Learning Goal: Describe the sampling distribution for sample proportions and use it to identify unusual (and more common) sample results.

Introduction: In this activity we will continue to investigate ways to use random sampling to draw an inference about a population.

Last time we gathered many random samples from a population (our class) in order to examine how random samples behave and to estimate the proportion of the population that was pierced.

This time we will begin with a hypothesis about the population. We will examine the behavior of random samples from the population when the hypothesis is true.

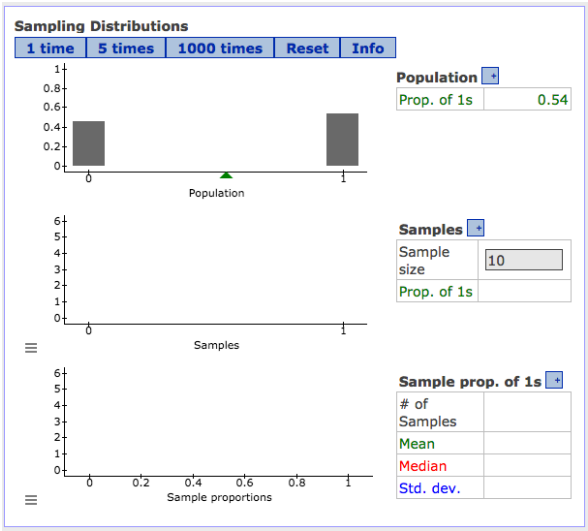
Here is a fact: According to the 2010 U.S. Census, about 54% of California residents were born in California.

Let's see if what is true for all California residents is also true for LMC students. In other words, our hypothesis is that 54% <u>of LMC students</u> were born in California.

In this activity we will use a simulation that is similar to what we have already done. But this time we will be investigating the variability in samples assuming our hypothesis about the LMC population is true.

Answer the following questions as your instructor does a simulation in StatCrunch.

Questions to answer during the demonstration



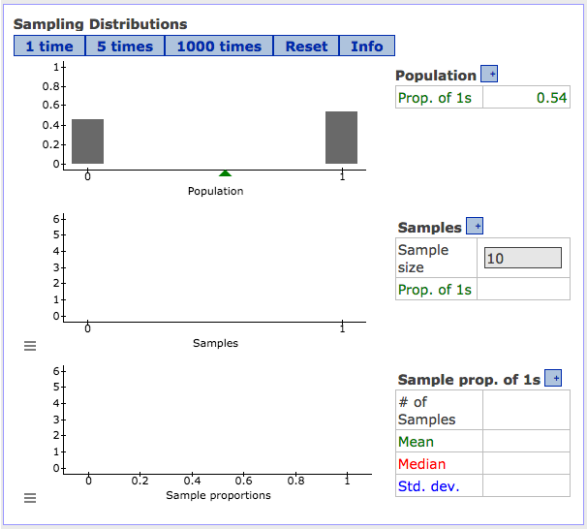
1) Your instructor set up the simulation by choosing *binary* because this categorical variable has two options: Born in CA (yes or no). Your instructor also entered p=0.54 to represent the hypothesis about the proportion of the LMC student population born in CA.

   a) How is this hypothesis represented in the graph of the population?

   b) What does a 0 represent? What does a 1 represent?

2) Your instructor will now use the simulation to select a random sample of 10 LMC students from the population that has 54% born in CA (p=0.54).

a) What proportion of the students in this random sample was born in CA? (Fill in the appropriate blank in the StatCrunch output.)



b) Make a graph in the StatCrunch printout to represent this sample.

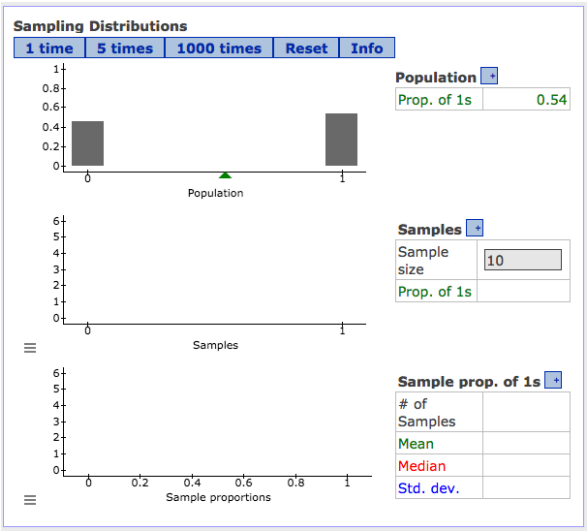
c) How many students were surveyed in this sample?

d) Put a dot in the bottom graph to represent this sample.

e) How much error is there in this random sample? (How far does it deviate from our hypothesized population proportion of 0.54?)

3) Your instructor will now use the simulation to select a 2nd random sample of 10 students from the LMC population.

a) What proportion of the students in this second random sample was born in CA? (Fill in the appropriate blank in the StatCrunch output.)



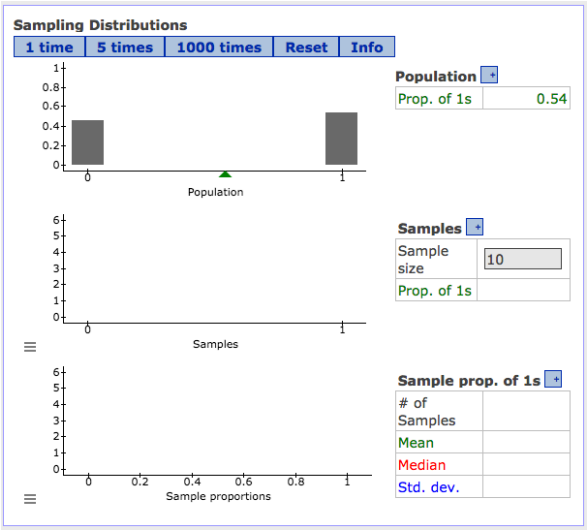
b) Make a graph in the StatCrunch printout to represent this sample.

c) How many students were surveyed in this sample?

d) Put a dot in the bottom graph to represent this sample. Also, put a dot to represent the 1st sample so that the bottom graph now has two dots.

e) How much error is there in this second random sample? (How far does it deviate from our hypothesized population proportion of 0.54?)

4) And again … your instructor will use the simulation again to select a third sample.



a) As we continue to select random samples, the sample proportions will vary. What stays the same in the simulation and does not change?

b) As before, fill in the sample proportion for the 3rd random sample and make a graph of the sample results in the StatCrunch printout.

c) Represent this sample proportion, along with the two previous one in the bottom graph.

d) Which of three sample proportions has the greatest amount of error? Which has the least?

e) In the StatCrunch printout, fill in the mean, median and standard deviation based on your instructor's display. What do each of these numbers tell us?

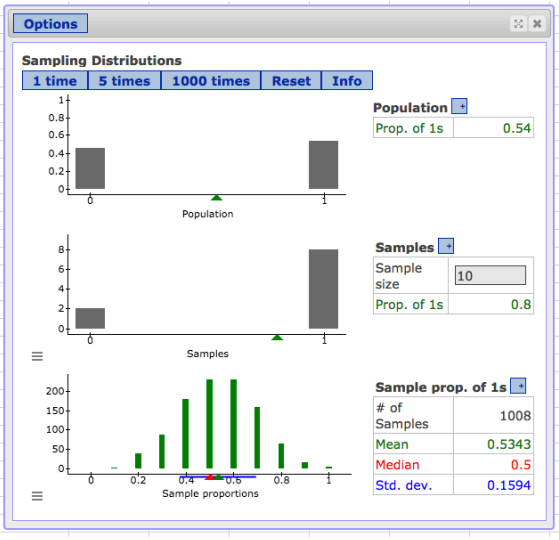
5) Your instructor will now select five additional samples. Answer these questions based on what you see in the simulation.

a) What are we hypothesizing is true about the population?

b) How many random samples have we collected? How many students are in each sample?

c) What do you think the blue line is telling us in the graph of the distribution of sample proportions?

6) Your instructor will now select 1,000 additional samples. The distribution of sample proportions shown here will be very similar to your instructor's results.



a) How many random samples have we collected?
How many students are in each sample?

b) Describe the shape of the distribution of sample proportions at this point.

c) What is the mean of the sampling distribution now? Why does this make sense?

d) On average approximately how much error is there in these samples?

e) Give an interval to represent typical sample proportions.

f) Identify a few sample proportions that are unusual.

7) Your instructor will select a random sample of 10 students from the class.

a) Is the class' sample proportion unusual when compared to the samples in the simulation? How do you know?

b) How much error is there in our class sample proportion relative to our hypothesized population proportion of 0.54? Is this much error unusual when sampling from a population of students in which 54% are born in CA?

c) Do you think our class sample could have come from a population in which 54% are born in California? Or do you think our class sample comes from a population with a different proportion of native Californians?