Arkansas Tech University
MATH 3513: Applied Statistics I
Dr. Marcel B. Finan

## 5.3 Small-sample Confidence Interval for a Population Mean

The concept of confidence intervals for the population mean discussed in Section 5.1 requires samples with large size ($n > 30$). This process does not apply for small samples. However, for small samples the normal distribution is being replaced by the so called the **Student's t distribution** in finding confidence intervals for a population mean.
Before we proceed with our discussion, we present the concept of **degrees of freedom.** In statistics, the number of degrees of freedom is the number of values in the final calculation of a statistic that are free to vary. For example, suppose that a final calculation of a statistic boils down to an equation of the form $a + b + c + d = 10$. The variables $a, b$, and $c$ can be assigned any number. However, the variable $d$ depends on the values of $a, b$, and $c$. In this case, we say that the number of degrees of freedom is 3.

**Student's $t$ distribution**
Let $\overline{X}$ be the mean of a simple random sample of size $n < 30$ taken from an approximately normally distributed population. The **Student's $t$ distribution** with **degrees of freedom** $n - 1$ is the distribution of the random variable

$$t_{n-1} = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

where $\mu$ is the population mean and $s$ is the sample standard deviation.
The pdf of the Student's distribution is different for different degrees of freedom. Figure 5.3.1 presents plots of the pdf for several choices of the degrees of freedom.

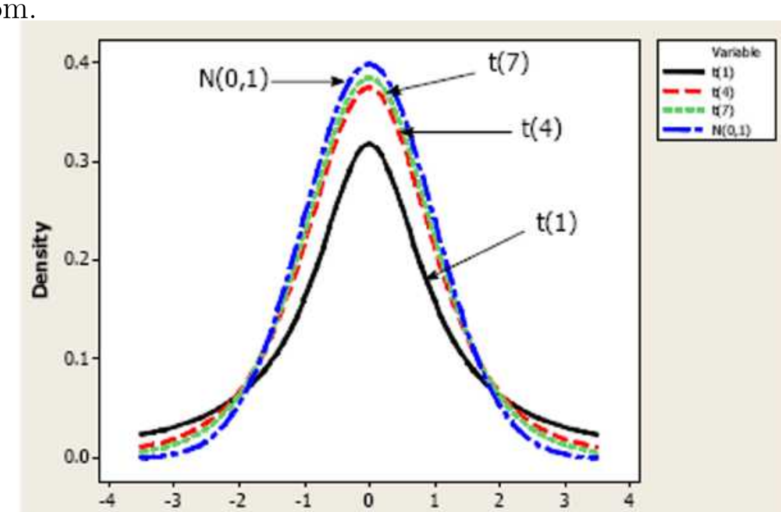


Figure 5.3.1

From Figure 5.3.1, we see that the $t$ curves are more spread out than the standard normal distribution, but the amount of extra spread decreases as $n-1$ increases.
Table A.3 (in Appendix A of your book), called a $t$ **table**, provides probabilities associated with the Student's $t$ distribution.

**Example 5.3.1**
Find the value for the $t_{12}$ distribution whose upper-tail probability is 0.025.

**Solution.**
Look down the column headed "0.025" to the row corresponding to 12 degrees of freedom. The value for $t_{12}$ is 2.179 ■

**Example 5.3.2**
Find the value for the $t_{12}$ distribution whose lower-tail probability is 0.025.

**Solution.**
Look down the column headed "0.025" to the row corresponding to 12 degrees of freedom. The value for $t_{12}$ is 2.179. This value cuts off an area, or probability, of 2.5% in the upper tail. The value whose lower-tail probability is 2.5% is $-2.179$ ■

**Confidence Intervals Using the Student's $t$ Distribution**
Consider a sample of size $n$ drawn from a normally distributed population. A $100(1-\alpha)\%$ confidence interval for the population mean $\mu$ is the interval

$$\left(\overline{X} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}, \overline{X} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\right)$$

where $\overline{X}$ is the mean of the sample, $s$ its standard deviation and $t_{n-1,\alpha/2}$ is the value which cuts off an area of $\frac{\alpha}{2}$ in the right-hand tail. See Figure 5.3.2.

**Example 5.3.3**
A sample of taxicabs had the following consumption of gasoline, in gallons, on a particular day: 32, 16, 25, 22, 27, 19, 20. Assume that the distribution of gasoline usage is normal. Find the 95% confidence interval.

**Solution.**
The mean of the sample is

$$\overline{X} = \frac{32+16+25+22+27+19+20}{7} = 23$$

and the standard deviation is

$$s = \sqrt{\frac{(32-23)^2+(16-23)^2+(25-23)^2+(22-23)^2+(27-23)^2+(19-23)^2+(20-23)^2}{6}}$$
$$\approx 5.416.$$

Also, using Table A.3, we find $t_{6,0.0025} = 2.447$. Hence, the 95% confidence interval is

$$\left(23 - 2.447 \times \frac{5.416}{\sqrt{7}}, 23 + 2.447 \times \frac{5.416}{\sqrt{7}}\right) \approx (18, 28) \blacksquare$$

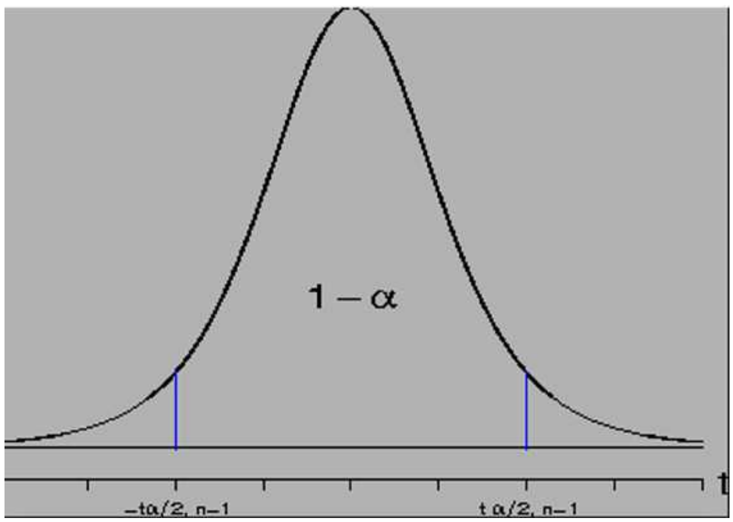


Figure 5.3.2

**<u>When is it Appropriate to use the Student $t$ Distribution?</u>**

One must decide whether a population is approximately normal by examining the sample. A reasonable way to proceed is to construct a boxplot or dotplot of the sample. If these plots do not reveal a strong asymmetry or any outliers then in most cases the Student's $t$ distribution will be reliable. See Examples 5.19 and 5.20 in the textbook.

The Student's $t$ distribution can be used to compute one-sided confidence intervals. The formulas are analogous to those used with large samples. Namely, we have either $\left(\overline{X} - t_{n-1,\alpha}\frac{s}{\sqrt{n}}, \infty\right)$ or $\left(-\infty, \overline{X} + t_{n-1,\alpha}\frac{s}{\sqrt{n}}\right)$.

**Remark 5.3.1**
If $\sigma$ is known then use the standard normal distribution instead of the Student's $t$ distribution.