1. Use the following to answer questions a-d:

A random sample of 200 students shows that 62% of students use the Student Health Center at some point during their time on campus, with a margin of error of $\pm 4\%$. Based on this information, identify each of the following as plausible or not for the percent of the entire student body that use the Student Health Center at some point during their time on campus.

a. 50%
A) Plausible B) Not Plausible
Ans: B Difficulty: Medium L.O.: 3.2.1; 3.2.2

b. 60%
A) Plausible B) Not Plausible
Ans: A Difficulty: Medium L.O.: 3.2.1; 3.2.2

c. 65%
A) Plausible B) Not Plausible
Ans: A Difficulty: Medium L.O.: 3.2.1; 3.2.2

d. 72%
A) Plausible B) Not Plausible
Ans: B Difficulty: Medium L.O.: 3.2.1; 3.2.2

Use the following to answer questions 2-7:

Identify each of the following as either a parameter or a statistic, and give the correct notation.

2. Correlation between height and armspan (distance from fingertip to fingertip when arms are extended to the sides) for all players on the Chicago Bulls basketball team, using data from all players currently on the team
Ans: parameter, ρ
Difficulty: Medium L.O.: 3.1.1

3. Proportion of students at your university that smoke, based on data from your class.
Ans: Statistic, $\hat{p}$
Difficulty: Medium L.O.: 3.1.1

4. Correlation between price of a textbook and the number of pages, based on 25 textbooks selected from the bookstore.
Ans: Statistic, r
Difficulty: Medium L.O.: 3.1.1

5. Average commute time for employees at a small company, based on interviews with all employees.
   Ans: Parameter, μ
   Difficulty: Medium L.O.: 3.1.1

6. Average gas price in Minnesota, based on prices at randomly selected gas stations throughout the state.
   Ans: Statistic, $\bar{x}$
   Difficulty: Medium L.O.: 3.1.1

7. Proportion of students at a university that are part-time, based on data on all students enrolled at the university.
   Ans: Parameter, *p*
   Difficulty: Moderate L.O.: 3.1.1

8. Use the following to answer questions a-e:

Identify whether each of the following samples is a possible bootstrap sample from this original sample: 20, 24, 19, 23, 18

a. 24, 18, 23
   A) Possible B) Not Possible
   Ans: B Difficulty: Medium L.O.: 3.3.1

b. 24, 19, 24, 20,23
   A) Possible B) Not Possible
   Ans: A Difficulty: Medium L.O.: 3.3.1

c. 20, 24, 21, 19, 18
   A) Possible B) Not Possible
   Ans: B Difficulty: Medium L.O.: 3.3.1

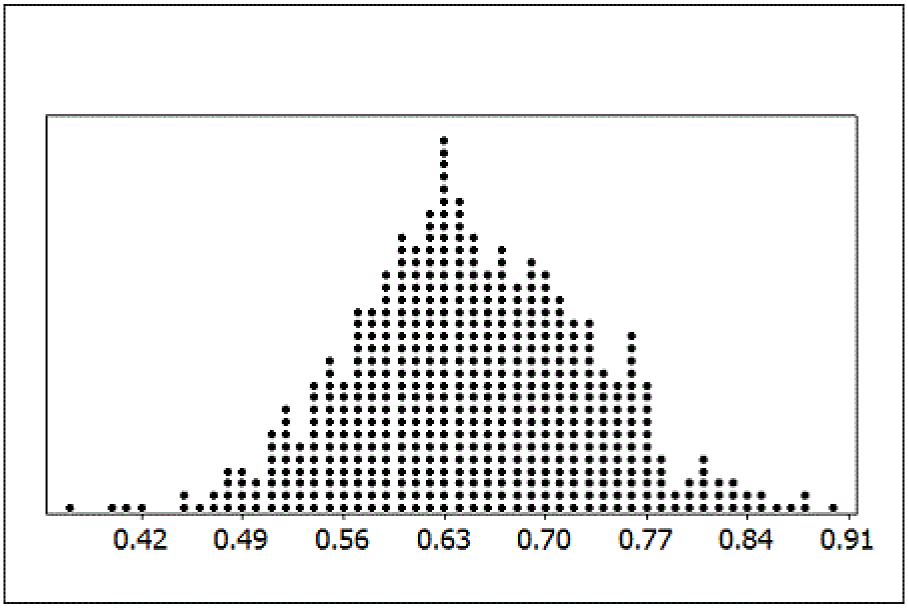
d. 20, 20, 20, 20, 20
   A) Possible B) Not Possible
   Ans: A Difficulty: Medium L.O.: 3.3.1

e. 18, 19, 20, 23, 24
   A) Possible B) Not Possible
   Ans: A Difficulty: Medium L.O.: 3.3.1

9. Briefly explain the distinction between a parameter and a statistic.
   Ans: A parameter is a number that describes some aspect of a population, and a statistic is computed from the data in a sample. A parameter is fixed while a statistic varies from sample to sample.
   Difficulty: Medium L.O.: 3.1.1

Use the following to answer questions 10-17:

The sampling distribution shows sample proportions from samples of size n = 35.



10. What does one dot on the sampling distribution represent?
    Ans: Each dot represents the sample proportion ($\hat{p}$) from a sample of size n = 35.
    Difficulty: Medium L.O.: 3.1.3

11. Estimate the population proportion from the dotplot.
    A) 0.56 B) 0.63 C) 0.70 D) 0.91
    Ans: B Difficulty: Medium L.O.: 3.1.4

12. Estimate the standard error of the sample proportions.
    A) 0.07 B) 0.63 C) 0.14 D) 0.01
    Ans: A Difficulty: Medium L.O.: 3.1.5

13. Using the sampling distribution, how likely is $\hat{p}$ = 0.65?
    A) Reasonably likely to occur from a sample of this size
    B) Unusual but might occur occasionally
    C) Extremely unlikely to ever occur
    Ans: A Difficulty: Medium L.O.: 3.1.3

14. Using the sampling distribution, how likely is $\hat{p}$ = 0.45?
    A) Reasonably likely to occur from a sample of this size
    B) Unusual but might occur occasionally
    C) Extremely unlikely to ever occur
    Ans: B Difficulty: Medium L.O.: 3.1.3

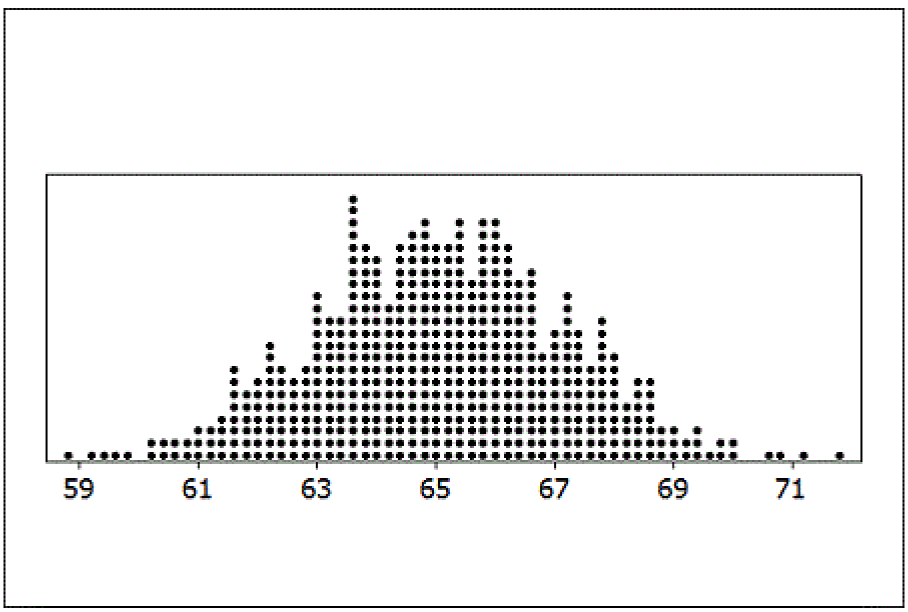
15. Using the sampling distribution, how likely is $\hat{p} = 0.98$?
    A) Reasonably likely to occur from a sample of this size
    B) Unusual but might occur occasionally
    C) Extremely unlikely to ever occur
    Ans: C Difficulty: Medium L.O.: 3.1.3

16. If samples of size $n = 65$ had been used instead of $n = 35$, which of the following would be true?
    A) The sample statistics would be centered at a larger proportion.
    B) The sample statistics would be centered at roughly the same proportion.
    C) The sample statistics would be centered at a smaller proportion.
    Ans: B Difficulty: Medium L.O.: 3.1.6

17. If samples of size $n = 65$ had been used instead of $n = 35$, which of the following would be true?
    A) The sample statistics would have more variability.
    B) The variability in the sample statistics would be about the same.
    C) The sample statistics would have less variability.
    Ans: C Difficulty: Medium L.O.: 3.1.6

Use the following to answer questions 18-25:

The sampling distribution shows sample means from samples of size $n = 50$.



18. What does one dot on the sampling distribution represent?
    Ans: Each individual dot represents the sample mean ($\bar{x}$) from a sample of size $n = 50$.
    Difficulty: Medium L.O.: 3.1.3

19. Estimate the population mean from the dotplot.
    A) 62 B) 63 C) 65 D) 67
    Ans: C Difficulty: Medium L.O.: 3.1.4

20. Estimate the standard error of the sample means.
    A) 1 B) 2 C) 3 D) 5
    Ans: B Difficulty: Medium L.O.: 3.1.5

21. Using the sampling distribution, how likely is $\bar{x} = 55.6$?
    A) Reasonably likely to occur from a sample of this size
    B) Unusual but might occur occasionally
    C) Extremely unlikely to ever occur
    Ans: C Difficulty: Medium L.O.: 3.1.3

22. Using the sampling distribution, how likely is $\bar{x} = 64.2$?
    A) Reasonably likely to occur from a sample of this size
    B) Unusual but might occur occasionally
    C) Extremely unlikely to ever occur
    Ans: A Difficulty: Medium L.O.: 3.1.3

23. Using the sampling distribution, how likely is $\bar{x} = 68.7$?
    A) Reasonably likely to occur from a sample of this size
    B) Unusual but might occur occasionally
    C) Extremely unlikely to ever occur
    Ans: B Difficulty: Medium L.O.: 3.1.3

24. If samples of size $n = 30$ had been used instead of $n = 50$, which of the following would be true?
    A) The sample means would be centered at a larger value.
    B) The sample means would be centered at the same value.
    C) The sample means would be centered at a smaller value.
    Ans: B Difficulty: Medium L.O.: 3.1.6

25. If samples of size $n = 30$ had been used instead of $n = 50$, which of the following would be true?
    A) The sample means would have more variability.
    B) The variability in the sample statistics would be about the same.
    C) The sample means would have less variability.
    Ans: A Difficulty: Medium L.O.: 3.1.6

Use the following to answer questions 26-27:

In an October 2012 survey of 7,786 randomly selected adults living in Germany, 5,840 said they exercised for at least 30 minutes three or more times per week.

26. Identify, with the proper notation, the quantity being estimated.
    Ans: p = proportion of German adults who exercise for 30 minutes three or more times per week.
    Difficulty: Medium L.O.: 3.1.1

27. Using the correct notation, give the value of the best estimate of the population parameter. Round your answer to two decimal places.
    Ans: $\hat{p}$ = 5,840/7,786 = 0.75 (proportion of the sample that say they exercise for 30 minutes three or more times per week)
    Difficulty: Medium L.O.: 3.1.2

Use the following to answer questions 28-32:

In an August 2012 Gallup survey of 1,012 randomly selected U.S. adults (age 18 and over), 53% said that they were dissatisfied with the quality of education students receive in kindergarten through grade 12. They also report that the "margin of sampling error is plus or minus 4%."

28. What is the population of interest?
    Ans: U.S. adults (age 18 and over)
    Difficulty: Easy L.O.: 1.2.1; 3.1.0

29. What is the sample being used?
    Ans: 1,012 randomly selected U.S. adults
    Difficulty: Easy L.O.: 1.2.1; 3.1.0

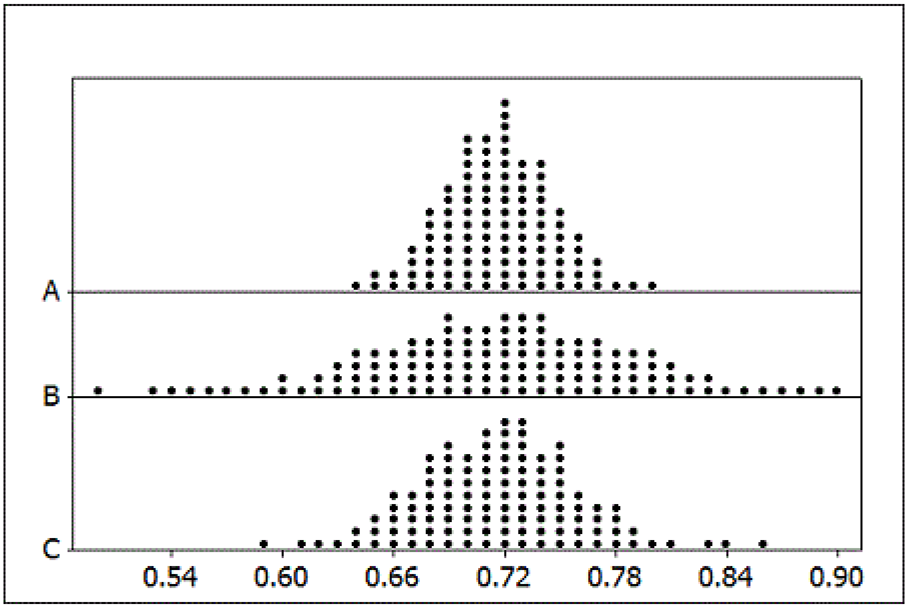
30. What is the population parameter of interest, and what is the correct notation for this parameter?
    Ans: p = proportion of U.S. adults who are dissatisfied with the quality of education students receive in kindergarten through grade 12
    Difficulty: Medium L.O.: 3.1.1

31. What is the relevant statistic?
    Ans: $\hat{p}$ = proportion of the sample of 1,012 randomly selected U.S. adults who are dissatisfied = 0.53
    Difficulty: Medium L.O.: 3.1.1

32. Find an interval estimate for the parameter of interest. Interpret it in terms of dissatisfaction in the quality of education students receive. Use two decimal places in your answer.
    Ans: 0.53 ± 0.04 = 0.49 to 0.57
    We are 95% sure that the true proportion of U.S. adults who are dissatisfied with the quality of education students receive in kindergarten through grade 12 is between 0.49 and 0.57 (i.e., 49% and 57%).
    Difficulty: Medium L.O.: 3.2.1; 3.2.4

33. Use the following to answer questions a-c:

According to U.S. Census data, 71.6% of Americans are age 21 and over. The provided figure shows possible sampling distributions for the proportion of a sample age 21 and over, for samples of size n = 50, n = 125, and n = 250.



Match the sample sizes (n = 50, n = 125, and n = 250) to their sampling distribution.

a. Sample A: n = ______________
   Ans: n = 250
   Difficulty: Medium L.O.: 3.1.6

b. Sample B: n = ______________
   Ans: n = 50
   Difficulty: Medium L.O.: 3.1.6

c. Sample C: n = ______________
   Ans: n = 125
   Difficulty: Medium L.O.: 3.1.6

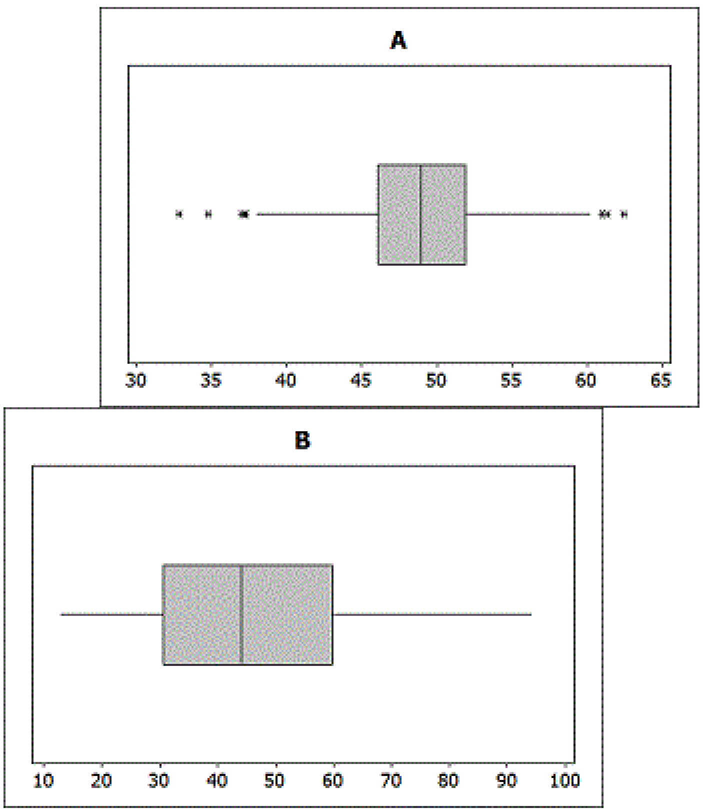
Use the following to answer questions 34-36:

According to ESPN.com, the average number of yards per game for all NFL running backs with at least 50 attempts in the 2011 season was 49 yards/game. A sample of 20 running backs from the 2011 season averaged 46.54 yards/game.

34. Is 49 yards/game a parameter or statistic?
    A) Parameter B) Statistic
    Ans: A Difficulty: Medium L.O.: 3.1.1

35. Is 46.54 yards/game a parameter or statistic?
    A) Parameter B) Statistic
    Ans: B Difficulty: Medium L.O.: 3.1.1

36. Two boxplots are shown. One boxplot corresponds to the yards/game for a random sample of $n = 20$ running backs. The other boxplot represents the values in a sampling distribution of 1,000 means of yards/game for samples of size $n = 20$.



Which boxplot represents the sample? Which boxplot represents the sampling distribution? Briefly explain how you know.
Ans: Boxplot A is the sampling distribution while Boxplot B is a single sample. The values plotted in A are sample means, which will have less variability than the data on individual running backs.
Difficulty: Challenging L.O.: 3.1.0

37. Use the following to answer questions a-e:

Identify if each of the following statements is a proper interpretation of a 95% confidence interval.

a. I am 95% sure that this interval will contain the population parameter.
A) Correct B) Incorrect
Ans: A Difficulty: Medium L.O.: 3.2.4

b. I am 95% sure that this interval will contain the sample statistic.
A) Correct B) Incorrect
Ans: B Difficulty: Moderate L.O.: 3.2.4

c. 95% of the population values will fall within this interval.
A) Correct B) Incorrect
Ans: B Difficulty: Medium L.O.: 3.2.4

d. The probability that the population parameter is in this interval is 0.95.
A) Correct B) Incorrect
Ans: B Difficulty: Medium L.O.: 3.2.4

e. 95% of the possible samples from this population will have sample statistics in *this particular interval*.
A) Correct B) Incorrect
Ans: B Difficulty: Challenging L.O.: 3.2.4

38. In 2010, the Centers for Disease Control and Prevention estimated 9.4% of children under the age of 18 had asthma. They reported the standard error to be 0.35%. Assuming that the sampling distribution is symmetric and bell-shaped, find a 95% confidence interval.
Ans: $0.094 \pm 2 \cdot 0.0035 \Rightarrow 0.094 \pm 0.007 \Rightarrow 0.087$ to 0.101 (or, 8.7% to 10.1%)
Difficulty: Medium L.O.: 3.2.3

39. Use the following to answer questions a-c:

A sample of 148 college students reports sleeping an average of 6.85 hours on weeknights, with a margin of error of 0.35 hours. Based on this information, identify each of the following as plausible or not for the average amount of sleep college students get on weeknights.

a. 6.6 hours
A) Plausible B) Not plausible
Ans: A Difficulty: Medium L.O.: 3.2.2

b. 7.5 hours
A) Plausible B) Not plausible
Ans: B Difficulty: Medium L.O.: 3.2.2

c. 8 hours
A) Plausible B) Not plausible
Ans: B Difficulty: Medium L.O.: 3.2.2

40. Decreasing the confidence level (say, from 95% to 85%) will cause the width of a typical confidence interval to
A) increase. B) decrease. C) remain the same.
Ans: B Difficulty: Medium L.O.: 3.4.3

Use the following to answer questions 41-46:

An Internet provider contacts a random sample of 300 customers and asks how many hours per week the customers use the Internet. It found the average amount of time spent on the Internet per week to be about 7.2 hours.

41. Define the parameter of interest, using the proper notation.
Ans: μ = mean number of hours per week all customers use the Internet
Difficulty: Medium L.O.: 3.1.1

42. Use the information from the sample to give the best estimate of the population parameter.
Ans: $\bar{x}$ = 7.2 hours
Difficulty: Easy L.O.: 3.1.2

43. Describe how to use the data to select one bootstrap sample. What statistic is recorded from this sample?
Ans: A bootstrap sample of size 300 (the same size as the original sample) would be generated by sampling from the original sample with replacement (i.e., each time a value is selected from the original sample it is "returned" to the sample and can be selected again). The bootstrap statistic would be the sample mean of the bootstrap sample.
Difficulty: Challenging L.O.: 3.3.1; 3.4.2

44. The standard error is about 0.458. Find and interpret a 95% confidence interval for the parameter. Round the margin of error to two decimal places.
Ans: $7.2 \pm 2*0.458 \Rightarrow 7.2 \pm 0.92 \Rightarrow 6.28$ to 8.12 hours

We are 95% sure that the mean number of hours all customers (of this provider) spend on the Internet each week is between 6.28 and 8.12 hours.
Difficulty: Medium L.O.: 3.2.3; 3.2.4

45. Percentiles of the bootstrap distribution are provided. Use the percentiles to report a 95% confidence interval for the parameter.

1%	2.5%	5%	10%	25%	50%	75%	90%	95%	97.5%
6.174	6.322	6.438	6.593	6.866	7.17	7.481	7.78	7.947	8.082

Ans: 6.322 hours to 8.082 hours
Difficulty: Medium L.O.: 3.4.1

46. Percentiles of the bootstrap distribution are provided. Use the percentiles to report a 90% confidence interval for the parameter.

1%	2.5%	5%	10%	25%	50%	75%	90%	95%	97.5%
6.174	6.322	6.438	6.593	6.866	7.17	7.481	7.78	7.947	8.082

Ans: 6.438 hours to 7.947 hours
Difficulty: Medium L.O.: 3.4.1

Use the following to answer questions 47-49:

Suppose that a 95% confidence interval for the slope of a regression line based on a sample of size n = 100 and the percentiles of the slopes for 1,000 bootstrap samples goes from 2.50 to 2.80. For each change described (with all else staying the same), indicate which of the three confidence intervals would be the most likely result.

47. Decrease the sample size to n = 60.
A) 2.53 to 2.77 (narrower) B) 2.50 to 2.80 (the same) C) 2.46 to 2.84 (wider)
Ans: C Difficulty: Challenging L.O.: 3.4.3

48. Increase the confidence level to 99%.
A) 2.53 to 2.77 (narrower) B) 2.50 to 2.80 (the same) C) 2.46 to 2.84 (wider)
Ans: C Difficulty: Challenging L.O.: 3.4.3

49. Increase the number of bootstrap samples to 5,000
A) 2.53 to 2.77 (narrower) B) 2.50 to 2.80 (the same) C) 2.46 to 2.84 (wider)
Ans: B Difficulty: Challenging L.O.: 3.3.0; 3.4.2

50. Use the following to answer questions a-d:

In a poll conducted before the Republican presidential primary in Michigan, 134 of 420 randomly chosen likely voters indicated that they planned to vote for Mitt Romney.

a. Compute a sample statistic from these data. Use appropriate notation to label your statistic. Report your answer with three decimal places.
   Ans: $\hat{p}$ = 134/420 = 0.319
   Difficulty: Medium L.O.: 3.1.2

b. Suppose that an article describing the poll says that the margin of error for the statistic is 0.045. Use this information to find an interval estimate.
   Ans: $0.319 \pm 0.045 \Rightarrow 0.274$ to 0.364
   Difficulty: Medium L.O.: 3.2.1

c. What quantity is the interval estimate in (b) trying to capture? Identify with appropriate notation and words.
   Ans: p = proportion of likely Michigan voters who plan to vote for Mitt Romney
   Difficulty: Medium L.O.: 3.1.1; 3.2.0

d. Interpret the confidence interval from (b).
   Ans: We are 95% sure that the proportion of likely Michigan voters that plan to vote for Romney is between 0.274 and 0.364.
   Difficulty: Medium L.O.: 3.2.4

51. Suppose that a student collects pulse rates from a random sample of 200 students at her college and finds a 90% confidence interval goes from 65.5 to 71.8 beats per minute. Is the following statement an appropriate interpretation of this interval? If not, explain why not.

"90% of the students at my college have mean pulse rates between 65.5 and 71.8 beats per minute."

Ans: This is not an appropriate interpretation of the confidence interval. Confidence intervals are constructed to learn about a parameter (a summary of a population), not the individual values in the population (which is what this interpretation is referring to). The correct interpretation of the interval is that "We are 95% sure that the mean pulse rate of all students at her college is between 65.5 and 71.8 beats per minute" - the interval is for a summary of the population values (in this case, the mean pulse rate).
Difficulty: Challenging L.O.: 3.2.4

Use the following to answer questions 52-56:

Suppose we are interested in comparing the proportion of male students who smoke to the proportion of female students who smoke. We have a random sample of 150 students (60 males and 90 females) that includes two variables: Smoke = "yes" or "no" and Gender = "female (F)" or "male (M)". The two-way table below summarizes the results.

	Smoke = Yes	Smoke = No	Sample Size
Gender = M	9	51	60
Gender = F	9	81	90

52. If the parameter of interest is the difference in proportions, p_m - p_f, where p_m and p_f represent the proportion of smokers in each gender, find a point estimate for this difference in proportions based on the data in the table. Report your answer with two decimal places.
    Ans: $\hat{p}_m - \hat{p}_f$ = 9/60 - 9/90 = 0.15 - 0.10 = 0.05
    Difficulty: Medium L.O.: 3.1.2

53. Describe how to use the data to construct a bootstrap distribution. What value should be recorded for each of the bootstrap samples.
    Ans: For each bootstrap sample, randomly select 60 males with replacement from the original sample of 60 males and randomly select 90 females with replacement from the original sample of 90 females. To compute the bootstrap statistic for this sample, compute the proportion of smokers among the bootstrap samples of males and females and find the difference ($\hat{p}_m - \hat{p}_f$). Repeat many times (say 1,000 or more times).
    Difficulty: Medium L.O.: 3.3.1; 3.4.2

54. Use technology to construct a bootstrap distribution with at least 1,000 samples and estimate the standard error.
    Ans: Answers will vary slightly: SE = 0.056 (based on 5,000 bootstrap samples in Statkey)
    Difficulty: Medium L.O.: 3.3.3; 3.3.4

55. Use the estimate of the standard error to construct a 95% confidence interval for the difference in the proportion of smokers between male and female students, p_m - p_f. Round the margin of error to three decimal places. Provide an interpretation of the interval in the context of this data situation.
    Ans: Answers will vary slightly: $0.05 \pm 2*0.056 \Rightarrow 0.05 \pm 0.112 \Rightarrow$ -0.062 to 0.162

    We are 95% sure that the difference in the proportion of smokers between male and female students is between -0.062 and 0.162.
    Difficulty: Medium L.O.: 3.2.3; 3.3.5

56. Use percentiles of your bootstrap distribution to provide a 98% confidence interval for the difference in the proportion of smokers between male and female students. State which percentiles you are using.
Ans: Answers will vary slightly: Using the 1%- and 99%-tiles, a 98% confidence interval for the difference in proportions is -0.078 to 0.189 (based on 5,000 bootstrap samples)
Difficulty: Medium L.O.: 3.4.1

Use the following to answer questions 57-62:

Suppose we are interested in comparing the proportion of male students who smoke to the proportion of female students who smoke. We have a random sample of 150 students (60 males and 90 females) that includes two variables: Smoke = "yes" or "no" and Gender = "female (F)" or "male (M)". The two-way table below summarizes the results.

	Smoke = Yes	Smoke = No	Sample Size
Gender = M	9	51	60
Gender = F	9	81	90

57. If the parameter of interest is the difference in proportions, p_m - p_f, where p_m and p_f represent the proportion of smokers in each gender, find a point estimate for this difference in proportions based on the data in the table. Report your answer with two decimal places.
Ans: $\hat{p}_m - \hat{p}_f$ = 9/60 - 9/90 = 0.15 - 0.10 = 0.05
Difficulty: Medium L.O.: 3.1.2

58. Describe how to use the data to construct a bootstrap distribution. What value should be recorded for each of the bootstrap samples.
Ans: For each bootstrap sample, randomly select 60 males with replacement from the original sample of 60 males and randomly select 90 females with replacement from the original sample of 90 females. To compute the bootstrap statistic for this sample, compute the proportion of smokers among the bootstrap samples of males and females and find the difference ($\hat{p}_m - \hat{p}_f$). Repeat many times (say 1,000 or more times).
Difficulty: Medium L.O.: 3.3.1; 3.4.2

59. Where should the bootstrap distribution be centered?
A) 0.0 B) 0.05 C) 0.10 D) 0.15
Ans: B Difficulty: Medium L.O.: 3.3.2

60. Describe how you would estimate the standard error from the bootstrap distribution.
Ans: Use the standard deviation of the statistics in the bootstrap sample.
Difficulty: Medium L.O.: 3.3.4

61. The standard error is estimated to be 0.056. Find and interpret (in the context of this data situation) a 95% confidence interval for the difference in the proportion of smokers between male and female students, p_m - p_f. Round the margin of error to three decimal places.
    Ans: $0.05 \pm 2*0.056 \Rightarrow 0.05 \pm 0.112 \Rightarrow -0.062$ to 0.162

    We are 95% sure that the difference in the proportion of smokers between male and female students is between -0.062 and 0.162.
    Difficulty: Medium L.O.: 3.2.3

62. Percentiles of the bootstrap distribution (based on 5,000 samples) are provided. Use the percentiles to provide a 98% confidence interval for the difference in the proportion of smokers between male and female students. State which percentiles you are using.

1%	2.5%	5%	10%	25%	75%	90%	95%
-0.078	-0.056	-0.039	-0.022	0.011	0.083	0.122	0.144

    Ans: Use the 1%- and 99%-iles: -0.078 to 0.189
    Difficulty: Medium L.O.: 3.4.1

63. Use the following to answer questions a-e:

To create a confidence interval from a bootstrap distribution using percentiles, we keep the middle values and chop off a certain percentage from each tail. Indicate what percent of values must be chopped off from each tail for each confidence level.

a. 95%
   Ans: 2.5%
   Difficulty: Medium L.O.: 3.4.1

b. 88%
   Ans: 6%
   Difficulty: Medium L.O.: 3.4.1

c. 99%
   Ans: 0.5%
   Difficulty: Medium L.O.: 3.4.1

d. 80%
   Ans: 10%
   Difficulty: Medium L.O.: 3.4.1

e. 96%
   Ans: 2%
   Difficulty: Medium L.O.: 3.4.1

Use the following to answer questions 64-69:

There are 24 students enrolled in an introductory statistics class at a small university. As an in-class exercise the students were asked how many hours of television they watch each week. Their responses, broken down by gender, are summarized in the provided table. Assume that the students enrolled in the statistics class are representative of all students at the university.

Male	3	1	12	12	0	4	10	4	5	5	2	10	10	$\bar{x}_1 = 6$
Female	10	3	2	10	3	2	0	1	6	1	5			$\bar{x}_2 = 3.91$

64. If the parameter of interest is the difference in means, $\mu_m - \mu_f$, where μ_m and μ_f are the mean number of hours spent watching television for males and females at this university, find a point estimate of the parameter based on the available data. Report your answer with two decimal places.
Ans: $\bar{x}_m - \bar{x}_f$ = 6 - 3.91 = 2.09
Difficulty: Medium L.O.: 3.1.2

65. Describe how to use the data to construct a bootstrap distribution. What value should be recorded for each of the bootstrap samples?
Ans: For each bootstrap sample, randomly select 13 males with replacement from the original sample of 13 males and randomly select 11 females with replacement from the original sample of 11 females. Compute the sample mean for males and females and find the difference in those sample means; this difference in sample means is the bootstrap statistic. Repeat this process many times.
Difficulty: Medium L.O.: 3.3.1; 3.4.2

66. Use technology to construct a bootstrap distribution with at least 1,000 samples and estimate the standard error.
Ans: Answers will vary: SE = 1.511 (based on 5,000 bootstrap samples)
Difficulty: Medium L.O.: 3.3.3

67. Use the estimate of the standard error to construct a 95% confidence interval for the difference in the mean number of hours spent watching television for males and females at this university. Round the margin of error to two decimal places.
Ans: Answers may vary: $2.09 \pm 2*1.511 \Rightarrow 2.09 \pm 3.02 \Rightarrow$ -0.93 to 5.11
Difficulty: Medium L.O.: 3.2.3; 3.3.5

68. Interpret your 95% confidence interval in the context of this data situation.
Ans: We are 95% sure that the difference in mean number of hours of TV for males and females at this university is between -0.93 and 5.11 hours.
Difficulty: Medium L.O.: 3.2.4

69. Use percentiles of your bootstrap distribution to provide a 95% confidence interval for the difference in the mean number of hours spent watching television for males and females at this university. Indicate which percentiles you are using.
Ans: Answers may vary: -0.888 to 4.972 (using the 2.5%- and 97.5%-iles)
Difficulty: Medium L.O.: 3.4.1

Use the following to answer questions 70-76:

There are 24 students enrolled in an introductory statistics class at a small university. As an in-class exercise the students were asked how many hours of television they watch each week. Their responses, broken down by gender, are summarized in the provided table. Assume that the students enrolled in the statistics class are representative of all students at the university.

Male	3	1	12	12	0	4	10	4	5	5	2	10	10	$\bar{x}_1 = 6$
Female	10	3	2	10	3	2	0	1	6	1	5			$\bar{x}_2 = 3.91$

70. If the parameter of interest is the difference in means, $\mu_m - \mu_f$, where μ_m and μ_f are the mean number of hours spent watching television for males and females at this university, find a point estimate of the parameter based on the available data. Report your answer with two decimal places.
Ans: $\bar{x}_m - \bar{x}_f$ = 6 - 3.91 = 2.09
Difficulty: Medium L.O.: 3.1.2

71. Describe how to use the data to construct a bootstrap distribution. What value should be recorded for each of the bootstrap samples?
Ans: For each bootstrap sample, randomly select 13 males with replacement from the original sample of 13 males and randomly select 11 females with replacement from the original sample of 11 females. Compute the sample mean for males and females and find the difference in those sample means; this difference in sample means is the bootstrap statistic. Repeat this process many times.
Difficulty: Medium L.O.: 3.3.1; 3.4.2

72. Where should the bootstrap distribution be centered?
A) 0 B) 2.09 C) 3.91 D) 6
Ans: B Difficulty: Medium L.O.: 3.3.2

73. Describe how you would estimate the standard error from the bootstrap distribution.
Ans: Use the standard deviation of the bootstrap distribution.
Difficulty: Medium L.O.: 3.3.4

74. The standard error is estimated to be 1.511 (based on 5,000 bootstrap samples). Find a 95% confidence interval for the difference in the mean number of hours spent watching television for males and females at this university. Round the margin of error to two decimal places.
Ans: $2.09 \pm 2*1.511 \Rightarrow 2.09 \pm 3.02 \Rightarrow$ -0.93 to 5.11
Difficulty: Medium L.O.: 3.2.3; 3.3.5

75. Interpret your 95% confidence interval in the context of this data situation.
Ans: We are 95% sure that the difference in mean number of hours of TV watched per week for males and females at this university is between -0.93 and 5.11 hours.
Difficulty: Medium L.O.: 3.2.4

76. Percentiles of the bootstrap distribution (based on 5,000 samples) are provided. Use the percentiles to provide a 95% confidence interval for the difference in the mean number of hours spent watching television for males and females at this university. Indicate which percentiles you are using.

1%	2.5%	5%	10%	25%	75%	90%	95%
-1.497	-0.888	-0.395	0.189	1.105	3.136	4.056	4.573

Ans: -0.888 to 4.972 (using the 2.5%- and 97.5%-iles)
Difficulty: Medium L.O.: 3.4.1

77. A sample of size 46 with a mean of 13.6 is to be used to construct a confidence interval for μ. A bootstrap distribution based on 1,000 samples is created. Where will the bootstrap distribution be centered?
A) 46 B) 13.6 C) μ D) 1,000
Ans: B Difficulty: Medium L.O.: 3.3.2

Use the following to answer questions 78-84:

November 6, 2012 was election day. Many of the major television networks aired coverage of the incoming election results during the primetime hours. The provided table displays the amount of time (in minutes) spent watching election coverage for a random sample of 25 U.S. adults.

123	120	45	30	40	86	36	52	86
2	70	155	70	168	156	107	126	66
71	97	73	90	69	5	68		

78. What is the population parameter of interest? Define using the appropriate notation.
Ans: μ = mean amount of time (in minutes) U.S. adults spent watching election coverage on election night
Difficulty: Medium L.O.: 3.1.1

79. Use the data from the sample to estimate the parameter of interest. Report your answer with two decimal places.
Ans: $\bar{x}$ = 80.44 minutes
Difficulty: Medium L.O.: 3.1.2

80. Describe how to use the data to construct a bootstrap distribution. What value should be recorded for each of the bootstrap samples?
Ans: For each bootstrap sample, randomly generate a sample of 25 with replacement from the original sample of 25. Compute the sample mean for each sample to use as the bootstrap statistic. Repeat many times.
Difficulty: Medium L.O.: 3.3.1; 3.4.2

81. Use technology to construct a bootstrap distribution with at least 1,000 samples and estimate the standard error.
Ans: Answers may vary: SE = 8.769 (based on 5,000 bootstrap samples)
Difficulty: Medium L.O.: 3.3.3; 3.3.4

82. Use the estimate of the standard error to construct a 95% confidence interval for the mean amount of time (in minutes) U.S. adults spent watching election coverage on election night. Use three decimal places in your answer.
Ans: Answers may vary: $80.44 \pm 2*8.769 \Rightarrow 80.44 \pm 17.538 \Rightarrow 62.902$ to 97.978 minutes
Difficulty: Medium L.O.: 3.2.3; 3.3.4

83. Use the percentiles of your bootstrap distribution to provide a 92% confidence interval for the mean amount of time (in minutes) U.S. adults spent watching election coverage on election night. Indicate which percentiles you are using.
Ans: Answers will vary: 65.160 to 95.78 (based on 5,000 bootstrap samples); use the 4%- and 96%-iles.
Difficulty: Medium L.O.: 3.4.1

84. Interpret your 92% confidence interval in the context of this data situation.
Ans: We are 92% sure that the mean amount of time (in minutes) U.S. adults spent watching election coverage on election night is between 65.16 and 95.78 minutes.
Difficulty: Medium L.O.: 3.2.4

Use the following to answer questions 85-92:

November 6, 2012 was election day. Many of the major television networks aired coverage of the incoming election results during the primetime hours. The provided table displays the amount of time (in minutes) spent watching election coverage for a random sample of 25 U.S. adults.

123	120	45	30	40	86	36	52	86
2	70	155	70	168	156	107	126	66
71	97	73	90	69	5	68		

85. What is the population parameter of interest? Define using the appropriate notation.
Ans: μ = mean amount of time (in minutes) U.S. adults spent watching election coverage on election night
Difficulty: Medium L.O.: 3.1.1

86. Use the data from the sample to estimate the parameter of interest. Report your answer with two decimal places.
Ans: $\bar{x}$ = 80.44 minutes
Difficulty: Medium L.O.: 3.1.2

87. Describe how to use the data to construct a bootstrap distribution. What value should be recorded for each of the bootstrap samples?
Ans: For each bootstrap sample, randomly generate a sample of 25 with replacement from the original sample of 25. Compute the sample mean for each sample to use as the bootstrap statistic. Repeat many times.
Difficulty: Medium L.O.: 3.3.1; 3.4.2

88. Where should the bootstrap distribution be centered?
A) 25 B) 60 C) 80.44 D) 100
Ans: C Difficulty: Medium L.O.: 3.3.2

89. Describe how you would estimate the standard error from the bootstrap distribution.
Ans: Use the standard deviation of the bootstrap distribution.
Difficulty: Medium L.O.: 3.3.4

90. The standard error is estimated to be 8.769 (based on 5,000 bootstrap samples). Find a 95% confidence interval for the mean amount of time (in minutes) U.S. adults spent watching election coverage on election night. Round the margin of error to two decimal places.
Ans: $80.44 \pm 2*8.769 \Rightarrow 80.44 \pm 17.54 \Rightarrow 62.90$ to 97.98 minutes
Difficulty: Medium L.O.: 3.2.3; 3.3.5

91. Percentiles of the bootstrap distribution (based on 5,000 samples) are provided. Use the percentiles to provide a 92% confidence interval for the mean amount of time (in minutes) U.S. adults spent watching election coverage on election night. Indicate which percentiles you are using.

2%	4%	6%	8%	92%	94%	96%	98%
63.000	65.160	66.880	68.240	92.740	94.080	95.780	98.54

Ans: 65.160 to 95.780 minutes (use the 4%- and 96%-iles)
Difficulty: Medium L.O.: 3.4.1

92. Interpret your 92% confidence interval in the context of this data situation.
    Ans: We are 92% sure that the mean amount of time (in minutes) U.S. adults spent watching election coverage on election night is between 65.16 and 95.78 minutes.
    Difficulty: Medium L.O.: 3.2.4

Use the following to answer questions 93-100:

A 2009 study to investigate the dominant paws in cats was described in *Animal Behaviour* (Volume 78, Issue 2). The researchers used a random sample of 42 domestic cats. In this study, each cat was shown a treat (5 grams of tuna), and while the cat watched, the food was placed inside a jar. The opening of the jar was small enough that the cat could not stick its head inside to remove the treat. The researcher recorded the paw that was first used by the cat to try to retrieve the treat. This was repeated 100 times for each cat (over a span of several days). The paw used most often was deemed the dominant paw (note that one cat used both paws equally and was classified as "ambidextrous"). Of the 42 cats studied, 20 were classified as "left-pawed".

93. What is the population parameter of interest? Define using the appropriate notation.
    Ans: p = proportion of domestic cats that are "left-pawed"
    Difficulty: Medium L.O.: 3.1.1

94. Use the data from the sample to estimate the parameter of interest. Report your answer with three decimal places.
    Ans: $\hat{p}$ = 20/42 = 0.476
    Difficulty: Easy L.O.: 3.1.2

95. Describe how to use the data to construct a bootstrap distribution. What value should be recorded for each of the bootstrap samples?
    Ans: For each bootstrap sample, randomly generate a sample of size 42 cats with replacement from the original sample of 42 cats. For each bootstrap sample, compute the sample proportion of "left-pawed" cats - this is the bootstrap statistic.
    Difficulty: Medium L.O.: 3.3.1; 3.4.2

96. Use technology to construct a bootstrap distribution with at least 1,000 samples and estimate the standard error.
    Ans: Answers will vary: SE = 0.078 (based on 5,000 bootstrap samples)
    Difficulty: Medium L.O.: 3.3.3; 3.3.4

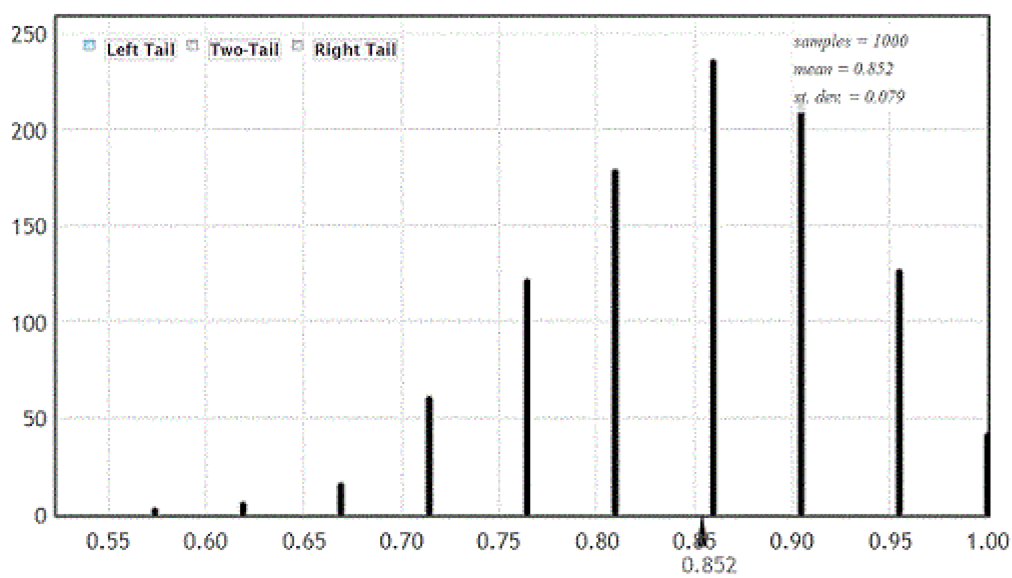
97. Use the estimate of the standard error to construct a 95% confidence interval for the proportion of domestic cats that are "left-pawed". Round the margin of error to three decimal places.
    Ans: Answers will vary: $0.476 \pm 2*0.078 \Rightarrow 0.476 \pm 0.156 \Rightarrow 0.320$ to 0.632
    Difficulty: Medium L.O.: 3.2.3; 3.3.5

98. Use the percentiles of your bootstrap distribution to provide a 99% confidence interval for the parameter. Indicate the percentiles that you use.
Ans: Answers will vary: 0.262 to 0.667 (using the 0.5%- and 99.5-iles)
Difficulty: Medium L.O.: 3.4.1

99. Provide an interpretation of your 99% confidence interval in the context of this data situation.
Ans: We are 99% sure that the proportion of domestic cats that are "left-pawed" is between 0.262 and 0.667.
Difficulty: Medium L.O.: 3.2.4

100. The researchers were also interested in comparing the proportion of "left-pawed" cats for male and female cats. Of the 21 male cats in the sample, 19 were classified as "left-pawed" while only 1 of the 21 female cats was considered to be "left-pawed".

A bootstrap distribution (based on 1,000 bootstrap samples) for difference in the proportion of "left-pawed" cats is provided. Would it be appropriate to use this bootstrap distribution to construct a confidence interval for the difference in the proportion of male and female cats that are "left-pawed"? Briefly explain.



Ans: No. The bootstrap distribution is not quite symmetric
Difficulty: Challenging L.O.: 3.4.4

Use the following to answer questions 101-109:

A 2009 study to investigate the dominant paws in cats was described in *Animal Behaviour* (Volume 78, Issue 2). The researchers used a random sample of 42 domestic cats. In this study, each cat was shown a treat (5 grams of tuna), and while the cat watched, the food was placed inside a jar. The opening of the jar was small enough that the cat could not stick its head inside to remove the treat. The researcher recorded the paw that was first used by the cat to try to retrieve the treat. This was repeated 100 times for each cat (over a span of several days). The paw used most often was deemed the dominant paw (note that one cat used both paws equally and was classified as "ambidextrous"). Of the 42 cats studied, 20 were classified as "left-pawed".

101. What is the population parameter of interest? Define using the appropriate notation.
Ans: p = proportion of domestic cats that are "left-pawed"
Difficulty: Medium L.O.: 3.1.1

102. Use the data from the sample to estimate the parameter of interest. Round your answer to three decimal places.
Ans: $\hat{p}$ = 20/42 = 0.476
Difficulty: Easy L.O.: 3.1.2

103. Describe how to use the data to construct a bootstrap distribution. What value should be recorded for each of the bootstrap samples?
Ans: For each bootstrap sample, randomly generate a sample of size 42 cats with replacement from the original sample of 42 cats. For each bootstrap sample, compute the sample proportion of "left-pawed" cats - this is the bootstrap statistic.
Difficulty: Medium L.O.: 3.3.1; 3.4.2

104. Where should the bootstrap distribution be centered?
A) 0.476 B) 20 C) 42 D) 0.95
Ans: A Difficulty: Medium L.O.: 3.3.2

105. Describe how you would estimate the standard error from the bootstrap distribution.
Ans: Use the standard deviation of the bootstrap distribution.
Difficulty: Medium L.O.: 3.3.4

106. The standard error is estimated to be 0.078 (based on 5,000 bootstrap samples). Find a 95% confidence interval for the proportion of domestic cats that are "left-pawed". Round the margin of error to three decimal places.
Ans: $0.476 \pm 2*0.078 \Rightarrow 0.476 \pm 0.156 \Rightarrow 0.32$ to 0.632
Difficulty: Medium L.O.: 3.2.3; 3.3.5

107. Percentiles of the bootstrap distribution (based on 5,000 samples) are provided. Use the percentiles to provide a 99% confidence interval for the parameter. Indicate the percentiles that you use.

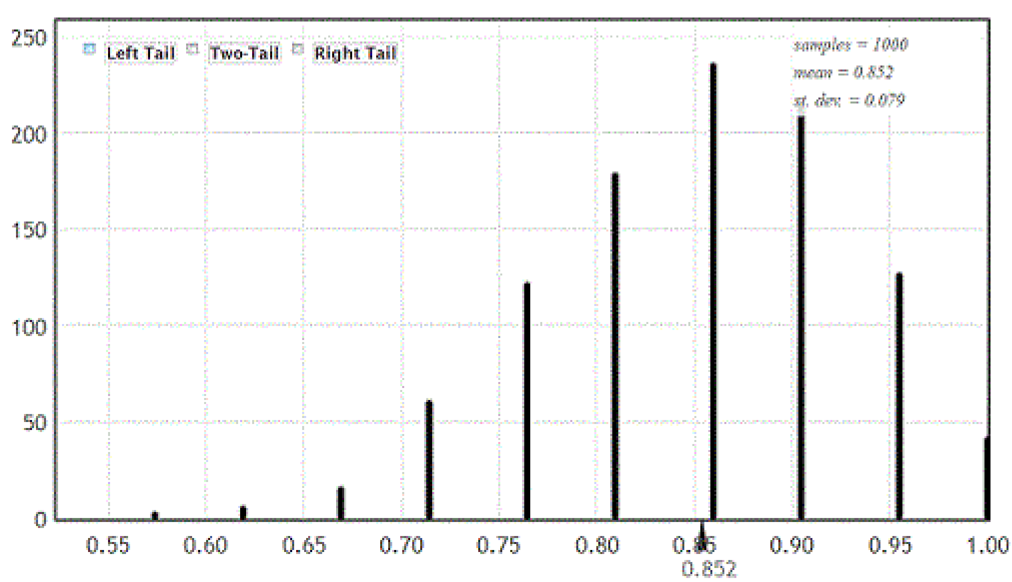
0.5%	1%	2.5%	5%	95%	97.5%	99%	99.5%
0.262	0.286	0.310	0.357	0.595	0.619	0.643	0.667

Ans: 0.262 to 0.667 (use 0.5%- and 99.5%-iles)
Difficulty: Medium L.O.: 3.4.1

108. Provide an interpretation of your 99% confidence interval in the context of this data situation.
Ans: We are 99% sure that the proportion of domestic cats that are "left-pawed" is between 0.262 and 0.667.
Difficulty: Medium L.O.: 3.2.4

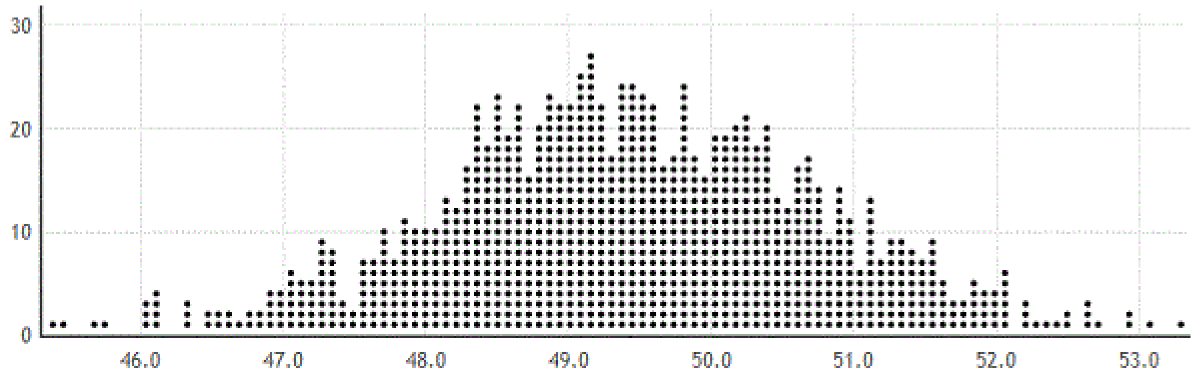
109. The researchers were also interested in comparing the proportion of "left-pawed" cats for male and female cats. Of the 21 male cats in the sample, 19 were classified as "left-pawed" while only 1 of the 21 female cats was considered to be "left-pawed".

A bootstrap distribution (based on 1,000 bootstrap samples) for difference in the proportion of "left-pawed" cats is provided. Would it be appropriate to use this bootstrap distribution to construct a confidence interval for the difference in the proportion of male and female cats that are "left-pawed"? Briefly explain.



Ans: No. The bootstrap distribution is not quite symmetric.
Difficulty: Challenging L.O.: 3.4.4

110. A bootstrap distribution, based on 1,000 bootstrap samples is provided. Use the distribution to estimate a 99% confidence interval for the population mean. Explain how you arrived at your answer.



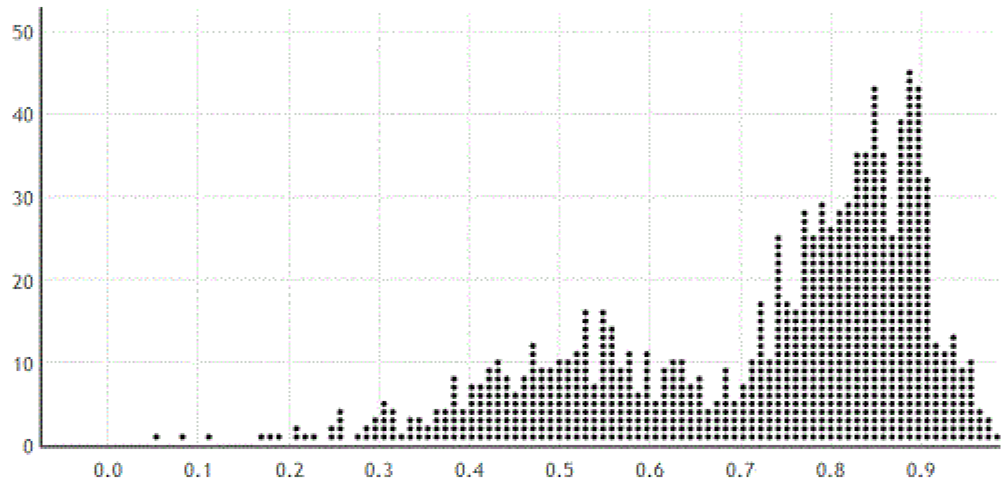
Ans: Answers may vary slightly: Since we are interested in a 99% confidence interval, 0.5% of the points should be in each tails of the distribution (i.e., 0.5% of 1,000 is 1,000(0.005) = 5 dots). The approximate cut-offs that leave 5 dots in each tail are 46 and 52.6. (In Statkey the actual cut-offs are 45.994 and 52.642.)

Difficulty: Challenging L.O.: 3.4.1

Use the following to answer questions 111-112:

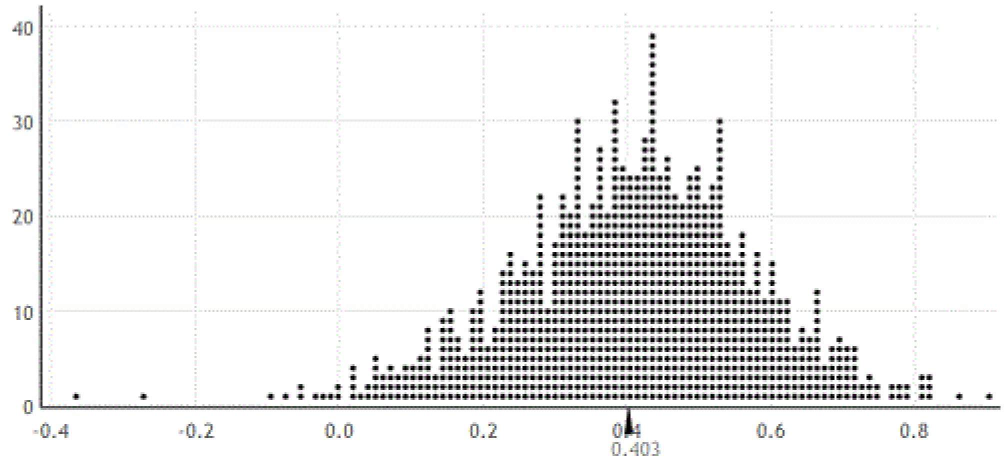
A biologist collected data on a random sample of porcupines. She wants to estimate the correlation between the body mass of a porcupine (in grams) and the length of the porcupine (in cm).

111. Her sample consists of 20 porcupines. A bootstrap distribution for the correlation between body mass and length (based on 1,000 samples) is provided. Would it be appropriate to use this bootstrap distribution to estimate a 95% confidence interval for the correlation between body mass and length of porcupines? Explain briefly.



Ans: No. The bootstrap distribution is extremely skewed.
Difficulty: Medium L.O.: 3.4.4

112. The biologist noted that two of the porcupines were much smaller than the others, and thus they were likely not "adults". Since she is only interested in adult porcupines, the biologist wants to use the 18 adults to estimate the correlation between body mass and body length. The sample correlation is 0.407. Her bootstrap distribution is provided. The standard error is estimated to be 0.165.



If appropriate, construct and interpret a 95% confidence interval for the correlation between body mass and body length for adult porcupines (with the margin of error rounded to three decimal places). If not appropriate, explain why not.

Ans: Because the bootstrap distribution is roughly symmetric, it is appropriate to use the bootstrap distribution to construct a confidence interval.

$0.407 \pm 2*0.165 \Rightarrow 0.407 \pm 0.33 \Rightarrow 0.077$ to 0.737

We are 95% sure that the correlation between body mass and body length for adult porcupines is between 0.077 and 0.737.

Difficulty: Medium L.O.: 3.2.3; 3.2.4; 3.3.5; 3.4.4

Use the following to answer questions 113-118:

In a survey conducted by the Gallup organization September 6-9, 2012, 1,017 adults were asked "In general, how much trust and confidence do you have in the mass media - such as newspapers, TV, and radio - when it comes to reporting the news fully, accurately, and fairly?" 81 said that they had a "great deal" of confidence, 325 said they had a "fair amount" of confidence, 397 said they had "not very much" confidence, and 214 said they had "no confidence at all".

113. Suppose the parameter of interest is the proportion of U.S. adults who have "no confidence at all" in the media. Use the data to find an estimate of this parameter. Report your answer with two decimal places.

Ans: $\hat{p}$ = 214/(81+325+397+214) = 214/1,017 = 0.21

Difficulty: Medium L.O.: 3.1.2

114. Describe how to use the data to construct a bootstrap distribution. What value should be recorded for each of the bootstrap samples?
Ans: For a single bootstrap sample, generate a sample of 1,017 responses with replacement from the original sample of responses. For each sample, compute the sample proportion of responses that are "no confidence at all" - this is the bootstrap statistic. Repeat many times.
Difficulty: Medium L.O.: 3.3.1; 3.4.2

115. Use technology to construct a bootstrap distribution with at least 1,000 samples and estimate the standard error.
Ans: Answers may vary: SE = 0.013 (based on 5,000 bootstrap samples)
Difficulty: Medium L.O.: 3.3.3; 3.3.4

116. Use the estimate of the standard error to construct a 95% confidence interval for the proportion of U.S. adults who have no confidence in the media. Round the margin of error to three decimal places.
Ans: Answers may vary: $0.21 \pm 2*0.013 \Rightarrow 0.21 \pm 0.026 \Rightarrow 0.184$ to 0.236
Difficulty: Medium L.O.: 3.2.3; 3.3.5

117. Provide an interpretation of your 95% confidence interval in the context of this data situation.
Ans: We are 95% sure that the proportion of U.S. adults who have no confidence in the media is between 0.184 and 0.236.
Difficulty: Medium L.O.: 3.2.4

118. Use percentiles of your bootstrap distribution to provide a 95% confidence interval for the proportion of U.S. adults who have no confidence in the media. Indicate the percentiles that you use.
Ans: Answers may vary: 0.186 to 0.235 (based on 5,000 bootstrap samples) using the 2.5%- and 97.5%-tiles.
Difficulty: Medium L.O.: 3.4.1

Use the following to answer questions 119-124:

In a survey conducted by the Gallup organization September 6-9, 2012, 1,017 adults were asked "In general, how much trust and confidence do you have in the mass media - such as newspapers, TV, and radio - when it comes to reporting the news fully, accurately, and fairly?" 81 said that they had a "great deal" of confidence, 325 said they had a "fair amount" of confidence, 397 said they had "not very much" confidence, and 214 said they had "no confidence at all".

119. Suppose the parameter of interest is the proportion of U.S. adults who have "no confidence at all" in the media. Use the data to find an estimate of this parameter. Report your answer with two decimal places.
Ans: $\hat{p} = 214/(81+325+397+214) = 214/1017 = 0.21$
Difficulty: Medium L.O.: 3.1.2

120. Describe how to use the data to construct a bootstrap distribution. What value should be recorded for each of the bootstrap samples?

Ans: For a single bootstrap sample, generate a sample of 1,017 responses with replacement from the original sample of responses. For each sample, compute the sample proportion of responses that are "no confidence at all" - this is the bootstrap statistic. Repeat many times.

Difficulty: Medium L.O.: 3.3.1; 3.4.2

121. Describe how you would estimate the standard error from the bootstrap distribution.

Ans: Use the standard deviation of the bootstrap distribution.

Difficulty: Medium L.O.: 3.3.4

122. The estimate of the standard error is 0.013. Use the estimate of the standard error to construct a 95% confidence interval for the proportion of U.S. adults who have no confidence in the media. Round the margin of error to three decimal places.

Ans: $0.21 \pm 2*0.013 \Rightarrow 0.21 \pm 0.026 \Rightarrow 0.184$ to 0.236

Difficulty: Medium L.O.: 3.2.3; 3.3.5

123. Provide an interpretation of your 95% confidence interval in the context of this data situation.

Ans: We are 95% sure that the proportion of U.S. adults who have no confidence in the media is between 0.184 and 0.236.

Difficulty: Medium L.O.: 3.2.4

124. Percentiles of the bootstrap distribution (based on 5,000 samples) are provided. Use the percentiles to provide a 95% confidence interval for the proportion of U.S. adults who have no confidence in the media. Indicate the percentiles that you use.

1%	2.5%	5%	10%	90%	95%	97.5%	99%
0.181	0.186	0.190	0.194	0.227	0.231	0.235	0.239

Ans: 0.186 to 0.235 (using the 2.5%- and 97.5%-tiles)

Difficulty: Medium L.O.: 3.4.1

125. A sample of $n = 10$ Illinois gas stations taken on August 8, 2012 had an average price of \$3.975 per gallon. The national average at this time was \$3.63. If we want to use the sample data to construct a 95% confidence interval for the average gas price in Illinois (on August 8, 2012), where would the bootstrap distribution be centered?

A) 3.63 B) 3.80 C) 3.975 D) 10

Ans: C Difficulty: Medium L.O.: 3.3.2

126. In a dotplot of a bootstrap distribution, the number of dots should match the size of the original sample.

Ans: False Difficulty: Medium L.O.: 3.3.0; 3.4.2

127. A bootstrap distribution will be centered at the value of the original statistic.
Ans: True Difficulty: Medium L.O.: 3.3.2