Basics of Hypothesis Testing

# Overview

Generating and testing hypotheses is the basis of nearly everything you do as a social scientist. This guide will discuss some of the basics of hypothesis testing that you might come across in readings for your courses or might use in your own research. As this is meant to be an introductory overview, we’ll mainly cover some basic concepts, including the foundations of hypothesis testing and a simple process for conducting a hypothesis test, regardless of your data.

There are a number of different varieties of hypothesis tests, and while this guide will not provide a detailed overview of all of them, I will include some information about the general properties of hypothesis tests and resources to learn more about the particularities of different varieties.

# Foundations

Before getting into the specific process of conducting a hypothesis test, let’s cover some of the basics about what a hypothesis is, and what it means to test one.

## What constitutes a hypothesis?

A hypothesis is a statement of an expected relationship between variables which simply state that variables *x* and *y* are not associated merely by chance. Such statements are generally informed by past research or by some observation of relationships that appear in the ‘real’ world, and can express an expected direction of a relationship (i.e. “if *x* occurs, *y* will increase/decrease”) or simply predict a non-coincidental relationship. One example of a simple hypothesis that might appear in political science is that *politicians will make more visits to their home districts during election years than off-cycle years*.

Sometimes, it is helpful to think about a hypothesis as a type of “if-then” statement, where you imagine manipulating one variable (the independent variable) and predict what will happen to another (the dependent variable). Using this method, we might think about the sample hypothesis above as follows: i*f an election is held during this year, then the number of days a politician spends in their home districts will increase in number.*

As with most political science, a hypothesis should be grounded in some form of theory. This can be a well-established theory like the democratic peace theory that appears often in some IR circles, something that you formulate on your own based on past research or logic, or can be informed by more sophisticated models of behavior, using techniques such as game theory to predict certain relationships. As you read the work of others and progress through your own research, it can be helpful to assess these underlying theories to ensure they “make sense” or that the hypotheses they/you generate are consistent with their theoretical foundations.

## What is hypothesis testing?

Hypothesis testing is a form of inferential statistics that can be used to draw conclusions about an entire population based on a random sampling thereof. Using random samples, we estimate properties of an entire population by evaluating data against two mutually-exclusive hypotheses, called the *null* and *alternative* hypotheses. Null hypotheses (sometimes annotated as ) tend to be pretty bland: they propose that there is no relationship or effect. Because of this, they may not always be explicitly stated in a published article, but rest assured that for every hypothesis, there is a null hypothesis of no effect at play. Using the example above, for instance, the null hypothesis might be expressed as “*politicians visit their home districts the same number of times each year.*” Note that while the null hypothesis tends to say that there is “no effect” or “no difference,” these terms are fairly broad, and can be thought of as “not enough of an effect to make a conclusion” rather than “absolutely no effect whatsoever.” A politician that flies home one more time on election years than off-cycle years (whether to campaign or participate in her hometown’s peculiar Leap Day festivities) is not enough of a difference to conclude that politicians behave differently in election years, but her colleague that increases visits tenfold in election years might constitute a significant difference.

While not very exciting, these hypotheses serve as the basis for hypothesis testing by providing a ‘status quo’ condition against which we can test evidence to determine if it is strong enough to draw conclusions. The null hypothesis is compared against a research hypothesis, known as the *alternative hypothesis* ( or ) which states simply that there *is* an effect/relationship at play.

The aim of hypothesis testing is to determine if there is sufficient evidence to reject the null hypothesis. Note that in science, we do not claim to “prove” either the null or alternative hypothesis, but merely whether there is sufficient evidence to allow us to reject or fail to reject the null.

## Broad types of hypothesis tests

While there are a number of varieties of hypothesis tests that you can use depending on your data and research question, these can be broadly classified along a few dimensions.

First, your hypothesis may make claims about a population mean or proportion. In our sample hypothesis, we may use slightly different methods depending on whether we want to know if a politician visits home more often during election years or if a certain proportion of politicians’ visits to their home districts occur in election years. So first, determine if you’re interested in studying a population mean, a difference in means, or proportion.

Second, consider whether your alternative hypothesis makes an explicit claim about the nature of an effect or relationship, or merely hypothesizes that such an effect exists. The latter is known as a *two- tailed hypothesis*. Two-tailed hypotheses allow you to test in both positive and negative directions (i.e. politicians change their travel behavior during elections), so if you’re merely trying to determine if a difference exists between samples, this can be a useful tool. These hypotheses don’t provide as much explanatory power as the alternative *one-tailed hypothesis*, which explicitly states the nature or direction of an effect (i.e. politicians visit *more* in election years than others). Here you can specify whether an effect should be greater or less than 0. Many political science papers will use some version of a one-tailed test.

Finally, determine if your data reflects a single sample or more. A *one-sample hypothesis test* uses one category to draw inferences about a single population. If, however, you are interested in the difference between samples grouped by some common feature, then a *two-sample* approach may be necessary. For instance, if you change the hypothesis above to specify that *democrats* travel home more often than *republicans* during election years, then a two-sample test allows you to estimate relationships for each group and draw conclusions about partisan differences in behavior. Alternatively, you may be interested in evaluating an effect over time, perhaps by comparing samples before and after an intervention. These tests are known as *paired* sample tests because the two samples are linked – in theory the same units of analysis observed at different times.

# Basic Process

Once you have your data, have clearly stated both the null and alternative hypothesis, and have determined the type of hypothesis test that is best for you, you can proceed to test whether or not you can reject a null hypothesis of no relationship/effect. In this section, I’ll walk through the basic process of a hypothesis test, which can be summarized in a few simple steps. Some differences exist for each type of test and for different types of data, and I will provide resources for more information on some of the major types of tests, but the basic process for any hypothesis test is as follows:

## 1: Set your parameters

In order to determine if the results of an experiment are sufficiently distinct from random chance, you must first determine how out of the ordinary a set of results must be before we can comfortably say that they are not just a coincidence. In statistics, this value is known as a “significance level,” or , or the probability of Type I Error (false positive). In most political science research, we use , but occasionally you might see values of .1 or .01.

## 2: Set a test statistic

Using the data you gathered and the parameter set beforehand, calculate a test statistic comparing the sample values against the parameter. Depending on the type of data this process and even the type of test statistic you should use will differ, but this process can usually be easily completed in R or other statistical software.

Here are some of the common test statistics used in hypothesis tests:

-*t-*statistic: Most common test statistic, but determined in different ways depending on the type of hypothesis test. For example, a one-sample t test is calculated as the proportion of the difference between the sample mean and known (or null) mean to its standard error. A t-value of 0 signifies no difference between sample and null. In R, you can calculate t-statistics using the t.test() function. For information about additional specifications to use for different types of data/tests, see <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/t.test>

-*z-*statistic: Similar to the t-statistic, but requires that you know population values, so is less practical in most political science settings.

-*F*-statistic: May be used in certain types of datasets, such as categorical data, or in conjunction with other statistical tools such as linear regression or ANOVA..

## 3: Use the test statistic to calculate the p-value

A p-value is the probability of finding the observed values (or even more extreme values) when the null hypothesis is true. When this probability is smaller than the pre-determined significance level, i.e. , we can be more comfortable concluding that the null hypothesis is not supported. When , then it is not possible to reject the null hypothesis.

## 4: Interpret results and make conclusions

Once you have calculated test-statistics and *p-*values, you can use this information to either conclude whether the

# Resources for more information

This guide is meant to provide a basic overview of the concepts behind hypothesis testing and to help you as you begin to navigate a hypothesis test of your own. Because hypothesis tests can vary considerably depending on your specific hypotheses and the nature of your data, I strongly suggest reviewing some of the following resources:

* Penn State University Statistics
  + Basic overview of different types of hypothesis tests: <https://online.stat.psu.edu/stat500/lesson/6a>
  + Analysis of Variance (ANOVA): <https://online.stat.psu.edu/stat502/lesson/1/1.2>
* A more comprehensive introduction to hypothesis testing, including more advanced topics such as Bayesian statistics: <https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/>
* <https://blog.minitab.com/blog/adventures-in-statistics-2/understanding-hypothesis-tests-significance-levels-alpha-and-p-values-in-statistics>
* T-tests for different types of data:
  + <https://blog.minitab.com/blog/adventures-in-statistics-2/understanding-t-tests-1-sample-2-sample-and-paired-t-tests>

# Final thoughts

Depending on your research question, hypothesis tests alone will generally not provide convincing evidence in support of your hypothesis. This is because hypothesis testing merely compares samples against expectations and so does not account for confounding conditions. Often, a hypothesis test serves as a useful starting point when developing a research paper: they can be used to determine whether or not a hypothesis is worth studying. Often, you will use the findings of a hypothesis test to decide whether to continue investigating a relationship or not.

Furthermore, a hypothesis should be firmly rooted in theory. Statistics can’t tell you anything more than an apparent relationship between a few numbers, the task of determining if that relationship is sensible rests with you. It is important to think through a hypothesis before you begin testing it: is it rooted in past research? Is it a logical extension to an existing theory? If you create your own theory, or use formal models to generate hypotheses, are the assumptions of that theory realistic? Is the data that you’re using appropriate for the question you intend to answer? These types of questions will help ensure that your own hypothesis testing is on sure footing.

If you can ensure that your hypotheses are sensible and have a sense for where a hypothesis test fits into your research project, these can be useful techniques for determining whether available data supports your ideas, can illuminate relationships and effects that might not be otherwise apparent, and can guide your research forward.