HYPOTHESIS TESTING

"It is a capital mistake to theorize before one has data."
Sir Arthur Conan Doyle

Hypothesis Testing/Null Hypothesis Significance Testing

Hypothesis:

Independent Variable:

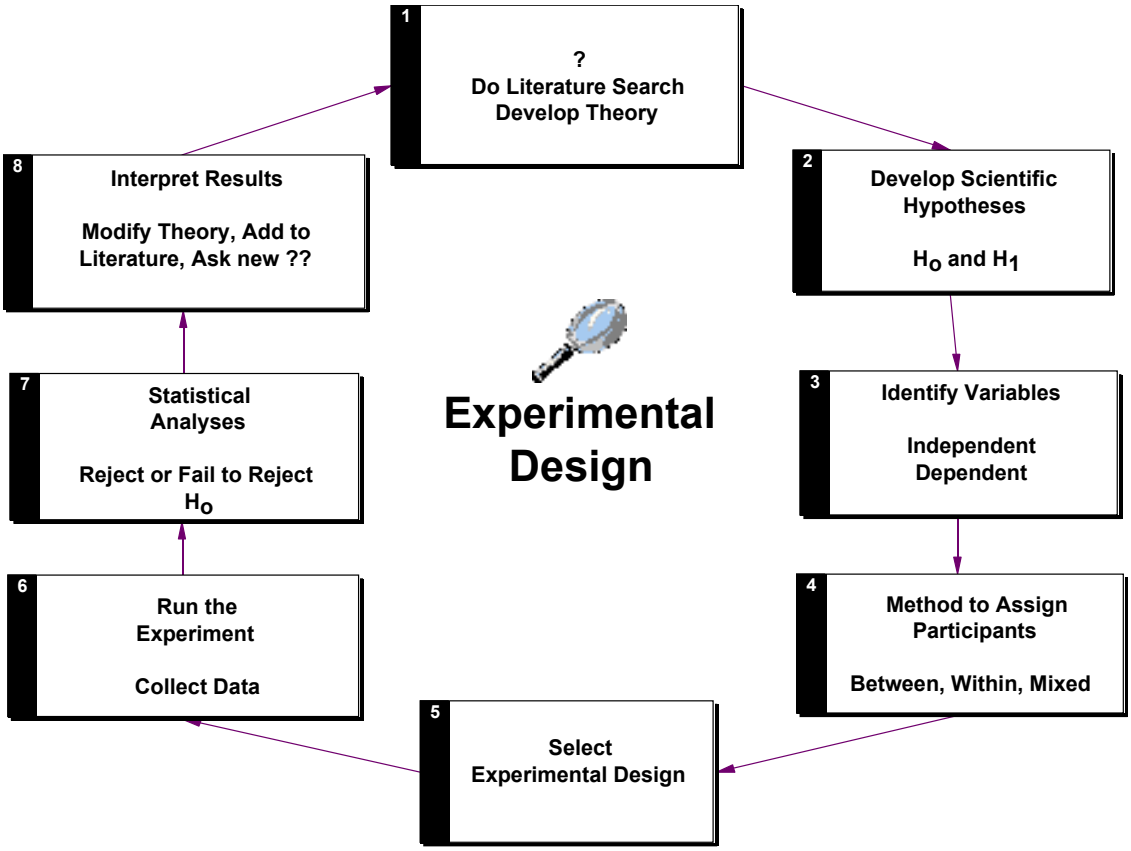
Dependent Variable:

Subject Variables:

Statistical Inference:

Use of the Term "Prove":

Experimental Design



The Process of Statistical Inference

Research Example: Effects of Prenatal Alcohol on Birth Weight

- Random sample of $n = 16$ pregnant rats
- Mother rats given daily doses of alcohol
- At birth, 1 pup is selected from each litter to produce a sample of n = 16 newborn rats
- Researcher would like to compare the sample with the general population of rats:
  → Newborn rats not exposed to alcohol: $\mu = 18$ grams, $\sigma = 4$
  → Sample of rats exposed to alcohol: $M = 15$ grams

1. State the null and alternative hypotheses.

Null hypothesis:

Alternative hypothesis:

One-tailed tests:

Two-tailed tests:

2. Set the criteria for a decision.

3. Collect sample data.

4. Compute sample statistics.

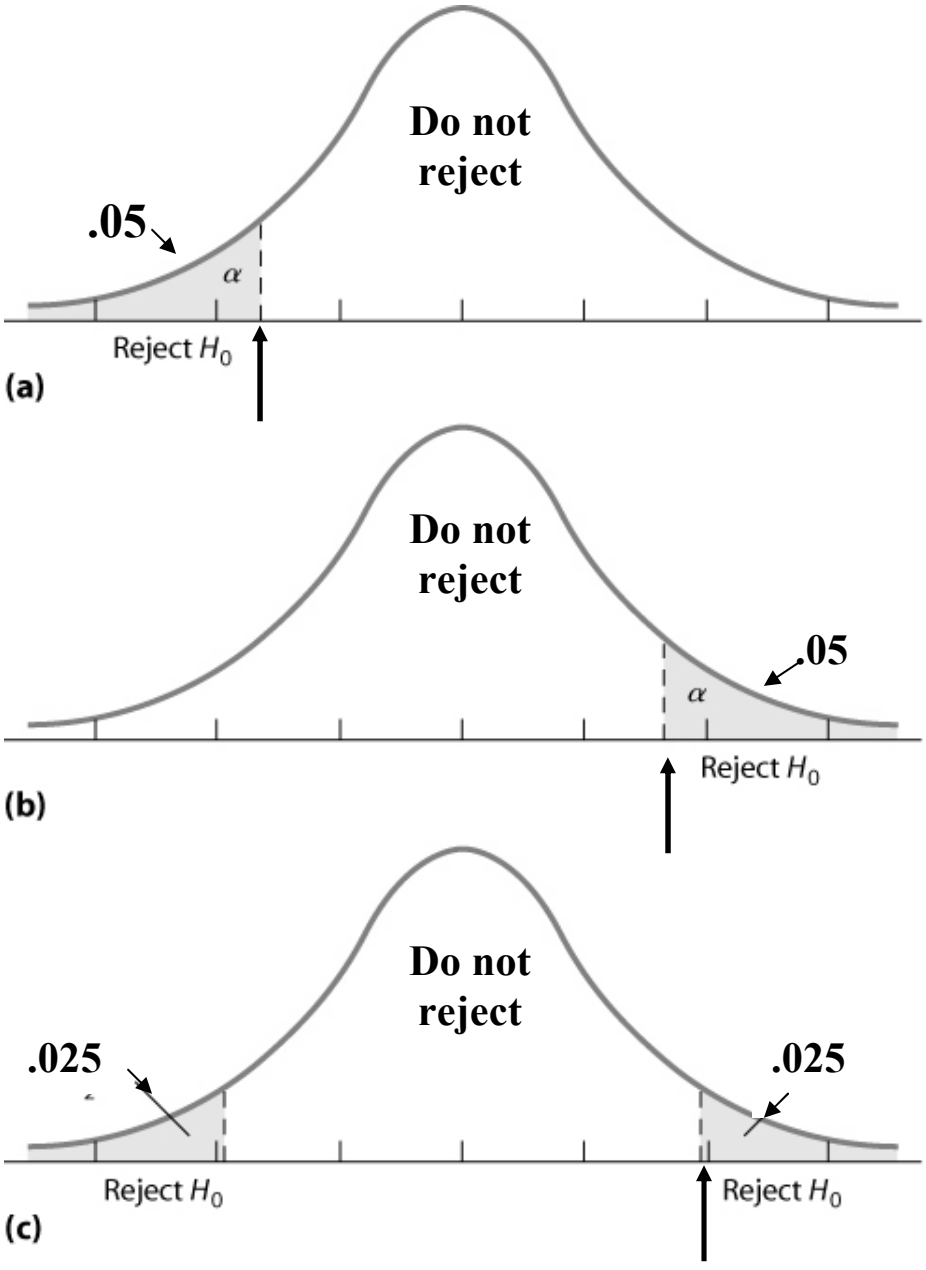
5. Select the right statistical test & execute it.

$$z = \frac{\text{sample mean (M)} - \text{hypothesized population mean } (\mu)}{\text{standard error between M \& } \mu}$$

$$z = \frac{\text{obtained difference}}{\text{difference due to chance}}$$

6. Calculate the probability, p, that our results could have been obtained if the null hypothesis were true in order to make a decision.

ONE –TAILED AND TWO-TAILED STATISTICAL TESTS



FOUR POSSIBLE OUTCOMES OF HYPOTHESIS TESTING

The Decision Matrix

		True state of affairs	
		H_0 is true	**H_0 is False**
Decision	**Reject the null**	Type I error α	Correct decision Power, 1-β
	Do not reject the null	Correct decision 1- α	Type II error β

Type I error (α):

- We erroneously reject the null hypothesis (H_0) when it is true. In experimental designs, we conclude that the population means are <u>not</u> equal, when they actually are, i.e. that the independent variable had an effect, when it actually did not.

Correct decision (1- α):

- We correctly do not reject the true H_0 when it is true. In experimental designs, we correctly conclude that the population means are =, when they actually are, i.e. that the IV did not have an effect, when it did not.

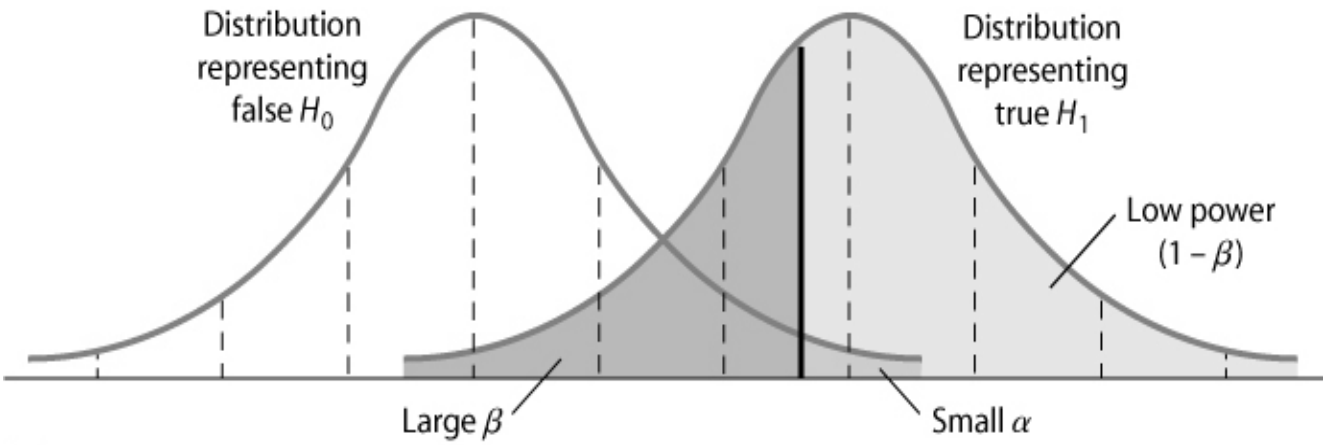
Correct decision, Power (1-β):

- We correctly reject the H_0 when it is false. We detect an effect that did occur. In experimental designs, we correctly conclude that the population means are =, when they indeed are, and that the IV had an effect, when it did.

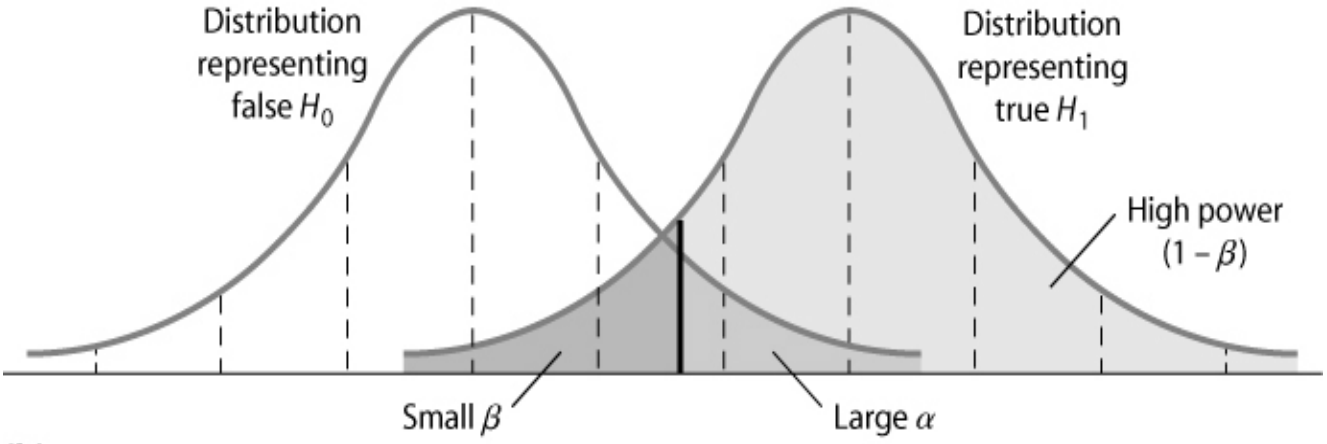
Type II error (β):

- We erroneously do <u>not</u> reject the H_0 when it is false. We fail to detect a significant result when that is actually present.
- In experimental designs, we wrongly conclude that the population means <u>are</u> equal, when they in fact are not, i.e. that the IV did <u>not</u> have an effect, when it actually did.

STATISTICAL POWER



(a)



(b)

THE DECISION MATRIX

TRUE STATE OF AFFAIRS

		Null Hypothesis True	**Null Hypothesis False**
DECISION	**Reject Null Hypothesis**	***Type I Error*** α	***Correct Decision (Power)*** $1 - \beta$
	Fail to Reject Null Hypothesis	***Correct Decision*** $1 - \alpha$	***Type II Error*** β

Name the type of error:

1. The dentist says you do not have any cavities, but you do.

2. The mechanic says you need a new battery and the battery is fine.

3. The smoke detector goes off and there is no fire.

4. The gas gauge says empty and the tank is empty.

IDENTIFYING TYPE I AND TYPE II ERRORS

A Investigators are researching a new vaccine to prevent the onset of AIDS in HIV positive subjects. 100 HIV+ subjects volunteered to participate. Half of the subjects (N = 50) were randomly assigned to Condition 1, which received the experimental vaccine. The other half (N = 50) were assigned to Condition 2, which received a placebo. Medically, the vaccine results looked very encouraging, but statistically significant differences between the groups were not attained. The researchers failed to reject the null hypothesis ($p < .01$), and the public is denied possible treatment.

Is the above scenario a potential Type I or Type II error? Why? (Hint: explain your decision by referring to the definitions of Type I or Type II errors)

Comment on the seriousness of committing a Type I or Type II error for this scenario.

B Educational consultants were hired by a progressive California school district to study the effectiveness of competitive vs. cooperative learning. The school district was known for its highly competitive philosophy, for which it was criticized, despite its successful student test scores on national assessments. The educational consultants researched several school districts nationwide and in Japan. They found that school districts that operated with a cooperative model had significantly higher test scores and higher student satisfaction than the competitive schools, i.e. investigators rejected the null hypothesis. The competitive school district is strongly considering a bond measure in the next election to fund a change to the cooperative model with estimated costs of approximately $250,000,000.

Is the above scenario a potential Type I or Type II error? Why?

Comment on the seriousness of committing a Type I or Type II error for this scenario.

C Silicone breast implants have been popular for many years for the purposes of breast reconstruction and breast enlargement. Since no evidence had been collected by the drug manufacturing company or the public, it was incorrectly assumed that they presented no harm to the public health, i.e. failure to reject the null hypothesis. Currently, due to reports of serious complications from thousands of women who elected to have this surgery, silicon implants are no longer being produced by major drug manufacturers.

Is the above scenario a potential Type I or Type II error? Why?

Comment on the seriousness of committing a Type I or Type II error for this scenario.

D In the 1930's, surgical procedures in which the neural connection between the prefrontal lobes was severed were performed frequently. Lobotomies were used on chronically hospitalized mental patients. It was considered a quick and easy treatment for mental disorder. We now know that lobotomized patients may sink into a confused stupor or become vegetable-like. (The null hypothesis was incorrectly rejected). Today, lobotomies are rarely performed in psychological clinical practice.

Is the above scenario a potential Type I or Type II error? Why?

Comment on the seriousness of committing a Type I or Type II error for this scenario.

E A researcher compared the number of cavities of children who had used either toothpaste brand X or toothpaste brand Y for a year. At the end of the year, the researcher found that the children who had used brand x had significantly fewer cavities than the children who had used brand Y. The difference was significant at the .05 level.

1. What is the null hypothesis?

2. What is the research hypothesis?

3. What would be the Type I error? What is the probability of a Type I error?

4. What would be the Type II error?

F The effect of a monetary incentive on performance on a cognitive task was investigated. The researcher predicted that greater monetary incentives would result in higher performance. Participants were told that they would receive 5¢, 25¢, or 50¢ for each word puzzle they correctly solved. A statistical test showed that there was a significant effect of monetary incentive on performance at the .01 level. The greater the incentive, the more puzzles solved.

1. What is the null hypothesis?

2. What is the research hypothesis?

3. What would be the Type I error? What is the probability of a Type I error?

4. What would be the Type II error?

G. A researcher investigated whether job applicants with popular names are viewed more favorably than equally qualified applicants with less popular names. Participants in one group read a resume of a job applicant with a popular name. Participants in the other group read the same resume, but the applicant had an unpopular name. There were five participants in each group. The results showed that the difference in evaluation of the applicants wasn't significant ($p=.22$, $\alpha=.05$).

1. What is the null hypothesis?

2. What is the research hypothesis?

3. What would be the Type I error? What is the probability of a Type I error?

4. What would be the Type II error?