EDP308: STATISTICAL LITERACY

The University of Texas at Austin, Fall 2020

RAZ: Rebecca A. Zárate, MA

Overview

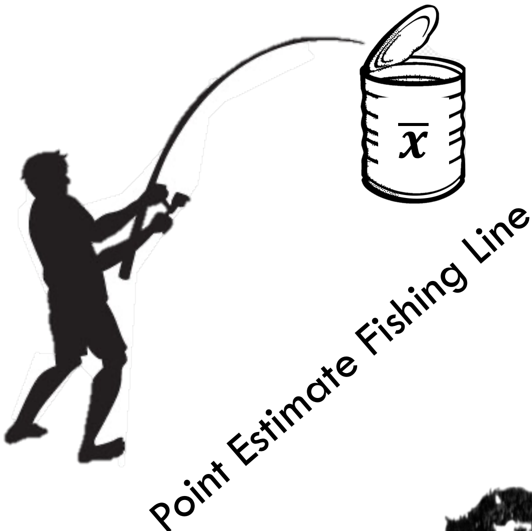
- Calculating a Confidence Interval
  - Point Estimate
  - Level of Confidence (z-critical)
  - Standard Error
  - Margin of Error
- Factors that Influence a Confidence Interval
  - Sample Size
  - Variance
  - Level of Desired Confidence
- Confidence Intervals in R

CI Calculations

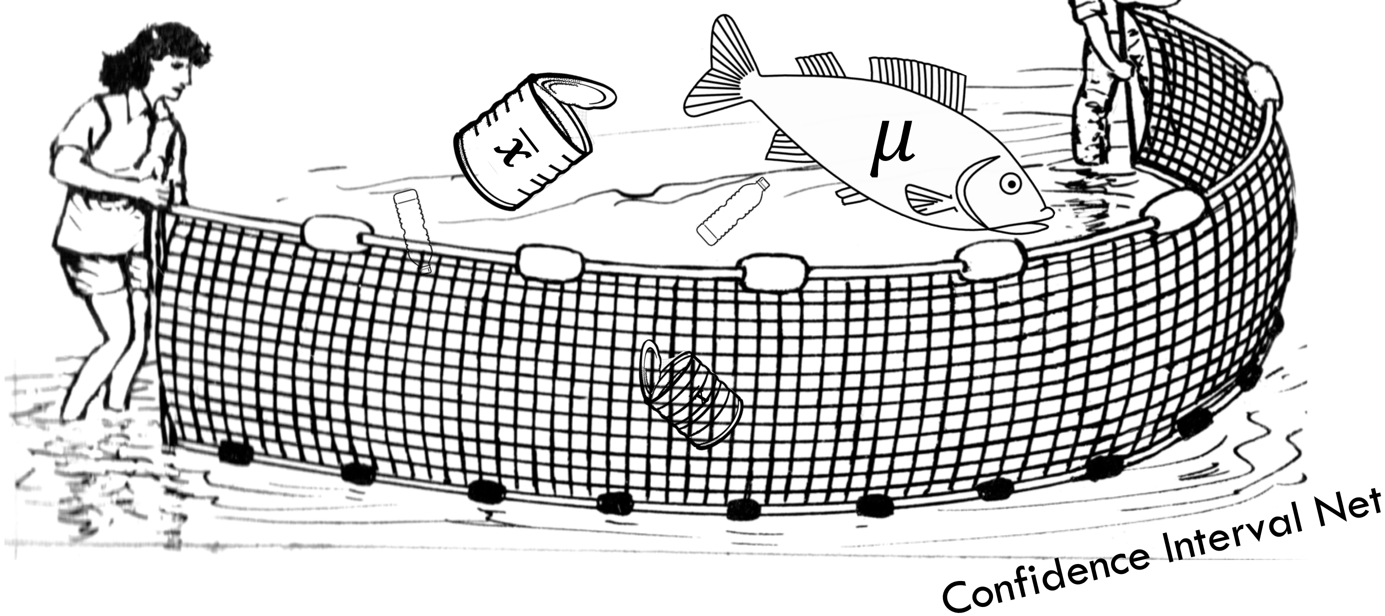
Confidence Intervals

- Why do we calculate confidence intervals?
- Rarely is a single point estimate ($\bar{x}$) the exact parameter value we are looking for (μ)
  - More likely that it is a little off, maybe a little too high or a little too low
- To increase the chances we capture the truth (μ), we put a little wiggle room, a little cushion around our point estimate
  - Ex. $\bar{x} = 65$" vs. 95% CI [64.5", 66.5"]

Two Ways to Fish for Truth



If you are trying to capture the Truth Tuna, casting a wider net increases your chances.



Calculations for Confidence Interval

Margin of Error:
The amount of error based on SE and a desired level of confidence in the original number scale

$$CI = \bar{x} \pm \underbrace{z_{critical} * \frac{\sigma}{\sqrt{n}}}$$

Point Estimate:
Statistic (ex. sample mean) in the original number scale

z-critical:
The level of confidence I want (ex. 95%, z=1.96) in its corresponding z-score

Standard Error (SE):
Standard deviation of the sampling distribution

Calculations for Confidence Interval

Margin of Error:
The amount of error based on SE and a desired level of confidence in the original number scale

$$CI = Point\ Estimate \pm z_{critical} * Standard\ Error$$

Point Estimate:
Statistic (ex. sample mean value) in the original number scale

z-value:
The level of confidence I want (ex. 95%, z=1.96) in its corresponding z-score

Standard Error (SE):
Standard deviation of the sampling distribution

Try it.

- I gave out a quiz to the stats literacy course. The standard deviation for the entire class was σ = 8. Now, we take a sample, n = 9, from our stats literacy population and calculate a mean, $\bar{x}$ = 81. Calculate a 95% confidence internal for this sample statistic.

$\bar{x} = 81$

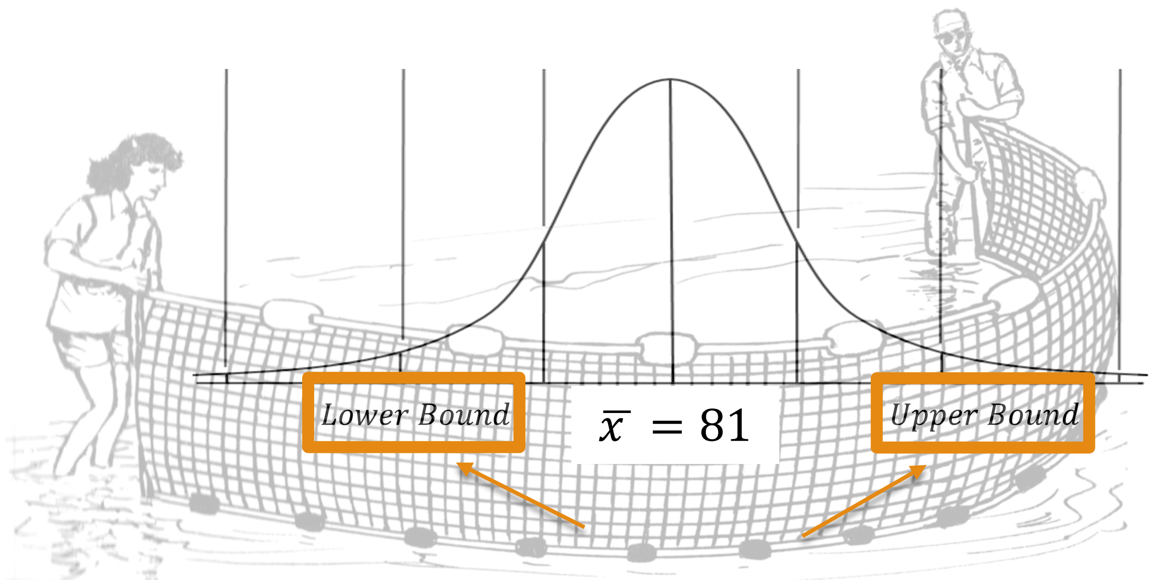
$\sigma = 8$

$n = 9$

$\sigma_{\bar{x}}$ (SE)=

MOE =

CI =



Find the lower and higher values that true μ might be

Try it.

$$CI = Point\ Estimate\ \pm z_{critical} * Standard\ Error$$

- I gave out a quiz to the stats literacy course. The standard deviation for the entire class was σ = 8. Now, we take a sample, n = 9, from our stats literacy population and calculate a mean, $\bar{x}$ = 81. Calculate a 95% confidence internal for this sample statistic.

$\bar{x} = 81$

$\sigma = 8$

$n = 9$

$\sigma_{\bar{x}}$ (SE)$= 2.67$

MOE =

CI =

$$\frac{\sigma}{\sqrt{n}} = \sigma_{\bar{x}} \qquad \frac{8}{\sqrt{9}} = 2.67$$

Standard Error ($\sigma_{\bar{x}}$)

This is like the standard deviation of the sampling distribution.

Try it.

$$CI = Point\ Estimate \pm z_{critical} * Standard\ Error$$

- I gave out a quiz to the stats literacy course. The standard deviation for the entire class was σ = 8. Now, we take a sample, n = 9, from our stats literacy population and calculate a mean, $\bar{x} = 81$. Calculate a 95% confidence internal for this sample statistic.

$\bar{x} = 81$

$\sigma = 8$

$n = 9$

$\sigma_{\bar{x}}$ (SE) $= 2.67$

$\text{MOE} = 5.23$

$\text{CI} =$

$$z_{critical} * \sigma_{\bar{x}} = Margin\ of\ Error$$

$$1.96 * 2.67 = 5.23$$

For 95% confidence, we need a z critical of 1.96

Try it.

$$CI = Point\ Estimate\ \pm z_{critical} * Standard\ Error$$

- I gave out a quiz to the stats literacy course. The standard deviation for the entire class was σ = 8. Now, we take a sample, n = 9, from our stats literacy population and calculate a mean, $\bar{x}$ = 81. Calculate a 95% confidence internal for this sample statistic.

$$\bar{x} = 81$$

$$\sigma = 8$$

$$n = 9$$

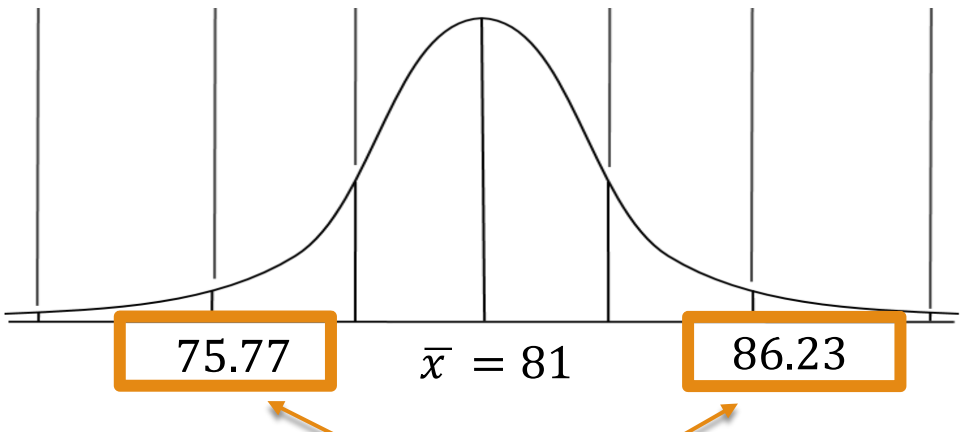
$$\sigma_{\bar{x}}\ (SE) = 2.67$$

$$MOE = 5.23$$

$$CI = [75.77, 86.23]$$

$$[75.77, 86.23] = 81 \pm 5.23$$

$$CI = Point\ Estimate\ \pm Margin\ of\ Error$$



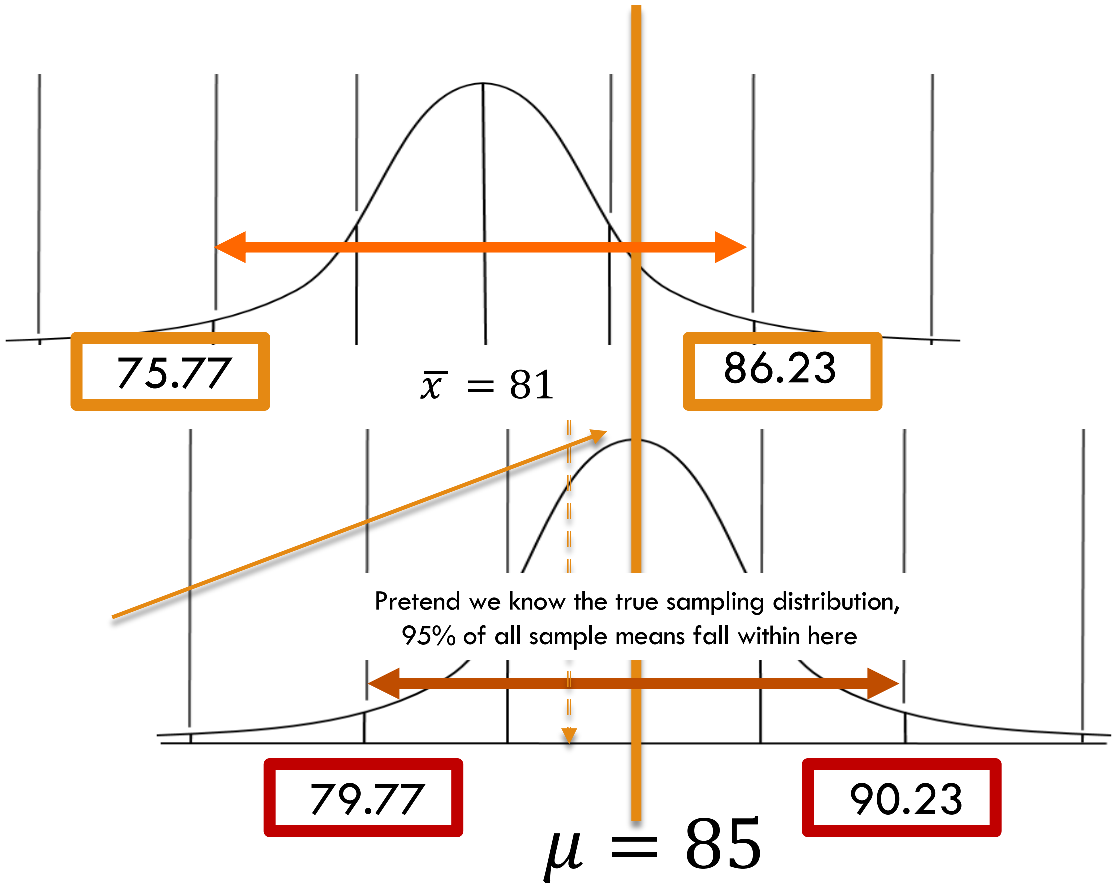
Find the lower and higher values that true μ might be

Try it.

$$CI = Point\ Estimate\ \pm z_{critical} * Standard\ Error$$

Now I reveal that the true mean is 85

Did we capture the truth in our CI?



Try it.

$$CI = Point\ Estimate\ \pm z_{critical} * Standard\ Error$$

Did our confidence interval contain the true population parameter μ?

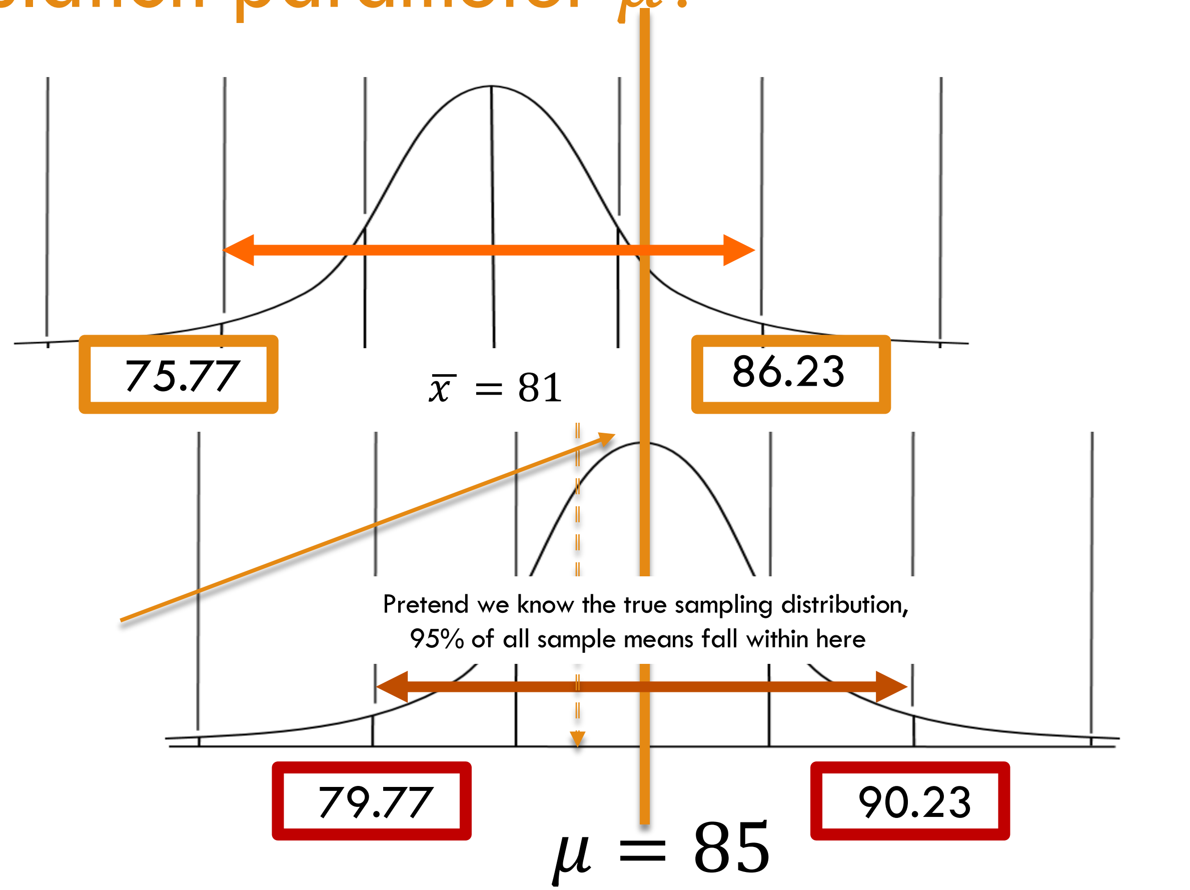
$$\mu = 85$$

Truth

$$[75.77, 86.23]$$

Our Estimate

Yes, 85 is in our CI!
Hooray!

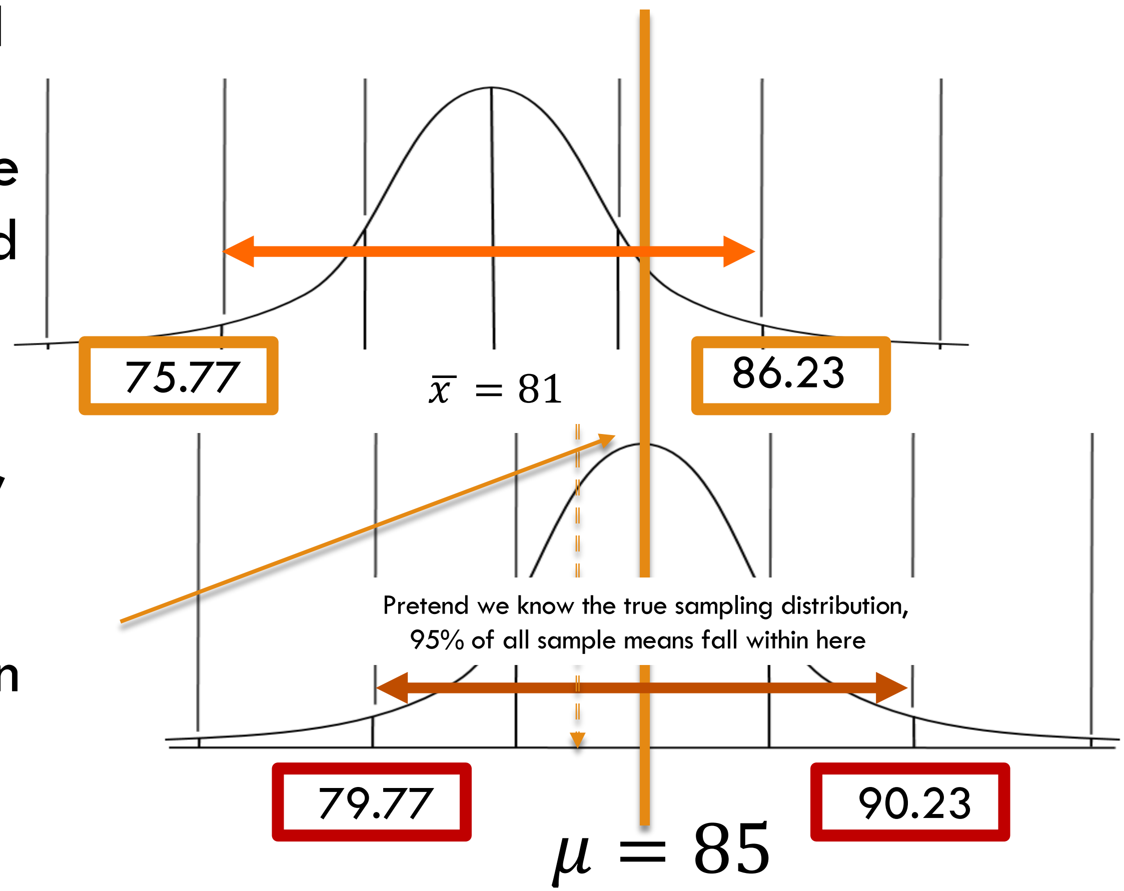


Try it.

95% CI [75.77, 86.23]

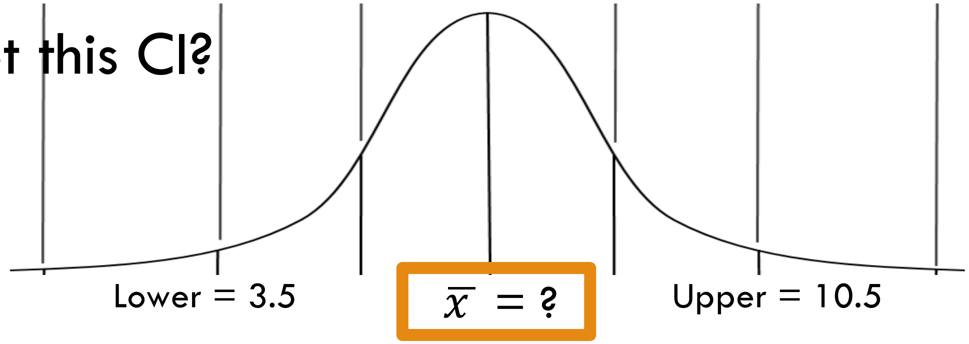
$$CI = Point\ Estimate \pm z_{critical} * Standard\ Error$$

- How do we interpret this confidence interval?
  - If we repeated the sample again and again, 95 out of 100 times, the true μ will be captured in our confidence interval.
  - Said another way, we are 95% confident the true μ value is between [75.77, 86.23].



Try it.

- You find a research article that reports a 95% confidence interval for the number of hours college students sleep based on a sample of 9 students with a known σ. They reported a 95% CI of [3.5, 10.5]
  - What is z?
  - What is the point estimate?
  - What is the Margin of Error (MOE)?
  - What is the standard error (SE)?
  - How would you interpret this CI?



Try it.

- You find a research article that reports a 95% confidence interval for the number of hours college students sleep based on a sample of 9 students. They reported a 95% CI of [3.5, 10.5]
  - What is z?
- Because σ is known, we use the unit normal z-table and we see the critical z value associated with 95% confidence is 1.96

$$z = 1.96$$

Try it.

$$CI = Point\ Estimate\ \pm z_{critical} * Standard\ Error$$

- You find a research article that reports a 95% confidence interval for the number of hours college students sleep based on a sample of 9 students. They reported a 95% CI of [3.5, 10.5]
  - What is z?
    - $z = 1.96$
  - What is the point estimate?
    - Looking for the middle of the CI

Lower Bound Upper Bound

$$\frac{3.5 + 10.5}{2} = 7$$

Point Estimate, $\bar{x} = 7$

$$[3.5 \xleftrightarrow{-MOE} 7 \xleftrightarrow{+MOE} 10.5]$$

Point Estimate

Lower Bound Upper Bound

Try it.

$$CI = Point\ Estimate\ \pm z_{critical}\ * Standard\ Error$$

- You find a research article that reports a 95% confidence interval for the number of hours college students sleep based on a sample of 9 students. They reported a 95% CI of [3.5, 10.5]
  - What is z?
    - $z = 1.96$
  - What is the point estimate?
    - $\bar{x} = 7$
  - What is the Margin of Error (MOE)?
    - Solve either equation, both give you MOE

$$\bar{x} - Lower\ Bound = MOE$$

$$Upper\ Bound - \bar{x} = MOE$$

$$MOE = 3.5$$

7 – MOE = 3.5

(rearranged) 7-3.5 = 3.5

10.5 – MOE = 3.5

(rearranged) 10.5 – 7 = 3.5

Try it.

$$CI = Point\ Estimate\ \pm z_{critical} * Standard\ Error$$

- You find a research article that reports a 95% confidence interval for the number of hours college students sleep based on a sample of 9 students. They reported a 95% CI of [3.5, 10.5]
  - What is z?
    - $z = 1.96$
  - What is the point estimate?
    - $\bar{x} = 7$
  - What is the Margin of Error (MOE)?
    - $MOE = 3.5$
  - What is the standard error (SE)?
    - $\sigma_{\bar{x}} = 1.79$

$$MOE = z * \sigma_{\bar{x}}$$

Rearrange

$$\frac{MOE}{z} = \sigma_{\bar{x}}$$

$$\frac{3.5}{1.96} = 1.79$$

Try it.

$CI = Point\ Estimate\ \pm z_{critical} * Standard\ Error$

- You find a research article that reports a 95% confidence interval for the number of hours college students sleep based on a sample of 9 students. They reported a 95% CI of [3.5, 10.5]
  - What is z?
    - $z = 1.96$
  - What is the point estimate?
    - $\bar{x} = 7$
  - What is the Margin of Error (MOE)?
    - $MOE = 3.5$
  - What is the standard error (SE)?
    - $\sigma_{\bar{x}} = 1.79$

How would you interpret this CI?

$$[3.5, 10.5] = 7 \pm 1.96 * 1.79$$

Try it.

$$CI = Point\ Estimate\ \pm z_{critical} * Standard\ Error$$

- You find a research article that reports a 95% confidence interval for the number of hours college students sleep based on a sample of 9 students. They reported a 95% CI of [3.5, 10.5]
- How would you interpret this CI?
  - “If we take repeated samples of n = 9 and compute a 95% confidence interval each time, approximately 95% of the intervals would contain the true number of hours college students sleep.
  - “We are 95% confident that the true number of hours college students sleep is between 3.5 and 10.5 hours.”

Factors that Influence CI

Factors that Influence CI

$$CI = \bar{x} \pm \underbrace{z * \underbrace{\frac{\sigma}{\sqrt{n}}}_{\text{SE}}}_{\text{MOE}}$$

Based on the equation above, what factors influence how wide or narrow your CI is?

Factors that Influence CI

What influences how wide or narrow our CI is?

$$CI = \bar{x} \pm \underbrace{z * \underbrace{\frac{\sigma}{\sqrt{n}}}_{SE}}_{MOE}$$

Narrower Confidence Interval

$[13 \xleftrightarrow{-MOE} 15 \xleftrightarrow{+MOE} 17]$

Lower Bound | Point Estimate | Upper Bound

Wider Confidence Interval

$[5 \xleftrightarrow{-MOE} 15 \xleftrightarrow{+MOE} 25]$

Lower Bound | Point Estimate | Upper Bound

Factors that Influence CI

The Margin of Error is what widens or narrows your CI, but what affects your MOE?

$$CI = \bar{x} \pm \underbrace{z * \overbrace{\frac{\sigma}{\sqrt{n}}}^{SE}}_{MOE}$$

Narrower Confidence Interval

$[13 \xleftrightarrow{-MOE} 15 \xleftrightarrow{+MOE} 17]$

Lower Bound — Point Estimate — Upper Bound

Wider Confidence Interval

$[5 \xleftrightarrow{-MOE} 15 \xleftrightarrow{+MOE} 25]$

Lower Bound — Point Estimate — Upper Bound

Factors that Influence CI

- The Margin of Error is affected by:
  - Sample size (n)
  - The variability in the POPULATION (σ)
  - The desired level of confidence (ex. 95%)
    - The z-critical value (ex. 1.96)

$$CI = \bar{x} \pm \underbrace{z * \underbrace{\frac{\sigma}{\sqrt{n}}}_{\text{SE}}}_{\text{MOE}}$$

Sample Size (n)

- As sample size increases:
  - What happens to the standard error?
  - What happens to the confidence interval (wider or narrower)?
  - What happens to the MOE?

$$CI = \bar{x} \pm \underbrace{z * \overbrace{\frac{\sigma}{\sqrt{n}}}^{SE}}_{MOE}$$

Sample Size (n)

$$CI = \bar{x} \pm \underbrace{z * \overbrace{\frac{\sigma}{\sqrt{n}}}^{SE}}_{MOE}$$

- As sample size (n) increases
  - The larger the sample size, the smaller the standard error
  - MOE decreases as n increase
  - The more narrow your confidence interval

$\uparrow$ **n** then $\downarrow \sigma_{\bar{x}}$ $\quad \downarrow MOE$ $\quad \downarrow Width$

Inverse Relationships

As n goes up…

…these go down.

Height Example With Different N's

- Are you more likely to miss the true μ with a small sample size or a large sample size?
- You are much more likely to be "off" if you have a small sample size… So to make sure you "catch the truth" you need to make your more or less range (the confidence interval) wider with small sample sizes…
  - 95% CI with **sample size of 36**, [65", 67"] vs.
  - 95% CI with **sample size of 4**, [63", 69"]
- Notice how the interval gets WIDER with a smaller sample size… There's more uncertainty with small sample sizes.

Variability (σ)

- As population variability (standard deviation) increases:
  - What happens to the standard error?
  - What happens to the confidence interval (wider or narrower)?
  - What happens to the MOE?

$$CI = \bar{x} \pm \underbrace{z * \overbrace{\frac{\sigma}{\sqrt{n}}}^{\text{SE}}}_{\text{MOE}}$$

Variability (σ)

$$CI = \bar{x} \pm \underbrace{z * \overbrace{\frac{\sigma}{\sqrt{n}}}^{\text{SE}}}_{\text{MOE}}$$

- As the variability in the POPULATION (σ) increases:
  - The standard error increase
  - MOE increase as variability increases
  - The confidence interval widens

$\uparrow \sigma$ then $\uparrow \sigma_{\bar{x}}$ $\uparrow MOE$ $\uparrow Width$

Direct Relationships

As σ goes up… …these go up.

Variability (σ)

- As population variability (standard deviation) increases:
  - What happens to the standard error?
  - What happens to the confidence interval (wider or narrower)?
  - What happens to the MOE?

$$CI = \bar{x} \pm \underbrace{z * \overbrace{\frac{\sigma}{\sqrt{n}}}^{SE}}_{MOE}$$

Level of Confidence ($z_{critical}$)

- If you increase your desired level of confidence (ex. from 95% to 99%):
  - What happens to the standard error?
  - What happens to the confidence interval (wider or narrower)?
  - What happens to the MOE?

$$CI = \bar{x} \pm \underbrace{z * \overbrace{\frac{\sigma}{\sqrt{n}}}^{SE}}_{MOE}$$

90% confidence has z = 1.65
95% confidence has z = 1.96
99% confidence has z = 2.58

Level of Confidence ($z_{critical}$)

- If you increase your desired level of confidence (ex. from 95% to 99%):
  - The standard error is not affected
  - The margin of error increases with higher levels of confidence
  - The confidence interval widens as you increase your desired level of confidence

$\uparrow z_{critical}$	$\sigma_{\bar{x}}$	MOE	$Width$
	—	$\uparrow$	$\uparrow$
then	No Effect	Direct Relationships	
As $z_{critical}$ goes up…	…this stay the same, and…	…these go up.	

Precision vs. Confidence

- The more precise, the less confidence
- The more confident, the less precise
  - I am 99% confident that the average age of this class is 19 or somewhere between
    - z = 2.58, CI: [17, 21]

      - How precise is this estimate?
  - I am 95% confident that the average age of this class is 19 or somewhere between
    - z = 1.96, CI: [17.5, 20.5]

      - How precise is this estimate?
  - I am 80% confident that the average age of this class is 19 or somewhere between
    - z = 1.28, CI: [18, 20]

      - How precise is this estimate?

*Made up data and intervals for example.

Confidence Intervals

- Gives us a range of what the truth (μ) might be
- Help increase our chance of catching the truth
  - In case our initial point estimate ($\bar{x}$) is off
    - Which is usually is…
- Later, they will help us with inferential statistics…

Reality Check

- If we are trying to calculate some estimates about something unknown, how likely is it that we will know a populations σ (standard deviation)?
  - Not very likely…
    - So we can't use the normal z-distribution anymore…

But we do have another option…

Up Next…

- We get a bit more realistic and start using a different distribution where we don't need to know the standard deviation (σ) for the population
- Now we will start using…

The t-distribution

Confidence Intervals in R

Confidence Intervals in R

- Calculating a confidence interval when σ is known.
  - I gave out a quiz to the stats literacy course. The standard deviation for the entire class was $\sigma = 8$. Now, we take a sample, $n = 9$, from our stats literacy population and calculate a mean, $\bar{x} = 81$. Calculate a 95% confidence internal for this sample statistic.

````
```{r}
############################################
### Confidence Intervals for Known Sigma ###
############################################

average <- 81
sd <- 8
n <- 9

# Critical values depends on level of confidence
# The most common are 95% and 99%, choose the appropriate one

critical_z_95 <- 1.96 #for 95% CI
critical_z_99 <- 2.58 #for 99% CI

# Calculate the Margin of Error
moe <- critical_z_95 * sd/sqrt(n)

# Subtract the Margin of Error to the mean to get the lower bound
# Add the Margin of Error to the mean to get the upper bound

lower_bound <- average - moe
upper_bound <- average + moe
```
````