Note: C.I. for $\mu_x$ of Type 2 fish is wider (i.e. our estimate for $\mu_x$ is less reliable/precise) Why?

- The conf. level is higher
- Sample std. dev. (s) is larger.
- Even though $n$ is larger (which shrinks the C.I.), the increase in $n$ is not enough to compensate for the increase in conf. level and s.

The formula for C.I. can be used to decide what minimum sample size is necessary, even before taking any sample! But you need to specify what is meant by necessary.

For example, say, you want your estimate of $\mu_x$ to be within some range $\pm B$ (for Bound). Then

$$\frac{Z^*s}{\sqrt{n}} = B \implies n_{\text{min}} = \left(\frac{Z^*s}{B}\right)^2$$

Note that $B$ is different from conf. level, or $z^*$. It has the dimensions of $\mu_x$ itself.

**Example**

What min. sample size is required for a margin of error of 0.03 of $\mu_x$?

$$n = \left(\frac{Z^*s}{B}\right)^2 = \left(\frac{1.96 (1.27)}{.03}\right)^2 = 6,885 \text{ Type I Fish.}$$

$$\left(\frac{2.575 (1.71)}{.03}\right)^2 = 2,543 \text{ Type I Fish.}$$

If you have no sample to provide an estimate of $s$, then you guess it! It's not hard. For example, if we're dealing with people's height, then $s$ is a few inches.
This type of C.I. is called **2-sided**.

Sometimes, though, we want to find only an **upper bound**, or a lower bound, for \( \mu_x \).

Then, we need to compute a 1-sided C.I. (or Conf. Bound):

**Upper Conf. Bound**: \( \bar{x} + z^* \frac{s_x}{\sqrt{n}} \)  

**Lower Conf. Bound**: \( \bar{x} - z^* \frac{s_x}{\sqrt{n}} \)

\( z^* \) is different from 2-sided \( z^* \)'s:
- 90\%: 1.28  
- 95\%: 1.645  
- 99\%: 2.33  

*Table I or last line in Table IV.*

1-sided C.I. (or conf. bounds) are useful when we want to see if the **true mean** is greater (or smaller) than some value.

**Interpretation (IMPORTANT!)**

1. Suppose 95\% upper conf. bound for \( \mu_x \) is 0.3. Then
   1) We are 95\% confident that \( \mu_x < 0.3 \)
   2) There is a 95\% prob. that a random upper conf. bound will be greater than \( \mu_x \).

Make sure you see why!
So far, we've done 1- and 2-sided C.I. for $\mu_x$.

2-Sided \[ \bar{x} \pm z^* \frac{\sigma_x}{\sqrt{n}} \quad \text{All different } z^* \]

Lower conf. bound: \[ \bar{x} - z^* \frac{\sigma_x}{\sqrt{n}} \]
upper: \[ \bar{x} + z^* \frac{\sigma_x}{\sqrt{n}} \]

what about C.I. for pop. proportion $\pi_x$?

To build the C.I. for $\pi_x$, we need the Sampl. dist. of $p$, the sample proportion. [Recall $\bar{x} \sim N(\mu_x, \frac{\sigma_x}{\sqrt{n}})$.

In a hw, you showed

$\pi_x \equiv E[p] = \pi_x \quad \text{pop. proportion.}$

$\sigma_p^2 = \frac{\pi_x(1-\pi_x)}{n} \quad \text{note resemblance to } \frac{\sigma_x^2}{n},$

where $\sigma_x^2 = \frac{\pi_x(1-\pi_x)}{n}$ [Binomial with $n=1$, also called Bernoulli($\pi$)].

even without knowing the sample. distv. itself. Additionally:

CLT: $p \sim N\left( \mu = \mu_p = \pi_x, \sigma = \sigma_p = \frac{\sqrt{\pi_x(1-\pi_x)}}{\sqrt{n}} \right)$

If $n \pi > 5$, $n(1-\pi) > 5$

Therefore, we can again compute the prob. that the sample prop. is between 2 numbers ($a < p < b$):

\[ \text{prob}(a < p < b) = \text{prob}\left( \frac{a - \mu_p}{\sigma_p} < \frac{p - \mu_p}{\sigma_p} < \frac{b - \mu_p}{\sigma_p} \right) \]

\[ = \text{prob}\left( \frac{a - \pi_x}{\sqrt{\pi_x(1-\pi_x)/n}} < z < \frac{b - \pi_x}{\sqrt{\pi_x(1-\pi_x)/n}} \right) \]

\[ = \text{table I}. \]
Now that we know the sampling distribution of $\hat{p}$, we can build a C.I. for $\pi$.

$\Rightarrow$ CLT $\Rightarrow$ If $n$ = large, then $p \sim N \left( \pi, \sqrt{\frac{\pi(1-\pi)}{n}} \right)$

$\Rightarrow$ What, then, has a std. normal dist? $z = \frac{\hat{p} - \pi_x}{\sqrt{\frac{\pi_x(1-\pi_x)}{n}}}$

Start with self-evident fact

Recall $\text{prob}(-1.96 < \frac{\hat{p} - \pi_x}{\frac{\pi_x(1-\pi_x)}{\sqrt{n}}} < 1.96) = 0.95$

$\Rightarrow \pi_x \Rightarrow 95\% \ C.I. \ for \ \pi_x$.]

$\text{prob}(-z^* < \frac{\hat{p} - \pi_x}{\frac{\pi_x(1-\pi_x)}{\sqrt{n}}} < z^*) = \text{conf. level}$

$\Rightarrow$ quadratic eqn in $\pi_x$. This is

$\Rightarrow \pi_x \Rightarrow$ why the C.I. for $\pi_x$ is a messy eqn. $\downarrow$

C.I. for $\pi_x$: $\frac{1}{1 + \frac{z^2}{n}} \left[ (\hat{p} + \frac{z^2}{2n}) \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2}{4n^2}} \right]$

Good News: If $n$ = large ($> 30$), Then $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Finally, 1-sided C.I. affects $z^*$ only.

$\pi_x$ denotes the proportion (e.g. of "goods") in pop. In the coin-tossing analog $\pi_x$ is the prob. of a head on a given toss. Note that this is all perfectly consistent, because the prob. of drawing a single "good" out of the population (i.e. prob of heads on a toss) is equal to the proportion of goods in pop.
In Summary:

1) \( E[X] = \mu_x \quad E[P] = \mu_p = \mu_x \)
   
   \( \text{Var}(X) = \sigma^2_x = \frac{\sigma^2}{n} \) \quad \text{Var}(P) = \sigma^2_p = \frac{\sigma^2}{n} \)

   quite independently of the distr. of \( X \) (i.e. the pop.)

   The right-hand sides are all pop. parameters.

Q: What to do if you don’t know what they are?

A1: Assume the pop.

A2: Estimate \( \mu_x \) with \( \bar{x} \), \( \sigma_x \) with \( s \)

1. \( \mu_x \) with \( p \) (or \( 1/2 \)). Why? (CE.2)

2) If pop = normal with params \( \mu, \sigma \), i.e. \( f(x) = \mathcal{N}(\mu, \sigma) \)

   Then \( \bar{x} \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}}) \) \quad x \sim \mathcal{N}(\mu, \sigma)

3) Even if pop is not normal, as long as \( n \) = large

4) Then, we can compute \( \text{prob}(a < \bar{x} < b) \) and \( \text{prob}(a < x < b) \)

5) We can build C.I.s for \( \mu_x \) and \( \mu_x \), both 1-sided and 2-sided.

We used CLT to derive the sample distr. of \( \bar{x} \), and from that, the C.I. for \( \mu_x \) and \( \mu_x \).

But even if we have no idea what the sample distr. of our statistics is, we can still build it empirically, using the bootstrap idea (in Lab).
Example: A past survey from 80 people:

- Lab is good: 17 (21.25% ≈ 0.21 = 17/80
- Lab is bad: 48 (60.00% ≈ 0.60 = 48/80
- No opinion: 15 (18.75% ≈ 0.19 = 15/80

Only part of the class voted, but assuming that the voters are a random sample from the whole class, we can find the true proportion of students who like the lab, etc.

Let's build the 95% C.I. for the various pop. proportions:

In this case, the population of students consists of 3 categories: "Lab is Good", "Lab is Bad", "No opinion". So, we have 3 proportions for which to build C.I.'s:

\[ \hat{p}_1: \text{True prop. of students who think Lab is Good} \]
\[ \hat{p}_2: \text{"Lab is Bad"} \]
\[ \hat{p}_3: \text{"No opinion"} \]

Note that these 3 props are NOT independent, because \(\hat{p}_1 + \hat{p}_2 + \hat{p}_3 = 1\). This will become important for C.I.s we will compute later.

<table>
<thead>
<tr>
<th>Good</th>
<th>Bad</th>
<th>No opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.21 ± 0.09</td>
<td>0.60 ± 0.11</td>
<td>0.19 ± 0.07</td>
</tr>
</tbody>
</table>

\[ 0.21 \pm 1.96 \sqrt{\frac{0.21 \cdot (1-0.21)}{80}} \quad \hat{p}_1 \]
\[ 0.60 \pm 1.96 \sqrt{\frac{0.60 \cdot (1-0.6)}{80}} \quad \hat{p}_2 \]
\[ 0.19 \pm 1.96 \sqrt{\frac{0.19 \cdot (1-0.19)}{80}} \quad \hat{p}_3 \]

We are 95% confident that the proportion of students who like the lab is here.
Suppose we are developing a new composite material for building airplane wings. We take a sample of size 100 of the material and test its breaking strength under a set of standard conditions. The sample mean and sample standard deviation of the breaking strength are 20 and 5, respectively.

a) What type of confidence interval is appropriate for this problem (2-sided interval, an upper confidence bound, or a lower confidence bound)? Explain. Hint: A small breaking point is a very bad thing!

b) Compute it for this data, and provide two interpretations. Use a confidence level of 95%.

Show that the $z$ for a 95% lower conf. bound for a population mean is 1.645. In other words, show that the 95% lower conf. bound for the pop. mean is $\bar{x} - 1.645 \frac{s}{\sqrt{n}}$. 