APPENDIX B

Developing Equations for Computing Regression Coefficients

Let a dependent variable Y and an independent variable x be related by an estimated regression function

$$Y = a + bx \tag{B.1}$$

Let Y_i be an estimate and y_i be an observed value of Y for a corresponding value x_i for x. Estimates of a and b can be obtained by minimizing the sum of the squares of the differences (R) between Y_i and y_i for a set of observed values, where

$$R = \sum_{i=1}^{n} (y_i - Y_i)^2 \tag{B.2}$$

Substituting $(a + bx_i)$ for Y_i in EqB.2, we obtain

$$R = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \tag{B.3}$$

Differentiating R partially with respect to a, then with respect to b, and equating each to zero, we obtain

$$\frac{\partial R}{\partial a} = -2 \sum_{i=1}^{n} (y_i - a - bx_i) = 0 \tag{B.4}$$

$$\frac{\partial R}{\partial b} = -2 \sum_{i=1}^{n} x_i(y_i - a - bx_i) = 0 \tag{B.5}$$

From EqB.4, we obtain

$$\sum_{i=1}^{n} y_i = na + b \sum_{i=1}^{n} x_i \tag{B.6}$$

giving

$$a = \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{b}{n}\sum_{i=1}^{n} x_i \tag{B.7}$$

From EqB.5, we obtain

$$\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 \tag{B.8}$$

Substituting for a, we obtain

$$b = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} \tag{B.9}$$

where

n = number of sets of observation
x_i = ith observation for x
y_i = ith observation for y

Eqations B.8 and B.9 may be used to obtain estimated values for a and b in EqB.1. To test the suitability of the regression function obtained, the coefficient of determination, R^2 (which indicates to what extent values of Y_i obtained from the regression function agree with observed values y_i) is determined from the expression

$$R^2 = \frac{\sum_{i=1}^{n} (Y_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{B.10}$$

which is also written as

$$R^2 = \frac{\sum_{i=1}^{n}\left(x_i y_i - n\bar{x}\,\bar{y}\right)^2}{\sum_{i=1}^{n} R\left(x_i^2 - n\bar{x}\right)\left(\sum_{i} y_i^2 - n\bar{y}^2\right)} \tag{B.11}$$

The closer the value of R^2 is to 1, the more suitable the estimated regression function for the data.