Introduction to t-Confidence Intervals with Examples

REVIEW

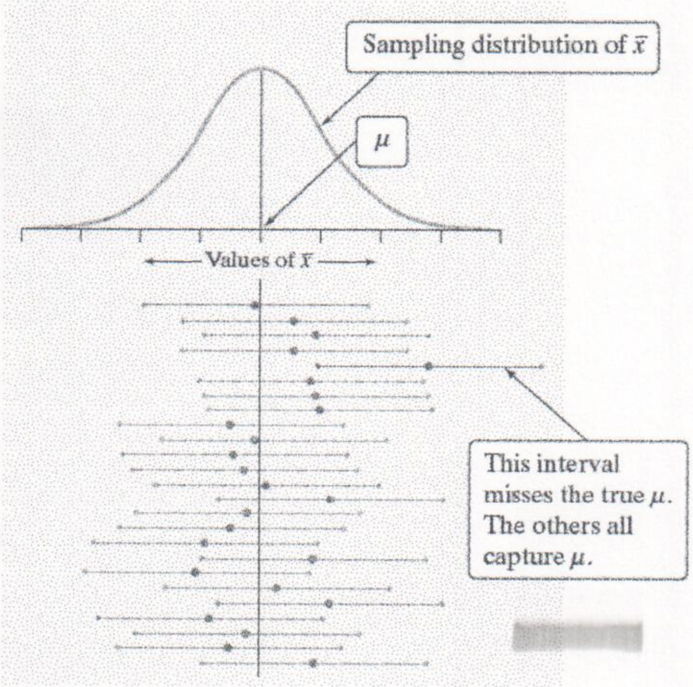
A **confidence interval** for a parameter has two parts:

1. An interval calculated from the data, which has the general form

   estimate ± margin of error

   The **margin of error** is the measure of uncertainty and is calculated as the product of the standard deviation of the statistic and a critical value based on the chosen level of confidence, C% (such as 90%, 95%, or 99%).

2. A **confidence level**, C%, which gives the probability that the interval will capture the true parameter value in repeated samples. That is, the confidence level is the success rate for the method.


Sampling distribution of $\bar{x}$
μ
Values of $\bar{x}$
This interval misses the true μ. The others all capture μ.

Interpreting the Confidence Interval $\bar{x}$ − margin of error

We are C% confident that the interval (↓________, ________↑) contains the true mean measure you want to know about – from context of the situation.

↑ $\bar{x}$ + margin of error

Introduction to t-Confidence Intervals with Examples

CONFIDENCE INTERVAL FOR A POPULATION MEAN WITH UNKNOWN σ

Example 1

Suppose we want to estimate the true mean systolic blood pressure of all long-haul truck drivers at a large trucking company. From a random sample of 41 long-haul truck drivers, the mean systolic blood pressure is $\bar{x} = 130$ mm Hg with a standard deviation of s_x = 12 mm Hg.

a. Define the parameter.

Mean systolic blood pressure of all long-haul truck drivers at the company.

b. Estimate the parameter with a point estimate

130 mm Hg

c. If the population standard deviation were known, what would be the approximate sampling distribution of the sample mean?

Since $n \geqq 30$ $(n=41)$, the sampling distribution is Normal.

d. What is the distribution of $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$? Standard Normal

e. In this study, the population standard deviation is **not** known and must be approximated. Use sample standard deviation to approximate the standard deviation of the sampling distribution.

$S = 12$ mm Hg

Approximate $\sigma_{\bar{x}} = \frac{12}{\sqrt{41}} = 1.875$

f. IMPORTANT QUESTION

If we replace σ with s_x, do you think that $\frac{\bar{x}-\mu}{\frac{s_x}{\sqrt{n}}}$ still has a standard normal distribution?

NO

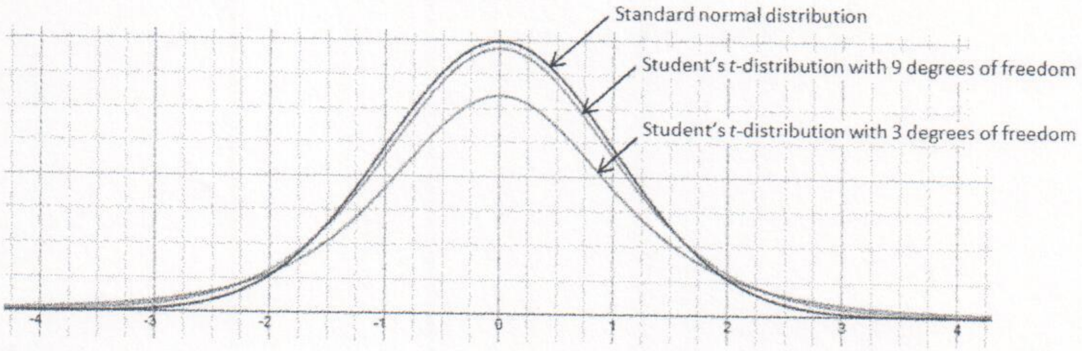
THE STUDENT'S *t*-DISTRIBUTION

When the population standard deviation is unknown (and hence, estimated from the sample data), we use the statistic s_X to estimate the parameter σ. As a result of this additional source of variation, the statistic $\frac{\bar{x}-\mu}{\frac{s_X}{\sqrt{n}}}$ no longer has a normal distribution. Instead, under certain conditions, this statistic has a **Student's *t*-distribution**. Several notes about the history and properties of the Student's *t*-distribution are worthy of mention.

Properties of the Student's *t*-Distribution

- W. S. Gosset published the Student's *t*-distribution in 1908. Gosset was a Guinness brewery employee and published his work under the pseudonym "Student."
- In his publications, Gosset assumed that random samples were taken from a Normal population. As the sample size increases, the normality assumption becomes less important.

- Like the standard normal distribution, the *t*-distribution is symmetric about its mean of 0.
- Unlike the standard normal distribution, the t-distribution has a standard deviation greater than 1. This is due to the additional variation from the use of s_X. However, as the sample size increases, the standard deviation gets closer and closer to 1. Also, as the sample size increases, the normality assumption becomes less and less important and the distribution of gets closer and closer to that of a standard normal distribution.
- The *t*-distribution is less peaked at the mean and thicker at the tails than the standard normal distribution.
- The *t*-distribution has a separate shape for each sample size, which is in turn characterized by **degrees of freedom**. As the degrees of freedom increase, the shape of the t-distribution gets closer and closer to that of the standard normal distribution.
- The figure below compares the density function for the *t*-distribution with 3 degrees of freedom, the t-distribution with 9 degrees and the standard normal distribution.



A *t*-CONFIDENCE INTERVAL FOR A POPULATION MEAN WITH UNKNOWN σ

Introduction to t-Confidence Intervals with Examples

When the sampling distribution is known to be Normal (population is Normal or n ≥ 30) and the population standard deviation is unknown (and hence, estimated from the sample data from a SRS), the margin of error for a confidence interval for the unknown parameter μ is

$$t^*(\frac{s_x}{\sqrt{n}})$$

where t^* is the critical value for the Student's *t*-distribution with $n-1$ degrees of freedom and C% of the area between $-t^*$ and t^*.

- Since $\frac{s_x}{\sqrt{n}}$ is now an estimate of the standard deviation of the sampling distribution, we define this quantity as the **standard error** of the mean.
- As before, we define $t^*(\frac{s_x}{\sqrt{n}})$ to be the **margin of error** of the estimation.

Summary of the conditions for this inference procedure

1. A random sample has been selected from the population.
2. The population has an approximate Normal distribution OR the sample size is large ($n \geq 30$).
3. The population standard deviation is unknown.

Finding t^* with n-1 degrees of freedom and confidence level C

Example 2

A table or calculator will be used to obtain the critical values for a Student's *t*-distribution. Remember that for a confidence interval for a population mean, the degrees of freedom is one less than the sample size ($df = n - 1$).

- Find the critical value for a 95% confidence interval using a sample of size $n = 5$.

  $df =$ 4 $t^* =$ 2.776

- Find the critical value for a 99% confidence interval using a sample of size $n = 10$.

  $df =$ 9 $t^* =$ 3.25

Example 3

Suppose we want to estimate the true mean systolic blood pressure of all long-haul truck drivers at a large trucking company with a 95% confidence interval. From a simple random sample of 41 long-haul truck drivers, the mean systolic blood pressure is $\bar{x}$ = 130 mm Hg with a standard deviation of s_x = 12 mm Hg. Construct and interpret the 95% confidence interval.

Solution

Choose the correct inference procedure: State the confidence level and define the unknown parameter of interest

We will calculate a *95% confidence interval* for a *population mean with an unknown σ*. The standard deviation will be estimated from the sample data.

Check the conditions

- A simple random sample was selected. ✓
- We don't know whether the population distribution of systolic blood pressures is Normal. Because the sample size is large (*n* = 41 > 30), we can use a *t*-distribution.

Carry out the inference procedure

To find the critical value t^*, we use a t-distribution with *df* = 41 – 1 = 40.

For a 95% level of confidence, t^* = 2.021.

$$\bar{x} \pm t^*\left(\frac{s_x}{\sqrt{n}}\right) = 130 \pm (2.021)\frac{12}{\sqrt{41}} = 130 \pm 3.79 = (126.21, 133.79).$$

State the conclusion by interpreting the confidence interval

We are 95% confident that the interval (126.2 mm Hg, 133.8 mm Hg) contains true mean systolic blood pressure of all long-haul truckers at the company.

Example 4

In the calculation of the confidence interval in example 4, identify the value of each of the following.

a. The point estimate of the parameter. 130

b. The critical value. 2.021

c. The (estimated) standard deviation† of the point estimate. 1.875 $12/\sqrt{41}$

d. The margin of error. 3.79

†An estimated standard deviation is called a standard error.

DETERMINING A SAMPLE SIZE

Suppose we want to find the minimum sample size required to estimate the true mean systolic blood pressure to within 2.5 mm Hg? We could set the margin of error, M, equal to $t^*(\frac{s_x}{\sqrt{n}})$ and solve for n. This time, the sample size would be in terms of the sample standard deviation s_X (which is not a problem). But there is a "Catch 22." The formula would also require knowing the value of t^*. However, finding the value of t^* requires knowledge of the sample size n (so as to determine the degrees of freedom).

To get around this circular problem, we simply use the value of z^* in place of t^*. Since our goal is to obtain only a ballpark estimate of the minimum sample size, the use of this approximation is acceptable. Thus, we will use the formula

$$n = \left(\frac{z^* s_x}{M}\right)^2$$

to approximate the minimum sample size necessary to estimate the sample mean to within a given margin of error M. Notice that this formula uses the sample standard deviation to estimate the population standard deviation.

Finally, you are reminded that when using the above formula, always remember to round the sample size **up** to the nearest whole number so that the margin of error is less strict than M.

Example 5

Suppose we want to find the minimum sample size required to estimate the true mean systolic blood pressure to within 2.5 mm Hg with a 95% level of confidence. Use the sample standard deviation of 12 mm Hg.

Solution

We will use the formula $n = \left(\frac{z^* s_x}{M}\right)^2$ where s_X = 12 mm Hg, M = 2.5 mm Hg, and z^* = 1.960 (the critical value of a 95% confidence interval).

$$n = \left(\frac{z^* s_x}{M}\right)^2 = n = \left(\frac{(1.96)(12)}{2.5}\right)^2 = 88.51.$$

As expected, the result is not a whole number. Round this number up to 89 long-haul truckers.

Answer

A minimum sample of size 89 long-haul truckers is needed to estimate the true mean systolic blood pressure to 2.5 mm Hg of the actual value with a confidence level of 95%.

THE MATCHED PAIRS DESIGN FOR A POPULATION MEAN WITH UNKNOWN σ

Example 6

Trace metals found in wells affect the taste of drinking water, and high concentrations can pose a health risk. Researchers measured the concentration of zinc (in mg/liter) near the top and the bottom of 10 randomly selected wells in a large region. The data are provided in the table below.

Well	1	2	3	4	5	6	7	8	9	10
Bottom	0.430	0.266	0.567	0.531	0.707	0.716	0.651	0.581	0.469	0.723
Top	0.415	0.238	0.390	0.410	0.605	0.609	0.632	0.523	0.411	0.611
Difference	0.015	0.028	0.177	0.121	0.102	0.107	0.019	0.058	0.058	0.112

a. Construct a 95% confidence interval for the mean difference μ in the zinc concentrations from these two locations.

Calculate $\bar{x}$ and s of the "Difference" row.

$\bar{x} = 0.0797$

$s = 0.0526$

critical value $t_{0.95,9} = 2.262$

$\bar{x} \pm \frac{(2.262)(0.0526)}{\sqrt{10}} = 0.0797 \pm 0.0376$

(0.04, 0.12)

b. Interpret the confidence interval.

We are 95% confident that the interval (0.04, 0.12) contains the true mean difference in zinc concentrations from the bottom to the top of the well.

c. Does your interval in part (a) give convincing evidence of a difference in zinc concentrations at the top and bottom of the wells in the region? Justify your answer.

If there was no difference in zinc concentrations between the two locations, the mean difference would be 0.

Our interval contains values that may capture the true mean difference, and since our interval does not contain 0, we conclude that there is likely a difference in zinc concentrations between the two locations in the well.