

Confidence Intervals for Population Means (Section 9.2)

So, we found a sample mean but what can we say about the population?

We will be doing nearly the same stuff as we did in the first section of this chapter. However, there are some significant differences. The main one is that we do *not* use the normal distribution. We will be using a similar, bell-shaped curve called **Student's *t*-distribution**.

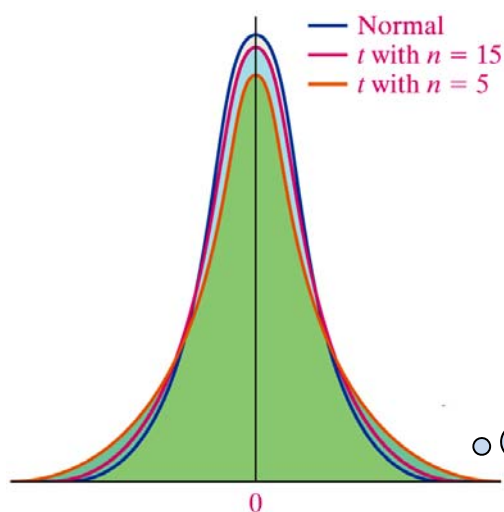
To make a long story not so long, using the method discussed in the previous section to find the confidence interval for $\bar{x}$ would be “point estimate $\pm$ margin of error” or rather $\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$.

This is a problem because this assumes we know the standard deviation of the population, which we would *not*. We can substitute the standard deviation of the sample, giving us $\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$.

However, by using s instead of σ , we introduce too much variation and we see that the

z -score of $\bar{x}$ (now $z_{\bar{x}} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$) is *not* normally distributed.

At first, this seems like we hit a brick wall. Luckily, a man by the name of William Gossett in the early 1900's solved this problem by creating what is known as **Student's *t*-distribution**. It is similar to the normal curve but has a little more variation. We see the normal curve here compared with two curves of the *t*-distribution.



Gossett was working for a brewery at the time. He published under a pseudonym so his competitors would not know.

The *t*-curve has fatter tails but a less-high center point. The *t*-curve depends on the value of the sample size.

Recall: Definition: A **point estimate** is the value of a (sample) statistic that estimates the value of a (population) parameter.

In this section, the point estimate we are playing with is $\bar{x}$, the sample mean. It is said to estimate μ , the population mean.

Student's t -Distribution:

Suppose that a simple random sample of size n is taken from a population. If the population from which the sample is drawn follows a normal distribution, the distribution of

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$
 follows Student's t -distribution with $n - 1$ degrees of freedom, where $\bar{x}$ is

the sample mean, s is the sample standard deviation, and μ is the population mean.

Properties of the t -Distribution:

There are similarities to the normal distribution as we will see. However, there are significant differences.

1. The t -distribution is different for different degrees of freedom. The degrees of freedom will be calculated as $n - 1$.
2. The t -distribution is centered at 0 and is symmetric about 0.
3. The area under the curve is 1. The area under the curve to the right of 0 is equal to the area under the curve to the left of 0. This area is $\frac{1}{2}$.
4. As t increases without bound (gets larger and larger), the graph approaches, but never reaches, the horizontal axis. As t decreases without bound (gets more and more negative), the graph approaches, but never reaches, the horizontal axis.
5. The area in the tails of the t -distribution is a little greater than the area in the tails of the standard normal distribution, because we are using s as an estimate of σ , thereby introducing further variability into the t -statistic.
6. As the sample size n increases, the density curve of t gets closer to the standard normal density curve. This result occurs because, as the sample size n increases, the values of s get closer to the values of σ , by the Law of Large Numbers.

Determining t -values:

Just as we looked up z -scores on a table, we will look up t -values. We will use the notation t_α to denote the t -value whose area under the curve to the right of t_α is α . We will need to use the sample size, n . As mentioned earlier, the degrees of freedom will be calculated as $n - 1$.

expl 1: Find the t -value such that the area under the t -distribution curve to the right is 0.025, assuming 19 degrees of freedom. Use the partial table given below.

Table of t-Distribution Areas

df	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.706	15.894	31.821	63.657	127.321	318.309	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.089	22.327	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.215	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850

Label your t -value as $t_{0.025}$.

Draw a quick t -curve with your t -value labeled and the area 0.025 shaded.

To look up a t -value, find the row with the correct degrees of freedom and the column with the correct value of α .

Table VII in the book, on page A-14, will give you these same values.

The table entry is the t -value such that the area under the curve to the right is α (column heading).

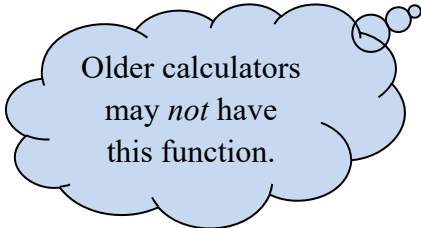
Using the TI calculator:

There is a function on the TI calculator that can be used for this. However, you must enter the area to the *left* of the desired t -value. Let's redo example 1.

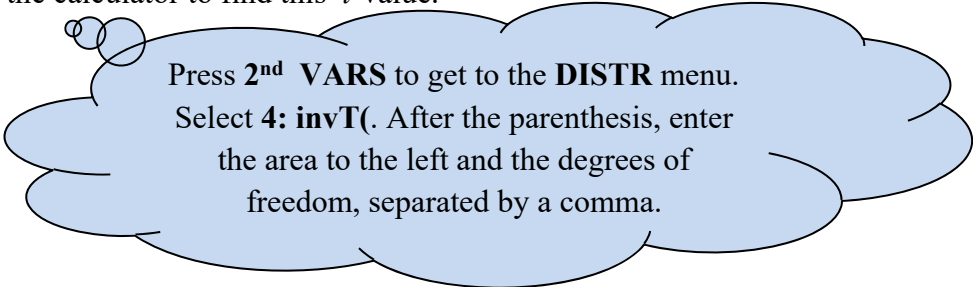
expl 2: Find the t -value such that the area under the t -distribution curve to the right is 0.025, assuming 19 degrees of freedom. Follow these steps to do this on the calculator.

a.) If the area under the curve to the right of the t -value is 0.025, what is the area to the left under the curve?

b.) Use the **invT** feature on the calculator to find this t -value.



Older calculators may *not* have this function.



Press **2nd VARS** to get to the **DISTR** menu.
Select **4: invT(**. After the parenthesis, enter the area to the left and the degrees of freedom, separated by a comma.

Now, the reason we need t -values is that we want to create confidence intervals as we did in the previous section.

Constructing a $(1 - \alpha)100\%$ Confidence Interval for μ :

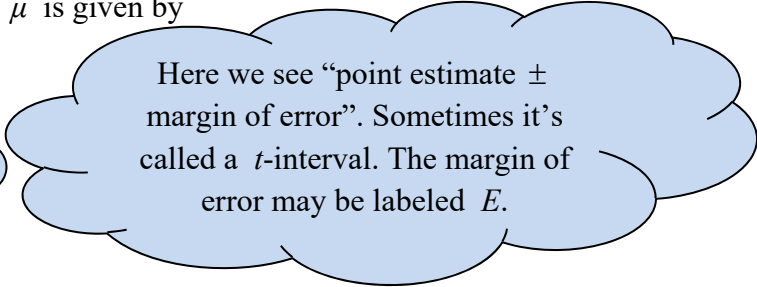
We must verify that the following is true before continuing.

- sample data come from a simple random sample or randomized experiment,
- sample data has no serious outliers,
- sample size is small relative to the population size ($n \leq 0.05N$), and
- the data come from a population that is normally distributed, or the sample size is large.

A $(1 - \alpha) \cdot 100\%$ confidence interval for μ is given by

$$\text{Lower bound: } \bar{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$\text{Upper bound: } \bar{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

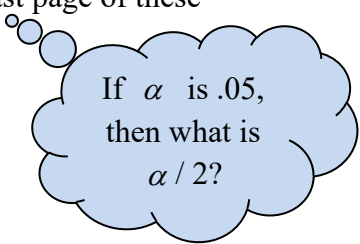


Here we see “point estimate $\pm$ margin of error”. Sometimes it's called a t -interval. The margin of error may be labeled E .

where $t_{\alpha/2}$ is the critical value with $n - 1$ degrees of freedom.

expl 3: A sample of sixteen 2011 Ford Focus cars has a mean gas mileage of 36.8 miles per gallon. The sample standard deviation is 2.92 miles per gallon. A normal probability plot has shown that the data could come from a normal population. In addition, a boxplot has shown that there are no outliers. Construct a 95% confidence interval for the mean gas mileage of all 2011 Ford Focus cars. Follow the steps below.

a.) Find the t -value needed on the table. A complete table is given on the last page of these notes.



If α is .05,
then what is
 $\alpha / 2$?

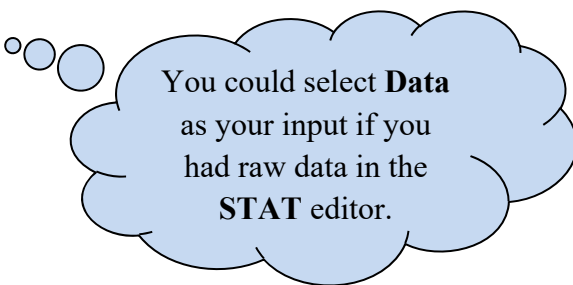
b.) Calculate the margin of error as $t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$.

c.) Form the 95% confidence interval and complete the sentence. Include units.

We are 95% confident that the true mean gas mileage of 2011 Ford Focus cars is between _____ and _____.

Finding a Confidence Interval Using the TI Calculator:

Press the **STAT** button and arrow over to **TESTS**. Arrow down to select **TInterval....** For **Inpt:** (input), we will choose **Stats**. This allows us to enter the values of the mean, standard deviation, and sample size. Do this, pressing **ENTER** after each. Enter the confidence level (**C-Level:**) in decimal form and then highlight **Calculate** at the bottom, pressing **ENTER**.



You could select **Data**
as your input if you
had raw data in the
STAT editor.

Determining the Sample Size n :

The sample size required to estimate the population mean, μ , with a level of confidence $(1 - \alpha) \cdot 100\%$ with a specified margin of error, E , is given by $n = \left(\frac{z_{\alpha/2} \cdot s}{E} \right)^2$. We round n up as before.

Using $t_{\alpha/2}$ would require us already knowing n . So, we use $z_{\alpha/2}$.

expl 4: Consider again the Ford Focuses. How large a sample do we need if we want to form a 95% confidence interval to estimate the mean within 0.5 miles per gallon?

Notice we use $z_{\alpha/2}$ here. What is $z_{0.025}$?

Instructions for STATCRUNCH:

Constructing a Confidence Interval for a Population Mean:

1. If needed, enter raw data in column **var1** and rename column.
2. Select **Stat > T Stats > One Sample**. Choose **With Data** or **With Summary** as appropriate.
3. If you have raw data, select the column under **Select column(s):***. If you have summary information, enter the **Sample mean**, **Sample std. dev** (standard deviation), and **Sample size** at the top. Select **Confidence interval for μ** and enter the **Level** desired. Click **Compute!**.

Determining Sample Size for Specified Margin of Error:

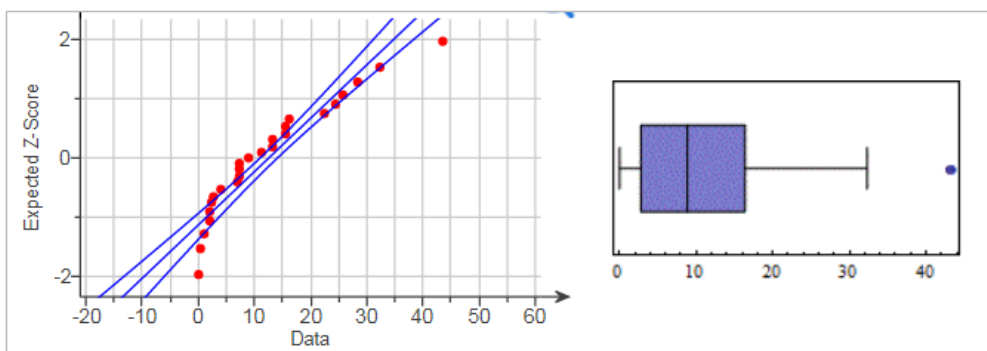
1. Select **Stat > T Stats > One Sample > Width/Sample Size**.
2. Under **Required parameters**, enter the **Confidence level** and (sample) **std. dev.**; Under **Enter one**, enter twice the margin of error needed as the **Width** and clear the **Sample size** cell. Click **Compute**.

The output will not be exactly what we got above due to rounding error.

When can we *not* make a confidence interval?:

Recall the population must be normal or the sample size large to be justified in using the t -distribution to make a confidence interval. Also, there should be no serious outliers in the data. We use the concepts of previous sections to determine if the data comes from a normal population or has outliers.

expl 5: The data from a simple random sample with 25 observations was used to construct the plots given below. The normal probability plot that was constructed has a correlation coefficient of 0.947. Judge whether a t -interval could be constructed using the data in the sample. Follow the steps below.



a.) Compare the given correlation coefficient to the critical value in the partial table below. What is your conclusion about the normality of the population?

Table VI for Assessing Normality	
Sample size, n	Critical Value
5	0.880
15	0.939
20	0.951
25	0.959
30	0.960

Recall, for the data to be normal, the correlation coefficient must be greater than the critical value.

b.) Consult the boxplot to determine if the data set has outliers. If it does, the confidence interval should *not* be performed. Are we justified in making a confidence interval using the t -distribution? Why?

Table of t -Distribution Areas: Given the degrees of freedom (first column, labeled **df**) and area to the right under the curve (title row, given as $\alpha/2$), look up the appropriate t -value. Be mindful of decimal points!

Table of t-Distribution Areas												
df	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.706	15.894	31.821	63.657	127.321	318.309	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.089	22.327	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.215	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
31	0.682	0.853	1.054	1.309	1.696	2.040	2.144	2.453	2.744	3.022	3.375	3.633
32	0.682	0.853	1.054	1.309	1.694	2.037	2.141	2.449	2.738	3.015	3.365	3.622
33	0.682	0.853	1.053	1.308	1.692	2.035	2.138	2.445	2.733	3.008	3.356	3.611
34	0.682	0.852	1.052	1.307	1.691	2.032	2.136	2.441	2.728	3.002	3.348	3.601
35	0.682	0.852	1.052	1.306	1.690	2.030	2.133	2.438	2.724	2.996	3.340	3.591
36	0.681	0.852	1.052	1.306	1.688	2.028	2.131	2.434	2.719	2.990	3.333	3.582
37	0.681	0.851	1.051	1.305	1.687	2.026	2.129	2.431	2.715	2.985	3.326	3.574
38	0.681	0.851	1.051	1.304	1.686	2.024	2.127	2.429	2.712	2.980	3.319	3.566
39	0.681	0.851	1.050	1.304	1.685	2.023	2.125	2.426	2.708	2.976	3.313	3.558
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
70	0.678	0.847	1.044	1.294	1.667	1.994	2.093	2.381	2.648	2.899	3.211	3.435
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
90	0.677	0.846	1.042	1.291	1.662	1.987	2.084	2.368	2.632	2.878	3.183	3.402
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z	0.674	0.842	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.090	3.291
df	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005

Area in Right Tail

(source: <http://www.chegg.com>)

If you do *not* see the correct degrees of freedom (**df**), use the closest value. Use the values of **z** given at the bottom if **df** is greater than 1,000.