

10.19: Estimating a Population Mean (2 of 3)

Learning Objectives

- Construct a confidence interval to estimate a population mean when conditions are met. Interpret the confidence interval in context.
- Interpret the meaning of a confidence level associated with a confidence interval.
- Adjust the margin of error by making changes to the confidence level or sample size.

On the previous page, we used a confidence interval to estimate the population mean, μ . For this confidence interval, we had to supply a guess for the population standard deviation, σ , based on previous studies. It may have occurred to you that if we do not know μ , it is unlikely that we know σ . So we now take a different approach. We estimate σ using the sample standard deviation, s .

This is the same type of adjustment we used in *Inference for One Proportion* when we had to adjust our model of the sampling distribution. The standard error of the sampling distribution is $\sqrt{p(1-p)/n}$. (If we knew p , then we wouldn't need to build a confidence interval.) We approximate p by the sample proportion, \hat{p} .

Our process for adjusting the confidence interval for estimating μ is similar. We use the sample standard deviation, s , to estimate σ . The standard error for the sampling distribution σ/\sqrt{n} becomes s/\sqrt{n} . So we adjust the margin of error in the confidence interval formula, but this adjustment is not as straightforward as our work with proportions.

This estimate for σ introduces more uncertainty in the process. The problem is worse with smaller samples because the sample standard deviations vary more. For small samples, s is a worse approximation for σ . Unfortunately, this approximation makes the normal model a bad fit and inappropriate for determining critical values. We instead use what is called a **t-model** for this purpose.

Introduction to the T-Model

Here is the formula for the T-score. We also include the z-score for comparison. The formulas are very similar.

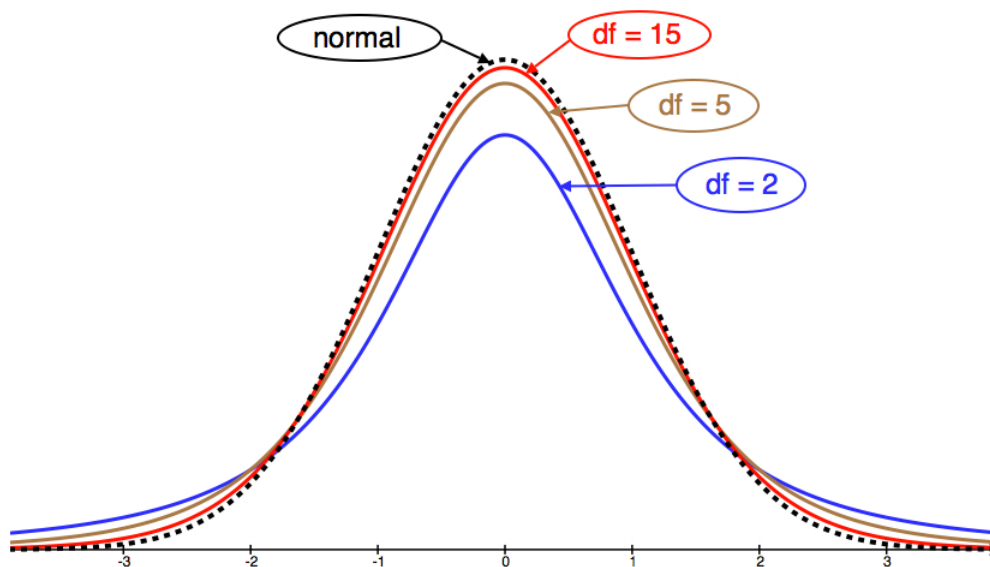
$$Z\text{-score} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad T\text{-score} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

The distribution of z-scores is the standard normal curve, with mean of 0 and standard deviation of 1. The distribution of T-scores depends on the sample size, n . There is a different T-model for every n . So the T-model is a family of curves.

Instead of referring to n to specify which T-model to use, we refer to the **degrees of freedom**, or df for short. For Topics 10.2 and 10.3, the number of degrees of freedom is 1 less than the sample size. That is, $df = n - 1$.

In summary, a normal model is defined by its mean and standard deviation. A T-model is a family of curves defined by the degrees of freedom.

Let's take a look at a few T-model curves (for various df) to see how they compare to the normal model.



We can see from the picture that as df grows, the T-model gets closer to the standard normal model.

Similarities between T-model and standard normal model:

- Symmetric with a central peak, bell-shaped.
- Centered at 0.
- The larger the degrees of freedom, the closer the T-model is to the standard normal model.

Difference between T-model and standard normal model:

- The T-model has more spread than the standard normal model.
- The T-model has more probability in the tails and less in the center than the standard normal model. We can see this in the fatter tails and lower central peak of the T-model.

When is a T-model a good fit for the sampling distribution of sample means?

Check these conditions before using the T-model:

- Use the T-model if σ (the population standard deviation) is unknown. If σ is known, then use the normal model instead of the T-model.
- Use the T-model if variable values are normally distributed in the population. If this is not true, then make sure the sample size is large (more than 30).

Example

Cable Strength

A group of engineers developed a new design for a steel cable. They need to estimate the amount of weight the cable can hold. The weight limit will be reported on cable packaging.

The engineers take a random sample of 45 cables and apply weights to each of them until they break. The mean breaking weight for the 45 cables is $\bar{x} = 768.2$ lb. The standard deviation of the breaking weight for the sample is $s = 15.1$ lb.

What should the engineers report as the mean amount of weight held by this type of cable?

Let's use these sample statistics to construct a 95% confidence interval for the mean breaking weight of this type of cable.

Checking conditions:

Since we do not know the standard deviation of breaking weights of all of the cables (the population parameter σ), we use the sample standard deviation (s) as an approximation for σ . Since we don't know σ , we must use the T-distribution to model the sampling distribution of means.

Is the T-model a good fit for the sampling distribution?

Yes, because the conditions are met:

- σ is unknown.
- The sample size is large enough.

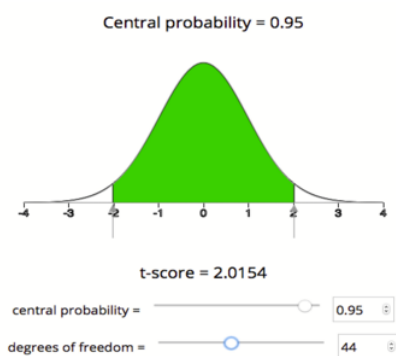
Finding the standard error:

As usual, we start by estimating the standard error. This estimate comes from the formula $\frac{s}{\sqrt{n}}$. However, since we don't know σ , we use $s = 15.1$ as an approximation for σ . So our estimate for the standard error of all sample means is $\frac{s}{\sqrt{n}} = \frac{15.1}{\sqrt{45}} \approx 2.25$.

Finding the margin of error:

To find the margin of error, we need to find the critical T-value that corresponds to a 95% confidence level. This is just like the critical Z-value when we built confidence intervals for proportions, except that it comes from the T-model instead of the standard normal model.

We will use technology to find the critical T-value. There are a number of tools for doing this. Some books will also give you the option to use printed tables of values. Here we will use a simulation that gives the T-model based on degrees of freedom. We want the T-values that cut off the central 95% of the area under the curve. It will look as follows.



Using the simulation, we see that the critical T-value for a 95% confidence interval with 44 degrees of freedom is $T_c = 2.015$, which means our margin of error for this confidence interval is

$$\begin{array}{l} \text{Estimated standard error} \\ \text{Margin of error} \end{array} = \begin{array}{l} \frac{s}{\sqrt{n}} \approx 2.25 \\ T_c \cdot \frac{s}{\sqrt{n}} \approx 2.015(2.25) \approx 4.53 \end{array}$$

Note: For 95% confidence, the empirical rule approximates the critical Z-value as 2. The empirical rule is based on the normal model. Using the T-model for $df = 44$, the critical T-value (2.015) is very close to 2. This makes sense because for larger df , the T-model is very close to the standard normal model. We will see that the critical T-value differs more from the critical Z-value when the sample sizes are small.

Finding the confidence interval:

We have all the pieces to build the confidence interval. In our example, the confidence interval is

$$\begin{array}{l} \begin{array}{l} \text{Estimated mean} \\ \text{Sample mean} \end{array} \pm \begin{array}{l} \text{Margin of error} \\ \text{Margin of error} \end{array} \\ 768.2 \pm 4.53 \quad (763.7, 772.7) \end{array}$$

Conclusion:

We are 95% confident that the mean breaking weight for all cables of this type is between 763.7 lb and 772.7 lb.

Confidence intervals at the 95% confidence level are common in practice. But 95% is not the only confidence level we use. Particularly in situations that involve safety issues, such as the previous example, people often prefer to estimate population means with 99% confidence intervals. Let's do some exploration with technology to see how changes in the confidence level affect the confidence interval.

[Click here to open this simulation in its own window.](#)

A link to an interactive elements can be found at the bottom of this page.

Try It

How Much Alcohol Do College Students Drink?

According to the website www.collegedrinkingprevention.gov, “About 25 percent of college students report academic consequences of their drinking including missing class, falling behind, doing poorly on exams or papers, and receiving lower grades overall.” A statistics student is curious about drinking habits of students at his college. He wants to estimate the mean number of alcoholic drinks consumed each week by students at his college. He plans to use a 90% confidence interval. He surveys a random sample of 71 students. The sample mean is 3.93 alcoholic drinks per week. The sample standard deviation is 3.78 drinks.

<https://assessments.lumenlearning.co...sessments/3684>

<https://assessments.lumenlearning.co...sessments/3753>

<https://assessments.lumenlearning.co...sessments/3685>

<https://assessments.lumenlearning.co...sessments/3686>

<https://assessments.lumenlearning.co...sessments/3687>

Try It

<https://assessments.lumenlearning.co...sessments/3754>

<https://assessments.lumenlearning.co...sessments/3755>

Try It

<https://assessments.lumenlearning.co...sessments/3691>

Contributors and Attributions

CC licensed content, Shared previously

- Concepts in Statistics. **Provided by:** Open Learning Initiative. **Located at:** <http://oli.cmu.edu>. **License:** *CC BY: Attribution*

This page titled [10.19: Estimating a Population Mean \(2 of 3\)](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Lumen Learning](#).