Chapter 9: Test the Equality and Estimate the Difference between Two Means, Variances, or Proportions (Independent Samples); Test and Estimate the Average of the Differences (Dependent Samples)

Learning Objectives
Upon successful completion of Chapter 9, you will be able to:

- Test the equality of two independent means and estimate the difference between two independent means:
  a. using $z$
  b. using $t$
- For paired, dependent data:
  a. test the average of their differences
  b. estimate the average of their differences
- For qualitative data:
  a. Test the equality of two proportions
  b. Estimate the difference in two proportions
- Test the equality of two variances for independent samples

I. Independent and Dependent Samples

A. Identifying Independent and Dependent Samples

- **Dependent Samples** are samples that are paired or matched in some way (There is an implied or stated connection that matches or pairs data.)
- **Independent Samples** are samples that are NOT related.

B. Types of Dependent Samples

1. The **same subjects** are used in a pre/post situation or before/after treatment or in two different situations.
2. Different “individuals” are paired based on some relevant characteristic.
   - Individuals within the same family are paired because of genetic similarities.
   - Both candidates for president in each election year are paired.
C. Examples

- John tests two different random samples of twenty-five 7th graders for reading comprehension. The first group used a computer assisted program; the second group used the regular program. Does this example involve independent or dependent samples?

  The samples are independent since there is no connection between the two groups of 7th graders.

- Sally tests twenty-five 7th graders for reading comprehension before and after they used a computer assisted program. Does this example involve independent or dependent samples?

  The samples are dependent since the same group is used “before” and “after” the students used a computer assisted program.

D. Issues with Dependent Samples

(Matching does control many variables influencing the experiment.)

- When subjects are matched according to one variable, the matching process does not eliminate the influence of other variables.
- When the same subjects are used for a pre/post study, sometimes the knowledge that they are participating in a study can influence the results.

II. Review: Hypothesis Testing Steps

<table>
<thead>
<tr>
<th>CRITICAL VALUE</th>
<th>P-VALUE METHOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. State the hypotheses</td>
<td>1. State the hypotheses</td>
</tr>
<tr>
<td>2. Find the test value</td>
<td>2. Find the test value</td>
</tr>
<tr>
<td>3. Find the c.v. and rejection region</td>
<td>3. Find p</td>
</tr>
<tr>
<td>4. Reject $H_0$ whenever the test value is in the rejection region</td>
<td>4. Rule:</td>
</tr>
<tr>
<td></td>
<td>$p \leq \alpha$ reject $H_0$</td>
</tr>
<tr>
<td></td>
<td>$p &gt; \alpha$ fail to reject $H_0$</td>
</tr>
<tr>
<td>5. Record an appropriate conclusion related to the problem</td>
<td></td>
</tr>
</tbody>
</table>
III. Two Independent Means

*Inferences About Two Independent Means*

A. z-Test for the Equality and z-Confidence Interval Estimate for their Difference

I. Assumptions

- The two samples must be independent, i.e., there is no relationship between the subjects in the two samples.
- Both population standard deviations must be known AND if a sample size is less than 30, those samples must be drawn from a normally distributed or approximately normally distributed populations.

II. z-Hypothesis Test for Equality of Two Independent Means

a) Formula

\[
H_0: \mu_1 = \mu_2 \text{ OR } H_0: \mu_1 - \mu_2 = 0
\]

\[
z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}
\]

\[
z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \text{ (calculator input)}
\]
**Note:** Before algebraic simplification, the formula is still basically:

\[
test\;value = \frac{\text{observed value} - \text{hypothesized value}}{\text{standard error}}
\]

- Observed difference: \((\bar{X}_1 - \bar{X}_2)\) – using the difference in sample means
- Hypothesized difference: \((\mu_1 - \mu_2)\) – using the difference in proposed population means
- Standard error of difference: \(\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}\) – using population variances \(\sigma_1^2\) and \(\sigma_2^2\)

**b) Problems**

- **Example:**
  A medical researcher wishes to see whether the average pulse rates of smokers is higher than the average pulse rates of non-smokers. The population standard deviation for the pulse rates of smokers is 5 \((\sigma_1 = 5)\) and the population standard deviation for the pulse rate of non-smokers is 6 \((\sigma_2 = 6)\). Samples of 100 smokers and 130 non-smokers are selected. The results are shown below.

Can the researcher conclude, at \(\alpha = 0.05\), that smokers have higher pulse rates than non-smokers? Use the p-value method.

\[
\begin{array}{|c|c|}
\hline
\text{Smokers} & \text{Non-smokers} \\
\hline
\bar{X}_1 = 90 & \bar{X}_2 = 88 \\
\hline
n_1 = 100 & n_2 = 130 \\
\hline
\sigma_1 = 5 & \sigma_2 = 6 \\
\hline
\end{array}
\]

1. \(H_0: \mu_1 = \mu_2\) \hspace{1cm} \(H_0: \mu_1 > \mu_2\) (claim)

2. \[z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(90 - 88)}{\sqrt{\left(\frac{5^2}{100} + \frac{6^2}{130}\right)}} = 2.56\]

3. (Just for understanding, note the c.v. = 1.65)

\[
p = P(z > 2.56) = normalcdf (2.56, 10) = .0052 \text{ (calculator)}
\]
Since $p = .00560$, which is less than .05, the level of significance, reject the hypothesis.

4. $P(z > 2.56) = 1 - .9948 = .0052$ (table)

5. The data supports higher pulse rates for smokers than non-smokers.

**Question 1**

Use the data below from the “Ages and Travel Study” to test the claim that the average age for domestic travel is greater than the average age for foreign travel. The population standard deviations and the data averages are below. Use $\alpha = 0.02$ and the critical value method.

<table>
<thead>
<tr>
<th>Domestic</th>
<th>Foreign</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{X}_1 = 61.2$ yrs</td>
<td>$\bar{X}_2 = 59.4$ yrs</td>
</tr>
<tr>
<td>$\sigma_1 = 5.9$</td>
<td>$\sigma_1 = 7.9$</td>
</tr>
<tr>
<td>$n_1 = 84$</td>
<td>$n_2 = 34$</td>
</tr>
</tbody>
</table>

III. z-Confidence Interval Estimate for the Difference Between Two Means

a) Formula

If the population standard deviations are known AND if either sample is less than 30, it must be selected from a normally distributed or approximately normally distributed population.

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

b) Problems

- **Example:**
  Find a 95% confidence interval estimate for the difference in the average flexibility rating of men and women.

<table>
<thead>
<tr>
<th></th>
<th>$n$</th>
<th>$\bar{X}$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>45</td>
<td>18.64</td>
<td>3.29</td>
</tr>
<tr>
<td>Women</td>
<td>31</td>
<td>20.99</td>
<td>2.07</td>
</tr>
</tbody>
</table>

**Note:**

- Both samples are larger than 30.
- Use $z$ since the population standard deviation is given for both groups.
\[
(20.99 - 18.64) \pm 1.96 \sqrt{\left(\frac{3.29^2}{45} + \frac{2.07^2}{31}\right)}
\]

2.35 ± 1.21

1.14 < \mu_w - \mu_m < 3.56

**Note**: This is the confidence interval estimate if the average for the men was subtracted from the average for the women. If the average for the women was subtracted from the average for the men, the confidence interval would become:

\[
(18.64 - 20.99) \pm 1.96 \sqrt{\left(\frac{2.07^2}{31} + \frac{3.29^2}{45}\right)}
\]

−2.35 ± 1.21

−3.56 < \mu_m - \mu_w < −1.14

Although both answers are correct, it is often better to choose the first since the interval has positive endpoints.

**Question 2**

A study comparing attitudes toward death was conducted in which organ donors (individuals who had signed organ donor cards) were compared with non-donors. The study is reported in the journal *Death Studies*. Templer’s Death Anxiety Scale (DAS) was administered to both groups. On this scale, high scores indicate high anxiety concerning death. The results were reported as follows:

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Pop Std Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organ Donors</td>
<td>35</td>
<td>5.36</td>
<td>2.91</td>
</tr>
<tr>
<td>Non-Organ Donors</td>
<td>69</td>
<td>7.62</td>
<td>3.45</td>
</tr>
</tbody>
</table>

Find a 90% confidence interval estimate for the differences in the average ratings on the Death Anxiety Scale for the organ donors and for non–organ donors.

**B. t-Test for the Equality of Two Independent Means and t-Confidence Interval Estimate for their Difference**

I. Assumptions

- The two samples must be independent, i.e., there is no relationship between the subjects in the two samples.
- One or both of the population standard deviations are not known AND if a sample size is less than 30, those samples must be drawn from a normally distributed or approximately normally distributed populations.
II. t-Hypothesis Test for Equality of Two Independent Means

a) **Formula** The difference between means when the two samples are *independent*, one or both of the population standard deviations are not known AND if a sample size is less than 30, that sample must be drawn from a normally distributed or approximately normally distributed population:

\[
t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}
\]

Where the degrees of freedom are equal to the smaller of \(n_1 - 1, n_2 - 1\).

**Note**: the variances are assumed to be unequal and not pooled or combined in any way.

b) **Problems**

- **Example**: John is comparing the average salaries for accountants in two different cities using data from randomly selected companies in each city. The data are shown below. Assume that salaries for accountants are normally distributed in both of these cities. Is there a significant difference in the average salaries at the significance level \(\alpha = 0.05\)?

<table>
<thead>
<tr>
<th>Chicago</th>
<th>New York</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\bar{X}_C) = $83,256</td>
<td>(\bar{X}_N) = $88,354</td>
</tr>
<tr>
<td>(S_C) = $3,256</td>
<td>(S_N) = $2,341</td>
</tr>
<tr>
<td>(n_C) = 15</td>
<td>(n_N) = 10</td>
</tr>
</tbody>
</table>

1. \(H_0: \mu_C = \mu_N\) (claim) \(H_1: \mu_C \neq \mu_N\)

2. \(t = \frac{\bar{X}_C - \bar{X}_N}{\sqrt{\frac{S_C^2}{n_C} + \frac{S_N^2}{n_N}}} = \frac{83256 - 88354}{\sqrt{3256^2/15 + 2341^2/10}} = -4.55\)

3. \(df = smaller\{9, 14\} = 9\) \(c.v. = \pm 2.262\)

4. c.v. method: the test value \(-4.55 < c.v. = -2.262\) or the test value is in the rejection region. Reject the null hypothesis.

5. The data is sufficient to refute or reject equal average salaries for accounts in New York compared to the average salaries in Chicago.
Question 3

Seth Young thinks that people walk faster when they are departing (getting on a plane) for a trip than when they are returning from their trip (getting off a plane). Assume walking times are normally distributed for each group. Use the findings for the randomly selected passengers in each category summarized in the table that follows to conduct the test. Use a 0.05 level of significance.

<table>
<thead>
<tr>
<th>Direction of Travel</th>
<th>Departure</th>
<th>Arrival</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean speed (ft per mn)</td>
<td>269</td>
<td>260</td>
</tr>
<tr>
<td>St Deviation (ft per mn)</td>
<td>53</td>
<td>34</td>
</tr>
<tr>
<td>Sample size</td>
<td>15</td>
<td>20</td>
</tr>
</tbody>
</table>

III. t-Confidence Interval Estimates for the Difference Between Two Independent Means

a) Formula

When the two samples are independent, one or both population standard deviations are not given, and the samples are drawn from normally distributed populations if the sample size is less than 30:

\[
(X_1 - X_2) \pm t\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^{\frac{1}{2}}
\]

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

b) Problems

- **Example:**

  Peter spends an average of 21.0 minutes helping each of 12 friends prepare for the statistics final exam. The standard deviation was 5.6 minutes. A student tutor in the Learning Center spent an average of 27.0 minutes helping 14 people prepare for the exam. The standard deviation was 4.3 minutes. Find the 98% confidence interval for the difference in the average time spent by Peter and by the student Learning Center tutor. Assume these times are normally distributed.

  \[\bar{X}_g = 21, n_g = 12, s_g = 5.6 \quad \bar{X}_v = 27, n_v = 14, s_v = 4.3\]

  \[d.f. = \text{smaller}\{11, 13\} = 11\]  
  For the t-distribution, c.v. = ± 2.718

  98% Confidence Interval:

  \[
  (27 - 21) \pm 2.718 \sqrt{\frac{5.6^2}{12} + \frac{4.3^2}{14}}
  \]

  6 ± 5.39

  6 ± 5

  \[1 < \mu_v - \mu_g < 11\]
Question 4
The time (in minutes) it took six white mice to learn to run a simple maze and the times it took six brown mice to learn to run the same maze are given here. Find the 95% confidence interval for the difference of the average times for white/brown mice. Assume maze running times are normally distributed.

<table>
<thead>
<tr>
<th>White mice</th>
<th>18</th>
<th>24</th>
<th>20</th>
<th>13</th>
<th>15</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brown mice</td>
<td>25</td>
<td>16</td>
<td>19</td>
<td>14</td>
<td>16</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>White Mice</th>
<th>Brown Mice</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{X}_w = 17.00$</td>
<td>$\bar{X}_B = 16.67$</td>
</tr>
<tr>
<td>$s_w = 4.56$</td>
<td>$s_B = 5.05$</td>
</tr>
<tr>
<td>$n_w = 6$</td>
<td>$n_B = 6$</td>
</tr>
</tbody>
</table>

Review the Difference between Independent and Dependent Samples (Go back to the beginning of the Guide and reread section I)

IV. Dependent Samples: Hypothesis tests for the average of the Differences and Confidence Interval Estimate for the average of the Differences (t-Distribution for Dependent Samples)

A. Assumptions
- The Paired Data values
- Match pairs of data values
- It must make sense to use the difference of paired values

B. Hypothesis Tests – Dependent Samples
   I. Kinds of Hypotheses Tests:
      Two-tailed
      $H_0: \mu_D = 0$
      $H_1: \mu_D \neq 0$
      
      Left-tailed
      $H_0: \mu_D = 0$
      $H_1: \mu_D < 0$
      
      Right-tailed
      $H_0: \mu_D = 0$
      $H_1: \mu_D > 0$

   Note: $\mu_D$ is the expected mean of the differences of the matched pairs.

II. Methods to Find the Average of the Differences and the Standard Deviation of the Difference
   a) Table Procedure – long method; not recommended unless you like a lot of work!
      1. Subtract the paired data values to find their differences, $D$.
      2. Find the mean of the differences $\overline{D}$.
      3. Find the standard deviation of the differences, $S_D$.
      4. Find the estimated standard error of the differences, $S_{\overline{D}}$.

      $S_{\overline{D}} = \sqrt{\frac{n\Sigma D^2 - (\Sigma D)^2}{n(n - 1)}}$
5. Find the test value, \( t \).

\[
t = \frac{(\bar{D} - \mu_D)}{(S_D/\sqrt{n})}
\]

*Note:* \( n \) is the number of data pairs.

b) Calculator Procedure (input the differences)
   1. Subtract each paired data set.
   2. Input the difference in L1.
   4. \( \bar{X} \) is \( \bar{D} \) and \( SX \) is \( S_D \).

c) Calculator Procedure (input the data)
   1. Input the After data in L1.
   2. Input the Before data in L2.
   3. While in the Edit mode, put your cursor on L3, i.e., move it above the first entry value.
   4. Set L3 = L1 – L2. Press ENTER. Note the L3 column should now have the values that are the differences in the L1 and L2 columns.
   5. Use STAT – CALC – one variable stats L3.
   6. \( \bar{X} \) is \( \bar{D} \) and \( SX \) is \( S_D \).

III. Test Statistic for the Paired Difference Hypothesis \( t \)-Test

\[
t = \frac{(\bar{D} - \mu_D)}{(S_D/\sqrt{n})}
\]

with \( d.f. = n - 1 \), where \( n \) is the number of data pairs.

where \( \bar{D} = \frac{\Sigma D}{n} \) is the average of the differences

where \( S_D = \sqrt{\frac{n\Sigma D^2 - (\Sigma D)^2}{n(n-1)}} \) is the standard deviation.

C. Confidence Interval for the Mean of the Difference

\[
\bar{D} - t \frac{S_D}{\sqrt{n}} < \mu_D < \bar{D} + t \frac{S_D}{\sqrt{n}} \quad \text{d.f.} = n - 1
\]

*Note:* First subtract the paired scores. Then, find the average of the differences and the standard deviation of the differences.
**Example:**

As an aid for improving students’ study habits, 9 students were randomly selected to attend a seminar on the importance of education in life. The table shows the number of hours each student studied per week before and after the seminar. At $\alpha = 0.10$, did attending the seminar increase the number of hours the students studied per week?

<table>
<thead>
<tr>
<th>Subject</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>Before</td>
<td>9</td>
<td>12</td>
<td>6</td>
<td>15</td>
<td>3</td>
<td>18</td>
<td>10</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>After</td>
<td>9</td>
<td>17</td>
<td>9</td>
<td>20</td>
<td>2</td>
<td>21</td>
<td>15</td>
<td>22</td>
<td>6</td>
</tr>
</tbody>
</table>

First, find the difference in each pair of data values. Then determine:

$$\Sigma D = -28 \quad \Sigma D^2 = 176 \quad \overline{D} = \frac{\Sigma D}{n} = -3.11$$

**Note:** The work can be done using a calculator or the table below.

<table>
<thead>
<tr>
<th>Before</th>
<th>After</th>
<th>$D$</th>
<th>$D^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>9</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>17</td>
<td>-5</td>
<td>25</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>-3</td>
<td>9</td>
</tr>
<tr>
<td>15</td>
<td>20</td>
<td>-5</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>21</td>
<td>-3</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>-5</td>
<td>25</td>
</tr>
<tr>
<td>13</td>
<td>22</td>
<td>-9</td>
<td>81</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

$$S_D = \sqrt{n \frac{\Sigma D^2 - (\Sigma D)^2}{n(n-1)}} \quad S_D = 3.33$$

$$S_D = \sqrt{(9 \cdot 176 - (-28)^2) \div (9 \cdot 8)} = 3.33$$

**Note 1:** These values can also be determined using a calculator. Please refer back to the calculator directions on page 10 part b and c. It will save a great deal of time in this section.

**Note 2:** If the program increases the number of hours a student studies, Before data minus After program data would have a negative average.

$$\overline{D} = -3.11 \quad S_D = 3.33$$

1. $H_0: \mu_D = 0 \quad H_1: \mu_D < 0$ (claim)
2. $t = \frac{D - \mu_D}{S_D / \sqrt{n}} = \frac{-3.11}{\frac{3.33}{\sqrt{9}}} = -3.11 \div \left(3.33 \div \sqrt{9}\right) = -2.8$
3. $c.v. = -1.397$ (left-tailed test negative c.v.) $d.f. = 8$
4. Since the test value -2.8 is smaller than the critical value -1.397, reject the null hypothesis.
5. The data does support an increase in the number of hours a student would study after attending the seminar.
Question 5

A composition teacher wishes to see whether the new grammar program will reduce grammatical errors in 2-page essays. Use the data below with $\alpha = 0.025$ to decide if the number of errors has been reduced.

<table>
<thead>
<tr>
<th>Student</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Errors before</td>
<td>12</td>
<td>9</td>
<td>0</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Errors after</td>
<td>9</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

*Note:* If the program is effective, the number of errors after the program would be less than the number of errors before the program. Thus, the average of the difference of the Before minus After should be positive.

$\overline{D} = 1.5 \quad S_D = 1.64$

$\Sigma D = 9 \quad \Sigma D^2 = 27 \quad \overline{D} = \frac{\Sigma D}{n} = \frac{9}{6} = 1.5$

<table>
<thead>
<tr>
<th>Before</th>
<th>After</th>
<th>$D$</th>
<th>$D^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>9</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

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

$S_D = \sqrt{(6 \cdot 27 - 9^2) ÷ (6 \cdot 5)} = 1.64$

*Note:* These values can also be determined using a calculator. Please refer back to the calculator directions on page 9 part b and c. It will save a great deal of time in this section.

$H_0: \mu_D = 0 \quad H_1: \mu_D > 0$
Question 6

A researcher compares the pulse rates of eight sets of identical twins to see whether there was a difference. The rates are given in the table in number of beats per minute. At $\alpha = 0.01$, is there a significant difference in the pulse rates of twins?

Use the p-value method.

<table>
<thead>
<tr>
<th>Before</th>
<th>After</th>
<th>D</th>
<th>$D^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>87</td>
<td>83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>86</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: These values can also be determined using a calculator. Please refer back to the calculator directions. It will save a great deal of time in this section.

V. Two Population Proportions

A. Test for Equality (of Proportions) Formula

$$H_0: p_1 = p_2 \quad H_1: p_1 \neq p_2 \quad \text{or} \quad H_1: p_1 > p_2 \quad \text{or} \quad H_1: p_1 < p_2$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\overline{p} \overline{q} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}{\overline{p} \overline{q} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}}$$

where $\overline{p} = \frac{(n_1\hat{p}_1 + n_2\hat{p}_2)}{(n_1 + n_2)} = \frac{(X_1 + X_2)}{(n_1 + n_2)}$

$\hat{p}_1 = \frac{X_1}{n_1}$ estimates $p_1$ \hspace{1cm} $\hat{p}_1$ and $\hat{p}_2$ are determined from samples

$p_2 = \frac{X_2}{n_2}$ estimates $p_2$ \hspace{1cm} $p_1$ and $p_2$ are values for the populations

Note: Since the proportions are assumed to be equal, their difference is zero and the values for $\hat{p}_1$ and $\hat{p}_2$ are pooled to form the standard error.
• **Example:**

Labor statistics indicate that 77% of cashiers and servers are women. A random sample in a large metropolitan area found that 112 of 150 cashiers and 150 of 200 servers were women. At the **0.05 level of significance**, is there a difference between the proportion of servers and the proportion of cashiers who are women?

1. \( H_0: p_1 = p_2 \) (claim)  \( H_1: p_1 \neq p_2 \)
   \[ \hat{p}_1 = \frac{X_1}{n_1} = \frac{112}{150} = 0.7467 \]
   \[ \hat{p}_2 = \frac{X_2}{n_2} = \frac{150}{200} = 0.75 \]

2. \[ \bar{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{112 + 150}{150 + 200} = 0.749 \]
   \[ q = 1 - \bar{p} = 1 - 0.749 = 0.251 \]

2. \[ z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p} \cdot q \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{(0.7467 - 0.75)}{\sqrt{0.749 \cdot 0.251 \left( \frac{1}{150} + \frac{1}{200} \right)}} = -0.07 \]

3. For \( \alpha = 0.05 \) with a *two-tailed test*, c.v. = \( \pm 1.96 \)

4. Since the test value = -0.07 is not further out in the tail past -1.96, do not reject the null hypothesis. Since the test value is negative, we checked the left tail.

5. The data does not reject or refute equal proportions of servers and cashiers are women.

---

**Question 7**

In a sample of 200 men, 130 said they used seat belts. In a sample of 300 women, 63 said they used seat belts. Test if men say they use safety belts more than women do. Use \( \alpha = 0.01 \) and the P-value method.

**Question 8**

A survey found that in a sample of 75 families, 26 own dogs. A survey done 15 years ago found that in a sample of 60 families, 26 own dogs. At \( \alpha = 0.05 \), has the proportion of dog owners changed over this 15 year period? Use the c.v. method.
B. Confidence Interval Estimate for the Difference between Two Proportions

Formula

\[ (\hat{p}_1 - \hat{p}_2) \pm z \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} \]

\[ \hat{p}_1 = \frac{x_1}{n_1}, \quad \hat{p}_2 = \frac{x_2}{n_2} \]

\[ \hat{q}_1 = 1 - \hat{p}_1, \quad \hat{q}_2 = 1 - \hat{p}_2, \quad \text{and} \quad n_1 \text{ and } n_2 \text{ are the respective sample sizes.} \]

Note: the proportions are not pooled to form the standard error since we are estimating their difference and are assuming that they are different.

Example:

Find a 95% confidence interval estimate for the difference in the proportions of male and female statistics student who work at least 20 hours each week.

\[ n_M = 15, \quad x_M = 9, \quad n_F = 28, \quad x_F = 16 \]

\[ \hat{p}_M = \frac{9}{15} = 0.6000, \quad \hat{p}_F = \frac{16}{18} = 0.5714 \]

\[ (0.6 - 0.5714) \pm 1.96 \sqrt{\frac{0.6 \cdot 0.4}{15} + \frac{0.5714 \cdot (1 - 0.5714)}{28}} \]

\[ (0.6 - 0.5714) \pm 1.96 \sqrt{\frac{0.6 \cdot 0.4}{15} + \frac{0.5714 \cdot 0.4286}{28}} \]

\[ 0.0286 \pm 0.3083 \]

\[-0.2797 < \hat{p}_M - \hat{p}_F < 0.3369 \]

\[-28.0\% < \hat{p}_M - \hat{p}_F < 33.7\% \]

Question 9

A nutritionist wants to estimate the difference in the proportion of individuals who have at most an eighth grade education and consume more than the USDA allowance of 300 mg of cholesterol and the proportion of individuals who have at least some college and consume too much cholesterol. From his interviews with 320 randomly selected individuals with at most an eighth grade education, he finds 114 consume too much cholesterol. From his interviews with 350 individuals with at least some college, he finds 112 of them consume too much cholesterol. Construct a 90% confidence interval estimate for the difference between the two proportions.

\[ n_E = 320, \quad x_E = 114, \quad n_C = 350, \quad x_C = 112 \]

\[ \hat{p}_E = \frac{114}{320} = 0.3563, \quad \hat{p}_C = \frac{112}{350} = 0.3200 \]
VI. Equality of Two Variances

A. Characteristics of the $F$ Distribution

1. Never negative, because variances are always positive or zero.
2. Positively skewed.
3. The mean value of $F$ is approximately equal to 1.
4. The $F$ Distribution is a family of curves based on the degrees of freedom of the variance of the numerator and the degrees of freedom of the variance of the denominator.

B. The $F$ Distribution to Test the Equality of Two Variances

- Both samples are randomly drawn from normally distributed populations AND the samples are independent.
- The $F$ distribution is the ratio of two sample variances. **In order to use the tables in the textbook, always call $S_1$ the larger variance.**

I. Formula: always use the larger variance for $S_1$

$$F = \frac{S_1^2}{S_2^2}$$

Degrees of freedom for the numerator: $(n_1 – 1)$
Degrees of freedom for the denominator: $(n_2 – 1)$

Where $n_1$ is the sample size for the sample with the larger variance and $n_2$ is the sample size for the sample with the smaller variance.

II. Notes and Reminders for the Use of the $F$ Test

- The larger variance should always be designated as $S_1^2$ and be placed in the numerator of the formula.
- For $F$, the critical value is always on the right.
- The mode is approximately at 1.0.
- For a two-tailed test, divide $\alpha$ by 2.
- The critical value is always on the right side of the $F$ curve.
• If the standard deviations instead of the variances are given in the problem, square them in the formula for the $F$ test.

• When the degrees of freedom are not in Table H, use the closest (smaller) value.

III. Critical Values for $F$ are in Table H

• Textbook Reference: Eighth Edition
  - pg. 773 for $\alpha = 0.005$, pg. 774 for $\alpha = 0.01$
  - pg. 775 for $\alpha = 0.25$, pg. 776 for $\alpha = 0.50$
  - pg. 777 for $\alpha = 0.10$

• For a two-tailed test, use the $\alpha/2$ table.

• For the one-tailed test, use the $\alpha$ table.

• Textbook Reference: Seventh Edition
  - pg. 786 for $\alpha = 0.005$, pg. 787 for $\alpha = 0.01$
  - pg. 788 for $\alpha = 0.025$, pg. 789 for $\alpha = 0.05$
  - pg. 790 for $\alpha = 0.10$

• Use the degrees of freedom for the numerator $(n_1 - 1)$ to find the column and the degrees of freedom for the denominator $(n_2 - 1)$ to find the row.

• The critical region is always on the right side of the $F$ curve.

IV. P-value Method for the $F$-Distribution

Since the larger variance is in the numerator, the test value is always larger than one on the $F$ graphs.

• For a one-tailed test, use:
  
  $p = P(F > \text{test value})$ (table)
  
  $p = Fcdf(test\ value, 1000, d.f.\ N., d.f.\ D)$(calculator)

• For a two-tailed test, use:
  
  $p = 2\ P(F > \text{test value})$ (table)
  
  $p = 2\ Fcdf(test\ value, 1000, d.f.\ N., d.f.\ D)$(calculator)

Be careful to use the degrees of freedom for the sample with the larger variance across the top of the table and list it first in the calculator.
Example:
The PSU bookstore was checking book prices. Assume book prices are normally distributed. The standard deviation for the price of 15 math books is 6.17. The standard deviation for the 25 history books is 13.12. At $\alpha = 0.10$, can we conclude that there is a difference in the variances?

1. $H_0: \sigma_1^2 = \sigma_2^2 \quad H_1: \sigma_1^2 \neq \sigma_2^2$ (claim)

2. $F = \frac{s_1^2}{s_2^2} = \frac{13.12^2}{6.17^2} = 4.52$

Note: Both 13.12 and 6.17 are squared since they are standard deviations. The larger variance is placed in the numerator.

3. Since this is a two-tailed test, use the table for $\alpha = \frac{0.10}{2} = 0.05$ to find the c.v.

4. $cv$ method: reject the null hypothesis because the test value $4.52 > cv = 2.35$ or the test value is in the rejection region.

P-value method: $p = 2Fcdf (4.52, 100, 24, 14) = .0051 < .10 = \alpha$

Reject the hypothesis.

5. The data is sufficient to support a difference in the variances of the prices of math and history books.

Question 10
A tax collector wishes to see if the variances of the values of tax-exempt properties are different for two large cities. The values of the tax-exempt properties in millions of dollars for the samples are shown. Assume these prices are normally distributed. At $\alpha = 0.05$, are the variances different? Use either the p-value method or the critical value method.

<table>
<thead>
<tr>
<th>City A</th>
<th>City B</th>
</tr>
</thead>
<tbody>
<tr>
<td>113</td>
<td>82</td>
</tr>
<tr>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>44</td>
<td>12</td>
</tr>
<tr>
<td>31</td>
<td>20</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>19</td>
<td>16</td>
</tr>
<tr>
<td>7</td>
<td>81</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
</tr>
</tbody>
</table>

$S_A = 27.88 \quad S_B = 72.74 \quad n_A = 14 \quad n_B = 16$

Note: These values are determined for each group separately using a calculator.
VII. Summary

- Means, variances, and proportions are population parameters that are often compared.

- The comparison of means can be made with the z-test if the samples are independent, the population variances are known, and the sample sizes are either larger than 30 or the samples are drawn from normally distributed populations.

- If either the population variance is not known and if one or both sample sizes are less than 30 then that sample is drawn from a normally distributed population, use the t-test.

- For dependent samples, use the average of the difference with the t-test.

- A z-test is used to compare two proportions.

- For independent samples drawn from normally distributed populations, the F test is used to determine whether or not the variances or standard deviations are equal.

Works Cited
Answer: Question 1
Use the data below from the “Ages and Travel Study” to test the claim that the average age for domestic travel is greater than the average age for foreign travel. The population standard deviations and the data averages are below. Use $\alpha = 0.02$ and the critical value method.

<table>
<thead>
<tr>
<th>Domestic</th>
<th>Foreign</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{X}_1 = 61.2$ yrs</td>
<td>$\bar{X}_2 = 59.4$ yrs</td>
</tr>
<tr>
<td>$\sigma_1 = 5.9$</td>
<td>$\sigma_1 = 7.9$</td>
</tr>
<tr>
<td>$n_1 = 84$</td>
<td>$n_2 = 34$</td>
</tr>
</tbody>
</table>

1. $H_0: \mu_1 = \mu_2$  $H_0: \mu_1 > \mu_2$ (claim)

2. $z = \frac{(\bar{X}_1 - \bar{X}_2)(\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(61.2 - 59.4) - 0}{\sqrt{\frac{5.9^2}{84} + \frac{7.9^2}{34}}} = 1.20$

3. c.v. = 2.05
4. Since the test value 1.20 is not larger than c.v. = 2.05, fail to reject the null hypothesis
5. The data does not support an older domestic traveling group

Answer: Question 2
A study comparing attitudes toward death was conducted in which organ donors (individuals who had signed organ donor cards) were compared with non-donors. The study is reported in the journal *Death Studies*. Templer’s Death Anxiety Scale (DAS) was administered to both groups. On this scale, high scores indicate high anxiety concerning death. The results were reported as follows:

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Pop Std Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organ Donors</td>
<td>35</td>
<td>5.36</td>
<td>2.91</td>
</tr>
<tr>
<td>Non-Organ Donors</td>
<td>69</td>
<td>7.62</td>
<td>3.45</td>
</tr>
</tbody>
</table>

Find a 90% confidence interval estimate for the differences in the average ratings on the Death Anxiety Scale for the organ donors and non–organ donors.

*Note:* Use $z$ since the population standard deviation is given for both groups and both sample sizes are larger than 30.

$$(7.62 - 5.36) \pm 1.65 \sqrt{\frac{3.45^2}{69} + \frac{2.91^2}{35}}$$

$2.26 \pm 1.06$

$1.20 < \mu_m - \mu_w < 3.32$
Answer: Question 3

Seth Young thinks that people walk faster when they are departing (getting on a plane) for a trip than when they are returning from their trip (getting off a plane). Use the findings for the randomly selected passengers in each category summarized in the table that follows to conduct the test. Use a 0.05 level of significance.

<table>
<thead>
<tr>
<th>Direction of Travel</th>
<th>Departure</th>
<th>Arrival</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean speed (ft per mn)</td>
<td>269</td>
<td>260</td>
</tr>
<tr>
<td>St Deviation (ft per mn)</td>
<td>53</td>
<td>34</td>
</tr>
<tr>
<td>Sample size</td>
<td>15</td>
<td>20</td>
</tr>
</tbody>
</table>

1. \( H_0: \mu_D = \mu_A \) \( H_1: \mu_D > \mu_A \) (claim)

2. \( t = \frac{\overline{x}_D - \overline{x}_A}{\sqrt{\frac{s_D^2}{n_D} + \frac{s_A^2}{n_A}}} = 0.57 \)

3. \( df = smaller\{14,19\} = 14 \) \( c.v. = +1.761 \)

4. c.v. method: the test value = .57 is not greater than c.v. = 1.761 the test value is not in the rejection region. Do not reject the null hypothesis.

5. This data does not support people walking faster when they are departing (getting on a plane) for a trip than when they are returning from their trip (getting off a plane).

Answer: Question 4

The times (in minutes) it took six white mice to learn to run a simple maze and the times it took six brown mice to learn to run the same maze are given here. Find the 95% confidence interval for the difference of the average times for white/brown mice.

<table>
<thead>
<tr>
<th>White mice</th>
<th>18</th>
<th>24</th>
<th>20</th>
<th>13</th>
<th>15</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brown mice</td>
<td>25</td>
<td>16</td>
<td>19</td>
<td>14</td>
<td>16</td>
<td>10</td>
</tr>
</tbody>
</table>

\[
(\overline{X}_B - \overline{X}_W) \pm t \sqrt{\frac{s_B^2}{n_B} + \frac{s_W^2}{n_W}}
\]

\[
(16.67 - 17) \pm 2.571 \sqrt{\frac{5.05^2}{6} + \frac{4.56^2}{6}}
\]

\[-.33 \pm 7.14\]

\[-7.5 < \mu_B - \mu_W < 6.8\]
Answer: Question 5
A composition teacher wishes to see whether the new grammar program will reduce grammatical errors in 2-page essays. Use the data below with $\alpha = 0.025$ to decide if number of errors has been reduced.

<table>
<thead>
<tr>
<th>Student</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Errors before</td>
<td>12</td>
<td>9</td>
<td>0</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Errors after</td>
<td>9</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

**Note:** If the program is effective, the number of errors after the program would be less than the number of errors before the program. Thus, the average of the difference of the Before minus After should be positive.

$\bar{D} = 1.5$  \hspace{1cm} $S_D = 1.64$  

\[ \sum D = 9 \quad \sum D^2 = 27 \quad \bar{D} = \frac{\sum D}{n} = \frac{9}{6} = 1.5 \]

<table>
<thead>
<tr>
<th>Before</th>
<th>After</th>
<th>D</th>
<th>$D^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>9</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ S_D = \sqrt{\frac{n\sum D^2 - (\sum D)^2}{n(n-1)}} \]

\[ S_D = \sqrt{\left( (6 \cdot 27 - 9^2) / (6 \cdot 5) \right) } = 1.64 \]

**Note:** These values can also be determined using a calculator. Please refer back to the calculator directions. It will save a great deal of time in this section.

1. $H_0: \mu_D = 0 \quad H_1: \mu_D > 0$ (claim)
2. $t = \frac{\bar{D} - \mu_D}{(S_D / \sqrt{n})} = \frac{1.5 - 0}{(1.64 / \sqrt{6})} = 1.5 \div \left( \frac{1.64}{\sqrt{6}} \right) = 2.24$
3. $c.v. = 2.571 \quad d.f. = 5$
4. Since the test value 2.24 is not greater than the c.v. = 2.571, do not reject the null hypothesis.
5. **The data does not support a reduction in errors.**
Answer: Question 6

A researcher compares the pulse rates of eight sets of identical twins to see whether there was a difference. The rates are given in the table in number of beats per minute. At $\alpha = 0.01$, is there a significant difference in the pulse rates of twins? Use the p-value method.

$\Sigma D = 10 \quad \Sigma D^2 = 104 \quad \overline{D} = \frac{\Sigma D}{n} = \frac{10}{8} = 1.25$

<table>
<thead>
<tr>
<th>Before</th>
<th>After</th>
<th>D</th>
<th>$D^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>87</td>
<td>83</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>92</td>
<td>95</td>
<td>-3</td>
<td>9</td>
</tr>
<tr>
<td>78</td>
<td>79</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>83</td>
<td>83</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>88</td>
<td>86</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>90</td>
<td>93</td>
<td>-3</td>
<td>9</td>
</tr>
<tr>
<td>84</td>
<td>80</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>93</td>
<td>86</td>
<td>7</td>
<td>49</td>
</tr>
</tbody>
</table>

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

$S_D = \sqrt{(8 \cdot 104 - 10^2) \div (8 \cdot 7)} = 3.62$

$\overline{D} = 1.25 \quad S_D = 1.64$

Note: These values can also be determined using a calculator. Please refer back to the calculator directions. It will save a great deal of time in this section.

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

2. $t = \frac{D-\mu_D}{S_D/\sqrt{n}} = \frac{1.25-0}{\frac{3.62}{\sqrt{8}}} = 1.25 \div \left(3.62 \div \sqrt{18}\right) = 0.977$

3. $0.20 < P$-value $< 0.50$ (table)

$P = 2tcdf(.98, 100, 7) = .3587$ (calculator)

4. Since $p$ is not less than 0.01, do not reject the null hypothesis.

5. The data does not support a difference in the pulse rates of twins.
Answer: Question 7
In a sample of 200 men, 130 said they used seat belts. In a sample of 300 women, 63 said they used seat belts. Test if men say they use safety belts more than women do. Use $\alpha = 0.01$ and the P-value method.

1. $H_0: P_1 = P_2 \quad H_1: P_1 > P_2$ (claim)

2. $\hat{p}_1 = \frac{X_1}{n_1} = \frac{130}{200} = 0.65 \quad \hat{p}_2 = \frac{X_2}{n_2} = \frac{63}{300} = 0.21$

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{(130 + 63)}{(200 + 300)} = 0.386 \quad \bar{q} = 1 - \bar{p} = 1 - 0.386 = 0.614$$

$$z = \frac{(0.65 - 0.21)}{\sqrt{0.386 \cdot 0.614 \left( \frac{1}{200} + \frac{1}{300} \right)}} = 9.90$$

3. P-value = $P (z > 9.90) = .0001$ (table)

4. Since $p$ is less than the level of significance equal to 0.01, reject the null hypothesis.

5. The data supports a more frequent use of safety belts by men than women.

Answer: Questions 8
A survey found that in a sample of 75 families, 26 own dogs. A survey done 15 years ago found that in a sample of 60 families, 26 own dogs. At $\alpha = 0.05$, has the proportion of dog owners changed over this 15 year period? Use the c.v. method.

1. $H_0: P_1 = P_2 \quad H_1: P_1 \neq P_2$ (claim)

2. \[
z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\bar{p} \bar{q} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{(0.3467 - 0.4333)}{\sqrt{0.3852 \cdot 0.6148 \left( \frac{1}{75} + \frac{1}{60} \right)}} = -1.0274
\]

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{(26 + 26)}{(75 + 60)} = 0.3852$$

$$\bar{q} = 1 - \bar{p} = 1 - 0.3852 = 0.6148$$

$$\hat{p}_1 = \frac{26}{75} = 0.3467 \quad \hat{p}_2 = \frac{26}{60} = 0.4333$$

3. For $\alpha = 0.05$ with a two-tailed test, c.v. = ± 1.96

4. Since the test value -1.0274 is not further out in the tail past c.v. = -1.96, do not reject the null hypothesis.

5. The data does not support a change in the proportions of families with dogs during the last 15 years.
Answer: Question 9

A nutritionist wants to estimate the difference in the proportion of individuals who have at most an eighth grade education and consume more than the USDA allowance of 300 mg of cholesterol and the proportion of individuals who have at least some college and consume too much cholesterol. From his interviews with 320 randomly selected individuals with at most an eighth grade education, he finds 114 consume too much cholesterol. From his interviews with 350 individuals with at least some college, he finds 112 of them consume too much cholesterol. Construct a 90% confidence interval estimate for the difference between the two proportions.

\[ n_E = 320 \quad x_E = 114 \quad n_C = 350 \quad x_C = 112 \]

\[ \hat{p}_E = \frac{114}{320} = 0.3563 \quad \hat{p}_C = \frac{112}{350} = 0.3200 \]

\[ (.3563 - .3200) \pm 1.645 \sqrt{\frac{.3563 \cdot (1-.3563)}{320} + \frac{.3200 \cdot (1-.3200)}{350}} \]

\[ (.036 \pm .060) \]

\[ -0.024 < \hat{p}_E - \hat{p}_C < 0.096 \]

\[ -2.40\% < \hat{p}_E - \hat{p}_C < 9.6\% \]
Answer: Question 10

A tax collector wishes to see if the variances of the values of tax-exempt properties are different for two large cities. The values of the tax-exempt properties in millions of dollars for the samples are shown. At $\alpha = 0.05$, are the variances different?

<table>
<thead>
<tr>
<th>City A</th>
<th>City B</th>
</tr>
</thead>
<tbody>
<tr>
<td>113</td>
<td>82</td>
</tr>
<tr>
<td>25</td>
<td>295</td>
</tr>
<tr>
<td>44</td>
<td>12</td>
</tr>
<tr>
<td>31</td>
<td>20</td>
</tr>
</tbody>
</table>

$S_A = 27.88$  $S_B = 72.74$  $n_A = 14$  $n_B = 16$

Critical Value Method:

1. $H_0: \sigma_A^2 = \sigma_B^2$  $H_1: \sigma_A^2 \neq \sigma_B^2$

2. $F = \frac{s_1^2}{s_2^2} = \frac{72.74^2}{27.88^2} = 6.81$

   Notice both standard deviations are squared and the larger (72.74) is placed in the numerator (top of the fraction).

3. For a two-tailed test, $\alpha/2 = .025$  $d.f.N. = 15$  $d.f.D = 13$

4. The test value is in the rejection region. Reject the null hypothesis.

5. The data supports a difference in the variances of tax exempt properties for these two cities.

P-value Method:

$S_1 = 72.74$  $S_2 = 25.97$

1. $H_0: \sigma_A^2 = \sigma_B^2$  $H_1: \sigma_A^2 \neq \sigma_B^2$

2. $F = \frac{s_1^2}{s_2^2} = \frac{72.74^2}{27.88^2} = 6.81$

3. $P = 2P(F > 6.81) = 2Fcdf(6.81,1000,15,13) = .0013$

4. Since $p = .0013 < \alpha = .05$, reject the null hypothesis.

5. The data supports a difference in the variances for tax-exempt properties in these two cities.