Questions about the Assignment

The correlation (*r*) between two variables is a sample statistic

If we are testing to see if there is a correlation between height and salary, would it be a left-tail, right-tail, or two-tail test?

A hypothesis is a claim.

Race and Divorce:
Are white people more likely than non-white people to get divorced?

Does the likelihood of getting divorced differ by race?

Hypothesis Testing III

Statistical significance

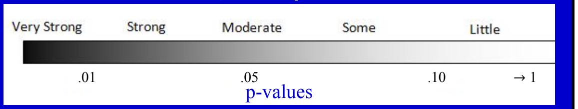
Errors

Power

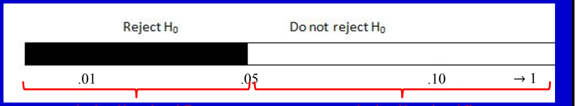
Significance and sample size

Statistical Conclusions

Strength of evidence against H_0:



Formal decision of hypothesis test [based on $\alpha = 0.05$]:



statistically significant
[$\text{p-value} < \alpha$]

not statistically significant
[$\text{p-value} \geq \alpha$]

Red Wine and Weight Loss

Rv, an ingredient in red wine and grapes, has been shown to promote weight loss in primates.

A sample of primates had various measurements taken before and after receiving Rv supplements for 4 weeks.

Red Wine and Weight Loss

In the test to see if the mean metabolism rate is higher after treatment, the p-value is 0.013.

Using $\alpha = 0.05$, is this difference statistically significant? (i.e., Should we reject H_0: which claims no difference?)

[✓ A. Yes] The p-value is lower than $\alpha = 0.05$, so the results are statistically significant and we reject H_0.

B. No

Red Wine and Weight Loss

In the test to see if the mean body mass is lower after treatment, the p-value is 0.013.

Using $\alpha = 0.01$, is this difference statistically significant? (i.e., Should we reject H_0: which claims no difference?)

A. Yes

[✓ B. No] The p-value is not lower than $\alpha = 0.01$, so the results are not statistically significant and we do not reject H_0.

Formal Decisions

Suppose many researchers around the world are all investigating the same research question. If the null hypothesis is true, using $\alpha = .05$, what proportion of hypothesis tests will incorrectly reject the null?

- A. None
- B. 95%
- [✓ C. 5%]
- D. It depends

5% of tests will get a p-value less than 0.05, just by random chance, if the null hypothesis is true

Errors

There are four possibilities:

Decision

		Reject H_0	Do not reject H_0
Truth	H_0 true	Type I Error	☺
	H_0 false	☺	Type II Error

Type I Error: Rejecting a true H_0

Type II Error: Failing to reject a false H_0

Red Wine and Weight Loss

In the test ($\alpha = .05$) to see if Rv is associated with metabolism rate, the p-value was 0.013 and we rejected H_0, which claimed that there was no association between Rv and metabolism rate.

If Rv actually is *not* associated with metabolism, a Type I Error would have been made. We rejected a true H_0.

In the test ($\alpha = .01$) to see if Rv is associated with body mass, the p-value is 0.013 and we did not reject the H_0, which claimed that there was not association between Rv and body mass.

If Rv actually is associated with body mass, a Type II Error would have been made. We failed to reject a false H_0.

Analogy to Law

H_o → innocent

H_a → guilty

A person is innocent until proven guilty.

Evidence must be beyond the shadow of a doubt.

α → the shadow of a doubt

Types of mistakes in a verdict?

- Convicting an innocent person ← Type I Error
- Releasing a guilty person ← Type II Error

Similarly, when there is not enough evidence to convict the defendant (i.e., accept H_a), the defendant is not declared innocent (i.e., H_0 is true), just not guilty (i.e., reject H_a).

Errors

Usually, we have no way of knowing whether an error has been made, unless we (or another researcher) conduct another study.

Similarly, we have no way of knowing whether our confidence interval actually contains the true population parameter.

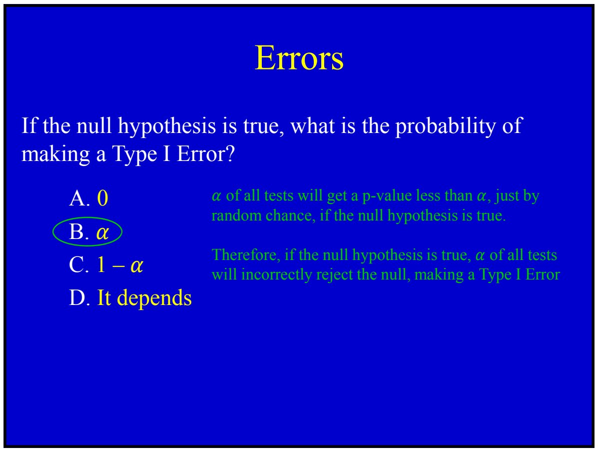
With a 95% confidence interval, 5% of the time our confidence interval will not contain the true population parameter.

Significance Level

Why would you use a smaller significance level, like $\alpha = 0.01$?

To make the test for significance more stringent.

Making α smaller reduces the likelihood of rejecting H_0 when it is actually true and H_a is not true (i.e., making a Type I Error).



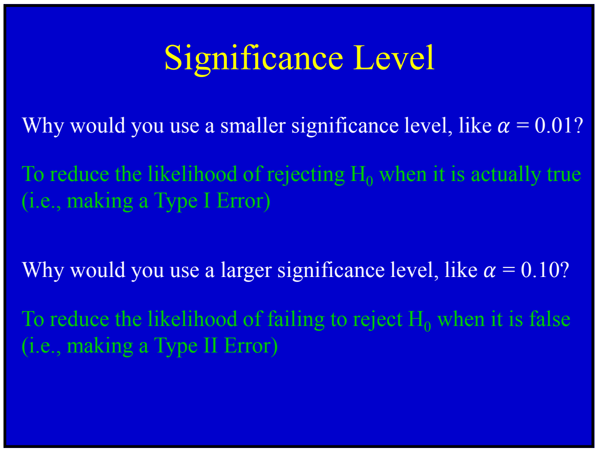
Errors

If the null hypothesis is true, what is the probability of making a Type I Error?

A. 0
[✓ B. α]
C. $1 - \alpha$
D. It depends

α of all tests will get a p-value less than α, just by random chance, if the null hypothesis is true.

Therefore, if the null hypothesis is true, α of all tests will incorrectly reject the null, making a Type I Error



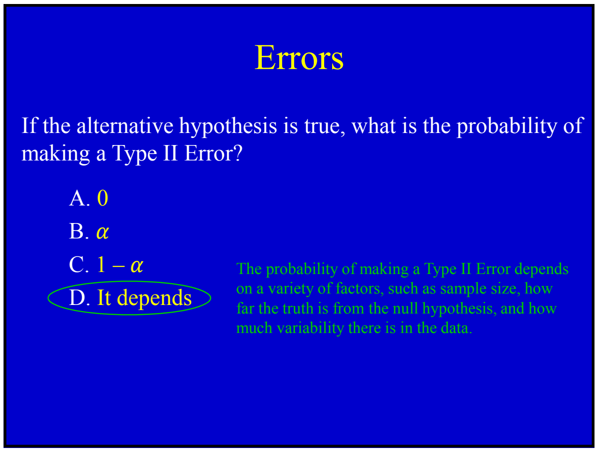
Significance Level

Why would you use a smaller significance level, like $\alpha = 0.01$?

To reduce the likelihood of rejecting H_0 when it is actually true (i.e., making a Type I Error)

Why would you use a larger significance level, like $\alpha = 0.10$?

To reduce the likelihood of failing to reject H_0 when it is false (i.e., making a Type II Error)

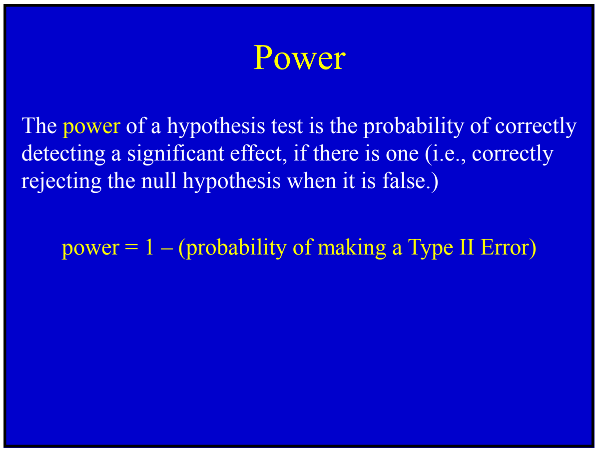


Errors

If the alternative hypothesis is true, what is the probability of making a Type II Error?

A. 0
B. α
C. $1 - \alpha$
[✓ D. It depends]

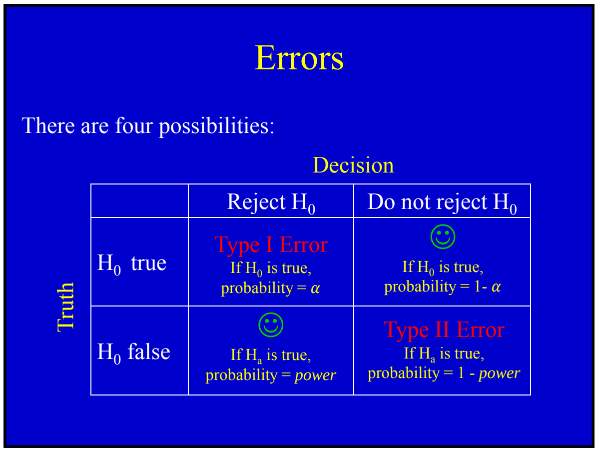
The probability of making a Type II Error depends on a variety of factors, such as sample size, how far the truth is from the null hypothesis, and how much variability there is in the data.



Power

The power of a hypothesis test is the probability of correctly detecting a significant effect, if there is one (i.e., correctly rejecting the null hypothesis when it is false.)

power = 1 – (probability of making a Type II Error)

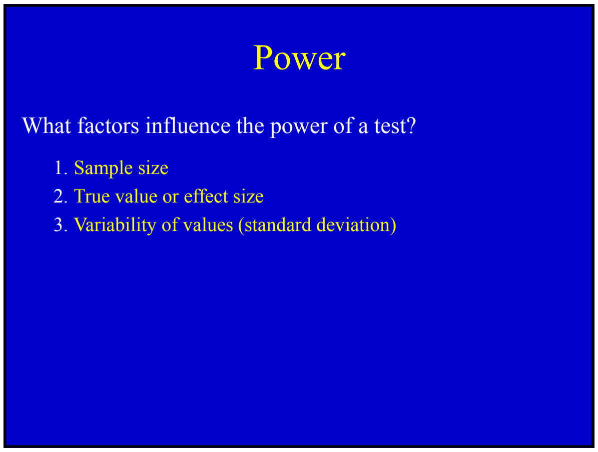


Errors

There are four possibilities:

Decision

		Reject H_0	Do not reject H_0
Truth	H_0 true	Type I Error If H_0 is true, probability = α	☺ If H_0 is true, probability = 1- α
	H_0 false	☺ If H_a is true, probability = *power*	Type II Error If H_a is true, probability = 1 - *power*



Power

What factors influence the power of a test?

1. Sample size
2. True value or effect size
3. Variability of values (standard deviation)

Power

What can you do to increase the power of your test?

[✓ A. Increase the sample size]
B. Make the true value farther from the null value
C. Decrease the standard deviation of your variables
D. Any of the above

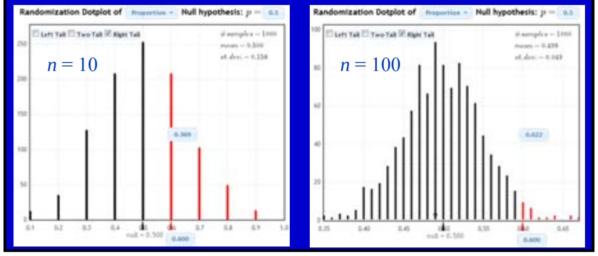
Usually, the only quantity you have control over is the sample size.
You certainly can't change the truth.
Sometimes there are ways to decrease the standard deviation, but not always.

Significance and Sample Size

www.lock5stat.com/statkey

$\hat{p} = 0.6$
$H_0: p = 0.5$
$H_a: p > 0.5$

Increasing the sample size decreases the standard deviation of the sampling distribution and reduces the proportion of sample statistics as extreme as our observed sample statistic which makes it easier to get significant results.



Significance and Sample Size

As the sample size decreases, the chance of making a Type II error (i.e., failing to reject H_0, when it is false)

A. decreases
[✓ B. increases]

As the sample size decreases, it becomes harder to get significant results, because more possibilities are plausible under random chance.

Significance and Sample Size

Failing to get significant results, does NOT mean H_0 is true.

This is particularly true for small sample sizes. Unless the true population parameter is very far from the null value, it is hard to find significant results if the sample size is small.

With small sample sizes, Type II Errors are very likely!

Statistical vs. Practical Significance

With small sample sizes, even large differences or effects may not be statistically significant.

With large sample sizes, even small differences or effects can be statistically significant.

A statistically significant result is not always practically significant, especially with large sample sizes.

Statistical vs. Practical Significance

Example: A weight loss program recruited 10,000 people for a randomized experiment.

A difference in average weight loss of 0.5lbs was found to be *statistically* significant.

Is a loss of ½ a pound *practically* significant?

Significance and Causation

The p-value alone tells you whether there is a significant association between two variables, but NOT whether this is a causal association.

The data collection method tells you whether causal conclusions can be made, but not whether an association is significant.

If the study is a randomized experiment AND the p-value indicates statistically significant results, then you can conclude that the explanatory variable has a causal effect on the response variable.

Constructing Hypothesis Tests for the Final Research Project

1. Construct research question
2. Define the parameter(s) of interest
3. State H_0 and H_a
4. Set significance level (α) *[usually 0.05 if unspecified]*
5. Collect data
6. Generate descriptive statistics
7. Calculate the appropriate observed sample statistic
8. Create a randomization sampling distribution (where H_0 is true)
9. Calculate the p-value of the observed sample statistic
10. Assess the strength of evidence against H_0
11. Make a formal decision based on the significance level
12. Interpret the conclusion in context

Assignment

Part I: 4.78, 4.120 and 4.130

Part II: Nothing for Part II

Summary

In making formal decisions, if the p-value is less than α, reject H_0 in favor of H_a; otherwise do not reject H_0.

There are two types of errors that can be made in hypothesis testing:
Rejecting a true null (Type I Error)
Failing to reject a false null (Type II Error).

Decreasing your significance level (α), reduces your risk of rejecting a true null.

Increasing your sample size, reduces your risk of failing to reject a false null.

Increasing your sample size, increases your chance of finding a significant result, if one exists.

p-value and Strength of Evidence

The smaller the p-value the…

[✓ A. stronger the evidence against the null]
B. stronger the evidence for the null
C. stronger the evidence against the alternative

If the p-value is small, then it would be very rare to get results as extreme as those observed, if the null hypothesis were true. This suggests that the null hypothesis is probably not true!

Videogames and GPA

If your roommate brings a videogame to college, will that lower your GPA?

What is the null and alternative hypothesis?

A. H_0: $p_v - p_{nv} = 0$, H_a: $p_v - p_{nv} < 0$
[✓ B. H_0: $\mu_v - \mu_{nv} = 0$, H_a: $\mu_v - \mu_{nv} < 0$]
C. H_0: $p_v - p_{nv} < 0$, H_a: $p_v - p_{nv} = 0$
D. H_0: $\mu_v - \mu_{nv} < 0$, H_a: $\mu_v - \mu_{nv} = 0$

GPA is quantitative, so we are dealing with means.
The equality is always in the null hypothesis.

Videogames and GPA

If your roommate brings a videogame to college, will that lower your GPA?

A study at Berea college conducted this test and obtained a p-value of 0.036. What does this mean?

A. The probability that H_0 is true is 0.036
B. The probability that H_0 is false is 0.036
C. The probability of seeing a difference in mean GPA as extreme as that in the observed sample is 0.036
[✓ D. The probability of seeing a difference in mean GPA as extreme as that in the observed sample is 0.036, if H_0 is true.]

This is the definition of the p-value.

Videogames and GPA

In the study about roommates bringing videogames to college and GPA, the p-value for the difference in mean GPA is 0.036. Using $\alpha = 0.05$, what would you conclude?

[✓ A. People with roommates who bring videogames have significantly lower GPAs.]
B. People with roommates who bring videogames do not have lower GPAs.
C. Nothing.

The p-value is less than 0.05 so the results are statistically significant.

Videogames and GPA

Based on this p-value, can you conclude that having a roommate who brings a videogame to campus *causes* you to have a lower GPA, on average?

A. Yes
B. No
[✓ C. It depends]

Yes, if the students were randomly assigned roommates. This would be an example of a natural randomized experiment and thus you could make conclusions about causality.

No, if the students could select their roommates. Since this would not be an randomized experiment, you cannot make conclusions about causality. Why?

Videogames and GPA

The study also tested whether students who bring a videogame to college themselves have lower GPAs, on average. The p-value of this test is 0.068. Using $\alpha = 0.05$, what would you conclude?

A. People who bring videogames have significantly lower GPAs.
B. People who bring videogames do not have lower GPAs.
C. In this study, the difference in mean GPA between students who bring videogames and those who don't is not statistically significant.
D. Nothing
[✓ E. Either C. or D.]

The results are not statistically significant because the p-value is greater than 0.05. If the results are not statistically significant, we can't conclude anything.

Red Wine and Weight Loss

In the test to see if locomotor activity changes after treatment, the p-value is 0.007.

Using $\alpha = 0.05$, is this difference statistically significant? (i.e., should we reject H_0: no difference?)

[✓ A. Yes]
B. No

The p-value is lower than $\alpha = 0.05$, so the results are statistically significant and we reject H_0.

Red Wine and Weight Loss

In the test to see if mean food intake changes after treatment, the p-value is 0.035.

Using $\alpha = 0.05$, is this difference statistically significant? (should we reject H_0: no difference?)

[✓ A. Yes]
B. No

The p-value is lower than $\alpha = 0.05$, so the results are statistically significant and we reject H_0.

Exam 1

- Exam 1:
  - In-class portion: Wednesday, 2/22
  - Lab portion: Thursday, 2/23
- In-class portion: (75%)
  - Open only to a calculator and one double sided page of notes *prepared by you*
  - Emphasis on conceptual understanding
- Lab portion: (25%)
  - Open to everything except communication of any form with other humans
  - Emphasis on actually analyzing data

Practice

- Last year's in-class and lab midterms, with solutions, are available on the course website (under documents)
- Full solutions to ALL the essential synthesis and review problems from Units 1 and 2 are available on the course website
- Doing problems is the key to success!!!

Keys to In-Class Exam Success

- ***Work lots of practice problems!***
- Take last year's exams under realistic conditions (time yourself, do it all before looking at the solutions, etc.)
- Prepare a good cheat sheet and use it when working problems
- Read the corresponding sections in the book if there are concepts you are still confused about

Keys to Lab Exam Success

- Primarily, make sure you know how to summarize, visualize, create an interval, and conduct a test for any one variable or relationship between two variables.
- Beyond that, make sure you are comfortable with the content from the labs
- Open-book does NOT mean you don't have to study. You will not have time to look up every command you need during the exam.