



# HYPOTHESIS TESTING

Frances Chumney, PhD



TRANSCRIPT:

This is Dr. Chumney. The focus of this lecture is hypothesis testing – both what it is, how hypothesis tests are used, and how to conduct hypothesis tests.

# CONTENT OUTLINE

- Logic of Hypothesis Testing
- Error & Alpha
- Hypothesis Tests
- Effect Size
- Statistical Power

HYPOTHESIS TESTING



TRANSCRIPT:


In this lecture, we will talk about both theoretical and applied concepts related to hypothesis testing.

*how we conceptualize hypotheses*

# LOGIC OF HYPOTHESIS TESTING

---

HYPOTHESIS TESTING



TRANSCRIPT:

Let's begin the lecture with a summary of the logic process that underlies hypothesis testing.

# HYPOTHESIS TESTING LOGIC

## ➤ Hypothesis Test

*statistical method that uses sample data to evaluate a hypothesis about a population*

## ➤ The Logic

- ❖ State a hypothesis about a population, usually concerning a population parameter
- ❖ Predict characteristics of a sample
- ❖ Obtain a random sample from the population
- ❖ Compare obtained data to prediction to see if they are consistent

HYPOTHESIS TESTING



### TRANSCRIPT:

It is often impossible or otherwise not feasible to collect data on every individual within a population. Therefore, researchers rely on samples to help answer questions about populations. Hypothesis testing is a statistical procedure that allows researchers to use sample data to draw inferences about the population of interest.

Hypothesis testing is one of the most commonly used inferential procedures. Hypothesis testing will combine many of the concepts we have already covered, including z-scores, probability, and the distribution of sample means.

To conduct a hypothesis test, we first state a hypothesis about a population, predict the characteristics of a sample of that population (that is, we predict that a sample will be representative of the population), obtain a sample, then collect data from that sample and analyze the data to see if it is consistent with our hypotheses.

# STEPS IN HYPOTHESIS TESTING

## ➤ Step 1: State the Hypotheses

### ❖ Null Hypothesis ( $H_0$ )

*in the general population there is no change, no difference, or no relationship;  
the independent variable will have no effect on the dependent variable*

#### ○ Example

- All dogs have four legs.
- There is no difference in the number of legs dogs have.

### ❖ Alternative Hypothesis ( $H_1$ )

*in the general population there is a change, a difference, or a relationship;  
the independent variable will have an effect on the dependent variable*

#### ○ Example

- 20% of dogs have only three legs.

HYPOTHESIS TESTING



## TRANSCRIPT:

The process of hypothesis testing begins by stating a hypothesis about the unknown population. Actually we state two opposing hypotheses.

The first hypothesis we state – the most important one – is the null hypothesis. The null hypothesis states that the treatment has no effect. In general the null hypothesis states that there is no change, no difference, no effect, and otherwise no relationship between the independent and dependent variables. Because we are hypothesizing that nothing is happening, it is called the **null** hypothesis.

The second hypothesis we state is the alternative hypothesis. The alternative hypothesis is the opposite of the null, and states that the treatment does have an effect on the dependent variable.

## STEP 1: STATE THE HYPOTHESES (EXAMPLE)

### ➤ Example

### How to Ace a Statistics Exam

little known facts about  
the positive impact of  
alcohol on memory  
during “cram” sessions

HYPOTHESIS TESTING



#### TRANSCRIPT:

Here is an example that we will use throughout this lecture. Let's suppose a researcher conducts a study in which half of the participants are provided with a brochure on this topic. The researchers observe, then, the amount of alcohol consumed by participants the night before a statistics exam.

## STEP 1: STATE THE HYPOTHESES (EXAMPLE)

- Dependent Variable
  - ❖ Amount of alcohol consumed the night before a statistics exam
- Independent/Treatment Variable
  - ❖ Intervention: Pamphlet (treatment group) or No Pamphlet (control group)
- Null Hypothesis ( $H_0$ )
  - ❖ No difference in alcohol consumption between the two groups the night before a statistics exam.
- Alternative Hypothesis ( $H_1$ )
  - ❖ The treatment group will consume more alcohol than the control group.

HYPOTHESIS TESTING



### TRANSCRIPT:

The independent variable in this research is whether or not a participant was provided the “informational” pamphlet to read. The dependent variable is the amount of alcohol consumed by each student the night before the exam.

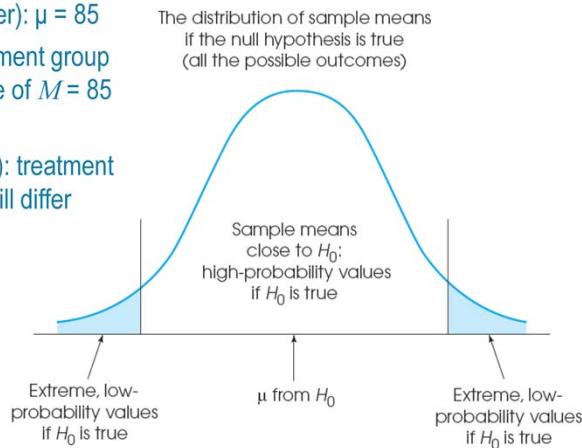
The null hypothesis for this research project might be that there is no difference in alcohol consumption between the two groups the night before a statistics exam.

The alternative hypothesis for this research project might be that the treatment group – that is, participants who were provided a copy of the pamphlet – will consume more alcohol.

## STEP 2: SET CRITERIA FOR DECISION

### ➤ Example

- ❖ Exam 1 (Previous Semester):  $\mu = 85$
- ❖ Null Hypothesis ( $H_0$ ): treatment group will have mean exam score of  $M = 85$  ( $\sigma = 8$ )
- ❖ Alternative Hypothesis ( $H_1$ ): treatment group mean exam score will differ from  $M = 85$



HYPOTHESIS TESTING



### TRANSCRIPT:

Data collected from samples are eventually used to evaluate the credibility of null hypotheses. The data will either support the null hypothesis or refute it. If there is a big discrepancy between the data and the hypothesis, we conclude that the hypothesis is wrong.

To formalize the decision process, we use the null hypothesis to predict the kind of sample mean that out to be obtained from our sample. More specifically, we determine exactly what sample means are consistent with the null hypothesis, and what sample means are at odds with the null hypothesis.

In this example, the null hypothesis states that the treatment has no effect and the population mean is still 85. If this is true, then the sample mean should have a value around 85. Therefore, a sample mean near 85 is consistent with the null hypothesis.

On the other hand, a sample mean that is very different from 85 is not consistent with the null hypothesis. To determine exactly what values are “near” 85 and values are “very different from” 85, we have to examine all of the possible sample means that could be obtained if the null hypothesis is true. In this figure, we have a distribution that represent possible mean exam scores for our example. According to the null hypothesis, this distribution is centered at mean of 85.

The distribution is divided into two sections:

1. Sample means that are likely to be obtained if the null hypothesis is true
2. Sample means that are very unlikely to be obtained if the null hypothesis is true (values that are very different from the null hypothesis)

When we examine this distribution, something we always have to keep in mind is that high-probability samples are located in the center of the distribution and have sample means close to the value specified in the null hypothesis. And the low-probability samples are located in the extreme tails.

Once the distribution is divided into these sections, we compare our sample data with the values in the distribution to determine whether our sample mean is consistent with the null hypothesis.

## STEP 2: SET CRITERIA FOR DECISION

### ➤ Alpha Level/Level of Significance

*probability value used to define the (unlikely) sample outcomes if the null hypothesis is true; e.g.,  $\alpha = .05$ ,  $\alpha = .01$ ,  $\alpha = .001$*

### ➤ Critical Region

*extreme sample values that are very unlikely to be obtained if the null hypothesis is true*

- ❖ Boundaries determined by alpha level
- ❖ If sample data falls within this region (the shaded tails), reject the null hypothesis

HYPOTHESIS TESTING



#### TRANSCRIPT:

To find the boundaries that separate the high-probability samples from the low-probability samples, we must define exactly what we mean by “low” and “high” probability. We do this by selecting a probability value – which we call the alpha level or level of significance – for the hypothesis test.

The alpha level is a small value that is used to identify the low-probability samples. Most of the time, we set alpha to be .05 (5%), but we sometimes use .01 (1%) and .001 (.1%).

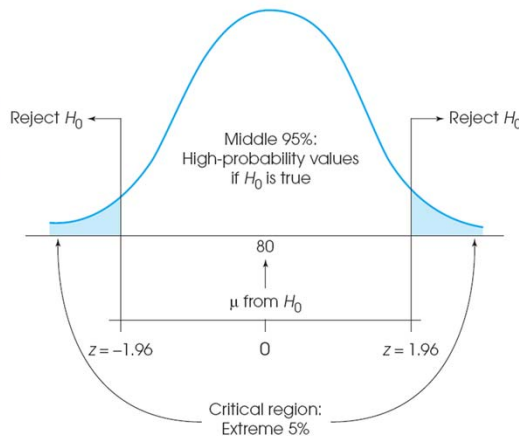
If we select an alpha level of .05, that tells us that the most extreme 5% of the scores are the ones we consider the least likely to occur.

The extreme unlikely values identified by the alpha level create the critical region. These values are in the outer-most tails of the distribution and identify outcomes that are not consistent with the null hypothesis. That means that when the null hypothesis is true – when there is no difference or change – the values that fall in these extreme tails are very unlikely to occur.

## STEP 2: SET CRITERIA FOR DECISION

### ➤ Critical Region Boundaries

- ❖ Assume normal distribution
- ❖ Alpha Level + Unit Normal Table
- ❖ Example: if  $\alpha = .05$ , boundaries extreme 5%
  - 2.5% in each tail (2-tailed)



HYPOTHESIS TESTING



#### TRANSCRIPT:

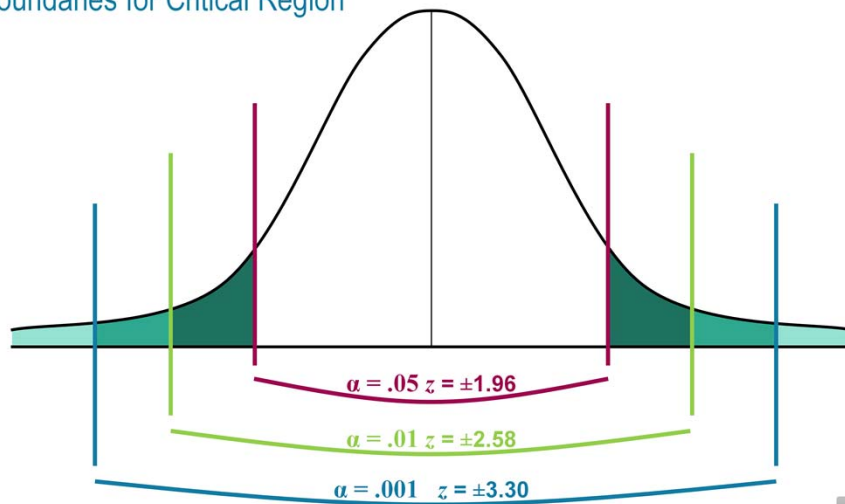
To determine the exact location for the boundaries that define the critical region, we use the alpha-level probability and the unit normal table. We assume a normal distribution of sample means and the unit normal table tells us exactly where the boundaries of the critical region fall.

If we set alpha to be .05, we are saying that we want to separate the most extreme 5% of the scores from the middle 95%. Because a normal distribution is symmetrical, this means we have to split that 5% between the two extreme ends, marking of 2.5% on both end of the distribution – the shaded parts in this figure. 2.5% is equal to a proportion of 0.0250.

In the unit normal table, if we look up a proportion of 0.0250 in Column C (the tail), we find the z-score is  $z = 1.96$ . So for any normal distribution, the extreme 5% split between the tails of the distribution is all scores greater than  $z = 1.96$  and less than  $z = -1.96$ .

## STEP 2: SET CRITERIA FOR DECISION

### ➤ Boundaries for Critical Region



TRANSCRIPT:

And here, we see where those critical boundaries fall for alpha levels of .05, .01, and .001.

We can look for these values in the unit normal table and see exactly where these values come from.

For example #2, we are going to go with a critical value of  $\alpha = .001$  b/c we want to be extra confident in our results before we start advertising them just because the intervention might be considered controversial or something.

### STEP 3: COLLECT, COMPUTE

- Collect data
- Compute sample mean
- Transform sample mean  $M$  to  $z$ -score

$$z = \frac{M - \mu}{\sigma_M}$$

- Example #2

$$z = \frac{95 - 85}{1.13} = \frac{10}{1.13} = 8.85$$

HYPOTHESIS TESTING



#### TRANSCRIPT:

The third step in hypothesis testing is the collection of data and computation of sample statistics. For this step, we select the number of participants we are going to use, and we either measure them on some characteristic, or we expose them to the treatment or control, or whatever it is we are using based on the design of our experiment.

In the alcohol and stats exam example from a few slides ago, we would select maybe 50 statistics students, give them the information pamphlet, and then wait for their next statistics exam. We would then have our data – their exam scores.

We calculate the mean of the exam scores, and we compare it to the population mean specified in our null hypothesis. Just for this example, we are going to assume a mean of 95.

## STEP 4: MAKE A DECISION

- Compare  $z$ -score with boundary of critical region for selected level of significance
- If...
  - ❖  $z$ -score falls in the tails, our mean is significantly different from  $H_0$ 
    - Reject  $H_0$
  - ❖  $z$ -score falls between the tails, our mean is not significantly different from  $H_0$ 
    - Fail to reject  $H_0$

HYPOTHESIS TESTING



### TRANSCRIPT:

Once we have our  $z$ -score we must compare it to the boundary of a critical region to determine whether to reject the null hypothesis. If the  $z$ -score falls into the tails, the mean is significantly different, and the null hypothesis should be rejected. If, however, the  $z$ -score falls between the tails of the distribution, we can conclude that the mean is not significantly different, and we would fail to reject the null hypothesis.

## HYPOTHESIS TESTING: **AN EXAMPLE (2-TAIL)**

### ➤ How to Ace a Statistics Exam...

- ❖ Population:  $\mu = 85$ ,  $\sigma = 8$
- ❖ Hypotheses
  - $H_0$ : Sample mean will not differ from  $M = 85$
  - $H_1$ : Sample mean will differ from  $M = 85$
- ❖ Set Criteria (Significance Level/Alpha Level)
  - $\alpha = .05$

HYPOTHESIS TESTING



### TRANSCRIPT:

Returning to our example – let's work through this process with a specific example. Let's suppose that we know our population mean is 85 and the population standard deviation is 8.

Our null hypothesis is that the sample mean will not differ from the population mean – that is, we expect our sample mean to be 85 because that is the value of the population mean. Our alternative hypothesis, then, is that our sample mean will be different from 85.

Then, we have to determine the significance level or alpha level for our test. For the purposes of this example, we will assume an alpha level of 0.05.

## HYPOTHESIS TESTING: **EXAMPLE (2-TAIL)**

### ➤ How to Ace a Statistics Exam...

#### ❖ Collect Data & Compute Statistics

- Intervention to 9 students
- Mean exam score,  $M = 90$

$$\sigma_M = \frac{\sigma}{\sqrt{9}} = \frac{8}{3} = 2.67$$

$$z = \frac{90 - 85}{2.67} = \frac{5}{2.67} = 1.87$$

HYPOTHESIS TESTING



TRANSCRIPT:

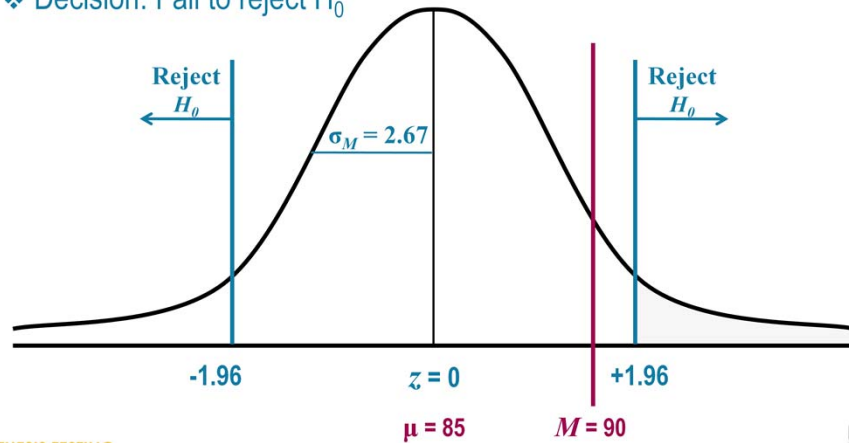
Suppose we conduct this research study with a sample of 9 individuals who read the brochure, and we find that the mean exam score for those persons is 90.

The first value we should calculate is the standard error of the mean, for which we plug in the standard deviation and sample size numbers provided. Then, we use that value to calculate a z-score.

## HYPOTHESIS TESTING: EXAMPLE (2-TAIL)

### ➤ How to Ace a Statistics Exam...

❖ Decision: Fail to reject  $H_0$



TRANSCRIPT:

This figure shows a normal distribution with the mean marked, as well as one standard deviation above and below the mean. If our sample mean falls further from the center of the distribution than + or minus 1.96, it will fall into one of the two extremes of the distribution. In this case, we would reject the null hypothesis. As we can see in this illustration, the sample mean falls between the two tails, which indicates that it is not significantly different from the population mean, so we would fail to reject the null hypothesis.

## REVISITING Z-SCORE STATISTICS

### ➤ A Test Statistic

- ❖ Single, specific statistic
- ❖ Calculated from the sample data
- ❖ Used to test  $H_0$

### ➤ Rule of Thumb...

- ❖ Large values of  $z$ 
  - Sample data probably DID NOT occur by chance – result of IV
- ❖ Small values of  $z$ 
  - Sample data probably DID occur by chance – not result of IV

HYPOTHESIS TESTING



### TRANSCRIPT:

When we talk about z-scores, we need to keep in mind that they do serve as a test statistic – a single, specific statistic created from the sample data that we can use to test our null hypothesis.

As a general rule of thumb, large values of  $z$  statistics indicate a large discrepancy between the sample data and the null hypothesis. In this situation, sample data are very unlikely to have occurred by chance alone. Therefore, it is due to our independent variable, whatever that may be.

And in contrast, small values of the  $z$  statistic indicate a small discrepancy between the sample data and the null hypothesis. In these situations, the sample data could have occurred by chance alone, they are close to the population mean, and so we conclude that our independent variable – whatever it is – had no effect on the sample data we collected.

*uncertainty leads to error*

## ERROR & ALPHA

---

HYPOTHESIS TESTING



TRANSCRIPT:

Because hypothesis testing is an inferential process, there is quite a bit of room for error to occur. Alpha is a pre-determined value that represents how wrong the researcher is willing to be.

# UNCERTAINTY & ERROR

## ➤ Hypothesis Testing = Inferential Process

- ❖ LOTS of room for error

## ➤ Types of Error

- ❖ Type I Error
- ❖ Type II Error

HYPOTHESIS TESTING



### TRANSCRIPT:

Hypothesis testing is an inferential process.... We are using limited information obtained from a sample as the basis for a general conclusion. Because we are making such generalizations, there is always the possibility that we will draw the wrong conclusion because sample data is not always exactly like the population. Remember – that is what error is.

There are two kinds of error that we are going to deal with in this chapter: Type I Errors and Type II Errors, and we are going to talk about each of them.

# TYPE 1 ERRORS

*error that occurs when the null hypothesis is rejected even though it is really true; the researcher identifies a treatment effect that does not really exist (a false positive)*

- Common Cause & Biggest Problem
  - ❖ Sample data are misleading due to sampling error
  - ❖ Significant difference reported in literature even though it isn't real
- Type I Errors & Alpha Level
  - ❖ Alpha level = probability of committing a Type I Error
  - ❖ Lower alphas = less chances of Type I Error

HYPOTHESIS TESTING



## TRANSCRIPT:

It is possible that the data will lead you to reject the null hypothesis – that is, determine that there is an effect – even when there really is no effect. This can happen because an effect is observed in the sample that does not exist in the population. This is called a Type I error, and is also sometimes called a false positive.

The biggest problem with Type I errors is that they may result in literature or even practices based on the expectation or acceptance that a particular treatment is effective when it is not.

The alpha level or probability level for a hypothesis test is the probability that the test will lead to a Type I error. Essentially, if we use the standard alpha level of 0.05, we are saying that we are accepting that there is a 5% chance that any significant effects we find are not significant at all.

## TYPE II ERRORS

*error that occurs when the null hypothesis is not rejected even it is really false; the researcher does not identify a treatment effect that really exists (a false negative)*

- Common Cause & Biggest Problem
  - ❖ Sample mean is not in critical region even though there is a treatment effect
  - ❖ Overlook effectiveness of interventions
- Type II Errors & Probability
  - ❖  $\beta$  = probability of committing a Type II Error

HYPOTHESIS TESTING



### TRANSCRIPT:

When an intervention has an effect but the statistical test does not detect it, a Type II error has occurred. This can happen when we observe effects that are too small to be statistically significant even though they are meaningful in practice. Beta is the probability of committing a Type II error.

The biggest danger of a Type II error is that it prevents us from recognizing when an intervention works.

# TYPE I & TYPE II ERRORS

## ➤ Experimenter's Decision

	Actual Situation	
	No Effect, $H_0$ True	Effect Exists, $H_0$ False
Reject $H_0$	Type I Error	😊
Retain $H_0$	😊	Type II Error

HYPOTHESIS TESTING



### TRANSCRIPT:

This table presents the possible outcomes of any decision you could make regarding a hypothesis based on the results of a statistical test. If you choose to reject the null hypothesis and conclude that there is an effect when there really is no effect in the population, a Type I error has been committed. When you choose to not reject the null hypothesis and conclude that there is no effect when there really is an effect in the population, a Type II error has been committed.

# SELECTING AN ALPHA LEVEL

- Functions of Alpha Level
  - ❖ Critical region boundaries
  - ❖ Probability of a Type I error
- Primary Concern in Alpha Selection
  - ❖ Minimize risk of Type I Error without maximizing risk of Type II Error
- Common Alpha Levels
  - ❖  $\alpha = .05$ ,  $\alpha = .01$ ,  $\alpha = .001$

HYPOTHESIS TESTING



## TRANSCRIPT:

The alpha level selected for a statistical hypothesis test very important. Not only does the alpha level indicate the chance the researcher is willing to take for committing a Type I error, but the alpha level also sets the boundaries of a critical region.


A critical region – as a reminder – is an area or areas of the space under the curve. In instances where the sample mean falls into the critical region or regions, the null hypothesis should be rejected.

*testing null hypotheses*

# HYPOTHESIS TESTS

---

HYPOTHESIS TESTING



TRANSCRIPT:

You should now understand the basic principles that underscore hypothesis testing, so we will begin talking about the practice of conducting hypothesis tests.

## HYPOTHESIS TESTS: INFLUENTIAL FACTORS

- Magnitude of difference between sample mean and population mean (in z-score formula, larger difference  $\Rightarrow$  larger numerator)

$$z = \frac{M - \mu}{\sigma_M} \qquad \sigma_M = \frac{\sigma}{\sqrt{n}}$$

- Variability of scores (influences  $\sigma_M$ ; more variability  $\Rightarrow$  larger  $\sigma_M$ )
- Sample size (influences  $\sigma_M$ ; larger sample size  $\Rightarrow$  smaller  $\sigma_M$ )

HYPOTHESIS TESTING



### TRANSCRIPT:

There are some characteristics that can influence the results of a hypothesis test more significantly than others. Specifically, a hypothesis test will be influenced by the size of the difference between a sample mean and the population mean. Also, a sample statistic with a large standard deviation indicating more variability will lead to larger expected values of  $M$ . Finally sample size could impact the results of a hypothesis test, as smaller samples lead to larger expected values of  $M$ .

## HYPOTHESIS TESTS: **ASSUMPTIONS**

- Random Sampling
- Independent Observations
- Value of  $\sigma$  is Constant
  - ❖ Despite treatment
- Normal sampling distribution

HYPOTHESIS TESTING

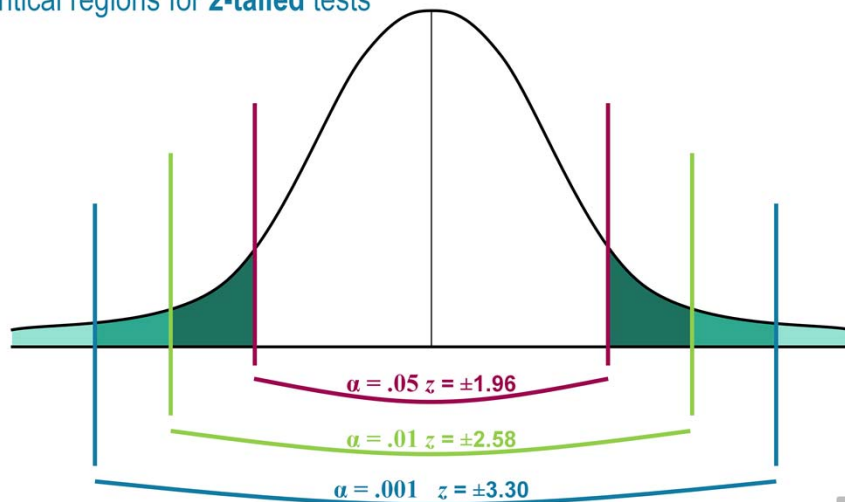


### TRANSCRIPT:

There are some assumptions inherent to hypothesis testing. Specifically, hypothesis testing assumes a random sample, independent observations, a constant value for standard deviation, and a normal sampling distribution. If any of these assumptions are violated, the risk of making an error increases.

## NON-DIRECTIONAL HYPOTHESIS TESTS

### ➤ Critical regions for 2-tailed tests



TRANSCRIPT:

To this point, we have only talked about non-directional hypothesis tests, which are also referred to as two-tailed tests. The reason we call them two-tailed tests is that the critical region occurs in both tails of the distribution.

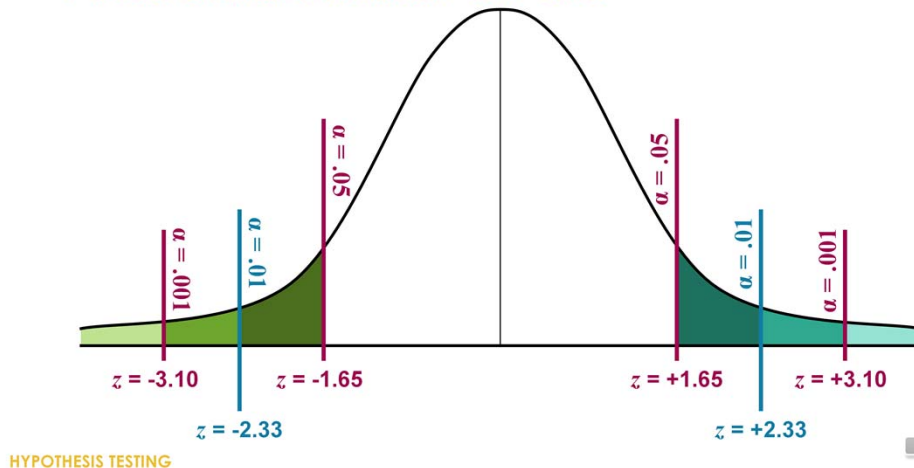
Here we have critical regions marked at both ends of the distribution at the .05, .01, and .001 alpha levels. The areas that are shaded in this figure represent the critical areas, which means that if the sample mean were to fall into one of the colored areas, the null hypothesis might be rejected.

It is a good idea to just make a note of the z-scores specified on this slide as being the boundaries for the various critical regions based on various levels of alpha and a two-tailed hypothesis test.

# DIRECTIONAL HYPOTHESIS TESTS

## ➤ Critical regions for 1-tailed tests

❖ Blue **or** Green tail of distribution – **NOT BOTH**



TRANSCRIPT:

In a 1-tailed hypothesis test (also called a directional test), our null hypothesis specifies which end of the distribution we are interested in. This means that we acknowledge critical regions at one end of the distribution or the other, but not both ends at the same time. Notice that the z scores used to define the critical regions have changed. For example, the z-score that marks the boundary of the critical region for a 2-tailed test is 1.96, but is 1.65 for a one-tailed test. The reason is that a two-tailed test takes the same amount of space under the curve and splits it to both ends of the distribution, and that changes where the critical area boundaries are located within the distribution.

Again, I recommend making a note of the z-scores specified on this slide as being the boundaries for the various critical regions based on various levels of alpha and a one-tailed hypothesis test.

# ALTERNATIVE HYPOTHESES

- Alternative Hypotheses for **2-tailed** tests
  - ❖ Do not specify direction of difference
  - ❖ Do not hypothesize whether sample mean should be lower or higher than population mean
  
- Alternative Hypotheses for **1-tailed** tests
  - ❖ Specify a difference
  - ❖ Hypothesis specifies whether sample mean should be lower or higher than population mean

HYPOTHESIS TESTING



## TRANSCRIPT:

Two-tailed alternative hypotheses are easy to spot because they specify that there is a difference, but they do not say specifically the direction of the difference. That is, they do not specify whether the sample mean is higher or lower than the population mean.

One-tailed alternative hypotheses specify that there is a difference, but also specify the direction of the difference. That is, the alternative hypothesis indicates whether the sample mean should be higher or lower than the population mean.

# NULL HYPOTHESES

- Null Hypotheses for **2-tailed** tests
  - ❖ Specify no difference between sample & population
  
- Null Hypotheses for **1-tailed** tests
  - ❖ Specify the opposite of the alternative hypothesis
  - ❖ Example #2
    - $H_0: \mu \leq 85$  (There is no increase in test scores.)
    - $H_1: \mu > 85$  (There is an increase in test scores.)

HYPOTHESIS TESTING



## TRANSCRIPT:

A null hypothesis for a two-tailed test specifies that there is no difference between a sample and a population.

A null hypothesis for a one-tailed test is not as neutral. Instead, it specifies the opposite of the alternative hypothesis for a one-tailed test.

In the example here, our  $H_0$  or null hypothesis is that there is no increase in test scores. The alternative hypothesis is that there is an increase in test scores. Because the alternative hypothesis specifies the direction of an expected change (increase), this would require a one-tailed hypothesis test.

## HYPOTHESIS TESTS: AN EXAMPLE (1-TAIL)

### ➤ How to Ace a Statistics Exam...

- ❖ Population:  $\mu = 85$ ,  $\sigma = 8$
- ❖ Hypotheses
  - $H_0$ : Sample mean will be less than or equal to  $M = 85$
  - $H_1$ : Sample mean be greater than  $M = 85$
- ❖ Set Criteria (Significance Level/Alpha Level)
  - $\alpha = .05$

HYPOTHESIS TESTING



### TRANSCRIPT:

Here is an example of a one-tailed hypothesis test. This goes back to our example of a researcher trying to convince students that drinking will increase retention of statistics information.

We know that the population mean is 85 and the population standard deviation is 8. The null hypothesis is that the sample mean will be less than or equal to 85. The alternative hypothesis is that the sample mean will be greater than 85. Suppose we decide to use an alpha level where the probability of making an incorrect decision is 0.05.

## HYPOTHESIS TESTS: AN EXAMPLE (1-TAIL)

### ➤ How to Ace a Statistics Exam...

#### ❖ Collect Data & Compute Statistics

- Intervention to 9 students
- Mean exam score,  $M = 90$

$$\sigma_M = \frac{\sigma}{\sqrt{9}} = \frac{8}{3} = 2.67$$

$$z = \frac{M - \mu}{\sigma_M} = \frac{90 - 85}{2.67} = \frac{5}{2.67} = 1.87$$

HYPOTHESIS TESTING



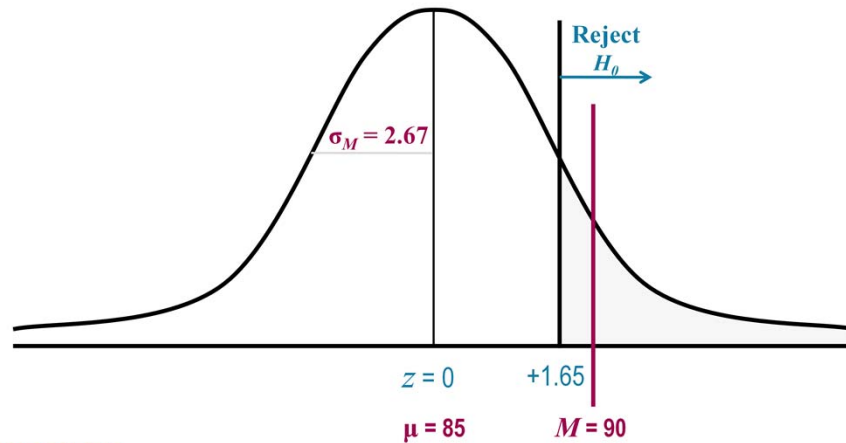
TRANSCRIPT:

Once we have this information gathered together, we are ready to collect our data and let's suppose that our participants earned a mean score of 90 on the exam. To calculate the sigma-sub-m, we divide the standard deviation of the population by the square root of the sample size. Then we convert our observed mean of 90 into a z-score.

## HYPOTHESIS TESTS: AN EXAMPLE (1-TAIL)

➤ How to Ace a Statistics Exam...

❖ Decision: Reject  $H_0$



TRANSCRIPT:

The last step is to decide whether or not to reject the null hypothesis. In this figure, we see that the boundary for the critical region for this one-tailed test has been marked at  $z=1.65$ . This means that any value that falls further above the mean than that particular point falls into the critical region, and the decision to reject the null hypothesis should be made.

Remember that on the last slide, we calculated a z-score for our sample mean of 90 to be 1.87. When we compare  $z=1.87$  to  $z=1.65$ , we see that our z-score falls into the critical region, and this means we should reject the null hypothesis.

*estimating the magnitude of an effect*

## EFFECT SIZE



HYPOTHESIS TESTING 

TRANSCRIPT:

A hypothesis test can help us to determine if an effect is statistically significant, but it cannot tell us anything about the size of that effect, or whether that effect is likely to be strong enough to make a difference in a real-world or clinical setting. For help with interpreting the results of a hypothesis test in this way, we rely on effect sizes.

## EFFECT SIZE

- Problem with hypothesis testing
  - ❖ Significance  $\neq$  Meaningful/Important/Big Effect
    - Significance is relative comparison: treatment effect compared to standard error
- Effect Size
  - statistic that describes the magnitude of an effect*
- Measures size of treatment effect in terms of (population) standard deviation

HYPOTHESIS TESTING



### TRANSCRIPT:

An effect size is a statistic that describes the magnitude of an effect. Effect sizes are sometimes interpreted as indicators of the “real-world” or practical effects of a treatment. This is an important concept, especially in the case of large samples. As a general rule of thumb, the larger a sample is, the more likely you are to find statistical differences.

## EFFECT SIZE: COHEN'S $D$

### ➤ Not influenced by sample size

$$\text{Cohen's } d = \frac{\text{mean difference}}{\text{standard deviation}}$$

### ➤ Evaluating Cohen's $d$

- ❖  $d = 0.2$  – Small Effect (mean difference  $\approx 0.2$  standard deviation)
- ❖  $d = 0.5$  – Medium Effect (mean difference  $\approx 0.5$  standard deviation)
- ❖  $d = 0.8$  – Large Effect (mean difference  $\approx 0.8$  standard deviation)

### ➤ Calculated the same for 1-tailed and 2-tailed tests

HYPOTHESIS TESTING



#### TRANSCRIPT:

One of the most common effect sizes is Cohen's  $d$ . Cohen's  $d$  is a reliable effect size because it is not influenced by the sample size. Whereas a hypothesis test based on large samples will increase the odds of rejecting the null hypothesis and claiming a difference when one really does not exist, effect sizes cannot help researchers to make determinations about hypothesis tests or hypotheses, but they do help describe whether significant effects might actually be meaningful.

Cohen's  $d$  is calculated as the ratio between the mean difference and the standard deviation.

Cohen's  $d$  values cannot be negative, and there is no limit to how large they can be. An effect size between 0 and 0.2 is considered a small effect; an effect size between 0.2 and 0.5 is considered a medium or moderate effect, and an effect size larger than 0.5 is considered large – especially if that effect size approaches or exceeds 0.8. It is important to note that these are general guidelines and not concrete rules. While these guidelines are generally accepted within the scientific community as the cutoff scores for interpreting Cohen's  $d$ , they are not based on much empirical evidence.


Finally, one last note on Cohen's  $d$  effect size – the effect size is not dependent on or otherwise influenced by whether a hypothesis test is 1-tailed or 2-tailed; the calculation is the same either way.

*probability of correctly rejecting a false null hypothesis*

# STATISTICAL POWER

---

HYPOTHESIS TESTING



TRANSCRIPT:

And finally, a quick overview of the concept of statistical power.

# STATISTICAL POWER

*the probability of correctly rejecting a null hypothesis when it is not true; the probability that a hypothesis test will identify a treatment effect when if one really exists*

- A priori
  - ❖ Calculate power before collecting data
  - ❖ Determine probability of finding treatment effect
- Power is influenced by...
  - ❖ Sample size
  - ❖ Expected effect size
  - ❖ Significance level for hypothesis test ( $\alpha$ )

HYPOTHESIS TESTING



## TRANSCRIPT:

The power of a statistical test is the probability that the test will correctly reject a false null hypothesis. As the definition implies, the more powerful a statistical test is, the more readily it will detect a treatment effect when one really exists.

Power is influenced by sample size – that is, a larger sample will better represent the population, and the power of the statistical test will increase as a result.

The effect size is the magnitude of an effect. As the effect size or magnitude of the effect increases, so does the likelihood of detecting it, which is essentially an increase in power.

Changing the significance level or alpha level of a test can decrease its power. Because lowering the alpha level reduces the chances that significant effects will be found, doing so translates to a loss of power to identify meaningful change that is happening.