Econometrics I TA Session

Wang Xin

June 2, 2022

Contents

1 Introduction to Hypothesis Testing

We begin the analysis with the regression model as a statement of a proposition

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

Assuming that it satisfies the classical linear model (CLM) assumptions.We know that OLS produces unbiased estimators of $\boldsymbol{\beta}$.

In the first few lessons, we mainly studied how to derive estimates of $\boldsymbol{\beta}$ and some properties of them. That is only one step when we carry a research, called 'fitting'. However, a relevant empirical question is whether the equation specified appears to be consistent with the data, that is, $\mathbf{y}$. Therefore, the estimates of $\boldsymbol{\beta}$ need to be examined. The "test" at this point, is whether $\hat{\boldsymbol{\beta}}$ in the least squares regression is zero or not.

In most application, our primary interest lies in testing the null hypothesis against the alternative hypothesis

$$H_0 : \boldsymbol{\beta} = \mathbf{0} \ versus \ H_1 : \boldsymbol{\beta} \neq \mathbf{0}$$

The hypothesis H_0 is referred to as the null hypothesis, while H_1 is referred to as the alternative hypothesis. Often the null hypothesis represents no change or no difference from the past, while the alternative represents change or difference. The alternative is often referred to as the research worker's hypothesis.

2 Testing Hypotheses about a Single Population Parameter: The t Test

2.1 A deduction of 't statistic'

Suppose that

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u \tag{1}$$

In this section, we study how to test hypotheses about a particular β_j. The null hypothesis is

$$H_0 : \beta_j = 0 \tag{2}$$

For a full understanding of hypothesis testing, one must remember that the β_j are unknown features of the population, and we will never know them with certainty. Nevertheless, we can hypothesize about the value of β_j and then use statistical inference to test our hypothesis.

It is important to understand what (2) means and to be able to describe this hypothesis in simple language for a particular application. Since β_j measures the partial effect of x_j on (the expected value of) y, after controlling for all other independent variables, (2) means that, once $x_1, x_2, \cdots, x_{j-1}, x_{j+1}, \cdots, x_k$ have been accounted for, x_j has no effect on the expected value of y. We cannot state the null hypothesis as "x_j does have a partial effect on y" because this is true for any value of β_j other than zero. Classical testing is suited for testing simple hypotheses like (2).

In order to carry hypotheses tests, we need to construct t-distribution for the standardized estimators:

From the Central Limit Theorem, we know that

$$\frac{\hat{\beta}_j - \beta_j}{\sigma / \sqrt{n}} \sim N(0, 1)$$

Moreover, the estimator of σ^2, s^2, has been proved to be unbiased. Since $\mathbf{u}$ is under normality assumption and $\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is symmetric and idempotent,

$$s^2 = \frac{1}{n-k-1} \hat{\mathbf{u}}'\hat{\mathbf{u}} = \frac{1}{n-k-1} \mathbf{u}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{u}$$

$$\frac{(n-k-1)s^2}{\sigma^2} \sim \chi^2(n-k-1) \tag{3}$$

$\dfrac{(n-k-1)s^2}{\sigma^2}$ is of χ^2 distributed.

From the definition of t-distribution, if the constant σ in $std(\hat{\beta}_j) = \sigma / \sqrt{n}$ has been replaced with the random variable s^2, a new estimator can be:

$$\frac{\dfrac{\hat{\beta}_j - \beta_j}{\sigma / \sqrt{n}}}{\sqrt{\dfrac{(n-k-1)s^2}{\sigma^2} / (n-k-1)}} = \frac{\hat{\beta}_j - \beta_j}{s / \sqrt{n}} = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t(n-k-1) \tag{4}$$

2

The statistic we use to test (2) (against any alternative) is called "the" t statistic or "the" t ratio of β_j and is defined as

$$t_{\hat{\beta}_j} \equiv \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \quad (5)$$

where β_j is 0 under (2).

To determine a rule for rejecting H_0, we need to decide on the relevant alternative hypothesis. For example

$$H_1 : \beta_j \neq 0$$

We must first decide on a significance level ("level" for short) or the probability of rejecting H_0 when it is in fact true. For concreteness, suppose we have decided on a 5% significance level, as this is the most popular choice. Thus, we are willing to mistakenly reject H_0 when it is true 5% of the time. Now, while $t_{\hat{\beta}_j}$ has a t distribution under H_0.

The definition of "sufficiently large," with a 5% significance level, which means 2.5% in each tail, is the 97.5th percentile in a t-distribution with $n - k - 1$ degrees of freedom; denote this by c. In other words, the rejection ruleis that H_0 is rejected in favor of H_1 at the 5% significance level if

$$|t_{\hat{\beta}_j}| > c \quad (6)$$

2.2   A Simple Example

Consider to estimate a model explaining college GPA,

$$colGPA = \beta_0 + \beta_1 hsGPA + \beta_2 ACT + \beta_3 skipped + u$$

The variables above include the college grade point average ($colGPA$), high school GPA ($hsGPA$), achievement test score (ACT), and the average number of lectures missed per week ($skipped$) for a sample of 141 students from a large university.

Using statistic software, the parameters can be estimated

$$\widehat{colGPA} = 1.39 + 0.412 hsGPA + 0.015 ACT - 0.083 skipped$$
$$\quad\ (0.33)\quad\ \ (0.094)\qquad\quad (0.011)\quad\ \ (0.026)$$
$$n = 141, \ R^2 = 0.234$$

We can easily compute t statistics to see which variables are statistically significant, using a two-sided alternative in each case. The 5% critical value is about 1.96, since the degrees of freedom ($141 - 4 = 137$) is large enough to use the standard normal approximation. The 1% critical value is about 2.58.

The t statistic on $hsGPA$ is 4.38, which is significant at very small significance levels. Thus, we say that "$hsGPA$ is statistically significant at any conventional significance level." The t statistic on ACT is 1.36, which is not statistically significant at the 10% level against a two-sided alternative. The coefficient on ACT is also practically small: a

10-point increase in ACT, which is large, is predicted to increase $colGPA$ by only 0.15 points. Thus, the variable ACT is practically, as well as statistically, insignificant.

The coefficient on *skipped* has a t statistic of $-0.083/0.026 = -3.19$, so *skipped* is statistically significant at the 1% significance level $(3.19 > 2.58)$. This coefficient means that another lecture missed per week lowers predicted $colGPA$ by about 0.083. Thus, holding $hsGPA$ and ACT fixed, the predicted difference in $colGPA$ between a student who misses no lectures per week and a student who misses five lectures per week is about 0.42. Remember that this says nothing about specific students; rather, 0.42 is the estimated average across a subpopulation of students.

In this example, for each variable in the model, we could argue that a one-sided alternative is appropriate. The variables $hsGPA$ and skipped are very significant using a two-tailed test and have the signs that we expect, so there is no reason to do a one-tailed test. On the other hand, against a one-sided alternative $(\beta_3 > 0)$, ACT is significant at the 10% level but not at the 5% level. This does not change the fact that the coefficient on ACT is pretty small.

3 Testing Multiple Linear Restrictions: The F Test

Still, we begin with the regression model (1).

Using t statistic, we have just tested hypotheses involving a single restriction. Frequently, we wish to test multiple hypotheses about the underlying parameters $\beta_0, \beta_1, \cdots, \beta_k$.

In terms of the parameters of the model, the null hypothesis is stated as

$$H_0 : \beta_3 = \beta_4 = \cdots = \beta_k = 0 \tag{7}$$

if (7) is true, then $x_3, x_4, \cdots, x_k$ have no effect on y after x_1 and x_2 have been controlled for and therefore should be excluded from the model.

Then the appropriate alternative is simply

$$H_1 : H_0 \ is \ not \ true.$$

The alternative holds if at least one of β_3, β_4 or β_k is different from zero. (Any or all could be different from zero.)

It is tempting to test (7) by using the t statistics on the variables $x_3, x_4, \cdots, x_k$ to determine whether each variable is individually significant. This option is not appropriate. A particular t statistic tests a hypothesis that puts no restrictions on the other parameters. Besides, we would have three outcomes to contend with—one for each t statistic. Furthermore, using separate t statistics to test a multiple hypothesis like (7) can be very misleading. We need a way to test the exclusion restrictions jointly.

For convenience, H_0 can be rewrited as

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{r} \tag{8}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_k)'$, $\mathbf{R} = \begin{pmatrix} 0,0,0,1,0,0,\cdots,0 \\ 0,0,0,0,1,0,\cdots,0 \\ \ddots \\ 0,0,0,0,0,0,\cdots,1 \end{pmatrix}$ is a $(k-2) \times (k+1)$ matrix, and $\mathbf{r} = (0,0,\cdots,0)'$ is a $(k-2) \times 1$ vector, according to (7).

At last few lessons, we have shown that

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1})$$

Hence we can derive that

$$\mathbf{R}\hat{\boldsymbol{\beta}} \sim N(\mathbf{R}\boldsymbol{\beta}, \sigma^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')$$

After normalizing $\mathbf{R}\hat{\boldsymbol{\beta}}$ and taking its square, we can find

$$\frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - r)'(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - r)}{\sigma^2} \sim \chi^2(G),$$

and $rank(\mathbf{R}) = G$.

According to the definition of F-distribution and equation (3), we can also replace σ with s^2

$$\frac{\dfrac{(\mathbf{R}\hat{\boldsymbol{\beta}} - r)'(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - r)}{\sigma^2}/G}{\dfrac{(n-k-1)s^2}{\sigma^2}/(n-k-1)} \sim F(G, n-k-1)$$

which can be rewritten as

$$\frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - r)'(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - r)/G}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k-1)} \sim F(G, n-k-1) \qquad (9)$$