Chapter 5: Confidence Intervals

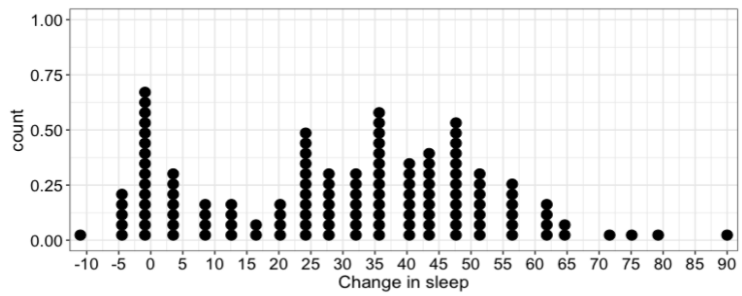
Investigation: Researchers would like to estimate the average change in sleep duration for adults trying a new experimental sleep aid. The researchers find that over the length of the study:

- The 132 participants report an average of 30.38 minutes more sleep a night while on the sleep aid.
- The standard deviation for change in sleep duration is 21.50 minutes.

Unit of observation:

Population of interest:

Variable (and type):



Testing vs. Estimating

- When do we test?
    - Hypothesis Testing examines one specific _____________ ______________ and helps us decide if it is probabilistically reasonable.
    - Is there evidence that this sleep aid increases sleep on average?
    - Is their evidence that this die is biased toward lower numbers?
- When would we estimate?
    - In some cases, we may wish to complement the hypothesis test question with an ___________ that communicates where the ______________ likely is.
    - By how much would we estimate nightly sleep increases while on this medication?
    - Approximately how much bias does this die seem to have?
    - It might also be that there just isn't a candidate parameter that is interesting to test—in which case, we might only present an interval estimate for the parameter!

How do we Estimate?

- Let's start by thinking about our best single guess for the parameter. In statistics, we often call this the _________ _____________.
    - Symbolically, the parameter we are trying to estimate in the sleep example is $\underline{\mu}$.
    - What is our point estimate for that parameter?

- But just the point estimate isn't all that helpful. Let's create an interval to express our _____________.
    - Let's start by calculating the standard error of our statistic to quantify about how much error we expect our estimate to have.

So then, we might say that we estimate the mean is ________ give or take about _______ minutes.

But *how many* standard errors we should extend to feel reasonably confident about our interval estimate?

- One approach to quantifying the likelihood of capturing the parameter is to make a _______________ assumption.
- Since our sample size is large, and since the data don't suggest a highly skewed variable distribution, let's assume that our *sample mean* comes from a distribution that is approx. normally distributed.
- Using properties of the normal distribution…

68% of possible $\bar{x}$'s will be within **1** standard error of μ

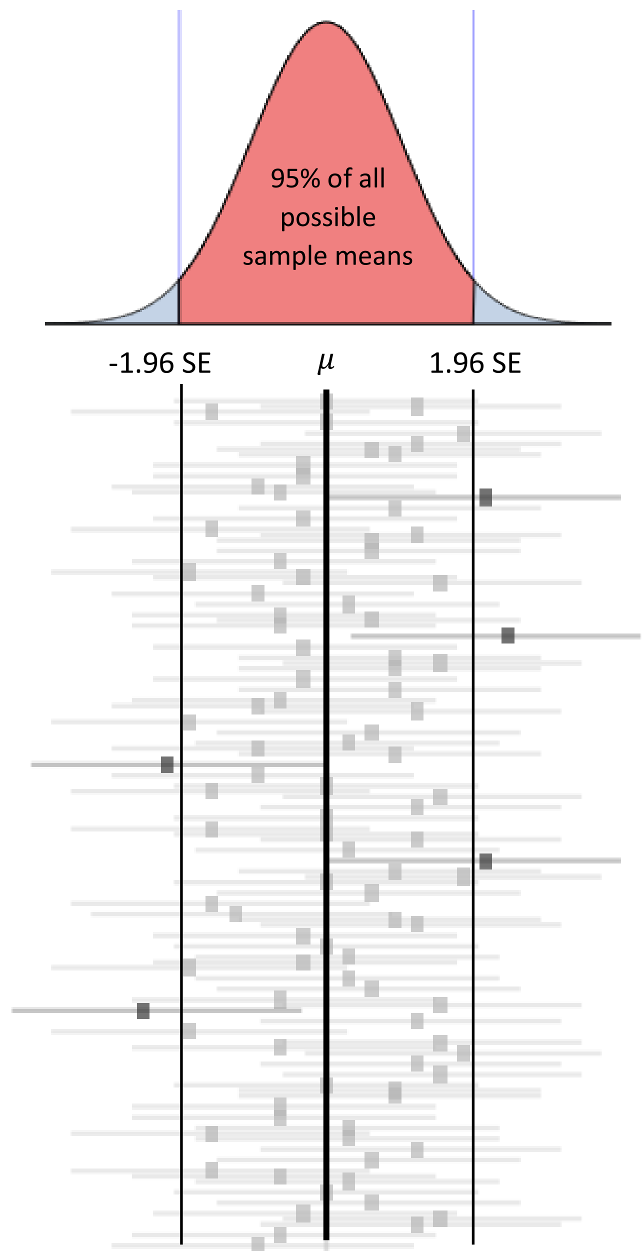
90% of possible $\bar{x}$'s will be within **1.645** standard errors of μ

95% of possible $\bar{x}$'s will be within **1.960** standard errors of μ

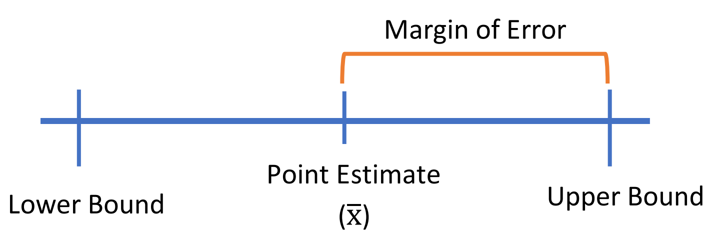
98% of possible $\bar{x}$'s will be within **2.326** standard errors of μ

99% of possible $\bar{x}$'s will be within **2.576** standard errors of μ

These will be provided to you in exam questions!



The actual distance we extend out from our point estimate is called the _______________________ and will be some number of standard errors in length.



Keep in mind that this method works if we assume our sample is **representative** of our population of interest. We'll talk about that more in Chapter 11!

Confidence interval for μ: $\bar{x} \pm z_{C\%} * SE_{\bar{x}}$

…where $z_{C\%}$ represents the number of ____________ _________ we would need to extend for a C% confidence level.

Let's now calculate a **95%** confidence interval for our parameter of interest in the sleep aid study.

Interpreting Confidence Intervals

When interpreting a confidence interval, we assume that our sample is representative of some population that we wish to generalize to. With this in mind, consider these two statements:

1. We are 95% confident that the average increase in sleep duration for these 132 people who took this sleep aid is between 26.7 and 34.0 minutes.
2. We are 95% confident that the average increase in sleep duration among all people who might take this sleep aid is between 26.7 and 34.0 minutes.

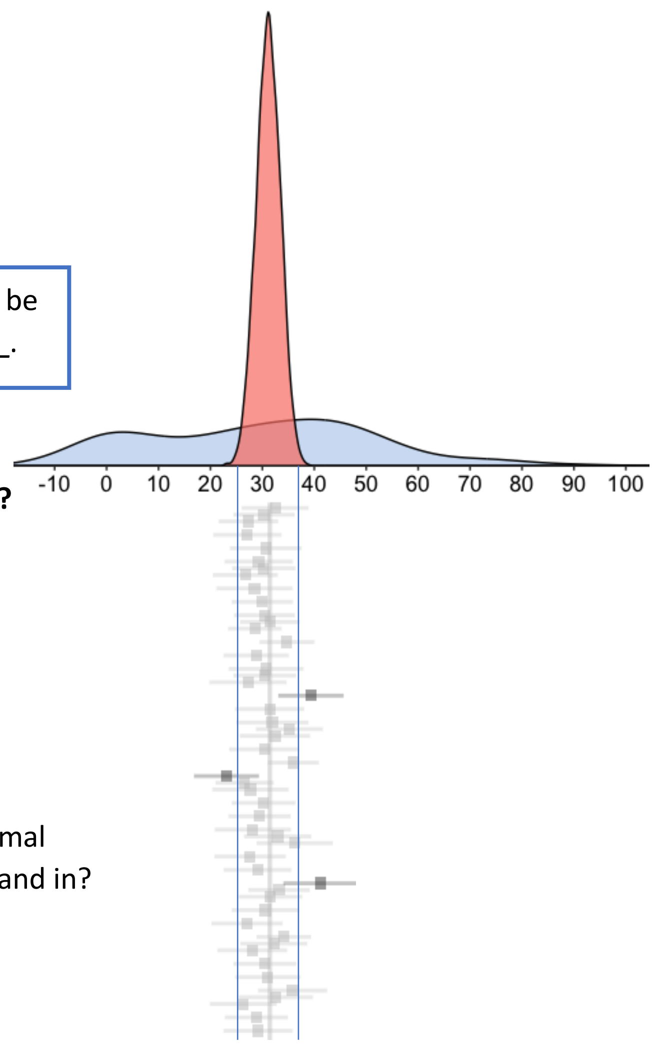
What difference do you notice between these wordings? Which one reflects the purpose of creating a confidence interval?

Now consider another statement:

3. We are confident that 95% of sleep aid users will report an increase in sleep duration of between 26.7 and 34.0 minutes

How is this statement different from the previous set?

When writing a confidence interval interpretation, we want to be clear that we are estimating the position of a ______________.



How would we approach answering a question like statement 3?

If the ***population*** we are sampling from happens to be ***normally distributed***, we could use z-scores and properties of the normal distribution like we did at the beginning of Chapter 4!

95% of individuals will land within $\mu \pm 1.96 * \sigma$

or we might approximate that with $\bar{x} \pm 1.96 * s$

Would this be a situation where we can use properties of the normal distribution to find the range that 95% of individual users would land in?

Digging Deeper: Why don't we say "95% chance" when we interpret a confidence interval?

- Saying something like "there is a 95% chance that the parameter is in our interval" subtly suggests that our parameter is a variable that moves around. However, the **parameter is fixed**!
- It's actually our ***interval*** that ***varies***. The position of our interval depends on the random sample and sample mean we happen to have. *See figures on previous two pages!*
- Once we have our interval, there is no "chance" process left. *Either it is in the interval, or it isn't!*
- For that reason, we want the probability in our statement to reflect…
  - The chance that we will get an interval that contains the parameter, rather than…
  - The chance that the parameter will be in the interval we have already calculated.

A more complete interpretation for a confidence interval would be: *"When following this approach, we are 95% confident that we will get an interval that will contain the parameter."*

Something like "We are 95% confident that the parameter is between X and Y" is *typically* considered acceptable, as we *could* now understand "confident" as referring to confidence in the process.

Reflection Questions

5.1. When building a CI, what name do we give to our single best estimate for a parameter?

5.2. What is the margin of error in the context of a confidence interval? How might you work backwards to calculate the margin of error given the interval bounds of a confidence interval?

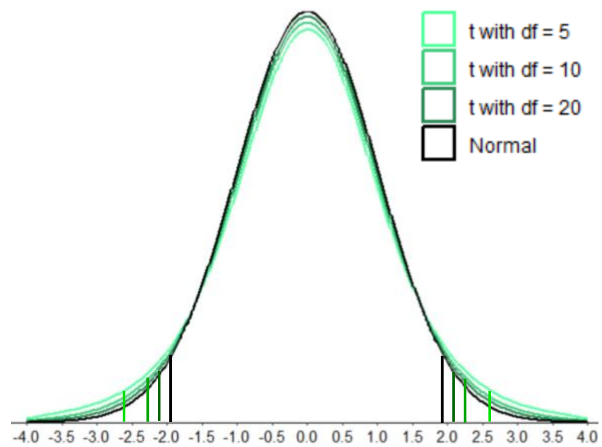
5.3. What is "weird" about this statement: "I'm 95% confident that my *sample statistic* will be inside my confidence interval." Does that statement capture the intent behind building a confidence interval?

5.4. **"**In a normally distributed population, 95% of individuals will be within about two standard errors of the mean." Do you agree with that statement? If not, how might you fix it to be correct?

T-intervals vs. Z-intervals for means

- As with hypothesis testing, we may often need to estimate σ with our sample standard deviation (s).
  - For large sample sizes, this discrepancy should be negligible, but at smaller sample sizes, we will want to make an appropriate adjustment to our method!

$$\mathrm{SE}_{\bar{x}} \approx \frac{s}{\sqrt{n}}$$



Building the t-interval

- Our standard error calculations will now *vary* (potentially quite a bit!) with each sample.
- If we still want to be 95% confident that our interval generating process will include the parameter, then we need to extend slightly more than 1.96 standard errors out from $\bar{x}$
- The number of approximated standard errors we now extend will be referred to as a **t-score** rather than a z-score.
- This t-score will depend both on the ______________ ________ and on the __________ ______ associated with that standard error calculation.
- Note again that when identifying t-distributions, we use the language of **degrees of freedom,** since its relation to sample size does change slightly in different situations.

t-interval for the Population Mean = $\bar{x} \pm t_{C\%} * s/\sqrt{n}$ (where t has df = n – 1)

z-interval for the Population Mean = $\bar{x} \pm z_{C\%} * \sigma/\sqrt{n}$ (σ known or well approximated)

- As an example, $t_{95\%} > 1.96$, but the magnitude of difference depends on how large a sample size we have.
  - For df = 5, $t_{95\%} = 2.571$. For df = 10, $t_{95\%} = 2.228$. For df = 20, $t_{95\%} = 2.086$
  - …as n increases, $t_{95\%}$ will approach _______ as "s" better approximates "σ"
- Using the **t distribution applet** from the art of stat web apps page, we can…

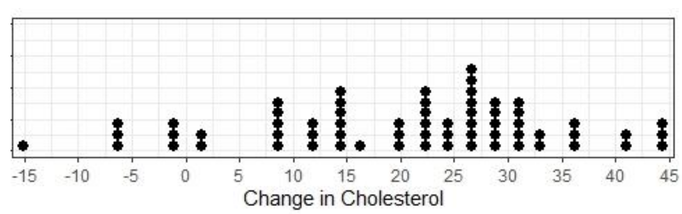
Find the t-score needed to create a 95% confidence interval for a mean using a sample of size 25.	Find the t-score needed to create a 99% confidence interval for a mean using a sample of size 51.

Assumptions for Creating z or t-intervals for a Mean

- z-intervals are appropriate if…
  - ✓ σ is known or well approximated (perhaps from using a very large sample)
  - ✓ The distribution of $\bar{x}$ is normally distributed.
    - Population is normally distributed, or sample size large enough for CLT to apply.
    - **For this class**, we'll say $n \geq 30$ is a good benchmark when we have no reason to think there is a large skew, and that we need n very large ($n \geq 100$) in cases where there is a large skew.
- t-intervals are appropriate if…
  - ✓ σ is being estimated by s
  - ✓ The distribution of $\bar{x}$ is normally distributed.
    - Population is normally distributed, or sample size large enough for CLT to apply.
    - **For this class**, we'll say $n \geq 30$ is a good benchmark when we have no reason to think there is a large skew, and that we need n very large ($n \geq 100$) in cases where there is a large skew.
- **And in general…**we assume our sample is representative of the population we are generalizing to!

Investigation: 64 patients who are considered representative of the high-cholesterol population over 50 are trying new medication for cholesterol. They had an average cholesterol drop of 19.0 mg/dL with a standard deviation of 14.3 mg/dL.

What type of confidence interval should we create? Are assumptions met?



Create a 95% confidence interval for the average cholesterol drop of patients over 50 with high cholesterol. *According to the t distribution applet, $t_{95\%}$ when df = 64 is **1.998***

Point Estimate

Margin of Error

Confidence interval

If you had a patient over 50 with high cholesterol, could you use this interval to say that you are 95% sure *his* cholesterol drop would be between these two values?

Estimating Proportions

Honeybee Dance Party! Honeybees sometimes dance when they return to the hive to indicate the presence of nectar for the hive and where to find that nectar. Might cocaine affect their dance frequency? And what might this tell us about the effects of cocaine on insects?

Researchers from the Australian National University, in partnership with the University of Illinois, observed how being dabbed with cocaine might affect the dance behavior of honeybees. They observed a group of honeybees taking a combined 51 trips to and from the hive over the course of an hour (some bees made 0 trips, and some made multiple trips). Approximately 39 of these trips resulted in a dance by that particular honeybee. Findings are reported in this HoneyBee Article

Unit of observation:

Variable (and type):

Parameter of interest:

Setting up the Basics again!

- Symbolically, the parameter we are trying to estimate in the bee example is
- What is our point estimate for that parameter?

- Once again, the point estimate may have some error! So let's estimate about how much error we expect in our point estimate by calculating the standard error.
- One change here though…when we did hypothesis testing, we could plug in our null hypothesized proportion to avoid the $s \approx \sigma$ uncertainty that we run into with numeric data.
- Since we're estimating rather than hypothesizing, we'll need to use an approximated formula with $\hat{p}$.

$$SE_{\hat{p}} \approx \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} =$$

The z-interval (Wald interval) option

- Since the distribution of possible sample proportions tends to follow a normal distribution as n increases, let's first try a z-interval method to estimate the proportion. Let's calculate a **90%** confidence interval using this approach! ($z_{90\%} = 1.645$)

Confidence Interval (Wald Interval) for π: $\hat{p} \pm Z_{C\%} * SE_{\hat{p}}$

Exploring Confidence Intervals: Open the **Explore Coverage Simulator** from the Art of Stat Web apps and let's make some observations about how confidence intervals behave.

Stay on the **Confidence Interval for a Proportion** tab up top.

1. What do you notice when you increase the level of confidence for your intervals? *Which part of the equation is affected by this?*

2. What do you notice when you generate confidence intervals from larger sample sizes? *Which part of the equation is affected by this?*

3. When we talk about coverage with confidence intervals, we're checking if the percentage of intervals we generate that contain the parameter actually matches the confidence level we chose! Under what situations does the coverage of Wald Intervals seem to be rather poor?

Reflection Questions

5.5. When estimating a mean, when might we need to construct a t-interval rather than a z-interval?

5.6. In general, when would a parametric interval estimation method (t or z) be appropriate for a mean?

5.7. As our sample size increases, how is the width of a confidence interval typically affected?

5.8. If we specify a higher confidence level, how is the width of a confidence interval typically affected?

Confidence Intervals for a Proportion Revisited

- As with z-tests for a proportion, a z-interval (Wald interval) approach only works well in ___________ sample situations.
  - Typically, we would want a minimum sample size of $n \geq 100$
  - Also, the proportion shouldn't be too close to 0 or 1. One way to check that is with a modified 10/10 rule: check that there are at least 10 "successes" and 10 "failures" in our data.

- Otherwise (as with the coverage simulation), it tends to underperform the confidence level we set!
- The **Clopper-Pearson interval** represents one option, based on properties of the ___________ distribution!
  - Essentially, we work backwards to consider what value the true proportion might take C% of the time to produce this number of successes.
  - However, the Clopper-Pearson interval tends to be too _______________ at lower sample sizes, meaning that the interval will often be wider and capture the parameter ______ _________ than we request with our confidence level!

- The **Agresti-Coull interval** is another z-interval approach that makes a simple adjustment to the Wald interval. It results in fairly consistent coverage at both smaller and larger samples!
  - The idea is simple—since Wald intervals underperform when the sample proportion is too close to 0 and 1, then simply add 2 successes and 2 failures to the calculation.

So rather than building our interval from $\hat{p} = \frac{x}{n}$ We will instead calculate $\tilde{p} = \frac{x+2}{n+4}$

- Likewise, we'll use $\tilde{p}$ in the numerator and $\tilde{n} = n + 4$ in the denominator of our SE calculation.
- By making this +2 success and +2 failure adjustment, the coverage of our interval will improve at smaller samples while having very little effect at larger samples.

Return to our Honeybee Dance party.

How might we instead use an Agresti-Coull interval to estimate the proportion of cocaine-influenced trips that result in dances by these honeybees with 90% confidence?

$\tilde{p}$

$SE_{\tilde{p}}$

Margin of Error:

Confidence Interval:

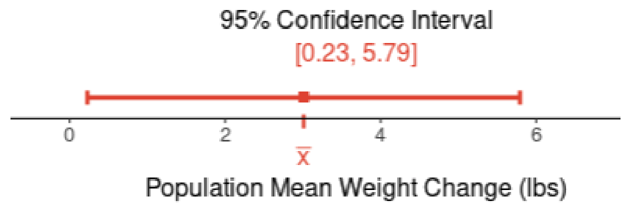
For more info about confidence interval options and coverage, here is a **great article**!

Relationship between Confidence Intervals and Hypothesis Testing

Let's take a look at the **Inference for a Population Mean** applet on the Art of Stat Web apps page. Choose the **Weight Change** dataset from the dropdown.

This data represents an investigation to estimate weight change for participants after completing a muscle building program. Shown first is a 95% confidence interval to estimate what the average change might be at a population level based on what we observed in this representative sample of 29 people.

Might this interval give us an indication about whether the average weight change might reasonably be 0?



If our 95% confidence interval ___________ include the null, then the p-value from a non-directional test using that null would be _____ than…

One way to think about this relationship is that confidence intervals represent the set of values we could set as our null hypothesis that we would ________________ when α = 100 – C%.

Reflection Questions

5.9. In what situations are Wald intervals especially unreliable for estimating a proportion (or rather, in what situations are they at least somewhat reliable)?

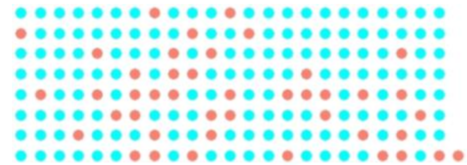
5.10. Why is an Agresti-Coull interval generally considered a better choice than a Wald interval? How is it constructed differently?

5.11. We're conducting a hypothesis test to determine whether the mean gas price in Chicago changed since the last election (it was officially recorded as $3.36). After collecting gas prices from a representative sample of 46 Chicago gas stations and completing a two-sided hypothesis test, we got a p-value of 0.021. If we were to use this same data to construct a 95% confidence interval for the true mean, would we expect $3.36 to be inside that interval? What if we created a 99% confidence interval?

Chapter 5 Additional Practice (Videos available in the Ch 5 module on Canvas!)

Practice: A particular variant of the BRCA1 gene has been linked to breast cancer. Researchers are trying to estimate the true risk of breast cancer among women with this gene variant by age 60.

185 women with this gene variant were studied over time, and 43 of them developed breast cancer by Age 60. Assume these 185 women are a representative sample of women with this gene variant.



Symbolically, what parameter are we trying to estimate? What is our point estimate for that parameter?

Using an Agresti-Coull interval, calculate the adjusted standard error using $\tilde{p}$ and $\tilde{n}$

Using an Agresti-Coull method, find a **98%** confidence interval for the proportion of women with the BRCA1 gene variant who will eventually develop breast cancer by age 60.

In drafting an abstract for a conference, your co-author writes "We are 98% confident that women with this BRCA1 variant will develop breast cancer by age 60." What's wrong with this statement? How could you rewrite it?

In related research, someone finds a 98% confidence interval for the risk of ovarian cancer based on a different gene variant. The interval is (0.084, 0.131). What is the point estimate and margin of error for this interval?

Practice: How long does it take semi-truck drivers to deliver supplies to a Champaign store when driving from a Chicago warehouse?

Let's say that we take a sample of 40 trips and recorded the amount of time in minutes it took them to complete the drive. The average drive time for that sample was 168 minutes with a *sample standard deviation* of 9 minutes.

Create a 95% confidence interval for the true average drive time. *According to the t-simulator, we'll need to use a t-score of 2.023.*

Now consider if we had created **98%** confidence interval with the same data.

First, **predict**. Which do you think will be wider, this 98% confidence interval or the 95% confidence interval from before? *Think conceptually or mathematically about what is changing here!*

Now, calculate the 98% confidence interval for the sample of 40. *Use t = 2.426*

Now let's say that we just took a sample of **20 trips**. Their average drive time was also 168 minutes with sample standard deviation 9 minutes.

First, **predict.** If we created a 95% confidence interval for the true average drive time, would it be larger or smaller than the interval using 40 trips? *Think conceptually or mathematically about what is changing here!*

Now check your work by calculating this new confidence interval. *Use t = 2.093*