

 Lagunita is retiring and will shut down at 12 noon Pacific Time on March 31, 2020. A few courses may be open for self-enrollment for a limited time. We will continue to offer courses on other online learning platforms; visit <http://online.stanford.edu>.

Course > Inference: Hypothesis Testing Overview > Hypothesis Testing Overview >
Hypothesis Testing: General Idea and Logic

 Bookmark this page

Hypothesis Testing: General Idea and Logic

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.

The purpose of this section is to gradually build your understanding about how statistical hypothesis testing works. We start by explaining the general logic behind the process of hypothesis testing. Once we are confident that you understand this logic, we will add some more details and terminology.

General Idea and Logic of Hypothesis Testing

To start our discussion about the idea behind statistical hypothesis testing, consider the following example:

Example

A case of suspected cheating on an exam is brought in front of the disciplinary committee at a certain university.

There are **two** opposing **claims** in this case:

- The **student's claim**: I did not cheat on the exam.
- The **instructor's claim**: The student did cheat on the exam.

Adhering to the principle "**innocent until proven guilty**," the committee asks the instructor for **evidence** to support his claim. The instructor explains that the exam had two versions, and shows the committee members that on three separate exam questions, the student used in his solution numbers that were given in the other version of the exam.

The committee members all agree that **it would be extremely unlikely to get evidence like that if the student's claim of not cheating had been true**. In other words, the committee members all agree that the instructor brought forward strong enough evidence to reject the student's claim, and conclude that the student did cheat on the exam.

What does this example have to do with statistics?

While it is true that this story seems unrelated to statistics, it captures all the elements of hypothesis testing and the logic behind it. Before you read on to understand why, it would be useful to read the example again. Please do so now.

Statistical hypothesis testing is defined as:

Assessing evidence provided by the data in favor of or against some claim about the population.

Here is how the process of statistical hypothesis testing works:

1. We have **two claims** about what is going on in the population. Let's call them for now **claim 1** and **claim 2**. Much like the story above, where the student's claim is challenged by the instructor's claim, claim 1 is challenged by claim 2.
(**Comment:** as you'll see in the examples that follow, these claims are usually about the value of population parameter(s) or about the existence or nonexistence of a relationship between two variables in the population).
2. We choose a sample, collect relevant data and summarize them (this is similar to the instructor collecting evidence from the student's exam).
3. We figure out how likely it is to observe data like the data we got, had claim 1 been true. (Note that the wording "how likely ..." implies that this step requires some kind of probability calculation). In the story, the committee members assessed how likely it is to observe the evidence like that which the instructor provided, had the student's claim of not cheating been true.
4. Based on what we found in the previous step, we make our decision:
 - If we find that if claim 1 were true it would be extremely unlikely to observe the data that we observed, then we have strong evidence against claim 1, and we reject it in favor of claim 2.
 - If we find that if claim 1 were true observing the data that we observed is not very unlikely, then we do not have enough evidence against claim 1, and therefore we cannot reject it in favor of claim 2.

In our story, the committee decided that it would be extremely unlikely to find the evidence that the instructor provided had the student's claim of not cheating been true. In other words, the members felt that it is extremely unlikely that it is just a coincidence that the student used the numbers from the

other version of the exam on three separate problems. The committee members therefore decided to reject the student's claim and concluded that the student had, indeed, cheated on the exam. (Wouldn't you conclude the same?)

Hopefully this example helped you understand the logic behind hypothesis testing. To strengthen your understanding of the process of hypothesis testing and the logic behind it, let's look at three statistical examples.

Example: 1

A recent study estimated that 20% of all college students in the United States smoke. The head of Health Services at Goodheart University suspects that the proportion of smokers may be lower there. In hopes of confirming her claim, the head of Health Services chooses a random sample of 400 Goodheart students, and finds that 70 of them are smokers.

Let's analyze this example using the 4 steps outlined above:

1. Stating the claims:

There are two claims here:

- **claim 1:** The proportion of smokers at Goodheart is 0.20.
- **claim 2:** The proportion of smokers at Goodheart is less than 0.20.

Claim 1 basically says "nothing special goes on in Goodheart University; the proportion of smokers there is no different from the proportion in the entire country." This claim is challenged by the head of Health Services, who suspects that the proportion of smokers at Goodheart is lower.

2. Choosing a sample and collecting data:

A sample of $n = 400$ was chosen, and summarizing the data revealed that the sample proportion of smokers is $\hat{p} = \frac{70}{400} = 0.175$.

While it is true that 0.175 is less than 0.20, it is not clear whether this is strong enough evidence against claim 1.

3. Assessment of evidence:

In order to assess whether the data provide strong enough evidence against claim 1, we need to ask ourselves: How surprising is it to get a sample proportion as low as $\hat{p} = 0.175$ (or lower), assuming claim 1 is true?

In other words, we need to find how likely it is that in a random sample of size $n = 400$ taken from a population where the proportion of smokers is $p = 0.20$ we'll get a sample proportion as low as $\hat{p} = 0.175$ (or lower).

It turns out that the probability that we'll get a sample proportion as low as $\hat{p} = 0.175$ (or lower) in such a sample is roughly 0.106 (do not worry about how this was calculated at this point).

4. Conclusion:

Well, we found that if claim 1 were true there is a probability of 0.106 of observing data like that observed.

Now you have to decide ...

Do you think that a probability of 0.106 makes our data rare enough (surprising enough) under claim 1 so that the fact that we **did** observe it is enough evidence to reject claim 1?

Or do you feel that a probability of 0.106 means that data like we observed are not very likely when claim 1 is true, but they are not unlikely enough to conclude that getting such data is sufficient evidence to reject claim 1.

Basically, this is your decision. However, it would be nice to have some kind of guideline about what is generally considered surprising enough.



Open Learning Initiative [↗](#)
Unless otherwise noted this work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License [↗](#).

© All Rights Reserved