CIs are possibly the most useful forms of inference because they give a range of “reasonable” values for a parameter.

But sometimes we want to know whether one particular value for a parameter is “reasonable.” In this case, a popular form of inference is the hypothesis test.

We use data to test a claim (about a parameter) called the null hypothesis.

Example 1: We claim the proportion of USC students who travel home for Christmas is 0.95.

Example 2: We claim the mean nightly hotel price for hotels in SC is no more than $65.

Null hypothesis (denoted H₀) often represents “status quo”, “previous belief” or “no effect”.

Alternative hypothesis (denoted Hₐ) is usually what we seek evidence for.

We will reject H₀ and conclude Hₐ if the data provide convincing evidence that Hₐ is true.

Evidence in the data is measured by a test statistic.
A test statistic measures how far away the corresponding sample statistic is from the parameter value(s) specified by H₀.

If the sample statistic is extremely far from the value(s) in H₀, we say the test statistic falls in the “rejection region” and we reject H₀ in favor of Hₐ.

Example 2: We assumed the mean nightly hotel price in SC is no more than $65, but we seek evidence that the mean price is actually greater than $65. We randomly sample 64 hotels and calculate the sample mean price \( \bar{X} \). Let \( Z = \frac{\bar{X} - 65}{\sigma/\sqrt{n}} \) be our “test statistic” here.

Note: If this Z value is much bigger than zero, then we have evidence against H₀: \( \mu \leq 65 \) and in favor of Hₐ: \( \mu > 65 \).

Suppose we’ll reject H₀ if \( Z > 1.645 \).

If \( \mu \) really is 65, then Z has a standard normal distribution. (Why?)

Picture:
If we reject $H_0$ whenever $Z > 1.645$, what is the probability we reject $H_0$ when $H_0$ really is true?

$$P(Z > 1.645 \mid \mu = 65) =$$

This is the probability of making a Type I error (rejecting $H_0$ when it is actually true).

$P$(Type I error) = “level of significance” of the test (denoted $\alpha$).

We don’t want to make a Type I error very often, so we choose $\alpha$ to be small:

The $\alpha$ we choose will determine our rejection region (determines how strong the sample evidence must be to reject $H_0$).

In the previous example, if we choose $\alpha = .05$, then $Z > 1.645$ is our rejection region.
Hypothesis Tests of the Population Mean

In practice, we don’t know \( \sigma \), so we don’t use the \( Z \)-statistic for our tests about \( \mu \).

Use the t-statistic: 
\[
t = \frac{\overline{X} - \mu_0}{s / \sqrt{n}}
\]
where \( \mu_0 \) is the value in the null hypothesis.

This has a t-distribution (with \( n - 1 \) d.f.) if \( H_0 \) is true (if \( \mu \) really equals \( \mu_0 \)).

Example 2: Hotel prices: 
\( H_0: \mu = 65 \)
\( H_a: \mu > 65 \)

Sample 64 hotels, get \( \overline{X} = $67 \) and \( s = $10 \). Let’s set \( \alpha = .05 \).

Rejection region:

Reject \( H_0 \) if \( t \) is bigger than 1.67.

Conclusion:
We never accept $H_0$; we simply “fail to reject” $H_0$.

This example is a one-tailed test, since the rejection region was in one tail of the t-distribution.

Only very large values of $t$ provided evidence against $H_0$ and for $H_A$.

Suppose we had sought evidence that the mean price was less than $72$. The hypotheses would have been:

$H_0$: $\mu = 72$

$H_A$: $\mu < 72$

Now very small values of $t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}}$ would be evidence against $H_0$ and for $H_A$.

Rejection region would be in left tail:
Rules for one-tailed tests about population mean

\[ H_0: \mu = \mu_0 \]
\[ H_a: \mu < \mu_0 \]  \ or \  \[ H_a: \mu > \mu_0 \]

Test statistic:  \[ t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}} \]

Rejection: \[ t < -t_{\alpha} \]  \ or \  \[ t > t_{\alpha} \]

Region: (where \( t_{\alpha} \) is based on \( n - 1 \) d.f.)

Rules for two-tailed tests about population mean

\[ H_0: \mu = \mu_0 \]
\[ H_a: \mu \neq \mu_0 \]

Test statistic:  \[ t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}} \]

Rejection: \[ t < -t_{\alpha/2} \]  \ or \  \[ t > t_{\alpha/2} \] (both tails)

Region: (where \( t_{\alpha/2} \) is based on \( n - 1 \) d.f.)
Example: We want to test (using $\alpha = .05$) whether or not the true mean height of male USC students is 70 inches.

Sample 26 male USC students. Sample data: $\bar{X} = 68.5$ inches, $s = 3.3$ inches.

**Assumptions of t-test (and CI) about $\mu$**
- We assume the data come from a population that is approximately normal.
- If this is not true, our conclusions from the hypothesis test may not be accurate (and our true level of confidence for the CI may not be what we specify).
- How to check this assumption?

- The t-procedures are robust: If the data are “close” to normal, the t-test and t CIs will be quite reliable.
Hypothesis Tests about a Population Proportion

We often wish to test whether a population proportion $p$ equals a specified value.

Example 1: We suspect a theater is letting underage viewers into R-rated movies. Question: Is the proportion of R-rated movie viewers at this theater greater than 0.25?

We test:

Recall: The sample proportion $\hat{p}$ is approximately

$N\left( p, \sqrt{\frac{pq}{n}} \right)$ for large $n$, so our test statistic for testing $H_0: p = p_0$

has a standard normal distribution when $H_0$ is true (when $p$ really is $p_0$).
Rules for one-tailed tests about population proportion

H₀: p = p₀
Hₐ: p < p₀  or  Hₐ: p > p₀

Test statistic:
\[ z = \frac{\hat{p} - p₀}{\sqrt{\frac{p₀q₀}{n}}} \]

Rejection
\[ z < -z_α \quad \text{or} \quad z > z_α \]

Region:

Rules for two-tailed tests about population proportion

H₀: p = p₀
Hₐ: p ≠ p₀

Test statistic:
\[ z = \frac{\hat{p} - p₀}{\sqrt{\frac{p₀q₀}{n}}} \]

Rejection
\[ z < -z_{α/2} \quad \text{or} \quad z > z_{α/2} \text{ (both tails)} \]

Region:

Assumptions of test (need large sample):

Need:
Example 1:
Test $H_0: p = 0.25$ vs. $H_a: p > 0.25$ using $\alpha = .01$.

We randomly select 60 viewers of R-rated movies, and 23 of those are underage.

Example 1(a): What if we had wanted to test whether the proportion of underage viewers was different from 0.25?
P-values

Recall that the significance level $\alpha$ is the desired \( P(\text{Type I error}) \) that we specify before the test.

The P-value (or “observed significance level”) of a test is the probability of observing as extreme (or more extreme) of a value of the test statistic than we did observe, if $H_0$ was in fact true.

The P-value gives us an indication of the strength of evidence against $H_0$ (and for $H_a$) in the sample.

This is a different (yet equivalent) way to decide whether to reject the null hypothesis:

- A small p-value (less than $\alpha$) = strong evidence against the null $\Rightarrow$ Reject $H_0$

- A large p-value (greater than $\alpha$) = little evidence against the null $\Rightarrow$ Fail to reject $H_0$

How do we calculate the P-value? It depends on the alternative hypothesis.
### One-tailed tests

<table>
<thead>
<tr>
<th>Alternative</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_a: &quot;&lt;&quot;$</td>
<td>Area to the left of the test statistic value in the appropriate distribution (t or z).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alternative</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_a: &quot;&gt;&quot;$</td>
<td>Area to the right of the test statistic value in the appropriate distribution (t or z).</td>
</tr>
</tbody>
</table>

### Two-tailed test

<table>
<thead>
<tr>
<th>Alternative</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_a: &quot;\neq&quot;$</td>
<td>2 times the “tail area” outside the test statistic value in the appropriate distribution (t or z). <strong>Double</strong> the tail area to get the P-value!</td>
</tr>
</tbody>
</table>
P-values for Previous Examples

Hotel Price Example:  \( H_0: \mu = 65 \) vs. \( H_a: \mu > 65 \)

Test statistic value:

Student height example:  \( H_0: \mu = 70 \) vs. \( H_a: \mu \neq 70 \)

Test statistic value:
Movie theater example: $H_0: p = 0.25$ vs. $H_a: p > 0.25$

Test statistic value:

What if we had done a two-tailed test of $H_0: p = 0.25$ vs. $H_a: p \neq 0.25$ at $\alpha = .01$?
Relationship between a CI and a (two-sided) hypothesis test:

- A test of $H_0: \mu = m^*$ vs. $H_a: \mu \neq m^*$ will reject $H_0$ if and only if a corresponding CI for $\mu$ does not contain the number $m^*$.

Example: A 95% CI for $\mu$ is (2.7, 5.5).

(1) At $\alpha = 0.05$, would we reject $H_0: \mu = 3$ in favor of $H_a: \mu \neq 3$?

(2) At $\alpha = 0.05$, would we reject $H_0: \mu = 2$ in favor of $H_a: \mu \neq 2$?

(3) At $\alpha = 0.10$, would we reject $H_0: \mu = 2$ in favor of $H_a: \mu \neq 2$?

(4) At $\alpha = 0.01$, would we reject $H_0: \mu = 3$ in favor of $H_a: \mu \neq 3$?
Power of a Hypothesis Test

- Recall the significance level \( \alpha \) is our desired
  \[ P(\text{Type I error}) = P(\text{Reject } H_0 \mid H_0 \text{ true}) \]

The other type of error in hypothesis testing:
Type II error =

\[ P(\text{Type II error}) = \beta \]

The power of a test is

- High power is desirable, but we have little control over it (different from \( \alpha \))

Calculating Power: The power of a test about \( \mu \)
depends on several things: \( \alpha, n, \sigma, \) and the true \( \mu \).

Example 1: Suppose we test whether the true mean nicotine contents in a population of cigarettes is greater than 1.5 mg, using \( \alpha = 0.01 \).

\[ H_0: \quad H_a: \]

We take a random sample of 36 cigarettes. Suppose we know \( \sigma = 0.20 \) mg. Our test statistic is
We reject $H_0$ if:

• Now, suppose $\mu$ is actually 1.6 (implying that $H_0$ is false). Let’s calculate the power of our test if $\mu = 1.6$:

This is just a normal probability problem!

• What if the true mean were 1.65?

Verify:

• The farther the true mean is into the “alternative region,” the more likely we are to correctly reject $H_0$.
Example 2: Testing $H_0: p = 0.9$ vs. $H_a: p < 0.9$ at $\alpha = 0.01$ using a sample of size 225.

Suppose the true $p$ is 0.8. Then our power is: