Inferences When Comparing Two Means

Dr. Tom Ilvento
FREC 408

Thus far...
- We have made an inference from a single sample mean and proportion to a population, using:
  - The sample mean (or proportion)
  - The sample standard deviation
  - Knowledge of the sampling distribution for the mean (proportion)
- And it matters if the sample size is large or small

Testing differences between two means or proportions
- The same strategy will apply for testing differences between two means or proportions
- With a few twists
  - Mean
    - Large sample
    - Small sample – pool the variance
  - Proportions
    - When testing \( H_0: \mu_1 = \mu_2 \): we need to check if \( p_1 = p_2 \)

What are independent, random samples?
- Independent samples means that each sample and the resulting variables do not influence the other sample
  - If we sampled the same subjects at two different times we would not have independent samples
  - If we sampled husband and wife, they would not be independent
- However, we have a strategy to assess change over time of the same subject – paired difference test

Testing differences between two means or proportions
- We will also need to come up with:
  - An estimator of the difference of two means/proportions
  - The standard error of the sampling distribution for our estimator
  - With two sample problems we have two sources of variability and sampling error
- We also must assume the samples are independent random samples

Decision Tree for Two Means

<table>
<thead>
<tr>
<th>Target</th>
<th>Assumptions</th>
<th>Test Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0: \mu_1 = \mu_2 )</td>
<td>Independent random samples Large sample size ((n_1, n_2 &gt; 30))</td>
<td>( z ), using sample variance</td>
</tr>
<tr>
<td>( \mu_1 - \mu_2 = 0 )</td>
<td>Independent random samples Small sample size |</td>
<td>( t ), using pooled variance ( S_p^2 )</td>
</tr>
</tbody>
</table>
### Decision Tree for two Proportions

<table>
<thead>
<tr>
<th>Testing</th>
<th>Assumptions</th>
<th>Test Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0: p_1 - p_2 = 0 )</td>
<td>Independent random samples</td>
<td>( z ), using ( \text{pooled} ) sample</td>
</tr>
<tr>
<td>Known that ( p_1 = p_2 ) under ( H_0 ) proportion ( \bar{p} )</td>
<td>Large sample size ( (n_1, n_2 &gt; 30) )</td>
<td>( z ) when ( D_0 \neq 0 )</td>
</tr>
</tbody>
</table>

### Example Problem
- Two groups of students were surveyed about lecture notes
  - 86 students in a promotional strategy class that required the purchase of the lecture notes
  - 35 students enrolled in a sales/retailing class which didn’t offer lecture notes
- At the end of the semester, students in both classes were asked if “Having a copy of the lecture was [would be] helpful in understanding the material.”
- The question was measured on a nine point scale where 1 = strongly disagree and 9 = strongly agree

### Lecture Notes problem
- Class with Lecture Notes
  - \( n_1 = 86 \)
  - \( \bar{x}_1 = 8.48 \)
  - \( s_1^2 = .94 \)
- Class without Lecture Notes
  - \( n_2 = 35 \)
  - \( \bar{x}_2 = 7.80 \)
  - \( s_2^2 = 2.99 \)

Do the samples provide sufficient evidence to conclude that there is a difference in mean responses of the two groups? Use \( \alpha = .01 \)

### Lecture Notes problem
- Null hypothesis: \( H_0 : (\mu_1, \mu_2) = ? \)
- Alternative Assumptions: Two independent samples, \( n \) is Large
- Test Statistic: \( z^* = \text{two-tailed test} \)
- Rejection Region: \( z_{.025} = 2.575 \)
- Calculation: \( z^* = \) ? \( z_{.05} \)
- Conclusion: ? \( H_0 : (\mu_1, \mu_2) = 0 \)

### We need to figure out the sampling distribution \( (\bar{x}_1 - \bar{x}_2) \)
- The mean of the sampling distribution for \( (\bar{x}_1 - \bar{x}_2) \)
- Will equal \( (\mu_1 - \mu_2) \)
- \( = D_0 \)
- We usually designate the expected difference as \( D_0 \) under the null hypothesis
- Most often we think of \( D_0 = 0 \); no difference

### Standard Error of the difference of two means
- The **Standard Error** of the difference of two means is given as:
  \[
  \sigma_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}
  \]

The sampling distribution of \( (\bar{x}_1 - \bar{x}_2) \) is approximately normal for large samples under the Central Limit Theorem
The Standard Error for the difference of two means

- Is based on two independent random samples
- We typically use the sample estimates of \( \sigma_1 \) and \( \sigma_2 \)
- Which are \( s_1 \) and \( s_2 \)

\[
\sigma_{(\bar{x}_1-\bar{x}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}
\]

The Test Statistic for our problem

\[
z^* = \frac{(8.48 - 7.80) - 0}{\sqrt{\frac{.94}{86} + \frac{2.99}{35}}}
\]

\[
z^* = (6.80 - 0)/.3104 = 2.1906
\]

99% Confidence Interval for the Difference of Two Means

\[
(\bar{x}_1 - \bar{x}_2) \pm z_{.01/2} \sigma_{(\bar{x}_1-\bar{x}_2)}
\]

\[
(8.48-7.80) \pm 2.575 \left(\sqrt{\frac{.94}{86} + \frac{2.99}{35}}\right)
\]

\[
.68 \pm 2.575(.3104)
\]

\[
.68 \pm .80
\]

\[
-.12 \text{ to } 1.48
\]

Notice the 99% C.I. contains the null hypothesis value, zero

Decision Tree for Two Means

<table>
<thead>
<tr>
<th>Target</th>
<th>Assumptions</th>
<th>Test Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0: \mu_1=\mu_2 )</td>
<td>Independent random samples</td>
<td>( z ), using sample variance</td>
</tr>
<tr>
<td>Large sample size (n1, n2&gt;30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( H_0: \mu_1=\mu_2 )</td>
<td>Independent random samples</td>
<td>( t ), using pooled variance</td>
</tr>
<tr>
<td>Small sample size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \mu_1-\mu_2 \neq 0 )</td>
<td>Both samples are approximately normal</td>
<td>Equal variances</td>
</tr>
<tr>
<td>Independent random samples</td>
<td>The population variances are equal</td>
<td></td>
</tr>
<tr>
<td>Populations appr. normal</td>
<td>( S_p^2 )</td>
<td></td>
</tr>
<tr>
<td>Equal variances</td>
<td>Random samples selected independently of each other</td>
<td></td>
</tr>
</tbody>
</table>

What about when n is small?

- We will use a t-test and the t distribution
- Assumptions
  - Both samples are approximately normal
  - The population variances are equal
  - Random samples selected independently of each other
The standard error for a small sample difference of means

- Since we assume $\sigma_1 = \sigma_2$
- thus $(s_1 = s_2)$
- We should pool our estimate of the standard error of the sampling distribution
- Using information from both sample estimates

POOLED ESTIMATE OF THE VARIANCE

- Then our formula will be a weighted average of $s_1$ and $s_2$

$$s^2_p = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

Note: the denominator reduces to $(n_1 + n_2 - 2)$ which is the d.f. for the t distribution

Next, we use the Pooled Estimate of the Variance to calculate the estimate of the standard error

$$\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{s^2_p + \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

What does pooling do for us?

- Pooling generates a weighted average as the estimate of the variance
- The weights are the sample sizes for each sample
- A pooled estimate is thought to be a better estimate if we can assume the variances are equal
- And our degrees of freedom are larger - d.f. = $n_1 + n_2 - 2$
- Which means the t-value will be smaller

Problem, Tapeworms in sheep

- An experiment was done to compare the mean number of tapeworms in the stomachs of sheep that had been treated for worms versus those not treated.
- There were 7 sheep in the Treatment group and 7 in the Control Group
- Is the number of tapeworms lower in the treatment group at $\alpha = .05$?

Tapeworms in sheep

- The means and standard deviations are:

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{x}_1 = 28.57$</td>
<td>$\bar{x}_2 = 40.0$</td>
</tr>
<tr>
<td>$s_1^2 = 198.62$</td>
<td>$s_2^2 = 215.33$</td>
</tr>
<tr>
<td>$n_1 = 7$</td>
<td>$n_2 = 7$</td>
</tr>
</tbody>
</table>
Tapeworms in sheep

We assume the variances are equal so we make a pooled estimate

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

\[ s_p = \sqrt{\frac{(7 - 1)198.62 + (7 - 1)215.33}{7 + 7 - 2}} = \frac{2483.7}{12} = 14.387 \]

Then we use our pooled estimate to estimate the standard error of the sampling distribution for the difference of means

\[ \hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \]

\[ \hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)} = 14.387 \sqrt{\frac{1}{7} + \frac{1}{7}} = 7.69 \]

Null hypothesis

- \( H_0: (\mu_1 - \mu_2) = 0 \)
- \( H_a: (\mu_1 - \mu_2) < 0 \) one-tailed test, lower

Assumptions

- Small independent samples, approx normal, variances are equal

Test Statistic

- \( t^* = \frac{(28.57 - 40.00)}{7.69} = -1.782 \)

Rejection Region

- \( t_{.05, 12 \text{ d.f.}} = -1.48 \)

Calculation

- \( t^* < -t_{.05, 12 \text{ d.f.}} \)
- \( -1.48 < -1.782 \)

Conclusion

We cannot reject \( H_0: (\mu_1 - \mu_2) = 0 \)

The 90% C.I. For the Tapeworm example

\[ (\bar{x}_1 - \bar{x}_2) \pm t_{1-\alpha/2, n_1+n_2-2 \text{ d.f.}} \cdot \hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)} \]

- \( (28.57 - 40.00), t_{12/2, n_1+n_2-2 \text{ d.f.}} [14.387 (1/7+1/7)^{-\frac{1}{2}}] \]
- \(-11.43 \pm 1.782(7.690) \)
- \(-11.43 \pm 13.704 \)
- \(-25.134 \text{ to } 2.273 \)

This C.I. contains zero – analogous to null hypothesis, one-tailed test, at \( \alpha = .05 \)
### What Do I need for a Difference of Means Tests?
- Two independent random samples
- Determine if I am dealing with a mean or proportion
- If a Mean, large sample or small sample
  - Large sample I need not assume the distribution of the populations, or anything about the variances
  - I calculate the standard error and make my test using a z-value
- Small sample I must assume
  - Independent random samples
  - From populations that are normally distributed
  - The variances are equal
- Small Sample Steps
  - Calculate pooled estimate of Variance
  - Use pooled estimate to calculate the standard error
  - Conduct the test use a t-value with n1 + n2 – 2 df

### Decision Tree for two Proportions

<table>
<thead>
<tr>
<th>Testing</th>
<th>Assumptions</th>
<th>Test Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0: p_1 - p_2 = 0 )</td>
<td>Independent random samples</td>
<td>z, using pooled sample proportion ( p )</td>
</tr>
<tr>
<td>( H_0: p_1 - p_2 = D_0 )</td>
<td>Independent random samples</td>
<td>When ( D_0 \neq 0 )</td>
</tr>
</tbody>
</table>

### What about the difference in proportions?
- Based on large sample only
- Same strategy as for the mean
  - We calculate the difference in the two sample proportions
  - Establish the sampling distribution for our estimator
  - Calculate a standard error of this sampling distribution
  - Conduct a test

### Differences of Proportions
- For the null hypothesis
  - \( E(\hat{p}_1 - \hat{p}_2) = (p_1 - p_2) = D_0 \)
- The Standard Error for \( (\hat{p}_1 - \hat{p}_2) \)
  \[
  \sigma_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}
  \]

### For Proportions
- Since this is a large sample problem, we could use the sample estimates of \( p_1 \) and \( p_2 \) to estimate the standard error
  \[
  \sigma_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}
  \]
For Proportions

- **Note:** Under the Null Hypothesis where $p_1 - p_2 = 0$
  - The book suggests that we use a weighted average for $p$, based on adding the total number of successes and divide by the sum of the two sample sizes (P477-78)
  - $p = (x_1 + x_2)/(n_1 + n_2)$
  - where $x = \#$ of successes
  
Note: the book uses $\hat{p}$ instead of $p$

---

So for Proportions

- **For a confidence interval**, use the sample estimates in this manner to generate the standard error – **since there is no assumption that the variances are equal**
  
  $$
  \sigma_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}
  $$

---

Gender gap in politics

- Over half the votes will be cast by women
- The question is if the Democratic platform is viewed more favorably by women
- Suppose 150 men and 150 women stated their party preferences for Republicans
- Data survey of 300 – 150 men and 150 women
  - 81 men favor Republicans
  - 70 women favor Republicans
- Conduct a test using $\alpha = .05$ that a lower proportion of women favor Republicans

---

Calculate the standard error

- The sample proportions are
  - Men $81/150 = .540$
  - Women $70/150 = .467$
  
- Pooled estimate
  - $p = (81+70)/(150+150) = .5033$
  - $q = 1 - .5033 = .4967$

---

Calculate the standard error

- Standard Error for the problem
  
  $$
  \sigma_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{[(.5033)(.4967)(1/150 + 1/150)]}^5
  $$
  
  $$
  \sigma_{(\hat{p}_1 - \hat{p}_2)} = [(249989)(.01333)]^5
  $$
  
  $$
  \sigma_{(\hat{p}_1 - \hat{p}_2)} = .057734
  $$
What will be the test statistic?

- \( z^* = [(0.540 - 0.467) - 0] / 0.057734 \)
  - Where the numerator shows the difference in voting Republican for men and women
  - Note that I subtract out the hypothesized value \( D_0 = 0 \)
  - I usually try to keep the \( z^* \) for a difference in proportion (or means) positive as a matter of convenience

Gender Gap in Politics

Null hypothesis: \( H_0: (p_m - p_w) = 0 \)
Alternative hypothesis: \( H_1: (p_m - p_w) > 0 \) one-tailed test, upper

Assumptions
- Large sample proportion, use normal, assume variances equal

Test Statistic
- \( z^* = [(0.540 - 0.467) - 0] / 0.057734 \)
- \( z_{0.05} = 1.645 \)

Calculation
- \( z^* = 1.264 \)
- \( z^* < z_{0.05} \)
- \( 1.264 < 1.645 \)

Conclusion
- Cannot reject \( H_0: (p_m - p_w) = 0 \)

Confidence Interval for the Gender in Politics Problem

- We will use \( \alpha = 0.10 \) to relate to the previous one-tailed hypothesis test of \( \alpha = 0.05 \)
- Standard error (not assuming equal variances) is:
  - \( s_{p_m-p_w} = \sqrt{(0.540)(0.460)/150 + (0.467)(0.533)/150} = 0.05758 \)
- 90% C.I.:
  - \( (0.540 - 0.467) \pm 1.645(0.05758) \)
  - \( 0.073 \pm 0.095 \)
  - \( 0.073 \) to \( 0.168 \)

Section 9.3: Paired Difference Test

- When you have a situation where we record a pre and post test for the same individual
- We cannot treat the samples as independent
- In these cases we can do a Paired Difference Test.
- It’s called “Matched Pairs Test” in the book. (page 467-70)

Paired Difference Test

- The strategy is relatively simple
- We simply create a new variable which is the difference of the pre-test from the post test
- This new variable can be thought as a single random sample
- For this new variable we calculate sample estimates of the mean and standard deviation

Paired Difference Test

- And then calculate C.I. or conduct a hypothesis test on this new variable
- Often times the mean difference is referred to as
  - \( \bar{D} \)
  - And the hypothesis is often:
    - \( H_0: \mu_D = 0 \)
    - This is no different than any single mean test, large sample or small sample
Example of paired difference data

<table>
<thead>
<tr>
<th>Patient</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Difference 2 - 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>-2</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Alzheimer’s study

- Twenty Alzheimer patients were asked to spell 24 homophone pairs given in random order.
- Homophones are words that have the same pronunciation as another word with a different meaning and different spelling.
- Nun and none; doe and dough
- The number of confusions were recorded
- The test was repeated one year later

Alzheimer’s study

- The researchers posed the following question:
  
  Do Alzheimer’s patients show a significant increase in mean homophone confusion over time?

- Use an alpha value of .05

Alzheimer’s study

Null hypothesis: H₀: μ₀ = 0
Alternative: H₁: μ₁ > 0 one-tailed test, upper small sample, normal
Test Statistic: t* = (1.65 - 0)/(3.2/√20)
Rejection Region: t₀.05, 19 d.f. = 1.729
Calculation: t* = 2.31
Conclusion: t* > t₀.05, 19 d.f.
2.31 > 1.729
Reject H₀: μ₀ = 0

Alzheimer’s study

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>4.15</td>
<td>5.80</td>
<td>1.65</td>
</tr>
<tr>
<td>Standard Error</td>
<td>0.78</td>
<td>0.94</td>
<td>0.72</td>
</tr>
<tr>
<td>Median</td>
<td>5.00</td>
<td>5.50</td>
<td>1.00</td>
</tr>
<tr>
<td>Mode</td>
<td>5.00</td>
<td>3.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>3.50</td>
<td>4.21</td>
<td>3.20</td>
</tr>
<tr>
<td>Sample Variance</td>
<td>12.24</td>
<td>17.75</td>
<td>10.24</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-0.85</td>
<td>0.13</td>
<td>0.08</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.41</td>
<td>0.64</td>
<td>0.48</td>
</tr>
<tr>
<td>Range</td>
<td>11</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
<td>-3</td>
</tr>
<tr>
<td>Maximum</td>
<td>11</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>Sum</td>
<td>83</td>
<td>116</td>
<td>33</td>
</tr>
<tr>
<td>Count</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>