Hypothesis Testing for the Population Proportion p: Summary of Issues

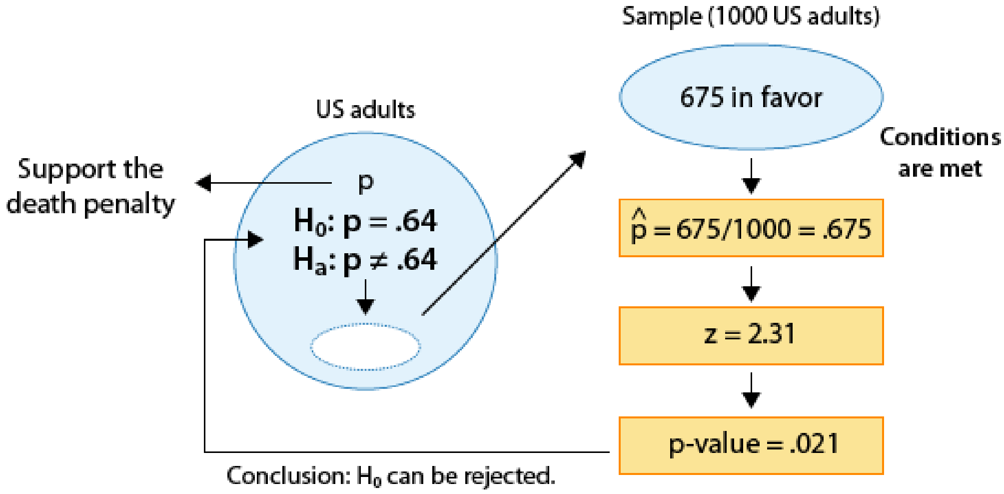
Learning Objective: Apply the concepts of: sample size, statistical significance vs. practical importance, and the relationship between hypothesis testing and confidence intervals.

Here is our final point on this subject:

When the data provide enough evidence to reject H_o, we can conclude (depending on the alternative hypothesis) that the population proportion is either less than, greater than or not equal to the null value p_0. However, we do not get a more informative statement about its actual value. It might be of interest, then, to follow the test with a 95% confidence interval that will give us more insight into the actual value of p.

Example

In our example 3,



we concluded that the proportion of U.S. adults who support the death penalty for convicted murderers has changed since 2003, when it was 0.64. It is probably of interest not only to know that the proportion has changed, but also to estimate what it has changed to. We've calculated the 95% confidence interval for p on the previous page and found that it is (0.645, 0.705).

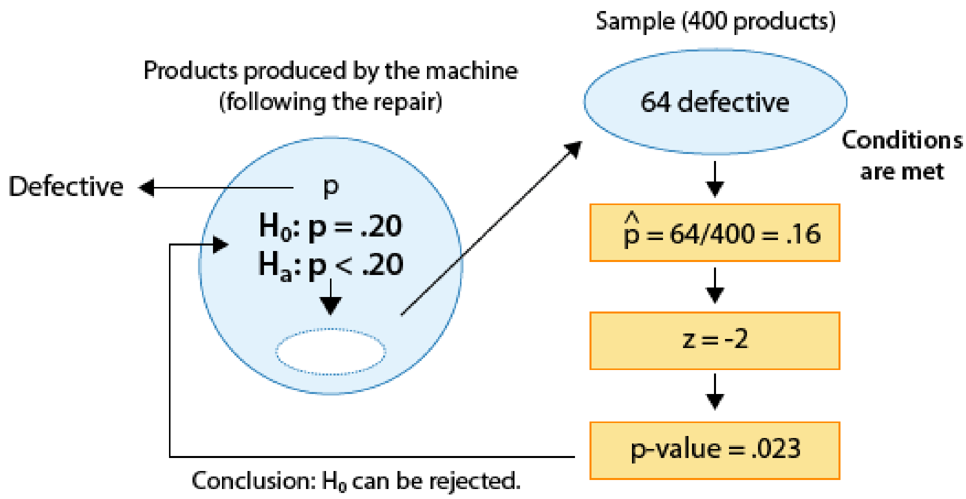
We can combine our conclusions from the test and the confidence interval and say:

Data provide evidence that the proportion of U.S. adults who support the death penalty for convicted murderers has changed since 2003, and we are 95% confident that it is now between 0.645 and 0.705. (i.e. between 64.5% and 70.5%).

Example

Let's look at our example 1 to see how a confidence interval following a test might be insightful in a different way.

Here is a summary of example 1:



We conclude that as a result of the repair, the proportion of defective products has been reduced to below 0.20 (which was the proportion prior to the repair). It is probably of great interest to the company not only to know that the proportion of defective has been reduced, but also estimate what it has been reduced to, to get a better sense of how effective the repair was. A 95% confidence interval for p in this case is:

$$0.16 \pm 2 \cdot \sqrt{\frac{0.16(1-0.16)}{400}} \approx 0.16 \pm 0.037 = (0.129,\ 0.197)$$

We can therefore say that the data provide evidence that the proportion of defective products has been reduced, and we are 95% sure that it has been reduced to somewhere between 12.9% and 19.7%. This is very useful information, since it tells us that even though the results were significant (i.e., the repair reduced the number of defective products), the repair might not have been effective enough, if it managed to reduce the number of defective products only to the range provided by the confidence interval. This, of course, ties back in to the idea of statistical significance vs. practical importance that we discussed earlier. Even though the results are significant (H_o was rejected), practically speaking, the repair might be considered ineffective.

Scenario: Safety of Airplane Drinking Water

The purpose of this activity is to give you hands-on practice in following up a test for the population proportion p in which H_o has been rejected with a confidence interval, and getting a sense of how the confidence interval is a natural and informative supplement to the test in these cases.

Background:

Recall from a previous activity the results of a study on the safety of airplane drinking water that was conducted by the U.S. Environmental Protection Agency (EPA). A study found that out of a random sample of 316 airplanes tested, 40 had coliform bacteria in the drinking water drawn from restrooms and kitchens. As a benchmark comparison, in 2003 the EPA found that about 3.5% of the U.S. population have coliform bacteria-infected drinking water. The question of interest is whether, based on the results of this study, we can conclude that drinking water on airplanes is more contaminated than drinking water in general. Let p be the proportion of contaminated drinking water on airplanes.

In a previous activity we tested H_o: $p = 0.035$ vs. H_a: $p > 0.035$ and found that the data provided extremely strong evidence to reject H_o and conclude that the proportion of contaminated drinking water in airplanes is larger than the proportion of contaminated drinking water in general (which is 0.035).

Now that we've concluded that, all we know about p is that we have very strong evidence that it is higher than 0.035. However, we have no sense of its magnitude. It will make sense to follow up the test by estimating p with a 95% confidence interval.

Learn By Doing (1/1 point)

Based on the data, find a 95% confidence interval for p and interpret it in context. Recall that the formula for that is:

Your Answer:

p^ +- 2 * sqrt(p*(1-p)/n) = 0.1266 +- 2 * sqrt(0.1266*0.8734 / 316) = 0.1266 +- 0.0374

Our Answer:

Since the sample proportion of contaminated drinking water is the 95% confidence interval for p is This means that, based on the data, we are 95% confident that the proportion of contaminated drinking water on airplanes is between 9% and 16.4%.

Resubmit Reset

Let's Summarize

Even though this unit is about the z-test for population proportion, it is loaded with very important ideas that apply to hypothesis testing in general. We've already summarized the details that are specific to the z-test for proportions, so the purpose of this summary is to highlight the general ideas.

The process of hypothesis testing has four steps:

I. Stating the null and alternative hypotheses (H_o and H_a).

II. Obtaining a random sample (or at least one that can be considered random) and collecting data. Using the data:

* **Check that the conditions** under which the test can be reliably used are met.

* **Summarize the data using a test statistic.**

The test statistic is a measure of the evidence in the data against H_o. The larger the test statistic is in magnitude, the more evidence the data present against H_o.

III. Finding the p-value of the test.

The p-value is the probability of getting data like those observed (or even more extreme) assuming that the null hypothesis is true, and is calculated using the null distribution of the test statistic. The p-value is a measure of the evidence against H_o. The smaller the p-value, the more evidence the data present against H_a.

IV. Making conclusions.

- Conclusions about the **significance of the results:**

If the p-value is small, the data present enough evidence to reject H_o (and accept H_a).

If the p-value is not small, the data do not provide enough evidence to reject H_o.

To help guide our decision, we use the significance level as a cutoff for what is considered a small p-value. The significance cutoff is usually set at 0.05, but should not be considered inviolable.

- Conclusions **in the context** of the problem.

Results that are based on a larger sample carry more weight, and therefore **as the sample size increases, results become more significant.**

Even a very small and practically unimportant effect becomes statistically significant with a large enough sample size. The **distinction between statistical significance and practical importance** should therefore always be considered.

For given data, the **p-value of the two-sided test is always twice as large as the p-value of the one-sided test**. It is therefore harder to reject H_o in the two-sided case than it is in the one-sided case in the sense that stronger evidence is required. Intuitively, the hunch or information that leads us to use the one-sided test can be regarded as a head-start toward the goal of rejecting H_o.

Confidence intervals can be used in order to carry out two-sided tests (at the 0.05 significance level). If the null value is not included in the confidence interval (i.e., is not one of the plausible values for the parameter), we have enough evidence to reject H_o. Otherwise, we cannot reject H_o.

If the results are significant, it might be of interest to **follow up the tests with a confidence interval** in order to get insight into the actual value of the parameter of interest.
