Unit 4 Lecture: Logic of Hypothesis Testing

I will begin this lecture by giving you some invaluable information: You MUST master the content of this unit in order to successfully complete the remainder of the course. If there is a Unit that demands study time, it is this one. In my opinion, the way to increase the probability of succeeding is to:

1. Read the required textbook readings to get the "overall big picture"
2. Read this lecture to make the "big picture" manageable
3. <u>Watch the Video Lectures</u> to get clear on the FOUR-STEP PROCESS of calculating a hypothesis test
4. Repeat steps 1-4 until you are able to complete the assignments

What is a Hypothesis Test?

As you might remember from Unit 1, researchers, if they could, would always study **populations**; but, populations are simply too large to allow researchers to observe and measure every single individual in it. For example, imagine that you were a researcher who wanted to study the population of 5-year-olds in the United States. How would you ever be able to observe and collect data on ALL 5-year-olds in the United States? I hope you can see the impossibility of the task.

Therefore, what researchers do, as you might remember, is to take a subset, a **sample**, of the population and make observations and collect data on the sample; that is, researchers would conduct their studies on samples not on populations. Once the researchers collect the data on their samples, they need to describe, organize, and interpret the data.

Up to this point in the course, we have focused on **descriptive statistics,** methods that can describe and organize data, but cannot "interpret" or "explain" data. Now, we will focus on statistical methods that DO allow researchers to interpret and explain their data. These statistical methods are called **inferential statistics**.

Hypothesis Tests are inferential statistical methods that allow a researcher to study samples and INFER the research results back to the population that was too big for them to study.

Important Terminology Before We Go On

One important thing to know is that a <u>hypothesis test is different from a hypothesis</u>. It will be helpful for you to understand (1) what a hypothesis is, (2) what an independent variable is, and (3) what a dependent variable is. These 3 "things" are interrelated.

<u>A hypothesis is a prediction of what a researcher predicts will happen under specific conditions</u>. The following is an example of a hypothesis (albeit simplified for instructional purposes.

A researcher hypothesizes "that meditation will have an effect on blood pressure."

Importantly, a hypothesis will have 2 variables: an independent variable and a dependent variable. In their experiments, researchers aim to show a cause and effect relationship between the independent variable and the dependent variable. You must learn to recognize a hypothesis and to name both the independent and dependent variable in it.

One way to figure out which variable is the independent variable (IV) and which is the dependent variable (DV) is to first be clear about the two variables of interest. What two variables in the above hypothesis is the researcher trying to "relate" to each other? Most students do not have too much trouble seeing that the two variables are meditation and blood pressure.

Now, the question is: Which is the independent variable (IV), and which is the dependent variable (DV)? One way that almost always works is to ask yourself –Which variable is going to CAUSE a change in the other variable? Will meditation cause some change in blood pressure or will blood pressure cause some change in meditation?

Clearly, meditation is going to **cause** some change in blood pressure; therefore, meditation is the *independent variable* (IV). The independent variable is always the one that is the CAUSE of whatever effect is observed.

The variable that will be affected by the independent variable, the variable that changes because of the meditation (the IV) is the dependent variable (DV); in this case, the DV is the blood pressure.

Also, it is important to know that the dependent variable is the variable that researchers **measure:** The variable on which they make observations and collect data. It is precisely those data that a hypothesis test can analyze in order to conclude whether or not there is support for the researcher's hypothesis. In other words, after a hypothesis test is used to analyze the experimental data, the researcher will be able to conclude whether or not meditation did have an effect (or cause a change in) blood pressure. And, importantly, if they do see a difference in blood pressure, was it caused by the meditation or was it likely to be a "just by chance" event?

I have highlights certain terms and information in yellow and certain terms and information in green. First, read everything as a whole (as you probably have just done). Secondly, read only the yellow. And, thirdly, read only the green. Reading it this way may help you to better understand the differences between an independent and dependent variable and how to recognize each one.

More on Hypothesis Testing

At this point, aim to get a general idea that all hypothesis tests will analyze differences between and among means from data collected in experiments. I am sure that you remember that "means" are arithmetic averages. One of the first things that researchers do when they collect the data from their experiments is to calculate means. Different experimental designs (which we will discuss later in this course) will result in different means.

Researchers may be interested in population means, sample means, control group means, experimental group means, etc. The research method that a researcher chooses →will determine the types of means that need to be analyzed, and the types of means that need to be analyzed →will determine which hypothesis test is the appropriate one to use to analyze the means. All this “talk” about means is something that we will discuss in this unit and also in the next two units.

So let us begin slowly…..and start by adding on to our general overview of hypothesis testing.

Different Types of Hypothesis Tests

In this course, you will learn the following hypothesis tests, all of which analyze (compare) specific means. While we only cover the Z-test in this unit, it is imperative that you study and understand this unit’s content so that you don’t have more than the needed amount of frustration in the next two units ☺

1. **Z-test**
   The Z-test will analyze (compare) a population mean (μ) to a sample mean (M)

2. **Independent-measures t-test**
   The independent-measures t-test will analyze (compare) a control group mean (M) with an experimental group mean (M)

3. **Dependent-measures t-test**
   The dependent-measures t-test will analyze (compare) a “before-treatment” group mean (M) to an “after-treatment” group mean (M)

4. **ANOVA**
   Notice that the first three tests only compare TWO means. When a researcher needs to compare more than two means, then ANOVA is necessary because it can look at differences of three means (M-M-M) or more.

Overall, notice that the TYPES of means you want to compare→ determines the appropriate hypothesis test to analyze the data

What all Hypothesis Tests Have in Common

The logic of hypothesis testing, which is the focus of this unit, is inherent (built in, true) of **ALL of the hypothesis tests that we will cover in this course;** therefore, again, the extent to which you master this unit will facilitate the next two units. I can’t seem to say that enough. Trust me on this.

Let’s continue…..

ALL hypothesis tests can be thought of as a fraction. A fraction consists of a numerator and a denominator. In **all** hypotheses tests, the numerator analyzes one thing and the denominator analyzes another thing. The mathematical result of **any** hypothesis test is nothing more than dividing the numerator by the denominator.

Let's take a look at the **general format that is seen in ALL hypotheses tests:**

$$\textbf{ALL HYPOTHESIS TESTS} = \frac{\textbf{Differences between means because of the IV}}{\textbf{Differences between means due to ERROR}}$$

The above "formula" is a very, very important idea to understand. As you can see from the general formula above, a hypothesis test can be expressed as a fraction.

The numerator calculates the difference between the means we believe is caused by the independent variable (IV). What this means is that when a researcher (or you) analyzes the data, some of the result of that analysis will allow the researcher (or you) to assume that the difference in the means (the means that go in the numerator) that you are seeing can be attributed to the independent variable (IV). To relate it to our earlier hypothesis above, we can assume that the differences are because the meditation caused the differences in blood pressure.

The denominator is referred to as "the ERROR" term. The denominator will also compare differences between means. But the denominator will calculate the differences between means that can be attributed to "just by chance" differences. These "just by chance" differences is what we mean by ERROR. "Just by chance" (ERROR) refers to differences between the means that are there for any reason other than differences caused by the IV.

A Flash from Past Course Information to Help Understand "just by chance"

Do you recall "sampling error" in Unit 1? Sampling error was the difference you should expect, just by chance, when you take a sample from the population. You haven't done anything to the sample, but yet you should expect some difference between the sample mean and the population mean. Do you recall this discussion? The "just by chance" error in the denominator of any hypothesis test has the same general idea: there will be a difference between the means that are compared in a hypothesis test that are there "just by chance" and not because of anything you have done.

TAKE NOTE: "Just by chance" is an important idea in hypothesis testing. As a matter of fact, the whole point of running a hypothesis test is to make sure that the results of a research studies are NOT a random or "just by chance" event, but rather the results can be attributed directly to the independent variable (IV).

Important MESSAGE from me to you:

If you cannot understand what is being discussed here, try reading the textbook one more time, and then come back to this lecture before moving on to the video lectures. While it may appear daunting to understand the content of this unit, you CAN do it; and it will facilitate the rest of this semester if you master this material.

Working with Research Scenarios

From this point on in this course, you will see many "research scenarios" such as this one:

Research Scenario:

You are a researcher who has developed a new math tutoring program. You hypothesize that your math tutoring program will have an effect on math ability. You also know that the adult **population** distribution for math ability is normal with a $\mu = 100$ and $\sigma=30$. You select a **sample** of $n = 25$ from this population to participate in your study. You spend six months teaching them math using your new math tutoring program. After your sample of $n=25$ completes the tutoring program, you give all 25 of them a standardized math test to measure their math ability. The standardized test is how you are measuring the DV (the dependent variable). Then you make your observations by looking at their test scores, and record your data: you record all 25 test scores. You calculate the sample mean (the mean of all 25 scores) and you get a mean of $M = 110$.

You would need to conduct a four-step hypothesis test in order to analyze your data and be able to conclude that your math tutoring program works ☺

In order to begin your 4 step hypothesis testing process, you first NEED TO DECIDE WHICH HYPOTHESIS TEST IS THE APPROPRIATE TEST TO ANALYZE YOUR DATA.

Beginning to Understand How to Choose the Appropriate Hypothesis Test

The first thing you must be able to do is decide which hypothesis test is appropriate to analyze the data. This is necessary before you can conduct the 4-step hypothesis test. Makes sense, right? If you don't know which test to use, how do you go on to the four-steps of doing it?

Here is how you should go about deciding which hypothesis test is appropriate:

First, ask yourself: What **<u>means</u>** do I need to compare?

> Answer: From the scenario and our discussion about it, you should be able to see that you want to compare a sample mean ($M = 110$) to a population mean ($\mu = 100$) because you want to see how your sample, with a $M=100$, performed on the math test when compared to the normal population, with a $\mu = 100$. So the comparison you want to make is a sample mean to a population mean.

Secondly, ask yourself: What is the appropriate hypothesis test when I want to compare a population mean, μ, to a sample mean, M ?

Answer: You have 2 choices: a Z-test or a Single- sample t-test.

How do you decide between those two?

Read the research scenario, and if it gives you the population standard deviation (σ), then, you would KNOW the population standard deviation, and that would tell you that the correct choice is a Z-test.

If you DON'T KNOW the population standard deviation (σ), then the correct choice would be the single-sample t-test.

We have not yet covered these tests, but we will cover the Z-test here, and the other tests in later units.

Thirdly, ask yourself: Do I need to do a **directional** (one-tail) Z-test or a **non-directional** (two-tail) test?

Answer: To make this decision, you need to read the hypothesis very carefully. The hypothesis in our research scenario reads, "the tutoring program will have an effect on math ability."

Notice that the researcher simply hypothesizes that there will be an affect. The researcher does NOT predict whether the affect will be an INCREASE or a DECREASE in math ability. While you may infer that the expected result is improved math ability, the hypothesis does not explicitly predict that.

Pay attention to this part:

When the hypothesis simply predicts "an effect," without a specific direction, then a non-directional (two-tail) test is the correct choice. If the hypothesis had read, "the tutoring program will INCREASE math ability," then the hypothesis is DIRECTIONAL and would require a directional (one-tail) test.

Because we want to compare a sample mean (M) to a population mean (μ) and the hypothesis is non-directional, the appropriate choice of hypothesis test is:

A non-directional (two-tail) Z-test. Once we have chosen the appropriate hypothesis test, we can calculate the 4-step process of the hypothesis test.

The four step process is explained in the Video Lecture

So, where are we at this point?

We are at the point at which (1) we have taken a sample of research participants from a population, (2) provided our participants with a tutoring program, (3) gave the entire sample a standardized math test, (4) recorded the data of our the 25 participant's performance on the math test, (5) calculated a mean for their 25 scores, M=110, and (6) chosen the correct hypothesis test to analyze the data.

Now, you would want to compare the means in our data, the sample mean of those who were taught by your new method (M=110) to the population mean of those who were not (μ=100) so that you can answer the most important question:

Did your new tutoring method of teaching math result in significantly better performance for those who took it (the sample) when compared to those who didn't (the population)? In other words, did the tutoring program have a significant effect on math ability as the (your) hypothesis predicted?

Again, how to conduct the 4-step hypothesis test is demonstrated in the VIDEO LECTURES ☺

Necessary Information Before Viewing the VIDEO LECTURE

At this point, I will only focus on Step 3 (the math part) of the 4-Step hypothesis testing process to (a) introduce you to the formula for the Z-test, (b) explain some further concepts, and (c) reinforce some of the concepts which have already been covered in this lecture.

This is a good place to return to the general "formula" for ALL hypotheses tests:

$$\textbf{ALL HYPOTHESIS TESTS} = \frac{\textbf{Differences between means because of the IV}}{\textbf{Differences between means due to ERROR}}$$

Remember that the above is a formula for ALL hypotheses tests. We are working the Z-test. Here is its formula:

$$Z = \frac{M - \mu}{\sigma_m}$$

Notice that the numerator has the two means we want to compare (1) the sample mean (M=110) and the population mean (μ = 100). At this point, we already know that the sample mean is 110 (M =110) and the population mean is 100 (μ = 100). Therefore, we can calculate the numerator (the difference in math ability that we believe is because of the new tutoring program):

$$M-\mu = 110-100 = 10$$

So, now, again, we know that the difference in math ability that can be attributed to the tutoring method (to the IV) is 10. If I asked you the question:

Did the tutoring program have an effect (Did it work)?

What would you say?

It may be tempting to say yes because, after all, there is a 10 point difference between those who got the new tutoring method and those who did not; but, the answer is: WE DON'T KNOW YET.

And this is the whole point of hypothesis testing:

We CANNOT say, at this point, that the tutoring program worked. Can you guess why?

The reason is because all we have done is calculated half of the hypothesis test (the numerator). And that, by itself isn't enough.

WE MUST COMPARE THE DIFFERENCE WE SEE AND BELIEVE IS DUE TO THE TUTORING PROGRAM (the numerator) to the DIFFERENCE THAT ARE THERE 'JUST BY CHANCE" (the denominator).

The "just by chance" differences are differences that are there between the sample mean and the population mean, but these differences are NOT caused by the tutoring program, but by, for example, any other "personal" differences such as IQ, age, education, etc. etc. These "by chance" differences is what we refer to as ERROR.

Each hypothesis test will have its own ERROR term with its own statistical notation. But for now, we are only focused on the test that we are interested in, which is the Z-test.

The ERROR term in the Z-test is called **STANDARD ERROR** and its statistical notation is:

σ_m (the 'm' is a subscript, not a mean)

and σ_m is calculated like this:

$$\frac{\sigma}{\sqrt{n}}$$

Let's remember where we are: at this point, we have calculated the difference between means in the numerator of the Z-test, and now we need to calculate the denominator of the Z-test.

We can calculate the standard error as follows (the numbers come from the research scenario above):

σ_m = $\frac{\sigma}{\sqrt{n}}$ = $\frac{30}{\sqrt{25}}$ = $\frac{30}{5}$ **Answer = 6**

Notice that the research scenario provided the value of the σ = 30 and that the sample size (n) was 25. Therefore if we divide σ (30) by the square root of n (5, because the square root of n=25 is 5), we get a standard error of $\sigma_m = 6$.

So now we know that the difference we can expect, just by chance (ERROR) is 6; and, the difference we believe is due to the tutoring program is 10. Before we can conclude that the new math tutoring method had an effect on math ability, we MUST COMPARE the NUMERATOR (difference we can attribute to the IV) to the DENOMINATOR (differences we attribute to "just by chance" differences) in the hypothesis test.

Again, we must compare the difference we think is due to the IV (the tutoring program) to the difference we believe is due to chance, or ERROR. The CRUX OF HYPOTHESIS TESTING is eliminating results that are no better than chance occurrence.

In order to conclude that the tutoring program did have an effect on math ability, we must show that the difference between the means due to the IV is SIGNIFICANTLY more than we can expect "just by chance."

And thus, we would need to complete the math part of the hypothesis test as follows:

$$Z = \frac{M-\mu}{\sigma_m} = \frac{110-100}{6} = 1.66$$

So, our hypothesis test results is a Z value = 1.66. We call this value (the 1.66) a test statistic. 1.66 is our Z-test statistic OUTCOME.

We now need to know whether this is a HIGH PROBABILITY OUTCOME or a LOW PROBABILITY OUTCOME.

If it is a low probability outcome, then we know that the results are better than "just by chance," and we conclude that there was an effect: our new math tutoring program DID have an effect on math ability. But we can be wrong by TYPE I ERROR.

If it is a high probability outcome, then we conclude that the results are no better than "just by chance," and we would conclude that the new tutoring method DID NOT have an effect on math ability. But we can also be wrong by TYPE II ERROR.

At this point, go to the VIDEOLECTURE. The video lecture will explain and demonstrate each step of the 4-Step Hypothesis Testing process, but it will be difficult to understand the video lecture if you do not have a good grasp of this lecture. The video lecture focuses on the 4 steps of hypothesis testing.

1. State the hypotheses (H_0 and H_1)
2. Set a critical region with an alpha level (α)
3. Do the math (largely explained in this lecture)
4. Make a decision (reject the null hypothesis or fail to reject the null hypothesis)