Stuff you now know:

- Dealing with ambiguity
- Random variable
- Histograms
- Comparative boxplots
- Quantiles
- Distributions
- Probability (e.g. from Poisson)
- Sample mean and variance
- Distr mean and variance
- Qqplots

- Scatterplots
- Correlation
- Regression (multiple, polynomial, ...)
- ANOVA (R^2, s_e ~ RMSE)
- Overfitting, collinearity, interaction

- Sampling distribution
- 1-sample Confidence Interval for ...
- 2-sample CI for ...
- t-distribution
Lecture 19 (Ch. 7, 8)

Here are the CI's we have derived:

\[ \text{1-sample} \]

- CI for \( \mu_x \):
  \[ \bar{x} \pm z^* \frac{\sigma_x}{\sqrt{n}} \text{ or } \frac{\bar{x} \pm t^* \frac{s_x}{\sqrt{n}}}{\sqrt{n}} \]

- CI for \( \sigma^2_x \):
  \[ p \pm z^* \sqrt{\frac{p(1-p)}{n}} \text{ or } \frac{\bar{x} \pm t^* \frac{s_x}{\sqrt{n}}}{\sqrt{n}} \]

\[ \text{2-sample} \]

- CI for \( \mu_1 - \mu_2 \):
  \[ \bar{x}_1 - \bar{x}_2 \pm z^* \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \text{ or } \frac{\bar{x}_1 - \bar{x}_2 \pm t^* \frac{s_1}{\sqrt{n}} + \frac{s_2}{\sqrt{n}}}{\sqrt{n}} \]

Random vs. observed: probabilistic vs. confidence interpretation

The formulas/math are easy. It's their interpretation that's HARD:
1) Confident... 2) Prob... 3) Corollary

(FYI: In addition to these 2-sided CIs, there is also UCB and LCB)

The formulas for these are derived the same way:

Start with \( P(-z^* < z < z^*) = \text{Conf. level} \)

\[ \text{solve for } \]

Substitute \( z = \frac{\bar{x} - \mu_x}{\sigma / \sqrt{n}} \sim \mathcal{N}(0,1) \text{ or } \frac{\bar{x} - \mu_x}{s / \sqrt{n}} \sim \mathcal{N}(0,1) \text{ as } \mu \text{ to get CI} \)

\[ \bar{x} \pm \frac{z^*}{\sqrt{\sigma^2 / n}} \]

\[ \bar{x}_1 - \bar{x}_2 \pm \frac{z^*}{\sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}} \]

\[ p_1 - p_2 \pm \frac{z^*}{\sqrt{p_1(1-p_1) / n_1 + p_2(1-p_2) / n_2}} \]
We required the 2 samples (in a 2-sample problem) to be independent. It happened when we wrote
\[ V[\bar{x}_1 - \bar{x}_2] = V[\bar{x}_1] + V[\bar{x}_2] + O = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \]
But there exist problems where the 2 samples are not independent.

**E.g. 1:** Does a certain pill increase IQ?

- Take 100 people, measure their IQ, w/o pill.
- Ask 100 people, to tell you how fast App X was.
- Time App X and App Y on each of 100 computers.

Such data are called "paired".

You can usually see/test this by looking at:

How do we build a C.I. for \( \mu_1 - \mu_2 \) from paired data?

- Make a new column:
- IQ before | IQ after | "d = x_1 - x_2"
- \[ d \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}} \]

Benefit: paired C.I.s are narrower (more precise).

The Math is trivial. Deciding paired vs. independent is NOT! - First!
Example Here is how we did the fish example in last lecture:

<table>
<thead>
<tr>
<th></th>
<th>$n$</th>
<th>$\bar{x}$</th>
<th>$s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I</td>
<td>56</td>
<td>9.15</td>
<td>1.27</td>
</tr>
<tr>
<td>Type II</td>
<td>61</td>
<td>3.08</td>
<td>1.71</td>
</tr>
</tbody>
</table>

Suppose we ask: Are the true/pop. means different?

$\mu_1 =$ pop. mean zinc in Type I 
$\mu_2 =$ pop. mean zinc in Type II

Important to define $\mu_1, \mu_2$ (the pop. parameters) clearly.

95% CI for $\mu_1 - \mu_2 : (9.15 - 3.08) \pm 1.96 \sqrt{\frac{(1.27)^2}{56} + \frac{(1.71)^2}{61}}$

$6.07 \pm 0.54 \Rightarrow [5.53, 6.61]$

There are many ways of collecting paired data, e.g.:

1) Take one Type I and one Type II, from $n$ lakes.
2) " ... " " ... " with the same weight.
3) ...

In every example, a Type I fish is somehow matched/paired with a Type II fish.

Benefit: In every example, the CI for $\mu_1 - \mu_2$ will be narrower than (ie. more precise)

Again, the first question you ask yourself should be whether or not the data are paired!
It may be tempting to think that a paired design does not exist for proportions. But it does. 2 genders (Boy/Girl), 2 computer types (Mac/Dell).

Here is an example data from an unpaired design:

<table>
<thead>
<tr>
<th></th>
<th>Boy with Mac</th>
<th>Girl with Mac</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mac</td>
<td>Mac</td>
</tr>
<tr>
<td></td>
<td>Dell</td>
<td>Mac</td>
</tr>
<tr>
<td></td>
<td>Dell</td>
<td>Dell</td>
</tr>
</tbody>
</table>

prop.of Boys \( \rightarrow \) prop.of Girl with Mac

\[ n_1 \quad \text{and} \quad n_2 \]

C.I. for \( \pi_1 - \pi_2 \): \( (\hat{\pi}_1 - \hat{\pi}_2) \pm Z \times \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}} \)

Here is an example data from a paired design:

<table>
<thead>
<tr>
<th></th>
<th>Husband with Mac</th>
<th>Wife with Mac</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mac</td>
<td>Mac</td>
</tr>
<tr>
<td></td>
<td>Dell</td>
<td>Mac</td>
</tr>
<tr>
<td></td>
<td>Dell</td>
<td>Dell</td>
</tr>
</tbody>
</table>

prop.of Husbands \( \rightarrow \) prop.of wives with Mac

\[ n \quad \text{and} \quad n \]

However, there is no simple formula for the C.I. of \( \pi_1 - \pi_2 \) when data are paired. (After all, we cannot make a new column of differences, because we cannot subtract "Mac" from "Mac", i.e. The data are categorical.)
Statistical Definition of Belief: A statement not based on data. So, technically, a belief is an assumption.

If decision-making (i.e. Reject or Not-reject) is the final goal of your study, then the machinery of computing C.I. can be massaged to form a more direct response. The revised methodology is called often called hypothesis testing (but, don't forget that we can test hypotheses with CIs, too). So, there are 3 hypothesis testing methods: 1) Using CIs, 2) using p-values (common approach), and 3) the Rejection Region approach; better for something called power. We are going to do the p-value approach.

The logic of the p-value methodology is very tricky!
It requires assuming a statement/belief about a pop. parameter, and then checking to see if evidence from data contradict the assumption. Recall that, without knowing pop parameters we cannot compute probs.

The question one asks is of the form "Does data provide sufficient evidence contrary to the assumption/belief?"
If Yes, then we reject the assumption/belief.
If No, then we cannot reject the assumption/belief, i.e., we just don't know!
Notice that "Cannot reject blah" is NOT the same as "Accept blah"!

One can also ask "Does the data provide sufficient evidence in support of some claim (but this time the claim is based on data, i.e. the opposite of the assumption/belief)?"

The reason we have to worry about this level of detail is this: There is no such thing as evidence for something that you have already assumed true! Evidence can only come from data, and we can only use evidence to reject an assumption (not to support it). Evidence (from data) cannot support an assumption/belief.
**Example**: Data says: \( n = 64, \ x = 34.4, \ S = 1.1. \)

Does the data provide evidence to support \( \mu_x \geq 34 \)?

\[ t = \frac{\bar{x} - \mu_x}{S/\sqrt{n}} = \frac{34.4 - 34}{1.1/\sqrt{64}} = \frac{1.2}{0.15} = 8.26 \]

\[ p = \text{Prob}(t > 8.26) \approx 0.000 \approx \text{very small} \]

FYI: Technically, we need a lower conf. bound (LCB) for questions like this. But we're skipping those this quarter. So, let's use our 2-sided CI:

**CI approach**: 95% CI for \( \mu_x \): \( 34.4 \pm 2 \frac{1.1}{\sqrt{64}} = [34.1, 34.7] \)

We are 95% confident that \( \mu_x \in [34.1, 34.7] \).

There is evidence that \( \mu_x \geq 34 \).

I.e. If someone claims \( \mu_x < 34 \), then we can Reject the claim.

A different way of arriving at that conclusion.

This time: Assume \( \mu_x < 34 \), and find evidence (from data) to the contrary.

Another way of saying this: Let \( H_0: \mu_x \leq 34 \), then assume \( H_0 = \text{True} \) called "Null hypothesis".

Q: What's contradictory?

A: Really large \( \bar{x}'s \) are contrary to \( \mu_x \geq 34 \). So, here is a measure of evidence (from data) contrary to \( \mu_x \geq 34 \): \( \text{Prob}(\bar{x} > x_{\text{obs}} | \mu_x = 34) \).

Let's start by computing that prob if \( \mu_x = 34 \) (the worst-case scenario for \( H_0 \)).

\[ \text{Prob}(\bar{x} > x_{\text{obs}} | \mu_x = 34) = \text{Prob} \left( \frac{\bar{x} - \mu_x}{S/\sqrt{n}} > \frac{x_{\text{obs}} - \mu_x}{S/\sqrt{n}} \right| \mu_x = 34 \)

\[ = \text{Prob}(t > 2.96, 63) \]

\[ = \text{Prob}(t > 2.91) \approx 0.0025 \]

\[ t_{\text{obs}} = \frac{x_{\text{obs}} - \mu_x}{S/\sqrt{n}} = \frac{34.4 - 34}{1.1/\sqrt{64}} = 8.26 \]

\[ p = \text{Prob}(t > 8.26) \approx 0.000 \approx \text{very small} \]
If we assume \( \mu_x \) is as big as it can get according to the claim, i.e., \( \mu_x = 34 \), then the probability of getting \( \bar{x}_{\text{obs}} \) larger than \( \bar{x}_{\text{obs}} \) is very small. That is a lot of evidence contrary to \( \mu_x = 34 \), because that probability already assumes \( \mu_x = 34 \). This is a tricky point, so read it again.

**Lesson 1:** smaller p-value \( \Rightarrow \) more evidence against \( \mu_x < 34 \). 

\[ \text{Very strange!} \] Just give it time!

So, for our example, the conclusion is there is evidence from data to reject \( \mu_x < 34 \) (i.e., the same conclusion as the CI method), all because the above probability (an example of a p-value) is small.

---

**Q:** Now, what if we relax \( \mu_x = 34 \) into \( \mu_x < 34 \)?

**A:** Repeat the above calculation,

\[ \text{if } \mu_x < 34, \Rightarrow \bar{x}_{\text{obs}} = \text{larger} \Rightarrow \text{p-value is smaller.} \]

So, if \( \text{prob}(\bar{x}_{\text{obs}} | \mu_x = 34) = \text{small} \), we can reject \( \mu_x > 34 \), not just \( \mu_x = 34 \).

In English: if data reject \( \mu_x > 34 \), then \( \mu_x < 34 \) is also rejected.

**Lesson 2:** It is sufficient to test \( \mu_x \leq 34 \). Don't forget, this value came from the belief (not data).

---

**Q:** Who decides what's a “sufficiently small” value for p-value?

**A:** You do! This “threshold probability” is labeled \( \alpha \).

It's called significance level. \[ \alpha = 1 - \text{Conf. level} \]

- \( 0.05 \) sign. level = 95% Conf. level.

Some common values are \( 0.05, 0.01, 0.001 \).

In summary, if p-value \( < \alpha \), then we reject \( H_0 \). Else, cannot reject \( H_0 \).

---

Next time, I'll formalize all of this even further, because there are still a few missing ingredients (e.g., the Alternative Hypothesis).
Consider the following data on $x_1$ and $x_2$ which was collected in a paired design:

$x_1 = c(-0.27, -0.14, 1.61, 0.09, 0.00, 2.07, 0.56, -1.67, -0.51, -0.54)$

$x_2 = c(-0.32, 0.20, 1.93, 0.54, 0.75, 1.77, 0.84, -0.29, -0.33, 0.17)$

a) Compute a 2-sided, 95% CI for the difference between the two true means. You may use R to do simple calculations, but use the CI formulas derived in class. BTW, you can "test" that $x_1$ and $x_2$ are paired by looking at their scatterplot:

```
plot(x1,x2)          # I see a linear association
```

b) Provide at least one interpretation of the observed CI, AND state the conclusion in English, i.e., the "corollary."

c) Consider the following data, which is the same as above, except the cases in $x_2$ have been randomly shuffled. Compute an appropriate 95% 2-sided CI.

$y_1 = c(-0.27, -0.14, 1.61, 0.09, 0.00, 2.07, 0.56, -1.67, -0.51, -0.54)$

$y_2 = c(0.20, 0.54, -0.33, 1.93, -0.32, 1.77, 0.75, 0.17, -0.29, 0.84)$

d) Provide one interpretation of the observed CI, AND state the conclusion in English, i.e., the "corollary."

e) Which one is narrower?

Consider the following sample observations: 2781, 2900, 3013, 2856, and 2888.
Suppose we want to test whether there is evidence contrary to the belief that $\mu < 3000$.

a) Compute the observed 95% 2-sided confidence interval (CI) for $\mu$.

b) Based on the above CI, is there evidence that $\mu$ is greater than 3000?

c) Write the appropriate null hypotheses.

d) Compute the p-value, recalling that it measures evidence from data contrary to the null hypothesis.

e) At alpha=0.05, state the conclusion "In English" (i.e., is there evidence that $\mu$ is greater than 3000?)