

 Lagunita is retiring and will shut down at 12 noon Pacific Time on March 31, 2020. A few courses may be open for self-enrollment for a limited time. We will continue to offer courses on other online learning platforms; visit <http://online.stanford.edu>.

Course > Inference: Hypothesis Testing for the Population Proportion > Issues in Hypothesis Testing > Hypothesis Testing for the Population Proportion p: Confidence Intervals

 Bookmark this page

Hypothesis Testing for the Population Proportion p: Confidence Intervals

Learning Objective: Apply the concepts of: sample size, statistical significance vs. practical importance, and the relationship between hypothesis testing and confidence intervals.

4. Hypothesis Testing and Confidence Intervals

The last topic we want to discuss is the relationship between hypothesis testing and confidence intervals. Even though the flavor of these two forms of inference is different (confidence intervals estimate a parameter, and hypothesis testing assesses the evidence in the data against one claim and in favor of another), there is a strong link between them.

We will explain this link (using the z-test and confidence interval for the population proportion), and then explain how confidence intervals can be used after a test has been carried out.

Recall that a confidence interval gives us a set of plausible values for the unknown population parameter. We may therefore examine a confidence interval to informally decide if a proposed value of population proportion seems plausible.

For example, if a 95% confidence interval for p , the proportion of all U.S. adults already familiar with Viagra in May 1998, was (0.61, 0.67), then it seems clear that we should be able to reject a claim that only 50% of all U.S. adults were familiar with the drug, since based on the confidence interval, 0.50 is not one of the plausible values for p .

In fact, the information provided by a confidence interval can be formally related to the information provided by a hypothesis test. (**Comment:** The relationship is more straightforward for two-sided alternatives, and so we will not present results for the one-sided cases.)

Suppose we want to carry out the **two-sided test**:

$$H_0 : p = p_0$$

$$H_a : p \neq p_0$$

using a significance level of 0.05.

An alternative way to perform this test is to find a 95% **confidence interval** for p and check:

If p_0 falls **outside** the confidence interval, **reject** H_0 .

If p_0 falls **inside** the confidence interval, **do not reject** H_0 .

In other words, if p_0 is not one of the plausible values for p , we reject H_0 .

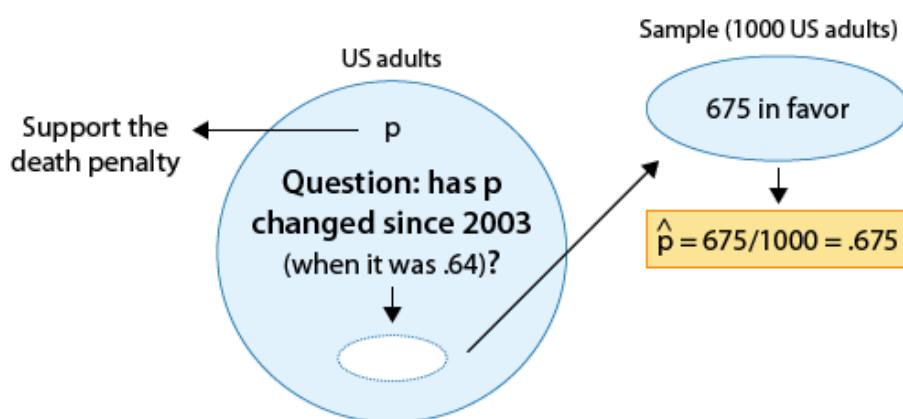
If p_0 is a plausible value for p , we cannot reject H_0 .

(**Comment:** Similarly, the results of a test using a significance level of 0.01 can be related to the 99% confidence interval.)

Let's look at two examples:

Example

Recall example 3, where we wanted to know whether the proportion of U.S. adults who support the death penalty for convicted murderers has changed since 2003, when it was 0.64.



We are testing:

$$H_0 : p = .64$$

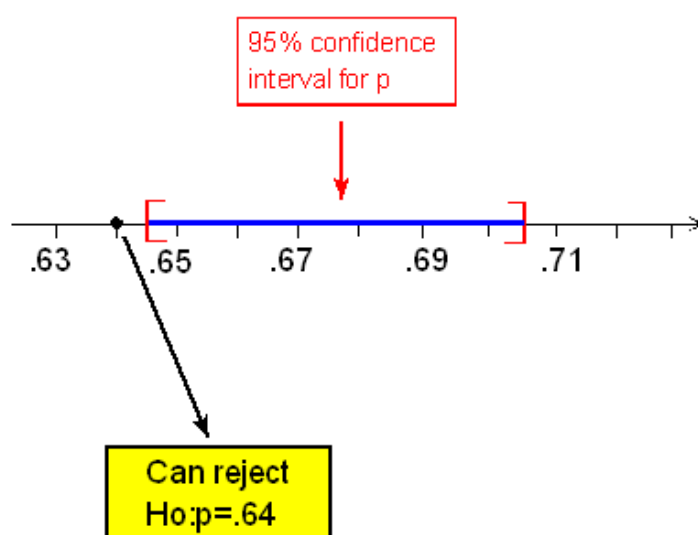
$$H_a : p \neq .64$$

and as the figure reminds us, we took a sample of 1,000 U.S. adults, and the data told us that 675 supported the death penalty for convicted murderers (i.e. $\hat{p} = 0.675$).

A 95% confidence interval for p, the proportion of **all** U.S. adults who support the death penalty, is:

$$0.675 \pm 2 \sqrt{\frac{0.675(1-0.675)}{1000}} \approx 0.675 \pm 0.03 = (0.645, 0.705)$$

Since the 95% confidence interval for p does not include 0.64 as a plausible value for p, we can reject H_0 and conclude (as we did before) that the proportion of U.S. adults who support the death penalty for convicted murderers has changed since 2003.



Example

You and your roommate are arguing about whose turn it is to clean the apartment. Your roommate suggests that you settle this by tossing a coin and takes one out of a locked box he has on the shelf. Suspecting that the coin might not be fair, you decide to test it first. You toss the coin 80 times, thinking to yourself that if, indeed, the coin is fair, you should get around 40 heads. Instead you get 48 heads. You are puzzled. You are not sure whether getting 48 heads out of 80 is enough evidence to conclude that the coin is unbalanced, or whether this a result that could have happened just by chance when the coin is fair.

Statistics can help you answer this question.

Let p be the true proportion (probability) of heads. We want to test whether the coin is fair or not:

$$H_0 : p = .5$$

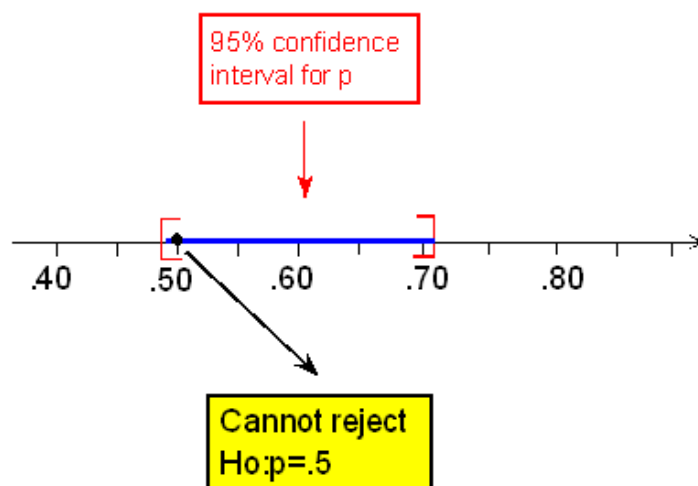
$$H_a : p \neq .5$$

The data we have are that out of $n = 80$ tosses, we got 48 heads, or that the sample proportion of heads is: $\hat{p} = \frac{48}{80} = 0.6$

The 95% confidence interval for p , the true proportion of heads for this coin, is:

$$0.6 \pm 2 \cdot \sqrt{\frac{0.6(1-0.6)}{80}} \approx 0.6 \pm 0.11 = (0.49, 0.71)$$

Since in this case 0.5 is one of the plausible values for p , we cannot reject H_0 . In other words, the data do not provide enough evidence to conclude that the coin is not fair.



Scenario: Online Credit Card Fraud

The UCLA Internet Report (February 2003) estimated that roughly 8.7% of Internet users are extremely concerned about credit card fraud when buying online. A study was designed in order to examine whether that proportion has changed since. Let p be the proportion of all Internet users who are concerned about credit card fraud. In this study we are therefore testing:

$$H_0 : p = .087$$

$$H_a : p \neq .087$$

Based on the collected data, a 95% confidence interval for p was found to be (0.08, 0.14).

Did I Get This

1/1 point (graded)

At the 0.05 significance level, we therefore:

- ☐ reject H_0 , since 0.087 falls inside the confidence interval.
- ☒ cannot reject H_0 , since 0.087 falls inside the confidence interval. ✓
- ☐ reject H_0 , since 0.087 falls outside the confidence interval.
- ☐ cannot reject H_0 , since 0.087 falls outside the confidence interval.

Answer

Correct:

Indeed, since 0.087 falls inside the 95% confidence interval for p , this means that we are 95% confidence that 0.087 is a plausible value for p . This is equivalent to not being able to reject $H_0: p = 0.087$ at the 0.05 significance level.

Submit**Scenario: Internet at Work for Personal Use**

The UCLA Internet Report (February 2003) estimated that roughly 60.5% of U.S. adults use the Internet at work for personal use. A follow-up study was conducted in order to explore whether that figure has changed since. Let p be the proportion of U.S. adults who use the Internet at work for personal use.

$$H_0 : p = .605$$

$$H_a : p \neq .605$$

Based on the collected data, the p-value of the test was found to be 0.001.

Did I Get This

1/1 point (graded)

Based on the collected data, the p-value of the test was found to be 0.001.

Which of the following could be the 95% confidence interval for p , based on the same data? Check all that apply.

- ☐ (0.595 , 0.639)

☐ (0.580 , 0.624)☒ (0.625 , 0.669)☒ (0.551 , 0.595)**Answer**

Correct:

Since the p-value of the test was 0.001, H_0 was rejected (at the 0.05 significance level, or an even lower level) the 95% confidence interval for p cannot include 0.605 as a plausible value. Therefore, (0.625 , 0.669) and (0.551 , 0.595) are both plausible confidence intervals.

[Submit](#)**Comment**

The context of the last example is a good opportunity to bring up an important point that was discussed earlier.

Even though we use 0.05 as a cutoff to guide our decision about whether the results are significant, we should not treat it as inviolable and we should always add our own judgment. Let's look at the last example again.

It turns out that the p-value of this test is 0.0734. In other words, it is maybe not extremely unlikely, but it is quite unlikely (probability of 0.0734) that when you toss a **fair** coin 80 times you'll get a sample proportion of heads of $48/80 = 0.6$ (or even more extreme). It is true that using the 0.05 significance level (cutoff), 0.0734 is not considered small enough to conclude that the coin is not fair. However, if you really don't want to clean the apartment, the p-value might be small enough for you to ask your roommate to use a different coin, or to provide one yourself!

Open Learning Initiative [↗](#)

[↗](#) Unless otherwise noted this work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License [↗](#).

© All Rights Reserved