ANSWER KEY: POPULATION MEAN CONFIDENCE INTERVALS FOR LARGE SAMPLES

This answer key provides solutions to the corresponding student activity sheet.

Population Mean Confidence Intervals for Large Samples

The data for these exercises are in the Minitab file ***CIForMean__LargeSample_Activity.mtw***.

Exercise 1

(**a**) Suppose that n = 100 of your favorite brand of candy bar are randomly sampled and their weights (in grams) are measured. A 95% confidence interval for the mean weight is 42 ≤ μ ≤ 48. Would a 99% confidence interval calculated from the same sample data be wider or tighter?

Tighter or **Wider**

Solution: The trade-off of gaining more precision for a confidence level is a wider confidence interval.

(b) Consider the following statement: Before the confidence interval is constructed, there is a 95% chance that μ is between 42 and 48. **Note:** If you answer true, then you are saying that μ is a random entity.

True or **False**

Solution: Parameters, such as the population mean μ, are not random entities. Although we may not know its value, the mean μ is fixed for a given population.

(c) Consider the following statement: Before the confidence interval is constructed, there is a 95% chance that the interval 42 and 48 contains μ. **Note:** If it is true, then you are saying that the confidence interval is a random entity.

True or False

Solution: The random entity in constructing a confidence interval for the population mean μ is the confidence interval. To construct a confidence interval, we select a random sample from the population. The interval endpoints are determined from this random sample.

(d) Consider the following statement: Suppose n = 100 candy bars are randomly sampled and a 95% confidence interval for μ is computed. If this process is repeated 1000 times, then approximately 950 of the confidence intervals will contain the true value of μ.

Is this statement correct?

True or False

Solution: This is exactly the point when constructing confidence intervals. There may not be exactly 950 intervals that contain μ, but the number will be close to 950.

Exercise 2

The *Central Limit Theorem*, the key to building a confidence interval for the population mean μ with a Z test, states that:

A. We can always use a normal curve to approximate the distribution of the sample mean $\bar{X}$.

B. If n is large (e.g. n > 30) and the original population is normal, then the distribution of the sample mean $\bar{X}$ can be approximated by a normal curve.

C. We can always use a normal curve to approximate the distribution of sample values X.

D. If n is large (e.g. n > 30) then the distribution of the sample mean $\bar{X}$ can be approximated by a normal curve even if the original distribution is not normal.

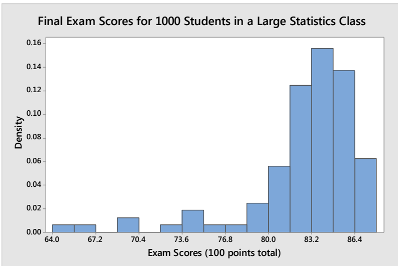
E. If n is large (e.g. n > 30), then the distribution of the sample values X can be approximated by a normal curve even if the original distribution is not normal.

Solution: The CLT is stated in answer D.

Exercise 3

True or False

You choose 36 scores from the population of all students' final exam scores (below) and calculate the mean $\bar{x}$ of their test scores. You repeat this process 100 times and plot the distribution $\bar{X}$ of the 100 sample means, the $\bar{x}$'s. The distribution $\bar{X}$ of the sample means will be approximately normal.



True or False

Solution: Although the n = 36 exam scores are taken from a population that is not normally distributed, the graph of their sample means ($\bar{x}$'s) will be normally distributed since the sample size is "large" (n > 30), according to the CLT.

Exercise 4

(a) Each tennis court contains a confidence interval for the true population mean resonance frequency (Hz) for all tennis rackets of a certain type using the same sample mean, sample standard deviation, and sample size, but not the same confidence level. What is the value of the sample mean resonance frequency $\bar{x}$ used to construct both of these confidence intervals?

Solution: The sample mean $\bar{x}$ for these confidence intervals is the center of the confidence interval. Therefore:

$$\bar{x} = \frac{114.4 + 115.6}{2} = \mathbf{115\ Hz}$$

(b) Both intervals were calculated from the same sample data, but the confidence level for one of these intervals is 90% and the other is 99%. Which interval, [114.4, 115.6] or [114.1, 115.9], has a confidence level of 90%? Why?

Solution: The 90% confidence interval is the narrower one, [114.4, 115.6]. In order to obtain additional confidence, we have to extend the width of the confidence interval. A 90% confidence interval has a tighter width since 90% of the area under a standard normal curve is "tighter" than 99% of the area under a standard normal curve.

Exercise 5

(a) Using a random sample of n = 36 previous concerts, the sample mean duration is $\bar{x}$ = 110 minutes, while the population standard deviation is σ = 20 minutes. Construct the 95% confidence interval by hand.

Solution: The sample size n is large (n = 36), so $\bar{X}$ is approximately normally distributed by the Central Limit Theorem (or exactly normal if the distribution of concert lengths is normal). Using a $z_{0.025}$ critical value, we obtain the following 95% confidence interval by hand:

$$\left[110 - 1.96 \cdot \frac{20}{\sqrt{36}}, 110 + 1.96 \cdot \frac{20}{\sqrt{36}}\right] = [\mathbf{103.467, 116.533}]\ \textbf{minutes}$$

(b) Following the Minitab instructions on the activity sheet, the interval is:

One-Sample Z

```
The assumed standard deviation = 20

 N    Mean  SE Mean       95% CI
36  110.00     3.33  (103.47, 116.53)
```

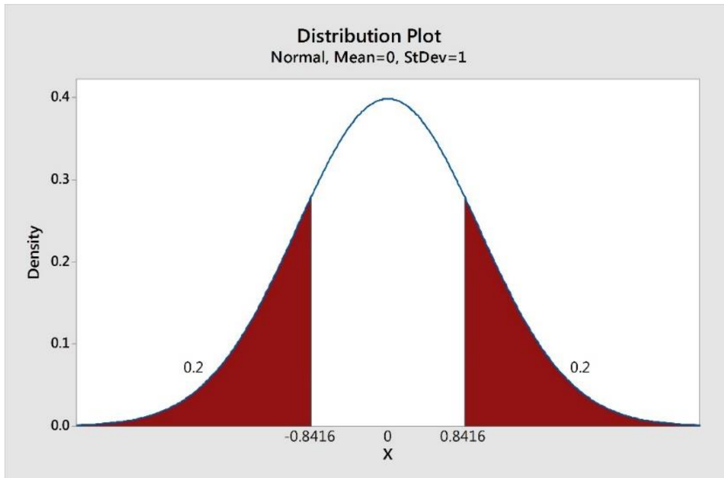
Exercise 6

A random sample of n = 50 drill bits is used to put holes into a steel doorframe. The lifetime of a drill bit is measured as the number of holes drilled before the bit fails. The average lifetime of a drill bit is 12.68 holes with a standard deviation of 6.83 holes. Which formula should be used to determine a 60% confidence interval for the mean lifetime of the drill bits?

A. 12.68 ± 0.25 * (6.83/√50)
B. 12.68 ± 0.25 * (6.83/50)
C. 12.68 ± 0.25 * (6.83^2/50)
D. 12.68 ± 0.84 * (6.83/√50)
E. 12.68 ± 0.84 * (6.83/50)
F. 12.68 ± 0.84 * (6.83^2/50)
G. 12.68 ± 1.28 * (6.83/√50)
H. 12.68 ± 1.28 * (6.83^2/√50)
I. 12.68 ± 1.28 * (6.83*√50)

Solution: By the CLT, the distribution of the sample means $\bar{X}$ is approximately normal (or exactly normal if the drill bit lifetime is normal). A 60% confidence interval requires a critical value of $z_{0.20}$, which is approximately 0.8416. This value can be determined using the standard normal table or Minitab (as shown in the ***Normal Distribution*** lesson). The confidence interval is constructed using a sample size of n = 50, and so the correct choice is:

$$\textbf{D. } \bar{\boldsymbol{x}} \pm \boldsymbol{z}_{0.20} \cdot \frac{\boldsymbol{s}}{\sqrt{\boldsymbol{n}}} = \mathbf{12.68} \pm \mathbf{0.8416} \cdot \frac{\mathbf{6.83}}{\sqrt{\mathbf{50}}}$$



Exercise 7

(a) A random sample of 64 cucumber-filled bags has sample mean weight $\bar{x}$ = 1.03 pounds with a sample standard deviation s = 0.08 lbs. Determine a 90% confidence interval for the population mean weight μ of these bags of cucumbers by hand. Then check your results using Minitab.

Solution: By the CLT, the distribution of the sample means $\bar{X}$ is approximately normal (or exactly normal if the bag weights are normal). A 90% confidence interval requires a critical value of $z_{0.05}$, which is approximately 1.645. This value can be determined using the standard normal table or Minitab (as shown in the ***Normal Distribution*** lesson). The confidence interval is constructed with a sample size of n = 64, and so the correct confidence interval is:

$$\bar{x} \pm z_{0.05} \cdot \frac{s}{\sqrt{n}} = 1.03 \pm 1.645 \cdot \frac{0.08}{\sqrt{64}} \rightarrow [\mathbf{1.014, 1.046}]$$

Using Minitab, we obtain:

One-Sample Z

```
The assumed standard deviation = 0.08


 N    Mean  SE Mean        90% CI
64  1.0300   0.0100  (1.0136, 1.0464)
```

(b) Your neighbor claims that the true mean weight of these bags is 1 pound. With 90% confidence, is the neighbor's claim true? Why or why not?

Solution: Since the 90% interval does not contain 1 pound, then we can say with 90% confidence that her claim is not true.

Exercise 8

For a confidence interval, as the confidence level increases (e.g. 90% → 95%), the reliability goes ____________ and the precision goes _______________.

A. up ... up B. down ... up C. down ... down **D. up ... down**

Solution: Although an increased confidence level results in a wider confidence interval with greater reliability, it also results in less precision.

Exercise 9

(a) A large box contains 10,000 ball bearings. A random sample of 120 bearings is chosen. The sample mean diameter is $\bar{x}$ = 10 mm, and the standard deviation is s = 0.24 mm. A 95% confidence interval for the true mean diameter of the 10,000 bearings is 10 ± 1.96 * 0.24 / $\sqrt{10000}$.

True or **False**

Solution: The correct sample size for the confidence interval is n = 120, not 10,000 which is the size of the population.

(b) For large n (i.e. n > 30), a two-sided 95% confidence interval for the population mean μ will contain the sample mean $\bar{x}$ approximately 95% of the time.

True or **False**

Solution: The confidence interval will ALWAYS contain sample mean $\bar{x}$. Based on how we calculate confidence intervals, the sample mean $\bar{x}$ is in the center of the interval.

(c) Suppose that a random sample of n = 50 bottles of a certain brand of cough syrup is selected, and a 95% confidence interval for the true mean alcohol content in these bottles is calculated as [7.8, 9.4] mg. A 99% confidence interval calculated from the same sample data would be wider.

True or False

Solution: A 99% confidence interval would be wider since it requires greater confidence. The appropriate z-score for a two-sided 95% confidence interval is $z_{0.025}$ = 1.96, while the z-score for a two-sided 99% confidence interval is $z_{0.005}$ = 2.58.

Exercise 10

For this exercise, students either collected data on their own mothers' ages or they are using the data provided in column C1 of ***CIForMean_LargeSample_Activity.mtw***. Below are solutions using the C1 data.

(a) In Minitab, determine the sample standard deviation s for Moms' ages. If you don't remember how to do this, see the ***Describing Data Numerically*** lesson.

Solution: Using Minitab, s=6.14:

Descriptive Statistics: Moms Ages (yrs)

```
                    Total
Variable            Count   Mean  SE Mean  StDev
Moms Ages (yrs)        33  51.79     1.07   6.14
```

(b) Using your class as a random sample of students who are "similar" to the large student population (e.g. student ages, similar majors, etc.), determine a two-sided 90% confidence interval for the true mean age μ of Moms. Construct this interval by hand or in Minitab. Recall that you'll need to use the sample standard deviation s in place of the unknown population standard deviation σ.

Solution: Calculating the interval by hand and using s = 6.14 in the place of σ, we get the following interval:

$$\left[51.79 - 1.645 * \frac{6.14}{\sqrt{33}}, 51.79 + 1.645 * \frac{6.14}{\sqrt{33}}\right] \cong \mathbf{[50.032, 53.548]\ years}$$

Minitab reports the interval as:

One-Sample Z: Moms Ages (yrs)

```
The assumed standard deviation = 6.14

Variable          N   Mean  StDev  SE Mean       90% CI
Moms Ages (yrs)  33  51.79   6.14     1.07  (50.03, 53.55)
```

(c) Does the above confidence interval actually contain the true mean age μ for Moms' ages for students in this category?

A. Yes B. No **C. We can't be sure**

Solution: There is a 90% chance that the interval captures the true mean age μ, and there is a 10% chance that it does not. Once the confidence interval is built, it either contains the true mean μ or it does not. There is a better chance that it does contain it, but we cannot be sure it does. We may have selected a sample that is one of the 10% of intervals that does not contain μ.

(d) Why can we assume that $\bar{X}$ is approximately normally distributed?

Solution: Since *n* is "large" (greater than 30), then $\bar{X}$ is normally distributed according to the Central Limit Theorem. Thus, we can build the confidence interval using a 1-sample *z* confidence interval for a population mean.

Exercise 11

(a) Let *X* represent the number of gallons of gas needed to fill your car's gas tank. In order to use the formula from the lesson to construct a confidence interval for the true mean number of gallons of gas μ to fill your gas tank, we have to assume that $\bar{X}$ is normally distributed. Why is this a reasonable assumption based on the boxplot and histogram? Please note that this question is asking about the shape of $\bar{X}$ and not *X*.

Solution: Our original data is symmetric and looks somewhat "normal." We don't need a "large" sample size to guarantee normality of $\bar{X}$. Since the original data is symmetric, then $n = 20$ is a reasonable sample size to guarantee that the sampling distribution of the mean will be normal. If the original distribution was highly skewed, then the sampling distribution for $\bar{X}$ with $n = 20$ identically skewed distributions may not be normal.

(b) The sample mean $\bar{x}$, sample standard deviation *s*, and standard error of the mean for the gas fill-up data are as follows:

Descriptive Statistics: Gas Fill-up (gal)

Variable	Total Count	Mean	SE Mean	StDev
Gas Fill-up (gal)	20	10.159	0.0207	0.0925

Construct and evaluate a two-sided 95% confidence interval for μ.

Solution:

$$10.189 \pm 1.96 \cdot \frac{0.113}{\sqrt{20}} \rightarrow [\mathbf{10.139}, \mathbf{10.239}]\ \textbf{gallons}$$

(c) Using exactly the same data set to construct a 99% confidence interval for μ, the width of the interval:

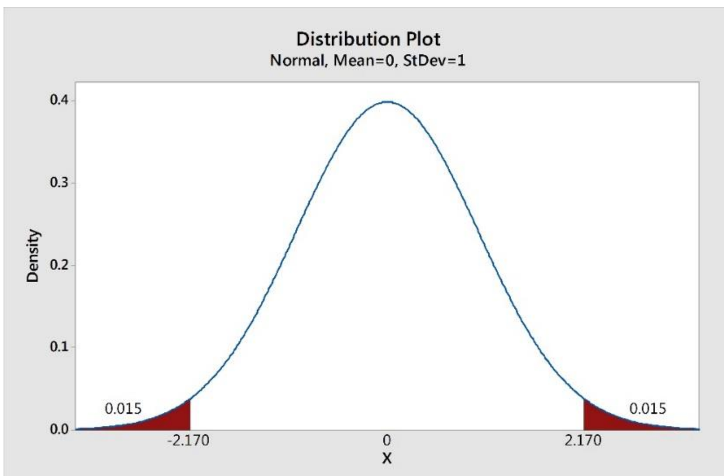
A. Increases B. Decreases C. Is not affected D. Not enough information

Exercise 12

For sample size *n* = 64, use a standard normal table or Minitab to fill in the *z*-score necessary to construct a 97% confidence interval for μ.

$$\bar{x} \pm \mathbf{2.432} * \frac{s}{\sqrt{64}}$$

Solution: A 97% confidence interval requires a critical value of $z_{0.015}$, which is approximately 2.17. This value can be determined using the standard normal table or Minitab (as shown in the ***Normal Distribution*** lesson).

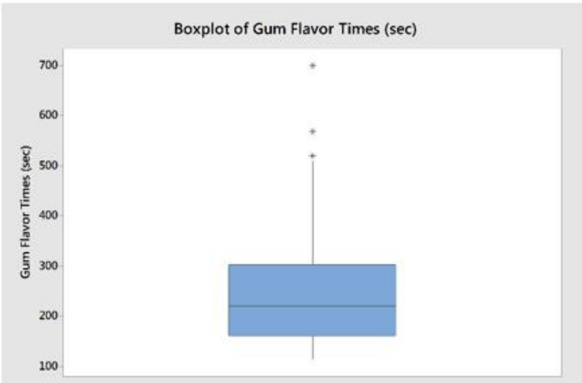


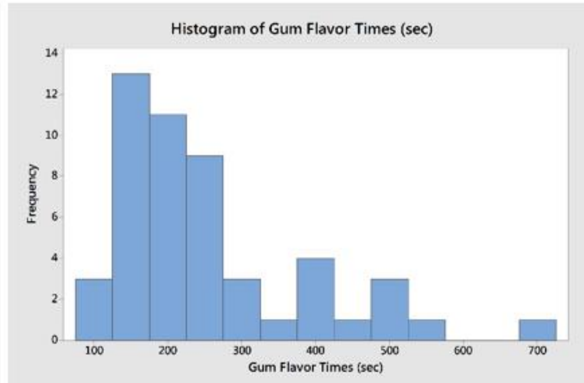
Exercise 13

If you do not have time to collect class data or do not have at least 30 students to perform this activity, then you can use the 'Gum Flavor Times (sec)' data found in C6 of ***CIForMean_LargeSample_Activity.mtw***. Below are solutions using the C6 data.

(a) In Minitab, construct both a boxplot and histogram of the data. If you do not remember how to do this, see the ***Describing Data Graphically*** lesson.

Solution:





(b) Describe the basic shape (symmetric, positively skewed, negatively skewed) and spread (amount of variability) of the data. It may not clearly fall into one specific category.

Solution: The data provided in C6 is positively skewed. The range of the data is approximately 600 seconds, which is 10 minutes. That's a large time discrepancy in measuring the length of chewing time before gum loses its flavor. It is not entirely surprising since the time until gum loses flavor is a subjective quantity. We did not provide an operational definition for measuring this time, so we cannot be sure what students are considering "loss of flavor." Also, the standard deviation is over 2 minutes at approximately 131.4 seconds.

(c) Based on your answer to **(b)** and the sample size, can we assume $\bar{X}$ is normally distributed? Why or why not?

Solution: Yes. Although the data provided in C6 is clearly positively skewed, the large sample size of n = 50 will allow $\bar{X}$ to take on the shape of a normal distribution.

(d) In Minitab, construct a two-sided 90% confidence interval for μ.

Solution:

One-Sample Z: Gum Flavor Time (sec)

```
The assumed standard deviation = 131.4


Variable                  N   Mean  StDev  SE Mean       90% CI
Gum Flavor Time (sec)   50  255.8  131.4     18.6  (225.2, 286.3)
```

(e) [BONUS] An investigator computes a 95% confidence interval for a population mean using a sample of size of n = 70. If she wishes to compute a 95% confidence interval that is half as wide, how large a sample does she need?

Solution: Let $\bar{x}$ represent the sample mean for the unknown data, and let σ represent the unknown population standard deviation. Just using the information that we are given, the confidence interval is:

$$\left[\bar{x} - 1.96 * \frac{\sigma}{\sqrt{70}}, \bar{x} + 1.96 * \frac{\sigma}{\sqrt{70}}\right]$$

The width of the confidence interval is:

$$\left(\bar{x} + 1.96 * \frac{\sigma}{\sqrt{70}}\right) - \left(\bar{x} - 1.96 * \frac{\sigma}{\sqrt{70}}\right) = \mathbf{2 * 1.96 * \frac{\boldsymbol{\sigma}}{\sqrt{70}}}$$

Half of this width is:

$$\frac{\mathbf{1}}{\mathbf{2}} * \left(2 * 1.96 * \frac{\sigma}{\sqrt{70}}\right) = \left(2 * 1.96 * \frac{\sigma}{\mathbf{2} * \sqrt{70}}\right) = \left(2 * 1.96 * \frac{\sigma}{\sqrt{\mathbf{4} * 70}}\right) = \left(2 * 1.96 * \frac{\sigma}{\sqrt{\mathbf{280}}}\right)$$

Thus, the sample size would need to be ***n* = 280**.