Section 2E – One Population Mean & Proportion Confidence Intervals

Vocabulary

Population: The collection of all people or objects to be studied.

Census: Collecting data from everyone in a population.

Sample: Collecting data from a small subgroup of the population.

Statistic: A number calculated from sample data in order to understand the characteristics of the data. For example, a sample mean average, a sample standard deviation, or a sample percentage.

Parameter: A number that describes the characteristics of a population like a population mean or a population percentage. Can be calculated from an unbiased census, but is often just a guess about the population.

Sampling Distribution: Take many random samples from a population, calculate a sample statistic like a mean or percent from each sample and graph all of the sample statistics on the same graph. The center of the sampling distribution is a good estimate of the population parameter.

Sampling Variability: Random samples values and sample statistics are usually different from each other and usually different from the population parameter.

Point Estimate: When someone takes a sample statistic and then claims that it is the population parameter.

Margin of Error: Total distance that a sample statistic might be from the population parameter. For normal sampling distributions and a 95% confidence interval, the margin of error is approximately twice as large as the standard error.

Standard Error: The standard deviation of a sampling distribution. The distance that typical sample statistics are from the center of the sampling distribution. Since the center of the sampling distributions is usually close to the population parameter, the standard error tells us how far typical sample statistics are from the population parameter.

Confidence Interval: Two numbers that we think a population parameter is in between. Can be calculated by either a bootstrap distribution or by adding and subtracting the sample statistic and the margin of error.

95% Confident: 95% of confidence intervals contain the population value and 5% of confidence intervals do not contain the population value.

90% Confident: 90% of confidence intervals contain the population value and 10% of confidence intervals do not contain the population value.

99% Confident: 99% of confidence intervals contain the population value and 1% of confidence intervals do not contain the population value.

Bootstrapping: Taking many random samples values from one original real random sample with replacement.

Bootstrap Sample: A simulated sample created by taking many random samples values from one original real random sample with replacement.

Bootstrap Statistic: A statistic calculated from a bootstrap sample.

Bootstrap Distribution: Putting many bootstrap statistics on the same graph in order to simulate the sampling variability in a population, calculate standard error, and create a confidence interval. The center of the bootstrap distribution is the original real sample statistic.

In the last section, we saw that if we have only one random sample from a population, we would not be able to find the population parameter exactly. The best we can do is create a confidence interval, which is two numbers that we think the population parameter is in between.

In this section, we will look at some of the famous formulas that statisticians use to estimate population parameters with confidence intervals. We will also look at sample data conditions in order to ensure the accuracy of the formula.

If our sampling distribution is normal, most one-population confidence interval formulas start from the following.

Sample Statistic ± Margin of Error

Early mathematicians and statisticians thought a lot about how to estimate the margin of error when you do not know the population parameter. The key was the sampling distribution. If a sampling distribution looked normal, then the empirical rule would suggest that the middle 95% would correspond to two standard deviations above and below the center. This gave rise to another common formula.

Sample Statistic ± (2 × Standard Error)

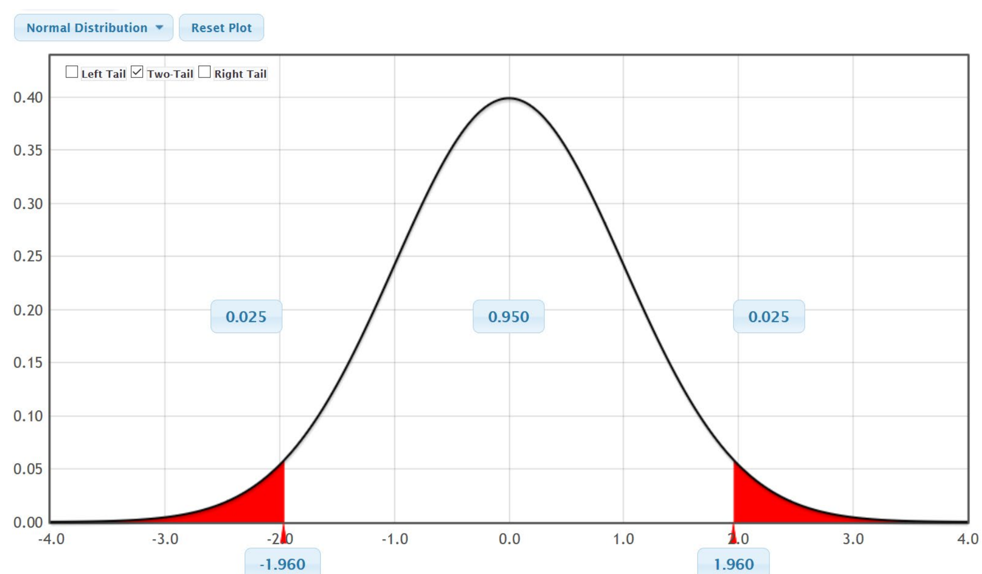
Critical Value Z-scores

What is the "2" representing in the following formula. It seems it is counting how many standard errors one is from the mean (center) of the sampling distribution. Does this remind you of a statistic we previously learned?

If you recall, the Z-score measures the number of standard deviations from the mean. So the "2" is really a Z-score. This gave rise to the idea of replacing the "2" with a Z-score. The Z-score can be adapted for 90%, 95% or 99%. Remember two standard deviation is just an approximation for 95%. If that is the case, can we get a more accurate number for 95%?

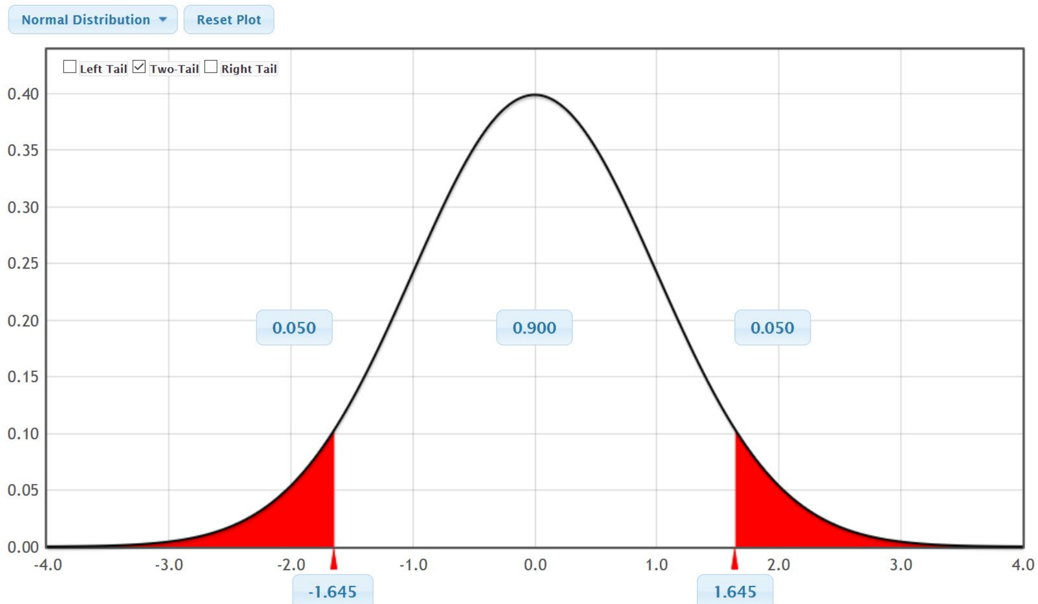
Using a normal calculator, we can calculate the Z-score for 90%, 95% and 99% confidence. These are very famous and are often referred to as "critical value Z-scores" or "Z_c" for short.

Go to www.lock5stat.com and open StatKey. Under the "theoretical distributions" menu, click on "normal". If the mean is zero and the standard deviation is one, then this will calculate Z-scores. Click the "two-tail" button. The first Z-score calculated is for 95%.



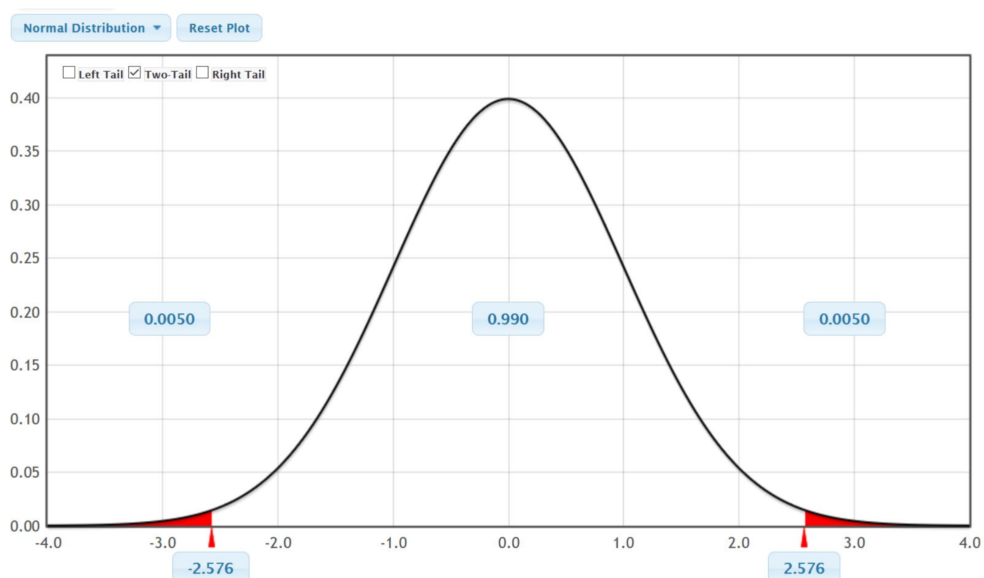
This is the most famous of all the critical value Z-scores. Remember, for the middle 95%, the empirical rule indicates that it will be "about" two standard deviations. It turns out, 1.96 standard deviations is more accurate. Notice that just like the confidence intervals have an upper limit and lower limit, so does the Z-score critical values.
For 95% confidence, we can replace the ±2 in the formula with ±1.96.

What about 90% confidence intervals? Go back to the normal calculator in StatKey and click on the "0.95" in the middle. Change it to 0.9 (90%).



Notice the Z-score for 90% confidence intervals is ±1.645. Notice that as the confidence interval decreases from 95% to 90%, the Z-score gets lower. This will cause the margin of error to decrease and the confidence interval to get narrower.

What about 99% confidence intervals? Go back to the normal calculator in StatKey and change the middle proportion into 0.99 (99%).



Notice the Z-score for 99% confidence intervals is ±2.576. Therefore, instead of being 1.645 standard errors away or 1.96 standard errors away, now we are 2.576 standard errors away. As the confidence interval increases from 95% to 99%, the Z-score gets larger. This will cause the margin of error to increase and the confidence interval to get wider.

Here are the famous critical value Z-scores.

- 90% confidence level: Z = ± 1.645
- 95% confidence level: Z = ± 1.96
- 99% confidence level: Z = ± 2.576

Let us summarize the progress of our one-population confidence interval formula. It is important to remember that these formulas only work if our sampling distribution looks normal. Z-scores calculate the number of standard deviations (standard errors) from the mean in a perfectly normal curve.

Sample Statistic ± Margin of Error

Sample Statistic ± (2 × Standard Error)

Sample Statistic ± (Z × Standard Error)

Statisticians discovered that as long as the sampling distribution was normal, the Z-scores were accurate for proportion (percentage) confidence intervals. The famous critical value Z-scores are still used to this day to calculate a confidence interval estimate of a population proportion (percentage).

<u>One-Population Proportion Confidence Interval</u>

Before computers were invented, it was very difficult to make sampling distributions. Yet it was vital to understanding sample statistics and calculating standard error. Early mathematicians and statisticians invented formulas to estimate the standard error.

Standard Error Estimation Formula for Proportions = $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Sample Proportion = p̂
Sample Size = n

So now, we can finish our estimation formula for a confidence interval estimate of the population proportion. In order to estimate the margin of error, we multiply the standard error by the number of standard errors (Z-score).

Sample Statistic ± Margin of Error
Sample Statistic ± (2 × Standard Error)
Sample Statistic ± (Z × Standard Error)

$\hat{p} \pm \left(Z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$

Example 1: Calculating the confidence interval for a proportion

A random sample of 54 bears in a region of California showed that 19 of them were female. Find the sample proportion and use the formula above to calculate a 95% confidence interval estimate for the population proportion of female bears in this region of California.

Sample Proportion ($\hat{p}$) = $\frac{Amount\ of\ Female\ Bears}{Sample\ Size} = \frac{19}{54} \approx 0.352$

Critical Value Z-score for 90% Confidence = ± 1.96

Now we will replace the Z-score with 1.96 and p̂ with 0.352 and *n* with 54 into our formula and work it out. Remember to follow order of operations. Notice the standard error estimate is 0.065 (6.5%) and the margin of error estimate is 0.127 (12.7%).

$$\hat{p} \pm \left(Z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

$$0.352 \pm \left(1.96\sqrt{\frac{0.352(1-0.352)}{54}}\right)$$

$$0.352 \pm \left(1.96\sqrt{\frac{0.352(0.648)}{54}}\right)$$

$$0.352 \pm (1.96 \times 0.065)$$

$$0.352 \pm (0.127)$$

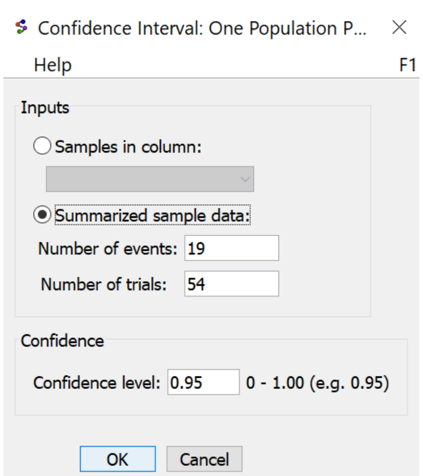
0.352 – 0.127 < Population Proportion of Female Bears (π) < 0.352 + 0.127

0.225 < Population Proportion of Female Bears (π) < 0.479

We are 95% confident that between 22.5% and 47.9% of all bears in this region of California are female.

Note: While it is important to understand formulas, data scientist today rely on computers to calculate confidence intervals. It is very difficult to calculate confidence intervals from large data sets with a formula and a calculator. The job of a data scientist, statistician, or data analyst is understand and explain the data, not to spend hours calculating something a computer can do in a split second.

To calculate this confidence interval with Statcato, we will click on the “statistics” menu and then “confidence intervals”. Click on one-population proportion and under summary data; enter 19 for the number of events and 54 for the number of trials. Set the confidence level to 0.95 and click OK.



Here is the Statcato printout. Notice the computer calculation is almost the same as the one we did with the formula and calculator. However, it took a lot less time.

Confidence Interval - One population proportion: confidence level = 0.95
Input: Summary data

Number of trials	Number of Events	Sample proportion	Margin of Error	95.0%CI
54	19	0.352	0.127	(0.2245, 0.4792)

Key Question: How accurate is this confidence interval?

This confidence interval relies on a Z-score and the standard error so the sampling distribution for sample proportions must be normal for this formula to be accurate. If we look at the section on the central limit theorem, we remember that for a sampling distribution for random sample proportions to be normal, we need at least ten success and at least ten failures. This gives rise to the assumptions or conditions required for certain confidence interval calculations. For the formula approach to be accurate, the following must be true. If any of these assumptions are not met, then the confidence interval may not be accurate.

<u>One-population Proportion Assumptions</u>

1. The categorical sample data should be collected randomly or be representative of the population.
2. Data values within the sample should be independent of each other.
3. There should be at least ten successes and at least ten failures.

Let us check these assumptions in the previous confidence interval for the proportion of female bears.

1. Random Categorical Data? *Yes. This data was random and gender is a categorical variable.*

2. Data values within the sample should be independent of each other. *This can be difficult to determine. It should not be the same bear measured multiple times. In addition, if one bear is female it should not change the probability of other bears being female. It is likely safe to assume these are true in this case.*

3. At least ten successes? *Yes. There were 19 female bears in the data, which is more than ten.*
   At least ten failures? *Yes. There were 54−19 = 35 bears that were not female in the data which is more than ten.*

Overall, it appears the data does satisfy the requirements for using the formula and so the confidence interval will be relatively accurate.

<u>Bootstrapping</u>

Is there a way to make a confidence interval if the data did not meet the assumptions?

It depends on which assumptions. One technique that is sometimes used is called "Bootstrapping". Bootstrapping does require the sample to be representative of the population. That usually means it was collected randomly with data values that are independent of each other. As long as you have those two assumptions, you can bootstrap.

<u>One-population Bootstrap Assumptions</u>

1. The sample data should be collected randomly or be representative of the population.
2. Data values within the sample should be independent of each other.

Bootstrapping does not use formula for standard error and critical values like Z-scores or T-scores. It calculates the middle 95%, 99% or 90% directly using a bootstrap sampling distribution. Since bootstrapping is not tied to formulas and critical values, it does not require the sampling distribution to be normal or to match up with a specific theoretical curve.

The idea of bootstrapping is to create a theoretical population by assuming that the population is just many copies of your one real representative random sample. In practice, bootstrapping uses computers to take thousands for random samples with replacement from your one representative random sample. It randomly selects a value from your data, but puts the value back before picking another value randomly. This allows us to get the same value in a bootstrap sample multiple times. It then calculates the statistic like the mean or proportion from all of the bootstrap samples. These are sometimes called "bootstrap statistics". Putting all the bootstrap statistics on the same graph gives a "bootstrap sampling distribution". If you find the computer find the cutoffs for the middle 95% of the bootstrap distribution, you have an estimated 95% confidence interval.

Bootstrapping: Taking many random samples values from one original real random sample with replacement.

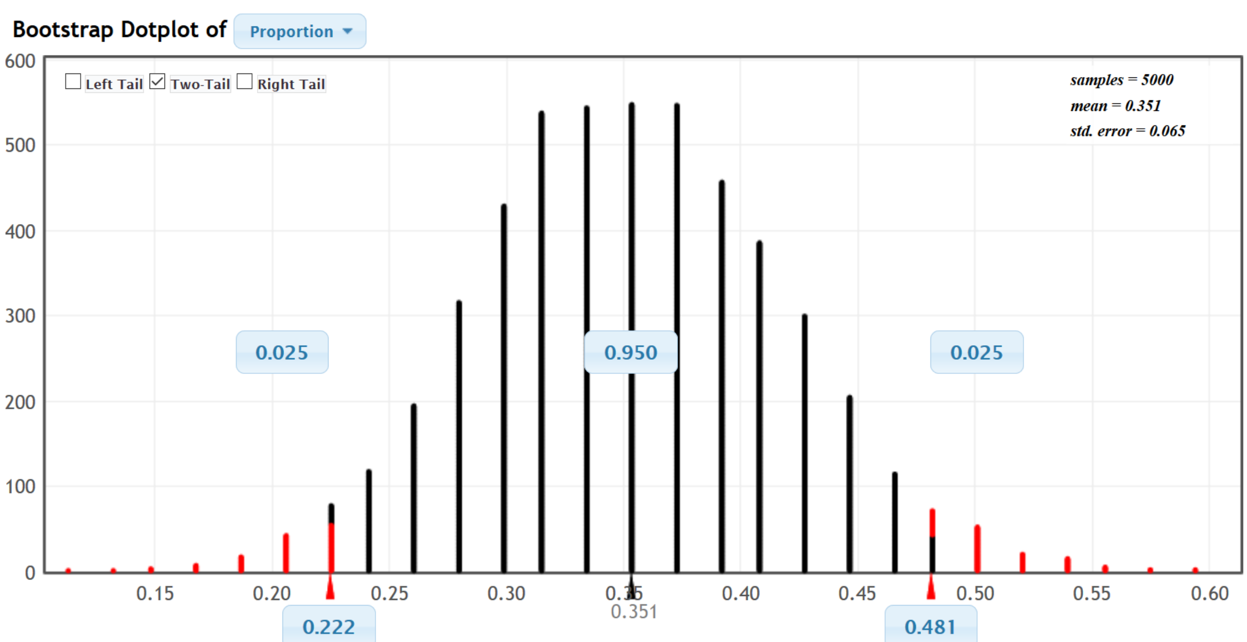
Bootstrap Sample: A simulated sample created by taking many random samples values from one original real random sample with replacement.

Bootstrap Statistic: A statistic calculated from a bootstrap sample.

Bootstrap Distribution: Putting many bootstrap statistics on the same graph in order to simulate the sampling variability in a population, calculate standard error, and create a confidence interval. The center of the bootstrap distribution is the original real sample statistic.

Female Bears Example

In the last example, we used the traditional Z critical value and standard error formula to create a confidence interval and estimate the population percentage of bears that are female. We could also use a bootstrap. Go to the "Bootstrap Confidence Interval" menu in StatKey at www.lock5stat.com and click on "CI for Single Proportion". Under "Edit Data" put in the random sample data count (19 female bears) and the total sample size (54 bears). Click "Generate 1000 Samples" a few times. Now click "Two-Tail". The default is 95%, but you can always change the middle proportion to 99% (0.99) or 90% (0.90) if needed. This problem was a 95% confidence interval, so we will leave the middle proportion as 0.95.



In a bootstrap confidence interval, the upper and lower limit of the confidence interval are found at the bottom right and left (0.222 and 0.481). Using these numbers, we are 95% confident that the population percentage of bears in this region of California that are female is between 22.2% and 48.1%. Notice that the upper limit, lower limit and standard error are very close to what we got by formula or Statcato. Notice that the shape of the bootstrap distribution is very normal. Though the bootstrap does not give us the margin of error like Statcato, we can use the formula we learned in the previous section. Remember the standard error and margin of error in this calculation are only reasonably accurate if the distribution is normal. Notice the margin of error is close to what we got by formula or Statcato.

Margin of Error = $\frac{(Upper\ Limit - Lower\ Limit)}{2} = \frac{(0.481-0.222)}{2} \approx 0.1295$

Key Notes about Bootstrapping

- A bootstrap distribution attempts to estimate and visualize the sampling variability in the population by creating a simulated population. Remember that standard error and margin of error are only accurate if the distribution is normal. So while we can estimate standard error and margin of error from a bootstrap, they may not be accurate if the bootstrap distribution is not normal.
- While a bootstrap distribution may be similar to a true sampling distribution from the population, there are important differences. The center of a bootstrap distribution is the sample statistic from the original real random data set. This makes the bootstrap ideal for estimating the confidence interval. A true sampling distribution is taking thousands of real samples from the population, so the center of a sampling distribution is the population parameter. We should not treat a true sampling distribution from the population the same as a bootstrap. If you have a sampling distribution, then the center can get a very accurate estimate of the population parameter. If you know the population parameter, you do not need a confidence interval. The middle 95% of a sampling distribution from an actual population is not a confidence interval.

Critical Value T-scores

In 1908, a statistician named William Gosset discovered that while Z-scores were very accurate for proportions, they were not very accurate when estimating mean averages, especially if the sample size was small. Small samples should have a larger margin of error than those indicated by Z-scores. To deal with this problem, he invented T-scores. His idea was that each sample size should have a different number of standard deviations. When Gosset invented the T-distribution, he worked for Guinness Beer and was not allowed to publish his work. He therefore published under the pseudonym "student". To this day, the T-distribution is often called the "Student T-Distribution" since it was invented by a then unknown author named "student".

T-scores are the same as Z-scores in the sense that they count the number of standard deviations or standard errors from the mean. However, they have a built in error correction for smaller data sets. For very large sample sizes, T-scores and Z-scores are about the same. For example, if we are using a 95% confidence level and our sample size is very large, then the T-score will be close to the Z-score of ± 1.96 standard deviations. When sample sizes are small, the T-scores become significantly greater than the Z-scores. This causes the margin of error to increase for small sample sizes. Remember, less random data should result in more error. We usually use Z-scores when estimating population proportions or percentages. We prefer to use T-scores when estimating population mean averages.

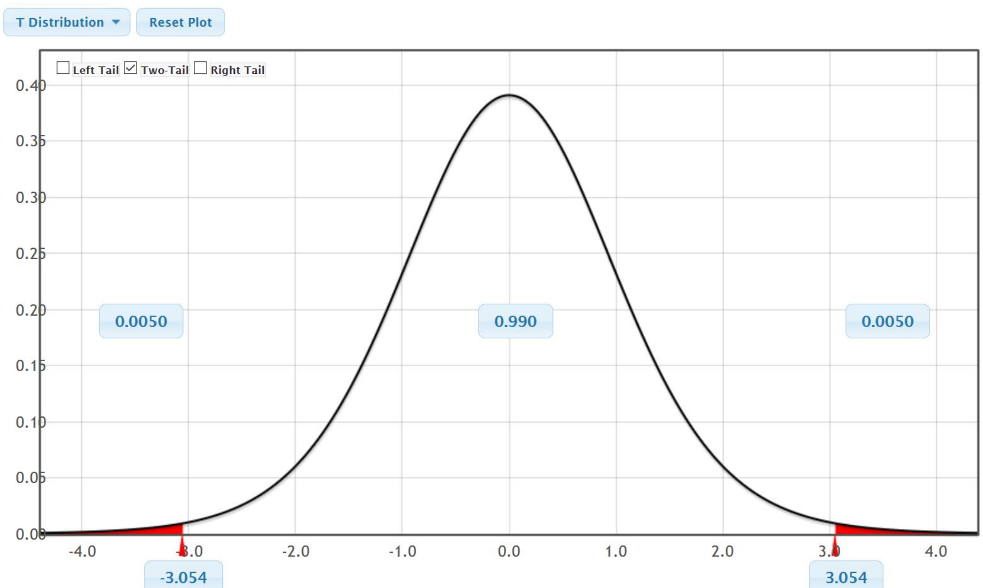
Note: You can use Z-scores for the mean if the sample size is large or if you know the population standard deviation exactly. However, we rarely know the population standard deviation with any certainty, especially when we do not even know the population mean. Also in large sample sizes, the T-scores are still accurate, so you might as well use the T-scores.

Degrees of Freedom

If you recall from previous sections, statistics like variance and standard deviation are based on a sum of squares divided by the degrees of freedom. For one sample, the degrees of freedom is usually equal to one less than the sample size ($df = n - 1$). Because of this, Gosset organized his T-scores not by sample size, but by degrees of freedom. Gosset calculated his T-scores with calculus and wrote them on charts. Before computers were invented, a statistician would first calculate the degrees of freedom and then look up the correct T-score on these charts. In modern times, T-scores can be easily calculated with computer programs like StatKey.

Example 1: Calculate the T-score critical value for a sample size n = 13 and a 99% confidence level.

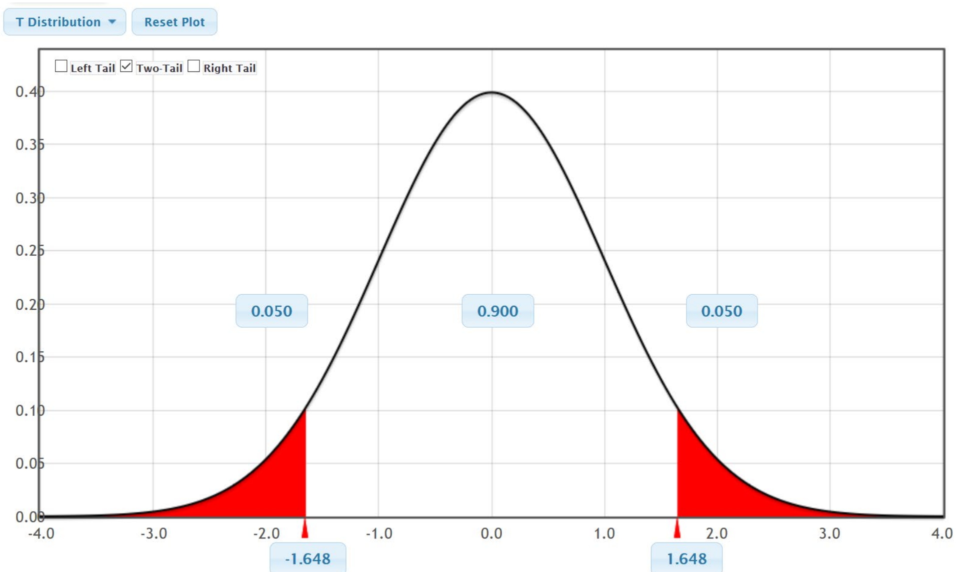
Go to www.lock5stat.com and click on "StatKey". Under the "theoretical distributions" menu, click on "t". Since the sample size is 13, the degrees of freedom will be df = 13−1 = 12. If we click on "two tail" and set the middle proportion to 0.99, we will get the following.



We see from the graph that critical value T-score for 99% confidence and 12 degrees of freedom is ± 3.054. Notice this is larger than the 99% confidence critical value Z-score (± 2.576). For smaller sample sizes, the T-scores are significantly greater than the Z-scores.

Example 2: Calculate the T-score critical value for a sample size n = 500 and a 90% confidence level.

Go to www.lock5stat.com and click on "StatKey". Under the "theoretical distributions" menu, click on "t". Since the sample size is 13, the degrees of freedom will be df = 500−1 = 499. If we click on "two tail" and set the middle proportion to 0.9, we will get the following.



We see from the graph that critical value T-score for 90% confidence and 499 degrees of freedom is ± 1.648. Notice this is very close to the 90% confidence critical value Z-score (± 1.645). For larger sample sizes, the T-scores and the Z-scores are about the same.

Summary of Critical Value T-scores

- T-scores (like Z-scores) count the number of standard deviations from the mean. In a sampling distribution of sample means, it counts how many standard errors we should be from the center of the sampling distribution for a given confidence level.
- T-scores are different for every sample size. They are usually organized by degrees of freedom. For one-population, the degrees of freedom is usually $df = n - 1$.
- T-scores are significantly larger than Z-scores for small sample sizes. The smaller the sample size, the larger the discrepancy between the T-score and Z-score.
- T-scores are about the same as Z-scores for large sample sizes.

One-Population Mean Confidence Interval

Let us look at the formula for calculating a one-population mean average confidence interval. Many computer programs to this day still use this formula.

Statisticians estimated the standard error for a sampling distribution for sample means with the following formula. The formula is surprisingly accurate and close to the standard error in an actual sampling distribution.

Standard Error Estimation Formula for Means = $\frac{s}{\sqrt{n}}$

Sample Standard Deviation = *s*
Sample Size = *n*

Here is the formula for a confidence interval estimate of the population mean. In order to estimate the margin of error, we multiply the standard error by the number of standard errors (T-score).

Sample Statistic ± Margin of Error
Sample Mean ± (T × Standard Error)

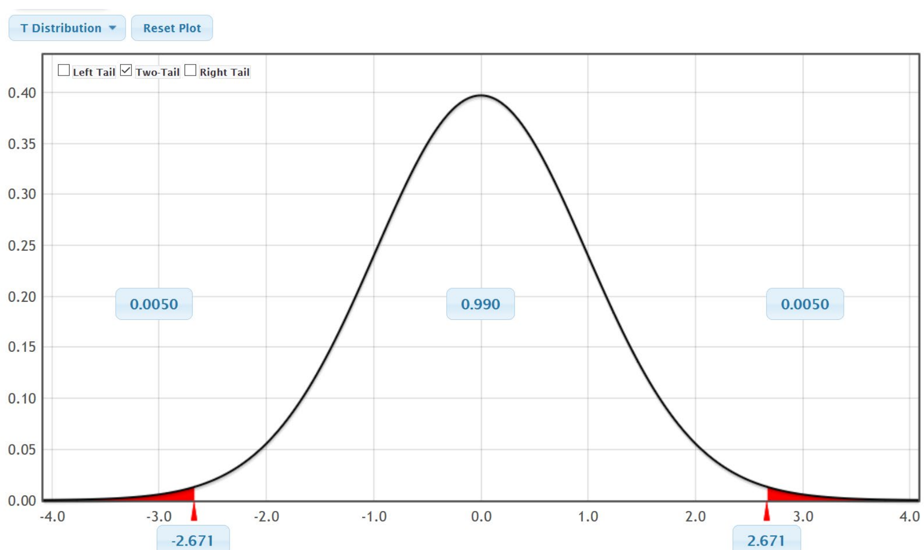
$\bar{x} \pm \left(T \frac{s}{\sqrt{n}}\right)$

Example 1: Calculating the confidence interval estimate of a population mean

A random sample of 54 bears in a region of California was taken. The weights of the bears showed a skewed right shape with a sample mean of 182.889 pounds and sample standard deviation of 121.801 pounds. Find the degrees of freedom and the critical value T-score. Then use the formula above to calculate a 99% confidence interval estimate for the population mean average weight of bears in this region of California.

Degrees of Freedom: df = n – 1 = 54 – 1 = 53.

Using the T-score calculator in StatKey we found that the critical Value T-score for 99% Confidence and 53 degrees of freedom is T = ± 2.671



Now we will replace the T-score with 2.671, $\bar{x}$ with 182.889, *n* with 54, and *s* with 121.801 into our formula and work it out. Remember to follow order of operations. Notice the standard error estimate is 16.575 pounds and the margin of error estimate is 44.272 pounds.

$\bar{x} \pm \left(T \frac{s}{\sqrt{n}}\right)$

$182.889 \pm 2.671 \times \frac{121.801}{\sqrt{54}}$

$182.889 \pm (2.671 \times 16.575)$

$182.889 \pm (44.272)$

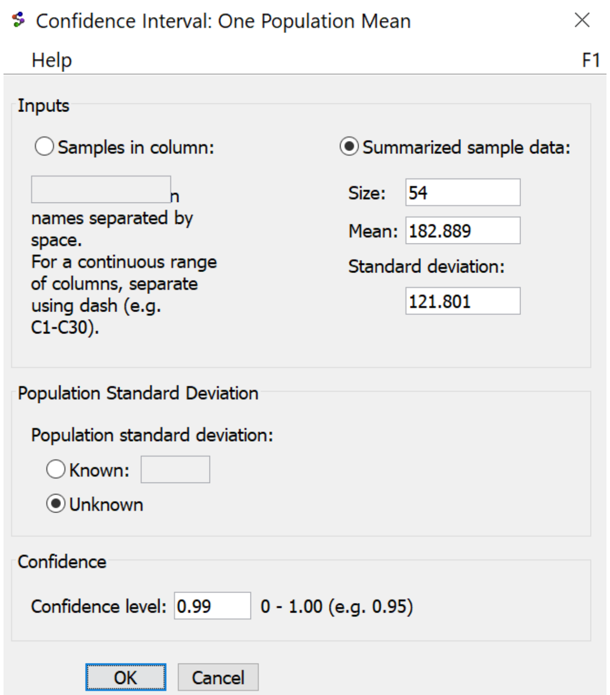
182.889 – 44.272 < Population Mean Average Weight of Bears in Pounds (μ) < 182.889 + 44.272

138.617 pounds < Population Mean Average Weight of Bears in Pounds (μ) < 272.161 pounds

We are 95% confident that the population mean average weight of bears in this region of California is in between 138.617 pounds and 272.161 pounds.

Note: While it is important to understand this formula, it is much easier to calculate this with a computer.

To calculate this confidence interval with Statcato, we will click on the "statistics" menu and then "confidence intervals". Click on "One-population mean". Under "Summary data", enter 182.889 for the mean, 121.801 for the standard deviation, and 54 for the number of trials. Set the confidence level to 0.99 and click OK. If we have the raw data, we could also put in the column "C1" where it says "samples in column".



Here is the Statcato printout. Notice the computer calculation is not exactly the same as the one we did with the formula and calculator. The computer did not round as much as we did. Computer calculations are usually much more accurate than calculator calculations because they tend to keep a lot more decimal places.

Confidence Interval - One population mean: confidence level = 0.99
Input: Summary data
σ unknown

Var	N	Mean	Stdev	Margin of Error	99.0%CI
summary	54.0	182.889	121.801	44.285	(138.6039, 227.1741)

It might be good to adjust our explanation sentence with the more accurate numbers from the computer.

We are 95% confident that the population mean average weight of bears in this region of California is in between 138.604 pounds and 272.174 pounds.

Key Question: How accurate is this confidence interval?

This confidence interval relies on a T-score and standard error so the sampling distribution for sample means must be normal for this formula to be accurate. If we look at the section on the central limit theorem, we remember that for a sampling distribution for random sample means to be normal, we need one of two things to be true. Either the data itself must be normal or the sample size must be at least 30. This gives rise to the assumptions or conditions required for mean average confidence interval calculations. For the formula approach to be accurate, the following must be true. If any of these assumptions are not met, then the confidence interval may not be accurate.

One-population Mean Assumptions

1. The quantitative sample data should be collected randomly or be representative of the population.
2. Data values within the sample should be independent of each other.
3. The sample size should be at least 30 or have a nearly normal shape.

Let us check these assumptions in the previous confidence interval for the mean average weight of bears.

1. Random Quantitative Data? *Yes. This data was random and weight in pounds is a quantitative variable.*

2. Data values within the sample should be independent of each other. *This can be difficult to determine. It should not be the same bear measured multiple times. These bears were probably tagged so they probably did not accidentally measure the same bear multiple times. Also, one bears weight should not change the probability of other bear having a certain weight. This data may not pass this assumption. Let us assume we see a bear that is eating well and is very heavy. Then there may be a higher probability of other bears being heavy in the same area.*

3. The sample data must be nearly normal or the sample size must be at least 30? *We see from the histogram that this data was skewed right, but the sample size was 54 (at least 30). Therefore, it does pass the 30 or normal requirement. Remember only one of the two need to be true for it to pass.*

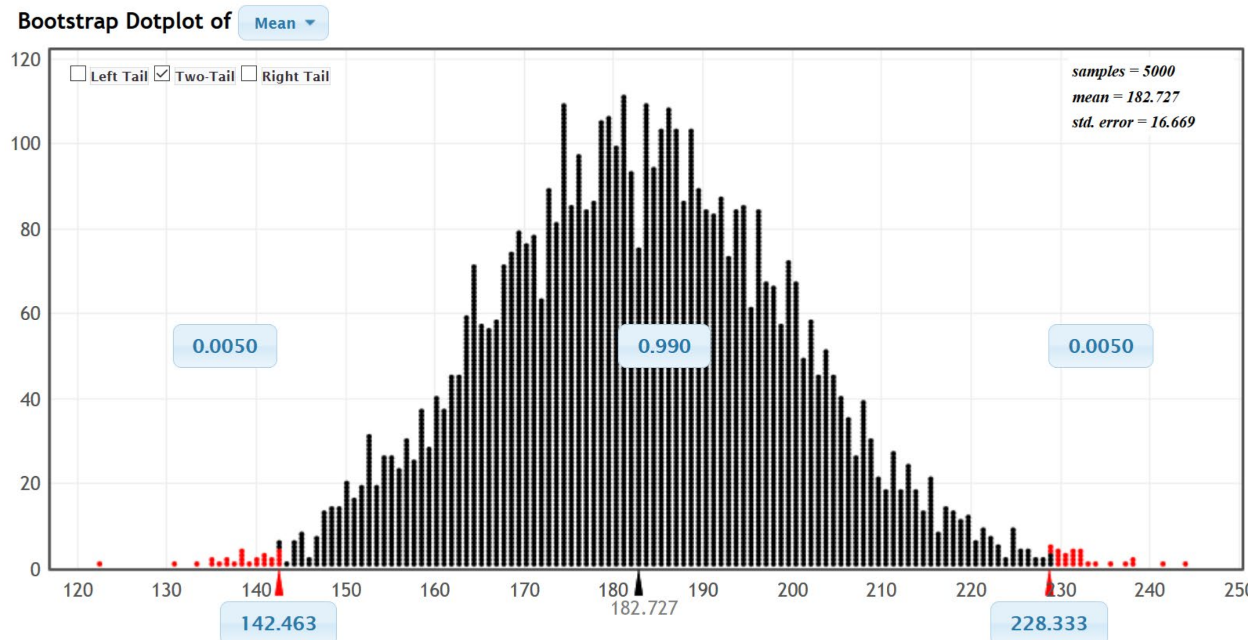
The data did satisfy the random requirement and the at least 30 or normal requirement. If the data does satisfy the independence assumption, then the data would satisfy the overall requirements for using the formula and so the confidence interval will be relatively accurate.

Could we have calculated this confidence interval with a bootstrap distribution?

Remember, the accuracy of a bootstrap is tied to the quality of the original sample data set. This data set was collected randomly but may fail the independence requirement.

Bear Weight Example

In this last example, we used the traditional T critical value and standard error formula to create a confidence interval and estimate the population mean average weight of bears. We could also use a bootstrap. First, go to the “Bear Data” at www.matt-teachout.org and copy the bear weight column of data. Now go to the “Bootstrap Confidence Interval” menu in StatKey at www.lock5stat.com and click on “CI for Single Mean, Median, St.Dev.” Under “Edit Data”, paste in the raw quantitative bear weight data. Make sure to check the “Header Row” box since this data set had a title and push “OK”. Click “Generate 1000 Samples” a few times. Now click “Two-Tail”. The default is 95%, but you can change the middle proportion to 99%. This problem was a 99% confidence interval, so we will change the middle proportion to 99% (0.99).



We see that the bootstrap distribution is normally distributed. The confidence interval has a lower limit of 142.463 pounds, an upper limit of 228.333 pounds and a standard error of 16.669. Notice these numbers are relatively close to the same numbers we got by formula and Statcato. Since the confidence interval is normal, we can use the margin of error back-solving formula to find the approximate margin of error.

Margin of Error = $\frac{(Upper\ Limit - Lower\ Limit)}{2} = \frac{(228.333-142.463)}{2} \approx 42.935$
