Chapter 5: Confidence Intervals

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:

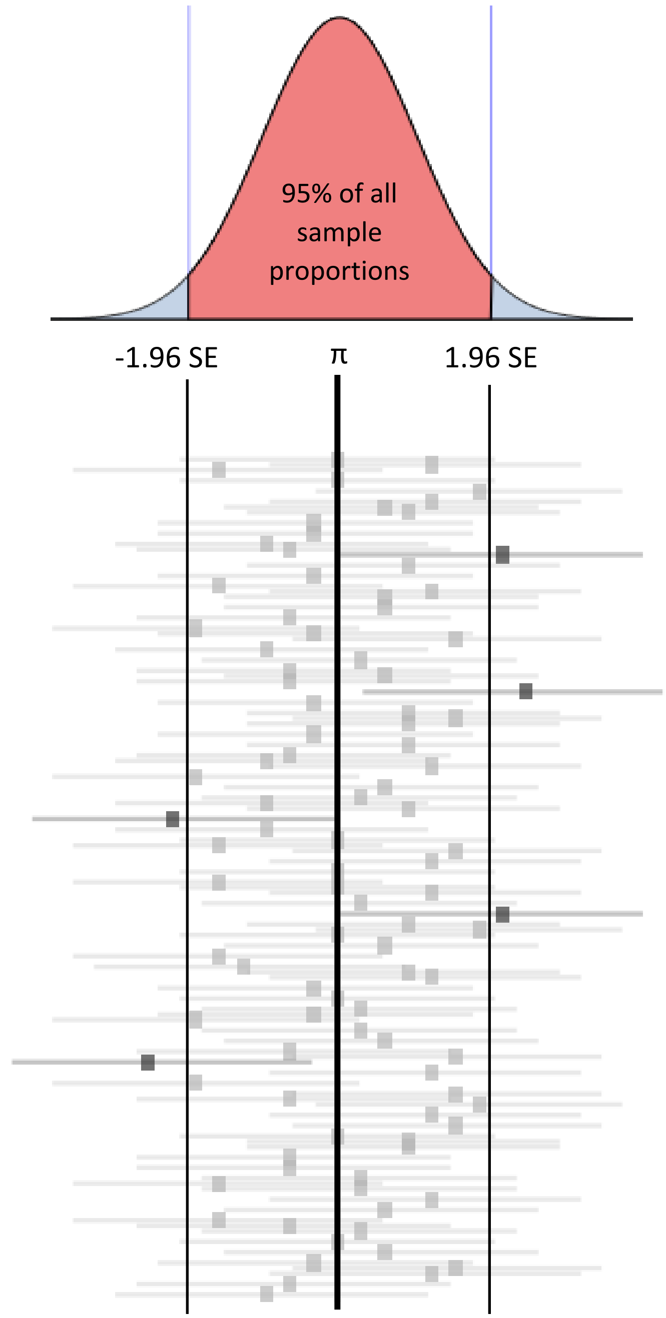
Testing vs. Estimating

- When do we test?
  - Hypothesis Testing examines one specific ______________ parameter and helps us decide if it is probabilistically reasonable.
  - Is there evidence that more than 50% of trips with cocaine result in a dance?
  - Is there evidence that the rate of dancing has increased? (comparing two proportions)
- When would we estimate?
  - Sometimes, we may not have a candidate parameter to test.
  - But even if we do test a candidate parameter, interval estimates can *complement* p-values by showcasing how far the true parameter likely is from that benchmark.
  - For example, p-values may help us determine how confident we are that a coin is biased, while a confidence interval helps us estimate how large that bias likely is.

How do we Estimate?

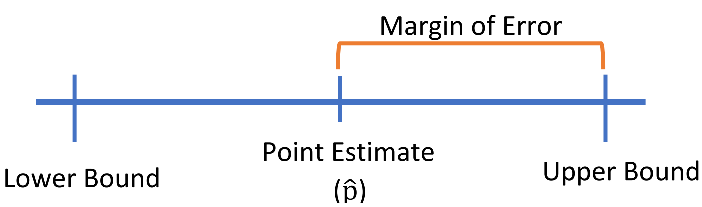
- Let's start by thinking about our best single guess for the parameter. In statistics, we often call this the ______________________.
  - Symbolically, the parameter we are trying to estimate in the bee example is $\boldsymbol{\pi}$.
  - What is our point estimate for that parameter?
- But just the point estimate isn't all that helpful. Let's create an interval to express our ______________.
  - Let's start with the ___________________ of this statistic as an estimate for the parameter. Let's calculate it below.

But *how many* standard errors we should extend out from our point estimate in building this interval?



- If we make a “parametric” assumption that $\hat{p}$ is normally distributed around π, then…
  - **90%** of all possible $\hat{p}$’s will be within **1.645** standard errors of π
  - **95%** of all possible $\hat{p}$’s will be within **1.960** standard errors of π
  - **98%** of all possible $\hat{p}$’s will be within **2.326** standard errors of π
  - **99%** of all possible $\hat{p}$’s will be within **2.576** standard errors of π

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 π: $\hat{p} \pm Z_{C\%} * SE_{\hat{p}}$

Z represents the number of ___________________ we would need to extend for a C% confidence level. There is then a 1 – C% probability that our confidence interval will **not** contain the parameter.

If trying to determine the z-score needed to construct any particular confidence interval, we could use the **normal distribution simulator** from the art of stat web apps page

- Go to “Find Percentile/Quantile”
- Take a two-tailed percentile
- Type in your desired confidence level

Assumptions for creating a z-interval for a proportion

Z-intervals for a proportion (also referred to as "Wald intervals") are fairly reliable in many contexts, but they do depend on a few assumptions

- ✓ The distribution of possible sample proportions is normally distributed
  - **The 10/10 Rule**: This is a reasonable assumption to make if our sample has **at least 10 of each response (e.g., >=10 "yes" and >=10 "no" responses)**
  - When this isn't true, the distribution of $\hat{p}$ might have some skew, or be too discrete to reasonably use a normal approximation.
- ✓ $\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$ is a reasonable estimate of $SE_{\hat{p}} = \frac{\sqrt{\pi(1-\pi)}}{\sqrt{n}}$
  - When we have a fairly large sample size (n >100), then this doesn't really matter. The difference due to using $\hat{p}$ will be very tiny.
  - When n is smaller, then Wald Intervals are typically "ok" as long as the 10/10 rule above still holds. But there are other methods too!
- ✓ **And in general…**we assume our sample is representative of the population we are generalizing to!

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.

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

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

Does the accuracy/coverage of Wald Intervals (z-intervals) change when the sample size is smaller, or when the true proportion is closer to 0 or 1?

Return to our Honeybee Dance party. Would a z-interval ("Wald Interval") be reasonably reliable to construct with the data we have?

What is the standard error for our point estimate?

What would be the 95% Margin of error for our point estimate?

Calculate the lower and upper bound for a **95%** confidence interval.

Let's say that the typical rate of honeybee dancing after a trip to and from this hive *without* cocaine was known to be 50%. How might our confidence interval give us an indication about whether the cocaine-laced rate of dancing might truly be higher than 50% or not?

If our 95% confidence interval _______________ include the null, then the p-value if testing that null would be __________ than ___________________

We can also see this by asking how many standard errors away from the null our sample proportion is.

Reflection Questions

5.1. How might a confidence interval (CI) provide additional information that we don't get from a p-value?

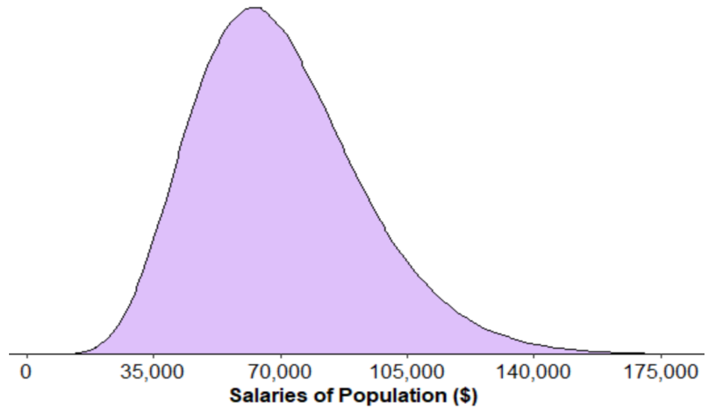
5.2. When building a CI, what name do we give to our single best estimate for a parameter?

5.3. 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?

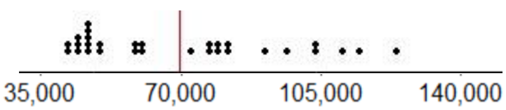
5.4. As our sample size increases, how is the width of a confidence interval typically affected?

Confidence Intervals for a Mean: μ

Investigation: We would like to examine the starting yearly income for students who have graduated from the U of I with a bachelor's degree in Statistics. Let's say that this distribution is well represented by the density curve below. This is a positively skewed distribution with μ = \$70,000 and σ = \$23,000.

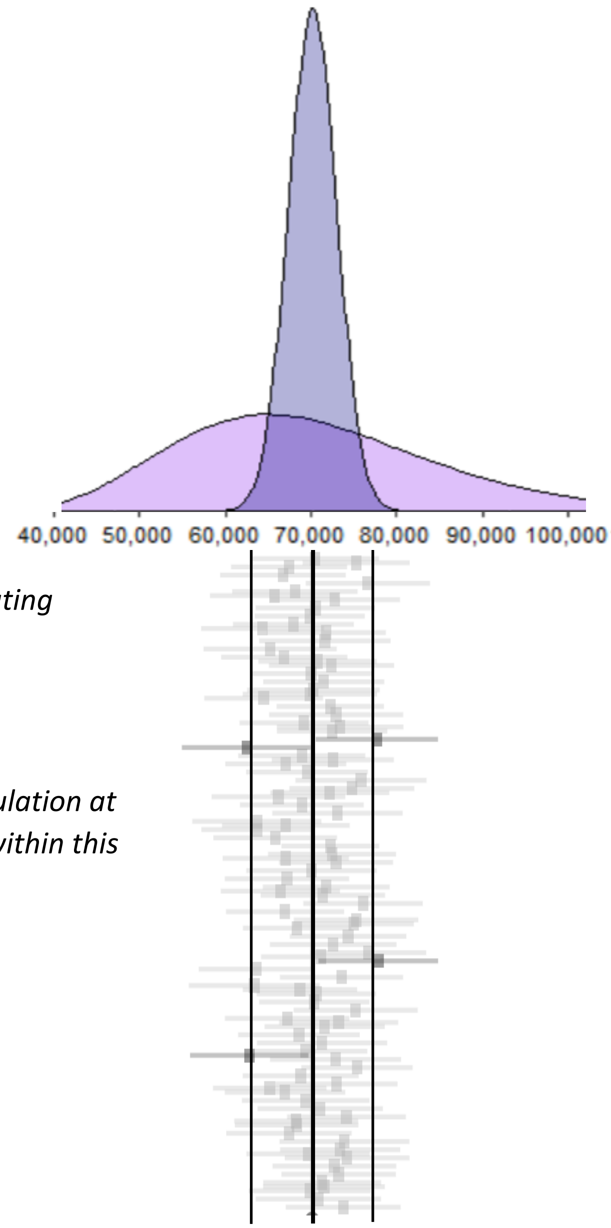


Now, let's imagine we sampled 30 recent graduates randomly from this population.



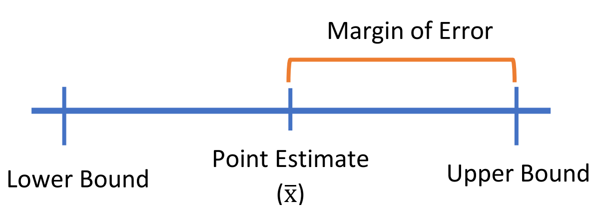
Let's consider how might follow a similar process to estimate μ using our sample information!

- Our **point estimate** represents our best single estimate for the parameter.
  - *Symbolically,* what is our point estimate for μ? $\underline{\bar{x}}$
- The **standard error** represents the expected error in our point estimate.
  - What is the standard error in our statistic as an estimate for this parameter?



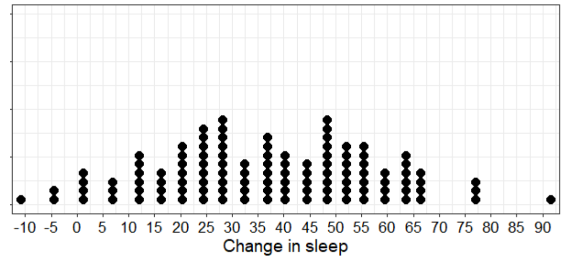
- Finally, let's consider *how many* standard errors we should extend out from our point estimate. This will help us determine the **margin of error** for our interval.
  - What would be our margin of error if we want 95% confidence in capturing μ? *Assume $\bar{x}$ is distributing normally.*

Consider this statement: *"If we choose a **graduate** from this population at random, we are 95% confident that **their starting salary** will be within this margin of error from \$70,000."* Is this a correct interpretation?



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 131 participants report an average of 36.14 minutes more sleep a night while on the sleep aid.
- The standard deviation for change in sleep duration is 18.00 minutes.



Assuming this is a representative sample of adults, use this information to **build a 98% confidence interval** for the average change in sleep duration for adults on this sleep aid.

Identify our point estimate for μ

Identify the standard error in our point estimate

Calculate the margin of error we will need if we want to be 98% confident in capturing the true mean change in sleep as a result of this medication. If done correctly, we should find the following interval.

(32.48, 39.80)

Interpreting Confidence Intervals

Which statement below correctly interprets this confidence interval in context?

I am 98% confident that an individual who takes this medication will experience between a 32.48 and 39.80 minute increase in their sleep.

I am 98% confident that the average change in sleep duration among all individuals who may take this medication is between 32.48 and 39.80 minutes.

I am 98% confident that the average change in sleep duration among these 131 individuals is between 32.48 and 39.80 minutes.

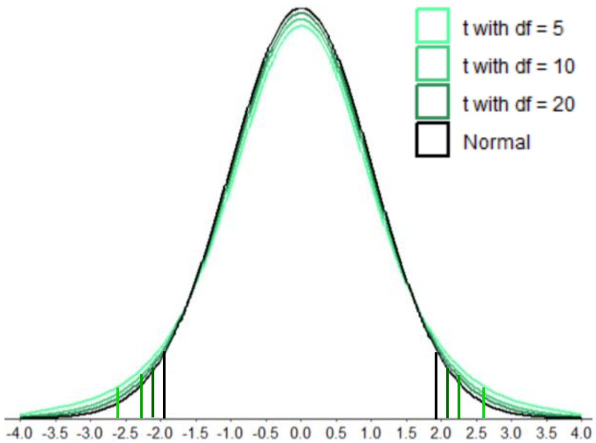
Confidence intervals are designed to identify the position of a ________________.

- Accounting for the estimation error when using s in place of σ.
  - When we calculate the standard error for the mean, we typically need to calculate it using the standard deviation of our sample (s) since we likely won't know σ.

$$\mathrm{SE}_{\bar{x}} \approx$$

  - 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!

- **Calculating t-intervals using the t-distribution**
  - Our standard error calculations will now *vary* (potentially quite a bit!) with each sample.
  - If I still want to be, say, 95% confident that my interval includes the parameter, then I might 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 standard error calculation.

t-interval for the Population Mean = $\bar{x} \pm t * s/\sqrt{n}$

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

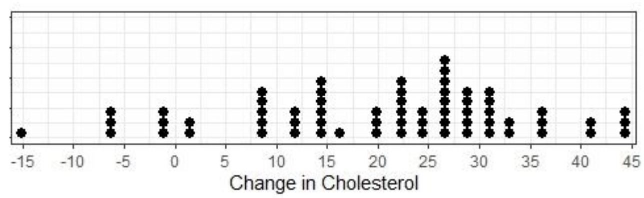
  - As an example, $t_{95\%} > 1.96$, but the magnitude of difference depends on how large a sample size we have.
    - For df = 5, $t_{95\%} = 2.571$. For df = 10, $t_{95\%} = 2.228$. For df = 20, $t_{95\%} = 2.086$
    - …as n increases, $t_{95\%}$ will approach 1.96 as "s" better approximates "σ"
  - Using the **t distribution simulator** 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 <u>Mean</u>

- ✓ z-intervals are appropriate if…
  - σ is known or well approximated (perhaps from using a very large sample)
  - The distribution of $\overline{x}$ is normally distributed.
    - Population is normally distributed, or sample size large enough for CLT to apply.
    - **For this class**, we'll say n > 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>100) in cases where there is a large skew.
- ✓ t-intervals are appropriate if…
  - σ is being estimated by s
  - The distribution of $\overline{x}$ is normally distributed.
    - Population is normally distributed, or sample size large enough for CLT to apply.
    - **For this class**, we'll say n > 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>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?

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

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?

Reflection Questions

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

5.6. What's a quick thing we might check with our data to determine whether a Wald interval (z-interval for a proportion) would be a somewhat reliable interval estimation method?

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

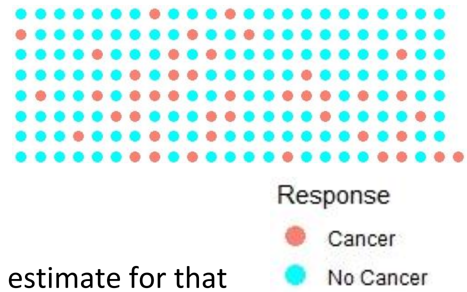
5.8. In general, when would a parametric interval estimation method (t or z) be appropriate for a mean?

5.9. 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?

Chapter 5 Additional Practice (if you need it!)

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.



Is it appropriate to create a Wald Interval in this context?

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

What is the expected error (standard error) in our point estimate as an estimate for the parameter?

Find a 95% 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 95% 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 95% 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*