ASSUMPTION 1 **Linear Regression Model:** The regression model is **linear in the parameters**, though it may or may not be linear in the variables. That is the regression model as shown in Eq. (2.4.2):

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{2.4.2}$$

As will be discussed in Chapter 7, this model can be extended to include more explanatory variables.

We have already discussed model (2.4.2) in Chapter 2. Since linear-in-parameter regression models are the starting point of the CLRM, we will maintain this assumption for most of this book.[8] Keep in mind that the regressand Y and the regressor X may be nonlinear, as discussed in Chapter 2.

ASSUMPTION 2 **Fixed X Values or X Values Independent of the Error Term:** Values taken by the regressor X may be considered fixed in repeated samples (the case of fixed regressor) or they may be sampled along with the dependent variable Y (the case of stochastic regressor). In the latter case, it is assumed that the X variable(s) and the error term are independent, that is, cov $(X_i, u_i) = 0$.

This can be explained in terms of our example given in Table 2.1 (page 35). Consider the various Y populations corresponding to the levels of income shown in the table. Keeping the value of income X fixed, say, at level \$80, we draw at random a family and observe its weekly family consumption Y as, say, \$60. Still keeping X at \$80, we draw at random another family and observe its Y value at \$75. In each of these drawings (i.e., repeated sampling), the value of X is fixed at \$80. We can repeat this process for all the X values shown in Table 2.1. As a matter of fact, the sample data shown in Tables 2.4 and 2.5 were drawn in this fashion.

Why do we assume that the X values are nonstochastic? Given that, in most social sciences, data usually are collected randomly on both the Y and X variables, it seems natural to assume the opposite—that the X variable, like the Y variable, is also random or stochastic. But initially we assume that the X variable(s) is nonstochastic for the following reasons:

First, this is done initially to simplify the analysis and to introduce the reader to the complexities of regression analysis gradually. *Second,* in experimental situations it may not be unrealistic to assume that the X values are fixed. For example, a farmer may divide his land into several parcels and apply different amounts of fertilizer to these parcels to see its effect on crop yield. Likewise, a department store may decide to offer different rates of discount on a product to see its effect on consumers. Sometimes we may want to fix the X values for a specific purpose. Suppose we are trying to find out the average weekly earnings of workers (Y) with various levels of education (X), as in the case of the data given in Table 2.6. In this case, the X variable can be considered fixed or nonrandom. *Third,* as we show in Chapter 13, even if the X variables are stochastic, the statistical results of linear regression based

[8] However, a brief discussion of nonlinear-in-parameter regression models is given in Chapter 14 for the benefit of more advanced students.