Chapter 9: Testing the Difference Between Two Means, Two Proportions, and Two Variances

Diana Pell

Section 9.2: Testing the Difference Between Two Means of Independent Samples

Note: Samples are independent samples when they are not related.

Formula for the $t$ Test for Testing the Difference Between Two Means, Independent Samples

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

where the degrees of freedom are equal to the smaller of $n_1 - 1$ or $n_2 - 1$.

Assumptions for the $t$ Test for Two Independent Means When $\sigma_1$ and $\sigma_2$ Are Unknown

1. The samples are random samples.
2. The sample data are independent of one another.
3. When the sample sizes are less than 30, the populations must be normally or approximately normally distributed.
Exercise 1. A researcher wishes to see if the average weights of newborn male infants are different from the average weights of newborn female infants. She selects a random sample of 10 male infants and finds the mean weight is 7 pounds 11 ounces and the standard deviation of the sample is 8 ounces. She selects a random sample of 8 female infants and finds that the mean weight is 7 pounds 4 ounces and the standard deviation of the sample is 5 ounces. Can it be concluded at $\alpha = 0.05$ that the mean weight of the males is different from the mean weight of the females? Assume that the variables are normally distributed.

1) $H_0: \mu_1 = \mu_2$ and $H_1: \mu_1 \neq \mu_2$ (claim)

2) $t$-test
   - Find C.V.
     - Table F
     - Two-tailed
     - $\alpha = 0.05$
     - d.f. = 7
   - $n_1 - 1 = 10 - 1 = 9$ (use smaller)
   - $n_2 - 1 = 8 - 1 = 7$

3) Test Value
   
   \[
   t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}
   \]
   
   \[
   t = \frac{(7\text{lb 11 oz} - 7\text{lb 4 oz}) - 0}{\sqrt{\frac{8^2}{10} + \frac{5^2}{8}}} = \frac{7}{3.086} = [2.268]
   \]

4) Do not reject the null hypothesis

5) Thus is not enough evidence to support the claim that the mean of the weights of the male infants is different from the mean of the weights of the female infants.
Confidence Intervals for the Difference of Two Means: Independent Samples

Table F

\[ \text{d.f.} = 7 \]
\[ \alpha = 0.05 \]
\[ t_{0.025} = 2.365 \]

Variances assumed to be unequal:

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

\[ \text{d.f.} = \text{smaller value of } n_1 - 1 \text{ or } n_2 - 1 \]

Exercise 2. Find the 95% confidence interval for the data in Exercise 1.

\[ (123 - 116) - 2.365 \cdot \sqrt{\frac{8^2}{10} + \frac{5^2}{8}} < \mu_1 - \mu_2 < (123 - 116) + 2.365 \cdot \sqrt{\frac{8^2}{10} + \frac{5^2}{8}} \]

\[ 7 - 7.3 < \mu_1 - \mu_2 < 7 + 7.3 \]

\[ -0.3 < \mu_1 - \mu_2 < 14.3 \]

Section 9.3: Testing the Difference Between Two Means: Dependent Samples

Note: Samples are considered to be dependent samples when the subjects are paired or matched in some way.

1. Will drug affect the reaction time of its users?
2. Subjects are given a test to find out their normal reaction times.
3. After taking the drug, the same subjects are tested again.
4. Means of the two tests are compared.
5. Since the same subjects are used in both lists, the samples are related.
When the samples are dependent, a special t test for dependent means is used. This test employs the difference in values of the matched pairs. The hypotheses are as follows:

<table>
<thead>
<tr>
<th>Two-tailed</th>
<th>Left-tailed</th>
<th>Right-tailed</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0: \mu_D = 0$</td>
<td>$H_0: \mu_D = 0$</td>
<td>$H_0: \mu_D = 0$</td>
</tr>
<tr>
<td>$H_1: \mu_D \neq 0$</td>
<td>$H_1: \mu_D &lt; 0$</td>
<td>$H_1: \mu_D &gt; 0$</td>
</tr>
</tbody>
</table>

$\mu_D = \text{expected mean of the difference of the matched pairs}$

Assumptions for the t Test for Two Means When the Samples Are Dependent

1. The sample or samples are random.
2. The sample data are dependent.
3. When the sample size or sample sizes are less than 30, the population or populations must be normally or approximately normally distributed.

Formulas for the t Test for Dependent Samples

$D = X_1 - X_2$

$t = \frac{D - \mu_D}{\sigma_D / \sqrt{n}}$

with d.f. = $n - 1$ and where

$D = \frac{\Sigma D}{n}$ and $\sigma_D = \sqrt{\frac{n \Sigma D^2 - (\Sigma D)^2}{n(n-1)}}$
Exercise 3. A random sample of nine local banks shows their deposits (in billions of dollars) 3 years ago and their deposits (in billions of dollars) today. At \( \alpha = 0.05 \), can it be concluded that the average in deposits for the banks is greater today than it was 3 years ago? Use \( \alpha = 0.05 \). Assume the variable is normally distributed.

<table>
<thead>
<tr>
<th>Bank</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>3 yrs ago</td>
<td>11.42</td>
<td>8.41</td>
<td>3.95</td>
<td>7.37</td>
<td>2.28</td>
<td>1.10</td>
<td>1.00</td>
<td>0.9</td>
<td>1.35</td>
</tr>
<tr>
<td>Today</td>
<td>16.69</td>
<td>9.44</td>
<td>6.53</td>
<td>5.58</td>
<td>2.92</td>
<td>1.88</td>
<td>1.78</td>
<td>1.5</td>
<td>1.22</td>
</tr>
</tbody>
</table>

1) \( H_0: \mu_D = 0 \) and \( \mu_D < 0 \) (claim)

2) \( d.f. = 9 - 1 = 8 \)

Use Table F

Left-tailed test

\( \alpha = 0.05 \)

C.V. = -1.860

3) Test value

<table>
<thead>
<tr>
<th>3 yrs ago ( (X_1) )</th>
<th>Today ( (X_2) )</th>
<th>( D = X_1 - X_2 )</th>
<th>( D^2 = (X_1 - X_2)^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.42</td>
<td>16.69</td>
<td>-5.27</td>
<td>27.7729</td>
</tr>
<tr>
<td>8.41</td>
<td>9.44</td>
<td>-1.03</td>
<td>1.0609</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

\( \sum D = -9.73 \) \( \sum D^2 = 40.5437 \)

Find \( S_D \):

\[ S_D = \sqrt{\frac{9(40.5437) - (-9.73)^2}{9(9-1)}} = 1.937 \]

Test value

\[ t = \frac{\bar{D} - \mu_0}{S_D / \sqrt{\frac{n}{9}}} = \frac{-1.081 - 0}{1.937 / \sqrt{9}} = -1.674 \]

\( \times \) \( \mu_0 = 0 \) if the hypothesis is \( \mu_0 = 0 \)

4) Do not reject the null hypothesis

5) There's not enough evidence to show that the deposits have increased over the last 3 yrs.
Exercise 4. A dietitian wishes to see if a person’s cholesterol level will change if the diet is supplemented by a certain mineral. Six randomly selected subjects were pretested, and then they took the mineral supplement for a 6-week period. The results are shown in the table. (Cholesterol level is measured in milligrams per deciliter.) Can it be concluded that the cholesterol level has been changed at $\alpha = 0.10$? Assume the variable is approximately normally distributed.

<table>
<thead>
<tr>
<th>Subject</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>before ($X_1$)</td>
<td>210</td>
<td>235</td>
<td>209</td>
<td>190</td>
<td>172</td>
<td>244</td>
</tr>
<tr>
<td>after ($X_2$)</td>
<td>190</td>
<td>170</td>
<td>210</td>
<td>199</td>
<td>173</td>
<td>226</td>
</tr>
</tbody>
</table>

1) $H_0: \mu_D = 0$ and $H_1: \mu_D \neq 0$ (claim)

2) df. = 6 - 1 = 5

$\alpha = 0.10$  

Table F

3) $D = X_1 - X_2$  

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$D = X_1 - X_2$</th>
<th>$D^2 = (X_1 - X_2)^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>210</td>
<td>190</td>
<td>20</td>
<td>400</td>
</tr>
<tr>
<td>235</td>
<td>170</td>
<td>65</td>
<td>4225</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>$\Sigma D = 100$</td>
<td>$\Sigma D^2 = 4890$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$S_D = \sqrt{\frac{\Sigma D^2 - \frac{1}{n}(\Sigma D)^2}{n(n-1)}} = 25.4$

Test value: $t = \frac{16.7 - 0}{25.4 / \sqrt{6}} = 1.610$

4) Do not reject the null hypothesis

5) There is not enough evidence to support the claim that the mineral changes a person's cholesterol level.
Exercise 5. Find the 90% confidence interval for the data in Exercise 4.

\[
16.7 - 2.015 \cdot \frac{25.4}{\sqrt{6}} < \mu_D < 16.7 + 2.015 \cdot \frac{25.4}{\sqrt{6}}
\]

\[
-4.2 < \mu_D < 37.6
\]
Hard body  
\[ \bar{x}_1 = 17.41 \]
\[ s_1 = 2.45 \]
\[ n_1 = 17 \]

Soft  
\[ \bar{x}_2 = 14.67 \]
\[ s_2 = 4.76 \]
\[ n_2 = 6 \]

\[ \alpha = 0.05 \]

1. \( H_0 : \mu_1 = \mu_2 \) and \( H_1 : \mu_1 \neq \mu_2 \) (claim)

2. C.V. = ±2.571

3. d.f. = 5

4. \( t = 1.351 \)

5. do not reject the null hypothesis

5. There is not enough evidence to support the claim that the means are not equal.

2. 95\% conf. int. for the difference of the means

\[ -2.48 < \mu_1 - \mu_2 < 7.96 \]

\[ (17.41 - 14.65) - 2.571 \cdot \sqrt{\frac{(2.45)^2}{17} + \frac{(4.76)^2}{6}} < \mu_1 - \mu_2 < \]
Chapter 9: Testing the Difference Between Two Means, Two Proportions, and Two Variances

Diana Pell

* test the difference in two sample proportions
* use Z test to test equality of two proportions
i.e. Is the proportion of men who exercise regularly less than the proportion of women who exercise regularly?

Section 9.4: Testing the Difference Between Proportions

**Formula for the z Test Value for Comparing Two Proportions**

\[
Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}
\]

where

\[
\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}, \quad \hat{p}_1 = \frac{X_1}{n_1}, \quad \hat{q} = 1 - \hat{p}, \quad \hat{p}_2 = \frac{X_2}{n_2}
\]

**Assumptions for the z Test for Two Proportions**

1. The samples must be random samples.
2. The sample data are independent of one another.
3. For both samples \(np \geq 5\) and \(nq \geq 5\).
Exercise 1. In the nursing home study mentioned in the chapter-opening Statistics Today, the researchers found that 12 out of 34 randomly selected small nursing homes had a resident vaccination rate of less than 80%, while 17 out of 24 randomly selected large nursing homes had a vaccination rate of less than 80%. At $\alpha = 0.05$, test the claim that there is no difference in the proportions of the small and large nursing homes with a resident vaccination rate of less than 80%.

1. $H_0 : \hat{p}_1 = \hat{p}_2$ and $H_1 : \hat{p}_1 \neq \hat{p}_2$  

2. $Z$-test  
   $Z = 0.05$  
   $Z = \frac{0.05}{0.025} = 2$  

3. $\overline{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{12 + 17}{34 + 24} = 0.5$ 

4. $\overline{q} = 1 - \overline{p} = 1 - 0.5 = 0.5$  

5. $Z = \frac{(\hat{p}_1 - \hat{p}_2) - (\hat{p}_1 - \hat{p}_2)}{\sqrt{\overline{p} \cdot \overline{q} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(0.35 - 0.71) - 0}{\sqrt{(0.5)(0.5)\left(\frac{1}{34} + \frac{1}{24}\right)}} = -2.70$  

4. Reject the null hypothesis  

5. There is enough evidence to reject the claim that there is no difference in the proportions of small and large nursing homes with a resident vaccination rate of less than 80%.
Exercise 2. A survey of 200 randomly selected male and female workers (100 in each group) found that 7\% of the male workers said that they worked more than 5 days per week while 11\% of the female workers said that they worked more than 5 days per week. At \( \alpha = 0.01 \), can it be concluded that the percentage of males who work more than 5 days per week is less than the percentage of female workers who work more than 5 days per week?

1) \( H_0 : p_1 = p_2 \) and \( H_1 : p_1 < p_2 \)

2) Table E, \( \alpha = 0.01 \)
   
   C.V. = -2.33

3) \( \hat{p}_1 = 0.07 \) and \( \hat{p}_2 = 0.11 \)
   
   \[ X_1 = \hat{p}_1 \cdot n_1 = (0.07)(100) = 7 \]
   
   \[ X_2 = \hat{p}_2 \cdot n_2 = (0.11)(100) = 11 \]

   \[ \bar{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{7 + 11}{100 + 100} = \frac{18}{200} = 0.09 \]

   \[ \bar{q} = 1 - \bar{p} = 1 - 0.09 = 0.91 \]

   \[ z = \frac{(0.07 - 0.11) - 0}{\sqrt{(0.09)(0.91) \left( \frac{1}{100} + \frac{1}{100} \right)}} = -0.99 \]

4) Do not reject the null hypothesis.

5) There is not enough evidence to support the claim that the prop. of men who say that they work more than 5 days a week is less than the prop. of women who say that they work more than 5 days a week.
Exercise 3. Find the 95% confidence interval for the difference of proportions for the data in Exercise 1.

\[ \hat{p}_1 = \frac{12}{34} = 0.35 \quad \hat{q}_1 = 0.65 \]

\[ \hat{p}_2 = \frac{17}{24} = 0.71 \quad \hat{q}_2 = 0.29 \]

\[ (0.35 - 0.71) - 1.96 \sqrt{\frac{(0.35)(0.65)}{34} + \frac{(0.71)(0.29)}{24}} < \hat{p}_1 - \hat{p}_2 < ... \]

\[ -0.602 < \hat{p}_1 - \hat{p}_2 < -0.118 \]