Review of Statistics
Part B: Hypothesis Testing

Thomas Wiemann
University of Chicago

Econometrics
Econ 21020

Updated: April 13, 2022

Recap

In Part A of the statistics review discussed estimation:

- ▷ Developed estimators via the sample analogue principle;
- ▷ Characterized estimators with finite and large sample properties.

Our analysis highlighted that an estimator $\hat{\theta}_n$ is a random variable and may thus differ from the true (fixed) parameter θ.

In Part B, we consider the question of whether the true parameter is equal to a particular value or within a particular set.

- ▷ For example, when interested in the expected returns to education,

$$\tau = E_U[g(1, U) - g(0, U)|W = 1] \tag{1}$$

  we may be particularly curious about whether $\tau > 0$.

The formal analysis of such questions is known as *hypothesis testing*.

Outline

1. Hypothesis Testing
    - ▷ Definitions
    - ▷ Two-Sided Hypothesis Testing
    - ▷ One-Sided Hypothesis Testing
2. Hypothesis Testing and Confidence Intervals

These notes benefit greatly from the exposition in Wasserman (2003) and the lecture notes of Prof. Max Tabord-Meehan.

Outline

1. **Hypothesis Testing**
   - ▷ **Definitions**
   - ▷ Two-Sided Hypothesis Testing
   - ▷ One-Sided Hypothesis Testing

2. Hypothesis Testing and Confidence Intervals

These notes benefit greatly from the exposition in Wasserman (2003) and the lecture notes of Prof. Max Tabord-Meehan.

Hypothesis Testing

Our analysis begins with defining a hypothesis to be tested.

Let θ denote the parameter of interest and Θ its possible values.

Consider a partition of Θ into two disjoint subsets Θ_0 and Θ_1 and that we wish to test

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_1. \tag{2}$$

Some terminology:

- H_0 is referred to as the *null hypothesis*;
- H_1 is referred to as the *alternative hypothesis*;
- When $\Theta_0 = \theta_0$ is a single element, H_0 is a *simple hypothesis*;
- When Θ_0 is a non-singleton set, H_0 is a *composite hypothesis*.

Hypothesis Testing (Contd.)

Example 1

Let Y denote hourly wages and W denote being a college graduate. Do college graduates earn upwards of $600 a week?

To formulate a corresponding hypothesis, let $\mu_{Y|1} \equiv E[Y|W = 1]$. Then

$$H_0 : \mu_{Y|1} \geq 600 \quad \text{versus} \quad H_1 : \mu_{Y|1} < 600.$$

Here H_0 is a composite hypothesis.

If we had instead asked, "Do college graduates earn $600 a week?", the corresponding hypothesis would be

$$H_0 : \mu_{Y|1} = 600 \quad \text{versus} \quad H_1 : \mu_{Y|1} \neq 600.$$

Here H_0 is a simple hypothesis.

Hypothesis Testing (Contd.)

Hypotheses pose economic questions in terms of statistical parameters.

- ▷ Now need a procedure to answer these questions.

For this purpose, define a *test statistic* T_n, which denotes a *known* function of the sample $X_1, \ldots, X_n \sim X$.

- ▷ $T_n(X_1, \ldots, X_n)$ is a function of random variables and hence random.

Hypothesis testing finds an appropriate region $\mathcal{R} \subset \text{supp } T_n$ such that

$$\begin{aligned} T_n \in \mathcal{R} \quad &\Rightarrow \quad \text{reject } H_0, \\ T_n \notin \mathcal{R} \quad &\Rightarrow \quad \text{don't reject } H_0. \end{aligned}$$

$\mathcal{R}$ is known as the *rejection region*. We exclusively consider $\mathcal{R}$ of the form

$$\mathcal{R}(c) = \{t \in \mathbb{R} \mid t > c\}, \tag{3}$$

for a *critical value* $c \in \mathbb{R}$. Note: "large" T_n is evidence against H_0.

Type I and Type II Errors

Because T_n is random, we are bound to make errors at some point.

Outcomes of Hypothesis Testing

	Don't Reject H_0	Reject H_0
H_0 true	correct	type I error
H_0 false	type II error	correct

We will need to trade-off type I error and type II errors in our analysis.

- The less likely we make type I errors, the more likely are type II errors (and vice versa).
- We often focus on controlling the probability of a type I error.

Why? Wasserman (2003) has a nice analogy: "*Hypothesis testing is like a legal trial. We assume someone is innocent unless the evidence strongly suggests that they are guilty. Similarly, we don't reject H_0 unless there is strong evidence against H_0.*"

Type I and Type II Errors (Contd.)

A test is characterized by its type I and type II error probabilities.

Definition 1 (Size and Power)

The *size* (or: *significance level*) of a test is $\alpha \in (0, 1)$ such that

$$\begin{aligned} \alpha = P\left(T_n \in \mathcal{R}(c_\alpha) \mid H_0 \text{ is true}\right) &= P\left(T_n > c_\alpha \mid H_0 \text{ is true}\right) \\ = P\left(\text{reject } H_0 \mid H_0 \text{ is true}\right) &= P\left(\text{type I error}\right). \end{aligned} \tag{4}$$

The *power* of a test is defined as

$$\begin{aligned} P\left(\text{type II error}\right) &= P\left(\text{don't reject } H_0 \mid H_0 \text{ is false}\right) \\ &= P\left(T_n \leq c_\alpha \mid H_0 \text{ is false}\right) \\ &= P\left(T_n \notin \mathcal{R}(c_\alpha) \mid H_0 \text{ is false}\right). \end{aligned} \tag{5}$$

In practice, we choose a critical value c_α s.t. our test has the desired size.

▷ This controls the probability of a type I error.

Type I and Type II Errors (Contd.)

In practice, economists often consider a size of $\alpha = 0.05$ appropriate.

- ▷ This is rather arbitrary: Is $1/20$ rare enough?
- ▷ Practitioners may disagree on the size they would like to consider.

The next definition allows for side-stepping the issue of pre-specified sizes.

Definition 2 (p-Value)

The *p-value* of a test is defined as

$$\inf\{\alpha \in (0,1) \mid T_n \in \mathcal{R}(c_\alpha)\}, \tag{6}$$

that is, the smallest size of the test such that H_0 would be rejected.

Small p-values are interpreted as evidence against H_0:

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

Importantly: Large p-values are not evidence in favor of H_0!

- ▷ Large p-values may also occur because our test has low power.

Outline

Two-Sided Hypothesis Testing

Let's make things more concrete: Consider a sample $X_1, \ldots, X_n \overset{iid}{\sim} X$.

- ▷ Suppose we are interested a parameter $\theta \in \mathbb{R}$ (e.g., $\theta = E[X]$), and that we developed an estimator $\hat{\theta}_n$ such that

$$\frac{\left(\hat{\theta}_n - \theta\right)}{se(\hat{\theta}_n)} \xrightarrow{d} N(0, 1). \tag{7}$$

Is θ equal to a particular value, say, θ_0?

For this purpose, we consider testing

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0. \tag{8}$$

We are now in need of an appropriate test statistic T_n and a corresponding critical value c_α such that the size of our test is $\alpha \in (0, 1)$.

Two-Sided Hypothesis Testing (Contd.)

Given the standard normal limit of (7), a natural choice of test statistic is

$$T_n = \left| \frac{\left(\hat{\theta}_n - \theta_0\right)}{se(\hat{\theta}_n)} \right| \tag{9}$$

- ▷ Recall that we reject H_0 if T_n is "large".
- ▷ Here, T_n increases in deviations of $\hat{\theta}_n$ from θ_0: Seems sensible!

The following theorem shows that T_n is indeed a useful test statistic:

Theorem 1

Let $\hat{\theta}_n$ be an estimator for θ such that (7) *holds. Then for T_n defined by* (9)*, it hold that*

$$P(T_n > z_{1-\frac{\alpha}{2}} \,|\, H_0 \text{ is true}) \to \alpha, \tag{10}$$

where $z_{1-\frac{\alpha}{2}} = \Phi^{-1}(1 - \frac{\alpha}{2})$ is the $1 - \frac{\alpha}{2}$ quantile of a standard normal.

Two-Sided Hypothesis Testing (Contd.)

Proof.

$$P(T_n > c \mid H_0) = P\left(\left|\frac{\hat{\theta}_n - \theta_0}{se(\hat{\theta}_n)}\right| > c \,\Big|\, H_0\right)$$

$$= P\left(\frac{\hat{\theta}_n - \theta_0}{se(\hat{\theta}_n)} > c \,\Big|\, H_0\right) + P\left(\frac{\hat{\theta}_n - \theta_0}{se(\hat{\theta}_n)} < -c \,\Big|\, H_0\right)$$

$$= 1 - P\left(\frac{\hat{\theta}_n - \theta_0}{se(\hat{\theta}_n)} \le c \,\Big|\, H_0\right) + P\left(\frac{\hat{\theta}_n - \theta_0}{se(\hat{\theta}_n)} < -c \,\Big|\, H_0\right)$$

$$\to 1 - \Phi(c) + \Phi(-c), \quad \text{because: } \frac{\hat{\theta}_n - \theta_0}{se(\hat{\theta}_n)} \xrightarrow{d} N(0,1)$$

$$= 1 - \Phi(c) + (1 - \Phi(c)) = 2(1 - \Phi(c))$$

Recall:

$$z_{1-\frac{\alpha}{2}} \equiv \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

$$\Rightarrow \Phi(z_{1-\frac{\alpha}{2}}) = \Phi\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) = 1 - \frac{\alpha}{2}$$

Take $c = z_{1-\frac{\alpha}{2}}$, then $2(1 - \Phi(z_{1-\frac{\alpha}{2}})) = 2\left(1 - \left(1 - \frac{\alpha}{2}\right)\right)$

$$= 2\,\frac{\alpha}{2} = \alpha$$

□

Note: *It's worth memorizing that when* $\alpha = 0.05$, *we have* $z_{1-\frac{\alpha}{2}} \approx 1.96$.

Two-Sided Hypothesis Testing (Contd.)

Example 2

Consider the test statistic T_n defined in equation (9). By Theorem 1, we reject $H_0 : \theta = \theta_0$ at significance level α when

$$T_n > z_{1-\frac{\alpha}{2}}. \tag{11}$$

Hence, the p-value is given by

$$\Rightarrow \Phi(T_n) > \Phi(z_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$$

$$\Rightarrow \alpha > 2(1 - \Phi(T_n))$$

$$\Rightarrow 2(1 - \Phi(T_n)) \equiv p\text{-value}$$

Outline

One-Sided Hypothesis Testing

Instead of the simple hypothesis considered in (18), suppose we test

$$H_0 : \theta \leq \theta_0 \quad \text{versus} \quad H_1 : \theta > \theta_0, \tag{12}$$

or

$$H_0 : \theta \geq \theta_0 \quad \text{versus} \quad H_1 : \theta < \theta_0. \tag{13}$$

Recall that we want large T_n to be evidence *against* H_0.

- For $H_0 : \theta \leq \theta_0$, choose

$$T_n = \frac{\left(\hat{\theta}_n - \theta_0\right)}{se(\hat{\theta}_n)} \tag{14}$$

- For $H_0 : \theta \geq \theta_0$, choose

$$T_n = \frac{-\left(\hat{\theta}_n - \theta_0\right)}{se(\hat{\theta}_n)} \tag{15}$$

One-Sided Hypothesis Testing

The next result shows that these are indeed useful test statistics:

Theorem 2

Let $\hat{\theta}_n$ be an estimator for θ such that (7) holds. Then for T_n defined by (14), it hold that

$$P(T_n > z_{1-\alpha} | H_0 \text{ is true}) \to \alpha, \qquad (16)$$

where $z_{1-\alpha} = \Phi^{-1}(1-\alpha)$ is the $1-\alpha$ quantile of a standard normal. An analogous result holds for T_n defined by (15).

Proof.

$$P\left(\frac{\hat{\theta}_n - \theta_0}{se(\hat{\theta}_n)} > c \,\Big|\, H_0\right) = 1 - P\left(\frac{\hat{\theta}_n - \theta_0}{se(\hat{\theta}_n)} \leq c \,\Big|\, H_0\right)$$

$$\to 1 - \Phi(c), \quad \text{because: } \frac{\hat{\theta}_n - \theta_0}{se(\hat{\theta}_n)} \xrightarrow{d} N(0,1)$$

Taking $c = z_{1-\alpha}$, implies $1 - \Phi(z_{1-\alpha}) = 1 - (1-\alpha) = \alpha$ □

***Note**: It's worth memorizing that when $\alpha = 0.05$, we have $z_{1-\alpha} \approx 1.64$.*

One-Sided Hypothesis Testing

Example 3

Consider the test statistic T_n defined in equation (14). By Theorem 2, we reject $H_0 : \theta = \theta_0$ at significance level α when

$$T_n > z_{1-\alpha}. \tag{17}$$

Hence, the p-value is given by

$$\Phi(T_n) > \Phi(z_{1-\alpha}) = 1-\alpha$$

$$\Rightarrow \quad \alpha > 1 - \Phi(T_n)$$

$$\Rightarrow \quad 1 - \Phi(T_n) \equiv p\text{-value}$$

Outline

1. Hypothesis Testing
   ▷ Definitions
   ▷ Two-Sided Hypothesis Testing
   ▷ One-Sided Hypothesis Testing
2. **Hypothesis Testing and Confidence Intervals**

Hypothesis Testing and Confidence Intervals

Consider the following thought experiment: Suppose you test

$$H_0 : \theta = \tilde{\theta}_0 \quad \text{versus} \quad H_1 : \theta \neq \tilde{\theta}_0, \tag{18}$$

for all possible values $\tilde{\theta}_0 \in \Theta$ using a test of size α.

- ▷ Whenever H_0 is not rejected, you write down the value of $\tilde{\theta}_0$.
- ▷ This gives the set (say, C_n) of $\tilde{\theta}_0$ for which H_0 would not be rejected.
- ▷ C_n summarizes the collection of hypotheses we would not reject.

It turns out that this newly constructed set C_n *is* the confidence interval discussed in Part A of the statistics review!

- ▷ This is known as the *duality between hypothesis testing and confidence intervals*.

This implies that we can use a $1 - \alpha$ confidence interval to test hypothesis at a significance level α.

- ▷ Step 1: Construct the $1 - \alpha$ confidence interval c_n;
- ▷ Step 2: Check whether $\theta_0 \in c_n$. If not, reject $H_0 : \theta = \theta_0$.

Hypothesis Testing and Confidence Intervals (Contd.)

To see this dual relationship, recall that we would include $\tilde{\theta}_0$ in the set C_n if our test of size α does not reject $H_0 : \theta = \tilde{\theta}_0$. That is, whenever

$$T_n \leq c_\alpha. \tag{19}$$

Take T_n as defined in Equation (9) so that $c_\alpha = z_{1-\frac{\alpha}{2}}$. Then

$$\left| \frac{\hat{\theta}_n - \theta_0}{se(\theta_n)} \right| \leq z_{1-\frac{\alpha}{2}} \Rightarrow -z_{1-\frac{\alpha}{2}} \leq \frac{\hat{\theta}_n - \theta_0}{se(\theta_n)} \leq z_{1-\frac{\alpha}{2}}$$

$$\Rightarrow \hat{\theta}_n - z_{1-\frac{\alpha}{2}} se(\hat{\theta}_n) \leq \theta_0 \leq \hat{\theta}_n + z_{1-\frac{\alpha}{2}} se(\hat{\theta}_n)$$

Hence, the set of $\tilde{\theta}_0$ for which we don't reject H_0 at significance level α is

$$C_n = \left[\hat{\theta}_n - z_{1-\frac{\alpha}{2}} se(\hat{\theta}_n), \hat{\theta}_n + z_{1-\frac{\alpha}{2}} se(\hat{\theta}_n) \right]$$

which is identical to our definition of the symmetric confidence interval.

Summary

This concludes our statistics review:

- ▷ Discussed the construction of estimators;
- ▷ Introduced tools to study the properties of estimators;
- ▷ Developed procedure for testing hypothesis about parameters.

Now we're fully equipped to delve into the analysis of causal questions!

- ▷ Can leverage our probability expertise for Task 1 (Definition) and Task 2 (Identification).
- ▷ Can leverage our statistics expertise for Task 3 (Estimation).

To get things started properly, we revisit the returns to education example in the next lecture.

References

Wasserman, L. (2003). *All of statistics*. Springer.