Hypothesis Testing for the Population Proportion p: Effect of Sample Size

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

More About Hypothesis Testing

The issues regarding hypothesis testing that we will discuss are:

1. The effect of sample size on hypothesis testing.

2. Statistical significance vs. practical importance. (This will be discussed in the activity following number 1.)

3. One-sided alternative vs. two-sided alternative—understanding what is going on.

4. Hypothesis testing and confidence intervals—how are they related?

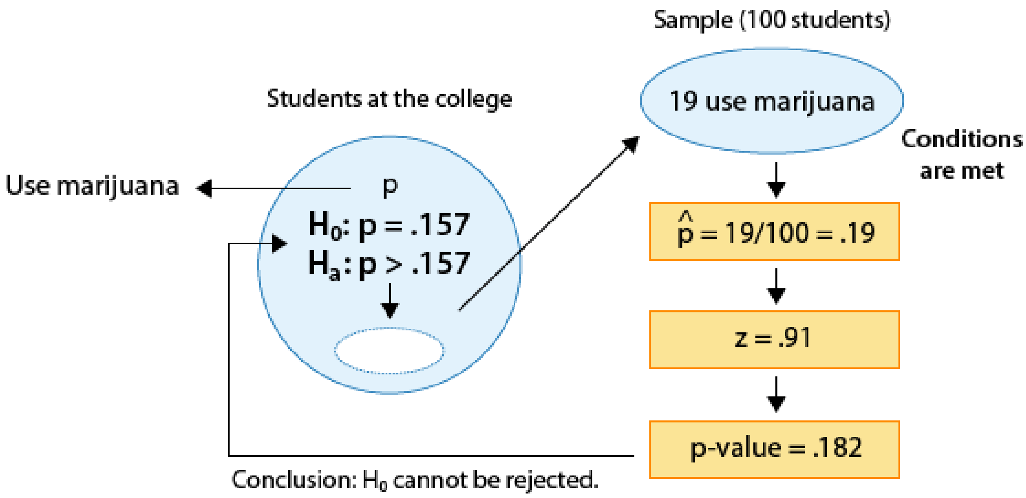
Let's start.

1. The Effect of Sample Size on Hypothesis Testing

We have already seen the effect that the sample size has on inference, when we discussed point and interval estimation for the population mean (μ) and population proportion (p). Intuitively ...

Larger sample sizes give us more information to pin down the true nature of the population. We can therefore expect the **sample** mean and **sample** proportion obtained from a larger sample to be closer to the population mean and proportion, respectively. As a result, for the same level of confidence, we can report a smaller margin of error, and get a narrower confidence interval. What we've seen, then, is that larger sample size gives a boost to how much we trust our sample results. In hypothesis testing, larger sample sizes have a similar effect. The following two examples will illustrate that a larger sample size provides more convincing evidence, and how the evidence manifests itself in hypothesis testing. Let's go back to our example 2 (marijuana use at a certain liberal arts college).

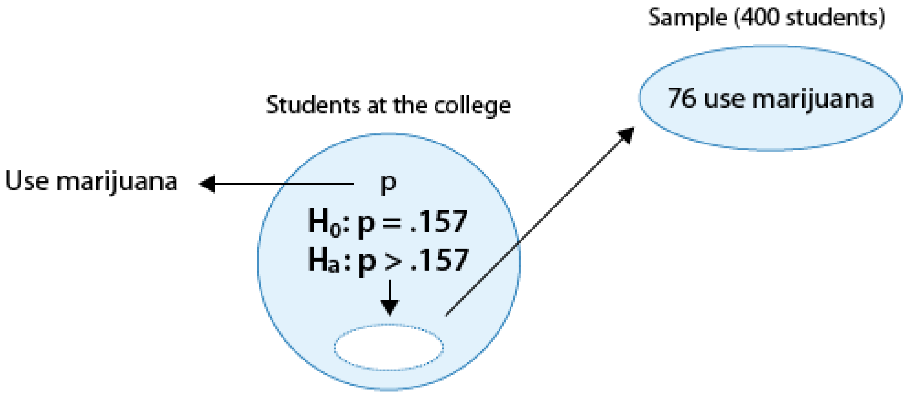
Example: 2



The data **do not** provide enough evidence that the proportion of marijuana users at the college is higher than the proportion among all U.S. college students, which is 0.157. So far, nothing new. Let's make small changes to the problem (and call it example 2*). The changes are highlighted and the problem is followed by a new figure that reflects the changes.

Example: 2*

There are rumors that students in 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 400 students from the college, 76 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? (reported by the Harvard School of Public Health).



We now have a larger sample (400 instead of 100), and also we changed the number of marijuana users (76 instead of 19).

Let's carry out the test in this case.

I. The question of interest did not change, so we are testing the same hypotheses:

H_o: p = 0.157

H_a: p > 0.157

II. We select a random sample of size **400** and find that 76 are marijuana users.

(Note that the data satisfy the conditions that allow us to use this test. Verify this yourself).

Let's summarize the data:

$$* \ \hat{p} = \frac{76}{400} = .19$$

This is the same sample proportion as in the original problem, so it seems that the data give us the same evidence, but when we calculate the test statistic, we see that actually this is not the case:

$$* \ z = \frac{.19 - .157}{\sqrt{\frac{.157(1-.157)}{400}}} \approx 1.81$$

Even though the sample proportion is the same (0.19), since here it is based on a larger sample (400 instead of 100), it is 1.81 standard deviations above the null value of 0.157 (as opposed to 0.91 standard deviations in the original problem).

III. For the p-value, we use statistical software to find p-value = 0.035.

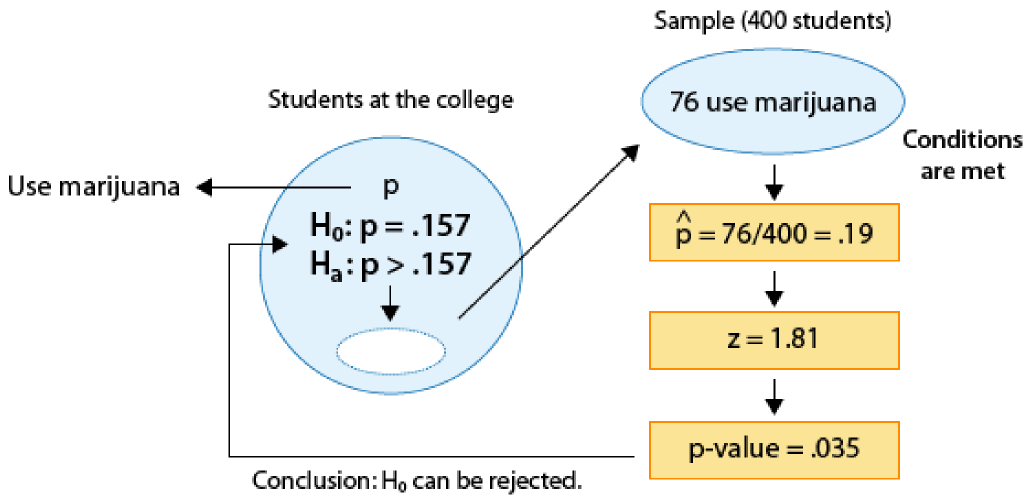
The p-value here is 0.035 (as opposed to 0.182 in the original problem). In other words, when H_o is true (i.e. when p = 0.157) it is quite unlikely (probability of 0.035) to get a sample proportion of 0.19 or higher based on a sample of size 400 (probability 0.035), and not very unlikely when the sample size is

100 (probability 0.182).

IV.

Our results here are significant. In other words, in example 2* the data provide enough evidence to reject H_o and conclude that the proportion of marijuana users at the college is higher than among all U.S. students.

Let's summarize with a figure:



What do we learn from these two examples?

We see that sample results that are based on a larger sample carry more weight.

In example 2, we saw that a sample proportion of 0.19 based on a sample of size of 100 was not enough evidence that the proportion of marijuana users in the college is higher than 0.157. Recall, from our general overview of hypothesis testing, that this conclusion (not having enough evidence to reject the null hypothesis) *doesn't* mean the null hypothesis is necessarily true (so, we never "accept" the null); it only means that the particular study didn't yield sufficient evidence to reject the null. It *might* be that the sample size was simply too small to detect a statistically significant difference.

However, in example 2*, we saw that when the sample proportion of 0.19 is obtained from a sample of size 400, it carries much more weight, and in particular, provides enough evidence that the proportion of marijuana users in the college is higher than 0.157 (the national figure). In *this* case, the sample size of 400 *was* large enough to detect a statistically significant difference.

The following activity will allow you to practice the ideas and terminology used in hypothesis testing when a result is not statistically significant.

Scenario: Support of U.S. Administration's Policies

Suppose that only 40% of the U.S. public supported the general direction of the previous U.S. administration's policies. To gauge whether the nationwide proportion, p, of support for the *current* administration is higher than 40%, a major polling organization conducts a random poll to test the hypotheses:

$H_o: p = 0.40$

$H_a: p > 0.40$

The results are reported to be **not statistically significant**, with a **p-value of 0.214**.

Learn By Doing

1/1 point (graded)

Based on the study, is the following statement is a valid conclusion or an invalid conclusion?

The results **do not reject** the null hypothesis:

- (selected) valid ✔
- invalid

Answer

Correct:

When the results are "not statistically significant," it means we don't have sufficient evidence to reject the null hypothesis, because any difference in the study was probably just by chance. In fact, the p-value is the probability of that chance difference, and a p-value of 0.214 is considerably larger than the traditional 0.05 level below which we would begin to reject the null hypothesis.

Submit

Learn By Doing

1/1 point (graded)

Based on the study, is the following statement is a valid conclusion or an invalid conclusion?

The results provide enough evidence to **accept** the null hypothesis:

- valid

invalid ✔

Answer
Correct:
We never "accept" the null hypothesis. In the case of the study described, the lack of statistical significance means there wasn’t sufficiently surprising evidence against the null, so the null hypothesis should merely remain the default conclusion until more evidence is gathered.

Submit

Learn By Doing

1/1 point (graded)
Based on the study, is the following statement is a valid conclusion or an invalid conclusion?

The results indicate that it **must** be true that the nationwide proportion, p, of support for the current administration **is equal to 40%**:

valid

invalid ✔

Answer
Correct:
We never "prove" the null hypothesis. In the case of the study described, the lack of statistical significance in the study indicates we don’t have sufficient evidence to reject the null hypothesis, but that doesn’t mean the null is true; the alternative hypothesis might actually be true, and it might simply be that the study wasn’t sensitive enough to detect a real difference.

Submit

Learn By Doing

1/1 point (graded)
Based on the study, is the following statement is a valid conclusion or an invalid conclusion?

The results indicate that the nationwide proportion, p, of support for the current administration might be greater than 40%, but the **sample size** in the study **might have been too small to detect a statistically significant difference**:

valid ✔

invalid

Answer

Correct:

Even though the null hypothesis can't be rejected by this study, the alternative hypothesis might actually be true; the study might simply not have been sensitive enough to detect the difference.

Submit