# Answer Key

**I. A.**

**1.** The population is all employee sick days. The sample is 500 randomly selected sick days. The characteristic of interest is whether each sick day occurred on a Friday.

**2.** The population is all of the people who live in the town. The sample is the 100 randomly selected town residents. The characteristic of interest is the number of years they have lived in the town.

**3.** The population is SAT scores of students who take the course at the learning center. The sample is the SAT scores of the 145 randomly selected students. The characteristic of interest is whether taking the prep course improves student SAT scores.

**4.** The population is all of the items that come off of the assembly line. The sample is every 5000$^{th}$ item that comes off the line in a 24-hour period. The characteristic of interest is whether each item is defective.

**5.** The population is American adults. The sample is 4565 American adults. The characteristic of interest is the link between BMI and the risk of getting the common cold.

**6.** The population is all children. The sample is 435 children. The characteristic of interest is whether the amount of sugar in the company's yogurt has a significant effect on its taste.

**I. B.**

**1.** This is an experiment since the farmer imposed a treatment, the nutritional supplement. There are two treatments. One treatment is giving the supplement to cows. The other treatment is not giving the supplement to cows.

**2.** This is an observational study because the principal gathered existing data from its natural setting.

**3.** This is a sample survey because the students were asked to respond to a specific question.

**4.** This is a sample survey because the voters were asked to respond to a specific question.

**5.** This is an observational study because the data was gathered from its natural setting.

**6.** This is an experiment because the researcher imposed a treatment, classical music. There are two treatments. One treatment is the group of students listening to classical music and the other treatment is the group of students not listening to classical music.

**I. C.**

**1.** Confounding could occur because kids have less allergies due to various other reasons, such as having parents who do not have allergies.

**2.** Confounding could occur because the sample of adults with depression could get better during the four-month period due to another factor, such as a change of seasons.

**3.** Confounding could occur because the frequency of leg cramps could be due to other factors, such as vitamin deficiencies or lack of sleep.

**4.** Confounding could occur because productivity could be different due to another factor, such as the type of work performed.

**5.** Confounding could occur because aggressive behavior in children could be related to other factors, such as poor nutrition.

**6.** Confounding could occur because educational success could be affected by other factors, such as the amount of educational capital in the school district.

**I. D.**

**1.** Answers will vary. There is no bias in this study.

**2.** Answers will vary. There is bias in this study because the voters in Albany who make less than $100,000 are not represented.

**3.** Answers will vary. There is bias in this study because the side effects of healthy patients could be different from the side effects of unhealthy patients.

**4.** Answers will vary. There is bias in this study because side effects of sick patients could be different from side effects of healthy patients.

**5.** Answers will vary. There is bias in this study because only people who visit the department of education's website have an opportunity to respond.

**6.** Answers will vary. There is bias in this study because the people who read the magazine are more likely to like the magazine.

**7.** Answers will vary. The sample could be biased because they only asked adults.

**8.** Answers will vary. The sample could be biased because only one class is surveyed. Also, math students may have different classroom needs for the cell phone than other subject-matter classes.

**9.** Answers will vary. The sample could be biased because only one class is surveyed. Also, English students probably do not need a graphing calculator.

**10.** Answers will vary. The sample could be biased because students who did not volunteer may be underrepresented.

**11.** Answers will vary. The sample could be biased because only 25 students responded. They may not be a representative sample of all students.

**12.** Answers will vary. The sample could be biased because only students passing her room responded. All students did not have the same opportunity to respond.

**I. E.**

**1.** Answers will vary. I chose an observational study because teacher salaries are public records in my city. To obtain a random sample, I could assign an identification number to each teacher and use a computer to randomly generate a sample of teachers.

**2.** Answers will vary. I chose an observational study because teacher salaries and teacher genders are public records in my city. To obtain a random sample, I could assign an identification number to each teacher, divide the population of teachers by gender, then use a computer to randomly select a sample from each gender group.

**3.** Answers will vary. I chose a sample survey to gather data because students are the best source for accurate information about their study habits. To obtain a random sample, I could randomly select one boy and one girl from every homeroom of every grade.

**4.** Answers will vary. I chose a sample survey to gather data because health records are private and the individuals have to consent to sharing information. To obtain a random sample, I could systematically select every 20th address in the town's directory and send the sample surveys to those addresses.

**5.** Answers will vary. I chose a cluster sample to gather data. To obtain a random sample, I could arrange all of the state's towns in clusters of three, then randomly select one cluster for analysis.

**6.** Answers will vary. I chose an experiment to gather data because two treatments can be used. One treatment is using shorter, more frequent study sessions and the other treatment is using longer, less frequent study sessions. To obtain a random sample, I could assign identification numbers to students and use a random number generator.

**II. A.**

**1.** Answers will vary. I chose a stratified random sampling to collect data. I could divide the population of the town into groups according to their geographic location (possibly north, south, east, west), and randomly select members from each group.

**2.** Answers will vary. I chose cluster sampling to collect data. During lunch, I could randomly choose one boy and one girl from each table of the lunch room.

**3.** Answers will vary. I chose a stratified random sampling to collect data. During the 9th grade lunch, I could randomly choose one boy and one girl from each table of the lunch room. I could repeat this method during the 11th grade lunch.

**4.** Answers will vary. I chose cluster sampling to collect data. I could divide the neighborhoods into groups of three, then randomly select one of the groups.

**5.** Answers will vary. I chose stratified random sampling to collect data. I could divide the city into neighborhoods, then randomly survey a small sample from each neighborhood.

**6.** Answers will vary. I chose random sampling to collect data. I could randomly select people that walk into the supermarket and ask them to participate in a taste test.

**II. B.**

**1.** Answers will vary. I chose 54, 75, 76, and 95. Fifty-four and 95 are the minimum and maximum values. Seventy-five and 76 are in the middle, between the minimum and maximum values.

**2.** Answers will vary. I chose 130, 203, 215, and 250. One hundred thirty and 250 are the minimum and maximum values. Two hundred three and 215 are in the middle, between the minimum and maximum values.

**3.** Answers will vary. I chose 2, 1, 0, and 3 because they are the first four values in the table and seem to be representative of the entire data set.

**4.** Answers will vary. I chose 7.95, 10.00, 13.27, and 14.75 because they seem to be representative of the mean of the entire data set.

**5.** Answers will vary. I chose 0.193, 0.215, 0.220, and 0.254 because they are the middle four values of the ordered data.

**6.** Answers will vary. I chose 54, 65, 76, and 85 because they are the middle four values of the ordered data.

**II. C.**

**1.** Answers will vary. The calculator generated the numbers 2, 3, 4, and 10. The random sample consists of Easton, Wu, Rodriguez, and Siegelman.

**2.** Answers will vary.

**3.** Answers will vary.

**4.** Answers will vary.

**5.** Answers will vary.

**6.** Answers will vary.

**II. D.**

**1a.** Answers will vary. My sample consists of the temperatures 81, 103, 97, and 98. I randomly selected one temperature from each continent.

**1b.** Answers will vary. My sample consists of the temperatures 92, 95, 81, 103, 94, 97, 101, and 98. I randomly selected two temperatures from each continent.

**2a.** Answers will vary. My sample consists of 4, 11, 7, 6, 8, 12, 8, 4, 9, 8, 0, and 14. I randomly selected the number of books read by 3 different students from each teacher's class.

**2b.** Answers will vary. My sample consists of 2, 3, 7, 11, 6, 13, 8, 3, 8, 4, 9, 4, 8, 4, 0, and 8. I randomly selected the number of books read by 4 different students from each teacher's class.

**3a.** Answers will vary. My sample consists of 213, 302, 312, 251, 234, 281, 324, 284, 184, 279, 314, 313, 192, 264, 319, and 256. I randomly selected the number of cars crossing at 4 different times from each day.

**3b.** Answers will vary. My sample consists of 124, 213, 302, 312, 297, 251, 249, 281, 342, 324, 284, 264, 184, 253, 279, 349, 313, 255, 192, 268, 319, 368, 305, and 279. I randomly selected the number of cars crossing at 6 different times from each day.

**4a.** Answers will vary. My sample consists of 16, 23, 24, 29, and 28. I randomly selected 1 volunteer day for each doctor and recorded the number of patient visits for that day.

**4b.** Answers will vary. My sample consists of 15, 26, 27, 18, 23, 27, 13, 24, 26, 19, 18, 26, 23, 27, and 18. I randomly selected 3 volunteer days for each doctor and recorded the number of patient visits for each day.

**5a.** Answers will vary. My sample consists of 426, 425, 482, 398, 431, 425, 324, 543, 274, 246, 134, and 176. I randomly selected 2 amounts of recycled materials from each of the 6 neighborhoods.

**5b.** Answers will vary. My sample consists of 426, 425, 441, 482, 324, 398, 431, 425, 486, 675, 573, 543, 274, 184, 234, 162, 176, and 186. I randomly selected 3 amounts of recycled materials from each of the 6 neighborhoods.

**6a.** Answers will vary. My sample consists of 42, 49, 38, 41, 56, 52, 57, 53, 86, 87, 93, 87, 112, 126, 118, and 116. I randomly selected 4 values from each of the 4 days of the week.

**6b.** Answers will vary. My sample consists of 43, 49, 48, 38, 46, 56, 57, 52, 57, 52, 86, 92, 93, 96, 87, 124, 126, 118, 116, and 128. I randomly selected 5 values from each of the 4 days of the week.

**II. E.**

**1.** Answers will vary. My two samples consisted of the polar bear weights from the San Diego Zoo, {892, 1216, 1375}, and the polar bear weights from the Philadelphia Zoo, {884, 1237, 1384}. The weights from any one of the zoos are acceptable as a cluster sample.

**2.** Answers will vary. My two samples consisted of the attendance numbers on Monday, {134, 142, 197, 223}, and the attendance numbers on Tuesday, {94, 134, 213, 227}. The attendance numbers from any one of the weekdays are acceptable as a cluster sample.

**3.** Answers will vary. My two samples consisted of the game durations for the A's, {123, 126, 136, 126, 148}, and the game durations for the Bulldogs, {108, 123, 129, 128, 149}. The times from any one of the teams are acceptable as a cluster sample.

**4.** Answers will vary. My two samples consisted of Anastasia's scores, {82, 72, 71, 78, 72}, and Hugo's scores, {85, 78, 69, 82, 73}. The scores from any one of the students are acceptable as a cluster sample.

**5.** Answers will vary. My two samples consisted of the weekday customers from week 2, {145, 88, 104, 174, 197}, and the weekday customers from week 3, {94, 96, 93, 163, 204}. The numbers of customers from any one of the weeks are acceptable as a cluster sample.

**6.** Answers will vary. My two samples consisted of the weekday stock prices from week 1, {2.89, 3.81, 1.48, 3.86, 2.27}, and the weekday stock prices from week 4, {2.42, 1.77, 0.96, 3.52, 0.91}. The prices from any one of the weeks are acceptable as a cluster sample.

**III. A.**

**1.** This is a 95% confidence interval because the margin of error is 2 times the standard deviation of the sampling distribution. $2(0.015) = 0.030$ or 3%

**2.** This is a 68% confidence interval because the margin of error is the same as the standard deviation of the sampling distribution.

**3.** This is a 99.7% confidence interval because the margin of error is 3 times the standard deviation of the sampling distribution. $3(0.02) = 0.06$ or 6%

**4.** This is a 95% confidence interval because the margin of error is 2 times the standard deviation of the sampling distribution. $2(7.65) = 15.3$

**5.** This is a 68% confidence interval because the margin of error is the same as the standard deviation of the sampling distribution.

**6.** This is a 99.7% confidence interval because the margin of error is 3 times the standard deviation of the sampling distribution. $3(0.193) \approx 0.58$

**III. B.**

**1.** The interval from 72.1% to 83.9% represents a 95% confidence interval for the population proportion.

**2.** The interval from 84% to 90% represents a 95% confidence interval for the population proportion.

**3.** The interval from 3% to 5% represents a 95% confidence interval for the population proportion.

**4.** The interval from 72% to 78% represents a 95% confidence interval for the population proportion.

**5.** The interval from 57.5% to 72.5% represents a 95% confidence interval for the population proportion.

**6.** The interval from 12.1% to 15.9% represents a 95% confidence interval for the population proportion.



**III. C.**

**1.** The interval from 34.3 minutes to 34.67 minutes represents a 95% confidence interval for the population mean.

**2.** The interval from 6.08 pounds to 6.12 pounds represents a 95% confidence interval for the population mean.

**3.** The interval from 3.46 hours to 3.54 hours represents a 95% confidence interval for the population mean.

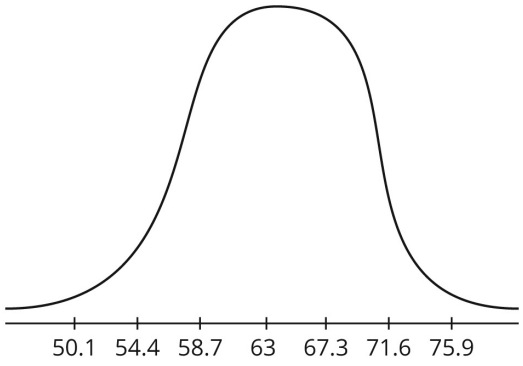
**4.** The interval from 20.6 inches to 21 inches represents a 95% confidence interval for the population mean.

**5.** The interval from 93.54 kWh to 94.46 kWh represents a 95% confidence interval for the population mean.

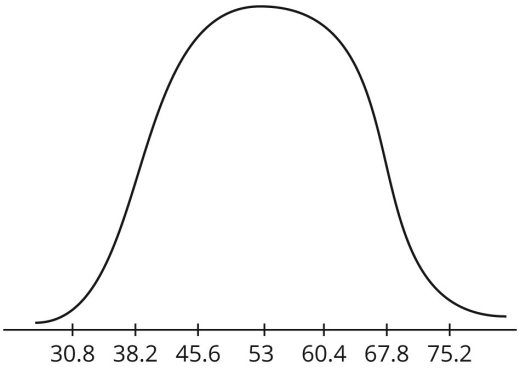
**6.** The interval from $21.60 to $22.40 represents a 95% confidence interval for the population mean.

**IV. A.**

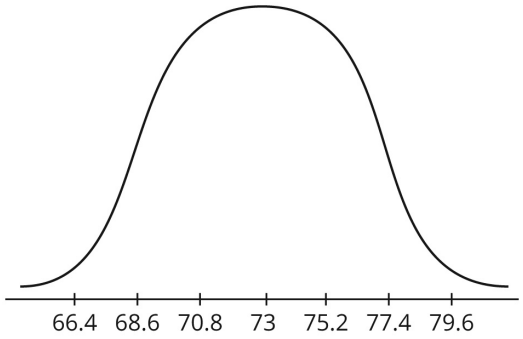
**1.** Sample proportion values less than 54.4% and greater than 71.6% are statistically significant because those values are outside of the 95% confidence interval.



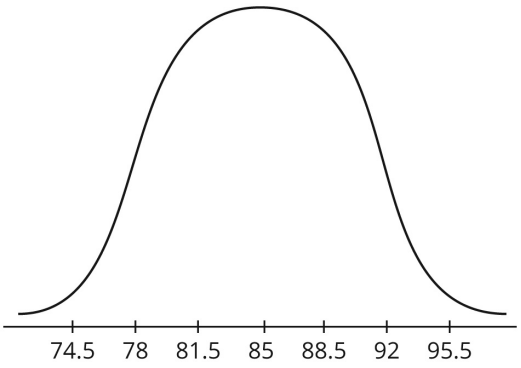

**2.** Sample proportion values less than 38.2% and greater than 67.8% are statistically significant because those values are outside of the 95% confidence interval.



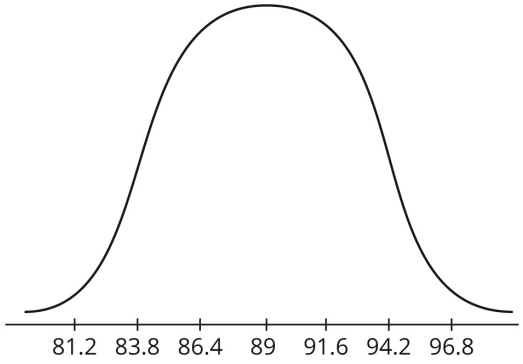

**3.** Sample proportion values less than 68.6% and greater than 77.4% are statistically significant because those values are outside of the 95% confidence interval.



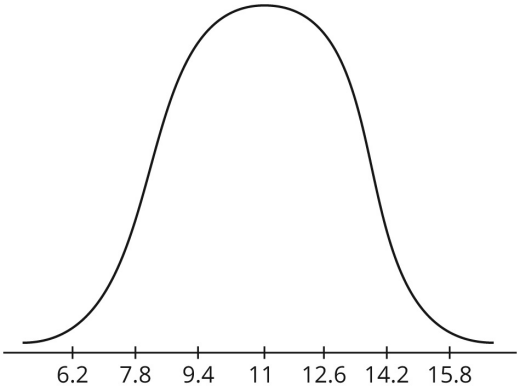

**4.** Sample proportion values less than 78% and greater than 92% are statistically significant because those values are outside of the 95% confidence interval.



**5.** Sample proportion values less than 83.8% and greater than 94.2% are statistically significant because those values are outside of the 95% confidence interval.




**6.** Sample proportion values less than 7.8% and greater than 14.2% are statistically significant because those values are outside of the 95% confidence interval.




**IV. B.**

**1.** The interval from 6.0% to 11.3% represents a 95% confidence interval for the population proportion of users who experienced constant headaches when using the nose spray. Because the manufacturer's claim, 5%, is not within the 95% confidence interval, the results are statistically significant and their claim is unlikely.

**2.** The difference between the two population proportion estimates of viewers is statistically significant because the two 95% confidence intervals do not overlap. The 95% confidence interval for the population proportion of Tuesday night viewership ranges from 41.9% to 48.1%.

The 95% confidence interval for the population proportion of Thursday night viewership ranges from 49.2% to 54.9%.

**3.** The difference between the two population proportion estimates of voters is not statistically significant because the two 95% confidence intervals overlap. The 95% confidence interval for the population proportion of voters supporting John before the ad is run ranges from 33.5% to 38.5%.

The 95% confidence interval for the population proportion of voters supporting John after the ad is run ranges from 35.4% to 40.6%.

**4.** The difference between the two population proportion estimates of TV viewers is not statistically significant because the two 95% confidence intervals overlap. The 95% confidence interval for the population proportion of girls who watch more than 2 hours of TV a day ranges from 63.1% to 73.3%.

The 95% confidence interval for the population proportion boys who watch more than 2 hours of TV a day ranges from 70.0% to 79.7%.

**5.** The interval from 14.3% to 17.1% represents a 95% confidence interval for the population proportion of rural college students who commute.

Because the percent of city college students who commute, 22%, is not within the 95% confidence interval, the results are statistically significant.

**6.** The difference between the two population proportion estimates of passing test scores is not statistically significant because the two 95% confidence intervals overlap.

The 95% confidence interval for the population proportion of passing test scores using the first method ranges from 68.9% to 76.5%.

The 95% confidence interval for the population proportion of passing test scores using the second method ranges from 70.2% to 76.9%.

**IV. C.**

**1.** The range of values for the population mean of each difference overlap, so there is not any statistical significance to the results. It is not likely that there is a correlation between using the additive and increasing gas mileage.

**2.** The data suggests that there is a statistically significant difference between the two shipping methods because the two confidence intervals do not overlap. It is likely that Method A is linked to shorter delivery times.

**3.** The data does suggest a possible link between class times and test scores because the two confidence intervals do not overlap. It cannot be stated that having to get up earlier will cause the students to do worse on the test, but there is statistical evidence of a link between test scores and class times.

**4.** The new method does not seem to decrease the wait time for riders because the two confidence intervals overlap. The type of method does not appear to be linked to the wait time for the riders.

**5.** The data does suggest a possible link between weight loss and working with a health coach because the two confidence intervals do not overlap. The adults who worked with a health coach lost more weight than those who did not work with a health coach.

**6.** The data does not suggest a possible link between gender and amount of sleep for adults because the two confidence intervals overlap. A correlation between gender and amount of sleep for adults is unlikely.