Important concepts and useful mathematical identities

In this course, most often regressor x is fixed; that is, it does not have statistical distribution. Estimator is a function of y observations. Its variance or covariance comes from the statistical distribution of y or random error ε.

The following identities are useful:

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\,\bar{y}$$

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i - \bar{x})y_i$$

$$\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x}) = \sum_{i=1}^{n}(x_i - \bar{x})x_i$$

Suppose U and W are random variables, a and b are real numbers.

$$\mathrm{Var}(a\mathrm{U} + b\mathrm{W}) = a^2\mathrm{Var}(\mathrm{U}) + b^2\mathrm{Var}(\mathrm{W}) + 2ab\mathrm{Cov}(\mathrm{U}, \mathrm{W})$$

In statistics, there is clear distinction between parameter and estimator.

Parameter is for population of interest, its value is unknown, to be estimated from the observations (or so-called data) of a (random) sample via "estimator".

Estimator is a function of y observations. Once y observations are obtained from a random sample, we can then calculate the value of the estimator (this value is called "estimate") by plugging them into the estimator formula. The estimate differs from sample to sample, the estimate has variance or standard error.

Example: The random variable Y has a mean (or expectation) denoted by μ (this is a parameter), from a population of interest. We derive an unbiased estimator called sample mean from the data of Y. That is, if the y values from n subjects randomly selected from that population are available, then the estimator of μ would be the sample mean (this is an estimator) given by

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad .$$

Once the values of y are obtained, the value of this sample mean estimator is obtained by plugging the n y values into this sample mean formula. As we know, another sample will give us a different set of y values, thus, the sample mean formula will yield a different value accordingly.

In (pure) math, $1 + 1 = 2$.

In statistics, $1 + 1$ is not necessarily 2 because each value has a standard deviation or standard error. This is why most often statistical description is like

3 ± 1.