

Lecture 14: Hypothesis Testing Part I

Chapter 4.3

1 / 22

Goals for Today

- ▶ Introduce Hypothesis Testing Framework
- ▶ Testing Hypotheses Using Confidence Intervals
- ▶ Types of Errors
- ▶ Testing Hypotheses Using p-Values

2 / 22

Statistical Hypothesis Testing

3 / 22

Example

We flip a coin many times and start to suspect that it is biased:

- ▶ H_0 : the coin is fair. i.e. the probability of heads is $p = 0.5$
- ▶ H_A : the coin is not fair. i.e. $p \neq 0.5$

4 / 22

Crucial Concept: Conclusions of Hypothesis Tests

5 / 22

Analogy: US Criminal Justice System

In the criminal justice system, the jury's verdict does NOT make any statement about the defendant being **innocent**, rather that there was not enough evidence to prove beyond a reasonable doubt that they were guilty.

6 / 22

Analogy: US Criminal Justice System

Let's compare criminal trials to hypothesis tests:

Truth:

- ▶ Truth about the defendant: innocent vs guilty
- ▶ Truth about the hypothesis: H_0 or H_A

Decision:

- ▶ Verdict: not guilty vs guilty
- ▶ Test outcome: "Do not reject H_0 " vs "Reject H_0 "

7 / 22

Testing Hypotheses Using Confidence Intervals

Example on page 173: The average 10 mile run time for the Cherry Blossom Run in 2006 μ_{2006} was 93.29 min. Researchers suspect μ_{2012} was different:

- ▶ H_0 : average time was the same. i.e. $\mu_{2012} = 93.29$
- ▶ H_A : average time was different. i.e. $\mu_{2012} \neq 93.29$

8 / 22

Testing Hypotheses Using Confidence Intervals

9 / 22

Decision Errors

10 / 22

Decision Errors

- ▶ Trade-off between these two error rates
 - ▶ procedures with lower type I error rates typically have higher type II error rates
 - ▶ vice-versa
- ▶ In other words, there is almost never a procedure that makes no type I errors and no type II errors. Some sort of balance between the two is required

11 / 22

Example: US Criminal Justice System

Defendants must be proven “guilty beyond a reasonable doubt” i.e. in theory they would rather let a guilty person go free, than put an innocent person in jail. So let:

- ▶ H_0 : the defendant is innocent
- ▶ H_A : the defendant is guilty

thus “rejecting H_0 ” = guilty verdict. i.e. putting them in jail

In this case:

- ▶ Type I error is putting an innocent person in jail (considered worse)
- ▶ Type II error is letting a guilty person go free.

12 / 22

Example: Airport Screening

An example of where type II error is much more serious: [airport screening](#). Let:

H_0 : passenger X does not have a bomb/weapon

H_A : passenger X has a bomb/weapon

Failing to reject H_0 when H_0 is false corresponds to not “patting down” passenger X when they really have a bomb/weapon. This is disastrous.

Hence the long lines at airport security.

13 / 22

Significance Level

Hypothesis testing is built around rejecting or failing to reject the null hypothesis
i.e. we do not reject H_0 unless we have [strong evidence](#).

As a rule of thumb, when H_0 is true, we do not want to incorrectly reject H_0 more than 5% of the time.
i.e. $\alpha = 0.05 = 5\%$ is the [significance level](#).

With 95% confidence intervals from earlier, we expect it to miss the true population parameter 5% of the time. This corresponds to $\alpha = 0.05$.

14 / 22

Thought experiment: p-Values

Say you flip a coin you think is fair 1000 times. Thus you expect 500 heads. Say you observe

- ▶ 501 heads? Do you think the coin is biased?
- ▶ 525 heads? Do you think the coin is biased?
- ▶ 900 heads? Do you think the coin is biased?

Intuitively, a **p-value** quantifies how **extreme** an observation is given the null hypothesis.

The smaller the p-value, the more **extreme** the observation, where the meaning of extreme depends on the context.

15 / 22

p-Value Definition

The **p-value** or **observed significance level** is the probability of observing a test statistic as extreme or more extreme (in favor of the alternative) as the one observed, assuming H_0 is true.

It is **NOT** the probability of H_0 being true. This is the most common misinterpretation of the **p-value**.

16 / 22

Exercise 4.28 on Page 177 on Sleep

A poll found that college students sleep about 7 hours a night. Researchers suspect that Reedies sleep more. They use a sample of $n = 110$ Reedies to investigate this claim at an $\alpha = 0.05$ level.

Exercise 4.28 on Page 177 on Sleep

Exercise 4.28 on Page 177 on Sleep

Exercise 4.28 on Page 177 on Sleep

Exercise 4.28 on Page 177 on Sleep

21 / 22

Next Time

- More Hypothesis Testing

22 / 22