Section 4.3

Determining Statistical Significance

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

Formal Decisions

A formal hypothesis test has only two possible conclusions:

1. The p-value is small: reject the null hypothesis in favor of the alternative
2. The p-value is not small: do not reject the null hypothesis

How small?

Significance Level

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

$$\text{p-value} < \alpha \quad \Rightarrow \quad \text{Reject } H_0$$

$$\text{p-value} \geq \alpha \quad \Rightarrow \quad \text{Do not Reject } H_0$$

Significance Level

- If the p-value is less than α, the results are *statistically significant*, and we reject the null hypothesis in favor of the alternative
- If the p-value is not less than α, the results are not statistically significant, and our test is inconclusive
- Often $\alpha = 0.05$ by default, unless otherwise specified

Elephant Example

H_0 : X is an elephant
H_a : X is not an elephant

Would you conclude, if you get the following data?

Although we can never be certain!

- X walks on two legs

Reject H_0; evidence that X is not an elephant

- X has four legs

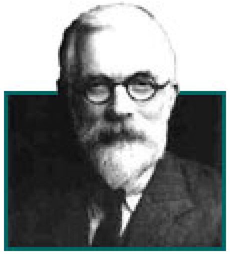
Do not reject H_0; we do not have sufficient evidence to determine whether X is an elephant

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



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 mean resting metabolic rate is higher after treatment, the p-value is 0.013.
- Using $\alpha = 0.05$,
  1) Is this difference statistically significant?
  2) Give a formal generic conclusion about H_0
  3) Give a conclusion in context

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

There is evidence that mean resting metabolic rate is higher after receiving resveratrol.

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 $\alpha = 0.05$,
  1) Is this difference statistically significant?
  2) Give a formal generic conclusion about H_0
  3) Give a conclusion in context

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

There is evidence that mean body mass is lower after receiving resveratrol.

Red Wine and Weight Loss

- In the test to see if locomotor activity changes after treatment, the p-value is 0.980.
- Using $\alpha = 0.05$,
    1) Is this difference statistically significant?
    2) Give a formal generic conclusion about H_0
    3) Give a conclusion in context

The p-value is not lower than $\alpha = 0.05$, so the results are not statistically significant and we do not reject H_0.

The data does not provide sufficient evidence to conclude that locomotor activity changes after treatment.

Red Wine and Weight Loss

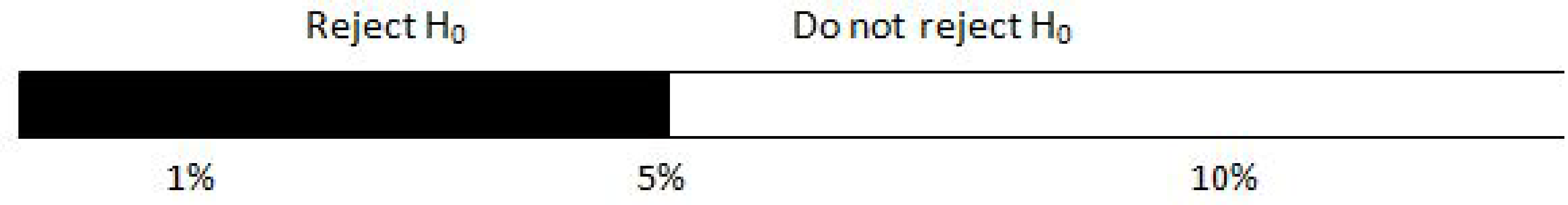
- In the test to see if the mean food intake changes after treatment, the p-value is 0.035.
- Using $\alpha = 0.05$,
  1) Is this difference statistically significant?
  2) Give a formal generic conclusion about H_0
  3) Give a conclusion in context

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

There is evidence that mean food intake is different for mice who after resveratrol.

Statistical Conclusions

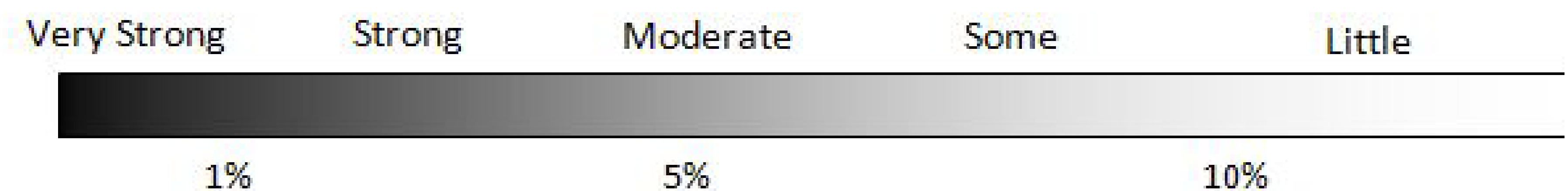
Formal decision of hypothesis test, based on $\alpha = 0.05$:



statistically significant

not statistically significant

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*

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

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

Multiple Sclerosis and Sunlight

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

- $H_0: p_{UV} - p_C = 0$
  $H_a: p_{UV} - p_C < 0$
- We have strong evidence against H_0
- Reject H_0
- We have strong evidence that UV light provides protection against MS, at least in mice.

Multiple Sclerosis and Sunlight

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

- $H_0: p_D - p_C = 0$
  $H_a: p_D - p_C < 0$
- We have little evidence against H_0
- Do not reject H_0
- We cannot conclude anything about Vitamin D and MS.

Errors

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
- A Type II Error is not rejecting a false null

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

Analogy to Law

H_o

H_a

A person is innocent until proven guilty.

Evidence must be beyond the shadow of a doubt.

p-value from data

α

Types of mistakes in a verdict?

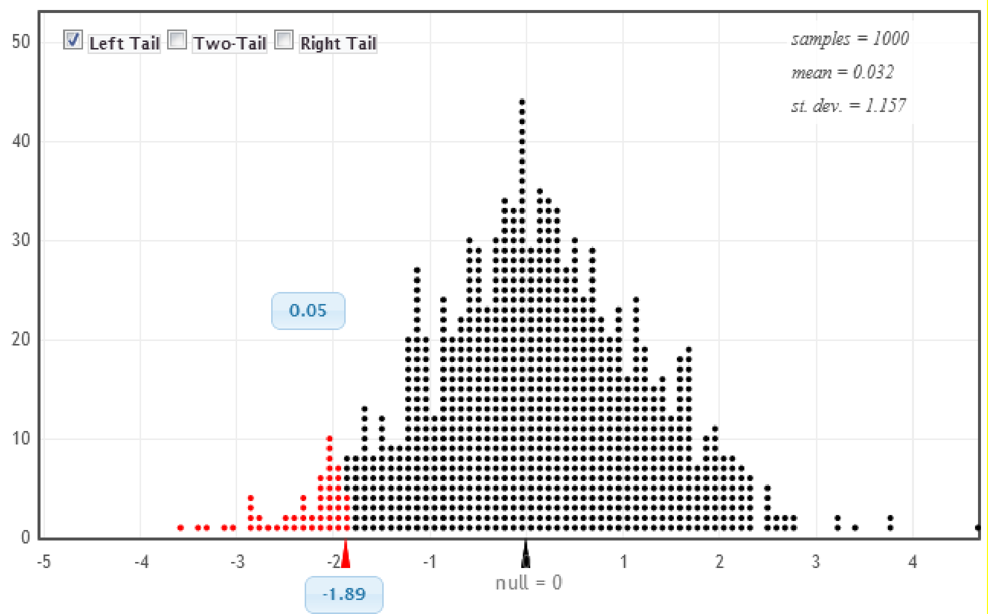
Convict an innocent ← Type I error

Release a guilty ← Type II error

Probability of Type I Error

- The probability of making a Type I error (rejecting a true null) is the significance level, α
- Randomization distribution, distribution of sample statistics if H_0 is true:

If H_0 is true and $\alpha = 0.05$, then 5% of statistics will be in tail (red), so 5% of the statistics will give p-values less than 0.05, so 5% of statistics will lead to rejecting H_0



Probability of Type II Error

- The probability of making a Type II Error (not rejecting a false null) depends on
  - Effect size (how far the truth is from the null)
  - Sample size
  - Variability
  - Significance level

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 $\alpha = 0.10$

Significance Level

Come up with a hypothesis testing situation in which you may want to…

- Use a smaller significance level, like $\alpha = 0.01$
- Use a larger significance level, like $\alpha = 0.10$

Summary

- Results are statistically significant if the p-value is less than the significance level, α
- In making formal decisions, reject H_0 if the p-value is less than α, otherwise do not reject H_0
- Not rejecting H_0 is NOT the same as accepting H_0
- There are two types of errors: rejecting a true null (Type I) and not rejecting a false null (Type II)