Section 2A

1.

N: Parameter describing the size of a population.

n: Statistic describing the size of a sample.

π or p: Parameter describing a population proportion.

$\hat{p}$: Statistic describing a sample proportion.

μ: Parameter describing a population mean average.

$\bar{x}$: Statistic describing a sample mean average.

σ: Parameter describing a population standard deviation.

s: Statistic describing a sample standard deviation.

σ^2: Parameter describing a population variance.

s^2: Statistic describing a sample variance.

ρ: Parameter describing a population correlation coefficient.

r: Statistic describing the correlation coefficient from sample data.

β_1: Parameter describing a population slope.

b_1: Statistic describing a slope from sample data.

3.

μ = 69.2 inches (parameter)
$\bar{x}$ = 69.5 inches (statistic)

5.

n = 300 students (statistic)
$\bar{x}$ = 101.9 IQ (statistic)
s = 14.8 IQ (statistic)

7.

μ = 12 units (parameter)
n = 160 students (statistic)
$\bar{x}$ = 12.37 units (statistic)

9.

N = 10,136,559 people (parameter)

11.

β_1=3 pounds per month (Parameter)

n = 54 bears (statistic)

b_1= 2.7055 pounds per month (Statistic)

13.

π = 0.78 (parameter)
n = 165 households (statistic)
$\hat{p}$ = 0.812 (statistic)

15.

μ = 100 IQ (parameter)
σ = 15 IQ (parameter)

$\bar{x}$ = 97.7 IQ (statistic)
s = 15.3 IQ (statistic)

17.

μ = 6.7 pH (parameter)
n = 53 lakes (statistic)
$\bar{x}$ = 6.591 pH (statistic)

19.

N = 59,530 people (parameter)

21.

n = 10 lions (statistic)
$\bar{x}$ = 437.2 pounds (statistic)
μ = 420 pounds (parameter)

23.

b_1= 6.23 mpg (Statistic)

β_1= 5.9 mpg (Parameter)

25.

n = 38 cars (statistic)
$\bar{x}$ = 177.289 displacement (statistic)
s = 88.877 displacement (statistic)

Sampling Distribution Act 1

1.

Answers will vary. If a student got a sample mean of 131.6 cents, then the margin of error would be 131.6 – 134.338 = −2.738 cents. The sample mean was 2.738 cents below the population mean.

3.

If all we know is one random sample, it will be virtually impossible to know the exact population mean. It is extremely difficult to determine a population parameter from a single random sample statistic.

5.

Answers will vary. If the middle 95% of sample means fell between 110 cents and 155 cents, then the approximate standard error would be (155 – 110)/4 ≈ 11.25 cents

---

Sampling Distribution Act 2

1.

We expect that a fair coin should have a population percentage of 50% or a population proportion of 0.5. If we flip the coin 30 times, we expect to get 15 tails.

Answers will vary. If a person got tails 19 times, then their sample proportion is approximately 0.633. That means their margin of error would be approximately 0.633 – 0.5 ≈ 0.133 or 13.3%. So our sample proportion was 0.133 above the population proportion. Or our sample percentage was 13.3% above the population percentage.

3.

If all we know is one random sample, it will be virtually impossible to know the exact population mean. It is extremely difficult to determine a population parameter from a single random sample statistic.

5.

Answers will vary. If the middle 95% of sample proportions fell between 0.33 and 0.74, then the approximate standard error would be (0.74 – 0.33)/4 ≈ 0.1025 or 10.25%

Section 2B

1.

To create a sampling distribution, we take lots of random samples from a population and calculate a statistic like the sample mean average or sample proportion from each of the random samples. We then put all of the statistics on a dot plot so we can see the shape and how much variability the sample statistics have.

3.

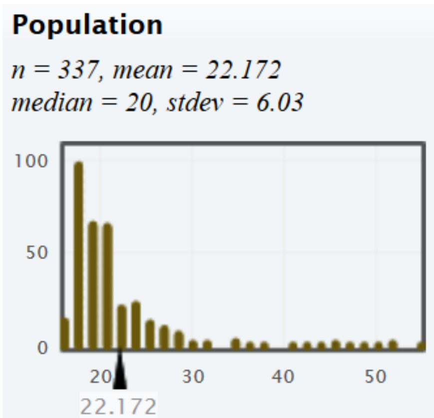
A point estimate is when a sample statistic is used as the approximate value of a population parameter. Point estimates often create confusion in articles because the articles rarely tell you the parameter they are quoting actually came from a sample. In other words, the population parameter quoted in the article is usually just a sample statistic and therefore has a margin of error and is off from the actual population parameter.

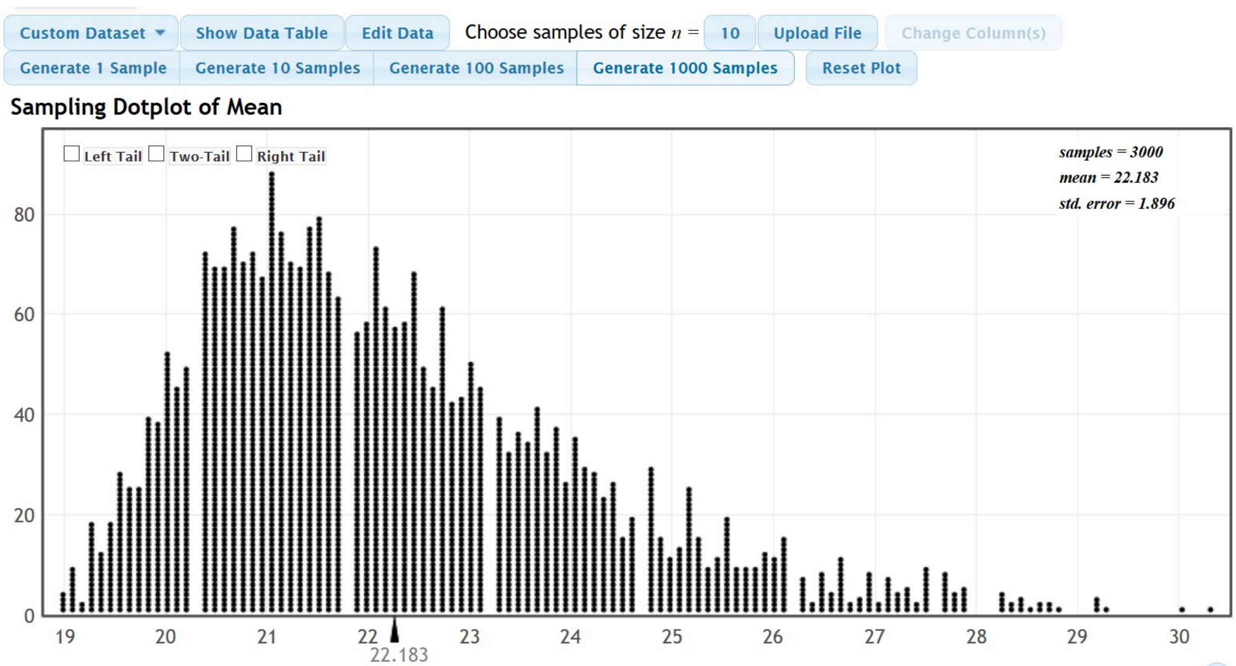
5.

The standard error is the standard deviation of the sampling distribution. If the sampling distribution is normal, then the standard error tells us how far typical sample statistics are from the population parameter. The mean and standard deviation are only accurate when data is normal. The standard error is a type of standard deviation so is only accurate when the sampling distribution is normal.

6. Do not confuse standard error with margin of error. Standard error gives us typical distance while margin of error is larger and accounts for statistics that are not typical.

7.





a) The shape of the population is skewed right and the population mean average age of the math 140 students in fall 2015 (μ) is 22.172 years.

b) The samples means were generally very different than the population mean of 22.172 years. In the example sampling distribution above, the sample means fell anywhere from 18.9 years to 30.3 years. (Answers will vary.)

c) The sample means were different than each other as well. In the example sampling distribution above, the sample means fell anywhere from 18.9 years to 30.3 years. (Answers will vary.)

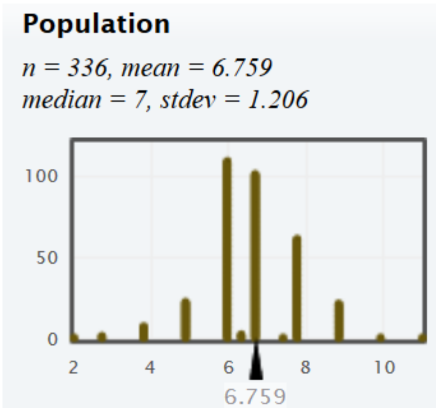
d) Answers will vary. The example sampling distribution above took 3000 samples all of size n = 10.

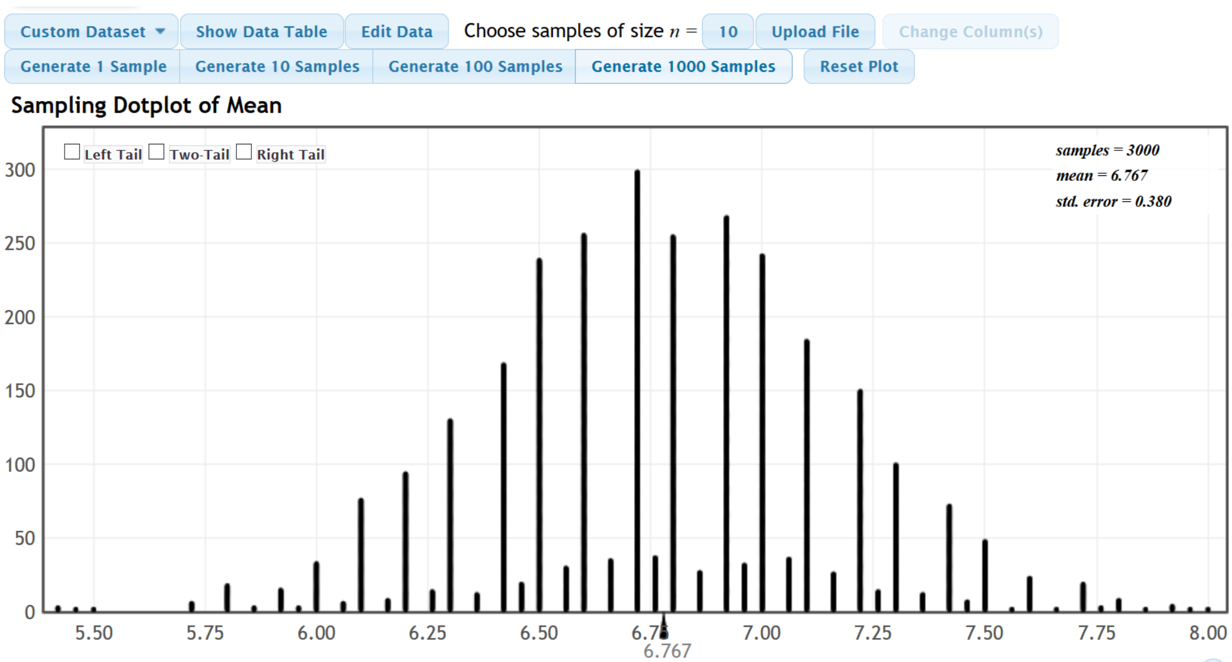
e) The example sampling distribution above was also skewed to the right.

f) Answers will vary. The mean average of all the sample means in the sampling distribution was 22.183 years and is pretty close to the population mean of 22.172 years.

g) Answers will vary. The standard error for the example sampling distribution above was 1.896 years. So typical sample means were within 1.896 years from the population mean.

9.





a) The shape of the population is nearly normal and the population mean average time math 140 students in fall 2015 (μ) sleep is 6.759 hours per night.

b) The samples means were generally very different than the population mean of 6.759 hours. In the example sampling distribution above, the sample means fell anywhere from 5.4 hours to 8 hours. (Answers will vary.)

c) The sample means were different than each other as well. In the example sampling distribution above, the sample means fell anywhere from 5.4 hours to 8 hours. (Answers will vary.)

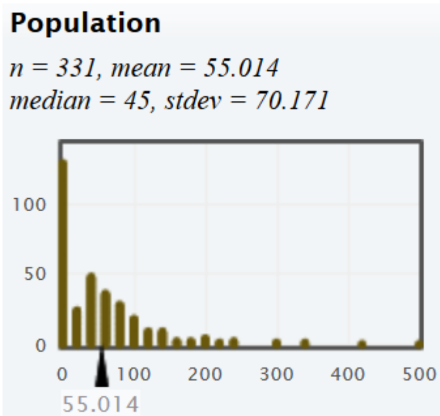
d) Answers will vary. The example sampling distribution above took 3000 samples all of size n = 10.

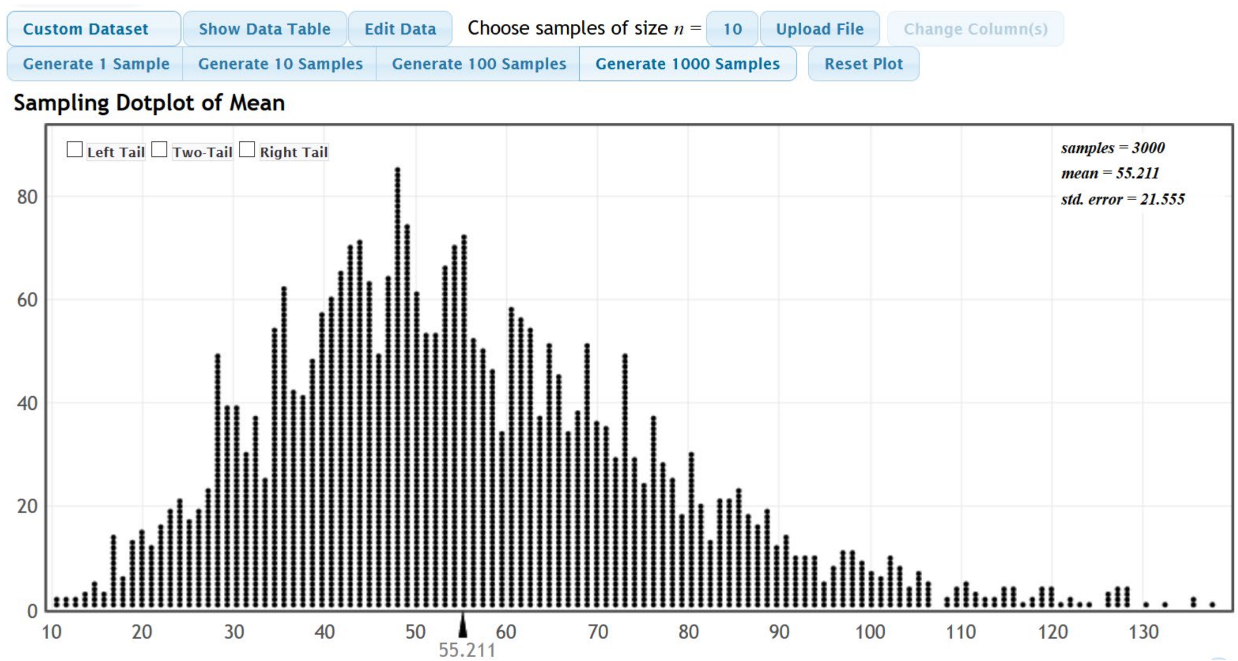
e) The example sampling distribution above was also nearly normal.

f) Answers will vary. The mean average of all the sample means in the sampling distribution was 6.767 hours and is pretty close to the population mean of 6.759 hours.

g) Answers will vary. The standard error for the example sampling distribution above was 0.380 hours. So typical sample means were within 0.380 hours from the population mean.

11.





a) The shape of the population is skewed right. The population mean average monthly cell phone bill for math 140 students in fall 2015 (μ) is $55.211.

b) The samples means were generally very different than the population mean of $55.211. In the example sampling distribution above, the sample means fell anywhere from $9.60 to $137.50. (Answers will vary.)

c) The sample means were different than each other as well. In the example sampling distribution above, the sample means fell anywhere from $9.60 to $137.50. (Answers will vary.)

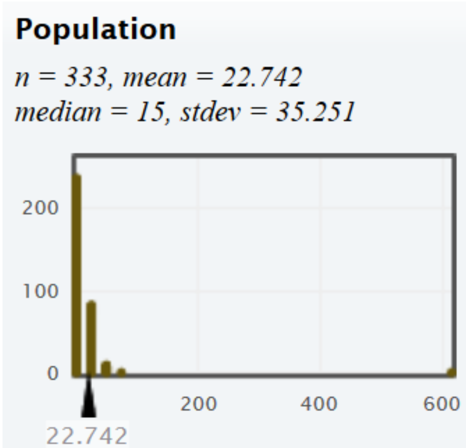
d) Answers will vary. The example sampling distribution above took 3000 samples all of size n = 10.

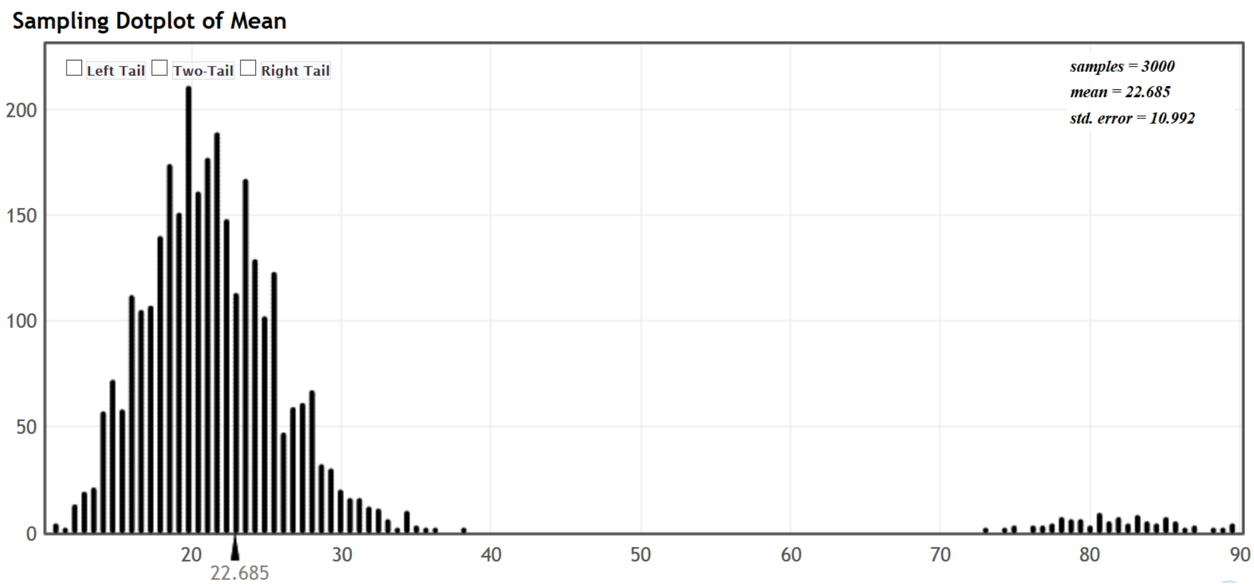
e) The example sampling distribution above was also skewed to the right.

f) Answers will vary. The mean average of all the sample means in the sampling distribution was $55.211 and is the same as the population mean of $55.211.

g) Answers will vary. The standard error for the example sampling distribution above was $21.555. So typical sample means were within $21.555 from the population mean.

13.





a) The shape of the population is skewed right. The population mean average time math 140 students in fall 2015 take to get to school (μ) is 22.742 minutes.

b) The samples means were generally very different than the population mean of 22.742 minutes. In the example sampling distribution above, the sample means fell anywhere from 10.4 minutes to 89.5 minutes. (Answers will vary.)

c) The sample means were different than each other as well. In the example sampling distribution above, the sample means fell anywhere from 10.4 minutes to 89.5 minutes. (Answers will vary.)

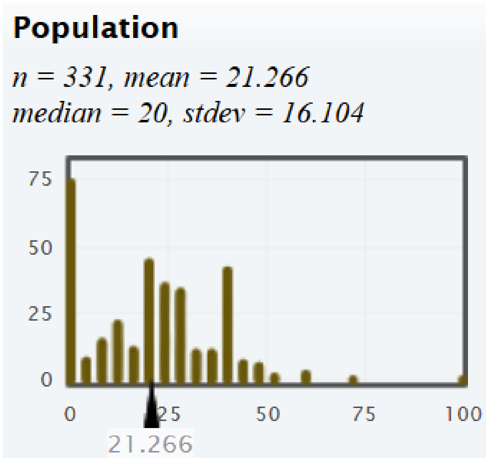
d) Answers will vary. The example sampling distribution above took 3000 samples all of size n = 10.

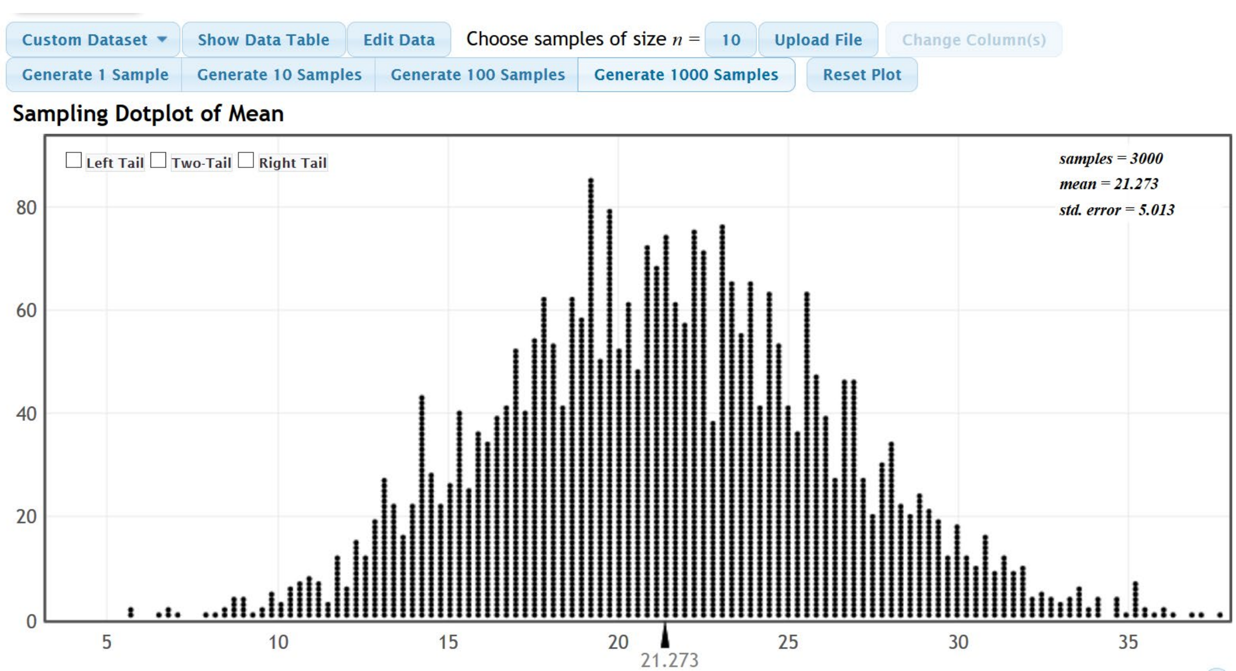
e) The example sampling distribution above was also skewed to the right.

f) Answers will vary. The mean average of all the sample means in the sampling distribution was 22.685 minutes and is very close to the population mean of 22.742 minutes.

g) Answers will vary. The standard error for the example sampling distribution above was 10.992 minutes. So typical sample means were within 10.992 minutes from the population mean.

15.





a) The shape of the population is skewed right. The population mean average time per week that math 140 students in fall 2015 work (μ) is 21.266 hours.

b) The sample means were generally very different than the population mean of 21.266 hours. In the example sampling distribution above, the sample means fell anywhere from 5.6 hours to 37.6 hours. (Answers will vary.)

c) The sample means were different than each other as well. In the example sampling distribution above, the sample means fell anywhere from 5.6 hours to 37.6 hours. (Answers will vary.)

d) Answers will vary. The example sampling distribution above took 3000 samples all of size n = 10.

e) The example sampling distribution above was nearly normal.

f) Answers will vary. The mean average of all the sample means in the sampling distribution was 21.273 hours and is very close to the population mean of 21.266 hours.

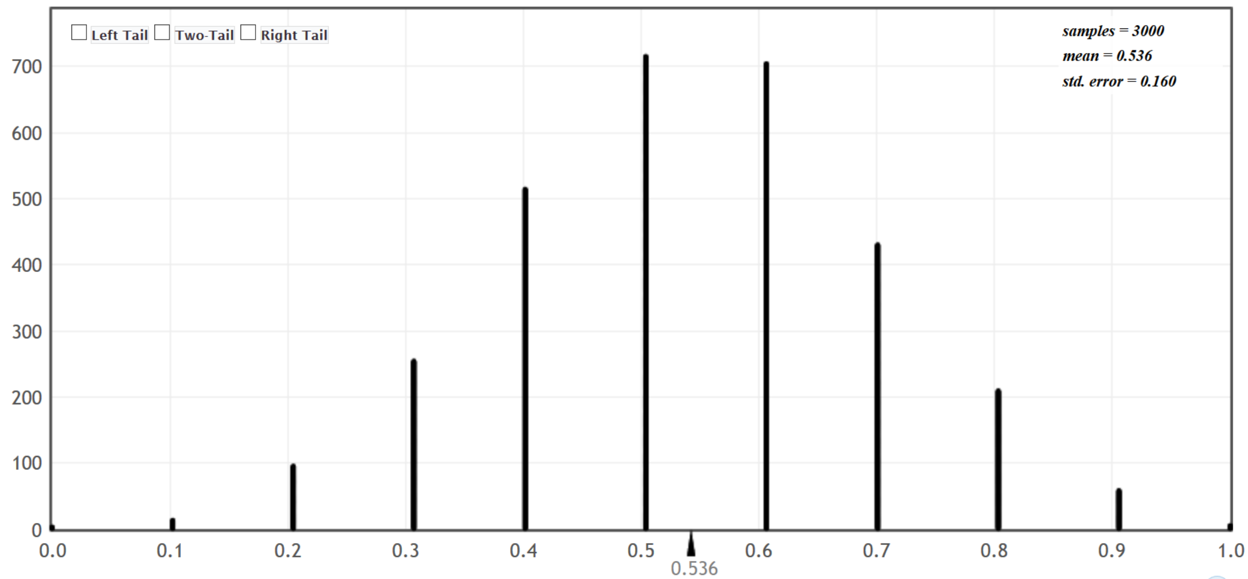
g) Answers will vary. The standard error for the example sampling distribution above was 5.013 hours. So typical sample means were within 5.013 hours from the population mean.

17.

Original Population

Proportion
0.537

Sampling Dotplot of Proportion



a) The sample proportions were generally very different than the population proportion of 0.537 (53.7%). In the example sampling distribution above, the sample proportions fell anywhere from 0.1 (10%) to 1.0 (100%). (Answers will vary.)

b) The sample proportions were different than each other as well. In the example sampling distribution above, the sample proportions fell anywhere from 0.1 (10%) to 1.0 (100%). (Answers will vary.)

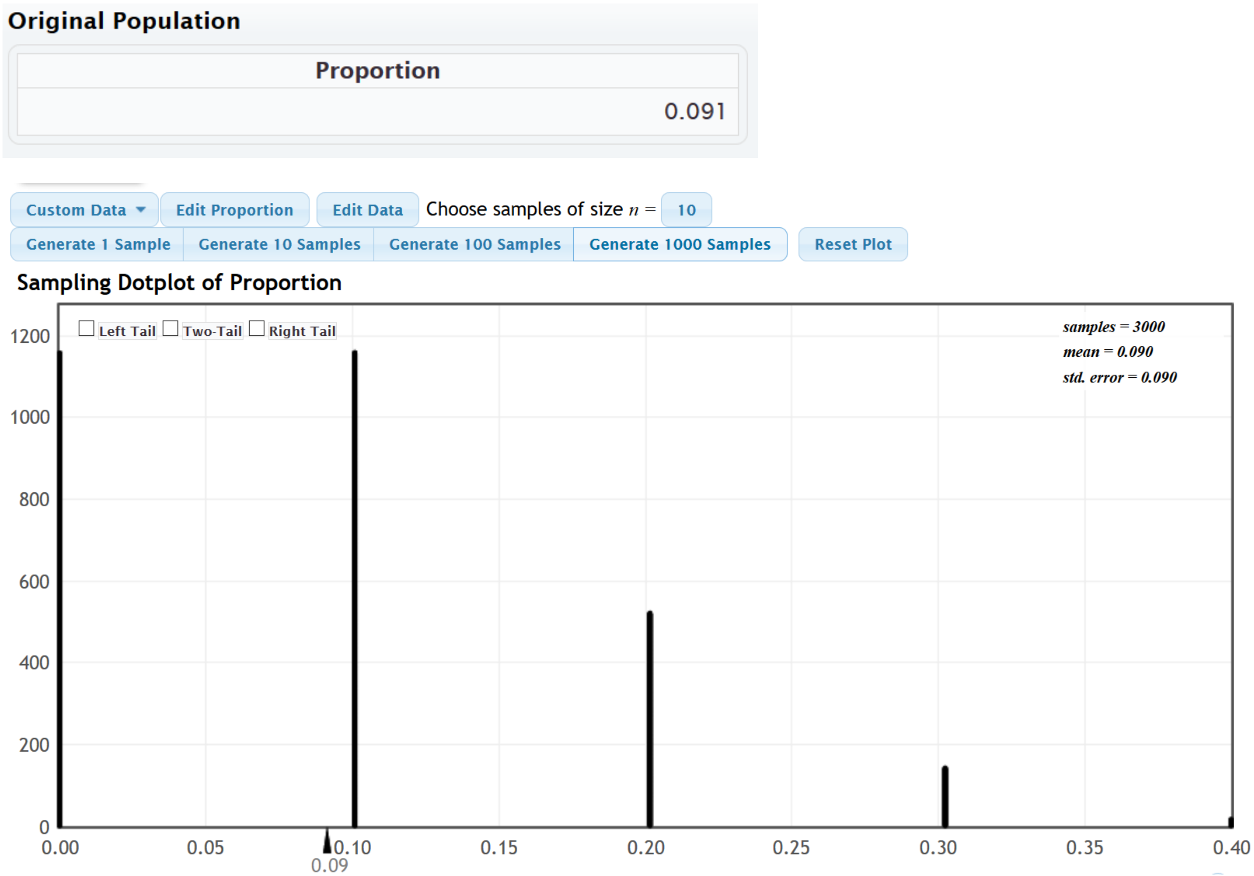
c) Answers will vary. The example sampling distribution above took 3000 samples all of size n = 10.

d) The example sampling distribution above was nearly normal.

e) Answers will vary. The mean average of all the sample proportions in the sampling distribution was 0.536 and is very close to the population proportion of 0.537.

f) Answers will vary. The standard error for the example sampling distribution above was 0.160. So typical sample proportions were within 0.160 (16%) from the population proportion.

19.



a) The sample proportions were generally very different than the population proportion of 0.091 (9.1%). In the example sampling distribution above, the sample proportions fell anywhere from 0 (0%) to 0.4 (40%). (Answers will vary.)

b) The sample proportions were different than each other as well. In the example sampling distribution above, the sample proportions fell anywhere from 0 (0%) to 0.4 (40%). (Answers will vary.)

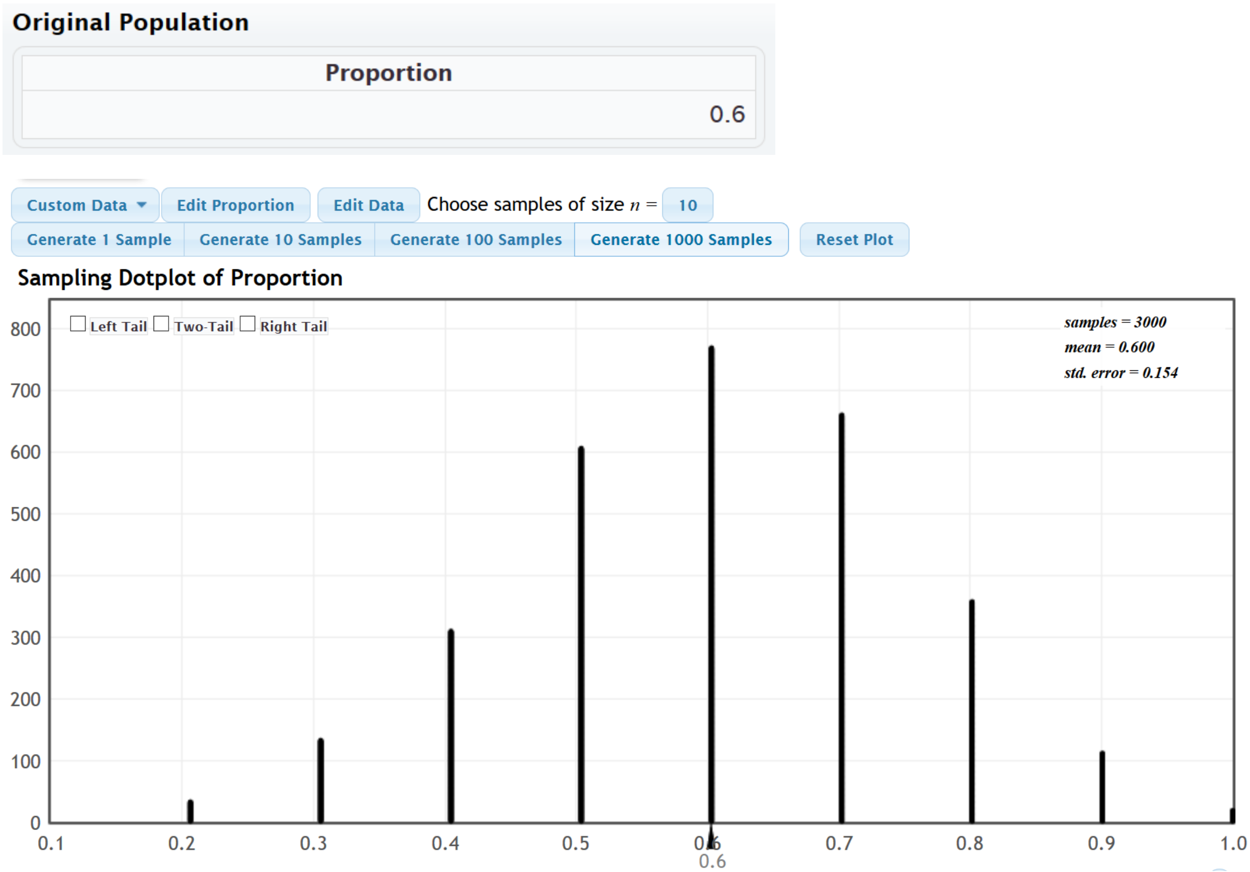
c) Answers will vary. The example sampling distribution above took 3000 samples all of size n = 10.

d) The example sampling distribution above used samples of size n = 10 and was skewed to the right.

e) Answers will vary. The mean average of all the sample proportions in the sampling distribution was 0.09 and is very close to the population proportion of 0.091.

f) Answers will vary. The standard error for the example sampling distribution above was 0.09. So typical sample proportions were within 0.09 (9%) from the population proportion.

21.



a) The sample proportions were generally very different than the population proportion of 0.6 (60%). In the example sampling distribution above, the sample proportions fell anywhere from 0.2 (20%) to 1.0 (100%). (Answers will vary.)

b) The sample proportions were different than each other as well. In the example sampling distribution above, the sample proportions fell anywhere from 0.2 (20%) to 1.0 (100%). (Answers will vary.)

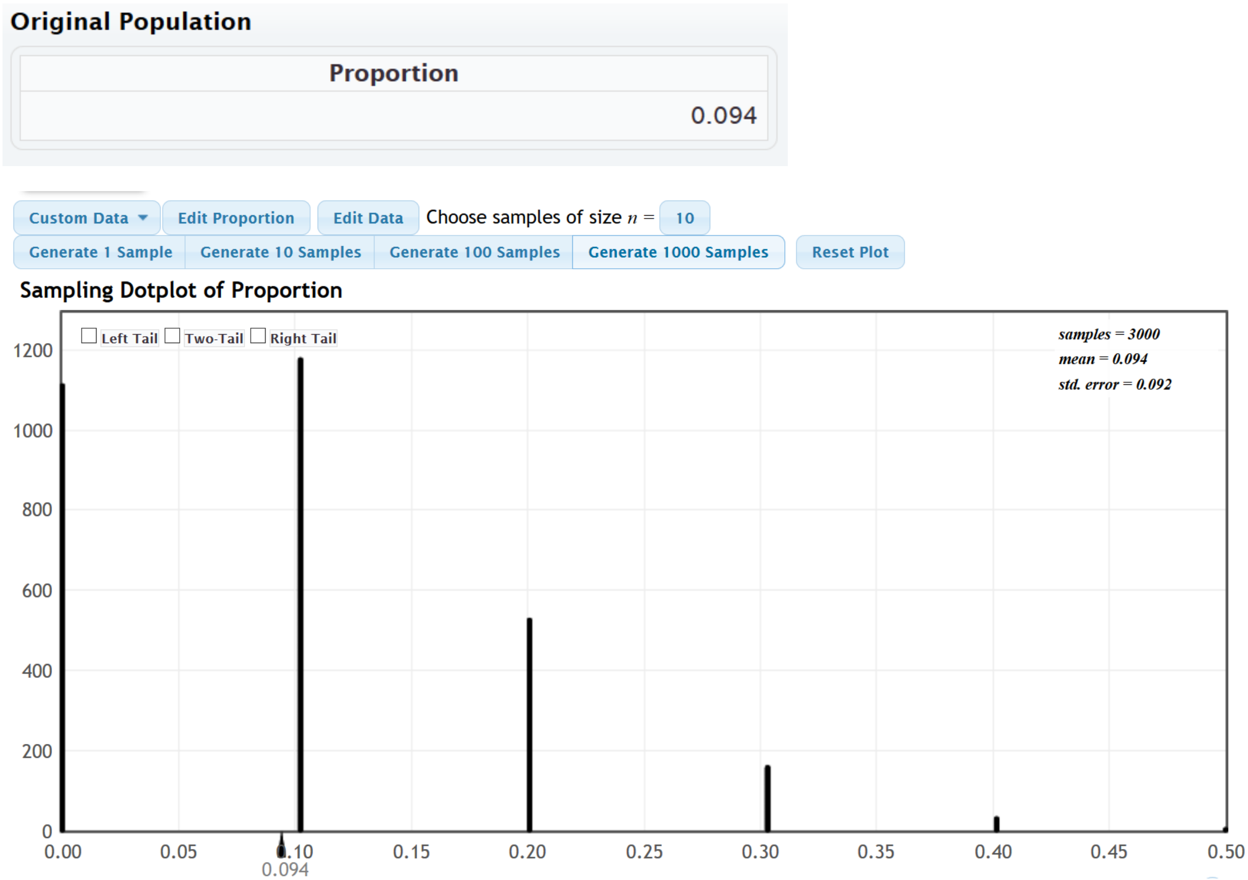
c) Answers will vary. The example sampling distribution above took 3000 samples all of size n = 10.

d) The example sampling distribution above used samples of size n = 10 and was nearly normal.

e) Answers will vary. The mean average of all the sample proportions in the sampling distribution was 0.6 and is the same as the population proportion of 0.6.

f) Answers will vary. The standard error for the example sampling distribution above was 0.154. So typical sample proportions were within 0.154 (15.4%) from the population proportion.

23.



a) The sample proportions were generally very different than the population proportion of 0.094 (9.4%). In the example sampling distribution above, the sample proportions fell anywhere from 0 (0%) to 0.4 (40%). (Answers will vary.)

b) The sample proportions were different than each other as well. In the example sampling distribution above, the sample proportions fell anywhere from 0 (0%) to 0.4 (40%). (Answers will vary.)

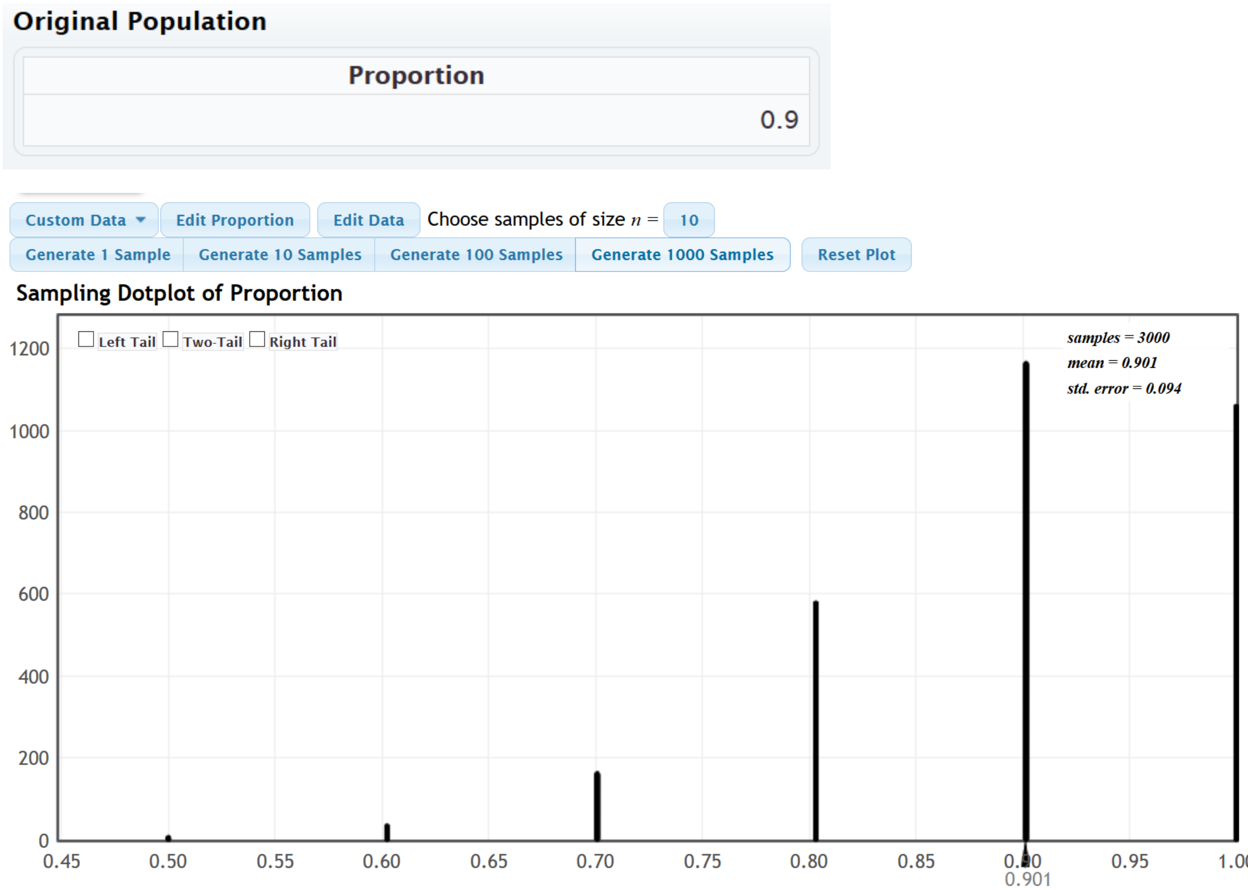
c) Answers will vary. The example sampling distribution above took 3000 samples all of size n = 10.

d) The example sampling distribution above used samples of size n = 10 and was skewed right.

e) Answers will vary. The mean average of all the sample proportions in the example sampling distribution above was 0.094 and is the same as the population proportion of 0.094.

f) Answers will vary. The standard error for the example sampling distribution above was 0.092. So typical sample proportions were within 0.092 (9.2%) from the population proportion.

25.



a) The sample proportions were generally very different than the population proportion of 0.9 (90%). In the example sampling distribution above, the sample proportions fell anywhere from 0.5 (50%) to 1.0 (100%). (Answers will vary.)

b) The sample proportions were different than each other as well. In the example sampling distribution above, the sample proportions fell anywhere from 0.5 (50%) to 1.0 (100%). (Answers will vary.)

c) Answers will vary. The example sampling distribution above took 3000 samples all of size n = 10.

d) The example sampling distribution above used samples of size n = 10 and was skewed left.

e) Answers will vary. The mean average of all the sample proportions in the example sampling distribution above was 0.901 and is very close to the population proportion of 0.90.

f) Answers will vary. The standard error for the example sampling distribution above was 0.094. So typical sample proportions were within 0.094 (9.4%) from the population proportion.

Section 2C

1.

Normality in sampling distributions are important for several reasons. We often use the mean average of all of the sample statistics in the distribution as an approximation of the population parameter. Mean averages are only accurate when they come from data that is normally distributed. We also use the standard deviation of the sampling distribution (standard error) as a measure of how far typical statistics are from the population parameter. Standard Deviation is also only accurate if it comes from data that is normally distributed. There are a multitude of formulas in statistics that use standard error to understand populations. That is why formulas are often tied to requirements that the sampling distribution be normal.

3.

The central limit theorem states that if the sample size is sufficiently large, then a sampling distribution for sample means or sample proportions will be normally distributed. The idea is to explore what conditions will indicate that our sampling distribution will be normal. Remember, normality is important when using the mean and standard deviation of the sampling distribution. For sample means from a non-normal population, we like our sample size to be at least 30. If a population was already normal, then we can have sample sizes below 30 for sample means. For sample proportions, we need to have at least ten successes and at least ten failures in our categorical data for the sampling distribution to be normal.

5.

Less data, results in more variability. So if we decrease the sample size and the population is not normal, then the standard error will get larger and the sampling distribution will look less and less normal.

7.

Yes. If the population was already normal, than the sampling distribution for sample means will look even more normal than the population. This happens because the standard error is smaller than the population standard deviation.

9.

a) The shape of the population is skewed right and the population mean average is 22.172 years.

b) This population is not normal. If we collected a random sample from this population, we must have a sample size of 30 or more.

11.

a) The shape of the population is skewed right and the population mean average is 55.014 dollars per month.

b) This population is not normal. If we collected a random sample from this population, we must have a sample size of 30 or more.

13.

a) The shape of the population is nearly normal and the population mean average is 66.511 inches.

b) This population is normal so any sampling distribution taken from this data will be even more normal. If we collected a random sample from this population, any sample size will give a normal sampling distribution. Though of course, more random data is better.

15.

a) $n = 10/0.091 \approx 109.89$

Always round sample size requirements up. So if we have a sample size of 110 or more, we are likely to have at least 10 successes.

b) $n = 10/(1 – 0.091) \approx 11.001$

Always round sample size requirements up. So if we have a sample size of 12 or more, we are likely to have at least 10 failures.

c) We will take the larger of the success and failure sample size requirements. We should recommend collecting data with a sample size of 110 or higher. In that case, we can expect the sampling distribution for sample proportions to be nearly normal.

17. 9.4% = 0.094

a) $n = 10/0.094 \approx 106.38$

Always round sample size requirements up. So if we have a sample size of 107 or more, we are likely to have at least 10 successes.

b) $n = 10/(1 – 0.094) \approx 11.038$

Always round sample size requirements up. So if we have a sample size of 12 or more, we are likely to have at least 10 failures.

c) We will take the larger of the success and failure sample size requirements. We should recommend collecting data with a sample size of 107 or higher. In that case, we can expect the sampling distribution for sample proportions to be nearly normal.

19. 10% = 0.1

a) $n = 10/0.1 = 100$

So if we have a sample size of 100 or more, we are likely to have at least 10 successes.

b) $n = 10/(1 – 0.1) \approx 11.111$

Always round sample size requirements up. So if we have a sample size of 12 or more, we are likely to have at least 10 failures.

c) We will take the larger of the success and failure sample size requirements. We should recommend collecting data with a sample size of 100 or higher. In that case, we can expect the sampling distribution for sample proportions to be nearly normal.

Section 2D

1.

4.9% ± 1.3%

Interval Notation: (3.6% , 6.2%) or (0.036 , 0.062)

Inequality Notation: $0.036 < \pi < 0.062$

Sentence: We are 95% confident that the population percentage of adults with this disease is between 3.6% and 6.2%.

3.

17.11 mm of Hg ± 3.31 mm of Hg

Interval Notation: (13.80 mm of Hg , 20.42 mm of Hg)

Inequality Notation: 13.80 mm of Hg < σ < 20.42 mm of Hg

Sentence: We are 90% confident that the population standard deviation for women's systolic blood pressure is between 13.80 mm of Hg and 20.42 mm of Hg.

5.

15.98 thousand dollars ± 3.78 thousand dollars

Interval Notation: (12.20 thousand dollars , 19.76 thousand dollars)

Inequality Notation: 12.20 thousand dollars < μ < 19.76 thousand dollars

Sentence: We are 90% confident that the population mean average price of a used mustang car is between 12.20 thousand dollars and 19.76 thousand dollars.

7.

172.55 pounds ± 11.272 pounds

Interval Notation: (161.278 pounds , 183.822 pounds)

Inequality Notation: 161.278 pounds < μ < 183.822 pounds

Sentence: We are 99% confident that the population mean average weight of men is between 161.278 pounds and 183.822 pounds.

9.

36.9% ± 1.44%

Interval Notation: (35.46% , 38.34%) or (0.3546 , 0.3834)

Inequality Notation: 0.3546 < π < 0.3834

Sentence: We are 95% confident that the population percentage of women in the U.S that are overweight is between 35.46% and 38.34%.

11.

Sentence: We are 95% confident that the population proportion of fat in the milk from Jersey cows is between 4.6% and 5.2%.

Sample proportion $\hat{p}$ = (0.052 + 0.046)/2 = 0.049

Margin of Error = (0.052 – 0.046)/2 = 3×10^{-3} = 0.003

13.

Sentence: We are 90% confident that the population proportion of people that will vote for the independent party candidate is between 6.8% and 8.3%.

Sample proportion $\hat{p}$ = (0.083 + 0.068)/2 = 0.0755

Margin of Error = (0.083 – 0.068)/2 = 7.5×10^{-3} = 0.0075

15.

Sentence: We are 99% confident that the population standard deviation for the heights of men is between 2.34 inches and 2.87 inches.

Sample standard deviation s = (2.87 + 2.34)/2 = 2.605 inches

Margin of Error = (2.87 – 2.34)/2 = 0.265 inches

17.

We are 99% confident that the population mean average pH of lakes in Florida is between 6.118 and 7.064.

Sample mean $\overline{x}$ = (7.064 + 6.118)/2 = 6.591

Margin of Error = (7.064 – 6.118)/2 = 0.473

19.

We are 95% confident that the population mean average price of apartments in Manhattan, NY is between $2514.36 and $3798.64.

Sample mean $\overline{x}$ = (3798.64 + 2514.36)/2 = $3156.50

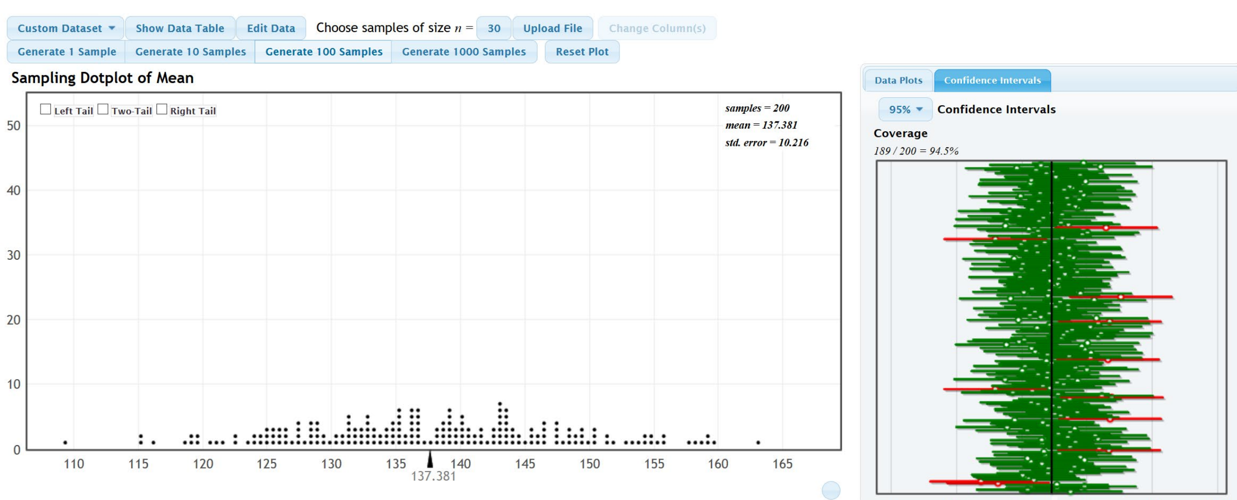
Margin of Error = (3798.64 – 2514.36)/2 = $642.14

21.

By making so many random samples, we see sampling variability at work. Each random sample is different. Each random sample has a different mean and different individual values. This results in a different confidence interval. Some random sample means are so vastly different than the population mean, that the confidence interval created was red and did not even contain the population mean. Other random sample means that were closer to the population mean and the confidence intervals created from these samples were green and did contain the population mean. So sampling variability suggests that not all confidence intervals made from random samples will contain the population mean.

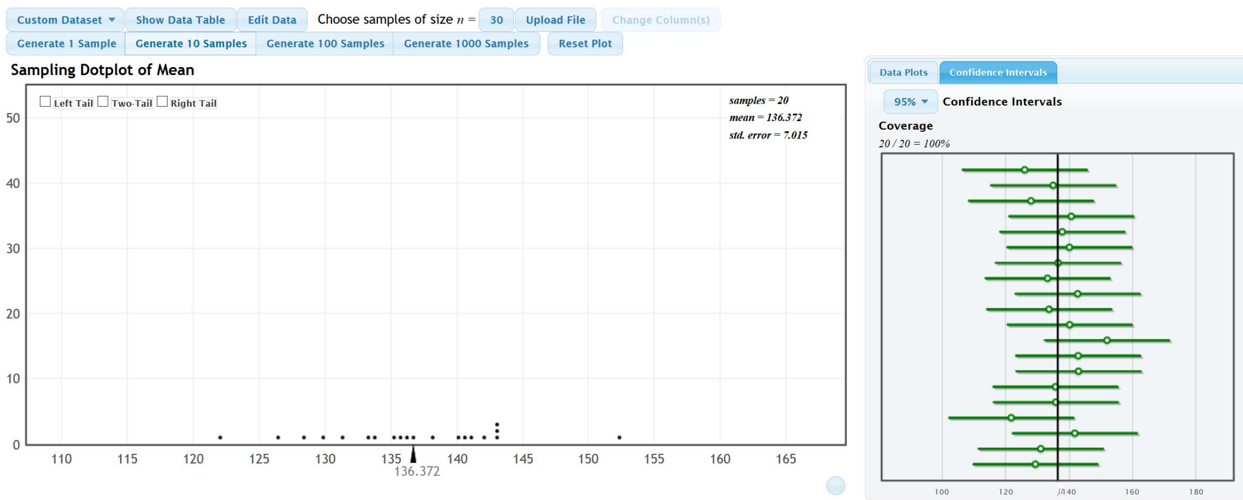
23.

Answers will vary. The example printout below, we took a total of 200 random samples and created 200 different confidence intervals. 189 of the confidence interval contained the population mean. So 94.5% of the confidence intervals contained the population mean. Notice that this is pretty close to the 95% confidence level.



25.

Answers will vary. For smaller number of samples the percentage that contain the population value is farther off of 95%. As the number of samples increased, the percentage got closer to 95%. The example printout below shows that at 20 samples, the percentage was still 100%, but the printout above shows that at 200 samples were at 94.5% which is much closer to the 95% confidence level. This is not surprising, since more random data usually results in less variability. The percentage from 200 random data sets generally will be more accurate than 20 data sets.

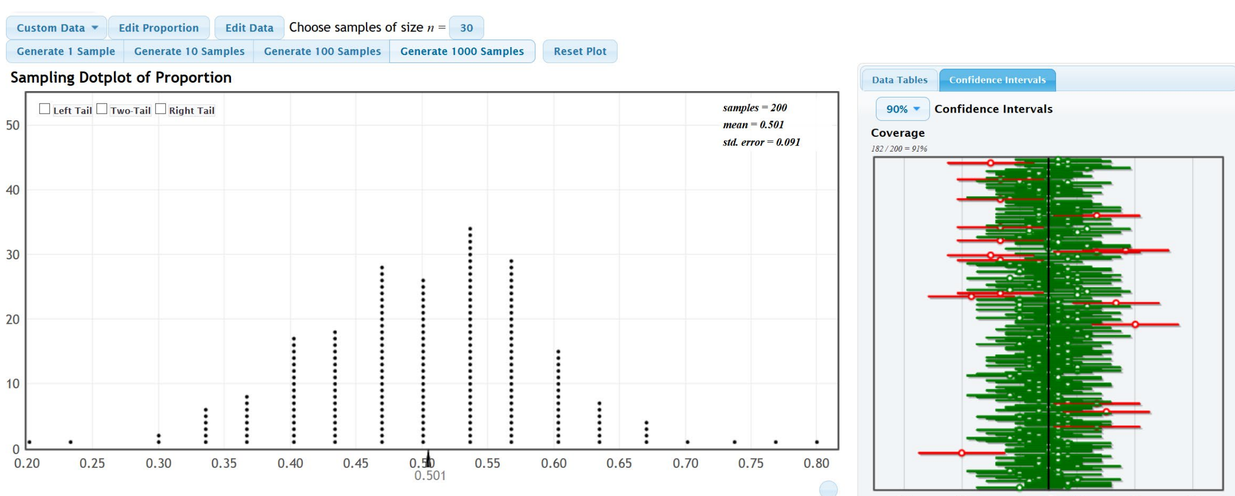


27.

By making so many random samples, we see sampling variability at work. Each random sample is different. Each random sample has a different proportion and different individual values. This results in a different confidence interval. Some random sample proportions are so vastly different than the population proportion, that the confidence interval created was red and did not contain the population proportion. Other random sample proportions that were closer to the population proportion and the confidence intervals created from these samples were green and did contain the population proportion. So sampling variability suggests that not all confidence intervals made from random samples will contain the population proportion.

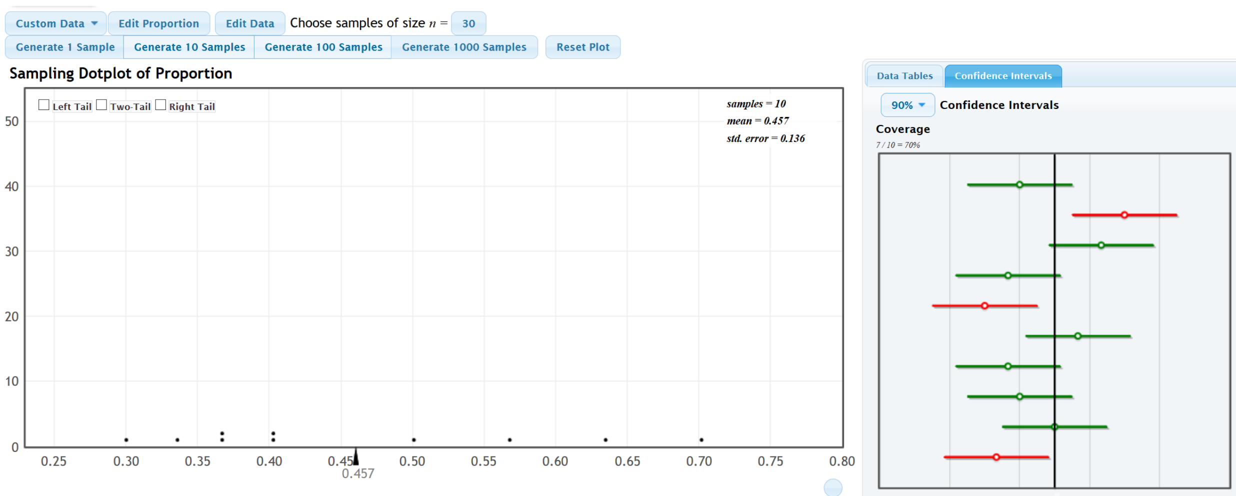
29.

Answers will vary. The example printout below, we took a total of 200 random samples and created 200 different confidence intervals. 182 of the confidence interval contained the population mean. So 91% of the confidence intervals contained the population proportion. Notice that this is pretty close to the 90% confidence level.



31.

Answers will vary. For smaller number of samples the percentage that contain the population proportion is farther off of 90%. As the number of samples increased, the percentage got closer to 90%. The example printout below shows that at 10 samples, the percentage was 70%, but the printout above shows that at 200 samples was 91% which is much closer to the 90% confidence level. This is not surprising, since more random data usually results in less variability. The percentage from 200 random data sets generally will be more accurate than 10 data sets.



Section 2E

1.

<u>One-population Proportion Assumptions</u>

- The categorical sample data should be collected randomly or be representative of the population.
- Data values within the sample should be independent of each other.
- There should be at least ten successes and at least ten failures.

3.

<u>One-population Bootstrap Assumptions</u>

- The sample data should be collected randomly or be representative of the population.
- Data values within the sample should be independent of each other.

5.

a)

The sample data did not pass all of the assumptions.

Random Sample? Yes. Assuming it is a random sample.

Individual rats independent? Yes. A small random sample of rats out of a large population will likely be independent of each other.

At least ten success? Yes. There were 23 rats that showed empathy.

At least ten failures? No. There were only 7 rats that did not show empathy.

b)

We estimate that the sample proportion can be 0.199 off from the population proportion.

c)

We are 99% confident that the population proportion of rats that show empathy is between 0.5678 and 0.9656.

OR

We are 99% confident that the population percentage of rats that show empathy is between 56.78% and 96.56%.

7.

a)

The sample data did pass all of the assumptions.

Random Sample? Yes. Given in #6 that it is a random sample.

Individual tests independent? Yes. A small random sample of lie detector tests out of a large population will likely be independent of each other. We would not want all of the tests done on the same machine though.

At least ten success? Yes. The machine caught the lie 31 times.

At least ten failures? Yes. The machines did not catch the lie 17 times.

b)

We estimate that the sample proportion can be 0.135 off from the population proportion.

c)

We are 95% confident that the population proportion of lies caught by lie detector machines is between 0.5105 and 0.7811.

OR

We are 95% confident that the population percentage of lies caught by lie detector machines is between 51.05% and 78.11%.

9.

a)

This data does not meet the assumptions. The confidence interval may not represent the population very well.

Random Sample? Yes. Given in the problem.

Individuals independent? Maybe not. The population size may be rather small compared to the sample. Individual cereals may be related.

At least 10 successes? No. There were only 4 cereals made by Quaker.

At least 10 failures? Yes. There were 20 cereals in the sample data not made by quaker.

b)

The population proportion could be as much as 0.125 from the sample proportion.

c)

We are 90% confident that the population proportion of cereals made by Quaker is in between 0.0415 and 0.2918.

OR

We are 90% confident that the population percentage of cereals made by Quaker is in between 4.15% and 29.18%.

Since the sample data did not meet the assumptions, the confidence interval may not be accurate.

11.

a)

Random Sample? Yes. Given in the problem.

Individuals independent? Probably not. The population size is rather small and certain companies may tend to put more or less sugar in their cereals.

At least ten successes? Yes. There were 10 cereals with high sugar content.

At least ten failures? Yes. There were 14 cereals that did not have high sugar content.

If the cereals are independent, this would pass the assumptions.

b)

The sample proportion of cereals with high sugar can be 0.197 off from the population proportion.

c)

We are 95% confident that the population proportion of cereals with high sugar content is between 0.2194 and 0.6139.

OR

We are 95% confident that the population percentage of cereals with a high sugar content is between 21.94% and 61.39%.

13.

a)

Random sample? Yes. Given in #12.

Individuals Independent? Probably. The population of all students that take ACT is rather large. Individuals in a small random sample will likely be independent.

At least 30 or normal? Yes. The data was skewed left, but because the sample size is 45 and over 30 it will pass the "30 or normal" requirement.

b)

The sample mean ACT score of 20.8 could be as much as 2.472 points off from the population mean.

c)

We are 90% confident that the population mean average ACT score is between 18.3284 points and 23.2716 points.

15.

a)

Random Sample? Yes. Given in #14.

Individuals Independent? Probably. The population is very large so a small sample of 50 randomly selected people will likely be independent of each other.

At least 30 or normal? Yes. The histogram looks normal (bell shaped). Also, the sample size was 50 which is over 30.

b)

The sample mean average body temperature of 98.26 °F could be as much as 0.217 °F off from the population mean.

c)

We are 95% confident that the population mean average body temperature is between 98.0426 °F and 98.4774 °F.

17.

a)

The data does not pass the assumptions for a mean average confidence interval. The confidence interval may not represent the population mean very well.

Random Sample? Yes. Given in #16.

Individuals Independent? Probably not. The population size is rather small and certain companies may put more or less sugar in their cereals.

At least 30 or normal? No. The sample size was only 24 and the data looks uniform (or bimodal) and is not normal.

b)

The sample mean amount of sugar in these cereals of 7.208 grams could be as much as 2.656 grams off from the population mean amount of sugar in all cereals.

c)

We are 99% confident that the population mean average amount of sugar in cereals is between 4.5527 grams and 9.8639 grams.

This confidence interval may not be accurate since the data used did not pass the assumptions.

19.

a)

If the individual cereals are independent of each other, the data would pass the assumptions for a mean average confidence interval.

Random Sample? Yes. Given in #18.

Individuals Independent? Probably not. The population size is rather small and certain companies may have more or less carbs in their cereals.

At least 30 or normal? Yes. The sample size was only 24. However, the histogram looks normal (bell shaped) so it does pass the “30 or normal” requirement.

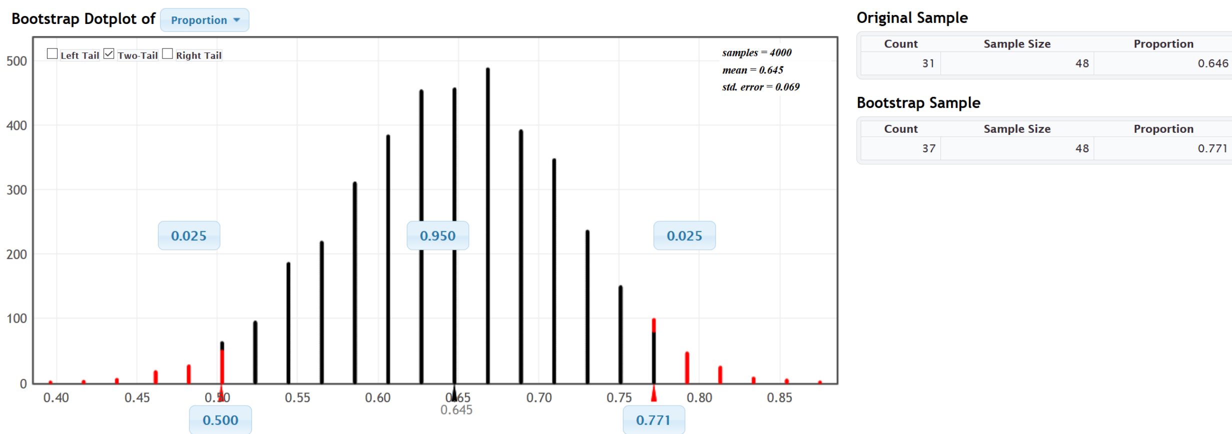
b)

The sample mean amount of carbs in these cereals of 15.043 grams could be as much as 2.061 grams off from the population mean amount of carbs in all cereals.

c)

We are 99% confident that the population mean average amount of carbs in cereals is between 12.9824 grams and 17.1036 grams.

21.



a)

Yes. Meets the assumptions for a bootstrap.

Random Sample? Yes. Given in the problem.

Individuals independent? Probably. A small sample out of a large population would probably be independent as long as the data did not rely on one machine.

b)

Answers will vary. In bootstrap above, there were 4000 bootstrap samples.

c)

Answers will vary. The bootstrap distribution above looks normal (bell shaped).

d)

Answers will vary. The confidence interval in the bootstrap distribution above is (0.500 , 0.771). They are very close to the traditional formula approach in #7.

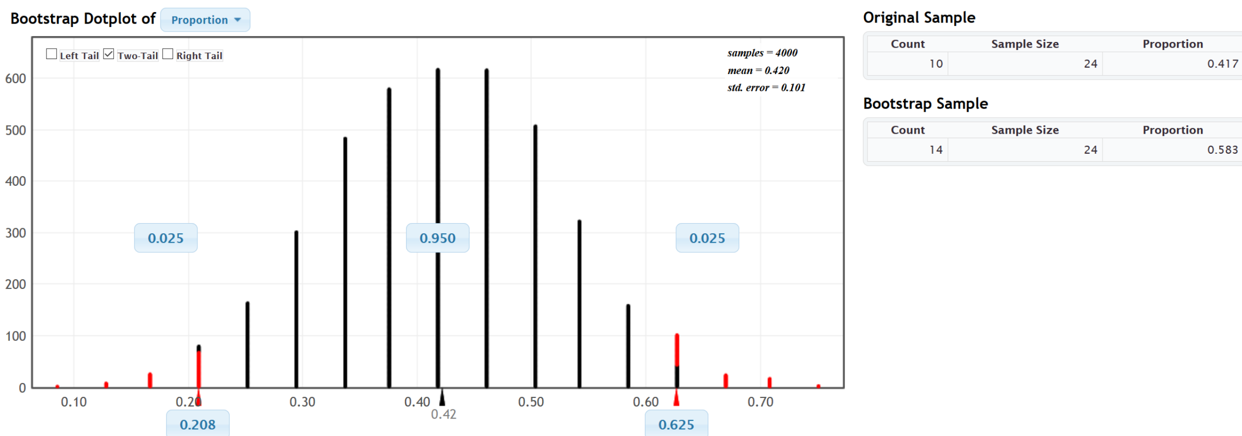
e) Answers will vary. Here are the sentences for the example above.

We are 95% confident that the population proportion of lies that are detected is between 0.500 and 0.771.

OR

We are 95% confident that the population percentage of lies that are detected is between 50.0% and 77.1%.

23.



a)

Might not meet the assumptions for a bootstrap.

Random Sample? Yes. Given in the problem.

Individuals independent? Probably not. The population is rather small and some companies be incline to put similar amounts of sugar in their cereals.

b)

Answers will vary. In bootstrap above, there were 4000 bootstrap samples.

c)

Answers will vary. The bootstrap distribution above looks normal (bell shaped).

d)

Answers will vary. The confidence interval in the bootstrap distribution above is (0.208 , 0.625). They are very close to the traditional formula approach in #11.

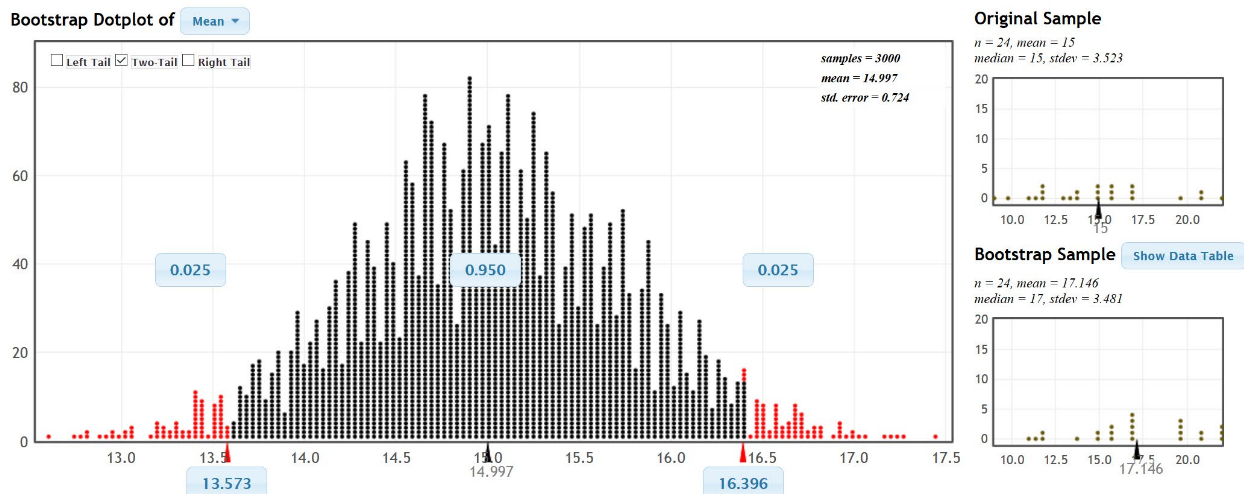
e) Answers will vary. Here are the sentences for the example above.

We are 95% confident that the population proportion of cereals with high sugar content is between 0.208 and 0.625.

OR

We are 95% confident that the population percentage of cereals with high sugar content is between 20.8% and 62.5%.

25.



a)

Might not meet the assumptions for a bootstrap.

Random Sample? Yes. Given in the problem.

Individuals independent? Probably not. The population is rather small and some companies be incline to have similar amounts of carbs in their cereals.

b)

Answers will vary. In bootstrap above, there were 3000 bootstrap samples.

c)

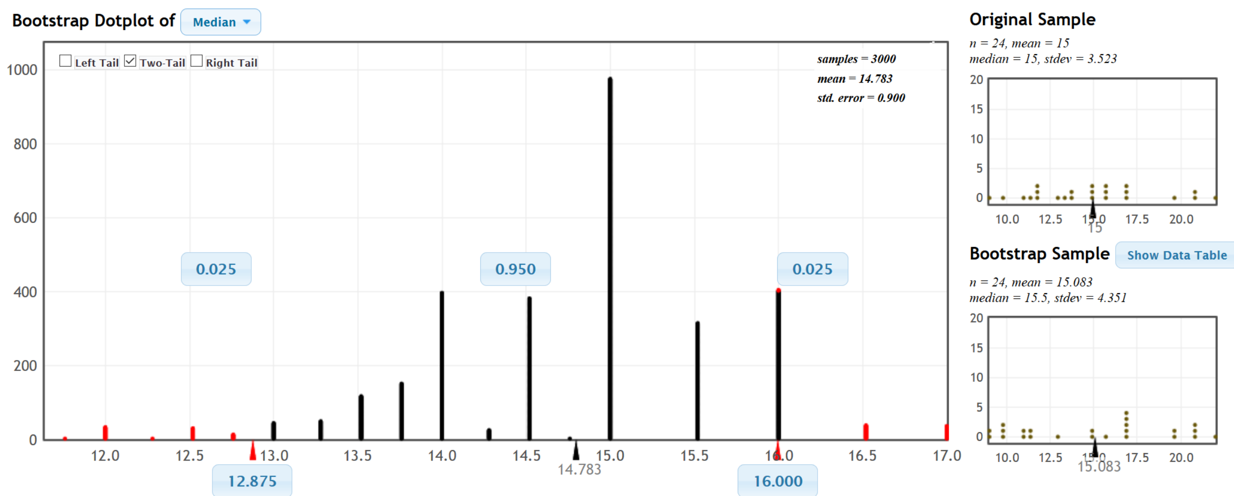
Answers will vary. The bootstrap distribution above looks normal (bell shaped).

d)

Answers will vary. The confidence interval in the bootstrap distribution above is (13.573 carbs , 16.396 carbs). They are very close to the traditional formula approach in #19.

e) Answers will vary. Here are the sentences for the example above.

We are 95% confident that the population mean amount of carbs in cereals is between 13.573 carbs and 16.396 carbs.



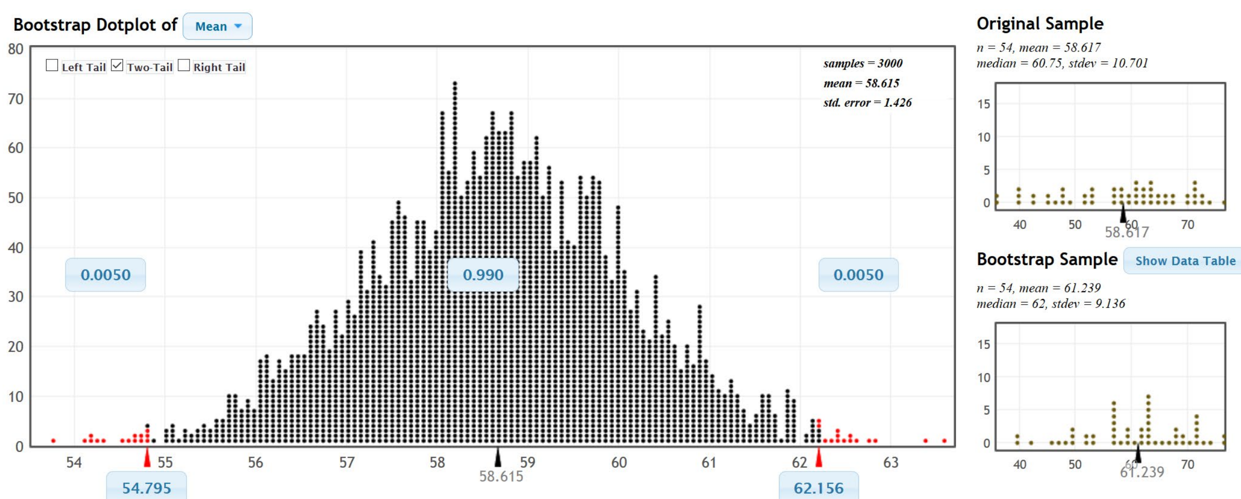
f) Answers will vary. The median bootstrap example above looks skewed left.

g) Answers will vary. The median bootstrap example above has a confidence interval of (12.875 carbs, 16.000 carbs.

h) Answers will vary. Here is the sentence for the median bootstrap example above.

We are 95% confident that the population median average amount of carbs in cereals is between 12.875 carbs and 16.000 carbs.

27.



a)

Yes. It does meet the assumptions for a bootstrap.

Random Sample? Yes. Given in the problem.

Individuals independent? Probably. The population of all bears is rather large. A random sample of 54 bears from all over will probably not be related. It would fail independence if this sample was taken from the same area.

b)

Answers will vary. In bootstrap above, there were 3000 bootstrap samples.

c)

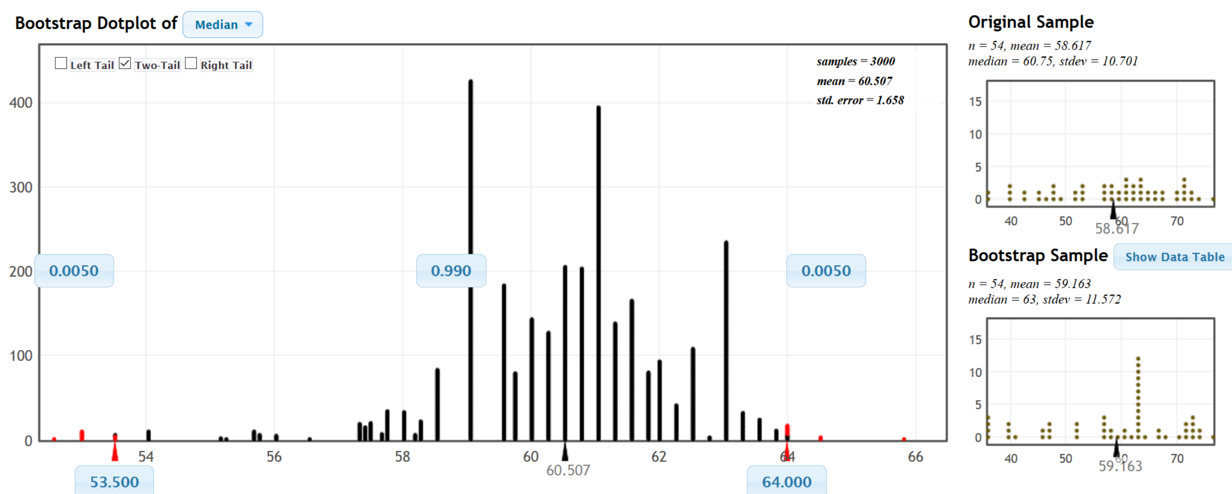
Answers will vary. The bootstrap distribution above looks normal (bell shaped).

d)

Answers will vary. The confidence interval in the bootstrap distribution above is (54.795 inches, 62.156 inches).

e) Answers will vary. Here are the sentences for the example above.

We are 99% confident that the population mean average length of this species of bear is between 54.795 inches and 62.156 inches.



f) Answers will vary. The median bootstrap example looks tri-modal.

g) Answers will vary. The median bootstrap example above has a confidence interval of (53.5 inches, 64.0 inches.

h) Answers will vary. Here is the sentence for the median bootstrap example above.

We are 95% confident that the population median average length of this species of bear is between 53.5 inches and 64.0 inches.

Section 2F

1.

<u>Two-population Proportion Assumptions</u>

- The two categorical samples should be collected randomly or be representative of the population.
- Data values within each sample should be independent of each other.
- Data values between the samples should be independent of each other.
- There should be at least ten successes and at least ten failures.

2.

<u>Two-population Mean Assumptions</u> (Matched Pair)

- The quantitative ordered pair sample data should be collected randomly or be representative of the population.
- Data values within the sample should be independent of each other.
- There should be at least thirty ordered pairs or the differences should have a nearly normal shape.

Two-population Mean Assumptions (Not Matched Pair)

- The two quantitative samples should be collected randomly or be representative of the population.
- Data values within each sample should be independent of each other.
- Data values between the two samples should be independent of each other.
- The sample sizes should be at least 30 or have a nearly normal shape.

3.

Two-population Proportion Bootstrap Assumptions

- The two categorical samples should be collected randomly or be representative of the population.
- Data values within each sample should be independent of each other.
- Data values between the samples should be independent of each other.

Two-population Bootstrap Mean Assumptions (Matched Pair)

- The quantitative ordered pair sample data should be collected randomly or be representative of the population.
- Data values within the sample should be independent of each other.

Two-population Mean Bootstrap Assumptions (Not Matched Pair)

- The two quantitative samples should be collected randomly or be representative of the population.
- Data values within each sample should be independent of each other.
- Data values between the two samples should be independent of each other.

5.

a)

Population 1 is significantly lower than population 2 since the confidence interval is (negative, negative). This shows there is a negative difference between the populations.

b)

Population 1 could be between 0.068 (6.8%) and 0.115 (11.5%) lower than population 2.

c)

We are 95% confident that the population proportion for population 1 is between 0.068 and 0.115 lower than population 2.

OR

We are 95% confident that the population percentage for population 1 could be between 6.8% and 11.5% lower than population 2.

7.

a)

There is no significant difference between population 1 and population 2 since the confidence interval is (negative, positive). This means we do not know if the difference is positive or negative. Either population might be larger.

b)

We are 90% confident that there is no significant difference between the population proportions for population 1 and population 2.

OR

We are 90% confident that there is no significant difference between the population percentages for population 1 and population 2.

9.

a)

Population 1 is significantly higher than population 2 since the confidence interval is (positive, positive). This shows there is a positive difference between the populations.

b)

The population proportion for population 1 could be between 0.049 (4.9%) and 0.058 (5.8%) higher than population 2.

c)

We are 99% confident that the population proportion for population 1 is between 0.049 and 0.058 higher than population 2.

OR

We are 99% confident that the population percentage for population 1 is between 4.9% and 5.8% higher than population 2.

11.

a)

There is no significant difference between population 1 and population 2 since the confidence interval is (negative, positive). This means we do not know if the difference is positive or negative. Either population might be larger.

b)

We are 95% confident that there is no significant difference between the population proportions for population 1 and population 2.

OR

We are 95% confident that there is no significant difference between the population percentages for population 1 and population 2.

13.

a)

This data was matched pair since it was the same people measured twice.

<u>Two-population Mean Assumptions</u> (Matched Pair)

- The quantitative ordered pair sample data should be collected randomly or be representative of the population. Yes. The data was collected randomly.
- Data values within the sample should be independent of each other. Yes. Since the population size is very large, a small sample of individuals from that population will probably not be related.
- There should be at least thirty ordered pairs or the differences should have a nearly normal shape. Yes. The sample size was only 28, but the differences were normal. This does pass the “30 or normal” requirement.

b)

The sample mean difference was 5.8 ACT points. Since the confidence interval was (positive, positive), this sample mean difference is significant.

c)

Since the confidence interval was (positive, positive), the ACT scores after the prep class (population 1) is significantly higher than the ACT scores before the prep class. We think that the population mean average ACT scores after the prep class (population 1) could be between 4.4159 points and 7.1841 points higher than the ACT scores before the prep class (population 2).

d)

We are 90% confident that the population mean average ACT scores after the prep class (population 1) is between 4.4159 points and 7.1841 points higher than the ACT scores before the prep class (population 2).

15.

a)

This data was not matched pair since it was two independent groups of people.

Two-population Mean Assumptions (Not Matched Pair)

- The two quantitative samples should be collected randomly or be representative of the population. Yes. The two samples were collected randomly.
- Data values within each sample should be independent of each other. Yes. Each sample was random from a large population so individuals are not likely to be related.
- Data values between the two samples should be independent of each other. Yes. Since the population is large, the people that live with smokers are not likely to be related to the people that do not live with smokers.
- The sample sizes should be at least 30 or have a nearly normal shape. Yes. Though the shapes are unknown, both sample sizes were over 30. So both samples pass the "30 or normal" requirement.

b)

The difference between the sample means was 5.6 mg/mL. Since the confidence interval was (negative, negative), this difference is significant.

c)

Since the confidence interval was (negative, negative), the population mean average cotinine for those that do not live with smokers (population 1) is significantly lower than for those that live with smokers. We think that the population mean average cotinine level for those that do not live with smokers (population 1) could be between 18.5696 ng/mL and 24.0304 ng/mL lower than for those that live with smokers (population 2).

d)

We are 95% confident that the population mean average cotinine level for those that do not live with smokers (population 1) is between 18.5696 ng/mL and 24.0304 ng/mL lower than for those that live with smokers (population 2).

17.

a)

This data was matched pair since it was the same people measured twice.

Two-population Mean Bootstrap Assumptions (Matched Pair)

- The two quantitative samples should be collected randomly or be representative of the population. Yes. The two samples were collected randomly.
- Data values within each sample should be independent of each other. Yes. Individual data was collected randomly from a large population so individuals are not likely to be related.

b)

The sample mean difference was -43.375 mm of Hg. Since the confidence interval was (negative, negative), this difference is significant.

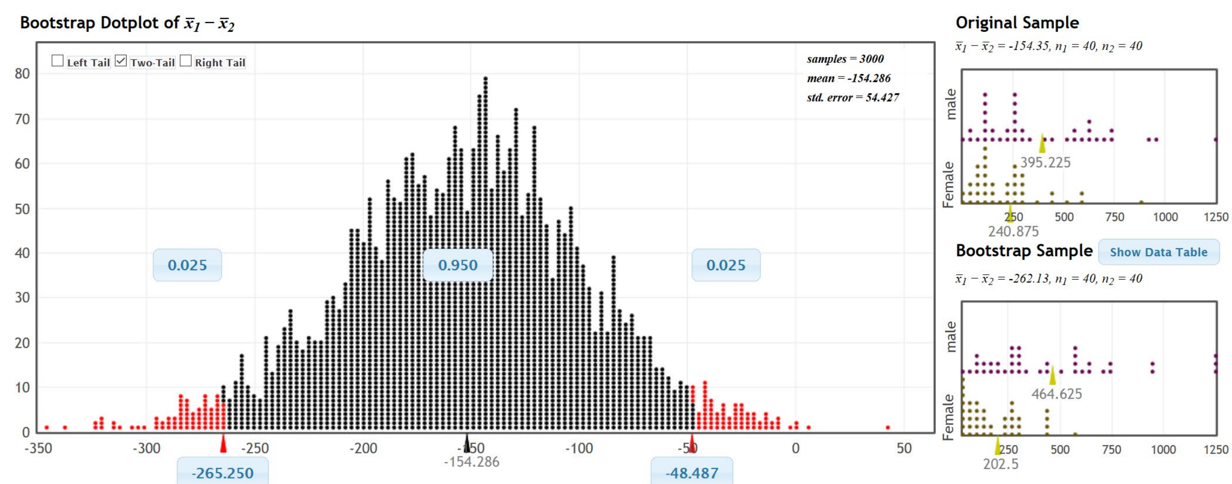
c)

Since the confidence interval was (negative, negative), the population mean average diastolic blood pressure (population 1) is significantly lower than the systolic blood pressure. We think that the population mean average diastolic blood pressure (population 1) is between 40.162 mm of Hg and 47.0 mm of Hg lower than the systolic blood pressure (population 2).

d)

We are 95% confident that the population mean average diastolic blood pressure (population 1) is between 40.162 mm of Hg and 47.0 mm of Hg lower than the systolic blood pressure (population 2).

19.



a)

This data was not matched pair since it was two independent groups of people.

<u>Two-population Mean Bootstrap Assumptions</u> (Not Matched Pair)

- The two quantitative samples should be collected randomly or be representative of the population. Yes. The two samples were collected randomly.
- Data values within each sample should be independent of each other. Yes. Each sample was random from a large population so individuals are not likely to be related.
- Data values between the two samples should be independent of each other. Yes. Since the population is large, the females and males are not likely to be related.

b)

Answers will vary. For the bootstrap example above, the difference between the sample means was 154.35 mg/dL. Since the confidence interval was (negative, negative), this difference is significant.

c)

Answers will vary. The confidence interval in the example above was (negative, negative), the population mean average cholesterol for women (population 1) is significantly lower than for men (population 2). We think that the population mean average cholesterol for women (population 1) could be between 48.487 mg/dL and 265.25 mg/dL lower than for men (population 2).

d)

We are 95% confident that the population mean average cholesterol for women (population 1) is between 48.487 mg/dL and 265.25 mg/dL lower than for men (population 2).

Section 2G

1.

<u>One-Population Variance or Standard Deviation Assumptions</u>

- The quantitative sample data should be collected randomly or be representative of the population.
- Data values within the sample should be independent of each other.
- The sample data must be normal.

3.

a)

The sample data does not pass the assumptions for a variance or standard deviation confidence interval.

<u>One-Population Variance or Standard Deviation Assumptions</u>

- The quantitative sample data should be collected randomly or be representative of the population. Yes. The data was collected randomly. This was given in #2.
- Data values within the sample should be independent of each other. Yes. Since this is a small random sample from a large population, individuals are unlikely to be related.
- The sample data must be normal. No. The data was skewed left.

b)

We are 90% confident that the population variance for ACT scores is between 70.8423 and 143.8391. These numbers may not be very accurate since the data did not meet the assumptions.

c)

We are 90% confident that the population standard deviation for ACT scores is between 8.4168 ACT points and 11.9933 ACT points. These numbers may not be very accurate since the data did not meet the assumptions.

5.

a)

The sample data does pass the assumptions for both variance and standard deviation confidence intervals.

<u>One-Population Variance or Standard Deviation Assumptions</u>

- The quantitative sample data should be collected randomly or be representative of the population. Yes. The data was collected randomly. This was given in #4.
- Data values within the sample should be independent of each other. Yes. Since this is a small random sample from a large population, individuals are unlikely to be related.
- The sample data must be normal. Yes. The sample data was normal (bell shaped).

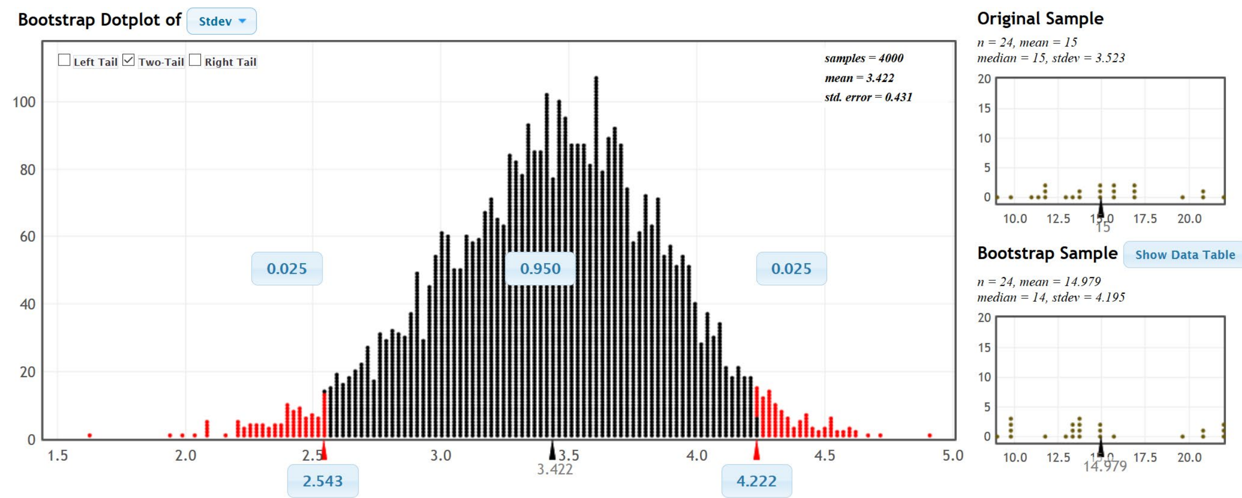
b)

We are 99% confident that the population variance for human body temperature is between 0.3666 and 1.0524.

c)

We are 99% confident that the population standard deviation for human body temperature is between 0.6054 °F and 1.0258 °F.

7.



a)

The sample data may not pass the assumptions for standard deviation bootstrap confidence intervals.

<u>One-Population Variance or Standard Deviation Assumptions</u>

- The quantitative sample data should be collected randomly or be representative of the population. Yes. The data was collected randomly.
- Data values within the sample should be independent of each other. Probably not. The population size is rather small and individual cereals may be related to each other since they are made by the same companies.

b)

Answers will vary. The example bootstrap above had 4000 bootstrap samples.

c)

Answers may vary. The example bootstrap distribution above looks normal (bell shaped).

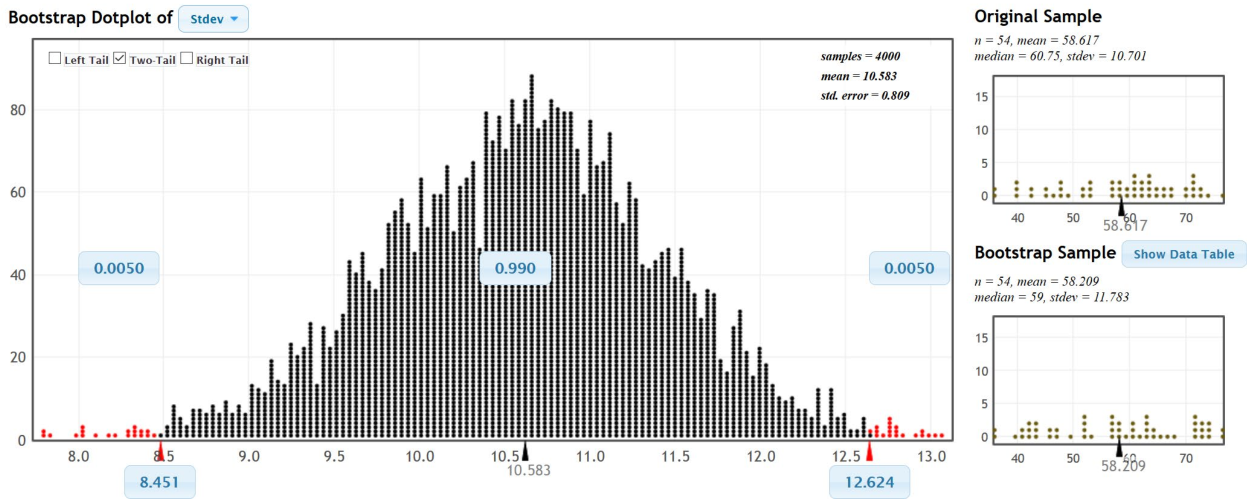
d)

Answers will vary. The example bootstrap confidence interval above was (2.543 carbs, 4.222 carbs).

e)

We are 95% confident that the population standard deviation for the amount of carbs in cereals is between 2.543 carbs and 4.222 carbs. These numbers may not be accurate since the sample data did not pass the assumptions.

9.



a)

The sample data does pass the assumptions for standard deviation bootstrap confidence intervals.

One-Population Variance or Standard Deviation Assumptions

- The quantitative sample data should be collected randomly or be representative of the population. Yes. The data was collected randomly.
- Data values within the sample should be independent of each other. Probably. Since it was a small random sample from a large population from different areas the bears are likely to be independent of each other.

b)

Answers will vary. The example bootstrap above had 4000 bootstrap samples.

c)

Answers may vary. The example bootstrap distribution above looks normal (bell shaped).

d)

Answers will vary. The example bootstrap confidence interval above was (8.451 inches, 12.624 inches).

e)

Answers will vary. Here is the sentence for the example bootstrap above.

We are 99% confident that the population standard deviation for the lengths of this species of bear is between 8.451 inches and 12.624 inches.

Chapter 2 Review

1.

N: Parameter describing the size of a population.

n: Statistic describing the size of a sample.

π or p: Parameter describing a population proportion.

p̂: Statistic describing a sample proportion.

μ: Parameter describing a population mean average.

x̅: Statistic describing a sample mean average.

σ: Parameter describing a population standard deviation.

s: Statistic describing a sample standard deviation.

σ^2: Parameter describing a population variance.

s^2: Statistic describing a sample variance.

ρ: Parameter describing a population correlation coefficient.

r: Statistic describing the correlation coefficient from sample data.

β_1: Parameter describing a population slope.

b_1: Statistic describing a slope from sample data.

2.

a)

$\bar{x}$ = 101.9 (statistic)
s = 14.8 (statistic)
s^2 = 219.04 (statistic)

μ = 100 (parameter)
σ = 15 (parameter)
σ^2 = 225 (parameter)

b)

ρ = 0 (parameter)
β_1 =

β_1=20 (parameter)

r = 0.338 (statistic)
$b_1 = 13.79\ (statistic)$

c)

$\hat{p}$ = 0.03 (statistic)
π = 0.015 (parameter)

d)

n = 238 (statistic)
N = 5,000,000 (parameter)

3.

<u>One-population Mean Assumptions</u>

- The quantitative sample data should be collected randomly or be representative of the population.
- Data values within the sample should be independent of each other.
- The sample size should be at least 30 or have a nearly normal shape.

4.

<u>One-Population Variance or Standard Deviation Assumptions</u>

- The quantitative sample data should be collected randomly or be representative of the population.
- Data values within the sample should be independent of each other.
- The sample data must be normal.

5.

<u>One-population Proportion Assumptions</u>

- The categorical sample data should be collected randomly or be representative of the population.
- Data values within the sample should be independent of each other.
- There should be at least ten successes and at least ten failures.

6.

<u>Two-population Mean Assumptions</u> (Matched Pair)

- The quantitative ordered pair sample data should be collected randomly or be representative of the population.
- Data values within the sample should be independent of each other.
- There should be at least thirty ordered pairs or the differences should have a nearly normal shape.

<u>Two-population Mean Assumptions</u> (Not Matched Pair)

- The two quantitative samples should be collected randomly or be representative of the population.
- Data values within each sample should be independent of each other.
- Data values between the two samples should be independent of each other.
- The sample sizes should be at least 30 or have a nearly normal shape.

7.

<u>Two-population Proportion Assumptions</u>

- The two categorical samples should be collected randomly or be representative of the population.
- Data values within each sample should be independent of each other.
- Data values between the samples should be independent of each other.
- There should be at least ten successes and at least ten failures.

8.

<u>Bootstrap and Randomized Simulation Assumptions</u>

- The sample data should be collected randomly or be representative of the population.
- Data values within each sample should be independent of each other.
- If multiple samples were collected that were not matched pair, then the data values between the samples should be independent of each other.

9.

Vocabulary

Population: The collection of all people or objects to be studied.

Census: Collecting data from everyone in a population.

Sample: Collecting data from a small subgroup of the population.

Statistic: A number calculated from sample data in order to understand the characteristics of the data. For example, a sample mean average, a sample standard deviation, or a sample percentage.

Parameter: A number that describes the characteristics of a population like a population mean or a population percentage. Can be calculated from an unbiased census, but is often just a guess about the population.

Sampling Distribution: Take many random samples from a population, calculate a sample statistic like a mean or percent from each sample and graph all of the sample statistics on the same graph. The center of the sampling distribution is a good estimate of the population parameter.

Sampling Variability: Random samples values and sample statistics are usually different from each other and usually different from the population parameter.

Point Estimate: When someone takes a sample statistic and then claims that it is the population parameter.

Margin of Error: Total distance that a sample statistic might be from the population parameter. For normal sampling distributions and a 95% confidence interval, the margin of error is approximately twice as large as the standard error.

Standard Error: The standard deviation of a sampling distribution. The distance that typical sample statistics are from the center of the sampling distribution. Since the center of the sampling distributions is usually close to the population parameter, the standard error tells us how far typical sample statistics are from the population parameter.

Confidence Interval: Two numbers that we think a population parameter is in between. Can be calculated by either a bootstrap distribution or by adding and subtracting the sample statistic and the margin of error.

95% Confident: 95% of confidence intervals contain the population value and 5% of confidence intervals do not contain the population value.

90% Confident: 90% of confidence intervals contain the population value and 10% of confidence intervals do not contain the population value.

99% Confident: 99% of confidence intervals contain the population value and 1% of confidence intervals do not contain the population value.

Bootstrapping: Taking many random samples values from one original real random sample with replacement.

Bootstrap Sample: A simulated sample created by taking many random samples values from one original real random sample with replacement.

Bootstrap Statistic: A statistic calculated from a bootstrap sample.

Bootstrap Distribution: Putting many bootstrap statistics on the same graph in order to simulate the sampling variability in a population, calculate standard error, and create a confidence interval. The center of the bootstrap distribution is the original real sample statistic.

10.

a)

We are 99% confident that the population mean weight is between 55.6 pounds and 69.4 pounds.

b)

We are 90% confident that the population proportion is in between 0.352 and 0.411.

OR

We are 90% confident that the population percentage is in between 35.2% and 41.1%.

c)

We are 95% confident that the population standard deviation is between 3.1 pounds and 4.7 pounds.

d)

We are 99% confident that the population variance is between 461.8 square inches and 591.3 square inches.

e)

Since the confidence interval is (positive, positive), this indicates that population 1 is significantly larger than population 2. The confidence interval indicates that population 1 is between 13.2 kg and 14.8 kg larger.

Sentence: We are 95% confident that the population mean average weight of population 1 is between 13.2 kg and 14.8 kg larger than population 2.

f)

Since the confidence interval is (negative, positive), this indicates that there is no significant difference between the population mean averages for population 1 and population 2. They are very close and we cannot tell which population is larger.

Sentence: We are 90% confident that there is no significant difference between the population mean averages for population 1 and population 2.

g)

Since the confidence interval is (negative, positive), this indicates that there is no significant difference between the population proportions for population 1 and population 2. The population percentages are very close and we cannot tell which population is larger.

Sentence: We are 95% confident that there is no significant difference between the population proportions for population 1 and population 2.

OR

We are 95% confident that there is no significant difference between the population percentages for population 1 and population 2.

h)

Since the confidence interval is (negative, negative), this indicates that population 1 is significantly smaller than population 2. The confidence interval indicates that population 1 is between 0.057 (5.7%) and 0.072 (7.2%) smaller than population 2.

Sentence: We are 99% confident that the population proportion for population 1 is between 0.057 and 0.072 less than population 2.

OR

We are 99% confident that the population percentage for population 1 is between 5.7% and 7.2% less than population 2.

11.

A sampling distribution is created by taking hundreds or thousands of random samples from a population. We then calculate a sample statistic like the mean, standard deviation or proportion and put all of the thousands of random sample statistics on the same graph. The center of the sampling distribution is often very close to the actual population parameter. The standard error is the standard deviation of the sampling distribution and helps us understand how far typical statistics may be from the center (population parameter). Overall the sampling distribution teaches us about sampling variability. This is the principle that random samples are usually very different from each other and are often very far off from the population parameter.

12.

If the sampling distribution is normal, then we can multiply the critical value Z-score or T-score times the standard error to get the margin of error. If we take the original random sample statistic and add and subtract the margin of error, we will get the upper and lower limits of the confidence interval.

13.

$\frac{s^2(n-1)}{\chi^2_{upper}}$ < Population Variance (σ^2) < $\frac{s^2(n-1)}{\chi^2_{lower}}$

To create a confidence interval for the population variance, multiply the sample variance s^2 by the degrees of freedom n-1 and then divide by the chi-squared critical values.

14.

a)

William Gosset

b)

Guiness Beer

c)

Needed better accuracy for smaller data sets.

d)

Guiness Beer had a strict no publishing policy and would have fired him.

e)

“student”

f)

When the sample size is small, the T-scores are significantly greater than the Z-scores. This accounts for more variability in smaller data sets.

g)

When the sample size is large, the T-scores and Z-scores are about the same.

h)

One and two-population proportion (percentage) confidence intervals use Z-scores.

i)

One and two-population mean average confidence intervals use T-scores.

j)

For one quantitative data set with a sample size “n”, the degrees of freedom is “n – 1”.

15.

The central limit theorem states that a sampling distribution made of sample means will be normal (bell shaped) if the sample size is sufficiently large. The mean and standard deviation of a sampling distribution are important calculations in inferential statistics but are only accurate if the sampling distribution is normal. The central limit theorem discusses what conditions need to be met for the sampling distributions to be normal. These are the foundations of our assumptions for the confidence intervals. For quantitative sample data that is not normal, we will need a sample size of at least 30 to ensure the sampling distribution will look nearly normal. For quantitative sample data that is already normal, we can use a sample below 30. For categorical data, we will need at least 10 successes and failures for the sampling distribution for proportions to look normal.

16.

Suppose we want to make a confidence interval, but the data does not pass the assumptions for our sampling distribution to look normal. This would mean our traditional formulas involving standard error, Z-scores and T-scores may not be very accurate. Bootstrapping is a technique for calculating a confidence interval directly without the traditional formula. In a bootstrap, we take thousands of random samples with replacement from the one random sample. We then calculate the statistic from each of the bootstrap samples and put all of the thousand bootstrap statistics on the bootstrap distribution. In a sense, we have created a simulated sampling distribution but not from the population. Find the middle 95%, 90% or 99% markers from the distribution and you have your confidence interval. Bootstrapping still requires the data to be collected randomly and individual observations should be independent, but it does not require the same central limit theorem assumptions. The bootstrap distribution does not have to look normal since we are calculating the middle 95%, 90% or 99% directly.
