Module 5: Introduction to Hypothesis Testing

Jon Starkweather, PhD
jonathan.starkweather@unt.edu
Consultant
Research and Statistical Support

UNT UNIVERSITY OF NORTH TEXAS
Discover the power of ideas.

Introduction to Statistics for the Social Sciences

RSS
Research and Statistical Support

The RSS short courses

The Research and Statistical Support (RSS) office at the University of North Texas hosts a number of “Short Courses”. A list of them is available at:

http://www.unt.edu/rss/Instructional.htm

Outline

Outline

Outline

Outline

RSS
Research and Statistical Support

Outline

Outline

RSS
Research and Statistical Support

Outline

RSS
Research and Statistical Support

Hypothesis Testing

Hypothesis Testing

- Hypothesis testing is a systematic procedure for deciding whether the results of a research study, which examines a sample, support a particular theory or practical innovation, which applies to a population.

Hypothesis Testing

- Hypothesis testing is a systematic procedure for deciding whether the results of a research study, which examines a sample, support a particular theory or practical innovation, which applies to a population.
- The type of hypothesis testing done in much of the social sciences is called Null Hypothesis Significance Testing (NHST).

RSS Research and Statistical Support

Decisions, Decisions...

- What is Null Hypothesis Significance Testing all about?

Research and Statistical Support

Decisions, Decisions...

- What is Null Hypothesis Significance Testing all about?
- Use an inferential procedure to examine the credibility of a hypothesis about a population based on the probability of sample data.

RSS
Research and Statistical Support

Decisions, Decisions...

- What is Null Hypothesis Significance Testing all about?
- Use an inferential procedure to examine the credibility of a hypothesis about a population based on the probability of sample data.
  - Notice; NHST is not about testing the probability of a hypothesis or theory!

RSS
Research and Statistical Support

Decisions, Decisions...

- What is Null Hypothesis Significance Testing all about?
- Use an inferential procedure to examine the credibility of a hypothesis about a population based on the probability of sample data.
  - Notice; NHST is not about testing the probability of a hypothesis or theory!
- NHST gets its name from the use of the Null hypothesis.

Decisions, Decisions...

- What is Null Hypothesis Significance Testing all about?
- Use an inferential procedure to examine the credibility of a hypothesis about a population based on the probability of sample data.
  - Notice; NHST is not about testing the probability of a hypothesis or theory!
- NHST gets its name from the use of the Null hypothesis.
  - Originally referred to a hypothesis of *no* difference.

RSS Research and Statistical Support

Decisions, Decisions...

- What is Null Hypothesis Significance Testing all about?
- Use an inferential procedure to examine the credibility of a hypothesis about a population based on the probability of sample data.
  - Notice; NHST is not about testing the probability of a hypothesis or theory!
- NHST gets its name from the use of the Null hypothesis.
  - Originally referred to a hypothesis of *no* difference.
  - Modern interpretation and use allows more precise specification.

RSS
Research and Statistical Support

Symbols of NHST

- Typically we use population symbols when expressing hypotheses.

Symbols of NHST

- Typically we use population symbols when expressing hypotheses.
- The Null hypothesis uses the symbol H_0 and is read as ‘H-oh’ or ‘H-naught’.

Symbols of NHST

- Typically we use population symbols when expressing hypotheses.
- The Null hypothesis uses the symbol H_0 and is read as 'H-oh' or 'H-naught'.
    - Example Null hypothesis in symbols: $H_0 : \mu_1 = \mu_2$

RSS
Research and Statistical Support

Symbols of NHST

- Typically we use population symbols when expressing hypotheses.
- The Null hypothesis uses the symbol H_0 and is read as ‘H-oh’ or ‘H-naught’.
    - Example Null hypothesis in symbols: $H_0 : \mu_1 = \mu_2$
    - Which is read as, we hypothesize the mean of population 1 is equal to the mean of population 2.

Symbols of NHST

- Typically we use population symbols when expressing hypotheses.
- The Null hypothesis uses the symbol H_0 and is read as ‘H-oh’ or ‘H-naught’.
  - Example Null hypothesis in symbols: $H_0 : \mu_1 = \mu_2$
  - Which is read as, we hypothesize the mean of population 1 is equal to the mean of population 2.
- The alternative hypothesis uses the symbol H_1 and is read as ‘H-one’.

Symbols of NHST

- Typically we use population symbols when expressing hypotheses.
- The Null hypothesis uses the symbol H_0 and is read as ‘H-oh’ or ‘H-naught’.
  - Example Null hypothesis in symbols: $H_0 : \mu_1 = \mu_2$
  - Which is read as, we hypothesize the mean of population 1 is equal to the mean of population 2.
- The alternative hypothesis uses the symbol H_1 and is read as ‘H-one’.
  - Example alternative hypothesis in symbols: $H_1 : \mu_1 \neq \mu_2$

Symbols of NHST

- Typically we use population symbols when expressing hypotheses.
- The Null hypothesis uses the symbol H_0 and is read as 'H-oh' or 'H-naught'.
  - Example Null hypothesis in symbols: $H_0 : \mu_1 = \mu_2$
  - Which is read as, we hypothesize the mean of population 1 is equal to the mean of population 2.
- The alternative hypothesis uses the symbol H_1 and is read as 'H-one'.
  - Example alternative hypothesis in symbols: $H_1 : \mu_1 \neq \mu_2$
  - Which is read as, we hypothesize the mean of population 1 is not equal to the mean of population 2.

RSS
Research and Statistical Support

Steps of NHST (one version among many)

1. Define the populations of interest and restate the research question as null and alternative hypotheses about the populations.

$$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 \neq \mu_2$$

Steps of NHST (one version among many)

1. Define the populations of interest and restate the research question as null and alternative hypotheses about the populations.

   $$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 \neq \mu_2$$

2. Determine the characteristics of the comparison distribution.

RSS Research and Statistical Support

Steps of NHST (one version among many)

1. Define the populations of interest and restate the research question as null and alternative hypotheses about the populations.

$$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 \neq \mu_2$$

2. Determine the characteristics of the comparison distribution.
3. Determine the cutoff sample score on the comparison distribution at which the null should be rejected (typically associated with $p = .05$).

Steps of NHST (one version among many)

1. Define the populations of interest and restate the research question as null and alternative hypotheses about the populations.

$$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 \neq \mu_2$$

2. Determine the characteristics of the comparison distribution.
3. Determine the cutoff sample score on the comparison distribution at which the null should be rejected (typically associated with $p = .05$).
4. Determine your sample's score on the comparison distribution (i.e., compute your sample statistic)

RSS
Research and Statistical Support

Steps of NHST (one version among many)

1. Define the populations of interest and restate the research question as null and alternative hypotheses about the populations.

$$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 \neq \mu_2$$

2. Determine the characteristics of the comparison distribution.
3. Determine the cutoff sample score on the comparison distribution at which the null should be rejected (typically associated with $p = .05$).
4. Determine your sample's score on the comparison distribution (i.e., compute your sample statistic)
5. Compare and make a decision. To reject or not to reject? That is the question!

RSS
Research and Statistical Support

Logic of NHST

- If we believe something to be different, why do we start by hypothesizing that things are the same?

RSS
Research and Statistical Support

Logic of NHST

- If we believe something to be different, why do we start by hypothesizing that things are the same?
- Falsification

RSS
Research and Statistical Support

Logic of NHST

- If we believe something to be different, why do we start by hypothesizing that things are the same?
- Falsification
  - Very difficult to prove things “true” but, it is not difficult to show things are “not true”.

RSS
Research and Statistical Support

Logic of NHST

- If we believe something to be different, why do we start by hypothesizing that things are the same?
- Falsification
  - Very difficult to prove things “true” but, it is not difficult to show things are “not true”.
- Provides a basis for statistical testing.

Logic of NHST

- If we believe something to be different, why do we start by hypothesizing that things are the same?
- Falsification
  - Very difficult to prove things "true" but, it is not difficult to show things are "not true".
- Provides a basis for statistical testing.
  - Gives us somewhere to start (i.e., something to test).

Logic of NHST

- If we believe something to be different, why do we start by hypothesizing that things are the same?
- Falsification
  - Very difficult to prove things “true” but, it is not difficult to show things are “not true”.
- Provides a basis for statistical testing.
  - Gives us somewhere to start (i.e., something to test).
  - Provides the *first* piece of evidence for our empirical decisions.

RSS
Research and Statistical Support

More on this ‘falsification’ thingy...

- No amount of confirmation can achieve certainty for some types of statements.

More on this 'falsification' thingy...

- No amount of confirmation can achieve certainty for some types of statements.
    - Not verifiable:

More on this 'falsification' thingy...

- No amount of confirmation can achieve certainty for some types of statements.
  - Not verifiable:
    - There are no pink swans.

More on this 'falsification' thingy...

- No amount of confirmation can achieve certainty for some types of statements.
  - Not verifiable:
    - There are no pink swans.
    - All swans are purple.

RSS
Research and Statistical Support

More on this ‘falsification’ thingy...

- No amount of confirmation can achieve certainty for some types of statements.
  - Not verifiable:
    - There are no pink swans.
    - All swans are purple.
    - You would have to collect all swans to verify these statements.

More on this 'falsification' thingy...

- No amount of confirmation can achieve certainty for some types of statements.
  - Not verifiable:
    - There are no pink swans.
    - All swans are purple.
    - You would have to collect all swans to verify these statements.
  - Verifiable:

RSS
Research and Statistical Support

More on this ‘falsification’ thingy...

- No amount of confirmation can achieve certainty for some types of statements.
  - Not verifiable:
    - There are no pink swans.
    - All swans are purple.
    - You would have to collect all swans to verify these statements.
  - Verifiable:
    - There are blue swans.

RSS
Research and Statistical Support

More on this ‘falsification’ thingy...

- No amount of confirmation can achieve certainty for some types of statements.
  - Not verifiable:
    - There are no pink swans.
    - All swans are purple.
    - You would have to collect all swans to verify these statements.
  - Verifiable:
    - There are blue swans.
    - Not all swans are green.

RSS
Research and Statistical Support

More on this 'falsification' thingy...

- No amount of confirmation can achieve certainty for some types of statements.
  - Not verifiable:
    - There are no pink swans.
    - All swans are purple.
    - You would have to collect all swans to verify these statements.
  - Verifiable:
    - There are blue swans.
    - Not all swans are green.
    - You would need only one example to verify these statements.

More on this 'falsification' thingy...

- No amount of confirmation can achieve certainty for some types of statements.
  - Not verifiable:
    - There are no pink swans.
    - All swans are purple.
    - You would have to collect all swans to verify these statements.
  - Verifiable:
    - There are blue swans.
    - Not all swans are green.
    - You would need only one example to verify these statements.
- Confirmed theory is not truth, but merely conjecture.

RSS
Research and Statistical Support

More on this ‘falsification’ thingy...

- No amount of confirmation can achieve certainty for some types of statements.
  - Not verifiable:
    - There are no pink swans.
    - All swans are purple.
    - You would have to collect all swans to verify these statements.
  - Verifiable:
    - There are blue swans.
    - Not all swans are green.
    - You would need only one example to verify these statements.
- Confirmed theory is not truth, but merely conjecture.
  - Nothing is *proven* true, only *supported* by evidence.

More on this 'falsification' thingy...

- No amount of confirmation can achieve certainty for some types of statements.
  - Not verifiable:
    - There are no pink swans.
    - All swans are purple.
    - You would have to collect all swans to verify these statements.
  - Verifiable:
    - There are blue swans.
    - Not all swans are green.
    - You would need only one example to verify these statements.
- Confirmed theory is not truth, but merely conjecture.
  - Nothing is *proven* true, only *supported* by evidence.
- But, theory can be *falsified* with more certainty.

RSS
Research and Statistical Support

Example: Dog IQ scores

Research question: Are dogs on cartoons smarter than 'regular' dogs? In order to investigate, we use Scooby Doo as a representative of dogs on cartoons and we measure Scooby Doo's IQ ($X = 123$).

Example: Dog IQ scores

Research question: Are dogs on cartoons smarter than 'regular' dogs? In order to investigate, we use Scooby Doo as a representative of dogs on cartoons and we measure Scooby Doo's IQ (X = 123).

- **Step 1:** Define the populations and state the null and alternative hypotheses.

RSS
Research and Statistical Support

Example: Dog IQ scores

Research question: Are dogs on cartoons smarter than 'regular' dogs? In order to investigate, we use Scooby Doo as a representative of dogs on cartoons and we measure Scooby Doo's IQ (X = 123).

- **Step 1:** Define the populations and state the null and alternative hypotheses.
  - Population 1: Dogs on cartoons (represented by Scooby).
  - Population 2: Dogs not on cartoons.

$$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 > \mu_2$$

RSS
Research and Statistical Support

Example: Dog IQ scores

Research question: Are dogs on cartoons smarter than 'regular' dogs? In order to investigate, we use Scooby Doo as a representative of dogs on cartoons and we measure Scooby Doo's IQ (X = 123).

- **Step 1:** Define the populations and state the null and alternative hypotheses.
  - Population 1: Dogs on cartoons (represented by Scooby).
  - Population 2: Dogs not on cartoons.

$$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 > \mu_2$$

- **Step 2:** Determine the comparison distribution.

RSS
Research and Statistical Support

Example: Dog IQ scores

Research question: Are dogs on cartoons smarter than 'regular' dogs? In order to investigate, we use Scooby Doo as a representative of dogs on cartoons and we measure Scooby Doo's IQ (X = 123).

- **Step 1:** Define the populations and state the null and alternative hypotheses.
  - Population 1: Dogs on cartoons (represented by Scooby).
  - Population 2: Dogs not on cartoons.

$$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 > \mu_2$$

- **Step 2:** Determine the comparison distribution.
  - Population distribution of IQ among dogs not on cartoons is known to be normal and has $\mu = 100, \sigma = 15$.

RSS
Research and Statistical Support

Example: Dog IQ scores

Research question: Are dogs on cartoons smarter than 'regular' dogs? In order to investigate, we use Scooby Doo as a representative of dogs on cartoons and we measure Scooby Doo's IQ (X = 123).

- **Step 1:** Define the populations and state the null and alternative hypotheses.
  - Population 1: Dogs on cartoons (represented by Scooby).
  - Population 2: Dogs not on cartoons.

$$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 > \mu_2$$

- **Step 2:** Determine the comparison distribution.
  - Population distribution of IQ among dogs not on cartoons is known to be normal and has $\mu = 100, \sigma = 15$.
  - Therefore, we can use the standard normal distribution (Z-score distribution) as our comparison distribution.

RSS
Research and Statistical Support

- **Step 3:** Determine the cutoff sample score, also called the *critical value*, associated with our probability cutoff (.05).

- **Step 3:** Determine the cutoff sample score, also called the *critical value*, associated with our probability cutoff (.05).
  - $p = .05$ corresponds to a Z-score = 1.64; this is our critical score on the comparison distribution.

Research and Statistical Support

- **Step 3:** Determine the cutoff sample score, also called the *critical value*, associated with our probability cutoff (.05).
  - $p = .05$ corresponds to a Z-score = 1.64; this is our critical score on the comparison distribution.
- Recall the table of Z-score values from the previous module.

http://www.sjsu.edu/faculty/gerstman/EpiInfo/z-table.htm

- **Step 3:** Determine the cutoff sample score, also called the *critical value*, associated with our probability cutoff (.05).
  - $p = .05$ corresponds to a Z-score = 1.64; this is our critical score on the comparison distribution.
- Recall the table of Z-score values from the previous module.

  http://www.sjsu.edu/faculty/gerstman/EpiInfo/z-table.htm
- The $p = .05$ is how we represent the top 5 percent of the comparison distribution as a cutoff point.

- **Step 3:** Determine the cutoff sample score, also called the *critical value*, associated with our probability cutoff (.05).
  - $p = .05$ corresponds to a Z-score = 1.64; this is our critical score on the comparison distribution.
- Recall the table of Z-score values from the previous module.

  http://www.sjsu.edu/faculty/gerstman/EpiInfo/z-table.htm

- The $p = .05$ is how we represent the top 5 percent of the comparison distribution as a cutoff point.
- The Z-score associated with $p = .05$ at the higher end of the Standard Normal Curve is 1.64.

RSS
Research and Statistical Support

- **Step 3:** Determine the cutoff sample score, also called the *critical value*, associated with our probability cutoff (.05).
  - $p = .05$ corresponds to a Z-score = 1.64; this is our critical score on the comparison distribution.
- Recall the table of Z-score values from the previous module.

  `http://www.sjsu.edu/faculty/gerstman/EpiInfo/z-table.htm`
- The $p = .05$ is how we represent the top 5 percent of the comparison distribution as a cutoff point.
- The Z-score associated with $p = .05$ at the higher end of the Standard Normal Curve is 1.64.
- Another way of saying this is: the Z-score, 1.64, represents the cutoff point between the lower 95% (or more precisely 94.95%) and the *top* 5% of scores.

RSS Research and Statistical Support

Critical Value

- In this example, we are saying 95% of the (dogs not on cartoons) population of IQ scores are below this point.
  - Population 2: dogs not on cartoons.

Critical Value

- In this example, we are saying 95% of the (dogs not on cartoons) population of IQ scores are below this point.
    - Population 2: dogs not on cartoons.
- Recall, we are testing whether or not dogs on cartoons (represented by Scooby Doo) are significantly smarter than dogs not on cartoons.

RSS
Research and Statistical Support

Critical Value

- In this example, we are saying 95% of the (dogs not on cartoons) population of IQ scores are below this point.
  - Population 2: dogs not on cartoons.
- Recall, we are testing whether or not dogs on cartoons (represented by Scooby Doo) are significantly smarter than dogs not on cartoons.
- If we find Scooby's IQ is greater than our cutoff (if Scooby's Z-score is greater than 1.64) then we have evidence he is *significantly* 'smarter' than dogs not on cartoons.

RSS
Research and Statistical Support

Critical Value

- In this example, we are saying 95% of the (dogs not on cartoons) population of IQ scores are below this point.
  - Population 2: dogs not on cartoons.
- Recall, we are testing whether or not dogs on cartoons (represented by Scooby Doo) are significantly smarter than dogs not on cartoons.
- If we find Scooby's IQ is greater than our cutoff (if Scooby's Z-score is greater than 1.64) then we have evidence he is *significantly* 'smarter' than dogs not on cartoons.
- Which is to say, evidence that he does not come from population 2, but instead he represents a different population, namely population 1.

RSS
Research and Statistical Support

- **Step 4:** Determine your sample's score on the comparison distribution (i.e., compute your statistic).

- **Step 4:** Determine your sample's score on the comparison distribution (i.e., compute your statistic).

$$\frac{X-\mu}{\sigma} = \frac{123-100}{15} = 1.53$$

- **Step 4:** Determine your sample's score on the comparison distribution (i.e., compute your statistic).

$$\frac{X-\mu}{\sigma} = \frac{123-100}{15} = 1.53$$

- Calculated Z-score (Z-calc) = 1.53, which when we look in the Z-score table corresponds to $p = .063$ (.9370 or 93.70%).

RSS
Research and Statistical Support

- **Step 4:** Determine your sample's score on the comparison distribution (i.e., compute your statistic).

$$\frac{X-\mu}{\sigma} = \frac{123-100}{15} = 1.53$$

- Calculated Z-score (Z-calc) = 1.53, which when we look in the Z-score table corresponds to $p = .063$ (.9370 or 93.70%).
- To find the p value; look for 1.5 on the left and 0.03 on the top (of the table linked below) which corresponds to a Z-score of 1.53

http://www.sjsu.edu/faculty/gerstman/EpiInfo/z-table.htm

RSS
Research and Statistical Support

- **Step 5:** Compare and make a decision. To reject or not to reject? That is the question!

- **Step 5:** Compare and make a decision. To reject or not to reject? That is the question!
- Compare our critical (crit) value to our calculated (calc) value (Z-scores or *p* values)

- **Step 5:** Compare and make a decision. To reject or not to reject? That is the question!
- Compare our critical (crit) value to our calculated (calc) value (Z-scores or *p* values)
  - $Z_{crit} = 1.64 > Z_{calc} = 1.53$
  - $p_{crit} = .05 < p_{calc} = .063$

- **Step 5:** Compare and make a decision. To reject or not to reject? That is the question!
- Compare our critical (crit) value to our calculated (calc) value (Z-scores or *p* values)
  - $Z_{crit} = 1.64 > Z_{calc} = 1.53$
  - $p_{crit} = .05 < p_{calc} = .063$
- Although a bit confusing, the second is the same as the first and is typically used when reporting results (e.g., research articles).

- **Step 5:** Compare and make a decision. To reject or not to reject? That is the question!
- Compare our critical (crit) value to our calculated (calc) value (Z-scores or *p* values)
  - $Z_{crit} = 1.64 > Z_{calc} = 1.53$
  - $p_{crit} = .05 < p_{calc} = .063$
- Although a bit confusing, the second is the same as the first and is typically used when reporting results (e.g., research articles).
  - Both indicate that Scooby's IQ was not far enough from the mean of population 2 to distinguish him as a member of population 1.

- **Step 5:** Compare and make a decision. To reject or not to reject? That is the question!
- Compare our critical (crit) value to our calculated (calc) value (Z-scores or *p* values)
  - $Z_{crit} = 1.64 > Z_{calc} = 1.53$
  - $p_{crit} = .05 < p_{calc} = .063$
- Although a bit confusing, the second is the same as the first and is typically used when reporting results (e.g., research articles).
  - Both indicate that Scooby's IQ was not far enough from the mean of population 2 to distinguish him as a member of population 1.
  - Thus, we have no evidence for population 1 and the idea that dogs on cartoons are smarter than dogs not on cartoons.

- **Step 5:** Compare and make a decision. To reject or not to reject? That is the question!
- Compare our critical (crit) value to our calculated (calc) value (Z-scores or *p* values)
  - $Z_{crit} = 1.64 > Z_{calc} = 1.53$
  - $p_{crit} = .05 < p_{calc} = .063$
- Although a bit confusing, the second is the same as the first and is typically used when reporting results (e.g., research articles).
  - Both indicate that Scooby's IQ was not far enough from the mean of population 2 to distinguish him as a member of population 1.
  - Thus, we have no evidence for population 1 and the idea that dogs on cartoons are smarter than dogs not on cartoons.
- Fail to reject the null hypothesis; we never say we *accept* the null or alternative hypothesis.

RSS
Research and Statistical Support

But what does the *p* value mean?

Probability obtained tells us:

But what does the *p* value mean?

Probability obtained tells us:

- If the null hypothesis were true, the probability of obtaining a sample statistic of the kind observed.

But what does the *p* value mean?

Probability obtained tells us:

- If the null hypothesis were true, the probability of obtaining a sample statistic of the kind observed.
- Stated another way: probability tells us how extreme is the result we got; assuming the null is true.

But what does the *p* value mean?

Probability obtained tells us:

- If the null hypothesis were true, the probability of obtaining a sample statistic of the kind observed.
- Stated another way: probability tells us how extreme is the result we got; assuming the null is true.
- If it is more extreme than our stated critical *p* value (i.e., lower probability than .05), then we reject the null hypothesis.

But what does the *p* value mean?

Probability obtained tells us:

- If the null hypothesis were true, the probability of obtaining a sample statistic of the kind observed.
- Stated another way: probability tells us how extreme is the result we got; assuming the null is true.
- If it is more extreme than our stated critical *p* value (i.e., lower probability than .05), then we reject the null hypothesis.
- If it is less extreme than our stated critical *p* value (i.e., greater probability than .05), then we **fail to** reject the null hypothesis.

But what does the *p* value mean?

Probability obtained tells us:

- If the null hypothesis were true, the probability of obtaining a sample statistic of the kind observed.
- Stated another way: probability tells us how extreme is the result we got; assuming the null is true.
- If it is more extreme than our stated critical *p* value (i.e., lower probability than .05), then we reject the null hypothesis.
- If it is less extreme than our stated critical *p* value (i.e., greater probability than .05), then we **fail to** reject the null hypothesis.
- We never accept either hypothesis, nor do we reject the alternative hypothesis. We either reject or fail to reject the null hypothesis.

RSS
Research and Statistical Support

More on probability obtained.

- It is important to recognize that the *p* values are what tie our calculated sample statistic to the comparison distribution and vice-versa.

More on probability obtained.

- It is important to recognize that the *p* values are what tie our calculated sample statistic to the comparison distribution and vice-versa.
- If Z-calc is larger than Z-crit, we reject the null because, it indicates our sample (representing population 1) is more extreme than our $p = .05$ critical value (representing population 2).

RSS
Research and Statistical Support

More on probability obtained.

- It is important to recognize that the *p* values are what tie our calculated sample statistic to the comparison distribution and vice-versa.
- If Z-calc is larger than Z-crit, we reject the null because, it indicates our sample (representing population 1) is more extreme than our $p = .05$ critical value (representing population 2).
  - Interpretation: Population 1 is significantly different from population 2 ($p < .05$).

RSS
Research and Statistical Support

More on probability obtained.

- It is important to recognize that the *p* values are what tie our calculated sample statistic to the comparison distribution and vice-versa.
- If Z-calc is larger than Z-crit, we reject the null because, it indicates our sample (representing population 1) is more extreme than our $p = .05$ critical value (representing population 2).
  - Interpretation: Population 1 is significantly different from population 2 ($p < .05$).
- If Z-crit is larger than Z-calc, we **fail to** reject because, it indicates our sample (population 1) is not extreme enough to differentiate it from population 2.

More on probability obtained.

- It is important to recognize that the *p* values are what tie our calculated sample statistic to the comparison distribution and vice-versa.
- If Z-calc is larger than Z-crit, we reject the null because, it indicates our sample (representing population 1) is more extreme than our $p = .05$ critical value (representing population 2).
  - Interpretation: Population 1 is significantly different from population 2 ($p < .05$).
- If Z-crit is larger than Z-calc, we **fail to** reject because, it indicates our sample (population 1) is not extreme enough to differentiate it from population 2.
  - Interpretation: Population 1 is not significantly different from population 2 ($p > .05$); they are the same (or only population 2 exists).

RSS
Research and Statistical Support

One-tailed vs. Two-tailed

- Note, our example alternative hypothesis was directional; dictating a ‘one-tailed’ test. We were interested in whether or not population 1 was significantly **greater than** population 2.

One-tailed vs. Two-tailed

- Note, our example alternative hypothesis was directional; dictating a 'one-tailed' test. We were interested in whether or not population 1 was significantly **greater than** population 2.

$$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 > \mu_2$$

RSS
Research and Statistical Support

One-tailed vs. Two-tailed

- Note, our example alternative hypothesis was directional; dictating a ‘one-tailed’ test. We were interested in whether or not population 1 was significantly **greater than** population 2.

$$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 > \mu_2$$

  - A precise alternative hypothesis.

One-tailed vs. Two-tailed

- Note, our example alternative hypothesis was directional; dictating a ‘one-tailed’ test. We were interested in whether or not population 1 was significantly **greater than** population 2.

$$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 > \mu_2$$

  - A precise alternative hypothesis.
- We were not asking the question “is population 1 significantly **different** from population 2”, which is a more general, non-directional hypothesis and dictates a two-tailed test.

RSS
Research and Statistical Support

One-tailed vs. Two-tailed

- Note, our example alternative hypothesis was directional; dictating a 'one-tailed' test. We were interested in whether or not population 1 was significantly **greater than** population 2.

$$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 > \mu_2$$

  - A precise alternative hypothesis.

- We were not asking the question "is population 1 significantly **different** from population 2", which is a more general, non-directional hypothesis and dictates a two-tailed test.

$$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 \neq \mu_2$$

RSS Research and Statistical Support

One-tailed vs. Two-tailed

- Note, our example alternative hypothesis was directional; dictating a 'one-tailed' test. We were interested in whether or not population 1 was significantly **greater than** population 2.

$$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 > \mu_2$$

  - A precise alternative hypothesis.
- We were not asking the question "is population 1 significantly **different** from population 2", which is a more general, non-directional hypothesis and dictates a two-tailed test.

$$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 \neq \mu_2$$

  - Because it is less specific; population 1 could be greater or less than population 2.

RSS
Research and Statistical Support

One-tailed vs. Two-tailed

- In a one-tailed test, the .05 critical level we typically use gets applied to one side (high or low) of the distribution.

RSS
Research and Statistical Support

One-tailed vs. Two-tailed

- In a one-tailed test, the .05 critical level we typically use gets applied to one side (high or low) of the distribution.
    - Note, one-tailed test hypotheses can go either direction, *greater than* or *less than*; it simply depends on how you state your alternative hypothesis, which is often a function of how much you know about the topic you are examining.

One-tailed vs. Two-tailed

- In a one-tailed test, the .05 critical level we typically use gets applied to one side (high or low) of the distribution.
  - Note, one-tailed test hypotheses can go either direction, *greater than* or *less than*; it simply depends on how you state your alternative hypothesis, which is often a function of how much you know about the topic you are examining.
- When doing a two-tailed test, we are interested in statistics associated with extreme values at either end of the comparison distribution.

One-tailed vs. Two-tailed

- In a one-tailed test, the .05 critical level we typically use gets applied to one side (high or low) of the distribution.
  - Note, one-tailed test hypotheses can go either direction, *greater than* or *less than*; it simply depends on how you state your alternative hypothesis, which is often a function of how much you know about the topic you are examining.
- When doing a two-tailed test, we are interested in statistics associated with extreme values at either end of the comparison distribution.
  - Take .05 and split it; .025 at both ends = .05 as the critical level we generally use to determine statistical significance.

RSS
Research and Statistical Support

One-tailed vs. Two-tailed

- Directional Alternative Hypotheses

$$H_1 : \mu_1 > \mu_2 \qquad H_1 : \mu_1 < \mu_2$$

One-tailed vs. Two-tailed

- Directional Alternative Hypotheses

$$H_1 : \mu_1 > \mu_2 \qquad H_1 : \mu_1 < \mu_2$$

  - Using $p = .05$ all lumped at one end of the distribution.

RSS
Research and Statistical Support

One-tailed vs. Two-tailed

- Directional Alternative Hypotheses

$$H_1 : \mu_1 > \mu_2 \qquad H_1 : \mu_1 < \mu_2$$

  - Using $p = .05$ all lumped at one end of the distribution.
  - Must be specified prior to collecting data.

RSS
Research and Statistical Support

One-tailed vs. Two-tailed

- Directional Alternative Hypotheses

$$H_1 : \mu_1 > \mu_2 \qquad H_1 : \mu_1 < \mu_2$$

  - Using $p = .05$ all lumped at one end of the distribution.
  - Must be specified prior to collecting data.
- Non-directional Alternative Hypotheses

$$H_1 : \mu_1 \neq \mu_2$$

One-tailed vs. Two-tailed

- Directional Alternative Hypotheses

$$H_1 : \mu_1 > \mu_2 \qquad H_1 : \mu_1 < \mu_2$$

  - Using $p = .05$ all lumped at one end of the distribution.
  - Must be specified prior to collecting data.

- Non-directional Alternative Hypotheses

$$H_1 : \mu_1 \neq \mu_2$$

  - Using $p = .05$ split in half, with one half ($p = .025$) at each end of the distribution.

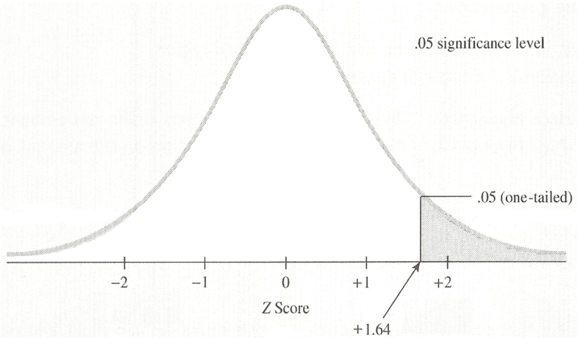
Examples follow below.

RSS
Research and Statistical Support

Directional Alternative Hypothesis

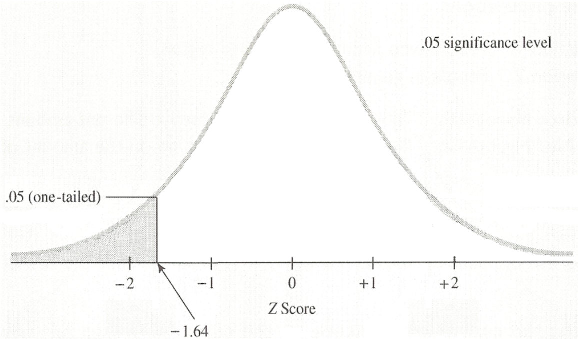
Population 1 **greater** than population 2

$H_1 : \mu_1 > \mu_2$



Directional Alternative Hypothesis

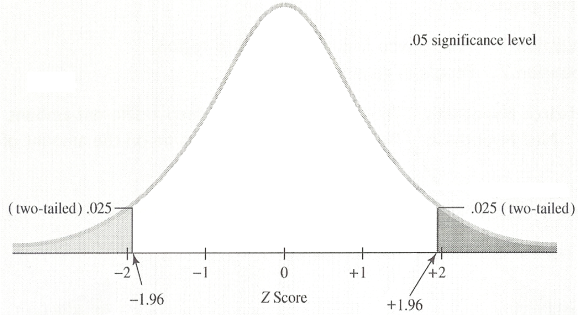
Population 1 **less** than population 2
$H_1 : \mu_1 < \mu_2$



Non-directional Alternative Hypothesis

Population 1 **different** than population 2

$H_1 : \mu_1 \neq \mu_2$



Decision Errors

- When the *right procedures* and *calculations* lead to the *wrong* decisions.

Decision Errors

- When the *right procedures* and *calculations* lead to the *wrong* decisions.
- **Type I** error: Rejecting the null hypothesis when in fact it is true.

RSS
Research and Statistical Support

Decision Errors

- When the *right procedures* and *calculations* lead to the *wrong* decisions.
- **Type I** error: Rejecting the null hypothesis when in fact it is true.
    - Also called Alpha error
    - Symbol: α

RSS
Research and Statistical Support

Decision Errors

- When the *right procedures* and *calculations* lead to the *wrong* decisions.
- **Type I** error: Rejecting the null hypothesis when in fact it is true.
    - Also called Alpha error
    - Symbol: α
- **Type II** error: Failing to reject the null when in fact it is false.

Decision Errors

- When the *right procedures* and *calculations* lead to the *wrong* decisions.
- **Type I** error: Rejecting the null hypothesis when in fact it is true.
  - Also called Alpha error
  - Symbol: α
- **Type II** error: Failing to reject the null when in fact it is false.
  - Also called Beta error
  - Symbol: β

RSS Research and Statistical Support

Decision Possibilities

		State of the World	
		H_0 true	H_0 false
Research Decision	Reject H_0	Type I error	Correct rejection
	Fail to reject H_0	Correct fail to reject	Type II error

RSS
Research and Statistical Support

More on Type I Error

Social sciences tend to focus more emphasis on Type I error than on Type II error.

More on Type I Error

Social sciences tend to focus more emphasis on Type I error than on Type II error.

- When initially designing a study, the Alpha level should be set early on (prior to data collection).

More on Type I Error

Social sciences tend to focus more emphasis on Type I error than on Type II error.

- When initially designing a study, the Alpha level should be set early on (prior to data collection).
    - Alpha is the probability of committing a Type I error.

RSS Research and Statistical Support

More on Type I Error

Social sciences tend to focus more emphasis on Type I error than on Type II error.

- When initially designing a study, the Alpha level should be set early on (prior to data collection).
  - Alpha is the probability of committing a Type I error.
- Typically, social science uses the conventional .05.

RSS
Research and Statistical Support

More on Type I Error

Social sciences tend to focus more emphasis on Type I error than on Type II error.

- When initially designing a study, the Alpha level should be set early on (prior to data collection).
    - Alpha is the probability of committing a Type I error.
- Typically, social science uses the conventional .05.
- This means, we are willing to run the risk of having a sample which is extreme enough (5% of the total population) to reject the null when the null is in fact true.

More on Type I Error

Social sciences tend to focus more emphasis on Type I error than on Type II error.

- When initially designing a study, the Alpha level should be set early on (prior to data collection).
  - Alpha is the probability of committing a Type I error.
- Typically, social science uses the conventional .05.
- This means, we are willing to run the risk of having a sample which is extreme enough (5% of the total population) to reject the null when the null is in fact true.
  - If Scooby's IQ was 160 but all the other cartoon dogs were especially dumb, we would reject the null when in fact there was no difference between the IQ of cartoon dogs (population 1) and the IQ dogs not in cartoons (population 2).

Type I Error continued

- The problem with Alpha is that we never know if we have committed the error, unless we fail to reject the null – in which case, we know for certain that we have not committed a Type I error.

RSS
Research and Statistical Support

Type I Error continued

- The problem with Alpha is that we never know if we have committed the error, unless we fail to reject the null – in which case, we know for certain that we have not committed a Type I error.
  - But, we may have then committed a Type II error.

RSS
Research and Statistical Support

Type I Error continued

- The problem with Alpha is that we never know if we have committed the error, unless we fail to reject the null – in which case, we know for certain that we have not committed a Type I error.
  - But, we may have then committed a Type II error.
- If Scooby was especially dumb, but all the other cartoon dogs were especially bright, then we would have failed to reject the null, when in fact there was a significant difference in IQ between dogs in cartoons and dogs not in cartoons.

RSS
Research and Statistical Support

Decision Probabilities

		State of the World: H_0 true	State of the World: H_0 false
Research Decision	Reject H_0	Type I $p = \alpha$	$p = 1 - \beta$ $= \text{Power}$
Research Decision	Fail to reject H_0	$p = 1 - \alpha$	Type II $p = \beta$

RSS
Research and Statistical Support

NHST Concerns and Controversies

First, much of this module will be revisited and reinforced in later modules.

NHST Concerns and Controversies

First, much of this module will be revisited and reinforced in later modules.

- The null is never true, strictly speaking.

RSS Research and Statistical Support

NHST Concerns and Controversies

First, much of this module will be revisited and reinforced in later modules.

- The null is never true, strictly speaking.
  - There will always be some (numerical) difference between any two populations, so why do we use a 'null' hypothesis?

NHST Concerns and Controversies

First, much of this module will be revisited and reinforced in later modules.

- The null is never true, strictly speaking.
  - There will always be some (numerical) difference between any two populations, so why do we use a 'null' hypothesis?
  - It provides a starting point; in later modules we will see that statistical significance is not "everything" and should not be the *only* thing we consider when making empirical decisions.

Concerns and Controversies continues on the next slide.

Concerns and Controversies continued

- Most controversies come from mis-interpretation of or over-reliance on the *p* value.

Concerns and Controversies continued

- Most controversies come from mis-interpretation of or over-reliance on the *p* value.
  - It does not represent the probability of the Null hypothesis or the Alternative hypothesis.

Concerns and Controversies continued

- Most controversies come from mis-interpretation of or over-reliance on the *p* value.
  - It does not represent the probability of the Null hypothesis or the Alternative hypothesis.
  - Nor does it say anything about how likely your result is to replicate.

RSS
Research and Statistical Support

Concerns and Controversies continued

- Most controversies come from mis-interpretation of or over-reliance on the *p* value.
  - It does not represent the probability of the Null hypothesis or the Alternative hypothesis.
  - Nor does it say anything about how likely your result is to replicate.
  - One study alone should never be used to make serious decisions; replication of findings across multiple samples must be done in order to have strong evidence for (and confidence in) research findings.

RSS Research and Statistical Support

Summary of Module 5

Module 5 covered the following topics:

- Rules and Steps of Hypothesis Testing

Summary of Module 5

Module 5 covered the following topics:

- Rules and Steps of Hypothesis Testing
- Individual Z-test.

RSS
Research and Statistical Support

Summary of Module 5

Module 5 covered the following topics:

- Rules and Steps of Hypothesis Testing
- Individual Z-test.
- *p* values.

RSS
Research and Statistical Support

Summary of Module 5

Module 5 covered the following topics:

- Rules and Steps of Hypothesis Testing
- Individual Z-test.
- *p* values.
- 1-tailed and 2-tailed tests.

Summary of Module 5

Module 5 covered the following topics:

- Rules and Steps of Hypothesis Testing
- Individual Z-test.
- *p* values.
- 1-tailed and 2-tailed tests.
- Decision Errors.

RSS Research and Statistical Support

Summary of Module 5

Module 5 covered the following topics:

- Rules and Steps of Hypothesis Testing
- Individual Z-test.
- *p* values.
- 1-tailed and 2-tailed tests.
- Decision Errors.
- Concerns and Controversies of NHST.

RSS Research and Statistical Support

Summary of Module 5

Module 5 covered the following topics:

- Rules and Steps of Hypothesis Testing
- Individual Z-test.
- *p* values.
- 1-tailed and 2-tailed tests.
- Decision Errors.
- Concerns and Controversies of NHST.

All of these topics will be revisited consistently in future modules.

RSS Research and Statistical Support

This concludes Module 5

Next time Module 6.

- Next time we'll begin covering Hypothesis Testing with means of Samples.
- Until next time; have a nice day.

These slides initially created on: October 6, 2010
These slides last updated on: October 7, 2010

- The bottom date shown is the date this Adobe.pdf file was created; LaTeX[1] has a command for automatically inserting the date of a document's creation.

RSS Research and Statistical Support

[1] This document was created in LaTeX using the Beamer package