Hypothesis Testing: Terminology

Learning Objective: Explain the logic behind and the process of hypotheses testing. In particular, explain what the p-value is and how it is used to draw conclusions.

More Details and Terminology

Now that we understand the general idea of how statistical hypothesis testing works, let's go back to each of the steps and delve slightly deeper, getting more details and learning some terminology.

Hypothesis testing step 1: Stating the claims.

In all three examples, our aim is to decide between two opposing points of view, Claim 1 and Claim 2. In hypothesis testing, **Claim 1** is called the **null hypothesis** (denoted "$\mathbf{H_0}$"), and **Claim 2** plays the role of the **alternative hypothesis** (denoted "$\mathbf{H_a}$"). As we saw in the three examples, the null hypothesis suggests nothing special is going on; in other words, there is no change from the status quo, no difference from the traditional state of affairs, no relationship. In contrast, the alternative hypothesis disagrees with this, stating that something is going on, or there is a change from the status quo, or there is a difference from the traditional state of affairs. The alternative hypothesis, H_a, usually represents what we want to check or what we suspect is really going on.

Let's go back to our three examples and apply the new notation:

In example 1:

- **H_0:** The proportion of smokers at Goodheart is 0.20.
- **H_a:** The proportion of smokers at Goodheart is less than 0.20.

In example 2:

- **H_0:** The mean concentration in the shipment is the required 245 ppm.
- **H_a:** The mean concentration in the shipment is not the required 245 ppm.

In example 3:

- **H_0:** Performance on the SAT is not related to gender (males and females score the same).
- **H_a:** Performance on the SAT is related to gender - males score higher.

Scenario: Smoking and Bachelor's Degree

According to the Centers for Disease Control and Prevention, the proportion of U.S. adults age 25 or older who smoke is 0.22. A researcher suspects that the rate is lower among U.S. adults 25 or older who have a bachelor's degree or higher education level.

Learn By Doing

1/1 point (graded)
What is the null hypothesis in this case?

- The proportion of smokers among U.S. adults 25 or older who have a bachelor's degree or higher is less than 0.22.
- **The proportion of smokers among U.S. adults 25 or older who have a bachelor's degree or higher is 0.22.** ✔
- The proportion of smokers among U.S. adults 25 or older who have a bachelor's degree or higher is not 0.22.
- There is no relationship between education level and smoking habits.

Answer
Correct:
Indeed, the null hypothesis claims that "nothing special is going on" or "there is no change from the status quo," which, in this context, means that there is nothing special about the smoking habits of adults age 25 and older who have a bachelor's degree or higher. They are the same as in the entire population of adults (25+).

Submit

Learn By Doing

1/1 point (graded)

What is the alternative hypothesis in this case?

- (●) The proportion of smokers among U.S. adults 25 or older who have a bachelor's degree or higher is less than 0.22. ✔
- ( ) The proportion of smokers among U.S. adults 25 or older who have a bachelor's degree or higher is 0.22.
- ( ) The proportion of smokers among U.S. adults 25 or older who have a bachelor's degree or higher is not 0.22.
- ( ) There is a relationship between level of education level and smoking habits.

Answer

Correct:

Indeed, the alternative hypothesis usually represents what we want to check, or what we suspect is really going on. In this case, what the researcher suspects is going on is that the proportion of smokers among U.S. adults age 25 and older who have a bachelor's degree or higher is lower than the rate of all adults in this age group.

Submit

Scenario: IQ and Parents

A study investigated whether there are differences between the mean IQ level of people who were reared by their biological parents and those who were reared by someone else.

Learn By Doing

1/1 point (graded)

What is the null hypothesis in this case?

- The mean IQ level of people who were reared by their biological parents is higher than the mean of those who were not.
- The mean IQ level of people who were reared by their biological parents is lower than the mean of those who were not.
- There is no difference between the mean IQ levels of people who were reared by their natural parents and those who were not. ✔
- There is a difference between the mean IQ levels of people who were reared by their natural parents and those who were not.

Answer
Correct:
Indeed the null hypothesis claims that "nothing special is going on," which, in this context, means that there is no difference between the mean IQ levels of the two groups.

Submit

Learn By Doing

1/1 point (graded)
What is the alternative hypothesis in this case?

- The mean IQ level of people who were reared by their biological parents is higher than the mean of those who were not.
- The mean IQ level of people who were reared by their biological parents is lower than the mean of those who were not.
- There is no difference between the mean IQ level of people who were reared by their natural parents and those who were not.
- There is a difference between the mean IQ level of people who were reared by their natural parents and those who were not. ✔

Answer
Correct:
Indeed, the alternative hypothesis represents what we suspect or want to check. In this case, we want to check whether there is a difference between the mean IQ levels of the two groups.

Submit

Scenario: Smoking and Education Level

Data were collected in order to determine whether there is a relationship between a person's level of education and whether or not the person is a smoker.

Did I Get This

1/1 point (graded)
What is the null hypothesis in this case?

- The more educated a person is, the less likely he or she is to smoke.
- Whether a person is a smoker or not is not related to his or her level of education. ✔
- There is no difference between the smoking rates of those who have a bachelor's degree or higher and those who do not have a high school diploma.
- Whether a person is a smoker or not is related to his or her level of education.

Answer
Correct:
Indeed the null hypothesis claims that "nothing special is going on," which, in this context, means that there is no relationship between level of education and smoking habits.

Submit

Did I Get This

1/1 point (graded)
What is the alternative hypothesis in this case?

- The more educated a person is, the less likely he or she is to smoke.
- Whether a person is a smoker or not is not related to his or her level of education.

- ○ There is no difference between the smoking rates of those who have a bachelor's degree or higher and those who do not have a high school diploma.
- ◉ Whether a person is a smoker or not is related to his or her level of education. ✔

Answer
Correct:
Indeed, the alternative hypothesis represents what we suspect or want to check. In this case we want to check whether smoking habits are related to level of education.

Submit

Hypothesis testing step 2: Choosing a sample and collecting data.

This step is pretty obvious. This is what inference is all about. You look at sampled data in order to draw conclusions about the entire population. In the case of hypothesis testing, based on the data, you draw conclusions about whether or not there is enough evidence to reject H_o.

There is, however, one detail that we would like to add here. In this step we collect data and **summarize** it. Go back and look at the second step in our three examples. Note that in order to summarize the data we used simple sample statistics such as the sample proportion ($\hat{p}$), sample mean ($\overline{x}$) and the sample standard deviation (s).

In practice, you go a step further and use these sample statistics to summarize the data with what's called a **test statistic**. We are not going to go into any details right now, but we will discuss test statistics when we go through the specific tests.

Hypothesis testing step 3: Assessing the evidence.

As we saw, this is the step where we calculate how likely is it to get data like that observed when H_o true. In a sense, this is the heart of the process, since we draw our conclusions based on this probability. If this probability is very small (see example 2), then that means that it would be very surprising to get data like that observed if H_0 were true. The fact that we **did** observe such data is therefore evidence against H_0, and we should reject it. On the other hand, if this probability is not very small (see example 3) this means that observing data like that observed is not very surprising if H_0 were true, so the fact that we observed such data does not provide evidence against H_o. This crucial probability, therefore, has a special name. It is called the **p-value** of the test.

In our three examples, the p-values were given to you (and you were reassured that you didn't need to worry about how these were derived):

- Example 1: p-value = 0.106

- Example 2: p-value = 0.0007
- Example 3: p-value = 0.29

Obviously, the smaller the p-value, the more surprising it is to get data like ours when H_0 is true, and therefore, the stronger the evidence the data provide against H_0. Looking at the three p-values of our three examples, we see that the data that we observed in example 2 provide the strongest evidence against the null hypothesis, followed by example 1, while the data in example 3 provides the least evidence against H_0.

Comments:

Right now we will not go into specific details about p-value calculations, but just mention that since the p-value is the probability of getting **data** like those observed when H_0 is true, it would make sense that the calculation of the p-value will be based on the data summary, which, as we mentioned, is the test statistic. Indeed, this is the case. In practice, we will mostly use software to provide the p-value for us.

It should be noted that in the past, before statistical software was such an integral part of intro stats courses it was common to use critical values (rather than p-values) in order to assess the evidence provided by the data. While this courses focuses on p-values, we will provide some details about the critical values approach later in this module for those students who are interested in learning more about it.
