Review Mechanics of OLS and Matrices

A. Colin Cameron U.C.-Davis

The first question does OLS manually using summation notation.
The second question does OLS using Stata or some other econometrics package.
The third question does OLS using matrix algebra. The fourth and fifth questions use matrix algebra.
The essentials are multiplication of conformable matrices and being able to take an inverse.

1. Let y denote ounces of butter consumed daily by a household and let x denote number of household members. A sample of four households has (x, y) equal to $(1, 1)$, $(2, 3)$, $(2, 4)$, and $(3, 4)$.

(a) Compute by hand $\bar{x}$, $\bar{y}$, $\sum_{i=1}^{4}(x_i - \bar{x})^2$, $\sum_{i=1}^{4}(y_i - \bar{y})^2$, and $\sum_{i=1}^{4}(x_i - \bar{x})(y_i - \bar{y})$.

(b) Compute the OLS estimates $\widehat{\beta}_1$ and $\widehat{\beta}_2$ from regression of ounces of butter on an intercept and household size.
[Note: For OLS $\widehat{y} = \widehat{\beta}_1 + \widehat{\beta}_2 x_i$; $\widehat{\beta}_2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$; $\widehat{\beta}_1 = \bar{y} - \widehat{\beta}_2 \bar{x}$.]
You should find $\widehat{\beta}_2 = 1.5$ and $\widehat{\beta}_1 = 0$.

(c) Compute $\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$ and hence R^2.

(d) Compute the sample correlation $r_{xy} = \dfrac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \times \sum_{i=1}^{N}(y_i - \bar{y})^2}}$ between ounces of butter and household size and compare to R^2 given in part (c).

(e) Compute $s^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$ and hence $\widehat{\mathrm{V}}[\widehat{\beta}_2] = s^2 / \sum_{i=1}^{n}(x_i - \bar{x})^2$.

(f) Test $H_0 : \beta_2 = 0$ against $H_a : \beta_2 \neq 0$ at significance level 0.05.

2. This question analyzes the same data set as the previous question, using Stata.
[If you do not have access to Stata then use an econometrics package that you do have access to.]
To save your Stata session output, once in Stata give the command `log using ass1log, text replace`. And when finished give the command `log close`.
Some help on Stata is available on my website but is not needed for this question. Just start up Stata on a PC or in a computer lab.

(a) Input the data using the Stata command `input` (for help in Stata give command `help input`).

(b) Then list the data using the Stata command `list`.

(c) Then get descriptive statistics using the Stata command `summarize`.
From these obtain $\bar{x}$, $\bar{y}$, $\sum_{i=1}^{4}(x_i - \bar{x})^2$ and $\sum_{i=1}^{4}(y_i - \bar{y})^2$.
[Hint: For the last two recall the definition of the sample variance $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$].

(d) Then regress ounces of butter on household size using the Stata command `regress` (for help in Stata give command `help regress`).

(e) Compare your OLS intercept and slope coefficient estimates and R^2 to those in the previous question.

(f) Compute the correlation of ounces of butter and household size using the Stata command `correlate` and compare to your answer in the previous question.

3. Consider the matrices $\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 4 \end{bmatrix}$.

Calculate the following by hand i.e. **do not** use a matrix package such as Matlab or Stata.

(a) Compute $\mathbf{X}'\mathbf{X}$. ($\mathbf{X}$ times the transpose of $\mathbf{X}$).

(b) Compute $(\mathbf{X}'\mathbf{X})^{-1}$. (The inverse of $\mathbf{X}'\mathbf{X}$).

(c) Compute $\mathbf{X}'\mathbf{y}$.

(d) Compute $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

(e) Compute $s^2 = \frac{1}{2}(\mathbf{y} - \widehat{\mathbf{y}})'(\mathbf{y} - \widehat{\mathbf{y}})$ where $\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$.

(f) Compute $\mathrm{V}[\widehat{\boldsymbol{\beta}}] = s^2(\mathbf{X}'\mathbf{X})^{-1}$.

(g) Compute $R^2 = 1 - (\mathbf{y} - \widehat{\mathbf{y}})'(\mathbf{y} - \widehat{\mathbf{y}})/(\mathbf{y} - \bar{y}\mathbf{e})'(\mathbf{y} - \bar{y}\mathbf{e})$ where $\bar{y}$ is the average of the elements of $\mathbf{y}$ and $\mathbf{e}$ is a 4×1 vector of ones.

(h) Compare your results to those in questions 1 and 2. They should be the same.

4. For the square matrix $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ answer the following from first principles.

(a) Is the matrix (i) triangular; (ii) symmetric; (iii) orthogonal; (iv) idempotent.

(b) Find the determinant and inverse of matrix A.

(c) For $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ find $\mathbf{x}'\mathbf{A}\mathbf{x}$ and hence state whether $\mathbf{A}$ is positive definite, positive semi-definite, negative definite, negative semi-definite, or none of these.

(d) Find the rank of $\mathbf{A}$ by finding the maximum number of linearly independent columns.

5. Consider the linear regression model

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

where $\mathbf{X}$ is an $N \times K$ matrix of constants and $\mathrm{E}[\mathbf{u}] = \mathbf{0}$ and $\mathrm{V}[\mathbf{u}] = \boldsymbol{\Omega}$, and the estimator

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}.$$

(a) Show that $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{u}$.

(b) Hence show that $\widehat{\boldsymbol{\beta}}$ is unbiased.

(c) Hence show that $\mathrm{V}[\widehat{\boldsymbol{\beta}}] = \mathrm{E}[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'] = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\boldsymbol{\Omega}\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}$.

(d) If $\mathbf{u}$ is normally distributed is $\widehat{\boldsymbol{\beta}}$ normally distributed?