Hypothesis Testing of Regression Parameters

Tadao Hoshino （星野匡郎）

ver. 2018 Fall Semester

Review of the Previous Lecture

- Consider a linear regression model

$$Y_i = \beta_{00} + X_i\beta_{01} + \varepsilon_i, \quad i = 1, ..., n$$

where (β_{00}, β_{01}) are the "true" regression coefficients such that

$$E(Y|X) = \beta_{00} + X\beta_{01} \ \ (\iff E(\varepsilon|X) = 0).$$

- The OLS estimator of (β_{00}, β_{01}):

$$\hat{\beta}_{n0} = \bar{Y}_n - \bar{X}_n\hat{\beta}_{n1}$$
$$\hat{\beta}_{n1} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_iY_i - \bar{X}_n\bar{Y}_n}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2}$$

Under certain conditions, we can show that

- the OLS estimator is **unbiased**:

$$E(\hat{\beta}_{n0}) = \beta_{00}, \;\; E(\hat{\beta}_{n1}) = \beta_{01}$$

- the OLS estimator is **consistent**:

$$\hat{\beta}_{n0} \xrightarrow{P} \beta_{00}, \;\; \hat{\beta}_{n1} \xrightarrow{P} \beta_{01}$$

  and

- as n increases to infinity, $\sqrt{n}(\hat{\beta}_{n1} - \beta_{01})$ is **distributed as normal** with mean 0 and variance $V(X)^{-1}\sigma^2$:

$$\sqrt{n}(\hat{\beta}_{n1} - \beta_{01}) \xrightarrow{d} N\left(0, \frac{\sigma^2}{V(X)}\right).$$

Hypothesis Testing: When the Variance Is Known

Hypothesis Testing with Known Variance

- We continue to consider the case of a simple linear regression:

$$Y_i = \beta_{00} + X_i\beta_{01} + \varepsilon_i, \;\; i = 1, ..., n$$

- We would like to know whether the explanatory variable X is actually a determinant of Y.
- Note that the estimate $\hat{\beta}_{n1}$ inevitably contains some estimation error under finite sample size.

We need to show that the true regression coefficient β_{01} (i.e., the marginal effect of X on Y) is not equal to zero.

- We first consider the case where the value of the variance $V(X)^{-1}\sigma^2$ is assumed to be known.

Hypothesis Testing with Known Variance

- First, note that

$$\sqrt{n}(\hat{\beta}_{n1} - \beta_{01}) \xrightarrow{d} N(0, v) \text{ is equivalent to } \frac{\hat{\beta}_{n1} - \beta_{01}}{\sqrt{v/n}} \xrightarrow{d} N(0, 1)$$

  where $v = V(X)^{-1}\sigma^2$.
- The above result implies that, as n increases, the variance of $\frac{\hat{\beta}_{n1} - \beta_{01}}{\sqrt{v/n}}$ converges to one.
- Recall that for any random variable X, the variance of $\frac{X - E(X)}{\sqrt{V(X)}}$ equals to one, where $\sqrt{V(X)}$ is the standard deviation of X.
- Thus, we can interpret $\sqrt{v/n}$ as the standard deviation of $\hat{\beta}_{n1}$.
- The standard deviation of an estimator is called the standard error.

Hypothesis Testing with Known Variance

- For simplicity, we write the standard error of $\hat{\beta}_{n1}$ as

$$se = \sqrt{v/n}.$$

- The standard error se is a measure of the expected dispersion of the OLS slope estimator $\hat{\beta}_{n1}$ around the true β_{01}. The smaller the standard error, the more efficient the estimator.
- Note that the standard error is proportional to $n^{-1/2}$.
- Hence, if we want to double the accuracy of the estimator (i.e., halve the standard error), the sample size needs to be (not doubled but) quadrupled.[1]

[1] This would be intuitively understandable. Increasing the sample size from 100 to 1000 may produce a drastic improvement in accuracy of the estimate, but increasing from 10000 to 11000 will only lead to a small perturbation in the estimate.

Hypothesis Testing with Known Variance

- We define the following statistic:

$$Z_n = \frac{\hat{\beta}_{n1} - \beta_{01}}{se}$$

  Since we have assumed that se is known, the only unknown parameter is β_{01}.

- If β_{01} were known, Z_n is computable, and because $Z_n \xrightarrow{d} N(0,1)$, we have

$$\Pr(-1.9599 \leq Z_n \leq 1.9599) \approx 0.95$$

  for sufficiently large n.

- In other words, the probability of occurring an event $\{|Z_n| > c\}$ for some $c \geq 1.9599$ is at most only 5%.

Hypothesis Testing with Known Variance

- However, since β_{01} is unknown in reality, we hypothesize that

$$\beta_{01} = 0$$

  The hypothesis states that the explanatory variable X is not a determinant of the dependent variable Y.

- This hypothesis is called the null hypothesis and is often referred to as H_0. The null hypothesis is usually expected to be false.
- The negation of the null hypothesis is called the alternative hypothesis, which is often referred to as H_1.

$$H_0 : \beta_{01} = 0, \;\; H_1 : \beta_{01} \neq 0$$

Hypothesis Testing with Known Variance

- If H_0 is true, the statistic Z_n can be simplified as

$$Z_n = \frac{\hat{\beta}_{n1} - \beta_{01}}{se} = \frac{\hat{\beta}_{n1}}{se}$$

- Thus, under H_0, we must have

$$\Pr\left(-1.9599 \leq \frac{\hat{\beta}_{n1}}{se} \leq 1.9599\right) \approx 0.95$$

That is, the probability of occurring an event $\{|\hat{\beta}_{n1}/se| > 1.9599\}$ must be about 5% under H_0.

Hypothesis Testing with Known Variance

- In general, let T_n be a test statistic:

$$T_n = T_n((Y_1, \mathbf{X}_1), ..., (Y_n, \mathbf{X}_n))$$

The hypothesis testing consists of the decision rule

Reject the null hypothesis H_0 if $T_n > c$.

- The threshold value c is called the critical value, which is pre-specified by researcher.
- In the above example, $T_n = |\hat{\beta}_{n1}/se|$, and $c = 1.9599$.

Framework of Hypothesis Testing

- Suppose that for a number $\alpha \in (0, 1)$ we have

$$\Pr(T_n > c_\alpha) = \alpha$$

  under the null hypothesis H_0.[2]

- Thus, if the computed value of T_n is larger than c_α, assuming that H_0 is true, such event occurs with probability at most α.

  $=>$ If we observe $\{T_n > c_\alpha\}$ for small enough α, we can conclude the H_0 is unacceptable (unrealistic).

- This threshold probability α is referred to as the significance level, and then we say the null hypothesis H_0 is rejected at α level of significance.

---

[2]Once α is given, the value of c_α is automatically determined. (e.g., $\alpha = 0.05 \Rightarrow c_\alpha = 1.9599$.)

Framework of Hypothesis Testing

- In econometrics, researchers often use either the 0.05 level (5% level) or the 0.01 level (1% level) of significance.
- Are 5% and 1% probability small enough to conclude as unrealistic? $=>$ The choice of significance level is quite subjective, and depends on the researcher's judgment.
- In our example, if $|\hat{\beta}_{n1}/se|$ is larger than 1.9599, we can reject the null hypothesis $H_0 : \beta_{01} = 0$ at the 5% significance level:

$$\text{If } \underbrace{|\hat{\beta}_{n1}/se|}_{T_n} > \underbrace{1.9599}_{c_{0.05}} \Rightarrow \text{At the 5\% significance level, } X \text{ affects } Y.$$

- Similarly,

$$\text{If } \underbrace{|\hat{\beta}_{n1}/se|}_{T_n} > \underbrace{2.5758}_{c_{0.01}} \Rightarrow \text{At the 1\% significance level, } X \text{ affects } Y.$$

Procedure of hypothesis testing

STEP 1. Decide the significance level α. (usually either 5% or 1%)

STEP 2. Develop a null hypothesis H_0, and compute the test statistic T_n under H_0.

STEP 3. Assuming that H_0 is true, compute the critical value c_α.

STEP 4. If $\{T_n > c_\alpha\}$ is observed, reject H_0 at α level of significance, otherwise accept H_0.

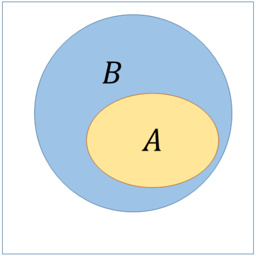
Hypothesis Testing as a "Proof by Contraposition"

Proof by contraposition

Suppose we want to prove the statement $A \Rightarrow B$.[a]
Assume $\neg B$, which is the logical negation of B, and show that $\neg B$ leads to the negation $\neg A$ of the original assumption A. Then, $A \Rightarrow B$ is true.
In short, the statement $A \Rightarrow B$ is logically equivalent to the statement $\neg B \Rightarrow \neg A$.

---

[a] $A \Rightarrow B$ is read as "A implies B", which means "if A is true then B is true".



$$
\begin{aligned}
& A \Rightarrow B \\
\Leftrightarrow\ & B \text{ includes } A \\
& (B \supseteq A) \\
\Leftrightarrow\ & \bar{A} \text{ includes } \bar{B} \\
& (\bar{A} \supseteq \bar{B}) \\
\Leftrightarrow\ & \neg B \Rightarrow \neg A
\end{aligned}
$$

Example

Let $A : \{x^2 + y^2 \leq 1\}$ and $B : \{x \leq 1\}$. Prove $A \Rightarrow B$.

- It suffices to show $\neg B \Rightarrow \neg A$.
- Assume $\neg B : \{x > 1\}$. Then, since $x^2 > 1$ and $y^2 \geq 0$, we must have

$$x^2 + y^2 > 1,$$

which is the negation of A. Thus, $A \Rightarrow B$. ■

Hypothesis testing of the regression parameter

- Let

$$A : \{\beta_{01} = 0\}$$
$$B : \{\hat{\beta}_{n1}/se \text{ is close to zero.}\}$$

  Note: $A = H_0$ and $\neg A = H_1$.

- We know that $A \Rightarrow B$ is true by the central limit theorem. Thus, $\neg B \Rightarrow \neg A$ is also true.
- If the computed value of $\hat{\beta}_{n1}/se$ is sufficiently away from zero such that $|\hat{\beta}_{n1}/se| > c$ for some $c > 0$, it is very unlikely that B is true, and thus $\neg B$ is more credible.
- Because $\neg B$ implies that A is not true, we can reject A.

Type I Error and Type II Error

Type I Error

A false rejection of the null hypothesis H_0 (rejecting H_0 although H_0 is true) is called a Type I error.

- The probability of a Type I error is

$$\Pr(\text{Reject } H_0 | H_0 \text{ is true}) = \Pr(T_n > c_\alpha | H_0 \text{ is true})$$

Note that this probability is exactly the significance level α, i.e., α is the maximum tolerable probability of Type I error.

Type II Error

A false acceptance of the null hypothesis H_0 (accepting H_0 although H_0 is false) is called a Type II error.

- The probability of a Type II error is

$$\begin{aligned}\Pr(\text{Accept } H_0 | H_1 \text{ is true}) &= \Pr(T_n \leq c_\alpha | H_1 \text{ is true}) \\ &= 1 - \Pr(T_n > c_\alpha | H_1 \text{ is true})\end{aligned}$$

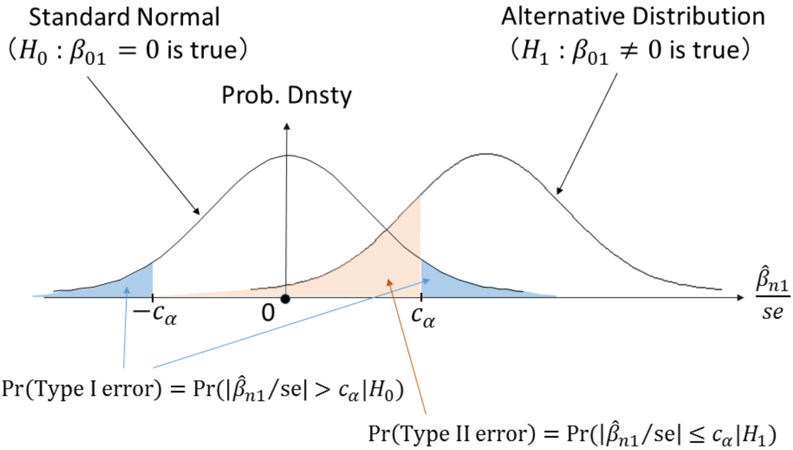
Type I Error and Type II Error

- Given the two possible states of the world (H_0 or H_1) and the two possible decisions (Accept H_0 or Reject H_0), there are four possible pairs of states and decisions as summarized below:

	Accept H_0	**Reject** H_0
H_0 **true**	correct decision	Type I error
H_1 **true**	Type II error	correct decision

- It is important to note that the probability of Type I error is decreasing in c_α, but that of Type II error is increasing in c_α: there is always a trade-off between Type I error and Type II error.

Type I Error and Type II Error



* Decreasing the probability of Type I error will increase the probability of Type II error, and vice versa.

Setup of the experiment

- Data generating process (DGP):

$$Y_i = \beta_{00} + X_i\beta_{01} + \varepsilon_i, \;\; i = 1, ..., 500$$

where $\beta_{00} = 1$, $X \sim N(0,1)$, $\varepsilon \sim N(0,1)$, and β_{01} is selected from

$$\beta_{01} \in \{0, 0.1, 0.2, 0.3\}.$$

When $\beta_{01} = 0$, H_0 is true; otherwise H_1 is true.
- The standard error of the OLS slope estimator $\hat{\beta}_{n1}$ is equal to $1/\sqrt{n}$. Thus, the test statistic is $T_n = |\hat{\beta}_{n1}|/\sqrt{n}$.

Procedure of the experiment

STEP 1. Draw X_i and ε_i from $N(0,1)$, and compute Y_i for a given β_{01}.

STEP 2. Calculate the OLS slope estimate $\hat{\beta}_{n1}$, and compute T_n.

STEP 3. Check whether $T_n > c_\alpha$ (rejection of H_0) or not, in which α is chosen from $\alpha \in \{0.05, 0.01\}$.

STEP 4. Repeat STEPs 1 - 3 many times, say 10,000 times, and compute the proportion of rejection of H_0:

- For $\beta_{01} = 0$, [prop. of rejection of H_0] $\approx \Pr(\text{Type I error})$
- For $\beta_{01} \neq 0$, $[1 -$ prop. of rejection of $H_0] \approx \Pr(\text{Type II error})$

R code used in the experiment

```
N    <- 500   # sample size
nrep <- 10000 # number of repetitions

exper <- function(beta, c_alpha){

      rec <- numeric(nrep)

      for(i in 1:nrep){
            X  <- rnorm(N)
            Y  <- 1 + X*beta + rnorm(N)
            betahat <- lm(Y ~ X)$coef[2]
            se <- sqrt(1/N)
            rec[i] <- (abs(betahat/se) > c_alpha)
            }

      mean(rec)

      }

exper(0, 1.9599)        # Type I error
1 - exper(0.1, 1.9599) # Type II error
```

Results of the experiment

	Type I error	Type II error		
	$\beta_{01} = 0$	$\beta_{01} = 0.1$	0.2	0.3
$\alpha = 0.05$ ($c_\alpha = 1.9599$)	0.0501	0.3888	0.0063	0
$\alpha = 0.01$ ($c_\alpha = 2.5758$)	0.0115	0.6400	0.0271	0

- As the theory suggests, the proportion of Type I error is almost equal to the significance level α.
- When β_{01} is not zero, but is close to zero, the proportion of Type II error is large. As β_{01} gets away from zero, the proportion of Type II error decreases to zero.
- We can observe the trade-off between Type I and Type II errors.

Hypothesis Testing: When the Variance Is Unknown

- In order to implement the testing procedure described above, we must know the variance v or the standard error se of $\hat{\beta}_{n1}$.
- In reality, both se and v are unknown because $V(X)$ and σ^2 are unknown. (recall: $v = V(X)^{-1}\sigma^2$, and $se = \sqrt{v/n}$).
- Fortunately, we can estimate $V(X)$ and σ^2 easily using the sample data.

Estimation of $V(X)$

- The estimation of $V(X)$ is straightforward. One can estimate it either by

$$\text{sample variance: } V_n(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

or by

$$\text{unbiased variance: } V_n(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

- Both are consistent estimators of $V(X)$.

Estimation of σ^2

- Recall that we have assumed

$$Y = \beta_{00} + X\beta_{01} + \varepsilon, \;\; E(\varepsilon^2) = \sigma^2$$

- Letting $(\hat{\beta}_{n0}, \hat{\beta}_{n1})$ be the OLS estimator of (β_{00}, β_{01}), compute the residual as follows:

$$\hat{\varepsilon}_i = Y_i - \hat{\beta}_{n0} - X_i\hat{\beta}_{n1}, \;\; i = 1, ..., n$$

- Then, similarly to the estimation of $V(X)$, σ^2 can be estimated either by

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 \quad \text{or by} \quad \hat{\sigma}_n^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

  The estimator in the right-hand side is an unbiased estimator of σ^2.

- Consequently, v can be estimated by

$$\hat{v}_n = V_n(X)^{-1}\hat{\sigma}_n^2,$$

and let $\widehat{se}_n = \sqrt{\hat{v}_n/n}$.

- Slutsky's theorem: Suppose that $\widehat{se}_n \xrightarrow{P} se$ and

$$\frac{\hat{\beta}_{n1} - \beta_{01}}{se} \xrightarrow{d} N(0,1).$$

Then, we have

$$\frac{\hat{\beta}_{n1} - \beta_{01}}{\widehat{se}_n} \xrightarrow{d} N(0,1).$$

- Let
$$t_n = \frac{\hat{\beta}_{n1} - \beta_{01}}{\widehat{se}_n}.$$
As the size of sample increases, the distribution of t_n can be approximated by the standard normal $N(0, 1)$.
- This statistic t_n is called the t-value (or t-statistic).
- Under the null hypothesis $H_0 : \beta_{01} = 0$, the t-value is given by
$$t_n = \frac{\hat{\beta}_{n1}}{\widehat{se}_n}.$$
Thus, for sufficiently large n, if the t_n is larger than 1.9599 in absolute value, we can reject H_0 at the 5% significance level.

- The above testing procedure based on the t-statistic is called the t-test.
- Note that the null-hypothesis does not need to be $H_0 : \beta_{01} = 0$. For example, if we want to test $H_0 : \beta_{01} = 1$, the corresponding t-statistic becomes

$$t_n = \frac{\hat{\beta}_{n1} - 1}{\widehat{se}_n}.$$

  Then, if this value is larger than 1.9599 in absolute value, we can reject $\beta_{01} = 1$ at the 5% significance level.
- The default null hypothesis in most statistical softwares is $H_0 : \beta_{01} = 0$.

t-value and p-value

p-value

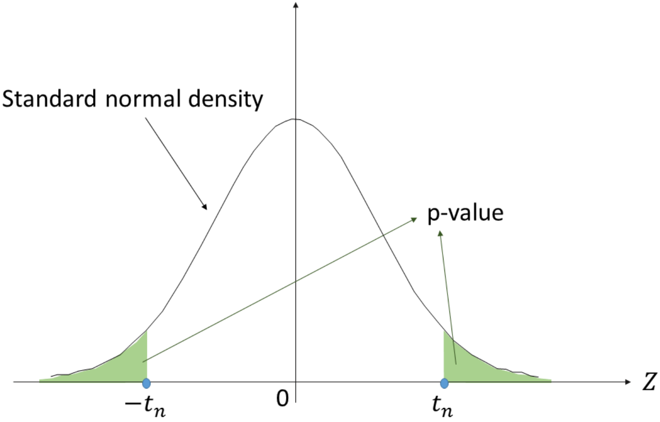
p-value

Let Z be a standard normal random variable, and t_n be a t-statistic computed under a null hypothesis H_0. Then the probability

$$p = \Pr(|Z| \geq |t_n|)$$

is called the p-value. When the p-value p is smaller than α, the null hypothesis H_0 can be rejected at α level of significance.

t-value and p-value



* For example, if $t_n \approx 1.9599$, $p \approx 0.05$.

Glossary I