This lab looks at the sampling distribution of the sample proportion \( \hat{p} \) and probabilities associated with sampling from a population with a categorical variable. Proportions are used to summarize information about categorical variables (what proportion of people belong to a particular category). To look at the distribution for the sample proportion, \( \hat{p} \), we will sample from a population of 250 Statistics 101 students. The characteristic we are interested in is the proportion of students with blue eyes. By taking many samples (called repeated sampling) from the population and looking at the distribution of the sample statistics generated the distribution for the sample statistic is obtained.

**Activity 1:** Refer to the table titled Eye Color for Population of 250 Statistics 101 Students. This table contains a listing of the eye colors of the population members. Rather than list the names of the population members, this table numbers them by rows numbered \{00, 01, \ldots, 09, 10, \ldots, 24\} and columns numbered \{0,1,2,\ldots, 9\}. For example, student 057 (Row 05 and Column 7) has brown eyes.

a. Use the random number table to select a simple random sample of 10 students from this population. Write the student numbers and eye colors on the answer sheet. Calculate the proportion of the students in your sample with blue eyes. Note: You can make more efficient use of the 3 digit random numbers doing the following. For numbers between 000 and 249 go directly to the table. For numbers between 250 and 499 subtract 250 and then go to the table. For numbers between 500 and 749 subtract 500 and then go to the table. For numbers between 750 and 999 subtract 750 and then go to the table.

b. Take another random sample but this time of size 25 from the population. For this sample, just keep track of the number of students in the sample with blue eyes. Write the proportion of the students in your sample with blue eyes on the answer sheet.

**Activity 2:** Doing the random sampling by hand is tedious. We can use JMP to do the random sampling for us provided we have a population to sample from in the form of a JMP data. On the course web site is a JMP data table, called eyecolor.JMP, with information about eye color for a population of 1070 individuals.

a. Use the Distribution platform to find the proportion of this population with blue eyes. Turn in the JMP output with your answers.

b. Go back to the course web site and right-click on the file bluesampleprop.JSL. Choose the **Save Link As** option from the menu and save the file to the computer’s desktop. Go to the desktop and double-click on the file to run the program. You don’t have to do anything except let the program do its thing. This program will take 100 samples of size 10, 100 samples of size 25 and 100 samples of size 50 from this population and determine the proportion of individuals in each of the samples with blue eyes. Once the script finishes running, you will see a data table with three columns. The first column contains the sample proportions from samples of size 10, the second column contains the sample proportions from sample of size 25, and the last column contains the sample proportions of samples of size 50. Use the Distribution
menu in JMP to obtain a histogram of these values and use this information to answer the following questions.

c. What are the mean values of the sample proportions from the three sample sizes? What values should each of these means be close to? Why?

d. What are the standard deviation values of the sample proportions from the three sample sizes? What values should each of these standard deviations be close to? Why?

e. What is the shape of the histogram of the sample proportion values for each of the three sample sizes? Are there any differences in the shapes as the sample size increases?

f. Add a normal quantile plot to each distribution output. Describe what you see in the normal quantile plot for each sample size. Could the normal distribution be used to model the distribution of the sample proportions for any of the three sample sizes? If so, which ones.

**Activity 3.** In the first activity in lab this week you looked at selecting a random sample of 10 from the population of 250 students and recording the proportion of students in your sample with blue eyes. In this exercise we will look at how to use probability to see how likely it is to get each of the possible values of the sample proportion, \( \hat{p} \). Our population has 31.2% blue eyed people and 68.8% of people with non-blue eyes. One probability rule is that for independent trials;

\[
\text{Prob}(A \text{ and } B) = \text{Prob}(A) \times \text{Prob}(B)
\]

Note that this expands to any number of independent trials;

\[
\text{Prob}(A \text{ and } B \text{ and } C \text{ and } D) = \text{Prob}(A) \times \text{Prob}(B) \times \text{Prob}(C) \times \text{Prob}(D)
\]

a. In a random sample of 10, in order to get a value of \( \hat{p} = 0 \) you have to see none of the 10 people with blue eyes. That means that all 10 of the people chosen would have to have non-blue eyes. Write a probability expression for the event that none of the 10 people have blue eyes and compute the probability.

b. In a random sample of 10, in order to get a value of \( \hat{p} = 0.1 \) you have to see exactly one person with blue eyes. One way to do this is for the first person selected to have blue eyes and the remaining nine people to have non-blue eyes. Write a probability expression for this event and compute the probability.

c. Of course the event described in b. is not the only way to have exactly one person in a sample of 10 have blue eyes. Name another way we could get a random sample with \( \hat{p} = 0.1 \). What is the probability associated with this new event.

d. How many different ways can you get a random sample with exactly one person with blue eyes?
e. Using b., c. and d., what is the probability that \( \hat{p} = 0.1 \) for a random sample of 10 people from the population with 31.2% blue eyed people?

f. What you are calculating are binomial probabilities. This is something JMP does very easily. Go to JMP and create a new data table with three columns. Label the first column # Blue, the second column p-hat, and the third column Probability. In the first column put the numbers from 0 to 10 (you will have 11 rows). For the second column use the Cols – Formula and enter the formula # Blue divided by 10. For the third column use the Cols – Formula – Probability – Binomial Probability and enter 0.312 for p, 10 for n, and click on the # Blue column for k. The formula should look like:

\[
\text{Binomial Probability}(0.312, 10, \# \text{ Blue})
\]

What is the probability that \( \hat{p} = 0 \)? What is the probability that \( \hat{p} = 0.1 \)?

g. In order to create a distribution for the values of \( \hat{p} \) add another column to your JMP table labeled Frequency. Use Cols – Formula – Probability*100,000,000 to fill this column. Use Analyze – Distribution with p-hat in Y, Columns and Frequency in Freq and Click on OK. For your JMP output, go to Histogram Options (red pull down menu next to p-hat) and de-select Vertical. Also select a Prob axis. Right click on the horizontal axis of the histogram and select Axis Settings. Make the Minimum 0, the Maximum 1, and the Increment 0.1. Use the JMP output to answer the following questions. Turn in the JMP output.

i. Describe the shape of the distribution.

ii. Compare the mean to the median. What does this comparison tell you about the shape of the distribution?

iii. What is the mean of the distribution? How does this relate to the proportion of people with blue eyes in the population?

iv. What is the standard deviation of the distribution? How does this relate to the proportion of people with blue eyes in the population?

h. Repeat parts of f. and g. to construct the probability distribution for \( \hat{p} \) with \( n=25 \) instead of 10. You will have to create a new data table. Be careful to correctly calculate the value of \( \hat{p} \) (remember that it should go from 0 to 1). Describe the shape, center and spread and relate the center and spread to the proportion of people in the population with blue eyes.
Stat 101 L: Laboratory 10 – Answer Sheet

Names: _________________________     _________________________

_________________________     _________________________

_________________________     _________________________

Activity 1:

a.  

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

\( n = 10, \quad \hat{p} = \) 

b.  

\( n = 25, \quad \hat{p} = \) 

Activity 2:

a.  Value of \( p \)?

b.  

c.  

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Mean</th>
<th>Close to?</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

d.  

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Standard deviation</th>
<th>Close to?</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
e. 

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Shape of histogram</th>
<th>Changes in shape?</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

f. 

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Normal quantile plot</th>
<th>Normal?</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Activity 3:**

a. \[ P(\text{no one with blue eyes in sample of 10}) = \]

b. \[ P(\text{first person with blue eyes and 9 other people with non blue eyes in sample of 10}) = \]

c. Another way to get one person with blue eyes and probability associated with that event.

d. Number of ways to get exactly one person with blue eyes in a sample of 10?
e. \( P(\hat{p} = 0.1) = \)

f. \( P(\hat{p} = 0) = P(\hat{p} = 0.1) = \)

g.  
i. Describe the shape of the distribution.

ii. Compare the mean to the median. What does this comparison tell you about the shape of the distribution?

iii. What is the mean of the distribution? How does this relate to the proportion of people with blue eyes in the population?

iv. What is the standard deviation of the distribution? How does this relate to the proportion of people with blue eyes in the population?

h. \( P(\hat{p} = 0) = P(\hat{p} = 0.1) = \)

i. Describe the shape of the distribution.

ii. Compare the mean to the median. What does this comparison tell you about the shape of the distribution?

iii. What is the mean of the distribution? How does this relate to the proportion of people with blue eyes in the population?

iv. What is the standard deviation of the distribution? How does this relate to the proportion of people with blue eyes in the population?