Parameter, Estimator, Estimate

H.M. James Hung

Regression analyses covered in this course are linear or nonlinear in parameters, not in regressors. For example, in a simple linear regression model

$$y = \beta_0 + \beta_1 x + \varepsilon$$

- β_0 and β_1 are parameters whose values are most often unknown and to be estimated from the data to be collected from a sample.

- Before the data are collected, we construct the so-called <u>estimators</u>, denoted by $\hat{\beta}_0$ and $\hat{\beta}_1$. They are functions of the data on (y, x). For instance, before the data are collected from 100 subjects, an estimator for β_1 could be

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{100} y_i x_i}{\sum_{i=1}^{100} x_i^2}$$

(this estimator may be good when $\boldsymbol{\beta_0} = 0$)

- After the data of the 100 subjects are collected, we plug the 100 values of (y , x) in the estimator $\hat{\beta}_1$. The value of $\hat{\beta}_1$ is then obtained and called <u>estimate</u> of β_1 .

It is critically important to distinguish between them when we write statistics.

As another note, we need to be careful about the following common errors.

- $y = \beta_0 + \beta_1 x$ is erroneous when x is a non-random regressor, because y is a random variable that has a statistical distribution, but $\beta_0 + \beta_1 x$ does not have a statistical distribution unless x is a random regressor.

- Variance is a parameter that needs to be estimated in most instances covered in this course. Thus, no distinction between variance and estimated variance or variance estimator is an error.

  The same kind of distinction as for the β's above is required when writing statistics. So is covariance or correlation.