesis is H_0: $\mu = \mu_0$, and the alternative hypothesis is

H_a: $\mu \neq \mu_0$ (Two-tailed) or H_a: $\mu < \mu_0$ (Left-tailed) or H_a: $\mu > \mu_0$ (Right-tailed)

Step 2 Decide on the significance level, α.

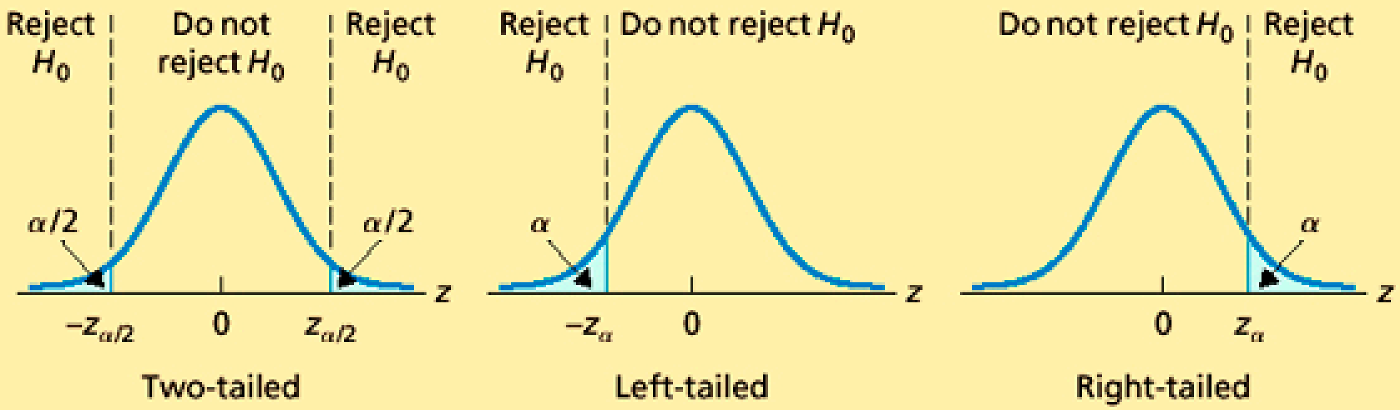
Step 3 Compute the value of the test statistic

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

Step 4 The critical value(s) are

$$\pm z_{\alpha/2} \text{ (Two-tailed)} \quad \text{or} \quad -z_{\alpha} \text{ (Left-tailed)} \quad \text{or} \quad z_{\alpha} \text{ (Right-tailed)}$$

Use Table II to find the critical value(s).



Step 5 If the value of the test statistic falls in the rejection region, reject H_0; otherwise, do not reject H_0.

Step 6 Interpret the results of the hypothesis test.

The hypothesis test is exact for normal populations and is approximately correct for large samples from nonnormal populations.

Example

Example: The mean water pressure in the main water pipe from a town well should be kept at 56 psi. Anything less and several homes will have an insufficient supply, and anything greater could burst the pipe. Suppose the water pressure is checked at 47 random times. The sample mean is 57.1. (Assume $\sigma = 7$). Is there any evidence to suggest the mean water pressure is different from 56? Use $\alpha = 0.01$

Solution:

1. *The Set-Up*
   a. Describe the parameter of concern:
      The mean water pressure in the main pipe

   b. State the null and alternative hypotheses
      H_o: $\mu = 56$
      H_a: $\mu \neq 56$

Solution Continued

2. *The Hypothesis Test Criteria*

   a. Check the assumptions:
      A sample of n = 47 is large enough for the CLT to apply

   b. Identify the test statistic
      The test statistic is z*

   c. Determine the level of significance: α = 0.01 (given)

3. *The Sample Evidence*

   a. The sample information: $\bar{x} = 57.1, \quad n = 47$

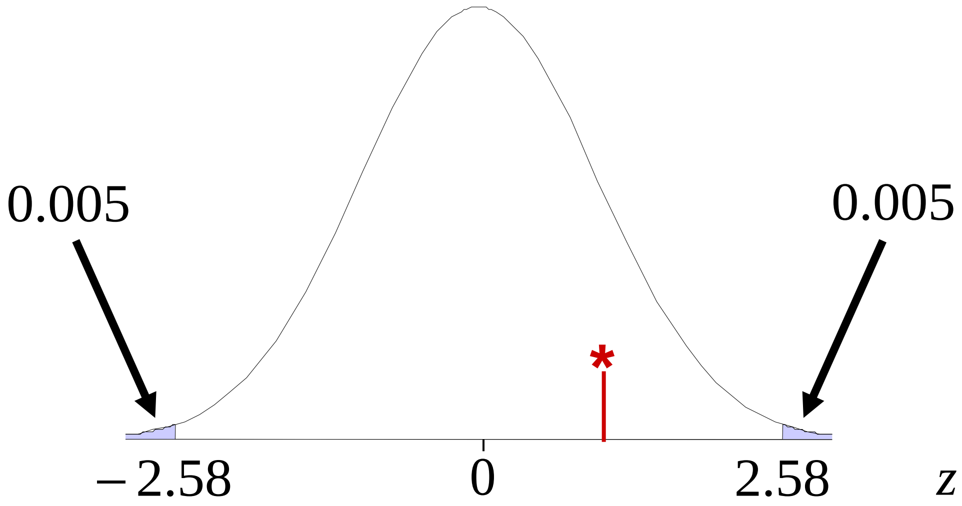
   b. Calculate the value of the test statistic:

$$z^* = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{57.1 - 56}{7 / \sqrt{47}} = 1.077$$

Solution Continued

4. *The Probability Distribution*

a. Determine the critical regions and the critical values



b. Determine whether or not the calculated test statistic is in the critical region

The calculated value of z, z* = 1.077, is in the *noncritical* region

Solution Continued

5. *The Results*

a. State the decision about H_o:

Fail to reject H_o

b. State the conclusion about H_a:

There is no evidence to suggest the water pressure is different from 56 at the 0.01 level of significance

Example

Example: An elementary school principal claims students receive no more than 30 minutes of homework each night. A random sample of 36 students showed a sample mean of 36.8 minutes spent doing homework (assume $\sigma = 7.5$). Is there any evidence to suggest the mean time spent on homework is greater than 30 minutes? Use $\alpha = 0.05$

Solution:

1. *The Set-Up*

The parameter of concern: μ, the mean time spent doing homework each night

H_o: $\mu = 30$ ($\leq$)
H_a: $\mu > 30$

Solution Continued

2. *The Hypothesis Test Criteria*

   a. The sample size is n = 36, the CLT applies

   b. The test statistic is $z*$

   c. The level of significance is given: $\alpha = 0.01$

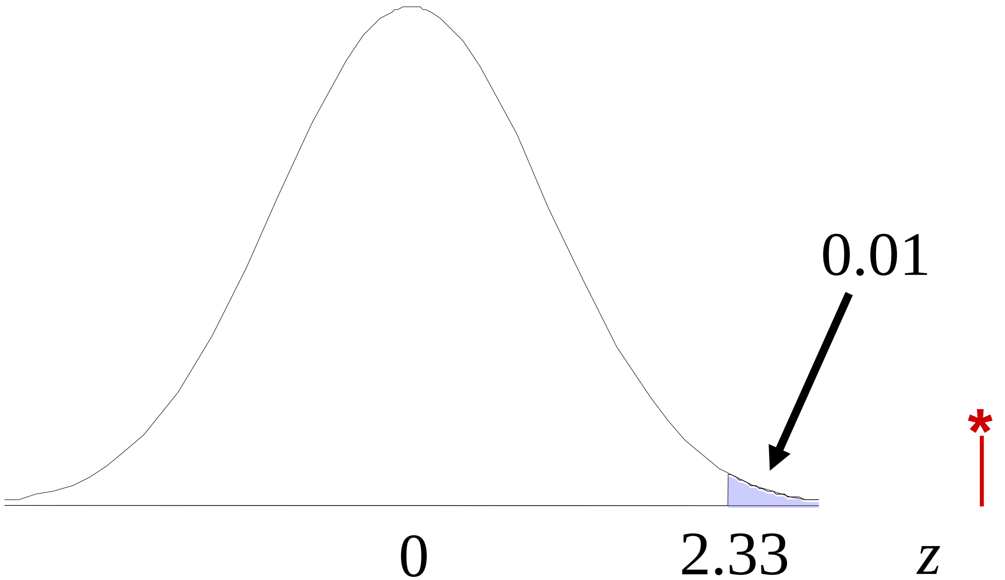
3. *The Sample Evidence*

$$\bar{x} = 36.8, \quad n = 36$$

$$z* = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{36.8 - 30}{7.5 / \sqrt{36}} = 5.44$$

Solution Continued

4. *The Probability Distribution*



The calculated value of z, $z^* = 5.44$, is in the *critical region*

Example

Example: A company advertises the net weight of its cereal is 24 ounces. A consumer group suspects the boxes are underfilled. They cannot check every box of cereal, so a sample of cereal boxes will be examined. A decision will be made about the true mean weight based on the sample mean of size 40 (the sample mean is 23.95 ounces). Test the consumer group's claim at 2% significant level. Assume $\sigma = 0.2$

Example Continued

- **Example Continued:** Weight of cereal boxes

Recall: H_o: μ = 24 (≥) (at least 24) H_a: μ < 24 (less than 24)

2. *The Hypothesis Test Criteria*

   a. Check the assumptions

   The weight of cereal boxes is probably mound shaped. A sample size of 40 should be sufficient for the CLT to apply. The sampling distribution of the sample mean can be expected to be normal.

   b. Identify the probability distribution and the test statistic to be used

   To test the null hypothesis, ask how many standard deviations away from μ is the sample mean

$$\text{test statistic}: \quad z^* = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

Solution Continued

c. Determine the level of significance

Let $\alpha = 0.02$

3. *The Sample Evidence*

a. Collect the sample information

A random sample of 40 cereal boxes is examined

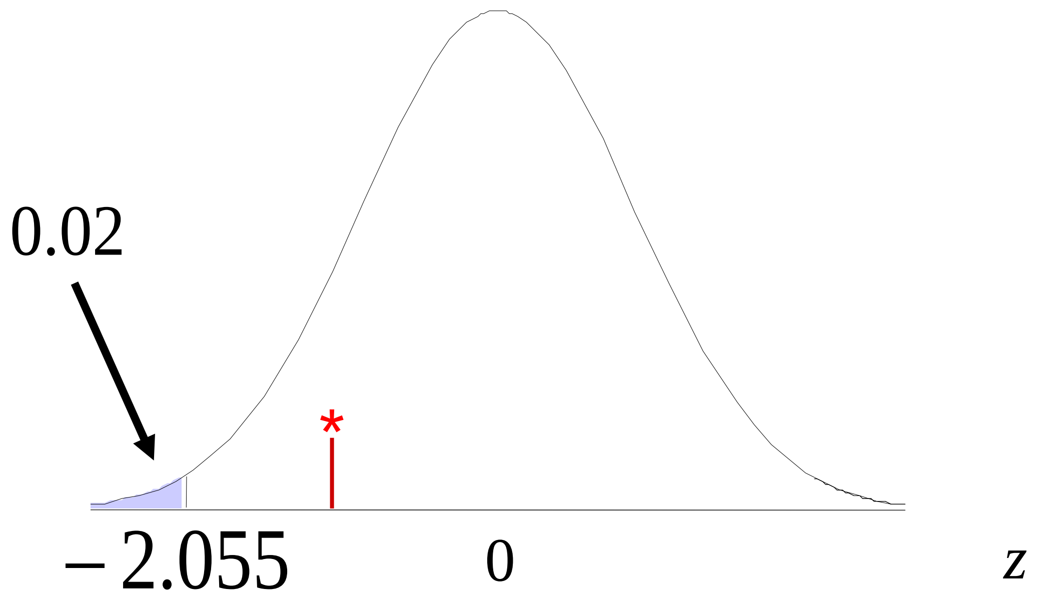
$\bar{x} = 23.95$ and $n = 40$

b. Calculate the value of the test statistic ($\sigma = 0.2$)

$$z* = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{23.95 - 24}{0.2 / \sqrt{40}} = -1.5811$$

Solution Continued

4. *The Probability Distribution*



The calculated value of z, $z^* = -1.5811$, is outside the *critical region*. *Hence accept H_0.*

Key Fact, Again

- Small sample ($n > 15$) – population must be normal.
- Moderate sample ($30 > n \geq 15$) – beware of outliers or the case that population deviate far from normal.
- Large sample ($n \geq 30$) – beware of outliers and if there is justifications to remove the outliers, then we should do so, before the test. Else, we should observe the inferences of outliers.

P-Value Approach Procedure

The One-Sample z-Test for a Population Mean (P-Value Approach)

Assumptions

1. Normal population or large sample
2. σ known

Step 1 The null hypothesis is H_0: $\mu = \mu_0$, and the alternative hypothesis is

H_a: $\mu \neq \mu_0$ (Two-tailed) or H_a: $\mu < \mu_0$ (Left-tailed) or H_a: $\mu > \mu_0$ (Right-tailed)

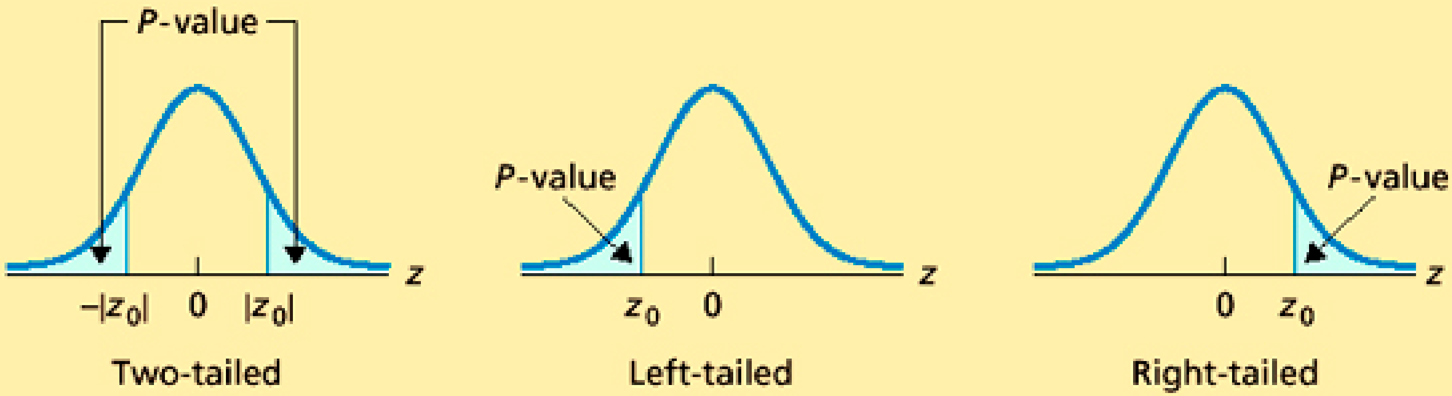
Step 2 Decide on the significance level, α.

Step 3 Compute the value of the test statistic

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

and denote that value z_0.

Step 4 Use Table II to obtain the *P*-value.



Step 5 If $P \leq \alpha$, reject H_0; otherwise, do not reject H_0.

Step 6 Interpret the results of the hypothesis test.

The hypothesis test is exact for normal populations and is approximately correct for large samples from nonnormal populations.

Example

Example: The mean age of all shoppers at a local jewelry store is 37 years (with a standard deviation of 7 years). In an attempt to attract older adults with more disposable income, the store launched a new advertising campaign. Following the advertising, a random sample of 47 shoppers showed a mean age of 39.3. Is there sufficient evidence to suggest the advertising campaign has succeeded in attracting older customers?

Solution:

1. *The Set-Up*

   a. Parameter of concern: the mean age, μ, of all shoppers

   b. The hypotheses:

   H_o: $\mu = 37$ ($\leq$)

   H_a: $\mu > 37$

Solution Continued

2. *The Hypothesis Test Criteria*

   a. The assumptions: The distribution of the age of shoppers is unknown. However, the sample size is large enough for the CLT to apply.

   b. The test statistic: The test statistic will be $z*$

   c. The level of significance: none given
   We will find a p-value

3. *The Sample Evidence*

   a. Sample information: $n = 47, \quad \bar{x} = 39.3$

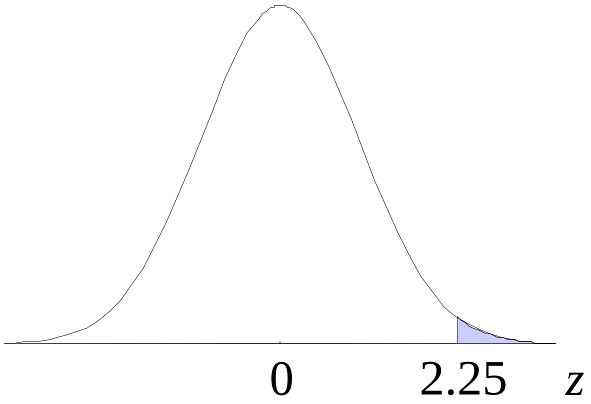
   b. Calculated test statistic: $z* = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \dfrac{39.3 - 37}{7 / \sqrt{47}} = 2.25$

Solution Continued

4. *The Probability Distribution*

   a. The *p*-value:

$$\begin{aligned} p\text{ - value} &= P(z > z^*) \\ &= P(z > 2.25) \\ &= 1 - 0.9878 \\ &= 0.0122 \end{aligned}$$



   b. Determine whether or not the *p*-value is smaller than α

   A comparison is not possible, no α given

5. *The Results*

Because the *p*-value is so small (**P** < 0.05), there is evidence to suggest the mean age of shoppers at the jewelry store is greater than 37

p-Value

The **idea of the *p*-value** is to express the degree of belief in the null hypothesis:

1. When the *p*-value is minuscule (like 0.0001), the null hypothesis would be rejected by everyone because the sample results are very unlikely for a true H_o
2. When the *p*-value is fairly small (like 0.01), the evidence against H_o is quite strong and H_o will be rejected by many
3. When the *p*-value begins to get larger (say, 0.02 to 0.08), there is too much probability that data like the sample involved could have occurred even if H_o were true, and the rejection of H_o is not an easy decision
4. When the *p*-value gets large (like 0.15 or more), the data is not at all unlikely if the H_o is true, and no one will reject H_o

p-Value Advantages & Disadvantage

Advantages of *p*-value approach:

1. The results of the test procedure are expressed in terms of a continuous probability scale from 0.0 to 1.0, rather than simply on a reject or fail to reject basis
2. A *p*-value can be reported and the user of the information can decide on the strength of the evidence as it applies to his/her own situation
3. Computers can do all the calculations and report the *p*-value, thus eliminating the need for tables

Disadvantage:

1. Tendency for people to put off determining the level of significance

Example

Example: The active ingredient for a drug is manufactured using fermentation. The standard process yields a mean of 26.5 grams (assume $\sigma = 3.2$). A new mixing technique during fermentation is implemented. A random sample of 32 batches showed a sample mean 27.1. Is there any evidence to suggest the new mixing technique has changed the yield?

Solution:

1. *The Set-Up*

   a. The parameter of interest is the mean yield of active ingredient, μ

   b. The null and alternative hypotheses:

   H_0: $\mu = 26.5$

   H_a: $\mu \neq 26.5$

Solution Continued

2. *The Hypothesis Test Criteria*

   a. Assumptions: A sample of size 32 is large enough to satisfy the CLT

   b. The test statistic: z^*

   c. The level of significance: find a *p*-value

3. *The Sample Evidence*

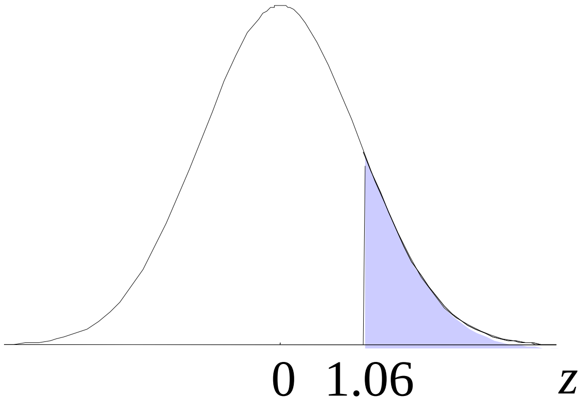
   a. From the sample: $n = 32, \quad \bar{x} = 27.1$

   b. The calculated test statistic: $z^* = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \dfrac{27.1 - 26.5}{3.2 / \sqrt{32}} = 1.06$

Solution Continued

4. *The Probability Distribution*
   a. The *p*-value:

$$\begin{aligned} p\text{ - value} &= 2 \times P(z >| z^* |) \\ &= 2 \times P(z > 1.06) \\ &= 2 \times (1 - 0.8554) \\ &= 2 \times 0.1446 = 0.2892 \end{aligned}$$



   b. The *p*-value is large
      There is no α given in the statement of the problem

5. *The Results*

Because the *p*-value is large (**P** = 0.2892), there is no evidence to suggest the new mixing technique has changed the mean yield

Idea of a p-value

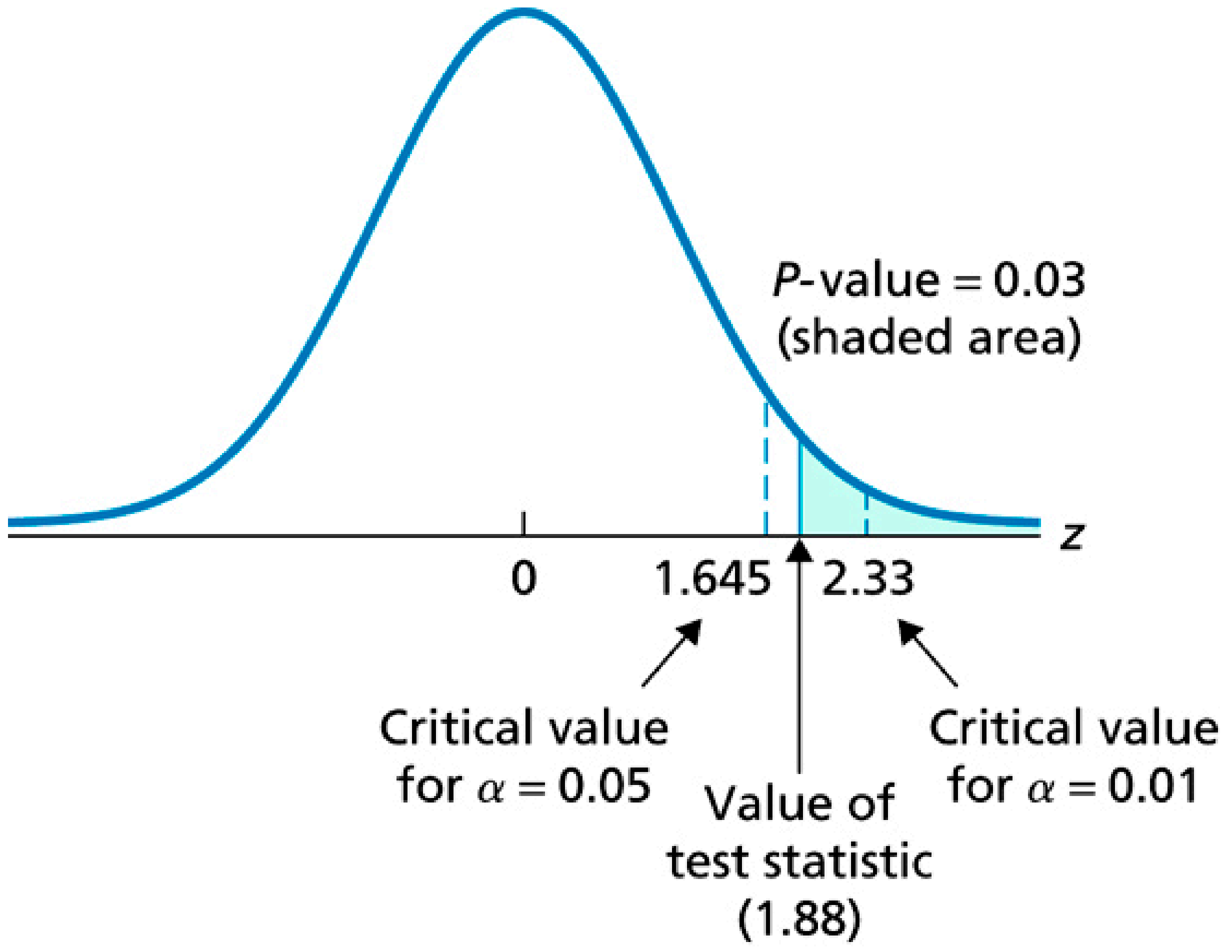


Table 9.11

Critical-Value Approach	*P*-Value Approach
Step 1 State the null and alternative hypotheses.	**Step 1** State the null and alternative hypotheses.
Step 2 Decide on the significance level, α.	**Step 2** Decide on the significance level, α.
Step 3 Compute the value of the test statistic.	**Step 3** Compute the value of the test statistic.
Step 4 Determine the critical value(s).	**Step 4** Determine the P-value.
Step 5 If the value of the test statistic falls in the rejection region, reject H_0; otherwise, do not reject H_0.	**Step 5** If $P \leq \alpha$, reject H_0; otherwise, do not reject H_0.
Step 6 Interpret the result of the hypothesis test.	**Step 6** Interpret the result of the hypothesis test.

Examples

- Read some of the examples in the textbook


Tonight's English
as a second
language class
has been
cancelled.
©2010 Scott Hilburn/Distributed by Universal Uclick

Section 9.7

Type II Error Probabilities; Power

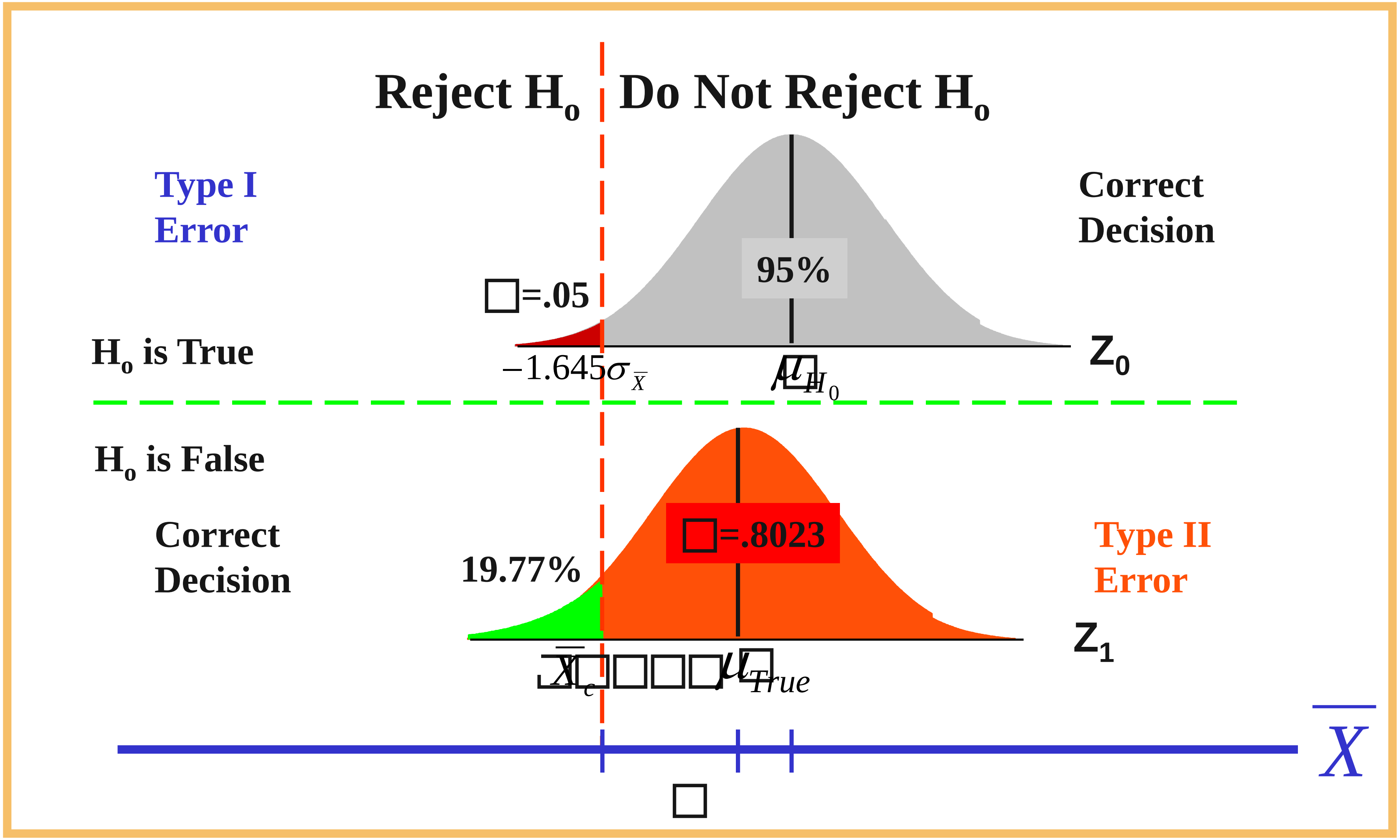
Type II error in Hypothesis Testing

POWER

The probability that a fixed level α significance test will reject H_0 when a particular alternative value of the parameter is true is called the **power** of the test.

- The power is computed as 1 - β, 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 $H_1 : \mu \neq 50$ 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 $\mu = 52$.

Type II Error Demonstration



Type II Error Demonstration

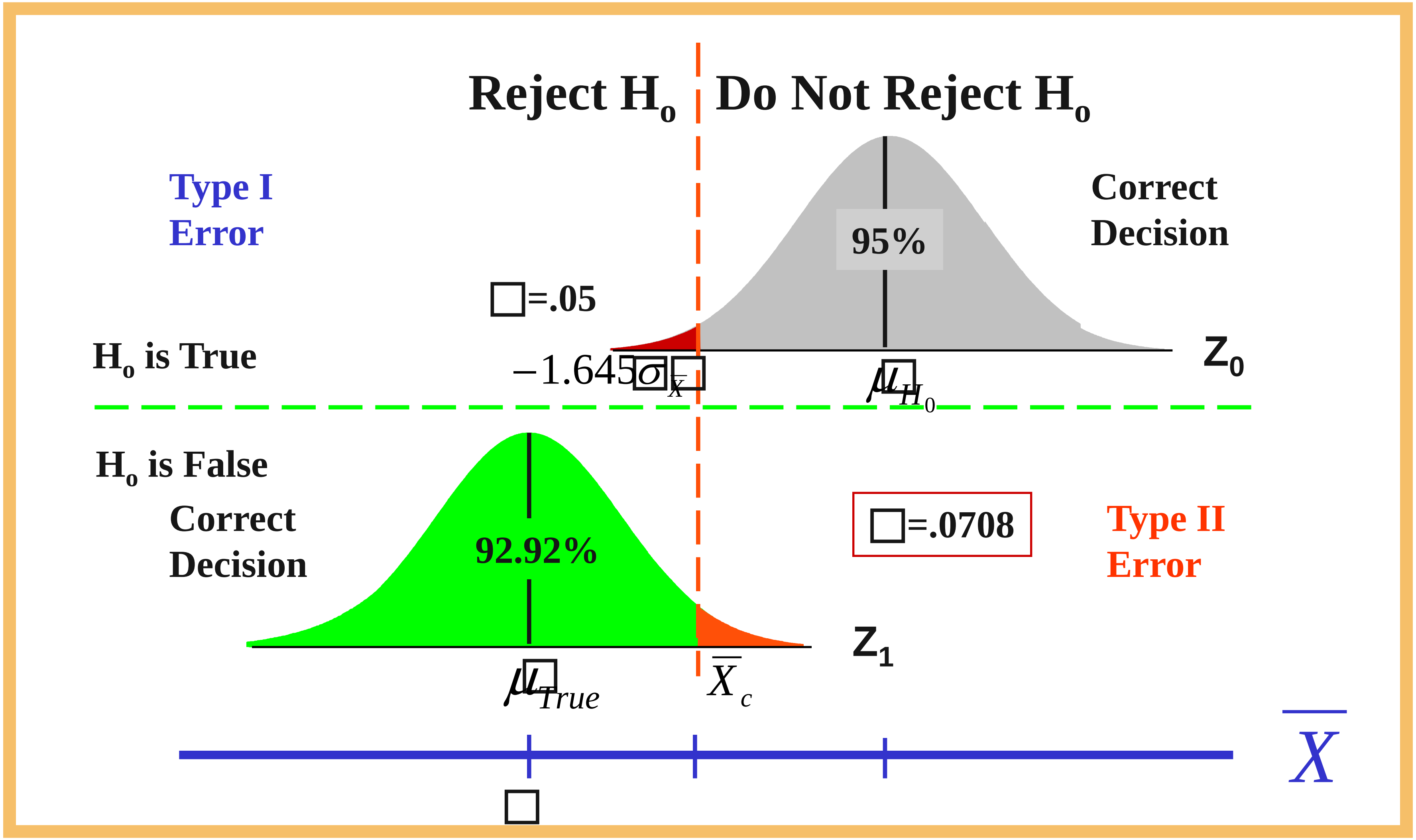


Figure 9.28

Decision criterion for the gas mileage illustration ($\alpha = 0.05$, $n = 30$)

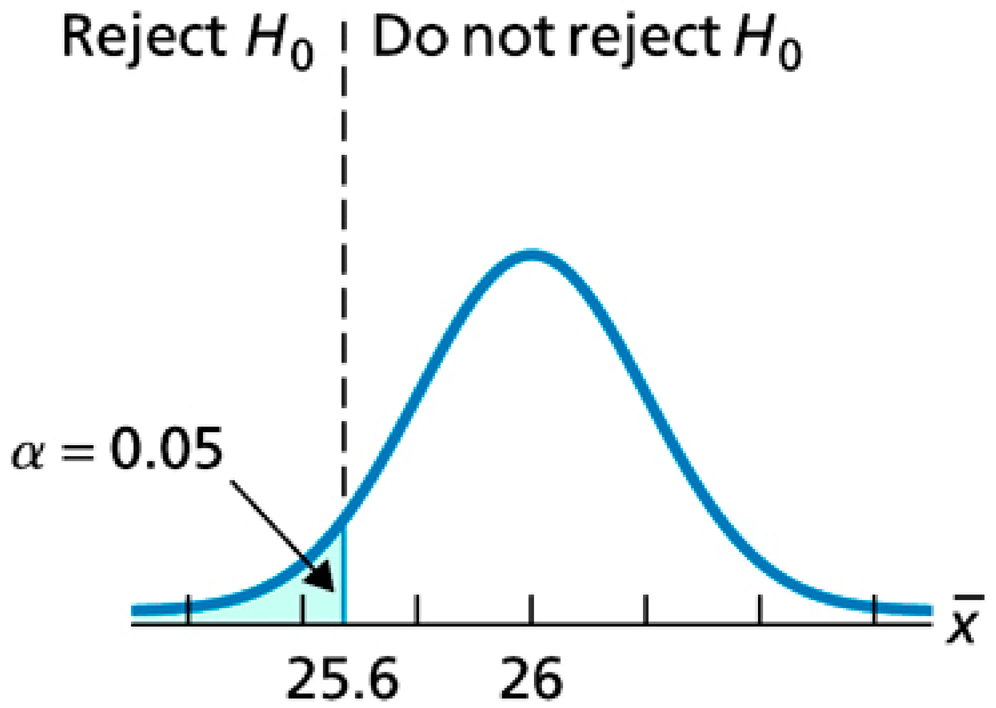
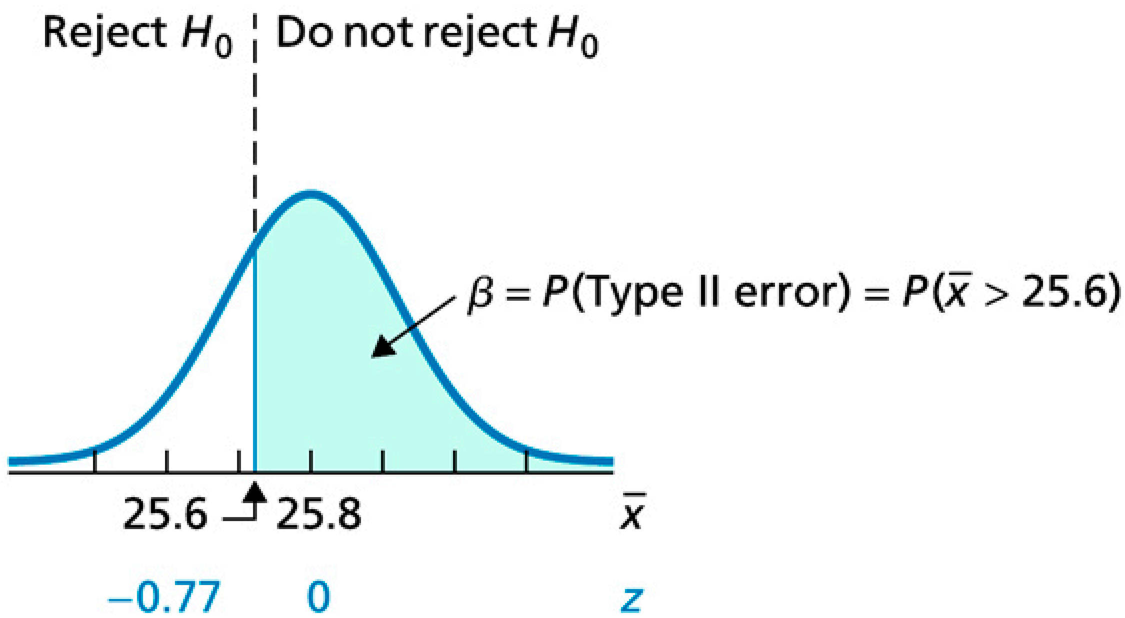


Figure 9.29

Determining the probability of a Type II error if $\mu = 25.8$ mpg



z-score computation:

$$\bar{x} = 25.6 \longrightarrow z = \frac{25.6 - 25.8}{0.26} = -0.77$$

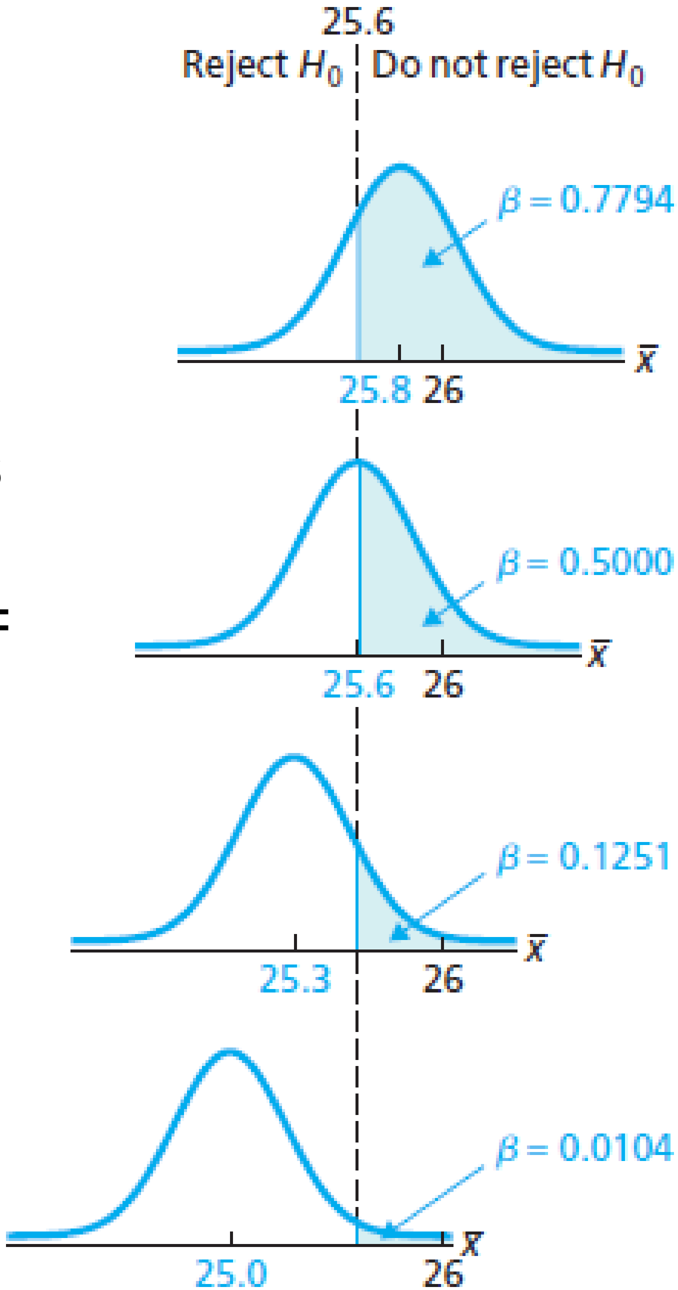
Area to the left of z:

0.2206

Shaded area = 1 − 0.2206 = 0.7794

Figure 9.31

Type II error probabilities for μ = 25.8, 25.6, 25.3, and 25.0 (α = 0.05, n = 30)



Definition 9.6

Power

The **power** of a hypothesis test is the probability of not making a Type II error, that is, the probability of rejecting a false null hypothesis. We have

$$\text{Power} = 1 - P(\text{Type II error}) = 1 - \beta .$$

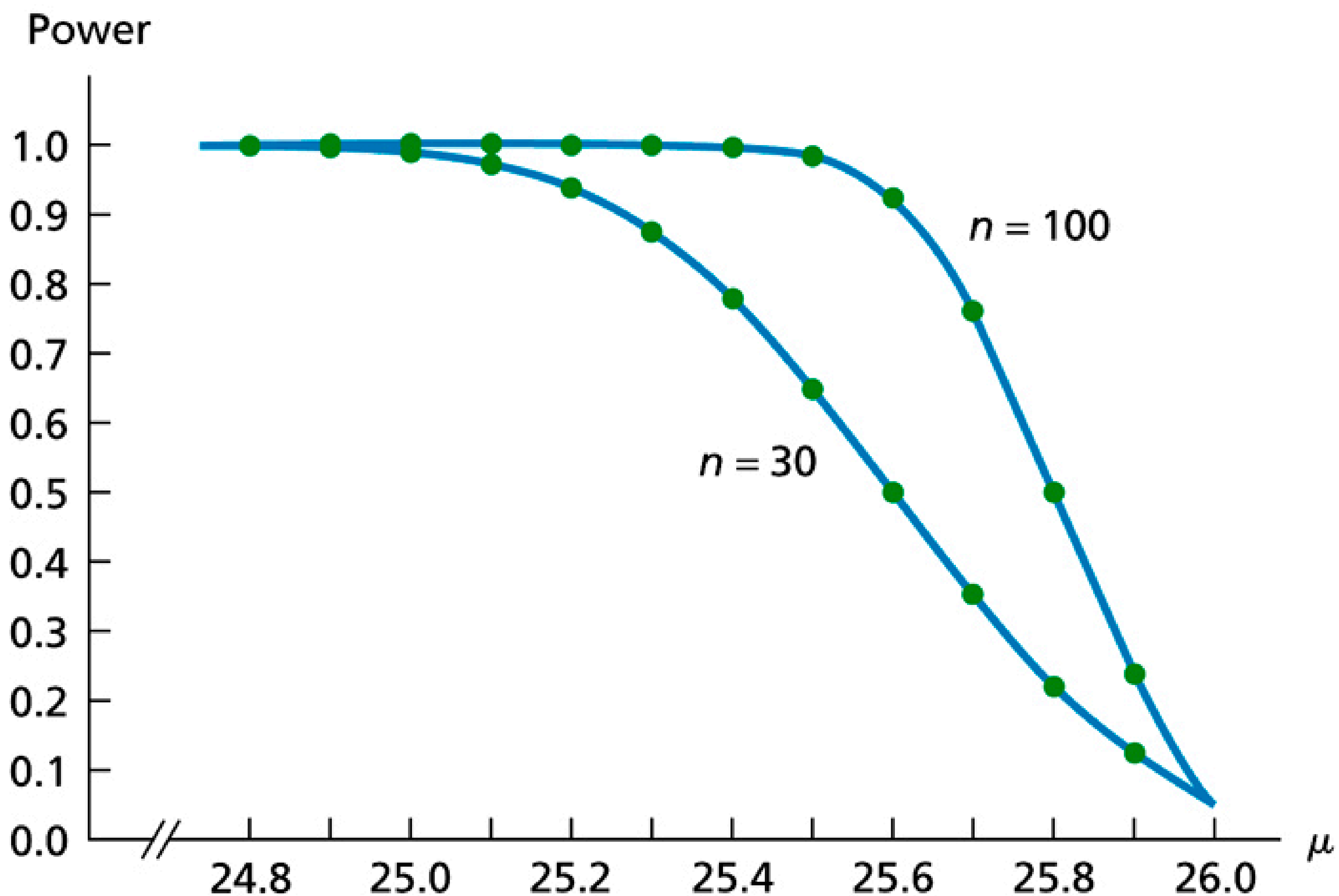
Table 9.16

Selected Type II error probabilities and powers for the gas mileage illustration ($\alpha = 0.05, n = 30$)

True mean μ	P (Type II error) β	Power $1 - \beta$
25.9	0.8749	0.1251
25.8	0.7794	0.2206
25.7	0.6480	0.3520
25.6	0.5000	0.5000
25.5	0.3520	0.6480
25.4	0.2206	0.7794
25.3	0.1251	0.8749
25.2	0.0618	0.9382
25.1	0.0274	0.9726
25.0	0.0104	0.9896
24.9	0.0036	0.9964
24.8	0.0010	0.9990

Figure 9.34

Power curves for the gas mileage illustration when $n = 30$ and $n = 100$ ($\alpha = 0.05$)



Example: Iron Deficiency. A statement reported that the recommended daily allowance of iron for adult females under the age of 51 is 18mg. A hypothesis is to be performed to decide whether the adult females under the age of 51 are, on average, getting less than the is recommended daily allowance 18mg of iron, at 1% significance level. Suppose we assume the population standard deviation is 4.2mg and 45 samples will be randomly chosen.

Determine the type II error for this test, if the mean recommended daily allowance of iron for adult females under the age of 51 is

1. 15.50mg
2. 17.25 mg

Example: Serving Time. According to a report, the mean length of imprisonment for motor-vehicle theft in Australia is 16.7 months. Now we want to perform a hypothesis test at the 5% significance level to decide whether the mean length of imprisonment for motor-vehicle theft in Sydney differs from the national mean? Suppose we assume the population standard deviation is 6 months and the sample size used to conduct the test is 100.

What is the power of this test, if the mean length of imprisonment for motor-vehicle theft in Sydney is

1. 14.5 months.
2. 17.0 months.