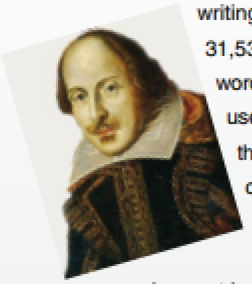
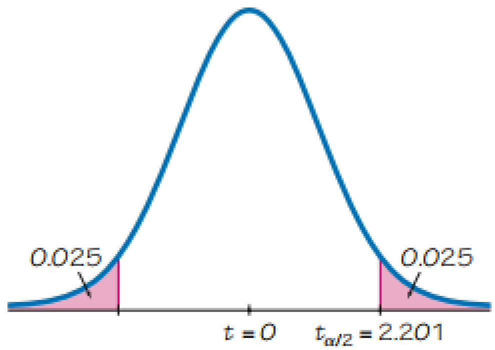
Shakespeare's Vocabulary



According to Bradley Efron and Ronald Thisted, Shakespeare's writings included 31,534 different words. They used probability theory to conclude that Shakespeare probably knew at least another 35,000 words that he didn't use in his writings. The problem of estimating the size of a population is an important problem often encountered in ecology studies, but the result given here is another interesting application. (See "Estimating the Number of Unseen Species: How Many Words Did Shakespeare Know?", in *Biometrika*, Vol. 63, No. 3.)

Figure 7-4 Critical Value $t_{\alpha/2}$



Confidence intervals can be easily constructed with technology or they can be manually constructed by using the following procedure.

Procedure for Constructing a Confidence Interval for μ

1. Verify that the two requirements in the preceding box are satisfied.
2. With σ unknown (as is usually the case), use $n - 1$ degrees of freedom and use technology or a t distribution table to find the critical value $t_{\alpha/2}$ that corresponds to the desired confidence level (as in the preceding Example 1).
3. Evaluate the margin of error using $E = t_{\alpha/2} \cdot s/\sqrt{n}$.
4. Using the value of the calculated margin of error E and the value of the sample mean $\bar{x}$, find the values of the confidence interval limits: $\bar{x} - E$ and $\bar{x} + E$. Substitute those values in one of these three general formats for the confidence interval:

$$\bar{x} - E < \mu < \bar{x} + E$$

or

$$\bar{x} \pm E$$

or

$$(\bar{x} - E, \bar{x} + E)$$

5. Round the resulting confidence interval limits using the round-off rule given in the preceding box.

Interpreting a Confidence Interval Be careful to interpret confidence intervals correctly. If we obtain a 95% confidence interval such as $58.1 < \mu < 63.3$, there is a correct interpretation and many incorrect interpretations.

Correct: "We are 95% confident that the interval from 58.1 to 63.3 actually does contain the true value of μ." This means that if we were to select many different samples of the same size and construct the corresponding confidence intervals, in the long run 95% of them would actually contain the value of μ. (This correct interpretation refers to the *success rate of the process* being used to estimate the population mean.)

Incorrect: Because μ is a fixed constant, it would be incorrect to say "there is a 95% chance that μ will fall between 58.1 and 63.3." It would also be incorrect to say that "95% of all data values are between 58.1 and 63.3," or that "95% of sample means fall between 58.1 and 63.3." Other possible incorrect interpretations are limited only by the imagination of the reader.