CONFIDENCE INTERVALS WITH UNKNOWN VARIANCE
  - Large Sample Case
  - Student $t$-Distribution

REGRESSION APPLICATIONS
  - Confidence Intervals for Betas
  - Confidence Intervals for Conditional Means

For next time:
  - Devore, Sections 12.3, 12.4, 13.4
• **STUDENT $t$-DISTRIBUTION**

*For any random sample $(X_1, \ldots, X_n)$ from a normal distribution with mean, $\mu$, the standardized sample mean statistic

$$T_{n-1} = \frac{\bar{X}_n - \mu}{S_n / \sqrt{n}}$$

has the $t$-distribution with $n-1$ degrees of freedom.*

• **Comparison with NORMAL ($\alpha = .025$)**

$$t_{\alpha,10} = 2.23$$
$$t_{\alpha,20} = 2.09$$
$$t_{\alpha,100} = 1.98$$
$$t_{\alpha,\infty} = 1.96$$
CUTTING MACHINE (Lecture 12)

If $\alpha = .05$, then a 95% confidence interval for $\mu = E(X)$ is now given by:

$$(1) \quad \mu \in \left[ \bar{X}_n \pm t_{.025,n-1} \frac{s_n}{\sqrt{n}} \right] = \left[ \bar{X}_n - t_{.025,n-1} \frac{s_n}{\sqrt{n}}, \bar{X}_n + t_{.025,n-1} \frac{s_n}{\sqrt{n}} \right]$$

In the cutting machine case, where $n = 20$ it follows that

$$(2) \quad t_{.025,n-1} = t_{.025,19} = 2.093 \quad [\text{Table A5}]$$

so that the same standard error ($s_n = .04$)

$$(3) \quad s(\bar{X}_n) = \frac{s_n}{\sqrt{n}} = \frac{.04}{\sqrt{20}} = .009$$

now yields a sampling error of

$$(4) \quad t_{.025,n-1} \frac{s_n}{\sqrt{n}} = (2.093)(.009) = .019$$

with associated 95% confidence interval

$$(5) \quad \left[ \bar{X}_n \pm t_{.025,n-1} \frac{s_n}{\sqrt{n}} \right] = [15.04 \pm .019] = [15.021,15.059]$$

Note that the very small change in confidence intervals here is due to the precision of the cuts ($s_n = .04$).
CONVENTIONS FOR THE $t$-DISTRIBUTION

<table>
<thead>
<tr>
<th>Notation</th>
<th>JMPIN</th>
<th>Devore</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{\alpha,\nu}$</td>
<td>t Quantile</td>
<td>Table A.5</td>
</tr>
<tr>
<td>$\alpha(t,\nu)$</td>
<td>t Distribution</td>
<td>Table A.8</td>
</tr>
</tbody>
</table>

- **Sample Mean** $\nu = n - 1$
- **Simple Regression** $\nu = n - 2$
- **Multiple Regression** $\nu = n - (k + 1)$
STOPPING DISTANCE PROBLEM

To evaluate the safety features of a new surface composition for roads, tests are run on the mean (full-brake) stopping distances for standard passenger vehicles using this surface.

DATA: A sample of \( n = 30 \) vehicles are tested for stopping distance at 60 mph, yielding data \((x_1, \ldots, x_n)\) with sample mean, \( \bar{x}_n = 285.35 \), and standard deviation, \( s_n = 13.82 \).

DECISION RULE: Traffic engineers will consider the new surface to be acceptable only if they can be at least 95% confident that the mean stopping distance, \( \mu \), does not exceed 300 ft (a typical tolerance level for most road surfaces).

QUESTION: Should the road surface be accepted on the basis of this data?
STOPPING DISTANCE PROBLEM

For the case of unknown variance the key probability,

\[
Pr\left( \mu \leq \bar{x}_n + z_\alpha \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha
\]

is now replaced by

\[
Pr\left( \mu \leq \bar{x}_n + t_{\alpha,n-1} \frac{s_n}{\sqrt{n}} \right) = 1 - \alpha
\]

So for \( n = 30 \) and \( \alpha = .05 \), it follows that

\[
t_{\alpha,n-1} = t_{.05,29} = 1.699 \quad [\text{Table A5}]
\]

which together with \( \bar{x}_n = 285.35 \) and \( s_n = 13.82 \) yields

\[
\bar{x}_n + t_{\alpha,n-1} \frac{s_n}{\sqrt{n}} = 285.35 + (1.699)\frac{13.82}{\sqrt{30}} = 289.6
\]

Thus the traffic engineers can be 95% confident that the true mean stopping distance, \( \mu \), satisfies

\[
\mu \leq 289.6 < 300
\]

and may conclude that this road surface is acceptable in terms of their safety criterion.
CRITICAL CONFIDENCE LEVELS FOR THE t-DISTRIBUTION

For a given sample \((x_1, \ldots, x_n)\) and threshold value, \(\mu^*\), for \(\mu\), if \(\bar{x}_n < \mu^*\) then one may ask “How confident can we be that the true value of \(\mu\) does not exceed \(\mu^*\)?”. Equivalently, what is the largest value of \(1 - \alpha\) (smallest value of \(\alpha\)) such that \(\bar{x}_n + t_{\alpha,n-1}(s_n / \sqrt{n}) \leq \mu^*\)? The “knife-edge” case yields:

\[
\bar{x}_n + t_{\alpha,n-1}(s_n / \sqrt{n}) = \mu^* \iff t_{\alpha,n-1} = \frac{\mu^* - \bar{x}_n}{s_n / \sqrt{n}}
\]

So in terms of Table A.8, \(\alpha^* = \alpha(t, n-1)\) where \(t = \frac{\mu^* - \bar{x}_n}{s_n / \sqrt{n}}\)

and the Critical Confidence Level is again given by:

\[
C^* = 100(1 - \alpha^*)
\]
To determine exactly how confident these engineers can be that road surface is safe, we can set $\mu^* = 300$ and solve the relevant “knife edge” problem:

\[ \bar{x}_n + t_{\alpha,n-1} \left( s_n / \sqrt{n} \right) = 300 \]

\[ \Rightarrow t_{\alpha,n-1} = \frac{300 - \bar{x}_n}{s_n / \sqrt{n}} = \frac{300 - 285.32}{13.82 / \sqrt{30}} = 5.806 \]

Thus in this case (by Table A.8):

\[ \alpha^* = \alpha(t,n-1) = \alpha(5.806,29) < 0.0005 \]

\[ \Rightarrow C^* = 100(1 - \alpha^*) > 99.9 \]

So the traffic engineers can in fact be more than 99.9% confident that this road surface is safe.
CONFIDENCE INTERVALS
WITH VARIANCE UNKNOWN

TWO-SIDED

Choose $100(1 - \alpha)$ Level

Sampling Error $= t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}$

Confidence Interval for $\mu$:

$\left[ \bar{X}_n \pm t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} \right]$

Typical Values (n=10):

<table>
<thead>
<tr>
<th>$1 - \alpha$</th>
<th>$\alpha / 2$</th>
<th>$t_{\alpha/2,9}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.99</td>
<td>.005</td>
<td>3.25</td>
</tr>
<tr>
<td>.95</td>
<td>.025</td>
<td>2.26</td>
</tr>
<tr>
<td>.90</td>
<td>.05</td>
<td>1.83</td>
</tr>
</tbody>
</table>

ONE-SIDED

Choose $100(1 - \alpha)$ Level

Sampling Error $= t_{\alpha, n-1} \frac{s}{\sqrt{n}}$

Confidence Interval(s) for $\mu$:

$[ \mu \geq \bar{X}_n - t_{\alpha, n-1} \frac{s}{\sqrt{n}} ]$

$[ \mu \leq \bar{X}_n + t_{\alpha, n-1} \frac{s}{\sqrt{n}} ]$

Typical Values (n=10):

<table>
<thead>
<tr>
<th>$1 - \alpha$</th>
<th>$\alpha$</th>
<th>$t_{\alpha,9}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.99</td>
<td>.01</td>
<td>2.82</td>
</tr>
<tr>
<td>.95</td>
<td>.05</td>
<td>1.83</td>
</tr>
<tr>
<td>.90</td>
<td>.10</td>
<td>1.38</td>
</tr>
</tbody>
</table>