Lecture 9: Hypothesis Tests and Confidence Intervals in Multiple Regression

Zheng Tian

Outline

1. Hypothesis Tests and Confidence Intervals For a Single Coefficient

2. Tests of Joint Hypotheses

1. Hypothesis Tests and Confidence Intervals For a Single Coefficient

2. Tests of Joint Hypotheses

The basic multiple regression model

Consider the following model

$$\mathbf{Y} = \beta_0 \boldsymbol{\iota} + \beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2 + \cdots + \beta_k \mathbf{X}_k + \mathbf{u} \tag{1}$$

- $\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k,$ and $\mathbf{u}$ are $n \times 1$ vectors of the dependent variable, regressors, and errors
- $\beta_0, \beta_1, \beta_2, \ldots,$ and β_k are parameters.
- $\boldsymbol{\iota}$ is the $n \times 1$ vector of 1s.

Review of $\mathrm{var}(\hat{\beta}|X)$

- The homoskedasticity-only covariance matrix if u_i is homoskedastic

$$\mathrm{var}(\hat{\beta}|\mathbf{X}) = \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1} \tag{2}$$

- The heteroskedasticity-robust covariance matrix if u_i is heteroskedastic

$$\mathrm{var_h}(\hat{\beta}|\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{\Sigma}(\mathbf{X}'\mathbf{X})^{-1} \tag{3}$$

where $\mathbf{\Sigma} = \mathbf{X}'\mathbf{\Omega}\mathbf{X}$, and $\mathbf{\Omega} = \mathrm{var}(\mathbf{u}|\mathbf{X})$.

The multivariate normal distribution of $\hat{\beta}$

We know that if the least squares assumptions hold, $\hat{\boldsymbol{\beta}}$ has an asymptotic multivariate normal distribution as

$$\hat{\boldsymbol{\beta}} \xrightarrow{d} N(\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}) \tag{4}$$

where $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}} = \mathrm{var}(\hat{\boldsymbol{\beta}}|\mathbf{X})$ for which use Equation (2) for the homoskedastic case and Equation (3) for the heteroskedastic case.

The estimator of $\text{var}(\hat{\beta}|X)$

The estimator of σ_u^2

$$s_u^2 = \frac{1}{n-k-1} \sum_{i=1}^{n} \hat{u}_i^2 \tag{5}$$

Thus, the estimator of the homoskedasticity-only covariance matrix is

$$\widehat{\text{var}(\hat{\beta})} = s_u^2 (\mathbf{X}'\mathbf{X})^{-1} \tag{6}$$

The estimator of $\boldsymbol{\Sigma}$

$$\widehat{\boldsymbol{\Sigma}} = \frac{n}{n-k-1} \sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i' \hat{u}_i^2 \tag{7}$$

where $\mathbf{X}_i$ is the vector of the i^{th} observation of $(k+1)$ regressors. The heteroskedasticity-consistent (robust) covariance matrix estimator is

$$\widehat{\text{var}_{\text{h}}(\hat{\beta})} = (\mathbf{X}'\mathbf{X})^{-1} \widehat{\boldsymbol{\Sigma}} (\mathbf{X}'\mathbf{X})^{-1} \tag{8}$$

The estimator of $SE(\hat{\beta}_j)$

We can get the standard error of $\hat{\beta}_j$ as the square root of the j[th] diagonal element of $\widehat{\mathrm{var}(\hat{\boldsymbol{\beta}})}$ for homoskedasticity and $\widehat{\mathrm{var}_{\mathrm{h}}(\hat{\boldsymbol{\beta}})}$ for heteroskedasticity. That is,

- Homoskedasticity-only standard error: $SE(\hat{\beta}_j) = \left(\left[\widehat{\mathrm{var}(\hat{\boldsymbol{\beta}})} \right]_{(j,j)} \right)^{\frac{1}{2}}$
- Heteroskedasticity-robust standard error: $SE(\hat{\beta}_j) = \left(\left[\widehat{\mathrm{var}_{\mathrm{h}}(\hat{\boldsymbol{\beta}})} \right]_{(j,j)} \right)^{\frac{1}{2}}$

The t-statistic

We can perform a two-sided hypothesis test as

$$H_0 : \beta_j = \beta_{j,0} \text{ vs. } H_1 : \beta_j \neq \beta_{j,0}$$

- We still use the t-statistic, computed as $t = (\hat{\beta}_j - \beta_{j,0})/SE(\hat{\beta}_j)$, where $SE(\hat{\beta}_j)$ is the standard error of $\hat{\beta}_j$.
- Under the null hypothesis, we have, in large samples, $t \overset{a}{\sim} N(0,1)$. Therefore, the p-value can still be computed as $2\Phi(-|t^{act}|)$.
- The null hypothesis is rejected at the 5% significant level when the p-value is less than 0.05, or equivalently, if $|t^{act}| > 1.96$. (Replace the critical value with 1.64 at the 10% level and 2.58 at the 1% level.)

Confidence intervals for a single coefficient

The confidence intervals for a single coefficient can be constructed as before using the t-statistic.

Given large samples, a 95% two-sided confidence interval for the coefficient β_j is

$$\left[\hat{\beta}_j - 1.96SE(\hat{\beta}_j),\ \hat{\beta}_j + 1.96SE(\hat{\beta}_j)\right]$$

Application to test scores and the student-teacher ratio

The estimated model can be written as follows

$$\widehat{TestScore} = \underset{(8.7)}{686.0} - \underset{(0.43)}{1.10} \times STR - \underset{(0.031)}{0.650} \times PctEl$$

- We test $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$. The t-statistic for this test can be computed as $t = (-1.10 - 0)/0.43 = -2.54 < -1.96$, and the p-value is $2\Phi(-2.54) = 0.011 < 0.05$. Based on either the t-statistic or the p-value, we can reject the null hypothesis at the 5% level.
- The confidence interval that contains the true value of β_1 with a 95% probability can be computed as $-1.10 \pm 1.96 \times 0.43 = (-1.95, -0.26)$.

Adding expenditure per pupil to the equation

Now we add a new explanatory variable in the regression, *Expn*, that is the expenditure per pupil in the district in thousands of dollars.

$$\widehat{TestScore} = \underset{(15.5)}{649.6} - \underset{(0.48)}{0.29} \times STR + \underset{(1.59)}{3.87} \times Expn - \underset{(0.032)}{0.656} \times PctEl$$

- The magnitude of the coefficient on STR decreases from 1.10 to 0.29 after *Expn* is added.
- The standard error of the coefficient on STR increases from 0.43 to 0.48 after *Expn* is added.
- Consequently, in the new model, the t-statistic for the coefficient becomes $t = -0.29/0.48 = -0.60 > -1.96$ so that we cannot reject the zero hypothesis at the 5% level. (neither can we at the 10% level).

How can we interpret such changes?

- The decrease in the magnitude of the coefficient reflects that expenditure per pupil is an important factor that carry over some influence of student-teacher ratio on test scores.

  In other words, holding expenditure per pupil and the percentage of English-learners constant, reducing class sizes by hiring more teachers have only small effect on test scores

- The increase in the standard error reflects that *Expn* and *STR* are correlated so that there is imperfect multicollinearity in this model. In fact, the correlation coefficient between the two variables is 0.48, which is relatively high.

1 Hypothesis Tests and Confidence Intervals For a Single Coefficient

2 Tests of Joint Hypotheses

The unrestricted model

Consider the following multiple regression model

$$\mathbf{Y} = \beta_0 \boldsymbol{\iota} + \beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2 + \cdots + \beta_k \mathbf{X}_k + \mathbf{u} \tag{9}$$

We call Equation (9) as the full model or the unrestricted model because β_0 to β_k can take any value without restrictions.

Joint hypothesis: a case of two zero restrictions

- Question: Are the coefficients on the first two regressors zero?
- Joint hypotheses

$$H_0 : \beta_1 = 0, \beta_2 = 0, \text{ vs. } H_1 : \text{either } \beta_1 \neq 0 \text{ or } \beta_2 \neq 0 \text{ (or both)}$$

- This is a joint hypothesis because $\beta_1 = 0$ and $\beta_2 = 0$ must hold at the same time. So if either of them is invalid, the null hypothesis is rejected as a whole.

The restricted model with two zero restrictions

- If the null hypothesis is true, we have

$$\mathbf{Y} = \beta_0 + \beta_3\mathbf{X}_3 + \beta_4\mathbf{X}_4 + \cdots + \beta_k\mathbf{X}_k + \mathbf{u} \tag{10}$$

  We call Equation (10) as the restricted model because we impose two restrictions $\beta_1 = 0$ and $\beta_2 = 0$.
- To test these two restrictions jointly means that we need to use a single statistic to test these restrictions simultaneously. That statistic is F-statistic.

Why not use t-statistic and test individual coefficients one at a time?

- Let us test the null hypothesis above using t-statistics for β_1 and β_2 separately. That is, t_1 is the t-statistic for $\beta_1 = 0$ and t_2 is the t-statistic for $\beta_2 = 0$.
- Compute the t-statistics t_1 for $\beta_1 = 0$ and t_2 for $\beta_2 = 0$. We call this "one-at-a-time" testing procedure.

What's the problem with the one-at-a-time procedure

How can we reject the null hypothesis with this procedure?

Using the one-at-a-time procedure, at the 5% significance level, we can reject the null hypothesis of $H_0 : \beta_1 = 0$ and $\beta_2 = 0$ when either $|t_1| > 1.96$ or $|t_2| > 1.96$ (or both). In other words, the null is not rejected only when both $|t_1| \leq 1.96$ and $|t_2| \leq 1.96$.

What is the probability of committing Type I error?

Assume t_1 and t_2 to be independent. Then,

$$\Pr(|t_1| \leq 1.96 \ \& \ |t_2| \leq 1.96) = \Pr(|t_1| \leq 1.96)\Pr(|t_2| \leq 1.96) = 0.95^2 = 90.25\%$$

So the probability of rejecting the null when it is true is $1 - 90.25\% = 9.75\%$. We may reject the null hypothesis with a higher probability than what we have pre-specified with the significant level.

Joint hypothesis involving one coefficient for each restriction

q restrictions

$$H_0 : \beta_1 = \beta_{1,0},\ \beta_2 = \beta_{2,0},\ \ldots,\ \beta_q = \beta_{q,0} \text{ versus}$$
$$H_1 : \text{at least one restriction does not hold}$$

The restricted model

Suppose that we are testing the q zero hypotheses, that is, q restrictions, $\beta_1 = \beta_2 = \cdots = \beta_q = 0$. The restricted model is

$$\mathbf{Y} = \beta_0 + \beta_{q+1}\mathbf{X}_{q+1} + \beta_{q+2}\mathbf{X}_{q+2} + \cdots + \beta_k\mathbf{X}_k + \mathbf{u} \tag{11}$$

Joint linear hypotheses

Joint hypotheses include linear hypotheses like the followings

1
$$H_0: \beta_1 = \beta_2 \text{ vs. } H_1: \beta_1 \neq \beta_2$$

2
$$H_0: \beta_1 + \beta_2 = 1 \text{ vs. } H_1: \beta_1 + \beta_2 \neq 1$$

3
$$\begin{aligned} &H_0: \beta_1 + \beta_2 = 0,\ 2\beta_2 + 4\beta_3 + \beta_4 = 3 \text{ vs.} \\ &H_1: \text{at least one restriction does not hold} \end{aligned}$$

A general form of joint hypotheses

We can use a matrix form to represent all linear hypotheses regarding the coefficients in Equation (9) as follows

$$H_0: \mathbf{R}\boldsymbol{\beta} = \mathbf{r} \text{ vs. } H_1: \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r} \tag{12}$$

where $\mathbf{R}$ is a $q \times (k+1)$ matrix with the full row rank, $\boldsymbol{\beta}$ represent the $k+1$ regressors, including the intercept, and $\mathbf{r}$ is a $q \times 1$ vector of real numbers.

Examples of $\mathbf{R}\beta = \mathbf{r}$

- For $H_0 : \beta_1 = 0, \beta_2 = 0$

$$\mathbf{R} = \begin{array}{c c} & \begin{array}{cccccc} \beta_0 & \beta_1 & \beta_2 & \beta_3 & \cdots & \beta_k \end{array} \\ \begin{array}{c} R1 \\ R2 \end{array} & \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \end{pmatrix} \end{array} \text{ and } \mathbf{r} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- For $H_0 : \beta_1 + \beta_2 = 0,\ 2\beta_2 + 4\beta_3 + \beta_4 = 3,\ \beta_1 = 2\beta_3 + 1$

$$\mathbf{R} = \begin{array}{c c} & \begin{array}{ccccccc} \beta_0 & \beta_1 & \beta_2 & \beta_3 & \beta_4 & \cdots & \beta_k \end{array} \\ \begin{array}{c} R1 \\ R2 \\ R3 \end{array} & \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 2 & 4 & 1 & \cdots & 0 \\ 0 & 1 & 0 & -2 & 0 & \cdots & 0 \end{pmatrix} \end{array} \text{ and } \mathbf{r} = \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix}$$

The general form of the F-statistic

To test the null hypothesis

$$H_0 : \mathbf{R}\beta = \mathbf{r}$$

we compute the F-statistic

$$F = \frac{1}{q}(\mathbf{R}\hat{\beta} - \mathbf{r})' \left[\mathbf{R}\widehat{\mathrm{var}(\hat{\beta})}\mathbf{R}' \right]^{-1} (\mathbf{R}\hat{\beta} - \mathbf{r}) \tag{13}$$

- $\hat{\beta}$ is the estimated coefficients by OLS and $\widehat{\mathrm{var}(\hat{\beta})}$ is the estimated covariance matrix.
- For homoskedastic errors, we can compute $\widehat{\mathrm{var}(\hat{\beta})}$ as in Equation (6)
- For heteroskedastic errors, we can compute $\widehat{\mathrm{var}_{\mathrm{h}}(\hat{\beta})}$ as in Equation (8)

The F distribution, the critical value, and the p-value

The F distribution

If the least square assumptions hold, under the null hypothesis, the F-statistic is asymptotically distributed as F distribution with degree of freedom (q, ∞). That is,

$$F \overset{a}{\sim} F(q, \infty)$$

The critical value and the p-value of F test.

The 5% critical value of the F test using F-statistic is c_α such that

$$\Pr(F < c_\alpha) = 0.95$$

And the p-value of F test can be computed as

$$p\text{-value} = \Pr(F > F^{act})$$

The F-statistic when $q = 2$

The F-statistic for testing the null hypothesis of $H_0 : \beta_1 = 0, \beta_2 = 0$ can be proved to take the following form,

$$F = \frac{1}{2} \frac{t_1^2 + t_2^2 - 2\hat{\rho}_{t_1,t_2} t_1 t_2}{1 - \hat{\rho}_{t_1,t_2}^2} \tag{14}$$

- For simplicity, suppose t_1 and t_2 are independent so that $\hat{\rho}_{t_1,t_2} = 0$. Then $F = \frac{1}{2}(t_1^2 + t_2^2)$.
- Under the null hypothesis, both t_1 and t_2 have standard normal distribution asymptotically. Then $t_1^2 + t_2^2$ has a chi-squared distribution with 2 degrees of freedom.
- It follows that $F = \frac{1}{2}(t_1^2 + t_2^2)$ has asymptotically distributed as $F(2, \infty)$.
- The discussion about F-statistic in Equation (14) will become complicated when $\hat{\rho}_{t_1,t_2} \neq 0$.