Introductory Econometrics

Tutorial 4

PART A: To prepare for this week's homework read the matrix algebra review on Moodle and the lecture notes for week 3.

.

PART B: This part will be covered in the tutorial. It is still a good idea to attempt these questions before the tutorial.

1. Suppose that

$$\underset{n\times 3}{\mathbf{X}} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \vdots & & \vdots \\ x_{n1} & x_{n2} & x_{n3} \end{bmatrix}$$

and

$$\underset{3\times 1}{\widehat{\boldsymbol{\beta}}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}$$

Show that $\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$ is an $n \times 1$ vector which is a linear combination (a weighted sum) of the columns of $\mathbf{X}$ with weights given by the elements of $\widehat{\boldsymbol{\beta}}$. That is:

$$\widehat{\mathbf{y}} = \text{first column of } \mathbf{X}\times\hat{\beta}_1 + \text{second column of } \mathbf{X}\times\hat{\beta}_2 + \text{third column of } \mathbf{X}\times\hat{\beta}_3$$

In fact this is not specific to $\mathbf{X}$ having 3 columns. It is true for any $n \times k$ matrix $\mathbf{X}$ and $k \times 1$ vector $\widehat{\boldsymbol{\beta}}$. Because of this, in linear regression, the predicted value of the dependent variable $\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$ is a linear combination of the columns of the matrix of independent variables $\mathbf{X}$.

2 This question is based on question C4 in Chapter 2 of the textbook. It is based on data on monthly salary and other characteristics of a random sample of 935 individuals. These data are in the file wage2.wf1. We concentrate on *wage* as the dependent variable and the IQ as the independent variable.

(a) Run a regression of *wage* on a constant only. Verify that the OLS estimate of the intercept is the sample mean of *wage* and the standard error of the regression is the sample standard deviation of *wage*.

(b) *Estimation, interpretation of the slope coefficient and R^2 of the regression*: Estimate a simple regression model where a one-point increase in *IQ* changes *wage* by a constant dollar amount. Use this model to find the predicted increase in *wage* for an increase in *IQ* of 15 points. Does *IQ* explain most of the variation in *wage*? What is the relationship between the R^2 of this regression and the sample correlation coefficient between wage and IQ?