Section 10.2 Exercises

STEP **4.** *Remember: We are no longer reminding you to use the four-step process in exercises that require you to perform inference.*

31. Cholesterol 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. Suppose we select independent SRSs of 25 men aged 20 to 34 and 36 boys aged 14 and calculate the sample mean cholesterol levels $\bar{x}_M$ and $\bar{x}_B$.

(a) What is the shape of the sampling distribution of $\bar{x}_M - \bar{x}_B$? Why?

(b) Find the mean of the sampling distribution. Show your work.

(c) Find the standard deviation of the sampling distribution. Show your work.

32. How tall? 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. Suppose we select independent SRSs of 16 young men and 9 young women and calculate the sample mean heights $\bar{x}_M$ and $\bar{x}_W$.

(a) What is the shape of the sampling distribution of $\bar{x}_M - \bar{x}_W$? Why?

(b) Find the mean of the sampling distribution. Show your work.

(c) Find the standard deviation of the sampling distribution. Show your work.

In Exercises 33 to 36, determine whether or not the conditions for using two-sample t procedures are met.

33. 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 displays the data.

```
Females    Males
        0 | 4
        0 | 555677778
   333  1 | 0000124
    95  1 |
  4332  2 | 2
    66  2 |
   410  3 |
     8  3 | 58
        4 |
     9  4 |
   100  5 |
     7  5 |
```

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

34. 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.
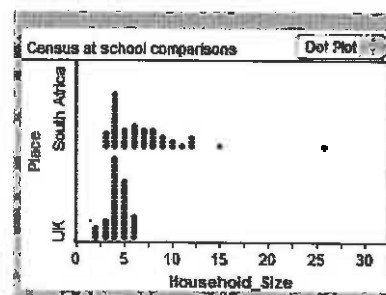
35. Literacy rates Do males have higher average literacy rates than females in Islamic countries? The table below shows the percent of men and women who were literate in the major Islamic nations at the time of this writing.[27] (We omitted countries with populations of less than 3 million.)

Country	Male (%)	Female (%)
Afghanistan	43	13
Algeria	80	60
Azerbaijan	99.9	99.7
Bangladesh	61	52
Egypt	80	64
Indonesia	94	86.8
Iran	84	70
Iraq	86	71
Jordan	96	89
Kazakhstan	100	99
Kyrgyzstan	99.3	98.1
Lebanon	93	82
Libya	96	83
Malaysia	92	85
Morocco	69	44
Pakistan	68.6	30.3
Saudi Arabia	90	81

| Country	Male (%)	Female (%)
Syria	86	74
Tajikistan	100	100
Tunisia	83	65
Turkey	98	90
Turkmenistan	99.3	98.3
Uzbekistan	100	99
Yemen	81	47

36. **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 400 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 histograms shown. Note that J is for journal and M is for magazine.
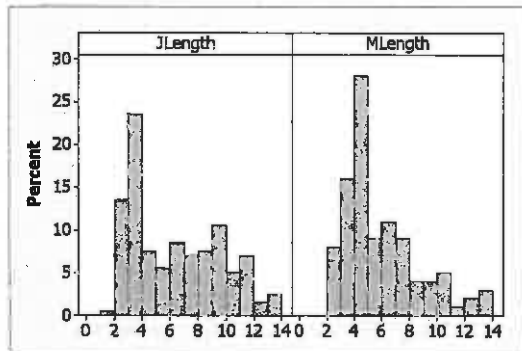
37. **Is red wine better than white wine?** Observational studies suggest that moderate use of alcohol by adults reduces heart attacks and that red wine may have special benefits. One reason may be that red wine contains polyphenols, substances that do good things to cholesterol in the blood and so may 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:[28]

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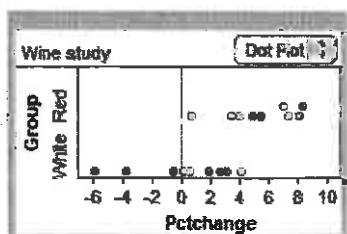
(b) Construct and interpret a 90% confidence interval for the difference in mean percent change in polyphenol levels for the red wine and white wine treatments.

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

38. **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:[29]
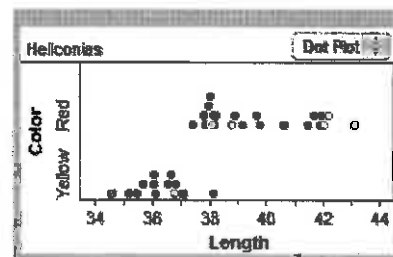
H. *caribaea* red							
41.90	42.01	41.93	43.09	41.17	41.69	39.78	40.57
39.63	42.18	40.66	37.87	39.16	37.40	38.20	38.07
38.10	37.97	38.79	38.23	38.87	37.78	38.01	

H. *caribaea* yellow							
36.78	37.02	36.52	36.11	36.03	35.45	38.13	37.10
35.17	36.82	36.66	35.68	36.03	34.57	34.63	

(a) A Fathom dotplot of the data is shown below. Write a few sentences comparing the distributions.

(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.

39. **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:[30]

Group	n	$\bar{x}$	s_x
Males	675	$1884.52	$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? 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.

40. **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.[31] 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. The use of two-sample t procedures is still 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.

41. **Baby birds** Do birds learn to time their breeding? Blue titmice 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

pg 645

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.[32] 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. Here are the data (days after 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

(a) Do the data provide convincing evidence to confirm the researchers' belief?

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

42. **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.[33] 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

(a) Do these data provide convincing evidence that DDT affects the mean relative height of the second spike's electrical response?

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

43. **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. 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?

44. **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:[34]

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

The researchers believed that the knee velocity would be higher for skilled rowers. Do the data provide convincing evidence to support this belief?

45. **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.[35]
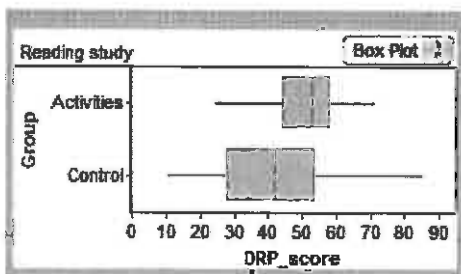

Reading study

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? Give appropriate evidence to justify your answer.

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

46. **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.[36]
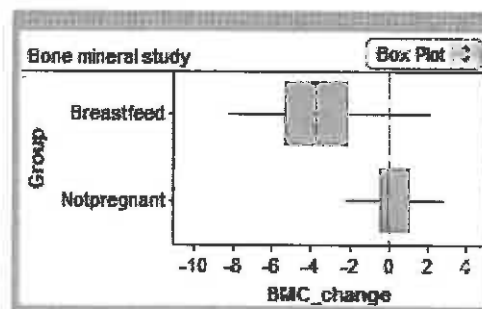
Bone mineral study

Bone mineral study

	Group	
	Breastfeed	Notpregnant
BMC_change	-3.58723	0.309091
	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? Give appropriate evidence to justify your answer.

(c) Can we conclude that breast-feeding causes a mother's bones to weaken? Why or why not?

47. Who talks more—men or women? Refer to Exercise 43. Construct and interpret a 95% confidence interval for the difference in mean number of words spoken in a day. Explain how this interval provides more information than the significance test in Exercise 43.

48. DDT in rats Refer to Exercise 42. Construct and interpret a 95% confidence interval for the difference in mean relative height of the second spike's electrical response. Explain how this interval provides more information than the significance test in Exercise 42.

49. 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 *not* $H_0: \mu_1 - \mu_2 = 0$.)

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

50. 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 *not* $H_0: \mu_1 - \mu_2 = 0$.)

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

51. 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.[37] 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 $\bar{x}_I - \bar{x}_E$.


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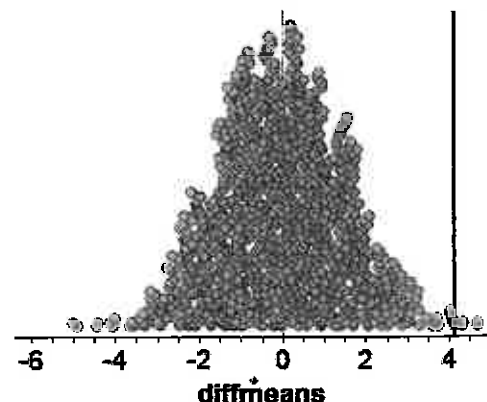
(a) Why did researchers randomly assign the subjects to the two treatment groups?

(b) In the actual experiment, $\bar{x}_I - \bar{x}_E = 4.15$. 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.

52. 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.[38]

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 $\bar{x}_A - \bar{x}_D$.

diffmean
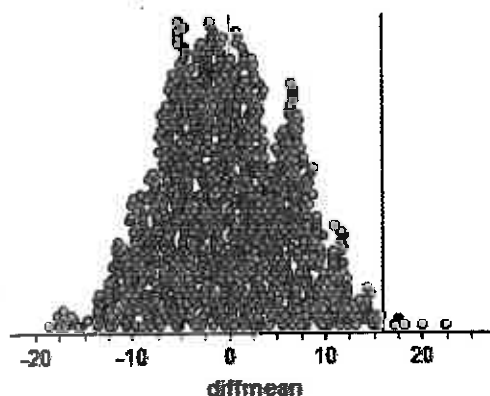
(a) Explain why the researchers didn't let the subjects choose whether to be in the sleep deprivation group or the unrestricted sleep group.

(b) In the actual experiment, $\bar{x}_A - \bar{x}_D = 15.92$. 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.

53. **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.[39]

(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.

54. **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.[40]

(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) Separate random samples of male and female college professors are 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.

Exercises 55 and 56 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.[41] 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	$\bar{x}$	s_x	$\bar{x}$	s_x
Coached	427	500	92	529	97	29	59
Uncoached	2733	506	101	527	101	21	52

55. **Coaching and SAT scores** Let's first ask if students who are coached increased their scores significantly.

(a) You could use the information on the Coached line to carry out either a two-sample t test comparing Try 1 with Try 2 for coached students or a paired t test using Gain. Which is the correct test? Why?

(b) Carry out the proper test. What do you conclude?

56. **Coaching and SAT scores** What we really want to know is whether coached students improve more than uncoached students, and whether any advantage is large enough to be worth paying for. Use the information above to answer these questions:

(a) How much more do coached students gain on the average? Construct and interpret a 99% confidence interval.

(b) Does the interval in part (a) give convincing evidence that coached students gain more, on average, than uncoached students? Explain.

(c) Based on your work, what is your opinion: do you think coaching courses are worth paying for?

Multiple choice: Select the best answer for Exercises 57 to 60.

57. There are two common methods for measuring the concentration of a pollutant in fish tissue. Do the two methods differ, on average? You apply both methods to each fish in 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 58 to 60 refer to the following setting. A study of road rage asked random 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 phone numbers. The researchers performed a test of the following hypotheses: $H_0: \mu_M = \mu_F$ versus $H_a: \mu_M \neq \mu_F$.

58. Which of the following describes a Type II error in the context of this study?

   (a) Finding convincing evidence that the true means are different for males and females, when in reality the true means are the same

   (b) Finding convincing evidence that the true means are different for males and females, when in reality the true means are different

   (c) Not finding convincing evidence that the true means are different for males and females, when in reality the true means are the same

   (d) Not finding convincing evidence that the true means are different for males and females, when in reality the true means are different

   (e) Not finding convincing evidence that the true means are different for males and females, when in reality there is convincing evidence that the true means are different

59. The P-value for the stated hypotheses is 0.002. Interpret this value in the context of this study.

   (a) Assuming that the true mean road rage score is the same for males and females, there is a 0.002 probability of getting a difference in sample means.

   (b) Assuming that the true mean road rage score is the same for males and females, there is a 0.002 probability of getting an observed difference at least as extreme as the observed difference.

   (c) Assuming that the true mean road rage score is different for males and females, there is a 0.002 probability of getting an observed difference at least as extreme as the observed difference.

   (d) Assuming that the true mean road rage score is the same for males and females, there is a 0.002 probability that the null hypothesis is true.

   (e) Assuming that the true mean road rage score is the same for males and females, there is a 0.002 probability that the alternative hypothesis is true.

60. Based on the P-value in Exercise 59, which of the following must be true?

   (a) A 90% confidence interval for $\mu_M - \mu_F$ will contain 0.

   (b) A 95% confidence interval for $\mu_M - \mu_F$ will contain 0.

   (c) A 99% confidence interval for $\mu_M - \mu_F$ will contain 0.

   (d) A 99.9% confidence interval for $\mu_M - \mu_F$ will contain 0.

   (e) It is impossible to determine whether any of these statements is true based only on the P-value.

In each part of Exercises 61 and 62, state 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. **Which inference method?**

   (a) 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.

   (b) 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?

   (c) 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.

   (d) Do experienced computer game players earn higher scores when they play with someone present to cheer them on or when they play alone? Fifty teenagers with experience playing a particular computer game have volunteered for a study. We randomly assign 25 of them to play the game alone and the other 25 to play the game with a supporter present. Each player's score is recorded.

62. Which inference method?

(a) 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 find out how much men and women differ on this measure.

(b) Are more than 75% of Toyota owners generally satisfied with their vehicles? Let's design a study to find out. We'll select a random sample of 400 Toyota owners. Then we'll ask each individual in the sample: "Would you say that you are generally satisfied with your Toyota vehicle?"

(c) Are male college students more likely to binge drink than female college students? The Harvard School of Public Health surveys random samples of male and female undergraduates at four-year colleges and universities about whether they have engaged in binge drinking.

(d) A bank wants to know which of two incentive plans will most increase the use of its credit cards and by how much. It offers each incentive to a group of current credit card customers, determined at random, and compares the amount charged during the following six months.

63. Quality control (2.2, 5.3, 6.3) Many manufacturing companies use statistical techniques to ensure that the products they make meet standards. One common way to do this is to take a random sample of products at regular intervals throughout the production shift. Assuming that the process is working properly, the mean measurement $\bar{x}$ from a random sample varies according to a Normal distribution with mean $\mu_{\bar{x}}$ and standard deviation $\sigma_{\bar{x}}$. For each question that follows, assume that the process is working properly.

(a) What's the probability that at least one of the next two sample means will fall more than $2\sigma_{\bar{x}}$ from the target mean $\mu_{\bar{x}}$? Show your work.

(b) What's the probability that the first sample mean that is greater than $\mu_{\bar{x}} + 2\sigma_{\bar{x}}$ is the one from the fourth sample taken?

Plant managers are trying to develop a criterion for determining when the process is not working properly. One idea they have is to look at the 5 most recent sample means. If at least 4 of the 5 fall outside the interval $(\mu_{\bar{x}} - \sigma_{\bar{x}}, \mu_{\bar{x}} + \sigma_{\bar{x}})$, they will conclude that the process isn't working.

(c) Find the probability that at least 4 of the 5 most recent sample means fall outside the interval, assuming the process is working properly. Is this a reasonable criterion? Explain.

64. Information online (8.2, 10.1) A random digit dialing sample of 2092 adults found that 1318 used the Internet.[42] Of the users, 1041 said that they expect businesses to have Web sites that give product information; 294 of the 774 nonusers said this.

(a) Construct and interpret a 95% confidence interval for the proportion of all adults who use the Internet.

(b) Construct and interpret a 95% confidence interval to compare the proportions of users and nonusers who expect businesses to have Web sites that give product information.

65. Coaching and SAT scores: Critique (4.1, 4.3) The data in Exercises 55 and 56 came from a random sample of students who took the SAT twice. The response rate was 63%, which is fairly good for non-government surveys.

(a) Explain how nonresponse could lead to bias in this study.

(b) We can't be sure that coaching actually *caused* the coached students to gain more than the uncoached students. Explain briefly but clearly why this is so.