

# Concepts in Hypothesis Testing

Dr. Austin R Brown

Kennesaw State University

# Why Should I Care About Hypothesis Testing?

- ▶ You've probably read or heard things on the Internet where claims are made such as:
  - ▶ Coffee drinkers tend to live longer than non-coffee drinkers
  - ▶ Psilocybin tends to improve PTSD symptoms
  - ▶ Breast milk fed babies tend to be healthier than formula fed babies
- ▶ How do researchers come to these conclusions? How are they able to make these claims?
  - ▶ They are using empirical data to answer their research questions through the use of various methods, which include a general procedure called *hypothesis testing*

# Understanding Hypotheses

- ▶ We had talked earlier about two different types of hypotheses that we have which are part of the scientific method and research process:
1. **Research Hypothesis:** Our best guess as to the answer of a research question based on our prior knowledge, experience, or review or relevant literature.
  2. **Statistical Hypotheses:** A set of two complementary (read, opposite) hypotheses called the null and alternative hypotheses which map our context and research hypothesis to a quantitative/mathematical context.
    - ▶ The alternative hypothesis is a mathematical representation of the research hypothesis, sometimes denoted  $H_A$
    - ▶ The null hypothesis is the exact opposite of the alternative hypothesis, sometimes denoted  $H_0$

# Specifying the Hypotheses

- ▶ Sound a little confusing? Well, let's practice with some research scenarios!

## Specifying the Hypotheses: Examples

- ▶ Suppose we are examining the recovery timelines of division one college basketball players who have experienced ACL tears over the last 5 years. We hypothesize that, thanks to modern surgery techniques as well as improvements in rehabilitation methods, the mean amount of time (in days) that it takes an athlete to return to play is now less than 240 days (8 months).
- ▶ With this information, let's specify our research question, research hypothesis, and statistical hypotheses!

# Specifying the Hypotheses: Examples

- ▶ Specifying the Research Hypotheses:
- ▶ **Population of Interest:** Division 1 College Basketball Players who have experienced ACL tears over the past 5 years.
- ▶ **Variable of Interest:** Recovery time, measured in days (quantitative variable)

# Specifying the Hypotheses: Examples

- ▶ Characteristic of the Variable to be Analyzed: Mean
- ▶ General Analytical Method: Comparison

## Specifying the Hypotheses: Examples

- ▶ Now, the research hypothesis is: the mean amount of time (in days) that it takes an athlete to return to play is less than 240 days.
- ▶ Specifying the research hypothesis is important because it is the alternative hypothesis:

$$H_A : \mu < 240$$



## Specifying the Hypotheses: Examples

- ▶ The null hypothesis is the exact opposite:

$$H_0 : \mu \geq 240$$

- ▶ See? Not so bad! Let's consider another example.

## Specifying the Hypotheses: Examples

- ▶ Suppose I wanted to know if the mean amount of time spent studying per week (in hours) differs between STEM and non-STEM graduate majors at KSU. I hypothesize there is a difference.
- ▶ **Populations of Interest:** STEM graduate majors and non-STEM graduate majors at KSU
- ▶ **Variables of Interest:** (1) Major (categorical) (2) Time Spent Studying (quantitative)

# Specifying the Hypotheses: Examples

- ▶ Characteristic of the Variables to be Analyzed: Means
- ▶ General Analytical Method: Comparison
- ▶ Our research hypothesis is: the means will differ.

## Specifying the Hypotheses: Examples

- ▶ Thus, our statistical hypotheses are:

$$H_A : \mu_{\text{STEM}} \neq \mu_{\text{Non-STEM}}$$

$$H_0 : \mu_{\text{STEM}} = \mu_{\text{Non-STEM}}$$

# Steps in Hypothesis Testing

- ▶ As you can tell from these examples, the null and alternative hypotheses are dependent upon the type of question being asked.
- ▶ This is not an arbitrary difference. The way our hypotheses are set up dictates how we calculate *test statistics* which in turn allows us to draw conclusions.
- ▶ What in the world is a “test statistic?” More on that in a moment!

# Steps in Hypothesis Testing

- ▶ A test statistic is:

$$\frac{\text{What was observed in the data} - \text{What would be expected if } H_0 \text{ were true}}{\text{Standardizing Constant}}$$

- ▶ So if the data more strongly support the null hypothesis, we'd expect the difference in the numerator to be small and thus the value of the test statistic to be close to 0.
  - ▶ If the data more strongly support the alternative, we'd expect the opposite to be the case.

# Steps in Hypothesis Testing

► All hypothesis tests consist of three core components:

1. The hypotheses,  $H_0$  and  $H_A$ .
2. A test statistic
3. A *critical region*, which is sometimes also referred to as a *rejection region*.
  - This is an area in the domain of the sampling distribution of the test statistic (yes, test statistics also have sampling distributions) where, if the value of the test statistic happens to fall, we conclude that we have more evidence in favor of the alternative hypothesis.
  - If the value of the test statistic does not fall within the critical region, then we conclude that we have more evidence in favor of the null hypothesis.

# Steps in Hypothesis Testing

- ▶ Whenever we are performing a hypothesis test, its conclusion involves us, the researcher, making a decision. Based on the value of the test statistic, we either conclude that we have more *statistical evidence* in favor of  $H_0$  from the data or that we have more *statistical evidence* in favor of  $H_A$  from the sample data.
  - ▶ If the results of the test indicate that the data more strongly support  $H_A$ , then this implies the research hypothesis has substantial statistical support.
  - ▶ If the results of the test indicate that the data more strongly support  $H_0$ , then this implies the research hypothesis does not have substantial statistical support.
- ▶ Clearly though, because we are working with only a subset of the population (a sample), there is a possibility that the conclusion we draw may be incorrect.



# Hypothesis Test Errors

- ▶ We can think about hypothesis tests like a criminal court case. Here, the role of the prosecution is to provide evidence to the jury “beyond a reasonable doubt” that the defendant is in fact guilty of the crimes of which they are accused.
- ▶ At the end of the hearings, the jury deliberates and makes a decision as to the accused’s guilt. The defendant can either be determined to be guilty or not guilty.
- ▶ But of course, we know that juries have been known to make wrong decisions – either letting a guilty person go free or wrongly convicting an innocent person.

# Hypothesis Test Errors

- ▶ If the jury convicts a person who was actually innocent, they've made an error. In hypothesis testing, if we conclude that we have more evidence in favor of the alternative when in fact the null is actually true, we have made a **Type 1 Error**.
  - ▶ It is common to say that we *reject* the null hypothesis rather than saying we have more evidence in favor of the alternative.
- ▶ If a jury acquits a person who was actually guilty, they've again made an error. In hypothesis testing, if we conclude that we have more evidence in favor of the null hypothesis when the alternative is actually true, we have made a **Type 2 Error**.
  - ▶ It is common to say that we have *failed to reject* the null hypothesis rather than saying we have more evidence in favor of the null.

# Hypothesis Test Errors

- ▶ In any given hypothesis test, there is always some non-zero probability of making either one of these errors.
- ▶ The probability of making a Type 1 error is typically denoted as  $\alpha$ .
- ▶ The probability of making a Type 2 error is typically denoted as  $\beta$ .
- ▶ Note, both of these probabilities are conditional.  
 $P(\text{Reject } H_0 | H_0 \text{ True}) = \alpha$  and  
 $P(\text{Fail to Reject } H_0 | H_A \text{ True}) = \beta$ .

# Hypothesis Test Errors

- ▶ As you can imagine, we'd like to keep both of these probabilities as low as possible. So how do we do that?
- ▶ For a Type 1 error, it's pretty easy. The researcher decides on what value of  $\alpha$  they are comfortable with.  $\alpha = 0.05$  is traditionally used.
- ▶ Controlling Type 2 error is a little more tricky. Generally, we can do so by having a sufficiently large sample and by using the appropriate tests for the type of data we are working with.

# The P-Value Method of Hypothesis Testing

- ▶ Remember our critical region? The critical region is defined by  $\alpha$ !
- ▶  $\alpha$  gives us a critical value which partitions the domain of the sampling distribution into the critical and non-critical regions.
- ▶ If the value of our test statistic exceeds this critical value(s), we conclude that we have more evidence in favor of the alternative hypothesis (in other words, we reject  $H_0$ ).
- ▶ If the value of our test statistic does not exceed this critical value, we make the opposite conclusion which is commonly referred to as failing to reject  $H_0$ .

# The P-Value Method of Hypothesis Testing

- ▶ In real research, we typically do not compare the value of our test statistic to a critical value to draw conclusions about our statistical hypotheses.
- ▶ We instead use something called a “p-value” which is likely the most misunderstood concept in statistics and research.
- ▶ The p-value is a probability associated with the test statistic we calculate.
  - ▶ But what is the probability??

# The P-Value Method of Hypothesis Testing

- ▶ The p-value associated with a given hypothesis test is the probability of observing a greater valued test statistic in a world where the null hypothesis is true.
- ▶ If our p-value falls below our  $\alpha$  (i.e.,  $p < \alpha = 0.05$ ), then we say the data more strongly support  $H_A$ .
  - ▶ Otherwise, we conclude they more strongly support  $H_0$ .
- ▶ Why does this make sense?

# The P-Value Method of Hypothesis Testing

- ▶ If  $p = 0.01$ , for example, then this means if we were to replicate the same study say 10,000 times, the proportion of test statistics which had values greater than the one we observed in our sample would be about 0.01.
- ▶ Which is to say, it feels unlikely we would have observed what we did in a world where  $H_0$  is true.
  - ▶ Thus, the data seem to indicate we live in a world where  $H_A$  is true.
- ▶ But again, this doesn't mean we have proven  $H_A$  to be definitively true and  $H_0$  to be definitively false. We merely have evidence that might be the case, not that it is.



# Effect Sizes

- ▶ Recall, hypothesis tests tell us whether what we observe in the data are *statistically* meaningful.
- ▶ They don't always tell us if what we are observing in the data are *practically* or *contextually* meaningful.
- ▶ To this end, whenever we perform a hypothesis test, we also calculate a value called an *effect size*, which can be thought of as the quantification of practical meaning.
  - ▶ In other words, sometimes the result of our hypothesis test can indicate that we have statistical support for the research hypothesis, but we may not have practical support.
- ▶ Each hypothesis test we perform has an associated effect size.

# Effect Sizes

- ▶ At the conclusion of every study, we need to determine two things:
- ▶ Using the value of the test statistic and/or its associated p-value, do we have statistical support for the research hypothesis?
- ▶ Using the value of the effect size, do we have practical support for the research hypothesis?
- ▶ Both are important to know because they tell us different things!!