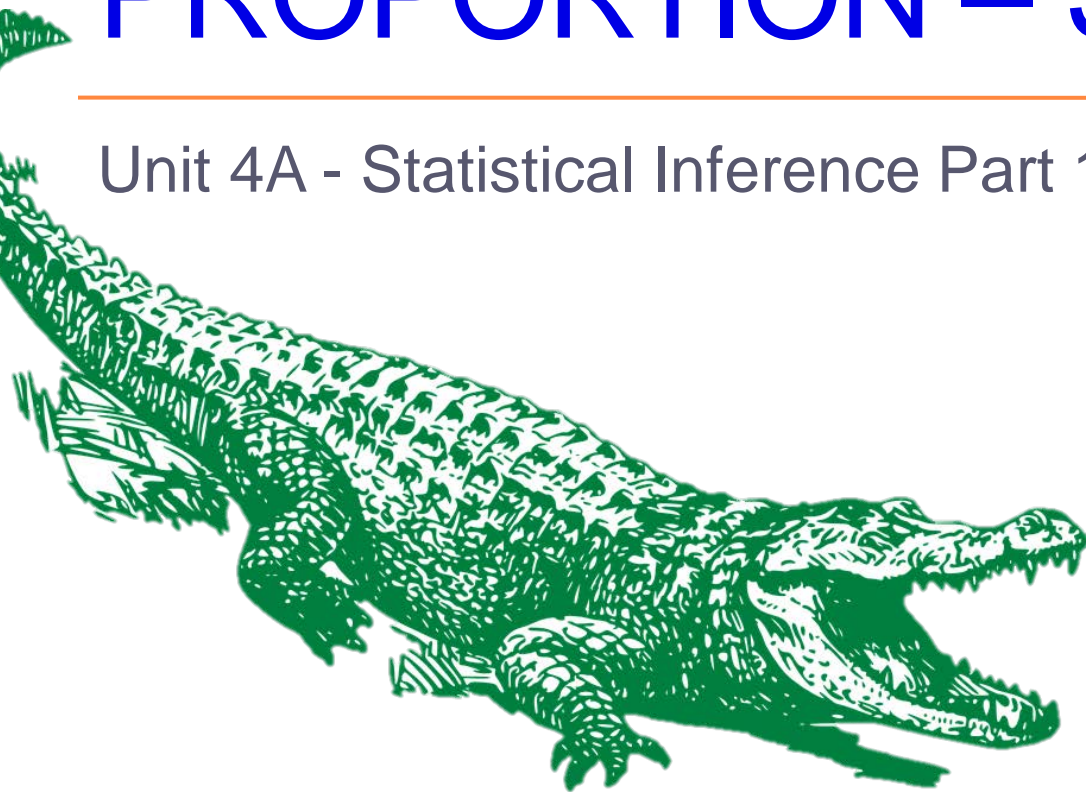
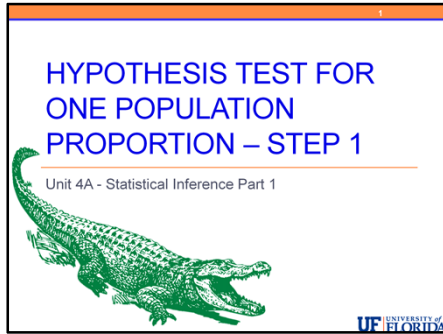


# HYPOTHESIS TEST FOR ONE POPULATION PROPORTION – STEP 1

---

Unit 4A - Statistical Inference Part 1





Now we will look at the details by hand for one situation, that for one population proportion. This will be the only type of problem we will work manually in this course.

This test is widely known as the “**z-test for the population proportion (p).**”

Our goal is to use this example to give you the tools you need to understand how this process works. After working a few problems, you should review the earlier material again. You will likely need to review the terminology and concepts a few times.

This is the easiest method to use to illustrate the process since we can use z-scores and the normal distribution instead of another more complex distribution.

After this test, we will rely on the software to find p-values (and confidence intervals) for us.

This method will be appropriate when we have a sample proportion measured from a random sample taken from the population. The variable will be categorical in nature and we are interested in estimating the true proportion in a certain category.

In this part of our discussion on hypothesis testing, we will go into details that we did not go into before. More specifically, we will use this test to introduce the idea of a **test statistic**, and details about **how p-values are calculated**.


# Steps in Hypothesis Testing

- STEP 1. State the **null** and **alternative hypotheses**.
- STEP 2. Check conditions; **summarize** with **test statistic**
- STEP 3. Find **p-value**
  - Probability of observing data like those observed (or more extreme) assuming  $H_0$  is true
- STEP 4. Based on p-value, decide whether we have enough evidence to:
  - Reject  $H_0$  (and accept  $H_a$ ) OR Fail to Reject  $H_0$
  - AND THEN most importantly ... **draw our conclusions in context**

2

### Steps in Hypothesis Testing

- STEP 1. State the **null** and **alternative hypotheses**.
- STEP 2. Check conditions; **summarize** with **test statistic**
- STEP 3. Find **p-value**
  - Probability of observing data like those observed (or more extreme) assuming  $H_0$  is true
- STEP 4. Based on p-value, decide whether we have enough evidence to:
  - Reject  $H_0$  (and accept  $H_a$ ) OR Fail to Reject  $H_0$
  - AND THEN most importantly ... **draw our conclusions in context**



Recall that there are basically 4 steps in the process of hypothesis testing:

STEP 1. State the **null** and **alternative hypotheses**.

STEP 2. Collect data, check conditions, and **summarize** with **test statistic**

STEP 3. Find **p-value**

Probability of observing data like those observed (or more extreme) assuming  $H_0$  is true

STEP 4. Based on p-value, decide whether we have enough evidence to:

Reject  $H_0$  (and accept  $H_a$ ) OR Fail to Reject  $H_0$


AND THEN most importantly ... **draw our conclusions in context**

# One Sample Z-Test for a Population Proportion

- Null Hypothesis:
  - $H_0: p = p_0$
  
- Alternative Hypothesis is one of:
  - $H_a: p < p_0$
  - $H_a: p > p_0$
  - $H_a: p \neq p_0$

One Sample Z-Test for a Population Proportion

- Null Hypothesis:
  - $H_0: p = p_0$
- Alternative Hypothesis is one of:
  - $H_a: p < p_0$
  - $H_a: p > p_0$
  - $H_a: p \neq p_0$



At the moment, we want to focus on setting up the hypotheses in this case. Although you may feel you can always set up the hypotheses easily in words, we want to be able to write the hypotheses in terms of the parameter of interest in the given problem.

Here that is the population proportion.

We want to test whether the population proportion,  $p$  is equal to a specified value – which we will call  $p_0$  (as the null hypothesis).

Versus, the alternative hypothesis – can be that  $p$  is less than  $p_0$ ,  $p$  is greater than  $p_0$  or  $p$  is simply different from  $p_0$ .

The first two are called one-tailed tests as they are interested in only one direction (less than or greater than).

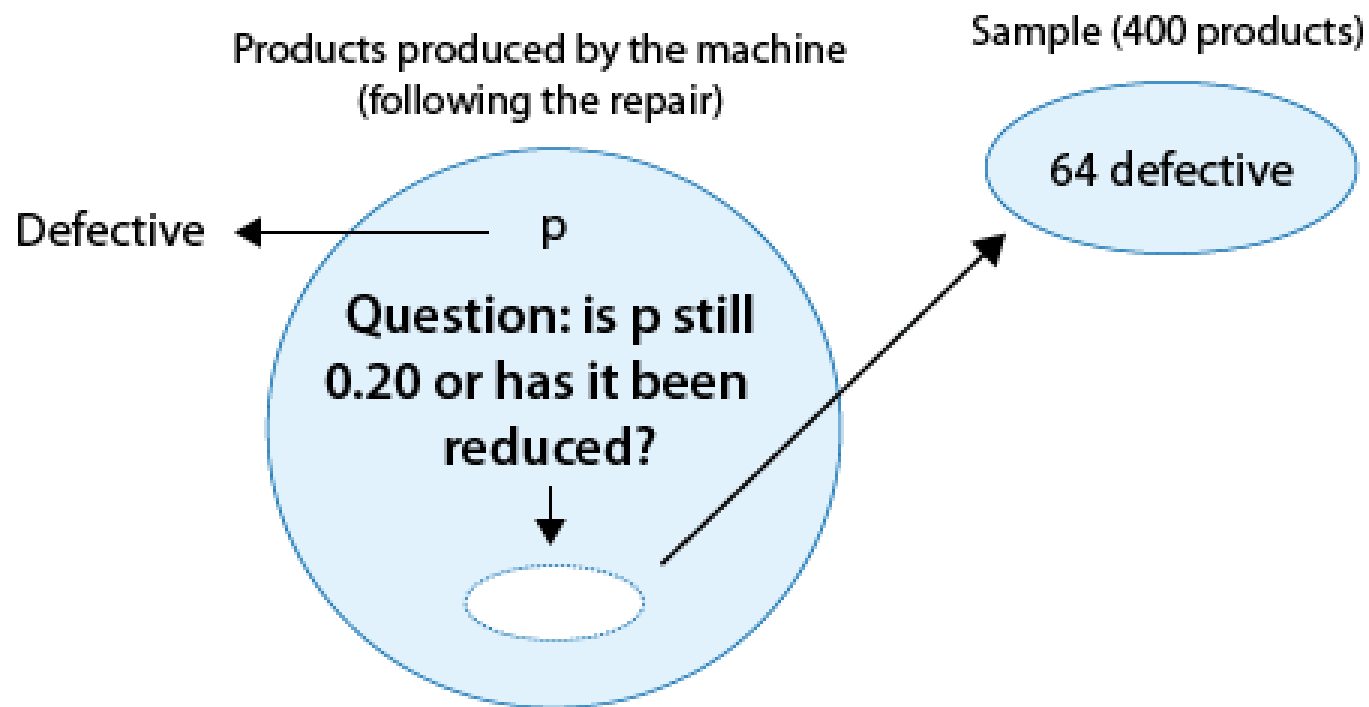
The last, where we have  $H_a: p$  is not equal to  $p_0$  is a two-tailed test. This type of test is almost always used in practice, even when only one direction is really of interest and for most of the semester we will focus on this type of test.

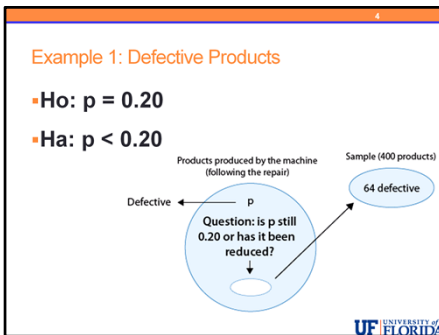
But... for this first situation where we are learning to conduct the process manually, we will learn how to deal with all three of these possible alternative hypotheses.

Let's start by introducing the three examples, which will be the leading examples in our discussion and setting up the hypotheses for each.

## Example 1: Defective Products

- **$H_0: p = 0.20$**
- **$H_a: p < 0.20$**





A machine is known to produce 20% defective products, and is therefore sent for repair.

After the machine is repaired, 400 products produced by the machine are chosen at random and 64 of them are found to be defective.

Do the data provide enough evidence that the proportion of defective products produced by the machine ( $p$ ) has been **reduced** as a result of the repair?

The figure displays the information, as well as the question of interest.

The question of interest helps us formulate the null and alternative hypotheses in terms of  $p$ , the proportion of defective products produced by the machine following the repair: We have

- $H_0: p = 0.20$  (No change; the repair did not help).

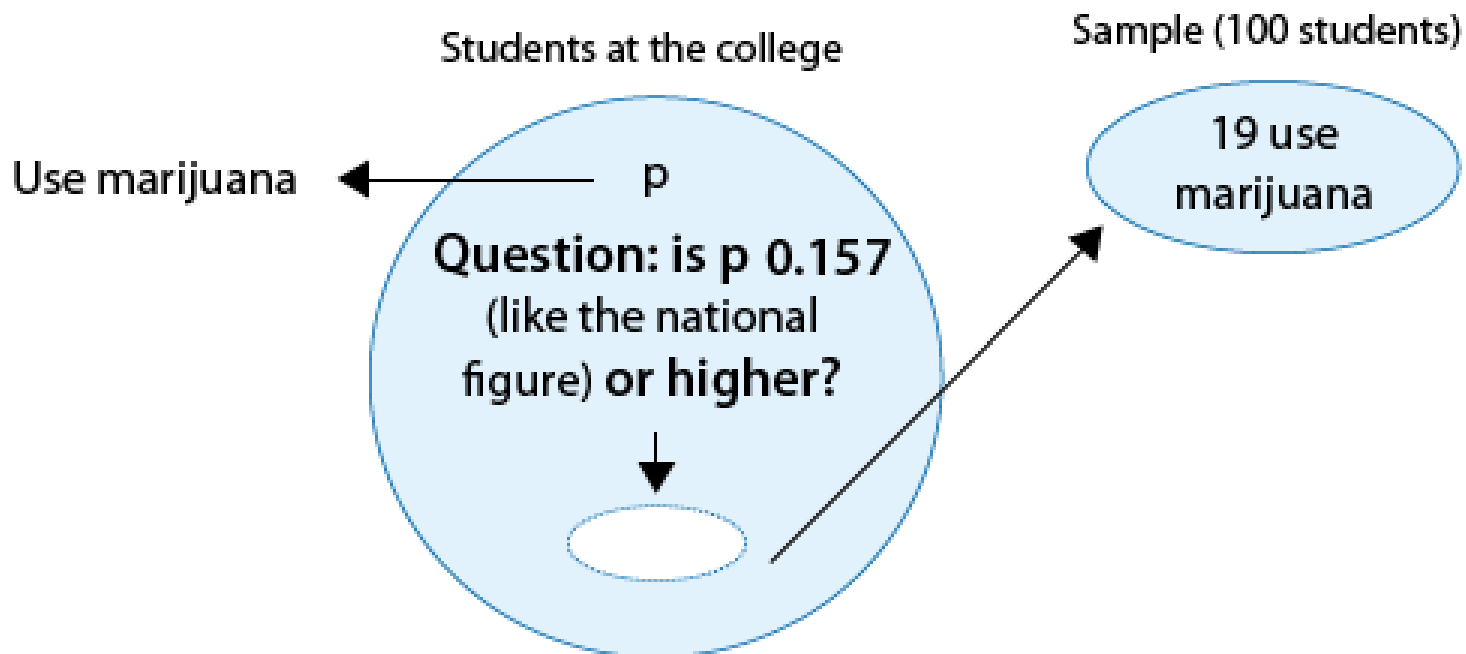
**Versus**

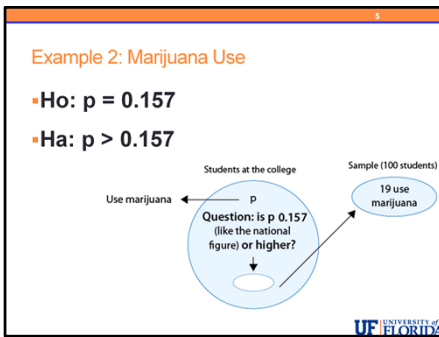
- $H_a: p < 0.20$  (The repair was effective at reducing the proportion of defective parts).



## Example 2: Marijuana Use

- **Ho:  $p = 0.157$**
- **Ha:  $p > 0.157$**





There are rumors that students at a certain liberal arts college are more inclined to use drugs than U.S. college students in general.

Suppose that in a simple random sample of 100 students from the college, 19 admitted to marijuana use.

Do the data provide enough evidence to conclude that the proportion of marijuana users among the students in the college ( $p$ ) is **higher** than the national proportion, which is 0.157? (This number is reported by the Harvard School of Public Health.)

Again, the figure displays the information as well as the question of interest.

As before, we can formulate the null and alternative hypotheses in terms of  $p$ , the proportion of students in the college who use marijuana:

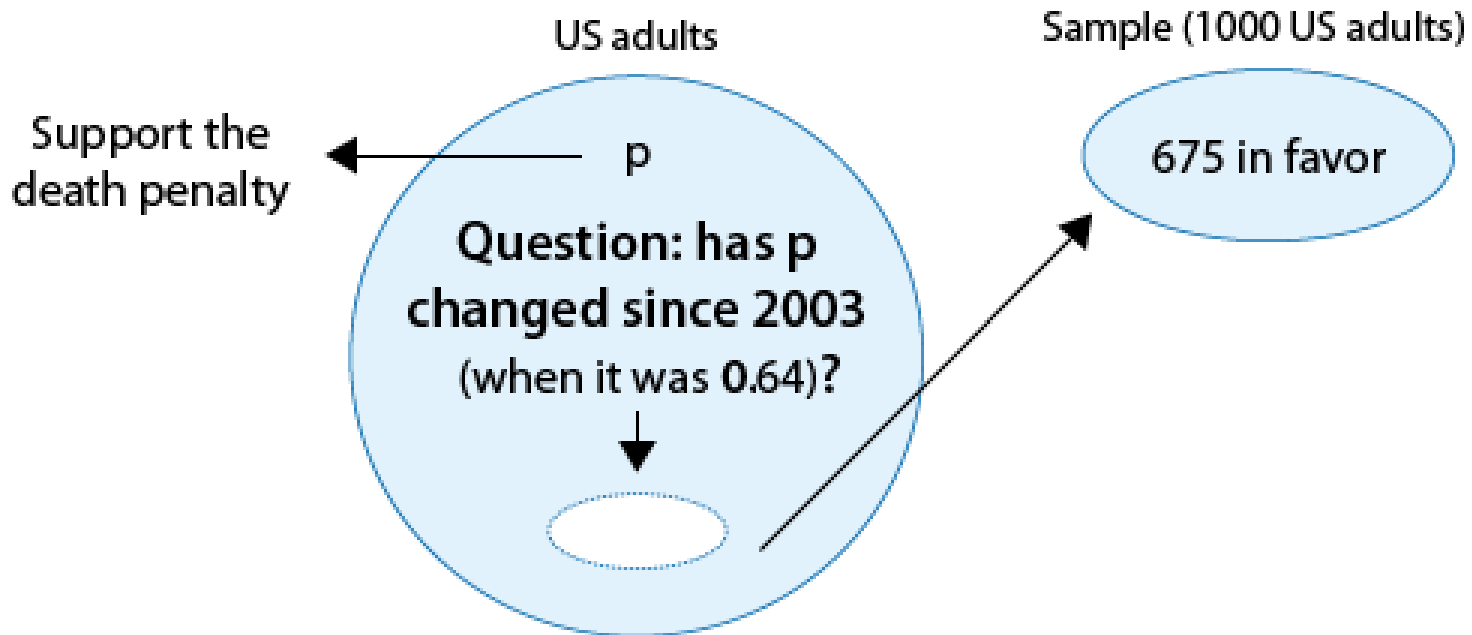
- $H_0: p = 0.157$  (same as among all college students in the country).

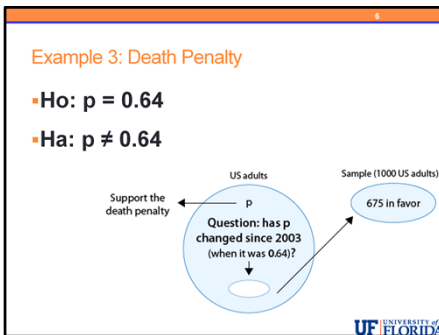
**Versus**

- $H_a: p > 0.157$  (higher than the national figure).

## Example 3: Death Penalty

- **$H_0: p = 0.64$**
- **$H_a: p \neq 0.64$**





Polls on certain topics are conducted routinely in order to monitor changes in the public's opinions over time. One such topic is the death penalty.

In 2003 a poll estimated that 64% of U.S. adults support the death penalty for a person convicted of murder. In a more recent poll, 675 out of 1,000 U.S. adults chosen at random were in favor of the death penalty for convicted murderers.

Do the results of this poll provide evidence that the proportion of U.S. adults who support the death penalty for convicted murderers ( $p$ ) **changed** between 2003 and the later poll?

Here is a figure that displays the information, as well as the question of interest.

Again, we can formulate the null and alternative hypotheses in term of  $p$ , the proportion of U.S. adults who support the death penalty for convicted murderers.

- $H_0: p = 0.64$  (No change from 2003).

### Versus

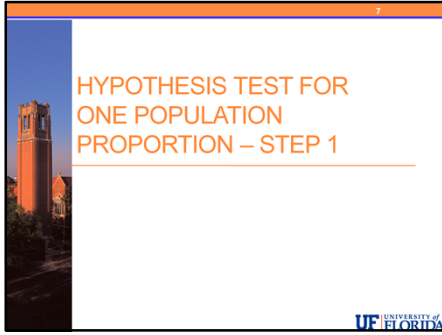
- $H_a: p \neq 0.64$  (Some change since 2003)

For most of the semester we will use these two-sided tests which will make setting up hypotheses somewhat easier but in this section, it will be important to correctly define the alternative hypothesis as the direction (less than, greater than, or different from) will play a key role in how we calculate our p-values.



# HYPOTHESIS TEST FOR ONE POPULATION PROPORTION – STEP 1

---



For one population proportion, we have discussed how to complete STEP 1 by setting up our hypotheses.

Next we will look at how to calculate the test statistic for STEP 2.