Math 107

Hypothesis Testing: Significance
(Sections 4.3 & 4.5)

Statistical Significance

p-value and H_0

Which of the following p-values gives the strongest evidence *against* H_0?

a) 0.005
b) 0.1
c) 0.32
d) 0.56
e) 0.94

p-value and H_0

Which of the following p-values gives the strongest evidence *against* H_0?

a) 0.22
b) 0.45
c) 0.03
d) 0.80
e) 0.71

p-value and H_0

Two different studies obtain two different p-values. Study A obtained a p-value of 0.002 and Study B obtained a p-value of 0.2. Which study obtained stronger evidence *against* the null hypothesis?

a) Study A
b) Study B

Formal Decisions

- If the p-value is small:
  - **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 not small:
  - **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

Significance Level

The **significance level**, α, is the threshold below which the p-value is deemed small enough to reject the null hypothesis

p-value $< \alpha \quad \Rightarrow$

p-value $\geq \alpha \quad \Rightarrow$

Red Wine and Weight Loss

- 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.

Red Wine and Weight Loss

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

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

a) Yes
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 0.007.

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

a) Yes
b) No

Red Wine and Weight Loss

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

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

a) Yes
b) No

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?)

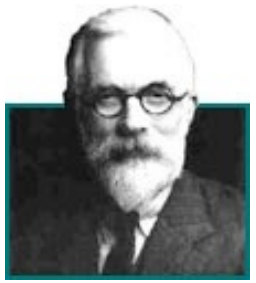
a) Yes
b) No

Never Accept H_0

- "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!

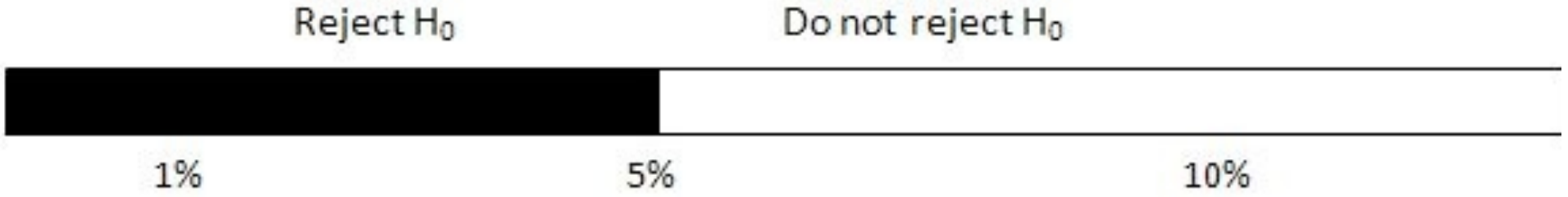
"*For the logical fallacy of believing that a hypothesis has been proved to be true, merely because it is not contradicted by the available facts, has no more right to insinuate itself in statistical than in other kinds of scientific reasoning…*"

-Sir R.A. Fisher

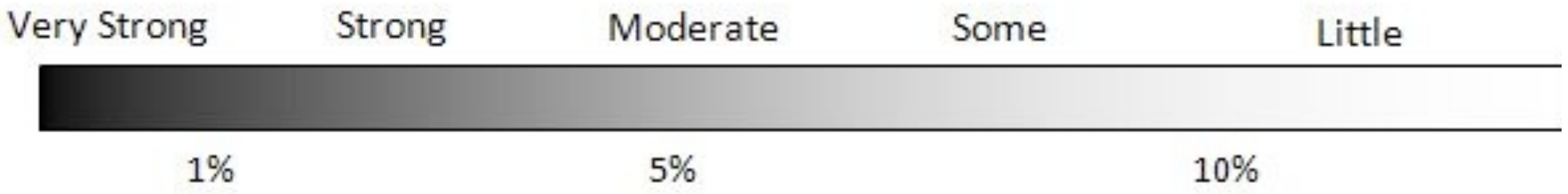


Statistical Conclusions

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



Informal strength of evidence against H_0:



Multiple Sclerosis and Sunlight

- It is believed that sunlight offers some protection against multiple sclerosis, but the reason is unknown
- Researchers randomly assigned mice to one of:
  - Control (nothing)
  - Vitamin D Supplements
  - UV Light
- All mice were injected with proteins known to induce a mouse form of MS, and they observed which mice got MS

Seppa, Nathan. "Sunlight may cut MS risk by itself", *Science News,* April 24, 2010 pg 9, reporting on a study appearing March 22, 2010 in the *Proceedings of the National Academy of Science.*

Multiple Sclerosis and Sunlight

For each situation below, write down

- *Null and alternative hypotheses*
- *Informal description of the strength of evidence against H_0*
- *Formal decision about H_0, using $\alpha = 0.05$*
- *Conclusion in the context of the question*

1. In testing whether UV light provides protection against MS (UV light vs control group), the p-value is 0.002.

2. In testing whether Vitamin D provides protection against MS (Vitamin D vs control group), the p-value is 0.47.

Statistical Errors

Errors

There are four possibilities:

		Decision	
		Reject H	Do not reject H
Truth	H		
	H		

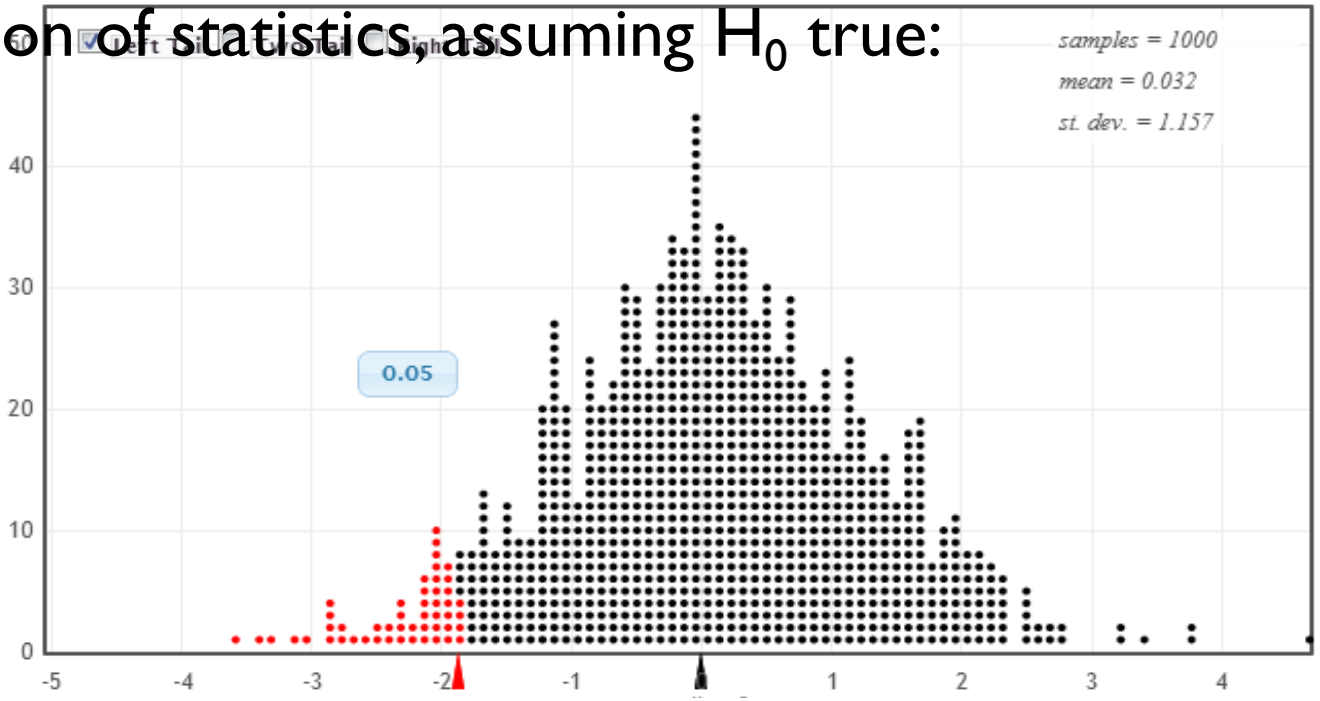
- A Type I Error is rejecting a true null (false positive)
- A Type II Error is not rejecting a false null (false negative)

Red Wine and Weight Loss

- In the test to see if resveratrol is associated with food intake, the p-value is 0.035.
  - ➡ If resveratrol ***is not*** associated with food intake, a **Type I Error** would have been made
- In the test to see if resveratrol is associated with locomotor activity, the p-value is 0.980.
  - ➡ If resveratrol ***is*** associated with locomotor activity, a **Type II Error** would have been made

Probability of Type I Error

Distribution of statistics, assuming H_0 true:

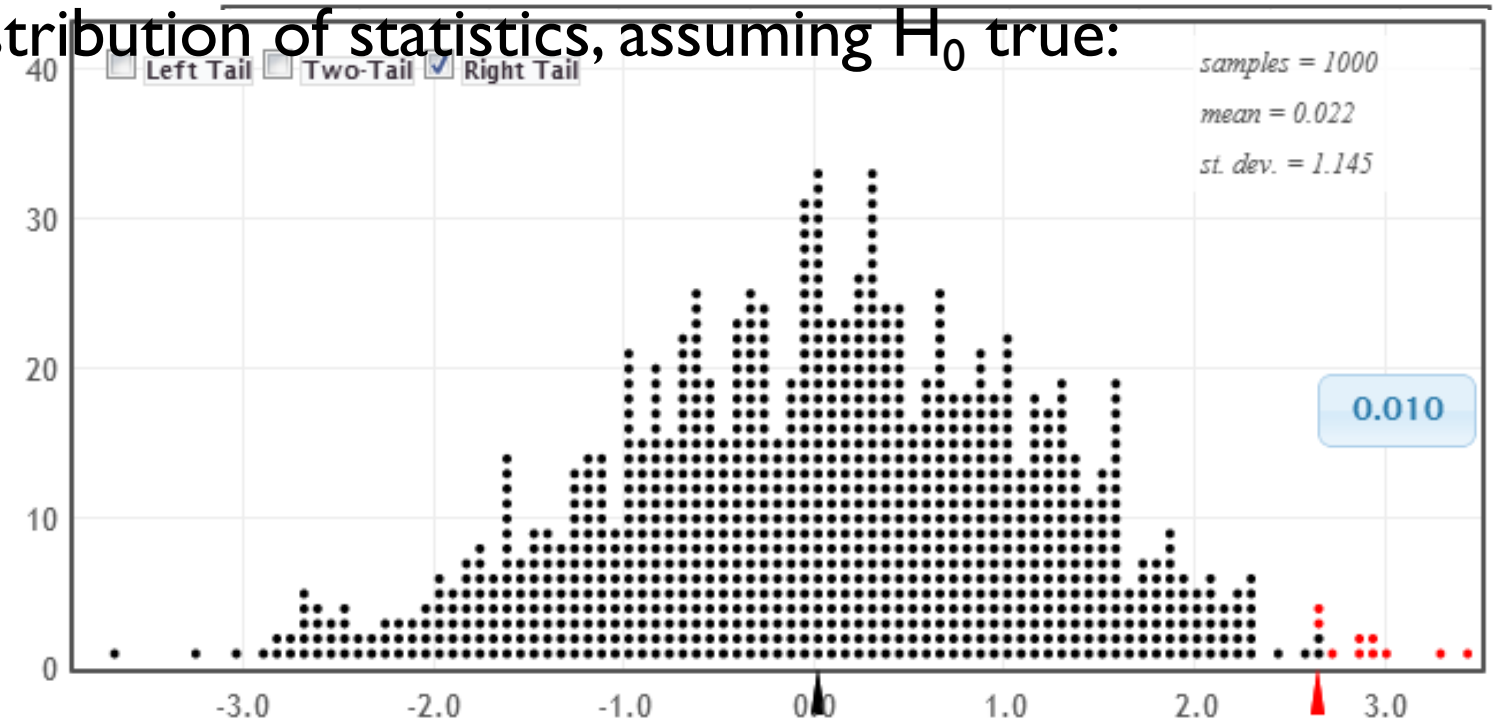


If the null hypothesis is true:

- 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

Probability of Type I Error

Distribution of statistics, assuming H_0 true:



If the null hypothesis is true:

- 1% of statistics will be in the most extreme 1%
- 1% of statistics will give p-values less than 0.01
- 1% of statistics will lead to rejecting H_0 at $\alpha = 0.01$
- If $\alpha = 0.05$, there is a 1% chance of a Type I error

Probability of Type II Error

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

Option 1: Decrease the significance level

Option 2: Increase the sample size

Probability of Errors

- The probability of making a Type I error (rejecting a true null) if the null is true is the significance level, α
- The probability of making a Type II error (not rejecting a false null) if the alternative is true depends on the significance level and the sample size (among other things)
- α should be chosen depending how bad it is to make a Type I or Type II error

Choosing α

- By default, usually $\alpha = 0.05$
- 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 $\alpha = 0.10$

Statistical vs Practical Significance

- 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

Statistical vs Practical Significance

- Example: 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 ½ a pound practically significant?