Note that unlike β_0, β_1 and β_2, the parameters $\gamma_1, \ldots, \gamma_q$ are themselves random. Determination of the extent of hospital-to-hospital variation in the bill amount — rather than the specific values of $\gamma_1, \ldots, \gamma_q$ — may be the object of a study. Since a combination of fixed and random parameters are involved, a model of this kind is called a *mixed effects linear model.* A general form of this model is discussed in Section 8.3.□

Example 1.5.4 (Simultaneous equations model) A model for the demand for food (d) is

$$d = \beta_0 + \beta_1 r + \beta_2 r_o + \beta_3 s + \epsilon, \tag{1.5.4}$$

where r is the price of food, r_o is the price of other commodities, s is the income and ϵ is an error term. On the other hand, the price r can be modelled as

$$r = \alpha_0 + \alpha_1 d + \alpha_2 w + \delta, \tag{1.5.5}$$

where w is an indicator of weather condition and δ is another error term. Note that both the equations are important for the study of demand and price of food, and the roles of response and explanatory variables are interchanged in the two equations. Models described by simultaneous linear equations of this kind are very important in econometrics. We shall not deal with such models in this book. The interested reader may refer to Fomby et al. (1984, Chapters 19–24). □

Inference related to the models described in this section has a close connection with that for the model (1.1.1). Several other issues such as model building, diagnostics, and validation are also common to all these 'linear models'. The model (1.1.1) — by virtue of being the simplest — serves as a common reference or benchmark for the other models.

1.6 Uses of the linear model

An important application of the linear model is in regression analysis where the average response is explained via other observed variables (called *regressors* in this context). Building a working model to describe the relationship among the variables is sometimes an end in itself. On the other hand, the model may be used as a vehicle for various types of analyses, as described below.

The focus of the analysis may be on a particular regressor. For example, we may wish to find out specifically how the human longevity is linked to the cholesterol level in the blood. One may build a linear model for longevity so that other potentially influential variables such as gender, environmental factors, marital status and health factors other than the blood cholesterol level are also included. The model would have the form (1.1.1) where y is the longevity, x_1 is the blood cholesterol level and $x_2, \ldots, x_p$ represent the other regressors. According to this model, the parameter β_1 is the rate of change of average longevity with the blood cholesterol level, with the other factors held constant. Analysis of data on the basis of the linear model would produce an estimate of β_1, along with an estimate of the associated estimation error. Methods of obtaining these estimates are discussed in Chapters 4 and 7.

Regression analysis on the basis of a linear model is also carried out in order to examine statistically certain empirical beliefs regarding the model. For instance, in the context of Example 1.1.2, one may wish to test the statement that the duration of stay in the intensive care unit affects the hospital bill at least three times as much as the duration of stay outside the intensive care unit. The quantitative form of this statement is the hypothesis $\beta_2 \geq 3\beta_1$, which may be tested on the basis of available data. Tests of statistical hypotheses of this kind are discussed in Chapters 5 and 7.

Another important use of the linear regression model is in the area of prediction. The value of the main variable of interest (the response variable) may be impossible to obtain at the time of analysis or may involve expensive measurement. A few more readily available variables may be identified as explanatory variables or predictors. Data on all the variables are collected in order to fit a linear regression model. Unobserved values of the response may be predicted on the basis of the fitted model and the values of the explanatory variables corresponding to the unobserved response. See Chapters 5 and 7 for methods of prediction in the linear model. Sometimes unobserved values of the 'explanatory variables' are predicted on the basis of the corresponding response and a fitted model. This reverse prediction problem is called *calibration* (see Brown, 1993).

Apart from the situation where the observer has no control over the explanatory variables, *designed experiments* are also used to measure the effects of certain explanatory variables or to test empirical beliefs statistically. The linear model can be used by the experimenter as a basis for choosing the values of the controllable variables so that the answers to the crucial questions are best answered. Chapter 6 briefly outlines the basic issues of experimental design.

If one of the explanatory variables is controllable, then one might ask: Which value of this variable will produce a desired level of response (within a certain margin of error)? This task, which is related to calibration, is called the problem of *control*. The linear model provides a framework for solving this problem (see Press, 1971, Chapter 14).

A somewhat related question in polynomial regression is the following: How should one choose the explanatory variable so that the expected response is optimized? A similar question may be asked when the response is modelled as a polynomial in *several* variables. Typically there are constraints on the range of the explanatory variables. In the context of the assumed model, the problem reduces to finding the maximum or minimum of the estimated *response surface* within a certain range of the variables. An example without range constraints is given in Exercise 1.15. Various methods for response surfaces are described in detail by Khuri and Cornell (1996) and Box and Draper (1987).

The linear model is also used as a basis for imputation of missing data. The idea is to fill in the void using information from related variables, as in the case of prediction (see Titterington and Sedransk, 1987, for details). Diagnostic tools developed in the context of the linear model are sometimes used to detect other defects in the data such as bad or incorrect data (see Belsley et al., 1980).

1.7 A tour through the rest of the book

Chapters 2 and 3 provide a brief summary of various linear-algebraic and statistical concepts that are needed for the later chapters. The