EDUR 7130
Presentation 5a

1. Inferential Statistics: Hypothesis Testing

Inferential statistics is based upon **hypothesis testing**:

- procedure for making inferences from sample to target population
- helps determine whether differences or relationships found are real or due to sampling error

One always tests the ***null*** hypothesis with hypothesis testing.

The null states

(a) two or more variables are not related, or
(b) that no differences exist between groups on the DV.

Examples

(a) There is no relationship between hours studied and test scores.
(b) There is no difference in test scores between females and males.

If the null is ***rejected***, that means the data collected suggest that (a) a relationship is found between variables or (b) differences on the DV found between groups.

If one ***fails to reject*** the null (i.e. null is not rejected), that means data collect suggest (a) no relationship between variables or (b) no differences on the DV found between groups.

If null is rejected, that is called a ***significant*** or ***statistically significant*** finding.

If the null is not rejected, that is termed ***not significant***, ***not statistically significant***, or ***insignificant***.

The term ***significant*** means

- the null is rejected, and
- does ***not*** imply importance or large effect obtained.

If you hear in the news that researchers report a significant finding between smoking and lung cancer, what does this mean?

It means the data collected by the researchers indicate the two variables may be related, so the null hypothesis of no relationship was rejected, but it does not mean an important or large effect was found. That is a separate judgment requiring professional assessment of the size of the relationship or difference observed. Significant does not mean importance.

Note that rejecting the null does not imply something important has been found, because the null can be rejected easily whenever the sample is large. Consider this example:

	Math Achievement Means	n
Experimental	85.13%	15,238
Control	85.12%	14,879

Is this significant difference of 0.01% important? Trivial differences can be significant if the sample is large enough, so significant does not mean something is important; instead, significant means only that the null hypothesis was rejected.

Recap:

- While hypothesis testing is done with sample data, results of hypothesis are inferred to the population.
- If the null hypothesis is rejected, one labels this rejection as "statistically significant" and concludes that there is a difference in the DV across groups (if the IV is qualitative), or that a relationship exists between IV and DV (if the IV is quantitative).
- If the null hypothesis is not rejected, then one concludes there is no difference across groups on the DV or that no relationship exists between IV and DV.
- Significant = null hypothesis rejected
- Non-significant = null hypothesis not rejected

Decisions and Errors with Hypothesis Testing

Table 1: Hypothesis Testing Decisions and Errors

		Population Reality	
		Ho is true **(Example: no difference In mathematics scores between boys and girls)**	**Ho is false** **(Example: true difference in mathematics scores between boys and girls)**
Sample Situation: Hypothesis Testing Decision by Researcher	**Reject Ho**	Mistake; **Type 1 Error** (probability of this error is **alpha α**, normally set at .05 or 5%)	Correct Decision
	Fail to Reject Ho (retain Ho)	Correct Decision	Mistake; **Type 2 Error** (probability of this error is **beta β**, normally set at .20 or lower if possible)

Table can be found here
http://www.bwgriffin.com/gsu/courses/edur7130/content/inferential_statistics.htm

Example 1 Research Scenario (No Real Difference)

Is there a difference in mathematics achievement between 3rd grade boys and girls?

Null Hypothesis

Ho: There is no difference in mathematics achievement between boys and girls.

Alternative Hypothesis (adopted if null is rejected)

Ha: There is a difference in mathematics achievement between boys and girls.

Assume there is ***no real difference*** in mathematics achievement for the **population** of 3rd grade females and males.

Sample Results: We Reject Ho

- We reject Ho, so we believe the sample data indicate a difference exists, and we rejected Ho in favor of Ha (the alternative hypothesis).
- We then generalize from sample to population and claim that a difference exists between boys and girls regarding mathematics achievement in the population.

Is this Example 1 conclusion in error? Is it possible we have mistakenly rejected the null hypothesis? If yes, what is this error called in hypothesis testing?

		Population Reality	
		Ho is true **(no difference in mathematics scores between boys and girls)**	**Ho is false** **(true difference in mathematics scores between boys and girls)**
Sample Situation: Hypothesis Testing Decision by Researcher	**Reject Ho**	Mistake; **Type 1 Error** (probability of this error is **alpha α**, normally set at .05 or 5%)	Correct Decision
	Fail to Reject Ho (retain Ho)	Correct Decision	Mistake; **Type 2 Error** (probability of this error is **beta β**, normally set at .20 or lower if possible)

Answer

If a true Ho is mistakenly rejected:

- In the population, the null is true (correct), so no difference in DV between groups exists in the population.
  - Current example, no difference in math achievement between females and males in population of 3rd grade students
  - So the null should not be rejected
- In sample null appears to be false, so we reject it.
  - Sample data appear to show difference in math achievement between 3rd grade females and males, so
  - We reject Ho in sample since sample data do not support Ho

- This is a Type 1 error – incorrectly rejecting null when it should not be rejected

Alpha (α) = probability of committing a Type 1 error (usually set at .05 or .01)

Recap:

- If one rejects Ho, one may have made a correct decision, or one may have committed an error, and we call this error a Type 1 error
- Type 1 error = rejecting a true Ho (the null is true in the population, but appears false in the sample).
- The mistake here is that researchers will claim to have found something (a difference or relationship) when in fact a difference or relationship does not exist in the population.
- Researcher control probability of Type 1 error by setting probability typically at .05 or .01

Example 2 Research Scenario (Real Difference)

Is there a difference in mathematics achievement between 3rd grade boys and girls?

Null Hypothesis

Ho: There will be no difference in mathematics achievement between boys and girls.

Alternative Hypothesis (adopted if null is rejected)

Ha: There will be a difference in mathematics achievements between boys and girls.

Now assume there is a ***real difference*** in mathematics achievement between 3rd grade females and males in the **population**.

Sample Results: We Fail to Reject Ho

- We fail to reject Ho, so we believe the sample data indicate no difference exists, and we adopt Ho rather than Ha (the alternative hypothesis).
- We then generalize from sample to population and claim that no difference exists between boys and girls regarding mathematics achievement in the population.

Is this Example 2 conclusion in error? Is it possible we have mistakenly failed to reject the null hypothesis? If yes, what is this error called in hypothesis testing?

		Population Reality	
		Ho is true **(no difference in mathematics scores between boys and girls)**	**Ho is false** **(true difference in mathematics scores between boys and girls)**
Sample Situation: Hypothesis Testing Decision by Researcher	**Reject Ho**	Mistake; **Type 1 Error** (probability of this error is **alpha α**, normally set at .05 or 5%)	Correct Decision
	Fail to Reject Ho (retain Ho)	Correct Decision	Mistake; **Type 2 Error** (probability of this error is **beta β**, normally set at .20 or lower if possible)

Answer

If we mistakenly fail to reject a false Ho:

- In population the null is false (incorrect), so a difference in DV exists between groups in population.
  - Current example, there is a difference in math achievement between females and males in population of 3rd grade students
  - So null should be rejected
- In sample null appears to be true, so we fail to reject it.

    - Sample data appear to show no difference in math achievement between 3rd grade females and males, so
    - we fail to reject Ho in sample since sample data do support Ho
- This is a Type 2 error – incorrectly failing to reject the null when it should be rejected

Beta (β) = probability of committing a Type 2 error (often set at .20 or lower if possible, but difficult to control this probability).

Type 2 error is more difficult to prevent than Type 1 error, so the probability level is set higher to conserve resources (e.g., prevent need for very large sample sizes).

Recap:

- If one fails to rejects Ho, one may have made a correct decision, or one may have committed an error, and we call this error a Type 2 error
- Type 2 error = failing to reject a false Ho (the null is false, incorrect, in the population, but appears to be true or correct in the sample).
- The mistake here is that researchers will claim that no difference or relationship exists when in fact a difference or relationship does exist in the population.
- Difficult to control probability of Type 2 error, but attempts are often made to hold it at .20 or lower

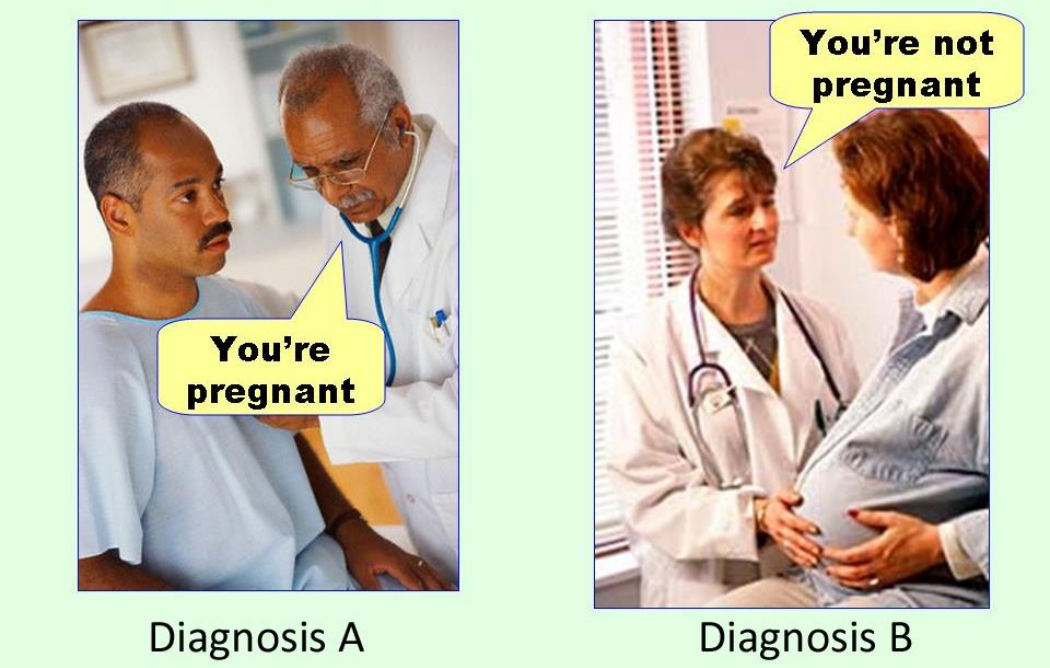
Medical Example of Type 1 and Type 2 Errors

Null Hypothesis:

Ho: The patient is not pregnant

Question 1

Which of the following diagnoses would be Type 1 and Type 2 errors?



Type 1 error: Saying there is an effect (a difference or relationship) when there is not in reality.
Type 2 error: Saying there is no effect (a difference or relationship) when there is an effect in reality.

Answer

Diagnosis A – Type 1 error, saying there is an effect (pregnant) when there is not really an effect in the population.

Diagnosis B – Type 2 error, claiming there is no effect when, in fact, there is an effect in the population.

Example 3

If there is no real difference in the population (for example, assume there is no real difference between females and males in motivation to read), and in a random sample of 10 females and 10 males we fail to reject the null hypothesis of no difference, have we made an error or a correct decision with our sample data? If an error, what type of error?

		Population Reality	
		Ho is true **(no difference in mathematics scores between boys and girls)**	**Ho is false** **(true difference in mathematics scores between boys and girls)**
Sample Situation: Hypothesis Testing Decision by Researcher	**Reject Ho**	Mistake; **Type 1 Error** (probability of this error is **alpha α**, normally set at .05 or 5%)	Correct Decision
	Fail to Reject Ho (retain Ho)	Correct Decision	Mistake; **Type 2 Error** (probability of this error is **beta β**, normally set at .20 or lower if possible)

Answer

This was a correct decision. If in the population of females and males there is no difference in motivation to read, and if we take a random sample of 10 females and 10 males and we fail to reject Ho in our sample, we have made a correct decision. In other words, in our sample we found no difference and that is consistent with the population, so we made a correct inference from sample to population.

Example 4

Now let's suppose there is a real difference in motivation to read between females and males in the population. We execute a study; we take a random sample of 25 females and 25 males. In our sample we fail to reject Ho; that is, in our study we found no difference between females and males in terms of their motivation to read. So we conclude there is no difference in motivation to read between females and males; i.e., we infer from our sample to the population no difference in motivation to read. Have we made an error or a correct decision in our sample? If an error, what type of error?

		Population Reality	
		Ho is true **(no difference in mathematics scores between boys and girls)**	**Ho is false** **(true difference in mathematics scores between boys and girls)**
Sample Situation: Hypothesis Testing Decision by Researcher	**Reject Ho**	Mistake; **Type 1 Error** (probability of this error is **alpha α**, normally set at .05 or 5%)	Correct Decision
	Fail to Reject Ho (retain Ho)	Correct Decision	Mistake; **Type 2 Error** (probability of this error is **beta β**, normally set at .20 or lower if possible)

Answer

This is a Type 2 error. We failed to detect a true difference (or a real relationship) in the population.

Example 5

Suppose there is no real difference in motivation to read between females and males in the population. We perform a study and take a random sample of 15 females and 15 males. In our sample we reject Ho; i.e., we find a difference between females and males in terms of their motivation to read. Based upon our sample we conclude there is a difference in motivation to read between females and males in the population; we infer from our sample to the population a difference in motivation to read. Have we made an error or a correct decision in our sample? If an error, what type of error?

		Population Reality	
		Ho is true **(no difference in mathematics scores between boys and girls)**	**Ho is false** **(true difference in mathematics scores between boys and girls)**
Sample Situation: Hypothesis Testing Decision by Researcher	**Reject Ho**	Mistake; **Type 1 Error** (probability of this error is **alpha α**, normally set at .05 or 5%)	Correct Decision
	Fail to Reject Ho (retain Ho)	Correct Decision	Mistake; **Type 2 Error** (probability of this error is **beta β**, normally set at .20 or lower if possible)

Answer

This is a Type 1 error--- our sample says there is a difference (or a relationship), yet there really is no difference or no relationship in the population, so we have made a Type 1 error (rejecting Ho when it is true).

Example 6

Medical researchers claimed to have found a significant benefit for those who take aspirin once a day in terms of lowering the probability of a heart attack. What type of error is possible for this study?

		Population Reality	
		Ho is true **(no difference in mathematics scores between boys and girls)**	**Ho is false** **(true difference in mathematics scores between boys and girls)**
Sample Situation: Hypothesis Testing Decision by Researcher	**Reject Ho**	Mistake; **Type 1 Error** (probability of this error is **alpha α**, normally set at .05 or 5%)	Correct Decision
	Fail to Reject Ho (retain Ho)	Correct Decision	Mistake; **Type 2 Error** (probability of this error is **beta β**, normally set at .20 or lower if possible)

Answer

Only a Type 1 error is possible since Ho rejected in the sample. Recall that the word "significant" means the null is rejected.

Note: Example 6 will probably appear on Test 2 (or something very much like it).

Example 7

Researchers, using National Educational Longitudinal data, claim there is no significant relation between hours spent watching TV and student performance on standardized tests. What type of error is possible for this study?

		Population Reality	
		Ho is true **(no difference in mathematics scores between boys and girls)**	**Ho is false** **(true difference in mathematics scores between boys and girls)**
Sample Situation: Hypothesis Testing Decision by Researcher	**Reject Ho**	Mistake; **Type 1 Error** (probability of this error is **alpha α**, normally set at .05 or 5%)	Correct Decision
	Fail to Reject Ho (retain Ho)	Correct Decision	Mistake; **Type 2 Error** (probability of this error is **beta β**, normally set at .20 or lower if possible)

Answer

No significant relation means researchers failed to reject Ho, so the only error possible when one fails to reject Ho with sample data is a Type 2 error.

[Note to instructor – about 25 minutes to cover above material – time remains for inferential test discussion]