

MODULE 2 LIVE LECTURE

---

# INFERENTIAL STATISTICS AND ANALYTICS

# INFERENCEAL STATISTICS AND ANALYTICS

---

- ▶ Module 2 Confidence interval
- ▶ Module 2 homework
- ▶ Quiz 2
- ▶ Course project phase 2
- ▶ Module 2 Live classroom Grading
- ▶ Summary

# INFERENCEAL STATISTICS AND ANALYTICS

---

- ▶ ***Module 2 Confidence interval***
- ▶ Module 2 homework
- ▶ Quiz 2
- ▶ Course project phase 2
- ▶ Module 2 Live classroom Grading
- ▶ Summary

# MODULE 2 REVIEW OF ESTIMATES AND SAMPLE SIZES

Module 01 - Review of Basic Statistics

Module 02 - Estimates and Sample Size

Review of Estimates and Sample Sizes

Learning Widget

Module 02 Discussion - Interpreting Confidence Intervals

Module 02 Homework

Course Project - Phase 2

Module 02 Quiz

Module 02 Live Classroom

Module 03 - Hypothesis Testing with One Sample

## Review of Estimates and Sample Sizes

Module 02 - Estimates and Sample Size

Looking at Estimates and Sample Sizes

Inferential statistics are useful when we attempt to make generalizations from the data we have available to us. As a result, selecting a strong sample from which we can make draw conclusions is important. For example, what can the shopping habits of 300,000 Americans tell us about the entire country's population?

We know that statistics can sometimes be challenging, especially for students who are learning it for the first time. We also know that statistics are very important in analytics, which we define and communication and decision-making based on this type of data. As a result, **it's important for you to attend your weekly live classrooms so you can test your learning**

- Finding a sample size necessary to estimate a population proportion.

+ Calculating the Point Estimate

+ Confidence Intervals

+ Interpreting a Confidence Interval

+ Finding Critical Values

+ Calculating the Margin of Error

+ Constructing a Confidence Interval for p

+ Determining Sample Size

---

▶ **Confidence Intervals for the *Unknown population Mean* and the *known population Standard Deviation***

- ▶ When estimating a population mean, the **population standard deviation** is *known*. In this case, we construct the confidence interval using the *standard normal distribution*.

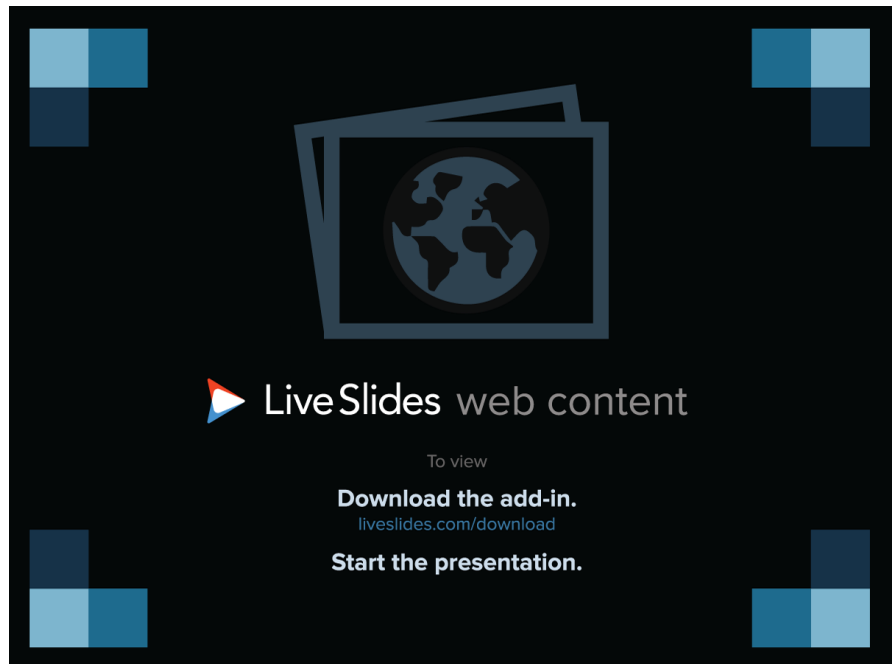
- ▶ ***How to construct the confidence interval for the unknown population mean and the *known* population standard deviation?---***>>

---

## ► **Confidence Intervals for Unknown population Mean and *known population Standard Deviation***

- When estimating a population mean, the standard deviation is **known**. In this case, we construct the confidence interval using the **standard normal distribution**.

- **How to construct the confidence interval for unknown population mean and *known population standard deviation*?---**►►



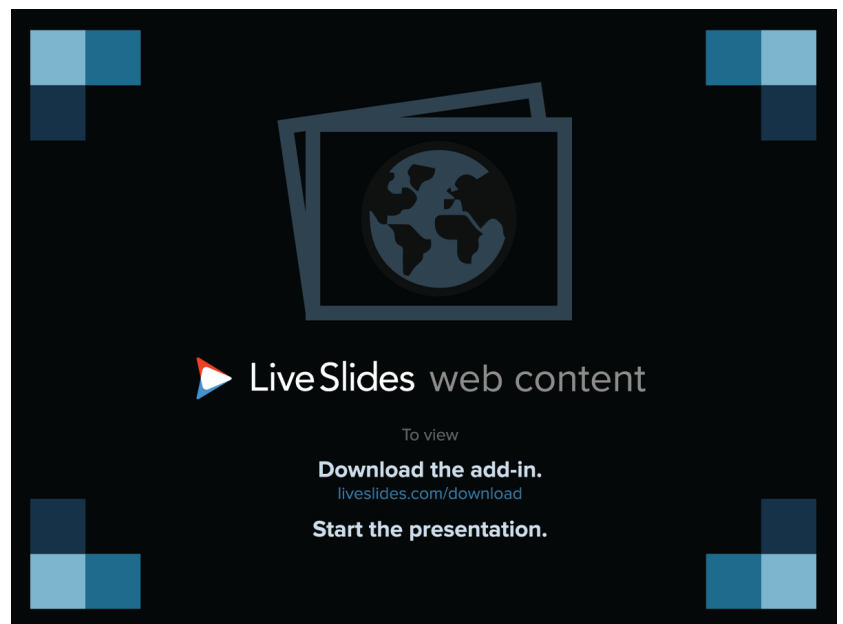
<https://www.youtube.com/watch?v=KG921rfbTDw&index=3&list=PLvxOuBpazmsMdPBRxBTvwLv5Lhuk0tuXh>

- 
- ▶ **Confidence Intervals for the Unknown population Mean and the *Unknown population* Standard Deviation**
  - ▶ When estimating a population mean, it is rare that the **population standard deviation** is known. In this case, we construct the confidence interval using the **Student's t distribution** instead of a standard normal distribution.
  - ▶ *How to construct the confidence interval for the unknown population mean and the **unknown** population standard deviation?---*>>

---

## ► **Confidence Intervals for Unknown population Mean and Unknown population Standard Deviation**

- When estimating a population mean, it is rare that the population standard deviation is known. In this case, we construct the confidence interval using the **Student's t distribution** instead of a standard normal distribution.
- **How to construct the confidence interval for unknown population mean and unknown population standard deviation?---**►►



<https://www.youtube.com/watch?v=bFefxSE5bmo&index=7&list=PLvxOuBpazmsMdPBRxBTvwLv5Lhuk0tuXh>



# INFERENCEAL STATISTICS AND ANALYTICS

---

- ▶ **Module 2 Confidence interval**
- ▶ **Module 2 homework**
- ▶ Quiz 2
- ▶ Course project phase 2
- ▶ Module 2 Live classroom Grading
- ▶ Summary

## MODULE 2 HOMEWORK

---

Problem 1: The local news provided poll results from 2000 adults who interviewed job applicants. The results showed that 35% of the adults said their biggest issue with interviewers is that they do not know company history.

a) What important piece of information was omitted from the statement above?  
Confidence interval

Confidence interval estimate of the sample proportion  $\hat{p}$ .  
eg. the confidence level is 90%

b) What is meant by the statement that "the margin of error is  $\pm 4$  percentage points"?  
When using 35% to estimate the value of the population percentage, the maximum likely difference between 35% and the true population percentage is 4 points.  
 $35\% - 4\% < \hat{p} < 35\% + 4\%$

c) the sample proportion  $\hat{p} = 35\%$ ,  $\hat{q} = 1 - \hat{p}$   
the population proportion  $p$  is unknown.

d) If the confidence level is 95%, what is the value of  $\alpha$ ?  
 $\alpha = 1 - 95\%$

## MODULE 2 HOMEWORK

---

Problem 2: In a poll of 555 randomly selected students, 40% stated that they enjoyed statistics. Answer the following questions:

a. Identify the number of students who say that they enjoy statistics

$$555 \times 40\% = 222$$

b. Construct a 95% confidence interval estimate of the percentage of all students who say that they enjoy statistics.

Confidence Interval: Proportion One Sample

Confidence Level: 0.95

Sample Size, n: 555

Number of Successes, x: 222

Evaluate

Margin of error,  $E = 0.0407574$

95% Confidence Interval (using normal approx):  
 $0.3592426 < p < 0.4407574$

Wilson Score Confidence Interval:  
 $0.3600645 < p < 0.4413103$

Print Copy

## MODULE 2 HOMEWORK

---

Problem 3: The following information provided below shows the output from the results of performing a confidence interval for a population mean. Answer the following questions:

|                     |                 |
|---------------------|-----------------|
| Confidence Interval | (233.4, 256.65) |
| $\bar{x}$           | 245.025         |
| $S_x$               | 36.357546       |
| n                   | 40              |

a. Identify the best point estimate of  $\mu$ .

The sample mean  $\bar{x}$  is the best point estimate of the population mean  $\mu$ .

b. Find the degrees of freedom.  
the degrees of freedom  $df=n-1$

d. Find the margin of error E  
Confidence interval (233.4, 256.65)

$$\bar{x} - E = 233.4$$

$$\bar{x} + E = 256.65$$

$$\bar{x} = 245.025$$

$$E = 256.65 - 245.025$$

c. Find the critical value  $t_{\alpha/2}$  that corresponds to  $n = 40$ .

$$E = t_{\alpha/2} \frac{S_x}{\sqrt{n}}$$

$$E = 256.65 - 245.025 \quad S_x = 36.35754604 \quad n=40$$

Problem 4: The cholesterol levels of 40 women were sampled and a 95% confidence interval estimate was obtained below. The units of measurement for the interval below are  $917.562 < \sigma^2 < 2254.129$

- a. Identify the confidence interval for the population standard deviation  $\sigma$ . Include the appropriate units of measure. Find the degrees of freedom.

$$\sqrt{917.562} < \sigma < \sqrt{2254.129}$$

The units of  $\sigma$  is mg/dL

## MODULE 2 HOMEWORK

---

Problem 5: You want to estimate the mean amount of time college students spend on the Internet each month. How many college students must you survey to be 95% confident that your sample mean is within 15 minutes of the population mean? Assume that the standard deviation of the population of monthly time spent on the Internet is 210 minutes.

The required sample size is  $n = \left(\frac{Z_{\alpha/2}\sigma}{E}\right)^2$

Margin of Error  $E=15$

Population standard deviation  $\sigma = 210$

The critical Value  $Z_{\alpha/2} = 1.959962$

You can use Statdisk to estimate the sample size.

Please go to Statdisk-> Analysis->Sample Size Determination-> Estimate Mean

Sample Size: Estimate Mean

|                                |      |   |
|--------------------------------|------|---|
| Confidence level:              | 0.95 | Required sample size is: 753<br><br>Assumed either infinite population or the population was sampled with replacement |
| Margin of Error, E:            | 15   |   |
| Population Standard Deviation: | 210  |   |
| Population Size:<br>(if known) |      |   |

Evaluate      Print      Copy

# INFERENCEAL STATISTICS AND ANALYTICS

---

- ▶ ***Module 2 Confidence interval***
- ▶ ***Module 2 homework***
- ▶ ***Quiz 2***
- ▶ Course project phase 2
- ▶ Module 2 Live classroom Grading
- ▶ Summary

## QUIZ 2

---

- 1) Use the given degree of confidence and sample data to construct a confidence interval for the population proportion  $p$ .  $n = 195$ ,  $x = 162$ ; 90% confidence

Confidence Level:

0.90

Sample Size,  $n$ :

195

Number of Successes,  $x$ :

162

Evaluate

Margin of error,  $E = 0.0441661$

90% Confidence Interval (using normal approx):  
 $0.7866031 < p < 0.8749354$

Wilson Score Confidence Interval:  
 $0.7821469 < p < 0.8703386$

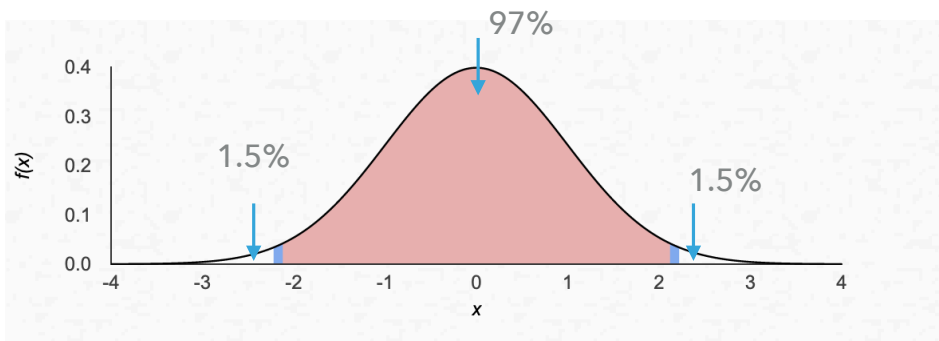
Print

Copy



## QUIZ 2

2) Find the value of  $z_{\alpha/2}$  that corresponds to a confidence level of 97%.



Normal Distribution

Enter one value, then click Evaluate to find the other value:

z Value:

Cumulative area from the left:

Evaluate

z Value: 2.170088  
Prob Dens: 0.0378706

Cumulative Probs  
Left: 0.985000  
Right: 0.015000  
2 Tailed: 0.030000  
Central: 0.970000  
As Table A-2: 0.985000

Print Copy

**Cumulative area from the left**

**=1.5%+97%**

**=98.5%**

**=0.985**

## QUIZ 2

---

3) 2000 people are selected randomly from a certain population and it is found that 389 people in the sample are over 6 feet tall. What is the point estimate of the proportion of people in the population who are over 6 feet tall?

$$\hat{p} = \frac{x}{n} \quad x = 389 \quad n = 2000$$

$$\hat{p} = \frac{389}{2000}$$

4) Thirty randomly selected students took the calculus final. If the sample mean was 95 and the standard deviation was 6.6, construct a 99% confidence interval for the mean score of all students.

Confidence Interval: Mean-One Sample

Confidence Level: 0.99

**Use Summary Statistics** Use Data

Sample Size, n: 30

Sample Mean: 95

Sample Standard Deviation, s: 6.6

Population Standard Deviation: (if known)

Evaluate

Margin of error, E =

99% Confident the po within the range: 91.67859 < mean < 98.

## QUIZ 2

---

- 5) Use the given data to find the minimum sample size required to estimate the population proportion.  
Margin of error: 0.028; confidence level: 99%;  $\hat{p}$  and  $\hat{q}$  unknown

The margin of error denoted E is the error in the difference between the sample proportion and population proportion p. It is found by  $E = z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$  the critical value by the standard deviation of the sample proportion.

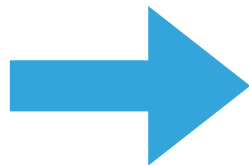
The formula to calculate the margin of error is .

$$n = \frac{[z_{\alpha/2}]^2 \hat{p}\hat{q}}{E^2} \text{ when } \hat{p} \text{ is known and } n = \frac{[z_{\alpha/2}]^2 0.25}{E^2} \text{ when } \hat{p} \text{ is unknown}$$

Confidence level 99%

$$z=2.576$$

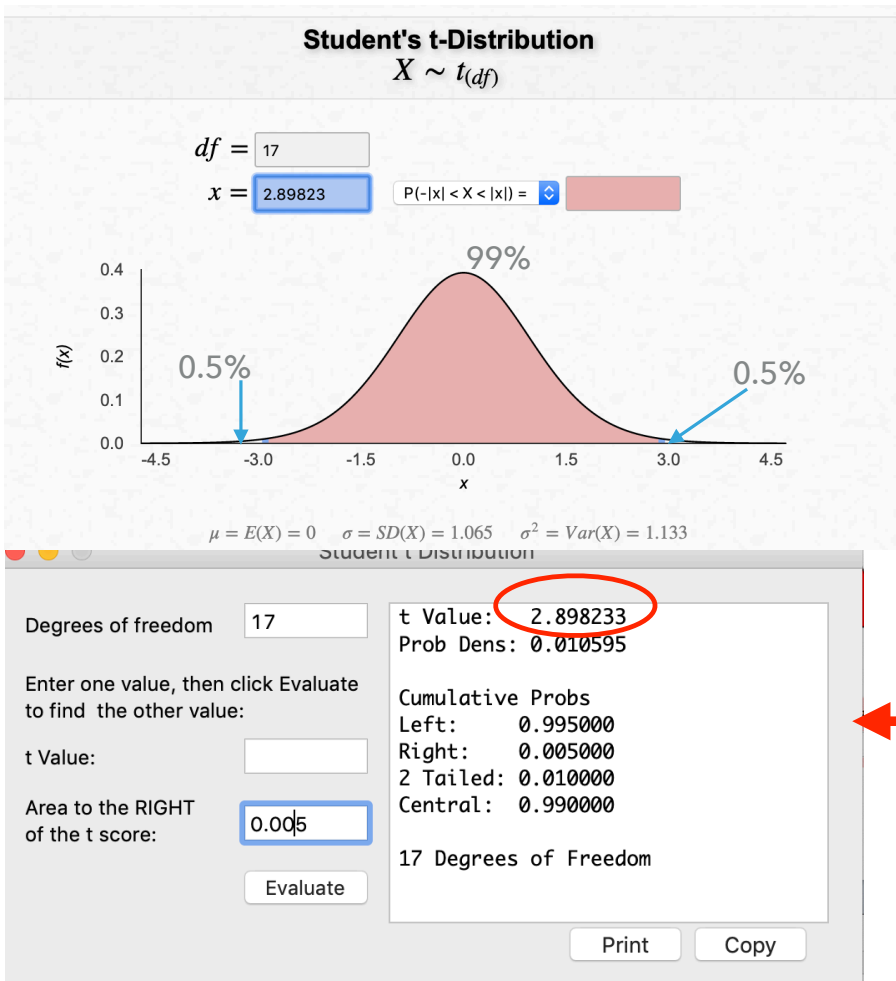
$$E=0.028$$



$$n = \frac{2.576^2 \times 0.25}{E^2}$$

## QUIZ 2

6) Find the appropriate critical value for the following: 99% confidence level ;  $n = 18$ ;  $\sigma$  is unknown; population appears to be normally distributed.



**Cumulative area**

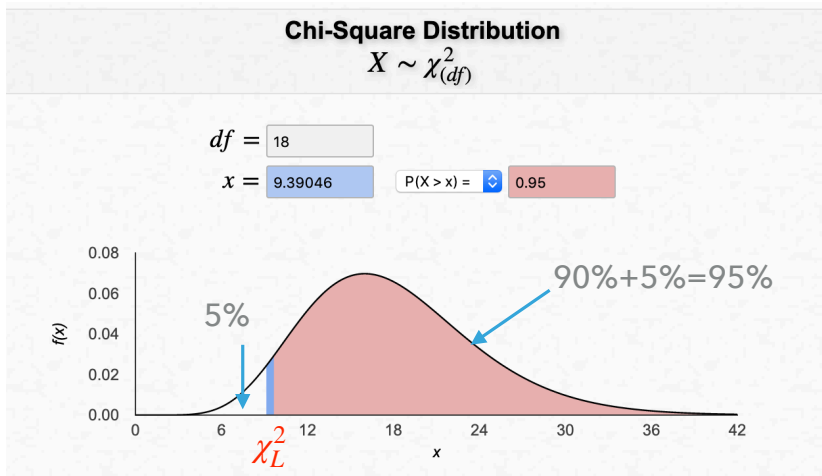
**from the right of the t score**

**=0.5%=0.005**

**$df = n - 1 = 18 - 1 = 17$**

## QUIZ 2

7) Find the critical value  $\chi_L^2$  corresponding to a sample size of 19 and a confidence level of 90 percent if the test is two-tailed.



The figure shows a Chi-Square Distribution calculator interface. The degrees of freedom is 18. The area to the right of the chi-sq value is 0.95. The critical value is 9.390446. The probability density is 0.0267664. The cumulative probabilities are Left: 0.050000 and Right: 0.950000. The degrees of freedom is 18. The interface includes buttons for Evaluate, Print, and Copy.

Area to the right  
of the chi-sq value  
=5%+90%  
=95%  
=0.95  
df=n-1=19-1=18

## QUIZ 2

---

9) A survey reported that 40% of people save loose change to pay bills. The margin of error was given as 3.1 percentage points. Which confidence interval corresponds to the given information?

$$\text{Margin of error} = t\left(\frac{s}{\sqrt{n}}\right) = 0.031$$

$$\text{Confidence Interval } p = \bar{p} \pm t\left(\frac{s}{\sqrt{n}}\right) = 0.4 \pm 0.031$$

$$0.4 - 0.031 < p < 0.4 + 0.031$$

# INFERENCEAL STATISTICS AND ANALYTICS

---

- ▶ ***Module 2 Confidence interval***
- ▶ ***Module 2 homework***
- ▶ ***Quiz 2***
- ▶ ***Course project phase 2***
- ▶ Module 2 Live classroom Grading
- ▶ Summary

## COURSE PROJECT PHASE 2

---

- ▶ Construct 95% confidence intervals for the population mean. Assume that your data is normally distributed and **the population standard deviation is unknown**.
  - ▶ Step 1: Find  $n, \bar{x}, s$ 
    - ▶ option 1(mean=62339,n=350, sample standard deviation s=19360)
    - ▶ option 2(mean=62.54,n=65, sample standard deviation s=9.257)
  - ▶ Step 2: Find the critical value t for a 95% confidence level. Use the **t-distribution** table to find the critical value. Alpha=0.05 in two tails with df=n-1 degrees of freedom. For example:
    - ▶ Option 1 a 95% confidence interval with 349 degrees of freedom, t=1.967
    - ▶ Option 2 a 95% confidence interval with 64 degrees of freedom, t=1.998
  - ▶ Step 3: Find the margin of error, E.
  - ▶ Step 4: Construct the confidence interval

$$E = t \frac{s}{\sqrt{n}}$$

$$\bar{x} - E < \mu < \bar{x} + E$$



## COURSE PROJECT PHASE 2

- ▶ Construct 95% confidence intervals for the population mean. Assume that your data is normally distributed and **the population standard deviation is unknown**.
- ▶ Use Statisk to find the margin errors and confidence intervals
- ▶ option 1(mean=62339,n=350, sample standard deviation s=19360)  
option 2(mean=62.54,n=65, sample standard deviation s=9.257)

Confidence Interval: Me

Confidence Level:

Sample Size, n:

Sample Mean:

Sample Standard Deviation, s:

Population Standard Deviation:   
(if known)

Confidence Interval: M

Confidence Level:

Sample Size, n:

Sample Mean:

Sample Standard Deviation, s:

Population Standard Deviation:   
(if known)

## COURSE PROJECT PHASE 2

---

- ▶ Construct 99% confidence intervals for the population mean. Assume that your data is normally distributed and **the population standard deviation is unknown**.
  - ▶ Step 1: Find  $n, \bar{x}, s$ 
    - ▶ option 1(mean=62339,n=350, sample standard deviation s=19360)
    - ▶ option 2(mean=62.54,n=65, sample standard deviation s=9.257)
  - ▶ Step 2: Find the critical value t for a 99% confidence level. Use the **t-distribution** table to find the critical value. Alpha=0.01 in two tails with df=n-1 degrees of freedom. For example:
    - Option 1 a 99% confidence interval with 349 degrees of freedom, t=2.590
    - Option 2 a 99% confidence interval with 64 degrees of freedom, t=2.655
  - ▶ Step 3: Find the margin of error, E.
  - ▶ Step 4: Construct the confidence interval

$$E = t \frac{s}{\sqrt{n}}$$

$$\bar{x} - E < \mu < \bar{x} + E$$

## COURSE PROJECT PHASE 2

- ▶ Construct 99% confidence intervals for the population mean. Assume that your data is normally distributed and **the population standard deviation is unknown**.
- ▶ Use Statisk to find the margin errors and confidence intervals
- ▶ option 1(mean=62339,n=350, sample standard deviation s=19360)  
option 2(mean=62.54,n=65, sample standard deviation s=9.257)

Confidence Interval: Mean

Confidence Level:

Sample Size, n:

Sample Mean:

Sample Standard Deviation, s:

Population Standard Deviation:

(if known)

Confidence Interval: Mean

Confidence Level:

Sample Size, n:

Sample Mean:

Sample Standard Deviation, s:

Population Standard Deviation:

(if known)

# INFERENCEAL STATISTICS AND ANALYTICS

---

- ▶ ***Module 2 Confidence interval***
- ▶ ***Module 2 homework***
- ▶ ***Quiz 2***
- ▶ ***Course project phase 2***
- ▶ ***Module 2 Live classroom Grading***
- ▶ Summary

## MODULE 2 LIVE CLASSROOM GRADING

---

The live classroom session archive will be added after the session ends. The following is how you will receive your points for the module 2 live classroom session:

Question: **What distribution will we use in the course project phase 2? the Standard Normal distribution or Student's t distribution?**

Please go to [module 2 live classroom](#), enter your response.

***You have until midnight CST on Sunday to confirm that you have viewed the live classroom session archive.***

# INFERENCEAL STATISTICS AND ANALYTICS

---

- ▶ ***Module 2 Confidence interval***
- ▶ ***Module 2 homework***
- ▶ ***Quiz 2***
- ▶ ***Course project phase 2***
- ▶ ***Module 2 Live classroom Grading***
- ▶ ***Summary***

## SUMMARY

---

- ▶ When estimating a population mean, the population standard deviation is **known**. In this case, we construct the confidence interval using the **standard normal distribution**.
- ▶ When estimating a population mean, the population standard deviation is **unknown**. In this case, we construct the confidence interval using the **Student's t distribution** instead of a standard normal distribution.
- ▶ Module 2 Question: ***What distribution will we use in the course project phase 2? The Standard Normal Distribution or Student's t distribution?***