Learning Goal: Recognize the logic behind a hypothesis test and how it relates to the P-value.
Learning Objective: Interpret a P-value as a probability in the context of a statistical study.

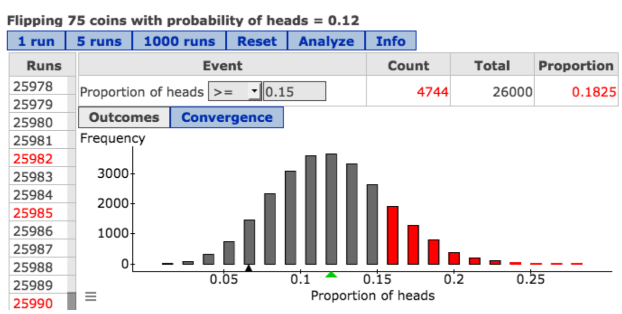
Introduction:
We now know how to use the P-value to draw a conclusion from a hypothesis test. In this short activity we will focus on the meaning of the P-value.

Let's return to a familiar example: A 2011 California Student Survey found that in any 30-day period, almost 12 percent of students in the 9th grade admit to using drugs at least once at school.
Source: http://www.huffingtonpost.com/marsha-rosenbaum/drug-education_b_3906983.html

To investigate this issue in your community, the school board hires a research firm to conduct a survey of a random sample of 9th graders in the district to determine if drug usage is greater than the statewide rate of 12%. Seventy-five students respond to the anonymous survey and 11 admit to drug usage at school during the last month. (11 out of 75 is about 15%)

The sample data is used to test the hypotheses:
Ho: p=0.12 (12% of the population of 9th graders have used drugs at school)
Ha: p>0.12 (More than 12% of the population of 9th graders have used drugs at school)

Here is the StatCrunch image from the simulation of selecting 26,000 random samples of 75 students.



The P-value is about 0.18.

What can we conclude?

In this activity we will focus on the **meaning of the P-value**, instead of focusing on how to use it to draw a conclusion.

What does P-value = 0.18 mean?

Here is one way to express it: *If the population proportion is 0.12, we expect to see sample proportions vary from this. But will sample proportions as large or larger than 0.15 occur very often? How often? What's the probability? The probability is 0.18.*

Here is another way to describe the P-value with a bit more context: *In random sampling from a population of 9th graders with 12% drug usage rate at school, we expect to see variability. With random samples of 75 students, there is an 18% chance that a sample will have 15% or more reported using drugs at school.*

Here is another way to describe the P-value. This one is very concise: *There is an 18% chance that a random sample of 75 ninth-graders will have 15% or more reported using drugs at school when we sample from a population of ninth-graders in which 12% are using drugs.*

Notice that each of these descriptions includes the following information:

- the null hypothesis
- the description of the samples included in the P-value
- the probability

The last two descriptions are better because they add context about

- the population
- the variable

1) Go back to the three interpretations of the P-value 0.18 and identify the null hypothesis, the description of the samples included in the P-value and the probability in each interpretation.

In general, a P-value is the probability that a sample statistic (such as $\hat{p}$) is as extreme or more extreme than the statistic from the observed sample, if the null hypothesis is true.

When you interpret a P-value, don't use this general interpretation. Instead describe it in a way that references the context of the problem, as shown in the examples above.

Group work:

2) "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

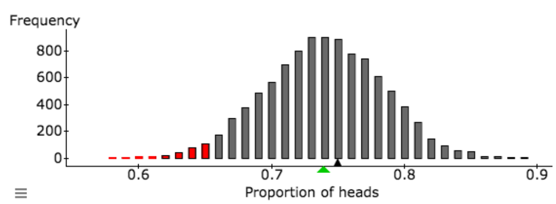
A Principal of a local high school hopes that cheating is less prevalent at her school. She hires a research firm to conduct a confidential survey of a random sample of 100 students at her school. Sixty-five percent of the 100 students admit to cheating on exams.

In a test of the hypotheses: Ho: p = 0.74, Ha: p < 0.74, the P-value is 0.027.

a) In the following interpretation of the P-value, identify the null hypothesis, the description of the samples included in the P-value, and the probability.

   There is a 2.7% chance that a random sample of 100 students will have less than 65% admitting to cheating on exams if 74% of the student population is cheating on exams.

b) In the simulated distribution of sample proportions, circle the samples included in the P-value. Label the hypothesized population proportion and the P-value.



3) *Do you feel that the death penalty acts as a deterrent to the commitment of murder, that it lowers the murder rate, or not?* According to Gallup polls, 64% of adults in 2011 answered "no, does not".
Source: http://www.gallup.com/poll/1606/death-penalty.aspx

Suppose that in a random sample of adults this year, 68% answer "no, does not."

Has the percentage of the public with opinion increased since 2011?

We test the following hypotheses: Ho: $p = 0.64$, Ha: $p > 0.64$. The P-value is 0.20.

Which of the following interpretations of the P-value are accurate and complete? For those that are not accurate, explain why.

- The probability that more than 64% of adults will answer “no” this year is 0.20.
- The probability that more than 68% of adults will answer “no” this year is 0.20.
- If random samples are greater than 68%, then there is a 20% chance that the null hypothesis is true this year.
- There is a 20% chance that random samples will have more than 68% answering “no” if 64% of the population has this opinion.
- If the null hypothesis is true, then the probability that random samples will have $\hat{p}$ greater than 0.68 is 0.20.

4) In a Gallup poll about the cause of record highs in 2015 temperatures, 49% of adults answered “human-caused climate change,” 46% chose “natural changes in the Earth’s temperatures,” and 5% had no opinion.

Suppose that in a random sample of adults this year, 55% attribute warmer weather patterns to “human-caused climate change.”

Is the percentage of the public with this opinion higher this year than in 2015? We test the following hypotheses: Ho: $p = 0.49$, Ha: $p > 0.49$. The P-value is 0.03. Interpret the P-value as a probability statement.