Week 6: Hypothesis Testing

4. Decision making with hypothesis testing

Stat 140 - 04

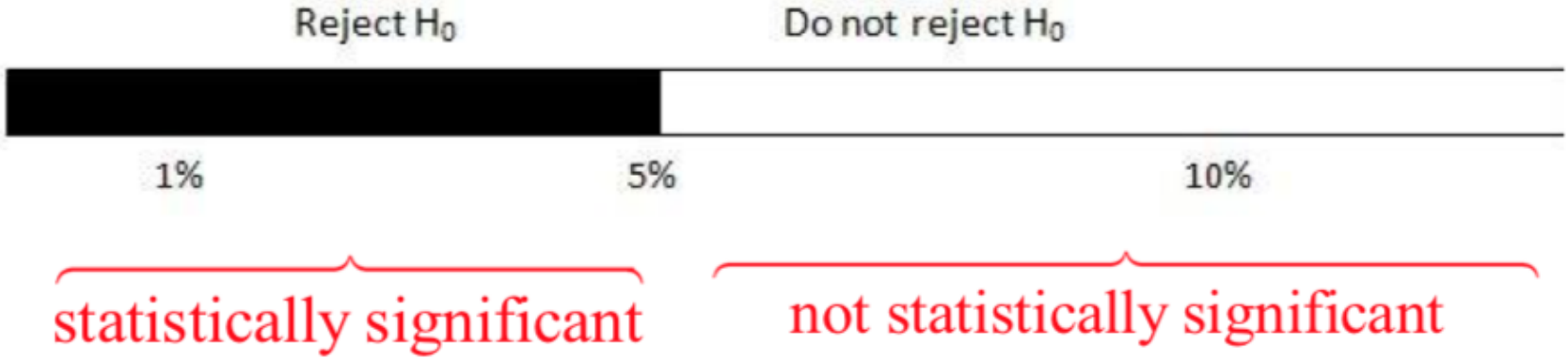
Mount Holyoke College

Dr. Shan Shan

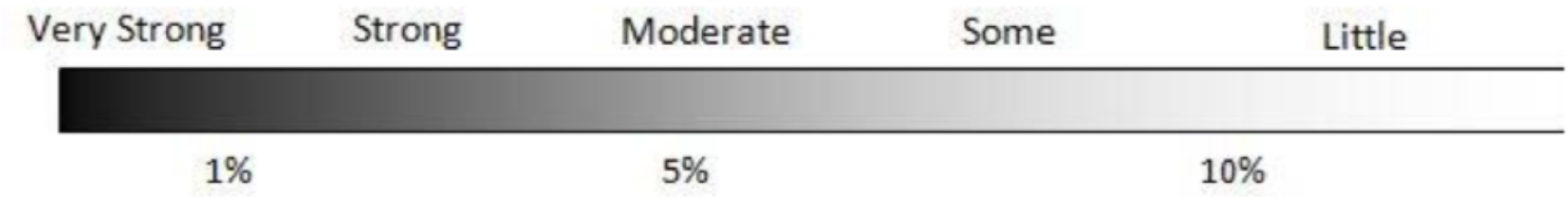
Slides posted at *http://sshanshans.github.io/stat140*

The smaller the p-value,
the stronger the evidence against H_0

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



Informal strength of evidence against H_0:



Fix a significance level α

- If the p-value is less than α
  - **Reject** H_0
  - the sample would be extreme if H_0 were true
  - the results are statistically significant
  - we have evidence for H_A
- If the p-value is greater than α
  - **Do not reject** H_0
  - the sample would not be too extreme if H_0 were true
  - the results are not statistically significant
  - the test is inconclusive; either H_0 or H_A may be true

Resveratrol, an ingredient in red wine and grapes, has been shown to promote weight loss in rodents, and has recently been investigated in primates (specifically, the Grey Mouse Lemur).

A sample of lemurs had various measurements taken before and after receiving resveratrol supplementation for 4 weeks

BioMed Central (2010, June 22). "Lemurs lose weight with 'life-extending' supplement resveratrol". Science Daily.

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

Poll question

Using $\alpha = 0.05$, is this difference statistically significant? (should we reject H_0?)

a. Yes
b. No

In the test to see if the mean body mass is lower after treatment, the p-value is 0.007.

Poll question

Using $\alpha = 0.05$, is this difference statistically significant? (should we reject H_0?)

a. Yes
b. No

In the test to see if locomotor activity changes after treatment, the p-value is 0.980.

Poll question

Using $\alpha = 0.05$, is this difference statistically significant? (should we reject H_0?)

a. Yes
b. No

"For the logical fallacy of believing that a hypothesis has been proved to be true, merely because it is not contradicted by the available facts, has no more right to insinuate itself in statistical than in other kinds of scientific reasoning..."

-Sir R. A. Fisher

- "Do not reject H_0" is not the same as "accept H_0"!
- Lack of evidence against H_0 is NOT the same as evidence for H_0!

Never Accept H_0

In a hypothesis test of

$$H_0 : \mu = 10 \text{ vs } H_a : \mu < 10$$

the p-value is 0.002.

Poll question

With $\alpha = 0.05$, what can we conclude?

a. Reject H0
b. Do not reject H0
c. Reject Ha
d. Do not reject Ha
e. We have insufficient evidence to conclude anything

In a hypothesis test of

$$H_0 : \mu = 10 \text{ vs } H_a : \mu < 10$$

the p-value is 0.002.

Poll question

With $\alpha = 0.05$, what can we conclude?

a. There is evidence that $\mu = 10$
b. There is evidence that $\mu < 10$
c. We have insufficient evidence to conclude anything

In a hypothesis test of

$$H_0 : \mu = 10 \text{ vs } H_a : \mu < 10$$

the p-value is 0.21.

Poll question

With $\alpha = 0.01$, what can we conclude?

a. Reject H0
b. Do not reject H0
c. Reject Ha
d. Do not reject Ha
e. We have insufficient evidence to conclude anything

In a hypothesis test of

$$H_0 : \mu = 10 \text{ vs } H_a : \mu < 10$$

the p-value is 0.21.

Poll question

With $\alpha = 0.01$, what can we conclude?

a. There is evidence that $\mu = 10$
b. There is evidence that $\mu < 10$
c. We have insufficient evidence to conclude anything

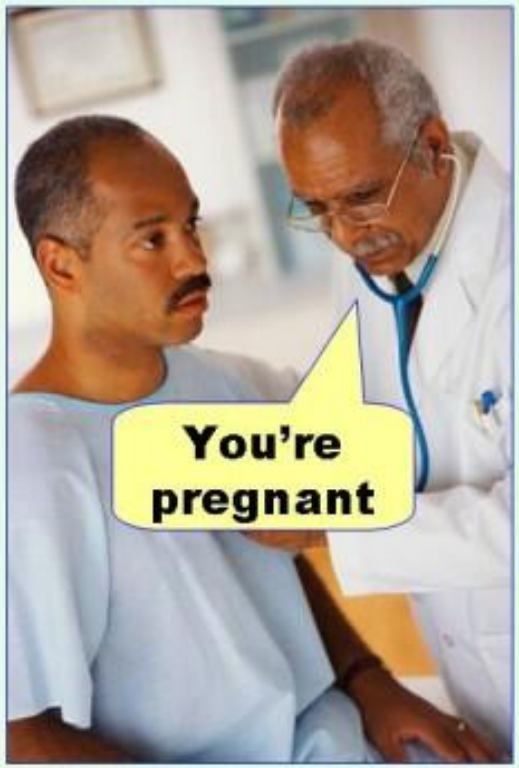
There are four possibilities:

Truth \ Decision	Reject H_0	Do not reject H_0
H_0 true	TYPE I ERROR	☺
H_0 false	☺	TYPE II ERROR

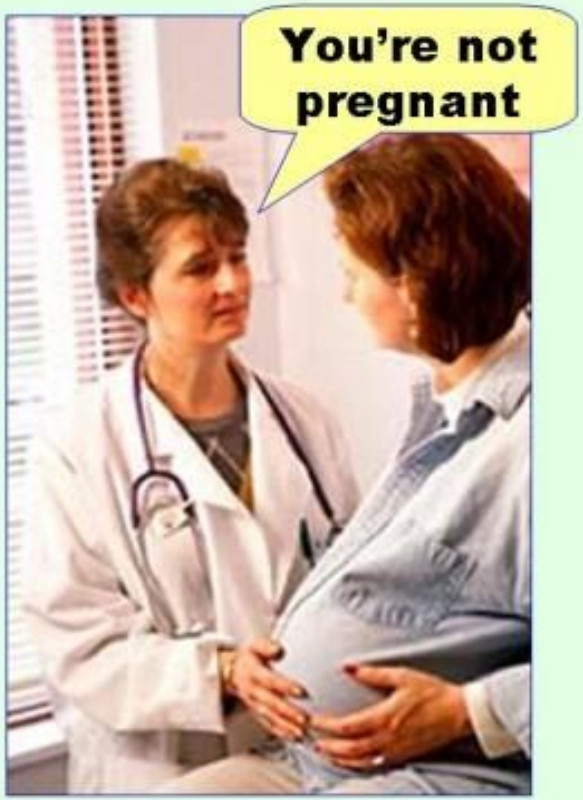
- A Type I Error is rejecting a true null (false positive)
- A Type II Error is not rejecting a false null (false negative)

- A person is innocent until proven guilty.
- Evidence must be beyond the shadow of a doubt.
- Types of mistakes in a verdict?
  - Convict an innocent (Type I Error)
  - Release a guilty (Type II Error)

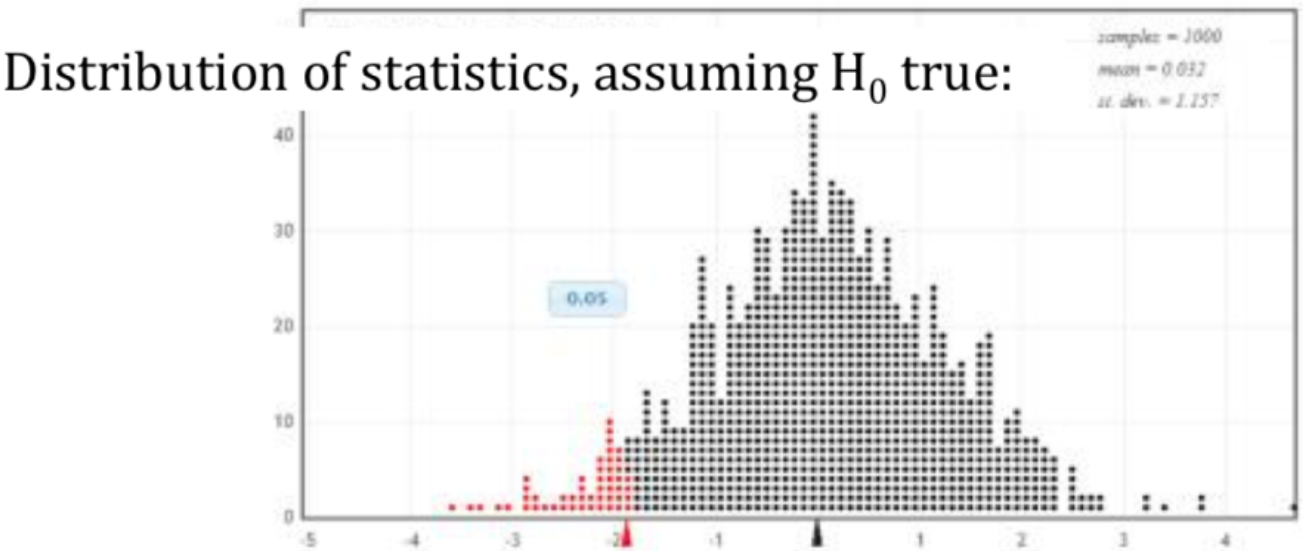
Type I error
(false positive)



Type II error
(false negative)



The probability of making a Type I error (rejecting a true null) is the significance level, α



If the null hypothesis is true:

- 5% of statistics will be in the most extreme 5%
- 5% of statistics will give p-values less than 0.05
- 5% of statistics will lead to rejecting H_0 at $\alpha = 0.05$
- If $\alpha = 0.05$, there is a 5% chance of a Type I error

How can we reduce the probability of making a Type II Error (not rejecting a false null)?

Option 1:

a. Decrease the significance level
b. Increase the significance level

Option 2:

a. Decrease the sample size
b. Increase the sample size

α should be chosen depending how bad it is to make a Type I or Type II error

- If a Type I error (rejecting a true null) is much worse than a Type II error, we may choose a smaller α, like $\alpha = 0.01$
- If a Type II error (not rejecting a false null) is much worse than a Type I error, we may choose a larger α, like $\alpha = 0.10$

By default, usually $\alpha = 0.05$

- With small sample sizes, even large differences or effects may not be significant
- With large sample sizes, even a very small difference or effect can be significant
- A statistically significant result is not always practically significant, especially with large sample sizes

Suppose a weight loss program recruits 10,000 people for a randomized experiment.

A difference in average weight loss of only 0.5 lbs could be found to be statistically significant

Suppose the experiment lasted for a year. Is a loss of 1/2 a pound practically significant?

http://www.newscientist.com/article/dn13754-breakfast-cereals-boost-chances-of-conceiving-boys.html

Are certain foods in your diet associated with whether or not you conceive a boy or a girl?

To study this, researchers asked women about their eating habits, including asking whether or not they ate 133 different foods regularly

A significant difference was found for breakfast cereal (mothers of boys eat more), prompting the headline "Breakfast Cereal Boosts Chances of Conceiving Boys".

Poll question

If a pregnant woman eats breakfast cereal every morning. Do you think this helped to boost the chances of having boys?

a. Yes
b. No
c. Impossible to tell