c Change the real value of p to .2 and simulate at least 200 tests. Click the button "Show Summary." Does anything look wrong?

10.17 A survey published in the *American Journal of Sports Medicine*[2] reported the number of meters (m) per week swum by two groups of swimmers—those who competed exclusively in breaststroke and those who competed in the individual medley (which includes breaststroke). The number of meters per week practicing the breaststroke was recorded for each swimmer, and the summary statistics are given below. Is there sufficient evidence to indicate that the average number of meters per week spent practicing breaststroke is greater for exclusive breaststrokers than it is for those swimming individual medley?

	Specialty	
	Exclusively Breaststroke	Individual Medley
Sample size	130	80
Sample mean (m)	9017	5853
Sample standard deviation (m)	7162	1961
Population mean	μ_1	μ_2

a State the null and alternative hypotheses.

b What is the appropriate rejection region for an $\alpha = .01$ level test?

c Calculate the observed value of the appropriate test statistic.

d What is your conclusion?

e What is a practical reason for the conclusion you reached in part (d)?

10.18 The hourly wages in a particular industry are normally distributed with mean \$13.20 and standard deviation \$2.50. A company in this industry employs 40 workers, paying them an average of \$12.20 per hour. Can this company be accused of paying substandard wages? Use an $\alpha = .01$ level test.

10.19 The output voltage for an electric circuit is specified to be 130. A sample of 40 independent readings on the voltage for this circuit gave a sample mean 128.6 and standard deviation 2.1. Test the hypothesis that the average output voltage is 130 against the alternative that it is less than 130. Use a test with level .05.

10.20 The Rockwell hardness index for steel is determined by pressing a diamond point into the steel and measuring the depth of penetration. For 50 specimens of an alloy of steel, the Rockwell hardness index averaged 62 with standard deviation 8. The manufacturer claims that this alloy has an average hardness index of at least 64. Is there sufficient evidence to refute the manufacturer's claim at the 1% significance level?

10.21 Shear strength measurements derived from unconfined compression tests for two types of soils gave the results shown in the following table (measurements in tons per square foot). Do the soils appear to differ with respect to average shear strength, at the 1% significance level?

Soil Type I	Soil Type II
$n_1 = 30$	$n_2 = 35$
$\overline{y}_1 = 1.65$	$\overline{y}_2 = 1.43$
$s_1 = 0.26$	$s_2 = 0.22$

2. *Source:* Kurt Grote, T. L. Lincoln, and J. G. Gamble, "Hip Adductor Injury in Competitive Swimmers," *American Journal of Sports Medicine* 32(1) (2004): 104.

stated that they always voted in presidential elections. Do the results of this sample provide sufficient evident to indicate that the percentage of adults who say that they always vote in presidential elections is different than the percentage reported in *American Demographics*? Test using $\alpha = .01$.

10.26 According to the *Washington Post*, nearly 45% of all Americans are born with brown eyes, although their eyes don't necessarily stay brown.[6] A random sample of 80 adults found 32 with brown eyes. Is there sufficient evidence at the .01 level to indicate that the proportion of brown-eyed adults differs from the proportion of Americans who are born with brown eyes?

10.27 The state of California is working very hard to ensure that all elementary age students whose native language is not English become proficient in English by the sixth grade. Their progress is monitored each year using the California English Language Development test. The results for two school districts in southern California for the 2003 school year are given in the accompanying table.[7] Do the data indicate a significant difference in the 2003 proportions of students who are fluent in English for the two districts? Use $\alpha = .01$.

District	Riverside	Palm Springs
Number of students tested	6124	5512
Percentage fluent	40	37

10.28 The commercialism of the U.S. space program has been a topic of great interest since Dennis Tito paid \$20 million to ride along with the Russian cosmonauts on the space shuttle.[8] In a survey of 500 men and 500 women, 20% of the men and 26% of the women responded that space should remain commercial free.

a Does statistically significant evidence exist to suggest that there is a difference in the population proportions of men and women who think that space should remain commercial free? Use a .05 level test.

b Why is a statistically significant difference in these population proportions of *practical* importance to advertisers?

10.29 A manufacturer of automatic washers offers a model in one of three colors: A, B, or C. Of the first 1000 washers sold, 400 were of color A. Would you conclude that customers have a preference for color A? Justify your answer.

10.30 A manufacturer claimed that at least 20% of the public preferred her product. A sample of 100 persons is taken to check her claim. With $\alpha = .05$, how small would the sample percentage need to be before the claim could legitimately be refuted? (Notice that this would involve a one-tailed test of the hypothesis.)

10.31 What conditions must be met for the Z test to be used to test a hypothesis concerning a population mean μ?

10.32 In March 2001, a Gallup poll asked, "How would you rate the overall quality of the environment in this country today—as excellent, good, fair or poor?" Of 1060 adults nationwide, 46% gave a rating of excellent or good. Is this convincing evidence that a majority of the nation's adults think the quality of the environment is fair or poor? Test using $\alpha = .05$.

6. *Source:* "Seeing the World Through Tinted Lenses," *Washington Post*, March 16, 1993, p. 5.

7. *Source:* Cadonna Peyton, "Pupils Build English Skills," *Press-Enterprise* (Riverside, Calif.), March 19, 2004, p. B-1.

8. *Source:* Adapted from "Toplines: To the Moon?" *American Demographics*, August 2001, p. 9.

EXAMPLE **10.9** Suppose that the vice president of Example 10.5 wants to test $H_0 : \mu = 15$ against $H_a : \mu = 16$ with $\alpha = \beta = .05$. Find the sample size that will ensure this accuracy. Assume that σ^2 is approximately 9.

Solution Because $\alpha = \beta = .05$, it follows that $z_\alpha = z_\beta = z_{.05} = 1.645$. Then

$$n = \frac{(z_\alpha + z_\beta)^2\sigma^2}{(\mu_a - \mu_0)^2} = \frac{(1.645 + 1.645)^2(9)}{(16 - 15)^2} = 97.4.$$

Hence, $n = 98$ observations should be used to meet the requirement that $\alpha \approx \beta \approx .05$ for the vice president's test. ■

Exercises

10.37 Refer to Exercise 10.19. If the voltage falls as low as 128, serious consequences may result. For testing $H_0 : \mu = 130$ versus $H_a : \mu = 128$, find the probability of a type II error, β, for the rejection region used in Exercise 10.19.

10.38 Refer to Exercise 10.20. The steel is sufficiently hard to meet usage requirements if the mean Rockwell hardness measure does not drop below 60. Using the rejection region found in Exercise 10.20, find β for the specific alternative $\mu_a = 60$.

10.39 Refer to Exercise 10.30. Calculate the value of β for the alternative $p_a = .15$.

10.40 Refer to Exercise 10.33. The political researcher should have designed a test for which β is tolerably low when p_1 exceeds p_2 by a meaningful amount. For example, find a common sample size n for a test with $\alpha = .05$ and $\beta \leq .20$ when in fact p_1 exceeds p_2 by .1. [*Hint:* The maximum value of $p(1 - p)$ is .25.]

10.41 Refer to Exercise 10.34. Using the rejection region found there, calculate β when $\mu_a = 5.5$.

10.42 In Exercises 10.34 and 10.41, how large should the sample size be if we require that $\alpha = .01$ and $\beta = .05$ when $\mu_a = 5.5$?

10.43 A random sample of 37 second graders who participated in sports had manual dexterity scores with mean 32.19 and standard deviation 4.34. An independent sample of 37 second graders who did not participate in sports had manual dexterity scores with mean 31.68 and standard deviation 4.56.

a Test to see whether sufficient evidence exists to indicate that second graders who participate in sports have a higher mean dexterity score. Use $\alpha = .05$.

b For the rejection region used in part (a), calculate β when $\mu_1 - \mu_2 = 3$.

10.44 Refer to Exercise 10.43. Find the sample sizes that give $\alpha = .05$ and $\beta = .05$ when $\mu_1 - \mu_2 = 3$. (Assume equal-size samples for each group.)

refrain from accepting a single θ value as being *the* true value. Additional comments regarding hypothesis testing are contained in Section 10.7.

Our previous discussion focused on the duality between two-sided confidence intervals and two-sided hypothesis tests. In the exercises that follow this section, you will be asked to demonstrate the correspondence between large-sample, one-sided hypothesis tests of level α and the construction of the appropriate upper or lower bounds with confidence coefficients $1-\alpha$. If you desire an α-level test of $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$ (an upper-tail test), you should accept the alternative hypothesis if θ_0 is less than a $100(1-\alpha)\%$ *lower confidence bound* for θ. If the appropriate alternative hypothesis is $H_a : \theta < \theta_0$ (a lower-tail test), you should reject $H_0 : \theta = \theta_0$ in favor of H_a if θ_0 is larger than a $100(1-\alpha)\%$ *upper confidence bound* for θ.

Exercises

10.45 Refer to Exercise 10.21. Construct a 99% confidence interval for the difference in mean shear strengths for the two soil types.

- **a** Is the value $\mu_1 - \mu_2 = 0$ inside or outside this interval?
- **b** Based on the interval, should the null hypothesis discussed in Exercise 10.21 be rejected? Why?
- **c** How does the conclusion that you reached compare with your conclusion in Exercise 10.21?

10.46 A large-sample α-level test of hypothesis for $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$ rejects the null hypothesis if

$$\frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} > z_\alpha.$$

Show that this is equivalent to rejecting H_0 if θ_0 is less than the large-sample $100(1-\alpha)\%$ lower confidence bound for θ.

10.47 Refer to Exercise 10.32. Construct a 95% lower confidence bound for the proportion of the nation's adults who think the quality of the environment is fair or poor.

- **a** How does the value $p = .50$ compare to this lower bound?
- **b** Based on the lower bound in part (a), should the alternative hypothesis of Exercise 10.32 be accepted?
- **c** Is there any conflict between the answer in part (b) and your answer to Exercise 10.32?

10.48 A large-sample α-level test of hypothesis for $H_0 : \theta = \theta_0$ versus $H_a : \theta < \theta_0$ rejects the null hypothesis if

$$\frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} < -z_\alpha.$$

Show that this is equivalent to rejecting H_0 if θ_0 is greater than the large-sample $100(1-\alpha)\%$ upper confidence bound for θ.

10.49 Refer to Exercise 10.19. Construct a 95% upper confidence bound for the average voltage reading.

a Do the data provide sufficient evidence to indicate that the decrease in the mean FVC over the decade for the men on the physical fitness program is less than 3.8 dl? Find the attained significance level for the test.

b Refer to part (a). If you choose $\alpha = .05$, do the data support the contention that the mean decrease in FVC is less than 3.8 dl?

c Test to determine whether the FVC drop for women on the physical fitness program was less than 3.1 dl for the decade. Find the attained significance level for the test.

d Refer to part (c). If you choose $\alpha = .05$, do the data support the contention that the mean decrease in FVC is less than 3.1 dl?

10.54 Do you believe that an exceptionally high percentage of the executives of large corporations are right-handed? Although 85% of the general public is right-handed, a survey of 300 chief executive officers of large corporations found that 96% were right-handed.

a Is this difference in percentages statistically significant? Test using $\alpha = .01$.

b Find the p-value for the test and explain what it means.

10.55 A check-cashing service found that approximately 5% of all checks submitted to the service were bad. After instituting a check-verification system to reduce its losses, the service found that only 45 checks were bad in a random sample of 1124 that were cashed. Does sufficient evidence exist to affirm that the check-verification system reduced the proportion of bad checks? What attained significance level is associated with the test? What would you conclude at the $\alpha = .01$ level?

10.56 A pharmaceutical company conducted an experiment to compare the mean times (in days) necessary to recover from the effects and complications that follow the onset of the common cold. This experiment compared persons on a daily dose of 500 milligrams (mg) of vitamin C to those who were not given a vitamin supplement. For each treatment category, 35 adults were randomly selected, and the mean recovery times and standard deviations for the two groups were found to be as given in the accompanying table.

	Treatment	
	No Supplement	500 mg Vitamin C
Sample size	35	35
Sample mean	6.9	5.8
Sample standard deviation	2.9	1.2

a Do the data indicate that the use of vitamin C reduces the mean time required to recover? Find the attained significance level.

b What would the company conclude at the $\alpha = .05$ level?

10.57 A publisher of a newsmagazine has found through past experience that 60% of subscribers renew their subscriptions. In a random sample of 200 subscribers, 108 indicated that they planned to renew their subscriptions. What is the p-value associated with the test that the current rate of renewals differs from the rate previously experienced?

10.58 In a study to assess various effects of using a female model in automobile advertising, each of 100 male subjects was shown photographs of two automobiles matched for price, color, and size but of different makes. Fifty of the subjects (group A) were shown automobile 1 with a female model and automobile 2 with no model. Both automobiles were shown without the model to the other 50 subjects (group B). In group A, automobile 1 (shown with the model) was judged to be more expensive by 37 subjects. In group B, automobile 1 was judged to be (or less), well under the legal 1000-pound limit, and desires σ to be less than 40. Tests were run on a random sample of $n = 40$ helmets, and the sample mean and variance were found to be equal to 825 pounds and 2350 pounds^2, respectively.

a If $\mu = 800$ and $\sigma = 40$, is it likely that any helmet subjected to the standard external force will transmit a force to a wearer in excess of 1000 pounds? Explain.

b Do the data provide sufficient evidence to indicate that when subjected to the standard external force, the helmets transmit a mean force exceeding 800 pounds?

c Do the data provide sufficient evidence to indicate that σ exceeds 40?

10.79 The manufacturer of a machine to package soap powder claimed that her machine could load cartons at a given weight with a range of no more than .4 ounce. The mean and variance of a sample of eight 3-pound boxes were found to equal 3.1 and .018, respectively. Test the hypothesis that the variance of the population of weight measurements is $\sigma^2 = .01$ against the alternative that $\sigma^2 > .01$.

a Use an $\alpha = .05$ level of significance. What assumptions are required for this test?

b What can be said about the attained significance level using a table in the appendix?

c **Applet Exercise** What can be said about the attained significance level using the appropriate applet?

10.80 Under what assumptions may the F distribution be used in making inferences about the ratio of population variances?

10.81 From two normal populations with respective variances σ_1^2 and σ_2^2, we observe independent sample variances S_1^2 and S_2^2, with corresponding degrees of freedom $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$. We wish to test $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_a : \sigma_1^2 \neq \sigma_2^2$.

a Show that the rejection region given by

$$\left\{F > F_{\nu_2,\alpha/2}^{\nu_1} \quad \text{or} \quad F < \left(F_{\nu_1,\alpha/2}^{\nu_2}\right)^{-1}\right\},$$

where $F = S_1^2/S_2^2$, is the same as the rejection region given by

$$\left\{S_1^2/S_2^2 > F_{\nu_2,\alpha/2}^{\nu_1} \quad \text{or} \quad S_2^2/S_1^2 > F_{\nu_1,\alpha/2}^{\nu_2}\right\}.$$

b Let S_L^2 denote the larger of S_1^2 and S_2^2 and let S_S^2 denote the smaller of S_1^2 and S_2^2. Let ν_L and ν_S denote the degrees of freedom associated with S_L^2 and S_S^2, respectively. Use part (a) to show that, under H_0,

$$P\left(S_L^2/S_S^2 > F_{\nu_S,\alpha/2}^{\nu_L}\right) = \alpha.$$

Notice that this gives an equivalent method for testing the equality of two variances.

10.82 Exercises 8.83 and 10.73 presented some data collected in a 1993 study by Susan Beckham and her colleagues. In this study, measurements of anterior compartment pressure (in millimeters of mercury) were taken for ten healthy runners and ten healthy cyclists. The researchers also obtained pressure measurements for the runners and cyclists at maximal O_2 consumption. The data summary is given in the accompanying table.

hypotheses), how do we proceed? Suppose that we use the method illustrated in Example 10.23 to find a uniformly most powerful test for $H_0' : \theta = \theta_0$ versus $H_a : \theta > \theta_0$. If θ_1 is a fixed value of θ that is less than θ_0 and we use the same test for $H_0'' : \theta = \theta_1$ versus H_a, typically, α will decrease and power(θ_a) will remain unchanged for all θ_a in H_a. In other words, if we have a good test for discriminating between H_0' and H_a, the same test will be even better for discriminating between H_0'' and H_a. For tests with composite null hypotheses of the form $H_0 : \theta \leq \theta_0$ (or $H_0 : \theta \geq \theta_0$), we define the significance level α to be the probability of a type I error when $\theta = \theta_0$; that is, $\alpha =$ power(θ_0). Generally, this value for α is the maximum value of the power function for $\theta \leq \theta_0$ (or $\theta \geq \theta_0$). Using this methodology, we can show that the test derived in Example 10.23 for testing $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$ is also the uniformly most powerful α-level test for testing $H_0 : \theta \leq \theta_0$ versus $H_a : \theta > \theta_0$.

In Example 10.23, we derived the uniformly most powerful test for $H_0 : \mu = \mu_0$ versus $H_a : \mu > \mu_0$ and found it to have rejection region $\{\overline{y} > \mu_0 + z_\alpha \sigma/\sqrt{n}\}$. If we wished to test $H_0 : \mu = \mu_0$ versus $H_a : \mu < \mu_0$, analogous calculations would lead us to $\{\overline{y} < \mu_0 - z_\alpha \sigma/\sqrt{n}\}$ as the rejection region for the test that is uniformly most powerful for all $\mu_a < \mu_0$. Therefore, if we wish to test $H_0 : \mu = \mu_0$ versus $H_a : \mu \neq \mu_0$, no single rejection region yields the most powerful test for all values of $\mu_a \neq \mu_0$. Although there are some special exceptions, in most instances there do not exist uniformly most powerful two-tailed tests. Thus, there are many null and alternative hypotheses for which uniformly most powerful tests do not exist.

The Neyman–Pearson lemma is useless if we wish to test a hypothesis about a single parameter θ when the sampled distribution contains other unspecified parameters. For example, we might want to test $H_0 : \mu = \mu_0$ when the sample is taken from a normal distribution with unknown variance σ^2. In this case, $H_0 : \mu = \mu_0$ does not uniquely determine the form of the distribution (since σ^2 could be any nonnegative number), and it is therefore *not* a simple hypothesis. The next section presents a very general and widely used method for developing tests of hypotheses. The method is particularly useful when unspecified parameters (called *nuisance parameters*) are present.

Exercises

10.88 Refer to Exercise 10.2. Find the power of the test for each alternative in (a)–(d).

- **a** $p = .4$.
- **b** $p = .5$.
- **c** $p = .6$.
- **d** $p = .7$.
- **e** Sketch a graph of the power function.

10.89 Refer to Exercise 10.5. Find the power of test 1 for each alternative in (a)–(e).

- **a** $\theta = .1$.
- **b** $\theta = .4$.
- **c** $\theta = .7$.
- **d** $\theta = 1$.
- **e** Sketch a graph of the power function.