10/13/2022

Hypothesis Testing Introduction

It's going to take a couple of lectures to lay out all the ins and outs of hypothesis testing. We will begin with the big picture today, and then in coming lectures work our way through additional concepts, some we discuss today in greater detail, and performing hypothesis tests in specific contexts.

Hypothesis testing is the heart of statistical inference. In hypothesis testing we are trying to answer the question at the very heart of scientific inquiry: do we have enough evidence to believe our new claim rather than continue to believe our prior assumptions? What does it mean for the evidence to be good enough? What is the probability that we are wrong about making our claims? How do we balance our prior assumptions with the new evidence? Virtually everything we will do from now on in statistics will be connected to a hypothesis test. We will learn some additional descriptive tools, particularly in the second semester, but even then, it will be in the service of analyzing data, making claims and judging the quality of the evidence we've collected.

When we conduct a hypothesis test, we are comparing two claims. One claim is based on our assumptions or prior knowledge, and one claim is based on the evidence we've collected. Our claims are set up in opposition to each other. Sometimes choosing the claim for our null hypothesis—our fallback position in the face of weak evidence—must be selected based on managing things like risk, not just our best prior knowledge. An example of this would be in the casing of testing water supplies for safety, say, in Flint, MI. You want to have strong evidence that the water is safe. You don't want to tell people it's safe and then find out later you were wrong because then it's too late. So, the safe assumption is that the water is contaminated, and then establish with good evidence that it is not. What the null hypothesis has in common all the time though, is that it is never based on the data we've collected.

The null hypothesis we can think of, as I said above, as the fallback position. What would you conclude if you had no evidence? What is the best thing, or the safest thing (broadly construed) to believe in the absence of evidence? This hypothesis is important in part because it is going to be the root of our test. It is the assumption we will make about the world, the standard by which we will compare our new data. The fundamental question will be: if the null hypothesis is true, what is the probability we could collect the data we have? What is the probability we could obtain these results purely by chance? We will use our knowledge of sampling distributions to make this determination.

In most cases, we can use confidence intervals to also test hypotheses. Although that is not the approach we will take in most cases. When we create a confidence interval, we are saying the true population mean is in a given range. But what if we collect another sample whose mean is outside that range? One interpretation could be that the mean of the new sample differs just by chance. But another interpretation is that the new sample is actually from a different population. The further away the new sample is from the center of confidence interval, the more likely the different population interpretation is to be true, because the chance of the first interpretation gets smaller and smaller. One question we'll have to settle on early on in the process will be how unlikely is unlikely enough? How small does the probability that the new sample meets the assumptions of the null before we conclude the evidence is strong enough to alter our beliefs? We'll discuss how confidence intervals and hypothesis testing is linked, and where they differ.

Many aspects of hypothesis testing are standardized, but there will be reasons to deviate from those standards in some cases.

There is one "real life" situation where we conduct a kind of hypothesis test, and that is in the legal system. When someone is arrested, they go to criminal court, and they are presumed innocent until proven guilty. The presumption of innocence is the null hypothesis in this context. The prosecution must collect sufficient evidence to overcome "reasonable doubt". We can think of this as the prosecution establishing that the evidence is strong enough to conclude that the chance that the accused is innocent is very unlikely. We can't quantify this evidence in the same way we can a sample mean, but the idea is the same. Civil courts have similar idea although the standard of evidence is lower (this means that they accept a greater chance that they could be wrong because the consequences of a civil trial is just money and not prison). We will sometimes refer back to this analogy, since everyone is familiar with the idea, to clarify some concepts.

The notation for the null hypothesis is typically H_0. In mathematical hypothesis tests, the null hypothesis will include a statement about our assumptions, such as that the mean is 0 (or some other value), the proportion is 50%, that two means are equal, or some other similar statement depending on the specific test we are conducting. As we introduce each type of hypothesis, we'll see what to do for each type of test. The null hypothesis may be stated with an $=$ sign, or inequalities that contain equal signs like $\leq, \geq$. To make things simpler, I will just always use an $=$ sign (since we are stating a property that we are directly using in our mathematics). Different authors may adopt a different convention, so you may see the other ones various texts.

The alternative hypothesis can be notated in several ways: H_a, H_A, H_1 are all somewhat common (a for alternative, 1 in contrast to 0). Any of these are fine, but you will see different notations in different texts. The alternative hypothesis likewise will describe a relationship with a particular statistic consistent with null hypothesis (if we use a statement about a mean in the null hypothesis, we can't make a statement about a proportion in the alternative, only something about a mean). The alternative hypothesis will also never include an equal sign. It can only contain $\neq, <, >$. The number expressed in the statement cannot be from the sample. It can only be in opposition to the null: the equality changes to inequality, but the number used in the statement does not change.

For example:

$$H_0: \mu = 0$$
$$H_a: \mu \neq 0$$

The $\neq$ sign indicates that this a two-tailed test (similar to your typical confidence interval). We are interested in the sample mean being far away from 0 (whatever our null hypothesis value is). We don't care whether the number is more or less. The probabilities come from both tails of the distribution. We find the critical values in this case similarly to two-sided confidence intervals.

Or, we can have the following:

$$H_0: \mu \leq 0 \text{ or } H_0: \mu = 0$$
$$H_a: \mu > 0$$

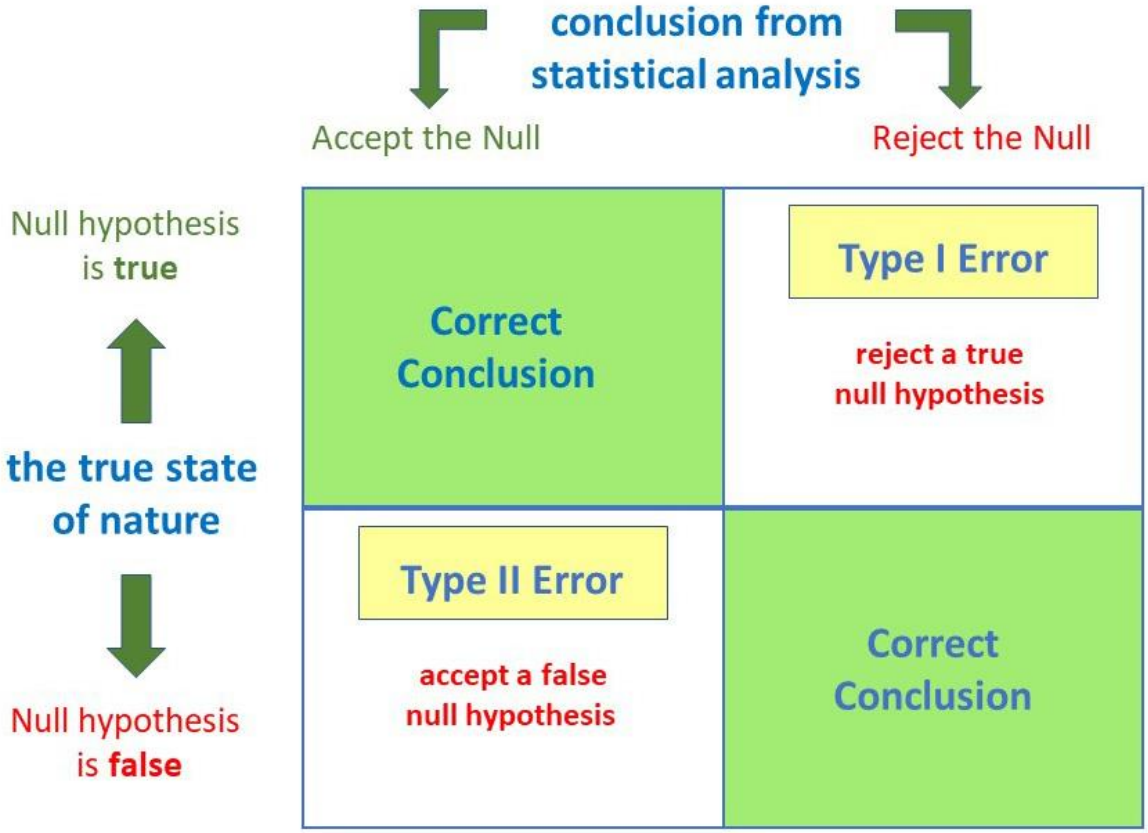
$$H_0: \mu \geq 0 \text{ or } H_0: \mu = 0$$
$$H_a: \mu < 0$$

If we use the inequality notation in the null (with the equality marker), the alternative hypothesis is always pointing in the opposite direction. We cannot prove equality because we are making an estimate (that's why it's in the null), but we can say it's higher or lower than the assumption in the null hypothesis. These cases are called one-tailed tests. We are only concerned with results that are in one

tail of the distribution, not the other end. We will find the critical values here similar to what we did for one-sided confidence intervals.

Which type of test you do is going to depend on the way the question is asked (in practice, we should try to formulate our research question to test before we collect data). If the problem statement wants to know if the data is "the same" or "different", then we are looking at a two-tailed test and we'll use $\neq$ in our alternative hypothesis statement. If the problem statement says more than or less than, then use the appropriate inequality in the alternative (if you are not using the $=$ sign in all cases, then set the null hypothesis in the opposite direction to the alternative you wish to prove).

There are two types of errors we can make in hypothesis testing. We could reject the null hypothesis incorrectly, or we could fail to reject the null hypothesis even though it is false. The first is called a Type I error. The second is called a Type II error.



We will say more about these errors in a future lecture. We generally set the chance of a Type I error, but then the Type II error will depend on that and our data.

It's worth noting that these errors are not "mistakes" in the usual sense. They are simply the fault of working with random variables and probabilities. They will happen sometimes even when we do everything "right".

The probability that unites confidence intervals and hypothesis testing is α. The confidence level in our confidence intervals is $1-\alpha$, and the significance level in our hypothesis testing is just α. If the confidence interval includes 95% of the sampling distribution, then 5% is left out. You may recall the use of α in our notation for the critical value for our confidence interval calculations. The significance level is the probability of a Type I error.

We can also talk about the power of a test, which is given by $1-\beta$. The value β is the chance of a Type II error. For now, it's worth knowing these ideas exist. We'll worry about calculating β when we have more familiarity with hypothesis testing.

Let's look at some examples and just set up the null and alternative hypotheses for these one-sample tests.

If a problem does not say so, the significance level is assumed to be 5%. For each example, we want to state the null and alternative hypothesis (is it a mean or proportion? What assumption are we making in the background that we are comparing our data to? Is the test one-tailed or two-tailed?)

Example.
A researcher thinks that if knee surgery patients go to physical therapy twice a week (instead of 3 times), their recovery period **will be longer**. Average recovery times for knee surgery patients is 8.2 weeks.

$$H_0: \mu = 8.2$$
$$H_a: \mu > 8.2$$

Example.
Blood glucose levels for obese patients have a mean of 100 with a standard deviation of 15. A researcher thinks that a diet high in raw cornstarch will **have an effect** on blood glucose levels. A sample of 30 patients who have tried the raw cornstarch diet have a mean glucose level of 140. Test the hypothesis that the raw cornstarch had an effect.

$$H_0: \mu = 100$$
$$H_a: \mu \neq 100$$

Example.
The mortgage department of a large bank is interested in the nature of loans of first-time borrowers. This information will be used to tailor their marketing strategy. They believe that 50% of first-time borrowers take out smaller loans than other borrowers. They perform a hypothesis test to determine if the percentage is **the same or different** from 50%. They sample 100 first-time borrowers and find 53 of these loans are smaller that the other borrowers. For the hypothesis test, they choose a 5% level of significance.

$$H_0: p = 0.50$$
$$H_a: p \neq 0.50$$

References:

1. https://faculty.ksu.edu.sa/sites/default/files/probability_and_statistics_for_engineering_and_the_sciences.pdf
2. https://assets.openstax.org/oscms-prodcms/media/documents/IntroductoryStatistics-OP_i6tAI7e.pdf
3. https://www.simplypsychology.org/type_I_and_type_II_errors.html
4. https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/
5. https://opentextbc.ca/introbusinessstatopenstax/chapter/full-hypothesis-test-examples/