Statistics 104 – Laboratory 9

This lab first looks at the distribution of the sample proportion \( \hat{p} \) by sampling from a population of 250 statistics students. It then looks at using a sample proportion to make inferences about the population proportion. The characteristic we are interested in is the proportion of the population with blue eyes.

1. We can simulate the construction of the distribution of a sample statistic by taking many samples (called repeated sampling) from the population and looking at the distribution of the sample statistics obtained. Refer to the table titled “Eye Color for a Population of 250 Statistics Students.” This table contains a listing of the eye color for the population members. Rather than list the names of the population members, this table numbers them by rows numbered (00, 01, . . ., 09, 10, . . ., 24) and columns numbered (0,1,2,. . ., 9). For example, student 037 (Row 03 and Column 7) has green eyes.

   a) Use the random number table to select a simple random sample of 25 students from this population. Write the student numbers and eye colors on the answer sheet. Calculate the sample proportion of blue eyes.

   Note: You can make more efficient use of 3 digit random numbers doing the following.
   - For numbers between 000 and 249 go directly to the table of heights.
   - For numbers between 250 and 499 subtract 250 and then go to the table of heights.
   - For numbers between 500 and 749 subtract 500 and then go to the table of heights.
   - For numbers between 750 and 999 subtract 750 and then go to the table heights.

   b) Take three other random samples of size 25 from the population. This requires different random numbers than the ones used in a). Use JMP this time by opening a New Data Table. Highlight Column 1 and go to Rows – Add Rows and add 25 rows. Then go to Cols – Formula – Random – Random Integer and put 250 in for n1. This will give you a random sample of 25 values from 001 to 250 = 000. You do not have to record the student numbers just simply keep track of the eye colors on the answer sheet and calculate the sample proportion of blue eyes for each sample. When you have finished getting your four random samples and associated sample proportions, go to the front of the room and write your sample proportions on the blackboard next to your group number.

This population of statistics students has a population proportion \( p = 0.312 \). We are interested in the distribution of the sample proportion for random samples of size \( n = 25 \).
c) What is the mean of the theoretical distribution of the sample proportion for samples of size \( n = 25 \)?

d) What is the standard deviation of the theoretical distribution of the sample proportion for samples of size \( n = 25 \)?

e) Draw a histogram for the sample proportions written on the blackboard. Describe the shape of the distribution. Does the distribution look like a normal model? Explain briefly.

f) If we increased the size of the sample to \( n = 100 \), describe the shape of the theoretical distribution of the sample proportion and give the mean and standard deviation.

2. For the next part of the lab you will use your sample proportions for your samples of size 25 to investigate the confidence interval for the population proportion \( p \).

   a) For each of your samples, construct a 95% confidence interval for the population proportion of blue-eyed students.

   b) How many of the four intervals contain the population proportion \( p = 0.312 \)? Put this number on the blackboard next to your group number.

   c) When all groups have put this information on the blackboard compute the percentage of intervals for the entire lab that contain the population proportion \( p = 0.312 \). Theoretically, what should this percentage be?

3. Combine your four samples into one sample of \( n = 100 \).

   a) Calculate a 95% confidence interval for the population proportion of blue eyes using this larger sample.

   b) Compare the width of this interval to the widths of the intervals based on samples of size \( n = 25 \).

4. For the last part of the lab you will look at simulating drawing samples from this population at random and investigate how different choices for the sample size and confidence level affect the confidence intervals constructed from these samples. You will use the applet on the web at

   http://statweb.calpoly.edu/chance/applets/Confsim/Confsim.html

You can get to this URL from the course web page via the link to Confidence Interval for a Population Proportion under Computing. In the applet, the method should be Proportions – Wald. Enter 0.312 for the value of \( \pi \). Note: this applet uses the Greek letter \( \pi \) for the population proportion.

   a) Enter the value of 25 for \( n \) (a random sample of size 25). The conf level should be 95 (a confidence level of 95%) and leave the number of Intervals at 1, then click on Sample. Does the interval displayed contain \( p \)? Press the sample button 9 more times. Report the running total for the intervals that contain \( p = 0.312 \).
b) The running total percentage in a) is a short-run result. Let’s look at a slightly longer run. Click on Reset and change the Intervals box from 1 to 100, and click on Sample. What is the difference between the confidence intervals displayed in red and those displayed in green? (If you don’t see any red intervals, click on Sample again.) What is the running total for intervals that contain \( p = 0.312 \).

c) Click the Sample button 9 more times. What is the running total now? What should it be? Explain briefly.

d) Is the Success/Failure condition satisfied for these values of \( p = 0.312 \) and \( n = 25 \)? Explain briefly.

e) Reset the applet. Enter the number 25 for \( n \), 1 in the Intervals box and 95 for the confidence level. Click on Sample. Report the values for the lower and upper limit of this confidence interval (right click on the interval). Note: if you happen to get a value of \( \hat{p} \) that is 0 or 1, Reset the applet and Sample again. Now change the value of \( n \) to 100 and click on Sample. Report the values for \( \hat{p} \) and the lower and upper limit of this confidence interval. How do the two intervals compare?

f) Reset the applet. Leave that sample size at \( n = 100 \) but put 250 in the Intervals box. Click Sample 4 times. What is the running total for the intervals that contain \( \pi \)? What should it be? Explain briefly.

g) Is the Success/Failure condition satisfied for these values of \( p = 0.312 \) and \( n = 100 \)? Explain briefly.

h) Reset the applet. Enter the number 100 for \( n \), 1 in the Intervals box and 95 for the confidence level. Click on Sample. Note: if you happen to get a value of \( \hat{p} \) that is 0 or 1, Reset the applet and Sample again. Report the values for \( \hat{p} \) and the lower and upper limit of this confidence interval. Now change the confidence level from 95% to 80% and click on Recalculate. Report the values for \( \hat{p} \) and the lower and upper limit of this confidence interval. How do the two intervals compare?

i) A narrow confidence interval is desirable. Give two things that you can do to get a narrower confidence interval. Which of these two ways do you think is better? Explain briefly.
Stat 104 – Laboratory 9
Group Answer Sheet

Names of Group Members: ____________________, ____________________
____________________, ____________________

1.

a) Sample 1

<table>
<thead>
<tr>
<th>Number</th>
<th>Color</th>
<th>Number</th>
<th>Color</th>
<th>Number</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sample proportion of blue eyes:

<table>
<thead>
<tr>
<th>Sample 2</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>blue</td>
<td>brown</td>
<td>green</td>
<td>hazel</td>
<td>other</td>
</tr>
<tr>
<td>Sample 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b)

\[ \hat{p} = \]  
\[ \hat{p} = \]  
\[ \hat{p} = \]

c) What is the mean of the theoretical distribution of the sample proportion for samples of size \( n = 25 \)?

d) What is the standard deviation of the theoretical distribution of the sample proportion for samples of size \( n = 25 \)?
e) Draw a histogram for the sample proportions on the blackboard. Describe the shape of the distribution. Does the distribution look like a normal model? Explain briefly.

f) If we increased the size of the sample to \( n = 100 \), describe the shape of the theoretical distribution of the sample proportion and give the mean and standard deviation.

2. For the next part of the lab you will use your sample proportions for your samples of size 25 to investigate the confidence interval for the population proportion \( p \).

   a) For each of your samples, construct a 95% confidence interval for the population proportion of blue-eyed students.

   \[
   \hat{p} - z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \text{ to } \hat{p} + z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}
   \]

   Does CI contain \( p = 0.312 \)?

<table>
<thead>
<tr>
<th>Sample</th>
<th>( \hat{p} - z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} )</th>
<th>( \hat{p} + z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} )</th>
<th>Does CI contain ( p = 0.312 )?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample 4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

   b) 

   c) When all groups have put this information on the blackboard compute the percentage of intervals for the entire lab that contain the population proportion \( p = 0.312 \). Theoretically, what should this percentage be?
3. Combine the four samples into one sample of $n = 100$.

   a) Calculate a 95% confidence interval for the population proportion of blue eyes using this larger sample.

   b) Compare the width of this interval to the widths of the intervals based on samples of size $n = 25$.

4. For the last part of the lab you will look at simulating drawing samples from this population at random and investigate how different choices for the sample size and confidence level affect the confidence intervals constructed from these samples.

   a) Does the interval displayed contain $p=0.312$?

      Report the running total for the intervals that contain $p=0.312$.

   b) What is the difference between the confidence intervals displayed in red and those displayed in green?

      What is the running total for intervals that contain $p=0.312$?

   c) What is the running total now?

      What should it be? Explain briefly.

   d) Is the Success/Failure condition satisfied for these values of $p=0.312$ and $n=25$? Explain briefly.
e)  

<table>
<thead>
<tr>
<th></th>
<th>n=25</th>
<th></th>
<th>n=100</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower</td>
<td>( \hat{p} )</td>
<td>Upper</td>
<td>Lower</td>
<td>( \hat{p} )</td>
</tr>
</tbody>
</table>

How do the two intervals compare?

f) What is the running total for the intervals that contain \( p=0.312 \)?

What should it be? Explain briefly.

g) Is the Success/Failure condition satisfied for these values of \( p=0.312 \) and \( n=100 \)? Explain briefly.

h)  

<table>
<thead>
<tr>
<th></th>
<th>95%</th>
<th></th>
<th>80%</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower</td>
<td>( \hat{p} )</td>
<td>Upper</td>
<td>Lower</td>
<td>( \hat{p} )</td>
</tr>
</tbody>
</table>

How do the two intervals compare?

i) A narrow confidence interval is desirable. Give two things that you can do to get a narrower confidence interval. Which of these two ways do you think is better? Explain briefly.