Learning Goals:

- Use a mathematical model of the normal curve to represent the distribution of sample proportions for a given scenario.
- Use a z-score and the standard normal model to estimate probabilities of specified events.

Introduction:

In the last activity we examined how the size of the random samples affects the average amount that the sample proportions vary from the population proportion. We discovered that large samples do a better job estimating the population proportion in the long run. In other words, large samples tend to have sample proportions that are closer to the population proportion, which means smaller amounts of error on average.

Why do we care about this?

- Because we use samples to estimate population proportions. So we need to understand how much error to expect on average.

- Also when we want to test a hypothesis about a population, we need some way of judging whether the real sample data could have come from the population described in our hypothesis. Therefore, we need to know how much sampling error will arise naturally from the sampling process in order to determine if the sample proportion from the real data is unusual or not.

Up to this point we have used simulations to estimate how much random samples behave. Let's summarize what we have learned about the distribution of sample proportions from these simulations. Fill in the blanks.

- *Center:* Proportions from random samples will vary but in the long run, sample proportions average out to the ____; therefore, the mean of the sample proportions equals ___________________________.
- *Spread:* Larger random samples will vary (*circle one*: less, more, the same amount) compared to smaller samples. Therefore, increasing the sample size will (*circle one*: increase, decrease, not affect) the standard deviation of sample proportions.
- *Shape:* For larger samples, the shape of the distribution of sample proportions is approximately ________________________.

At a time when technology was less advanced (or non-existent!), statisticians were not able to run simulations. Instead, they developed mathematical models to describe the distribution of sample proportions. A mathematical model is an equation and its associated curve.

Here are the features of the mathematical model we will use to describe the distribution of sample proportions:

- *Center:* Mean of the sample proportions is p, the population proportion.

- *Spread:* Standard deviation of the sample proportions depends on the population proportion, p, and the size of the sample, n. The standard deviation equals $\sqrt{\frac{p(1-p)}{n}}$.

- *Shape:* For large samples, the mathematical model is a normal curve. The normal curve is a good model for the distribution of sample proportions only when $np \geq 10$ and $n(1-p) \geq 10$. If these conditions are not met, then we have to rely on simulations instead of the normal curve.

Examples:

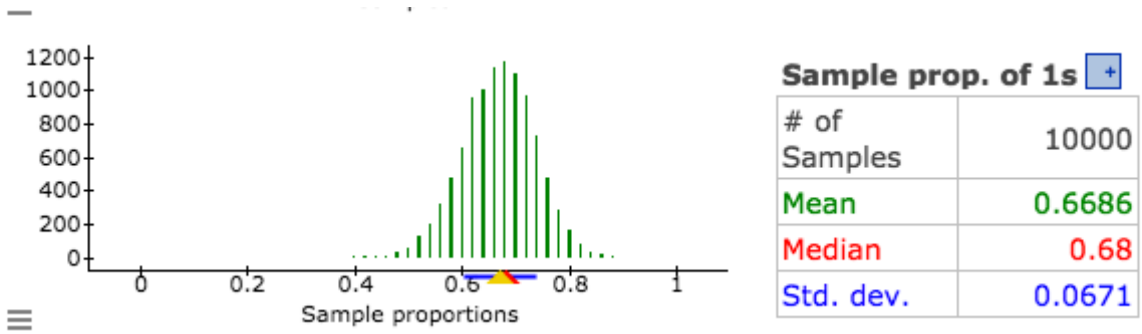
1) Use the mathematical model for the distribution of sample proportions to answer the following questions:

   Suppose that on the next ballot there is proposition on legalizing marijuana. In order for the proposition to pass, 67% of voters must vote yes. Prior to the election, a local newspaper conducts a poll of 50 randomly selected voters.

   a) Is a normal model a good fit for the distribution of sample proportions?

   b) In the long run, how much error do we expect on average in this situation based on the mathematical model?

c) We ran a simulation in StatCrunch with p=0.67 and n=50. Explain how the distribution of sample proportions from the simulation relates to the normal model.



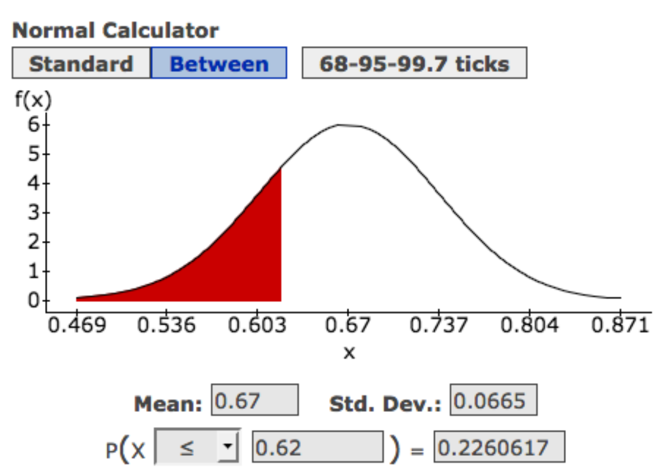
2) Now suppose prior to the election, the newspaper conducts a poll of 50 randomly selected voters and finds that 62% plan to vote yes on the proposition to legalize marijuana.

Since a 67% majority is needed to pass the proposition, this poll suggests that the proposition will not pass. However, if the 5% error is within expected amount of error for this situation, then this poll does not necessarily mean that the proposition will not pass. In other words, a sample with 62% voting YES may be fairly likely when sampling from a population with 67% voting YES. Let's see if this is the case.

a) Is 62% an unusual result for a random sample of 50 people when sampling from a population that has 67% supporting the proposition? How do you know?

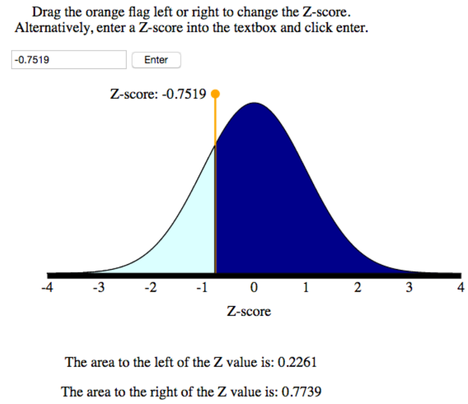
b) What is the probability that a random sample of 50 voters will have 62%, or even fewer, planning to vote YES? (Use the StatCrunch Normal Calculator to find the probability.)

How does this probability support your answer in (a)?



c) In OLI probability is determined using the standard normal curve for Z-scores.

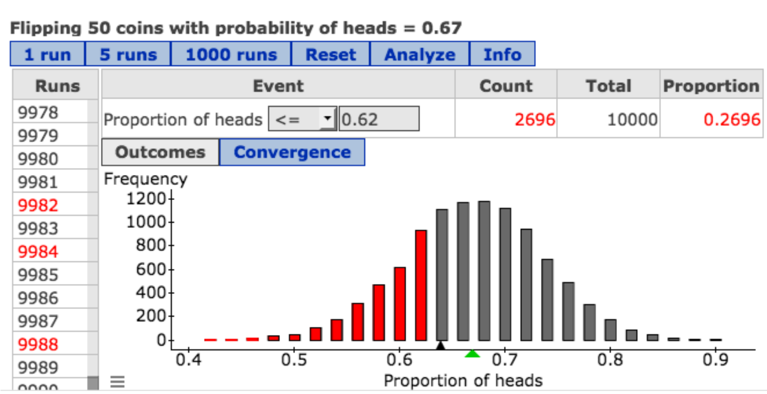
- Verify that the Z-score for a sample with 62% voting YES is approximately z = -0.7519.



- Why do we use the area to the left of z=-0.7519 to estimate the probability that 62% or fewer plan to vote YES?

- How does the probability using the standard normal curve compare to the probability using the normal curve in (b)?

d) We conducted a simulation to estimate the same probability. In this simulation a coin flip represents a vote. A head represents a YES vote. The coin is weighted so that 67% of the time it lands on a head. This represents 67% of the population voting YES. A sample consists of 50 coin tosses to represent a random sample of 50 people. The proportion of heads in 50 tosses represents the sample proportion voting YES. In the simulation, we selected 10,000 random samples (which means the coin was flipped 50 x 10,000 = 500,000 times!)



- Circle the samples in the distribution with proportions less than or equal to 0.62.
- According to the simulation, what is the probability that a random sample will have 62% or fewer planning to vote YES?

Notice that the probability estimates differ slightly when we use a mathematical model of the normal curve and when we use a simulation. This is because the normal model is a good fit for the distribution of sample proportions, but not a perfect fit! Anyway, both are estimates that come from different approaches, so don't get hung up on nailing down the probability exactly because this is not possible. What is important is to be able to use the probability estimate to draw a conclusion. In this case, both answers suggest that it is not that unusual to get a sample proportion of 62% or fewer from a population that would pass the proposition at 67%.

Group work:

1) In describing the LMC student population, suppose that a reporter writes, "The majority of LMC students have not used the CORE for academic assistance." The LMC Experience conducts a poll of 30 randomly selected LMC students to investigate whether this is true. In the poll, 30% (9 of the 30) report never using services in the CORE.

   a) Is 30% an unusual result if the reporter's claim is true? How do you know?

   b) Verify that a normal model fits the distribution of sample proportions in this case.

   c) Use the StatCrunch Normal Calculator (or the OLI z-score applet) to estimate the probability that a random sample of 30 students will have 30%, or even fewer, who have used services in the CORE, assuming that the reporter's statement about all LMC students is true. (Also sketch the normal curve and show the mean and standard deviation, or your z-score calculation.)

      (StatCrunch instructions: From the StatCrunch log in page, choose Open StatCrunch to open an blank spreadsheet. Under **Stat**, **Calculators**, choose **Normal**. Set the mean and standard deviation according to the formulas. Set the inequality as indicated in the probability question and enter the sample proportion. Hit Compute.)

2) In an article titled "Tatoos Becoming More Accepted at Work", CBS News reported in 2007 that 23% of college students were tattooed. Let's use this as a hypothesis for the proportion the population of LMC students who are tattooed.

   a) Suppose we randomly select 30 LMC students and find that about 33% are tattooed (10 out of 30). Is this an unusual result if our hypothesis is true? How do you know?

   b) Verify that a normal model is NOT a good fit for the distribution of sample proportions in this situation.

   c) Because the normal model is not a good fit, we must use a simulation to estimate probabilities. Conduct a coin-flipping simulation in StatCrunch to answer the following question: What is the probability that 33% or more are tattooed in a random sample of 30 if our hypothesis is true?

      Explain how you used the results of the simulation to estimate the probability. Include enough information about the set-up of the simulation that someone else could replicate your work.

      (StatCrunch instructions: From the StatCrunch log in page, choose Open StatCrunch to open an blank spreadsheet. Under **Applets**, **Simulation**, choose **Coin flipping**. Under **Simulate coin tosses**, set the probability of heads to the hypothesized population proportion. Under **Tally heads in tosses**, click **Proportion**. Set the inequality as indicated in the probability question and enter the sample proportion.)