Difference between two proportions (Independent Samples):
hypothesis tests and confidence intervals.

1. Alcohol abuse has been described by college presidents as the number one problem on campus and it is an important cause of death in young adults. How common is it? A survey of 17,096 students in U.S. four-year colleges collected information on drinking behavior and alcohol-related problems. The researchers defined “frequent binge drinking” as having five or more drinks in a row three or more times in the past two weeks. Data were also summarized by gender. Here is a summary:

<table>
<thead>
<tr>
<th>Population</th>
<th>n</th>
<th># of binge drinkers</th>
<th>( \hat{p} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (men)</td>
<td>7,180</td>
<td>1,630</td>
<td>0.227</td>
</tr>
<tr>
<td>2 (women)</td>
<td>9,916</td>
<td>1,684</td>
<td>0.170</td>
</tr>
<tr>
<td>Total</td>
<td>17,096</td>
<td>3,314</td>
<td>0.194</td>
</tr>
</tbody>
</table>

a. Display the data with Pie charts and a side-by-side Bar Graph. Then describe the graphs in a paragraph.

**ANSWER:**

**Male student binge drinkers versus non-binge drinkers in a sample of 7,180 males.**

**Female student binge drinkers versus non-binge drinkers in a sample of 9,916 females.**

![Pie charts for male and female binge drinkers](image)

We can see that both samples have more non-binge drinkers than drinkers. Less than \( \frac{1}{4} \) of the sample are binge drinkers for both males and females. There are slightly more binge drinkers for males when compared with females.
That both samples have more non-binge drinkers than drinkers is seen again with the side-by-side bar graph. The sample of males has slightly more binge drinkers (or slightly less non-binge drinkers) than the sample of females. There are around 20% of binge drinkers for both males and females. Alternatively, we can see that there are around 80% of non-binge drinkers for both males and females.

b. Define

Population 1: MALE students in U.S. four-year colleges
Population 2: FEMALE students in U.S. four-year colleges
Trait: are binge drinkers

\( p_1 \): The proportion of all MALE students in U.S. four-year colleges that are binge drinkers.

\( p_2 \): The proportion of all FEMALE students in U.S. four-year colleges that are binge drinkers.
c. Are all of the assumptions satisfied so that we may use two-sample methods for proportions?

**ANSWER:**

Yes. All conditions are satisfied as seen below.

1) SRS: We will assume that the survey was administered randomly to both males and females.

2) The number of binge drinkers in the sample of males was $1,630 \geq 10$.
   The number of binge drinkers in the sample of females was $1,684 \geq 10$.
   This can also be seen with the computations
   
   \[
   n_1 \hat{p}_1 = (7,180)(0.227) = 1630 \geq 10 \text{ and }
   n_2 \hat{p}_2 = (9,916)(0.170) = 1684 \geq 10.
   \]

3) The number of non-binge drinkers in the sample of males was $5,550 \geq 10$.
   The number of non-binge drinkers in the sample of females was $8,232 \geq 10$.
   This can also be seen with the computations
   
   \[
   n_1(1 - \hat{p}_1) = 7,180(1 - 0.227) = 7,180(0.773) = 5,550 \geq 10 \text{ and }
   n_2(1 - \hat{p}_2) = 9,916(1 - 0.170) = 9,916(0.830) = 8,232 \geq 10.
   \]

4) $10n_1 = 10(7,180) = 71,800$. There are at least 71,800 male students in U.S. four-year colleges.
   $10n_2 = 10(9,916) = 99,160$. There are at least 99,160 female students in U.S. four-year colleges.

5) It is safe to say that the population of female students and the population of male students in U.S. four-year colleges are independent from each other.

---

d. The sample proportions are certainly quite different, but we will perform a significance test to see if the difference is large enough to lead us to believe that the population proportions are not equal. We will test at the 5% level.

**ANSWER:**

1) \[ H_0 : p_1 = p_2 \]
   \[ H_a : p_1 \neq p_2 \]

2) \[ \alpha = 0.05 \]

3) \[
\begin{array}{|c|c|c|}
\hline
\text{Population 1} & \text{Sample Size} & \text{Sample Proportion} \\
\hline
n_1 = 7,180 & \hat{p}_1 = 0.227 \\
\hline
\text{Population 2} & n_2 = 9,916 & \hat{p}_2 = 0.170 \\
\hline
\end{array}
\]
4) 
\[ \hat{p} = \frac{1.630 + 1.684}{7.180 + 9.916} = \frac{3.314}{17.096} = 0.194 \]

Observe that \( \hat{p} \) was the estimate on the bottom line of the data summary given at the very beginning of our story problem situation.

Test Statistic:
\[ z = \frac{(0.227 - 0.170)}{\sqrt{(0.194)(1 - 0.194)\left(\frac{1}{7180} + \frac{1}{9916}\right)}} = \frac{0.057}{0.006126} = 9.3 \]

P-value = \( 2 \text{normalcdf}(9.3, 1\text{E99}) = 1.011\text{E}-20 \)

NOTE:
- We multiply by 2 because we have a two-tailed hypothesis test (not equal sign in the Alternate Hypothesis).
- The P-value may be slightly different depending on round-off error. In any case it should be very small!

5) P-value \( \alpha \)
\[ 1.011\text{E}-20 \leq 0.05 \]
reject \( H_0 \)

6) There is significant evidence that the proportion of all male students at U.S. four-year colleges that are binge drinkers is not equal to the proportion of all female students at U.S. four-year colleges that are binge drinkers.

2. Refer to the previous exercise. We will find a 95% confidence interval for the difference between the proportion of men and women who are frequent binge drinkers.

**ANSWER:**
For 95% confidence we have \( z^* = 1.960 \). So that
\[ (0.227 - 0.170) \pm 1.960 \sqrt{\frac{0.227(1 - 0.227)}{7180} + \frac{0.170(1 - 0.170)}{9916}} \]
\[ 0.57 \pm 0.012 \]
\[ (0.045, 0.069) \]

The confidence interval takes \( \hat{p}_1 - \hat{p}_2 \) and uses it to estimate \( p_1 - p_2 \). So…
We are 95% confident that the difference between the proportion of male frequent binge drinkers and female frequent binge drinkers at U.S. four-year colleges is between 0.045 and 0.069.
Alternatively, we can report with 95% confidence that the men are 5.7% more likely to be frequent binge drinkers than women at U.S. four-year colleges, with a margin of error of 1.2%.