

CONFIDENCE INTERVALS: CONFIDENCE INTERVAL, SINGLE POPULATION MEAN, STANDARD DEVIATION UNKNOWN, STUDENT'S-T*

Susan Dean

Barbara Illowsky, Ph.D.

This work is produced by OpenStax-CNX and licensed under the Creative Commons Attribution License 3.0[†]

Abstract

Confidence Interval, Single Population Mean, Population Standard Deviation Unknown, Student-t is part of the collection coll0555 written by Barbara Illowsky and Susan Dean with contributions from Roberta Bloom.

In practice, we rarely know the population **standard deviation**. In the past, when the sample size was large, this did not present a problem to statisticians. They used the sample standard deviation s as an estimate for σ and proceeded as before to calculate a **confidence interval** with close enough results. However, statisticians ran into problems when the sample size was small. A small sample size caused inaccuracies in the confidence interval.

William S. Gossett (1876-1937) of the Guinness brewery in Dublin, Ireland ran into this problem. His experiments with hops and barley produced very few samples. Just replacing σ with s did not produce accurate results when he tried to calculate a confidence interval. He realized that he could not use a normal distribution for the calculation; he found that the actual distribution depends on the sample size. This problem led him to "discover" what is called the **Student's-t distribution**. The name comes from the fact that Gosset wrote under the pen name "Student."

Up until the mid 1970s, some statisticians used the **normal distribution** approximation for large sample sizes and only used the Student's-t distribution for sample sizes of at most 30. With the common use of graphing calculators and computers, the practice is to use the Student's-t distribution whenever s is used as an estimate for σ .

If you draw a simple random sample of size n from a population that has approximately a normal distribution with mean μ and unknown population standard deviation σ and calculate the t-score $t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}$, then the t-scores follow a **Student's-t distribution with $n - 1$ degrees of freedom**. The t-score has the same interpretation as the **z-score**. It measures how far \bar{x} is from its mean μ . For each sample size n , there is a different Student's-t distribution.

*Version 1.24: Jun 18, 2012 9:44 am -0500

[†]<http://creativecommons.org/licenses/by/3.0/>

The **degrees of freedom**, $n - 1$, come from the calculation of the sample standard deviation s . In Chapter 2, we used n deviations ($x - \bar{x}$ **values**) to calculate s . Because the sum of the deviations is 0, we can find the last deviation once we know the other $n - 1$ deviations. The other $n - 1$ deviations can change or vary freely. **We call the number $n - 1$ the degrees of freedom (df).**

Properties of the Student's-t Distribution

- The graph for the Student's-t distribution is similar to the Standard Normal curve.
- The mean for the Student's-t distribution is 0 and the distribution is symmetric about 0.
- The Student's-t distribution has more probability in its tails than the Standard Normal distribution because the spread of the t distribution is greater than the spread of the Standard Normal. So the graph of the Student's-t distribution will be thicker in the tails and shorter in the center than the graph of the Standard Normal distribution.
- The exact shape of the Student's-t distribution depends on the "degrees of freedom". As the degrees of freedom increases, the graph Student's-t distribution becomes more like the graph of the Standard Normal distribution.
- The underlying population of individual observations is assumed to be normally distributed with unknown population mean μ and unknown population standard deviation σ . The size of the underlying population is generally not relevant unless it is very small. If it is bell shaped (normal) then the assumption is met and doesn't need discussion. Random sampling is assumed but it is a completely separate assumption from normality.

Calculators and computers can easily calculate any Student's-t probabilities. The TI-83,83+,84+ have a tcdf function to find the probability for given values of t. The grammar for the tcdf command is tcdf(lower bound, upper bound, degrees of freedom). However for confidence intervals, we need to use **inverse** probability to find the value of t when we know the probability.

For the TI-84+ you can use the invT command on the DISTRibution menu. The invT command works similarly to the invnorm. The invT command requires two inputs: **invT(area to the left, degrees of freedom)** The output is the t-score that corresponds to the area we specified.

The TI-83 and 83+ do not have the invT command. (The TI-89 has an inverse T command.)

A probability table for the Student's-t distribution can also be used. The table gives t-scores that correspond to the confidence level (column) and degrees of freedom (row). (The TI-86 does not have an invT program or command, so if you are using that calculator, you need to use a probability table for the Student's-t distribution.) When using t-table, note that some tables are formatted to show the confidence level in the column headings, while the column headings in some tables may show only corresponding area in one or both tails.

A Student's-t table (See the Table of Contents **15. Tables**) gives t-scores given the degrees of freedom and the right-tailed probability. The table is very limited. **Calculators and computers can easily calculate any Student's-t probabilities.**

The notation for the Student's-t distribution is (using T as the random variable) is

- $T \sim t_{df}$ where $df = n - 1$.
- For example, if we have a sample of size $n=20$ items, then we calculate the degrees of freedom as $df=n-1=20-1=19$ and we write the distribution as $T \sim t_{19}$

If the population standard deviation is not known, the error bound for a population mean is:

- $EBM = t_{\frac{\alpha}{2}} \cdot \left(\frac{s}{\sqrt{n}} \right)$
- $t_{\frac{\alpha}{2}}$ is the t-score with area to the right equal to $\frac{\alpha}{2}$
- use $df = n - 1$ degrees of freedom
- s = sample standard deviation

The format for the confidence interval is:

$$(\bar{x} - \text{EBM}, \bar{x} + \text{EBM}).$$

The TI-83, 83+ and 84 calculators have a function that calculates the confidence interval directly. To get to it,

Press **STAT**

Arrow over to **TESTS**.

Arrow down to **8:TIInterval** and press **ENTER** (or just press **8**).

Example 1

Suppose you do a study of acupuncture to determine how effective it is in relieving pain. You measure sensory rates for 15 subjects with the results given below. Use the sample data to construct a 95% confidence interval for the mean sensory rate for the population (assumed normal) from which you took the data.

The solution is shown step-by-step and by using the TI-83, 83+ and 84+ calculators.

8.6; 9.4; 7.9; 6.8; 8.3; 7.3; 9.2; 9.6; 8.7; 11.4; 10.3; 5.4; 8.1; 5.5; 6.9

Solution

- You can use technology to directly calculate the confidence interval.
- The first solution is step-by-step (Solution A).
- The second solution uses the Ti-83+ and Ti-84 calculators (Solution B).

Solution A

To find the confidence interval, you need the sample mean, \bar{x} , and the EBM.

$$\bar{x} = 8.2267 \quad s = 1.6722 \quad n = 15$$

$$\text{df} = 15 - 1 = 14$$

$$\text{CL} = 0.95 \quad \text{so} \quad \alpha = 1 - \text{CL} = 1 - 0.95 = 0.05$$

$$\frac{\alpha}{2} = 0.025 \quad t_{\frac{\alpha}{2}} = t_{.025}$$

The area to the right of $t_{.025}$ is 0.025 and the area to the left of $t_{.025}$ is $1 - 0.025 = 0.975$

$t_{\frac{\alpha}{2}} = t_{.025} = 2.14$ using $\text{invT}(.975, 14)$ on the TI-84+ calculator.

$$\text{EBM} = t_{\frac{\alpha}{2}} \cdot \left(\frac{s}{\sqrt{n}} \right)$$

$$\text{EBM} = 2.14 \cdot \left(\frac{1.6722}{\sqrt{15}} \right) = 0.924$$

$$\bar{x} - \text{EBM} = 8.2267 - 0.9240 = 7.3$$

$$\bar{x} + \text{EBM} = 8.2267 + 0.9240 = 9.15$$

The 95% confidence interval is **(7.30, 9.15)**.

We estimate with 95% confidence that the true population mean sensory rate is between 7.30 and 9.15.

Solution B

Using a function of the TI-83, TI-83+ or TI-84 calculators:

Press **STAT** and arrow over to **TESTS**.

Arrow down to **8:TIInterval** and press **ENTER** (or you can just press **8**). Arrow to **Data** and press **ENTER**.

Arrow down to **List** and enter the list name where you put the data.

Arrow down to **Freq** and enter **1**.

Arrow down to **C-level** and enter **.95**

Arrow down to **Calculate** and press **ENTER**.

The 95% confidence interval is **(7.3006, 9.1527)**

NOTE: When calculating the error bound, a probability table for the Student's-t distribution can also be used to find the value of t. The table gives t-scores that correspond to the confidence level (column) and degrees of freedom (row); the t-score is found where the row and column intersect in the table.

**With contributions from Roberta Bloom

Glossary

Definition 1: Confidence Interval (CI)

An interval estimate for an unknown population parameter. This depends on:

- The desired confidence level.
- Information that is known about the distribution (for example, known standard deviation).
- The sample and its size.

Definition 2: Confidence Level (CL)

The percent expression for the probability that the confidence interval contains the true population parameter. For example, if the CL = 90%, then in 90 out of 100 samples the interval estimate will enclose the true population parameter.

Definition 3: Degrees of Freedom (df)

The number of objects in a sample that are free to vary.

Definition 4: Error Bound for a Population Mean (EBM)

The margin of error. Depends on the confidence level, sample size, and known or estimated population standard deviation.

Definition 5: Normal Distribution

A continuous random variable (RV) with pdf $f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$, where μ is the mean of the distribution and σ is the standard deviation. Notation: $X \sim N(\mu, \sigma)$. If $\mu = 0$ and $\sigma = 1$, the RV is called **the standard normal distribution**.

Definition 6: Standard Deviation

A number that is equal to the square root of the variance and measures how far data values are from their mean. Notation: s for sample standard deviation and σ for population standard deviation.

Definition 7: Student's-t Distribution

Investigated and reported by William S. Gossett in 1908 and published under the pseudonym Student. The major characteristics of the random variable (RV) are:

- It is continuous and assumes any real values.
- The pdf is symmetrical about its mean of zero. However, it is more spread out and flatter at the apex than the normal distribution.
- It approaches the standard normal distribution as n gets larger.
- There is a "family" of t distributions: every representative of the family is completely defined by the number of degrees of freedom which is one less than the number of data.