Contents

Hypothesis Testing

1. Introduction

In everyday life, we often have to make decisions based on incomplete information. These may be decisions that are important to us such as, "Will I improve my biology grades if I spend more time studying vocabulary?" or "Should I become chemistry major to increase my chances of getting into med school?" This section is about the use of hypothesis testing to help us with these decisions. Hypothesis testing is a kind of statistical inference that involves asking a question, collecting data, and then examining what the data tells us about how to proceed.

In a formal hypothesis test, hypotheses are always statements about the population. The hypothesis might involve statements about the average values (means) of some variable in the population. For example, we may want to know if the average time that college freshmen spend studying each week is really 20 hours per week. We may want to compare this average time spent studying for freshmen that earned a GPA of 3.0 or higher and those that did not.

2. What is a Hypothesis

It is an educated guess about something in the world around you either by experiment or observation. For example:

- A new medicine we think might work
- A better way to administer test.
- A possible location of new species

3. Developing Null and Alternative Hypotheses

In statistical hypothesis testing, there are always two hypotheses. There is an initial research hypothesis of which truth is unknown.

Example 1: *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. An average recovery times for knee surgery patients is 8.2 weeks.*

The hypothesis statement in this question is that the researcher believes the average recovery time is more than 8.2 weeks. It can be written in mathematical terms as:
H_1: $\mu > 8.2$

Next we need to state null hypothesis. That's what will happen if the researcher is *wrong*. In the above example, if the researcher is wrong then the recovery time is less than or

equal to 8.2 weeks. In math, that's:
H_0 μ ≤ 8.2

The research hypothesis proposed by researcher is called research hypothesis and given the symbol H1.The hypothesis to be tested is called the **null hypothesis** and given the symbol H0. The null hypothesis state that the average recovery time is less than 8.2. Now to prove our research hypothesis we have to disprove null hypothesis.

Example2: *A researcher thinks that college freshmen do not study 20 hours per week. So the research hypothesis is*

H1 : μ != 20

Now we state out null hypothesis assuming researcher is wrong. If we were to test the hypothesis that college freshmen study 20 hours per week, we would express our null hypothesis as:

H0 : μ = 20

We test the null hypothesis against an alternative hypothesis, which is given the symbol H1.

Example 3
We have a medicine that is being manufactured and each pill is supposed to have 14 milligrams of the active ingredient. What are our null and alternative hypotheses?

Solution:
So we are trying to prove that each pill does not have 14 mg of active ingredient.
H0 : μ = 14
Ha : μ != 14
Our null hypothesis states that the population has a mean equal to 14 milligrams. Our alternative hypothesis states that the population has a mean that is different than 14 milligrams.

Example 4
The school principal wants to test if it is true what teachers say – that high school juniors use the computer an average 3.2 hours a day. What are our null and alternative hypotheses?
Solution
H0 : μ = 3:2
Ha : μ != 3:2
Our null hypothesis states that the population has a mean equal to 3.2 hours. Our alternative hypothesis states that the population has a mean that differs from 3.2 hours.

ALTERNATE HYPOTHESIS
A statement that is accepted if the sample data provide sufficient evidence that the null hypothesis is false.

NULL HYPOTHESIS
A statement about the value of a population parameter developed for the purpose of testing numerical evidence.

4. Deciding Whether to Reject the Null Hypothesis

The alternative hypothesis can be supported only by rejecting the null hypothesis. To reject the null hypothesis means to find a large enough difference between your sample mean and the hypothesized (null hypothesis) mean that it raises real doubt that the true population mean is 20. If the difference between the (null) hypothesized mean and the sample mean is very large, we reject the null hypothesis. If the difference is very small, we do not. In each hypothesis test, we have to decide in advance what the magnitude of that difference must be to allow us to reject the null hypothesis. Below is an overview of this process. Notice that if we fail to find a large enough difference to reject, we fail to reject the null hypothesis. Those are your only two alternatives. When a hypothesis is tested, a statistician must decide on how much of a difference between means is necessary in order to reject the null hypothesis.

5. Significance Level

The Four Possible Outcomes in Hypothesis Testing

	Researcher	
Null Hypothesis	Does Not Reject H_0	Rejects H_0
H_0 is true	Correct decision	Type I error
H_0 is false	Type II error	Correct decision

Remember that there will be some sample means that are extremes – that is going to happen about 5% of the time, since 95% of all sample means fall within about two standard deviations of the mean. What happens if we run a hypothesis test and we get an extreme sample mean? It won't look like our (null hypothesis) hypothesized mean, even if it comes from that distribution. We would be likely to reject the null hypothesis. But we would be wrong. When we decide to reject or not reject the null hypothesis, we have four possible scenarios:
a. A true hypothesis is rejected. (In correct decision)
b. A true hypothesis is not rejected. (Correct decision)
c. A false hypothesis is not rejected. (In correct decision)
d. A false hypothesis is rejected. (Correct decision)
If a hypothesis is true and we do not reject it (Option 2) or if a false hypothesis is rejected (Option 4), we have made the correct decision.

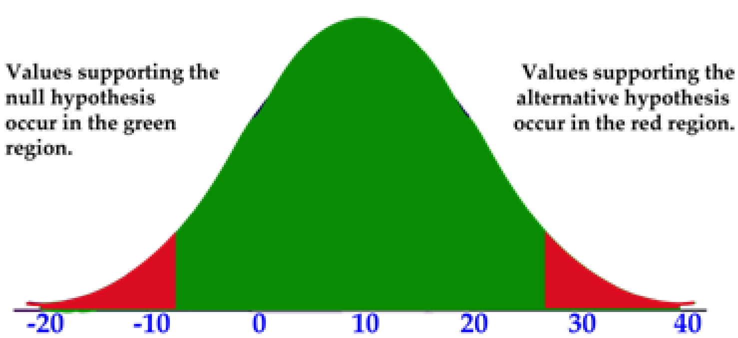
But if we reject a true hypothesis (Option 1) or a false hypothesis is not rejected (Option 3) we have made an error.

The general approach to hypothesis testing focuses on the Type I error: rejecting the null hypothesis when it may be true. Guess what? The level of significance, also known as the alpha level, IS the probability of making a Type I error. At the 0:05 level, the decision to reject the hypothesis may be incorrect 5% of the time.
Statisticians first choose a level of significance or alpha (a) level for their hypothesis test. Similar to the significance level you used in constructing confidence intervals, this alpha level tells us how improbable a sample mean must be for it to be deemed "significantly different" from the hypothesized mean. The most frequently used levels of significance are 0:05 and 0:01: An alpha level of 0.05 means that we will consider our sample mean to be significantly different from the hypothesized mean if the chances of observing that sample mean are less than 5%. Similarly, an alpha level of 0.01 means that we will consider our sample mean to be significantly different from the hypothesized mean if the chances of observing that sample mean are less than 1%.

6. Two Tailed hypothesis testing

A hypothesis test can be one-tailed or two-tailed. The examples above are all two-tailed hypothesis tests except first one. We indicate that the average study time is either 20 hours per week, or it is not. Computer use averages 3.2 hours per week, or it does not. We do not specify whether we believe the true mean to be higher or lower than the hypothesized mean.
We just believe it must be different.

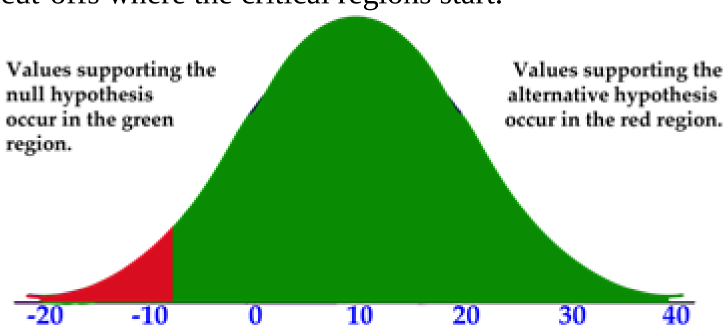
In a two-tailed test, you will reject the null hypothesis if your sample mean falls in either tail of the distribution. For this reason, the alpha level (let's assume .05) is split across the two tails. The curve below shows the critical regions for a two-tailed test. These are the regions under the normal curve that, together, sum to a probability of 0.05. Each tail has a probability of 0.025. The z-scores that designate the start of the critical region are called the critical values. If the sample mean taken from the population falls within these critical regions, or "rejection regions," we would conclude that there was too much of a difference and we would reject the null hypothesis. However, if the mean from the sample falls in the middle of the distribution (in between the critical regions) we would fail to reject the null hypothesis.



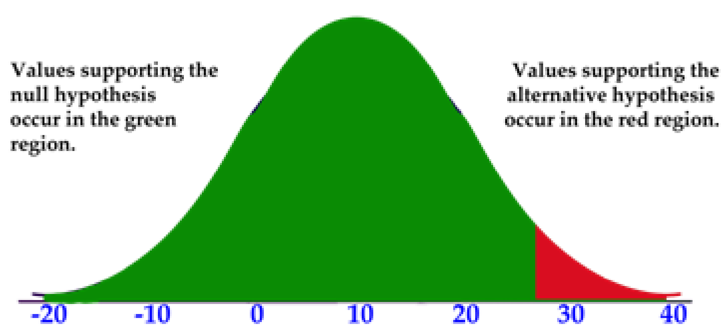
A two-tailed test has tails on both ends of the graph. This is a test where the null hypothesis is a claim of a specific value. For example: $H_0 : X = 5$

7. One-Tailed Hypothesis Test

We would use a single-tail hypothesis test when the direction of the results is anticipated or we are only interested in one direction of the results. For example, a single-tail hypothesis test may be used when evaluating whether or not to adopt a new textbook. We would only decide to adopt the textbook if it improved student achievement relative to the old textbook.
When performing a single-tail hypothesis test, our alternative hypothesis looks a bit different. We use the symbols of greater than or less than. For example, let's say we were claiming that the average SAT score of graduating seniors was GREATER than 1,110. Remember, our own personal hypothesis is the alternative hypothesis. Then our null and alternative hypothesis could look something like:
H0 : $\mu <= 1100$
H1 : $\mu > 1100$
In this scenario, our null hypothesis states that the mean SAT scores would be less than or equal to 1,100 while the alternate hypothesis states that the SAT scores would be greater than 1,100. A single-tail hypothesis test also means that we have only one critical region because we put the entire critical region into just one side of the distribution. When the alternative hypothesis is that the sample mean is greater, the critical region is on the right side of the distribution (see below). When the alternative hypothesis is that the sample is smaller, the critical region is on the left side of the distribution. To calculate the critical regions, we must first find the critical values or the cut-offs where the critical regions start.



A left-tailed test only has a tail on the left side of the graph



A right-tailed test only has a tail on the right side of the graph:

8. Type I and Type II Errors

You should be able to recognize what each type of error looks like in a particular hypothesis test. For example, suppose you are testing whether listening to rock music helps you improve your memory of 30 random objects.
Assume further that it doesn't. A Type I error would be concluding that listening to rock music did help memory (but you are wrong). A Type I error will only occur when your null hypothesis is false. Let's assume that listening to rock music does improve memory. In this scenario, if you concluded that it didn't, you would be wrong again. But this time you would be making a Type II error—failing to find a significant difference when one in fact exists. It is also important that you realize that the chance of making a Type I error is under our direct control. Often we establish the alpha level based on the severity of the consequences of making a Type I error. If the consequences are not that serious, we could set an alpha level at 0:10 or 0:20: In other words, we are comfortable making a decision where we could falsely reject the null hypothesis 10 to 20% of the time. However, in a field like medical research, we would set the alpha level very low (at 0:001 for example) if there was potential bodily harm to patients.

9. Lesson Summary

a. Hypothesis testing involves making educated guesses about a population based on a sample drawn from the population. We generate null and alternative hypotheses based on the mean of the population to test these guesses.
b. We establish critical regions based on level of significance or alpha (a) levels. If the value of the test statistic falls in these critical regions, we are able to reject it.
c. When we make a decision about a hypothesis, there are four different outcome and possibilities and two different types of errors. A Type I error is when we reject the null hypothesis when it is true and a Type II error is when we do not reject the null hypothesis, even when it is false.

10. Review Questions

1. If the difference between the hypothesized population mean and the mean of the sample is large, we ___ the null hypothesis. If the difference between the hypothesized population mean and the mean of the sample is small, we ___ the null hypothesis.
2. At the Chrysler manufacturing plant, there is a part that is supposed to weigh precisely 19 pounds. The engineers take a sample of parts and want to know if they meet the weight specifications. What are our null and alternative hypotheses?

3. In a hypothesis test, if difference between the sample mean and the hypothesized mean divided by the standard error falls in the middle of the distribution and in between the critical values, we ___ the null hypothesis. If this number falls in the critical regions and beyond the critical values, we ___ the null hypothesis.
4. Sacramento County high school seniors have an average SAT score of 1,020. From a random sample of 144 Sacramento High School students we find the average SAT score to be 1,100 with a standard deviation of 144: We want to know if these high school students are representative of the overall population. What are our

5. Please fill in the types of errors missing from the table below:

Decision Made	Null Hypothesis is true	Null Hypothesis is False
Reject Null Hypothesis	**(1)**	**(3)**
Do not Reject Null Hypothesis	**(2)**	**(4)**

Review Answers

1. Reject, Fail to Reject
2. H0 : μ = 19, Ha : μ 6= 19
3. Fail to Reject, Reject
4. H0 : μ = 1020;Ha : μ 6= 1020;Z = 6:67
5. Type I error, Type II error