

8.8: Hypothesis Testing (2 of 5)

Learning Objectives

- Recognize the logic behind a hypothesis test and how it relates to the P-value.

In this section, our focus is hypothesis testing, which is part of inference. On the previous page, we practiced stating null and alternative hypotheses from a research question. Forming the hypotheses is the first step in a hypothesis test. Here are the general steps in the process of hypothesis testing. We will see that hypothesis testing is related to the thinking we did in *Linking Probability to Statistical Inference*.

Step 1: Determine the hypotheses.

The hypotheses come from the research question.

Step 2: Collect the data.

Ideally, we select a random sample from the population. The data comes from this sample. We calculate a statistic (a mean or a proportion) to summarize the data.

Step 3: Assess the evidence.

Assume that the null hypothesis is true. Could the data come from the population described by the null hypothesis? Use simulation or a mathematical model to examine the results from random samples selected from the population described by the null hypothesis. Figure out if results similar to the data are likely or unlikely. Note that the wording “likely or unlikely” implies that this step requires some kind of probability calculation.

Step 4: State a conclusion.

We use what we find in the previous step to make a decision. This step requires us to think in the following way. *Remember that we assume that the null hypothesis is true.* Then one of two outcomes can occur:

- One possibility is that results similar to the actual sample are extremely unlikely. This means that the data do not fit in with results from random samples selected from the population described by the null hypothesis. In this case, it is unlikely that the data came from this population, so we view this as strong evidence against the null hypothesis. We reject the null hypothesis in favor of the alternative hypothesis.
- The other possibility is that results similar to the actual sample are fairly likely (not unusual). This means that the data fit in with typical results from random samples selected from the population described by the null hypothesis. In this case, we do not have evidence against the null hypothesis, so we cannot reject it in favor of the alternative hypothesis.

Example

Data Use on Smart Phones

According to an article by Andrew Berg (“Report: Teens Texting More, Using More Data,” *Wireless Week*, October 15, 2010), Nielsen Company analyzed cell phone usage for different age groups using cell phone bills and surveys. Nielsen found significant growth in data usage, particularly among teens, stating that “94 percent of teen subscribers self-identify as advanced data users, turning to their cellphones for messaging, Internet, multimedia, gaming, and other activities like downloads.” The study found that the mean cell phone data usage was 62 MB among teens ages 13 to 17. A researcher is curious whether cell phone data usage has increased for this age group since the original study was conducted. She plans to conduct a hypothesis test.

Step 1: Determine the hypotheses.

The null hypothesis is often a statement of “no change,” so the null hypothesis will state that there is no change in the mean cell phone data usage for this age group since the original study. In this case, the alternative hypothesis is that the mean has increased from 62 MB.

H_0 : The mean data usage for teens with smart phones is still 62 MB.

H_a : The mean data usage for teens with smart phones is greater than 62 MB.

Step 2: Collect the data.

The next step is to obtain a sample and collect data that will allow the researcher to test the hypotheses. The sample must be representative of the population and, ideally, should be a random sample. In this case, the researcher must randomly sample teens who use smart phones.

For the purposes of this example, imagine that the researcher randomly samples 50 teens who use smart phones. She finds that the mean data usage for these teens was 75 MB with a standard deviation of 45 MB. Since it is greater than 62 MB, this sample mean provides some evidence in favor of the alternative hypothesis. But the researcher anticipates that samples will vary when the null hypothesis is true. So how much of a difference will make her doubt the null hypothesis? Does she have evidence strong enough to reject the null hypothesis?

Step 3: Assess the evidence.

To assess the evidence, the researcher needs to know how much variability to expect in random samples when the null hypothesis is true. She begins with the assumption that H_0 is true – in this case, that the mean data usage for teens is still 62 MB. She then determines how unusual the results of the sample are: *If the mean for all teens with smart phones actually is 62 MB, what is the chance that a random sample of 50 teens will have a sample mean of 75 MB or higher?* Obviously, this probability depends on how much variability there is in random samples of this size from this population.

The probability of observing a sample mean at least this high if the population mean is 62 MB is approximately 0.023 (later topics explain how to calculate this probability). The probability is quite small. It tells the researcher that if the population mean is actually 62 MB, a sample mean of 75 MB or higher will occur only about 2.3% of the time. This probability is called the **P-value**.

Note: The P-value is a conditional probability, discussed in the module *Relationships in Categorical Data with Intro to Probability*. The condition is the assumption that the null hypothesis is true.

Step 4: Conclusion.

The small P-value indicates that it is unlikely for a sample mean to be 75 MB or higher if the population has a mean of 62 MB. It is therefore unlikely that the data from these 50 teens came from a population with a mean of 62 MB. The evidence is strong enough to make the researcher doubt the null hypothesis, so she rejects the null hypothesis in favor of the alternative hypothesis. *The researcher concludes that the mean data usage for teens with smart phones has increased since the original study. It is now greater than 62 MB. ($P = 0.023$)*

Comment

Notice that the P-value is included in the preceding conclusion, which is a common practice. It allows the reader to see the strength of the evidence used to draw the conclusion.

How Small Does the P-Value Have to Be to Reject the Null Hypothesis?

A small P-value indicates that it is unlikely that the actual sample data came from the population described by the null hypothesis. More specifically, a small P-value says that there is only a small chance that we will randomly select a sample with results at least as extreme as the data if H_0 is true. The smaller the P-value, the stronger the evidence against H_0 .

But how small does the P-value have to be in order to reject H_0 ?

In practice, we often compare the P-value to 0.05. We reject the null hypothesis in favor of the alternative if the P-value is less than (or equal to) 0.05.

Note: This means that sampling variability will produce results at least as extreme as the data 5% of the time. In other words, in the long run, 1 in 20 random samples will have results that suggest we should reject H_0 even when H_0 is true. This variability is just due to chance, but it is unusual enough that we are willing to say that results this rare suggest that H_0 is not true.

Statistical Significance: Another Way to Describe Unlikely Results

When the P-value is less than (or equal to) 0.05, we also say that the difference between the actual sample statistic and the assumed parameter value is **statistically significant**. In the previous example, the P-value is less than 0.05, so we say the difference between the sample mean (75 MB) and the assumed mean from the null hypothesis (62 MB) is statistically significant. You will also see this described as a **significant difference**. A significant difference is an observed difference that is too large to attribute to chance. In other words, it is a difference that is unlikely when we consider sampling variability alone. If the difference is statistically significant, we reject H_0 .

Other Observations about Stating Conclusions in a Hypothesis Test

In the example, the sample mean was greater than 62 MB. This fact alone does not suggest that the data supports the alternative hypothesis. We have to determine that the data is not only larger than 62 MB but larger than we would expect to see in a random sampling if the population mean is 62 MB. We therefore need to determine the P-value. If the sample mean was less than or equal to 62 MB, it would not support the alternative hypothesis. We don't need to find a P-value in this case. The conclusion is clear without it.

We have to be very careful in how we state the conclusion. There are only two possibilities.

- We have enough evidence to reject the null hypothesis and support the alternative hypothesis.
- We do not have enough evidence to reject the null hypothesis, so there is not enough evidence to support the alternative hypothesis.

If the P-value in the previous example was greater than 0.05, then we would not have enough evidence to reject H_0 and accept H_a . In this case our conclusion would be that “there is not enough evidence to show that the mean amount of data used by teens with smart phones has increased.” Notice that this conclusion answers the original research question. It focuses on the alternative hypothesis. It does *not* say “the null hypothesis is true.” We never accept the null hypothesis or state that it is true. When there is not enough evidence to reject H_0 , the conclusion will say, in essence, that “there is not enough evidence to support H_a .” But of course we will state the conclusion in the specific context of the situation we are investigating.

We compared the P-value to 0.05 in the previous example. The number 0.05 is called the **significance level** for the test, because a P-value less than or equal to 0.05 is statistically significant (unlikely to have occurred solely by chance). The symbol we use for the significance level is α (the lowercase Greek letter alpha). We sometimes refer to the significance level as the α -level. We call this value the significance level because if the P-value is less than the significance level, we say the results of the test showed a significance difference.

If the P-value $\leq \alpha$, we reject the null hypothesis in favor of the alternative hypothesis.

If the P-value $> \alpha$, we fail to reject the null hypothesis.

In practice, it is common to see 0.05 for the significance level. Occasionally, researchers use other significance levels. In particular, if rejecting H_0 will be controversial or expensive, we may require stronger evidence. In this case, a smaller significance level, such as 0.01, is used. As with the hypotheses, we should choose the significance level before collecting data. It is treated as an agreed-upon benchmark prior to conducting the hypothesis test. In this way, we can avoid arguments about the strength of the data.

We look more at how to choose the significance level later. On this page we continue to use a significance level of 0.05.

First let's look at some exercises that focus on the P-value and its meaning. Then we'll try some that cover the conclusion.

Learn By Doing

For many years, working full-time has meant working 40 hours per week. Nowadays, it seems that corporate employers expect their employees to work more than this amount. A researcher decides to investigate this hypothesis.

H_0 : The average time full-time corporate employees work per week is 40 hours.

H_a : The average time full-time corporate employees work per week is more than 40 hours.

To substantiate his claim, the researcher randomly selects 250 corporate employees and finds that they work an average of 47 hours per week with a standard deviation of 3.2 hours.

<https://assessments.lumenlearning.co...sessments/3600>

According to the Centers for Disease Control (CDC), roughly 21.5% of all high school seniors in the United States have used marijuana. (The data were collected in 2002. The figure represents those who smoked during the month prior to the survey, so the actual figure might be higher.) A sociologist suspects that the rate among African American high school seniors is lower. In this case, then,

H_0 : The rate of African American high-school seniors who have used marijuana is 21.5% (same as the overall rate of seniors).

H_a : The rate of African American high-school seniors who have used marijuana is lower than 21.5%.

To check his claim, the sociologist chooses a random sample of 375 African American high school seniors, and finds that 16.5% of them have used marijuana.

<https://assessments.lumenlearning.co...sessments/3601>

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 [8.8: Hypothesis Testing \(2 of 5\)](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Bill Pelz](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.