- **Normal/Large Sample:** The population distribution is Normal *or* the sample size is large ($n \geq 30$). When the sample size is small ($n < 30$), examine a graph of the sample data for any possible departures from Normality in the population. You should be safe using a t distribution as long as there is no strong skewness and no outliers are present.

- The **one-sample t test for a mean** uses the test statistic

$$t = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}}$$

with P-values calculated from the t distribution with $n - 1$ degrees of freedom.

- Confidence intervals provide additional information that significance tests do not—namely, a set of plausible values for the parameter μ. A two-sided test of $H_0: \mu = \mu_0$ at significance level α gives the same conclusion as a $100(1 - \alpha)\%$ confidence interval for μ.
- Analyze **paired data** by first taking the difference within each pair to produce a single sample. Then use one-sample t procedures.
- There are three factors that influence the sample size required for a statistical test: significance level, effect size, and the desired power of the test.
- Very small differences can be highly significant (small P-value) when a test is based on a large sample. A statistically significant difference need not be practically important.
- Many tests run at once will probably produce some significant results by chance alone, even if all the null hypotheses are true.

9.3 TECHNOLOGY CORNERS

TI-Nspire Instructions in Appendix B; HP Prime instructions on the book's Web site.

19. Computing P-values from t distributions on the calculator page 579
20. One-sample t test for a mean on the calculator page 582

Section 9.3 Exercises

65. **Attitudes** The Survey of Study Habits and Attitudes (SSHA) is a psychological test that measures students' attitudes toward school and study habits. Scores range from 0 to 200. Higher scores indicate more positive attitudes. The mean score for U.S. college students is about 115. A teacher suspects that older students have better attitudes toward school. She gives the SSHA to an SRS of 45 of the over 1000 students at her college who are at least 30 years of age. Check the conditions for carrying out a significance test of the teacher's suspicion.

66. **Anemia** Hemoglobin is a protein in red blood cells that carries oxygen from the lungs to body tissues. People with fewer than 12 grams of hemoglobin per deciliter of blood (g/dl) are anemic. A public health official in Jordan suspects that Jordanian children are at risk of anemia. He measures a random sample of

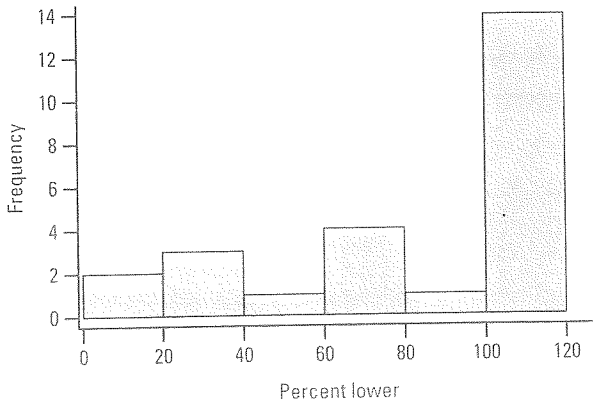
50 children. Check the conditions for carrying out a significance test of the official's suspicion.

67. **Ancient air** The composition of the earth's atmosphere may have changed over time. To try to discover the nature of the atmosphere long ago, we can examine the gas in bubbles inside ancient amber. Amber is tree resin that has hardened and been trapped in rocks. The gas in bubbles within amber should be a sample of the atmosphere at the time the amber was formed. Measurements on 9 specimens of amber from the late Cretaceous era (75 to 95 million years ago) give these percents of nitrogen:[20]

63.4 65.0 64.4 63.3 54.8 64.5 60.8 49.1 51.0

Explain why we should not carry out a one-sample t test in this setting.

68. **Paying high prices?** A retailer entered into an exclusive agreement with a supplier who guaranteed to provide all products at competitive prices. The retailer eventually began to purchase supplies from other vendors who offered better prices. The original supplier filed a lawsuit claiming violation of the agreement. In defense, the retailer had an audit performed on a random sample of 25 invoices. For each audited invoice, all purchases made from other suppliers were examined and compared with those offered by the original supplier. The percent of purchases on each invoice for which an alternative supplier offered a lower price than the original supplier was recorded.[21] For example, a data value of 38 means that the price would be lower with a different supplier for 38% of the items on the invoice. A histogram and some computer output for these data are shown below. Explain why we should not carry out a one-sample t test in this setting.



Summary statistics

Column	n	Mean	Std. Dev.	Std. Err.	Median	Min	Max	Q1	Q3
pctlower	25	77.76	32.6768	6.5353603	100	0	100	68	100

69. **Attitudes** In the study of older students' attitudes from Exercise 65, the sample mean SSHA score was 125.7 and the sample standard deviation was 29.8.

(a) Calculate the test statistic.

(b) Find the P-value using Table B. Then obtain a more precise P-value from your calculator.

70. **Anemia** For the study of Jordanian children in Exercise 66, the sample mean hemoglobin level was 11.3 mg/dl and the sample standard deviation was 1.6 mg/dl.

(a) Calculate the test statistic.

(b) Find the P-value using Table B. Then obtain a more precise P-value from your calculator.

71. **One-sided test** Suppose you carry out a significance test of $H_0: \mu = 5$ versus $H_a: \mu < 5$ based on a sample of size $n = 20$ and obtain $t = -1.81$.

(a) Find the P-value for this test using Table B or technology. What conclusion would you draw at the 5% significance level? At the 1% significance level?

(b) Redo part (a) using an alternative hypothesis of $H_a: \mu \neq 5$.

72. **Two-sided test** The one-sample t statistic from a sample of $n = 25$ observations for the two-sided test of

$$H_0: \mu = 64$$
$$H_a: \mu \neq 64$$

has the value $t = -1.12$.

(a) Find the P-value for this test using Table B or technology. What conclusion would you draw at the 5% significance level? At the 1% significance level?

(b) Redo part (a) using an alternative hypothesis of $H_a: \mu < 64$.

73. **Construction zones** Every road has one at some point—construction zones that have much lower speed limits. To see if drivers obey these lower speed limits, a police officer uses a radar gun to measure the speed (in miles per hours, or mph) of a random sample of 10 drivers in a 25 mph construction zone. Here are the data:

pg 580

27 33 32 21 30 30 29 25 27 34

(a) Is there convincing evidence that the average speed of drivers in this construction zone is greater than the posted speed limit?

(b) Given your conclusion in part (a), which kind of mistake—a Type I error or a Type II error—could you have made? Explain what this mistake would mean in context.

74. **Heat through the glass** How well materials conduct heat matters when designing houses, for example. Conductivity is measured in terms of watts of heat

power transmitted per square meter of surface per degree Celsius of temperature difference on the two sides of the material. In these units, glass has conductivity about 1. The National Institute of Standards and Technology provides exact data on properties of materials. Here are measurements of the heat conductivity of 11 randomly selected pieces of a particular type of glass:[22]

1.11 1.07 1.11 1.07 1.12 1.08 1.08 1.18 1.18 1.18 1.12

(a) Is there convincing evidence that the mean conductivity of this type of glass is greater than 1?

(b) Given your conclusion in part (a), which kind of mistake—a Type I error or a Type II error—could you have made? Explain what this mistake would mean in context.

75. **Healthy bones** The recommended daily allowance (RDA) of calcium for women between the ages of 18 and 24 years is 1200 milligrams (mg). Researchers who were involved in a large-scale study of women's bone health suspected that their participants had significantly lower calcium intakes than the RDA. To test this suspicion, the researchers measured the daily calcium intake of a random sample of 36 women from the study who fell in the desired age range. The Minitab output below displays the results of a significance test.

One-Sample T: Calcium intake (mg)

Test of mu = 1200 vs < 1200

Variable	N	Mean	StDev	SE Mean	T	P
Calcium	36	856.2	306.7	51.1	−6.73	0.000

(a) Do these data give convincing evidence to support the researchers' suspicion? Justify your answer.

(b) Interpret the *P*-value in context.

76. **Taking stock** An investor with a stock portfolio worth several hundred thousand dollars sued his broker due to the low returns he got from the portfolio at a time when the stock market did well overall. The investor's lawyer wants to compare the broker's performance against the market as a whole. He collects data on the broker's returns for a random sample of 36 weeks. Over the 10-year period that the broker has managed portfolios, stocks in the Standard & Poor's 500 index gained an average of 0.95% per week. The Minitab output below displays the results of a significance test.

One-Sample T: Return (percent)

Test of mu = 0.95 vs < 0.95

Variable	N	Mean	StDev	SE Mean	T	P
Return (percent)	36	−1.441	4.810	0.802	−2.98	0.003

(a) Do these data give convincing evidence to support the lawyer's case? Justify your answer.

(b) Interpret the *P*-value in context.

77. **Pressing pills** A drug manufacturer forms tablets by compressing a granular material that contains the active ingredient and various fillers. The hardness of a sample from each batch of tablets produced is measured to control the compression process. The target value for the hardness is $\mu = 11.5$. The hardness data for a random sample of 20 tablets are

pg 583

11.627	11.613	11.493	11.602	11.360
11.374	11.592	11.458	11.552	11.463
11.383	11.715	11.485	11.509	11.429
11.477	11.570	11.623	11.472	11.531

Is there convincing evidence at the 5% level that the mean hardness of the tablets differs from the target value?

78. **Filling cola bottles** Bottles of a popular cola are supposed to contain 300 milliliters (ml) of cola. There is some variation from bottle to bottle because the filling machinery is not perfectly precise. An inspector measures the contents of six randomly selected bottles from a single day's production. The results are

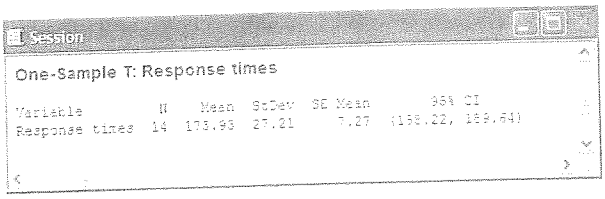
299.4 297.7 301.0 298.9 300.2 297.0

Do these data provide convincing evidence that the mean amount of cola in all the bottles filled that day differs from the target value of 300 ml?

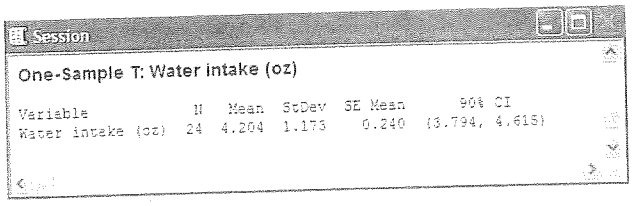
79. **Pressing pills** Refer to Exercise 77. Construct and interpret a 95% confidence interval for the population mean μ. What additional information does the confidence interval provide?

80. **Filling cola bottles** Refer to Exercise 78. Construct and interpret a 95% confidence interval for the population mean μ. What additional information does the confidence interval provide?

81. **Fast connection?** How long does it take for a chunk of information to travel from one server to another and back on the Internet? According to the site internettrafficreport.com, a typical response time is 200 milliseconds (about one-fifth of a second). Researchers collected data on response times of a random sample of 14 servers in Europe. A graph of the data reveals no strong skewness or outliers. The following figure displays Minitab output for a one-sample *t* interval for the population mean. Is there convincing evidence at the 5% significance level that the site's claim is incorrect? Justify your answer.



82. **Water!** A blogger claims that U.S. adults drink an average of five 8-ounce glasses of water per day. Skeptical researchers ask a random sample of 24 U.S. adults about their daily water intake. A graph of the data shows a roughly symmetric shape with no outliers. The figure below displays Minitab output for a one-sample t interval for the population mean. Is there convincing evidence at the 10% significance level that the blogger's claim is incorrect? Justify your answer.



83. **Tests and CIs** The P-value for a two-sided test of the null hypothesis $H_0: \mu = 10$ is 0.06.

(a) Does the 95% confidence interval for μ include 10? Why or why not?

(b) Does the 90% confidence interval for μ include 10? Why or why not?

84. **Tests and CIs** The P-value for a two-sided test of the null hypothesis $H_0: \mu = 15$ is 0.03.

(a) Does the 99% confidence interval for μ include 15? Why or why not?

(b) Does the 95% confidence interval for μ include 15? Why or why not?

85. **Right versus left** The design of controls and instruments affects how easily people can use them. A student project investigated this effect by asking 25 right-handed students to turn a knob (with their right hands) that moved an indicator. There were two identical instruments, one with a right-hand thread (the knob turns clockwise) and the other with a left-hand thread (the knob must be turned counterclockwise). Each of the 25 students used both instruments in a random order. The following table gives the times in seconds each subject took to move the indicator a fixed distance.[23] Note that smaller times are better.

pg 586

Subject	Right thread	Left thread
1	113	137
2	105	105
3	130	133
4	101	108
5	138	115
6	118	170
7	87	103
8	116	145
9	75	78
10	96	107
11	122	84
12	103	148
13	116	147
14	107	87
15	118	166
16	103	146
17	111	123
18	104	135
19	111	112
20	89	93
21	78	76
22	100	116
23	89	78
24	85	101
25	88	123

(a) Explain why it was important to randomly assign the order in which each subject used the two knobs.

(b) The project designers hoped to show that right-handed people find right-hand threads easier to use, on average. Carry out a test at the 5% significance level to investigate this claim.

86. **Floral scents and learning** We hear that listening to Mozart improves students' performance on tests. Maybe pleasant odors have a similar effect. To test this idea, 21 subjects worked two different but roughly equivalent paper-and-pencil mazes while wearing a mask. The mask was either unscented or carried a floral scent. Each subject used both masks, in a random order. The table below gives the subjects' times (in seconds) with both masks.[24] Note that smaller times are better.

Subject	Unscented	Scented
1	30.60	37.97
2	48.43	51.57
3	60.77	56.67
4	36.07	40.47

Subject	Unscented	Scented
5	68.47	49.00
6	32.43	43.23
7	43.70	44.57
8	37.10	28.40
9	31.17	28.23
10	51.23	68.47
11	65.40	51.10
12	58.93	83.50
13	54.47	38.30
14	43.53	51.37
15	37.93	29.33
16	43.50	54.27
17	87.70	62.73
18	53.53	58.00
19	64.30	52.40
20	47.37	53.63
21	53.67	47.00

(a) Explain why it was important to randomly assign the order in which each subject used the two masks.

(b) Do these data provide convincing evidence that the floral scent improved performance, on average?

87. **Growing tomatoes** Researchers suspect that Variety A tomato plants have a higher average yield than Variety B tomato plants. To find out, researchers randomly select 10 Variety A and 10 Variety B tomato plants. Then the researchers divide in half each of 10 small plots of land in different locations. For each plot, a coin toss determines which half of the plot gets a Variety A plant; a Variety B plant goes in the other half. After harvest, they compare the yield in pounds for the plants at each location. The 10 differences in yield (Variety A − Variety B) are recorded. A graph of the differences looks roughly symmetric and single-peaked with no outliers. A paired t test on the differences yields $t = 1.295$ and P-value $= 0.1138$.

(a) State appropriate hypotheses for the paired t test. Be sure to define your parameter.

(b) What are the degrees of freedom for the paired t test?

(c) Interpret the P-value in context. What conclusion should the researchers draw?

(d) Describe a Type I error and a Type II error in this setting. Which mistake could researchers have made based on your answer to part (c)?

88. **Music and memory** Does listening to music while studying hinder students' learning? Two AP® Statistics students designed an experiment to find out. They selected a random sample of 30 students from their medium-sized high school to participate. Each subject was given 10 minutes to memorize two different lists of 20 words, once while listening to music and once in silence. The order of the two word lists was determined at random; so was the order of the treatments. The difference in the number of words recalled (music − silence) was recorded for each subject. A paired t test on the differences yielded $t = -3.01$ and P-value $= 0.0027$.

(a) State appropriate hypotheses for the paired t test. Be sure to define your parameter.

(b) What are the degrees of freedom for the paired t test?

(c) Interpret the P-value in context. What conclusion should the students draw?

(d) Describe a Type I error and a Type II error in this setting. Which mistake could students have made based on your answer to part (c)?

89. **The power of tomatoes** Refer to Exercise 87. Explain two ways that the researchers could have increased the power of the test to detect $\mu = 0.5$.

90. **Music and memory** Refer to Exercise 88. Which of the following changes would give the test a higher power to detect $\mu = -1$: using $\alpha = 0.01$ or $\alpha = 0.10$? Explain.

91. **Significance and sample size** A study with 5000 subjects reported a result that was statistically significant at the 5% level. Explain why this result might not be particularly large or important.

92. **Sampling shoppers** A marketing consultant observes 50 consecutive shoppers at a supermarket, recording how much each shopper spends in the store. Explain why it would not be wise to use these data to carry out a significance test about the mean amount spent by all shoppers at this supermarket.

93. **Do you have ESP?** A researcher looking for evidence of extrasensory perception (ESP) tests 500 subjects. Four of these subjects do significantly better ($P < 0.01$) than random guessing.

(a) Is it proper to conclude that these four people have ESP? Explain your answer.

(b) What should the researcher now do to test whether any of these four subjects have ESP?

94. **Ages of presidents** Joe is writing a report on the backgrounds of American presidents. He looks up the ages of all the presidents when they entered office. Because Joe took a statistics course, he uses these numbers to perform a significance test about the mean age of all U.S. presidents. Explain why this makes no sense.

Multiple choice: Select the best answer for Exercises 95 to 102.

95. The reason we use t procedures instead of z procedures when carrying out a test about a population mean is that

(a) z requires that the sample size be large.

(b) z requires that you know the population standard deviation σ.

(c) z requires that the data come from a random sample or randomized experiment.

(d) z requires that the population distribution be perfectly Normal.

(e) z can only be used for proportions.

96. You are testing $H_0: \mu = 10$ against $H_a: \mu < 10$ based on an SRS of 20 observations from a Normal population. The t statistic is $t = -2.25$. The P-value

(a) falls between 0.01 and 0.02.

(b) falls between 0.02 and 0.04.

(c) falls between 0.04 and 0.05.

(d) falls between 0.05 and 0.25.

(e) is greater than 0.25.

97. You are testing $H_0: \mu = 10$ against $H_a: \mu \neq 10$ based on an SRS of 15 observations from a Normal population. What values of the t statistic are statistically significant at the $\alpha = 0.005$ level?

(a) $t > 3.326$ (d) $t < -3.326$ or $t > 3.326$

(b) $t > 3.286$ (e) $t < -3.286$ or $t > 3.286$

(c) $t > 2.977$

98. After checking that conditions are met, you perform a significance test of $H_0: \mu = 1$ versus $H_a: \mu \neq 1$. You obtain a P-value of 0.022. Which of the following must be true?

(a) A 95% confidence interval for μ will include the value 1.

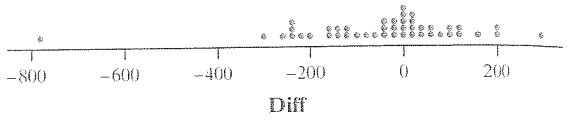
(b) A 95% confidence interval for μ will include the value 0.

(c) A 99% confidence interval for μ will include the value 1.

(d) A 99% confidence interval for μ will include the value 0.

(e) None of these is necessarily true.

99. Does Friday the 13th have an effect on people's behavior? Researchers collected data on the number of shoppers at a sample of 45 nearby grocery stores on Friday the 6th and Friday the 13th in the same month. The dotplot and computer output below summarize the data on the difference in the number of shoppers at each store on these two days (subtracting in the order 6th minus 13th).[25]



N	Mean	Median	TrMean	StDev	SEMean	Min	Max	Q1	Q3
45	-46.5	-11.0	-37.4	178.0	26.1	-774.0	302.0	-141.0	53.5

Researchers would like to carry out a test of $H_0: \mu_d = 0$ versus $H_a: \mu_d \neq 0$, where μ_d is the true mean difference in the number of grocery shoppers on these two days. Which of the following conditions for performing a paired t test are clearly satisfied?

I. Random II. 10% III. Normal/Large Sample

(a) I only (c) III only (e) I, II, and III

(b) II only (d) I and II only

100. The most important condition for sound conclusions from statistical inference is that

(a) the data come from a well-designed random sample or randomized experiment.

(b) the population distribution be exactly Normal.

(c) the data contain no outliers.

(d) the sample size be no more than 10% of the population size.

(e) the sample size be at least 30.

101. Vigorous exercise helps people live several years longer (on average). Whether mild activities like slow walking extend life is not clear. Suppose that the added life expectancy from regular slow walking is just 2 months. A statistical test is more likely to find a significant increase in mean life expectancy if

(a) it is based on a very large random sample and a 5% significance level is used.

(b) it is based on a very large random sample and a 1% significance level is used.

(c) it is based on a very small random sample and a 5% significance level is used.

(d) it is based on a very small random sample and a 1% significance level is used.

(e) the size of the sample doesn't have any effect on the significance of the test.

102. A researcher plans to conduct a significance test at the $\alpha = 0.01$ significance level. She designs her study to have a power of 0.90 at a particular alternative value of the parameter of interest. The probability that the researcher will commit a Type II error for the particular alternative value of the parameter at which she computed the power is

(a) 0.01. (b) 0.10. (c) 0.89. (d) 0.90. (e) 0.99.

103. **Is your food safe?** (5.1) "Do you feel confident or not confident that the food available at most grocery stores is safe to eat?" When a Gallup Poll asked this question, 87% of the sample said they were confident.[26] Gallup announced the poll's margin of error for 95% confidence as ±3 percentage points. Which of the following sources of error are included in this margin of error? Explain.

(a) Gallup dialed landline telephone numbers at random and so missed all people without landline phones, including people whose only phone is a cell phone.

(b) Some people whose numbers were chosen never answered the phone in several calls or answered but refused to participate in the poll.

(c) There is chance variation in the random selection of telephone numbers.

104. **Spinning for apples** (6.3 or 7.3) In the "Ask Marilyn" column of *Parade* magazine, a reader posed this question: "Say that a slot machine has five wheels, and each wheel has five symbols: an apple, a grape, a peach, a pear, and a plum. I pull the lever five times. What are the chances that I'll get at least one apple?" Suppose that the wheels spin independently and that the five symbols are equally likely to appear on each wheel in a given spin.

(a) Find the probability that the slot player gets at least one apple in one pull of the lever. Show your method clearly.

(b) Now answer the reader's question. Show your method clearly.

FRAPPY! Free Response AP® Problem, Yay!

The following problem is modeled after actual AP® Statistics exam free response questions. Your task is to generate a complete, concise response in 15 minutes.

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

Anne reads that the average price of regular gas in her state is $4.06 per gallon. To see if the average price of gas is different in her city, she selects 10 gas stations at random and records the price per gallon for regular gas at each station. The data, along with the sample mean and standard deviation, are listed in the table below.

Station	Price
1	4.13
2	4.01
3	4.09
4	4.05
5	3.97
6	3.99
7	4.05
8	3.98
9	4.09
10	4.02
Mean	4.038
SD	0.0533

Do the data provide convincing evidence that the average price of gas in Anne's city is different from $4.06 per gallon?

After you finish, you can view two example solutions on the book's Web site (www.whfreeman.com/tps5e). Determine whether you think each solution is "complete," "substantial," "developing," or "minimal." If the solution is not complete, what improvements would you suggest to the student who wrote it? Finally, your teacher will provide you with a scoring rubric. Score your response and note what, if anything, you would do differently to improve your own score.