Lecture 17:
One Way ANOVA
API-201Z

Maya Sen

Harvard Kennedy School
http://scholar.harvard.edu/msen
Announcements

▶ Midterm #2 one week from Thursday

▶ Review session next Tuesday afternoon will be taped

▶ Because of Veteran's Day, shifting my OH to Tuesday noon to 2pm (Taubman 356)

▶ Have posted readings for Thursday – Oregon health care case study
Announcements

- Midterm #2 one week from Thursday
Announcements

- Midterm #2 one week from Thursday
- Review session next Tuesday afternoon will be taped
Announcements

- Midterm #2 one week from Thursday
- Review session next Tuesday afternoon will be taped
- Because of Veteran’s Day, shifting my OH to Tuesday noon to 2pm (Taubman 356)
Announcements

- Midterm #2 one week from Thursday
- Review session next Tuesday afternoon will be taped
- Because of Veteran’s Day, shifting my OH to Tuesday noon to 2pm (Taubman 356)
- Have posted readings for Thursday – Oregon health care case study
Roadmap

- Finish up paired tests
- One-Way Analysis of Variance (ANOVA)
- Multiple comparisons and Bonferroni corrections
- Multiple comparisons corrections will be last topic covered on Midterm #2
- Leaves one common type of test (Chi Square tests) for final, along with regression
Roadmap

- Finish up paired tests
Roadmap

- Finish up paired tests
- One-Way Analysis of Variance (ANOVA)

Multiple comparisons and Bonferroni corrections will be last topic covered on Midterm #2
Leaves one common type of test (Chi Square tests) for final, along with regression
Roadmap

- Finish up paired tests
- One-Way Analysis of Variance (ANOVA)
- Multiple comparisons and Bonferroni corrections
Roadmap

- Finish up paired tests
- One-Way Analysis of Variance (ANOVA)
- Multiple comparisons and Bonferroni corrections
- Multiple comparisons corrections will be last topic covered on Midterm #2
Roadmap

- Finish up paired tests
- One-Way Analysis of Variance (ANOVA)
- Multiple comparisons and Bonferroni corrections
- Multiple comparisons corrections will be last topic covered on Midterm #2
- Leaves one common type of test (Chi Square tests) for final, along with regression
Paired Tests for Proportions

For paired data, we have to take into account the fact that we have dependence between groups.

For sample means, it's straightforward. Take the difference between the groups as a new quantity, use that to re-calculate the standard deviation, and conduct a hypothesis test.

However, for proportions we sometimes don't have the entire table of individual observations.

Usually only have a contingency table.

So we estimate the standard error slightly differently.
Paired Tests for Proportions

- For paired data, we have to take into account the fact that we have dependence between groups.
Paired Tests for Proportions

- For paired data, we have to take into account the fact that we have dependence between groups.
- For sample means, straightforward → take difference between groups as new quantity, use that to re-calculate standard deviation, conduct hypothesis test.
Paired Tests for Proportions

For paired data, we have to take into account the fact that we have dependence between groups. For sample means, straightforward: take the difference between groups as a new quantity, use that to re-calculate the standard deviation, conduct a hypothesis test. However, for proportions, we sometimes don’t have the entire table of individual observations.
Paired Tests for Proportions

- For paired data, we have to take into account the fact that we have dependence between groups.
- For sample means, straightforward → take difference between groups as new quantity, use that to re-calculate standard deviation, conduct hypothesis test.
- However, for proportions we sometimes don’t have the entire table of individual observations.
- Usually only have a contingency table.
Paired Tests for Proportions

- For paired data, we have to take into account the fact that we have dependence between groups.
- For sample means, straightforward → take difference between groups as a new quantity, use that to re-calculate the standard deviation, conduct hypothesis test.
- However, for proportions we sometimes don’t have the entire table of individual observations.
- Usually only have a contingency table.
- So we estimate the standard error slightly differently.
Proportion example from public opinion

Ex) Public opinion example (difference in proportion) from GSS data on government oversight given suspected terrorist activity

Each person asked 2 questions: (1) ok for gov't to tap phone, or (2) ok for gov't to conduct random stops

Results as follows:

<table>
<thead>
<tr>
<th>Q1: Tap Phone</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q2: Random Stop on St</td>
<td>494</td>
<td>335</td>
</tr>
<tr>
<td></td>
<td>126</td>
<td>537</td>
</tr>
</tbody>
</table>

Question: Does the true proportion answering yes to the first question differ significantly from the second question?
Proportion example from public opinion

- Ex) Public opinion example (difference in proportion) from GSS data on government oversight given suspected terrorist activity.
Proportion example from public opinion

- Ex) Public opinion example (difference in proportion) from GSS data on government oversight given suspected terrorist activity
- Each person asked 2 questions: (1) ok for gov’t to tap phone, or (2) ok for gov’t to conduct random stops

<table>
<thead>
<tr>
<th>Q2: Random Stop</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1: Tap Phone</td>
<td>Yes</td>
<td>494</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>126</td>
</tr>
</tbody>
</table>

Question: Does the true proportion answering yes to the first question differ significantly from the second question?
Proportion example from public opinion

- Ex) Public opinion example (difference in proportion) from GSS data on government oversight given suspected terrorist activity
- Each person asked 2 questions: (1) ok for gov’t to tap phone, or (2) ok for gov’t to conduct random stops
- Results as follows:

<table>
<thead>
<tr>
<th>Question 2: Random Stop</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1: Tap Phone</td>
<td>Yes</td>
<td>494</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>126</td>
</tr>
</tbody>
</table>
Proportion example from public opinion

- Ex) Public opinion example (difference in proportion) from GSS data on government oversight given suspected terrorist activity
- Each person asked 2 questions: (1) ok for gov’t to tap phone, or (2) ok for gov’t to conduct random stops
- Results as follows:

<table>
<thead>
<tr>
<th>Q1: Tap Phone</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>494</td>
<td>335</td>
</tr>
<tr>
<td>No</td>
<td>126</td>
<td>537</td>
</tr>
</tbody>
</table>

Question: Does the true proportion answering yes to the first question differ significantly from the second question?
Ex) Public opinion example (difference in proportion) from GSS data on government oversight given suspected terrorist activity

Each person asked 2 questions: (1) ok for gov’t to tap phone, or (2) ok for gov’t to conduct random stops

Results as follows:

<table>
<thead>
<tr>
<th>Q1: Tap Phone</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>494</td>
<td>335</td>
</tr>
<tr>
<td>No</td>
<td>126</td>
<td>537</td>
</tr>
</tbody>
</table>

Question: Does the true proportion answering yes to the first question differ significantly from the second question?
Proportion example from public opinion

Calculate some basic point estimates of “yes” answers:

Question #1: Proportion of people in sample who believe authorities should be able to tap phones:
\[ \hat{\pi}_1 = \frac{494}{1492} = 0.556 \]

Question #2: Proportion of people in sample who believe authorities should be able to randomly stop and search people on street:
\[ \hat{\pi}_2 = \frac{494}{1492} = 0.416 \]
Proportion example from public opinion

- Calculate some basic point estimates of “yes” answers:

  - Question #1: Proportion of people in sample who believe authorities should be able to tap phones: 
    \[ \hat{\pi}_1 = \frac{494 + 335}{1492} = 0.556 \]
  
  - Question #2: Proportion of people in sample who believe authorities should be able to randomly stop and search people on street: 
    \[ \hat{\pi}_2 = \frac{494 + 126}{1492} = 0.416 \]
Proportion example from public opinion

- Calculate some basic point estimates of “yes” answers:
- Question #1: Proportion of people in sample who believe authorities should be able to tap phones:

\[ \hat{\pi}_1 = \frac{494 + 335}{1492} = 0.556 \]

- Question #2: Proportion of people in sample who believe authorities should be able to randomly stop and search people on street:

\[ \hat{\pi}_2 = \frac{494 + 126}{1492} = 0.416 \]
Proportion example from public opinion

- Calculate some basic point estimates of “yes” answers:
- Question #1: Proportion of people in sample who believe authorities should be able to tap phones:
  \[ \hat{\pi}_1 = \frac{494 + 335}{1492} = 0.556 \]
Proportion example from public opinion

- Calculate some basic point estimates of “yes” answers:
- Question #1: Proportion of people in sample who believe authorities should be able to tap phones:
  \[ \hat{\pi}_1 = \frac{494 + 335}{1492} = 0.556 \]

- Question #2: Proportion of people in sample who believe authorities should be able to randomly stop and search people on street:
Proportion example from public opinion

- Calculate some basic point estimates of “yes” answers:
- Question #1: Proportion of people in sample who believe authorities should be able to tap phones:
  \[ \hat{\pi}_1 = \frac{494 + 335}{1492} = 0.556 \]

- Question #2: Proportion of people in sample who believe authorities should be able to randomly stop and search people on street:
  \[ \hat{\pi}_2 = \frac{494 + 126}{1492} = 0.416 \]
Proportion example from public opinion
Proportion example from public opinion

- Steps 1 & 2 of hypothesis test the same as non-paired proportions test
Proportion example from public opinion

- Steps 1 & 2 of hypothesis test the same as non-paired proportions test
- Step 3: This is where paired data differ from independent data
Proportion example from public opinion

- Steps 1 & 2 of hypothesis test the same as non-paired proportions test
- Step 3: This is where paired data differ from independent data
- Independent data: Assume no covariance between groups
Proportion example from public opinion

- Steps 1 & 2 of hypothesis test the same as non-paired proportions test
- Step 3: This is where paired data differ from independent data
- Independent data: Assume no covariance between groups
- Paired (non-independent) data: Must adjust standard error to accommodate covariance
Proportion example from public opinion

- Steps 1 & 2 of hypothesis test the same as non-paired proportions test
- Step 3: This is where paired data differ from independent data
- Independent data: Assume no covariance between groups
- Paired (non-independent) data: Must adjust standard error to accommodate covariance
- In means case → Just looked at difference (\( \bar{X}_d \))
Proportion example from public opinion

- Steps 1 & 2 of hypothesis test the same as non-paired proportions test
- Step 3: This is where paired data differ from independent data
- Independent data: Assume no covariance between groups
- Paired (non-independent) data: Must adjust standard error to accommodate covariance
- In means case → Just looked at difference ($\bar{X}_d$)
- In proportions case → Oftentimes don’t have that data
Proportion example from public opinion

- Interested in distribution of $\pi_1 - \pi_2$
- By CLT, this should be normally distributed
- From earlier lecture, if independent, then $\pi_1 - \pi_2 \sim N(\pi_1 - \pi_2, \pi_1(1-\pi_1)n_1 + \pi_2(1-\pi_2)n_2)$
- However, here dependent, so $\pi_1 - \pi_2 \sim N(\pi_1 - \pi_2, \pi_1\pi_2 - (\pi_1 - \pi_2)^2/n)$
- where these refer to cell proportions (not conditional proportions)
- (Proof in appendix)
Proportion example from public opinion

- Interested in distribution of $\hat{\pi}_1 - \hat{\pi}_2$

- By CLT, this should be normally distributed

- From earlier lecture, if independent, then $\hat{\pi}_1 - \hat{\pi}_2 \sim N(\pi_1 - \pi_2, \pi_1(1-\pi_1)n_1 + \pi_2(1-\pi_2)n_2)$

- However, here dependent, so $\hat{\pi}_1 - \hat{\pi}_2 \sim N(\pi_1 - \pi_2, \pi_1\pi_2 - (\pi_1\pi_2 - \pi_2\pi_2)^2/n)$

- where these refer to cell proportions (not conditional proportions)

- (Proof in appendix)
Proportion example from public opinion

- Interested in distribution of $\hat{\pi}_1 - \hat{\pi}_2$
- By CLT, this should be normally distributed
Proportion example from public opinion

- Interested in distribution of $\hat{\pi}_1 - \hat{\pi}_2$
- By CLT, this should be normally distributed
- From earlier lecture, if independent, then $\hat{\pi}_1 - \hat{\pi}_2$

However, here dependent, so $\hat{\pi}_1 - \hat{\pi}_2$
Proportion example from public opinion

- Interested in distribution of $\hat{\pi}_1 - \hat{\pi}_2$
- By CLT, this should be normally distributed
- From earlier lecture, if independent, then $\hat{\pi}_1 - \hat{\pi}_2$

$$\sim N(\pi_1 - \pi_2, \frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2})$$
Proportion example from public opinion

- Interested in distribution of $\hat{\pi}_1 - \hat{\pi}_2$
- By CLT, this should be normally distributed
- From earlier lecture, if independent, then $\hat{\pi}_1 - \hat{\pi}_2$
  
  $$\sim N(\pi_1 - \pi_2, \frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2})$$

- However, here dependent, so $\hat{\pi}_1 - \hat{\pi}_2$
Proportion example from public opinion

- Interested in distribution of \( \hat{\pi}_1 - \hat{\pi}_2 \)
- By CLT, this should be normally distributed
- From earlier lecture, if independent, then \( \hat{\pi}_1 - \hat{\pi}_2 \)
  \[
  \sim N(\pi_1 - \pi_2, \frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2})
  \]

- However, here dependent, so \( \hat{\pi}_1 - \hat{\pi}_2 \)
  \[
  \sim N(\pi_1 - \pi_2, \frac{\pi_{12} + \pi_{21} - (\pi_{12} - \pi_{21})^2}{n})
  \]
Proportion example from public opinion

- Interested in distribution of $\hat{\pi}_1 - \hat{\pi}_2$
- By CLT, this should be normally distributed
- From earlier lecture, if independent, then $\hat{\pi}_1 - \hat{\pi}_2$
  \[
  \sim N(\pi_1 - \pi_2, \frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2})
  \]
- However, here dependent, so $\hat{\pi}_1 - \hat{\pi}_2$
  \[
  \sim N(\pi_1 - \pi_2, \frac{\pi_{12} + \pi_{21} - (\pi_{12} - \pi_{21})^2}{n})
  \]
- where these refer to cell proportions (not conditional proportions)
Proportion example from public opinion

- Interested in distribution of $\hat{\pi}_1 - \hat{\pi}_2$
- By CLT, this should be normally distributed
- From earlier lecture, if independent, then $\hat{\pi}_1 - \hat{\pi}_2$

$$\sim N(\pi_1 - \pi_2, \frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2})$$

- However, here dependent, so $\hat{\pi}_1 - \hat{\pi}_2$

$$\sim N(\pi_1 - \pi_2, \frac{\pi_{12} + \pi_{21} - (\pi_{12} - \pi_{21})^2}{n})$$

- where these refer to cell proportions (not conditional proportions)
- (Proof in appendix)
Proportion example from public opinion

This gives us a test statistic:

\[ z = \hat{\pi}_1 - \hat{\pi}_2 - (\pi_1 - \pi_2) \sqrt{\hat{\pi}_1 \hat{\pi}_2 + \hat{\pi}_2 \hat{\pi}_1} - (\hat{\pi}_1 \hat{\pi}_2 - \hat{\pi}_2 \hat{\pi}_1)^2 \frac{1}{n} \]

When the null is true, \( \pi_1 - \pi_2 = 0 \), we have some options:

1) Use this expression for the standard error
2) Simplify test using McNemar's Test for comparing dependent proportions (medicine/public health)

\[ z = n_{12} - n_{21} \sqrt{n_{12} + n_{21}} \]

where this approximately comes from standard Normal.

Intuition borrows from Binomial distribution – explanation of McNemar's test statistic in Appendix.
Proportion example from public opinion

- This gives us a test statistic:

\[
\hat{\pi}_1 - \hat{\pi}_2 - (\pi_1 - \pi_2) \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}} - (\hat{\pi}_1 - \hat{\pi}_2)^2 \frac{1}{n_1 n_2}
\]

- When the null is true, \(\pi_1 - \pi_2 = 0\), we have some options:
  1. Use this expression for the standard error
  2. Simplify test using McNemar's Test for comparing dependent proportions (medicine/public health)

\[
\frac{n_{12} - n_{21}}{\sqrt{n_{12} + n_{21}}}
\]

- where this approximately comes from standard Normal

Intuition borrows from Binomial distribution – explanation of McNemar's test statistic in Appendix
This gives us a test statistic:

\[ z = \frac{\hat{\pi}_1 - \hat{\pi}_2 - (\pi_1 - \pi_2)}{\sqrt{\frac{\hat{\pi}_{12} + \hat{\pi}_{21} - (\hat{\pi}_{12} - \hat{\pi}_{21})^2}{n}}} \]
Proportion example from public opinion

- This gives us a test statistic:

\[
z = \frac{\hat{\pi}_1 - \hat{\pi}_2 - (\pi_1 - \pi_2)}{\sqrt{\frac{\hat{\pi}_{12} + \hat{\pi}_{21} - (\hat{\pi}_{12} - \hat{\pi}_{21})^2}{n}}}
\]

- When the null is true, \( \pi_1 - \pi_2 = 0 \), we have some options
Proportion example from public opinion

- This gives us a test statistic:

\[
Z = \frac{\hat{\pi}_1 - \hat{\pi}_2 - (\pi_1 - \pi_2)}{\sqrt{\frac{\hat{\pi}_{12} + \hat{\pi}_{21} - (\hat{\pi}_{12} - \hat{\pi}_{21})^2}{n}}}
\]

- When the null is true, \( \pi_1 - \pi_2 = 0 \), we have some options:
  1) Use this expression for the standard error

Intuition borrows from Binomial distribution – explanation of McNemar's test statistic in Appendix.
Proportion example from public opinion

- This gives us a test statistic:

\[ z = \frac{\hat{\pi}_1 - \hat{\pi}_2 - (\pi_1 - \pi_2)}{\sqrt{\hat{\pi}_{12} + \hat{\pi}_{21} - (\hat{\pi}_{12} - \hat{\pi}_{21})^2 / n}} \]

- When the null is true, \( \pi_1 - \pi_2 = 0 \), we have some options
  - 1) Use this expression for the standard error
  - 2) Simplify test using McNemar’s Test for comparing dependent proportions (medicine/public health)
Proportion example from public opinion

- This gives us a test statistic:

\[
z = \frac{\hat{\pi}_1 - \hat{\pi}_2 - (\pi_1 - \pi_2)}{\sqrt{\frac{\hat{\pi}_{12} + \hat{\pi}_{21} - (\hat{\pi}_{12} - \hat{\pi}_{21})^2}{n}}}
\]

- When the null is true, \(\pi_1 - \pi_2 = 0\), we have some options
  - 1) Use this expression for the standard error
  - 2) Simplify test using McNemar’s Test for comparing dependent proportions (medicine/public health)

\[
z = \frac{n_{12} - n_{21}}{\sqrt{n_{12} + n_{21}}}
\]

Intuition borrows from Binomial distribution – explanation of McNemar’s test statistic in Appendix
Proportion example from public opinion

- This gives us a test statistic:

\[
z = \frac{\hat{\pi}_1 - \hat{\pi}_2 - (\pi_1 - \pi_2)}{\sqrt{\frac{\hat{\pi}_{12} + \hat{\pi}_{21} - (\hat{\pi}_{12} - \hat{\pi}_{21})^2}{n}}}
\]

- When the null is true, \( \pi_1 - \pi_2 = 0 \), we have some options
  - 1) Use this expression for the standard error
  - 2) Simplify test using McNemar’s Test for comparing dependent proportions (medicine/public health)

\[
z = \frac{n_{12} - n_{21}}{\sqrt{n_{12} + n_{21}}}
\]

- where this approximately comes from standard Normal
Proportion example from public opinion

- This gives us a test statistic:

\[ z = \frac{\hat{\pi}_1 - \hat{\pi}_2 - (\pi_1 - \pi_2)}{\sqrt{\hat{\pi}_{12} + \hat{\pi}_{21} - (\hat{\pi}_{12} - \hat{\pi}_{21})^2}} \]

- When the null is true, \(\pi_1 - \pi_2 = 0\), we have some options
  1) Use this expression for the standard error
  2) Simplify test using McNemar’s Test for comparing dependent proportions (medicine/public health)

\[ z = \frac{n_{12} - n_{21}}{\sqrt{n_{12} + n_{21}}} \]

- where this approximately comes from standard Normal
- Intuition borrows from Binomial distribution – explanation of McNemar’s test statistic in Appendix
Proportion example from public opinion

Calculating McNemar's test statistic:

\[ z = \frac{n_{12} - n_{21}}{\sqrt{n_{12} + n_{21}}} \]

where this approx comes from a standard normal distribution.
Calculating McNemar’s test statistic:
Calculating McNemar’s test statistic:

\[ z = \frac{n_{12} - n_{21}}{\sqrt{n_{12} + n_{21}}} \]

\[ = \frac{335 - 126}{\sqrt{335 + 126}} \approx 9.7341 \]
Calculating McNemar’s test statistic:

\[ z = \frac{n_{12} - n_{21}}{\sqrt{n_{12} + n_{21}}} \]

\[ = \frac{335 - 126}{\sqrt{335 + 126}} = 9.7341 \]

where this approx comes from a standard normal distribution
Step 4: Calculate \( p \)-value (using two-tailed test)

\[ p \text{-value} = 2 \times P(Z \leq -9.7341) \]

\( p \)-value < 0.01

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

Question: What is your conclusion?

(People seem to have different tolerance for gov't tapping phones vs random stops on street)
Proportion example from public opinion

- **Step 4**: Calculate $p$-value (using two-tailed test)

  - $p$-value = $2 \times P(Z \leq -9.7341)$
  - $p$-value < 0.01

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

  - Question: What is your conclusion?
  - (People seem to have different tolerance for gov't tapping phones vs random stops on street)
Proportion example from public opinion

- Step 4: Calculate $p$-value (using two-tailed test)
  
  $p$-value $= 2 \times P(Z \leq -9.7341)$
Step 4: Calculate $p$-value (using two-tailed test)

$p$-value $= 2 \times P(Z \leq -9.7341)$

$p$-value $< 0.01$
Proportion example from public opinion

- Step 4: Calculate $p$-value (using two-tailed test)
  - $p$-value = $2 \times P(Z \leq -9.7341)$
  - $p$-value < 0.01
- Step 5: Decide whether or not to reject the null hypothesis and interpret results
Proportion example from public opinion

- Step 4: Calculate $p$-value (using two-tailed test)
  - $p$-value = $2 \times P(Z \leq -9.7341)$
  - $p$-value < 0.01
- Step 5: Decide whether or not to reject the null hypothesis and interpret results
- Question: What is your conclusion?
Proportion example from public opinion

- Step 4: Calculate $p$-value (using two-tailed test)
  - $p$-value = $2 \times P(Z \leq -9.7341)$
  - $p$-value < 0.01
- Step 5: Decide whether or not to reject the null hypothesis and interpret results
- Question: What is your conclusion?
- (People seem to have different tolerance for gov’t tapping phones vs random stops on street)
Proportion example from public opinion

To calculate confidence interval, follow the same formula:

\[ \hat{\pi}_1 - \hat{\pi}_2 \pm z_{\alpha/2} \sqrt{\hat{\pi}_1(1-\hat{\pi}_1)/n - \hat{\pi}_2(1-\hat{\pi}_2)/n} \]

Using the full form of standard error:
Proportion example from public opinion

- To calculate confidence interval, follow same formula:
To calculate confidence interval, follow same formula

\[ \hat{\pi}_1 - \hat{\pi}_2 \pm z_{\alpha/2} SE[\hat{\pi}_1 - \hat{\pi}_2] \]
Proportion example from public opinion

- To calculate confidence interval, follow same formula

\[
\hat{\pi}_1 - \hat{\pi}_2 \pm z_{\alpha/2} SE[\hat{\pi}_1 - \hat{\pi}_2]
\]

- Using full form of standard error
Proportion example from public opinion

To calculate confidence interval, follow same formula

\[ \hat{\pi}_1 - \hat{\pi}_2 \pm z_{\alpha/2} \times SE[\hat{\pi}_1 - \hat{\pi}_2] \]

Using full form of standard error

\[ \hat{\pi}_1 - \hat{\pi}_2 \pm z_{\alpha/2} \sqrt{\frac{\hat{\pi}_{12} - \hat{\pi}_{21} - (\hat{\pi}_{12} - \hat{\pi}_{21})^2}{n}} \]
Proportion example from public opinion

Using our example

For 95% CI:

\[0.556 - 0.416 \pm 1.96 \sqrt{\frac{1492}{335} - 126 \frac{1492}{335} - \left(\frac{335}{1492} - 126\right)^2 1492}\]

\[0.14 \pm 0.018\]

So 95% is (0.122, 0.158)

Does it include 0?

What does this mean substantively?
Proportion example from public opinion

▶ Using our example

0.556 - 0.416 ± 1.96 \sqrt{\frac{335}{1492} - \frac{126}{1492}}

0.14 ± 0.018

So 95% is (0.122, 0.158)

Does it include 0?

What does this mean substantively?
Proportion example from public opinion

- Using our example
- For 95% CI:
Proportion example from public opinion

- Using our example
- For 95% CI:

\[
0.556 - 0.416 \pm 1.96 \sqrt{\frac{335}{1492} - \frac{126}{1492} - \left( \frac{335}{1492} - \frac{126}{1492} \right)^2} \frac{1492}{1492}
\]

0.14 ± 0.018

So 95% is (0.122, 0.158)

Does it include 0?

What does this mean substantively?
Using our example

For 95% CI:

\[
0.556 - 0.416 \pm 1.96 \sqrt{\frac{335}{1492} - \frac{126}{1492} - \left( \frac{335}{1492} - \frac{126}{1492} \right)^2} \frac{1492}{1492}
\]

0.14 ± 0.018

So 95% is (0.122, 0.158)
Using our example

For 95% CI:

\[ 0.556 - 0.416 \pm 1.96 \sqrt{\frac{\frac{335}{1492} - \frac{126}{1492} - \left( \frac{\frac{335}{1492} - \frac{126}{1492}}{1492} \right)^2}{1492}} \]

\[ 0.14 \pm 0.018 \]

So 95% is (0.122, 0.158)

Does it include 0?
Proportion example from public opinion

- Using our example
- For 95% CI:

\[ 0.556 - 0.416 \pm 1.96 \sqrt{\frac{335}{1492} - \frac{126}{1492} - \left( \frac{335}{1492} - \frac{126}{1492} \right)^2}{1492} \]

\[ 0.14 \pm 0.018 \]

- So 95% is (0.122, 0.158)
- Does it include 0?
- What does this mean substantively?
Switching to multiple comparisons
Switching to multiple comparisons

- We have spent the last few classes looking at tests for:
  - one and two means (independent, paired, pooled)
  - one and two proportions (independent or paired)
  - What happens if we want to compare observations from 3 or more independent populations?
Switching to multiple comparisons

- We have spent the last few classes looking at tests for:
  - one and two means (independent, paired, pooled)
Switching to multiple comparisons

- We have spent the last few classes looking at tests for:
  - one and two means (independent, paired, pooled)
  - one and two proportions (independent or paired)
Switching to multiple comparisons

- We have spent the last few classes looking at tests for:
  - one and two means (independent, paired, pooled)
  - one and two proportions (independent or paired)
- What happens if we want to compare observations from 3 or more independent populations?
Some examples
Some examples

▶ Economics: Does mean consumer debt differ meaningfully between five different countries?
Some examples

- Economics: Does mean consumer debt differ meaningfully between five different countries?
- Medicine: Does a medical treatment help blacks, Latino, Asian Americans differently?
Some examples

- Economics: Does mean consumer debt differ meaningfully between five different countries?
- Medicine: Does a medical treatment help blacks, Latino, Asian Americans differently?
- Education: Is there a difference in the average SAT scores across 4 high schools in Boston?
Some examples

- **Economics**: Does mean consumer debt differ meaningfully between five different countries?
- **Medicine**: Does a medical treatment help blacks, Latino, Asian Americans differently?
- **Education**: Is there a difference in the average SAT scores across 4 high schools in Boston?
- **Health**: Does mean weight loss differ over 6 months between subjects following 5 different diets?
Some examples

- Economics: Does mean consumer debt differ meaningfully between five different countries?
- Medicine: Does a medical treatment help blacks, Latino, Asian Americans differently?
- Education: Is there a difference in the average SAT scores across 4 high schools in Boston?
- Health: Does mean weight loss differ over 6 months between subjects following 5 different diets?

In all of these → want to compare population means across more than two groups
Life Expectancy Example
Life Expectancy Example

- Ex) Life expectancies from 193 countries around the world
Life Expectancy Example

- Ex) Life expectancies from 193 countries around the world
- Data based on World Bank data for 6 different continents
Life Expectancy Example

- Ex) Life expectancies from 193 countries around the world
- Data based on World Bank data for 6 different continents
- We can assume different continents (groups) independent (no country in more than one continent)
<table>
<thead>
<tr>
<th>Rank</th>
<th>Life Expectancy</th>
<th>Continent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>73.1</td>
<td>Africa</td>
</tr>
<tr>
<td>2</td>
<td>48.1</td>
<td>Asia</td>
</tr>
<tr>
<td>3</td>
<td>81.8</td>
<td>Oceania</td>
</tr>
<tr>
<td>4</td>
<td>77</td>
<td>Europe</td>
</tr>
<tr>
<td>5</td>
<td>75.1</td>
<td>North America</td>
</tr>
<tr>
<td>6</td>
<td>73.1</td>
<td>Africa</td>
</tr>
<tr>
<td>7</td>
<td>74.3</td>
<td>South America</td>
</tr>
</tbody>
</table>
## Life Expectancy Example

<table>
<thead>
<tr>
<th>Country</th>
<th>Average Life Expectancy</th>
<th>Continent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>73.1</td>
<td>Africa</td>
</tr>
<tr>
<td>2</td>
<td>48.1</td>
<td>Asia</td>
</tr>
<tr>
<td>3</td>
<td>81.8</td>
<td>Oceania</td>
</tr>
<tr>
<td>4</td>
<td>77</td>
<td>Europe</td>
</tr>
<tr>
<td>5</td>
<td>75.1</td>
<td>North America</td>
</tr>
<tr>
<td>6</td>
<td>73.1</td>
<td>Africa</td>
</tr>
<tr>
<td>7</td>
<td>74.3</td>
<td>South America</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Region</td>
<td>Life Expectancy</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Africa</td>
<td>57.54</td>
<td>7.97</td>
</tr>
<tr>
<td>Asia</td>
<td>72.02</td>
<td>6.33</td>
</tr>
<tr>
<td>Europe</td>
<td>78.11</td>
<td>3.93</td>
</tr>
<tr>
<td>Oceania</td>
<td>72.69</td>
<td>5.37</td>
</tr>
<tr>
<td>North Am</td>
<td>74.93</td>
<td>4.13</td>
</tr>
<tr>
<td>South Am</td>
<td>73.81</td>
<td>3.26</td>
</tr>
</tbody>
</table>
## Life Expectancy Example

<table>
<thead>
<tr>
<th></th>
<th>Africa</th>
<th>Asia</th>
<th>Europe</th>
<th>Oceania</th>
<th>North Am</th>
<th>South Am</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{X}_i$</td>
<td>57.54</td>
<td>72.02</td>
<td>78.11</td>
<td>72.69</td>
<td>74.93</td>
<td>73.81</td>
</tr>
<tr>
<td>$s_i$</td>
<td>7.97</td>
<td>6.33</td>
<td>3.93</td>
<td>5.37</td>
<td>4.13</td>
<td>3.26</td>
</tr>
<tr>
<td>$n_i$</td>
<td>52</td>
<td>50</td>
<td>42</td>
<td>13</td>
<td>25</td>
<td>11</td>
</tr>
</tbody>
</table>
Life Expectancy Example
Life Expectancy Example

Life Expectancy of Countries by Continent

[Box plot showing life expectancy by continent]
Life Expectancy Example

We could compare each possible pair using difference-in-means t-test at $\alpha = 0.05$ level.

African to Oceania, Africa to Europe, Oceania to Europe, etc.

Problem → Each hypothesis test has $P$(Type I Error) of 0.05.

What is the probability of a Type I error if we test all pairwise combinations of means?

15 possible combinations of tests →

$Pr$ none of them having a Type 1 error $= 0.95^{15}$

$Pr$ at least one has a Type 1 error, $1 - 0.95^{15}$, or around 54%

As number of groups compared increases → $P$(at least one Type I error) also increases.
Life Expectancy Example

- We could compare each possible pair using difference-in-means $t$-test at $\alpha = 0.05$ level

Problem

- Each hypothesis test has $P$(Type I Error) of 0.05

- What is the probability of a Type I error if we test all pairwise combinations of means?

- 15 possible combinations of tests

- $\Pr$ none of them having a Type 1 error = $0.95^{15}$

- So $\Pr$ at least one has a Type 1 error, $1 - 0.95^{15}$, or around 54%

- As number of groups compared increases $\rightarrow P$(at least one Type I error) also increases
Life Expectancy Example

- We could compare each possible pair using difference-in-means $t$-test at $\alpha = 0.05$ level
  - African to Oceania, Africa to Europe, Oceania to Europe, etc.

- Problem → Each hypothesis test has P(Type I Error) of 0.05
- What is the probability of a Type I error if we test all pairwise combinations of means?
- 15 possible combinations of tests → $\Pr$ none of them having a Type 1 error = $\frac{0.95}{15}$
- So $\Pr$ at least one has a Type 1 error, $1 - \frac{0.95}{15}$, or around 54%
- As number of groups compared increases → P(at least one Type I error) also increases
Life Expectancy Example

- We could compare each possible pair using difference-in-means $t$-test at $\alpha = 0.05$ level
  - African to Oceania, Africa to Europe, Oceania to Europe, etc.
- Problem $\rightarrow$ Each hypothesis test has $P(\text{Type I Error})$ of 0.05
Life Expectancy Example

- We could compare each possible pair using difference-in-means $t$-test at $\alpha = 0.05$ level
  - African to Oceania, Africa to Europe, Oceania to Europe, etc.
- Problem $\rightarrow$ Each hypothesis test has $P$(Type I Error) of 0.05
- What is the probability of a Type I error if we test all pairwise combinations of means?

\[ \Pr \text{ none of them having a Type 1 error} = 0.95^{15} \]
\[ \Pr \text{ at least one has a Type 1 error} = 1 - 0.95^{15}, \text{ or around } 54\% \]

As number of groups compared increases $\rightarrow P$(at least one Type I error) also increases
Life Expectancy Example

- We could compare each possible pair using difference-in-means $t$-test at $\alpha = 0.05$ level
  - African to Oceania, Africa to Europe, Oceania to Europe, etc.
- Problem → Each hypothesis test has $P(\text{Type I Error})$ of 0.05
- What is the probability of a Type I error if we test all pairwise combinations of means?
- 15 possible combinations of tests →

Pr none of them having a Type 1 error = $0.95^{15}$

So Pr at least one has a Type 1 error, $1 - 0.95^{15}$, or around 54%
Life Expectancy Example

- We could compare each possible pair using difference-in-means $t$-test at $\alpha = 0.05$ level
  - African to Oceania, Africa to Europe, Oceania to Europe, etc.
- Problem $\rightarrow$ Each hypothesis test has $P$(Type I Error) of 0.05
- What is the probability of a Type I error if we test all pairwise combinations of means?
- 15 possible combinations of tests $\rightarrow$
  - $Pr$ none of them having a Type 1 error $= 0.95^{15}$
Life Expectancy Example

- We could compare each possible pair using difference-in-means $t$-test at $\alpha = 0.05$ level
  - African to Oceania, Africa to Europe, Oceania to Europe, etc.
- Problem $\rightarrow$ Each hypothesis test has $P(\text{Type I Error})$ of 0.05
- What is the probability of a Type I error if we test all pairwise combinations of means?
- 15 possible combinations of tests $\rightarrow$
  - Pr none of them having a Type 1 error $= 0.95^{15}$
  - So Pr at least one has a Type 1 error, $1 - 0.95^{15}$, or around 54%
Life Expectancy Example

- We could compare each possible pair using difference-in-means \( t \)-test at \( \alpha = 0.05 \) level
  - African to Oceania, Africa to Europe, Oceania to Europe, etc.
- Problem \( \rightarrow \) Each hypothesis test has \( P(\text{Type I Error}) \) of 0.05
- What is the probability of a Type I error if we test all pairwise combinations of means?
- 15 possible combinations of tests \( \rightarrow \)
  - \( Pr \) none of them having a Type 1 error \( = 0.95^{15} \)
  - So \( Pr \) at least one has a Type 1 error, \( 1 - 0.95^{15} \), or around 54%
- As number of groups compared increases \( \rightarrow P(\text{at least one Type I error}) \) also increases
ANOVA

▶ Instead use one-way ANOVA (Analysis of Variance)
▶ Type of test frequently used in psychology, epidemiology, other fields that rely on experiments
▶ “one-way” → Exploring one characteristic (life expectancy)
▶ Could explore two characteristics (life expectancy, weight) w/ “two-way ANOVA” (more complicated)
▶ Here, use one-way ANOVA as a global test, which tests null hypothesis that population means are all equal
▶ Null hypothesis for this ANOVA test: $\mu_1 = \mu_2 = \mu_3 = ... = \mu_k$
▶ Alternative hypothesis: At last two of the population means are unequal
▶ Note: Can be all population means, some population means, or just two that differ
→ Null hypothesis generally pretty strong for global tests
ANOVA

- Instead use one-way ANOVA (Analysis of Variance)

Type of test frequently used in psychology, epidemiology, other fields that rely on experiments

"one-way" → Exploring one characteristic (life expectancy)

Could explore two characteristics (life expectancy, weight) with "two-way ANOVA" (more complicated)

Here, use one-way ANOVA as a global test, which tests null hypothesis that population means are all equal

Null hypothesis for this ANOVA test:

\[ \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k \]

Alternative hypothesis:

At least two of the population means are unequal

Note: Can be all population means, some population means, or just two that differ

→ Null hypothesis generally pretty strong for global tests
ANOVA

- Instead use one-way ANOVA (Analysis of Variance)
- Type of test frequently used in psychology, epidemiology, other fields that rely on experiments
Instead use one-way ANOVA (Analysis of Variance)

Type of test frequently used in psychology, epidemiology, other fields that rely on experiments

“one-way” → Exploring one characteristic (life expectancy)

Null hypothesis for this ANOVA test:

$\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$

Alternative hypothesis:

At least two of the population means are unequal

Note: Can be all population means, some population means, or just two that differ

Null hypothesis generally pretty strong for global tests
ANOVA

- Instead use one-way ANOVA (Analysis of Variance)
- Type of test frequently used in psychology, epidemiology, other fields that rely on experiments
  - “one-way” → Exploring one characteristic (life expectancy)
  - Could explore two characteristics (life expectancy, weight) w/ “two-way ANOVA” (more complicated)

Null hypothesis for this ANOVA test:

$\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$

Alternative hypothesis:

At last two of the population means are unequal

Note: Can be all population means, some population means, or just two that differ

Null hypothesis generally pretty strong for global tests
ANOVA

- Instead use **one-way ANOVA** (Analysis of Variance)
- Type of test frequently used in psychology, epidemiology, other fields that rely on experiments
  - “one-way” → Exploring one characteristic (life expectancy)
  - Could explore two characteristics (life expectancy, weight) w/ “two-way ANOVA” (more complicated)
- Here, use one-way ANOVA as a **global test**, which tests null that population means are all equal
Instead use one-way ANOVA (Analysis of Variance) Type of test frequently used in psychology, epidemiology, other fields that rely on experiments

- “one-way” → Exploring one characteristic (life expectancy)
- Could explore two characteristics (life expectancy, weight) w/ “two-way ANOVA” (more complicated)

Here, use one-way ANOVA as a global test, which tests null that population means are all equal Null hypothesis for this ANOVA test:
Instead use **one-way ANOVA** (Analysis of Variance)

Type of test frequently used in psychology, epidemiology, other fields that rely on experiments

- “one-way” → Exploring one characteristic (life expectancy)
- Could explore two characteristics (life expectancy, weight) w/ “two-way ANOVA” (more complicated)

Here, use one-way ANOVA as a global test, which tests null that population means are all equal

Null hypothesis for this ANOVA test:

- \( \mu_1 = \mu_2 = \mu_3 = \ldots \mu_k \)
Instead use **one-way ANOVA (Analysis of Variance)**

- Type of test frequently used in psychology, epidemiology, other fields that rely on experiments
  - “one-way” → Exploring one characteristic (life expectancy)
  - Could explore two characteristics (life expectancy, weight) w/ “two-way ANOVA” (more complicated)

- Here, use one-way ANOVA as a **global test**, which tests null that population means are all equal

- Null hypothesis for this ANOVA test:
  - $\mu_1 = \mu_2 = \mu_3 = \cdots \mu_k$

- Alternative hypothesis:
Instead use one-way ANOVA (Analysis of Variance)
Type of test frequently used in psychology, epidemiology, other fields that rely on experiments
  “one-way” → Exploring one characteristic (life expectancy)
  Could explore two characteristics (life expectancy, weight) w/ “two-way ANOVA” (more complicated)
Here, use one-way ANOVA as a global test, which tests null that population means are all equal
Null hypothesis for this ANOVA test:
  \[ \mu_1 = \mu_2 = \mu_3 = \ldots \mu_k \]
Alternative hypothesis:
  At last two of the population means are unequal
ANOVA

- Instead use **one-way ANOVA (Analysis of Variance)**
- Type of test frequently used in psychology, epidemiology, other fields that rely on experiments
  - “one-way” → Exploring one characteristic (life expectancy)
  - Could explore two characteristics (life expectancy, weight) w/ “two-way ANOVA” (more complicated)
- Here, use one-way ANOVA as a **global test**, which tests null that population means are all equal
  - Null hypothesis for this ANOVA test:
    - $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$
  - Alternative hypothesis:
    - At least two of the population means are unequal
    - Note: Can be all population means, some population means, or just two that differ
ANOVA

▶ Instead use one-way ANOVA (Analysis of Variance)
▶ Type of test frequently used in psychology, epidemiology, other fields that rely on experiments
  ▶ “one-way” → Exploring one characteristic (life expectancy)
  ▶ Could explore two characteristics (life expectancy, weight) w/ “two-way ANOVA” (more complicated)
▶ Here, use one-way ANOVA as a global test, which tests null that population means are all equal
▶ Null hypothesis for this ANOVA test:
  ▶ $\mu_1 = \mu_2 = \mu_3 = \ldots \mu_k$
▶ Alternative hypothesis:
  ▶ At last two of the population means are unequal
  ▶ Note: Can be all population means, some population means, or just two that differ
▶ → Null hypothesis generally pretty strong for global tests
ANOVA

Different from hypothesis tests (which rely on CLT, comparing test statistics to a standard normal)

ANOVA: Compare between-group and within-group variation

ANOVA tests rely on the fact that total variability composed of:

1. Variability between groups
   - Ex) Compare mean life expectancy for each continent to global mean life expectancy

2. Variability within groups
   - Ex) Compare life expectancy of individual countries to their continent's mean life expectancy

However: Both ANOVA and hypothesis tests rely on calculating test statistic, using that to reject or not reject null hypothesis.
ANOVA

- Different from hypothesis tests (which rely on CLT, comparing test statistics to a standard normal)

  1. Variability between groups
     - Ex) Compare mean life expectancy for each continent to global mean life expectancy
  2. Variability within groups
     - Ex) Compare life expectancy of individual countries to their continent's mean life expectancy

However: Both ANOVA and hypothesis tests rely on calculating test statistic, using that to reject or not reject null hypothesis.
ANOVA

- Different from hypothesis tests (which rely on CLT, comparing test statistics to a standard normal)
- ANOVA: Compare between-group and within-group variation

1. Variability between groups
   - Ex) Compare mean life expectancy for each continent to global mean life expectancy
2. Variability within groups
   - Ex) Compare life expectancy of individual countries to their continent's mean life expectancy

However: Both ANOVA and hypothesis tests rely on calculating test statistic, using that to reject or not reject null hypothesis
ANOVA

- Different from hypothesis tests (which rely on CLT, comparing test statistics to a standard normal)
- ANOVA: Compare between-group and within-group variation
- ANOVA tests rely on the fact that total variability composed of
ANOVA

- Different from hypothesis tests (which rely on CLT, comparing test statistics to a standard normal)
- ANOVA: Compare between-group and within-group variation
- ANOVA tests rely on the fact that total variability composed of
  1. Variability *between* groups

Ex) Compare mean life expectancy for each continent to global mean life expectancy
Ex) Compare life expectancy of individual countries to their continent's mean life expectancy

However: Both ANOVA and hypothesis tests rely on calculating test statistic, using that to reject or not reject null hypothesis
ANOVA

- Different from hypothesis tests (which rely on CLT, comparing test statistics to a standard normal)
- ANOVA: Compare between-group and within-group variation
- ANOVA tests rely on the fact that total variability composed of
  1. Variability *between* groups
     - Ex) Compare mean life expectancy for each continent to global mean life expectancy
  2. Variability *within* groups
     - Ex) Compare life expectancy of individual countries to their continent's mean life expectancy
ANOVA

- Different from hypothesis tests (which rely on CLT, comparing test statistics to a standard normal)
- ANOVA: Compare between-group and within-group variation
- ANOVA tests rely on the fact that total variability composed of
  1. Variability *between* groups
     - Ex) Compare mean life expectancy for each continent to global mean life expectancy
  2. Variability *within* groups
ANOVA

- Different from hypothesis tests (which rely on CLT, comparing test statistics to a standard normal)
- ANOVA: Compare between-group and within-group variation
- ANOVA tests rely on the fact that total variability composed of
  1. Variability *between* groups
     - Ex) Compare mean life expectancy for each continent to global mean life expectancy
  2. Variability *within* groups
     - Ex) Compare life expectancy of individual countries to their continent’s mean life expectancy
ANOVA

- Different from hypothesis tests (which rely on CLT, comparing test statistics to a standard normal)
- ANOVA: Compare between-group and within-group variation
- ANOVA tests rely on the fact that total variability composed of
  1. Variability between groups
     - Ex) Compare mean life expectancy for each continent to global mean life expectancy
  2. Variability within groups
     - Ex) Compare life expectancy of individual countries to their continent’s mean life expectancy
- However: Both ANOVA and hypothesis tests rely on calculating test statistic, using that to reject or not reject null hypothesis
Variability between groups

Variability between each continent and global mean is between-group sum of squares. This adds squared differences of (a) each group mean from (b) the global ("grand") mean. The between-group sum of squares for \( k \) groups is:

\[
\sum_{i=1}^{k} n_i (\bar{X}_i - \bar{X})^2
\]

Where:
- \( i \) is an index representing group (here, six continents)
- \( k = \) number of groups (six)
- \( \bar{X}_i \) is the mean of group \( i \)
- \( \bar{X} \) is the global ("grand") mean

Intuition: If group means are close to each other (and therefore to the grand mean) this will be small.
Variability between groups

- Variability between each continent and global mean is between-group sum of squares

\[ \sum_{i=1}^{k} n_i (\overline{X}_i - \overline{X})^2 \]

Where:
- \( i \) is an index representing group (here, six continents)
- \( k \) = # of groups (six)
- \( \overline{X}_i \) is the mean of group \( i \)
- \( \overline{X} \) is global (“grand”) mean

Intuition: If group means are close to each other (and therefore to grand mean) this will be small
Variability between groups

- Variability between each continent and global mean is between-group sum of squares
- Adds squared differences of (a) each group mean from (b) global ("grand") mean

\[
\text{between-group sum of squares for } k \text{ groups is:} \\
\sum_{i=1}^{k} n_i (\bar{X}_i - \bar{X})^2
\]

- \(i\) is an index representing group (here, six continents)
- \(k\) = # of groups (six)
- \(\bar{X}_i\) is the mean of group \(i\)
- \(\bar{X}\) is global ("grand") mean

Intuition: If group means are close to each other (and therefore to grand mean) this will be small
Variability between groups

- Variability between each continent and global mean is the between-group sum of squares.
- Adds squared differences of (a) each group mean from (b) global ("grand") mean.
- The between-group sum of squares for $k$ groups is:

$$\sum_{i=1}^{k} n_i \left( \bar{X}_i - \bar{X} \right)^2$$

- Where:
  - $i$ is an index representing group (here, six continents).
  - $k = \#\ of\ groups$ (six).
  - $\bar{X}_i$ is the mean of group $i$.
  - $\bar{X}$ is the global ("grand") mean.

Intuition: If group means are close to each other (and therefore to the grand mean), this will be small.
Variability between groups

- Variability between each continent and global mean is between-group sum of squares
- Adds squared differences of (a) each group mean from (b) global ("grand") mean
- The between-group sum of squares for $k$ groups is:

\[
\sum_{i=1}^{k} n_i (\bar{X}_i - \bar{X})^2
\]
Variability between groups

- Variability between each continent and global mean is between-group sum of squares
- Adds squared differences of (a) each group mean from (b) global ("grand") mean
- The between-group sum of squares for $k$ groups is:

$$\sum_{i=1}^{k} n_i (\bar{X}_i - \bar{X})^2$$

- Where:
Variability between groups

- Variability between each continent and global mean is between-group sum of squares
- Adds squared differences of (a) each group mean from (b) global (“grand”) mean
- The between-group sum of squares for $k$ groups is:

$$\sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X})^2$$

- Where:
  - $i$ is an index representing group (here, six continents)
Variability between groups

- Variability between each continent and global mean is between-group sum of squares
- Adds squared differences of (a) each group mean from (b) global ("grand") mean
- The between-group sum of squares for $k$ groups is:

$$
\sum_{i=1}^{k} n_i (\bar{X}_i - \bar{X})^2
$$

- Where:
  - $i$ is an index representing group (here, six continents)
  - $k = \#$ of groups (six)
Variability between groups

- Variability between each continent and global mean is the between-group sum of squares.
- Adds squared differences of (a) each group mean from (b) global ("grand") mean.
- The between-group sum of squares for \( k \) groups is:

\[
\sum_{i=1}^{k} n_i (\bar{X}_i - \bar{X})^2
\]

- Where:
  - \( i \) is an index representing group (here, six continents)
  - \( k = \# \) of groups (six)
  - \( \bar{X}_i \) is the mean of group \( i \)
Variability between groups

- Variability between each continent and global mean is between-group sum of squares.
- Adds squared differences of (a) each group mean from (b) global (“grand”) mean.
- The between-group sum of squares for $k$ groups is:

$$
\sum_{i=1}^{k} n_i (\bar{X}_i - \bar{X})^2
$$

- Where:
  - $i$ is an index representing group (here, six continents).
  - $k = \#$ of groups (six).
  - $\bar{X}_i$ is the mean of group $i$.
  - $\bar{X}$ is global (“grand”) mean.
Variability between groups

- Variability between each continent and global mean is between-group sum of squares
- Adds squared differences of (a) each group mean from (b) global ("grand") mean
- The between-group sum of squares for $k$ groups is:

$$\sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X})^2$$

- Where:
  - $i$ is an index representing group (here, six continents)
  - $k = \#$ of groups (six)
  - $\bar{X}_i$ is the mean of group $i$
  - $\bar{X}$ is global ("grand") mean
- Intuition: If group means are close to each other (and therefore to grand mean) this will be small
Variability within groups

The variability between individual countries within a continent is the within-group sum of squares. It adds squared differences of (a) each observation from (b) their group's mean.

$k \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \overline{X}_i)^2$

Where:
- $X_{ij}$ is an individual observation $j$ in group $i$.
- $n_i$ observations in group $i$.
- $k$ = number of continents.

Also referred to as Mean Squared Error (MSE).
Variability within groups

- The variability between individual countries within a continent is **within-group sum of squares**

\[
\sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2
\]

- Where \(X_{ij}\) is an individual observation
- \(n_i\) observations in group \(i\)
- \(k\) = number of continents

Also referred to as Mean Squared Error (MSE)
Variability within groups

- The variability between individual countries within a continent is \textit{within-group sum of squares}
- Adds squared differences of (a) each observation from (b) their group’s mean

\[
\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(X_{ij} - \bar{X}_i\right)^2
\]

- Where \(X_{ij}\) is an individual observation in group \(i\)
- \(n_i\) observations in group \(i\)
- \(k\) = number of continents
- Also referred to as Mean Squared Error (MSE)
Variability within groups

- The variability between individual countries within a continent is **within-group sum of squares**
- Adds squared differences of (a) each observation from (b) their group’s mean

\[
\sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2
\]
Variability within groups

- The variability between individual countries within a continent is **within-group sum of squares**

- Adds squared differences of (a) each observation from (b) their group’s mean

\[
\sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2
\]

- Where

\[
X_{ij} \quad \text{is an individual observation } j \text{ in group } i
\]

\[
n_i \quad \text{observations in group } i
\]

\[
k \quad \text{number of continents}
\]
Variability within groups

- The variability between individual countries within a continent is **within-group sum of squares**
- Adds squared differences of (a) each observation from (b) their group’s mean

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$$

Where
- $X_{ij}$ is an individual observation $j$ in group $i$
Variability within groups

- The variability between individual countries within a continent is **within-group sum of squares**
- Adds squared differences of (a) each observation from (b) their group’s mean

\[
\sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2
\]

- Where
  - \( X_{ij} \) is an individual observation \( j \) in group \( i \)
  - \( n_i \) observations in group \( i \)
Variability within groups

- The variability between individual countries within a continent
  is **within-group sum of squares**

- Adds squared differences of (a) each observation from (b)
  their group’s mean

\[
\sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2
\]

- Where
  - \(X_{ij}\) is an individual observation \(j\) in group \(i\)
  - \(n_i\) observations in group \(i\)
  - \(k = \# \text{ of continents}\)
Variability within groups

- The variability between individual countries within a continent is **within-group sum of squares**.
- Adds squared differences of (a) each observation from (b) their group’s mean.

\[
\sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2
\]

- Where
  - \( X_{ij} \) is an individual observation \( j \) in group \( i \).
  - \( n_i \) observations in group \( i \).
  - \( k = \# \) of continents.

- Also referred to as **Mean Squared Error (MSE)**.
Overall variability

A measure of overall variability in the dataset is called total Sum of Squares (SS). It adds squared differences of all individual observations across all groups from the grand mean.

\[ k \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} (X_{ij} - \bar{X})^2 \]
Overall variability

- A measure of overall variability in the dataset is called total Sum of Squares (SS)
Overall variability

- A measure of overall variability in the dataset is called **total Sum of Squares (SS)**
- Adds squared differences of all individual observations across all groups from global ("grand") mean
A measure of overall variability in the dataset is called total Sum of Squares (SS)

 Adds squared differences of all individual observations across all groups from global ("grand") mean

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2$$
ANOVA Table

Common to organize this info in an ANOVA table, which includes:

▶ Source of variation: (1) Between Group, (2) Within Group, or (3) Total
▶ Sum of Squares value
▶ Degrees of Freedom
▶ Mean Sum of Squares, which equals for each row
  
  Sum of Squares
  Degrees of Freedom

▶ ANOVA F-statistic (will explain)
▶ \( p \) -value (will explain)
Common to organize this info in an ANOVA table, which includes:

- Source of variation: (1) Between Group, (2) Within Group, or (3) Total
- Sum of Squares value
- Degrees of Freedom
- Mean Sum of Squares, which equals for each row
- ANOVA F-statistic (will explain)
- $p$-value (will explain)
ANOVA Table

Common to organize this info in an ANOVA table, which includes:

- Source of variation: (1) Between Group, (2) Within Group, or (3) Total
ANOVA Table

Common to organize this info in an ANOVA table, which includes:

- Source of variation: (1) Between Group, (2) Within Group, or (3) Total
- Sum of Squares value
ANOVA Table

Common to organize this info in an **ANOVA table**, which includes:

- Source of variation: (1) Between Group, (2) Within Group, or (3) Total
- Sum of Squares value
- Degrees of Freedom
Common to organize this info in an ANOVA table, which includes:

- Source of variation: (1) Between Group, (2) Within Group, or (3) Total
- Sum of Squares value
- Degrees of Freedom
- Mean Sum of Squares, which equals for each row
Common to organize this info in an ANOVA table, which includes:

- Source of variation: (1) Between Group, (2) Within Group, or (3) Total
- Sum of Squares value
- Degrees of Freedom
- Mean Sum of Squares, which equals for each row

\[
\frac{\text{Sum of Squares}}{\text{Degrees of Freedom}}
\]
Common to organize this info in an ANOVA table, which includes:

- Source of variation: (1) Between Group, (2) Within Group, or (3) Total
- Sum of Squares value
- Degrees of Freedom
- Mean Sum of Squares, which equals for each row

\[
\frac{\text{Sum of Squares}}{\text{Degrees of Freedom}}
\]

- ANOVA F-statistic (will explain)
ANOVA Table

Common to organize this info in an ANOVA table, which includes:
- Source of variation: (1) Between Group, (2) Within Group, or (3) Total
- Sum of Squares value
- Degrees of Freedom
- Mean Sum of Squares, which equals for each row

\[
\frac{\text{Sum of Squares}}{\text{Degrees of Freedom}}
\]

- ANOVA F-statistic (will explain)
- A \( p \)-value (will explain)
ANOVA Table

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean SS</th>
<th>F</th>
<th>-stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## ANOVA Table

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean SS</th>
<th>$F$-stat</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>Sum of Squares</td>
<td>df</td>
<td>Mean SS</td>
<td>F-stat</td>
<td>p-value</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------------------------</td>
<td>----</td>
<td>---------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>Between</td>
<td>$\sum n_i (\bar{X}_i - \bar{X})^2$</td>
<td></td>
<td>$\frac{\sum n_i (\bar{X}_i - \bar{X})^2}{\sum n_i}$</td>
<td>$\frac{\sum n_i (\bar{X}_i - \bar{X})^2}{\sum n_i}$</td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>$\sum \sum (X_{ij} - \bar{X}_i)^2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$\sum \sum (X_{ij} - \bar{X})^2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## ANOVA Table

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean SS</th>
<th>F-stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>$\sum n_i (\bar{X}_i - \bar{X})^2$</td>
<td>$k - 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>$\sum \sum (X_{ij} - \bar{X}_i)^2$</td>
<td>$n - k$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$\sum \sum (X_{ij} - \bar{X})^2$</td>
<td>$n - 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### ANOVA Table

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean SS</th>
<th>F-stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>11891.39</td>
<td>$k - 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>6701.02</td>
<td>$n - k$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>18592.41</td>
<td>$n - 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>Sum of Squares</td>
<td>df</td>
<td>Mean SS</td>
<td>F-stat</td>
<td>p-value</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------</td>
<td>----</td>
<td>---------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>Between</td>
<td>11891.39</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>6701.02</td>
<td>187</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>18592.41</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## ANOVA Table

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean SS</th>
<th>$F$-stat</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>11891.39</td>
<td>5</td>
<td>2378.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>6701.02</td>
<td>187</td>
<td>35.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>18592.41</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
To Conduct ANOVA Test

▶ Remember null hypothesis:
\[ \mu_1 = \mu_2 = \cdots = \mu_k \]

▶ And let's further assume groups have same population standard deviation, \( \sigma \)

▶ If null is true \( \rightarrow \) every group's \( X \)s come from same distribution:
\[ X_{ij} \sim (\mu, \sigma^2) \]

▶ But if null is not true \( \rightarrow \) each group's \( X \)s come from different distributions:
\[ X_{ij} \sim (\mu_i, \sigma^2) \]
To Conduct ANOVA Test

- Remember null hypothesis: $\mu_1 = \mu_2 = \ldots = \mu_k$
To Conduct ANOVA Test

- Remember null hypothesis: $\mu_1 = \mu_2 = \ldots = \mu_k$
- And let’s further assume groups have same population standard deviation, $\sigma$
To Conduct ANOVA Test

- Remember null hypothesis: $\mu_1 = \mu_2 = \ldots = \mu_k$
- And let’s further assume groups have same population standard deviation, $\sigma$
- If null is true $\rightarrow$ every group’s $X$s come from same distribution:
To Conduct ANOVA Test

- Remember null hypothesis: $\mu_1 = \mu_2 = \ldots = \mu_k$
- And let’s further assume groups have same population standard deviation, $\sigma$
- If null is true $\rightarrow$ every group’s $X$s come from same distribution:
  $$X_{ij} \sim (\mu, \sigma^2)$$
To Conduct ANOVA Test

- Remember null hypothesis: $\mu_1 = \mu_2 = ... = \mu_k$
- And let’s further assume groups have same population standard deviation, $\sigma$
- If null is true $\rightarrow$ every group’s $X$s come from same distribution:
  \[ X_{ij} \sim (\mu, \sigma^2) \]
- But if null is not true $\rightarrow$ each group’s $X$s come from different distributions:
To Conduct ANOVA Test

- Remember null hypothesis: $\mu_1 = \mu_2 = \ldots = \mu_k$
- And let’s further assume groups have same population standard deviation, $\sigma$
- If null is true $\rightarrow$ every group’s $X$s come from same distribution:
  $$X_{ij} \sim (\mu, \sigma^2)$$
- But if null is not true $\rightarrow$ each group’s $X$s come from different distributions:
  $$X_{ij} \sim (\mu_i, \sigma^2)$$
To Conduct ANOVA Test

Here is the key intuition:

1. Between group variability:
   - If null is true → only source of variance is population variability (so $\sigma^2$)
   - If null is false → $\bar{X}_i$ from different groups, you have variation come from differences in means (b/c $\mu_i$'s vary) as well as population variability ($\sigma^2$)

2. If null false → between group error larger than if null true

3. Within group variability:
   - Unaffected by null being true or false
   - Should be around $\sigma^2$ (if same across groups)

→ If null true, within group error and between group error should be close together
→ If null false, between group error $> \text{within group error}$, reflecting the fact that $\mu_i$ varies
To Conduct ANOVA Test

Here is the key intuition:

(1) Between group variability:
- If null is true → only source of variance is population variability (so $\sigma^2$)
- If null is false → $\bar{X}_i$ from different groups, you have variation come from differences in means (b/c $\mu_i$’s vary) as well as population variability ($\sigma^2$)
- If null false → between group error larger than if null true

(2) Within group variability:
- Unaffected by null being true or false
- Should be around $\sigma^2$ (if same across groups)
- → If null true, within group error and between group error should be close together
- → If null false, between group error $\gg$ within group error, reflecting the fact that $\mu_i$ varies
To Conduct ANOVA Test

Here is the key intuition:

1. Between group variability:
   - If null is true → only source of variance is population variability ($\sigma^2$).
   - If null is false → $\bar{X}_i$ from different groups, you have variation come from differences in means (b/c $\mu_i$'s vary) as well as population variability ($\sigma^2$).
   - If null false → between group error larger than if null true.

2. Within group variability:
   - Unaffected by null being true or false.
   - Should be around $\sigma^2$ (if same across groups).
   - If null true, within group error and between group error should be close together.
   - If null false, between group error $> \sigma^2$, reflecting the fact that $\mu_i$ varies.
To Conduct ANOVA Test

Here is the key intuition:

- (1) Between group variability:
  - If null is true → only source of variance is population variability (so $\sigma^2$).

- (2) Within group variability:
  - Unaffected by null being true or false.
  - Should be around $\sigma^2$ (if same across groups).
  - If null true, within group error and between group error should be close together.
  - If null false, between group error $>\text{within group error}$, reflecting the fact that $\mu_i$ varies.
To Conduct ANOVA Test

Here is the key intuition:

1. Between group variability:
   - If null is true → only source of variance is population variability (so $\sigma^2$)
   - If null is false → $\bar{X}_i$ from different groups, you have variation come from differences in means (b/c $\mu_i$'s vary) as well as population variability ($\sigma^2$)
To Conduct ANOVA Test

Here is the key intuition:

1. Between group variability:
   - If null is true → only source of variance is population variability (so \( \sigma^2 \))
   - If null is false → \( \bar{X}_i \) from different groups, you have variation come from differences in means (b/c \( \mu_i \)'s vary) as well as population variability (\( \sigma^2 \))
   - If null false → between group error larger than if null true
To Conduct ANOVA Test

Here is the key intuition:

▶ (1) Between group variability:
  ▶ If null is true → only source of variance is population variability (so $\sigma^2$)
  ▶ If null is false → $\bar{X}_i$ from different groups, you have variation come from differences in means (b/c $\mu_i$’s vary) as well as population variability ($\sigma^2$)
  ▶ If null false → between group error larger than if null true

▶ (2) Within group variability:
To Conduct ANOVA Test

Here is the key intuition:

▶ (1) Between group variability:
  ▶ If null is true → only source of variance is population variability (so $\sigma^2$)
  ▶ If null is false → $\bar{X}_i$ from different groups, you have variation come from differences in means (b/c $\mu_i$’s vary) as well as population variability ($\sigma^2$)
  ▶ If null false → between group error larger than if null true

▶ (2) Within group variability:
  ▶ Unaffected by null being true or false
To Conduct ANOVA Test

Here is the key intuition:

- **(1) Between group variability:**
  - If null is **true** → only source of variance is population variability (so $\sigma^2$)
  - If null is **false** → $\bar{X}_i$ from different groups, you have variation come from differences in means (b/c $\mu_i$’s vary) as well as population variability ($\sigma^2$)
  - If null false → between group error larger than if null true

- **(2) Within group variability:**
  - Unaffected by null being true or false
  - Should be around $\sigma^2$ (if same across groups)
To Conduct ANOVA Test

Here is the key intuition:

1. Between group variability:
   - If null is true → only source of variance is population variability (so $\sigma^2$)
   - If null is false → $\bar{X}_i$ from different groups, you have variation come from differences in means (b/c $\mu_i$’s vary) as well as population variability ($\sigma^2$)
   - If null false → between group error larger than if null true

2. Within group variability:
   - Unaffected by null being true or false
   - Should be around $\sigma^2$ (if same across groups)

→ If null true, within group error and between group error should be close together
To Conduct ANOVA Test

Here is the key intuition:

1. Between group variability:
   - If null is true → only source of variance is population variability (so $\sigma^2$)
   - If null is false → $\bar{X}_i$ from different groups, you have variation come from differences in means (b/c $\mu_i$’s vary) as well as population variability ($\sigma^2$)
   - If null false → between group error larger than if null true

2. Within group variability:
   - Unaffected by null being true or false
   - Should be around $\sigma^2$ (if same across groups)

→ If null true, within group error and between group error should be close together

→ If null false, between group error $\geq$ within group error, reflecting the fact that $\mu_i$ varies
To Conduct ANOVA Test

This intuition gives us the ANOVA test (sometimes called ANOVA F-Test)

Part of a larger suite of tests that compare a test statistic to the F probability distribution (F for R.A. Fisher)

(F distribution is the ratio of two \( \chi^2 \)-distributed RVs, where \( \chi^2 \) are squares of Normally distributed RVs – see Wikipedia)

F distribution completely determined by its two degrees of freedom (one corresponding to the numerator, the other to the denominator)
To Conduct ANOVA Test

- This intuition gives us the ANOVA test (sometimes called ANOVA F-Test)
To Conduct ANOVA Test

- This intuition gives us the ANOVA test (sometimes called ANOVA $F$-Test)
- Part of a larger suite of tests that compare a test statistic to the $F$ probability distribution ($F$ for R.A. Fisher)
To Conduct ANOVA Test

- This intuition gives us the ANOVA test (sometimes called ANOVA $F$-Test)
- Part of a larger suite of tests that compare a test statistic to the $F$ probability distribution ($F$ for R.A. Fisher)
- ($F$ distribution is the ratio of two $\chi^2$-distributed RVs, where $\chi^2$ are squares of Normally distributed RVs – see Wikipedia)
To Conduct ANOVA Test

- This intuition gives us the ANOVA test (sometimes called ANOVA \(F\)-Test)
- Part of a larger suite of tests that compare a test statistic to the \(F\) probability distribution (\(F\) for R.A. Fisher)
- \((F\) distribution is the ratio of two \(\chi^2\)-distributed RVs, where \(\chi^2\) are squares of Normally distributed RVs – see Wikipedia)
- \(F\) distribution completely determined by its two degrees of freedom (one corresponding to the numerator, the other to the denominator)
To Conduct ANOVA Test

Examples of F Distributions

- $d_1 = 1$ $d_2 = 1$
- $d_1 = 2$ $d_2 = 4$
- $d_1 = 5$ $d_2 = 4$
- $d_1 = 10$ $d_2 = 20$
To Conduct ANOVA Test

ANOVA F-test compares between group variance to within group variance. Large difference between two → reason to reject null. Why? Suggests mean between group differs. Only small (or no) difference between two → less reason to reject null. Why? Suggests mean between group is not differing (or not differing that much).
To Conduct ANOVA Test

- ANOVA $F$-test compares between group variance to within group variance.
To Conduct ANOVA Test

- ANOVA F-test compares between group variance to within group variance
- Large difference between two $\rightarrow$ reason to reject null
To Conduct ANOVA Test

- ANOVA $F$-test compares between group variance to within group variance
- Large difference between two $\rightarrow$ reason to reject null
  - Why? Suggests mean between group differs
To Conduct ANOVA Test

- ANOVA $F$-test compares between group variance to within group variance
- Large difference between two $\rightarrow$ reason to reject null
  - Why? Suggests mean between group differs
- Only small (or no) difference between two $\rightarrow$ less reason to reject null
To Conduct ANOVA Test

- ANOVA $F$-test compares between group variance to within group variance
- Large difference between two $\rightarrow$ reason to reject null
  - Why? Suggests mean between group differs
- Only small (or no) difference between two $\rightarrow$ less reason to reject null
  - Why? Suggests mean between group is not differing (or not differing that much)
To Conduct ANOVA Test

Step 1: Null and alternative hypotheses

- $H_0$: $\mu_1 = \mu_2 = \ldots = \mu_k$
- $H_1$: At least one unequal

Step 2: Collect data

Step 3: Calculate ANOVA $F$-statistic

Relies on previous intuition:

$$F = \frac{\text{Mean Sum of Squares Between Groups}}{\text{Mean Sum of Squares Within Groups}} = \frac{\text{MS}_b}{\text{MS}_w}$$

Closer $F$-statistic is to 1 $\rightarrow$ the more likely the null ($H_0$: $\mu_1 = \mu_2 = \ldots = \mu_k$) is to be true
To Conduct ANOVA Test

- Step 1: Null and alternative hypotheses

\[ H_0: \mu_1 = \mu_2 = \ldots = \mu_k \]

\[ H_1: \text{At least one unequal} \]

- Step 2: Collect data

- Step 3: Calculate ANOVA F-statistic

Relies on previous intuition: \[ F = \frac{\text{Mean Sum of Squares Between Groups}}{\text{Mean Sum of Squares Within Groups}} = \frac{MS_b}{MS_w} \]

Closer \( F \)-statistic is to 1 \( \rightarrow \) the more likely the null \( (H_0: \mu_1 = \mu_2 = \ldots = \mu_k) \) is to be true
To Conduct ANOVA Test

- Step 1: Null and alternative hypotheses
  - \( H_0 : \mu_1 = \mu_2 = ... = \mu_k \)
To Conduct ANOVA Test

- **Step 1: Null and alternative hypotheses**
  - $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$
  - $H_1 : \text{At least one unequal}$
To Conduct ANOVA Test

- **Step 1: Null and alternative hypotheses**
  - \( H_0 : \mu_1 = \mu_2 = \ldots = \mu_k \)
  - \( H_1 : \) At least one unequal

- **Step 2: Collect data**

- **Step 3: Calculate ANOVA**
  - Relies on previous intuition:
    \[
    F\text{-statistic} = \frac{\text{Mean Sum of Squares Between Groups}}{\text{Mean Sum of Squares Within Groups}} = \frac{MS_b}{MS_w}
    \]
  - Closer \( F\)-statistic is to 1 → the more likely the null \((H_0 : \mu_1 = \mu_2 = \ldots = \mu_k)\) is to be true
To Conduct ANOVA Test

▶ Step 1: Null and alternative hypotheses
  ▶ $H_0 : \mu_1 = \mu_2 = ... = \mu_k$
  ▶ $H_1 : \text{At least one unequal}$

▶ Step 2: Collect data

▶ Step 3: Calculate ANOVA $F$-statistic

$F$-statistic = \frac{\text{Mean Sum of Squares Between Groups}}{\text{Mean Sum of Squares Within Groups}}$

The closer the $F$-statistic is to 1, the more likely the null ($H_0 : \mu_1 = \mu_2 = ... = \mu_k$) is to be true.
To Conduct ANOVA Test

- **Step 1:** Null and alternative hypotheses
  - $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$
  - $H_1 : \text{At least one unequal}$

- **Step 2:** Collect data

- **Step 3:** Calculate ANOVA $F$-statistic

- Relies on previous intuition:
  
  \[ F = \frac{\text{Mean Sum of Squares Between Groups}}{\text{Mean Sum of Squares Within Groups}} = \frac{MS_b}{MS_w} \]

  Closer $F$-statistic is to 1 $\rightarrow$ the more likely the null ($H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$) is to be true
To Conduct ANOVA Test

- **Step 1: Null and alternative hypotheses**
  - $H_0 : \mu_1 = \mu_2 = ... = \mu_k$
  - $H_1 : \text{At least one unequal}$

- **Step 2: Collect data**

- **Step 3: Calculate ANOVA $F$-statistic**

- Relies on previous intuition:

$$F = \frac{\text{Mean Sum of Squares Between Groups}}{\text{Mean Sum of Squares Within Groups}}$$

$$= \frac{MS_b}{MS_w}$$
To Conduct ANOVA Test

- Step 1: Null and alternative hypotheses
  - $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$
  - $H_1 : \text{At least one unequal}$
- Step 2: Collect data
- Step 3: Calculate ANOVA $F$-statistic
- Relies on previous intuition:

\[
F = \frac{\text{Mean Sum of Squares Between Groups}}{\text{Mean Sum of Squares Within Groups}} = \frac{MS_b}{MS_w}
\]

- Closer $F$-statistic is to 1 $\rightarrow$ the more likely the null ($H_0 : \mu_1 = \mu_2 = \ldots = \mu_6$) is to be true
To Conduct ANOVA Test

In our example, $MS_B = 2378.28$ and $MS_W = 35.83$

$F = \frac{MS_B}{MS_W} = \frac{2378.28}{35.83} = 66.37$

Looks like between group error $>\text{within group error}$

Means that more variance between groups than within groups
→ intuition counsels against null
To Conduct ANOVA Test

▶ In our example, $MS_B = 2378.28$ and $MS_W = 35.83$
To Conduct ANOVA Test

- In our example, $MS_B = 2378.28$ and $MS_W = 35.83$

\[
F = \frac{2378.28}{35.83} = 66.37
\]
To Conduct ANOVA Test

▶ In our example, $MS_B = 2378.28$ and $MS_W = 35.83$

$$F = \frac{2378.28}{35.83} = 66.37$$

▶ Looks like between group error $> \text{within group error}$
To Conduct ANOVA Test

In our example, $MS_B = 2378.28$ and $MS_W = 35.83$

$$F = \frac{2378.28}{35.83} = 66.37$$

- Looks like between group error $>$ within group error
- Means that more variance between groups than within groups
  $\rightarrow$ intuition counsels against null
## ANOVA Table

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean SS</th>
<th>F-Stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>11891.39</td>
<td>5</td>
<td>2378.28</td>
<td>66.37</td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>6701.02</td>
<td>187</td>
<td>35.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>18592.41</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## ANOVA Table

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean SS</th>
<th>$F$-stat</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>11891.39</td>
<td>5</td>
<td>2378.28</td>
<td>66.37</td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>6701.02</td>
<td>187</td>
<td>35.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>18592.41</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The F-Test

Step 4: Calculate the p-value associated with the F-statistic using the F-distribution. Can use F-tables or STATA. Need degrees of freedom from the numerator (df1) and denominator (df2)

\[
\text{df} = (k - 1, n - k)
\]

\[
= (6 - 1, 193 - 6)
\]

\[
= (5, 187)
\]

And use this info to determine appropriate F-distribution.
The F-Test

- Step 4: Calculate $p$-value associated with $F$-statistic using the $F$-distribution

$\eta = (k - 1, n - k) = (6 - 1, 193 - 6) = (5, 187)$
The F-Test

- Step 4: Calculate $p$-value associated with $F$-statistic using the $F$-distribution
- Can use $F$-tables or STATA

\[
df = (k - 1, n - k) = (5, 187)
\]
The F-Test

- Step 4: Calculate $p$-value associated with $F$-statistic using the $F$-distribution
- Can use $F$-tables or STATA
- Need degrees of freedom from the numerator (df1) and denominator (df2)
The F-Test

▶ Step 4: Calculate $p$-value associated with $F$-statistic using the $F$-distribution
▶ Can use $F$-tables or STATA
▶ Need degrees of freedom from the numerator (df1) and denominator (df2)

$$df = (k - 1, n - k)$$
$$= (6 - 1, 193 - 6)$$
$$= (5, 187)$$
The F-Test

- Step 4: Calculate \( p \)-value associated with \( F \)-statistic using the \( F \)-distribution
- Can use \( F \)-tables or STATA
- Need degrees of freedom from the numerator (\( df_1 \)) and denominator (\( df_2 \))

\[
df = (k - 1, n - k)
\]

\[
= (6 - 1, 193 - 6)
\]

\[
= (5, 187)
\]

- And use this info to determine appropriate \( F \)-distribution
The F-Test
The F-Test

Probability $p$

$F^*$
# Table E

F critical values (continued)

<table>
<thead>
<tr>
<th>p</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>.100</td>
<td>3.01</td>
<td>2.62</td>
<td>2.42</td>
<td>2.29</td>
<td>2.20</td>
<td>2.13</td>
<td>2.08</td>
<td>2.04</td>
<td>2.00</td>
</tr>
<tr>
<td>.050</td>
<td>4.41</td>
<td>3.55</td>
<td>3.16</td>
<td>2.93</td>
<td>2.77</td>
<td>2.66</td>
<td>2.58</td>
<td>2.51</td>
<td>2.46</td>
</tr>
<tr>
<td>.025</td>
<td>5.98</td>
<td>4.56</td>
<td>3.95</td>
<td>3.61</td>
<td>3.38</td>
<td>3.22</td>
<td>3.10</td>
<td>3.01</td>
<td>2.93</td>
</tr>
<tr>
<td>.010</td>
<td>8.29</td>
<td>6.01</td>
<td>5.09</td>
<td>4.58</td>
<td>4.25</td>
<td>4.01</td>
<td>3.84</td>
<td>3.71</td>
<td>3.60</td>
</tr>
<tr>
<td>.001</td>
<td>15.38</td>
<td>10.39</td>
<td>8.49</td>
<td>7.46</td>
<td>6.81</td>
<td>6.35</td>
<td>6.02</td>
<td>5.76</td>
<td>5.56</td>
</tr>
<tr>
<td>.100</td>
<td>2.99</td>
<td>2.61</td>
<td>2.40</td>
<td>2.27</td>
<td>2.18</td>
<td>2.11</td>
<td>2.06</td>
<td>2.02</td>
<td>1.98</td>
</tr>
<tr>
<td>.050</td>
<td>4.38</td>
<td>3.52</td>
<td>3.13</td>
<td>2.90</td>
<td>2.74</td>
<td>2.63</td>
<td>2.54</td>
<td>2.48</td>
<td>2.42</td>
</tr>
<tr>
<td>.025</td>
<td>5.92</td>
<td>4.51</td>
<td>3.90</td>
<td>3.56</td>
<td>3.33</td>
<td>3.17</td>
<td>3.05</td>
<td>2.96</td>
<td>2.88</td>
</tr>
<tr>
<td>.010</td>
<td>8.18</td>
<td>5.93</td>
<td>5.01</td>
<td>4.50</td>
<td>4.17</td>
<td>3.94</td>
<td>3.77</td>
<td>3.63</td>
<td>3.52</td>
</tr>
<tr>
<td>.001</td>
<td>15.08</td>
<td>10.16</td>
<td>8.28</td>
<td>7.27</td>
<td>6.62</td>
<td>6.18</td>
<td>5.85</td>
<td>5.59</td>
<td>5.39</td>
</tr>
<tr>
<td>.100</td>
<td>2.97</td>
<td>2.59</td>
<td>2.38</td>
<td>2.25</td>
<td>2.16</td>
<td>2.09</td>
<td>2.04</td>
<td>2.00</td>
<td>1.96</td>
</tr>
<tr>
<td>.050</td>
<td>4.35</td>
<td>3.49</td>
<td>3.10</td>
<td>2.87</td>
<td>2.71</td>
<td>2.60</td>
<td>2.51</td>
<td>2.45</td>
<td>2.39</td>
</tr>
<tr>
<td>.025</td>
<td>5.87</td>
<td>4.46</td>
<td>3.86</td>
<td>3.51</td>
<td>3.29</td>
<td>3.13</td>
<td>3.01</td>
<td>2.91</td>
<td>2.84</td>
</tr>
<tr>
<td>.010</td>
<td>8.10</td>
<td>5.85</td>
<td>4.94</td>
<td>4.43</td>
<td>4.10</td>
<td>3.87</td>
<td>3.70</td>
<td>3.56</td>
<td>3.46</td>
</tr>
<tr>
<td>.001</td>
<td>14.82</td>
<td>9.95</td>
<td>8.10</td>
<td>7.10</td>
<td>6.46</td>
<td>6.02</td>
<td>5.69</td>
<td>5.44</td>
<td>5.24</td>
</tr>
<tr>
<td>.100</td>
<td>2.96</td>
<td>2.57</td>
<td>2.36</td>
<td>2.23</td>
<td>2.14</td>
<td>2.08</td>
<td>2.02</td>
<td>1.98</td>
<td>1.95</td>
</tr>
<tr>
<td>.050</td>
<td>4.32</td>
<td>3.47</td>
<td>3.07</td>
<td>2.84</td>
<td>2.68</td>
<td>2.57</td>
<td>2.49</td>
<td>2.42</td>
<td>2.37</td>
</tr>
<tr>
<td>.025</td>
<td>5.83</td>
<td>4.42</td>
<td>3.82</td>
<td>3.48</td>
<td>3.25</td>
<td>3.09</td>
<td>2.97</td>
<td>2.87</td>
<td>2.80</td>
</tr>
<tr>
<td>.010</td>
<td>8.02</td>
<td>5.78</td>
<td>4.87</td>
<td>4.37</td>
<td>4.04</td>
<td>3.81</td>
<td>3.64</td>
<td>3.51</td>
<td>3.40</td>
</tr>
<tr>
<td>.001</td>
<td>14.59</td>
<td>9.77</td>
<td>7.94</td>
<td>6.95</td>
<td>6.32</td>
<td>5.88</td>
<td>5.56</td>
<td>5.31</td>
<td>5.11</td>
</tr>
<tr>
<td>.100</td>
<td>2.95</td>
<td>2.56</td>
<td>2.35</td>
<td>2.22</td>
<td>2.13</td>
<td>2.06</td>
<td>2.01</td>
<td>1.97</td>
<td>1.93</td>
</tr>
<tr>
<td>.050</td>
<td>4.30</td>
<td>3.44</td>
<td>3.05</td>
<td>2.82</td>
<td>2.66</td>
<td>2.55</td>
<td>2.46</td>
<td>2.40</td>
<td>2.34</td>
</tr>
<tr>
<td>.025</td>
<td>5.79</td>
<td>4.38</td>
<td>3.78</td>
<td>3.44</td>
<td>3.22</td>
<td>3.05</td>
<td>2.93</td>
<td>2.84</td>
<td>2.76</td>
</tr>
<tr>
<td>.010</td>
<td>7.95</td>
<td>5.72</td>
<td>4.82</td>
<td>4.31</td>
<td>3.99</td>
<td>3.76</td>
<td>3.59</td>
<td>3.45</td>
<td>3.35</td>
</tr>
<tr>
<td>.001</td>
<td>14.38</td>
<td>9.61</td>
<td>7.80</td>
<td>6.81</td>
<td>6.19</td>
<td>5.76</td>
<td>5.44</td>
<td>5.19</td>
<td>4.99</td>
</tr>
</tbody>
</table>
The F-Test

To finish up the ANOVA table:

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean SS</th>
<th>F-stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>11891.39</td>
<td>5</td>
<td>2378.28</td>
<td>66.37</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Within</td>
<td>6701.02</td>
<td>187</td>
<td>35.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>18592.41</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The F-Test

- Step 4: Calculate \( p \)-value

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean SS</th>
<th>( F )-stat</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>11891.39</td>
<td>5</td>
<td>2378.28</td>
<td>66.37</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Within</td>
<td>6701.02</td>
<td>187</td>
<td>35.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>18592.41</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The F-Test

- Step 4: Calculate $p$-value
- $p$-value < 0.0001

To finish up the ANOVA table:

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean SS</th>
<th>$F$-stat</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>11891.39</td>
<td>5</td>
<td>2378.28</td>
<td>66.37</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Within</td>
<td>6701.02</td>
<td>187</td>
<td>35.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>18592.41</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The F-Test

- Step 4: Calculate $p$-value
- $p$-value < 0.0001
- To finish up the ANOVA table:
The F-Test

- Step 4: Calculate $p$-value
- $p$-value < 0.0001
- To finish up the ANOVA table:

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean SS</th>
<th>$F$-stat</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>11891.39</td>
<td>5</td>
<td>2378.28</td>
<td>66.37</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Within</td>
<td>6701.02</td>
<td>187</td>
<td>35.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>18592.41</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The F-Test

Step 5: Interpretation

- With $p$-value < 0.01, strong evidence to reject $H_0$: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_6$.
- Evidence that at least one of the continent's mean life expectancies differs from another.
The F-Test

- Step 5: Interpretation $p$-value and decide whether to reject null

  With $p$-value < 0.01, strong evidence to reject $H_0$: $\mu_1 = \mu_2 = \mu_3 = ... = \mu_6$

  Evidence that at least one of the continent’s mean life expectancies differs from another.
The F-Test

Step 5: Interpretation $p$-value and decide whether to reject null

- With $p$-value < 0.01, strong evidence to reject $H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots \mu_6$

The F-Test

- Step 5: Interpretation $p$-value and decide whether to reject null
- With $p$-value $< 0.01$, strong evidence to reject $H_0: \mu_1 = \mu_2 = \mu_3 = \ldots \mu_6$
- Evidence that at least one of the continent’s mean life expectancies differs from another
ANOVA F-test Assumptions

1. Groups have population distributions that are normal
   ▶ F-test not as accurate, but approximate when populations non-normal
   ▶ Becomes better with large sample sizes

2. Groups have identical standard deviations
   ▶ Modest departures ok
   ▶ Check largest standard deviation no more than twice that of the smallest standard deviation
   ▶ Note: Very robust to this if sample size across groups is same
ANOVA F-test Assumptions

Requires 2 strong assumptions

1. Groups have population distributions that are normal
   ▶ F-test not as accurate, but approximate when populations non-normal
   ▶ Becomes better with large sample sizes

2. Groups have identical standard deviations
   ▶ Modest departures ok
   ▶ Check largest standard deviation no more than twice that of the smallest standard deviation
   ▶ Note: Very robust to this if sample size across groups is same
ANOVA F-test Assumptions

Requires 2 strong assumptions

1. Groups have population distributions that are normal
ANOVA F-test Assumptions

Requires 2 strong assumptions

1. Groups have population distributions that are normal
   - $F$-test not as accurate, but approximate when populations non-normal

2. Groups have identical standard deviations
   - Modest departures ok
   - Check largest standard deviation no more than twice that of the smallest standard deviation
   - Note: Very robust to this if sample size across groups is same
ANOVA F-test Assumptions

Requires 2 strong assumptions

1. Groups have population distributions that are normal
   - $F$-test not as accurate, but approximate when populations non-normal
   - Becomes better with large sample sizes
ANOVA F-test Assumptions

Requires 2 strong assumptions

1. Groups have population distributions that are normal
   - $F$-test not as accurate, but approximate when populations non-normal
   - Becomes better with large sample sizes

2. Groups have identical standard deviations
ANOVA F-test Assumptions

Requires 2 strong assumptions

1. Groups have population distributions that are normal
   - $F$-test not as accurate, but approximate when populations non-normal
   - Becomes better with large sample sizes

2. Groups have identical standard deviations
   - Modest departures ok
ANOVA F-test Assumptions

Requires 2 strong assumptions

1. Groups have population distributions that are normal
   - $F$-test not as accurate, but approximate when populations non-normal
   - Becomes better with large sample sizes

2. Groups have identical standard deviations
   - Modest departures ok
   - Check largest standard deviation no more than twice that of the smallest standard deviation
ANOVA F-test Assumptions

Requires 2 strong assumptions

1. Groups have population distributions that are normal
   - $F$-test not as accurate, but approximate when populations non-normal
   - Becomes better with large sample sizes

2. Groups have identical standard deviations
   - Modest departures ok
   - Check largest standard deviation no more than twice that of the smallest standard deviation
   - Note: Very robust to this if sample size across groups is same
Multiple Comparisons

ANOVA $F$-test: Useful to test null that all population means equal to each other.

But all it really tests is whether significant difference between at least two means.

Next question: Which groups differ?

Could do bunch of pairwise tests but the overall error rate will still be high.

With 6 continents, we have 15 pairwise comparisons.

Probability of seeing at least one significant result due to chance is high (around 54%).

Instead: Do ANOVA and then do (ex post) comparisons between all means, correcting for multiple comparisons.

Multiple comparisons approach also useful in other contexts.
Multiple Comparisons

- ANOVA $F$-test: Useful to test null that all population means equal to each other

- But all it really tests is whether significant difference between at least two means

- Next question: Which groups differ?

- Could do bunch of pairwise tests but the overall error rate will still be high

- With 6 continents, we have 15 pairwise comparisons

- Probability of seeing at least one significant result due to chance is high (around 54%)

- Instead: Do ANOVA and then do (ex post) comparisons between all means, correcting for multiple comparisons

- Multiple comparisons approach also useful in other contexts
Multiple Comparisons

- **ANOVA F-test**: Useful to test null that *all* population means equal to each other.
- But all it really tests is whether significant difference between at least two means.
Multiple Comparisons

- ANOVA $F$-test: Useful to test null that all population means equal to each other
- But all it really tests is whether significant difference between at least two means
- Next question: Which groups differ?
Multiple Comparisons

- ANOVA $F$-test: Useful to test null that all population means equal to each other
- But all it really tests is whether significant difference between at least two means
- Next question: Which groups differ?
- Could do bunch of pairwise tests but the overall error rate will still be high
Multiple Comparisons

▶ ANOVA $F$-test: Useful to test null that all population means equal to each other
▶ But all it really tests is whether significant difference between at least two means
▶ Next question: Which groups differ?
▶ Could do bunch of pairwise tests but the overall error rate will still be high
  ▶ With 6 continents, we have 15 pairwise comparisons
Multiple Comparisons

- ANOVA $F$-test: Useful to test null that all population means equal to each other
- But all it really tests is whether significant difference between at least two means
- Next question: Which groups differ?
- Could do bunch of pairwise tests but the overall error rate will still be high
  - With 6 continents, we have 15 pairwise comparisons
  - Probability of seeing at least one significant result due to chance is high (around 54%)
Multiple Comparisons

- ANOVA $F$-test: Useful to test null that all population means equal to each other
- But all it really tests is whether significant difference between at least two means
- Next question: Which groups differ?
- Could do bunch of pairwise tests but the overall error rate will still be high
  - With 6 continents, we have 15 pairwise comparisons
  - Probability of seeing at least one significant result due to chance is high (around 54%)
- Instead: Do ANOVA and then do (ex post) comparisons between all means, correcting for **multiple comparisons**
Multiple Comparisons

- ANOVA $F$-test: Useful to test null that all population means equal to each other
- But all it really tests is whether significant difference between at least two means
- Next question: Which groups differ?
- Could do bunch of pairwise tests but the overall error rate will still be high
  - With 6 continents, we have 15 pairwise comparisons
  - Probability of seeing at least one significant result due to chance is high (around 54%)
- Instead: Do ANOVA and then do (ex post) comparisons between all means, correcting for multiple comparisons
- Multiple comparisons approach also useful in other contexts
Bonferroni Method of Multiple Comparisons

Bonferroni Method: Commonly used to compare groups after carrying out the global ANOVA test.

Spreads out the significance (say, 5%) across all the tests.

Ex) For our 15 comparisons, use $\alpha^* = 0.05/15 = 0.00333$ for each test instead of usual $\alpha = 0.05$.

Note: $\alpha^*$ known as Bonferroni correction.

Ex) Here, if we were calculating 15 difference-in-means confidence intervals, then redraws all 15 confidence intervals more conservatively so the probability that all intervals contain the 15 population differences is 95%.

As opposed to probability that each interval contains a specific difference is 5%.
Bonferroni Method of Multiple Comparisons

- **Bonferroni Method**: Commonly used to compare groups after carrying out the global ANOVA test.

  - Spreads out the significance (say, 5%) across all the tests.
  - Example: For our 15 comparisons, use $\alpha^* = \frac{0.05}{15} = 0.00333$ for each test instead of the usual $\alpha = 0.05$.
  - Note: $\alpha^*$ known as Bonferroni correction.
  - Example: Here, if we were calculating 15 difference-in-means confidence intervals, then redraws all 15 confidence intervals more conservatively so the probability that all intervals contain the 15 population differences is 95%.
  - As opposed to the probability that each interval contains a specific difference is 5%.
Bonferroni Method of Multiple Comparisons

- **Bonferroni Method**: Commonly used to compare groups after carrying out the global ANOVA test
- Spreads out the significance (say, 5%) across all the tests

\[
\alpha = \frac{0.05}{15} = 0.00333
\]

Note: \( \alpha \) known as Bonferroni correction

Ex) Here, if we were calculating 15 difference-in-means confidence intervals, then redraw all 15 confidence intervals more conservatively so the probability that all intervals contain the 15 population differences is 95%

As opposed to probability that each interval contains a specific difference is 5%
Bonferroni Method of Multiple Comparisons

- **Bonferroni Method**: Commonly used to compare groups after carrying out the global ANOVA test
- Spreads out the significance (say, 5%) across all the tests
- Ex) For our 15 comparisons, use $\alpha^* = 0.05/15 = 0.00333$ for each test instead of usual $\alpha = 0.05$
Bonferroni Method of Multiple Comparisons

- **Bonferroni Method**: Commonly used to compare groups after carrying out the global ANOVA test
- Spreads out the significance (say, 5%) across all the tests
- Ex) For our 15 comparisons, use $\alpha^* = 0.05/15 = 0.00333$ for each test instead of usual $\alpha = 0.05$
- Note: $\alpha^*$ known as **Bonferroni correction**
Bonferroni Method of Multiple Comparisons

- **Bonferroni Method**: Commonly used to compare groups after carrying out the global ANOVA test
- Spreads out the significance (say, 5%) across all the tests
- Ex) For our 15 comparisons, use $\alpha^* = 0.05/15 = 0.00333$ for each test instead of usual $\alpha = 0.05$
- Note: $\alpha^*$ known as Bonferroni correction
- Ex) Here, if we were calculating 15 difference-in-means confidence intervals, then redraws all 15 confidence intervals more conservatively so the probability that *all* intervals contain the 15 population differences is 95%
Bonferroni Method of Multiple Comparisons

- **Bonferroni Method**: Commonly used to compare groups after carrying out the global ANOVA test
- Spreads out the significance (say, 5%) across all the tests
- Ex) For our 15 comparisons, use \( \alpha^* = \frac{0.05}{15} = 0.00333 \) for each test instead of usual \( \alpha = 0.05 \)
- **Note**: \( \alpha^* \) known as Bonferroni correction
- Ex) Here, if we were calculating 15 difference-in-means confidence intervals, then redraws all 15 confidence intervals more conservatively so the probability that all intervals contain the 15 population differences is 95%
  - As opposed to probability that each interval contains a specific difference is 5%
Suppose we wanted to compare Groups 3 and 2.

In calculating the appropriate standard error, use the population pooled standard deviation

\[ SE(\bar{X}_3 - \bar{X}_2) = \sqrt{s_p^2 \left( \frac{1}{n_3} + \frac{1}{n_2} \right)} \]

Can estimate pooled standard deviation using within-group variability, \( MS_w \) (also called MSE):

\[ SE(\bar{X}_3 - \bar{X}_2) = \sqrt{MS_w \left( \frac{1}{n_3} + \frac{1}{n_2} \right)} \]
Bonferroni Method of Multiple Comparisons

- Suppose we wanted to compare Groups 3 and 2
Suppose we wanted to compare Groups 3 and 2

In calculating the appropriate standard error, use the population pooled standard deviation
Bonferroni Method of Multiple Comparisons

- Suppose we wanted to compare Groups 3 and 2
- In calculating the appropriate standard error, use the population pooled standard deviation

\[
SE(\bar{X}_3 - \bar{X}_2) = \sqrt{s_p^2 \left( \frac{1}{n_3} + \frac{1}{n_2} \right)}
\]
Bonferroni Method of Multiple Comparisons

- Suppose we wanted to compare Groups 3 and 2
- In calculating the appropriate standard error, use the population pooled standard deviation

\[
SE(\bar{X}_3 - \bar{X}_2) = \sqrt{s^2_p \left( \frac{1}{n_3} + \frac{1}{n_2} \right)}
\]

- Can estimate pooled standard deviation using within-group variability, \(MS_w\) (also called MSE):
Suppose we wanted to compare Groups 3 and 2

In calculating the appropriate standard error, use the population pooled standard deviation

\[
SE(\bar{X}_3 - \bar{X}_2) = \sqrt{s_p^2 \left( \frac{1}{n_3} + \frac{1}{n_2} \right)}
\]

Can estimate pooled standard deviation using within-group variability, \( MS_w \) (also called MSE):

\[
SE(\bar{X}_3 - \bar{X}_2) = \sqrt{MS_w \left( \frac{1}{n_3} + \frac{1}{n_2} \right)}
\]
Bonferroni Method of Multiple Comparisons

\[
\bar{X}_3 - \bar{X}_2 \pm t\left(n-k, \alpha^*\right) \sqrt{\frac{1}{n_3} + \frac{1}{n_2}} \sqrt{\text{MS}_w}
\]

For our analysis:

\[
78.11 - 72.02 \pm t\left(193, \frac{0.00333}{2}\right) \sqrt{35.83} \left(\frac{1}{42} + \frac{1}{50}\right)
\]

Which simplifies to

\[
= 6.09 \pm 3.676 \approx [2.41, 9.77]
\]

Interpret as a standard confidence interval, except note that this corrects for multiple comparisons (spreads out \(\alpha\) for all CIs calculated)
Bonferroni Method of Multiple Comparisons

- Use this SE with Bonferroni correction:

\[
\bar{X}_3 - \bar{X}_2 \pm t\left(n-k, \alpha^*\right) \sqrt{\frac{MS_w}{\left(\frac{1}{n_3} + \frac{1}{n_2}\right)}}
\]

For our analysis:

\[
78.11 - 72.02 \pm t\left(193, 0.00333 / 2\right) \sqrt{35.83 \left(\frac{1}{42} + \frac{1}{50}\right)}
\]

Which simplifies to

\[
= 6.09 \pm 2.934 \times \sqrt{1.57} = 6.09 \pm 3.676 = [2.41, 9.77]
\]

Interpret as a standard confidence interval, except note that this corrects for multiple comparisons (spreads out \(\alpha\) for all CIs calculated).
Bonferroni Method of Multiple Comparisons

- Use this SE with Bonferroni correction:

\[ \bar{X}_3 - \bar{X}_2 \pm t_{(n-k, \alpha^* \frac{1}{2})} \sqrt{MS_w \left( \frac{1}{n_3} + \frac{1}{n_2} \right)} \]

- For our analysis:

\[ 78.11 - 72.02 \pm t_{(193, 0.00333 / 2)} \sqrt{35.83 \left( \frac{1}{42} + \frac{1}{50} \right)} \]

- Which simplifies to

\[ = 6.09 \pm 2.934 \times \sqrt{1.57} = 6.09 \pm 3.676 = [2.41, 9.77] \]

- Interpret as a standard confidence interval, except note that this corrects for multiple comparisons (spreads out \( \alpha \) for all CIs calculated)
Bonferroni Method of Multiple Comparisons

- Use this SE with Bonferroni correction:

\[
\bar{X}_3 - \bar{X}_2 \pm t_{(n-k, \frac{\alpha^*}{2})} \sqrt{MS_w \left( \frac{1}{n_3} + \frac{1}{n_2} \right)}
\]

- For our analysis:

\[
78.11 - 72.02 \pm t_{(193, 0.00333/2)} \sqrt{35.83 \left( \frac{1}{42} + \frac{1}{50} \right)}
\]

Which simplifies to

\[
6.09 \pm 3.676 = [2.41, 9.77]
\]

Interpret as a standard confidence interval, except note that this corrects for multiple comparisons (spreads out \( \alpha \) for all CIs calculated).
Bonferroni Method of Multiple Comparisons

- Use this SE with Bonferroni correction:

\[ \bar{X}_3 - \bar{X}_2 \pm t_{(n-k, \alpha^*)} \sqrt{MS_w \left( \frac{1}{n_3} + \frac{1}{n_2} \right)} \]

- For our analysis:

\[ 78.11 - 72.02 \pm t_{(193, 0.00333/2)} \sqrt{35.83 \left( \frac{1}{42} + \frac{1}{50} \right)} \]

Which simplifies to

\[ = 6.09 \pm 3.676 \]

\[ = [2.41, 9.77] \]

Interpret as a standard confidence interval, except note that this corrects for multiple comparisons (spreads out \( \alpha \) for all CIs calculated).
Bonferroni Method of Multiple Comparisons

- Use this SE with Bonferroni correction:

\[
\bar{X}_3 - \bar{X}_2 \pm t_{(n-k, \frac{\alpha^*}{2})} \sqrt{MS_w \left( \frac{1}{n_3} + \frac{1}{n_2} \right)}
\]

- For our analysis:

\[
78.11 - 72.02 \pm t_{(193,0.00333/2)} \sqrt{35.83 \left( \frac{1}{42} + \frac{1}{50} \right)}
\]

- Which simplifies to

\[
= 6.09 \pm 2.934 \times \sqrt{1.57} = 6.09 \pm 3.676 = [2.41, 9.77]
\]

Interpret as a standard confidence interval, except note that this corrects for multiple comparisons (spreads out \(\alpha\) for all CIs calculated).
Bonferroni Method of Multiple Comparisons

- Use this SE with Bonferroni correction:

\[
\bar{X}_3 - \bar{X}_2 \pm t_{(n-k, \frac{\alpha^*}{2})} \sqrt{MS_w \left( \frac{1}{n_3} + \frac{1}{n_2} \right)}
\]

- For our analysis:

\[
78.11 - 72.02 \pm t_{(193, 0.00333/2)} \sqrt{35.83 \left( \frac{1}{42} + \frac{1}{50} \right)}
\]

- Which simplifies to

\[
= 6.09 \pm 2.934 \times \sqrt{1.57}
\]

\[
= 6.09 \pm 3.676
\]

\[
= [2.41, 9.77]
\]
Bonferroni Method of Multiple Comparisons

- Use this SE with Bonferroni correction:
  \[
  \bar{X}_3 - \bar{X}_2 \pm t_{(n-k, \frac{\alpha^*}{2})} \sqrt{MS_w \left( \frac{1}{n_3} + \frac{1}{n_2} \right)}
  \]

- For our analysis:
  \[
  78.11 - 72.02 \pm t_{(193, 0.00333/2)} \sqrt{35.83 \left( \frac{1}{42} + \frac{1}{50} \right)}
  \]

- Which simplifies to
  \[
  = 6.09 \pm 2.934 \times \sqrt{1.57}
  \]
  \[
  = 6.09 \pm 3.676
  \]
  \[
  = [2.41, 9.77]
  \]

- Interpret as a standard confidence interval, except note that this corrects for multiple comparisons (spreads out \( \alpha \) for all CIs calculated)
Next Time

- Oregon Health Experiment
Appendix: Standard Error for Paired Diff in Proportion

Note: Notation here slightly different than used in main part of slides

\[
\text{Var}[\hat{\pi}_2 - \hat{\pi}_1] = \text{Var}[\hat{\pi}_{01} - \hat{\pi}_{10}]
= \text{Var}[\hat{\pi}_{01}] + \text{Var}[\hat{\pi}_{10}] - 2\text{Cov}[\hat{\pi}_{01}, \hat{\pi}_{10}]
= \frac{1}{n} \pi_{01}(1 - \pi_{01}) + \frac{1}{n} \pi_{10}(1 - \pi_{10}) + \frac{2}{n} \pi_{01} \pi_{10}
= \frac{1}{n} \left( \pi_{01} + \pi_{10} - (\pi_{01} - \pi_{10})^2 \right)
\]
Appendix: McNemar’s Test

Under $H_0$: $n_{12}$ (or equivalently $n_{21}$) has a binomial distribution with parameters $n^* = n_{12} + n_{21}$ and "success" probability 0.5.

Under $H_0$: the mean of $n_{12}$ is $0.5n^*$.

Under $H_0$: the variance of $n_{12}$ is $0.5(1 - 0.5)n^*$.

Using CLT, we can approximate the binomial with Normal: $n_{12} \sim N(0.5n^*, 0.5(1 - 0.5)n^*)$.

Standardizing, we get $z = \frac{n_{12} - 0.5n^*}{\sqrt{0.5(1 - 0.5)n^*}}$.

Using $n^* = n_{12} + n_{21}$ can further simplify to $z = \frac{n_{12} - n_{21}}{\sqrt{n_{12} + n_{21}}}$.

This approximates standard Normal.
Appendix: McNemar’s Test

- Under $H_0$: $n_{12}$ (or equivalently $n_{21}$) has a binomial distribution with parameters $n^* = n_{12} + n_{21}$ and “success” probability 0.5.
Appendix: McNemar’s Test

- Under $H_0$: $n_{12}$ (or equivalently $n_{21}$) has a binomial distribution with parameters $n^* = n_{12} + n_{21}$ and “success” probability 0.5
  - Under $H_0$: the mean of $n_{12}$ is $0.5n^*$
Appendix: McNemar’s Test

- Under $H_0$: $n_{12}$ (or equivalently $n_{21}$) has a binomial distribution with parameters $n^* = n_{12} + n_{21}$ and “success” probability 0.5
  - Under $H_0$: the mean of $n_{12}$ is $0.5n^*$
  - Under $H_0$: the variance of $n_{12}$ is $0.5(1 - 0.5)n^*$

Using CLT, we can approximate the binomial with Normal:

$$n_{12} \sim \mathcal{N}\left(0.5n^*, 0.5\left(1 - 0.5\right)n^*\right)$$

Standardizing, we get

$$z = \frac{n_{12} - 0.5n^*}{\sqrt{0.5\left(1 - 0.5\right)n^*}}$$

Using $n^* = n_{12} + n_{21}$ can further simplify to

$$z = \frac{n_{12} - n_{21}}{\sqrt{n_{12} + n_{21}}}$$

where this approximates standard Normal.
Appendix: McNemar’s Test

- Under $H_0$: $n_{12}$ (or equivalently $n_{21}$) has a binomial distribution with parameters $n^* = n_{12} + n_{21}$ and “success” probability 0.5
  - Under $H_0$: the mean of $n_{12}$ is $0.5n^*$
  - Under $H_0$: the variance of $n_{12}$ is $0.5(1 - 0.5)n^*$
- Using CLT, we can approximate the binomial with Normal:
Appendix: McNemar’s Test

- Under $H_0$: $n_{12}$ (or equivalently $n_{21}$) has a binomial distribution with parameters $n^* = n_{12} + n_{21}$ and “success” probability 0.5
  - Under $H_0$: the mean of $n_{12}$ is $0.5n^*$
  - Under $H_0$: the variance of $n_{12}$ is $0.5(1 - 0.5)n^*$

- Using CLT, we can approximate the binomial with Normal:

$$n_{12} \sim N(0.5n^*, 0.5(1 - 0.5)n^*)$$
Appendix: McNemar’s Test

- Under $H_0$: $n_{12}$ (or equivalently $n_{21}$) has a binomial distribution with parameters $n^* = n_{12} + n_{21}$ and “success” probability 0.5
  - Under $H_0$: the mean of $n_{12}$ is $0.5n^*$
  - Under $H_0$: the variance of $n_{12}$ is $0.5(1 - 0.5)n^*$
- Using CLT, we can approximate the binomial with Normal:
  \[ n_{12} \sim N(0.5n^*, 0.5(1 - 0.5)n^*) \]
- Standardizing, we get
Appendix: McNemar’s Test

- Under $H_0$: $n_{12}$ (or equivalently $n_{21}$) has a binomial distribution with parameters $n^* = n_{12} + n_{21}$ and “success” probability 0.5
  - Under $H_0$: the mean of $n_{12}$ is $0.5n^*$
  - Under $H_0$: the variance of $n_{12}$ is $0.5(1 - 0.5)n^*$
- Using CLT, we can approximate the binomial with Normal:

\[
n_{12} \sim N(0.5n^*, 0.5(1 - 0.5)n^*)\]

- Standardizing, we get

\[
z = \frac{n_{12} - 0.5n^*}{\sqrt{0.5(1 - 0.5)n^*}}\]
Appendix: McNemar’s Test

- Under $H_0$: $n_{12}$ (or equivalently $n_{21}$) has a binomial distribution with parameters $n^* = n_{12} + n_{21}$ and “success” probability 0.5
  - Under $H_0$: the mean of $n_{12}$ is $0.5n^*$
  - Under $H_0$: the variance of $n_{12}$ is $0.5(1 - 0.5)n^*$
- Using CLT, we can approximate the binomial with Normal:
  \[ n_{12} \sim N(0.5n^*, 0.5(1 - 0.5)n^*) \]

- Standardizing, we get
  \[ z = \frac{n_{12} - 0.5n^*}{\sqrt{0.5(1 - 0.5)n^*}} \]

- Using $n^* = n_{12} + n_{21}$ can further simplify to
  \[ = \frac{n_{12} - n_{21}}{\sqrt{n_{12} + n_{21}}} \]
Appendix: McNemar’s Test

- Under $H_0$: $n_{12}$ (or equivalently $n_{21}$) has a binomial distribution with parameters $n^* = n_{12} + n_{21}$ and “success” probability 0.5
  - Under $H_0$: the mean of $n_{12}$ is $0.5n^*$
  - Under $H_0$: the variance of $n_{12}$ is $0.5(1 - 0.5)n^*$
- Using CLT, we can approximate the binomial with Normal:
  
  $$n_{12} \sim N(0.5n^*, 0.5(1 - 0.5)n^*)$$

- Standardizing, we get
  
  $$z = \frac{n_{12} - 0.5n^*}{\sqrt{0.5(1 - 0.5)n^*}}$$

- Using $n^* = n_{12} + n_{21}$ can further simplify to
  
  $$= \frac{n_{12} - n_{21}}{\sqrt{n_{12} + n_{21}}}$$

- where this approximates standard Normal