Goals for the day:

1. Words: confidence interval, confidence level, margin of error
2. R:
3. Big idea: We make a statement with confidence about the value of an unknown mean. (But alas we have to know the unknown standard deviation.)

Warning. We are following the book here which has an approach that is typical in engineering statistics texts. We are also using the terminology and approach of the fundamentals of engineering exam. But it is a mistake and a confusion that we will repair next week!

Words in a book.

Population All words in a book.

Variable length of word (in letters)

Parameter \( \mu \)

Sample \( n = 290 \) words

Statistic \( \bar{x} \)

The probability is 95% that \( \bar{x} \) will be within \( 1.96 \frac{\sigma}{\sqrt{n}} \) of \( \mu \). That means that the probability is 95% that \( \mu \) is within \( 1.96 \frac{\sigma}{\sqrt{n}} \) of \( \bar{x} \).

If the sample size \( n \) is large, the interval

\[
\left( \bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} \right)
\]

is an approximate 95% confidence interval for \( \mu \).

There are two approximations:

1. Central Limit Theorem says the distribution of \( \bar{X} \) is approximately normal.
2. We are using \( s \) to approximate \( \sigma \) since \( \sigma \) is unknown.

Write a 95% confidence interval for the length of words in the population:
In the M241 package are several datasets that come from samples from populations.

<table>
<thead>
<tr>
<th>Population</th>
<th>Variable</th>
<th>Sample size</th>
<th>variable name</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Calvin College students</td>
<td>Body Temperature</td>
<td>130</td>
<td>normtemp$Temp</td>
</tr>
<tr>
<td>Day of the year</td>
<td>Maximum wind velocity at San Diego Airpt</td>
<td>6209</td>
<td>wind$Wind</td>
</tr>
<tr>
<td>Calvin students taking GRE test</td>
<td>quantitative score</td>
<td>43</td>
<td>gre$quant</td>
</tr>
<tr>
<td>Glass manufacturing process</td>
<td>breaking strength</td>
<td>31</td>
<td>windowstrength$ksi</td>
</tr>
</tbody>
</table>

On the basis of these data, write 95% confidence intervals for

- The mean body temperature of Calvin students
- The mean maximum daily wind velocity at the Sandiego airport
- The mean quantitative GRE score of Calvin students that might take GRE test
- The mean breaking strength of glass produced by the process

Just as there are 95% confidence intervals, it is easy to conceive of 99% or 90% confidence intervals. A confidence interval for \( \mu \) has the form

\[
\bar{x} \pm (\text{critical value}) \frac{s}{\sqrt{n}}
\]

We can change the confidence level by changing the critical value from 1.96 to some other value. Examine how we got 1.96 and compute the critical values the following intervals. (You might want to use qnorm to find these values.)

- 90% confidence interval
- 99% confidence interval

If we use 10 samples to generate 90% confidence intervals, we might expect 9 of them to capture the mean of the population. We’ll do an experiment.

Recall that sr$GPA has the GPAs of 1,333 seniors as of February, 2005. Compute \( \mu \), the mean GPA of this population:

\[
\mu
\]

Select 10 different samples of size 30 from this population (sample(sr$GPA, 30)) and compute a 90% confidence interval for the mean from each sample. How many of your 10 intervals contained \( \mu \)?

The Central Limit Theorem gives us an approximate result. The approximation is better the larger \( n \) is. And the approximation is better for population distributions that are not highly skew and that do not have large outliers. We do not know the population distribution but can get an idea from the sample. In which of the examples above do you think the use of the Central Limit Theorem is most suspect?