3.4. Consider the sample regression

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$$

Imposing the restrictions (i) $\sum \hat{u}_i = 0$ and (ii) $\sum \hat{u}_i X_i = 0$, obtain the estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ and show that they are identical with the least-squares estimators given in Eqs. (3.1.6) and (3.1.7). This method of obtaining estimators is called the **analogy principle.** Give an intuitive justification for imposing restrictions (i) and (ii). (*Hint:* Recall the CLRM assumptions about u_i.) In passing, note that the analogy principle of estimating unknown parameters is also known as the **method of moments** in which sample moments (e.g., sample mean) are used to estimate population moments (e.g., the population mean). As noted in **Appendix A**, a **moment** is a summary statistic of a probability distribution, such as the expected value and variance.

3.5. Show that r^2 defined in (3.5.5) ranges between 0 and 1. You may use the Cauchy–Schwarz inequality, which states that for any random variables X and Y the following relationship holds true:

$$[E(XY)]^2 \leq E(X^2)E(Y^2)$$

3.6. Let $\hat{\beta}_{YX}$ and $\hat{\beta}_{XY}$ represent the slopes in the regression of Y on X and X on Y, respectively. Show that

$$\hat{\beta}_{YX}\hat{\beta}_{XY} = r^2$$

where r is the coefficient of correlation between X and Y.

3.7. Suppose in Exercise 3.6 that $\hat{\beta}_{YX}\hat{\beta}_{XY} = 1$. Does it matter then if we regress Y on X or X on Y? Explain carefully.

3.8. Spearman's rank correlation coefficient r_s is defined as follows:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

where $d =$ difference in the ranks assigned to the same individual or phenomenon and $n =$ number of individuals or phenomena ranked. Derive r_s from r defined in Eq. (3.5.13). *Hint:* Rank the X and Y values from 1 to n. Note that the sum of X and Y ranks is $n(n+1)/2$ each and therefore their means are $(n+1)/2$.

3.9. Consider the following formulations of the two-variable PRF:

$$\text{Model I:} \quad Y_i = \beta_1 + \beta_2 X_i + u_i$$

$$\text{Model II:} \quad Y_i = \alpha_1 + \alpha_2(X_i - \bar{X}) + u_i$$

a. Find the estimators of β_1 and α_1. Are they identical? Are their variances identical?

b. Find the estimators of β_2 and α_2. Are they identical? Are their variances identical?

c. What is the advantage, if any, of model II over model I?

3.10. Suppose you run the following regression:

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + \hat{u}_i$$

where, as usual, y_i and x_i are deviations from their respective mean values. What will be the value of $\hat{\beta}_1$? Why? Will $\hat{\beta}_2$ be the same as that obtained from Eq. (3.1.6)? Why?

3.19. *The relationship between nominal exchange rate and relative prices.* From annual observations from 1985 to 2005, the following regression results were obtained, where Y = exchange rate of the Canadian dollar to the U.S. dollar (CD/\$) and X = ratio of the U.S. consumer price index to the Canadian consumer price index; that is, X represents the relative prices in the two countries:

$$\hat{Y}_t = -0.912 + 2.250X_t \qquad r^2 = 0.440$$
$$\text{se} = \qquad\qquad 0.096$$

a. Interpret this regression. How would you interpret r^2?

b. Does the positive value of X_t make economic sense? What is the underlying economic theory?

c. Suppose we were to redefine X as the ratio of the Canadian CPI to the U.S. CPI. Would that change the sign of X? Why?

3.20. Table 3.6 gives data on indexes of output per hour (X) and real compensation per hour (Y) for the business and nonfarm business sectors of the U.S. economy for 1960–2005. The base year of the indexes is 1992 = 100 and the indexes are seasonally adjusted.

a. Plot Y against X for the two sectors separately.

b. What is the economic theory behind the relationship between the two variables? Does the scattergram support the theory?

c. Estimate the OLS regression of Y on X. Save the results for a further look after we study Chapter 5.

3.21. From a sample of 10 observations, the following results were obtained:

$$\sum Y_i = 1{,}110 \qquad \sum X_i = 1{,}700 \qquad \sum X_i Y_i = 205{,}500$$
$$\sum X_i^2 = 322{,}000 \qquad \sum Y_i^2 = 132{,}100$$

with coefficient of correlation $r = 0.9758$. But on rechecking these calculations it was found that two pairs of observations were recorded:

Y	X		Y	X
90	120	instead of	80	110
140	220		150	210

What will be the effect of this error on r? Obtain the correct r.

3.22. Table 3.7 gives data on gold prices, the Consumer Price Index (CPI), and the New York Stock Exchange (NYSE) Index for the United States for the period 1974–2006. The NYSE Index includes most of the stocks listed on the NYSE, some 1500-plus.

a. Plot in the same scattergram gold prices, CPI, and the NYSE Index.

b. An investment is supposed to be a hedge against inflation if its price and/or rate of return at least keeps pace with inflation. To test this hypothesis, suppose you decide to fit the following model, assuming the scatterplot in (a) suggests that this is appropriate:

$$\text{Gold price}_t = \beta_1 + \beta_2 \,\text{CPI}_t + u_t$$
$$\text{NYSE index}_t = \beta_1 + \beta_2 \,\text{CPI}_t + u_t$$

straightforward exercise in differential calculus. As shown in Appendix 3A, Section 3A.1, the process of differentiation yields the following equations for estimating β_1 and β_2:

$$\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i \tag{3.1.4}$$

$$\sum Y_i X_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2 \tag{3.1.5}$$

where n is the sample size. These simultaneous equations are known as the **normal equations.**

Solving the normal equations simultaneously, we obtain

$$\begin{aligned} \hat{\beta}_2 &= \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - \left(\sum X_i\right)^2} \\ &= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \\ &= \frac{\sum x_i y_i}{\sum x_i^2} \end{aligned} \tag{3.1.6}$$

where $\bar{X}$ and $\bar{Y}$ are the sample means of X and Y and where we define $x_i = (X_i - \bar{X})$ and $y_i = (Y_i - \bar{Y})$. *Henceforth, we adopt the convention of letting the lowercase letters denote deviations from mean values.*

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n \sum X_i^2 - \left(\sum X_i\right)^2} \\ &= \bar{Y} - \hat{\beta}_2 \bar{X} \end{aligned} \tag{3.1.7}$$

The last step in Equation 3.1.7 can be obtained directly from Eq. (3.1.4) by simple algebraic manipulations.

Incidentally, note that, by making use of simple algebraic identities, formula (3.1.6) for estimating β_2 can be alternatively expressed as

$$\begin{aligned} \hat{\beta}_2 &= \frac{\sum x_i y_i}{\sum x_i^2} \\ &= \frac{\sum x_i Y_i}{\sum X_i^2 - n\bar{X}^2} \\ &= \frac{\sum X_i y_i}{\sum X_i^2 - n\bar{X}^2} \end{aligned} \tag{3.1.8}$$ [2]

[2] *Note 1:* $\sum x_i^2 = \sum (X_i - \bar{X})^2 = \sum X_i^2 - 2\sum X_i \bar{X} + \sum \bar{X}^2 = \sum X_i^2 - 2\bar{X}\sum X_i + \sum \bar{X}^2$, since $\bar{X}$ is a constant. Further noting that $\sum X_i = n\bar{X}$ and $\sum \bar{X}^2 = n\bar{X}^2$ since $\bar{X}$ is a constant, we finally get $\sum x_i^2 = \sum X_i^2 - n\bar{X}^2$.

Note 2: $\sum x_i y_i = \sum x_i (Y_i - \bar{Y}) = \sum x_i Y_i - \bar{Y}\sum x_i = \sum x_i Y_i - \bar{Y}\sum (X_i - \bar{X}) = \sum x_i Y_i$, since $\bar{Y}$ is a constant and since the sum of deviations of a variable from its mean value [e.g., $\sum (X_i - \bar{X})$] is always zero. Likewise, $\sum y_i = \sum (Y_i - \bar{Y}) = 0$.