MODULE 5

CONFIDENCE INTERVALS

OUTLINE

OUTLINE

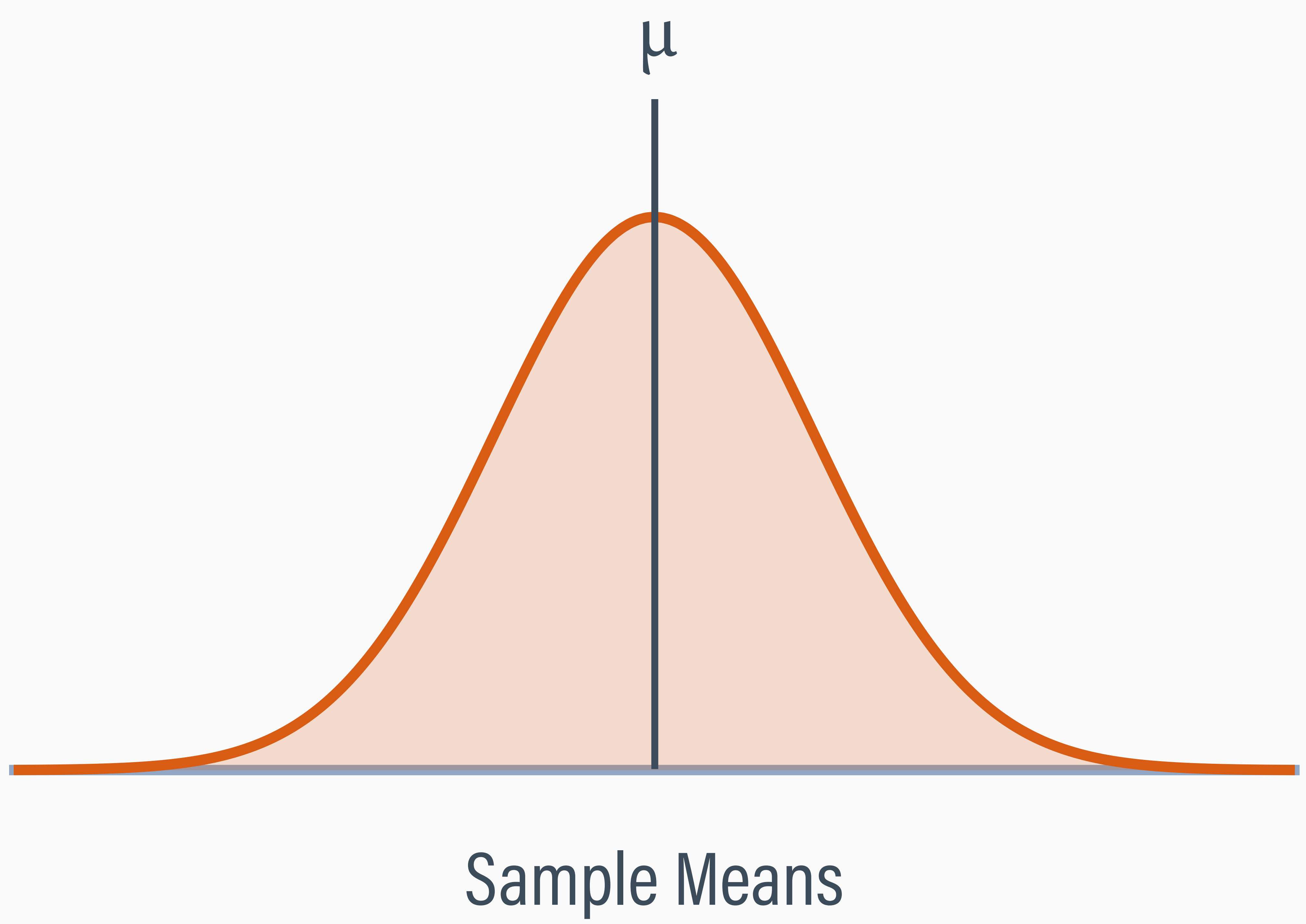
1. Quick review
2. Overview of confidence intervals
3. Constructing 95% confidence intervals
4. Properties of confidence intervals
5. Inference by eye
6. Study questions

QUICK REVIEW: SAMPLING ERROR

- The frequentist paradigm imagines a single population that spawns many hypothetical random samples of data (one parameter, many hypothetical estimates)
- The amount by which an estimate differs from the true population statistic is called sampling error
- Every hypothetical sample has a different amount of error

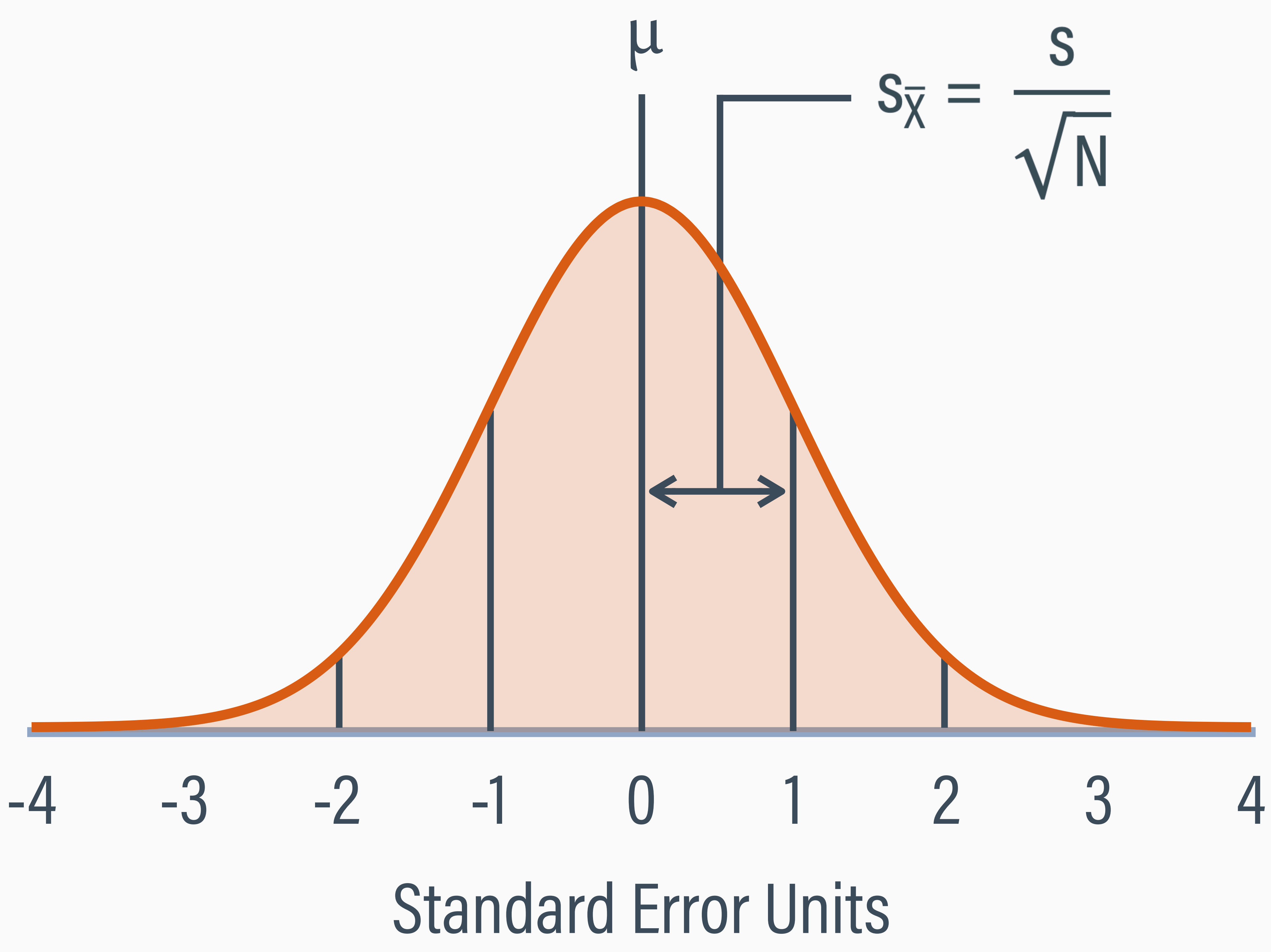
QUICK REVIEW: SAMPLING DISTRIBUTION

- The distribution of the estimates from many hypothetical samples is a sampling distribution
- With a large enough N, sample means follow a normal curve centered at the true mean
- Most estimates have small sampling errors, but a few have larger errors



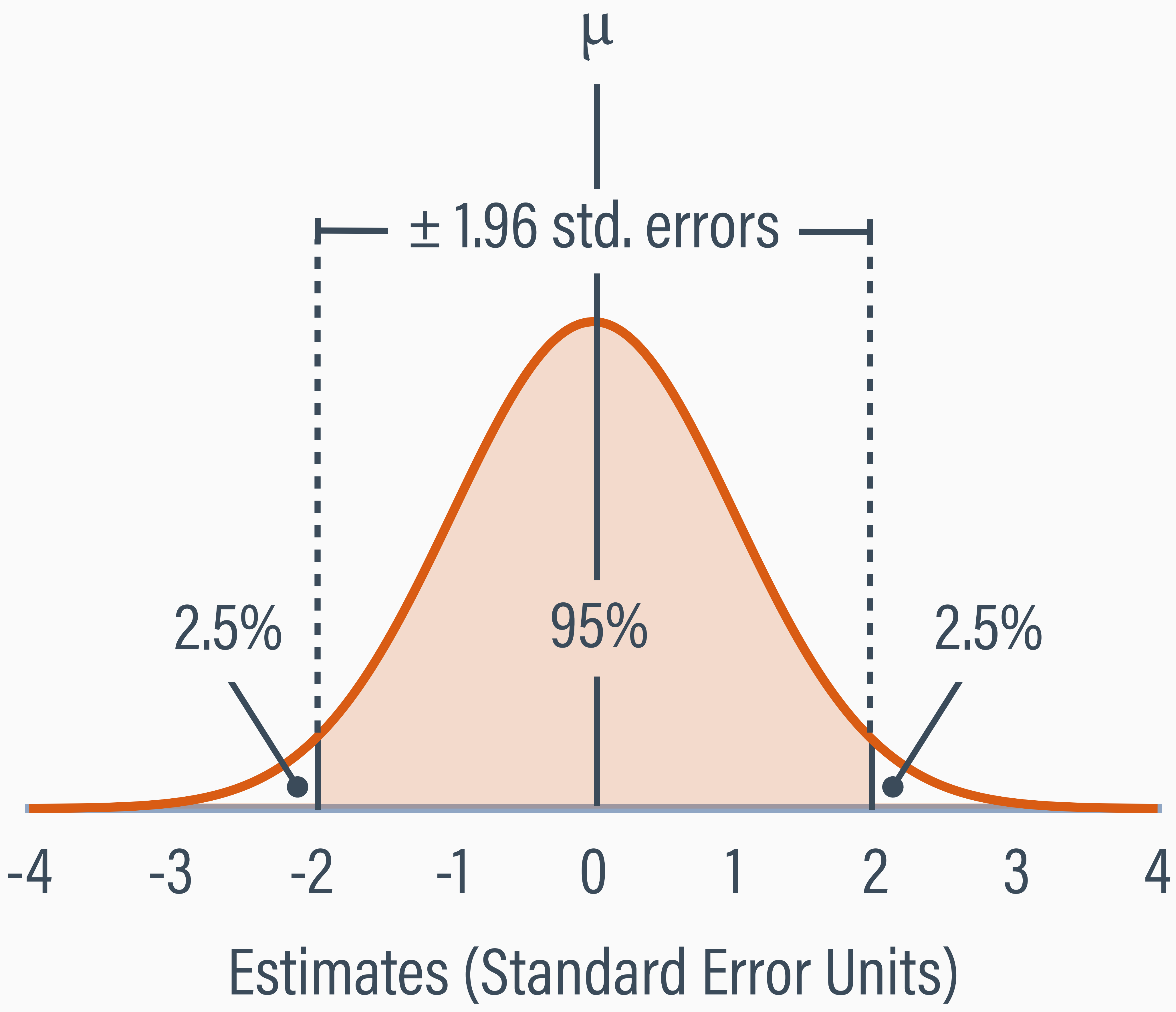
QUICK REVIEW: STANDARD ERROR

- The standard error is the average distance from a sample mean and the true mean
- $s_{\bar{X}}$ = standard deviation of the sample means
- The standard error is the average or expected amount of sampling error across many hypothetical samples



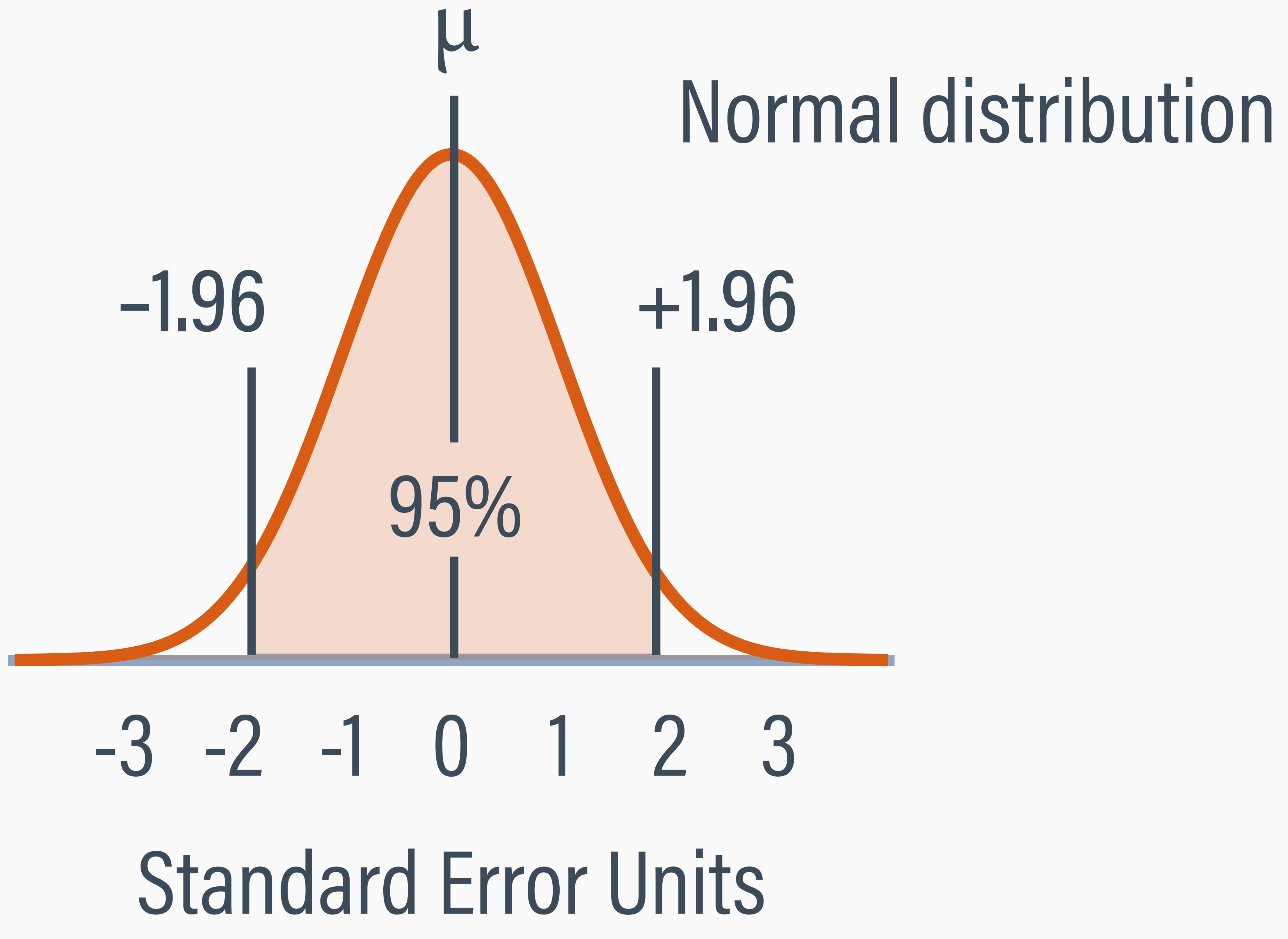
QUICK REVIEW: NORMAL CURVE RULE

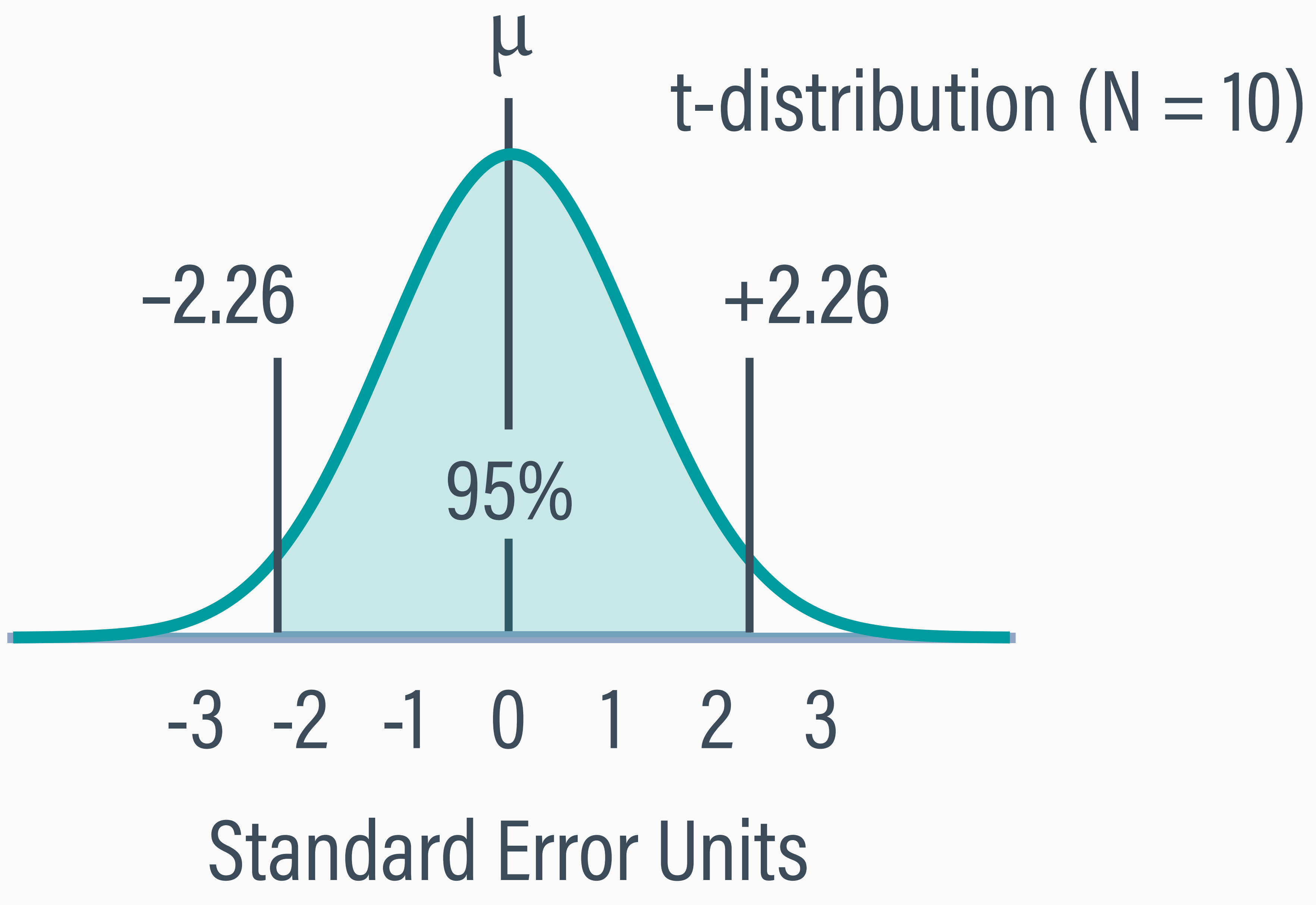
- The standard error is the standard deviation of many hypothetical sample means
- We can apply normal curve rules of thumb
- 95% of the means from large samples are within ± 1.96 standard errors of the true mean



QUICK REVIEW: T-DISTRIBUTION

- When using small samples, the normal curve is an inaccurate description of sampling error
- The t-distribution is a series of bell-shaped curves that stretch out (become more variable) as the N decreases
- Small samples are more likely to produce outlier estimates, and “stretching” the curve honors that





OUTLINE

1. Quick review
2. Overview of confidence intervals
3. Constructing 95% confidence intervals
4. Properties of confidence intervals
5. Small-sample adjustment with the t-distribution
6. Study questions

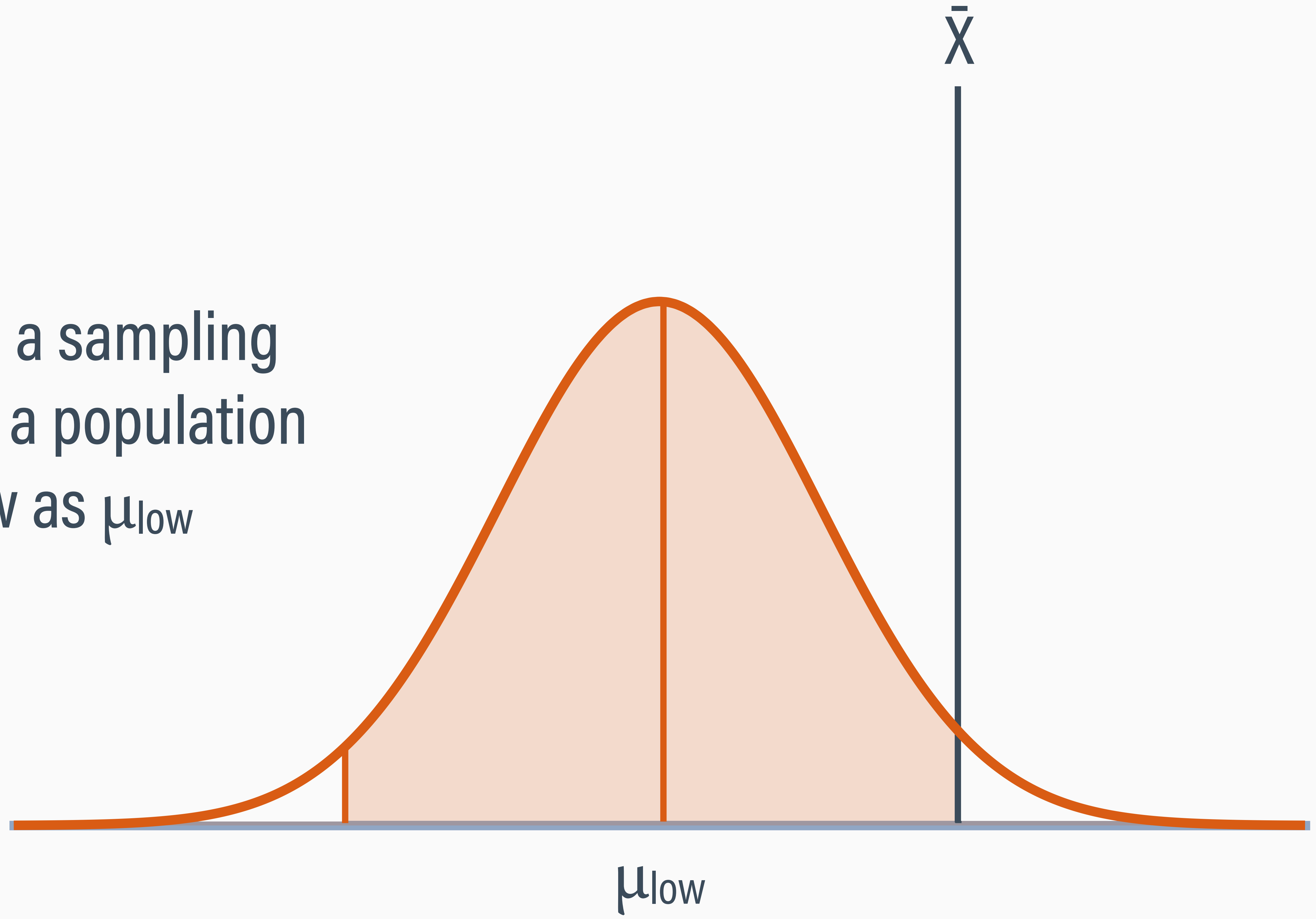
CONFIDENCE INTERVAL OVERVIEW

- We have a single sample of data that gives us an estimate $\bar{X}$
- The estimate could be too high ($\bar{X} > \mu$) or too low ($\bar{X} < \mu$)
- A confidence interval answers the following question: What are the lowest and highest values of μ that could have plausibly produced the $\bar{X}$ from our sample data?

LOWEST PLAUSIBLE POPULATION MEAN

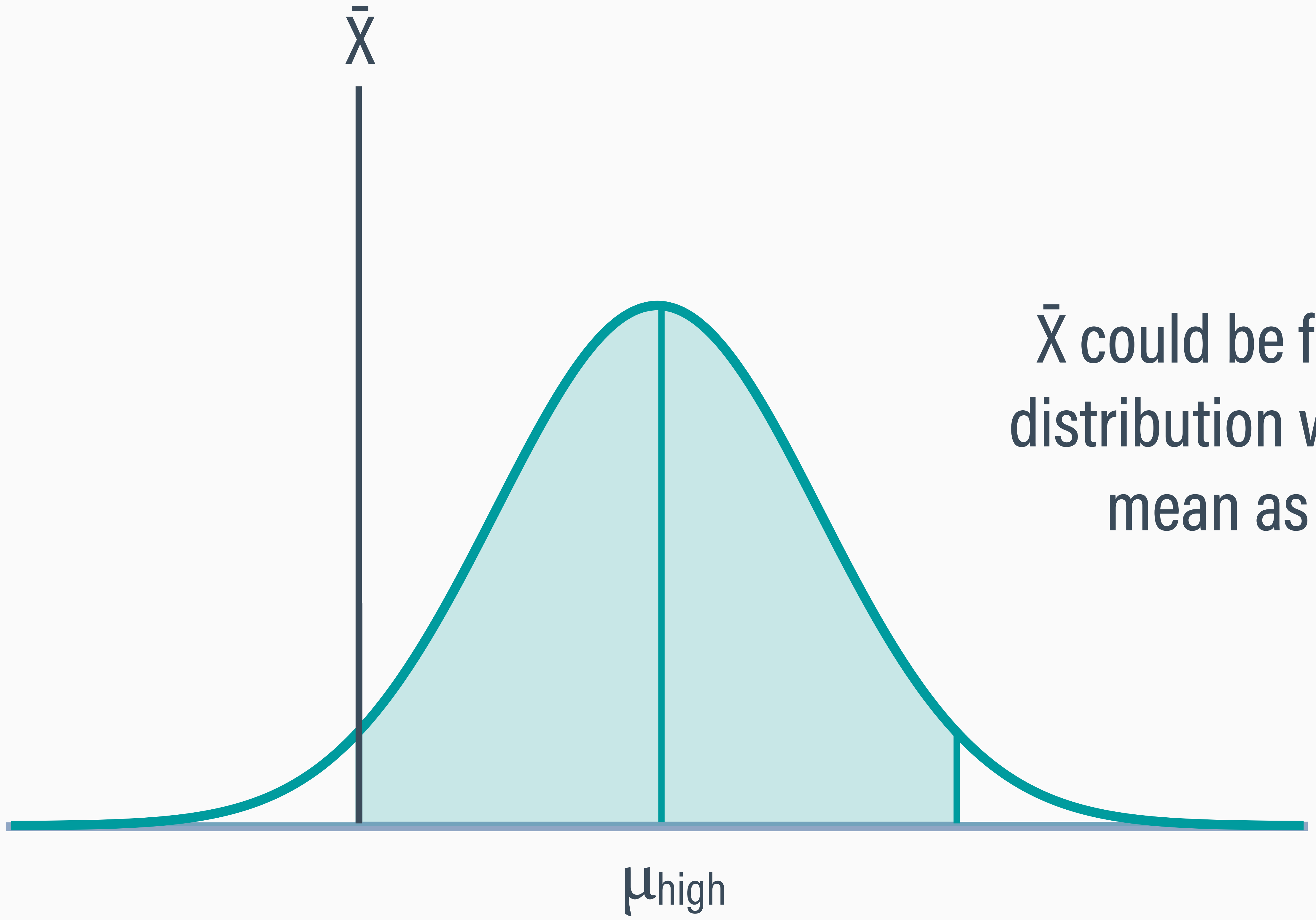
- If the estimate misses in the positive direction ($\bar{X} > \mu$), then $\bar{X}$ should lie far (but not too far) from the lowest plausible value of μ

$\bar{X}$ could be from a sampling distribution with a population mean as low as μ_{low}



HIGHEST PLAUSIBLE POPULATION MEAN

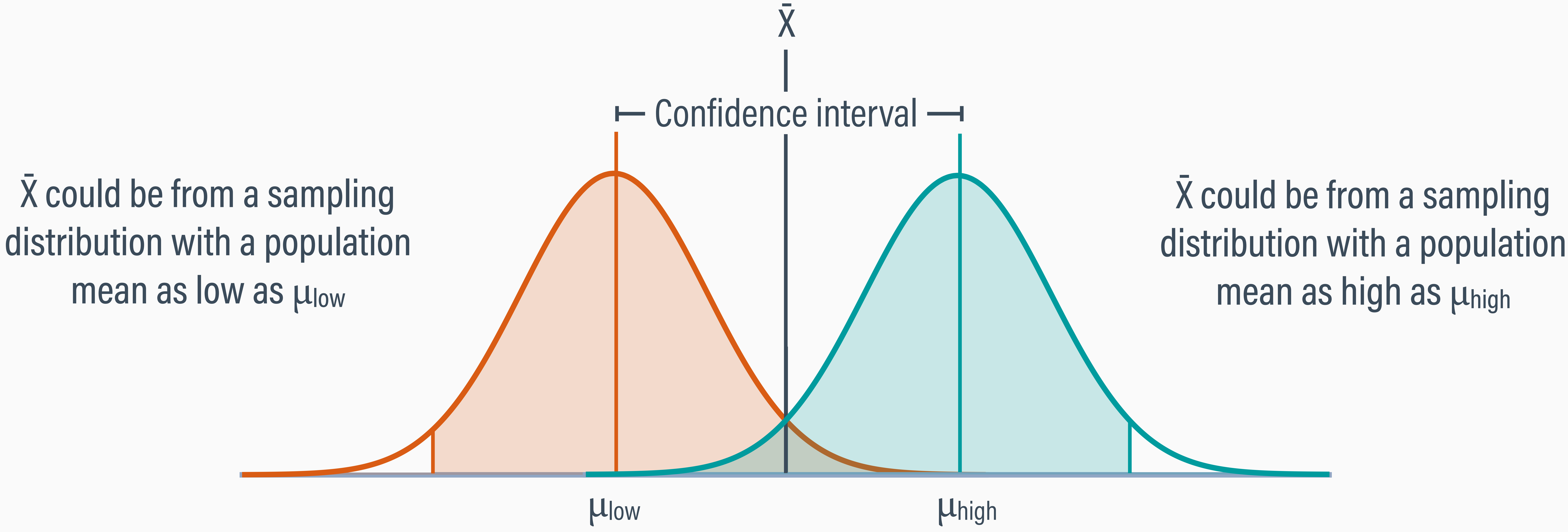
- If the estimate misses in the negative direction ($\bar{X} < \mu$), then $\bar{X}$ should lie far (but not too far) from the highest plausible value of μ



$\bar{X}$ could be from a sampling distribution with a population mean as high as μ_{high}

MOST EXTREME PARAMETER VALUES

- A confidence interval gives the two most extreme values of the population mean that could have reasonably produced the data



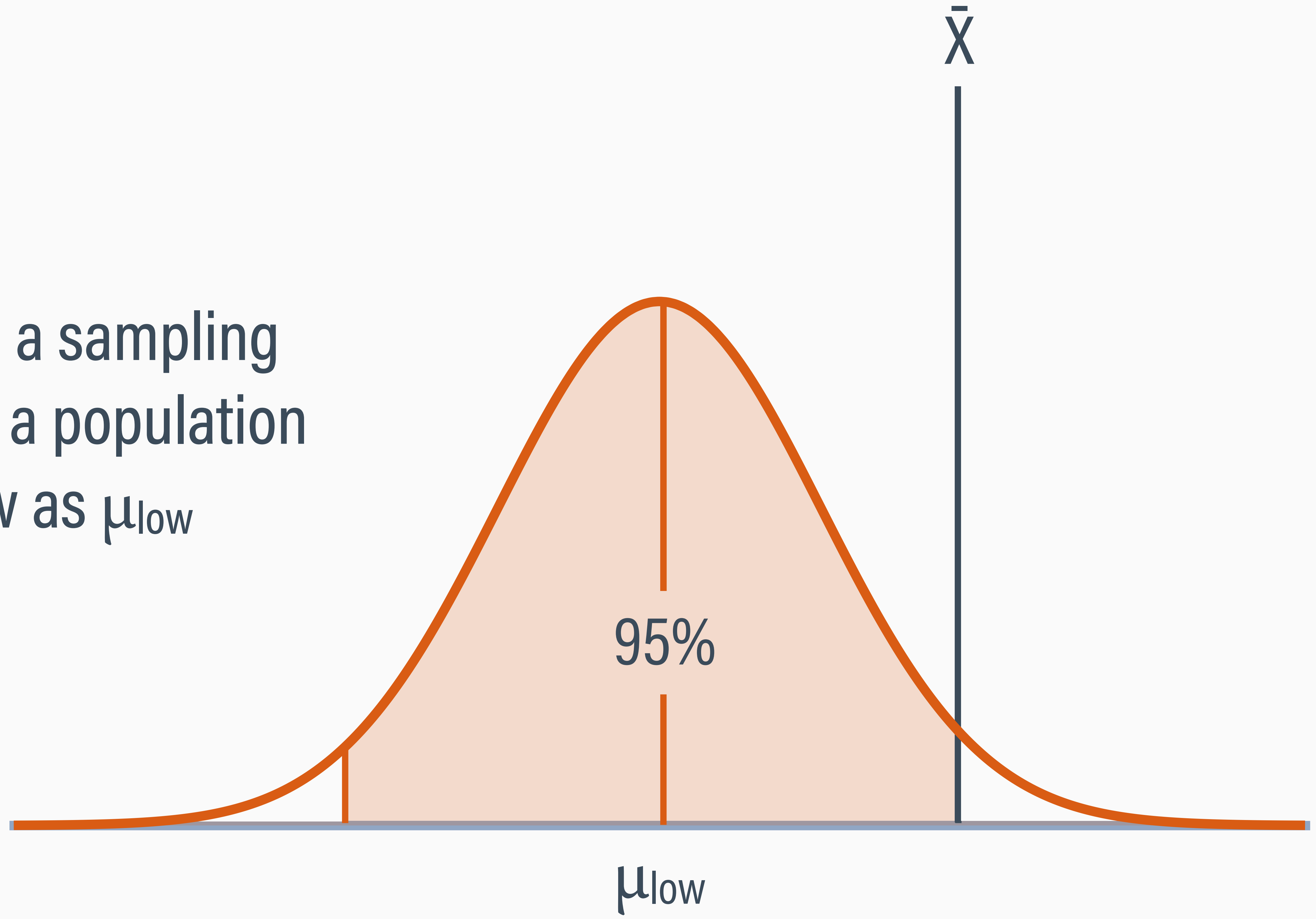
SETTING THE CONFIDENCE LEVEL

- An estimate should lie far (but not too far) from the lowest and highest plausible values of μ
- That distance is defined as a researcher-specified probability
- 95% confidence intervals predominate, but 90% and 99% are defensible and common in some applications

95% CONFIDENCE

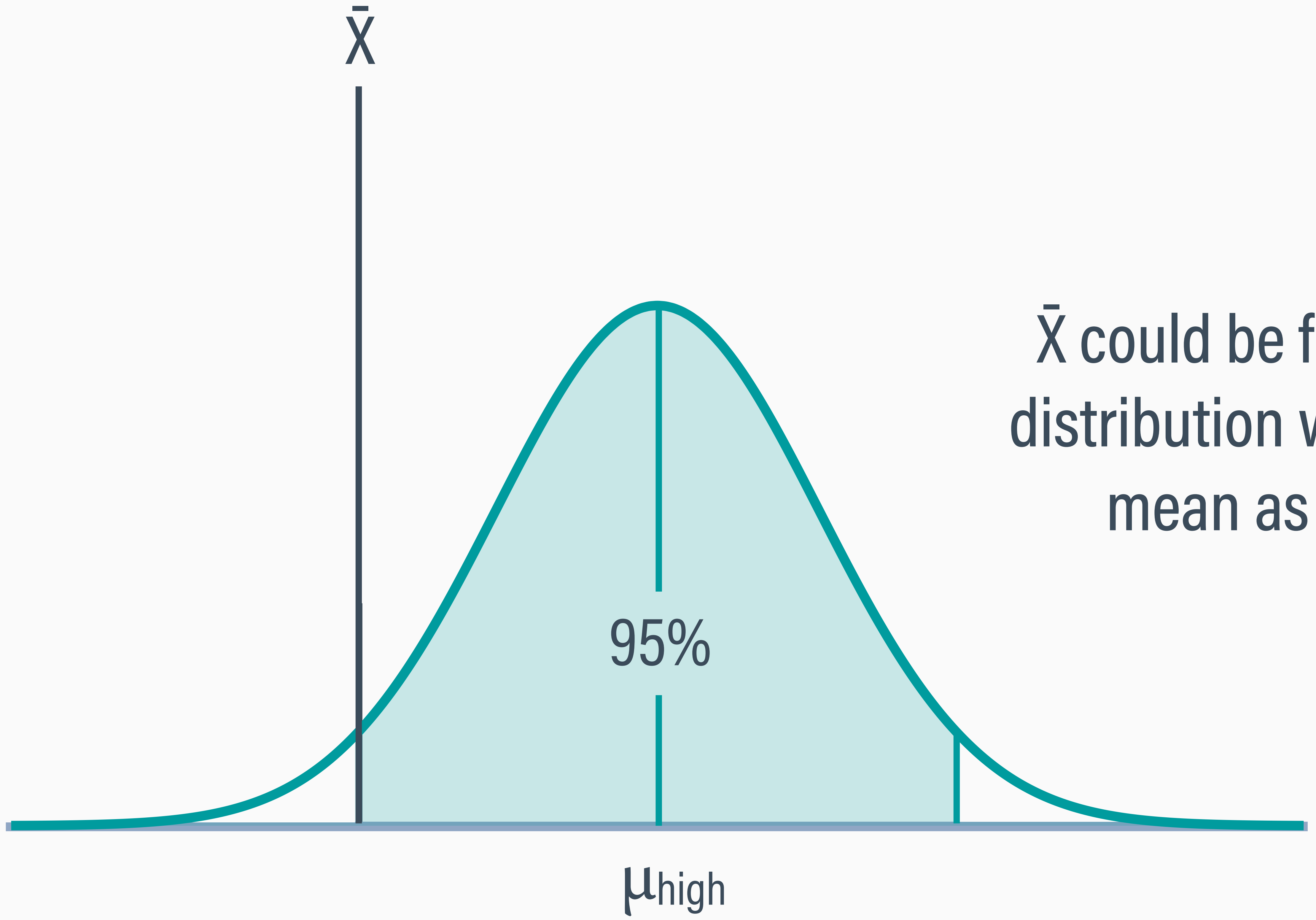
- 95% confidence implies that $\bar{X}$ lies on the upper limit of the middle 95% of all samples from a population with μ_{low} (a lower value of μ is unlikely but not impossible)

$\bar{X}$ could be from a sampling distribution with a population mean as low as μ_{low}



95% CONFIDENCE

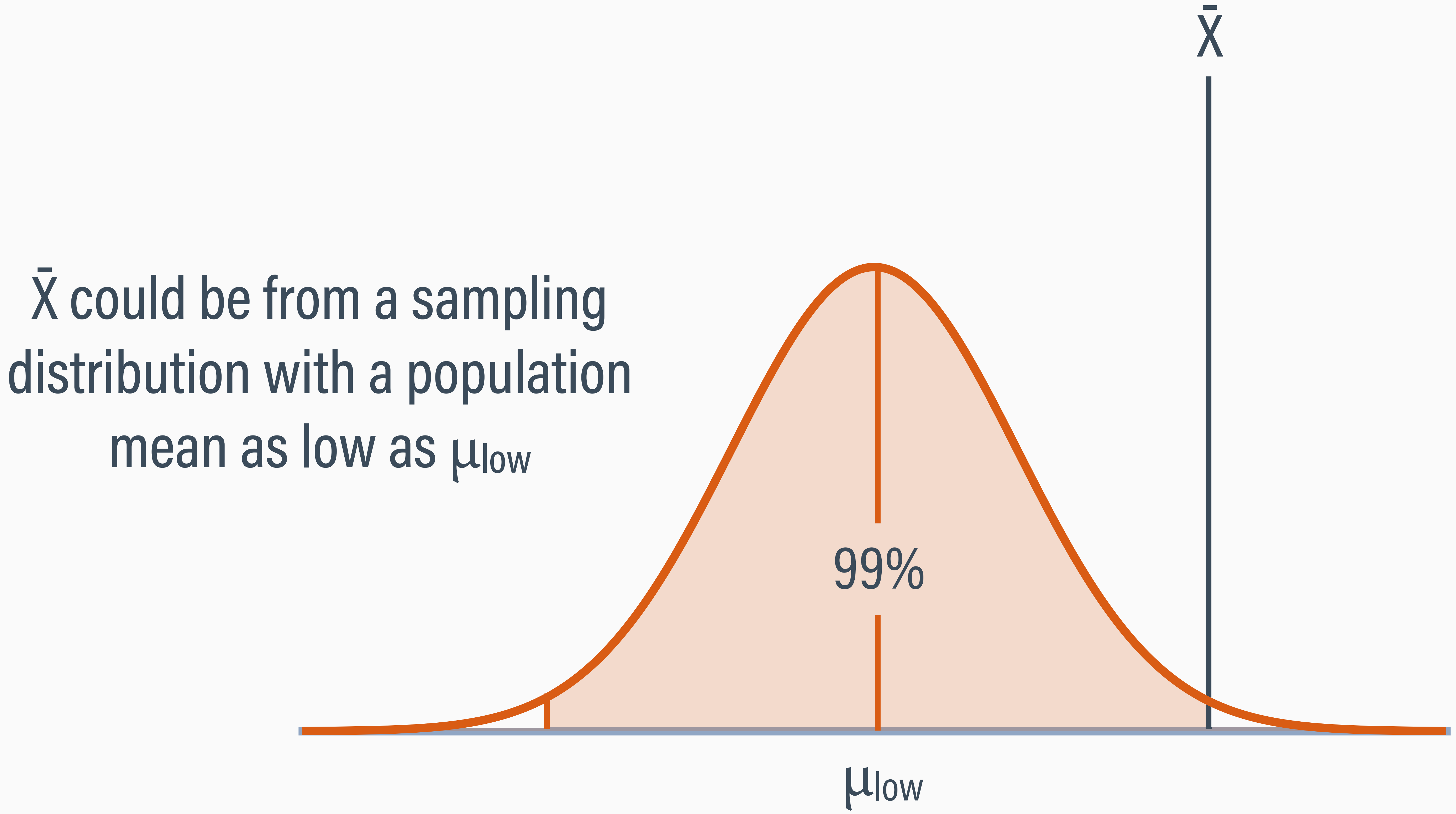
- 95% confidence implies that $\bar{X}$ lies on the lower limit of the middle 95% of all samples from a population with μ_{high} (a higher value of μ is unlikely but not impossible)



$\bar{X}$ could be from a sampling distribution with a population mean as high as μ_{high}

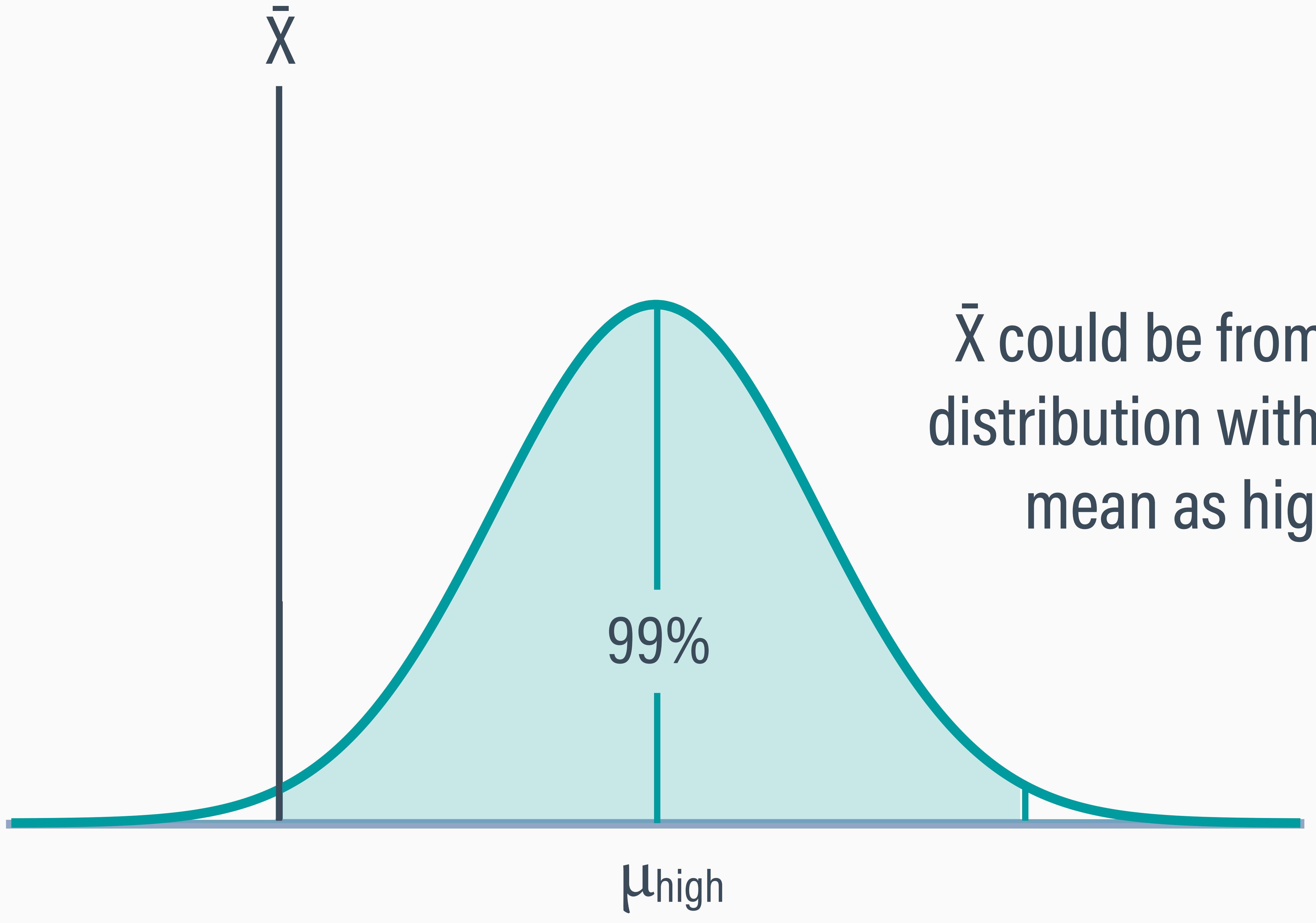
99% CONFIDENCE

- 99% confidence implies that $\bar{X}$ lies on the upper limit of the middle 99% of all samples from a population with μ_{low}



99% CONFIDENCE

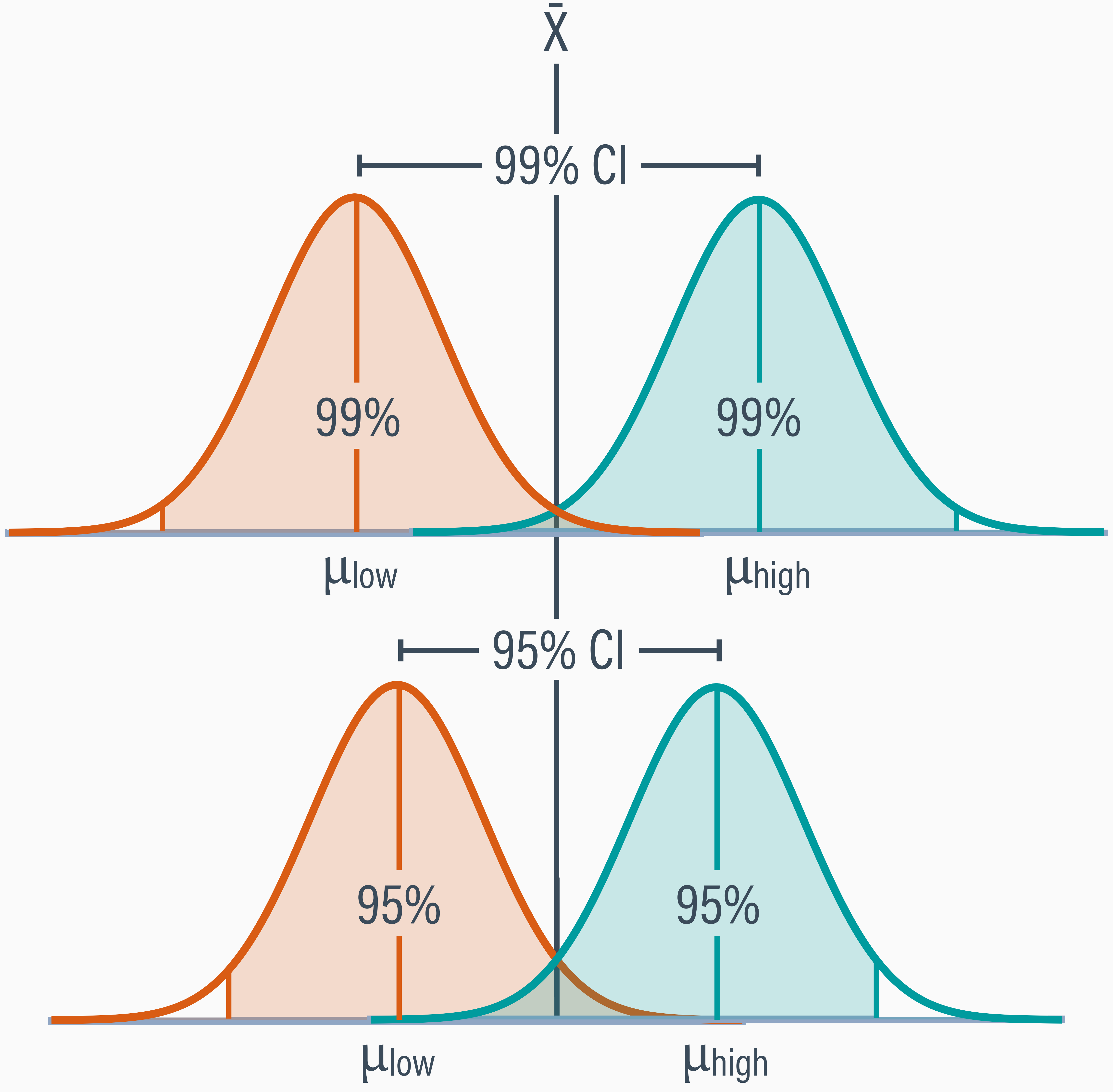
- 99% confidence implies that $\bar{X}$ lies on the lower limit of the middle 99% of all samples from a population with μ_{high}



$\bar{X}$ could be from a sampling distribution with a population mean as high as μ_{high}

95% VS. 99% INTERVALS

- 95% confidence intervals are the norm
- 99% intervals are wider (encompass a broader range of plausible μ) because they allow an estimate to be more distant from μ



OUTLINE

1. Quick review
2. Overview of confidence intervals
3. **Constructing 95% confidence intervals**
4. Properties of confidence intervals
5. Inference by eye
6. Study questions

SMOKING AND DRINKING CESSATION TRIAL

Pharmacological treatments that can concomitantly address cigarette smoking and heavy drinking stand to improve health care delivery for these highly prevalent co-occurring conditions. This superiority trial compared the combination of varenicline and naltrexone against varenicline alone for smoking cessation and drinking reduction among heavy-drinking smokers.

Lara Ray

ReJoyce Green

Ray, L.A., Green, R., Enders, C., et al. (2021). Efficacy of combining varenicline and naltrexone for smoking cessation and drinking reduction: A randomized clinical trial. *American Journal of Psychiatry*, *178*, 818–828.

KEY VARIABLES

Breath (alveolar) carbon monoxide

A measure of carbon monoxide in the lungs. Breath carbon monoxide is a biomarker of smoking behavior common in clinical trials. Higher scores reflect more frequent smoking.

Medication arm

Participants were randomly assigned to receive one of two meds: varenicline plus naltrexone or varenicline plus placebo pills

STEP 1: COMPUTE SAMPLE STATISTICS

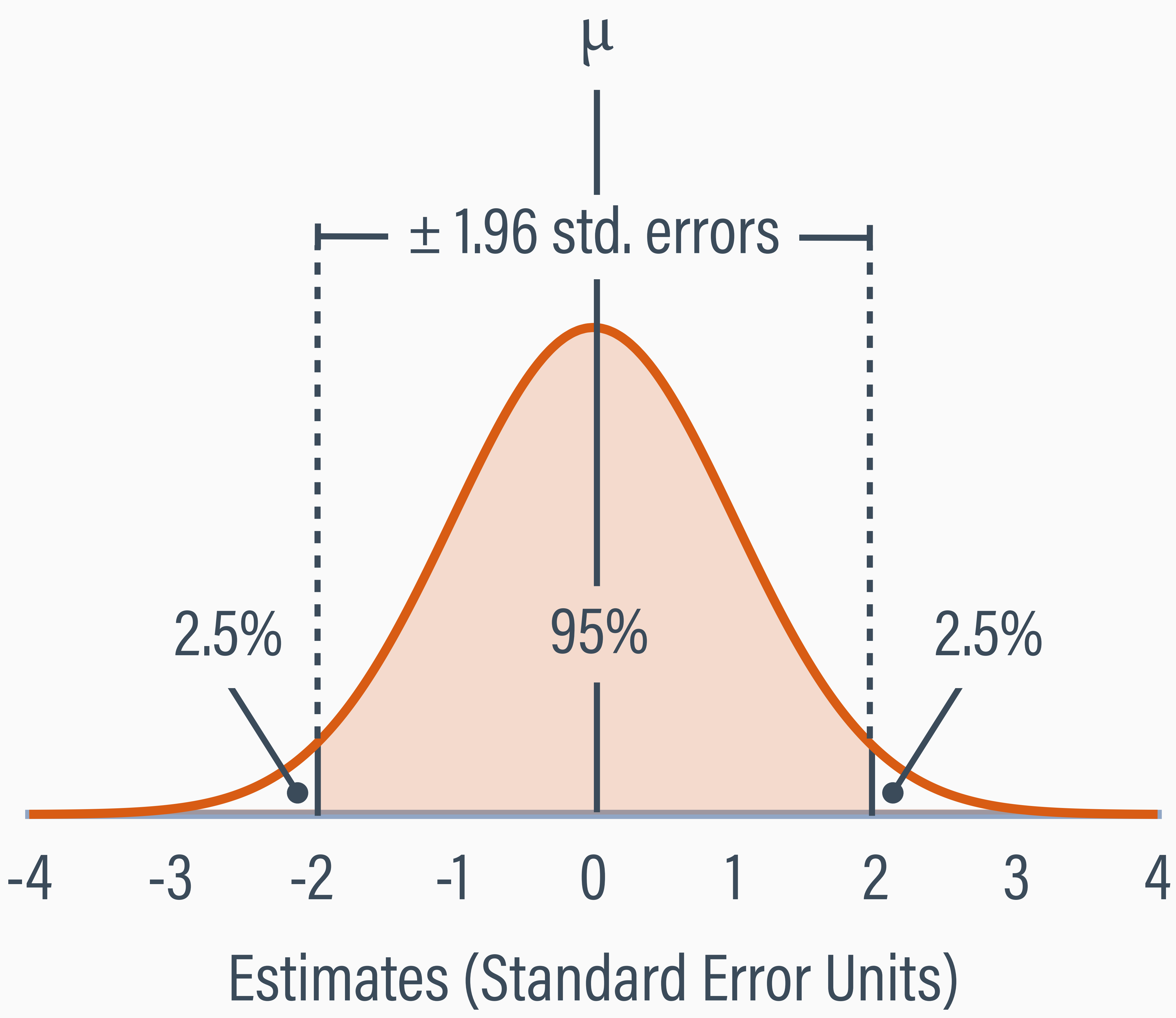
- The standard error is integral to computing confidence intervals
- The expected (average) sampling error for an estimate based on N = 165 is 0.47 breath CO points
- We expect this sample mean to be ±0.47 from the true mean

Mean ($\bar{X}$)	5.5
Std. Dev. (s)	5.96
N	165

$$s_{\bar{X}} = \frac{s}{\sqrt{N}} = \frac{5.96}{\sqrt{165}} = 0.47$$

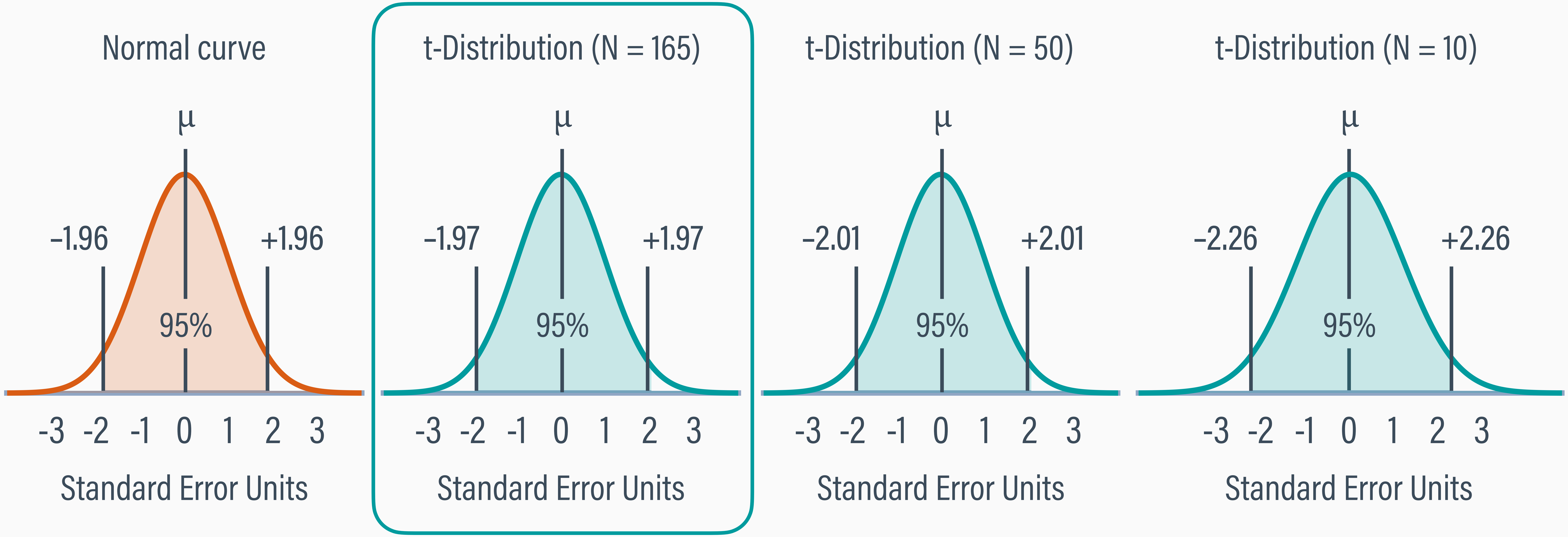
STEP 2: DETERMINE CRITICAL VALUES

- For 95% confidence, we need to determine the width of the 95% confidence range in standard error units (called a critical value)
- We can apply a normal curve approximation
- 95% of the means from large samples are within ± 1.96 standard errors of the true mean



T-DISTRIBUTION CRITICAL VALUES

- Software programs use a t-distribution where 95% critical values stretch out as the sample size decreases



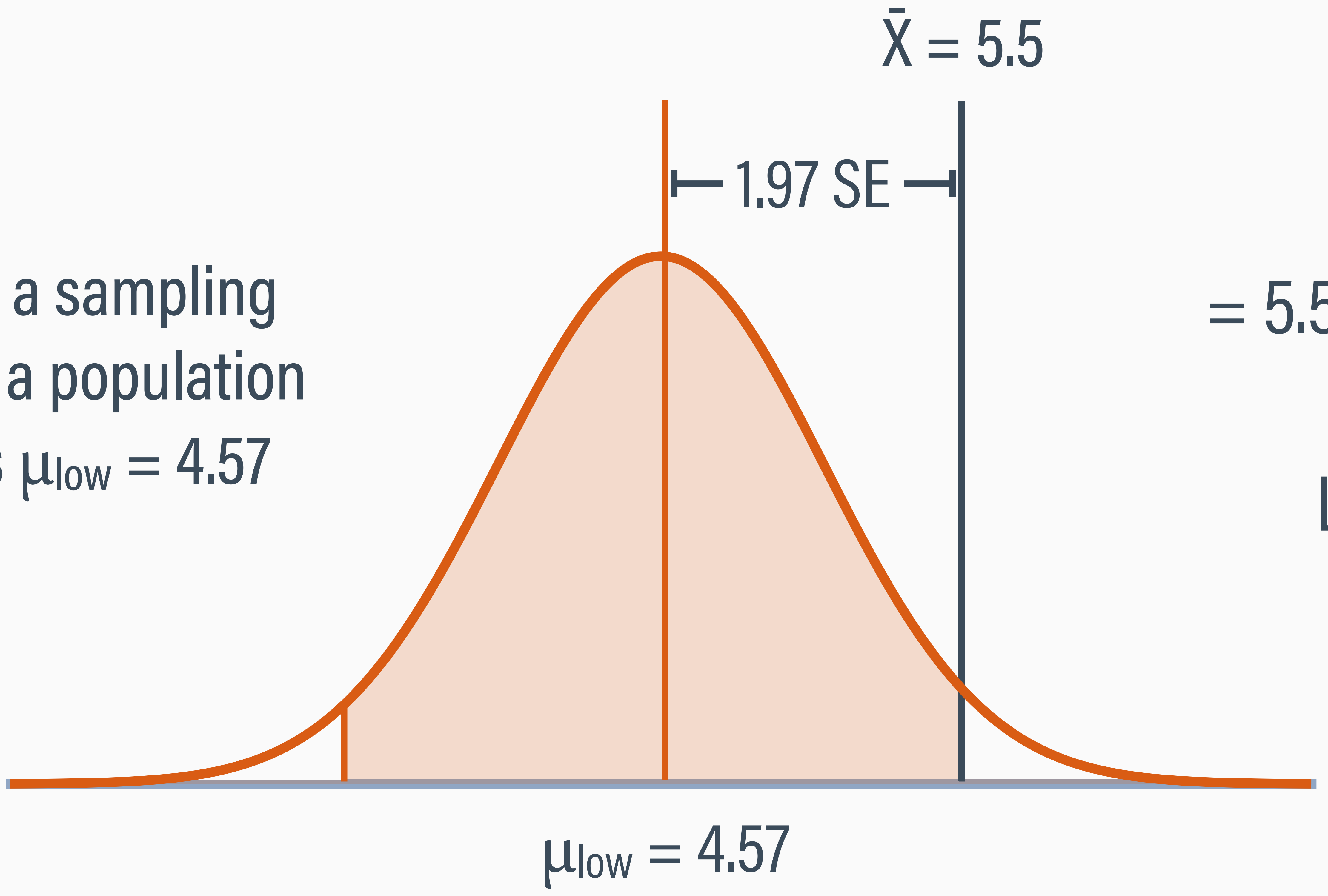
STEP 3: SOLVE FOR INTERVAL LIMITS

- We know that our estimate is 1.97 standard error units above (or below) the lowest (or highest) plausible values of μ
- With the standard error, we can easily solve for μ_{low} and μ_{high}
- $CI_{95\%} = \bar{X} \pm (\text{critical value} \times \text{standard error})$

LOWEST PLAUSIBLE POPULATION MEAN

- The lowest plausible value of μ lies 1.97 standard error units ($\approx 1.97 \times 0.47 = 0.93$ breath CO points) below the sample mean

$\bar{X}$ could be from a sampling distribution with a population mean as low as $\mu_{low} = 4.57$



$$CI_{95\%} = \bar{X} \pm (1.97 \times \text{standard error})$$

$$= 5.5 \pm (1.97 \times 0.47) = [Lower_{95\%}, Upper_{95\%}]$$

$$Lower_{95\%} = 5.5 - (1.97 \times 0.47) = 4.57$$

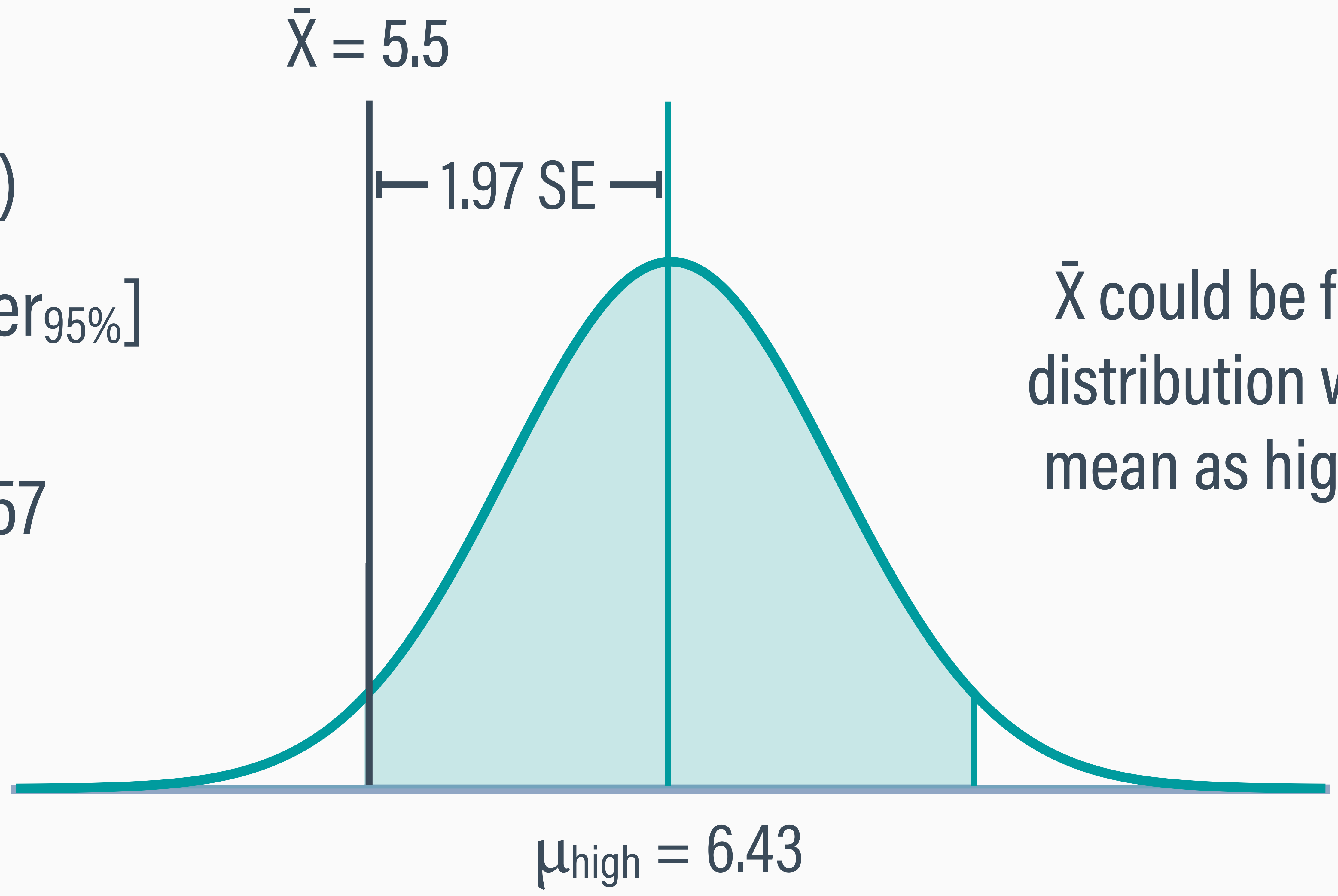
HIGHEST PLAUSIBLE POPULATION MEAN

- The highest plausible value of μ lies 1.97 standard error units ($\approx 1.97 \times 0.47 = 0.93$ breath CO points) above the sample mean

$$CI_{95\%} = \bar{X} \pm (1.97 \times \text{standard error})$$

$$= 5.5 \pm (1.97 \times 0.47) = [Lower_{95\%}, Upper_{95\%}]$$

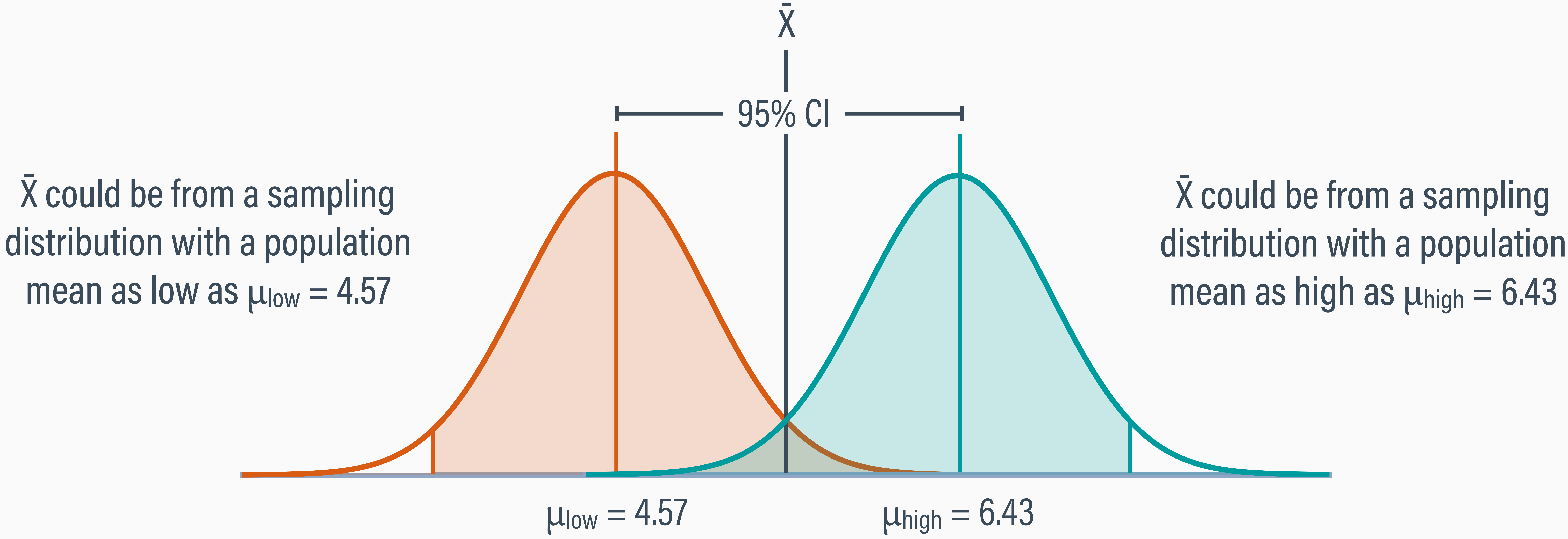
$$Upper_{95\%} = 5.5 + (1.97 \times 0.47) = 4.57$$



$\bar{X}$ could be from a sampling distribution with a population mean as high as $\mu_{high} = 6.43$

95% CONFIDENCE INTERVAL

- $CI_{95\%} = [4.57, 6.43]$ gives the two most extreme values of the population mean that could have reasonably produced the data



R OUTPUT

```
   Mean    Lower    Upper
5.527273 4.611307 6.443238
```

The clinical trial produced a mean and 95% confidence interval of $\bar{X} = 5.5$ and $CI_{95\%} = [4.57, 6.43]$. It is widely believed that breath CO levels of 5 reflect successful outcomes in clinical trials. In small groups of two or three, discuss whether this sample of of $N = 165$ participants could have reasonably originated from a population with a true mean of $\mu = 5$ (i.e., a population where the trial was deemed a success).

OUTLINE

1. Quick review
2. Overview of confidence intervals
3. Constructing 95% confidence intervals
4. **Properties of confidence intervals**
5. Inference by eye
6. Study questions

MEANING OF 95% CONFIDENCE

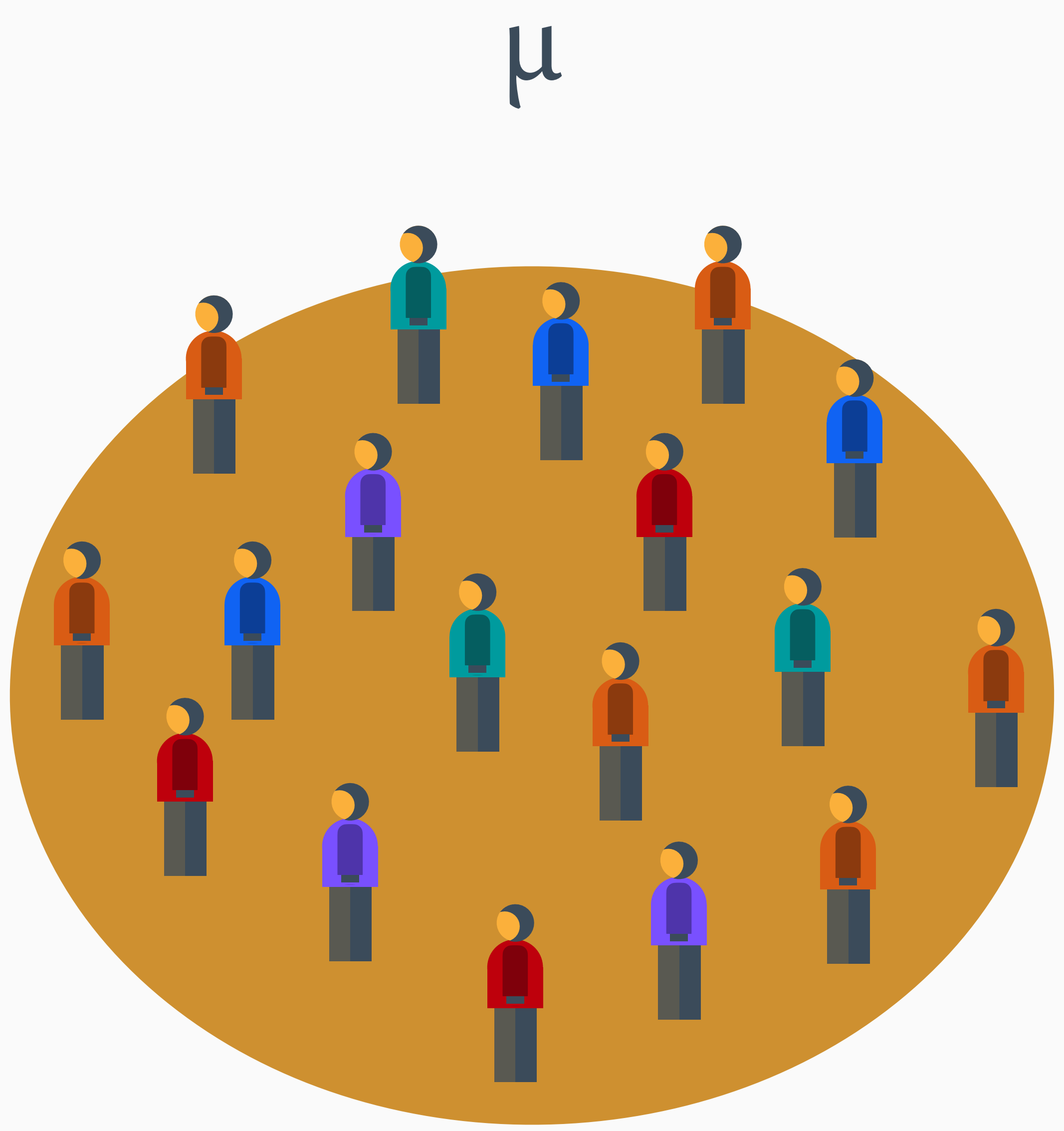
- Researchers often say things like, "we are 95% confident that the true mean falls somewhere in the interval $CI_{95\%} = [4.57, 6.43]$" or "there is a 95% probability that the true mean falls somewhere in the interval $CI_{95\%} = [4.57, 6.43]$"
- This language is a bit sloppy and imprecise because it does not convey what is meant by 95% confident or 95% probable

FREQUENTIST PARADIGM REVISITED

- The frequentist paradigm is defined by the idea that there is one population with unknown parameters (e.g., μ and σ)
- We imagine numerous hypothetical samples of size N from the population, each with its own unique estimates and 95% CIs
- The population-level statistics (parameters) are locked in at a single set of values, whereas the sample-level statistics (estimates) vary across different data sets

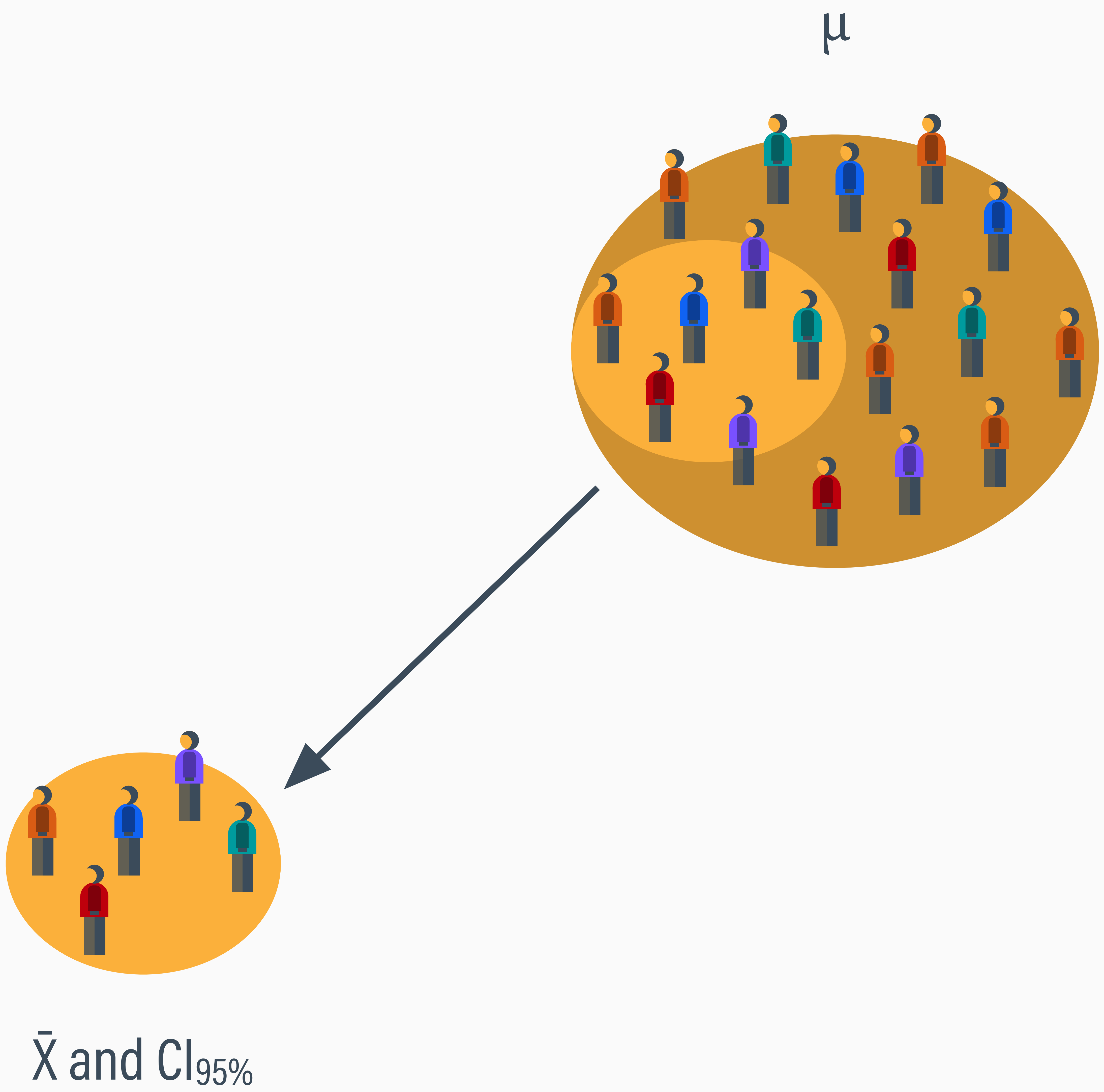
FREQUENTIST FRAMEWORK

One population with unknown parameter μ



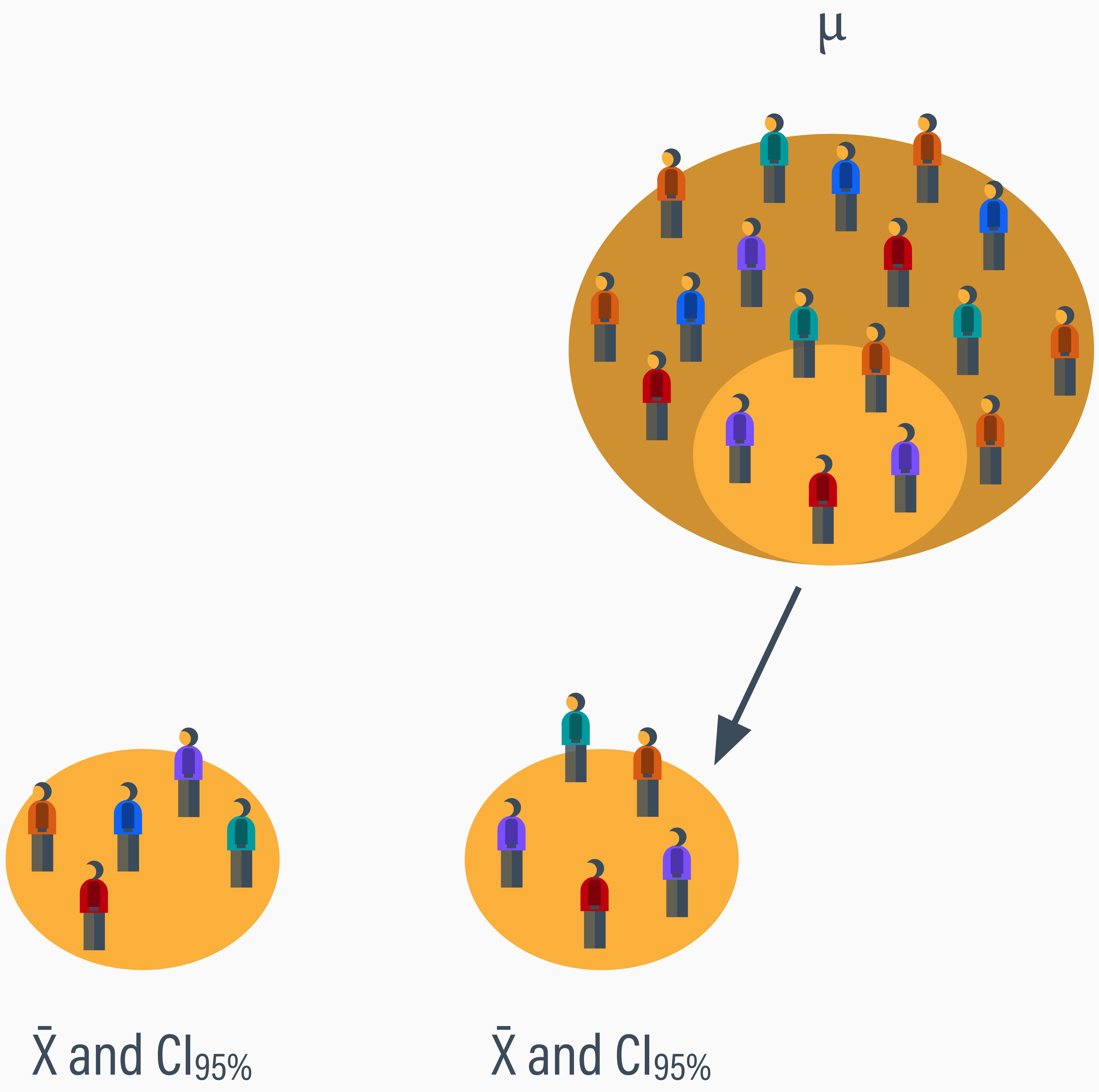
FREQUENTIST FRAMEWORK

One population with unknown parameters. Estimates and CIs vary across different hypothetical samples.



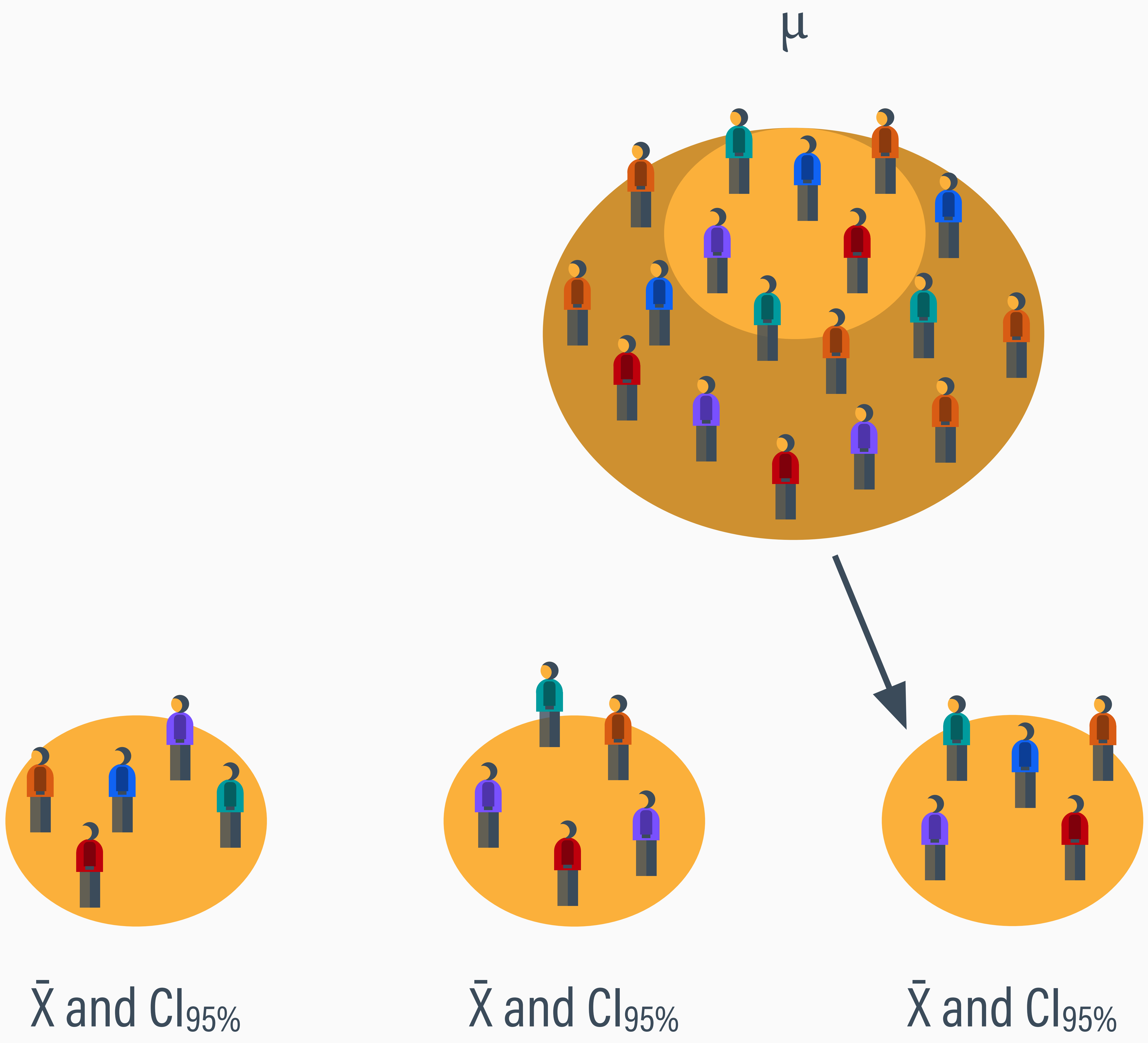
FREQUENTIST FRAMEWORK

One population with unknown parameters. Estimates and CIs vary across different hypothetical samples.



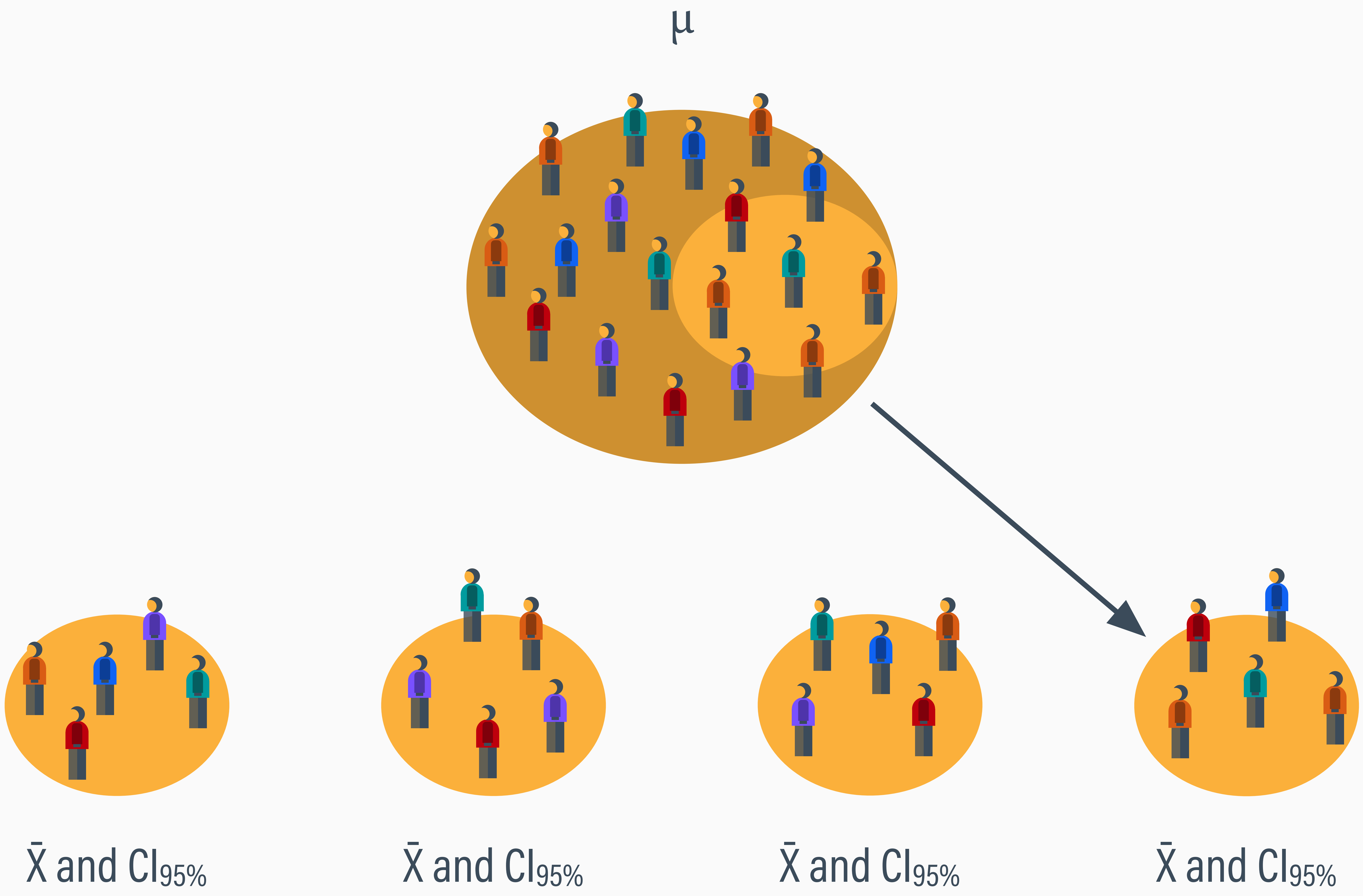
FREQUENTIST FRAMEWORK

One population with unknown parameters. Estimates and CIs vary across different hypothetical samples.

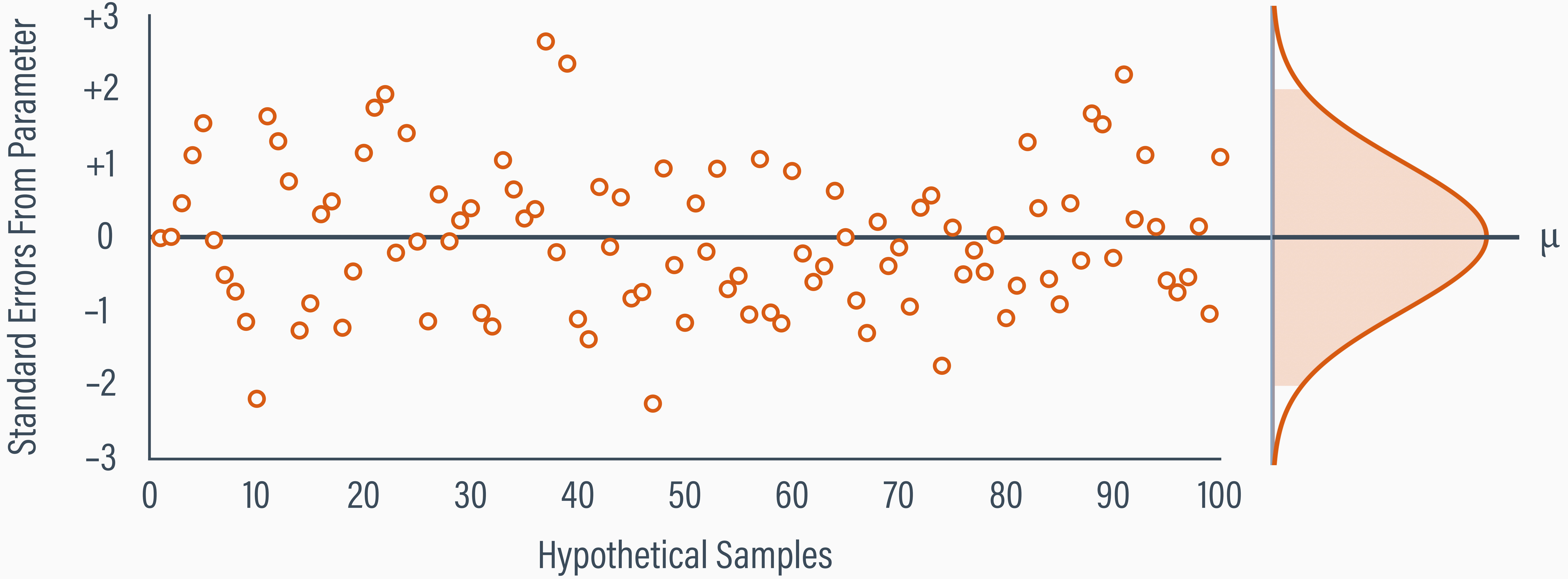


FREQUENTIST FRAMEWORK

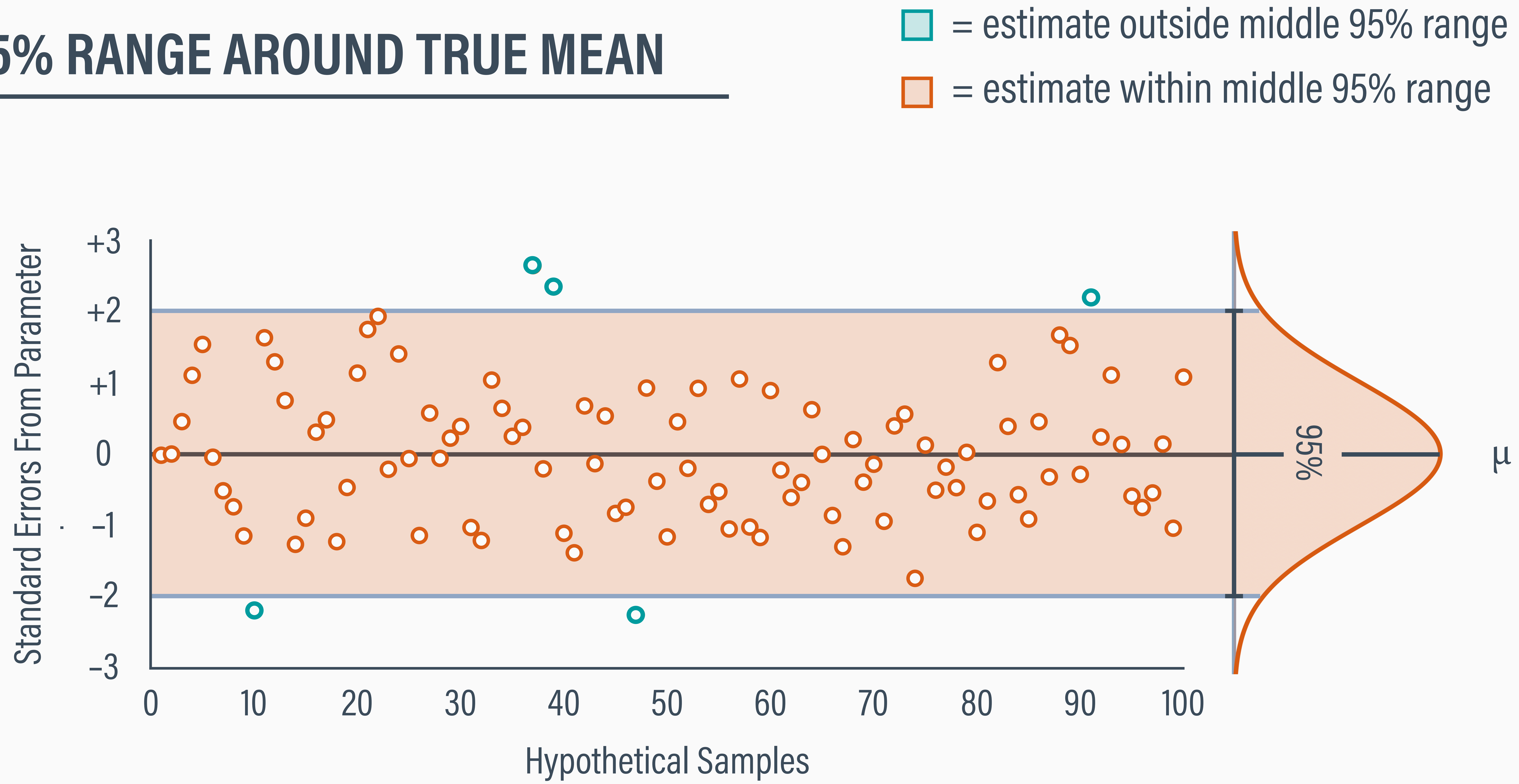
One population with unknown parameters. Estimates and CIs vary across different hypothetical samples.



ESTIMATES FROM 100 RANDOM SAMPLES



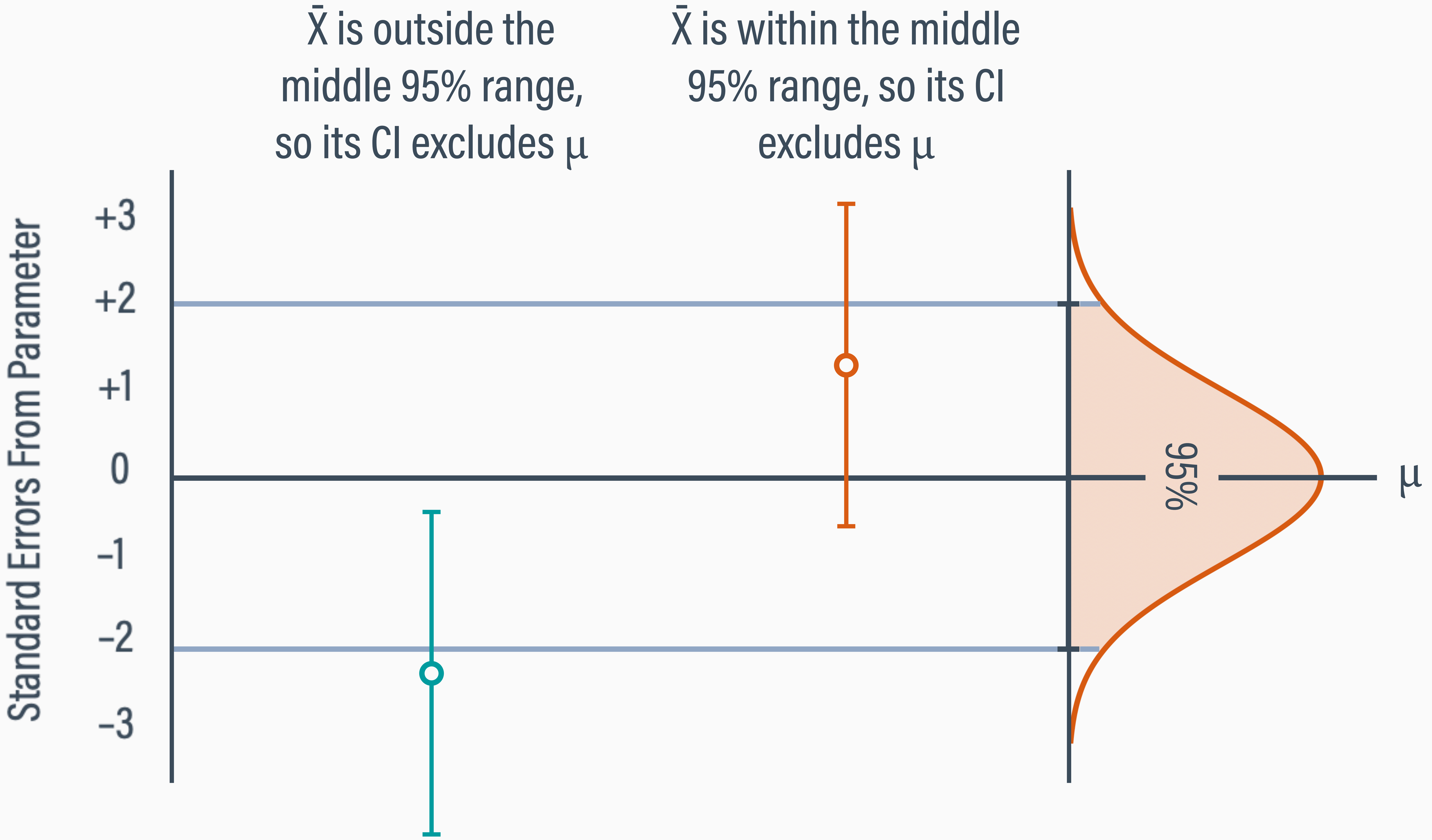
95% RANGE AROUND TRUE MEAN



CONFIDENCE = LONG RUN PROBABILITY

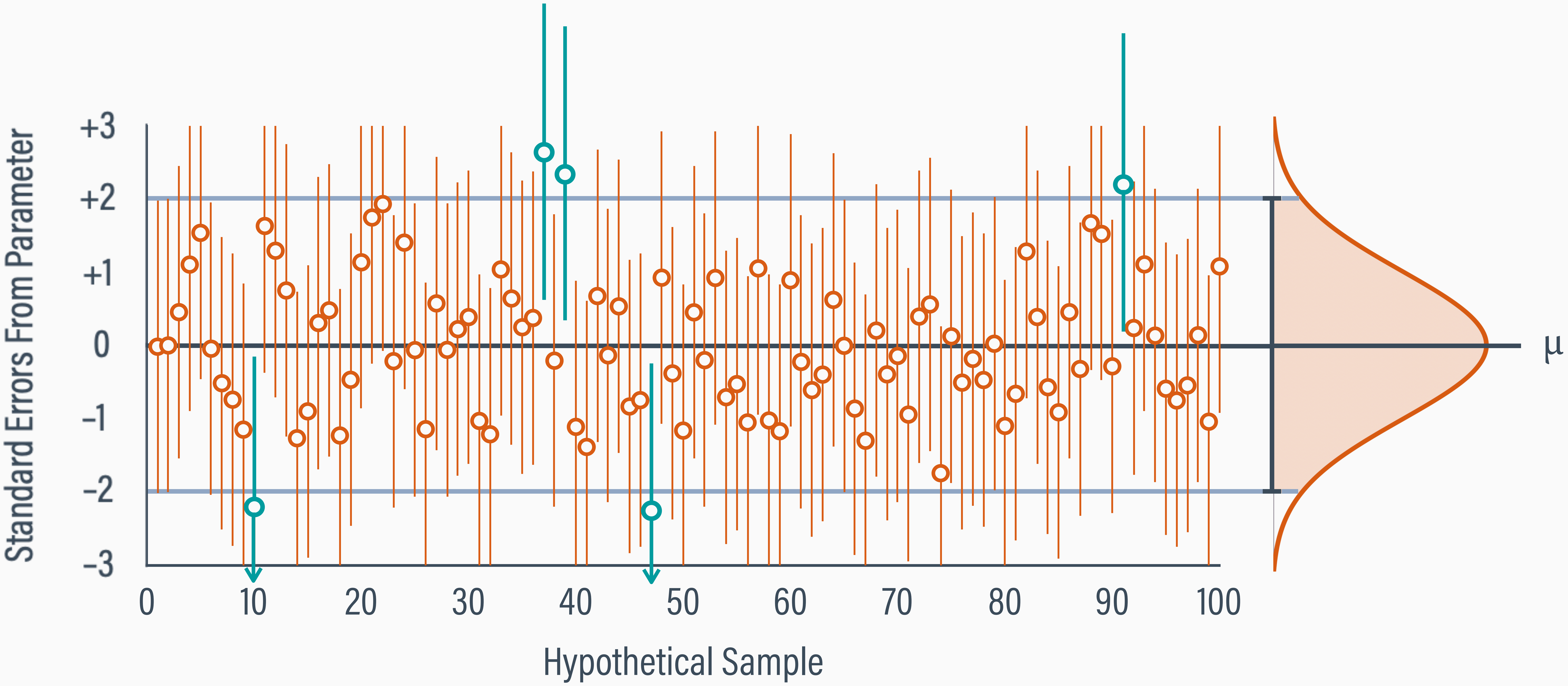
- Each sample has a unique mean and confidence interval
- In any given sample, the interval either contains the true population mean or it does not (the probability is either 0 or 1)
- Confidence and probability are concepts that unfold over many hypothetical random samples from a population
- Across many samples, there is a 95% probability that a confidence interval contains the true population statistic

INTERVALS FROM SINGLE SAMPLES



95% OF INTERVALS CONTAIN THE TRUE MEAN

On average, the 95 out of every 100 samples should give a CI that includes the true population statistic



The clinical trial produced a mean and 95% confidence interval of $\bar{X} = 5.5$ and $CI_{95\%} = [4.57, 6.43]$. Consider the following statement: “There is a 95% probability (we are 95% confident) that the true mean ranges between 4.57 and 6.43”. In small groups of two or three, discuss why this statement is incorrect (or misleading, at best).

OUTLINE

HYPOTHESIS TESTING WITH 95% INTERVALS

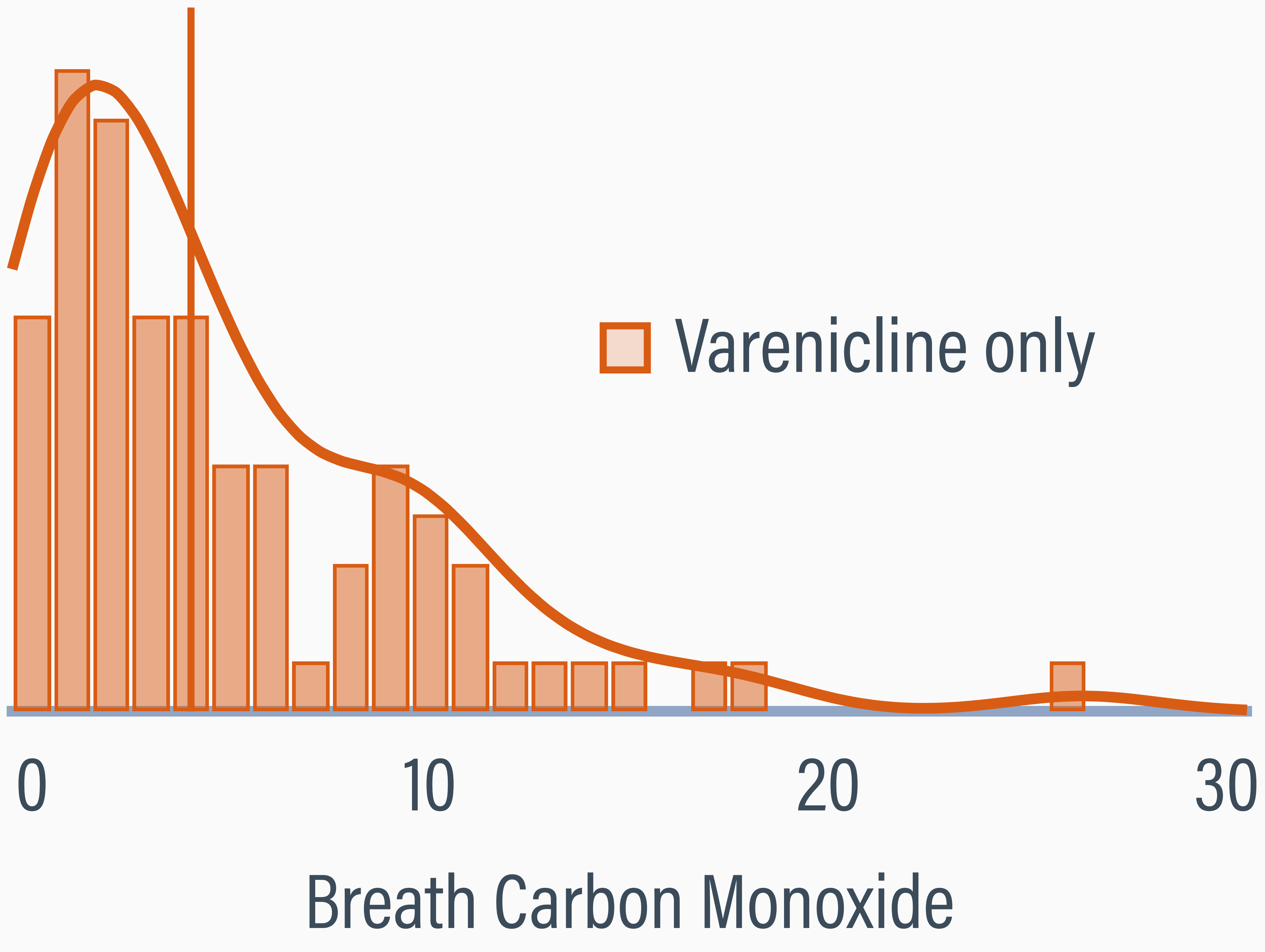
- The clinical trial produced $\bar{X} = 5.5$ and $CI_{95\%} = [4.57, 6.43]$
- It is widely believed that breath CO levels of 5 reflect successful outcomes in clinical trials
- The data support the hypothesis that the true mean is $\mu = 5$ because that value is within the interval (a population with $\mu = 5$ could have plausibly produced these data)

COMPARATIVE RESEARCH QUESTIONS

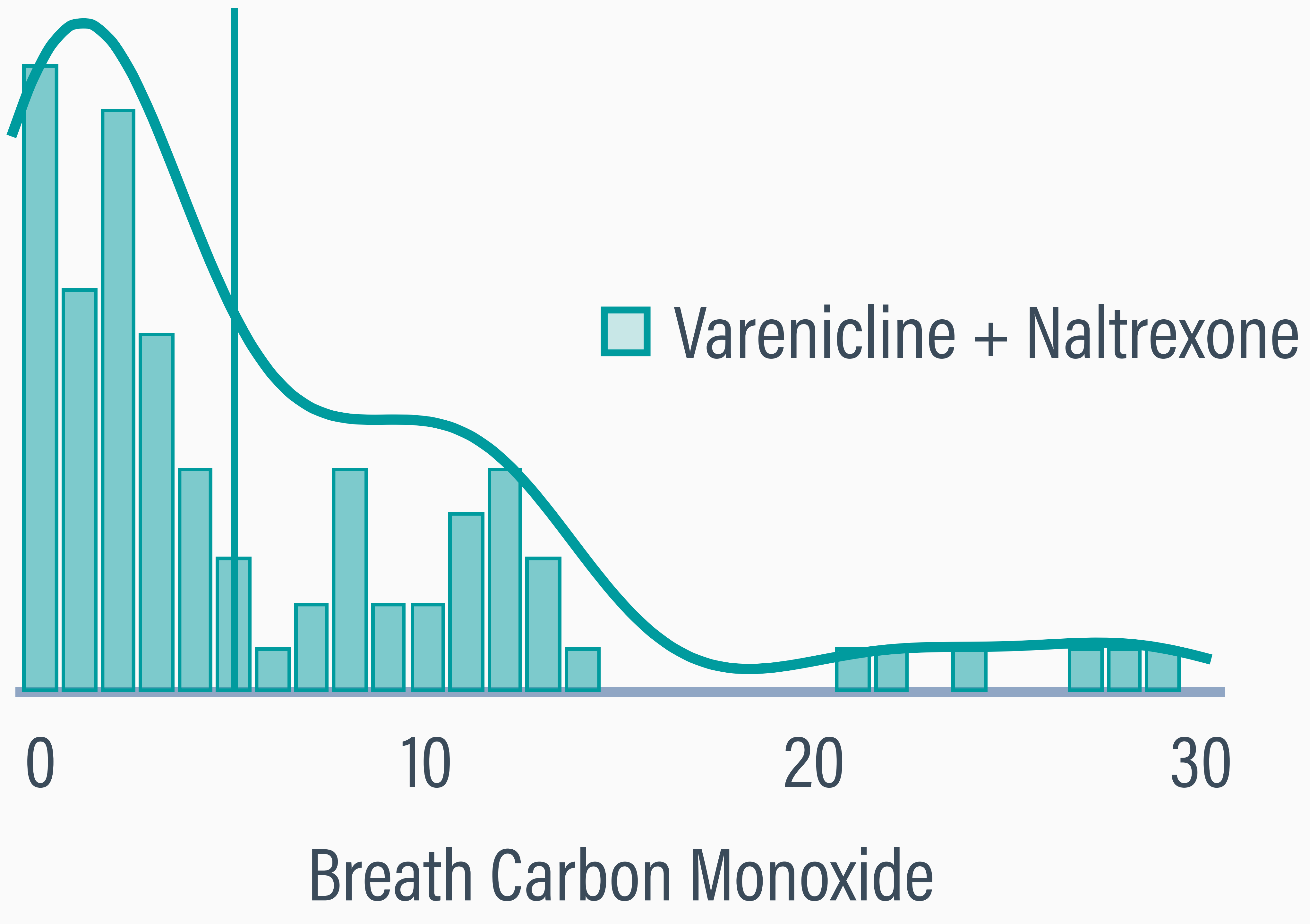
- Comparative research questions ask whether two or more groups (or occasions) differ from one another
- Question: Do participants in the two treatment groups differ in their smoking levels?
- We can answer this question by visually comparing group-specific 95% confidence intervals

GROUP MEANS AND INTERVALS

$\bar{X} = 5.02$
$s = 4.88$
$CI_{95\%} = [3.95, 6.09]$



$\bar{X} = 6.02$
$s = 6.86$
$CI_{95\%} = [4.52, 7.52]$

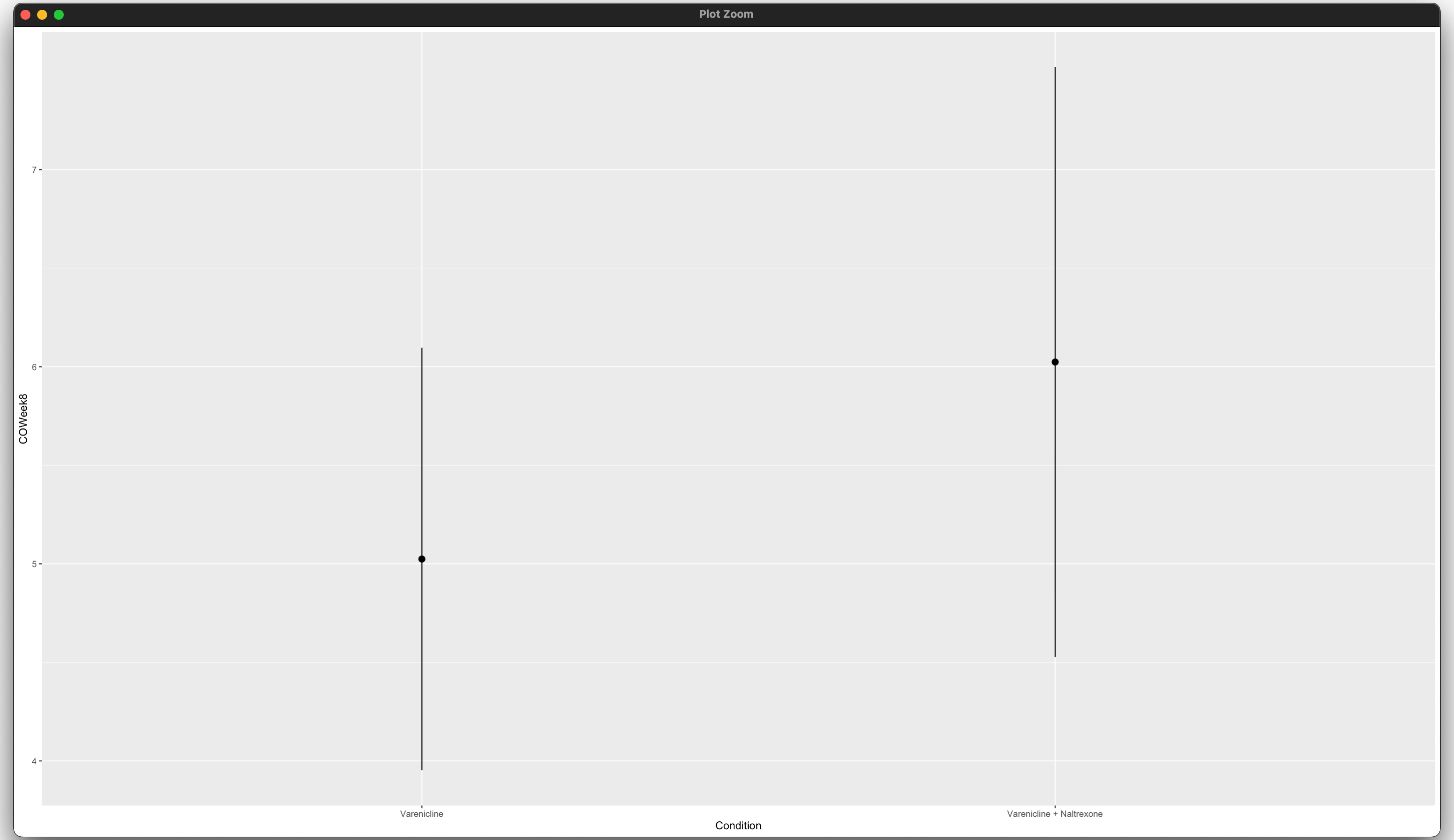


R OUTPUT

```
$Varenicline
    Mean    Lower    Upper 
5.024390 3.952419 6.096362 

$`Varenicline + Naltrexone`
    Mean    Lower    Upper 
6.024096 4.527132 7.521061 
```

R OUTPUT

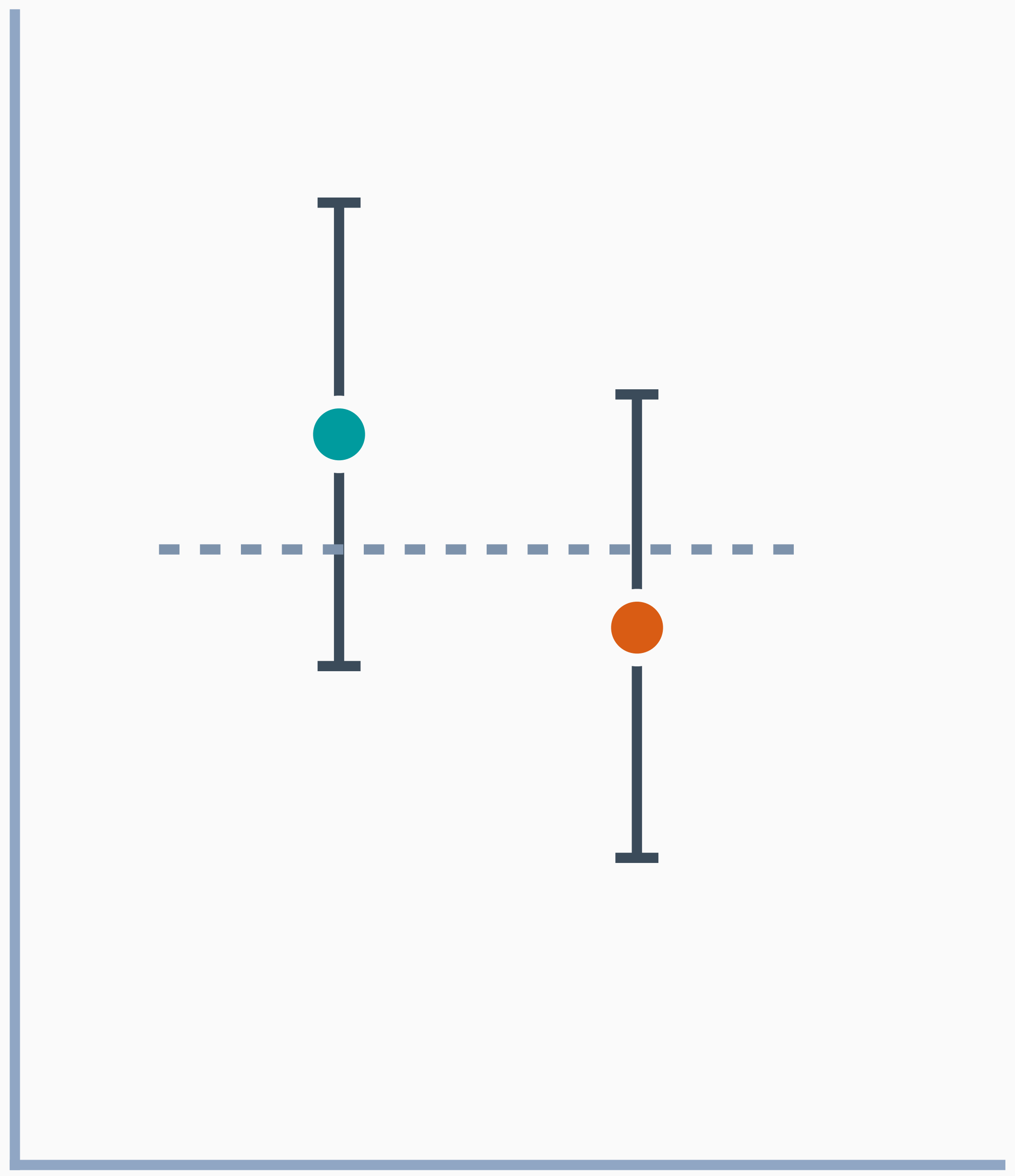

Plot Zoom
7
6
5
4
COWeek8
Varenicline
Varenicline + Naltrexone
Condition

INFERENCE BY EYE

- The amount of confidence interval overlap provides information about whether the difference between two groups is “statistically significant”
- If error bars overlap by less than 50%, the group means differ beyond chance expectations
- The difference is “significant at $p < .05$”

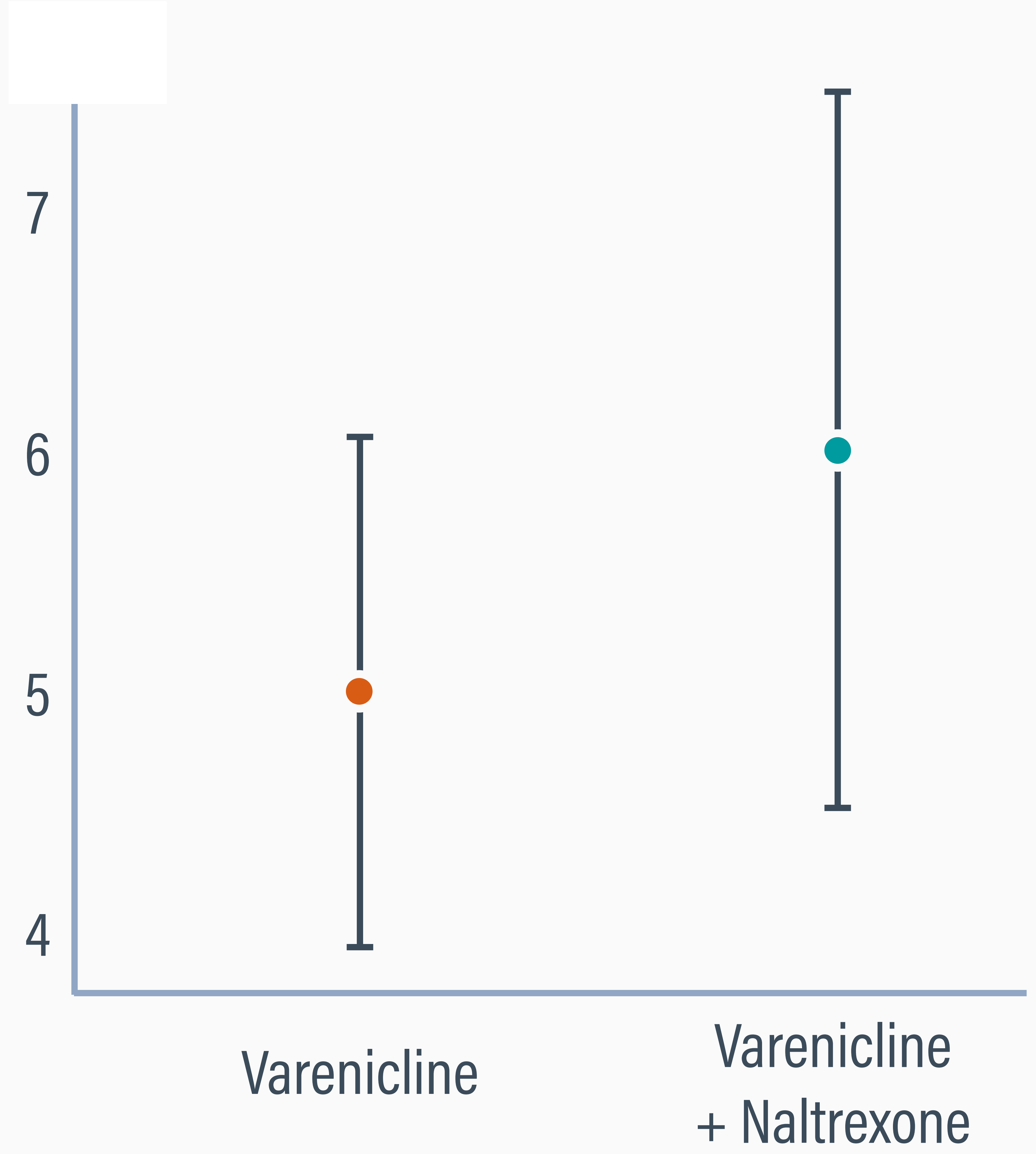
INFERENCE BY EYE

Nonsignificant mean difference ($p > .05$) | Indeterminate mean difference ($p = .05$) | Significant mean difference ($p < .05$)

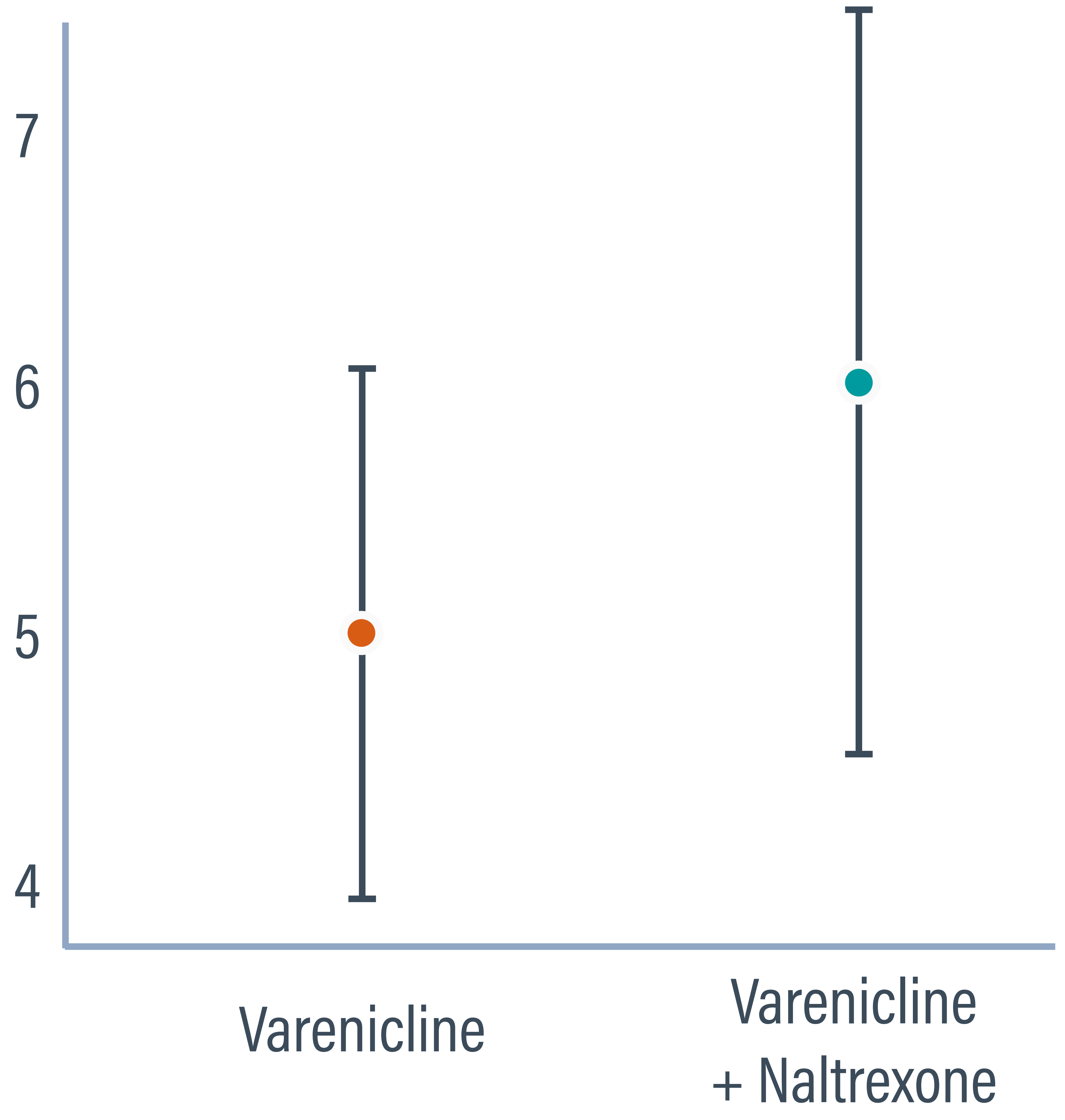


More than 50% overlap in 95% CIs | Exactly 50% overlap in 95% CIs | Less than 50% overlap in 95% CIs

One of the main goals of the research study is to determine whether the treatment groups differ. In small groups of two or three, apply the inference by eye rule. Do the groups exhibit a statistically meaningful difference?



The dual medication group’s 95% interval is noticeably wider, despite the groups being the same size. In small groups of two or three, discuss potential causes for this difference. Hint: look at the descriptive statistics.



OUTLINE

STUDY QUESTIONS (1)

A poll of 500 registered voters reports that 53% of respondents favor some type of government-backed universal health care. The 95% confidence interval is $CI_{95\%} = [49\%, 57\%]$.

1. Provide an interpretation of the confidence interval (I am not asking about its statistical properties). By precise, while describing it in a way that a layperson who reads the poll results could understand.

2. What would happen to the width (span, size) of the confidence interval if the sample size doubled to N = 1000?

STUDY QUESTIONS (2)

A poll of 500 registered voters reports that 53% of respondents favor some type of government-backed universal health care. The 95% confidence interval is $CI_{95\%} = [49\%, 57\%]$.

3. Suppose the researchers want to know if attitudes are perfectly neutral in the population (i.e., the population parameter is 50%). Discuss whether the confidence interval supports or refutes the hypothesis that attitudes toward health care are neutral.

4. What would happen to the size (width) of the confidence interval if the researchers decided that they wanted 99% confidence rather than 95%. What happens if they wanted 90% instead of 95%?

STUDY QUESTIONS (3)

5. On a 1 to 5 scale, a study asks participants to rate how appropriate it is to ask someone out on a date via text. The sample mean is $\bar{X} = 3.52$, and the confidence interval is $CI_{95\%} = [3.37, 3.68]$. Provide an interpretation of the confidence interval. In addition to the interpretation, explain what "95% confident" refers to in the frequentist paradigm.

6. Suppose the researchers want to know if attitudes toward texting are perfectly neutral (i.e., the population mean is 3 on a 1 to 5 scale). Discuss whether the confidence interval supports or refutes the hypothesis that attitudes are neutral.

STUDY QUESTIONS (4)

A poll of 500 registered voters reports that 60% of Democrat respondents favor some type of government-backed universal health care ($CI_{95\%}$ = [56%, 64%]). In contrast, 53% of Republican respondents favor such a plan ($CI_{95\%}$ = [57%, 59%]).

7. Sketch a graph that allows you to apply the inference by eye rule. Do the two groups have statistically different attitudes toward universal health care? Or are they similar?

8. What does it mean to say that you are "95% confident" (or there is a "95% probability") that the interval between 56% and 64% contains the true Democratic mean? Your answer should incorporate a discussion of the frequentist statistical framework.