Minitab

LESSON: POPULATION MEAN CONFIDENCE INTERVALS FOR LARGE SAMPLES
This lesson includes an overview of the subject, instructor notes, and example exercises using Minitab.

Population Mean Confidence Intervals for Large Samples

Lesson Overview

Given a large sample size *n* (where we will assume $n > 30$), we will construct confidence intervals for the mean μ of a population.

Prerequisites

This lesson requires knowledge from the ***Normal Distribution*** and ***Sampling Distribution of*** $\bar{X}$ (i.e. Central Limit Theorem) lessons. Students need to understand that the distribution of $\bar{X}$ is approximately normally distributed for a large sample size *n*. Otherwise, the process that we use in this lesson for constructing a confidence interval with a large sample will not make sense.

Learning Targets

This lesson teaches students how to:

- Interpret what a confidence interval says about the likelihood of the interval containing the parameter of interest.
- Construct a confidence interval by hand and use a standard normal table for a population mean given a large sample size *n*.
- Construct a confidence interval using Minitab for a population mean given a large sample size *n*.

Time Required

It will take the instructor at least 60 minutes in class to introduce this lesson if the confidence interval simulation exercise is included—it is highly recommended. This exercise will help students understand the correct interpretation of a confidence interval for a given confidence level. It will take about 15 minutes for students to generate their confidence intervals and write them on the board during class. The exercises on the activity sheet will also take 60 minutes, and they can be used as homework or quiz problems.

Materials Required

- Minitab 17 or Minitab Express
- Minitab worksheet of sample data, entitled ***CIForMean_LargeSample_Lesson.mtw***
- Students will need Minitab in class for the confidence interval simulation. If you do not have time or Minitab access in class, then you can use the simulated data provided at the end of this lesson.

Assessment

The activity sheet contains exercises for students to assess their understanding of the learning targets for this lesson.

Possible Extensions

This lesson introduces students to constructing confidence intervals for a population mean with a large sample size. The recommended follow-up lesson is ***Population Mean Hypothesis Testing for Large Samples***.

References

For the Activity: "The Blind Paper Cutter: Teaching about Variation, Bias, Stability, and Process Control," by Richard Stone. Available from The American Statistician, Volume 52, No. 3 (August 1998), pp. 244-247.

Miscellaneous

Cartoon Images: freepik.com, clipart panda, classroomclipart

Instructor Notes with Examples

Confidence Intervals

Recall that the **mean of a population**, or the average of an entire set of data, is **μ**. The **sample mean $\bar{x}$** is the mean of a sample from that population. Often for a given population, we don't know μ, but we can determine $\bar{x}$ by selecting a random sample of data from the population. This lesson uses $\bar{x}$ to determine an interval estimate for μ.

Scenario: We have a population, and we want to estimate the **population parameter** μ.

- You might ask, "Why don't we just compute the population mean rather than calculate an estimate for it?" Good question! Suppose your population of interest is Canada, and you want to know the mean age of the population.
  - Due to lack of time, energy, and money, you cannot obtain the age of every person in Canada.
  - You can select a sample (e.g. a simple random sample – see the ***Sampling*** lesson) and calculate the mean of that sample, $\bar{x}$.
- Your next question may be, "Why don't we just use the sample mean $\bar{x}$ to estimate the population mean μ?" Another good question!
  - We can – but the sample mean $\bar{x}$ may be *quite* different from the population mean μ, even if we obtained the sample correctly.
  - In addition, a single number estimate by itself, such as $\bar{x}$, provides no information about the precision and reliability of the estimate with respect to the larger population.
- Statisticians use the sample statistic $\bar{x}$ and the population or sample standard deviation to provide an interval of plausible estimates for the population parameter μ. This interval is called a **confidence interval**.

Definition: A ***confidence interval*** is an entire interval of plausible values for a population parameter, such as μ, based on observations obtained from a **random sample** of size ***n***.

Definition: The ***confidence level*** is a measure of the degree of reliability of the confidence interval.

Minitab Confidence Interval Simulation Exercise

A Minitab simulation is very helpful in understanding what a confidence interval says about the likelihood of an interval containing the parameter of interest. Explaining *how* the interval is being constructed is discussed later and is not required to understand the results of the simulation. The goal of this simulation is to see the relationship between the parameter of interest, the confidence level, and the confidence interval.

1. In Minitab, have each student construct ten columns, each with 100 data points sampled from a standard normal population.

Minitab 17

1 Choose **Calc > Random Data > Normal**.
2 In **Number of rows of data to generate**, enter *100*.
3 In **Store in columns**, enter *C1-C10*.
4 In **Mean**, enter *0*.
5 In **Standard deviation**, enter *1*.
6 Click **OK**.

Minitab Express

1 Open the generate random data dialog box.
   - Mac: **Data > Generate Random Data**
   - PC: **DATA > Random Data**

2 In **Number of columns to generate**, enter *10*.
3 In **Number of rows in each column**, enter *100*.
4 From **Distribution**, select **Normal**.
5 In **Mean**, enter *0*.
6 In **Standard deviation**, enter *1*.
7 Click **OK**.

2. Have students examine their columns of random data from a standard normal population.

   a) Are the sample values in the columns representative of numbers that they'd expect to see from a normal distribution with mean 0 and standard deviation 1?
   b) Ask if anyone has a sample value greater than 3 or less than -3 in any of their columns. From the 68–95–99.7 rule (see the ***Normal Distribution*** lesson), 99.7% of the data from a normal distribution lies within 3 standard deviations of the mean.

3. Have students compute 95% confidence intervals for the population mean for each column. How the intervals are constructed will be discussed later in this lesson.

Minitab 17

1 Choose **Stat > Basic Statistics > 1-Sample Z**.
2 In **One or more samples, each in a column**, enter *C1-C10*.

3 In **Known standard deviation**, enter 1. [Recall, we sampled data from a standard normal distribution with standard deviation 1.]
4 Click **OK**.

Minitab Express

1 Open the 1-Sample Z dialog box.
- Mac: **Statistics > 1-Sample Inference > Z**
- PC: **STATISTICS > One Sample > Z**

2 From the drop-down list, select **Sample data in a column**.
3 In **Sample**, enter *C1*.
4 In **Known standard deviation**, enter *1*. [Recall, we sampled data from a standard normal distribution with standard deviation 1.]
5 Click **OK**.
6 To re-run the analysis for **Sample** columns *C2, C3, …, C10*, select **Make Similar** above the Output Pane.

4. Have students write their ten 95% confidence intervals for the population mean μ on the front board, rounding the intervals to 2 decimal places to save class time.

   The appendix for this lesson includes one hundred 95% confidence intervals for the population mean μ, where the population of interest is a standard normal distribution. Below is one set of ten intervals, in which one of them does not contain the true population mean μ = 0.

Variable	N	Mean	StDev	95% CI
C1	100	0.032	1.069	(-0.164, 0.228)
C2	100	0.071	1.057	(-0.125, 0.267)
C3	100	0.114	1.084	(-0.082, 0.310)
C4	100	-0.070	1.012	(-0.266, 0.126)
C5	100	-0.091	0.903	(-0.287, 0.105)
C6	100	0.145	1.025	(-0.051, 0.341)
C7	100	0.216	1.035	(0.020, 0.412)
C8	100	0.056	1.049	(-0.140, 0.252)
C9	100	-0.006	1.033	(-0.202, 0.190)
C10	100	-0.044	1.016	(-0.240, 0.152)

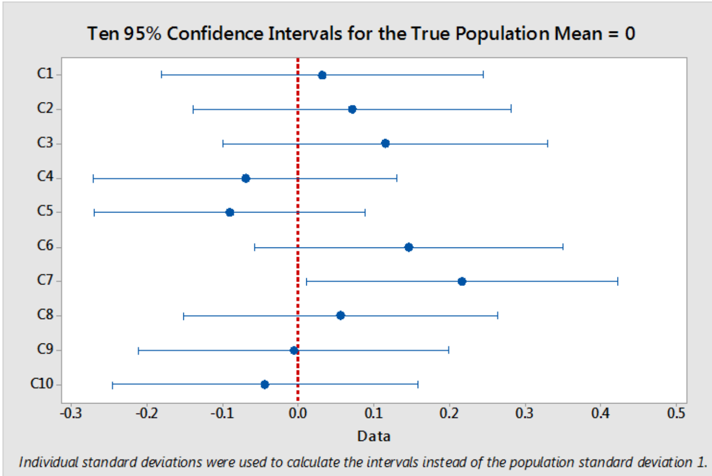
5. With the students, highlight and count the number of 95% confidence intervals for the population mean that don't contain the true mean, which we set as μ = 0. You should notice that approximately 95% of them will contain the population mean μ = 0, while approximately 5% of them will not.

6. Make sure the students see that the confidence intervals are unique since each random sample of size n = 100 is unique. It's important to reinforce in students' minds that the random part of the confidence interval is the interval, and not the parameter μ.

Minitab Confidence Interval Simulation Exercise Notes:

- On average, *approximately* 95% of the intervals will contain the true value of μ (which we set as 0). This is how to interpret a 95% confidence interval – there's a 95% chance, **before constructing the interval**, that the interval will actually contain the parameter of interest.
- If we had constructed 90% confidence intervals instead of 95%, then approximately 90% of the intervals would have contained the true value of μ. Similarly, if we had constructed 99% confidence intervals, then approximately 99% of the intervals would have contained the true value of μ. ***The trade-off for increased confidence is a wider interval***.
- The true μ (in this case μ = 0) is not changing. The population mean, although unknown, is *fixed*. We hope that our confidence interval captures μ. Before we cast our net for μ, we have a 95% chance of catching μ. Once we draw a random sample from the population of interest, the net has been tossed and either we have or have not captured μ.
- We *expect* approximately 95% of the intervals we construct to contain μ, but we also expect some variation. That is, in any group of 100 samples, it is possible to find only, say, 92 intervals that contain μ. In another group of 100 samples, we might find 97 that contain μ, and so forth. The 95% refers to the *overall* percentage of intervals that will contain the true population mean.

A confidence interval **may** or **may not** actually contain the true value of the population parameter. Each confidence interval computed for the same parameter will be different since it depends on the random sample taken from the population. Below is a graph of the ten 95% confidence intervals computed for the population mean in the simulation exercise. Only nine of them contain the true mean μ = 0, while the confidence interval for C7 does not.



Example 1

A survey claims that the average earnings for college students for a one-month internship is μ = \$4500. Your students tell you this amount seems high. To test this claim, you take a random sample of n = 100 students from your college population. The sample mean $\bar{x}$ = \$3975.

Using the random sample of 100 students' monthly internship salaries at your college, you calculate a 95% confidence interval for the mean to be [\$3525, \$4425]. Notice that this interval does not contain the survey result's stated mean of μ = \$4500. *What can we conclude?*

From the Minitab confidence interval simulation exercise, 95% of the intervals built in this random fashion will capture the true mean monthly earnings μ for summer college students at your college. One of the following statements must be true:

- The interval you computed just happens to be one of the 5% of confidence intervals that does not capture μ. You could take another sample to see if it reflects the results of the first confidence interval.
- The survey result's claim of students earning an average of \$4500 for a one-month internship is incorrect at least at the 5% level for your college.
- Your college does not reflect a random sample of the population of all college students. To test the survey result's claim, you would need to choose a random sample of students across the population of interest.

Important: For the rest of this lesson, large $\bar{X}$ represents a random variable and $\bar{x}$ is an actual value from that distribution. Small $\bar{x}$ is the sample mean computed with data. Large $\bar{X}$ is the distribution of all possible small $\bar{x}$'s.

Example 2

A quality control engineer works at a cereal company, and she wants to estimate the **true mean fill weight μ** of cereal boxes filled on a given day. She draws a simple random sample of 100 boxes from the population of all cereal boxes that are filled that day. The **population standard deviation** is known to be **σ = 0.1 oz.** for this process. She computes the **sample mean fill weight** to be $\bar{x}$ **= 12.05 oz.**

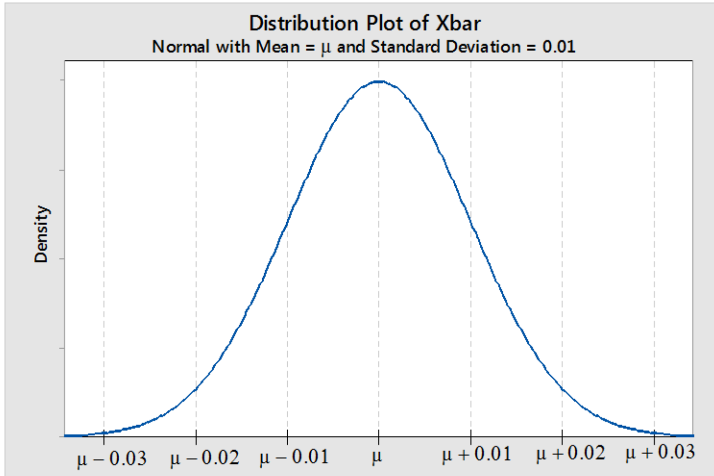
- Let μ represent the true (unknown) population mean fill weight. The population standard deviation is σ = 0.1 oz.
- Let x_1, x_2, ..., x_{100} be the individual weights of the simple random sample of 100 cereal boxes.
- The value $\bar{x}$ is a sample average weight of a random sample of 100 cereal boxes. Every random sample of 100 boxes will produce a different $\bar{x}$.
- Let $\bar{X}$ represent the distribution of all $\bar{x}$'s from different random samples from the population of all cereal boxes produced that day.
- From the ***Sampling Distribution of*** $\bar{X}$ lesson, the **Central Limit Theorem** says:

> Let X_1, X_2, X_3, ..., X_n be independent random variables with identical distributions with mean μ and standard deviation σ. If n is "large" enough ($n > 30$ suggested in most texts), then:
>
> The sampling distribution of the sample mean $\bar{X}$ is *approximately* normally distributed with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$ if the original distributions are non-normal.
>
> The larger the sample size n is, the more normally distributed the sampling distribution will be and the more tightly it will converge about the true population mean μ.

- The distribution of the sample mean $\bar{X}$ is *exactly* normally distributed with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$ if the original distributions are normal.
- Let $\mu_{\bar{X}}$ and $\sigma_{\bar{X}}$ represent the mean and standard deviation of the distribution of $\bar{X}$, respectively.
- For the cereal box fill weight example, $\bar{X}$ is:

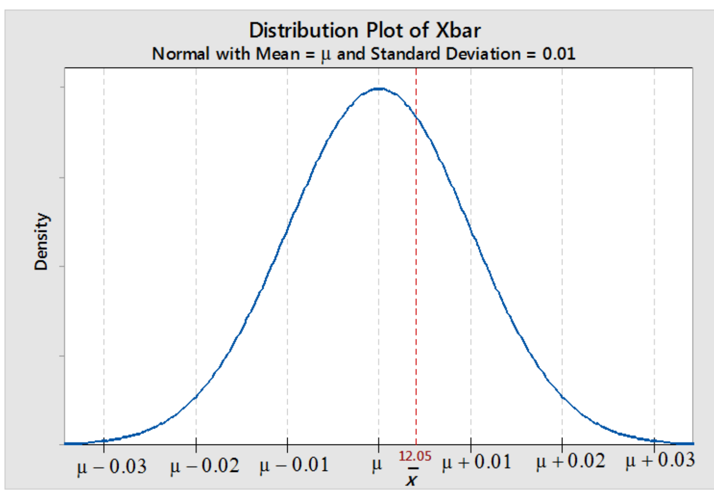
$$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_{100}}{100}$$

  with mean $\mu_{\bar{X}} = \mu$ oz. and standard deviation $\sigma_{\bar{X}} = \frac{0.1}{\sqrt{100}} = 0.01$ oz.
- Independent of the original distribution, the approximate distribution of $\bar{X}$ for many samples of size $n = 100$ has the following shape:



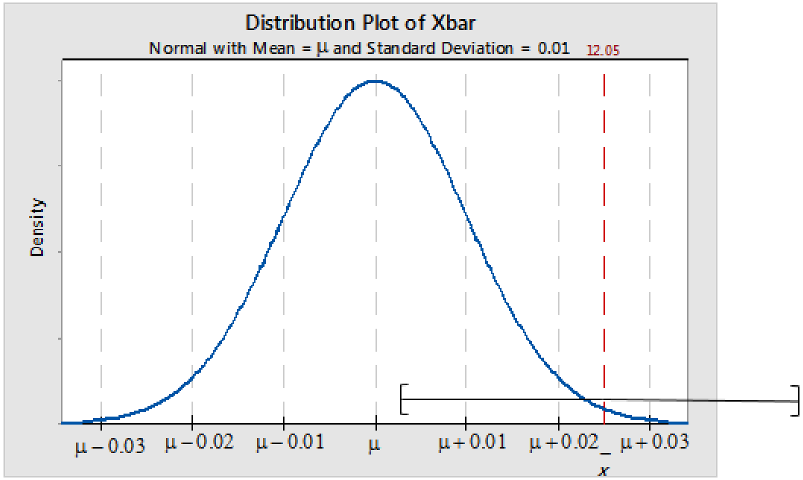
Let's consider two separate cases for the location of $\bar{x}$ = 12.05 on the distribution of $\bar{X}$ graph.

Case 1: Suppose the calculated sample average $\bar{x}$ lands within one standard deviation of the true population mean μ. We have no idea what μ is, but we hope to capture μ with a 95% confidence interval constructed with $\bar{x}$ as the center of the interval. Below is a hypothetical graph of this case.



Conclusion to Case 1: The 95% confidence interval captures μ as shown above.

Case 2: Suppose the calculated sample average $\bar{x}$ lands outside of two standard deviations from the true population mean μ. Again, we have no idea what μ is, but we hope to capture μ with a 95% confidence interval constructed with $\bar{x}$ as the center of the interval. Below is a hypothetical graph of this case.



Conclusion to Case 2: The 95% confidence interval does not capture μ as shown above.

- When constructing a 95% confidence interval, there is a 95% chance that we'll obtain a sample mean $\bar{x}$ that is within two standard deviations of the population mean. Constructing a confidence interval around sample means that fall within two standard deviations of the population mean will guarantee that we capture μ.
- There is a 5% chance that we'll obtain a sample mean $\bar{x}$ that is outside of two standard deviations of the population mean. Constructing a confidence interval around these sample means will not capture μ.

Simplistic view of calculating the confidence interval using the 68–95–99.7 rule:

- Although we don't know the value of μ, the 68-95-99.7 rules says that approximately 95% of data lies within two standard deviations of μ.
- The probability of drawing a sample with a mean $\bar{x}$ that is within two standard deviations of μ is 95%.
- If we obtain a sample mean $\bar{x}$ that is within two standard deviations of μ, then putting 95% confidence interval bounds around it will capture the true population mean μ.
- The probability of drawing a sample with a mean $\bar{x}$ that is more than two standard deviations away from μ is 5%.
- If we obtain a sample mean $\bar{x}$ that is more than two standard deviations away from μ, then putting 95% confidence interval bounds around it will not capture the true population mean μ.

To calculate an approximate 95% confidence interval in this example, we let $\bar{x}$ be at the center of the interval. To attempt to capture μ within the interval, we extend the upper interval bound to two standard deviations to the right of $\bar{x}$. The lower interval bound is set at two standard deviations to the left of $\bar{x}$. Symbolically, here is what we have:

- Center of confidence interval: $\bar{x} = 12.05 \text{ oz.}$

- Upper confidence interval bound: $\bar{x} + 2 * \sigma_{\bar{X}} = 12.05 + 2 * 0.01 = 12.07$ oz.
- Lower confidence interval bound: $\bar{x} - 2 * \sigma_{\bar{X}} = 12.05 - 2 * 0.01 = 12.03$ oz.

Our 95% confidence interval for the true population mean μ is approximately **[12.03, 12.07] oz.**

To compute this interval using Minitab:

1. Open the 1-Sample Z dialog box.
   - Minitab 17: **Stat > Basic Statistics > 1-Sample Z**
   - Minitab Express Mac: **Statistics > 1-Sample Inference > Z**
   - Minitab Express PC: **STATISTICS > One Sample > Z**
2. From the drop-down list, select **Summarized data**.
3. In **Sample size**, enter *100*.
4. In **Sample mean**, enter *12.05*.
5. In **Known standard deviation**, enter *0.1*.
6. Click **OK**.

One-Sample Z

```
The assumed standard deviation = 0.1

  N     Mean  SE Mean         95% CI
100  12.0500   0.0100  (12.0304, 12.0696)
```

Minitab reports the 95% confidence interval as: **[12.0304, 12.0696] oz.** This answer differs slightly from our previous calculation by hand because the probability of being between -2σ and +2σ on a normal curve is slightly greater than 0.95; it is actually 0.9545. We will improve our calculation by hand later in this lesson.

Question: Does the 95% confidence interval that we constructed for the true population mean μ actually contain μ?

Answer: Probably. The above 95% confidence interval for μ is computed by a procedure that succeeds in covering the population mean 95% of the time. There's a 95% chance that we captured μ and a 5% chance that we did not.

Example 3

Choose the best interpretation of a 95% confidence interval for the population mean μ.

A. If repeated random samples were taken and the 95% confidence interval was computed for each sample, 95% of the intervals would contain the population mean.

B. The probability that the population mean μ is in the confidence interval is 0.95.

C. 95% of the population distribution is contained in the confidence interval.

Answer: The correct answer is A; it follows the confidence interval simulation exercise that we did at the beginning of the lesson. Answer B is incorrect because it places the probability on μ, instead of on the confidence interval. Answer C is incorrect since the confidence interval for the population mean is built using sample means and not values from the population distribution. Using population distribution values would give us a confidence interval that is wider than the one for the population mean.

Confidence Levels

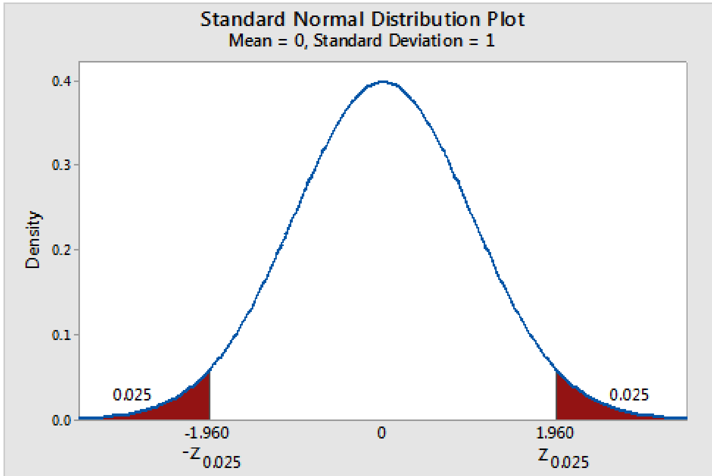
Let the Greek letter α be a number between 0 and 1, and let 100 * (1 – α)% denote the **confidence level**. For example, if α = 0.05, then the corresponding confidence level is 95%. If α = 0.01, then the confidence level is 99%.

Notation:

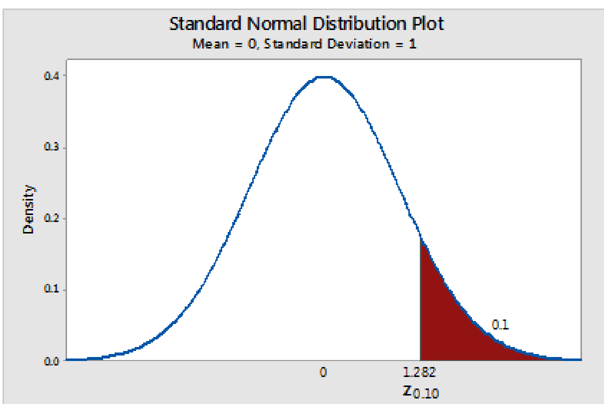
- Suppose we have a standard normal distribution *Z*.
- Let $z_{\alpha/2}$ denote a ***z*-score** with **α/2 probability** to its **right**. Note that $\alpha/2$ is simply a subscript of *z*, and we use this notation for writing a compact formula for a confidence interval later in this lesson. Texts use the same notation.
- Similarly let $-z_{\alpha/2}$ denote a ***z*-score** with **α/2 probability** to its **left.**

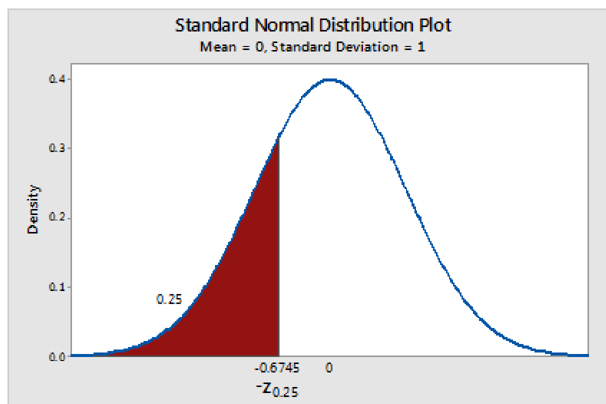
Examples:

- The value $z_{0.025}$ is the positive *z*-score that has α/2 = 0.025 probability to its right. Using the standard normal table or Minitab (as shown in the ***Normal Distribution*** lesson), the desired *z*-score is **1.96**.
- The value $-z_{0.025}$ is the negative *z*-score that has α/2 = 0.025 probability to its left. The desired *z*-score is **-1.96**.



- The value $z_{0.10}$ is the positive *z*-score that has α/2 = 0.1 probability to its right. The desired *z*-score is **1.282**.
- The value $-z_{0.25}$ is the negative *z*-score that has α/2 = 0.25 probability to its left. The desired *z*-score is **-0.6745**.





The most commonly used α's and their *z*-scores are listed in this table:

100*(1-α)% Two-Sided Confidence Interval	**α**	**α/2**	$\pm z_{\alpha/2}$
99% Two-sided Confidence Interval	0.01	0.005	**±2.58**
95% Two-sided Confidence Interval	0.05	0.025	**±1.96**
90% Two-sided Confidence Interval	0.10	0.05	**±1.645**

We are now ready for the formula for a confidence interval for the population mean μ given a large sample size *n*.

A ***two-tailed 100 * (1-α)% confidence interval for μ*** when n is large is:

$$\left[\bar{x} - z_{\alpha/2} * \sigma_{\bar{X}}, \bar{x} + z_{\alpha/2} * \sigma_{\bar{X}}\right] = \left[\bar{\boldsymbol{x}} - \boldsymbol{z}_{\alpha/2} * \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}, \bar{\boldsymbol{x}} + \boldsymbol{z}_{\alpha/2} * \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}\right]$$

If σ is unknown, replace it with the sample standard deviation s for large n ($n > 30$). We call $\frac{\sigma}{\sqrt{n}}$ the ***standard error of the mean*** and Minitab refers to it in output as *SE Mean*.

Example 4

Every child loves getting his or her first bicycle. However, many parents dread the painful task of reading through confusing instructions to put that bike together. To make sure a parent leaves enough time for bike building before its exciting reveal, wouldn't it be nice if the instruction manual provided a confidence interval for the true population mean time to put the bike together?

Suppose a bike manufacturer wants to construct a confidence interval for the true mean time to put its most popular bike together. Instead of asking buyers to send in a survey (reread the ***Sampling*** lesson if necessary) about construction times, they randomly select 50 potential buyers and have them assemble the bike and time themselves.

The mean time to assemble the bike based on the random sample of $n = 50$ people is $\bar{x} = 10.4$ minutes. The population standard deviation of the assembly times is known to be σ = 1.2 minutes.

(a) Determine a two-sided 95% confidence interval for the true population mean assembly time μ. What assumption are you making about the distribution of $\bar{X}$?

Solution: Although we do not know the population distribution for the assembly times, we can assume that $\bar{X}$ is approximately normally distributed since the **sample size is *n* = 50**. Since we are building a two-sided 95% confidence interval for the population mean, we need to split $\alpha = 0.05$ into two halves and allow 0.025 probability in each tail of the normal curve. The z-score corresponding to 0.025 probability in the right tail of a standard normal distribution is 1.96, and -1.96 for 0.025 probability in the left tail. Using the confidence interval formula above, we have:

$$\left[\bar{x} - z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}\right] \cong \left[10.4 - 1.96 * \frac{1.2}{\sqrt{50}}, 10.4 + 1.96 * \frac{1.2}{\sqrt{50}}\right] \cong \mathbf{[10.067, 10.733]\ minutes}$$

To compute this interval using Minitab:

1 Open the 1-Sample Z dialog box.
   - Minitab 17: **Stat > Basic Statistics > 1-Sample Z**
   - Minitab Express Mac: **Statistics > 1-Sample Inference > Z**
   - Minitab Express PC: **STATISTICS > One Sample > Z**

2 From the drop-down list, select **Summarized data**.
3 In **Sample size**, enter *50*.
4 In **Sample mean**, enter *10.4*.
5 In **Known standard deviation**, enter *1.2*.
6 Click **OK**.

Minitab yields the following 95% confidence interval:

One-Sample Z

```
The assumed standard deviation = 1.2

 N    Mean   SE Mean        95% CI
50  10.400    0.170   (10.067, 10.733)
```

(b) Determine a two-sided 99% confidence interval for the true population mean assembly time μ.

Solution: Since we are building a two-sided 99% confidence interval for the population mean, then we need to split α = 0.01 into two halves and allow 0.005 probability in each tail of the normal curve. The *z*-scores corresponding to 0.005 probability in the right tail and left tail of a standard normal distribution is 2.58 and -2.58, respectively. Using the confidence interval formula above, we have:

$$\left[\bar{x} - z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}\right] \cong \left[10.4 - 2.58 * \frac{1.2}{\sqrt{50}}, 10.4 + 2.58 * \frac{1.2}{\sqrt{50}}\right] \cong \mathbf{[9.962, 10.838]\ minutes}$$

Notice that we although we have greater reliability with a 99% confidence interval, compared to 95%, we now have a wider interval and therefore less precision.

Since we want a 99% confidence interval, we'll need to add an extra step to the Minitab instructions from part **(a)**.

1 Open the 1-Sample Z dialog box.
   - Minitab 17: **Stat > Basic Statistics > 1-Sample Z**
   - Minitab Express Mac: **Statistics > 1-Sample Inference > Z**
   - Minitab Express PC: **STATISTICS > One Sample > Z**

2 From the drop-down list, select **Summarized data**.
3 In **Sample size**, enter *50*.

4. In **Sample mean**, enter *10.4*.
5. In **Known standard deviation**, enter *1.2*.
6. Enter the desired confidence level.
    - Minitab 17: Click **Options**. In **Confidence level**, enter *99*. Click **OK**.
    - Minitab Express Mac: Click the **Options** tab. Under **Confidence level**, select *99*. Click **OK**.
    - Minitab Express PC: Click the **Options** tab. Under **Confidence level**, select *99*. Click **OK**.

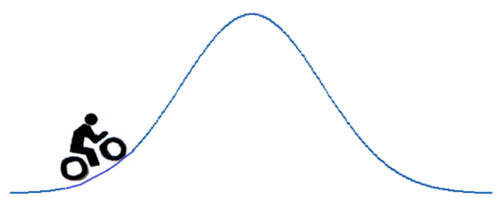
Minitab yields the following 99% confidence interval:

One-Sample Z

```
The assumed standard deviation = 1.2

 N    Mean  SE Mean       99% CI
50  10.400    0.170  (9.963, 10.837)
```

Example 5



On any given day, customers call the bike manufacturer with questions regarding assembly instructions for its most popular bike. The time between phone calls is recorded for n = 64 randomly selected calls and entered into the Minitab worksheet in the column named "Time Btw Calls (mins)." Determine a 90% confidence interval for the true population mean time between phone calls μ based on the sample data.

Solution: Like **Example 4**, we do not know the population distribution of the time between calls, but we can assume that $\bar{X}$ is approximately normally distributed since the **sample size is n = 64**. Since we are building a two-sided 90% confidence interval for the population mean, we need to split α = 0.10 into two halves and allow 0.05 probability in each tail of the normal curve. The z-score corresponding to 0.05 probability in the right tail of a standard normal distribution is 1.645, and -1.645 for 0.05 probability in the left tail.

In this example, we are given a set of data instead of summarized statistics. We can use Minitab to determine the sample mean $\bar{x}$ and the sample standard deviation s (see the ***Describing Data Numerically*** lesson). Since we don't know the population standard deviation σ, we can use s to estimate it, given the large sample size n. Here are the descriptive statistics from Minitab:

Descriptive Statistics: Time Btw Calls (mins)

Variable	Total Count	Mean	SE Mean	StDev	Minimum	Maximum
Time Btw Calls (mins)	64	3.176	0.336	2.687	0.103	13.132

Note that the standard error of the mean or "SE Mean" is equal to $\frac{s}{\sqrt{n}} \cong \frac{2.687}{\sqrt{64}} \cong 0.336$. Using the sample statistics, the 99% confidence interval is:

$$\left[\bar{x} - z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}\right] \cong [3.176 - 1.645 * 0.336, 3.176 + 1.645 * 0.336\,] \cong [\mathbf{2.623, 3.729}]\ \textbf{minutes}$$

We need to make a make a slight change to the Minitab instructions from the previous example for computing a confidence interval given a column of data.

Minitab 17

1 Choose **Stat > Basic Statistics > 1-Sample Z**.
2 In **One or more samples, each in a column**, enter *Time Btw Calls (mins)*.
3 In **Known standard deviation**, enter *2.687*.
4 Click **Options**. In **Confidence level**, enter *90*.
5 Click **OK** in each dialog box.

Minitab Express

1 Open the 1-Sample Z dialog box.
   - Mac: **Statistics > 1-Sample Inference > Z**
   - PC: **STATISTICS > One Sample > Z**
2 From the drop-down list, select **Sample data in a column**.
3 In **Sample**, enter *Time Btw Calls (mins)*.
4 In **Known standard deviation**, enter *2.687*.
5 Click the **Options** tab. Under **Confidence level**, select *90*.
6 Click **OK**.

Minitab yields the confidence interval:

One-Sample Z: Time Btw Calls (mins)

The assumed standard deviation = 2.687

Variable	N	Mean	StDev	SE Mean	90% CI
Time Btw Calls (mins)	64	3.176	2.687	0.336	(2.623, 3.728)

Appendix

This appendix contains one hundred confidence intervals for the population mean μ. The data was generated from a standard normal distribution with mean μ = 0 and standard deviation σ = 1.

In the simulation exercise, we know the population mean μ, but typically μ is unknown. Since we built 95% confidence intervals, we expect approximately 95 out of 100 of them to contain the true mean μ = 0. The highlighted intervals below do not contain the population mean μ = 0.

Anthony

Variable	N	Mean	StDev	95% CI
C1	100	0.067	0.925	(-0.129, 0.263)
C2	100	0.044	0.992	(-0.152, 0.240)
C3	100	0.009	0.972	(-0.187, 0.205)
C4	100	0.054	0.908	(-0.142, 0.250)
C5	100	-0.045	0.934	(-0.241, 0.151)
C6	100	0.230	1.113	(0.034, 0.426)
C7	100	0.030	0.898	(-0.166, 0.226)
C8	100	-0.241	1.061	(-0.437, -0.045)
C9	100	-0.189	0.980	(-0.385, 0.007)
C10	100	0.062	0.938	(-0.134, 0.258)

Barry

Variable	N	Mean	StDev	95% CI
C1	100	0.032	1.069	(-0.164, 0.228)
C2	100	0.071	1.057	(-0.125, 0.267)
C3	100	0.114	1.084	(-0.082, 0.310)
C4	100	-0.070	1.012	(-0.266, 0.126)
C5	100	-0.091	0.903	(-0.287, 0.105)
C6	100	0.145	1.025	(-0.051, 0.341)
C7	100	0.216	1.035	(0.020, 0.412)
C8	100	0.056	1.049	(-0.140, 0.252)
C9	100	-0.006	1.033	(-0.202, 0.190)
C10	100	-0.044	1.016	(-0.240, 0.152)

Cass

Variable	N	Mean	StDev	95% CI
C1	100	0.073	1.026	(-0.123, 0.269)
C2	100	-0.169	0.917	(-0.365, 0.027)
C3	100	-0.053	0.932	(-0.249, 0.143)
C4	100	0.137	1.000	(-0.059, 0.333)
C5	100	-0.081	0.988	(-0.277, 0.115)
C6	100	-0.035	0.914	(-0.231, 0.161)
C7	100	-0.105	1.072	(-0.301, 0.091)
C8	100	-0.015	1.027	(-0.211, 0.181)
C9	100	0.103	0.995	(-0.093, 0.299)
C10	100	0.115	0.903	(-0.081, 0.311)

Dave

Variable	N	Mean	StDev	95% CI
C1	100	0.074	0.903	(-0.122, 0.270)
C2	100	0.194	0.927	(-0.002, 0.390)
C3	100	-0.041	1.025	(-0.237, 0.155)
C4	100	-0.070	0.984	(-0.266, 0.126)
C5	100	-0.105	0.988	(-0.301, 0.091)
C6	100	-0.098	0.994	(-0.294, 0.098)
C7	100	0.049	1.037	(-0.147, 0.245)
C8	100	-0.031	1.022	(-0.227, 0.165)
C9	100	-0.044	0.983	(-0.240, 0.152)
C10	100	0.060	0.943	(-0.136, 0.256)

Eston

Variable	N	Mean	StDev	95% CI
C1	100	-0.054	1.018	(-0.250, 0.142)
C2	100	0.207	1.076	(0.011, 0.403)
C3	100	-0.046	0.956	(-0.242, 0.150)
C4	100	-0.095	1.013	(-0.291, 0.101)
C5	100	0.067	0.973	(-0.129, 0.263)
C6	100	-0.040	1.183	(-0.236, 0.156)
C7	100	-0.082	1.055	(-0.278, 0.114)
C8	100	-0.078	0.989	(-0.274, 0.118)
C9	100	0.028	0.997	(-0.168, 0.224)
C10	100	0.029	1.050	(-0.167, 0.225)

Finn

Variable	N	Mean	StDev	95% CI
C1	100	0.070	0.986	(-0.126, 0.266)
C2	100	-0.100	1.027	(-0.296, 0.096)
C3	100	0.111	0.847	(-0.085, 0.307)
C4	100	0.003	1.075	(-0.193, 0.199)
C5	100	0.109	1.083	(-0.087, 0.305)
C6	100	0.108	1.098	(-0.088, 0.304)
C7	100	0.250	0.875	(0.054, 0.446)
C8	100	-0.122	1.063	(-0.318, 0.074)
C9	100	-0.057	0.850	(-0.253, 0.139)
C10	100	0.140	1.082	(-0.056, 0.336)

Ginger

Variable	N	Mean	StDev	95% CI
C1	100	-0.011	0.947	(-0.207, 0.185)
C2	100	-0.018	1.017	(-0.213, 0.178)
C3	100	0.034	1.019	(-0.162, 0.230)
C4	100	-0.126	1.048	(-0.322, 0.070)
C5	100	-0.020	0.956	(-0.216, 0.176)
C6	100	-0.190	0.875	(-0.386, 0.006)
C7	100	-0.082	1.044	(-0.278, 0.114)
C8	100	0.086	1.043	(-0.110, 0.282)
C9	100	0.032	0.964	(-0.164, 0.228)
C10	100	0.088	0.890	(-0.108, 0.284)

Henry

Variable	N	Mean	StDev	95% CI
C1	100	0.214	1.018	(0.018, 0.410)
C2	100	-0.087	1.033	(-0.283, 0.109)
C3	100	-0.002	0.939	(-0.198, 0.194)
C4	100	0.060	1.043	(-0.136, 0.256)
C5	100	-0.026	1.032	(-0.222, 0.170)
C6	100	-0.024	0.978	(-0.220, 0.172)
C7	100	0.199	0.939	(0.003, 0.395)
C8	100	0.048	0.970	(-0.148, 0.244)
C9	100	0.109	1.090	(-0.087, 0.305)
C10	100	0.009	0.949	(-0.187, 0.205)

Ireland

Variable	N	Mean	StDev	95% CI
C1	100	-0.004	0.893	(-0.200, 0.192)
C2	100	0.193	1.077	(-0.003, 0.389)
C3	100	0.138	1.021	(-0.058, 0.334)
C4	100	-0.067	1.135	(-0.263, 0.129)
C5	100	-0.019	1.043	(-0.215, 0.177)
C6	100	-0.037	1.054	(-0.233, 0.159)
C7	100	-0.022	1.028	(-0.218, 0.174)
C8	100	0.112	0.955	(-0.084, 0.308)
C9	100	-0.025	1.111	(-0.221, 0.171)
C10	100	-0.037	0.878	(-0.233, 0.159)

Jack

Variable	N	Mean	StDev	95% CI
C1	100	0.094	0.931	(-0.102, 0.290)
C2	100	-0.031	0.945	(-0.227, 0.165)
C3	100	0.002	1.031	(-0.194, 0.198)
C4	100	-0.082	0.987	(-0.278, 0.114)
C5	100	0.084	0.994	(-0.112, 0.280)
C6	100	0.210	0.877	(0.014, 0.406)
C7	100	0.050	1.010	(-0.146, 0.246)
C8	100	0.126	1.086	(-0.070, 0.322)
C9	100	-0.009	0.909	(-0.205, 0.187)
C10	100	0.147	1.048	(-0.049, 0.343)

Conclusion

8 of the 100 95% confidence intervals above do NOT contain the true mean $\mu = 0$.