جامعة نيويورك أبوظبي
NYU | ABU DHABI

# PSYCH-UH 1004Q: Statistics for Psychology

## Class 11: confidence intervals

Prof. Jon Sprouse
Psychology

# Point estimates vs Interval estimates

A **point estimate** is a single number that is offered as an estimate of a statistic or a parameter.

Imagine you were playing a guessing game about the number of jelly beans in a jar. A point estimate would be to give a single number as your guess, like 87.

How confident are you that the true number is equal to your point estimate? Probably not very confident.

An **interval estimate** is a range (a lower bound and an upper bound) that is offered as an estimate for a statistic or a parameter.

An interval estimate would be to give a range as your guess, like 40-120.

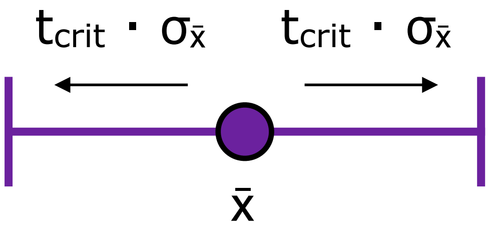
How confident are you that the true number lies within your range? Probably much more confident.

So, what we want is a way of making interval estimates that is grounded in probability…

## Confidence Intervals: a first introduction

# Confidence Intervals

**Confidence intervals** are one specific type of interval estimate.



upper bound: $\bar{x} + (t_{crit} \cdot \sigma_{\bar{x}})$

lower bound: $\bar{x} - (t_{crit} \cdot \sigma_{\bar{x}})$

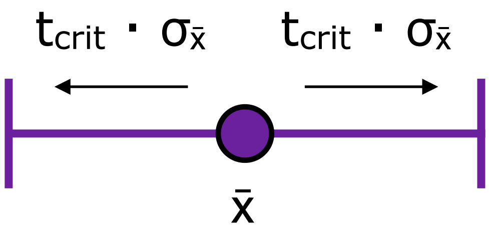
They were created by Jerzy Neyman, so they are frequentist, and they are specifically part of the Neyman-Pearson approach to NHT.

As a frequentist idea, CIs they are about repeating a process over and over again to derive a frequentist probability statement (long-run relative frequency).

Jerzy Neyman
(1894-1981)

As a Neyman-Pearson idea, CIs are about minimizing errors when you make decisions.

# The process for constructing a CI



upper bound: $\bar{x} + (t_{crit} \cdot \sigma_{\bar{x}})$

lower bound: $\bar{x} - (t_{crit} \cdot \sigma_{\bar{x}})$

1. Pick how often you'd like to not be wrong - **95%** is a common choice

2. Look up the critical *t*-value for 5% errors for your experiment's df. You must use the **two-tailed** *t* because the population mean could be in either direction.

3. Multiply that *t*-value by the **standard error** (estimated from your sample).

4. **Subtract** that number from your sample mean for the lower bound and **add** that number to your mean for the upper bound.

5. **Declare** that the population mean is in this interval. If we were to repeat this over and over, we'd be right 95% of the time!

Warning - the population mean may not be in the interval! We don't know the population mean, so we don't know. But under the N-P approach to NHT, we need to make a DECISION. In N-P, we just declare that it is there. (Just like we just declare that we have rejected the null hypothesis or not.)

# Some examples calculating a CI

| **Example** | **Numbers to calculate** | **Confidence Interval** |
|---|---|---|
| mean = 47<br>standard deviation = 30<br>df=24 | R-code: qt(.975, df=24)<br>$t_{crit}$ = 2.06<br>$\sigma_{\bar{x}}$ = 6 | **CI:** 47 ± 12.36<br>so, 34.64 to 59.36 |
| mean = 148<br>standard deviation = 20<br>df=15 | R-code: qt(.975, df=15)<br>$t_{crit}$ = 2.13<br>$\sigma_{\bar{x}}$ = 5 | **CI:** 148 ± 10.65<br>so, 137.35 to 158.65 |
| mean = 5.5<br>standard deviation = .75<br>df=8 | R-code: qt(.975, df=8)<br>$t_{crit}$ = 2.31<br>$\sigma_{\bar{x}}$ = .25 | **CI:** 5.5 ± 0.58<br>so, 4.92 to 6.08 |

Remember: we **do not know** if the population mean is within these intervals or not. We simply **declare** that it is. The goal is to minimize our total errors if we repeat the process over and over, not know whether we made an error or not in this specific instance.

# The four N-P outcomes

Calculating a CI and declaring that the population mean is within it will lead to four combinations of the state of the universe and your decision.

| | | **The population mean is:** | |
|---|---|---|---|
| | | outside the CI | inside the CI |
| **Your decision is:** | inside the CI | error | correct decision |
| | outside the CI | correct decision | error |

The cool thing is that if you follow this procedure for constructing CIs and declaring the population mean is in the CI, you will be correct 95% of the time (or whatever percentage you set), and only make an error 5% of the time (or 1 minus whatever percentage you set).

# In practice you only experience the top row

With CIs, you never claim that the mean is outside the CI, so you really only need to look at the top row of the chart:

| | | **The population mean is:** | |
|---|---|---|---|
| | | outside the CI | inside the CI |
| **Your decision is:** | inside the CI | error | correct decision |
| | outside the CI | correct decision | error |

The cool thing is that if you follow this procedure for constructing CIs and declaring the population mean is in the CI, you will be correct 95% of the time (or whatever percentage you set), and only make an error 5% of the time (or 1 minus whatever percentage you set).

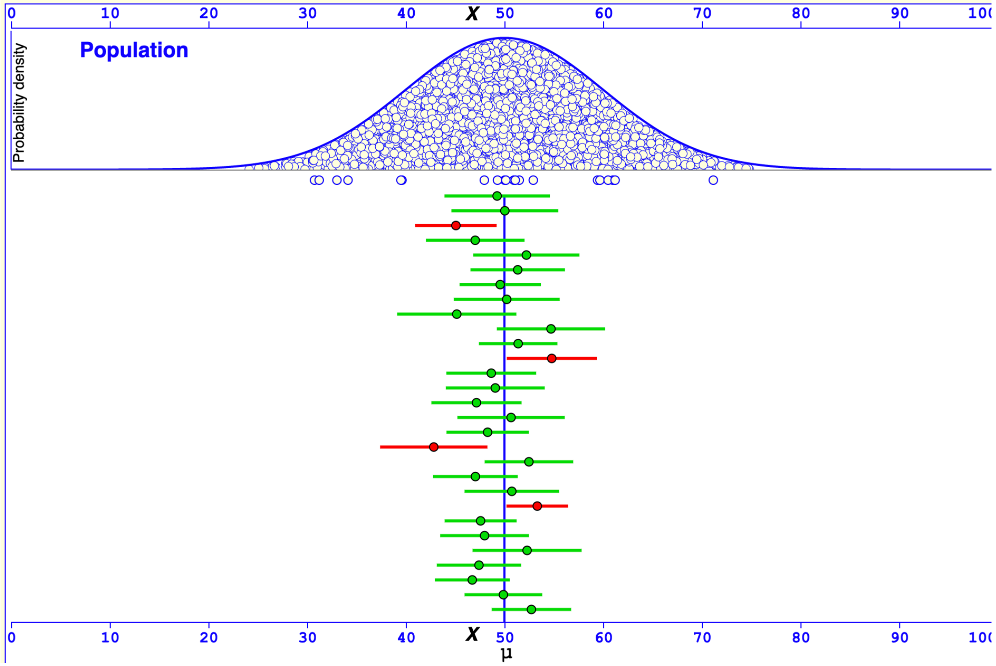
# To see the probability, we can watch a simulation

This is an interactive chart that simulates sampling repeatedly from a population, calculating a mean and standard deviation for the sample, and then drawing a confidence interval around that sample mean.

https://www.esci-dances.thenewstatistics.com/

This will allow us to watch how many CIs overlap the population mean.

You will see that it is the percentage that we set.



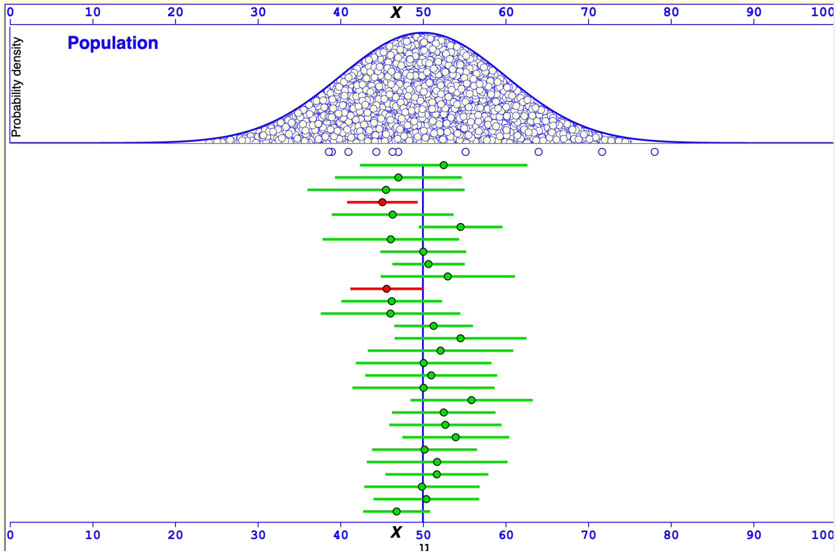
But, crucially, you can also see that some CIs do not contain the population mean. It is still just an estimate!

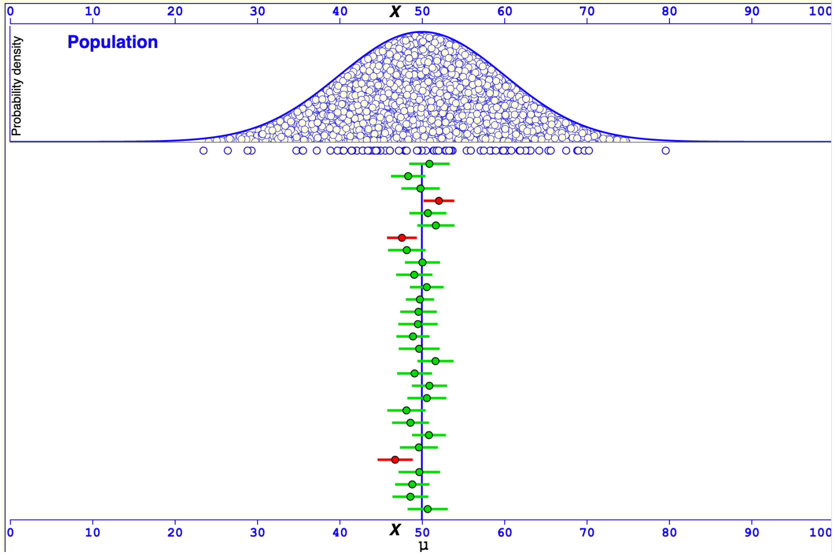
# The width of CIs

You can see from the equation that the width of CIs is dependent on the standard error of the sample. This will vary from sample to sample

upper bound: $\bar{x} + (t_{crit} \cdot \sigma_{\bar{x}})$

lower bound: $\bar{x} + (t_{crit} \cdot \sigma_{\bar{x}})$





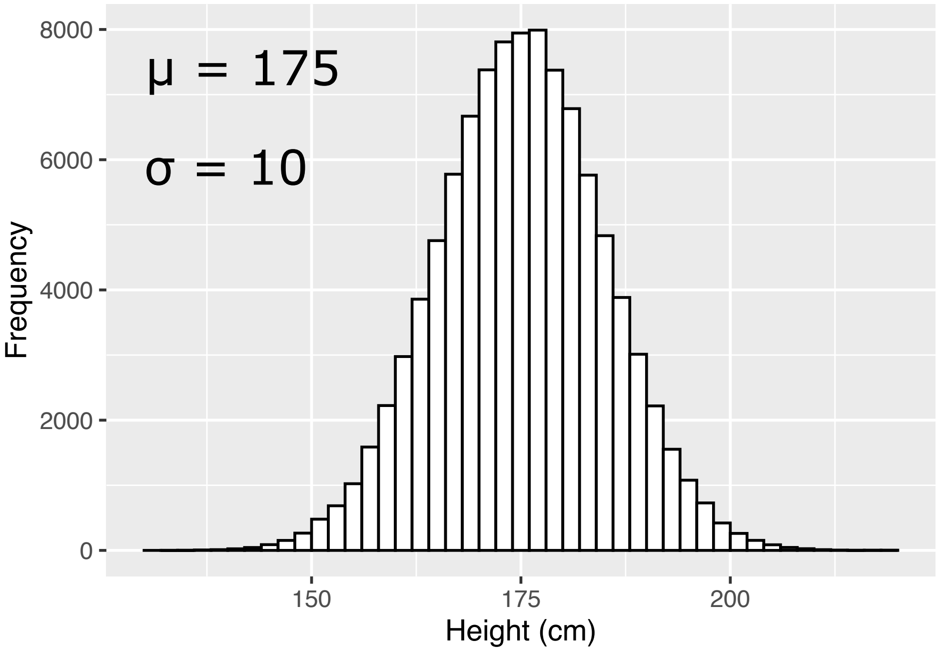
And it will also vary by sample size because standard error is divided by the square root of the sample size. So a larger sample size will lead to a smaller SE all other things equal.

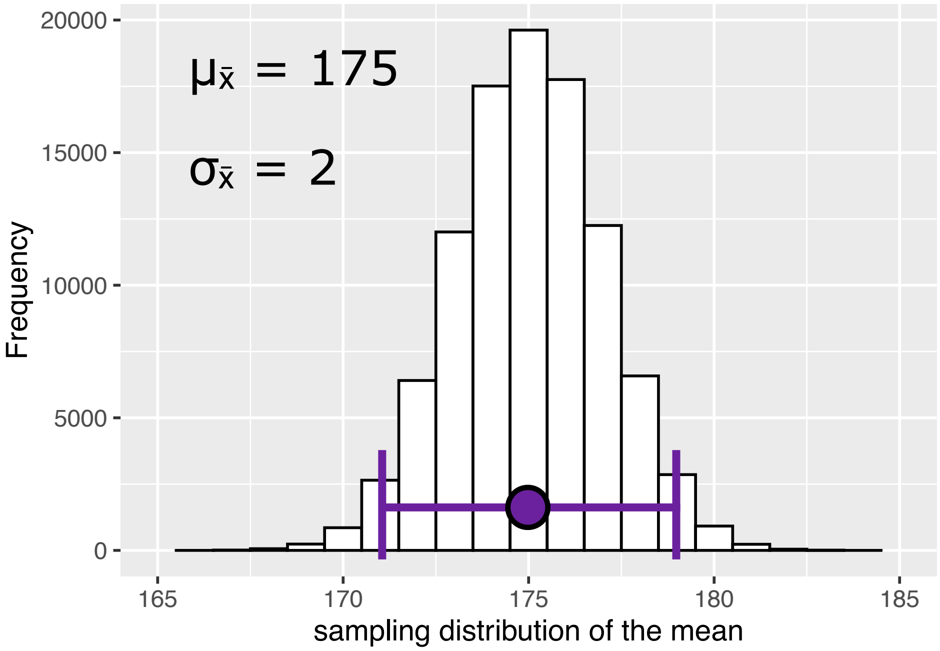
# Why does this process work?

(why does the process capture the population mean in 95% of cases?)

# Why does the process yield CIs containing the population mean 95% of the time?

For df=24, the critical *t* for a two-tailed alpha of .05 is 2.06. That is the t-value that marks .025 in either tail of the *t*-distribution. If we multiply it by the standard error (as we do in CIs!), it marks out 95% of the sampling distribution of the mean!



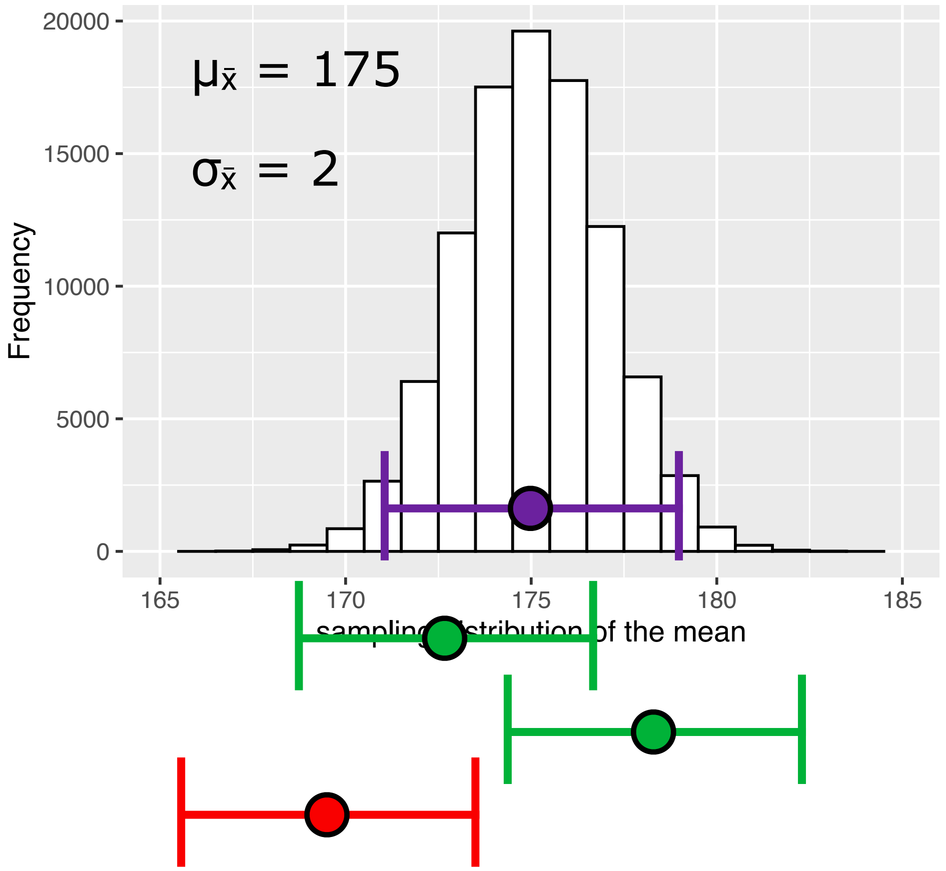


If we draw a 95% confidence interval around the mean of the sampling distribution of the mean (because we are talking about sample means!), you can see that it captures 95% of possible sample means!

# Why does the process yield CIs containing the population mean 95% of the time?

And now we can see that if we select other sample means, and draw a 95% CI around those, their CI will overlap the population mean if they are within that 95%.

In reality, things are a bit more complicated because the width of the CIs will vary based on the standard deviation of the specific sample, but this will be offset by changes in the mean.



On average, sample means that are within 95% of the distribution of the population mean will have CIs that overlap the population mean. So, on average, if you declare that your CIs always include the population mean, you will be correct 95% of the time.

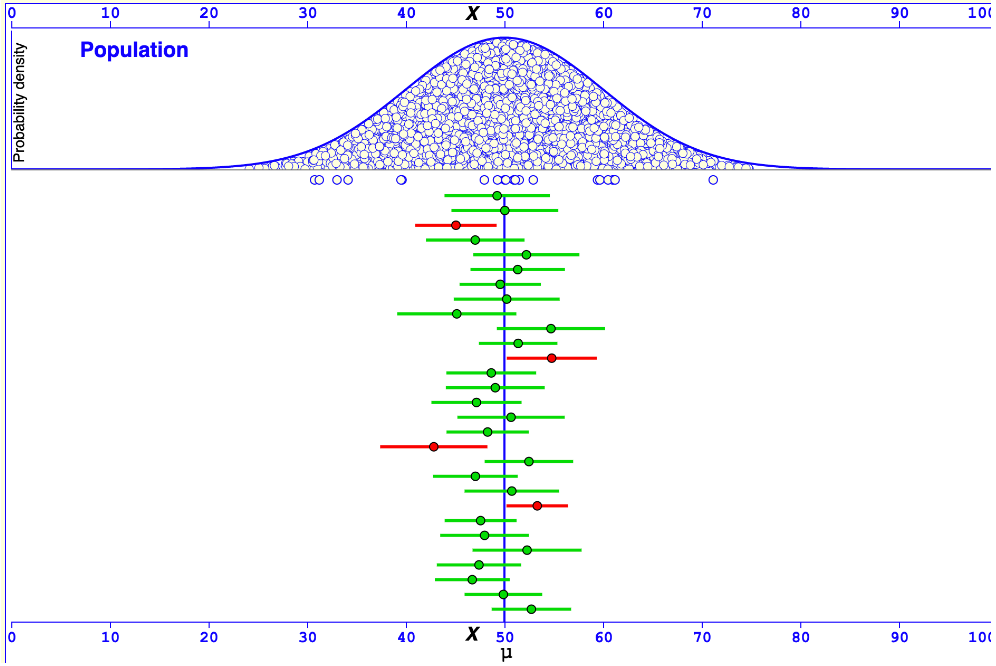
# To see the probability, we can watch a simulation

Let’s watch this again. It is worth seeing it so you can really feel how CIs work!

https://www.esci-dances.thenewstatistics.com/

This will allow us to watch how many CIs overlap the population mean.

You will see that it is the percentage that we set.



But, crucially, you can also see that some CIs do not contain the population mean. It is still just an estimate!

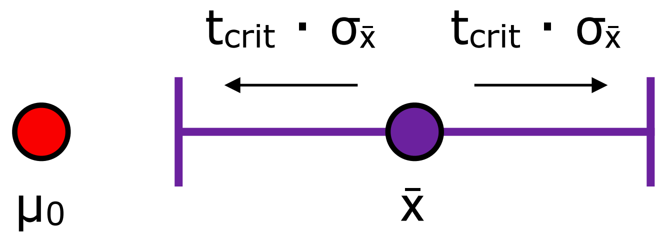
# A cool relationship between CIs and *t*-tests

(Or, why you can use CIs as a one-sample, two-tailed *t*-test to reject the null hypothesis at a given alpha.)

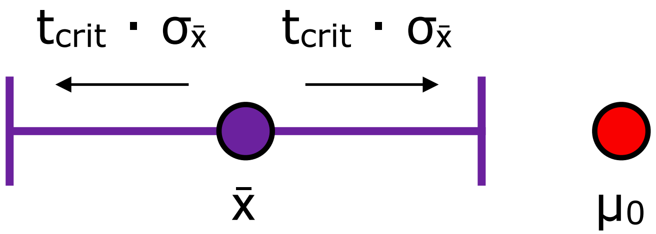
# Using CIs as two-tailed one-sample *t*-test

**Confidence intervals** are equivalent to a two-tailed one-sample *t*-test. To see this, let's imagine some scenarios in which you have calculated a CI from your sample mean (in violet) and you have a population mean under the null hypothesis (in red).

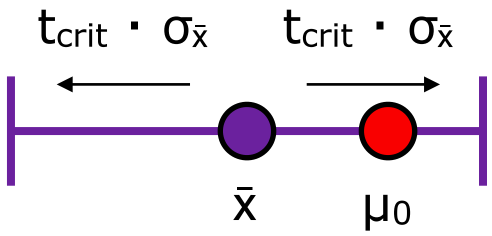
If the $\mu_0$ is less than the lower bound of the CI, then a one-sample *t*-test would be **significant at p<.05**.



If the $\mu_0$ is greater than the upper bound of the CI, then a one-sample *t*-test would be **significant at p<.05**.



If the $\mu_0$ is within the bounds of the CI, then a one-sample *t*-test would **not** be significant **at p<.05**.



# We can see this with simple algebra!

For df=24, the critical *t* for a two-tailed alpha of .05 is 2.06.

In a two-tailed *t*-test, we reject the null when *t*>2.06 or *t*<-2.06.

|  |  |  |
|---|---|---|
| Here are the formulae for the two outcomes that will be significant: | $\frac{\bar{x} - \mu_0}{s_{\bar{x}}} > 2.06$ | $\frac{\bar{x} - \mu_0}{s_{\bar{x}}} < -2.06$ |
| Multiply both sides by $s_{\bar{x}}$: | $\bar{x} - \mu_0 > 2.06 \cdot s_{\bar{x}}$ | $\bar{x} - \mu_0 < -2.06 \cdot s_{\bar{x}}$ |
| Subtract $\bar{x}$ from both sides: | $- \mu_0 > 2.06 \cdot s_{\bar{x}} - \bar{x}$ | $- \mu_0 < -2.06 \cdot s_{\bar{x}} - \bar{x}$ |
| Multiply both sides by -1: | $\mu_0 < \bar{x} - 2.06 \cdot s_{\bar{x}}$ | $\mu_0 > \bar{x} + 2.06 \cdot s_{\bar{x}}$ |
| So, just through basic algebra, we can equivalently say that we reject the null when: | $\mu_0 < \bar{x} - 2.06 \cdot s_{\bar{x}}$ | $\mu_0 > \bar{x} + 2.06 \cdot s_{\bar{x}}$ |

You will recognize these as the two equations for the lower and upper bounds of the 95% confidence interval. This shows that if the null hypothesis mean is below or above the interval, it is algebraically equivalent to being significant!

# Using CIs as *t*-tests in practice

**Example**

mean = 47
standard deviation = 30
df=24

**Numbers to calculate**

R-code: qt(.975, df=24)
$t_{crit} = 2.06$
$\sigma_{\bar{x}} = 6$

**Confidence Interval**

**CI:** 47 ± 12.36

We can equivalently say that we can reject the null hypothesis at an alpha of .05 (.025 in each tail) for any population mean above 59.36 or below 34.64.

**Example**

mean = 5.5
standard deviation = .75
df=8

**Numbers to calculate**

R-code: qt(.975, df=8)
$t_{crit} = 2.31$
$\sigma_{\bar{x}} = .25$

**Confidence Interval**

**CI:** 5.5 ± 0.58

We can equivalently say that we can reject the null hypothesis at an alpha of .05 (.025 in each tail) for any population mean above 6.08 or below 4.92.

CAUTION: confidence intervals are open to misinterpretation

# What you can conclude

The probability statement for confidence intervals is about the procedure of drawing the boundaries. It is not about the specific values contained within the boundaries.

What you can claim correctly: the procedure for drawing CIs means that 95% of the CIs drawn this way will contain the population mean.

What you can claim correctly: if a value is outside of the CI, a *t*-test would be significant at the equivalent alpha level (95% = .05 alpha, .025 in each tail) for a population mean of that value.

| **Example** | **Numbers to calculate** | **Confidence Interval** |
|---|---|---|
| mean = 47 | R-code: qt(.975, df=24) | **CI:** 47 ± 12.36 |
| standard deviation = 30 | $t_{crit}$ = 2.06 | |
| df=24 | $\sigma_{\bar{x}}$ = 6 | |

So, I can claim that if I repeat this procedure over and over, getting a different CI each time, 95% of them will contain the population mean. And, I can claim that a null hypothesis for a population mean above or below these bounds would be significant. That is all.

# What you cannot conclude

The probability statement for confidence intervals is about the procedure of drawing the boundaries. It is not about the specific values contained within the boundaries.

What you CAN'T claim: anything about the specific values in a single CI that is sitting in front of you.

| **Example** | **Numbers to calculate** | **Confidence Interval** |
|---|---|---|
| mean = 47 | R-code: qt(.975, df=24) | **CI:** 47 ± 12.36 |
| standard deviation = 30 | $t_{crit} = 2.06$ | |
| df=24 | $\sigma_{\bar{x}} = 6$ | |

You cannot say that there is a 95% chance that the mean is within this CI.

You cannot say that the specific values in this specific interval in front of you are more likely to be correct than the values outside of it.

This is counterintuitive! I know! It is the frustrating thing about CIs.

# Showing you why

It is **very tempting** to make the logical leap from that the fact that the procedure will capture the population mean 95% of the time to the incorrect claim that there is a 95% chance that the mean is within a specific CI.

But we can show that this is a logical fallacy with a contrived example.

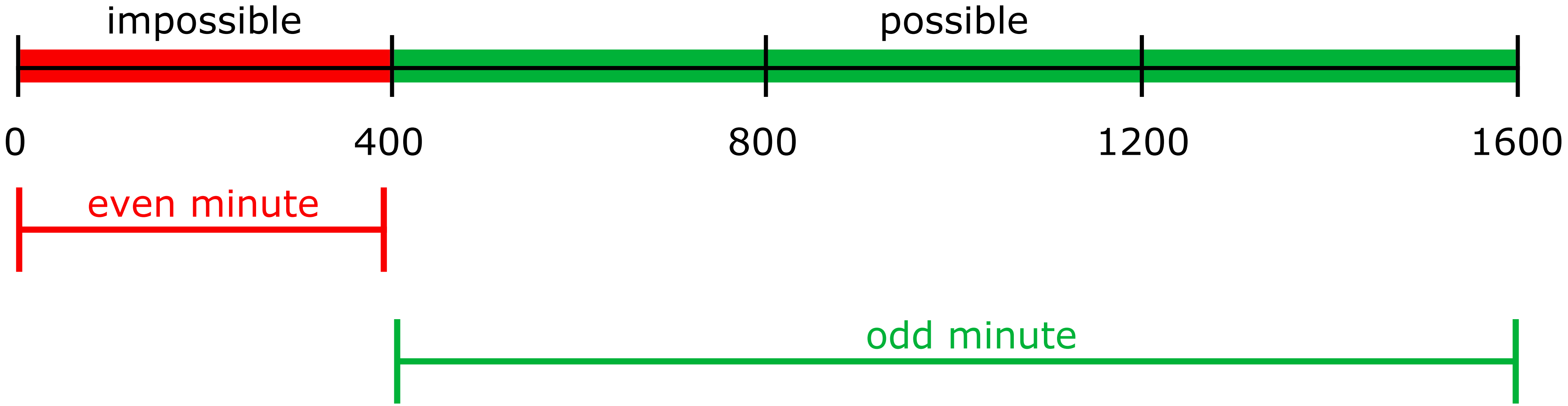
Let’s say that I want to draw an interval estimate for the mean SAT score at NYUAD. SAT scores can only range from 400 to 1600. That is how the test works.

Let’s say that I am allowed to take samples of students, and ask them their SAT scores. And imagine that I created the following interval procedure, called a **Jon Interval**.

| | |
|---|---|
| even minute | If the time on my watch ends in an even minute, the interval is 0-399 (lower = 0, upper = 399). |
| odd minute | If the time on my watch ends in an odd minute, the interval is the full range of the scale from 400-1600 (lower = 400, upper = 1600). |

# What is the probability of the procedure?

You can see that there are only two Jon Intervals: one that **never** contains the true mean of the SATs at NYUAD because it doesn’t contain real scores (0-399 are not possible scores!), and one that **always** contains the true mean because it contains all of the possible scores.



And you can also see that both have an equal probability of occurring over the long run (even vs odd minutes), so the long-run frequency that Jon Intervals will contain the true mean is:

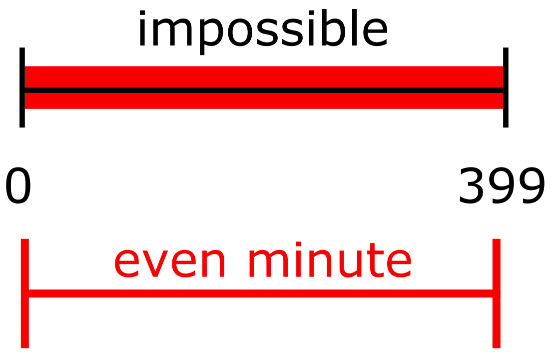
$$\frac{1}{2} = .5$$

In other words, Jon Intervals contain population mean 50% of the time. (Not good, but not bad either)

# What is the probability that a specific Jon Interval in front of me contains the mean?

Now let's look at the specific values of a specific interval sitting in front of us. There are only two possible intervals, so we can look at both of them in turn.

Let's start with the even-minute interval. And let's ask about the two claims that I told you <u>could not</u> be made about intervals.



Bad claim 1: that there is a 50% chance that the mean is within a specific CI in front of you.

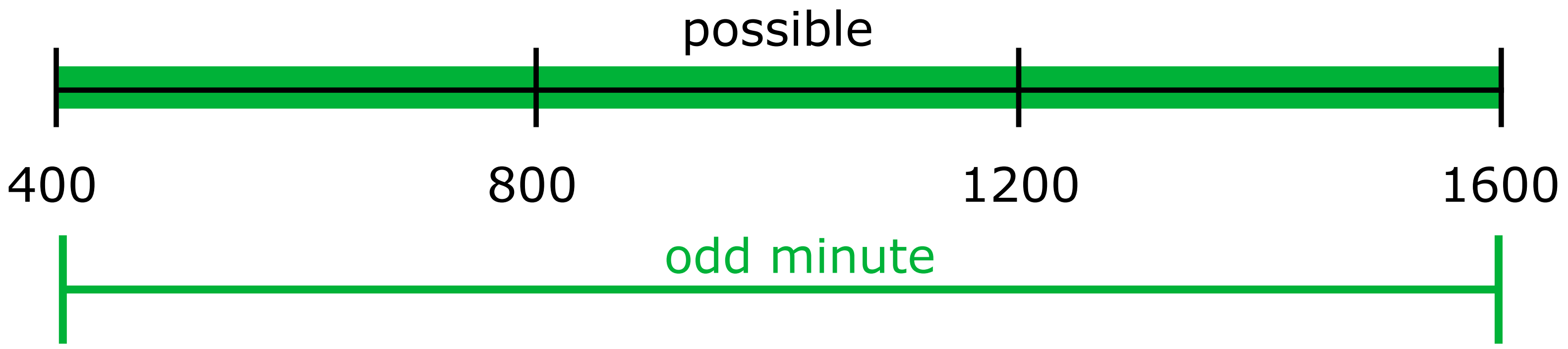
This is obviously **false**. We know that there is a 0% chance that the mean SAT score is in this interval because it does not contain possible scores.

Bad claim 2: that the specific values in this specific interval are more likely to be correct than the values outside of it.

This is also obviously false. These scores are <u>impossible</u>, so they are **less likely** than scores outside of it.

# What is the probability that a specific Jon Interval in front of me contains the mean?

Now let's look at the odd-minute interval:



Bad claim 1: that there is a 50% chance that the mean is within in a specific CI in front of you.

This is obviously **false**. We know that there is a 100% chance that the mean SAT score is in this interval because it contains **all** of the possible scores.

Bad claim 2: that the specific values in this specific interval are more likely to be correct than the values outside of it.

This is **true**, but in a completely uninformative way. This is a meaningless interval because it contains **all possible values**.

# What you can and can’t claim about CIs

The probability statement for confidence intervals (that it is 50% or 95% or whatever) is about the procedure of drawing the boundaries. It is not about the specific values contained within the boundaries.

What you can claim correctly: the procedure for drawing CIs means that 95% of the CIs drawn this way will contain the population mean.

What you can claim correctly: if a value is outside of the CI, a *t*-test would be significant at the equivalent alpha level (95% = .05 alpha, .025 in each tail) for a population mean of that value.

What you CAN'T claim: anything about the specific values in a single CI that is sitting in front of you.

You cannot say that there is a 95% chance that the mean is within this CI.

You cannot say that the specific values in this specific interval in front of you are more likely to be correct than the values outside of it.

# Most people make incorrect claims about CIs

It is very common for people to make incorrect claims about CIs:

People sometimes say there is a 95% chance that the value is within the specific CI in front of them.

People sometimes say that the values in the CI are more likely than ones outside it.

**What should you do?**

1. Remember probability with CIs is about frequentism - long-run relative frequency. It is not about belief (e.g., about the population mean).

2. Remember that CIs are about the process -- the process yields an interval containing the pop mean 95% of the time. It is not about a specific CI or a specific value in front of you. (Visualize the CI simulation)

3. Use CIs when you want to decide and declare that the population mean is within some interval (like stating your guess about a jar of jellybeans).

4. Use CIs when you want to determine which values would or would not be significant in a two-tailed one-sample *t*-test. CIs are *t*-tests.