Announcements

For the final exercise...

▶ Be finalizing your final exercise groups
▶ Be narrowing down good data sources
▶ Feel free to come to OH to talk about any coding issues, see also coding help online, at the Stats Dept, and at IQSS
▶ (Data download and cleaning can often take longer than you'd like!)
Announcements

▶ For the final exercise...
Announcements

- For the final exercise...
- Be finalizing your final exercise groups
Announcements

- For the final exercise...
- Be finalizing your final exercise groups
- Be narrowing down good data sources
Announcements

- For the final exercise...
- Be finalizing your final exercise groups
- Be narrowing down good data sources
- Feel free to come to OH to talk about any coding issues, see also coding help online, at the Stats Dept, and at IQSS
Announcements

▶ For the final exercise…
▶ Be finalizing your final exercise groups
▶ Be narrowing down good data sources
▶ Feel free to come to OH to talk about any coding issues, see also coding help online, at the Stats Dept, and at IQSS
▶ (Data download and cleaning can often take longer than you’d like!)
Roadmap

Continuing with statistical inference and with hypothesis tests

Last time: $z$-test with a single sample mean

Today: $t$-tests

Comparing two sample means

Type I and II errors

Practical versus Statistical Significance
Roadmap

- Continuing with statistical inference and with hypothesis tests
Roadmap

- Continuing with statistical inference and with hypothesis tests
- Last time: z-test with a single sample mean
Roadmap

- Continuing with statistical inference and with hypothesis tests
- Last time: z-test with a single sample mean
- Today:
Roadmap

- Continuing with statistical inference and with hypothesis tests
- Last time: $z$-test with a single sample mean
- Today:
  - $t$-tests
Roadmap

- Continuing with statistical inference and with hypothesis tests
- Last time: $z$-test with a single sample mean
- Today:
  - $t$-tests
  - Comparing two sample means
Roadmap

- Continuing with statistical inference and with hypothesis tests
- Last time: $z$-test with a single sample mean
- Today:
  - $t$-tests
  - Comparing two sample means
  - Type I and II errors
Roadmap

- Continuing with statistical inference and with hypothesis tests
- Last time: $z$-test with a single sample mean
- Today:
  - $t$-tests
  - Comparing two sample means
  - Type I and II errors
  - Practical versus Statistical Significance
Step 1: Constructing Null/Alternative Hypotheses

Let's review these steps looking at a single population mean (from last time)

**Null Hypothesis ($H_0$):** Some statement about the population parameters

- The "Devil's Advocate" hypothesis → Assumes whatever you seek to prove did not happen
- Usually "no effect" or "no difference" or "due to chance"
- Simplest case: comparing a single population mean to some benchmark
  - Ex) $H_0: \mu = 2.5$ or $H_0: \mu = -40$

**Alternative Hypothesis ($H_a$ or $H_1$):** The statement we suspect (or hope) is true instead of $H_0$

- $H_a: \mu \neq 2.5$ (two tailed)
- $H_a: \mu > 2.5$ (one tailed)
- $H_a: \mu < 2.5$ (one tailed)

→ Choice of one-tailed versus two-tailed test affects how we calculate $p$-value
Step 1: Constructing Null/Alternative Hypotheses

Let’s review these steps looking at a single population mean (from last time)
Step 1: Constructing Null/Alternative Hypotheses

Let’s review these steps looking at a single population mean (from last time)

▶ Null Hypothesis ($H_0$): Some statement about the population parameters

- $H_0$: $\mu = 2.5$ or $H_0$: $\mu = -40$

▶ Alternative Hypothesis ($H_a$ or $H_1$): The statement we suspect (or hope) is true instead of $H_0$

- $H_a$: $\mu \neq 2.5$ (two tailed)
- $H_a$: $\mu > 2.5$ (one tailed)
- $H_a$: $\mu < 2.5$ (one tailed)

→ Choice of one-tailed versus two-tailed test affects how we calculate p-value
Step 1: Constructing Null/Alternative Hypotheses

Let’s review these steps looking at a single population mean (from last time)

- **Null Hypothesis (H₀):** Some statement about the population parameters
  - The “Devil’s Advocate” hypothesis → Assumes whatever you seek to prove did not happen

- **Alternative Hypothesis (Hₐ or H₁):** The statement we suspect (or hope) is true instead of H₀
  - Hₐ: \( \mu \neq 2.5 \) (two tailed)
  - Hₐ: \( \mu > 2.5 \) (one tailed)
  - Hₐ: \( \mu < 2.5 \) (one tailed)

→ Choice of one-tailed versus two-tailed test affects how we calculate p-value
Step 1: Constructing Null/Alternative Hypotheses

Let’s review these steps looking at a single population mean (from last time)

- **Null Hypothesis ($H_0$):** Some statement about the population parameters
  - The “Devil’s Advocate” hypothesis $\rightarrow$ Assumes whatever you seek to prove did not happen
  - Usually “no effect” or “no difference” or “due to chance”
Step 1: Constructing Null/Alternative Hypotheses

Let’s review these steps looking at a single population mean (from last time)

▶ Null Hypothesis ($H_0$): Some statement about the population parameters
  ▶ The “Devil’s Advocate” hypothesis → Assumes whatever you seek to prove did not happen
  ▶ Usually “no effect” or “no difference” or “due to chance”
  ▶ Simplest case: comparing a single population mean to some benchmark

▶ Alternative Hypothesis ($H_a$ or $H_1$): The statement we suspect (or hope) is true instead of $H_0$
Step 1: Constructing Null/Alternative Hypotheses

Let’s review these steps looking at a single population mean (from last time)

- **Null Hypothesis** \( (H_0) \): Some statement about the population parameters
  - The “Devil’s Advocate” hypothesis → Assumes whatever you seek to prove did not happen
  - Usually “no effect” or “no difference” or “due to chance”
  - Simplest case: comparing a single population mean to some benchmark

Example: \( H_0 : \mu = 2.5 \) or \( H_0 : \mu = -40 \)
Step 1: Constructing Null/Alternative Hypotheses

Let’s review these steps looking at a single population mean (from last time)

- **Null Hypothesis** \((H_0)\): Some statement about the population parameters
  - The “Devil’s Advocate” hypothesis \(\rightarrow\) Assumes whatever you seek to prove did not happen
  - Usually “no effect” or “no difference” or “due to chance”
  - Simplest case: comparing a single population mean to some benchmark
  - Ex) \(H_0 : \mu = 2.5\) or \(H_0 : \mu = -40\)

- **Alternative Hypothesis** \((H_a\) or \(H_1)\) The statement we suspect (or hope) is true instead of \(H_0\)
Step 1: Constructing Null/Alternative Hypotheses

Let’s review these steps looking at a single population mean (from last time)

- **Null Hypothesis** \((H_0)\): Some statement about the population parameters
  - The “Devil’s Advocate” hypothesis → Assumes whatever you seek to prove did not happen
  - Usually “no effect” or “no difference” or “due to chance”
  - Simplest case: comparing a single population mean to some benchmark
    - Ex) \(H_0 : \mu = 2.5\) or \(H_0 : \mu = -40\)

- **Alternative Hypothesis** \((H_a\) or \(H_1)\) The statement we suspect (or hope) is true instead of \(H_0\)
  - \(H_a : \mu \neq 2.5\) (two tailed)
Step 1: Constructing Null/Alternative Hypotheses

Let’s review these steps looking at a single population mean (from last time)

- **Null Hypothesis** ($H_0$): Some statement about the population parameters
  - The “Devil’s Advocate” hypothesis → Assumes whatever you seek to prove did not happen
  - Usually “no effect” or “no difference” or “due to chance”
  - Simplest case: comparing a single population mean to some benchmark
    - Ex) $H_0 : \mu = 2.5$ or $H_0 : \mu = -40$

- **Alternative Hypothesis** ($H_a$ or $H_1$) The statement we suspect (or hope) is true instead of $H_0$
  - $H_a : \mu \neq 2.5$ (two tailed)
  - $H_a : \mu > 2.5$ (one tailed)
Step 1: Constructing Null/Alternative Hypotheses

Let’s review these steps looking at a single population mean (from last time)

- **Null Hypothesis ($H_0$):** Some statement about the population parameters
  - The “Devil’s Advocate” hypothesis $\rightarrow$ Assumes whatever you seek to prove did not happen
  - Usually “no effect” or “no difference” or “due to chance”
  - Simplest case: comparing a single population mean to some benchmark
    - Ex) $H_0: \mu = 2.5$ or $H_0: \mu = -40$

- **Alternative Hypothesis ($H_a$ or $H_1$)** The statement we suspect (or hope) is true instead of $H_0$
  - $H_a: \mu \neq 2.5$ (two tailed)
  - $H_a: \mu > 2.5$ (one tailed)
  - $H_a: \mu < 2.5$ (one tailed)
Step 1: Constructing Null/Alternative Hypotheses

Let’s review these steps looking at a single population mean (from last time)

- **Null Hypothesis ($H_0$):** Some statement about the population parameters
  - The “Devil’s Advocate” hypothesis → Assumes whatever you seek to prove did not happen
  - Usually “no effect” or “no difference” or “due to chance”
  - Simplest case: comparing a single population mean to some benchmark
  - Ex) $H_0: \mu = 2.5$ or $H_0: \mu = -40$

- **Alternative Hypothesis ($H_a$ or $H_1$)** The statement we suspect (or hope) is true instead of $H_0$
  - $H_a: \mu \neq 2.5$ (two tailed)
  - $H_a: \mu > 2.5$ (one tailed)
  - $H_a: \mu < 2.5$ (one tailed)

→ Choice of one-tailed versus two-tailed test affects how we calculate $p$-value
Step 3: Calculating a Test Statistic

- Suppose we are interested in $H_0: \mu = \mu_0$ where $\mu_0 = 2.5$
- Sample mean follows CLT, so: $\bar{X} \sim N(\mu, \sigma^2/n)$
- Which means we can normalize $Z = \bar{X} - \mu / \sigma / \sqrt{n}$
- Assume null hypothesis is true: $\mu = 2.5$
- Gives test statistic, $z$: $z = \bar{X} - 2.5 / \sigma / \sqrt{n}$
- In practice, we have to estimate $\sigma$ using sample standard deviation
Step 3: Calculating a Test Statistic

- Suppose we are interested in \( H_0 : \mu = \mu_0 \) where \( \mu_0 = 2.5 \)
Step 3: Calculating a Test Statistic

- Suppose we are interested in \( H_0 : \mu = \mu_0 \) where \( \mu_0 = 2.5 \)
- Sample mean follows CLT, so:

\[
\bar{X} \sim N(\mu, \sigma^2/n)
\]

which means we can normalize

\[
Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}
\]

Assume null hypothesis is true: \( \mu = 2.5 \)

Gives test statistic, \( z \):

\[
z = \frac{\bar{X} - 2.5}{\sigma/\sqrt{n}}
\]

In practice, we have to estimate \( \sigma \) using sample standard deviation.
Step 3: Calculating a Test Statistic

- Suppose we are interested in $H_0 : \mu = \mu_0$ where $\mu_0 = 2.5$
- Sample mean follows CLT, so:

\[ \bar{X} \sim N(\mu, \sigma^2 / n) \]
Step 3: Calculating a Test Statistic

- Suppose we are interested in $H_0 : \mu = \mu_0$ where $\mu_0 = 2.5$
- Sample mean follows CLT, so:
  $$\bar{X} \sim N(\mu, \sigma^2/n)$$

- which means we can normalize
Step 3: Calculating a Test Statistic

- Suppose we are interested in $H_0 : \mu = \mu_0$ where $\mu_0 = 2.5$
- Sample mean follows CLT, so:

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

- which means means we can normalize

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$
Step 3: Calculating a Test Statistic

▶ Suppose we are interested in $H_0 : \mu = \mu_0$ where $\mu_0 = 2.5$
▶ Sample mean follows CLT, so:

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

▶ which means means we can normalize

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

▶ Assume null hypothesis is true: $\mu = 2.5$
Step 3: Calculating a Test Statistic

- Suppose we are interested in $H_0 : \mu = \mu_0$ where $\mu_0 = 2.5$
- Sample mean follows CLT, so:

\[ \bar{X} \sim N(\mu, \sigma^2/n) \]

- which means we can normalize

\[ Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \]

- Assume null hypothesis is true: $\mu = 2.5$
- Gives test statistic, $z$: 

\[ z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \]
Step 3: Calculating a Test Statistic

- Suppose we are interested in $H_0: \mu = \mu_0$ where $\mu_0 = 2.5$
- Sample mean follows CLT, so:
  \[
  \bar{X} \sim N(\mu, \sigma^2/n)
  \]
- which means we can normalize
  \[
  Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}
  \]
- Assume null hypothesis is true: $\mu = 2.5$
- Gives test statistic, $z$:
  \[
  z = \frac{\bar{X} - 2.5}{\sigma/\sqrt{n}}
  \]
Step 3: Calculating a Test Statistic

- Suppose we are interested in $H_0 : \mu = \mu_0$ where $\mu_0 = 2.5$
- Sample mean follows CLT, so:
  \[ \bar{X} \sim N(\mu, \sigma^2/n) \]

- which means means we can normalize
  \[ Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \]

- Assume null hypothesis is true: $\mu = 2.5$
- Gives test statistic, $z$:
  \[ z = \frac{\bar{X} - 2.5}{\sigma/\sqrt{n}} \]

- In practice, we have to estimate $\sigma$ using sample standard deviation
Step 4: Determine $p$-value

Note: If null is true (which we assumed for purposes of calculating the test statistic), $z$ should come from standard normal.

Step 4: Use test statistic (compared to standard normal) to calculate $p$-value.

$p$-value: Given that null hypothesis true, what is probability of seeing test statistic as extreme as the one we got?
Step 4: Determine $p$-value

- Note: If null is true (which we assumed for purposes of calculating the test statistic), $z$ should come from standard normal
Step 4: Determine $p$-value

- Note: If null is true (which we assumed for purposes of calculating the test statistic), $z$ should come from standard normal
- Step 4: Use test statistic (compared to standard normal) to calculate $p$-value
Step 4: Determine $p$-value

- **Note:** If null is true (which we assumed for purposes of calculating the test statistic), $z$ should come from standard normal.
- **Step 4:** Use test statistic (compared to standard normal) to calculate $p$-value.
- **$p$-value:** Given that null hypothesis true, what is probability of seeing test statistic as extreme as the one we got?
Step 4: Determine $p$-value

**z TEST FOR A POPULATION MEAN**

To test the hypothesis $H_0: \mu = \mu_0$ based on an SRS of size $n$ from a population with unknown mean $\mu$ and known standard deviation $\sigma$, compute the test statistic

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

In terms of a standard Normal random variable $Z$, the $P$-value for a test of $H_0$ against

- $H_a: \mu > \mu_0$ is $P(Z \geq z)$
- $H_a: \mu < \mu_0$ is $P(Z \leq z)$
- $H_a: \mu \neq \mu_0$ is $2P(Z \geq |z|)$

These $P$-values are exact if the population distribution is Normal and are approximately correct for large $n$ in other cases.
Step 5: Reject or do not reject null hypothesis

- Given the p-value, consider whether to reject the null hypothesis.

Some rules of thumb regarding critical values:

<table>
<thead>
<tr>
<th>p-value</th>
<th>Accepted Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 0.10</td>
<td>No evidence to reject $H_0$</td>
</tr>
<tr>
<td>$0.05 &lt;$</td>
<td>Weak evidence to reject $H_0$</td>
</tr>
<tr>
<td>$0.01 &lt;$</td>
<td>Some evidence to reject $H_0$</td>
</tr>
<tr>
<td>$0.001 &lt;$</td>
<td>Strong evidence to reject $H_0$</td>
</tr>
<tr>
<td>$\leq 0.001$</td>
<td>Very strong evidence to reject $H_0$</td>
</tr>
</tbody>
</table>

Note: The smaller the p-value, the more extreme it is, given the null hypothesis. This gives you more reason to reject the null hypothesis.
Step 5: Reject or do not reject null

▶ Step 5: Given $p$-value, consider whether to reject the null hypothesis
Step 5: Reject or do not reject null

- Step 5: Given \( p \)-value, consider whether to reject the null hypothesis
- Some rules of thumb regarding critical values:
Step 5: Reject or do not reject null

- Step 5: Given $p$-value, consider whether to reject the null hypothesis
- Some rules of thumb regarding critical values:

<table>
<thead>
<tr>
<th>$p$-value</th>
<th>Accepted Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&gt; 0.10$</td>
<td>No evidence to reject $H_0$</td>
</tr>
<tr>
<td>$0.05 &lt; p \leq 0.10$</td>
<td>Weak evidence to reject $H_0$</td>
</tr>
<tr>
<td>$0.01 &lt; p \leq 0.05$</td>
<td>Some evidence to reject $H_0$</td>
</tr>
<tr>
<td>$0.001 &lt; p \leq 0.01$</td>
<td>Strong evidence to reject $H_0$</td>
</tr>
<tr>
<td>$p \leq 0.001$</td>
<td>Very strong evidence to reject $H_0$</td>
</tr>
</tbody>
</table>
Step 5: Reject or do not reject null

- Step 5: Given $p$-value, consider whether to reject the null hypothesis

- Some rules of thumb regarding critical values:

<table>
<thead>
<tr>
<th>$p$-value</th>
<th>Accepted Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$-value $&gt; 0.10$</td>
<td>No evidence to reject $H_0$</td>
</tr>
</tbody>
</table>
Step 5: Reject or do not reject null

- Step 5: Given $p$-value, consider whether to reject the null hypothesis
- Some rules of thumb regarding critical values:

<table>
<thead>
<tr>
<th>$p$-value</th>
<th>Accepted Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$-value $&gt; 0.10$</td>
<td>No evidence to reject $H_0$</td>
</tr>
<tr>
<td>$0.05 &lt; p$-value $\leq 0.10$</td>
<td>Weak evidence to reject $H_0$</td>
</tr>
</tbody>
</table>
Step 5: Reject or do not reject null

- Step 5: Given $p$-value, consider whether to reject the null hypothesis

- Some rules of thumb regarding critical values:

<table>
<thead>
<tr>
<th>$p$-value</th>
<th>Accepted Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$-value &gt; 0.10</td>
<td>No evidence to reject $H_0$</td>
</tr>
<tr>
<td>0.05 &lt; $p$-value ≤ 0.10</td>
<td>Weak evidence to reject $H_0$</td>
</tr>
<tr>
<td>0.01 &lt; $p$-value ≤ 0.05</td>
<td>Some evidence to reject $H_0$</td>
</tr>
</tbody>
</table>
Step 5: Reject or do not reject null

- Step 5: Given $p$-value, consider whether to reject the null hypothesis

- Some rules of thumb regarding critical values:

<table>
<thead>
<tr>
<th>$p$-value</th>
<th>Accepted Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$-value $&gt; 0.10$</td>
<td>No evidence to reject $H_0$</td>
</tr>
<tr>
<td>$0.05 &lt; p$-value $\leq 0.10$</td>
<td>Weak evidence to reject $H_0$</td>
</tr>
<tr>
<td>$0.01 &lt; p$-value $\leq 0.05$</td>
<td>Some evidence to reject $H_0$</td>
</tr>
<tr>
<td>$0.001 &lt; p$-value $\leq 0.01$</td>
<td>Strong evidence to reject $H_0$</td>
</tr>
</tbody>
</table>

Note: The smaller the $p$-value the more extreme it is, given the null hypothesis. This gives you more reason to reject the null hypothesis.
Step 5: Reject or do not reject null

- Step 5: Given $p$-value, consider whether to reject the null hypothesis
- Some rules of thumb regarding critical values:

<table>
<thead>
<tr>
<th>$p$-value</th>
<th>Accepted Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$-value $&gt; 0.10$</td>
<td>No evidence to reject $H_0$</td>
</tr>
<tr>
<td>$0.05 &lt; p$-value $\leq 0.10$</td>
<td>Weak evidence to reject $H_0$</td>
</tr>
<tr>
<td>$0.01 &lt; p$-value $\leq 0.05$</td>
<td>Some evidence to reject $H_0$</td>
</tr>
<tr>
<td>$0.001 &lt; p$-value $\leq 0.01$</td>
<td>Strong evidence to reject $H_0$</td>
</tr>
<tr>
<td>$p$-value $\leq 0.001$</td>
<td>Very strong evidence to reject $H_0$</td>
</tr>
</tbody>
</table>
Step 5: Reject or do not reject null hypothesis

- Step 5: Given \( p \)-value, consider whether to reject the null hypothesis

- Some rules of thumb regarding critical values:

<table>
<thead>
<tr>
<th>( p )-value</th>
<th>Accepted Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p )-value &gt; 0.10</td>
<td>No evidence to reject ( H_0 )</td>
</tr>
<tr>
<td>0.05 &lt; ( p )-value ≤ 0.10</td>
<td>Weak evidence to reject ( H_0 )</td>
</tr>
<tr>
<td>0.01 &lt; ( p )-value ≤ 0.05</td>
<td>Some evidence to reject ( H_0 )</td>
</tr>
<tr>
<td>0.001 &lt; ( p )-value ≤ 0.01</td>
<td>Strong evidence to reject ( H_0 )</td>
</tr>
<tr>
<td>( p )-value ≤ 0.001</td>
<td>Very strong evidence to reject ( H_0 )</td>
</tr>
</tbody>
</table>

- Note: The smaller the \( p \)-value the more extreme it is, given the null
Step 5: Reject or do not reject null

- Step 5: Given $p$-value, consider whether to reject the null hypothesis

- Some rules of thumb regarding critical values:

<table>
<thead>
<tr>
<th>$p$-value</th>
<th>Accepted Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$-value $&gt; 0.10$</td>
<td>No evidence to reject $H_0$</td>
</tr>
<tr>
<td>$0.05 &lt; p$-value $\leq 0.10$</td>
<td>Weak evidence to reject $H_0$</td>
</tr>
<tr>
<td>$0.01 &lt; p$-value $\leq 0.05$</td>
<td>Some evidence to reject $H_0$</td>
</tr>
<tr>
<td>$0.001 &lt; p$-value $\leq 0.01$</td>
<td>Strong evidence to reject $H_0$</td>
</tr>
<tr>
<td>$p$-value $\leq 0.001$</td>
<td>Very strong evidence to reject $H_0$</td>
</tr>
</tbody>
</table>

- Note: The smaller the $p$-value the more extreme it is, given the null

- Gives you more reason to reject the null
Two Potential Problems

1. Reasoning relies on CLT
   - CLT relies on sampling distributions approximating normal as \( n \) goes up
   - What if you have small \( n \)?
   - Using CLT questionable

2. We used sample parameters to approximate the standard normal
   - (We used sample standard deviation in place of population standard deviation)
   - But standardizing assumes you use the actual population parameters (which we don't know)
   - Thus, we introduced additional uncertainty
   - To take this into account, nearly all hypothesis testing uses Student's \( t \) distribution instead of normal
Two Potential Problems

Used standard normal in example, but might not be wise – why?

1. Reasoning relies on CLT
   - CLT relies on sampling distributions approximating normal as $n$ goes up
   - What if you have small $n$?
   - Using CLT questionable

2. We used sample parameters to approximate the standard normal
   - (We used sample standard deviation in place of population standard deviation)
   - But standardizing assumes you use the actual population parameters (which we don't know)
   - Thus, we introduced additional uncertainty
   - To take this into account, nearly all hypothesis testing uses Student's t distribution instead of normal
Two Potential Problems

Used standard normal in example, but might not be wise – why?
1. Reasoning relies on CLT

   ▶ CLT relies on sampling distributions approximating normal as \( n \) goes up
   ▶ What if you have small \( n \)?
   ▶ Using CLT questionable

   ▶ We used sample parameters to approximate the standard normal
     ▶ (We used sample standard deviation in place of population standard deviation)
     ▶ But standardizing assumes you use the actual population parameters (which we don't know)
     ▶ Thus, we introduced additional uncertainty
     ▶ To take this into account, nearly all hypothesis testing uses Student's \( t \) distribution instead of normal
Two Potential Problems

Used standard normal in example, but might not be wise – why?

1. Reasoning relies on CLT
   - CLT relies on sampling distributions approximating normal as \( n \) goes up

2. We used sample parameters to approximate the standard normal
   - (We used sample standard deviation in place of population standard deviation)
   - But standardizing assumes you use the actual population parameters (which we don’t know)
   - Thus, we introduced additional uncertainty
   - To take this into account, nearly all hypothesis testing uses Student’s \( t \) distribution instead of normal
Two Potential Problems

Used standard normal in example, but might not be wise – why?

1. Reasoning relies on CLT
   - CLT relies on sampling distributions approximating normal as \( n \) goes up
   - What if you have small \( n \)?

2. We used sample parameters to approximate the standard normal
   - (We used sample standard deviation in place of population standard deviation)
   - But standardizing assumes you use the actual population parameters (which we don't know)
   - Thus, we introduced additional uncertainty
   - To take this into account, nearly all hypothesis testing uses Student's \( t \) distribution instead of normal
Two Potential Problems

Used standard normal in example, but might not be wise – why?

1. Reasoning relies on CLT
   - CLT relies on sampling distributions approximating normal as $n$ goes up
   - What if you have small $n$?
   - Using CLT questionable
Two Potential Problems

Used standard normal in example, but might not be wise – why?

1. Reasoning relies on CLT
   - CLT relies on sampling distributions approximating normal as \( n \) goes up
   - What if you have small \( n \)?
   - Using CLT questionable

2. We used sample parameters to approximate the standard normal
Two Potential Problems

Used standard normal in example, but might not be wise – why?

1. Reasoning relies on CLT
   ▶ CLT relies on sampling distributions approximating normal as $n$ goes up
   ▶ What if you have small $n$?
   ▶ Using CLT questionable

2. We used sample parameters to approximate the standard normal
   ▶ (We used sample standard deviation in place of population standard deviation)
Two Potential Problems

Used standard normal in example, but might not be wise – why?

1. Reasoning relies on CLT
   ▶ CLT relies on sampling distributions approximating normal as $n$ goes up
   ▶ What if you have small $n$?
   ▶ Using CLT questionable

2. We used sample parameters to approximate the standard normal
   ▶ (We used sample standard deviation in place of population standard deviation)
   ▶ But standardizing assumes you use the actual population parameters (which we don’t know)
Two Potential Problems

Used standard normal in example, but might not be wise – why?

1. Reasoning relies on CLT
   - CLT relies on sampling distributions approximating normal as $n$ goes up
   - What if you have small $n$?
   - Using CLT questionable

2. We used sample parameters to approximate the standard normal
   - (We used sample standard deviation in place of population standard deviation)
   - But standardizing assumes you use the actual population parameters (which we don’t know)
   - Thus, we introduced additional uncertainty
Two Potential Problems

Used standard normal in example, but might not be wise – why?

1. Reasoning relies on CLT
   ▶ CLT relies on sampling distributions approximating normal as $n$ goes up
   ▶ What if you have small $n$?
   ▶ Using CLT questionable

2. We used sample parameters to approximate the standard normal
   ▶ (We used sample standard deviation in place of population standard deviation)
   ▶ But standardizing assumes you use the actual population parameters (which we don’t know)
   ▶ Thus, we introduced additional uncertainty

▶ To take this into account, nearly all hypothesis testing uses Student’s $t$ distribution instead of normal
Who was “Student”? 
Who was “Student”?
Who was “Student”?

[Image of a man with a moustache and glasses]
Who was "Student"?

**Vol. VI March, 1908 No. 1**

**BIOMETRIKA.**

**The Probable Error of a Mean.**

By Student.

*Introduction.*

Any experiment may be regarded as forming an individual of a "population" of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong. In a great number of cases the question finally turns on the value of a mean, either directly, or as the mean difference between the two quantities.

If the number of experiments be very large, we may have precise information
Who was “Student”??
Student’s $t$-distribution

- Similar in shape to Normal distribution, but with fatter tails
- For sample sizes $> 100$, $t$-distribution and $N(0, 1)$ distributions virtually identical
- Thus: Use $t$-distribution to be conservative, but inferences converge as $n$ goes up
Student’s $t$-distribution

- Similar in shape to Normal distribution, but with fatter tails
Student’s $t$-distribution

- Similar in shape to Normal distribution, but with fatter tails
- For sample sizes $> 100$, $t$ distribution and $N(0, 1)$ distributions virtually identical
Student’s $t$-distribution

- Similar in shape to Normal distribution, but with fatter tails
- For sample sizes $>100$, $t$ distribution and $N(0,1)$ distributions virtually identical
- Thus: Use $t$ distribution to be conservative, but inferences converge as $n$ goes up
Student’s $t$-distribution

▶ $t$-distribution shape determined by size of sample

▶ Exact shape requires knowing the degrees of freedom, $\nu$ or $\nu$

▶ Degrees of freedom takes into account # of observations and fact that you need data to estimate parameters

▶ For the sample mean, $\nu = n - 1$

▶ Thus, if 50 observations $\nu = 50 - 1$ (or 49)
Student’s $t$-distribution

- $t$ distribution shape determined by size of sample
  
  $t$ distribution shape determined by size of sample

- Degrees of freedom takes into account number of observations and fact that you need data to estimate parameters
  
  For the sample mean, $\nu = n - 1$

  Thus, if 50 observations, $\nu = 50 - 1$ (or 49)
Student’s $t$-distribution

- $t$ distribution shape determined by size of sample
- Exact shape requires knowing the **degrees of freedom**, $df$ or $ν$
Student’s $t$-distribution

- $t$ distribution shape determined by size of sample
- Exact shape requires knowing the degrees of freedom, $df$ or $\nu$
- Degrees of freedom takes into account # of observations and fact that you need data to estimate parameters
Student’s $t$-distribution

- $t$ distribution shape determined by size of sample
- Exact shape requires knowing the degrees of freedom, $df$ or $\nu$
- Degrees of freedom takes into account # of observations and fact that you need data to estimate parameters
- For the sample mean, $\nu = n - 1$
Student’s $t$-distribution

- $t$ distribution shape determined by size of sample
- Exact shape requires knowing the degrees of freedom, $df$ or $\nu$
- Degrees of freedom takes into account # of observations and fact that you need data to estimate parameters
- For the sample mean, $\nu = n - 1$
- Thus, if 50 observations $\nu = 50 - 1$ (or 49)
Student’s $t$-distribution

$Z \sim N(0, 1)$

$T$ with $df = 15$

$T$ with $df = 5$
Student's $t$-distribution

The probability density function (pdf) for the $t$-distribution is:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)\sqrt{\nu\pi} \times \Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi} \times \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}}$$

where $\nu$ is degrees of freedom.

Test statistic calculated similarly to before:

$$t_{df} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

and we compare this to the appropriate $t$-distribution (with $df$) as opposed to standard normal.
Student’s $t$-distribution

- The probability density function (pdf) for the $t$-distribution is:

$$
\begin{align*}
  f(x) &= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \\
  \text{where } \nu \text{ is degrees of freedom} \\
  \text{and } \Gamma(n) = (n-1)!
\end{align*}
$$

- Test statistic calculated similarly to before:

$$
  t_{df} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}
$$

and we compare this to the appropriate $t$-distribution (with $df$) as opposed to standard normal.
Student’s $t$-distribution

The probability density function (pdf) for the $t$-distribution is:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu \pi} \times \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}$$

where $\nu$ is degrees of freedom.
Student’s $t$-distribution

- The probability density function (pdf) for the $t$-distribution is:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \times \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

- where $\nu$ is degrees of freedom
Student’s *t*-distribution

- The probability density function (pdf) for the *t*-distribution is:

\[
f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \times \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}
\]

- where \(\nu\) is degrees of freedom
- and \(\Gamma(n) = (n - 1)!\)
Student’s $t$-distribution

- The probability density function (pdf) for the t-distribution is:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \times \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

- where $\nu$ is degrees of freedom
- and $\Gamma(n) = (n - 1)!$
- Test statistic calculated similarly to before:
Student’s $t$-distribution

- The probability density function (pdf) for the $t$-distribution is:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \times \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\left(\nu+1\right)/2}$$

- where $\nu$ is degrees of freedom
- and $\Gamma(n) = (n - 1)!$
- Test statistic calculated similarly to before:

$$t_{df} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$
Student’s $t$-distribution

- The probability density function (pdf) for the $t$-distribution is:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \times \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}$$

- where $\nu$ is degrees of freedom
- and $\Gamma(n) = (n - 1)!$
- Test statistic calculated similarly to before:

$$t_{df} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

- and we compare this to the appropriate $t$ distribution (with $df$) as opposed to standard normal
When to use z-test versus t-tests?

**z-tests:**
- Follow normal distribution
- Assume you know true population standard deviation
- More accurate than t-tests when population standard deviations ($\sigma$) known
- Converges to a t-test with larger sample sizes ($n > 30$)

**t-tests**
- Follow student's t-distribution
- More conservative than z-test, b/c of fatter tails
- So better when you use sample standard deviation (instead of $\sigma$)
- Appropriate w/ small samples ($n \leq 30$)
- B/c you usually don't know $\sigma$, t-tests more common than z-tests (used in STATA, R)
When to use z-test versus t-tests?

- **z-tests:**
  - Follow normal distribution
  - Assume you know true population standard deviation
  - More accurate than t-tests when population standard deviations (σ) known
  - Converges to a t-test with larger sample sizes (n > 30)

- **t-tests**
  - Follow student's t-distribution
  - More conservative than z-test, because of fatter tails
  - So better when you use sample standard deviation (instead of σ)
  - Appropriate with small samples (n ≤ 30)
  - Because you usually don't know σ, t-tests more common than z-tests (used in STATA, R)
When to use \textit{z}-test versus \textit{t}-tests?

\begin{itemize}
  \item \textit{z}-tests:
    \begin{itemize}
      \item Follow normal distribution
    \end{itemize}
\end{itemize}
When to use z-test versus t-tests?

- **z-tests:**
  - Follow normal distribution
  - Assume you know true population standard deviation

- **t-tests:**
  - Follow student's t-distribution
  - More conservative than z-test, because of fatter tails
  - So better when you use sample standard deviation (instead of \( \sigma \))
  - Appropriate with small samples (\( n \leq 30 \))
  - Because you usually don't know \( \sigma \), t-tests more common than z-tests (used in STATA, R)
When to use z-test versus t-tests?

- **z-tests:**
  - Follow normal distribution
  - Assume you know true population standard deviation
  - More accurate than t-tests when population standard deviations ($\sigma$) known

- **t-tests:**
  - Follow student's $t$-distribution
  - More conservative than z-test, because of fatter tails
  - So better when you use sample standard deviation (instead of $\sigma$)
  - Appropriate with small samples ($n \leq 30$)
  - Because you usually don't know $\sigma$, t-tests more common than z-tests (used in STATA, R)
When to use z-test versus t-tests?

- **z-tests:**
  - Follow normal distribution
  - Assume you know true population standard deviation
  - More accurate than t-tests when population standard deviations ($\sigma$) known
  - Converges to a t-test with larger sample sizes ($n > 30$)

- **t-tests**

- More conservative than z-test, because of fatter tails
- So better when you use sample standard deviation (instead of $\sigma$)
- Appropriate with small samples ($n \leq 30$)
- Because you usually don't know $\sigma$, t-tests more common than z-tests (used in STATA, R)
When to use $z$-test versus $t$-tests?

- **$z$-tests:**
  - Follow normal distribution
  - Assume you know true population standard deviation
  - More accurate than $t$-tests when population standard deviations ($\sigma$) known
  - Converges to a $t$-test with larger sample sizes ($n > 30$)

- **$t$-tests**
  - Follow student’s $t$-distribution
When to use $z$-test versus $t$-tests?

- **$z$-tests:**
  - Follow normal distribution
  - Assume you know true population standard deviation
  - More accurate than $t$-tests when population standard deviations ($\sigma$) known
  - Converges to a $t$-test with larger sample sizes ($n > 30$)

- **$t$-tests**
  - Follow student’s $t$-distribution
  - More conservative than $z$-test, b/c of fatter tails
When to use $z$-test versus $t$-tests?

- **$z$-tests:**
  - Follow normal distribution
  - Assume you know true population standard deviation
  - More accurate than $t$-tests when population standard deviations ($\sigma$) known
  - Converges to a $t$-test with larger sample sizes ($n > 30$)

- **$t$-tests**
  - Follow student’s $t$-distribution
  - More conservative than $z$-test, b/c of fatter tails
  - So better when you use sample standard deviation (instead of $\sigma$)
When to use z-test versus t-tests?

- **z-tests:**
  - Follow normal distribution
  - Assume you know true population standard deviation
  - More accurate than t-tests when population standard deviations ($\sigma$) known
  - Converges to a t-test with larger sample sizes ($n > 30$)

- **t-tests**
  - Follow student’s t-distribution
  - More conservative than z-test, b/c of fatter tails
  - So better when you use sample standard deviation (instead of $\sigma$)
  - Appropriate w/ small samples ($n \leq 30$)
When to use z-test versus t-tests?

- **z-tests:**
  - Follow normal distribution
  - Assume you know true population standard deviation
  - More accurate than t-tests when population standard deviations ($\sigma$) known
  - Converges to a t-test with larger sample sizes ($n > 30$)

- **t-tests**
  - Follow student’s $t$-distribution
  - More conservative than z-test, b/c of fatter tails
  - So better when you use sample standard deviation (instead of $\sigma$)
  - Appropriate w/ small samples ($n \leq 30$)

- B/c you usually don’t know $\sigma$, t-tests more common than z-tests (used in STATA, R)
Hypothesis Tests for Comparing Groups

Let's make this more realistic

Up to now: Testing whether \( \mu \) equals some benchmark (e.g., 2.5 ppl per household)

- OK when we have some benchmark to compare our sample to
- Ex) Given 100 jobs, are half (0.50) going to women?

More common: Interested in comparing 2 samples to each other, trying to make inferences about two population means

- Ex) Comparing 2 different observational samples (e.g., African-American vs white income)
- Ex) 2 different sets of experimental conditions (e.g., those receiving vaccines vs not)

Note: For now, important that 2 groups contain independent groups of people (each subject provides only one value)
Hypothesis Tests for Comparing Groups

- Let’s make this more realistic
Hypothesis Tests for Comparing Groups

- Let’s make this more realistic
- Up to now: Testing whether $\mu$ equals some benchmark (e.g., 2.5 ppl per household)

OK when we have some benchmark to compare our sample to

Ex) Given 100 jobs, are half (0.50) going to women?

More common: Interested in comparing 2 samples to each other, trying to make inferences about two population means

Ex) Comparing 2 different observational samples (e.g., African-American vs white income)

Ex) 2 different sets of experimental conditions (e.g., those receiving vaccines vs not)

Note: For now, important that 2 groups contain independent groups of people (each subject provides only one value)
Hypothesis Tests for Comparing Groups

- Let’s make this more realistic
- Up to now: Testing whether \( \mu \) equals some benchmark (e.g., 2.5 ppl per household)
  - OK when we have some benchmark to compare our sample to
Hypothesis Tests for Comparing Groups

- Let’s make this more realistic
- Up to now: Testing whether \( \mu \) equals some benchmark (e.g., 2.5 ppl per household)
  - OK when we have some benchmark to compare our sample to
  - Ex) Given 100 jobs, are half (0.50) going to women?
Hypothesis Tests for Comparing Groups

- Let’s make this more realistic
- Up to now: Testing whether \( \mu \) equals some benchmark (e.g., 2.5 ppl per household)
  - OK when we have some benchmark to compare our sample to
  - Ex) Given 100 jobs, are half (0.50) going to women?
- More common: Interested in comparing 2 samples to each other, trying to make inferences about two population means
Hypothesis Tests for Comparing Groups

▶ Let’s make this more realistic
▶ Up to now: Testing whether \( \mu \) equals some benchmark (e.g., 2.5 ppl per household)
  ▶ OK when we have some benchmark to compare our sample to
  ▶ Ex) Given 100 jobs, are half (0.50) going to women?
▶ More common: Interested in comparing 2 samples to each other, trying to make inferences about two population means
  ▶ Ex) Comparing 2 different observational samples (e.g., African-American vs white income)
Hypothesis Tests for Comparing Groups

- Let’s make this more realistic
- Up to now: Testing whether $\mu$ equals some benchmark (e.g., 2.5 ppl per household)
  - OK when we have some benchmark to compare our sample to
  - Ex) Given 100 jobs, are half (0.50) going to women?
- More common: Interested in comparing 2 samples to each other, trying to make inferences about two population means
  - Ex) Comparing 2 different observational samples (e.g., African-American vs white income)
  - Ex) 2 different sets of experimental conditions (e.g., those receiving vaccines vs not)
Hypothesis Tests for Comparing Groups

- Let’s make this more realistic
- Up to now: Testing whether \( \mu \) equals some benchmark (e.g., 2.5 ppl per household)
  - OK when we have some benchmark to compare our sample to
  - Ex) Given 100 jobs, are half (0.50) going to women?
- More common: Interested in comparing 2 samples to each other, trying to make inferences about two population means
  - Ex) Comparing 2 different observational samples (e.g., African-American vs white income)
  - Ex) 2 different sets of experimental conditions (e.g., those receiving vaccines vs not)
- Note: For now, important that 2 groups contain independent groups of people (each subject provides only one value)
Hypothesis Tests for Comparing Groups

Specifically: Our goal is to use two samples to draw conclusions about the difference between two population means $\mu_1$ and $\mu_2$.

This is known as a difference-in-means test.

What do we need?

- Two samples (independent, so no pairs)
- Sample means and sample standard deviations
- Number of observations in each sample, $n_1$ and $n_2$
Hypothesis Tests for Comparing Groups

Specifically: Our goal is to use two samples to draw conclusions about the difference between two population means $\mu_1$ and $\mu_2$. This is known as a difference-in-means test.

- We need two samples (independent, so no pairs)
- Sample means and sample standard deviations
- Number of observations in each sample, $n_1$ and $n_2$
Hypothesis Tests for Comparing Groups

▶ Specifically: Our goal is to use two samples to draw conclusions about difference between two population means $\mu_1$ and $\mu_2$
  ▶ That is, $\mu_1 - \mu_2$
Hypothesis Tests for Comparing Groups

- Specifically: Our goal is to use two samples to draw conclusions about difference between two population means \( \mu_1 \) and \( \mu_2 \)
  - That is, \( \mu_1 - \mu_2 \)
  - This is known as a difference-in-means test
Hypothesis Tests for Comparing Groups

- Specifically: Our goal is to use two samples to draw conclusions about difference between two population means $\mu_1$ and $\mu_2$
  - That is, $\mu_1 - \mu_2$
  - This is known as a difference-in-means test
- What do we need?
Hypothesis Tests for Comparing Groups

▶ Specifically: Our goal is to use two samples to draw conclusions about difference between two population means $\mu_1$ and $\mu_2$
  ▶ That is, $\mu_1 - \mu_2$
  ▶ This is known as a difference-in-means test

▶ What do we need?
  ▶ Two samples (independent, so no pairs)
Hypothesis Tests for Comparing Groups

- Specifically: Our goal is to use two samples to draw conclusions about difference between two population means \( \mu_1 \) and \( \mu_2 \)
  - That is, \( \mu_1 - \mu_2 \)
  - This is known as a difference-in-means test

- What do we need?
  - Two samples (independent, so no pairs)
  - Sample means and sample standard deviations
Hypothesis Tests for Comparing Groups

Specifically: Our goal is to use two samples to draw conclusions about difference between two population means $\mu_1$ and $\mu_2$

- That is, $\mu_1 - \mu_2$
- This is known as a difference-in-means test

What do we need?

- Two samples (independent, so no pairs)
- Sample means and sample standard deviations
- Number of observations in each sample, $n_1$ and $n_2$
Childhood Immunization Example

You study vaccination rates and childhood health outcomes.

Your research team has gathered important data:

- Sampled 91 countries with low vaccination rate (fewer than 90% of infants immunized) → average mortality rate of 92.42 (per 1k births)
- Sampled 97 countries with high vaccination rate (more than 90% of infants immunized) → average mortality rate of 24.97 (per 1k births)

Given the two samples, is there a meaningful difference in their population childhood mortality?
Childhood Immunization Example

- You study vaccination rates and childhood health outcomes
Childhood Immunization Example

- You study vaccination rates and childhood health outcomes
- Your research team has gathered important data:
  - Sampled 91 countries with low vaccination rate (fewer than 90% of infants immunized) → average mortality rate of 92.42
  - Sampled 97 countries with high vaccination rate (more than 90% of infants immunized) → average mortality rate of 24.97 (per 1k births)
  - Given the two samples, is there a meaningful difference in their population childhood mortality?
Childhood Immunization Example

- You study vaccination rates and childhood health outcomes
- Your research team has gathered important data:
  - Sampled 91 countries with low vaccination rate (fewer than 90% of infants immunized) → average mortality rate of 92.42
You study vaccination rates and childhood health outcomes.

Your research team has gathered important data:

- Sampled 91 countries with low vaccination rate (fewer than 90% of infants immunized) → average mortality rate of 92.42 per 1k births.
- Sampled 97 countries with high vaccination rate (more than 90% of infants immunized) → average mortality rate of 24.97 per 1k births.
Childhood Immunization Example

- You study vaccination rates and childhood health outcomes
- Your research team has gathered important data:
  - Sampled 91 countries with low vaccination rate (fewer than 90% of infants immunized) → average mortality rate of 92.42
  - Sampled 97 countries with high vaccination rate (more than 90% of infants immunized) → average mortality rate of 24.97 (per 1k births)
- Given the two samples, is there a meaningful difference in their population childhood mortality?
## Childhood Immunization Example

<table>
<thead>
<tr>
<th>Group</th>
<th>Country</th>
<th>Status</th>
<th>n</th>
<th>( \bar{x} )</th>
<th>s</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>90% immunized</td>
<td>91</td>
<td>92.42</td>
<td>73.21</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>&gt;90% immunized</td>
<td>97</td>
<td>24.98</td>
<td>30.92</td>
</tr>
</tbody>
</table>
Childhood Immunization Example

Full data look like:

<table>
<thead>
<tr>
<th>Group</th>
<th>Country</th>
<th>Status</th>
<th>n</th>
<th>( \bar{X} )</th>
<th>s</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>( \leq 90% ) immunized</td>
<td>91</td>
<td>92.42</td>
<td>73.21</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>( &gt; 90% ) immunized</td>
<td>97</td>
<td>24.98</td>
<td>30.92</td>
</tr>
</tbody>
</table>
Childhood Immunization Example

Full data look like:

<table>
<thead>
<tr>
<th>Group</th>
<th>Country Status</th>
<th>n</th>
<th>$\bar{X}$</th>
<th>s</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\leq 90%$ immunized</td>
<td>91</td>
<td>92.42</td>
<td>73.21</td>
</tr>
<tr>
<td>2</td>
<td>$&gt; 90%$ immunized</td>
<td>97</td>
<td>24.98</td>
<td>30.92</td>
</tr>
</tbody>
</table>
Childhood Immunization Example

What are our populations under study?

- All Group 1 children
- All Group 2 children

Are they independent?

- Are the subjects only in one group or the other?
  - Probably safe to assume here
  - Will consider paired observations later

What is parameter (or parameters) we are interested in finding out about?

- $\mu_1 =$ mean in countries with $\leq 90\%$ immunized
- $\mu_2 =$ mean in countries with $> 90\%$ immunized

Specifically, whether difference between $\mu_1$ and $\mu_2$

Now ready for difference in means test
Childhood Immunization Example

- What are our populations under study?

\[
\mu_1 = \text{mean in countries with } \leq 90\% \text{ immunized}
\]

\[
\mu_2 = \text{mean in countries with } > 90\% \text{ immunized}
\]

Specifically, whether difference between \( \mu_1 \) and \( \mu_2 \)

Now ready for difference in means test
Childhood Immunization Example

- What are our populations under study?
  - All Group 1 children
Childhood Immunization Example

- What are our populations under study?
  - All Group 1 children
  - All Group 2 children
Childhood Immunization Example

- What are our populations under study?
  - All Group 1 children
  - All Group 2 children
- Are they independent?

- What is parameter (or parameters) we are interested in finding out about?
  - \( \mu_1 \) = mean in countries with \( \leq 90\% \) immunized
  - \( \mu_2 \) = mean in countries with \( > 90\% \) immunized
- Specifically, whether difference between \( \mu_1 \) and \( \mu_2 \)
- Now ready for difference in means test
Childhood Immunization Example

- What are our populations under study?
  - All Group 1 children
  - All Group 2 children
- Are they independent?
  - Are the subjects only in one group or the other?

\[ \mu_1 = \text{mean in countries with } \leq 90\% \text{ immunized} \]
\[ \mu_2 = \text{mean in countries with } > 90\% \text{ immunized} \]

Now ready for difference in means test
Childhood Immunization Example

- What are our populations under study?
  - All Group 1 children
  - All Group 2 children

- Are they independent?
  - Are the subjects only in one group or the other?
  - Probably safe to assume here

\[ \mu_1 = \text{mean in countries with } \leq 90\% \text{ immunized} \]
\[ \mu_2 = \text{mean in countries with } > 90\% \text{ immunized} \]

Now ready for difference in means test.
Childhood Immunization Example

▶ What are our populations under study?
  ▶ All Group 1 children
  ▶ All Group 2 children

▶ Are they independent?
  ▶ Are the subjects only in one group or the other?
  ▶ Probably safe to assume here
  ▶ Will consider paired observations later
Childhood Immunization Example

- What are our populations under study?
  - All Group 1 children
  - All Group 2 children

- Are they independent?
  - Are the subjects only in one group or the other?
  - Probably safe to assume here
  - Will consider paired observations later

- What is parameter (or parameters) we are interested in finding out about?
  - $\mu_1 =$ mean in countries with $\leq 90\%$ immunized
  - $\mu_2 =$ mean in countries with $> 90\%$ immunized
  - Specifically, whether difference between $\mu_1$ and $\mu_2$
Childhood Immunization Example

- What are our populations under study?
  - All Group 1 children
  - All Group 2 children

- Are they independent?
  - Are the subjects only in one group or the other?
  - Probably safe to assume here
  - Will consider paired observations later

- What is parameter (or parameters) we are interested in finding out about?
  - $\mu_1 =$ mean in countries with $\leq$ 90% immunized
Childhood Immunization Example

- What are our populations under study?
  - All Group 1 children
  - All Group 2 children

- Are they independent?
  - Are the subjects only in one group or the other?
    - Probably safe to assume here
    - Will consider paired observations later

- What is parameter (or parameters) we are interested in finding out about?
  - $\mu_1 = \text{mean in countries with} \leq 90\% \text{ immunized}$
  - $\mu_2 = \text{mean in countries with} > 90\% \text{ immunized}$
Childhood Immunization Example

- What are our populations under study?
  - All Group 1 children
  - All Group 2 children

- Are they independent?
  - Are the subjects only in one group or the other?
  - Probably safe to assume here
  - Will consider paired observations later

- What is parameter (or parameters) we are interested in finding out about?
  - $\mu_1 = \text{mean in countries with } \leq 90\% \text{ immunized}$
  - $\mu_2 = \text{mean in countries with } > 90\% \text{ immunized}$
  - Specifically, whether difference between $\mu_1$ and $\mu_2$
Childhood Immunization Example

- What are our populations under study?
  - All Group 1 children
  - All Group 2 children

- Are they independent?
  - Are the subjects only in one group or the other?
  - Probably safe to assume here
  - Will consider paired observations later

- What is parameter (or parameters) we are interested in finding out about?
  - $\mu_1 = \text{mean in countries with } \leq 90\% \text{ immunized}$
  - $\mu_2 = \text{mean in countries with } > 90\% \text{ immunized}$
  - Specifically, whether difference between $\mu_1$ and $\mu_2$

- Now ready for difference in means test
Childhood Immunization Example
Childhood Immunization Example

Hypothesis testing steps:
Hypothesis testing steps:

- Step 1: Null and Alternative Hypotheses
Childhood Immunization Example

Hypothesis testing steps:
- Step 1: Null and Alternative Hypotheses
- Step 2: Collect data (already done)
Childhood Immunization Example

Hypothesis testing steps:
► Step 1: Null and Alternative Hypotheses
► Step 2: Collect data (already done)
► Step 3: Calculate a test statistic
Hypothesis testing steps:

- Step 1: Null and Alternative Hypotheses
- Step 2: Collect data (already done)
- Step 3: Calculate a test statistic
- Step 4: Calculate appropriate $p$-value (given one-tailed or two-tailed test)
Childhood Immunization Example

Hypothesis testing steps:

▶ Step 1: Null and Alternative Hypotheses
▶ Step 2: Collect data (already done)
▶ Step 3: Calculate a test statistic
▶ Step 4: Calculate appropriate $p$-value (given one-tailed or two-tailed test)
▶ Step 5: Decide whether to reject
Childhood Immunization Example

Step 1: Null and Alternative Hypotheses

For $H_0$:

Remember: Null hypothesis is usually no difference/effect, or random chance
Here: No difference in mortality in low vs high vaccinate rate countries

$H_0$: $\mu_1 - \mu_2 = 0$ (or $\mu_1 = \mu_2$)

Could also have $H_0$: $\mu_1 - \mu_2 = a$ (where $a$ is some constant)

For $H_a$:

$H_a$: $\mu_1 - \mu_2 \neq 0$ (or $\mu_1 \neq \mu_2$)
Step 1: Null and Alternative Hypotheses

- For $H_0$: Remember: Null hypothesis is usually no difference/effect, or random chance
  - Here: No difference in mortality in low vs high vaccinate rate countries

- $H_0$: $\mu_1 - \mu_2 = 0$ (or $\mu_1 = \mu_2$)

- Could also have $H_0$: $\mu_1 - \mu_2 = a$ (where $a$ is some constant)

- For $H_a$:
  - $H_a$: $\mu_1 - \mu_2 \neq 0$ (or $\mu_1 \neq \mu_2$)
Childhood Immunization Example

- Step 1: Null and Alternative Hypotheses
  - For $H_0$:
  - $H_0: \mu_1 - \mu_2 = 0$ (or $\mu_1 = \mu_2$)
  - Could also have $H_0: \mu_1 - \mu_2 = a$ (where $a$ is some constant)

- For $H_a$:
  - $H_a: \mu_1 - \mu_2 \neq 0$ (or $\mu_1 \neq \mu_2$)
Childhood Immunization Example

- Step 1: Null and Alternative Hypotheses
- For $H_0$:
  - Remember: Null hypothesis is usually no difference/effect, or random chance
Childhood Immunization Example

- Step 1: Null and Alternative Hypotheses
- For $H_0$:
  - Remember: Null hypothesis is usually no difference/effect, or random chance
  - Here: No difference in mortality in low vs high vaccinate rate countries
Step 1: Null and Alternative Hypotheses

For $H_0$:
- Remember: Null hypothesis is usually no difference/effect, or random chance
- Here: No difference in mortality in low vs high vaccinate rate countries
- $H_0: \mu_1 - \mu_2 = 0$ (or $\mu_1 = \mu_2$)
Step 1: Null and Alternative Hypotheses

For $H_0$:
- Remember: Null hypothesis is usually no difference/effect, or random chance
- Here: No difference in mortality in low vs high vaccinate rate countries
- $H_0: \mu_1 - \mu_2 = 0$ (or $\mu_1 = \mu_2$)
- Could also have $H_0: \mu_1 - \mu_2 = a$ (where $a$ is some constant)
Step 1: Null and Alternative Hypotheses

For $H_0$:
- Remember: Null hypothesis is usually no difference/effect, or random chance
- Here: No difference in mortality in low vs high vaccinate rate countries
- $H_0: \mu_1 - \mu_2 = 0$ (or $\mu_1 = \mu_2$)
- Could also have $H_0: \mu_1 - \mu_2 = a$ (where $a$ is some constant)

For $H_a$:
Childhood Immunization Example

- Step 1: Null and Alternative Hypotheses
- For $H_0$:
  - Remember: Null hypothesis is usually no difference/effect, or random chance
  - Here: No difference in mortality in low vs high vaccinate rate countries
    - $H_0: \mu_1 - \mu_2 = 0$ (or $\mu_1 = \mu_2$)
    - Could also have $H_0: \mu_1 - \mu_2 = a$ (where $a$ is some constant)
- For $H_a$:
  - $H_a: \mu_1 - \mu_2 \neq 0$ (or $\mu_1 \neq \mu_2$)
Childhood Immunization Example

Step 2: Collect sample data (presented in table)

Step 3: Calculate appropriate test statistic

Let's review steps in doing this
Childhood Immunization Example

- Step 2: Collect sample data (presented in table)
Childhood Immunization Example

- Step 2: Collect sample data (presented in table)
- Step 3: Calculate appropriate test statistic
Childhood Immunization Example

- Step 2: Collect sample data (presented in table)
- Step 3: Calculate appropriate test statistic
- Let’s review steps in doing this
Remember the CLT

What does the Central Limit Theorem (CLT) tell us?

1. The sums and means of random samples of observations have an approximately normal distribution.
2. This distribution becomes "more and more" normal the more observations are included in the sum or the mean.

So: $\bar{X}_1 - \bar{X}_2$ (which is a sum) has an approximate Normal distribution centered around the true population difference.

This is true regardless of the distributions that individual observations come from.
Remember the CLT

1. The sums and means of random samples of observations have an approximately normal distribution

True regardless of distributions that individual observations come from
Remember the CLT

- (1) The sums and means of random samples of observations have an approximately normal distribution
- (2) This distribution becomes “more and more” normal the more observations are included in the sum or the mean
Remember the CLT

- (1) The sums and means of random samples of observations have an approximately normal distribution.

- (2) This distribution becomes “more and more” normal the more observations are included in the sum or the mean.

- So: $\bar{X}_1 - \bar{X}_2$ (which is a sum) has an approximate Normal distribution centered around true population difference.
(1) The sums and means of random samples of observations have an approximately normal distribution.

(2) This distribution becomes “more and more” normal the more observations are included in the sum or the mean.

So: $\bar{X}_1 - \bar{X}_2$ (which is a sum) has an approximate Normal distribution centered around true population difference.

True regardless of distributions that individual observations come from.
Remember the CLT

\[ \bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \sigma^2_1/n_1 + \sigma^2_2/n_2) \]

Note: Variances add, since samples are independent

Note: If individual observations in each sample come from an exact Normal distribution, then \( \bar{X}_1 - \bar{X}_2 \) has exact Normal distribution
Remember the CLT

- Under CLT

\[
\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2)
\]

Note: Variances add, since samples are independent

Note: If individual observations in each sample come from an exact Normal distribution, then \(\bar{X}_1 - \bar{X}_2\) has exact Normal distribution
Remember the CLT

Under CLT

\[ \bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}) \]
Remember the CLT

- Under CLT

\[ \bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}) \]

- Note: Variances add, since samples are independent
Remember the CLT

Under CLT

\[ \bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}) \]

Note: Variances add, since samples are independent

Note: If individual observations in each sample come from an exact Normal distribution, then \( \bar{X}_1 - \bar{X}_2 \) has exact Normal distribution
Remember Standardizing

\[
\begin{align*}
\bar{X}_1 - \bar{X}_2 &\sim N(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2) \\
\text{then we can standardize by subtracting mean and dividing by standard error} \\
z &= \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}
\end{align*}
\]
Remember Standardizing

Given

\[ \bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \sigma^2_1 n_1 + \sigma^2_2 n_2) \]

then we can standardize by subtracting mean and dividing by standard error

\[ z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma^2_1 n_1 + \sigma^2_2 n_2}} \]
Given

\[ \bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}) \]
Given

\[ \bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}) \]

then we can standardize by subtracting mean and dividing by standard error.
Given

\[ \bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}) \]

then we can standardize by subtracting mean and dividing by standard error

\[ z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \]
Calculating Test Statistic

Then make key assumption that null is true, so $\mu_1 - \mu_2 = 0$

Substituting this and estimating using samples' standard deviations ($s$) gives us test statistic, $z$:

$$z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where $z$ would be from a standard normal distribution

Note: This is flexible, so can use this to test other differences:

Ex): $H_0: \mu_1 - \mu_2 = 3$

Ex): $H_0: \mu_1 - \mu_2 = 100$
Calculating Test Statistic

- Then make key assumption that null is true, so $\mu_1 - \mu_2 = 0$
Calculating Test Statistic

- Then make key assumption that null is true, so $\mu_1 - \mu_2 = 0$
- Substituting this and estimating using samples’ standard deviations ($s$) gives us test statistic, $z$:
Calculating Test Statistic

Then make key assumption that null is true, so $\mu_1 - \mu_2 = 0$

Substituting this and estimating using samples’ standard deviations ($s$) gives us test statistic, $z$:

$$z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$
Calculating Test Statistic

Then make key assumption that null is true, so $\mu_1 - \mu_2 = 0$

Substituting this and estimating using samples’ standard deviations ($s$) gives us test statistic, $z$:

$$z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where $z$ would be from a standard normal distribution
Calculating Test Statistic

- Then make key assumption that null is true, so $\mu_1 - \mu_2 = 0$
- Substituting this and estimating using samples’ standard deviations ($s$) gives us test statistic, $z$:

$$z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

$$= \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}}$$

- where $z$ would be from a standard normal distribution
- Note: This is flexible, so can use this to test other differences:
Calculating Test Statistic

- Then make key assumption that null is true, so \( \mu_1 - \mu_2 = 0 \)
- Substituting this and estimating using samples’ standard deviations (s) gives us test statistic, \( z \):

\[
z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}
\]

\[
= \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}
\]

- where \( z \) would be from a standard normal distribution
- Note: This is flexible, so can use this to test other differences:

- Ex) \( H_0 : \mu_1 - \mu_2 = 3 \)
Calculating Test Statistic

- Then make key assumption that null is true, so $\mu_1 - \mu_2 = 0$
- Substituting this and estimating using samples’ standard deviations ($s$) gives us test statistic, $z$:

$$
z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}
= \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}
$$

- where $z$ would be from a standard normal distribution
- Note: This is flexible, so can use this to test other differences:
  - Ex) $H_0 : \mu_1 - \mu_2 = 3$
  - Ex) $H_0 : \mu_1 - \mu_2 = 100$
Calculating Test Statistic

As before, we (1) may not have large enough sample size and (2) estimate using sample standard deviation.

→ Use Student's $t$ distribution instead of normal to be conservative.

Gives us $t$ statistic:

$$t_{df} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

Formula for $df$ more complicated than for a single mean.

Good approximation is smaller of $(n_1 - 1)$ and $(n_2 - 1)$.

More exact calculation for $df$ used by software packages.
Calculating Test Statistic

- As before, we (1) may not have large enough sample size and (2) estimate using sample standard deviation
Calculating Test Statistic

- As before, we (1) may not have large enough sample size and (2) estimate using sample standard deviation
- Use Student’s $t$ distribution instead of normal to be conservative

**Formula for $t$ statistic:**

$$t_{df} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

**Formula for $df$:**

More complicated than for a single mean

Good approximation is smaller of $(n_1 - 1)$ and $(n_2 - 1)$

More exact calculation for $df$ used by software packages
Calculating Test Statistic

- As before, we (1) may not have large enough sample size and (2) estimate using sample standard deviation
- → Use Student’s $t$ distribution instead of normal to be conservative
- Gives us $t$ statistic:

\[
t_{df} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}
\]

- Formula for $df$ more complicated than for a single mean
- Good approximation is smaller of $(n_1 - 1)\text{ and } (n_2 - 1)$
- More exact calculation for $df$ used by software packages
Calculating Test Statistic

- As before, we (1) may not have large enough sample size and (2) estimate using sample standard deviation
- Use Student’s $t$ distribution instead of normal to be conservative
- Gives us $t$ statistic:

$$t_{df} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$
Calculating Test Statistic

- As before, we (1) may not have large enough sample size and (2) estimate using sample standard deviation
- → Use Student’s $t$ distribution instead of normal to be conservative
- Gives us $t$ statistic:

$$t_{df} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

- Formula for $df$ more complicated than for a single mean
Calculating Test Statistic

- As before, we (1) may not have large enough sample size and (2) estimate using sample standard deviation
- → Use Student’s $t$ distribution instead of normal to be conservative
- Gives us $t$ statistic:

$$t_{df} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- Formula for $df$ more complicated than for a single mean
- Good approximation is *smaller* of $(n_1 - 1)$ and $(n_2 - 1)$
Calculating Test Statistic

- As before, we (1) may not have large enough sample size and (2) estimate using sample standard deviation
- Use Student’s $t$ distribution instead of normal to be conservative
- Gives us $t$ statistic:

$$t_{df} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- Formula for $df$ more complicated than for a single mean
- Good approximation is smaller of $(n_1 - 1)$ and $(n_2 - 1)$
- More exact calculation for $df$ used by software packages
Calculating Test Statistic

\[ t_{df} = \bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2) \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \]

\[ t_{90} = 92.42 - 24.98 \sqrt{\frac{73.21}{91} + \frac{30.92}{97}} = 8.13 \]
Calculating Test Statistic

- For our example:
Calculating Test Statistic

For our example:

\[
t_{df} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}
\]

\[
t_{90} = \frac{92.42 - 24.98}{\sqrt{\frac{73.21^2}{91} + \frac{30.92^2}{97}}}
\]

= 8.13
Childhood Immunization Example

Step 4: Calculate p-value

Again, consider the two-tailed versus one-tailed test

Here, $p$-value = 0.0000 for two-tailed test

Step 5: Decide whether or not to reject the null hypothesis and interpret results

What would you do here?

What is the substantive interpretation?
Childhood Immunization Example

- Step 4: Calculate p-value

Again, consider the two-tailed versus one-tailed test. Here, $p$-value = 0.0000 for two-tailed test.

Step 5: Decide whether or not to reject the null hypothesis and interpret results

What would you do here?

What is the substantive interpretation?
Childhood Immunization Example

- Step 4: Calculate p-value
- Again, consider the two-tailed versus one-tailed test
Step 4: Calculate p-value
Again, consider the two-tailed versus one-tailed test
    Here, $p$-value = 0.0000 for two-tailed test
Childhood Immunization Example

- Step 4: Calculate p-value
- Again, consider the two-tailed versus one-tailed test
  - Here, $p$-value = 0.0000 for two-tailed test
- Step 5: Decide whether or not to reject the null hypothesis and interpret results
Step 4: Calculate p-value

Again, consider the two-tailed versus one-tailed test

Here, \( p \)-value = 0.0000 for two-tailed test

Step 5: Decide whether or not to reject the null hypothesis and interpret results

What would you do here?
Childhood Immunization Example

- Step 4: Calculate p-value
- Again, consider the two-tailed versus one-tailed test
  - Here, $p$-value = 0.0000 for two-tailed test
- Step 5: Decide whether or not to reject the null hypothesis and interpret results
  - What would you do here?
  - What is the substantive interpretation?
Cases Where 2 Populations have Equal Variance

- Instances where groups are independent (no “pairs” in both groups), but observations come from same underlying distribution
- Most plausible with a randomized controlled trial
- Ex) Use a coin flip to assign subjects to treatment, control conditions
  → underlying standard deviation should be same in both groups
- In these cases: If we can assume $\sigma_1^2 = \sigma_2^2$ then estimation easier
  → pooled standard error can be used
- If it's not clear, we will tell you on problem sets and exams
- Rule of thumb: if ratio of larger to smaller standard deviation is less than 2, equal variance assumption reasonable
Cases Where 2 Populations have Equal Variance

Note: Instances where groups are independent (no “pairs” in both groups), but observations come from same underlying distribution

If it's not clear, we will tell you on problem sets and exams

Rule of thumb: if ratio of larger to smaller standard deviation is less than 2, equal variance assumption reasonable
Cases Where 2 Populations have Equal Variance

▶ Note: Instances where groups are independent (no “pairs” in both groups), but observations come from same underlying distribution
  ▶ Most plausible with a randomized controlled trial

▶ If it's not clear, we will tell you on problem sets and exams
▶ Rule of thumb: if ratio of larger to smaller standard deviation is less than 2, equal variance assumption reasonable
Cases Where 2 Populations have Equal Variance

- Note: Instances where groups are independent (no “pairs” in both groups), but observations come from same underlying distribution
  - Most plausible with a randomized controlled trial
  - Ex) Use a coin flip to assign subjects to treatment, control conditions → underlying standard deviation should be same in both groups

- If it's not clear, we will tell you on problem sets and exams
- Rule of thumb: if ratio of larger to smaller standard deviation is less than 2, equal variance assumption reasonable
Cases Where 2 Populations have Equal Variance

- Note: Instances where groups are independent (no “pairs” in both groups), but observations come from same underlying distribution
  - Most plausible with a randomized controlled trial
  - Ex) Use a coin flip to assign subjects to treatment, control conditions → underlying standard deviation should be same in both groups
- In these cases: If we can assume $\sigma_1^2 = \sigma_2^2$ then estimation easier → pooled standard error can be used
Cases Where 2 Populations have Equal Variance

- Note: Instances where groups are independent (no “pairs” in both groups), but observations come from same underlying distribution
  - Most plausible with a randomized controlled trial
  - Ex) Use a coin flip to assign subjects to treatment, control conditions → underlying standard deviation should be same in both groups
- In these cases: If we can assume $\sigma_1^2 = \sigma_2^2$ then estimation easier → pooled standard error can be used
- If it’s not clear, we will tell you on problem sets and exams
Cases Where 2 Populations have Equal Variance

- Note: Instances where groups are independent (no “pairs” in both groups), but observations come from same underlying distribution
  - Most plausible with a randomized controlled trial
  - Ex) Use a coin flip to assign subjects to treatment, control conditions → underlying standard deviation should be same in both groups

- In these cases: If we can assume $\sigma_1^2 = \sigma_2^2$ then estimation easier → pooled standard error can be used

- If it’s not clear, we will tell you on problem sets and exams

- Rule of thumb: if ratio of larger to smaller standard deviation is less than 2, equal variance assumption reasonable
Cases Where 2 Populations have Equal Variance

Pooled estimator of standard deviation:

\[ s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \]

(weighted average of the two standard deviations, with \( n_1 - 1 \) correction)

Test statistic becomes:

\[ t_{df} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \]

with \( df \) equal to \( n_1 + n_2 - 2 \).
Cases Where 2 Populations have Equal Variance

- Pooled estimator of standard deviation:

\[
\hat{s}_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}
\]

(weighted average of the two standard deviations, w/ \(n-1\) correction)

- Test statistic becomes:

\[
t_{df} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\hat{s}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}
\]

with \(df\) equal to \(n_1 + n_2 - 2\)
Cases Where 2 Populations have Equal Variance

- Pooled estimator of standard deviation:

\[ s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \]
Cases Where 2 Populations have Equal Variance

- Pooled estimator of standard deviation:

\[ s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \]

- (weighted average of the two standard deviations, w/ \(n - 1\) correction)
Cases Where 2 Populations have Equal Variance

- Pooled estimator of standard deviation:

\[ s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \]

- (weighted average of the two standard deviations, w/n \( - 1 \) correction)

- Test statistic becomes:
Cases Where 2 Populations have Equal Variance

- Pooled estimator of standard deviation:

\[ s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \]

- (weighted average of the two standard deviations, w/ \(n - 1\) correction)

- Test statistic becomes:

\[ t_{df} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \]
Cases Where 2 Populations have Equal Variance

- Pooled estimator of standard deviation:

\[ s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \]

- (weighted average of the two standard deviations, w/\( n - 1 \) correction)

- Test statistic becomes:

\[ t_{df} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \]

- with \( df \) equal to \( n_1 + n_2 - 2 \)
Childhood Immunization Example

Assume equal variance in immunizing/not immunizing countries (good assumption?)

\[ s_p = \sqrt{\left(\frac{n_1 - 1}{s_1^2}\right) + \left(\frac{n_2 - 1}{s_2^2}\right)} \]

\[ = \sqrt{\left(\frac{90}{73.21^2}\right) + \left(\frac{96}{30.92^2}\right)} \]

\[ = \sqrt{\left(\frac{91}{97}\right) - 2} = 55.559 \]

And test statistic becomes:

\[ t_{df} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \]

\[ t_{186} = 92.42 - 24.98 \]

\[ 55.559 \]

\[ = 8.317 \]

How would this compare to previous test statistic (8.13)?
Childhood Immunization Example

- Assume equal variance in immunizing/not immunizing countries (good assumption?)

\[
\text{Assume equal variance in immunizing/not immunizing countries (good assumption?)}
\]

\[
\begin{align*}
\sigma^2 &= \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}} \\
&= \sqrt{\frac{(90-1)73.21^2 + (96-1)30.92^2}{91 + 97 - 2}} \\
&= \sqrt{\frac{89 	imes 73.21^2 + 95 	imes 30.92^2}{186}} \\
&= 55.559
\end{align*}
\]

\[
\text{and test statistic becomes:}
\]

\[
t_{df} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}
\]

\[
t_{186} = \frac{92.42 - 24.98}{55.559 \sqrt{\frac{1}{91} + \frac{1}{97}}} = 8.317
\]

\[
\text{How would this compare to previous test statistic (8.13)?}
\]
Childhood Immunization Example

- Assume equal variance in immunizing/not immunizing countries (good assumption?)

\[
s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}
\]

\[
= \sqrt{\frac{(90)73.21^2 + (96)30.92^2}{91 + 97 - 2}} = 55.559
\]
Childhood Immunization Example

- Assume equal variance in immunizing/not immunizing countries (good assumption?)

\[
s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}
\]

\[
= \sqrt{\frac{(90)73.21^2 + (96)30.92^2}{91 + 97 - 2}} = 55.559
\]

- and test statistic becomes:
Childhood Immunization Example

- Assume equal variance in immunizing/not immunizing countries (good assumption?)

\[
s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}
\]

\[
= \sqrt{\frac{(90)73.21^2 + (96)30.92^2}{91 + 97 - 2}} = 55.559
\]

- and test statistic becomes:

\[
t_{df} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}
\]

\[
t_{186} = \frac{92.42 - 24.98}{55.559 \sqrt{\frac{1}{91} + \frac{1}{97}}} = 8.317
\]

- How would this compare to previous test statistic (8.13)?
To Summarize

Hypothesis tests for comparing two means:

▶ If population standard deviations ($\sigma_1$ and $\sigma_2$) known and unequal and if fairly large $n \rightarrow \text{z-test}$

▶ If $\sigma_1$ and $\sigma_2$ unknown and unequal $\rightarrow \text{t-test}$ (also called Welch's $\text{t-test}$)

▶ If $\sigma_1$ and $\sigma_2$ unknown but can be assumed equal $\rightarrow \text{t-test}$ w/ pooled standard errors

The problem should be clear on whether you can assume equal variance
To Summarize

Hypothesis tests for comparing two means:

- If population standard deviations ($\sigma_1$ and $\sigma_2$) known and unequal and if fairly large $n \rightarrow z$-test
- If $\sigma_1$ and $\sigma_2$ unknown and unequal $\rightarrow t$-test (also called Welch's $t$-test)
- If $\sigma_1$ and $\sigma_2$ unknown but can be assumed equal $\rightarrow t$-test w/ pooled standard errors

The problem should be clear on whether you can assume equal variance
To Summarize

Hypothesis tests for comparing two means:

- If population standard deviations ($\sigma_1$ and $\sigma_2$) known and unequal and if fairly large $n \rightarrow z$-test

- If $\sigma_1$ and $\sigma_2$ unknown and unequal $\rightarrow t$-test (also called Welch's $t$-test)

- If $\sigma_1$ and $\sigma_2$ unknown but can be assumed equal $\rightarrow t$-test w/ pooled standard errors

The problem should be clear on whether you can assume equal variance
To Summarize

Hypothesis tests for comparing two means:

- If population standard deviations ($\sigma_1$ and $\sigma_2$) known and unequal and if fairly large $n \rightarrow z$-test
- If $\sigma_1$ and $\sigma_2$ unknown and unequal $\rightarrow t$-test (also called Welch’s $t$-test)
To Summarize

Hypothesis tests for comparing two means:

- If population standard deviations ($\sigma_1$ and $\sigma_2$) known and unequal and if fairly large $n \to z$-test
- If $\sigma_1$ and $\sigma_2$ unknown and unequal $\to t$-test (also called Welch’s $t$-test)
- If $\sigma_1$ and $\sigma_2$ unknown but can be assumed equal $\to t$-test w/ pooled standard errors
To Summarize

Hypothesis tests for comparing two means:

- If population standard deviations (σ₁ and σ₂) known and unequal and if fairly large $n \rightarrow z$-test
- If σ₁ and σ₂ unknown and unequal $\rightarrow t$-test (also called Welch’s $t$-test)
- If σ₁ and σ₂ unknown but can be assumed equal $\rightarrow t$-test w/ pooled standard errors

$\rightarrow$ The problem should be clear on whether you can assume equal variance
Potential issues with Hypothesis Testing

1) Hypothesis Tests test the null hypothesis, not the alternate hypothesis
2) Cannot guarantee full certainty regarding whether null hypothesis is 100% false
3) Will not guarantee practical significance
Potential issues with Hypothesis Testing

Hypothesis tests are not perfect!
Potential issues with Hypothesis Testing

Hypothesis tests are not perfect!

1) Hypothesis Tests test the null hypothesis, not the alternate hypothesis
Potential issues with Hypothesis Testing

Hypothesis tests are not perfect!

1) Hypothesis Tests test the null hypothesis, not the alternate hypothesis
2) Cannot guarantee full certainty regarding whether null hypothesis is 100% false
Hypothesis tests are not perfect!

1) Hypothesis Tests test the null hypothesis, not the alternate hypothesis
2) Cannot guarantee full certainty regarding whether null hypothesis is 100% false
3) Will not guarantee practical significance
Potential issues with Hypothesis Testing

Hypothesis tests are not perfect!

1) Hypothesis Tests test the null hypothesis, not the alternate hypothesis

2) Cannot guarantee full certainty regarding whether null hypothesis is 100% false

3) Will not guarantee practical significance
1) Hypotheses Tests test the null hypothesis
1) Hypotheses Tests test the null hypothesis

- Interpret as to whether you do or do not reject the null, NOT whether you accept the alternative

- $p$-value is the probability of obtaining results as "extreme" or more given that the null hypothesis is true.
1) Hypotheses Tests test the null hypothesis

- Interpret as to whether you do or do not reject the null, NOT whether you accept the alternative
- (A bit like proof by contradiction)
1) Hypotheses Tests test the null hypothesis

- Interpret as to whether you do or do not reject the null, NOT whether you accept the alternative
- (A bit like proof by contradiction)
- True/False?
1) Hypotheses Tests test the null hypothesis

- Interpret as to whether you do or do not reject the null, NOT whether you accept the alternative
- (A bit like proof by contradiction)
- True/False?
  - The statement “p-value is .003” is equivalent to the statement “there is a 0.3% probability that null hypothesis is true”
1) Hypotheses Tests test the null hypothesis

- Interpret as to whether you do or do not reject the null, NOT whether you accept the alternative
- (A bit like proof by contradiction)
- True/False?
  - The statement “p-value is .003” is equivalent to the statement “there is a 0.3% probability that null hypothesis is true”
  - $p$-value is probability of obtaining results as ”extreme” or more given that null hypothesis is true
1) Hypotheses Tests test the null hypothesis

- Interpret as to whether you do or do not reject the null, NOT whether you accept the alternative
- (A bit like proof by contradiction)
- True/False?
  - The statement “p-value is .003” is equivalent to the statement “there is a 0.3% probability that null hypothesis is true”
  - $p$-value is probability of obtaining results as ”extreme” or more given that null hypothesis is true
  - If you reject the null hypothesis, it must mean the alternate hypothesis must be true
2) No full certainty about correctly rejecting null hypothesis tests given us $p$-values, or how extreme our test statistic is given null being true. However: Do not allow us to rule out null hypothesis with 100% certainty. Specifically, we're concerned about two scenarios:

1. We reject the null hypothesis, even though the null is true
2. We fail to reject the null hypothesis, even though the null is false
2) No full certainty about correctly rejecting null hypothesis tests given us p-values, or how extreme our test statistic is given null being true.
2) No full certainty about correctly rejecting null

- Hypothesis tests given us $p$-values, or how extreme our test statistic is given null being true
- However: Do not allow us to rule out null hypothesis w/ 100% certainty
2) No full certainty about correctly rejecting null

- Hypothesis tests given us $p$-values, or how extreme our test statistic is given null being true
- However: Do not allow us to rule out null hypothesis w/ 100% certainty
- Specifically, we’re concerned about two scenarios:
2) No full certainty about correctly rejecting null

- Hypothesis tests given us $p$-values, or how extreme our test statistic is given null being true
- However: Do not allow us to rule out null hypothesis w/ 100% certainty
- Specifically, we’re concerned about two scenarios:
  1. We reject the null hypothesis, even though the null is true
2) No full certainty about correctly rejecting null

- Hypothesis tests given us $p$-values, or how extreme our test statistic is given null being true
- However: Do not allow us to rule out null hypothesis w/ 100% certainty
- Specifically, we’re concerned about two scenarios:
  1. We reject the null hypothesis, even though the null is true
  2. We fail to reject the null hypothesis, even though the null is false
2) No full certainty about correctly rejecting null

- Hypothesis tests given us $p$-values, or how extreme our test statistic is given null being true
- However: Do not allow us to rule out null hypothesis w/ 100% certainty
- Specifically, we’re concerned about two scenarios:
  1. We reject the null hypothesis, even though the null is true
  2. We fail to reject the null hypothesis, even though the null is false
- (Both conditional statements)
Type I versus Type II Errors

$H_0$ is true
$H_0$ is not true

Reject $H_0$
Do not reject $H_0$
Type I versus Type II Errors

<table>
<thead>
<tr>
<th></th>
<th>$H_0$ is true</th>
<th>$H_0$ is not true</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject $H_0$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do not reject $H_0$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Type I versus Type II Errors

<table>
<thead>
<tr>
<th></th>
<th>$H_0$ is true</th>
<th>$H_0$ is not true</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject $H_0$</td>
<td></td>
<td>Good</td>
</tr>
<tr>
<td>Do not reject $H_0$</td>
<td>Good</td>
<td></td>
</tr>
</tbody>
</table>
Type I versus Type II Errors

<table>
<thead>
<tr>
<th></th>
<th>$H_0$ is true</th>
<th>$H_0$ is not true</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject $H_0$</td>
<td>Bad!</td>
<td>Good</td>
</tr>
<tr>
<td>Do not reject $H_0$</td>
<td>Good</td>
<td>Bad!</td>
</tr>
</tbody>
</table>
### Type I versus Type II Errors

<table>
<thead>
<tr>
<th></th>
<th>( H_0 ) is true</th>
<th>( H_0 ) is not true</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject ( H_0 )</td>
<td>Type 1</td>
<td>Good</td>
</tr>
<tr>
<td>Do not reject ( H_0 )</td>
<td>Good</td>
<td>Type 2</td>
</tr>
</tbody>
</table>
Type I Error

Type I: Null hypothesis rejected when in fact it is true

Akin to false negative

Mammogram analogy: Negative test, despite having the disease

\[ P(\text{Type I error}) = P(\text{Rejecting } H_0 | H_0 \text{ true}) = \alpha \]

\( \alpha \): Also referred to as level of significance or the critical value

Very common to set \( \alpha = 0.05, \alpha = 0.10, \text{ or } \alpha = 0.01 \)

Question: Why would we want to avoid Type I Error?

Question: How do you interpret a level of significance of 0.05?
Type I Error

- Type I: Null hypothesis rejected when in fact it is true

Akin to false negative
Mammogram analogy: Negative test, despite having the disease

\[ P(\text{Type I error}) = P(\text{Rejecting } H_0 | H_0 \text{ true}) = \alpha \]

\( \alpha \): Also referred to as level of significance or the critical value

Very common to set \( \alpha = 0.05 \), \( \alpha = 0.10 \), or \( \alpha = 0.01 \)

Question: Why would we want to avoid Type I Error?

Question: How do you interpret a level of significance of 0.05?
Type I Error

- Type I: Null hypothesis rejected when in fact it is true
- Akin to false negative

$P(\text{Type I error}) = P(\text{Rejecting } H_0 | H_0 \text{ true}) = \alpha$

$\alpha$: Also referred to as level of significance or the critical value

Very common to set $\alpha = 0.05, \alpha = 0.10, \text{ or } \alpha = 0.01$

Question: Why would we want to avoid Type I Error?

Question: How do you interpret a level of significance of 0.05?
Type I Error

- **Type I**: Null hypothesis rejected when in fact it is true
- Akin to false negative
  - Mammogram analogy: Negative test, despite having the disease

\[
P(\text{Type I error}) = P(\text{Rejecting } H_0 \mid H_0 \text{ true}) = \alpha
\]

- \(\alpha\): Also referred to as level of significance or the critical value
- Very common to set \(\alpha = 0.05\), \(\alpha = 0.10\), or \(\alpha = 0.01\)

Question: Why would we want to avoid Type I Error?

Question: How do you interpret a level of significance of 0.05?
Type I Error

- **Type I**: Null hypothesis rejected when in fact it is true
- Akin to false negative
  - Mammogram analogy: Negative test, despite having the disease
- \( P(\text{Type I error}) = P(\text{Rejecting } H_0|H_0 \text{ true}) = \alpha \)
- \( \alpha \): Also referred to as level of significance or the critical value
- Very common to set \( \alpha = 0.05, \alpha = 0.10, \text{ or } \alpha = 0.01 \)
- Question: Why would we want to avoid Type I Error?
- Question: How do you interpret a level of significance of 0.05?
Type I Error

- **Type I**: Null hypothesis rejected when in fact it is true
- Akin to false negative
  - Mammogram analogy: Negative test, despite having the disease
- \( P(\text{Type I error}) = P(\text{Rejecting } H_0|H_0 \text{ true}) = \alpha \)
- \( \alpha \): Also referred to as level of significance or the critical value
Type I Error

- Type I: Null hypothesis rejected when in fact it is true
- Akin to false negative
  - Mammogram analogy: Negative test, despite having the disease
- \( P(\text{Type I error}) = P(\text{Rejecting } H_0 | H_0 \text{ true}) = \alpha \)
- \( \alpha \): Also referred to as level of significance or the critical value
- Very common to set \( \alpha = 0.05, \alpha = 0.10, \text{ or } \alpha = 0.01 \)
Type I Error

- **Type I**: Null hypothesis rejected when in fact it is true
- Akin to false negative
  - Mammogram analogy: Negative test, despite having the disease
- \[ P(\text{Type I error}) = P(\text{Rejecting } H_0|H_0 \text{ true}) = \alpha \]
- \( \alpha \): Also referred to as level of significance or the critical value
- Very common to set \( \alpha = 0.05 \), \( \alpha = 0.10 \), or \( \alpha = 0.01 \)
- Question: Why would we want to avoid Type I Error?
Type I Error

- **Type I**: Null hypothesis rejected when in fact it is true
- Akin to false negative
  - Mammogram analogy: Negative test, despite having the disease
- $P(\text{Type I error}) = P(\text{Rejecting } H_0|H_0 \text{ true}) = \alpha$
- $\alpha$: Also referred to as level of significance or the critical value
- Very common to set $\alpha = 0.05$, $\alpha = 0.10$, or $\alpha = 0.01$
- Question: Why would we want to avoid Type I Error?
- Question: How do you interpret a level of significance of 0.05?
Type II Error

- Type II: Null hypothesis is not rejected when in fact it is false.
  - Akin to false positive.
  - Mammogram analogy: Positive test, despite not having the disease.

\[
P(\text{Type II error}) = P(\text{Not rejecting } H_0 | H_0 \text{ false})
\]

- Often set at \( P(\text{Type II error}) = 0.20 = \beta \)
- \( 1 - \beta = P(\text{Rejecting } H_0 | H_0 \text{ false}) \)
  - Known as the power of a test (more later).

- Sometimes considered less important.
- However: consider your specific problem
  - In some instances, either error may be more costly.

Vaccine example: which error concerns you the most?
Type II Error

- **Type II**: Null hypothesis is not rejected when in fact it is false
Type II Error

- **Type II**: Null hypothesis is not rejected when in fact it is false
- Akin to false positive
Type II Error

- **Type II**: Null hypothesis is not rejected when in fact it is false
- Akin to false positive
  - Mammogram analogy: Positive test, despite not having the disease

\[
P(\text{Type II error}) = P(\text{Not rejecting } H_0 | H_0 \text{ false})
\]

Often set at \( P(\text{Type II error}) = 0.20 = \beta \)

\[1 - \beta = P(\text{Rejecting } H_0 | H_0 \text{ false})\]

Known as the power of a test (more later)

Sometimes considered less important

However: consider your specific problem → in some instances, either error may be more costly

Vaccine example: which error concerns you the most?
Type II Error

- **Type II:** Null hypothesis is not rejected when in fact it is false
- Akin to false positive
  - Mammogram analogy: Positive test, despite not having the disease
- \( P(\text{Type II error}) = P(\text{Not rejecting } H_0|H_0 \text{ false}) \)
Type II Error

- **Type II**: Null hypothesis is not rejected when in fact it is false
- Akin to false positive
  - Mammogram analogy: Positive test, despite not having the disease
- \[ P(\text{Type II error}) = P(\text{Not rejecting } H_0|H_0 \text{ false}) \]
- Often set at \( P(\text{Type II error}) = 0.20 = \beta \)
Type II Error

- **Type II**: Null hypothesis is not rejected when in fact it is false
- Akin to false positive
  - Mammogram analogy: Positive test, despite not having the disease
- $P(\text{Type II error}) = P(\text{Not rejecting } H_0|H_0 \text{ false})$
- Often set at $P(\text{Type II error}) = 0.20 = \beta$
- $1 - \beta = P(\text{Rejecting } H_0|H_0 \text{ false})$
Type II Error

- **Type II**: Null hypothesis is not rejected when in fact it is false
- Akin to false positive
  - Mammogram analogy: Positive test, despite not having the disease
- \( P(\text{Type II error}) = P(\text{Not rejecting } H_0|H_0 \text{ false}) \)
- Often set at \( P(\text{Type II error}) = 0.20 = \beta \)
- \( 1 - \beta = P(\text{Rejecting } H_0|H_0 \text{ false}) \)
  - Known as the **power** of a test (more later)
Type II Error

- **Type II**: Null hypothesis is not rejected when in fact it is false
- Akin to false positive
  - Mammogram analogy: Positive test, despite not having the disease

\[
P(\text{Type II error}) = P(\text{Not rejecting } H_0 | H_0 \text{ false})
\]

- Often set at \( P(\text{Type II error}) = 0.20 = \beta \)
- \( 1 - \beta = P(\text{Rejecting } H_0 | H_0 \text{ false}) \)
  - Known as the power of a test (more later)
- Sometimes considered less important
Type II Error

- **Type II**: Null hypothesis is not rejected when in fact it is false
- Akin to false positive
  - Mammogram analogy: Positive test, despite not having the disease
- \( P(\text{Type II error}) = P(\text{Not rejecting } H_0|H_0 \text{ false}) \)
- Often set at \( P(\text{Type II error}) = 0.20 = \beta \)
- \( 1 - \beta = P(\text{Rejecting } H_0|H_0 \text{ false}) \)
  - Known as the power of a test (more later)
- Sometimes considered less important
- However: consider your specific problem → in some instances, either error may be more costly
**Type II Error**

- **Type II**: Null hypothesis is not rejected when in fact it is false
- Akin to false positive
  - Mammogram analogy: Positive test, despite not having the disease
- 
  \[ P(\text{Type II error}) = P(\text{Not rejecting } H_0|H_0 \text{ false}) \]

- Often set at \( P(\text{Type II error}) = 0.20 = \beta \)
- \( 1 - \beta = P(\text{Rejecting } H_0|H_0 \text{ false}) \)
  - Known as the **power** of a test (more later)

- Sometimes considered less important
- However: consider your specific problem → in some instances, either error may be more costly
- Vaccine example: which error concerns you the most?
3) Practice versus Statistical Significance

- A very small p-value provides very strong evidence to reject the null hypothesis.
- But: statistical significance $\neq$ practical significance.
3) Practice versus Statistical Significance

- A very small $p$-value provides very strong evidence to reject null hypothesis
3) Practice versus Statistical Significance

- A very small $p$-value provides very strong evidence to reject null hypothesis
- But: statistical significance $\neq$ practical significance
3) Practice versus Statistical Significance

Example:

$H_0$: $\mu_1 - \mu_2 = 0$

$H_a$: $\mu_1 - \mu_2 \neq 0$

Sample data provides a difference of 0.02 and $p$-value < 0.00001

Although $p$-value small, substantive difference between two means is estimated as only 0.02, or 2%

Such a small difference may not turn out to be substantively important – e.g., weight, cents, income

Often occurs with very large samples

Why? Easier to get small $p$-value w/ large $n$

→ Important to consider practical significance when sample size large (“Big Data”)
3) Practice versus Statistical Significance

- Example:

\[ H_0: \mu_1 - \mu_2 = 0 \]

\[ H_a: \mu_1 - \mu_2 \neq 0 \]

Sample data provides a difference of 0.02 and \( p \)-value < 0.00001

Although \( p \)-value small, substantive difference between two means is estimated as only 0.02, or 2%

Such a small difference may not turn out to be substantively important – e.g., weight, cents, income

Often occurs with very large samples

→ Important to consider practical significance when sample size large (“Big Data”)
3) Practice versus Statistical Significance

- Example:

- $H_0 : \mu_1 - \mu_2 = 0$
3) Practice versus Statistical Significance

- Example:
  - $H_0 : \mu_1 - \mu_2 = 0$
  - $H_a : \mu_1 - \mu_2 \neq 0$

- Sample data provides a difference of 0.02 and $p$-value < 0.00001

- Although $p$-value small, substantive difference between two means is estimated as only 0.02, or 2%

- Such a small difference may not turn out to be substantively important – e.g., weight, cents, income

- Often occurs with very large samples

- Why? Easier to get small $p$-value with large $n$

- → Important to consider practical significance when sample size large ("Big Data")
3) Practice versus Statistical Significance

- Example:
  - $H_0 : \mu_1 - \mu_2 = 0$
  - $H_a : \mu_1 - \mu_2 \neq 0$
  - Sample data provides a difference of 0.02 and $p$-value $< 0.00001$

- Although $p$-value small, substantive difference between two means is estimated as only 0.02, or 2%
- Such a small difference may not turn out to be substantively important – e.g., weight, cents, income
- Often occurs with very large samples
- Why? Easier to get small $p$-value with large $n$
- → Important to consider practical significance when sample size large ("Big Data")
3) Practice versus Statistical Significance

- Example:
  - $H_0: \mu_1 - \mu_2 = 0$
  - $H_a: \mu_1 - \mu_2 \neq 0$
  - Sample data provides a difference of 0.02 and $p$-value < 0.00001
  - Although $p$-value small, substantive difference between two means is estimated as only 0.02, or 2%

- Often occurs with very large samples

- Why? Easier to get small $p$-value with large $n$

- → Important to consider practical significance when sample size large ("Big Data")
3) Practice versus Statistical Significance

- Example:
  - $H_0: \mu_1 - \mu_2 = 0$
  - $H_a: \mu_1 - \mu_2 \neq 0$

  - Sample data provides a difference of 0.02 and $p$-value $< 0.00001$
  - Although $p$-value small, substantive difference between two means is estimated as only 0.02, or 2%
  - Such a small difference may not turn out to be substantively important – e.g., weight, cents, income

- Often occurs with very large samples

- Why? Easier to get small $p$-value w/ large $n$

- → Important to consider practical significance when sample size large (“Big Data”)
3) Practice versus Statistical Significance

- Example:
  - $H_0 : \mu_1 - \mu_2 = 0$
  - $H_a : \mu_1 - \mu_2 \neq 0$
  - Sample data provides a difference of 0.02 and $p$-value $< 0.00001$
  - Although $p$-value small, substantive difference between two means is estimated as only 0.02, or 2%
  - Such a small difference may not turn out to be substantively important – e.g., weight, cents, income
  - Often occurs with very large samples

- Why? Easier to get small $p$-value with large $n$
3) Practice versus Statistical Significance

- Example:
  - $H_0 : \mu_1 - \mu_2 = 0$
  - $H_a : \mu_1 - \mu_2 \neq 0$
  - Sample data provides a difference of 0.02 and $p$-value < 0.00001
  - Although $p$-value small, substantive difference between two means is estimated as only 0.02, or 2%
  - Such a small difference may not turn out to be substantively important – e.g., weight, cents, income
  - Often occurs with very large samples
  - Why? Easier to get small $p$-value w/ large $n$
3) Practice versus Statistical Significance

- Example:
  - \( H_0 : \mu_1 - \mu_2 = 0 \)
  - \( H_a : \mu_1 - \mu_2 \neq 0 \)
  - Sample data provides a difference of 0.02 and \( p \)-value < 0.00001
  - Although \( p \)-value small, substantive difference between two means is estimated as only 0.02, or 2%
  - Such a small difference may not turn out to be substantively important – e.g., weight, cents, income
  - Often occurs with very large samples
  - Why? Easier to get small \( p \)-value w/ large \( n \)
  - \( \rightarrow \) Important to consider practical significance when sample size large ("Big Data")
Next Time

- Hypothesis testing for proportions
- Introducing confidence intervals