The Basics of Hypothesis Tests and Their Interpretations

Daniel Goldman

August 2018

1 Abstract

This paper summarizes the nature of hypothesis tests, as well as their interpretations, including the importance of understanding the underlying phenomenon being tested. Some formal logic and statistics is assumed.

2 Introduction

Hypothesis testing can seem complicated, and some methods are fairly sophisticated, but the basics of hypothesis testing is fairly straight forward. This discussion explains the nature of hypothesis testing, and seeks to improve understanding of correct and incorrect methodologies surrounding their usage.

Before starting on hypothesis testing, let us go back to something more fundamental in mathematics: proof by contradiction. In proof by contradiction, we have a candidate statement, and we want to show that the statement is false. For instance, we might start with the statement "there is a largest prime number p ." This statement is then assumed true and We seek to reach a conclusion that results in a contradiction, perhaps "there is a prime number $p^* > p$."

Because we generally do not allow for contradictions, we conclude that our initial assumption must be wrong, and that the statement is false. However, if we fail to reach a contradiction, we learn nothing. A contradiction must be reached in order for us to learn anything about the initial statement.

Hypothesis testing works in a very similar way. It just is not absolute. We assume that our null hypothesis is true. We then make a number of observations, and estimate the probability of seeing those observations, under the assumption that our null hypothesis is true. We also assume that observations we make are not unusual: if we observe something, we assume that it is not an extremely rare observation.

So if the probability of seeing our observed result is rare, based on the assumed hypothesis, either the event is a rare event, or the hypothesis is false. And so, hypothesis testing is just a form of proof by contradiction. It is for this reason that we never accept the null hypothesis. It is just our initial assumption, used in an attempt to produce a proof by contradiction.

3 Alternative Hypotheses

Acceptance of alternative hypothesis can actually be more problematic than understanding why we reject or fail to reject a null hypothesis, and it involves the need to have a full understanding of the hypothesis that we are testing. Suppose we assume that, within a given population of people, the average height of an individual is 176cm, and we survey a group of people and obtain the following results.

111	188	154	238
113	98	190	136
151	162	158	176
164	120	136	92
93	116	202	130

Since no information on the standard deviation of the population is given, we use a T-test and We run our hypothesis test as follows, using a significance level of 0.05:

$h_0 : \mu = 176$
 $n = 20$
 $df = 19$

The sample mean and standard deviation are 146.4 and 39.3264 respectively, giving a p-value of 0.003244, which means that for a significance level of 0.05, we would reject the null hypothesis. This means that we can accept the alternative hypothesis, the actual mean height is not equal to 176cm, right? Not necessarily.

First, recall that our actual assumption, in order to engage in proof by contradiction, is that our hypothesis is true and that our observations are not unusual. Our significance level defines what we mean by an unusual observation. We have found a contradiction, and so our assumption is false. Via De Morgan's law, this means that either our hypothesis is false or that the observation was actually unusual.

Even if we continue to assume that our observation was indeed unusual, our hypothesis often contains more assumptions than we think. In this example, the T-test was used. However, the T-test requires a number of conditions to be met, including normality. Below is a histogram of data used in this example.

As we can see, the sample does not appear to be normal. It may still be that the population is normally distributed, but this result should give us pause. And indeed, for the example, the data was sampled from a uniform distribution, with mean 176, a minimum of 86 and a maximum of 266. And so even though we rejected the null hypothesis, it was incorrect to conclude that the mean of the population was not 176cm.

Now let us go back and test a new null hypothesis: $\mu = 176$ and the distribution is uniformly distributed. Since heights cannot be negative, the minimum for the distribution would be 0 and the maximum would be 352, as the mean of a uniform distribution is half the difference between the minimum and maximum. Now our p-value is $\frac{146.4}{352} = 0.4159$, and so p is much greater than $p_{\frac{\alpha}{2}}$, so we fail to reject the null hypothesis.

4 Real World Phenomenon

Admittedly, the example that I provided was pathological. I tricked the reader into thinking a certain way. However, hypotheses do not exist on their own. They are generated through our understanding of real world phenomenon. It is because of our understanding of how height varies through a population that we assume normality and other conditions necessary to perform our hypothesis test. So when our tests cause us to reject our null hypothesis, what it is telling us is not what is true, but rather that our understanding of the mechanism being tested is incorrect and needs further consideration.