

t Distribution

Prerequisites

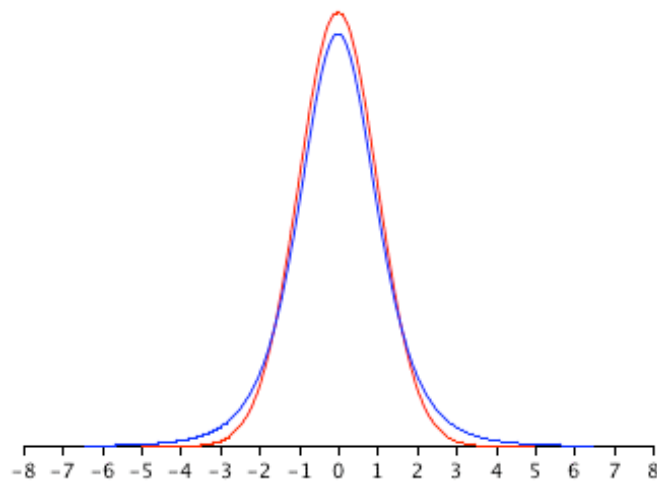
[Normal Distribution](#), [Areas Under Normal Distributions](#), [Degrees of Freedom](#), [Confidence Interval for the Mean](#)

In the introduction to normal distributions it was shown that 95% of the area of a normal distribution is within 1.96 standard deviations of the mean. Therefore if you randomly sampled a value from a normal distribution with a mean of 100, the probability it would be within 1.96σ of 100 is 0.95. Similarly, if you sample N values from the population, the probability that the sample mean (M) will be within $1.96 \sigma_M$ of 100 is 0.95.

Now consider the case in which you have a normal distribution but you do not know the standard deviation. You sample N values and compute the sample mean (M) and estimate the standard error of the mean (σ_M) with (s_M). What is the probability that M will be within $1.96 s_M$ of the population mean (μ)? This is a difficult problem because there are two ways in which M could be more than $1.96 s_M$ from μ : (1) M could, by chance, be either very high or very low and (2) s_M could, by chance, be very low. Intuitively, it makes sense that the probability of being within 1.96 standard errors of the mean should be smaller than in the case when the standard deviation is known (and cannot be underestimated). But exactly how much smaller? Fortunately, the way to work out this type of problem was solved in the early 20th century by W. S. Gossett who determined the distribution of a mean divided by an **estimate** of the standard error. This distribution is called the *student's t distribution* or sometimes just the *t distribution*. Gossett worked out the t distribution and associated statistical tests while working for a brewery in Ireland. Because of a contractual agreement with the brewery, he published the article under the pseudo name "Student." That is why the t test is called the "Student's t."

The t distribution is very similar to the normal distribution when the estimate of variance is based on many [degrees of freedom](#) but has relatively more scores in its tails when there are fewer degrees of freedom. Figure 1 shows the t distribution with 4 degrees of freedom and the standard normal distribution. Notice that the normal distribution has relatively more scores in the center of the distribution and the t distribution has relatively more in the tails. The t distribution is therefore [leptokurtic](#).

Figure 1. A comparison of the t distribution with 4 df (in blue) and the standard normal distribution (in red).



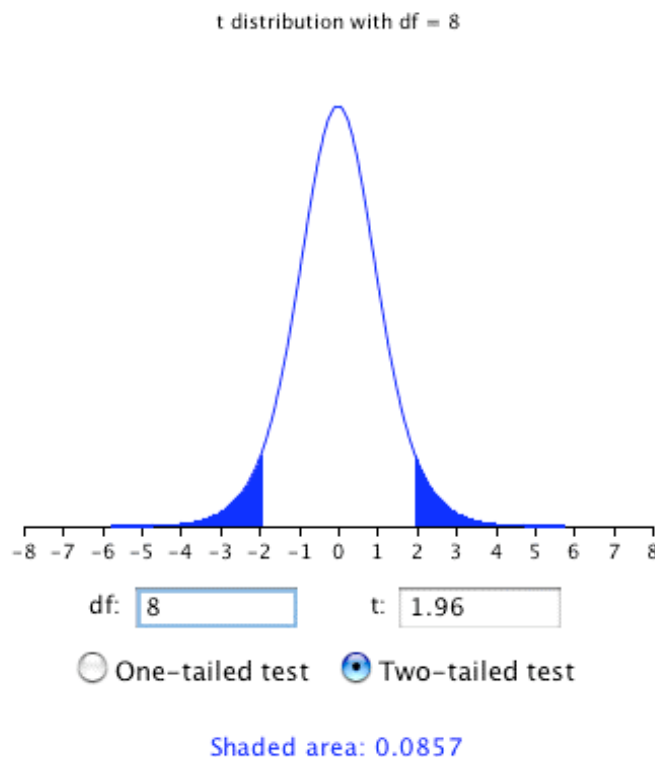
Since the t distribution is leptokurtic, the percentage of the distribution within 1.96 standard deviations of the mean is less than the 95% for the normal distribution. Table 1 shows the number of standard deviations from the mean required to contain 95% and 99% of the area of the t distribution for various degrees of freedom. These are the values of t that you use in a confidence interval. The corresponding values for the normal distribution are 1.96 and 2.58 respectively. Notice that with few degrees of freedom, the values of t are much higher than the corresponding values for a normal distribution and that the difference decreases as the degrees of freedom increase. The values in Table 1 can be obtained from the ["Find t for a confidence interval" calculator](#).

Table 1. Abbreviated t table.

df	0.95	0.99
2	4.303	9.925
3	3.182	5.841
4	2.776	4.604
5	2.571	4.032
8	2.306	3.355
10	2.228	3.169
20	2.086	2.845
50	2.009	2.678
100	1.984	2.626

Returning to the problem posed at the beginning of this section, suppose you sampled 9 values from a normal population and estimated the standard error of the mean (σ_M) with (s_M). What is the probability that M would be within $1.96s_M$ of μ ? Since the sample size is 9, there are $N - 1 = 8$ df. From Table 1 you can see that with 8 df the probability is 0.95 that the mean will be within $2.306 s_M$ of μ . As shown in Figure 2, the "[t distribution](#)" calculator can be used to find that 0.086 of the area of a t distribution is more than 1.96 standard deviations from the mean, so the probability that M would be less than $1.96s_M$ from μ is $1 - 0.086 = 0.914$.

Figure 2. Area more than 1.96 standard deviations from the mean in a t distribution with 8 df. Note that the two-tailed button is selected so that the area in both tails will be included.



As expected, this probability is less than 0.95 that would have been obtained if

σ_M had been known instead of estimated.

[Online: Calculator: Find t for a confidence interval](#)

[Online: Calculator: t distribution](#)