

# Central Limit Theorem and Confidence Intervals - Introduction



<https://github.com/learn-co-curriculum/dsc-confidence-intervals-intro-v2-2>



<https://github.com/learn-co-curriculum/dsc-confidence-intervals-intro-v2-2/issues/new/choose>

## Introduction

In this section, you'll be introduced to inferential statistics. In particular, we'll start with discussing techniques for describing the (un)certainty of a given numeric estimate in a formal, rigorous way.

## Managing Uncertainty When Reporting Numbers

Let's start with an example: you work for a furniture company that wants to develop a desk chair specifically designed for 12-year-old children. Your boss wants to know how tall the back of the chair should be, to accommodate this customer segment.

Obviously you are not able to get measurements for all 12-year-olds in the history of time. In fact, the actual children who are likely to use this chair are not even 12 years old yet, since the chair hasn't been developed and manufactured yet! So, it won't be possible to measure your population of interest directly. Instead, you'll need to use sampling and statistical knowledge to report a measurement to your boss. This is a form of **inferential** statistics, because you are drawing conclusions beyond just describing statistics of your sample data.

Let's say you perform some ergonomics studies on a sample of subjects and conclude that, on average, 15 inches is the best chair back height.

In addition to the measurement itself, you'll want to provide some additional information about uncertainty. Given the information you have collected, **how confident are you** that 15 inches is the appropriate measurement?

- One way you could hedge and avoid giving the "wrong" answer would be to say that "I am confident that the best chair back height is between 0 and 1 million inches". You would be technically correct, but that is not actually useful to your boss!
- On the other hand, if you arbitrarily choose some interval, like "14-16 inches", that might sound more reasonable, but you still wouldn't be able to ground that statement in actual statistical analysis.

**Confidence intervals** are a tool that allows us to give a more formal answer to this question, and they form the basis for the statistical testing (hypothesis testing) covered in the next section.

## Concepts in This Section

In order to understand and apply confidence intervals, this section covers three areas: the Central Limit Theorem, confidence intervals, and the t-distribution.

## The Central Limit Theorem

The Central Limit Theorem allows us to treat non-normal distributions as normal distributions, and provides a way for us to estimate parameters about a population.

## Confidence Intervals

A confidence interval is a range of values surrounding an estimated parameter. The width of the range depends on the variance of the data (more variance results in a wider confidence interval, less variance in a narrower confidence interval) as well as the **confidence level** (a higher confidence level results in a wider confidence interval, lower confidence level results in a narrower confidence interval).

Going back to the example above, based on the variance of the heights of 12-year-olds from our study as well as a 90% confidence level, we might produce an estimate like "12-18 inches". That is still a fairly wide range to work with, but it's more meaningful than saying "0 to 1 million inches" (which we are confident includes the right answer) but also communicates the true uncertainty better than a randomly-chosen "14-16 inches" would. That range of 12-18 inches is the **confidence interval**.

## The t-Distribution

Initial confidence intervals we'll construct will use the z-distribution introduced previously. This technique is somewhat limited because it requires that we know the standard deviation of the *population*.

The t-distribution allows us to work with samples where the population standard deviation is unknown (as well as smaller samples), in order to form confidence intervals.

## Summary


Inferential statistics are powerful because they allow you to make claims about data that you don't have access to! A key aspect of making these kinds of claims is communicating your uncertainty appropriately, and confidence intervals can help. Once you have mastered this concept, you will

move on to hypothesis testing and modeling techniques, which utilize the same general approach of making claims and communicating uncertainty about unknown values.

How do you feel about this lesson?



Have specific feedback?

**Tell us here!**  <https://github.com/learn-co-curriculum/dsc-confidence-intervals-intro-v2-2/issues/new/choose>