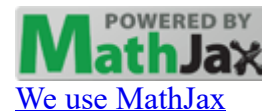


## Confidence Intervals for Variances and Standard Deviations



We have learned that estimates of population means can be made from sample means, and confidence intervals can be constructed to better describe those estimates. Similarly, we can estimate a population standard deviation from a sample standard deviation, and when the original population is normally distributed, we can construct confidence intervals of the standard deviation as well.

### The Theory

Variances and standard deviations are a very different type of measure than an average, so we can expect some major differences in the way estimates are made.

We know that the population variance formula, when used on a sample, does not give an unbiased estimate of the population variance. In fact, it tends to underestimate the actual population variance. For that reason, there are two formulas for variance, one for a population and one for a sample. The sample variance formula is an unbiased estimator of the population variance. (Unfortunately, the sample standard deviation is still a biased estimator.)

Also, both variance and standard deviation are nonnegative numbers. Since neither can take on a negative value, the domain of the probability distribution for either one is not  $(-\infty, \infty)$ , thus the normal distribution cannot be the distribution of a variance or a standard deviation. The correct PDF must have a domain of  $[0, \infty)$ . It can be shown that if the original population of data is normally distributed, then the expression  $\frac{(n-1)s^2}{\sigma^2}$  has a chi-square distribution with  $n-1$  degrees of freedom.

The chi-square distribution of the quantity  $\frac{(n-1)s^2}{\sigma^2}$  allows us to construct confidence intervals for the variance and the standard deviation (when the original population of data is normally distributed). For a confidence level  $1 - \alpha$ , we will have the inequality  $\chi_{1-\alpha/2}^2 \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi_{\alpha/2}^2$ . Solving this inequality for the population variance  $\sigma^2$ , and then the population standard deviation  $\sigma$ , leads us to the following pair of confidence intervals.

$$\frac{(n-1)s^2}{\chi_{\alpha/2}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{1-\alpha/2}^2}$$

$$\sqrt{\frac{(n-1)s^2}{\chi_{\alpha/2}^2}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{\chi_{1-\alpha/2}^2}}$$

It is worth noting that since the chi-square distribution is not symmetric, we will be obtaining confidence intervals that are not symmetric about the point estimate.

## Example

A statistician chooses 27 randomly selected dates, and when examining the occupancy records of a particular motel for those dates, finds a standard deviation of 5.86 rooms rented. If the number of rooms rented is normally distributed, find the 95% confidence interval for the population standard deviation of the number of rooms rented.

For a sample size of  $n = 27$ , we will have  $df = n - 1 = 26$  degrees of freedom. For a 95% confidence interval, we have  $\alpha = 0.05$ , which gives 2.5% of the area at each end of the chi-square distribution. We find values of  $\chi^2_{0.975} = 13.844$  and  $\chi^2_{0.025} = 41.923$ . Evaluating  $\frac{(n-1)s^2}{\chi^2}$ , we obtain 21.297 and 64.492. This leads to the inequalities  $21.297 \leq \sigma^2 \leq 64.492$  for the variance, and taking square roots,  $4.615 \leq \sigma \leq 8.031$  for the standard deviation.

---