Section 2G – One-Population Variance & Standard Deviation Confidence Intervals

It is often vital to estimate the standard deviation of a population. However, it can be very difficult to estimate with any accuracy, especially if we only have one random sample. Remember our principle of sampling variability. We saw in previous sections that sample standard deviations (s) will usually be very different from each other and can be very different from the population standard deviation (σ).

<u>One-Population Variance and Standard Deviation Confidence Intervals</u>

Recall that the population standard deviation is the square root of the population variance (σ^2). So we often estimate the population variance and then simply take the square root of the variance to get the standard deviation. The principle of sampling variability also applies to variance. Sample variances (s^2) will usually be very different from each other and may be very different from the population variance (σ^2).

Sampling distributions for variance are usually skewed to the right and rarely have a normal shape. Since the sampling distribution is not normal or symmetric, we cannot use the traditional formula approach of the sample statistic ± margin of error. That formula will not work.

It is important to note that even though the quantitative data itself may be normal, the sampling distribution for variance still may be skewed to the right. Statisticians discovered that as long as the quantitative data itself was normal, the sampling distribution for sample variance follows a Chi-Squared distribution with degrees of freedom ($df = n - 1$). So the formula for making a confidence interval to estimate population variance uses Chi-Squared critical values (χ^2). It is important to note that no matter what the sample size is, the sample data must be normal for this formula to work. If the data is not normal, we must resort to another technique like bootstrapping.

<u>Calculating Chi-Squared Critical Values</u>

Example 1: Calculate the Chi-Squared (χ^2) critical values for a sample size n = 10 and a 90% confidence level.

Go to www.lock5stat.com and click on "StatKey". Under the "theoretical distributions" menu, click on "χ^2". Since the sample size is 10, the degrees of freedom will be df = 10 − 1 = 9. If we click on "two tail" and set the middle proportion to 0.90, we will get the following. Variance is calculated with a sum of squares. That makes it impossible to ever be negative. That also means the upper and lower Chi-Squared critical values will be very different. The upper Chi-Squared critical value will be the larger number on the right and the lower Chi-Squared critical value will be the smaller number on the left. Both will be positive. You can see that Chi-Squared looks skewed to the right for nine degrees of freedom.
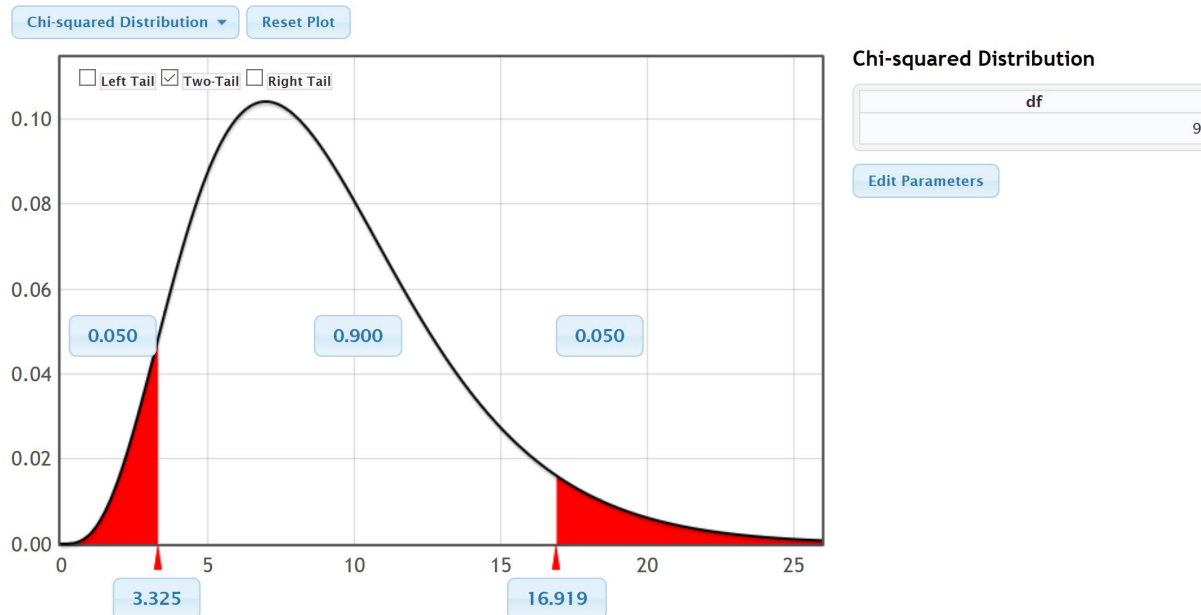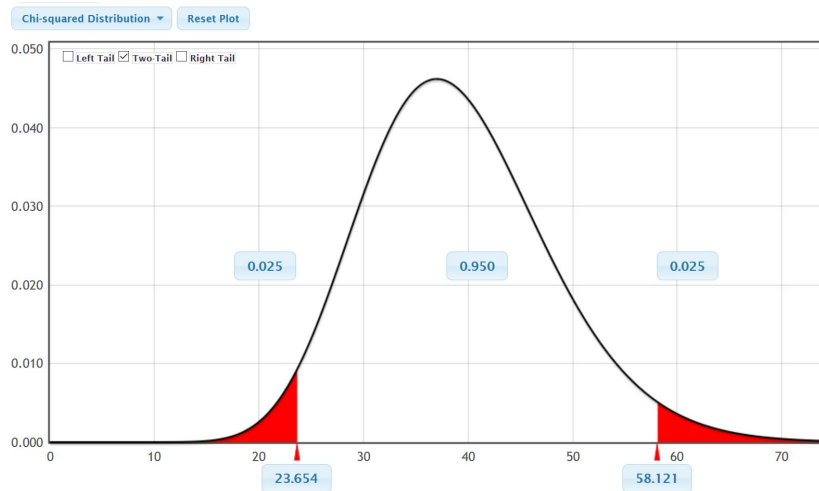
We see from the graph that upper critical value for 90% confidence and 9 degrees of freedom is 16.919 and the lower critical value for 90% confidence and 9 degrees of freedom is 3.325.

Example 2: Calculate the Chi-Squared (χ^2) critical values for a sample size n = 40 and a 95% confidence level.

Go to www.lock5stat.com and click on "StatKey". Under the "theoretical distributions" menu, click on "χ^2". Since the sample size is 40, the degrees of freedom will be df = 40 − 1 = 39. If we click on "two tail" and set the middle proportion to 0.95, we will get the following. Notice that the upper and lower Chi-Squared critical values will both be positive, but will be very different. Also, notice that as the degrees of freedom increases, the chi-squared distribution looks less skewed to the right.

We see from the graph that upper critical value for 95% confidence and 39 degrees of freedom is 58.121 and the lower critical value for 95% confidence and 39 degrees of freedom is 23.654.

Confidence Interval Formulas for Variance and Standard Deviation

Here is the confidence interval formulas for estimating population variance. Taking the square root gives us the formula for estimating population standard deviation as well. Notice that the upper critical value is on the left and lower critical value is on the right. When you divide by a larger number, the overall fraction is smaller.

$$\frac{s^2(n-1)}{\chi^2_{upper}} < \text{Population Variance } (\sigma^2) < \frac{s^2(n-1)}{\chi^2_{lower}}$$

$$\sqrt{\frac{s^2(n-1)}{\chi^2_{upper}}} < \text{Population Standard Deviation } (\sigma) < \sqrt{\frac{s^2(n-1)}{\chi^2_{lower}}}$$

Example 1: We measured the heights in inches of 40 randomly selected men. The data showed a normal shape. The sample standard deviation was 3.020 inches. Use the Chi-squared critical values and the formulas above to create a 95% confidence interval for the population variance and the population standard deviation.

We calculated the Chi-squared critical values in the previous example. The upper critical value was 58.121 and the lower critical value was 23.654.

Sample Variance $(s^2) = (3.020)^2 = 9.1204$

Degrees of Freedom $(n-1) = 40 - 1 = 39$

$$\frac{s^2(n-1)}{\chi^2_{upper}} < \text{Population Variance } (\sigma^2) < \frac{s^2(n-1)}{\chi^2_{lower}}$$

$$\frac{9.1204\,(40-1)}{58.121} < \text{Population Variance } (\sigma^2) < \frac{9.1204\,(40-1)}{23.654}$$

$$\frac{9.1204\,(39)}{58.121} < \text{Population Variance } (\sigma^2) < \frac{9.1204\,(39)}{23.654}$$

6.11992 < Population Variance (σ^2) < 15.03744

Variance Confidence Interval Sentence: We are 95% confident that the population variance for all men is between 6.11992 and 15.03744 square inches. *(Notice that variance is in square units since it is the standard deviation squared.)*

If we take the square root of our answers, we can get an estimate of the population standard deviation.

$\sqrt{6.11992}$ < Population Standard Deviation (σ) < $\sqrt{15.03744}$

2.47 inches < Population Standard Deviation (σ) < 3.88 inches

Standard Deviation Confidence Interval Sentence: We are 95% confident that the population standard deviation for all men is between 2.47 inches and 3.88 inches.

As with all calculations, it is much easier and more accurate to calculate these with a computer program.

In Statcato, we can go to the "Statistics" menu and click on confidence intervals. If we click on "1-population variance" and enter the sample size (40) and sample standard deviation 3.020 under summary data, we get the following. Notice you can also calculate the confidence interval from raw data or by entering the sample variance.

1-Population Variance   ✕

Help   F1

Inputs

◯ Samples in column:

⦿ Summarized sample data:

    Sample Size: 40

    ◯ Variance:

    ⦿ Standard deviation: 3.020

Confidence

Confidence level: 0.95   0 - 1.00 (e.g. 0.95)

[OK]   [Cancel]

Notice Statcato gave us almost the same confidence intervals for variance and standard deviation as we calculated with the formula.

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

N	Variance	Stdev	95.0%CI Variance	95.0%CI Stdev
40	9.120	3.02	(6.1200, 15.0372)	(2.4739, 3.8778)

Here are the assumptions for making a confidence interval to estimate population variance or standard deviation.

One-Population Variance or Standard Deviation Confidence Interval 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.
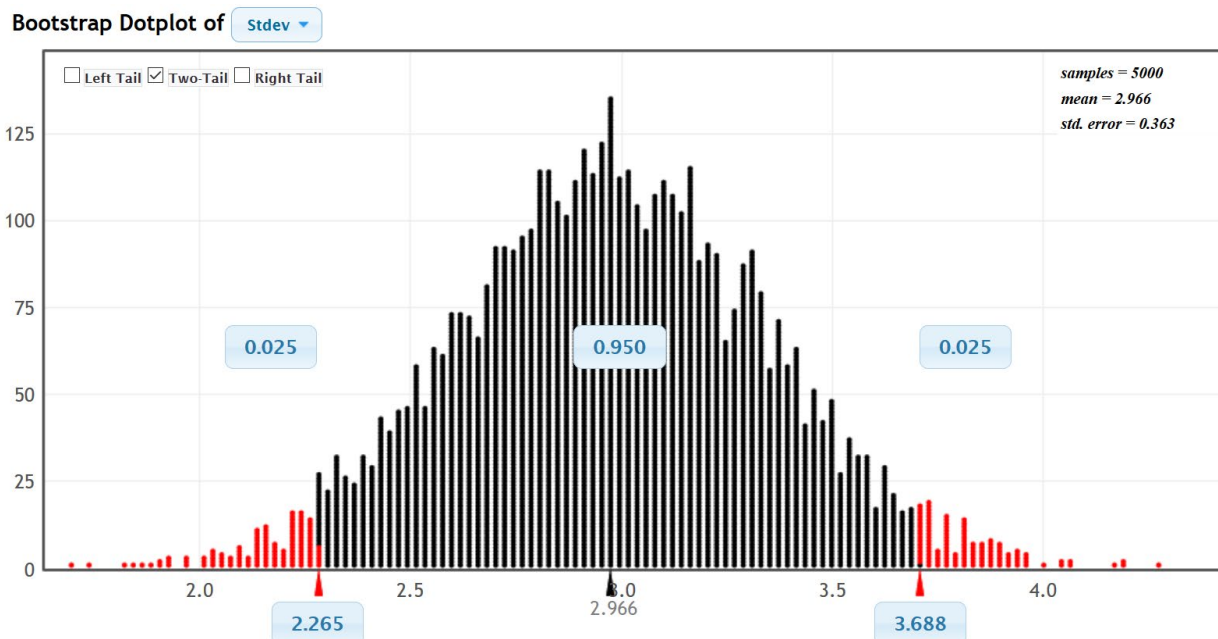
3. The sample data must be normal.

Does the men's height data meet these assumptions? Let us check them.

1. Is this random quantitative sample data or sample data that represents the population? Yes. Height is quantitative and this was a random sample.

2. Are the data values within the sample independent of each other? Yes. A random sample out of a large population will be unlikely to accidently get men that are family members. One man's height should not change the probability of another man's height.

3. Is the sample data normal? We did not see the histogram, but the problem did state that the data was normal.

Men's Height Bootstrap Example

Let us use a bootstrap distribution to estimate the confidence interval for population standard deviation for men's height. First go to the "Health Data" at www.matt-teachout.org and copy the men's height 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 men's height data. Make sure to check the "Header Row" box since this data set had a title and push "OK". Under the "Generate Dot plot of" menu, click on "St.Dev." (standard deviation). Now click "Generate 1000 Samples" a few times. Then click "Two-Tail". Make sure the middle proportion is 0.95 (95%).

Bootstrap Dotplot of [Stdev ▾]

☐ Left Tail ☑ Two-Tail ☐ Right Tail

samples = 5000
mean = 2.966
std. error = 0.363

125

100

75

0.025   0.950   0.025

50

25

0

2.0   2.5   3.0   3.5   4.0
                 2.966

2.265   3.688

Notice that the upper and lower limits of the confidence interval are close to what we got with formula or Statcato.