Hypothesis Testing for the Population Proportion p: Drawing Conclusions

Learning Objective: Carry out hypothesis testing for the population proportion and mean (when appropriate), and draw conclusions in context.

4. Drawing Conclusions Based on the P-Value

This last part of the four-step process of hypothesis testing is the same across all statistical tests, and actually, we've already said basically everything there is to say about it, but it can't hurt to say it again.

The p-value is a measure of how much evidence the data present against H_o. The smaller the p-value, the more evidence the data present against H_o.

We already mentioned that what determines what constitutes enough evidence against H_o is the **significance level** (α), a cutoff point below which the p-value is considered small enough to reject H_o in favor of H_a. The most commonly used significance level is 0.05.

It is important to mention again that this step has essentially two sub-steps:

(i) Based on the p-value, determine whether or not the results are significant (i.e., the data present enough evidence to reject H_o).

(ii) State your conclusions in the context of the problem.

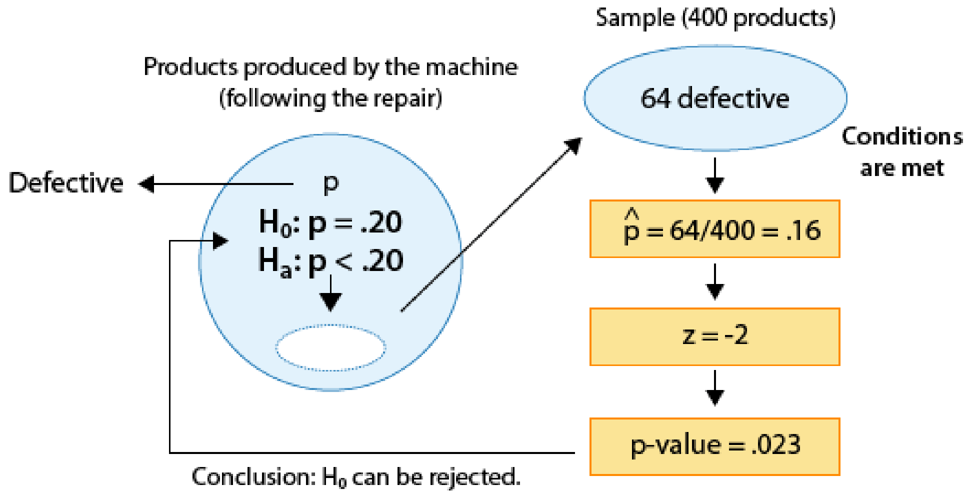
Let's go back to our three examples and draw conclusions.

Example: 1

(Has the proportion of defective products been reduced from 0.20 as a result of the repair?)

We found that the p-value for this test was 0.023.

Since 0.023 is small (in particular, 0.023 < 0.05), the data provide enough evidence to reject H_o and conclude that as a result of the repair the proportion of defective products has been reduced to below 0.20. The following figure is the complete story of this example, and includes all the steps we went through, starting from stating the hypotheses and ending with our conclusions:



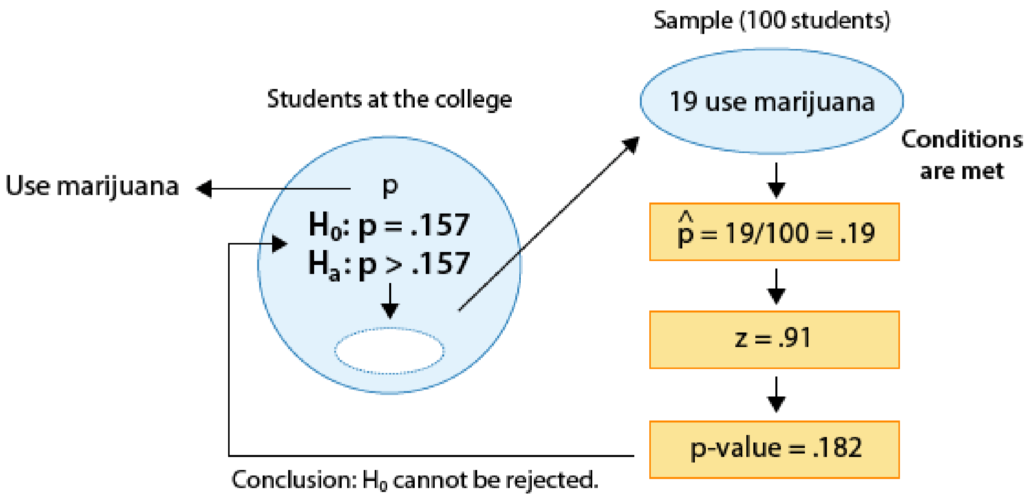
Example: 2

(Is the proportion of students who use marijuana at the college higher than the national proportion, which is 0.157?)

We found that the p-value for this test was 0.182.

Since 0.182 is **not** small (in particular, 0.182 > 0.05), the data do not provide enough evidence to reject H_o.

We therefore do **not** have enough evidence to conclude that the proportion of students at the college who use marijuana is higher than the national figure. Here is the complete story of this example:

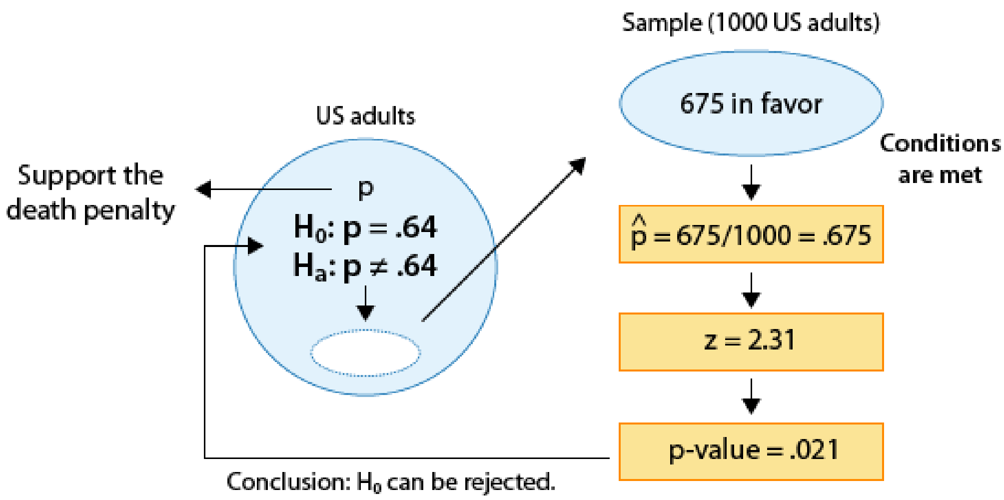


Example: 3

(Has the proportion of U.S. adults who support the death penalty for convicted murderers changed since 2003, when it was 0.64?)

We found that the p-value for this test was 0.021.

Since 0.021 is small (in particular, $0.021 < 0.05$), the data provide enough evidence to reject H_o, and we conclude that the proportion of adults who support the death penalty for convicted murderers has changed since 2003. Here is the complete story of this example:



Did I Get This

1/1 point (graded)

Two hypothesis tests were conducted.

In test I, a significance level of 0.05 was used, and the p-value was calculated to be 0.025.

In test II, a significance level of 0.01 was used, and the p-value was calculated to be 0.025.

Which of the following is true?

- In both test I and test II, H_o is rejected.
- In both test I and test II, H_o is not rejected.
- In test I, H_o is rejected, and in test II, H_o is not rejected. ✔
- In test I, H_o is not rejected, and in test II, H_o is rejected.

Answer
Correct:
Indeed, in test I, the p-value is below the significance level (0.025 less than 0.05) and therefore considered small enough to reject H_o. In test II, on the other hand, where the significance level is set at 0.01, the p-value is not small enough (0.025 greater than 0.01) to reject H_o.

Submit

Did I Get This

1/1 point (graded)
A test has a p-value of 0.0436. Which of the following is true? Check all that apply.

- [ ] For any level of significance that is below 0.0436, H_o is rejected.
- [x] For any level of significance that is below 0.0436, H_o is not rejected.
- [x] For any level of significance that is above 0.0436, H_o is rejected.
- [ ] For any level of significance that is above 0.0436, H_o is not rejected.

✔

Answer
Correct:
For any level of significance that is below 0.0436, H_o is not rejected and for any level of significance that is above 0.0436, H_o is rejected are both true statements.

Submit

Many Students Wonder ...

Question: Why is 5% is often selected as the significance level in hypothesis testing, and why 1% is the next most typical level.

Answer:
This is largely due to just convenience and tradition.

When Ronald Fisher (one of the founders of modern statistics) published one of his tables, he used a mathematically convenient scale that included 5% and 1%. Later, these same 5% and 1% levels were used by other people, in part just because Fisher was so highly esteemed. But mostly these are arbitrary levels.

The idea of selecting some sort of relatively small cutoff was historically important in the development of statistics; but it's important to remember that there is really a continuous range of increasing confidence towards the alternative hypothesis, not a single all-or-nothing value. There isn't much meaningful difference, for instance, between a p-value of 0.049 or 0.051, and it would be foolish to declare one case definitely a "real" effect and to declare the other case definitely a "random" effect. In either case, the study results were roughly 5% likely by chance if there's no actual effect.

Whether such a p-value is sufficient for us to reject a particular null hypothesis ultimately depends on the risk of making the wrong decision, and the extent to which the hypothesized effect might contradict our prior experience or previous studies.