L.7: Hypothesis Testing

Econometrics 1: ver. 2024 Fall Semester

Naoki Awaya

One-slide Review

- Simple linear regression model: $Y_i = \beta_0^* + X_i \beta_1^* + \epsilon_i$, $i = 1, \ldots, n$.
- The OLS slope estimator:

$$\hat{\beta}_{n1} = \text{sample cov}(Y, X)/\text{sample var}(X)$$

- Central limit theorem:

$$\sqrt{n}(\hat{\beta}_{n1} - \beta_1^*) \xrightarrow{d} N\left(0, \sigma^2/\mathbb{V}(X)\right)$$

where σ^2 and $\mathbb{V}(X)$ are the variances of ϵ and X, respectively.

Hypothesis Testing:
Case 1. When the Variance Is Known

Hypothesis Testing with Known Variance

- We continue to consider a simple linear regression model:

$$Y_i = \beta_0^* + X_i \beta_1^* + \epsilon_i, \quad i = 1, \ldots, n$$

- We would like to know whether the explanatory variable X is actually a determinant of Y.
- We would like to test

> whether the true regression coefficient β_1^* is zero or not.

- Note that the estimate $\hat{\beta}_{n1}$ contains some estimation error for finite sample size:
  - showing $\hat{\beta}_{n1} \neq 0$ is not sufficient
  - but as n gets large, the difference btwn $\hat{\beta}_{n1}$ and β_1^* diminishes

Hypothesis Testing with Known Variance

- Suppose that the variances σ^2 and $\mathbb{V}(X)$ are known quantities, and let

$$V_\beta = \sigma^2/\mathbb{V}(X).$$

- The central limit theorem gives that

$$\sqrt{n}(\hat{\beta}_{n1} - \beta_1^*) \xrightarrow{d} N\left(0, V_\beta\right).$$

- Then, dividing both sides by $\sqrt{V_\beta}$, the above result can be restated as

$$\frac{\hat{\beta}_{n1} - \beta_1^*}{\sqrt{V_\beta/n}} \xrightarrow{d} N(0,1).$$

(standardization)

Hypothesis Testing with Known Variance

- The denominator term $\sqrt{V_\beta/n}$ is interpreted as the standard deviation of $\hat{\beta}_{n1}$.
- Write $\mathrm{sd}_\beta = \sqrt{V_\beta/n}$ so that $\dfrac{\hat{\beta}_{n1} - \beta_1^*}{\mathrm{sd}_\beta} \xrightarrow{d} N(0,1)$.
- sd_β measures the dispersion of the OLS slope estimator $\hat{\beta}_{n1}$ around the true β_1^*.
- By the property of standard normal distribution, the 95% CI for β_1^* can be obtained by

$$
\begin{aligned}
0.95 &\approx \Pr\left(-1.96 \leq \frac{\hat{\beta}_{n1} - \beta_1^*}{\mathrm{sd}_\beta} \leq 1.96\right) \\
&= \Pr\left(\hat{\beta}_{n1} - 1.96 \cdot \mathrm{sd}_\beta \leq \beta_1^* \leq \hat{\beta}_{n1} + 1.96 \cdot \mathrm{sd}_\beta\right)
\end{aligned}
$$

Hypothesis Testing with Known Variance

- If the 95% CI does not include zero, we may conclude that β_1^* is likely non-zero.
- Define
$$Z_n = \frac{\hat{\beta}_{n1} - \beta_1^*}{\mathrm{sd}_\beta}.$$
Assuming that sd_β is known, the only unknown parameter is β_1^*.
- **If** β_1^* **were known**, Z_n is computable. Because $Z_n \xrightarrow{d} N(0,1)$,
$$\Pr(-1.96 \leq Z_n \leq 1.96) \approx 0.95.$$
- I.e., the probability of observing $\{|Z_n| > 1.96\}$ is approximately 5%.

Hypothesis Testing with Known Variance

- However, since β_1^* is unknown in reality, we hypothesize that

$$\beta_1^* = 0$$

  This hypothesis states that <u>X is not a determinant of Y</u>.
- This hypothesis is called the null hypothesis and is denoted as $\mathbb{H}_0$. The null hypothesis is usually expected to be false.
- The negation of the null hypothesis is called the alternative hypothesis, which is denoted as $\mathbb{H}_1$.

$$\begin{aligned} \text{Null hypothesis} \quad & \mathbb{H}_0 : \beta_1^* = 0 \quad (X \text{ does not affect } Y) \\ \text{Alternative hypothesis} \quad & \mathbb{H}_1 : \beta_1^* \neq 0 \quad (X \text{ does affect } Y) \end{aligned}$$

Hypothesis Testing with Known Variance

- **If** $\mathbb{H}_0 : \beta_1^* = 0$ **is true**, the statistic Z_n can be simplified as

$$Z_n = \frac{\hat{\beta}_{n1} - \beta_1^*}{\mathrm{sd}_\beta} = \frac{\hat{\beta}_{n1}}{\mathrm{sd}_\beta}$$

- Thus, **under** $\mathbb{H}_0$, we must have

$$\Pr\left(-1.96 \leq \frac{\hat{\beta}_{n1}}{\mathrm{sd}_\beta} \leq 1.96\right) \approx 0.95$$

$\Longrightarrow$ the probability of observing $\{|\hat{\beta}_{n1}/\mathrm{sd}_\beta| > 1.96\}$ is about 5% if $\mathbb{H}_0$ is true
$\Longrightarrow$ if one thinks 5% is small enough to conclude $\mathbb{H}_0$ is unrealistic, we should **reject** $\mathbb{H}_0$ if $\{|\hat{\beta}_{n1}/\mathrm{sd}_\beta| > 1.96\}$ is observed

Framework of Hypothesis Testing

Framework of Hypothesis Testing

- Let T_n be a general test statistic computed from data.
- Statistical hypothesis testing is based on the following decision rule:[1]

Reject the null hypothesis $\mathbb{H}_0$ if $\{T_n > c\}$ is observed, **accept** $\mathbb{H}_0$ if $\{T_n \leq c\}$ is observed.

- The threshold value c is called the critical value, which is pre-specified by researcher.
- In the above example, $T_n = |\hat{\beta}_{n1}/\mathrm{sd}_\beta|$, and $c = 1.96$.

[1] IMPORTANT: accepting the null hypothesis does NOT mean its correctness but only means that we do not have sufficient evidence to reject it.

Framework of Hypothesis Testing

- For a given probability $\alpha \in (0, 1)$, set c_α so that

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

  under the null hypothesis $\mathbb{H}_0$.[2]

- Thus, if we actually observe $\{T_n > c_\alpha\}$ with a very small α, we can conclude that $\mathbb{H}_0$ is "unrealistic".
- This threshold probability α is referred to as the significance level. When $\{T_n > c_\alpha\}$ is true, we say that

> the null hypothesis $\mathbb{H}_0$ is rejected at the $100\alpha\%$ significance level.

[2]Once α is given, the value of c_α is automatically determined by the distribution of T_n: e.g., $\alpha = 0.05 \Rightarrow c_\alpha = 1.96$. Even when the distribution of T_n is unknown, c_α is usually estimable from the asymptotic distribution or by simulation.

Framework of Hypothesis Testing

- Researchers often use either $\alpha = 0.05$ or 0.01 of significance.
- *Are 5% and 1% small enough to conclude as unrealistic?*
  - The choice of significance level is quite subjective, and depends on the context.

Hypothesis testing procedure

Step 1. Compute the test statistic T_n ($|\hat{\beta}_{n1}/\mathrm{sd}_\beta|$).

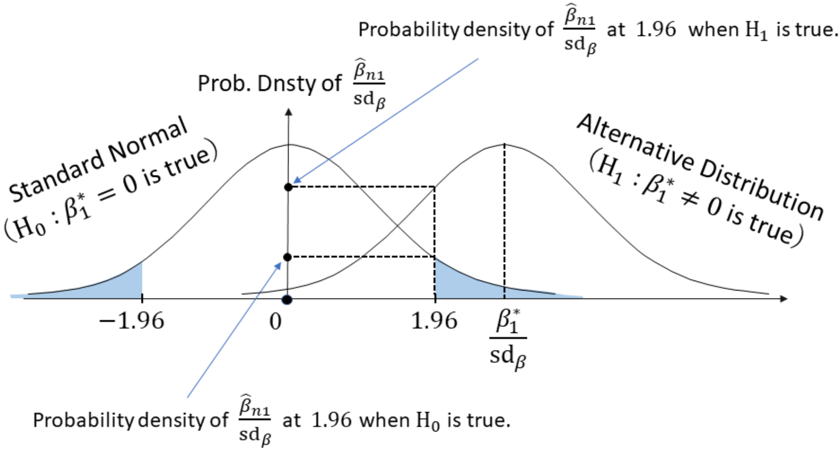
Step 2. Compute the critical value c_α by solving

$$\Pr(T_n > c \mid \mathbb{H}_0 \text{ is true}) = \alpha$$

with respect to c. ($c_{0.05} = 1.96$, $c_{0.01} = 2.58$)

Step 3. If $T_n > c_\alpha$, reject $\mathbb{H}_0$ at the $100\alpha\%$ significance level.

Framework of Hypothesis Testing



Hypothesis Testing as a "Proof by Contraposition"

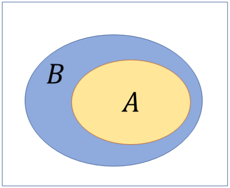
Proof by Contraposition

- Suppose we would like to prove the statement $A \Rightarrow B$.[a]
- Assume the negation of B, $\neg B$,[b] and show that $\neg B$ leads to the negation $\neg A$ of the original assumption A.
- Then, $A \Rightarrow B$ is true.

* In a nutshell, $A \Rightarrow B$ is logically equivalent to $\neg B \Rightarrow \neg A$.

---

[a] $A \Rightarrow B$ is read as "A implies B", which means "if A is true then B is true".
[b] $\neg B$ is read as "not B".



$$A \Rightarrow B$$
$$\Leftrightarrow B \text{ includes } A \Leftrightarrow \bar{A} \text{ includes } \bar{B}$$
$$(B \supseteq A) \qquad (\bar{A} \supseteq \bar{B})$$
$$\Leftrightarrow \neg B \Rightarrow \neg A$$

Hypothesis Testing as a "Proof by Contraposition"

Hypothesis testing for the regression coefficient

- We set

$$A : \{\beta_1^* = 0\} \; (= \mathbb{H}_0)$$
$$B : \{|\hat{\beta}_{n1}/\mathrm{sd}_\beta| \text{ is close to zero}\}$$

- We know that $A \Rightarrow B$ is true for large enough n by the central limit theorem. Thus, its contraposition $\neg B \Rightarrow \neg A$ is also true.
- If the computed value of $|\hat{\beta}_{n1}/\mathrm{sd}_\beta|$ is sufficiently away from zero even when n is large, B is not true, implying that A is not true, i.e., $\mathbb{H}_0$ is rejected.

Hypothesis Testing:

Case 2. When the Variance Is Unknown

Hypothesis Testing with Unknown Variance

- In order to implement the testing procedure described above, we must know the standard deviation sd_β of $\hat{\beta}_{n1}$.
- In reality, sd_β is unknown because $\mathbb{V}(X)$ and σ^2 are unknown.
  - Recall: $\mathrm{sd}_\beta = \sqrt{V_\beta / n}$ and $V_\beta = \sigma^2 / \mathbb{V}(X)$
- Fortunately, we can estimate both $\mathbb{V}(X)$ and σ^2 easily from the sample data.

Hypothesis Testing with Unknown Variance

Estimation of $\mathbb{V}(X)$

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

$$\text{sample variance estimator: } \mathbb{V}_n(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

or by

$$\text{unbiased variance estimator: } \mathbb{V}_n(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

- Both estimators are consistent for $\mathbb{V}(X)$.

Hypothesis Testing with Unknown Variance

Estimation of σ^2

- Recall that our model is

$$Y = \beta_0^* + X\beta_1^* + \epsilon, \;\; \mathbb{E}(\epsilon^2) = \sigma^2$$

- Letting $(\hat{\beta}_{n0}, \hat{\beta}_{n1})$ be the OLS estimator of (β_0^*, β_1^*), compute the residuals as follows:

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

- Then, similarly to the estimation of $\mathbb{V}(X)$, σ^2 can be consistently estimated by

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i^2.$$

Hypothesis Testing with Unknown Variance

- Consequently, V_β can be estimated by

$$\widehat{V}_\beta = \hat{\sigma}_n^2 / \mathbb{V}_n(X),$$

and let $\widehat{\mathrm{se}}_\beta = \sqrt{\widehat{V}_\beta / n}$. This estimator $\widehat{\mathrm{se}}_\beta$ is called the standard error.

- In addition, we have the following result:

Slutsky's theorem:

If $\widehat{\mathrm{se}}_\beta \xrightarrow{p} \mathrm{sd}_\beta$ and $\frac{\hat{\beta}_{n1} - \beta_1^*}{\mathrm{sd}_\beta} \xrightarrow{d} N(0,1)$, then $\frac{\hat{\beta}_{n1} - \beta_1^*}{\widehat{\mathrm{se}}_\beta} \xrightarrow{d} N(0,1)$.

- That is, the normality result still holds even when sd_β is replaced by its consistent estimator.

t-test

- Let

$$t_n = \frac{\hat{\beta}_{n1} - \beta_1^*}{\widehat{\mathrm{se}}_\beta}.$$

  If n is large, the distribution of t_n can be approximated by $N(0, 1)$.[3]
- This statistic t_n is called the t-value (or t-statistic).
- Under the null hypothesis $\mathbb{H}_0 : \beta_1^* = 0$, the t-value is obtained as

$$t_n = \frac{\hat{\beta}_{n1}}{\widehat{\mathrm{se}}_\beta}.$$

  Thus, for sufficiently large n,

if $|t_n| > 1.96$ (resp. $|t_n| > 2.58$), we can reject $\mathbb{H}_0$ at the 5% (resp. 1%) significance level.

[3] For finite n, t_n follows a t-distribution (ref. Lecture Note 4).

t-test

- The above testing procedure is called the t-test.
- Note that the null-hypothesis does not need to be $\mathbb{H}_0 : \beta_1^* = 0$. For example, if we would like to test $\mathbb{H}_0 : \beta_1^* = 1$, the corresponding t-statistic becomes
$$t_n = \frac{\hat{\beta}_{n1} - 1}{\widehat{\mathrm{se}}_\beta}.$$
- Then, if this value is larger than 1.96 in absolute value, we can reject $\mathbb{H}_0 : \beta_1^* = 1$ at the 5% significance level.

p-value

p-value

Let Z be a standard normal random variable, and t_n be the t-statistic computed under $\mathbb{H}_0 : \beta_1^* = 0$. Then the probability

$$p = \Pr(|Z| \geq \underbrace{|t_n|}_{= |\hat{\beta}_{n1}/\widehat{\mathsf{se}}_\beta|})$$

$$= \text{probability of observing } t\text{-values more extreme than } \hat{\beta}_{n1}/\widehat{\mathsf{se}}_\beta$$

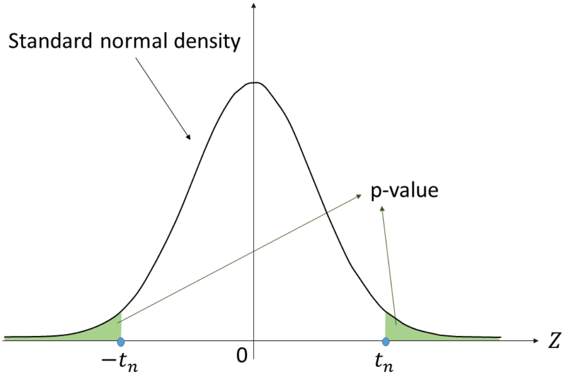
is called the p-value. If

$$p \leq \alpha$$

then $\mathbb{H}_0$ is rejected at the $100\alpha\%$ significance level.

- We can check whether the explanatory variable is significant or not just by checking its p-value.

t-value and p-value



- The p-value is calculated from the t-statistic: larger $t_n \Longrightarrow$ smaller p.
- For example, if $t_n = 1.96$, $p = 0.05$.

Type I Error and Type II Error

Type I Error and Type II Error

Type I Error

A false rejection of the null hypothesis $\mathbb{H}_0$ (rejecting $\mathbb{H}_0$ although $\mathbb{H}_0$ is true) is called a Type I error.

- The probability of a Type I error is

$$\Pr(\text{Reject } \mathbb{H}_0 | \mathbb{H}_0 \text{ is true}) = \Pr(T_n > c_\alpha | \mathbb{H}_0 \text{ is true})$$

- Note that <u>this probability is exactly the significance level α.</u>

In our regression setting,

Type I error

$=$ error of concluding that X is a determinant of Y although it is not.

Type I Error and Type II Error

Type II Error

A false acceptance of the null hypothesis $\mathbb{H}_0$ (accepting $\mathbb{H}_0$ although $\mathbb{H}_1$ is true) is called a Type II error.

- The probability of a Type II error is

$$\begin{aligned}\Pr(\text{Accept } \mathbb{H}_0 | \mathbb{H}_1 \text{ is true}) &= \Pr(T_n \leq c_\alpha | \mathbb{H}_1 \text{ is true}) \\ &= 1 - \Pr(T_n > c_\alpha | \mathbb{H}_1 \text{ is true})\end{aligned}$$

In our regression setting,

Type II error
$=$ error of concluding that X is not a determinant of Y although it actually is.

Type I Error and Type II Error

- Given the two possible states of the world ($\mathbb{H}_0$ or $\mathbb{H}_1$), there are four possible pairs of states and decisions:

	Accept H_0	Reject H_0
H_0 is true	**Correct decision** $X \nrightarrow Y$ (prediction) $X \nrightarrow Y$ (truth)	**Type 1 error** $X \rightarrow Y$ (prediction) $X \nrightarrow Y$ (truth)
H_1 is true	**Type 2 error** $X \nrightarrow Y$ (prediction) $X \rightarrow Y$ (truth)	**Correct decision** $X \rightarrow Y$ (prediction) $X \rightarrow Y$ (truth)

- Note that the probability of Type I error is decreasing in c_α, but that of Type II error is increasing in c_α.
- That is, it is impossible to reduce both Type I error and Type II error at the same time by manipulating c_α.

Type I Error and Type II Error

A simulation analysis:

```
>  beta0 <- 1
>  t_val <- function(n, beta1){
+   X <- rnorm(n)
+   e <- rnorm(n)
+   Y <- beta0 + X*beta1 + e
+   bhat1 <- cov(X,Y)/var(X)
+   bhat0 <- mean(Y) - mean(X)*bhat1
+   resid <- Y - bhat0 - X*bhat1
+   V_beta <- var(resid)/var(X)
+   se <- sqrt(V_beta/n)
+   bhat1/se
+   }
```

Type I Error and Type II Error

- Judging Type I error:

```
>  T1err <- function(n){
+   tt <- t_val(n, 0) # H0 is true, beta_1 = 0
+   ifelse(abs(tt) > 1.96, 1, 0) # Reject H0 at the 5%
+   }
```

- Judging Type II error:

```
>  T2err <- function(n){
+   tt <- t_val(n, 0.3) # H1 is true, beta_1 = 0.3
+   ifelse(abs(tt) <= 1.96, 1, 0) # Accept H0 at the 5%
+   }
```

Type I Error and Type II Error

```
> R <- matrix(0, 500, 3)
> for(i in 1:500){
+   R[i,1] <- T1err(50)
+   R[i,2] <- T1err(200)
+   R[i,3] <- T1err(1000)
+   }
> colMeans(R) # Frequency of Type 1 Error
```

```
## [1] 0.072 0.056 0.048
```

- Recall that the nominal significance level is 0.05.

Type I Error and Type II Error

```
> R <- matrix(0, 500, 3)
> for(i in 1:500){
+   R[i,1] <- T2err(50)
+   R[i,2] <- T2err(200)
+   R[i,3] <- T2err(1000)
+   }
> colMeans(R) # Frequency of Type 2 Error
```

```
## [1] 0.448 0.012 0.000
```

- The probability of Type II Error can be reduced by increasing n.