Chapter 6: Confidence Intervals (CIs)
- Review what we’ve learned
- Explore intuition
Confidence Interval for $\mu$

Let $Y_1, \ldots, Y_n$ be independent observations from a population with some $\mu$ and $\sigma$. The $100(1 - \alpha)\%$ CI for $\mu$ is given by

- $\bar{y} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$ if $n \geq 30$
- $\bar{y} \pm t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}$ if $n < 30$ and $Y_i$ normal
In previous lecture we learned that “the probability that $\mu$ is in the CI is about $1 - \alpha$.”

In general, this is the **wrong way to express this idea.**

Why?

Suppose an experiment is conducted and a confidence interval is generated (e.g. $\bar{y} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$). $\mu$ is either in the interval or it isn’t. (Binary outcome)
Wrong Interpretation

In previous lecture we learned that “the probability that $\mu$ is in the CI is about $1 - \alpha$.”

In general, this is the **wrong way to express this idea**.

Why?

Suppose an experiment is conducted and a confidence interval is generated (e.g. $\bar{y} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$). $\mu$ is either in the interval or it isn’t. (Binary outcome)
“In the long run, if the experiment and procedure that generated the CI were repeated many times, we could expect the proportion of such CIs containing \( \mu \) to be \( 1 - \alpha \).”
Typically, we use $\alpha = .05$, giving a 95% confidence interval.

Why? The inventor of CIs said “begin wrong once in twenty times isn’t so bad.”

Clearly we need to select $\alpha$ based on our application of interest.
Suppose the lifespan of a certain one-celled organism is denoted by $Y$ and has a true mean of $\mu = 61$ minutes. Suppose further that we don’t know the true value of $\mu$ but that we observe the lifespan of 15 such organisms. Find the 95% CI for $\mu$ and assess whether it contains the true value of $\mu$. Assume that $Y$ is from a normal population.
Example

Suppose we observed $\bar{y} = 70.33$ and $s = 69.37$. Because $n = 15 < 30$ and we are assuming that $Y \sim N(\mu, \sigma)$, the following formula can be used

$$\bar{y} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}$$

Since $\alpha = .05$, we need $t_{.025,14} = 2.14$. The 95% CI is then

$$70.33 \pm 2.14 \frac{69.37}{\sqrt{15}} = (32, 108.66)$$

which contains $\mu = 61$. 
Let’s repeat the experiment $M = 100,000$ times and see what proportion of the time the 95% CI contains $\mu = 61$.

Result: 91.2% of the CIs contain $\mu$. Are we satisfied? Why isn’t it 95%?
Let’s repeat the experiment $M = 100,000$ times and see what proportion of the time the 95% CI contains $\mu = 61$.

Result: 91.2% of the CIs contain $\mu$. Are we satisfied? Why isn’t it 95%?
Assumptions about $Y$

- We assumed that $Y$ was from a normal population in order to use the confidence interval.
- Let’s make a Normal Probability Plot to assess the normality of the original sample of size $n = 15$. 
Here are the 15 cell lifespans:

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>46.07</td>
<td>72.08</td>
<td>8.89</td>
<td>8.53</td>
<td>26.60</td>
<td></td>
</tr>
<tr>
<td>176.59</td>
<td>75.00</td>
<td>32.92</td>
<td>58.35</td>
<td>8.97</td>
<td></td>
</tr>
<tr>
<td>84.83</td>
<td>46.48</td>
<td>75.49</td>
<td>269.86</td>
<td>64.33</td>
<td></td>
</tr>
</tbody>
</table>

Sorted:

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>8.53</td>
<td>8.89</td>
<td>8.97</td>
<td>26.60</td>
<td>32.92</td>
<td></td>
</tr>
<tr>
<td>46.07</td>
<td>46.48</td>
<td>58.35</td>
<td>64.33</td>
<td>72.08</td>
<td></td>
</tr>
<tr>
<td>75.00</td>
<td>75.49</td>
<td>84.83</td>
<td>176.59</td>
<td>269.86</td>
<td></td>
</tr>
</tbody>
</table>
Here are the 15 cell lifespans:

```
46.07  72.08  8.89  8.53  26.60
176.59  75.00  32.92  58.35  8.97
84.83  46.48  75.49  269.86  64.33
```

Sorted:

```
8.53  8.89  8.97  26.60  32.92
46.07  46.48  58.35  64.33  72.08
75.00  75.49  84.83  176.59  269.86
```
Recall: For a Normal Probability Plot, we need to compute the values $z_\alpha$ for the values of $1 - \alpha$ equal to

$$\frac{1 - .5}{15}, \quad \frac{2 - .5}{15}, \ldots, \quad \frac{15 - .5}{15}$$

which are, from left to right,

$$-1.83, -1.28, -0.97, -0.73, -0.52, -0.34, -0.17, 0.00, 0.17, 0.34, 0.52, 0.73, 0.97, 1.28, 1.83$$
Plot the sorted observations against the values of $z_\alpha$. 

Normal Probability Plot of Cell Lifespans

$Z$ score
Sorted Observation
The distribution is probably not normal (why and in what way?).

We therefore have discovered why the 95% CI we built was actually a 91% CI.

This is a relatively innocuous example of a CI being less “confident” than advertised.
The **standard error** of the mean $\bar{Y}$ is given by

$$SE_{\bar{Y}} = \frac{s}{\sqrt{n}}$$
Confidence Interval Formulas with $SE_{\bar{Y}}$

- $\bar{y} \pm z_{\alpha/2}SE_{\bar{Y}}$ if $n \geq 30$
- $\bar{y} \pm t_{\alpha/2, n-1}SE_{\bar{Y}}$ if $n < 30$ and $Y_i$ normal

Point Estimate and Error Margin

In the CIs above, $\bar{y}$ is called the **point estimate**. Also, $z_{\alpha/2}SE_{\bar{Y}}$ is called the **error margin**. ($t$ in the case of the 2nd CI.)
Cls are usually viewed as

Point Estimate ± Error Margin
s is also a point estimate of $\sigma$.

We will later build confidence intervals for $\sigma$. 