8.6 If a right-tailed test is rejected at the 1% level of significance, do we need to test at the 5% level of significance? Why?

8.7 If the null hypothesis is not rejected, does that imply that it is true? Explain.

Learning the Mechanics

8.8 Consider a test of H_0: $\mu = 5$. In each of the following cases, give the rejection region for the test in terms of the z-statistic:

a. H_a: $\mu > 5$, $\alpha = .10$

b. H_a: $\mu > 5$, $\alpha = .05$

c. H_a: $\mu < 5$, $\alpha = .01$

d. H_a: $\mu \neq 5$, $\alpha = .10$

8.9 For each of the following rejection regions, sketch the sampling distribution for z and indicate the location of the rejection region.

a. $z > 1.96$

b. $z > 1.645$

c. $z > 2.575$

d. $z < -1.28$

e. $z < -1.645$ or $z > 1.645$

f. $z < -2.575$ or $z > 2.575$

g. For each of the rejection regions specified in parts **a–f**, what is the probability that a Type I error will be made?

Ⓓ Applet Exercise 8.1

Use the applet entitled *Hypotheses Test for a Mean* to investigate the frequency of Type I and Type II errors. For this exercise, use $n = 100$ and the normal distribution with mean 50 and standard deviation 10.

a. Set the null mean equal to 50 and the alternative to *not equal*. Run the applet one time. How many times was the null hypothesis rejected at level .05? In this case, the null hypothesis is true. Which type of error occurred each time the true null hypothesis was rejected? What is the probability of rejecting a true null hypothesis at level .05? How does the proportion of times the null hypothesis was rejected compare with this probability?

b. Clear the applet, then set the null mean equal to 47, and keep the alternative at *not equal*. Run the applet one time. How many times was the null hypothesis *not* rejected at level .05? In this case, the null hypothesis is false. Which type of error occurred each time the null hypothesis was *not* rejected? Run the applet several more times without clearing. Based on your results, what can you conclude about the probability of failing to reject the null hypothesis for the given conditions?

Applying the Concepts—Basic

8.10 **Walking to improve health.** In a study investigating a link between walking and improved health (*Social Science & Medicine*, Apr. 2014), researchers reported that adults walked an average of 5.5 days in the past month for the purpose of health or recreation. Specify the null and alternative hypotheses for testing whether the true average number of days in the past month that adults walked for the purpose of health or recreation is lower than 5.5 days.

8.11 **Americans' favorite sport.** *The Harris Poll* (Dec. 2013) conducted an online survey of American adults to determine their favorite sport. Your friend believes professional (NFL) football is the favorite sport for 45% of American adults. Specify the null and alternative hypotheses for testing this belief. Be sure to identify the parameter of interest.

8.12 **Infants' listening time.** Researchers writing in *Analysis of Verbal Behavior* (Dec. 2007) reported that the mean listening time of 16-month-old infants exposed to nonmeaningful monosyllabic words (e.g., "giff," "cham," "gack") is 8 seconds. Set up the null and alternative hypotheses for testing the claim.

8.13 **Play Golf America program.** In the Play Golf America program, teaching professionals at participating golf clubs provide a free 10-minute lesson to new customers. According to the Professional Golf Association (PGA), golf facilities that participate in the program gain, on average, \$2,400 in green fees, lessons, or equipment expenditures. A teaching professional at a golf club believes that the average gain in green fees, lessons, or equipment expenditures for participating golf facilities exceeds \$2,400.
NW

a. In order to support the claim made by the teaching professional, what null and alternative hypotheses should you test?

b. Suppose you select $\alpha = .05$. Interpret this value in the words of the problem.

c. For $\alpha = .05$, specify the rejection region of a large-sample test.

8.14 **Effectiveness of online courses.** The Sloan Survey of Online Learning, "Going the Distance: Online Education in the United States, 2011," reported that 68% of college presidents believe that their online education courses are as good as or superior to courses that utilize traditional face-to-face instruction.

a. Give the null hypothesis for testing the claim made by the Sloan Survey.

b. Give the rejection region for a two-tailed test using $\alpha = .01$.

8.15 **DNA-reading tool for quick identification of species.** A biologist and a zoologist at the University of Florida were the first scientists to test the effectiveness of a high-tech handheld device designed to instantly identify the DNA of an animal species (*PLOS Biology*, Dec. 2005). They used the DNA-reading device on tissue samples collected from mollusks with brightly colored shells. The scientists discovered that the error rate of the device is less than 5 percent. Set up the null and alternative hypotheses as if you want to support the findings.
NW

Applying the Concepts—Intermediate

8.16 **Calories in school lunches.** India's Mid-Day Meal scheme mandates that high schools that are part of this scheme must serve lunches that contain at least 700 calories and 20 grams of protein. Suppose a nutritionist believes that the true mean number of calories served at lunch at all high schools that are part of this scheme is less than 700 calories.

a. Identify the parameter of interest.

b. Specify the null and alternative hypotheses for testing this claim.

c. Describe a Type I error in the words of the problem.

d. Describe a Type II error in the words of the problem.

When publishing the results of a statistical test of hypothesis in journals, case stu[illegible]ies, reports, and so on, many researchers make use of p-values. Instead of selecting α beforehand and then conducting a test, as outlined in this chapter, the researcher computes (usually with the aid of a statistical software package) and reports the value of th[e] appropriate test statistic and its associated p-value. It is left to the reader of the repo[rt] to judge the significance of the result (i.e., the reader must determine whether to reje[ct] the null hypothesis in favor of the alternative hypothesis, based on the reported p-value. Usually, the null hypothesis is rejected if the observed significance level is *less than* th[e] fixed significance level, α, chosen by the reader. The inherent advantage of reporting te[st] results in this manner is twofold: (1) Readers are permitted to select the maximum valu[e] of α that they would be willing to tolerate if they actually carried out a standard test o[f] hypothesis in the manner outlined in this chapter, and (2) a measure of the degree o[f] significance of the result (i.e., the p-value) is provided.

Ethics IN Statistics

Selecting the value of α *after* computing the observed significance level (p-value) in order to guarantee a preferred conclusion is considered *unethical statistical practice*.

Reporting Test Results as p-Values: How to Decide Whether to Reject H_0

1. Choose the maximum value of α that you are willing to tolerate.
2. If the observed significance level (p-value) of the test is less than the chosen value of α, reject the null hypothesis. Otherwise, do not reject the null hypothesis.

Note: Some statistical software packages (e.g., SPSS) will conduct only two-tailed test[s] of hypothesis. For these packages, you obtain the p-value for a one-tailed test as show[n] in the box:

Converting a Two-Tailed p-Value from a Printout to a One-Tailed p-Value

$$p = \frac{\text{Reported } p\text{-value}}{2} \quad \text{if} \begin{cases} H_a \text{ is of form } > \text{ and } z \text{ is positive} \\ H_a \text{ is of form } < \text{ and } z \text{ is negative} \end{cases}$$

$$p = 1 - \left(\frac{\text{Reported } p\text{-value}}{2}\right) \quad \text{if} \begin{cases} H_a \text{ is of form } > \text{ and } z \text{ is negative} \\ H_a \text{ is of form } < \text{ and } z \text{ is positive} \end{cases}$$

Exercises 8.22–8.30

8.22 Consider the test of $H_0: \mu = 7$. For each of the following, find the p-value of the test:

a. $H_a: \mu > 7, z = 1.20$ [redacted]
b. $H_a: \mu < 7, z = -1.20$ [redacted]
c. $H_a: \mu \neq 7, z = 1.20$ [redacted]

8.23 If a hypothesis test were conducted using $\alpha = .05$, for which of the following p-values would the null hypothesis be rejected?

a. .06 b. .10
c. .01 d. .001
e. .251 f. .042 [redacted]

8.24 For each α and observed significance level (p-value) pair, indicate whether the null hypothesis would be rejected.

a. $\alpha = .05$, p-value $= .10$ [redacted]
b. $\alpha = .10$, p-value $= .05$ [redacted]
c. $\alpha = .01$, p-value $= .001$ [redacted]
d. $\alpha = .025$, p-value $= .05$ [redacted]
e. $\alpha = .10$, p-value $= .45$ [redacted]

8.25 In a test of the hypothesis $H_0: \mu = 50$ versus $H_a: \mu > 50$, a sample of $n = 100$ observations possessed mean $\bar{x} = 49.4$ and standard deviation $s = 4.1$. Find and interpret the p-value for this test. p [illegible]

8.26 In a test of $H_0: \mu = 100$ against $H_a: \mu > 100$, the sample data yielded the test statistic $z = 2.17$. Find and interpret the p-value for the test. p = .0150

8.27 In a test of the hypothesis $H_0: \mu = 10$ versus $H_a: \mu \neq 10$, a sample of $n = 50$ observations possessed mean $\bar{x} = 10.7$ and standard deviation $s = 3.1$. Find and interpret the p-value for this test. p = .1096

8.28 NW In a test of $H_0: \mu = 100$ against $H_a: \mu \neq 100$, the sample data yielded the test statistic $z = 2.17$. Find the p-value for the test. p = .03

8.29 In a test of $H_0: \mu = 75$ performed using the computer, SPSS reports a two-tailed p-value of .1032. Make the appropriate conclusion for each of the following situations:

a. $H_a: \mu < 75, z = -1.63, \alpha = .05$ Fail to reject H_0
b. $H_a: \mu < 75, z = -1.63, \alpha = .10$ Fail to reject H_0
c. $H_a: \mu > 75, z = 1.63, \alpha = .10$ Reject H_0
d. $H_a: \mu \neq 75, z = -1.63, \alpha = .01$ Fail to reject H_0

8.30 An analyst tested the null hypothesis $\mu \geq 20$ against the alternative hypothesis that $\mu < 20$. The analyst reported a p-value of .06. What is the smallest value of α for which the null hypothesis would be rejected? .06

Exercises 8.51–8.72

Understanding the Principles

8.51 When do we use the t-test instead of the z-test in small samples?

8.52 What are the assumptions behind a t-test in testing a hypothesis about a population mean?

Learning the Mechanics

8.53 For each of the following rejection regions, sketch the sampling distribution of t and indicate the location of the rejection region on your sketch:

a. $t > 1.64$, where df = 6

b. $t < -1.872$, where df = 12

c. $t < -2.161$ or $t > 2.161$, where df = 25

8.54 For each of the rejection regions defined in Exercise 8.53, what is the probability that a Type I error will be made?

8.55 A random sample of n observations is selected from a normal population to test the null hypothesis tha $\mu = 15$. Specify the rejection region for each of the following combinations of H_a, α, and n:

a. $H_a: \mu \neq 15; \alpha = .05; n = 16$

b. $H_a: \mu > 15; \alpha = .01; n = 26$

c. $H_a: \mu > 15; \alpha = .10; n = 12$

d. $H_a: \mu < 15; \alpha = .01; n = 15$

e. $H_a: \mu \neq 15; \alpha = .10; n = 22$

f. $H_a: \mu < 15; \alpha = .05; n = 6$

8.56 The following sample of six measurements was randomly selected from a normally distributed population: 2, 5, −2, 7, 2, 4.

a. Test the null hypothesis that the mean of the population is 2 against the alternative hypothesis, $\mu < 2$. Use $\alpha = .05$.

b. Test the null hypothesis that the mean of the population is 2 against the alternative hypothesis, $\mu \neq 2$. Use $\alpha = .05$.

c. Find the observed significance level for each test.

8.57 NW A sample of five measurements, randomly selected from a normally distributed population, resulted in the following summary statistics: $\bar{x} = 4.8$, $s = 1.3$.

a. Test the null hypothesis that the mean of the population is 6 against the alternative hypothesis, $\mu < 6$. Use $\alpha = .05$.

b. Test the null hypothesis that the mean of the population is 6 against the alternative hypothesis, $\mu \neq 6$. Use $\alpha = .05$.

c. Find the observed significance level for each test.

8.58 Suppose you conduct a t-test for the null hypothesis $H_0: \mu = 1{,}500$ versus the alternative hypothesis $H_a: \mu > 1{,}500$, based on a sample of 17 observations. The test results are $t = 1.89$, p-value = .038.

a. What assumptions are necessary for the validity of this procedure?

b. Interpret the results of the test.

c. Suppose the alternative hypothesis had been the two-tailed $H_a: \mu \neq 1{,}500$. If the t-statistic were unchanged, what would the p-value be for this test? Interpret the p-value for the two-tailed test.

Applying the Concepts—Basic

8.59 D TRAPS **Lobster trap placement.** Refer to the *Bulletin of Marine Science* (Apr. 2010) observational study of lobster trap placement by teams fishing for the red spiny lobster in Baja California Sur, Mexico, Exercise 7.41 (p. 363). Trap spacing measurements (in meters) for a sample of seven teams of red spiny lobster fishermen are reproduced in the accompanying table. Let μ represent the average of the trap spacing measurements for the population of red spiny lobster fishermen fishing in Baja California Sur, Mexico. In Exercise 7.41 you computed the mean and standard deviation of the sample measurements to be $\bar{x} = 89.9$ meters and $s = 11.6$ meters, respectively. Suppose you want to determine if the true value of μ differs from 95 meters.

93	99	105	94	82	70	86

Based on Shester, G. G. "Explaining catch variation among Baja California lobster fishers through spatial analysis of trap-placement decisions." *Bulletin of Marine Science*, Vol. 86, No. 2, April 2010 (Table 1), pp. 479–498.

a. Specify the null and alternative hypotheses for this test.

b. Since $\bar{x} = 89.9$ is less than 95, a fisherman wants to reject the null hypothesis. What are the problems with using such a decision rule?

c. Compute the value of the test statistic.

d. Find the approximate p-value of the test.

e. Select a value of α, the probability of a Type I error. Interpret this value in the words of the problem.

f. Give the appropriate conclusion, based on the results of parts **d** and **e**.

g. What conditions must be satisfied for the test results to be valid?

h. In Exercise 7.41 you found a 95% confidence interval for μ. Does the interval support your conclusion in part **f**?

8.60 D PAI **Music performance anxiety.** Refer to the *British Journal of Music Education* (Mar. 2014) study of performance anxiety by music students, Exercise 7.37 (p. 362). Recall that the Performance Anxiety Inventory (PAI) was used to measure music performance anxiety on a scale from 20 to 80 points. The table below gives PAI values for participants in eight different studies. A MINITAB printout of the data analysis is shown.

54	42	51	39	41	43	55	40

Source: Patston, T. "Teaching stage fright? – Implications for music educators." *British Journal of Music Education*, Vol. 31, No. 1, Mar. 2014 (adapted from Figure 1).

One-Sample T: PAI

```
Test of mu = 40 vs > 40

                                         95% Lower
Variable  N   Mean  StDev  SE Mean       Bound      T      P
PAI       8  45.63   6.59     2.33       41.21   2.41  0.023
```

a. Set up the null and alternative hypotheses for determining whether the mean PAI value, μ, for all similar studies of music performance anxiety exceeds 40.
b. Find the rejection region for the test, part a, using $\alpha = .05$.
c. Compute the test statistic.
d. State the appropriate conclusion for the test.
e. What conditions are required for the test results to be valid?
f. Locate the p-value for the test on the MINITAB printout and use it to make a conclusion. (Your conclusion should agree with your answer in part d.)
g. How would your conclusion change if you used $\alpha = .01$?

8.61 Dental anxiety study. Refer to the *BMC Oral Health* (Vol. 9, 2009) study of adults who completed the Dental Anxiety Scale, presented in Exercise 5.37 (p. 279). Recall that scores range from 0 (no anxiety) to 25 (extreme anxiety). Summary statistics for the scores of 15 adults who completed the questionnaire are $\bar{x} = 10.7$ and $s = 3.6$. Conduct a test of hypothesis to determine whether the mean Dental Anxiety Scale score for the population of college students differs from $\mu = 11$. Use $\alpha = .05$.

8.62 Crab spiders hiding on flowers. Refer to the *Behavioral Ecology* (Jan. 2005) experiment on crab spiders' use of camouflage to hide from predators (e.g., birds) on flowers, presented in Exercise 2.42 (p. 79). Researchers at the French Museum of Natural History collected a sample of 10 adult female crab spiders, each sitting on the yellow central part of a daisy, and measured the chromatic contrast between each spider and the flower. The data (for which higher values indicate a greater contrast, and, presumably, an easier detection by predators) are shown in the accompanying table. The researchers discovered that a contrast of 70 or greater allows birds to see the spider. Of interest is whether the true mean chromatic contrast of crab spiders on daisies is less than 70.

D SPIDER

57	75	116	37	96	61	56	2	43	32

Based on Thery, M., et al. "Specific color sensitivities of prey and predator explain camouflage in different visual systems." *Behavioral Ecology*, Vol. 16, No. 1, Jan. 2005 (Table 1).

a. Define the parameter of interest, μ.
b. Set up the null and alternative hypotheses of interest.
c. Find $\bar{x}$ and s for the sample data, and then use these values to compute the test statistic.
d. Give the rejection region for $\alpha = .10$.
e. State the appropriate conclusion in the words of the problem.

8.63 Cheek teeth of extinct primates. Refer to the *American Journal of Physical Anthropology* (Vol. 142, 2010) study of the characteristics of cheek teeth (e.g., molars) in an extinct primate species, Exercise 2.38 (p. 78). Recall that the researchers recorded the dentary depth of molars (in millimeters) for a sample of 18 cheek teeth extracted from skulls. These depth measurements are listed in the accompanying table. Anthropologists know that the mean dentary depth of molars in an extinct primate species—called Species A—is 15 millimeters. Is there evidence to indicate that the sample of 18 cheek teeth come from some other extinct primate species (i.e., some species other than Species A)? Use the SPSS printout at the bottom of the page to answer the question.

D MOLARS

18.12	16.55
19.48	15.70
19.36	17.83
15.94	13.25
15.83	16.12
19.70	18.13
15.76	14.02
17.00	14.04
13.96	16.20

Based on Boyer, D. M., Evans, A. R., and Jernvall, J. "Evidence of dietary differentiation among Late Paleocene–Early Eocene Plesiadapids (Mammalia, Primates)." *American Journal of Physical Anthropology*, Vol. 142, © 2010.

8.64 Radon exposure in Egyptian tombs. Refer to the *Radiation Protection Dosimetry* (Dec. 2010) study of radon exposure in Egyptian tombs, Exercise 7.39 (p. 362). The radon levels—measured in becquerels per cubic meter (Bq/m^3)—in the inner chambers of a sample of 12 tombs are listed in the table. For the safety of the guards and visitors, the Egypt Tourism Authority (ETA) will temporarily close the tombs if the true mean level of radon exposure in the tombs rises to 6,000 Bq/m^3. Consequently, the ETA wants to conduct a test to determine if the true mean level of radon exposure in the tombs is less than 6,000 Bq/m^3, using a Type I error probability of .10. An SAS analysis of the data is shown on p. 427. Specify all the elements of the test: H_0, H_a, test statistic, p-value, α, and your conclusion.

D TOMBS

50	910	180	580	7800	4000
390	12100	3400	1300	11900	1100

SPSS Output for Exercise 8.63

One-Sample Statistics

	N	Mean	Std Deviation	Std. Error Mean
M2Depth	18	16.4994	1.97042	.46443

One-Sample Test

	Test Value = 15					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
M2Depth	3.229	17	.005	1.49944	.5196	2.4793

SAS Output for Exercise 8.64

The TTEST Procedure

Variable: RADON

N	Mean	Std Dev	Std Err	Minimum	Maximum
12	3642.5	4486.9	1295.3	50.0000	12100.0

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
3642.5	-Infty	5968.6	4486.9	3178.5	7618.3

DF	t Value	Pr < t
11	-1.82	0.0480

Applying the Concepts—Intermediate

8.65 Yield strength of steel connecting bars. To protect against earthquake damage, steel beams are typically fitted and connected with plastic hinges. However, these plastic hinges are prone to deformations and are difficult to inspect and repair. An alternative method of connecting steel beams—one that uses high-strength steel bars with clamps—was investigated in *Engineering Structures* (July 2013). Mathematical models for predicting the performance of these steel connecting bars assume the bars have a mean yield strength of 300 megapascals (MPa). To verify this assumption, the researchers conducted material property tests on the steel connecting bars. In a sample of three tests, the yield strengths were 354, 370, and 359 MPa. Do the data indicate that the true mean yield strength of the steel bars exceeds 300 MPa? Test using $\alpha = .01$.

8.66 Pitch memory of amusiacs. Amusia is a congenital disorder that adversely impacts one's perception of music. Refer to the *Advances in Cognitive Psychology* (Vol. 6, 2010) study of the pitch memory of individuals diagnosed with amusia, Exercise 7.45 (p. 363). Recall that each in a sample of 17 amusiacs listened to a series of tone pairs and was asked to determine if the tones were the same or different. In the first trial, the tones were separated by 1 second; in a second trial, the tones were separated by 5 seconds. The difference in accuracy scores for the two trials was determined for each amusiac (where the difference is the score on the first trial minus the score on the second trial). The mean score difference was .11 with a standard deviation of .19.

a. In theory, the longer the delay between tones, the less likely one is to detect a difference between the tones. Consequently, the true mean score difference should exceed 0. Set up the null and alternative hypotheses for testing the theory.

b. Carry out the test, part a, using $\alpha = .05$. Is there evidence to support the theory?

8.67 Free recall memory strategy. (D RECALL) Psychologists who study memory often use a measure of "free recall" (e.g., the number of correctly recalled items in a list of to-be-remembered items). The strategy used to memorize the list—for example, category clustering—is often just as important. Researchers at Central Michigan University developed an algorithm for computing measures of category clustering in *Advances in Cognitive Psychology* (Oct. 2012). One measure, called ratio of repetition, was recorded for a sample of 8 participants in a memory study. These ratios are listed in the table. Test the theory that the average ratio of repetition for all participants in a similar memory study differs from .5. Select an appropriate Type I error rate for your test.

.25	.43	.57	.38	.38	.60	.47	.30

Source: Senkova, O., & Otani, H. "Category clustering calculator for free recall." *Advances in Cognitive Psychology*, Vol. 8, No. 4, Oct. 2012 (Table 3).

8.68 Increasing hardness of polyester composites. Polyester resins reinforced with fiberglass are used to fabricate wall panels of restaurants. It is theorized that adding cement kiln dust (CKD) to the polyester composite will increase wall panel hardness. In a study published in *Advances in Applied Physics* (Vol. 2, 2014), hardness (joules per squared centimeters) was determined for three polyester composite mixtures that used a 40% CKD weight ratio. The hardness values were reported as 83, 84, and 79 j/cm^2. Research has shown that the mean hardness value of polyester composite mixtures that use a 20% CKD weight ratio is $\mu = 76\ j/cm^2$. In your opinion, does using a 40% CKD weight ratio increase the mean hardness value of polyester composite mixtures? Support your answer statistically.

8.69 Minimizing tractor skidding distance. (D SKID) Refer to the *Journal of Forest Engineering* (July 1999) study of minimizing tractor skidding distances along a new road in a European forest, presented in Exercise 7.48 (p. 364). The skidding distances (in meters) were measured at 20 randomly selected road sites. The data are repeated in the accompanying table. Recall that a logger working on the road claims that the mean skidding distance is at least 425 meters. Is there sufficient evidence to refute this claim? Use $\alpha = .10$.

488	350	457	199	285	409	435	574	439	546
385	295	184	261	273	400	311	312	141	425

Based on Tujek, J., and Pacola, E. "Algorithms for skidding distance modeling on a raster Digital Terrain Model." *Journal of Forest Engineering*, Vol. 10, No. 1, July 1999 (Table 1).

8.70 Dissolved organic compound in lakes. (D LAKES) The level of dissolved oxygen in the surface water of a lake is vital to maintaining the lake's ecosystem. Environmentalists from the University of Wisconsin monitored the dissolved oxygen levels over time for a sample of 25 lakes in the state (*Aquatic Biology*, May 2010). To ensure a representative sample, the environmentalists focused on several lake characteristics, including dissolved organic compound (DOC). The DOC data (measured in grams per cubic meters) for the 25 lakes are listed in the table on p. 428. The population of Wisconsin lakes has a mean DOC value of 15 grams/m^3. Use a hypothesis test (at $\alpha = .10$) to make an inference about whether the sample is representative of all Wisconsin lakes for the characteristic dissolved organic compound.

Finally, we calculate the number of standard deviations (the z value) between the sampled and hypothesized values of the binomial proportion:

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0/n}} = \frac{.086 - .2}{\sqrt{(.2)(.8)/140}} = \frac{-.114}{.034} = -3.35$$

The implication is that the observed sample proportion is (approximately) 3.35 standard deviations below the null-hypothesized proportion, .2 (Figure 8.19). Therefore, we reject the null hypothesis, concluding at the .05 level of significance that the true failure rate of the new method for detecting breast cancer differs from .20. Since $\hat{p} = .086$, it appears that the new method is better (i.e., has a smaller failure rate) than the method currently in use. (To estimate the magnitude of the failure rate for the new method, a confidence interval can be constructed.)

The test of hypothesis about a population proportion p is summarized in the next box. Note that the procedure is entirely analogous to that used for conducting large-sample tests about a population mean.

Large-Sample Test of Hypothesis about p: Normal (z) Statistic

Test statistic: $z_c = \frac{(\hat{p} - p_0)}{\sigma_{\hat{p}}} = \frac{(\hat{p} - p_0)}{\sqrt{(p_0 q_0/n)}}$

	One-Tailed Tests		Two-Tailed Test
	$H_0: p = p_0$	$H_0: p = p_0$	$H_0: p = p_0$
	$H_a: p < p_0$	$H_a: p > p_0$	$H_a: p \neq p_0$
Rejection region:	$z_c < -z_\alpha$	$z_c > z_\alpha$	$\lvert z_c \rvert > z_{\alpha/2}$
p-value:	$P(z < z_c)$	$P(z > z_c)$	$2P(z > z_c)$ if z_c is positive $2P(z < z_c)$ if z_c is negative

Decision: Reject H_0 if $\alpha >$ p-value or if test statistic (z_c) falls in rejection region where $P(z > z_\alpha) = \alpha$, $P(z > z_{\alpha/2}) = \alpha/2$, and $\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 \mid H_0 \text{ true})$.

[*Note:* The symbol for the numerical value assigned to p under the null hypothesis is p_0.]

Conditions Required for a Valid Large-Sample Hypothesis Test for p

1. A random sample is selected from a binomial population.
2. The sample size n is large. (This condition will be satisfied if both $np_0 \geq 15$ and $nq_0 \geq 15$.)

Example 8.10

A Hypothesis Test for p—Proportion of Defective Batteries

Problem The reputations (and hence sales) of many businesses can be severely damaged by shipments of manufactured items that contain a large percentage of defectives. For example, a manufacturer of alkaline batteries may want to be reasonably certain that less than 5% of its batteries are defective. Suppose 300 batteries are randomly selected from a very large shipment; each is tested and 10 defective batteries are found. Does this outcome provide sufficient evidence for the manufacturer to conclude that the fraction defective in the entire shipment is less than .05? Use $\alpha = .01$.

Solution The objective of the sampling is to determine whether there is sufficient evidence to indicate that the fraction defective, p, is less than .05. Consequently, we will test the null hypothesis that $p = .05$ against the alternative hypothesis that $p < .05$. The elements of the test are

$H_0: p = .05$ (Fraction of defective batteries equals .05.)

$H_a: p < .05$ (Fraction of defective batteries is less than .05.)

Test statistic: $z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}}$

Rejection region: $z < -z_{.01} = -2.33$ (see Figure 8.20)

Statistics IN Action (continued)

On the MINITAB printout, x represents the number of the 250 people with colds that used more than 60 tissues. Note hat $x = 154$. This value is used to compute the test statis- ic $z = 3.67$, highlighted on the printout. The p-value of the est, also highlighted on the printout, is p-value $= .000$. Since his value is less than $\alpha = .05$, there is sufficient evidence (at $\alpha = .05$) to reject H_0; we conclude that the proportion of all KLEENEX® users who use more than 60 tissues when they have a cold exceeds .5. This conclusion again supports the company's decision to put more than 60 tissues in an anti-viral box of KLEENEX.

Data Set: TISSUES

Test and CI for One Proportion: USED60

```
Test of p = 0.5 vs p > 0.5

Event = MORE

                                        95% Lower
Variable      X    N  Sample p              Bound  Z-Value  P-Value
USED60      154  250  0.616000          0.565404     3.67    0.000

Using the normal approximation.
```

Figure SIA8.2
MINITAB test of $p = .5$ for KLEENEX survey

Exercises 8.73–8.93

Un

8.73 What type of data, quantitative or qualitative, is typically associated with making inferences about a population proportion p?

8.74 What conditions are required for a valid large-sample test for p?

Le

8.75 For the binomial sample sizes and null-hypothesized values of p in each part, determine whether the sample size is large enough to use the normal approximation methodology presented in this section to conduct a test of the null hypothesis H_0: $p = p_0$.
- **a.** $n = 500$, $p_0 = .05$
- **b.** $n = 100$, $p_0 = .99$
- **c.** $n = 50$, $p_0 = .2$
- **d.** $n = 20$, $p_0 = .2$
- **e.** $n = 10$, $p_0 = .4$

8.76 NW Suppose a random sample of 100 observations from a binomial population gives a value of $\hat{p} = .69$ and you wish to test the null hypothesis that the population parameter p is equal to .75 against the alternative hypothesis that p is less than .75.
- **a.** Noting that $\hat{p} = .69$, what does your intuition tell you? Does the value of $\hat{p}$ appear to contradict the null hypothesis?
- **b.** Use the large-sample z-test to test H_0: $p = .75$ against the alternative hypothesis H_a: $p < .75$. Use $\alpha = .05$. How do the test results compare with your intuitive decision from part **a**?
- **c.** Find and interpret the observed significance level of the test you conducted in part **b**.

8.77 Suppose the sample in Exercise 8.76 has produced $\hat{p} = .84$ and we wish to test H_0: $p = .9$ against the alternative H_a: $p < .9$.
- **a.** Calculate the value of the z statistic for this test.
- **b.** Note that the numerator of the z statistic ($\hat{p} - p_0 = .84 - .90 = -.06$) is the same as for Exercise 8.76. Considering this, why is the absolute value of z for this exercise larger than that calculated in Exercise 8.76?
- **c.** Complete the test, using $\alpha = .05$, and interpret the result.
- **d.** Find the observed significance level for the test and interpret its value.

8.78 A random sample of 100 observations is selected from a binomial population with unknown probability of success p. The computed value of $\hat{p}$ is equal to .74.
- **a.** Test H_0: $p = .65$ against H_a: $p > .65$. Use $\alpha = .01$.
- **b.** Test H_0: $p = .65$ against H_a: $p > .65$. Use $\alpha = .10$.
- **c.** Test H_0: $p = .90$ against H_a: $p \neq .90$. Use $\alpha = .05$.
- **d.** Form a 95% confidence interval for p.
- **e.** Form a 99% confidence interval for p.

8.79 Refer to Exercise 7.58 (p. 342), in which 50 consumers taste-tested a new snack food.
SNACK
- **a.** Test H_0: $p = .5$ against H_a: $p > .5$, where p is the proportion of customers who do not like the snack food. Use $\alpha = .10$.
- **b.** Report the observed significance level of your test.

Use the applet entitled *Hypotheses Test for a Proportion* to investigate the relationships between the probabilities of Type I and Type II errors occurring at levels .05 and .01. For this exercise, use $n = 100$, true $p = 0.5$, and alternative *not equal*.
- **a.** Set null $p = .5$. What happens to the proportion of times the null hypothesis is rejected at the .05 level and at the .01 level as the applet is run more and more times? What type of error has occurred when the null hypothesis is rejected in this situation? Based on your results, is this type of error more likely to occur at level .05 or at level .01? Explain.

b. Set null $p = .6$. What happens to the proportion of times the null hypothesis is *not* rejected at the .05 level and at the .01 level as the applet is run more and more times? What type of error has occurred when the null hypothesis is *not* rejected in this situation? Based on your results, is this type of error more likely to occur at level .05 or at level .01? Explain.

c. Use your results from parts **a** and **b** to make a general statement about the probabilities of Type I and Type II errors at levels .05 and .01.

Applet Exercise 8.6

Use the applet entitled *Hypotheses Test for a Proportion* to investigate the effect of the true population proportion p on the probability of a Type I error occurring. For this exercise, use $n = 100$ and alternative *not equal*.

a. Set true $p = .5$ and null $p = .5$. Run the applet several times, and record the proportion of times the null hypothesis is rejected at the .01 level.

b. Clear the applet and repeat part **a** for true $p = .1$ and null $p = .1$. Then repeat one more time for true $p = .01$ and null $p = .01$.

c. Based on your results from parts **a** and **b**, what can you conclude about the probability of a Type I error occurring as the true population proportion gets closer to 0?

Applying the Concepts—Basic

8.80 Paying for music downloads. If you use the Internet, have you ever paid to access or download music? This was one of the questions of interest in a recent *Pew Internet and American Life Project Survey* (Oct. 2010). In a representative sample of 755 adults who use the Internet, 506 stated that they have paid to download music. Let p represent the true proportion of all Internet-using adults who have paid to download music.

MUSIC

a. Compute a point estimate of p.

b. Set up the null and alternative hypotheses for testing whether the true proportion of all Internet-using adults who have paid to download music exceeds .7.

c. Compute the test statistic for part **b**.

d. Find the rejection region for the test if $\alpha = .01$.

e. Find the p-value for the test.

f. Make the appropriate conclusion using the rejection region.

g. Make the appropriate conclusion using the p-value.

8.81 Underwater sound-locating ability of alligators. Alligators have shown the ability to determine the direction of an airborne sound. But can they locate underwater sounds? This was the subject of research published in the *Journal of Herpetology* (Dec. 2014). Alligators inhabiting the flood control canals in the Florida Everglades were monitored for movement toward a sound produced from a submerged diving bell. Movements within a 180° arc of the direction toward the sound were scored as movements toward the sound; all movements in other directions were scored as movements away from the sound. Consequently, the researchers assumed that the proportion of movements toward the sound expected by chance is $180°/360° = .5$. In a sample of $n = 50$ alligators, 42 moved toward the underwater sound.

a. Give the null and alternative hypotheses for testing whether the true proportion of alligators that move toward the underwater sound is higher than expected by chance.

b. In a sample of $n = 50$ alligators, assume that 42 moved toward the underwater sound. Use this information to compute an estimate of the true proportion of alligators that move toward the underwater sound.

c. Compute the test statistic for this study.

d. Compute the observed significance level (p-value) of the test.

e. Make the appropriate conclusion in the words of the problem.

8.82 Gummy bears: red or yellow? *Chance* (Winter 2010) presented a lesson in hypothesis testing carried out by medical students in a biostatistics class. Students were blind-folded and then given a red-colored or yellow-colored gummy bear to chew. (Half the students were randomly assigned to receive the red gummy bear and half to receive the yellow bear. The students could not see what color gummy bear they were given.) After chewing, the students were asked to guess the color of the candy based on the flavor. Of the 121 students who participated in the study, 97 correctly identified the color of the gummy bear.

a. If there is no relationship between color and gummy bear flavor, what proportion of the population of students will correctly identify the color?

b. Specify the null and alternative hypotheses for testing whether color and flavor are related.

c. Carry out the test and give the appropriate conclusion at $\alpha = .01$. Use the p-value of the test to make your decision.

8.83 Dehorning of dairy calves. For safety reasons, calf dehorning has become a routine practice at dairy farms. A report by Europe's Standing Committee on the Food Chain and Animal Health (SANKO) stated that 80% of European dairy farms carry out calf dehorning. A later study, published in the *Journal of Dairy Science* (Vol. 94, 2011), found that in a sample of 639 Italian dairy farms, 515 dehorn calves. Does the *Journal of Dairy Science* study support or refute the figure reported by SANKO? Explain.

8.84 Teenagers' use of emoticons in school writing. Refer to the *Pew Internet and American Life Project* (Apr. 2008) survey of the writing habits of U.S. teenagers, Exercise 7.69 (p. 371). Recall that in a random sample of 700 teenagers, 448 admitted to using at least one informal element in school writing assignments. [*Note*: Emoticons, such as using the symbol ":)" to represent a smile, and abbreviations, such as writing "LOL" for "laughing out loud," are considered informal elements.] Is there evidence to indicate that less than 65% of all U.S. teenagers have used at least one informal element in school writing assignments? Test the relevant hypotheses using $\alpha = .05$.

8.85 TV subscription streaming. "Streaming" of television programs is trending upward. According to *The Harris Poll* (Aug. 26, 2013), over one-third of Americans qualify as "subscription streamers," i.e., those who watch streamed TV programs through a subscription service such as Netflix, Hulu Plus, or Amazon Prime. The poll included 2,242 adult TV viewers, of which 785 are subscription streamers. Based on this result, can you conclude that the true fraction of adult TV viewers who are subscription streamers differs from one-third? Carry out the test using a Type I error rate of .10.