Chapter 11  Asking and Answering Questions about the
Difference between Two Population Proportions

Section 11.1 Exercise Set 1

11.1: (a) Question type (Q): Estimation

Study type (S): Sample data

Type of data (T): One categorical variable

Number of samples or treatments (N): Two samples (one from the population of men and from the population of women)

This particular combination of answers to QSTN suggests that a large-sample confidence interval for a difference of two population proportions is appropriate.

(b) Using the 5-step process (EMC³):

Estimate (E): The difference in proportions living in poverty for men and women, \( p_1 - p_2 \), will be estimated, where \( p_1 \) is the true proportion of men living in poverty and \( p_2 \) is the true proportion of women living in poverty.

Method (M): Because the answers to the four key questions are (Q) estimation, (S) sample data, (T) one categorical variable, and (N) two samples, we can construct and interpret a 90% large-sample confidence interval for the difference in proportions living in poverty for men and women.

Check (C): We are told that the data are from large representative samples of men and women. In addition, \( n_1 \hat{p}_1 = 1,200(0.125) = 150 \), \( n_1 (1 - \hat{p}_1) = 1,200(1 - 0.125) = 1,050 \), \( n_2 \hat{p}_2 = 1,000(0.151) = 151 \), and \( n_2 (1 - \hat{p}_2) = 1,000(1 - 0.151) = 849 \), which are all at least 10. The two required conditions are satisfied.
Calculations (C):

\[
\left( \hat{p}_1 - \hat{p}_2 \right) \pm \left( z \text{ critical value} \right) \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \\
(0.125 - 0.151) \pm 1.645 \sqrt{\frac{0.125(1-0.125)}{1200} + \frac{0.151(1-0.151)}{1000}}
\]

\[-0.026 \pm 1.645(0.01481) \\
-0.026 \pm 0.024363 \\
(-0.050, -0.002) \]

Communicate Results (C):

Interpret confidence interval: We are 90% confident that the actual difference in proportions living in poverty for men and women is somewhere between –0.0504 and –0.0016. Because both endpoints of the confidence interval are negative, we believe that the percent of women living in poverty is greater than the percent of men living in poverty by somewhere between 0.2% and 5.0%.

Interpret confidence level: The method used to construct this interval estimate is successful in capturing the actual value of the difference in population proportions 90% of the time.

11.2:  (a) Yes. Let \( \hat{p}_1 = 0.20 \) and \( \hat{p}_2 = 0.15 \) and \( n_1 = n_2 = 1,112 \). Since

\[
n_1 \hat{p}_1 = 1,112(0.20) = 222.4, \quad n_1 (1 - \hat{p}_1) = 1,112(1 - 0.20) = 889.6, \\
n_2 \hat{p}_2 = 1,112(0.15) = 166.8, \quad \text{and} \quad n_2 (1 - \hat{p}_2) = 1,112(1 - 0.15) = 945.2
\]

are all at least 10, the sample sizes are large enough to use the large-sample confidence interval.

(b) The difference in proportions of Americans ages 12 and older who owned MP3 player in 2006 and the corresponding proportion for 2005, \( p_1 - p_2 \), will be estimated, where \( p_1 \) is the true proportion of Americans ages 12 and older who owned MP3 player in 2006 and \( p_2 \) is the true proportion of Americans ages 12 and older who owned MP3 player in 2005. Let \( \hat{p}_1 = 0.20 \) and \( \hat{p}_2 = 0.15 \).
\[(\hat{p}_1 - \hat{p}_2) \pm (z \text{ critical value}) \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\]

\[(0.20 - 0.15) \pm 1.96 \sqrt{\frac{0.20(1-0.20)}{1,112} + \frac{0.15(1-0.15)}{1,112}}\]

\[0.05 \pm 1.96(0.016079)\]

\[0.05 \pm 0.031515\]

\[(0.018, 0.082)\]

(c) Zero is not included in the confidence interval. This means that you can be confident that the proportion of Americans ages 12 and older who owned an MP3 player was greater in 2006 than in 2005 by somewhere between 0.018 and 0.082.

(d) We are 95% confident that the actual difference in proportions of Americans ages 12 and older who owned MP3 player in 2006 and the corresponding proportion for 2005 is somewhere between 0.018 and 0.082. Because both endpoints of the confidence interval are positive, we believe that the percent of Americans ages 12 and older who owned MP3 player in 2006 is greater than the corresponding proportion in 2005 by somewhere between 1.8% and 8.2%.

11.3: (a) Using the 5-step process (EMC³):

Estimate (E): The difference between the proportion of avid mountain bikers with low sperm count and the proportion for nonbikers, \( p_1 - p_2 \), will be estimated, where \( p_1 \) is the true proportion of avid mountain bikers with low sperm count and \( p_2 \) is the true proportion nonbikers with low sperm count.

Method (M): Because the answers to the four key questions are (Q) estimation, (S) sample data, (T) one categorical variable, and (N) two samples, we can construct and interpret a 95% large-sample confidence interval for the difference in proportions of avid mountain bikers with low sperm count and the proportion for nonbikers.

Check (C): We are told that the samples are representative of the populations of avid mountain bikers and nonbikers. In addition, \( n_1 \hat{p}_1 = 100(0.9) = 90 \), \( n_1 (1-\hat{p}_1) = 100(1-0.9) = 10 \), \( n_2 \hat{p}_2 = 100(0.26) = 26 \), and \( n_2 (1-\hat{p}_2) = 100(1-0.26) = 74 \), which are all at least 10. The two required conditions are satisfied.
Calculations (C):

\[
(\hat{p}_1 - \hat{p}_2) \pm (z \text{ critical value}) \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}
\]

\[
(0.90 - 0.26) \pm 1.96 \sqrt{\frac{0.90(1-0.90)}{100} + \frac{0.26(1-0.26)}{100}}
\]

\[
0.64 \pm 1.96(0.053141)
\]

\[
0.64 \pm 0.104157
\]

\[
(0.536, 0.744)
\]

Communicate Results (C):

Interpret confidence interval: We are 95% confident that the actual difference in the proportion of avid bikers with low sperm count and nonbikers with low sperm count is somewhere between 0.536 and 0.744. Because both endpoints of the confidence interval are positive, we believe that the percent of avid bikers with low sperm count is greater than the percent of nonbikers with low sperm count by somewhere between 53.6% and 74.4%.

Interpret confidence level: The method used to construct this interval estimate is successful in capturing the actual value of the difference in population proportions 95% of the time.

(b) No, because this was an observational study and it is not a good idea to draw cause-and-effect conclusions from an observational study.

Section 11.1 Exercise Set 2

11.4: Using the 5-step process (EMC³):

Estimate (E): The difference in the proportion of men who expect to get a raise or promotion this year and this proportion for women, \( p_1 - p_2 \), will be estimated, where \( p_1 \) is the true proportion of men who expect to get a raise or promotion this year and \( p_2 \) is the true proportion of women who expect to get a raise or promotion this year.

Method (M): Because the answers to the four key questions are (Q) estimation, (S) sample data, (T) one categorical variable, and (N) two samples, we can construct and interpret a 95% large-sample confidence interval for the difference in proportions of men and women who expect to get a raise or promotion this year.

Check (C): We are told that the two samples were independently selected representative samples. In addition, \( n_1\hat{p}_1 = 507(0.52) = 263.64 \), \( n_1(1 - \hat{p}_1) = 507(1 - 0.52) = 243.36 \),
\( n_2p_2 = 507(0.37) = 187.59 \), and \( n_2(1 - \hat{p}_2) = 507(1 - 0.37) = 319.41 \), which are all at least 10. The two required conditions are satisfied.

Calculations (C):

\[
(\hat{p}_1 - \hat{p}_2) \pm (z \text{ critical value}) \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}
\]

\[
(0.52 - 0.37) \pm 1.96 \sqrt{\frac{0.52(1 - 0.52)}{507} + \frac{0.37(1 - 0.37)}{507}}
\]

\[
0.15 \pm 1.96(0.030856)
\]

\[
0.15 \pm 0.060478
\]

\[
(0.0895, 0.2105)
\]

Communicate Results (C):

Interpret confidence interval: We are 95% confident that the actual difference in proportion of men who expect to get a raise or promotion this year and this proportion for women is somewhere between 0.0895 and 0.2105. Because both endpoints of the confidence interval are positive, we believe that the percent of men who expect to get a raise or promotion this year is greater than the percent of women who expect to get a raise or promotion this year by somewhere between 8.95% and 21.05%.

Interpret confidence level: The method used to construct this interval estimate is successful in capturing the actual value of the difference in population proportions 95% of the time.

11.5: (a) Yes. Let \( \hat{p}_1 = 0.43 \) and \( \hat{p}_2 = 0.25 \) and \( n_1 = n_2 = 200 \). Since \( n_1\hat{p}_1 = 200(0.43) = 86 \), \( n_1(1 - \hat{p}_1) = 200(1 - 0.43) = 114 \), \( n_2\hat{p}_2 = 200(0.25) = 50 \), and \( n_2(1 - \hat{p}_2) = 200(1 - 0.25) = 150 \) are all at least 10, the sample sizes are large enough to use the large-sample confidence interval.

(b) Using the 5-step process (EMC³):

Estimate (E): The difference in the proportion of people with a college degree who reported sunburn and the corresponding proportion for those without a high school degree, \( p_1 - p_2 \), will be estimated, where \( p_1 \) is the true proportion of people with a college degree who report sunburn and \( p_2 \) is the true proportion of those without a high school degree who report sunburn.

Method (M): Because the answers to the four key questions are (Q) estimation, (S) sample data, (T) one categorical variable, and (N) two samples, we can construct and
interpret a 90% large-sample confidence interval for the difference in the proportion of people with a college degree who reported sunburn and the corresponding proportion for those without a high school degree.

Check (C): We are told to suppose that the two samples were independently and randomly selected. In addition, \( \hat{p}_1 = 0.43 \) and \( \hat{p}_2 = 0.25 \) and \( n_1 = n_2 = 200 \). Since 
\[
\begin{align*}
1 \cdot n_1 \hat{p}_1 = 200(0.43) = 86 \quad n_1 (1-\hat{p}_1) = 200(1-0.43) = 114 \quad n_2 \hat{p}_2 = 200(0.25) = 50 \\
\text{and} \\
n_2 (1-\hat{p}_2) = 200(1-0.25) = 150,
\end{align*}
\]
which are all at least 10. The two required conditions are satisfied.

Calculations (C):
\[
\begin{align*}
(\hat{p}_1 - \hat{p}_2) \pm (z \text{ critical value}) \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \\
(0.43 - 0.25) \pm 1.645 \sqrt{\frac{0.43(1-0.43)}{200} + \frac{0.25(1-0.25)}{200}} \\
0.18 \pm 1.645(0.046508) \\
0.18 \pm 0.076506 \\
(0.1035, 0.2565)
\end{align*}
\]

Communicate Results (C):

Interpret confidence interval: We are 90% confident that the actual difference in the proportion of people with a college degree who reported sunburn and the corresponding proportion for those without a high school degree is somewhere between 0.1035 and 0.2565. Because both endpoints of the confidence interval are positive, we believe that the percent of people with a college degree who reported sunburn is greater than the corresponding percent for those without a high school degree by somewhere between 10.35 and 25.65 percentage points.

Interpret confidence level: The method used to construct this interval estimate is successful in capturing the actual value of the difference in population proportions 90% of the time.

(c) Zero is not included in the confidence interval. Therefore, because both endpoints of the confidence interval are positive, we believe that the percent of people with a college degree who reported sunburn is greater than the corresponding percent for those without a high school degree by somewhere between 10.35 and 25.65 percentage points.
(d) We are 90% confident that the actual difference in the proportion of people with a college degree who reported sunburn and the corresponding proportion for those without a high school degree is somewhere between 0.1035 and 0.2565.

11.6: Using the 5-step process (EMC³):

**Estimate (E):** The difference in the proportion of students who think that understanding science and having math skills are essential and this proportion for parents, \( p_1 - p_2 \), will be estimated, where \( p_1 \) is the true proportion of students who think that understanding science and having math skills are essential and \( p_2 \) is the true proportion of parents who think that understanding science and having math skills are essential.

**Method (M):** Because the answers to the four key questions are (Q) estimation, (S) sample data, (T) one categorical variable, and (N) two samples, we can construct and interpret a 95% large-sample confidence interval for the difference in the proportion of students who think that understanding science and having math skills are essential and this proportion for parents.

**Check (C):** We are told that the two samples (students and parents) were independently selected random samples. In addition, 
\[
\begin{align*}
n_1 \hat{p}_1 &= 1,342(0.50) = 671, \\
n_1 (1 - \hat{p}_1) &= 1,342(1-0.50) = 671, \\
n_2 \hat{p}_2 &= 1,379(0.62) = 854.98, \\
n_2 (1 - \hat{p}_2) &= 1,379(1-0.62) = 524.02,
\end{align*}
\]
which are all at least 10. The two required conditions are satisfied.

**Calculations (C):**

\[
\left( \hat{p}_1 - \hat{p}_2 \right) \pm z \text{ critical value} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}
\]

\[
\begin{align*}
(0.50 - 0.62) &\pm 1.96 \sqrt{\frac{0.50(1-0.50)}{1,342} + \frac{0.62(1-0.62)}{1,379}} \\
&= -0.12 \pm 1.96(0.018898) \\
&= -0.12 \pm 0.03704 \\
&= (-0.15704, -0.08296)
\end{align*}
\]

**Communicate Results (C):**

Interpret confidence interval: We are 95% confident that the actual difference in the proportion of students who think that understanding science and having math skills are essential and the proportion of parents who think that understanding science and having math skills are essential is somewhere between –0.157 and –0.0830. Because both endpoints of the confidence interval are negative, we believe that the percent of
parents who think that understanding science and having math skills are essential is greater than this percent for students by somewhere between 8.3 and 15.7 percentage points.

Interpret confidence level: The method used to construct this interval estimate is successful in capturing the actual value of the difference in population proportions 95% of the time.

Section 11.1 Additional Exercises

11.7: Using the 5-step process (EMC³):

Estimate (E): The difference between the proportion of high school students who believed marijuana was very distracting or extremely distracting in 2009 and this proportion in 2011, \( p_1 - p_2 \), will be estimated, where \( p_1 \) is the true proportion of high school students who believed marijuana was very distracting or extremely distracting in 2009 and \( p_2 \) is the true proportion of high school students who believed marijuana was very distracting or extremely distracting in 2011.

Method (M): Because the answers to the four key questions are (Q) estimation, (S) sample data, (T) one categorical variable, and (N) two samples, we can construct and interpret a 99% large-sample confidence interval for the difference in proportions of high school students who believed marijuana was very distracting or extremely distracting in 2009 and this proportion in 2011.

Check (C): We are told that the samples are representative of the populations high school students in 2009 and 2011. In addition, \( n_1 \hat{p}_1 = 2,300(0.78) = 1794 \), \( n_1 (1 - \hat{p}_1) = 2,300(1-0.78) = 506 \), \( n_2 \hat{p}_2 = 2,294(0.70) = 1605.8 \), and \( n_2 (1 - \hat{p}_2) = 2,294(1-0.70) = 688.2 \), which are all at least 10. The two required conditions are satisfied.
Calculated (C):

\[
(\hat{p}_1 - \hat{p}_2) \pm (z \text{ critical value}) \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}
\]

\[
(0.78 - 0.70) \pm 2.58 \sqrt{\frac{0.78(1-0.78)}{2,300} + \frac{0.70(1-0.70)}{2,294}}
\]

\[
0.08 \pm 2.58(0.01289)
\]

\[
0.08 \pm 0.033256
\]

\[
(0.047, 0.113)
\]

Communicate Results (C):

Interpret confidence interval: We are 99% confident that the actual difference in the proportion of high school students who believed marijuana was very distracting or extremely distracting in 2009 and this proportion in 2011 is somewhere between 0.047 and 0.113. Because both endpoints of the confidence interval are positive, we believe that the percent of high school students who believed marijuana was very distracting or extremely distracting in 2009 is greater than the percent of high school students who believed marijuana was very distracting or extremely distracting in 2011 by somewhere between 4.7% and 11.3%.

Interpret confidence level: The method used to construct this interval estimate is successful in capturing the actual value of the difference in population proportions 99% of the time.

11.8: Using the 5-step process (EMC³):

Estimate (E): The difference in the proportion of college graduates who were unemployed in 2008 and the proportion of college graduates who were unemployed in 2009, \( p_1 - p_2 \), will be estimated, where \( p_1 \) is the true proportion of college graduates who were unemployed in 2008 and \( p_2 \) is the true proportion of college graduates who were unemployed in 2009.

Method (M): Because the answers to the four key questions are (Q) estimation, (S) sample data, (T) one categorical variable, and (N) two samples, we can construct and interpret a 95% large-sample confidence interval for the difference in the proportion of college graduates who were unemployed in 2008 and the proportion of college graduates who were unemployed in 2009.

Check (C): We are told that the two samples are independently selected representative samples. In addition, \( n_1 \hat{p}_1 = 500(0.026) = 13 \), \( n_1 (1- \hat{p}_1) = 500(1-0.026) = 487 \),
\[ n_2 p_2 = 500(0.046) = 23, \text{ and } n_2(1 - \hat{p}_2) = 500(1 - 0.046) = 477, \text{ which are all at least 10.} \]
The two required conditions are satisfied.

Calculations (C):

\[
\left( \hat{p}_1 - \hat{p}_2 \right) \pm z \text{ critical value} = \left( \hat{p}_1(1 - \hat{p}_1) + \hat{p}_2(1 - \hat{p}_2) \right) \sqrt{\frac{0.026(1 - 0.026)}{500} + \frac{0.046(1 - 0.046)}{500}}
\]

\[
-0.02 \pm 1.96(0.011765)
\]

\[
-0.02 \pm 0.023059
\]

\[
(-0.04306, 0.00306)
\]

Communicate Results (C):

Interpret confidence interval: We are 95\% confident that the actual difference in the proportion of college graduates who were unemployed in 2008 and the proportion of college graduates who were unemployed in 2009 is somewhere between –0.04306 and 0.00306. Because the endpoints of the confidence interval have opposite signs, zero is included in the interval, and there may be no difference in the proportion of college graduates who were unemployed in 2008 and the proportion who were unemployed in 2009.

Interpret confidence level: The method used to construct this interval estimate is successful in capturing the actual value of the difference in population proportions 95\% of the time.
11.9: (a) Using the 5-step process (EMC^3):

Estimate (E): The difference between the proportion of high school graduates who were unemployed in 2008 and high school graduates who were unemployed in 2009, \( p_1 - p_2 \), will be estimated, where \( p_1 \) is the true proportion of high school graduates who were unemployed in 2008 and \( p_2 \) is the true proportion of high school graduates who were unemployed in 2009.

Method (M): Because the answers to the four key questions are (Q) estimation, (S) sample data, (T) one categorical variable, and (N) two samples, we can construct and interpret a 99% large-sample confidence interval for the difference between the proportion of high school graduates who were unemployed in 2008 and 2009.

Check (C): We are told that the samples are representative of the populations high school graduates in 2008 and 2009. In addition, \( n_1 \hat{p}_1 = 400(0.057) = 22.8 \), \( n_1 (1 - \hat{p}_1) = 400(1 - 0.057) = 377.2 \), \( n_2 \hat{p}_2 = 400(0.097) = 38.8 \), and \( n_2 (1 - \hat{p}_2) = 400(1 - 0.097) = 361.2 \), which are all at least 10. The two required conditions are satisfied.

Calculations (C):

\[
(\hat{p}_1 - \hat{p}_2) \pm z_{\text{critical value}} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}
\]

\[
(0.057 - 0.097) \pm 2.58 \sqrt{\frac{0.057(1 - 0.057)}{400} + \frac{0.097(1 - 0.097)}{400}}
\]

\[-0.040 \pm 2.58(0.018798)
\]

\[-0.040 \pm 0.048498
\]

\((-0.0885, 0.0085)\)

Communicate Results (C):

Interpret confidence interval: We are 99% confident that the actual difference in the proportion of high school graduates who were unemployed in 2008 and high school graduates who were unemployed in 2009 is somewhere between –0.0885 and 0.0085. Because the endpoints of the confidence interval have opposite signs, zero is included in the interval, and there may be no difference in the proportion of high school graduates that were unemployed in 2008 and the proportion that were unemployed in 2009.
Interpret confidence level: The method used to construct this interval estimate is successful in capturing the actual value of the difference in population proportions 99% of the time.

(b) Wider, because the confidence level in part (a) is greater than the confidence level in the previous exercise, and the sample sizes are smaller.

11.10: (a) Using the five-step process (EMC³):

Estimate (E): The proportion of children ages 6 months to 3 years who have a TV in their bedroom.

Method (M): Because the answers to the four key questions are (Q) estimation, (S) sample data, (T) one categorical variable, and (N) one sample, we can construct and interpret a 95% confidence interval for the proportion of children ages 6 months to 3 years who have a TV in their bedroom.

Check (C): We are told that a random sample of parents of children ages 6 months to 3 years was taken. In addition, \( n\hat{p} = 100(0.3) = 30 \) and \( n(1 - \hat{p}) = 100(1 - 0.30) = 70 \), which are both at least 10. The two required conditions are satisfied.

Calculations (C):

\[
\hat{p} \pm (z \text{ critical value}) \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}
\]

\[
0.30 \pm 1.96 \sqrt{\frac{0.30(1 - 0.30)}{100}}
\]

\[
0.30 \pm 1.96(0.045826)
\]

\[
0.30 \pm 0.089819
\]

\[
(0.21018, 0.38982)
\]

Communicate results (C):

Interpret confidence interval: We are 95% confident that the actual proportion of children ages 6 months to 3 years who have a TV in their bedroom is between 0.210 and 0.390.

Interpret confidence level: The method used to construct this interval estimate is successful in capturing the actual value of the proportion 95% of the time.

(b) Using the five-step process (EMC³):

Estimate (E): The proportion of children ages 3 to 6 years who have a TV in their bedroom.
Method (M): Because the answers to the four key questions are (Q) estimation, (S) sample data, (T) one categorical variable, and (N) one sample, we can construct and interpret a 95% confidence interval for the proportion of children ages 3 to 6 years who have a TV in their bedroom.

Check (C): We are told that a random sample of parents of children ages 3 to 6 years was taken. In addition, \( n \hat{p} = 100(0.43) = 43 \) and \( n(1 - \hat{p}) = 100(1 - 0.43) = 57 \), which are both at least 10. The two required conditions are satisfied.

Calculations (C):

\[
\hat{p} \pm (z \text{ critical value}) \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}
\]

\[
0.43 \pm 1.96 \sqrt{\frac{0.43(1 - 0.43)}{100}}
\]

\[
0.43 \pm 1.96(0.049508)
\]

\[
0.43 \pm 0.097036
\]

\[
(0.33296, 0.52704)
\]

Communicate results (C):

Interpret confidence interval: We are 95% confident that the actual proportion of children ages 3 to 6 years who have a TV in their bedroom is between 0.333 and 0.527.

Interpret confidence level: The method used to construct this interval estimate is successful in capturing the actual value of the proportion 95% of the time.

(c) Yes, the confidence intervals overlap. This suggests that the two population proportions might not be different from each other.

(d) Using the 5-step process (EMC³):

Estimate (E): The difference in the proportions of children ages 6 months to 3 years and 3 to 6 years who have a TV in their bedroom, \( p_1 - p_2 \), will be estimated, where \( p_1 \) is the true proportion of children ages 6 months to 3 years who have a TV in their bedroom and \( p_2 \) is the true proportion of children ages 3 to 6 years who have a TV in their bedroom.

Method (M): Because the answers to the four key questions are (Q) estimation, (S) sample data, (T) one categorical variable, and (N) two samples, we can construct and interpret a 95% large-sample confidence interval for the difference in the proportion of children ages 6 months to 3 years who have a TV in their bedroom and children ages 3 to 6 years who have a TV in their bedroom.
Check (C): We are told that the two samples are independently selected representative samples. In addition,  

\[ n_1 \hat{p}_1 = 100(0.30) = 30 \text{ and } n_1 (1 - \hat{p}_1) = 100(1 - 0.30) = 70, \]

\[ n_2 \hat{p}_2 = 100(0.43) = 43 \text{ and } n_2 (1 - \hat{p}_2) = 100(1 - 0.43) = 57, \]

which are all at least 10. The two required conditions are satisfied.

Calculations (C):

\[
(\hat{p}_1 - \hat{p}_2) \pm \left( z \text{ critical value} \right) \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}
\]

\[
(0.30 - 0.43) \pm 1.96 \sqrt{\frac{0.30(1-0.30)}{100} + \frac{0.43(1-0.43)}{100}}
\]

\[
= -0.13 \pm 1.96(0.067461)
\]

\[
= -0.13 \pm 0.132224
\]

\[
= (-0.262224, 0.002224)
\]

Communicate Results (C):

Interpret confidence interval: We are 95% confident that the actual difference in the proportions of children ages 6 months to 3 years and children 3 to 6 years who have a TV in their bedroom is somewhere between –0.262 and 0.002. Because the endpoints of the confidence interval have opposite signs, zero is included in the interval, and there may be no difference in the proportion of children ages 6 months to 3 years and children ages 3 to 6 years who have a TV in their bedroom.

Interpret confidence level: The method used to construct this interval estimate is successful in capturing the actual value of the difference in population proportions 95% of the time.

(e) Yes, the interval in part (d) is consistent with the answer in part (c). Both interpretations indicate that there might not be a significant difference in the proportion of children ages 6 months to 3 years and children ages 3 to 6 years who have a TV in their bedroom.

Section 11.2 Exercise Set 1

11.11: (a) \( H_0 : p_1 - p_2 = 0 \) versus \( H_a : p_1 - p_2 > 0 \), where \( p_1 \) is the actual proportion of workers who work over 40 hours per week who get less than 7 hours of sleep, and \( p_2 \) is the actual proportion of workers who work between 35-40 hours per week who get less than 7 hours of sleep.
(b) Yes, because there are more than 10 successes (those who get less than 7 hours of sleep) and 10 failures (those who do not get less than 7 hours of sleep) in each sample.

(c) The test statistic is \( z = 3.63 \) and the \( P \)-value = 0.000. Because the \( P \)-value of 0.000 is less than \( \alpha \) of 0.01, we reject \( H_0 \).

(d) There is convincing evidence that the proportion of workers who usually get less than 7 hours of sleep a night is higher for those who work more than 40 hours per week than for those who work between 35 and 40 hours per week.

11.12: Using the 5-step process (HMC³):

Hypotheses (H): We want to determine if there is evidence that the proportion of guests who reserve a room online who are satisfied is higher than the proportion of guests who reserve a room using a telephone system who are satisfied. Therefore, the population characteristics of interest are \( p_T \), the proportion of guests who reserve rooms using a telephone system who are satisfied, and \( p_O \), the proportion of guests who reserve rooms online who are satisfied. Test the hypotheses \( H_0 : p_T - p_O = 0 \) versus \( H_a : p_T - p_O < 0 \).

Method (M): Because the answers to the four key questions are (Q) hypothesis testing, (S) sample data, (T) one categorical variable, and (N) two samples, we can consider a large-sample test for a difference in population proportions. The test statistic is:

\[
    z = \frac{\hat{p}_T - \hat{p}_O}{\sqrt{\frac{\hat{p}_C(1-\hat{p}_C)}{n_T} + \frac{\hat{p}_C(1-\hat{p}_C)}{n_O}}}
\]

and we will use a significance level of \( \alpha = 0.05 \).

Check (C): We are told that the samples are independent and randomly selected. In addition, the sample of those who reserved a room using a telephone system contains 57 successes and 23 failures, and the sample of those who reserved a room online contains 50 successes and 10 failures; these four values are all at least 10. The two required conditions are satisfied.

Calculations (C): The sample proportions are \( \hat{p}_T = \frac{57}{80} = 0.7125 \) and \( \hat{p}_O = \frac{50}{60} = 0.8333 \).

The combined proportion is \( \hat{p}_C = \frac{n_T \hat{p}_T + n_O \hat{p}_O}{n_T + n_O} = \frac{(80)(0.7125) + (60)(0.8333)}{80 + 60} = 0.7643 \).

Therefore, the test statistic is

\[
    z = \frac{\hat{p}_T - \hat{p}_O}{\sqrt{\frac{\hat{p}_C(1-\hat{p}_C)}{n_T} + \frac{\hat{p}_C(1-\hat{p}_C)}{n_O}}} = \frac{0.7125 - 0.8333}{\sqrt{\frac{0.7643(1-0.7643)}{80} + \frac{0.7643(1-0.7643)}{60}}} = -1.667.
\]
This is a lower-tailed test (the inequality in \( H_0 \) is \(<\)), so the \( P \)-value is the area under the \( z \) curve and to the left of \(-1.667\). Therefore, the \( P \)-value is \( P(z \leq -1.667) = 0.048 \).

Communicate Results (C): Because the \( P \)-value of 0.048 is less than the selected significance level of \( \alpha = 0.05 \), we reject the null hypothesis. There is convincing evidence that the proportion of guests who are satisfied is higher for those who reserve a room online.

11.13: Using the 5-step process (HMC³):

Hypotheses (H): We want to determine if there is convincing evidence that the proportion of college graduates who experience sunburn is higher than the proportion of those without a high school degree who experience sunburn. Therefore, the population characteristics of interest are \( p_1 \), the proportion of college graduates who experience sunburn, and \( p_2 \), the proportion of those without a high school degree who experience sunburn. Test the hypotheses \( H_0 : p_1 - p_2 = 0 \) versus \( H_a : p_1 - p_2 > 0 \).

Method (M): Because the answers to the four key questions are (Q) hypothesis testing, (S) sample data, (T) one categorical variable, and (N) two samples, we can consider a large-sample test for a difference in population proportions. The test statistic is:

\[
z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_C(1-\hat{p}_C)}{n_1} + \frac{\hat{p}_C(1-\hat{p}_C)}{n_2}}} = \frac{0.43 - 0.25}{\sqrt{\frac{0.34(1-0.34)}{200} + \frac{0.34(1-0.34)}{200}}} = 3.800.
\]

This is an upper-tailed test (the inequality in \( H_a \) is \( > \)), so the \( P \)-value is the area under the \( z \) curve and to the right of 3.800. Therefore, the \( P \)-value is \( P(z \geq 3.800) \approx 0 \).

Communicate Results (C): Because the \( P \)-value of approximately 0 is less than the selected significance level of \( \alpha = 0.01 \), we reject the null hypothesis. There is
convincing evidence that the proportion experiencing sunburn is higher for college graduates than for those without a high school degree.

Section 11.2 Exercise Set 2

11.14: (a) \( H_0 : p_1 - p_2 = 0 \) versus \( H_a : p_1 - p_2 \neq 0 \), where \( p_1 \) is the actual proportion of passengers with post-flight respiratory symptoms for planes that do not recirculate air, and \( p_2 \) is the actual proportion of passengers with post-flight respiratory symptoms for planes that do recirculate air.

(b) Yes, because \( n_1 \hat{p}_1 = 517(0.208897) = 108 \), \( n_1 (1 - \hat{p}_1) = 517(1 - 0.208897) = 409 \), \( n_2 \hat{p}_2 = 583(0.188679) = 110 \), and \( n_2 (1 - \hat{p}_2) = 583(1 - 0.188679) = 473 \), which are all at least 10.

(c) The test statistic is \( z = 0.84 \) and the \( P \)-value = 0.401. Because the \( P \)-value of 0.401 is greater than \( \alpha \) of 0.01, we fail to reject \( H_0 \).

(d) There is not convincing evidence that the proportion of passengers with post-flight respiratory symptoms for planes that do not recirculate air differs from the proportion of passengers with post-flight respiratory symptoms for planes that do recirculate air.

11.15: Using the 5-step process (HMC³):

Hypotheses (H): We want to determine if there is convincing evidence that the proportion of all parents who think their teen checks social networking sites more than 10 times a day is less than the proportion of all teens who check more than 10 times a day. The population characteristics of interest are \( p_1 \), the proportion of all parents who think their teen checks social networking sites more than 10 times a day, and \( p_2 \), the proportion of all teens who check social networking sites more than 10 times a day. Test the hypotheses \( H_0 : p_1 - p_2 = 0 \) versus \( H_a : p_1 - p_2 < 0 \).

Method (M): Because the answers to the four key questions are (Q) hypothesis testing, (S) sample data, (T) one categorical variable, and (N) two samples, we can consider a large-sample test for a difference in population proportions. The test statistic is:

\[
z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_C(1 - \hat{p}_C)}{n_1} + \frac{\hat{p}_C(1 - \hat{p}_C)}{n_2}}}, \quad \text{and we will use a significance level of } \alpha = 0.01.
\]

Check (C): We are told that the samples were independently selected and representative of American teens and parents of American teens. In addition, the sample of parents of teens included 40 successes and 960 failures, and the sample of teens included 220
successes and 780 failures, which are all at least 10. The two required conditions are satisfied.

Calculations (C): The two sample proportions are \( \hat{p}_1 = \frac{40}{1,000} = 0.040 \) and \( \hat{p}_2 = \frac{220}{1,000} = 0.220 \). The combined proportion is

\[
\hat{p}_c = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{(1,000)(0.040) + (1,000)(0.220)}{1,000 + 1,000} = 0.13 .
\]

Therefore, the test statistic is

\[
z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}} = \frac{0.040 - 0.220}{\sqrt{\frac{0.13(1-0.13)}{1,000} + \frac{0.13(1-0.13)}{1,000}}} = -11.97 .
\]

This is a lower-tailed test (the inequality in \( H_a \) is \(<\)), so the \( P \)-value is the area under the \( z \) curve and to the left of \(-11.97\). Therefore, the \( P \)-value is \( P(z \leq -11.97) \approx 0 \).

Communicate Results (C): Because the \( P \)-value of approximately 0 is less than the selected significance level of \( \alpha = 0.01 \), we reject the null hypothesis. There is convincing evidence that the proportion all parents who think their teen checks social networking sites more than 10 times a day is less than the proportion of all teens who check more than 10 times a day.

11.16: No, I would not use the large-sample test for a difference in population proportions because this question is phrased as a one-proportion hypothesis test, not a difference in proportions.

Section 11.2 Additional Exercises

11.17: Using the 5-step process (HMC\(^3\)):

Hypotheses (H): We want to determine if there is convincing evidence that the proportion of young adults who think their parents would provide financial support and the proportion of parents who say they would provide financial support are different. Therefore, the population characteristics of interest are \( p_1 \), the proportion of all young adults who think their parents would provide financial support, and \( p_2 \), the proportion of parents of young adults who would provide financial support. Test the hypotheses

\( H_0 : p_1 - p_2 = 0 \) versus \( H_a : p_1 - p_2 \neq 0 \).

Method (M): Because the answers to the four key questions are (Q) hypothesis testing, (S) sample data, (T) one categorical variable, and (N) two samples, we can consider a
large-sample test for a difference in population proportions. The test statistic is:

\[ z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_C(1-\hat{p}_C)}{n_1} + \frac{\hat{p}_C(1-\hat{p}_C)}{n_2}}} \]

, and we will use a significance level of \( \alpha = 0.05 \).

Check (C): We are told that the samples are independent and randomly selected. In addition, \( n_1\hat{p}_1 = (600)(0.41) = 246 \), \( n_1(1-\hat{p}_1) = (600)(1-0.41) = 354 \),

\( n_2\hat{p}_2 = (300)(0.43) = 129 \), and \( n_2(1-\hat{p}_2) = (300)(1-0.43) = 171 \), which are all at least 10. The two required conditions are satisfied.

Calculations (C): The sample proportions are \( \hat{p}_1 = 0.41 \) and \( \hat{p}_2 = 0.43 \). The combined proportion is

\[ \hat{p}_C = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{(600)(0.41) + (300)(0.43)}{600 + 300} = 0.417. \]

Therefore, the test statistic is

\[ z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{0.417(1-0.417)}{600} + \frac{0.417(1-0.417)}{300}}} = -0.574. \]

This is a two-tailed test (the inequality in \( H_0 \) is \( \neq \)), so the P-value is the total area under the z curve and to the left of –0.574 and to the right of 0.574. Therefore, by the symmetry of the standard normal curve, the P-value is \( 2 \cdot P(z \leq -0.574) = 0.566 \).

Communicate Results (C): Because the P-value of 0.566 is greater than the selected significance level of \( \alpha = 0.05 \), we fail to reject the null hypothesis. There is not convincing evidence of a difference between the proportion of young adults who think that their parents would provide financial support for marriage and the proportion of parents who say they would provide financial support for marriage.

11.18 Using the 5-step process (HMC^3):

Hypotheses (H): We want to determine if there is convincing evidence that the proportion of parents of young adults who say they would help with buying a house or renting an apartment is significantly less than the proportion of young adults who think that their parents would help. The population characteristics of interest are \( p_1 \), the proportion of all parents of young adults who say they would help with buying a house or renting an apartment, and \( p_2 \), the proportion of young adults who think that their parents would help. Test the hypotheses \( H_0 : p_1 - p_2 = 0 \) versus \( H_a : p_1 - p_2 < 0 \).

Method (M): Because the answers to the four key questions are (Q) hypothesis testing, (S) sample data, (T) one categorical variable, and (N) two samples, we can consider a
large-sample test for a difference in population proportions. The test statistic is:

\[ z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_C(1 - \hat{p}_C)}{n_1} + \frac{\hat{p}_C(1 - \hat{p}_C)}{n_2}}} \], and we will use a significance level of \( \alpha = 0.05 \).

Check (C): We are told that the samples are independent and randomly selected. In addition, \( n_1\hat{p}_1 = (300)(0.27) = 81 \), \( n_1(1 - \hat{p}_1) = (300)(1 - 0.27) = 219 \),
\( n_2\hat{p}_2 = (600)(0.37) = 222 \), and \( n_2(1 - \hat{p}_2) = (600)(1 - 0.37) = 378 \), which are all at least 10. The two required conditions are satisfied.

Calculations (C): The sample proportions are \( \hat{p}_1 = 0.27 \) and \( \hat{p}_2 = 0.37 \). The combined proportion is \( \hat{p}_C = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{(300)(0.27) + (600)(0.37)}{300 + 600} = 0.3367 \). Therefore, the test statistic is

\[ z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_C(1 - \hat{p}_C)}{n_1} + \frac{\hat{p}_C(1 - \hat{p}_C)}{n_2}}} = \frac{0.27 - 0.37}{\sqrt{\frac{0.3367(1 - 0.3367)}{300} + \frac{0.3367(1 - 0.3367)}{600}}} = -2.993. \]

This is a lower-tailed test (the inequality in \( H_0 \) is \(<\)), so the \( P \)-value is the area under the \( z \) curve and to the left of \(-2.993\). Therefore, the \( P \)-value is \( P(z \leq -2.993) = 0.0014 \).

Communicate Results (C): Because the \( P \)-value of 0.0014 is less than the selected significance level of \( \alpha = 0.05 \), we reject the null hypothesis. There is convincing evidence that the proportion of parents of young adults who say they would help with buying a house or renting an apartment is significantly less than the proportion of young adults who think that their parents would help.

11.19: Since the data given are population characteristics, an inference procedure is not applicable. It is known that the rate of Lou Gehrig’s disease among soldiers sent to the war is higher than for those not sent to the war.

Are You Ready To Move On? Chapter 11 Review Exercises

11.20: (a) Using the 5-step process (EMC³):

Estimate (E): The difference in the proportion of teens using a cell phone while driving before the ban and the proportion after the ban, \( p_1 - p_2 \), will be estimated, where \( p_1 \) is the true proportion of teens who use cell phones while driving before the ban and \( p_2 \) is the true proportion of teens who use cell phones while driving after the ban.

Method (M): Because the answers to the four key questions are (Q) estimation, (S) sample data, (T) one categorical variable, and (N) two samples, we can construct and
interpret a 95% large-sample confidence interval for the difference in proportions of teens using a cell phone while driving before the ban and the proportion after the ban.

Check (C): We are told that the two samples were independently selected representative samples. In addition, \( n_1 \hat{p}_1 = 200(0.11) = 22 \), \( n_1 (1 - \hat{p}_1) = 200(1 - 0.11) = 178 \), \( n_2 \hat{p}_2 = 150(0.12) = 18 \), and \( n_2 (1 - \hat{p}_2) = 150(1 - 0.12) = 132 \), which are all at least 10. The two required conditions are satisfied.

Calculations (C):

\[
(\hat{p}_1 - \hat{p}_2) \pm (z \text{ critical value}) \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}
\]

\[
(0.11 - 0.12) \pm 1.96 \sqrt{\frac{0.11(1-0.11)}{200} + \frac{0.12(1-0.12)}{150}}
\]

\[
-0.01 \pm 1.96(0.034547)
\]

\[
-0.01 \pm 0.067712
\]

\[
(-0.0777, 0.0577)
\]

Communicate Results (C):

Interpret confidence interval: We are 95% confident that the actual difference in proportion of teens using a cell phone while driving before the ban and the proportion after the ban somewhere between –0.0777 and 0.0577. Because the endpoints of the confidence interval have opposite signs, zero is included in the interval, and there may be no difference in the proportion of teens using a cell phone while driving before the ban and the proportion after the ban.

Interpret confidence level: The method used to construct this interval estimate is successful in capturing the actual value of the difference in population proportions 95% of the time.

(b) Yes, zero is included in the confidence interval. This implies that there may be no difference in the proportion of teens using a cell phone while driving before the ban and the proportion after the ban.

11.21: Using the 5-step process (HMC$^3$):

Hypotheses (H): We want to determine if there is convincing evidence that the proportion of teens that approve of banning cell phone use and texting while driving is less than the proportion of parents of teens who approve. Therefore, the population characteristics of interest are \( p_1 \), the proportion of all teens who approve of banning cell phone use and texting while driving, and \( p_2 \), the proportion of all parents of teens who
approve of banning cell phone use and texting while driving. Test the hypotheses

\[ H_0 : p_1 - p_2 = 0 \text{ versus } H_a : p_1 - p_2 < 0. \]

Method (M): Because the answers to the four key questions are (Q) hypothesis testing, (S) sample data, (T) one categorical variable, and (N) two samples, we can consider a large-sample test for a difference in population proportions. The test statistic is:

\[
z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}} ,
\] and we will use a significance level of \( \alpha = 0.05 \).

Check (C): We are told to assume that the samples are representative of the two populations. In addition, \( n_1 \hat{p}_1 = (600)(0.74) = 444 \), \( n_1 (1 - \hat{p}_1) = (600)(1 - 0.74) = 156 \), \( n_2 \hat{p}_2 = (400)(0.95) = 380 \), and \( n_2 (1 - \hat{p}_2) = (400)(1 - 0.95) = 20 \), which are all at least 10. The two required conditions are satisfied.

Calculations (C): The sample proportions are \( \hat{p}_1 = 0.74 \) and \( \hat{p}_2 = 0.95 \). The combined proportion is \( \hat{p}_c = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2} = \frac{(600)(0.74) + (400)(0.95)}{600 + 400} = 0.824 \). Therefore, the test statistic is

\[
z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}} = \frac{0.74 - 0.95}{\sqrt{\frac{0.824(1-0.824)}{600} + \frac{0.824(1-0.824)}{400}}} = -8.543 \ .
\] This is a lower-tailed test (the inequality in \( H_a \) is \(<\)), so the \( P \)-value is the area under the \( z \) curve and to the left of \(-8.543 \). Therefore, the \( P \)-value is \( P(z \leq -8.543) \approx 0 \).

Communicate Results (C): Because the \( P \)-value of approximately 0 is less than the selected significance level of \( \alpha = 0.05 \), we reject the null hypothesis. There is convincing evidence that the proportion of teens who approve of the proposed laws is less than the proportion of parents of teens who approve.

11.22: Using the 5-step process (HMC³):

Hypotheses (H): We want to determine if there is convincing evidence that the proportion of Dutch boys who listen to music at high volume is greater than the proportion of Dutch girls who listen to music at high volume. The population characteristics of interest are \( p_1 \), the proportion of Dutch boys who listen to music at high volume, and \( p_2 \), the proportion of Dutch girls who listen to music at high volume. Test the hypotheses \( H_0 : p_1 - p_2 = 0 \text{ versus } H_a : p_1 - p_2 > 0 \).
Method (M): Because the answers to the four key questions are (Q) hypothesis testing, (S) sample data, (T) one categorical variable, and (N) two samples, we can consider a large-sample test for a difference in population proportions. The test statistic is:

\[
z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}} ,
\]
and we will use a significance level of \( \alpha = 0.01 \).

Check (C): We are told that the samples are independent and randomly selected. In addition, the number of Dutch boys who reported that they listen to music at high volume is 397, the number of Dutch boys who did not listen to music at high volume is 367, the number of Dutch girls who reported that they listen to music at high volume is 331, and the number of Dutch girls who did not listen to music at high volume is 417, which are all at least 10. The two required conditions are satisfied.

Calculations (C): The sample proportions are \( \hat{p}_1 = \frac{397}{764} = 0.5196 \) and \( \hat{p}_2 = \frac{331}{748} = 0.4425 \).

The combined proportion is \( \hat{p}_c = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{(764)(0.5196) + (748)(0.4425)}{764 + 748} = 0.4815 \).

Therefore, the test statistic is

\[
z = \frac{0.5196 - 0.4425}{\sqrt{\frac{0.4815(1-0.4815)}{764} + \frac{0.4815(1-0.4815)}{748}}} = 2.9999 .
\]
This is an upper-tailed test (the inequality in \( H_a \) is \( > \)), so the \( P \)-value is the area under the \( z \) curve and to the right of 2.9999. Therefore, the \( P \)-value is \( P(z \geq 2.9999) = 0.0014 \).

Communicate Results (C): Because the \( P \)-value of 0.0014 is less than the selected significance level of \( \alpha = 0.01 \), we reject the null hypothesis. There is convincing evidence that the proportion of Dutch boys who listen to music at high volume is greater than the proportion of Dutch girls who listen to music at high volume. Yes, the sample data support the authors’ conclusion.

11.23: (a) Using the 5-step process (HMC^3):

Hypotheses (H): We want to determine if there is convincing evidence that the proportion of those who think that newspapers are boring is different for teenage girls and boys. Therefore, the population characteristics of interest are \( p_G \), the proportion of all teenage girls who think that newspapers are boring, and \( p_B \), the proportion of all teenage boys who think that newspapers are boring. Test the hypotheses \( H_0 : p_G - p_B = 0 \) versus \( H_a : p_G - p_B \neq 0 \).
Method (M): Because the answers to the four key questions are (Q) hypothesis testing, (S) sample data, (T) one categorical variable, and (N) two samples, we can consider a large-sample test for a difference in population proportions. The test statistic is:

\[
z = \frac{\hat{p}_G - \hat{p}_B}{\sqrt{\frac{\hat{p}_C(1-\hat{p}_C)}{n_G} + \frac{\hat{p}_C(1-\hat{p}_C)}{n_B}}}, \text{ and we will use a significance level of } \alpha = 0.05.
\]

Check (C): We are told that the samples are representative of the two populations. In addition, \(n_G\hat{p}_G = (58)(0.41) = 23.78\), \(n_G(1-\hat{p}_G) = (58)(1-0.41) = 34.22\), \(n_B\hat{p}_B = (41)(0.44) = 18.04\), and \(n_B(1-\hat{p}_B) = (41)(1-0.44) = 22.96\), which are all at least 10. The two required conditions are satisfied.

Calculations (C): The sample proportions are \(\hat{p}_G = 0.41\) and \(\hat{p}_B = 0.44\). The combined proportion is \(\hat{p}_C = \frac{n_G\hat{p}_G + n_B\hat{p}_B}{n_G + n_B} = \frac{(58)(0.41) + (41)(0.44)}{58 + 41} = 0.422\). Therefore, the test statistic is

\[
z = \frac{\hat{p}_G - \hat{p}_B}{\sqrt{\frac{\hat{p}_C(1-\hat{p}_C)}{n_G} + \frac{\hat{p}_C(1-\hat{p}_C)}{n_B}}} = \frac{0.41 - 0.44}{\sqrt{\frac{0.422(1-0.422)}{58} + \frac{0.422(1-0.422)}{41}}} = -0.298.
\]

This is a two-tailed test (the inequality in \(H_a\) is \(\neq\)), so the \(P\)-value is the total area under the \(z\) curve and to the left of \(-0.298\) and to the right of \(0.298\). Therefore, using the symmetry of the normal curve, the \(P\)-value is \(2 \cdot P(z \leq -0.298) = 0.766\).

Communicate Results (C): Because the \(P\)-value of 0.766 is greater than the selected significance level of \(\alpha = 0.05\), we fail to reject the null hypothesis. There is not convincing evidence that the proportion of those who think that newspapers are boring is different for teenage girls and boys.

(b) Using the 5-step process (HMC³):

Hypotheses (H): We want to determine if there is convincing evidence that the proportion of those who think that newspapers are boring is different for teenage girls and boys. Therefore, the population characteristics of interest are \(p_G\), the proportion of all teenage girls who think that newspapers are boring, and \(p_B\), the proportion of all teenage boys who think that newspapers are boring. Test the hypotheses \(H_0: p_G - p_B = 0\) versus \(H_a: p_G - p_B \neq 0\).

Method (M): Because the answers to the four key questions are (Q) hypothesis testing, (S) sample data, (T) one categorical variable, and (N) two samples, we can consider a
large-sample test for a difference in population proportions. The test statistic is:

\[ z = \frac{\hat{p}_G - \hat{p}_B}{\sqrt{\frac{\hat{p}_C(1 - \hat{p}_C)}{n_G} + \frac{\hat{p}_C(1 - \hat{p}_C)}{n_B}}} \], and we will use a significance level of \( \alpha = 0.05 \).

Check (C): We are told that the samples are representative of the two populations. In addition, \( n_G \hat{p}_G = (2,000)(0.41) = 820 \), \( n_G (1 - \hat{p}_G) = (2,000)(1 - 0.41) = 1,180 \), \( n_B \hat{p}_B = (2,500)(0.44) = 1,100 \), and \( n_B (1 - \hat{p}_B) = (2,500)(1 - 0.44) = 1,400 \), which are all at least 10. The two required conditions are satisfied.

Calculations (C): The sample proportions are \( \hat{p}_G = 0.41 \) and \( \hat{p}_B = 0.44 \). The combined proportion is

\[ \hat{p}_C = \frac{n_G \hat{p}_G + n_B \hat{p}_B}{n_G + n_B} = \frac{(2,000)(0.41) + (2,500)(0.44)}{2,000 + 2,500} = 0.427 \].

Therefore, the test statistic is

\[ z = \frac{\hat{p}_G - \hat{p}_B}{\sqrt{\frac{\hat{p}_C(1 - \hat{p}_C)}{n_G} + \frac{\hat{p}_C(1 - \hat{p}_C)}{n_B}}} = \frac{0.41 - 0.44}{\sqrt{\frac{0.427(1 - 0.427)}{2,000} + \frac{0.427(1 - 0.427)}{2,500}}} = -2.022 \].

This is a two-tailed test (the inequality in \( H_a \) is \( \neq \)), so the \( P \)-value is the total area under the \( z \) curve and to the left of \(-2.022\) and to the right of \(2.022\). Therefore, using the symmetry of the normal curve, the \( P \)-value is \( 2 \cdot P(z \leq -2.022) = 0.043 \).

Communicate Results (C): Because the \( P \)-value of 0.043 is less than the selected significance level of \( \alpha = 0.05 \), we reject the null hypothesis. There is convincing evidence that the proportion of those who think that newspapers are boring is different for teenage girls and boys.

(c) Assuming that the population proportions are equal, you are much less likely to get a difference in sample proportions as large as the one given when the samples are very large than when the samples are relatively small.