複迴歸基本例題

(取自 Wooldridge Chap 3，Chap 4)

- 例題 4: 由「單迴歸」進入「複迴歸」

EXAMPLE 3.2 Hourly Wage Equation

Using the 526 observations on workers in WAGE1, we include *educ* (years of education), *exper* (years of labor market experience), and *tenure* (years with the current employer) in an equation explaining log(*wage*). The estimated equation is

$$\widehat{\log(wage)} = .284 + .092\, educ + .0041\, exper + .022\, tenure$$
$$n = 526. \quad [3.19]$$

As in the simple regression case, the coefficients have a percentage interpretation. The only difference here is that they also have a ceteris paribus interpretation. The coefficient .092 means that, holding *exper* and *tenure* fixed, another year of education is predicted to increase log(*wage*) by .092, which translates into an approximate 9.2% [100(.092)] increase in *wage*. Alternatively, if we take two people with the same levels of experience and job tenure, the coefficient on *educ* is the proportionate difference in predicted wage when their education levels differ by one year. This measure of the return to education at least keeps two important productivity factors fixed; whether it is a good estimate of the ceteris paribus return to another year of education requires us to study the statistical properties of OLS (see Section 3-3).

討論重點:

(1) 當解釋變數的數目增加時，R^2 有怎樣的變化?

(2) 與範例 2.10 比較，*educ* 的係數產生了怎樣的變化?

(3) 此兩道例題中，*educ* 的係數該如何解讀? (拉丁文 ceteris paribus 何意?)

● 例題 5: 承例題 4，增加「檢定」的討論

EXAMPLE 4.1 **Hourly Wage Equation**

Using the data in WAGE1 gives the estimated equation

$$\widehat{\log(wage)} = \underset{(.104)}{.284} + \underset{(.007)}{.092}\, educ + \underset{(.0017)}{.0041}\, exper + \underset{(.003)}{.022}\, tenure$$
$$n = 526,\ R^2 = .316,$$

where standard errors appear in parentheses below the estimated coefficients. We will follow this convention throughout the text. This equation can be used to test whether the return to *exper*, controlling for *educ* and *tenure*, is zero in the population, against the alternative that it is positive. Write this as H_0: $\beta_{exper} = 0$ versus H_1: $\beta_{exper} > 0$. (In applications, indexing a parameter by its associated variable name is a nice way to label parameters, since the numerical indices that we use in the general model are arbitrary and can cause confusion.) Remember that β_{exper} denotes the unknown population parameter. It is nonsense to write "H_0: .0041 = 0" or "H_0: $\hat{\beta}_{exper} = 0$."

Since we have 522 degrees of freedom, we can use the standard normal critical values. The 5% critical value is 1.645, and the 1% critical value is 2.326. The t statistic for $\hat{\beta}_{exper}$ is

$$t_{exper} = .0041/.0017 \approx 2.41,$$

and so $\hat{\beta}_{exper}$, or *exper*, is statistically significant even at the 1% level. We also say that "$\hat{\beta}_{exper}$ is statistically greater than zero at the 1% significance level."

The estimated return for another year of experience, holding tenure and education fixed, is not especially large. For example, adding three more years increases log(*wage*) by 3(.0041) = .0123, so wage is only about 1.2% higher. Nevertheless, we have persuasively shown that the partial effect of experience *is* positive in the population.

討論重點:

(1) 迴歸式中，欲檢定某個解釋變數是否對 y 變數有顯著影響，其「虛無假設」該怎麼寫才正確? (此處我們需區分有 hat 和沒有 hat 的差異)

(2) 括弧中的數字表「標準誤」(standard error)，如何用標準誤計算 t 統計量? (一般 t 統計量的臨界值是多少? 怎樣快速心算是否顯著?)

(3) 文中所謂的偏效果(partial effect)是什麼意思? (學習正確用詞)

(4) 如果某一項的係數檢定出來發現不顯著，那該作怎樣的處理?

附註:

t 統計量所遵守的 t 分配，自由度通常很大(正確數字為 df = n-k-1)，使其近似於一個常態分配(Z 分配)，因此我們常可以 Z 分配的臨界值作顯著性的概略判斷

(實務上顯著性的判斷仍仰賴 p-value 為主，軟體均會提供)