7 Multiple Regression: Estimation

7.1 INTRODUCTION

In *multiple regression*, we attempt to predict a *dependent* or *response* variable y on the basis of an assumed linear relationship with several *independent* or *predictor* variables $x_1, x_1, \ldots, x_k$. In addition to constructing a model for prediction, we may wish to assess the extent of the relationship between y and the x variables. For this purpose, we use the multiple correlation coefficient R (Section 7.7).

In this chapter, y is a continuous random variable and the x variables are fixed constants (either discrete or continuous) that are controlled by the experimenter. The case in which the x variables are random variables is covered in Chapter 10. In analysis-of-variance (Chapters 12–15), the x variables are fixed and discrete.

Useful applied expositions of multiple regression for the fixed-x case can be found in Morrison (1983), Myers (1990), Montgomery and Peck (1992), Graybill and Iyer (1994), Mendenhall and Sincich (1996), Ryan (1997), Draper and Smith (1998), and Kutner et al. (2005). Theoretical treatments are given by Searle (1971), Graybill (1976), Guttman (1982), Kshirsagar (1983), Myers and Milton (1991), Jørgensen (1993), Wang and Chow (1994), Christensen (1996), Seber and Lee (2003), and Hocking (1976, 1985, 2003).

7.2 THE MODEL

The multiple linear regression model, as introduced in Section 1.2, can be expressed as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon. \tag{7.1}$$

We discuss estimation of the β parameters when the model is linear in the β's. An example of a model that is linear in the β's but not the x's is the second-order

---

Linear Models in Statistics, *Second Edition*, by Alvin C. Rencher and G. Bruce Schaalje

response surface model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \varepsilon. \tag{7.2}$$

To estimate the β's in (7.1), we will use a sample of n observations on y and the associated x variables. The model for the ith observation is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, 2, \ldots, n. \tag{7.3}$$

The assumptions for ε_i or y_i are essentially the same as those for simple linear
가정. 단순선형 회귀 때와 동일.
regression in Section 6.1:

1. $E(\varepsilon_i) = 0$ for $i = 1, 2, \ldots, n$, or, equivalently, $E(y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$.
2. $\text{var}(\varepsilon_i) = \sigma^2$ for $i = 1, 2, \ldots, n$, or, equivalently, $\text{var}(y_i) = \sigma^2$.
3. $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ for all $i \neq j$, or, equivalently, $\text{cov}(y_i, y_j) = 0$.

Assumption 1 states that the model is correct, in other words that all relevant x's are included and the model is indeed linear. Assumption 2 asserts that the variance of y is constant and therefore does not depend on the x's. Assumption 3 states that the y's are uncorrelated with each other, which usually holds in a random sample (the observations would typically be correlated in a time series or when repeated measurements are made on a single plant or animal). Later we will add a normality assumption (Section 7.6), under which the y variable will be independent as well as uncorrelated.

하나라도 만족하지 못한다면, 그 estimator는 poor.

When all three assumptions hold, the least-squares estimators of the β's have some good properties (Section 7.3.2). If one or more assumptions do not hold, the estimators may be poor. Under the normality assumption (Section 7.6), the maximum likelihood estimators have excellent properties.

Any of the three assumptions may fail to hold with real data. Several procedures have been devised for checking the assumptions. These diagnostic techniques are discussed in Chapter 9. → 가정을 확인하는 방법을 다룸

Writing (7.3) for each of the n observations, we have

$$\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_k x_{1k} + \varepsilon_1 \\
y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_k x_{2k} + \varepsilon_2 \\
&\ \vdots \\
y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_k x_{nk} + \varepsilon_n.
\end{aligned}$$

These n equations can be written in matrix form as

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

or

$$\underset{n\times1}{\mathbf{y}} = \underset{n\times(k+1)}{\mathbf{X}}\overset{(k+1)\times1}{\boldsymbol{\beta}} + \underset{n\times1}{\boldsymbol{\varepsilon}}. \tag{7.4}$$

The preceding three assumptions on ε_i or y_i can be expressed in terms of the model in (7.4):

1. $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ or $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$.
2. $\text{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$ or $\text{cov}(\mathbf{y}) = \sigma^2\mathbf{I}$. (가정 2,3)

↳ 대각성분만 σ^2, 나머지는 0.

Note that the assumption $\text{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$ includes both the previous assumptions $\text{var}(\varepsilon_i) = \sigma^2$ and $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$.

The matrix $\mathbf{X}$ in (7.4) is $n \times (k+1)$. In this chapter we assume that $n > k+1$ and rank $(\mathbf{X}) = k+1$. If $n < k+1$ or if there is a linear relationship among the x's, for example, $x_5 = \sum_{j=1}^{4} x_j/4$, then $\mathbf{X}$ will not have full column rank. If the values of the x_{ij}'s are planned (chosen by the researcher), then the $\mathbf{X}$ matrix essentially contains the experimental design and is sometimes called the *design matrix*.

The β parameters in (7.1) or (7.4) are called *regression coefficients*. To emphasize their collective effect, they are sometimes referred to as *partial regression coefficients*. The word *partial* carries both a mathematical and a statistical meaning. Mathematically, the partial derivative of $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$ with respect to x_1, for example, is β_1. Thus β_1 indicates the change in $E(y)$ with a unit increase in x_1 when $x_2, x_3, \ldots, x_k$ are held constant. Statistically, β_1 shows the effect of x_1 on $E(y)$ in the presence of the other x's. This effect would typically be different from the effect of x_1 on $E(y)$ if the other x's were not present in the model. Thus, for example, β_0 and β_1 in

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

will usually be different from β_0^* and β_1^* in

$$y = \beta_0^* + \beta_1^* x_1 + \varepsilon^*.$$

위 두 식의 계수들이 같아짐. (뒤에 나옴)

[If x_1 and x_2 are orthogonal, that is, if $\mathbf{x}_1'\mathbf{x}_2 = 0$ or if $(\mathbf{x}_1 - \bar{x}_1\mathbf{j})'(\mathbf{x}_2 - \bar{x}_2\mathbf{j}) = 0$, where $\mathbf{x}_1$ and $\mathbf{x}_2$ are columns in the $\mathbf{X}$ matrix, then $\beta_0 = \beta_0^*$ and $\beta_1 = \beta_1^*$; see Corollary 1 to Theorem 7.9a and Theorem 7.10]. The change in parameters when an x is deleted from the model is illustrated (with estimates) in the following example.

TABLE 7.1 Data for Example 7.2

Observation Number	y	x_1	x_2
1	2	0	2
2	3	2	6
3	2	2	7
4	7	2	5
5	6	4	9
6	8	4	8
7	10	4	7
8	7	6	10
9	8	6	11
10	12	6	9
11	11	8	15
12	14	8	13

부자연스러운.

Example 7.2. [See Freund and Minton (1979, pp. 36–39)]. Consider the (contrived) data in Table 7.1. $\hat{\beta}_0, \hat{\beta}_1$

Using (6.5) and (6.6) from Section 6.2 and (7.6) in Section 7.3 (see Example 7.3.1), we obtain prediction equations for y regressed on x_1 alone, on x_2 alone, and on both x_1 and x_2:

$$
\begin{aligned}
① \quad \hat{y} &= 1.86 + 1.30x_1, \\
② \quad \hat{y} &= .86 + .78x_2, \\
③ \quad \hat{y} &= 5.37 + 3.01x_1 - 1.29x_2.
\end{aligned}
$$

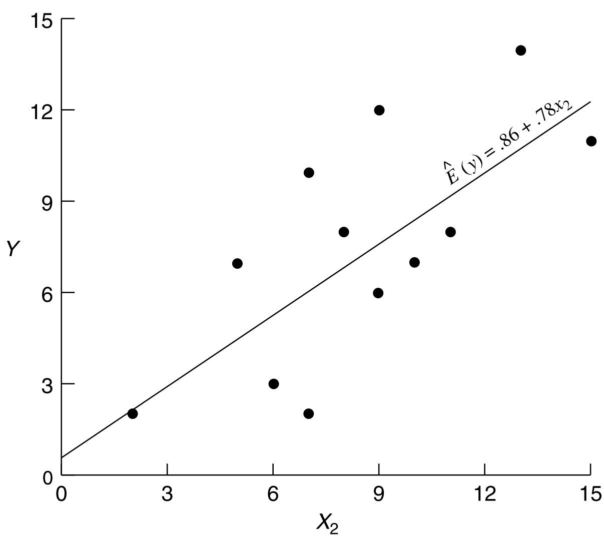

Figure 7.1 Regression of y on x_2 ignoring x_1. ②

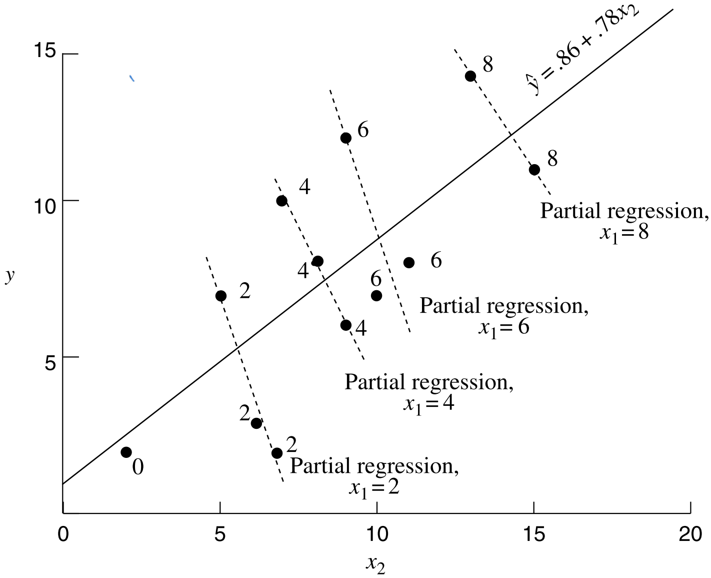

Figure 7.2 Regression of y on x_2 showing the value of x_1 at each point and partial regressions of y on x_2.

As expected, the coefficients change from either of the reduced models to the full model. Note the sign change as the coefficient of x_2 changes from .78 to -1.29.

The values of y and x_2 are plotted in Figure 7.1 along with the prediction equation $\hat{y} = .86 + .78x_2$. The linear trend is clearly evident.

In Figure 7.2 we have the same plot as in Figure 7.1, except that each point is labeled with the value of x_1. Examining values of y and x_2 for a fixed value of x_1 (2, 4, 6, or 8) shows a negative slope for the relationship. These negative relationships are shown as partial regressions of y on x_2 for each value of x_1. The partial regression coefficient $\hat{\beta}_2 = -1.29$ reflects the negative slopes of these four partial regressions.

Further insight into the meaning of the partial regression coefficients is given in Section 7.10. ∎