

# Chapter 6

---

## Multiple Regression I

Multiple regression analysis is one of the most widely used of all statistical methods. In this chapter, we first discuss a variety of multiple regression models. Then we present the basic statistical results for multiple regression in matrix form. Since the matrix expressions for multiple regression are the same as for simple linear regression, we state the results without much discussion. We conclude the chapter with an example, illustrating a variety of inferences and residual analyses in multiple regression analysis.

### 6.1 Multiple Regression Models

---

#### Need for Several Predictor Variables

When we first introduced regression analysis in Chapter 1, we spoke of regression models containing a number of predictor variables. We mentioned a regression model where the response variable was direct operating cost for a branch office of a consumer finance chain, and four predictor variables were considered, including average number of loans outstanding at the branch and total number of new loan applications processed by the branch. We also mentioned a tractor purchase study where the response variable was volume of tractor purchases in a sales territory, and the nine predictor variables included number of farms in the territory and quantity of crop production in the territory. In addition, we mentioned a study of short children where the response variable was the peak plasma growth hormone level, and the 14 predictor variables included gender, age, and various body measurements. In all these examples, a single predictor variable in the model would have provided an inadequate description since a number of key variables affect the response variable in important and distinctive ways. Furthermore, in situations of this type, we frequently find that predictions of the response variable based on a model containing only a single predictor variable are too imprecise to be useful. We noted the imprecise predictions with a single predictor variable in the Toluca Company example in Chapter 2. A more complex model, containing additional predictor variables, typically is more helpful in providing sufficiently precise predictions of the response variable.

In each of the examples just mentioned, the analysis was based on observational data because the predictor variables were not controlled, usually because they were not susceptible to direct control. Multiple regression analysis is also highly useful in experimental situations where the experimenter can control the predictor variables. An experimenter typically will wish to investigate a number of predictor variables simultaneously because almost always

more than one key predictor variable influences the response. For example, in a study of productivity of work crews, the experimenter may wish to control both the size of the crew and the level of bonus pay. Similarly, in a study of responsiveness to a drug, the experimenter may wish to control both the dose of the drug and the method of administration.

The multiple regression models which we now describe can be utilized for either observational data or for experimental data from a completely randomized design.

### First-Order Model with Two Predictor Variables

When there are two predictor variables  $X_1$  and  $X_2$ , the regression model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \quad (6.1)$$

is called a first-order model with two predictor variables. A first-order model, as we noted in Chapter 1, is linear in the predictor variables.  $Y_i$  denotes as usual the response in the  $i$ th trial, and  $X_{i1}$  and  $X_{i2}$  are the values of the two predictor variables in the  $i$ th trial. The parameters of the model are  $\beta_0$ ,  $\beta_1$ , and  $\beta_2$ , and the error term is  $\varepsilon_i$ .

Assuming that  $E\{\varepsilon_i\} = 0$ , the regression function for model (6.1) is:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \quad (6.2)$$

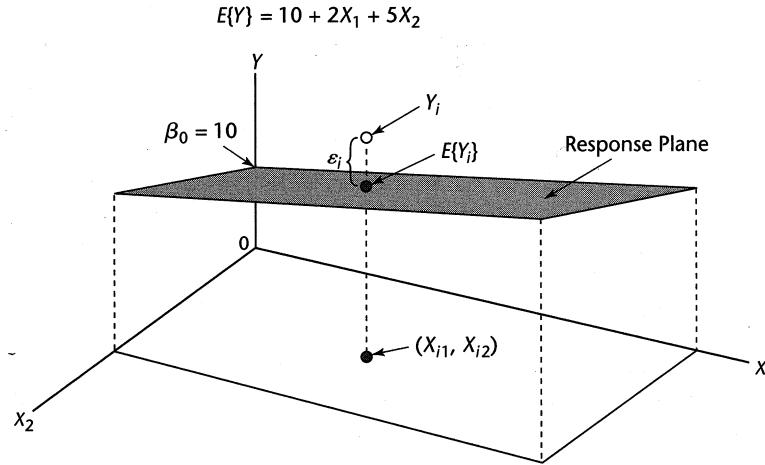
Analogous to simple linear regression, where the regression function  $E\{Y\} = \beta_0 + \beta_1 X$  is a line, regression function (6.2) is a plane. Figure 6.1 contains a representation of a portion of the response plane:

$$E\{Y\} = 10 + 2X_1 + 5X_2 \quad (6.3)$$

Note that any point on the response plane (6.3) corresponds to the mean response  $E\{Y\}$  at the given combination of levels of  $X_1$  and  $X_2$ .

Figure 6.1 also shows an observation  $Y_i$  corresponding to the levels  $(X_{i1}, X_{i2})$  of the two predictor variables. Note that the vertical rule in Figure 6.1 between  $Y_i$  and the response plane represents the difference between  $Y_i$  and the mean  $E\{Y_i\}$  of the probability distribution of  $Y$  for the given  $(X_{i1}, X_{i2})$  combination. Hence, the vertical distance from  $Y_i$  to the response plane represents the error term  $\varepsilon_i = Y_i - E\{Y_i\}$ .

**FIGURE 6.1**  
Response  
Function is a  
Plane—Sales  
Promotion  
Example.



Frequently the regression function in multiple regression is called a *regression surface* or a *response surface*. In Figure 6.1, the response surface is a plane, but in other cases the response surface may be more complex in nature.

**Meaning of Regression Coefficients.** Let us now consider the meaning of the regression coefficients in the multiple regression function (6.3). The parameter  $\beta_0 = 10$  is the  $Y$  intercept of the regression plane. If the scope of the model includes  $X_1 = 0, X_2 = 0$ , then  $\beta_0 = 10$  represents the mean response  $E\{Y\}$  at  $X_1 = 0, X_2 = 0$ . Otherwise,  $\beta_0$  does not have any particular meaning as a separate term in the regression model.

The parameter  $\beta_1$  indicates the change in the mean response  $E\{Y\}$  per unit increase in  $X_1$  when  $X_2$  is held constant. Likewise,  $\beta_2$  indicates the change in the mean response per unit increase in  $X_2$  when  $X_1$  is held constant. To see this for our example, suppose  $X_2$  is held at the level  $X_2 = 2$ . The regression function (6.3) now is:

$$E\{Y\} = 10 + 2X_1 + 5(2) = 20 + 2X_1 \quad X_2 = 2 \quad (6.4)$$

Note that this response function is a straight line with slope  $\beta_1 = 2$ . The same is true for any other value of  $X_2$ ; only the intercept of the response function will differ. Hence,  $\beta_1 = 2$  indicates that the mean response  $E\{Y\}$  increases by 2 with a unit increase in  $X_1$  when  $X_2$  is constant, no matter what the level of  $X_2$ . We confirm therefore that  $\beta_1$  indicates the change in  $E\{Y\}$  with a unit increase in  $X_1$  when  $X_2$  is held constant.

Similarly,  $\beta_2 = 5$  in regression function (6.3) indicates that the mean response  $E\{Y\}$  increases by 5 with a unit increase in  $X_2$  when  $X_1$  is held constant.

When the effect of  $X_1$  on the mean response does not depend on the level of  $X_2$ , and correspondingly the effect of  $X_2$  does not depend on the level of  $X_1$ , the two predictor variables are said to have *additive effects* or *not to interact*. Thus, the first-order regression model (6.1) is designed for predictor variables whose effects on the mean response are additive or do not interact.

The parameters  $\beta_1$  and  $\beta_2$  are sometimes called *partial regression coefficients* because they reflect the partial effect of one predictor variable when the other predictor variable is included in the model and is held constant.

### Example

The response plane (6.3) shown in Figure 6.1 is for a regression model relating test market sales ( $Y$ , in 10 thousand dollars) to point-of-sale expenditures ( $X_1$ , in thousand dollars) and TV expenditures ( $X_2$ , in thousand dollars). Since  $\beta_1 = 2$ , if point-of-sale expenditures in a locality are increased by one unit (1 thousand dollars) while TV expenditures are held constant, expected sales increase by 2 units (20 thousand dollars). Similarly, since  $\beta_2 = 5$ , if TV expenditures in a locality are increased by 1 thousand dollars and point-of-sale expenditures are held constant, expected sales increase by 50 thousand dollars.

### Comments

1. A regression model for which the response surface is a plane can be used either in its own right when it is appropriate, or as an approximation to a more complex response surface. Many complex response surfaces can be approximated well by a plane for limited ranges of  $X_1$  and  $X_2$ .

2. We can readily establish the meaning of  $\beta_1$  and  $\beta_2$  by calculus, taking partial derivatives of the response surface (6.2) with respect to  $X_1$  and  $X_2$  in turn:

$$\frac{\partial E\{Y\}}{\partial X_1} = \beta_1 \quad \frac{\partial E\{Y\}}{\partial X_2} = \beta_2$$

The partial derivatives measure the rate of change in  $E\{Y\}$  with respect to one predictor variable when the other is held constant. ■

### First-Order Model with More than Two Predictor Variables

We consider now the case where there are  $p - 1$  predictor variables  $X_1, \dots, X_{p-1}$ . The regression model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad (6.5)$$

is called a first-order model with  $p - 1$  predictor variables. It can also be written:

$$Y_i = \beta_0 + \sum_{k=1}^{p-1} \beta_k X_{ik} + \varepsilon_i \quad (6.5a)$$

or, if we let  $X_{i0} \equiv 1$ , it can be written as:

$$Y_i = \sum_{k=0}^{p-1} \beta_k X_{ik} + \varepsilon_i \quad \text{where } X_{i0} \equiv 1 \quad (6.5b)$$

Assuming that  $E\{\varepsilon_i\} = 0$ , the response function for regression model (6.5) is:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{p-1} X_{p-1} \quad (6.6)$$

This response function is a *hyperplane*, which is a plane in more than two dimensions. It is no longer possible to picture this response surface, as we were able to do in Figure 6.1 for the case of two predictor variables. Nevertheless, the meaning of the parameters is analogous to the case of two predictor variables. The parameter  $\beta_k$  indicates the change in the mean response  $E\{Y\}$  with a unit increase in the predictor variable  $X_k$ , when all other predictor variables in the regression model are held constant. Note again that the effect of any predictor variable on the mean response is the same for regression model (6.5) no matter what are the levels at which the other predictor variables are held. Hence, first-order regression model (6.5) is designed for predictor variables whose effects on the mean response are additive and therefore do not interact.

#### Comment

When  $p - 1 = 1$ , regression model (6.5) reduces to:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$$

which is the simple linear regression model considered in earlier chapters. ■

### General Linear Regression Model

In general, the variables  $X_1, \dots, X_{p-1}$  in a regression model do not need to represent different predictor variables, as we shall shortly see. We therefore define the general linear

regression model, with normal error terms, simply in terms of  $X$  variables:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad (6.7)$$

where:

$\beta_0, \beta_1, \dots, \beta_{p-1}$  are parameters

$X_{i1}, \dots, X_{i,p-1}$  are known constants

$\varepsilon_i$  are independent  $N(0, \sigma^2)$

$i = 1, \dots, n$

If we let  $X_{i0} \equiv 1$ , regression model (6.7) can be written as follows:

$$Y_i = \beta_0 X_{i0} + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad (6.7a)$$

where  $X_{i0} \equiv 1$ , or:

$$Y_i = \sum_{k=0}^{p-1} \beta_k X_{ik} + \varepsilon_i \quad \text{where } X_{i0} \equiv 1 \quad (6.7b)$$

The response function for regression model (6.7) is, since  $E\{\varepsilon_i\} = 0$ :

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{p-1} X_{p-1} \quad (6.8)$$

Thus, the general linear regression model with normal error terms implies that the observations  $Y_i$  are independent normal variables, with mean  $E\{Y_i\}$  as given by (6.8) and with constant variance  $\sigma^2$ .

This general linear model encompasses a vast variety of situations. We consider a few of these now.

**$p - 1$  Predictor Variables.** When  $X_1, \dots, X_{p-1}$  represent  $p - 1$  different predictor variables, general linear regression model (6.7) is, as we have seen, a first-order model in which there are no interaction effects between the predictor variables. The example in Figure 6.1 involves a first-order model with two predictor variables.

**Qualitative Predictor Variables.** The general linear regression model (6.7) encompasses not only quantitative predictor variables but also qualitative ones, such as gender (male, female) or disability status (not disabled, partially disabled, fully disabled). We use indicator variables that take on the values 0 and 1 to identify the classes of a qualitative variable.

Consider a regression analysis to predict the length of hospital stay ( $Y$ ) based on the age ( $X_1$ ) and gender ( $X_2$ ) of the patient. We define  $X_2$  as follows:

$$X_2 = \begin{cases} 1 & \text{if patient female} \\ 0 & \text{if patient male} \end{cases}$$

The first-order regression model then is as follows:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \quad (6.9)$$

where:

$X_{i1}$  = patient's age

$$X_{i2} = \begin{cases} 1 & \text{if patient female} \\ 0 & \text{if patient male} \end{cases}$$

The response function for regression model (6.9) is:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \quad (6.10)$$

For male patients,  $X_2 = 0$  and response function (6.10) becomes:

$$E\{Y\} = \beta_0 + \beta_1 X_1 \quad \text{Male patients} \quad (6.10a)$$

For female patients,  $X_2 = 1$  and response function (6.10) becomes:

$$E\{Y\} = (\beta_0 + \beta_2) + \beta_1 X_1 \quad \text{Female patients} \quad (6.10b)$$

These two response functions represent parallel straight lines with different intercepts.

In general, we represent a qualitative variable with  $c$  classes by means of  $c - 1$  indicator variables. For instance, if in the hospital stay example the qualitative variable disability status is to be added as another predictor variable, it can be represented as follows by the two indicator variables  $X_3$  and  $X_4$ :

$$X_3 = \begin{cases} 1 & \text{if patient not disabled} \\ 0 & \text{otherwise} \end{cases}$$

$$X_4 = \begin{cases} 1 & \text{if patient partially disabled} \\ 0 & \text{otherwise} \end{cases}$$

The first-order model with age, gender, and disability status as predictor variables then is:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i \quad (6.11)$$

where:

$$X_{i1} = \text{patient's age}$$

$$X_{i2} = \begin{cases} 1 & \text{if patient female} \\ 0 & \text{if patient male} \end{cases}$$

$$X_{i3} = \begin{cases} 1 & \text{if patient not disabled} \\ 0 & \text{otherwise} \end{cases}$$

$$X_{i4} = \begin{cases} 1 & \text{if patient partially disabled} \\ 0 & \text{otherwise} \end{cases}$$

In Chapter 8 we present a comprehensive discussion of how to model qualitative predictor variables and how to interpret regression models containing qualitative predictor variables.

**Polynomial Regression.** Polynomial regression models are special cases of the general linear regression model. They contain squared and higher-order terms of the predictor variable(s), making the response function curvilinear. The following is a polynomial regression model with one predictor variable:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i \quad (6.12)$$

Figure 1.3 on page 5 shows an example of a polynomial regression function with one predictor variable.

Despite the curvilinear nature of the response function for regression model (6.12), it is a special case of general linear regression model (6.7). If we let  $X_{i1} = X_i$  and  $X_{i2} = X_i^2$ , we can write (6.12) as follows:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

which is in the form of general linear regression model (6.7). While (6.12) illustrates a curvilinear regression model where the response function is quadratic, models with higher-degree polynomial response functions are also particular cases of the general linear regression model. We shall discuss polynomial regression models in more detail in Chapter 8.

**Transformed Variables.** Models with transformed variables involve complex, curvilinear response functions, yet still are special cases of the general linear regression model. Consider the following model with a transformed  $Y$  variable:

$$\log Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \quad (6.13)$$

Here, the response surface is complex, yet model (6.13) can still be treated as a general linear regression model. If we let  $Y'_i = \log Y_i$ , we can write regression model (6.13) as follows:

$$Y'_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

which is in the form of general linear regression model (6.7). The response variable just happens to be the logarithm of  $Y$ .

Many models can be transformed into the general linear regression model. For instance, the model:

$$Y_i = \frac{1}{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i} \quad (6.14)$$

can be transformed to the general linear regression model by letting  $Y'_i = 1/Y_i$ . We then have:

$$Y'_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

**Interaction Effects.** When the effects of the predictor variables on the response variable are not additive, the effect of one predictor variable depends on the levels of the other predictor variables. The general linear regression model (6.7) encompasses regression models with nonadditive or interacting effects. An example of a nonadditive regression model with two predictor variables  $X_1$  and  $X_2$  is the following:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i \quad (6.15)$$

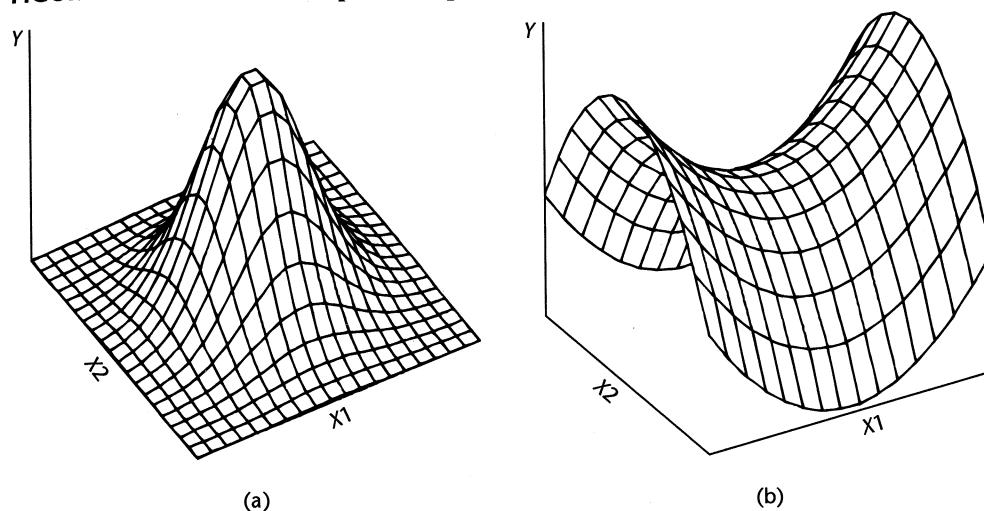
Here, the response function is complex because of the interaction term  $\beta_3 X_{i1} X_{i2}$ . Yet regression model (6.15) is a special case of the general linear regression model. Let  $X_{i3} = X_{i1} X_{i2}$  and then write (6.15) as follows:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

We see that this model is in the form of general linear regression model (6.7). We shall discuss regression models with interaction effects in more detail in Chapter 8.

**Combination of Cases.** A regression model may combine several of the elements we have just noted and still be treated as a general linear regression model. Consider the following regression model containing linear and quadratic terms for each of two predictor variables and an interaction term represented by the cross-product term:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i1}^2 + \beta_3 X_{i2} + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} + \varepsilon_i \quad (6.16)$$

**FIGURE 6.2 Additional Examples of Response Functions.**

Let us define:

$$Z_{i1} = X_{i1} \quad Z_{i2} = X_{i1}^2 \quad Z_{i3} = X_{i2} \quad Z_{i4} = X_{i2}^2 \quad Z_{i5} = X_{i1}X_{i2}$$

We can then write regression model (6.16) as follows:

$$Y_i = \beta_0 + \beta_1 Z_{i1} + \beta_2 Z_{i2} + \beta_3 Z_{i3} + \beta_4 Z_{i4} + \beta_5 Z_{i5} + \varepsilon_i$$

which is in the form of general linear regression model (6.7).

The general linear regression model (6.7) includes many complex models, some of which may be highly complex. Figure 6.2 illustrates two complex response surfaces when there are two predictor variables, that can be represented by general linear regression model (6.7).

**Meaning of Linear in General Linear Regression Model.** It should be clear from the various examples that general linear regression model (6.7) is not restricted to linear response surfaces. The term *linear model* refers to the fact that model (6.7) is linear in the parameters; it does not refer to the shape of the response surface.

We say that a regression model is linear in the parameters when it can be written in the form:

$$Y_i = c_{i0}\beta_0 + c_{i1}\beta_1 + c_{i2}\beta_2 + \cdots + c_{i,p-1}\beta_{p-1} + \varepsilon_i \quad (6.17)$$

where the terms  $c_{i0}$ ,  $c_{i1}$ , etc., are coefficients involving the predictor variables. For example, first-order model (6.1) in two predictor variables:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

is linear in the parameters, with  $c_{i0} = 1$ ,  $c_{i1} = X_{i1}$ , and  $c_{i2} = X_{i2}$ .

An example of a nonlinear regression model is the following:

$$Y_i = \beta_0 \exp(\beta_1 X_i) + \varepsilon_i$$

This is a nonlinear regression model because it cannot be expressed in the form of (6.17). We shall discuss nonlinear regression models in Part III.

## 6.2 General Linear Regression Model in Matrix Terms

We now present the principal results for the general linear regression model (6.7) in matrix terms. This model, as noted, encompasses a wide variety of particular cases. The results to be presented are applicable to all of these.

It is a remarkable property of matrix algebra that the results for the general linear regression model (6.7) in matrix notation appear exactly as those for the simple linear regression model (5.57). Only the degrees of freedom and other constants related to the number of  $X$  variables and the dimensions of some matrices are different. Hence, we are able to present the results very concisely.

The matrix notation, to be sure, may hide enormous computational complexities. To find the inverse of a  $10 \times 10$  matrix  $\mathbf{A}$  requires a tremendous amount of computation, yet it is simply represented as  $\mathbf{A}^{-1}$ . Our reason for emphasizing matrix algebra is that it indicates the essential conceptual steps in the solution. The actual computations will, in all but the very simplest cases, be done by computer. Hence, it does not matter to us whether  $(\mathbf{X}'\mathbf{X})^{-1}$  represents finding the inverse of a  $2 \times 2$  or a  $10 \times 10$  matrix. The important point is to know what the inverse of the matrix represents.

To express general linear regression model (6.7):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

in matrix terms, we need to define the following matrices:

(6.18a)

$$\mathbf{Y}_{n \times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

(6.18b)

$$\mathbf{X}_{n \times p} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,p-1} \end{bmatrix}$$

(6.18)

(6.18c)

$$\boldsymbol{\beta}_{p \times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}$$

(6.18d)

$$\boldsymbol{\varepsilon}_{n \times 1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Note that the  $\mathbf{Y}$  and  $\boldsymbol{\varepsilon}$  vectors are the same as for simple linear regression. The  $\boldsymbol{\beta}$  vector contains additional regression parameters, and the  $\mathbf{X}$  matrix contains a column of 1s as well as a column of the  $n$  observations for each of the  $p - 1$   $X$  variables in the regression model. The row subscript for each element  $X_{ik}$  in the  $\mathbf{X}$  matrix identifies the trial or case, and the column subscript identifies the  $X$  variable.

In matrix terms, the general linear regression model (6.7) is:

$$\mathbf{Y}_{n \times 1} = \mathbf{X}_{n \times p} \boldsymbol{\beta}_{p \times 1} + \boldsymbol{\varepsilon}_{n \times 1} \quad (6.19)$$

where:

$\mathbf{Y}$  is a vector of responses

$\boldsymbol{\beta}$  is a vector of parameters

$\mathbf{X}$  is a matrix of constants

$\boldsymbol{\epsilon}$  is a vector of independent normal random variables with expectation

$E\{\boldsymbol{\epsilon}\} = \mathbf{0}$  and variance-covariance matrix:

$$\sigma^2_{n \times n} \{ \boldsymbol{\epsilon} \} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}$$

Consequently, the random vector  $\mathbf{Y}$  has expectation:

$$E\{\mathbf{Y}\} = \mathbf{X}\boldsymbol{\beta}_{n \times 1} \quad (6.20)$$

and the variance-covariance matrix of  $\mathbf{Y}$  is the same as that of  $\boldsymbol{\epsilon}$ :

$$\sigma^2_{n \times n} \{ \mathbf{Y} \} = \sigma^2 \mathbf{I} \quad (6.21)$$

### 6.3 Estimation of Regression Coefficients

---

The least squares criterion (1.8) is generalized as follows for general linear regression model (6.7):

$$Q = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_{p-1} X_{ip-1})^2 \quad (6.22)$$

The least squares estimators are those values of  $\beta_0, \beta_1, \dots, \beta_{p-1}$  that minimize  $Q$ . Let us denote the vector of the least squares estimated regression coefficients  $b_0, b_1, \dots, b_{p-1}$  as  $\mathbf{b}$ :

$$\mathbf{b}_{p \times 1} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix} \quad (6.23)$$

The least squares normal equations for the general linear regression model (6.19) are:

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y} \quad (6.24)$$

and the least squares estimators are:

$$\mathbf{b}_{2 \times 1} = (\mathbf{X}'\mathbf{X})_{2 \times 2}^{-1} (\mathbf{X}'\mathbf{Y})_{2 \times 1} \quad (6.25)$$

The method of maximum likelihood leads to the same estimators for normal error regression model (6.19) as those obtained by the method of least squares in (6.25). The likelihood function in (1.26) generalizes directly for multiple regression as follows:

$$L(\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_{p-1} X_{ip-1})^2 \right] \quad (6.26)$$

Maximizing this likelihood function with respect to  $\beta_0, \beta_1, \dots, \beta_{p-1}$  leads to the estimators in (6.25). These estimators are least squares and maximum likelihood estimators and have all the properties mentioned in Chapter 1: they are minimum variance unbiased, consistent, and sufficient.

## 6.4 Fitted Values and Residuals

---

Let the vector of the fitted values  $\hat{Y}_i$  be denoted by  $\hat{\mathbf{Y}}$  and the vector of the residual terms  $e_i = Y_i - \hat{Y}_i$  be denoted by  $\mathbf{e}$ :

$$(6.27a) \quad \hat{\mathbf{Y}}_{n \times 1} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} \quad (6.27b) \quad \mathbf{e}_{n \times 1} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \quad (6.27)$$

The fitted values are represented by:

$$\hat{\mathbf{Y}}_{n \times 1} = \mathbf{X}\mathbf{b} \quad (6.28)$$

and the residual terms by:

$$\mathbf{e}_{n \times 1} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\mathbf{b} \quad (6.29)$$

The vector of the fitted values  $\hat{\mathbf{Y}}$  can be expressed in terms of the hat matrix  $\mathbf{H}$  as follows:

$$\hat{\mathbf{Y}}_{n \times 1} = \mathbf{H}\mathbf{Y} \quad (6.30)$$

where:

$$\mathbf{H}_{n \times n} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \quad (6.30a)$$

Similarly, the vector of residuals can be expressed as follows:

$$\mathbf{e}_{n \times 1} = (\mathbf{I} - \mathbf{H})\mathbf{Y} \quad (6.31)$$

The variance-covariance matrix of the residuals is:

$$\sigma^2 \{\mathbf{e}\}_{n \times n} = \sigma^2 (\mathbf{I} - \mathbf{H}) \quad (6.32)$$

which is estimated by:

$$\mathbf{s}^2_{n \times n}(\mathbf{e}) = MSE(\mathbf{I} - \mathbf{H}) \quad (6.33)$$

## 6.5 Analysis of Variance Results

### Sums of Squares and Mean Squares

The sums of squares for the analysis of variance in matrix terms are, from (5.89):

$$SSTO = \mathbf{Y}'\mathbf{Y} - \left(\frac{1}{n}\right) \mathbf{Y}'\mathbf{J}\mathbf{Y} = \mathbf{Y}' \left[ \mathbf{I} - \left(\frac{1}{n}\right) \mathbf{J} \right] \mathbf{Y} \quad (6.34)$$

$$SSE = \mathbf{e}'\mathbf{e} = (\mathbf{Y} - \mathbf{X}\mathbf{b})'(\mathbf{Y} - \mathbf{X}\mathbf{b}) = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y} = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y} \quad (6.35)$$

$$SSR = \mathbf{b}'\mathbf{X}'\mathbf{Y} - \left(\frac{1}{n}\right) \mathbf{Y}'\mathbf{J}\mathbf{Y} = \mathbf{Y}' \left[ \mathbf{H} - \left(\frac{1}{n}\right) \mathbf{J} \right] \mathbf{Y} \quad (6.36)$$

where  $\mathbf{J}$  is an  $n \times n$  matrix of 1s defined in (5.18) and  $\mathbf{H}$  is the hat matrix defined in (6.30a).

$SSTO$ , as usual, has  $n - 1$  degrees of freedom associated with it.  $SSE$  has  $n - p$  degrees of freedom associated with it since  $p$  parameters need to be estimated in the regression function for model (6.19). Finally,  $SSR$  has  $p - 1$  degrees of freedom associated with it, representing the number of  $X$  variables  $X_1, \dots, X_{p-1}$ .

Table 6.1 shows these analysis of variance results, as well as the mean squares  $MSR$  and  $MSE$ :

$$MSR = \frac{SSR}{p - 1} \quad (6.37)$$

$$MSE = \frac{SSE}{n - p} \quad (6.38)$$

The expectation of  $MSE$  is  $\sigma^2$ , as for simple linear regression. The expectation of  $MSR$  is  $\sigma^2$  plus a quantity that is nonnegative. For instance, when  $p - 1 = 2$ , we have:

$$E\{MSR\} = \sigma^2 + \frac{1}{2} \left[ \beta_1^2 \sum (X_{i1} - \bar{X}_1)^2 + \beta_2^2 \sum (X_{i2} - \bar{X}_2)^2 + 2\beta_1\beta_2 \sum (X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2) \right]$$

Note that if both  $\beta_1$  and  $\beta_2$  equal zero,  $E\{MSR\} = \sigma^2$ . Otherwise  $E\{MSR\} > \sigma^2$ .

**TABLE 6.1**  
ANOVA Table  
for General  
Linear  
Regression  
Model (6.19).

Source of Variation	SS	df	MS
Regression	$SSR = \mathbf{b}'\mathbf{X}'\mathbf{Y} - \left(\frac{1}{n}\right) \mathbf{Y}'\mathbf{J}\mathbf{Y}$	$p - 1$	$MSR = \frac{SSR}{p - 1}$
Error	$SSE = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y}$	$n - p$	$MSE = \frac{SSE}{n - p}$
Total	$SSTO = \mathbf{Y}'\mathbf{Y} - \left(\frac{1}{n}\right) \mathbf{Y}'\mathbf{J}\mathbf{Y}$	$n - 1$	

which is estimated by:

$$\mathbf{s}^2_{n \times n} \{\mathbf{e}\} = MSE(\mathbf{I} - \mathbf{H}) \quad (6.33)$$

## 6.5 Analysis of Variance Results

### Sums of Squares and Mean Squares

The sums of squares for the analysis of variance in matrix terms are, from (5.89):

$$SSTO = \mathbf{Y}'\mathbf{Y} - \left(\frac{1}{n}\right) \mathbf{Y}'\mathbf{J}\mathbf{Y} = \mathbf{Y}' \left[ \mathbf{I} - \left(\frac{1}{n}\right) \mathbf{J} \right] \mathbf{Y} \quad (6.34)$$

$$SSE = \mathbf{e}'\mathbf{e} = (\mathbf{Y} - \mathbf{X}\mathbf{b})'(\mathbf{Y} - \mathbf{X}\mathbf{b}) = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y} = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y} \quad (6.35)$$

$$SSR = \mathbf{b}'\mathbf{X}'\mathbf{Y} - \left(\frac{1}{n}\right) \mathbf{Y}'\mathbf{J}\mathbf{Y} = \mathbf{Y}' \left[ \mathbf{H} - \left(\frac{1}{n}\right) \mathbf{J} \right] \mathbf{Y} \quad (6.36)$$

where  $\mathbf{J}$  is an  $n \times n$  matrix of 1s defined in (5.18) and  $\mathbf{H}$  is the hat matrix defined in (6.30a).

$SSTO$ , as usual, has  $n - 1$  degrees of freedom associated with it.  $SSE$  has  $n - p$  degrees of freedom associated with it since  $p$  parameters need to be estimated in the regression function for model (6.19). Finally,  $SSR$  has  $p - 1$  degrees of freedom associated with it, representing the number of  $X$  variables  $X_1, \dots, X_{p-1}$ .

Table 6.1 shows these analysis of variance results, as well as the mean squares  $MSR$  and  $MSE$ :

$$MSR = \frac{SSR}{p - 1} \quad (6.37)$$

$$MSE = \frac{SSE}{n - p} \quad (6.38)$$

The expectation of  $MSE$  is  $\sigma^2$ , as for simple linear regression. The expectation of  $MSR$  is  $\sigma^2$  plus a quantity that is nonnegative. For instance, when  $p - 1 = 2$ , we have:

$$E\{MSR\} = \sigma^2 + \frac{1}{2} \left[ \beta_1^2 \sum (X_{i1} - \bar{X}_1)^2 + \beta_2^2 \sum (X_{i2} - \bar{X}_2)^2 + 2\beta_1\beta_2 \sum (X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2) \right]$$

Note that if both  $\beta_1$  and  $\beta_2$  equal zero,  $E\{MSR\} = \sigma^2$ . Otherwise  $E\{MSR\} > \sigma^2$ .

**TABLE 6.1**  
ANOVA Table  
for General  
Linear  
Regression  
Model (6.19).

Source of Variation	SS	df	MS
Regression	$SSR = \mathbf{b}'\mathbf{X}'\mathbf{Y} - \left(\frac{1}{n}\right) \mathbf{Y}'\mathbf{J}\mathbf{Y}$	$p - 1$	$MSR = \frac{SSR}{p - 1}$
Error	$SSE = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y}$	$n - p$	$MSE = \frac{SSE}{n - p}$
Total	$SSTO = \mathbf{Y}'\mathbf{Y} - \left(\frac{1}{n}\right) \mathbf{Y}'\mathbf{J}\mathbf{Y}$	$n - 1$	

## F Test for Regression Relation

To test whether there is a regression relation between the response variable  $Y$  and the set of  $X$  variables  $X_1, \dots, X_{p-1}$ , i.e., to choose between the alternatives:

$$\begin{aligned} H_0: \beta_1 &= \beta_2 = \dots = \beta_{p-1} = 0 \\ H_a: \text{not all } \beta_k &(k = 1, \dots, p-1) \text{ equal zero} \end{aligned} \quad (6.39a)$$

we use the test statistic:

$$F^* = \frac{MSR}{MSE} \quad (6.39b)$$

The decision rule to control the Type I error at  $\alpha$  is:

$$\begin{aligned} \text{If } F^* \leq F(1 - \alpha; p - 1, n - p), \text{ conclude } H_0 \\ \text{If } F^* > F(1 - \alpha; p - 1, n - p), \text{ conclude } H_a \end{aligned} \quad (6.39c)$$

The existence of a regression relation by itself does not, of course, ensure that useful predictions can be made by using it.

Note that when  $p - 1 = 1$ , this test reduces to the  $F$  test in (2.60) for testing in simple linear regression whether or not  $\beta_1 = 0$ .

## Coefficient of Multiple Determination

The coefficient of multiple determination, denoted by  $R^2$ , is defined as follows:

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO} \quad (6.40)$$

It measures the proportionate reduction of total variation in  $Y$  associated with the use of the set of  $X$  variables  $X_1, \dots, X_{p-1}$ . The coefficient of multiple determination  $R^2$  reduces to the coefficient of simple determination in (2.72) for simple linear regression when  $p - 1 = 1$ , i.e., when one  $X$  variable is in regression model (6.19). Just as before, we have:

$$0 \leq R^2 \leq 1 \quad (6.41)$$

where  $R^2$  assumes the value 0 when all  $b_k = 0$  ( $k = 1, \dots, p - 1$ ), and the value 1 when all  $Y$  observations fall directly on the fitted regression surface, i.e., when  $Y_i = \hat{Y}_i$  for all  $i$ .

Adding more  $X$  variables to the regression model can only increase  $R^2$  and never reduce it, because  $SSE$  can never become larger with more  $X$  variables and  $SSTO$  is always the same for a given set of responses. Since  $R^2$  usually can be made larger by including a larger number of predictor variables, it is sometimes suggested that a modified measure be used that adjusts for the number of  $X$  variables in the model. The *adjusted coefficient of multiple determination*, denoted by  $R_a^2$ , adjusts  $R^2$  by dividing each sum of squares by its associated degrees of freedom:

$$R_a^2 = 1 - \frac{\frac{SSE}{n-p}}{\frac{SSTO}{n-1}} = 1 - \left( \frac{n-1}{n-p} \right) \frac{SSE}{SSTO} \quad (6.42)$$

This adjusted coefficient of multiple determination may actually become smaller when another  $X$  variable is introduced into the model, because any decrease in  $SSE$  may be more than offset by the loss of a degree of freedom in the denominator  $n - p$ .

### Comments

1. To distinguish between the coefficients of determination for simple and multiple regression, we shall from now on refer to the former as the coefficient of simple determination.
2. It can be shown that the coefficient of multiple determination  $R^2$  can be viewed as a coefficient of simple determination between the responses  $Y_i$  and the fitted values  $\hat{Y}_i$ .
3. A large value of  $R^2$  does not necessarily imply that the fitted model is a useful one. For instance, observations may have been taken at only a few levels of the predictor variables. Despite a high  $R^2$  in this case, the fitted model may not be useful if most predictions require extrapolations outside the region of observations. Again, even though  $R^2$  is large,  $MSE$  may still be too large for inferences to be useful when high precision is required. ■

### Coefficient of Multiple Correlation

The coefficient of multiple correlation  $R$  is the positive square root of  $R^2$ :

$$R = \sqrt{R^2} \quad (6.43)$$

When there is one  $X$  variable in regression model (6.19), i.e., when  $p - 1 = 1$ , the coefficient of multiple correlation  $R$  equals in absolute value the correlation coefficient  $r$  in (2.73) for simple correlation.

## 6.6 Inferences about Regression Parameters

The least squares and maximum likelihood estimators in  $\mathbf{b}$  are unbiased:

$$E\{\mathbf{b}\} = \boldsymbol{\beta} \quad (6.44)$$

The variance-covariance matrix  $\sigma^2\{\mathbf{b}\}$ :

$$\sigma^2\{\mathbf{b}\}_{p \times p} = \begin{bmatrix} \sigma^2\{b_0\} & \sigma\{b_0, b_1\} & \cdots & \sigma\{b_0, b_{p-1}\} \\ \sigma\{b_1, b_0\} & \sigma^2\{b_1\} & \cdots & \sigma\{b_1, b_{p-1}\} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma\{b_{p-1}, b_0\} & \sigma\{b_{p-1}, b_1\} & \cdots & \sigma^2\{b_{p-1}\} \end{bmatrix} \quad (6.45)$$

is given by:

$$\sigma^2\{\mathbf{b}\} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \quad (6.46)$$

The estimated variance-covariance matrix  $s^2\{\mathbf{b}\}$ :

$$s^2\{\mathbf{b}\}_{p \times p} = \begin{bmatrix} s^2\{b_0\} & s\{b_0, b_1\} & \cdots & s\{b_0, b_{p-1}\} \\ s\{b_1, b_0\} & s^2\{b_1\} & \cdots & s\{b_1, b_{p-1}\} \\ \vdots & \vdots & \ddots & \vdots \\ s\{b_{p-1}, b_0\} & s\{b_{p-1}, b_1\} & \cdots & s^2\{b_{p-1}\} \end{bmatrix} \quad (6.47)$$

is given by:

$$s^2\{\mathbf{b}\} = MSE(\mathbf{X}'\mathbf{X})^{-1} \quad (6.48)$$

From  $s^2\{\mathbf{b}\}$ , one can obtain  $s^2\{b_0\}$ ,  $s^2\{b_1\}$ , or whatever other variance is needed, or any needed covariances.

### Interval Estimation of $\beta_k$

For the normal error regression model (6.19), we have:

$$\frac{b_k - \beta_k}{s\{b_k\}} \sim t(n-p) \quad k = 0, 1, \dots, p-1 \quad (6.49)$$

Hence, the confidence limits for  $\beta_k$  with  $1 - \alpha$  confidence coefficient are:

$$b_k \pm t(1 - \alpha/2; n-p)s\{b_k\} \quad (6.50)$$

### Tests for $\beta_k$

Tests for  $\beta_k$  are set up in the usual fashion. To test:

$$\begin{aligned} H_0: \beta_k &= 0 \\ H_a: \beta_k &\neq 0 \end{aligned} \quad (6.51a)$$

we may use the test statistic:

$$t^* = \frac{b_k}{s\{b_k\}} \quad (6.51b)$$

and the decision rule:

$$\begin{aligned} \text{If } |t^*| &\leq t(1 - \alpha/2; n-p), \text{ conclude } H_0 \\ \text{Otherwise} &\text{ conclude } H_a \end{aligned} \quad (6.51c)$$

The power of the  $t$  test can be obtained as explained in Chapter 2, with the degrees of freedom modified to  $n-p$ .

As with simple linear regression, an  $F$  test can also be conducted to determine whether or not  $\beta_k = 0$  in multiple regression models. We discuss this test in Chapter 7.

### Joint Inferences

The Bonferroni joint confidence intervals can be used to estimate several regression coefficients simultaneously. If  $g$  parameters are to be estimated jointly (where  $g \leq p$ ), the confidence limits with family confidence coefficient  $1 - \alpha$  are:

$$b_k \pm Bs\{b_k\} \quad (6.52)$$

where:

$$B = t(1 - \alpha/2g; n-p) \quad (6.52a)$$

In Chapter 7, we discuss tests concerning subsets of the regression parameters.

## 6.7 Estimation of Mean Response and Prediction of New Observation

---

### Interval Estimation of $E\{Y_h\}$

For given values of  $X_1, \dots, X_{p-1}$ , denoted by  $X_{h1}, \dots, X_{h,p-1}$ , the mean response is denoted by  $E\{Y_h\}$ . We define the vector  $\mathbf{X}_h$ :

$$\mathbf{X}_h = \begin{bmatrix} 1 \\ X_{h1} \\ \vdots \\ X_{h,p-1} \end{bmatrix} \quad (6.53)$$

so that the mean response to be estimated is:

$$E\{Y_h\} = \mathbf{X}'_h \boldsymbol{\beta} \quad (6.54)$$

The estimated mean response corresponding to  $\mathbf{X}_h$ , denoted by  $\hat{Y}_h$ , is:

$$\hat{Y}_h = \mathbf{X}'_h \mathbf{b} \quad (6.55)$$

This estimator is unbiased:

$$E\{\hat{Y}_h\} = \mathbf{X}'_h \boldsymbol{\beta} = E\{Y_h\} \quad (6.56)$$

and its variance is:

$$\sigma^2\{\hat{Y}_h\} = \sigma^2 \mathbf{X}'_h (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_h \quad (6.57)$$

This variance can be expressed as a function of the variance-covariance matrix of the estimated regression coefficients:

$$\sigma^2\{\hat{Y}_h\} = \mathbf{X}'_h \boldsymbol{\sigma}^2\{\mathbf{b}\} \mathbf{X}_h \quad (6.57a)$$

Note from (6.57a) that the variance  $\sigma^2\{\hat{Y}_h\}$  is a function of the variances  $\sigma^2\{b_k\}$  of the regression coefficients and of the covariances  $\sigma\{b_k, b_{k'}\}$  between pairs of regression coefficients, just as in simple linear regression. The estimated variance  $s^2\{\hat{Y}_h\}$  is given by:

$$s^2\{\hat{Y}_h\} = MSE(\mathbf{X}'_h (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_h) = \mathbf{X}'_h \mathbf{s}^2\{\mathbf{b}\} \mathbf{X}_h \quad (6.58)$$

The  $1 - \alpha$  confidence limits for  $E\{Y_h\}$  are:

$$\hat{Y}_h \pm t(1 - \alpha/2; n - p) s\{\hat{Y}_h\} \quad (6.59)$$

### Confidence Region for Regression Surface

The  $1 - \alpha$  confidence region for the entire regression surface is an extension of the Working-Hotelling confidence band (2.40) for the regression line when there is one predictor variable. Boundary points of the confidence region at  $\mathbf{X}_h$  are obtained from:

$$\hat{Y}_h \pm W s\{\hat{Y}_h\} \quad (6.60)$$

where:

$$W^2 = pF(1 - \alpha; p, n - p) \quad (6.60a)$$

The confidence coefficient  $1 - \alpha$  provides assurance that the region contains the entire regression surface over all combinations of values of the  $X$  variables.

### Simultaneous Confidence Intervals for Several Mean Responses

To estimate a number of mean responses  $E\{Y_h\}$  corresponding to different  $\mathbf{X}_h$  vectors with family confidence coefficient  $1 - \alpha$ , we can employ two basic approaches:

1. Use the Working-Hotelling confidence region bounds (6.60) for the several  $\mathbf{X}_h$  vectors of interest:

$$\hat{Y}_h \pm Ws\{\hat{Y}_h\} \quad (6.61)$$

where  $\hat{Y}_h$ ,  $W$ , and  $s\{\hat{Y}_h\}$  are defined in (6.55), (6.60a), and (6.58), respectively. Since the Working-Hotelling confidence region covers the mean responses for all possible  $\mathbf{X}_h$  vectors with confidence coefficient  $1 - \alpha$ , the selected boundary values will cover the mean responses for the  $\mathbf{X}_h$  vectors of interest with family confidence coefficient greater than  $1 - \alpha$ .

2. Use Bonferroni simultaneous confidence intervals. When  $g$  interval estimates are to be made, the Bonferroni confidence limits are:

$$\hat{Y}_h \pm Bs\{\hat{Y}_h\} \quad (6.62)$$

where:

$$B = t(1 - \alpha/2g; n - p) \quad (6.62a)$$

For any particular application, we can compare the  $W$  and  $B$  multiples to see which procedure will lead to narrower confidence intervals. If the  $\mathbf{X}_h$  levels are not specified in advance but are determined as the analysis proceeds, it is better to use the Working-Hotelling limits (6.61) since the family for this procedure includes all possible  $\mathbf{X}_h$  levels.

### Prediction of New Observation $Y_{h(\text{new})}$

The  $1 - \alpha$  prediction limits for a new observation  $Y_{h(\text{new})}$  corresponding to  $\mathbf{X}_h$ , the specified values of the  $X$  variables, are:

$$\hat{Y}_h \pm t(1 - \alpha/2; n - p)s\{\text{pred}\} \quad (6.63)$$

where:

$$s^2\{\text{pred}\} = MSE + s^2\{\hat{Y}_h\} = MSE(1 + \mathbf{X}'_h(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h) \quad (6.63a)$$

and  $s^2\{\hat{Y}_h\}$  is given by (6.58).

### Prediction of Mean of $m$ New Observations at $\mathbf{X}_h$

When  $m$  new observations are to be selected at the same levels  $\mathbf{X}_h$  and their mean  $\bar{Y}_{h(\text{new})}$  is to be predicted, the  $1 - \alpha$  prediction limits are:

$$\hat{Y}_h \pm t(1 - \alpha/2; n - p)s\{\text{predmean}\} \quad (6.64)$$

where:

$$s^2\{\text{predmean}\} = \frac{MSE}{m} + s^2\{\hat{Y}_h\} = MSE \left( \frac{1}{m} + \mathbf{X}'_h (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_h \right) \quad (6.64a)$$

### Predictions of $g$ New Observations

Simultaneous Scheffé prediction limits for  $g$  new observations at  $g$  different levels  $\mathbf{X}_h$  with family confidence coefficient  $1 - \alpha$  are given by:

$$\hat{Y}_h \pm Ss\{\text{pred}\} \quad (6.65)$$

where:

$$S^2 = gF(1 - \alpha; g, n - p) \quad (6.65a)$$

and  $s^2\{\text{pred}\}$  is given by (6.63a).

Alternatively, Bonferroni simultaneous prediction limits can be used. For  $g$  predictions with family confidence coefficient  $1 - \alpha$ , they are:

$$\hat{Y}_h \pm Bs\{\text{pred}\} \quad (6.66)$$

where:

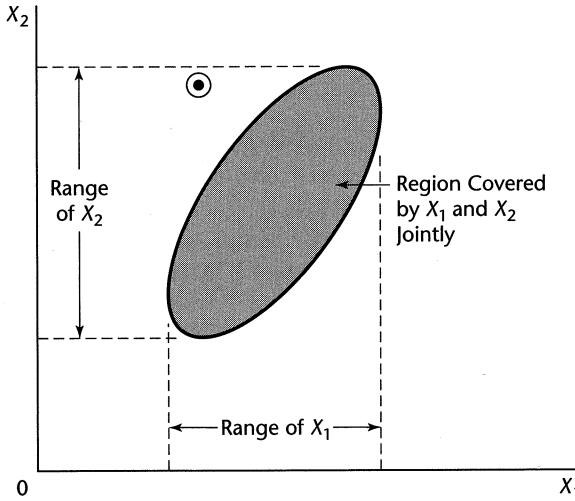
$$B = t(1 - \alpha/2g; n - p) \quad (6.66a)$$

A comparison of  $S$  and  $B$  in advance of any particular use will indicate which procedure will lead to narrower prediction intervals.

### Caution about Hidden Extrapolations

When estimating a mean response or predicting a new observation in multiple regression, one needs to be particularly careful that the estimate or prediction does not fall outside the scope of the model. The danger, of course, is that the model may not be appropriate when it is extended outside the region of the observations. In multiple regression, it is particularly easy to lose track of this region since the levels of  $X_1, \dots, X_{p-1}$  jointly define the region. Thus, one cannot merely look at the ranges of each predictor variable. Consider Figure 6.3,

**FIGURE 6.3**  
Region of  
Observations  
on  $X_1$  and  $X_2$   
Jointly,  
Compared with  
Ranges of  $X_1$   
and  $X_2$   
Individually.



where the shaded region is the region of observations for a multiple regression application with two predictor variables and the circled dot represents the values  $(X_{h1}, X_{h2})$  for which a prediction is to be made. The circled dot is within the ranges of the predictor variables  $X_1$  and  $X_2$  individually, yet is well outside the joint region of observations. It is easy to spot this extrapolation when there are only two predictor variables, but it becomes much more difficult when the number of predictor variables is large. We discuss in Chapter 10 a procedure for identifying hidden extrapolations when there are more than two predictor variables.

## 6.8 Diagnostics and Remedial Measures

Diagnostics play an important role in the development and evaluation of multiple regression models. Most of the diagnostic procedures for simple linear regression that we described in Chapter 3 carry over directly to multiple regression. We review these diagnostic procedures now, as well as the remedial measures for simple linear regression that carry over directly to multiple regression.

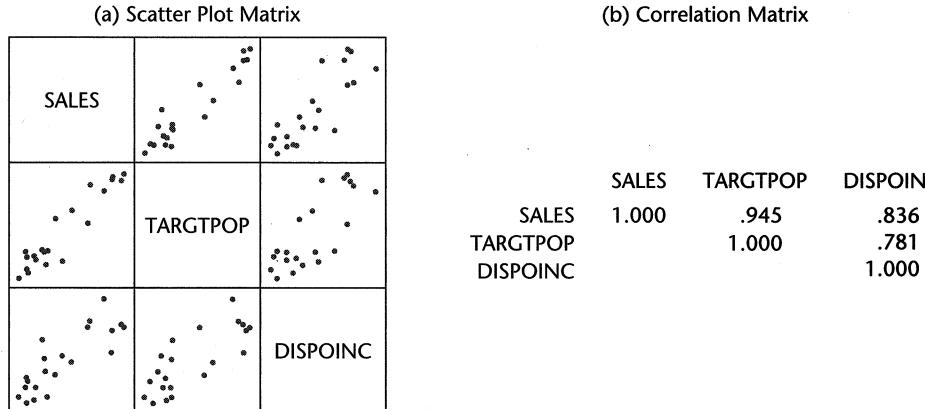
Many specialized diagnostics and remedial procedures for multiple regression have also been developed. Some important ones will be discussed in Chapters 10 and 11.

### Scatter Plot Matrix

Box plots, sequence plots, stem-and-leaf plots, and dot plots for each of the predictor variables and for the response variable can provide helpful, preliminary univariate information about these variables. Scatter plots of the response variable against each predictor variable can aid in determining the nature and strength of the bivariate relationships between each of the predictor variables and the response variable and in identifying gaps in the data points as well as outlying data points. Scatter plots of each predictor variable against each of the other predictor variables are helpful for studying the bivariate relationships among the predictor variables and for finding gaps and detecting outliers.

Analysis is facilitated if these scatter plots are assembled in a *scatter plot matrix*, such as in Figure 6.4. In this figure, the  $Y$  variable for any one scatter plot is the name found in

**FIGURE 6.4**  
SYGRAPH  
Scatter Plot  
Matrix and  
Correlation  
Matrix—  
Dwaine Studios  
Example.



its row, and the  $X$  variable is the name found in its column. Thus, the scatter plot matrix in Figure 6.4 shows in the first row the plots of  $Y$  (SALES) against  $X_1$  (TARGETPOP) and  $X_2$  (DISPOINC), of  $X_1$  against  $Y$  and  $X_2$  in the second row, and of  $X_2$  against  $Y$  and  $X_1$  in the third row. These variables are described on page 236. Alternatively, by viewing the first column, one can compare the plots of  $X_1$  and  $X_2$  each against  $Y$ , and similarly for the other two columns. A scatter plot matrix facilitates the study of the relationships among the variables by comparing the scatter plots within a row or a column. Examples in this and subsequent chapters will illustrate the usefulness of scatter plot matrices.

A complement to the scatter plot matrix that may be useful at times is the correlation matrix. This matrix contains the coefficients of simple correlation  $r_{Y1}, r_{Y2}, \dots, r_{Y,p-1}$  between  $Y$  and each of the predictor variables, as well as all of the coefficients of simple correlation among the predictor variables— $r_{12}$  between  $X_1$  and  $X_2$ ,  $r_{13}$  between  $X_1$  and  $X_3$ , etc. The format of the correlation matrix follows that of the scatter plot matrix:

$$\begin{bmatrix} 1 & r_{Y1} & r_{Y2} & \cdots & r_{Y,p-1} \\ r_{Y1} & 1 & r_{12} & \cdots & r_{1,p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{Y,p-1} & r_{1,p-1} & r_{2,p-1} & \cdots & 1 \end{bmatrix} \quad (6.67)$$

Note that the correlation matrix is symmetric and that its main diagonal contains 1s because the coefficient of correlation between a variable and itself is 1. Many statistics packages provide the correlation matrix as an option. Since this matrix is symmetric, the lower (or upper) triangular block of elements is frequently omitted in the output.

Some interactive statistics packages enable the user to employ *brushing* with scatter plot matrices. When a point in a scatter plot is brushed, it is given a distinctive appearance on the computer screen in each scatter plot in the matrix. The case corresponding to the brushed point may also be identified. Brushing is helpful to see whether a case that is outlying in one scatter plot is also outlying in some or all of the other plots. Brushing may also be applied to a group of points to see, for instance, whether a group of cases that does not fit the relationship for the remaining cases in one scatter plot also follows a distinct pattern in any of the other scatter plots.

### Three-Dimensional Scatter Plots

Some interactive statistics packages provide *three-dimensional scatter plots* or *point clouds*, and permit spinning of these plots to enable the viewer to see the point cloud from different perspectives. This can be very helpful for identifying patterns that are only apparent from certain perspectives. Figure 6.6 on page 238 illustrates a three-dimensional scatter plot and the use of spinning.

### Residual Plots

A plot of the residuals against the fitted values is useful for assessing the appropriateness of the multiple regression function and the constancy of the variance of the error terms, as well as for providing information about outliers, just as for simple linear regression. Similarly,

a plot of the residuals against time or against some other sequence can provide diagnostic information about possible correlations between the error terms in multiple regression. Box plots and normal probability plots of the residuals are useful for examining whether the error terms are reasonably normally distributed.

In addition, residuals should be plotted against each of the predictor variables. Each of these plots can provide further information about the adequacy of the regression function with respect to that predictor variable (e.g., whether a curvature effect is required for that variable) and about possible variation in the magnitude of the error variance in relation to that predictor variable.

Residuals should also be plotted against important predictor variables that were omitted from the model, to see if the omitted variables have substantial additional effects on the response variable that have not yet been recognized in the regression model. Also, residuals should be plotted against interaction terms for potential interaction effects not included in the regression model, such as against  $X_1 X_2$ ,  $X_1 X_3$ , and  $X_2 X_3$ , to see whether some or all of these interaction terms are required in the model.

A plot of the absolute residuals or the squared residuals against the fitted values is useful for examining the constancy of the variance of the error terms. If nonconstancy is detected, a plot of the absolute residuals or the squared residuals against each of the predictor variables may identify one or several of the predictor variables to which the magnitude of the error variability is related.

## **Correlation Test for Normality**

The correlation test for normality described in Chapter 3 carries forward directly to multiple regression. The expected values of the ordered residuals under normality are calculated according to (3.6), and the coefficient of correlation between the residuals and the expected values under normality is then obtained. Table B.6 is employed to assess whether or not the magnitude of the correlation coefficient supports the reasonableness of the normality assumption.

## **Brown-Forsythe Test for Constancy of Error Variance**

The Brown-Forsythe test statistic (3.9) for assessing the constancy of the error variance can be used readily in multiple regression when the error variance increases or decreases with one of the predictor variables. To conduct the Brown-Forsythe test, we divide the data set into two groups, as for simple linear regression, where one group consists of cases where the level of the predictor variable is relatively low and the other group consists of cases where the level of the predictor variable is relatively high. The Brown-Forsythe test then proceeds as for simple linear regression.

## **Breusch-Pagan Test for Constancy of Error Variance**

The Breusch-Pagan test (3.11) for constancy of the error variance in multiple regression is carried out exactly the same as for simple linear regression when the error variance increases or decreases with one of the predictor variables. The squared residuals are simply regressed against the predictor variable to obtain the regression sum of squares  $SSR^*$ , and the test proceeds as before, using the error sum of squares  $SSE$  for the full multiple regression model.

When the error variance is a function of more than one predictor variable, a multiple regression of the squared residuals against these predictor variables is conducted and the regression sum of squares  $SSR^*$  is obtained. The test statistic again uses  $SSE$  for the full multiple regression model, but now the chi-square distribution involves  $q$  degrees of freedom, where  $q$  is the number of predictor variables against which the squared residuals are regressed.

## F Test for Lack of Fit

The lack of fit  $F$  test described in Chapter 3 for simple linear regression can be carried over to test whether the multiple regression response function:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \cdots + \beta_{p-1} X_{p-1}$$

is an appropriate response surface. Repeat observations in multiple regression are replicate observations on  $Y$  corresponding to levels of each of the  $X$  variables that are constant from trial to trial. Thus, with two predictor variables, repeat observations require that  $X_1$  and  $X_2$  each remain at given levels from trial to trial.

Once the ANOVA table, shown in Table 6.1, has been obtained,  $SSE$  is decomposed into pure error and lack of fit components. The pure error sum of squares  $SSPE$  is obtained by first calculating for each replicate group the sum of squared deviations of the  $Y$  observations around the group mean, where a replicate group has the same values for each of the  $X$  variables. Let  $c$  denote the number of groups with distinct sets of levels for the  $X$  variables, and let the mean of the  $Y$  observations for the  $j$ th group be denoted by  $\bar{Y}_j$ . Then the sum of squares for the  $j$ th group is given by (3.17), and the pure error sum of squares is the sum of these sums of squares, as given by (3.16). The lack of fit sum of squares  $SSLF$  equals the difference  $SSE - SSPE$ , as indicated by (3.24).

The number of degrees of freedom associated with  $SSPE$  is  $n - c$ , and the number of degrees of freedom associated with  $SSLF$  is  $(n - p) - (n - c) = c - p$ . Thus, for testing the alternatives:

$$\begin{aligned} H_0: E\{Y\} &= \beta_0 + \beta_1 X_1 + \cdots + \beta_{p-1} X_{p-1} \\ H_a: E\{Y\} &\neq \beta_0 + \beta_1 X_1 + \cdots + \beta_{p-1} X_{p-1} \end{aligned} \quad (6.68a)$$

the appropriate test statistic is:

$$F^* = \frac{SSLF}{c - p} \div \frac{SSPE}{n - c} = \frac{MSLF}{MSPE} \quad (6.68b)$$

where  $SSLF$  and  $SSPE$  are given by (3.24) and (3.16), respectively, and the appropriate decision rule is:

$$\begin{aligned} \text{If } F^* &\leq F(1 - \alpha; c - p, n - c), \text{ conclude } H_0 \\ \text{If } F^* &> F(1 - \alpha; c - p, n - c), \text{ conclude } H_a \end{aligned} \quad (6.68c)$$

### Comment

When replicate observations are not available, an approximate lack of fit test can be conducted if there are cases that have similar  $\mathbf{X}_h$  vectors. These cases are grouped together and treated as pseudoreplicates, and the test for lack of fit is then carried out using these groupings of similar cases. ■