

SECTION 10.2 Exercises

In exercises that call for two-sample t procedures, you may use either of the two options for the degrees of freedom that we have discussed, unless you are told otherwise. Be sure to state what df you are using.

35. **Cholesterol** (6.2) The level of cholesterol in the blood for all men aged 20 to 34 follows a Normal distribution with mean 188 milligrams per deciliter (mg/dl) and standard deviation 41 mg/dl. For 14-year-old boys, blood cholesterol levels follow a Normal distribution with mean 170 mg/dl and standard deviation 30 mg/dl.

(a) Let M = the cholesterol level of a randomly selected 20- to 34-year-old man and B = the cholesterol level of a randomly selected 14-year-old boy. Describe the shape, center, and spread of the distribution of $M - B$.

(b) Find the probability that a randomly selected 14-year-old boy has higher cholesterol than a randomly selected man aged 20 to 34. Show your work.

36. **How tall?** (6.2) The heights of young men follow a Normal distribution with mean 69.3 inches and standard deviation 2.8 inches. The heights of young women follow a Normal distribution with mean 64.5 inches and standard deviation 2.5 inches.

(a) Let M = the height of a randomly selected young man and W = the height of a randomly selected young woman. Describe the shape, center, and spread of the distribution of $M - W$.

(b) Find the probability that a randomly selected young man is at least 2 inches taller than a randomly selected young woman. Show your work.

37. **Cholesterol** Refer to Exercise 35. Suppose we select independent SRSs of 25 men aged 20 to 34 and 36 boys aged 14 and calculate the sample mean heights \bar{x}_M and \bar{x}_B .

(a) Describe the shape, center, and spread of the sampling distribution of $\bar{x}_M - \bar{x}_B$.

(b) Find the probability of getting a difference in sample means $\bar{x}_M - \bar{x}_B$ that's less than 0 mg/dl. Show your work.

(c) Should we be surprised if the sample mean cholesterol level for the 14-year-old boys exceeds the sample mean cholesterol level for the men? Explain.

38. **How tall?** Refer to Exercise 36. Suppose we select independent SRSs of 16 young men and 9 young

women and calculate the sample mean heights \bar{x}_M and \bar{x}_B .

(a) Describe the shape, center, and spread of the sampling distribution of $\bar{x}_M - \bar{x}_B$.

(b) Find the probability of getting a difference in sample means $\bar{x}_M - \bar{x}_B$ that's greater than or equal to 2 inches. Show your work.

(c) Should we be surprised if the sample mean height for the young women is more than 2 inches less than the sample mean height for the young men? Explain.

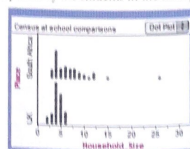
In Exercises 39 to 42, determine whether or not the conditions for using two-sample t procedures are met.

39. **Shoes** How many pairs of shoes do teenagers have? To find out, a group of AP Statistics students conducted a survey. They selected a random sample of 20 female students and a separate random sample of 20 male students from their school. Then they recorded the number of pairs of shoes that each respondent reported having. The back-to-back stemplot below displays the data.

Females	Males
0	4
0	3 5 5 7 7 7 7 8
3 5 5	1
9 5	1
4 3 3 2	2
6 6	2
4 1 0	3
8	3
4	5 8
9	
10 0	
7	

Key: 3|2 represents a male student with 22 pairs of shoes.

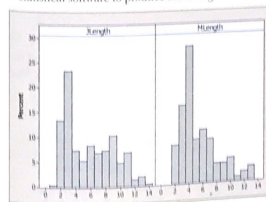
40. **Household size** How do the numbers of people living in households in the United Kingdom (U.K.) and South Africa compare? To help answer this question, we used CensusAtSchool's random data selector to choose independent samples of 50 students from each country. Here is a Fathom dotplot of the household sizes reported by the students in the survey.



41. **Literacy rates** Do males have higher average literacy rates than females in Islamic countries? The table below shows the percent of men and women at least 15 years old who were literate in 2008 in the major Islamic nations. (We omitted countries with populations of less than 3 million.) Data for a few nations, such as Afghanistan and Iraq, were not available.³⁰

Country	Female percent	Male percent
Algeria	66	94
Bangladesh	48	71
Egypt	58	88
Iran	77	97
Jordan	87	99
Kazakhstan	100	100
Lebanon	96	98
Libya	78	100
Malaysia	90	98
Morocco	43	84
Saudi Arabia	79	98
Syria	77	95
Tajikistan	100	100
Tunisia	69	97
Turkey	81	99
Uzbekistan	96	99
Yemen	41	93

42. **Long words** Mary was interested in comparing the mean word length in articles from a medical journal and an airline's in-flight magazine. She counted the number of letters in the first 200 words of an article in the medical journal and in the first 100 words of an article in the airline magazine. Mary then used Minitab statistical software to produce the histogram shown.



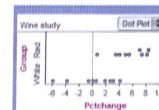
43. **Is red wine better than white wine?** Observational studies suggest that moderate use of alcohol by adults reduces heart attacks and that red wine may have special benefits. One reason may be that red wine contains polyphenols, substances that do good things to cholesterol in the blood and so may reduce the risk

of heart attacks. In an experiment, healthy men were assigned at random to drink half a bottle of either red or white wine each day for two weeks. The level of polyphenols in their blood was measured before and after the two-week period. Here are the percent changes in level for the subjects in both groups.³¹

Red wine:	3.5	8.1	7.4	4.0	0.7	4.9	8.4	7.0	5.5
White wine:	3.1	0.5	-3.8	4.1	-0.6	2.7	1.9	-5.9	0.1

(a) A Fathom dotplot of the data is shown below. Use the graph to answer these questions:

- Are the centers of the two groups similar or different? Explain.
- Are the spreads of the two groups similar or different? Explain.



(b) Construct and interpret a 90% confidence interval for the difference in mean percent change in polyphenol levels for the red wine and white wine treatments.

(c) Does the interval in part (b) suggest that red wine is more effective than white wine? Explain.

44. **Tropical flowers** Different varieties of the tropical flower *Heliconia* are fertilized by different species of hummingbirds.



Researchers believe that over time, the lengths of the flowers and the forms of the hummingbirds' beaks have evolved to match each other. Here are data on the lengths in millimeters for random samples of two

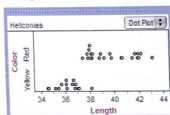
color varieties of the same species of flower on the island of Dominica.³²

H. caribaea red										
41.90	42.01	41.93	43.09	41.17	41.69	38.78	40.57			
39.63	42.18	40.66	37.87	39.16	37.40	38.20	38.07			
38.10	37.97	38.79	38.23	38.87	37.78	38.01				

H. caribaea yellow										
38.78	37.02	36.52	35.11	36.03	35.45	38.13	37.1			
35.17	36.82	36.66	35.68	36.03	34.57	34.63				

(a) A Fathom dotplot of the data is shown below. Use the graph to answer these questions:

- Are the centers of the two groups similar or different? Explain.
- Are the spreads of the two groups similar or different? Explain.



(b) Construct and interpret a 95% confidence interval for the difference in the mean lengths of these two varieties of flowers.

(c) Does the interval support the researchers' belief that the two flower varieties have different average lengths? Explain.

45. **Paying for college** College financial aid offices expect students to use summer earnings to help pay for college. But how large are these earnings? One large university studied this question by asking a random sample of 1296 students who had summer jobs how much they earned. The financial aid office separated the responses into two groups based on gender. Here are the data in summary form:³³

Group	n	\bar{x}	s_x
Males	675	\$1084.52	\$1368.37
Females	621	\$1360.39	\$1037.46

(a) How can you tell from the summary statistics that the distribution of earnings in each group is strongly skewed to the right? A graph of the data reveals no outliers. The use of two-sample t procedures is still justified. Why?

(b) Construct and interpret a 90% confidence interval for the difference between the mean

summer earnings of male and female students at this university.

(c) Interpret the 90% confidence level in the context of this study.

46. **Happy customers** As the Hispanic population in the United States has grown, businesses have tried to understand what Hispanics like. One study interviewed a random sample of customers leaving a bank. Customers were classified as Hispanic if they preferred to be interviewed in Spanish or as Anglo if they preferred English. Each customer rated the importance of several aspects of bank service on a 10-point scale.³⁴ Here are summary results for the importance of "reliability" (the accuracy of account records and so on):

Group	n	\bar{x}	s_x
Anglo	92	6.37	0.60
Hispanic	86	5.91	0.93

(a) The distribution of reliability ratings in each group is not Normal. A graph of the data reveals no outliers. The use of two-sample t procedures is justified. Why?

(b) Construct and interpret a 95% confidence interval for the difference between the mean ratings of the importance of reliability for Anglo and Hispanic bank customers.

(c) Interpret the 95% confidence level in the context of this study.

47. **Web business** You want to compare the daily sales for two different designs of Web pages for your Internet business. You assign the next 60 days to either Design A or Design B, 30 days to each.

(a) Describe how you would assign the days for Design A and Design B using the partial line of random digits provided below. Then use your plan to select the first three days for using Design A. Show your method clearly on your paper.

24005 52114 26224 39078

(b) Would you use a one-sided or a two-sided significance test for this problem? Explain your choice. Then set up appropriate hypotheses.

(c) If you plan to use Table B to calculate the P -value, what are the degrees of freedom?

(d) The t statistic for comparing the mean sales is 2.06. Using Table B, what P -value would you report? What would you conclude?

Exercises 48 to 50 refer to the following setting. Do birds learn to time their breeding? Blue titmouse eat caterpillars. The birds would like lots of caterpillars around when they

have young to feed, but they must breed much earlier. Do the birds learn from one year's experience when to time their breeding next year? Researchers randomly assigned 7 pairs of birds to have the natural caterpillar supply supplemented while feeding their young and another 6 pairs to serve as a control group relying on natural food supply. The next year, they measured how many days after the caterpillar peak the birds produced their nestlings.³⁵

48. **Did the randomization produce similar groups?** First, compare the two groups of birds in the first year. The only difference should be the chance effect of the random assignment. The study report says: "In the experimental year, the degree of synchronization did not differ between food-supplemented and control females." For this comparison, the report gives $t = -1.05$.

(a) What type of t statistic (one-sample, paired, or two-sample) is this? Justify your answer.

(b) Explain how this value of t leads to the quoted conclusion.

49. **Did the treatment have an effect?** The investigators expected the control group to adjust their breeding date the next year, whereas the well-fed supplemented group had no reason to change. The report continues: "But in the following year, food-supplemented females were more out of synchrony with the caterpillar peak than the controls." Here are the data (days behind caterpillar peak):

Control:	4.6	2.3	7.7	6.0	4.6	-1.2	
Supplemented:	15.5	11.3	5.4	16.5	11.3	11.4	7.7

Carry out an appropriate test and show that it leads to the quoted conclusion.

50. **Year-to-year comparison** Rather than comparing the two groups in each year, we could compare the behavior of each group in the first and second years. The study report says: "Our main prediction was that females receiving additional food in the nestling period should not change laying date the next year, whereas controls, which (in our area) breed too late in their first year, were expected to advance their laying date in the second year."

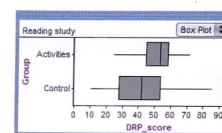
Comparing days behind the caterpillar peak in Years 1 and 2 gave $t = 0.63$ for the control group and $t = -2.63$ for the supplemented group.

(a) What type of t statistic (one-sample, paired, or two-sample) are these? Justify your answer.

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

(c) Explain why these t -values do not agree with the prediction.

51. **Teaching reading** An educator believes that new reading activities in the classroom will help elementary school pupils improve their reading ability. She recruits 44 third-grade students and randomly assigns them into two groups. One group of 21 students does these new activities for an 8-week period. A control group of 23 third-graders follows the same curriculum without the activities. At the end of the 8 weeks, all students are given the Degree of Reading Power (DRP) test, which measures the aspects of reading ability that the treatment is designed to improve. Comparative boxplots and summary statistics for the data from Fathom are shown below.³⁶



Reading study	Group	
	Activities	Control
DRP_score	51.4762	41.5217
	21	23
	11.0074	17.1487

S1 = mean ()
S2 = count ()
S3 = stdDev ()

(a) Based on the graph and numerical summaries, write a few sentences comparing the DRP scores for the two groups.

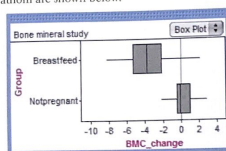
(b) Is the mean DRP score significantly higher for the students who did the reading activities? Carry out an appropriate test to support your answer.

(c) Can we conclude that the new reading activities caused an increase in the mean DRP score? Explain.

(d) Construct and interpret a 95% confidence interval for the difference in mean DRP scores. Explain how this interval provides more information than the significance test in part (b).

52. **Does breast-feeding weaken bones?** Breast-feeding mothers secrete calcium into their milk. Some of the calcium may come from their bones, so mothers may lose bone mineral. Researchers compared a random sample of 47 breast-feeding women with a random sample of 22 women of similar age who were neither pregnant nor lactating. They measured the percent change in the bone mineral content (BMC) of the women's spines over three months. Comparative

boxplots and summary statistics for the data from Fathom are shown below.³⁷



Bone mineral study	
	Group
	Breastfeed Notpregnant
BMC_change	-3.58723 0.308091
	47 22
	2.50561 1.29832
S1 = mean ()	
S2 = count ()	
S3 = stdDev ()	

- (a) Based on the graph and numerical summaries, write a few sentences comparing the percent changes in BMC for the two groups.
- (b) Is the mean change in BMC significantly lower for the mothers who are breast-feeding? Carry out an appropriate test to support your answer.
- (c) Can we conclude that breast-feeding causes a mother's bones to weaken? Why or why not?
- (d) Construct and interpret a 95% confidence interval for the difference in mean bone mineral loss. Explain how this interval provides more information than the significance test in part (b).

53. **Who talks more—men or women?** Researchers equipped random samples of 56 male and 56 female students from a large university with a small device that secretly records sound for a random 30 seconds during each 12.5-minute period over two days. Then they counted the number of words spoken by each subject during each recording period and, from this, estimated how many words per day each subject speaks. The female estimates had a mean of 16,177 words per day with a standard deviation of 7520 words per day. For the male estimates, the mean was 16,569 and the standard deviation was 9108.
- (a) Do these data provide convincing evidence of a difference in the average number of words spoken in a day by male and female students at this university? Carry out an appropriate test to support your answer.

- (b) Interpret the P-value from part (a) in the context of this study.

54. **DDT in rats** Poisoning by the pesticide DDT causes convulsions in humans and other mammals. Researchers seek to understand how the convulsions are caused. In a randomized comparative experiment, they compared 6 white rats poisoned with DDT with a control group of 6 unpoisoned rats. Electrical measurements of nerve activity are the main clue to the nature of DDT poisoning. When a nerve is stimulated, its electrical response shows a sharp spike followed by a much smaller second spike. The researchers measured the height of the second spike as a percent of the first spike when a nerve in the rat's leg was stimulated.³⁸ For the poisoned rats the results were

12.207 16.869 25.050 22.429 8.456 20.589

The control group data were

11.074 9.686 12.064 9.351 8.182 6.642

Computer output for a two-sample t test on these data from SAS software is shown below. (Note that SAS provides two-sided P-values.)

TTEST PROCEDURE					
Variable: SPIKE					
GROUP	N	MEAN	Std Dev	Std Error	
DDT	6	17.60000000	6.34014839	2.58835474	
CONTROL	6	9.49983333	1.95005932	0.79610839	
Variances					
	T	DF	Prob> T		
Unequal	2.9912	5.9	0.0247		
Equal	2.9912	10.0	0.0135		

- (a) Do these data provide convincing evidence that DDT affects the mean height of the second spike's electrical response? Carry out a significance test to help answer this question.

- (b) Interpret the P-value from part (a) in the context of this study.

55. **Competitive rowers** What aspects of rowing technique distinguish between novice and skilled competitive rowers? Researchers compared two randomly selected groups of female competitive rowers: a group of skilled rowers and a group of novices. The researchers measured many mechanical aspects of rowing style as the subjects rowed on a Stanford Rowing Ergometer. One important variable is the angular velocity of the knee, which describes the rate at which the knee joint opens as the legs push the body back on the sliding

seat. The data show no outliers or strong skewness. Here is the SAS computer output.³⁹

TTEST PROCEDURE					
Variable: KNEE					
GROUP	N	Mean	Std Dev	Std Error	
SKILLED	10	4.182	0.479	0.151	
NOVICE	8	3.010	0.959	0.339	
Variances					
	T	DF	Prob> T		
Unequal	3.1583	9.8	0.0104		
Equal	3.3918	16.0	0.0037		

- (a) The researchers believed that the knee velocity would be higher for skilled rowers. Use the computer output to carry out an appropriate test of this belief. (Note that SAS provides two-sided P-values.) What do you conclude?

- (b) Use technology to construct and interpret a 90% confidence interval for the mean difference between the knee velocities of skilled and novice female rowers.

- (c) If you had used Table B to construct the confidence interval described in part (b), how would the two results compare? Justify your answer without doing any calculations.

56. **Competitive rowers** The research in the previous exercise also wondered whether skilled and novice rowers differ in weight or other physical characteristics. Graphs of the data reveal no outliers or strong skewness. Here is the SAS computer output for weight in kilograms:

TTEST PROCEDURE					
Variable: WEIGHT					
GROUP	N	Mean	Std Dev	Std Error	
SKILLED	10	70.37	6.10	1.93	
NOVICE	8	68.45	9.04	3.20	
Variances					
	T	DF	Prob> T		
Unequal	11.8	0.6165			
Equal	0.5376	16.0	0.5982		

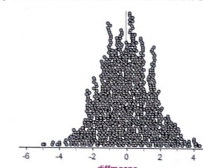
- (a) Calculate the missing t statistic in the computer output. Show your work.

- (b) Is there significant evidence of a difference in the mean weights of skilled and novice rowers? Justify your answer. (Note that SAS provides two-sided P-values.)

- (c) If the P-value had been computed using the more conservative number of degrees of freedom, would it have been greater than, equal to, or less than 0.6165? Justify your answer without doing any calculations.

57. **Rewards and creativity** Dr. Teresa Amabile conducted a study involving 47 college students, who were randomly assigned to two treatment groups. The 23 students in one group were given a list of statements about external reasons (E) for writing, such as public recognition, making money, or pleasing their parents. The 24 students in the other group were given a list of statements about internal reasons (I) for writing, such as expressing yourself and enjoying playing with words. Both groups were then instructed to write a poem about laughter. Each student's poem was rated separately by 12 different poets using a creativity scale.⁴⁰ The 12 poets' ratings of each student's poem were averaged to obtain an overall creativity score.

We used Fathom software to randomly reassign the 47 subjects to the two groups 1000 times, assuming the treatment received doesn't affect each individual's average creativity rating. The dotplot shows the approximate randomization distribution of $\bar{x}_I - \bar{x}_E$.



- (a) Why did researchers randomly assign the subjects to the two treatment groups?

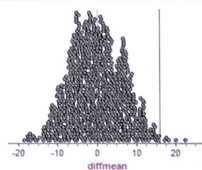
- (b) In the actual experiment, $\bar{x}_I - \bar{x}_E = 4.15$. What conclusion would you draw? Justify your answer with appropriate evidence.

- (c) Based on your conclusion in part (b), could you have made a Type I error or a Type II error? Justify your answer.

58. **Sleep deprivation** Does sleep deprivation linger for more than a day? Researchers designed a study using 21 volunteer subjects between the ages of 18 and 25. All 21 participants took a computer-based visual discrimination test at the start of the study. Then the subjects were randomly assigned into two groups. The 11 subjects in one group, D, were deprived of sleep for an entire night in a laboratory setting. The 10 subjects in the other group, A, were allowed unrestricted sleep for the night. Both groups were allowed as much sleep as they wanted for the next two nights. On Day 4, all the subjects took the same visual discrimination test on the computer. Researchers recorded the improvement in time (measured

in milliseconds) from Day 1 to Day 4 on the test for each subject.⁴¹

We used Fathom software to randomly reassign the 21 subjects to the two groups 1000 times, assuming the treatment received doesn't affect each individual's time improvement on the test. The dotplot shows the approximate randomization distribution of $\bar{x}_A - \bar{x}_B$.



- (a) Explain why the researchers didn't let the subjects choose whether to be in the sleep deprivation group or the unrestricted sleep group.
- (b) In the actual experiment, $\bar{x}_A - \bar{x}_B = 15.92$. This value is marked with a blue line in the figure. What conclusion would you draw? Justify your answer with appropriate evidence.
- (c) Based on your conclusion in part (b), could you have made a Type I error or a Type II error? Justify your answer.

59. Paired or unpaired? In each of the following settings, decide whether you should use paired t procedures or two-sample t procedures to perform inference. Explain your choice.⁴²

- (a) To test the wear characteristics of two tire brands, A and B, each brand of tire is randomly assigned to 50 cars of the same make and model.
- (b) To test the effect of background music on productivity, factory workers are observed. For one month, each subject works without music. For another month, the subject works while listening to music on an MP3 player. The month in which each subject listens to music is determined by a coin toss.
- (c) A study was designed to compare the effectiveness of two weight-reducing diets. Fifty obese women who volunteered to participate were randomly assigned into two equal-sized groups. One group used Diet A and the other used Diet B. The weight of each woman was measured before the assigned diet and again after 10 weeks on the diet.

60. Paired or unpaired? In each of the following settings, decide whether you should use paired t procedures or two-sample t procedures to perform inference. Explain your choice.⁴³

- (a) To compare the average weight gain of pigs fed two different rations, nine pairs of pigs were used. The pigs in each pair were littermates. A coin toss was used to decide which pig in each pair got Ration A and which got Ration B.
- (b) A random sample of college professors is taken. We wish to compare the average salaries of male and female teachers.
- (c) To test the effects of a new fertilizer, 100 plots are treated with the new fertilizer, and 100 plots are treated with another fertilizer. A computer's random number generator is used to determine which plots get which fertilizer.

In each of Exercises 61 to 64, say which inference procedure from Chapter 8, 9, or 10 you would use. Be specific. For example, you might say, "Two-sample z test for the difference between two proportions." You do not need to carry out any procedures.

- 61. Preventing drowning** Drowning in bathtubs is a major cause of death in children less than 5 years old. A random sample of parents was asked many questions related to bathtub safety. Overall, 85% of the sample said they used baby bathtubs for infants. Estimate the percent of all parents of young children who use baby bathtubs.
- 62. Driving too fast** How seriously do people view speeding in comparison with other annoying behaviors? A large random sample of adults was asked to rate a number of behaviors on a scale of 1 (no problem at all) to 5 (very severe problem). Do speeding drivers get a higher average rating than noisy neighbors?
- 63. Looking back on love** How do young adults look back on adolescent romance? Investigators interviewed 40 couples in their midtwenties. The female and male partners were interviewed separately. Each was asked about his or her current relationship and also about a romantic relationship that lasted at least two months when they were aged 15 or 16. One response variable was a measure on a numerical scale of how much the attractiveness of the adolescent partner mattered. You want to compare the men and women on this measure.
- 64. Dropping out** You have data from interviews with a random sample of students who failed to graduate from a particular college in 7 years and also from a random sample of students who entered at the same time and did graduate. You will use these data to compare the percents of students from rural backgrounds among dropouts and graduates.
- 65. A better drug?** In a pilot study, a company's new cholesterol-reducing drug outperforms the currently available drug. If the data provide convincing

evidence that the mean cholesterol reduction with the new drug is more than 10 milligrams per deciliter of blood (mg/dl) greater than with the current drug, the company will begin the expensive process of mass-producing the new drug. For the 14 subjects who were assigned at random to the current drug, the mean cholesterol reduction was 54.1 mg/dl with a standard deviation of 11.93 mg/dl. For the 15 subjects who were randomly assigned to the new drug, the mean cholesterol reduction was 68.7 mg/dl with a standard deviation of 13.3 mg/dl. Graphs of the data reveal no outliers or strong skewness.

- (a) Carry out an appropriate significance test. What conclusion would you draw? (Note that the null hypothesis is $H_0: \mu_1 - \mu_2 = 0$.)
- (b) Based on your conclusion in part (a), could you have made a Type I error or a Type II error? Justify your answer.

- 66. Down the toilet** A company that makes hotel toilets claims that its new pressure-assisted toilet reduces the average amount of water used by more than 0.5 gallon per flush when compared to its current model. To test this claim, the company randomly selects 30 toilets of each type and measures the amount of water that is used when each toilet is flushed once. For the current-model toilets, the mean amount of water used is 1.64 gal with a standard deviation of 0.29 gal. For the new toilets, the mean amount of water used is 1.09 gal with a standard deviation of 0.18 gal.
- (a) Carry out an appropriate significance test. What conclusion would you draw? (Note that the null hypothesis is $H_0: \mu_1 - \mu_2 = 0$.)
- (b) Based on your conclusion in part (a), could you have made a Type I error or a Type II error? Justify your answer.

Multiple choice: Select the best answer for Exercises 67 to 70.

- 67.** One major reason that the two-sample t procedures are widely used is that they are quite *robust*. This means that
- (a) t procedures do not require that we know the standard deviations of the populations.
- (b) t procedures work even when the Random, Normal, and Independent conditions are violated.
- (c) t procedures compare population means, a comparison that answers many practical questions.
- (d) confidence levels and P -values from the t procedures are quite accurate even if the population distribution is not exactly Normal.
- (e) confidence levels and P -values from the t procedures are quite accurate even if outliers and strong skewness are present.

- 68.** There are two common methods for measuring the concentration of a pollutant in fish tissue. Do the two methods differ on the average? You apply both methods to a random sample of 18 carp and use
- (a) the paired t test for μ_d
- (b) the one-sample z test for p .
- (c) the two-sample t test for $\mu_1 - \mu_2$.
- (d) the two-sample z test for $p_1 - p_2$.
- (e) none of these.

Exercises 69 and 70 refer to the following setting. A study of road rage asked samples of 596 men and 523 women about their behavior while driving. Based on their answers, each person was assigned a road rage score on a scale of 0 to 20. The participants were chosen by random digit dialing of telephone numbers.

- 69.** We suspect that men are more prone to road rage than women. To see if this is true, test these hypotheses for the mean road rage scores of all male and female drivers:
- (a) $H_0: \mu_M = \mu_F$ versus $H_a: \mu_M > \mu_F$
- (b) $H_0: \mu_M = \mu_F$ versus $H_a: \mu_M \neq \mu_F$
- (c) $H_0: \mu_M = \mu_F$ versus $H_a: \mu_M < \mu_F$
- (d) $H_0: \bar{x}_M = \bar{x}_F$ versus $H_a: \bar{x}_M > \bar{x}_F$
- (e) $H_0: \bar{x}_M = \bar{x}_F$ versus $H_a: \bar{x}_M < \bar{x}_F$
- 70.** The two-sample t statistic for the road rage study (male mean minus female mean) is $t = 3.18$. The P -value for testing the hypotheses from the previous exercise satisfies
- (a) $0.001 < P < 0.005$. (d) $0.002 < P < 0.01$.
- (b) $0.0005 < P < 0.001$. (e) $P > 0.01$.
- (c) $0.001 < P < 0.002$.

Exercises 71 to 74 refer to the following setting. Coaching companies claim that their courses can raise the SAT scores of high school students. Of course, students who retake the SAT without paying for coaching generally raise their scores. A random sample of students who took the SAT twice found 427 who were coached and 2733 who were uncoached.⁴⁴ Starting with their Verbal scores on the first and second tries, we have these summary statistics:

	Try 1			Try 2			Gain		
	n	\bar{x}	s_x	n	\bar{x}	s_x	\bar{x}	s_x	s_x
Coached	427	500	92	529	97	29	59		
Uncoached	2733	506	101	527	101	21	52		

- 71. Coaching and SAT scores** [10.1] What proportion of students who take the SAT twice are coached? To answer this question, Jannie decides to construct a 99% confidence interval. Her work is shown below. Explain what's wrong with Jannie's method.