Learning Goal: Test a hypothesis about a population proportion using a simulated sampling distribution or a normal model of the sampling distribution. State a conclusion in context.

Introduction:

Now we focus on the second type of inference: hypothesis testing. We are going to work on the logic of a hypothesis test now and address the more formal aspects of it later.

In hypothesis testing, we make a claim about a population proportion and use a sample proportion to test it. This is very similar to the thinking we did with simulations in the previous module.

Example: According to a 2013 survey of college students conducted by Citi and Seventeen Magazine, 20% of college students have a credit card when they start college. Is the percentage higher for community college students this year?

Here is our claim: More than 20% of community college freshmen have a credit card this year.

We start with the hypothesis that 20% of community college freshmen have a credit card. Our claim is that the percentage is higher. If our claim is true, then our hypothesis about the population is wrong.

To test the hypothesis, we select a random sample of community college freshmen. Suppose that the data shows that 25% of the sample has a credit card.

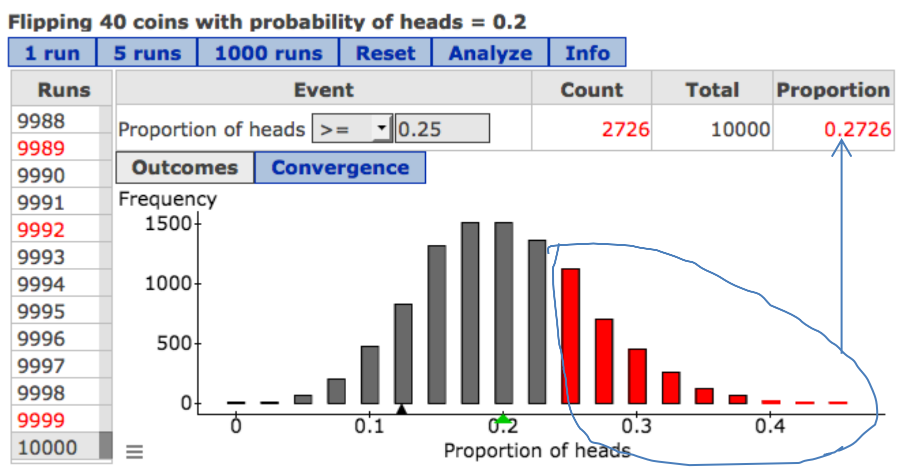
1) Does this sample support our claim? Why or why not?

But wait! Maybe when we collect random samples from a population with 20% owning a credit card, we will find that a sample proportion of 25% is not that unusual. Maybe the 5% error is not that big relative to the typical error we will see.

Surprise, surprise … we need to determine that amount of the variability expected in random samples drawn from a population where 20% have a credit card before we can draw any conclusions.

We now have two ways to examine variability: simulation or the normal model (if conditions are met.) We know that the amount of variability in random samples depends on the sample size. Let's examine what happens when we select samples of 40 freshmen (n=40). Then let's see what happens when we collect larger samples of 300 (n=300)

2) *Sample size of 40:* Here are the results from a coin-flipping simulation in StatCrunch with probability of a head set to 0.20. Each "run" is 40 coin flips representing a random sample of 40 community college freshmen.

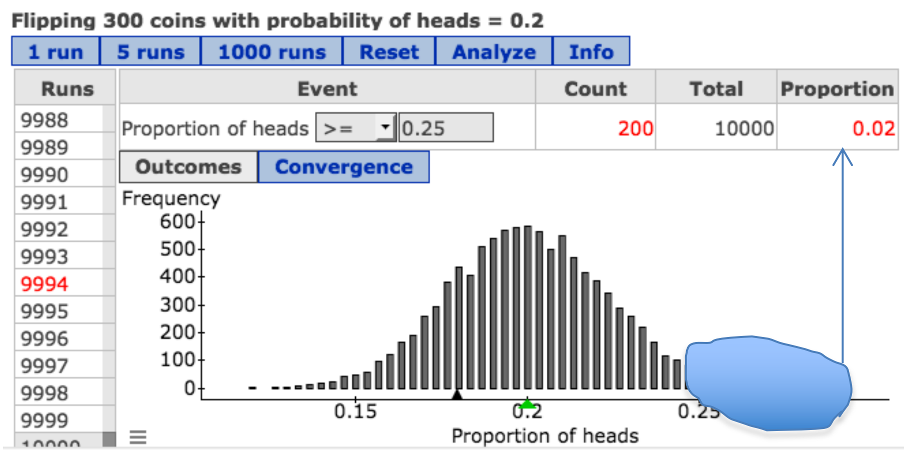


a) With a sample of 40 freshmen, is a proportion of 0.25 unusual when the population proportion is 0.20? How do you know?

b) Based on the simulation, what is the probability that a random sample of 40 has 25% or more with a credit card? How do you know?

c) What does this suggest about our claim? Does our sample data provide strong evidence that more than 20% of the entire population own a credit card? Why or why not?

3) *Sample size of 300:* Here are the results from a coin-flipping simulation in StatCrunch with probability of a head set to 0.20. Each "run" is 300 coin flips representing a random sample of 300 community college freshmen.



a) With a sample of 300 freshmen, is a proportion of 0.25 unusual when the population proportion is 0.20? How do you know?

b) What is the probability that a random sample of 300 has 25% or more with a credit card? How do you know?

c) What does this suggest about our claim? Does our sample data provide strong evidence that more than 20% of the entire population own a credit card? Why or why not?

Summary:

Are we right that more than 20% of the all community college freshmen own a credit card? In other words, is our claim true?

- If the sample proportion of 0.25 is fairly typical, then the associated probability is large. This means that it is not surprising to see random samples with 25% or more owning a credit card when 20% of the population owns one. In other words, even though more than 20% of the sample had a credit card, **the deviation is not large enough** to support our claim that the population proportion is also larger than 20%.

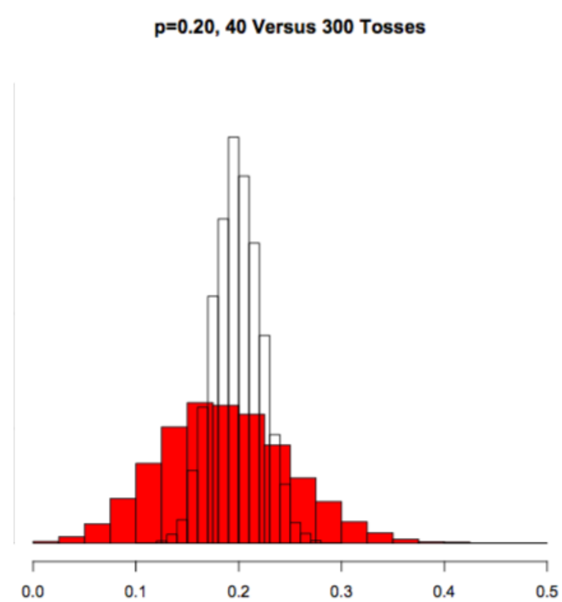
  Therefore, we conclude that our **claim is not true.**

- If the sample proportion of 0.25 is unusual, then the associated probability is small. This means that it is surprising to see random samples with 25% or more owning a credit card when 20% of the population owns one. In other words, this **deviation is large enough** to support our claim that the population proportion is also larger than 20%.

  Therefore, we conclude that our **claim is true.**

4) How can the same sample result (a sample proportion of 25% in this case) lead to different conclusions about the population proportion?

   We put both sampling distributions on the same axis for easier comparison. Use this image to explain why a sample proportion of 25% gives different conclusions when n=40 and n=300.



Group work:

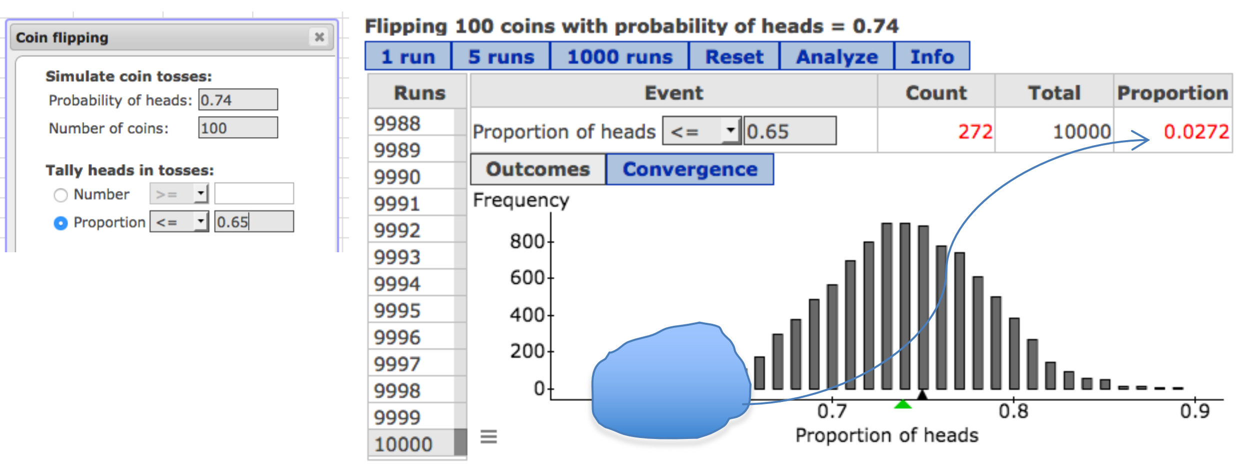
5) "Authoritative numbers are hard to come by, but according to a 2002 confidential survey of 12,000 high school students, 74 percent admitted cheating on an examination at least once in the past year."
Source: http://abcnews.go.com/Primetime/story?id=132376&page=1

A Principal of a local high school is appalled by this statistic. She wants to determine if the situation is any better at her school.

She hires a firm to conduct a confidential survey of a random sample of 100 students at her school. Sixty-five percent of the 100 students admit to cheating on exams.

a) Identify the following for this scenario:
   - The hypothesized population proportion
   - The sample proportion from the data
   - The sample size

b) The consultant runs a simulation and gets the following results. Explain how your answers in (a) relate to this simulation.



c) Is a sample proportion of 65% unusual when drawing random samples of 100 students from a population in which 74% cheat on exams? How do you know?

d) Based on this simulation, what is the probability that a random sample of students will have 65% or fewer cheating on exams if 74% of the population of students is cheating? How do you know?

e) What does this simulation suggest about the principal's hope that cheating is less prevalent at her school?

f) Could the principal have used a normal model instead of a simulation? Why or why not?

g) The principal wants an estimate of the proportion of students who cheat on exams at her school. Calculate a 95% confidence interval and interpret it. Does the confidence interval support your answer to (e)?