Lecture 12: Introduction to Hypothesis Testing

MATH 667-01
Statistical Inference
University of Louisville

October 24, 2017

Introduction

- We introduce hypothesis testing with important terminology and definitions in Section 8.1 of Casella and Berger (2002)[1].
- We also discuss types of errors associated with hypothesis tests and the power of a test as discussed in Section 8.3.
- Throughout this lecture, we illustrate the concepts in relation to the example introduced in Lecture 1.

[1] Casella, G. and Berger, R. (2002). *Statistical Inference, Second Edition*. Duxbury Press, Belmont, CA.

Terminology and Definitions

- *Definition L12.1* (Def 8.1.1 on p.373): A *hypothesis* is a statement about the population parameter.
- *Definition L12.2* (Def 8.1.2 on p.373): Two complementary hypotheses in a hypothesis testing problem are called the *null hypothesis* and the *alternative hypothesis*. They are denoted by H_0 and H_1, respectively.
- If θ is the population parameter and Θ is the parameter space, we write the hypotheses as $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^c$ where $\Theta = \Theta_0 \cup \Theta_0^c$.
- For example, in a judicial setting, we may have $\Theta = \{\text{innocent, guilty}\}$ and want to test H_0 : person is innocent versus H_1 : person is guilty.
- Or, if we have a parameter space $\Theta = (-\infty, \infty)$, we might test $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$.

A Hypothesis Testing Example

- *Example L12.1*: Suppose that there are ten marbles in a box, M of which are red and $10 - M$ of which are blue, and suppose that we will select a random sample of four marbles without replacement and observe the number of reds selected in the sample. We want to set up a hypothesis test to examine each of the following statements.

  (a) There are 6 red marbles in the box.

  (b) There are at least 6 marbles in the box.

  For each statement, what are the null and alternative hypotheses?

- *Answer to Example L12.1*: (a) $H_0 : M = 6$ versus $H_a : M \neq 6$ (b) $H_0 : M \geq 6$ versus $H_a : M < 6$

- *Definition L12.3* (Def 8.1.3 on p.374): A *hypothesis testing procedure* or *hypothesis test* is a rule that specifies:
    a. For which sample values the decision is made to accept H_0 as true.
    b. For which sample values H_0 is rejected and H_1 is accepted as true.

  The subset of the sample space for which H_0 will be rejected is called the *rejection region* or *critical region*. The complement of the rejection region is called the *acceptance region*.

A Hypothesis Testing Example

- *Example L12.2*: Suppose that there are ten marbles in a box, M of which are red and $10 - M$ of which are blue. Consider a hypothesis test procedure based on a random sample of four marbles selected without replacement where we reject $H_0 : M = 6$ versus $H_a : M \neq 6$ if and only if we observe either zero or four red marbles in the sample.
  (a) What is the acceptance region for this procedure?
  (b) What is the rejection region?
- *Answer to Example L12.2*: Let x be the observed number of red marbles selected in the sample. The null hypothesis is not rejected if x is in the acceptance region $\{1, 2, 3\}$.
  (b) The null hypothesis is rejected if x is in the rejection region $\{0, 4\}$.

- When performing a hypothesis test, we either make the decision to reject H_0 or to fail to reject H_0.
- So there are two possible types of errors:
  1. Rejecting H_0 when it is actually true.
  2. Failing to reject H_0 when H_a is true.
- The first is called a *Type I error*.
  The second is called a *Type II error*.
- The most common procedure in hypothesis testing is to control the Type I error (fix it to be a small positive number α).
- The **power** of a test to detect a particular alternative is the probability that a level α test will reject H_0 when the particular alternative value of the parameter is true; this is 1 – P(Type II error for the particular alternative value).

Four Possible Results of a Decision in a Significance Test

Reality About H_0	Decision: Fail to reject H_0	Decision: Reject H_0
H_0 is actually True	Correct Decision	Type I Error
H_0 is actually False	Type II Error	Correct Decision

- *Example L12.3*: Suppose that there are ten marbles in a box, M of which are red and $10 - M$ of which are blue. Consider a hypothesis test procedure based on a random sample of four marbles selected without replacement where we reject $H_0 : M = 6$ versus $H_a : M \neq 6$ if and only if we observe either zero or four red marbles in the sample.
  (a) What is the probability of a Type I error?
  (b) What is the probability of a Type II error when $M = 5$?
  (c) What is the power when $M = 2$?
- *Answer to Example L12.3*: Let X be the observed number of red marbles selected in the sample.
  (a)
$$\begin{aligned} P(\text{Type I error}) &= P_{M=6}(X = 0 \text{ or } X = 4) \\ &= P_{M=6}(X = 0) + P_{M=6}(X = 4) \\ &= \frac{\binom{6}{0}\binom{4}{4}}{\binom{10}{4}} + \frac{\binom{6}{4}\binom{4}{0}}{\binom{10}{4}} = \frac{1 \cdot 1}{210} + \frac{15 \cdot 1}{210} \\ &= \frac{16}{210} \approx .076 \end{aligned}$$

A Hypothesis Testing Example

- *Answer to Example L12.3(b) continued*:
  $P($Type II error when $M=5) = P_{M=5}(1 \le X \le 3)$
  $$= P_{M=5}(X=1) + P_{M=5}(X=2) + P_{M=5}(X=3)$$
  $$= \frac{\binom{5}{1}\binom{5}{3}}{\binom{10}{4}} + \frac{\binom{5}{2}\binom{5}{2}}{\binom{10}{4}} + \frac{\binom{5}{3}\binom{5}{1}}{\binom{10}{4}}$$
  $$= \frac{50}{210} + \frac{100}{210} + \frac{50}{210} = \frac{200}{210} \approx .952$$
- (c) The power when $M=2$ is

$$\begin{aligned} 1 - P_{M=2}(1 \le X \le 3) &= P_{M=2}(X=0 \text{ or } X=4) \\ &= P_{M=2}(X=0) \\ &= \frac{\binom{2}{0}\binom{8}{4}}{\binom{10}{4}} = \frac{1 \cdot 70}{210} = \frac{1}{3}. \end{aligned}$$

- *Definition L12.4* (Def 8.3.1 on p.383): The *power function* of a hypothesis test with rejection region R is the function of θ defined by $\beta(\theta) = P_\theta(\boldsymbol{X} \in R)$.
- *Definition L12.5* (Def 8.3.5 on p.385): For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a *size* α *test* if $\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$.
- *Definition L12.6* (Def 8.3.6 on p.385): For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a *level* α *test* if $\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$.

A Hypothesis Testing Example

- *Example L12.4*: Suppose that there are ten marbles in a box, M of which are red and $10 - M$ of which are blue and that we randomly select four marbles at random without replacement.
  (a) What is the power function of a test which rejects $H_0 : M = 6$ if and only if we observe either zero or four red marbles in the sample? What is the size of this test? Is this a level $.05$ test?
  (b) What is the power function of a test which rejects $H_0 : M \geq 6$ if and only if we observe either zero red marbles in the sample? What is the size of this test? Is this a level $.05$ test?

A Hypothesis Testing Example

- *Answer to Example L12.4*: (a) Here, we compute the power for all possible values of M. This can be done with the following command in R.

```
> dhyper(0,0:10,10:0,4)+dhyper(4,0:10,10:0,4)
 [1] 1.00000000 0.60000000 0.33333333 0.16666667
 [5] 0.07619048 0.04761905 0.07619048 0.16666667
 [9] 0.33333333 0.60000000 1.00000000
```

- So the power function is

$$\beta(M) = P_M(X=0 \text{ or } X=4) = \begin{cases} 1 & \text{if } M = 0, 10 \\ 3/5 & \text{if } M = 1, 9 \\ 1/3 & \text{if } M = 2, 8 \\ 1/6 & \text{if } M = 3, 7 \\ 8/105 & \text{if } M = 4, 6 \\ 20/21 & \text{if } M = 5 \end{cases} .$$

- The size of this test is $\frac{8}{105} = .076$, so it is not a level $.05$ test.

A Hypothesis Testing Example

- *Answer to Example L12.4 continued*: (b) Here, we compute the power for all possible values of M. This can be done with the following command in R.

```
> dhyper(0,0:10,10:0,4)
 [1] 1.000000000 0.600000000 0.333333333 0.166666667
 [5] 0.071428571 0.023809524 0.004761905 0.000000000
 [9] 0.000000000 0.000000000 0.000000000
```

- *Answer to Example L12.4(b) continued*: So the power function is $\beta(M) = P_M(X=0) = \begin{cases} 1 & \text{if } M=0 \\ 3/5 & \text{if } M=1 \\ 1/3 & \text{if } M=2 \\ 1/6 & \text{if } M=3 \\ 1/14 & \text{if } M=4 \\ 1/42 & \text{if } M=5 \\ 1/210 & \text{if } M=6 \\ 0 & \text{if } M \geq 7 \end{cases}.$
- The size of this test is $\sup_{M \geq 6} \beta(M) = \frac{1}{210} \approx .0048.$
  This is a level .05 test.
- The power function can be plotted using the R command

```
plot(dhyper(0,0:10,10:0,4),xlab="M",ylab="Power")
```

A Hypothesis Testing Example