

8.15: Hypothesis Test for a Population Proportion (3 of 3)

Learning Objectives

- Conduct a hypothesis test for a population proportion. State a conclusion in context.
- Interpret the P-value as a conditional probability in the context of a hypothesis test about a population proportion.
- Distinguish statistical significance from practical importance.
- From a description of a study, evaluate whether the conclusion of a hypothesis test is reasonable.

More about the P-Value

The P-value is a probability that describes the likelihood of the data if the null hypothesis is true. More specifically, the P-value is the probability that sample results are as extreme as or more extreme than the data if the null hypothesis is true. The phrase “as extreme as or more extreme than” means farther from the center of the sampling distribution in the direction of the alternative hypothesis.

More generally, we view the P-value a description of the strength of the evidence against the null hypothesis and in support of the alternative hypothesis. But the P-value is a probability about sample results, not about the null or alternative hypothesis.

One More Note about P-Values and the Significance Level

You may wonder why 5% is often selected as the significance level in hypothesis testing and why 1% is also a commonly used level. It 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 toward the alternative hypothesis, not a single all-or-nothing value. There isn't much meaningful difference, for instance, between the P-values 0.049 and 0.051, and it would be foolish to declare one case definitely a “real” effect and the other case definitely a “random” effect. In either case, the study results are 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.

Example

Sample Size and Hypothesis Testing

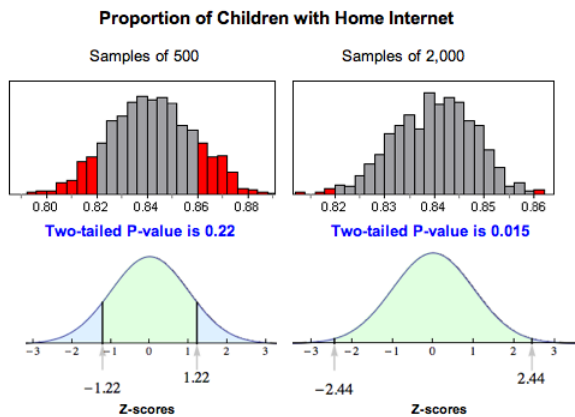
Consider our earlier example about teenagers and Internet access. According to the Kaiser Family Foundation, 84% of U.S. children ages 8 to 18 had Internet access at home as of August 2009. Researchers wonder if this number has changed since then. The hypotheses we tested were:

$$H_0: p = 0.84$$

$$H_a: p \neq 0.84$$

The original sample consisted of 500 children, and 86% of them had Internet access at home. The P-value was about 0.22, which was not strong enough to reject the null hypothesis. There was not enough evidence to show that the proportion of all U.S. children ages 8 to 18 have Internet access at home.

Suppose we sampled 2,000 children and the sample proportion was still 86%. Our test statistic would be $Z \approx 2.44$, and our P-value would be about 0.015. The larger sample size would allow us to reject the null hypothesis even though the sample proportion was the same.



Why does this happen? Larger samples vary less, so a sample proportion of 0.86 is more unusual with larger samples than with smaller samples if the population proportion is really 0.84. This means that if the alternative hypothesis is true, a larger sample size will make it more likely that we reject the null. Therefore, we generally prefer a larger sample as we have seen previously.

Drawing Conclusions from Hypothesis Tests

It is tempting to get involved in the details of a hypothesis test without thinking about how the data was collected. Whether we are calculating a confidence interval or performing a hypothesis test, the results are meaningless without a properly designed study. Consider the following exercises about how data collection can affect the results of a study.

Try It

<https://assessments.lumenlearning.co...essments/3964>

Let's Summarize

In this section, we looked at the four steps of a hypothesis test as they relate to a claim about a population proportion.

Step 1: Determine the hypotheses.

- The hypotheses are claims about the population proportion, p .
- The null hypothesis is a hypothesis that the proportion equals a specific value, p_0 .
- The alternative hypothesis is the competing claim that the parameter is less than, greater than, or not equal to p_0 .

Step 2: Collect the data.

Since the hypothesis test is based on probability, random selection or assignment is essential in data production. Additionally, we need to check whether the sample proportion can be $np \geq 10$ and $n(1 - p) \geq 10$.

Step 3: Assess the evidence.

- Determine the test statistic which is the z-score for the sample proportion. The formula is:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$
- Use the test statistic, together with the alternative hypothesis to determine the P-value. You can use a standard normal table (or Z-table) or technology (such as the simulations on the second page of this topic) to find the P-value.
- If the alternative hypothesis is greater than, the P-value is the area to the right of the test statistic. If the alternative hypothesis is less than, the P-value is the area to the left of the test statistic. If the alternative hypothesis is not equal to, the P-value is equal to double the tail area beyond the test statistic.

Step 4: Give the conclusion.

- A small P-value says the data is unlikely to occur if the null is true. If the P-value is less than or equal to the significance level, we reject the null hypothesis and accept the alternative hypothesis instead.
- If the P-value is greater than the significance level, we say we “fail to reject” the null hypothesis. We never say that we “accept” the null hypothesis. We just say that we don’t have enough evidence to reject it. This is equivalent to saying we don’t have enough evidence to support the alternative hypothesis.

- We write the conclusion in the context of the research question. Our conclusion is usually a statement about the alternative hypothesis (we accept H_a or fail to accept H_a) and should include the P-value.

Other Hypothesis Testing Notes

Remember that the P-value is the probability of seeing a sample proportion as extreme as the one observed from the data if the null hypothesis is true. The probability is about the random sample, not about the null or alternative hypothesis.

A larger sample size makes it more likely that we will reject the null hypothesis if the alternative is true. Another way of thinking about this is that increasing the sample size will decrease the likelihood of a type II error. Recall that a type II error is failing to reject the null hypothesis when the alternative is true.

Increasing the sample size can have the unintended effect of making the test sensitive to differences so small they don't matter. A statistically significant difference is one large enough that it is unlikely to be due to sampling variability alone. Even a difference so small that it is not important can be statistically significant if the sample size is big enough.

Finally, remember the phrase "garbage in, garbage out." If the data collection methods are poor, then the results of a hypothesis test are meaningless. No statistical methods can create useful information if our data comes from convenience or voluntary response samples. Additionally, the results of a hypothesis test apply only to the population from whom the sample was chosen.

Contributors and Attributions

CC licensed content, Shared previously

- Concepts in Statistics. **Provided by:** Open Learning Initiative. **Located at:** <http://oli.cmu.edu>. **License:** [CC BY: Attribution](#)

This page titled [8.15: Hypothesis Test for a Population Proportion \(3 of 3\)](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Lumen Learning](#).