1.3 LINEAR REGRESSION MODELS

If we denote the response variable by Y and the explanatory variables by $X_1, X_2, \ldots, X_K$, then a general model relating these variables is

$$E[Y|X_1 = x_1, X_2 = x_2, \ldots, X_K = x_K] = \phi(x_1, x_2, \ldots, x_K),$$

although, for brevity, we will usually drop the conditioning part and write $E[Y]$. In this book we direct our attention to the important class of linear models, that is,

$$\phi(x_1, x_2, \ldots, x_K) = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K,$$

which is linear in the parameters β_j. This restriction to linearity is not as restrictive as one might think. For example, many functions of several variables are approximately linear over sufficiently small regions, or they may be made linear by a suitable transformation. Using logarithms for the gravitational model, we get the straight line

$$\log F = \log \alpha - \beta \log d. \tag{1.1}$$

For the linear model, the x_i could be functions of other variables z, w, etc.; for example, $x_1 = \sin z$, $x_2 = \log w$, and $x_3 = zw$. We can also have $x_i = x^i$, which leads to a polynomial model; the linearity refers to the parameters, not the variables. Note that "categorical" models can be included under our umbrella by using *dummy (indicator)* x-variables. For example, suppose that we wish to compare the means of two populations, say, $\mu_i = E[U_i]$ $(i = 1, 2)$. Then we can combine the data into the single model

$$\begin{aligned} E[Y] &= \mu_1 + (\mu_2 - \mu_1)x \\ &= \beta_0 + \beta_1 x, \end{aligned}$$

where $x = 0$ when Y is a U_1 observation and $x = 1$ when Y is a U_2 observation. Here $\mu_1 = \beta_0$ and $\mu_2 = \beta_0 + \beta_1$, the difference being β_1. We can extend this idea to the case of comparing m means using $m - 1$ dummy variables.

In a similar fashion we can combine two straight lines,

$$U_j = \alpha_j + \gamma_j x_1 \qquad (j = 1, 2),$$

using a dummy x_2 variable which takes the value 0 if the observation is from the first line, and 1 otherwise. The combined model is

$$\begin{aligned} E[Y] &= \alpha_1 + \gamma_1 x_1 + (\alpha_2 - \alpha_1)x_2 + (\gamma_2 - \gamma_1)x_1 x_2 \\ &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3, \end{aligned} \tag{1.2}$$

say, where $x_3 = x_1 x_2$. Here $\alpha_1 = \beta_0$, $\alpha_2 = \beta_0 + \beta_2$, $\gamma_1 = \beta_1$, and $\gamma_2 = \beta_1 + \beta_3$.

In the various models considered above, the explanatory variables may or may not be random. For example, dummy variables are nonrandom. With random X-variables, we carry out the regression conditionally on their observed values, provided that they are measured exactly (or at least with sufficient accuracy). We effectively proceed as though the X-variables were not random at all. When measurement errors cannot be ignored, the theory has to be modified, as we shall see in Chapter 9.

1.4 EXPECTATION AND COVARIANCE OPERATORS

In this book we focus on vectors and matrices, so we first need to generalize the ideas of expectation, covariance, and variance, which we do in this section.

Let Z_{ij} $(i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n)$ be a set of random variables with expected values $E[Z_{ij}]$. Expressing both the random variables and their expectations in matrix form, we can define the general expectation operator of the matrix $\mathbf{Z} = (Z_{ij})$ as follows:

***Definition* 1.1**

$$E[\mathbf{Z}] = (E[Z_{ij}]).$$

THEOREM 1.1 *If* $\mathbf{A} = (a_{ij})$, $\mathbf{B} = (b_{ij})$, *and* $\mathbf{C} = (c_{ij})$ *are* $l \times m$, $n \times p$, *and* $l \times p$ *matrices, respectively, of constants, then*

$$E[\mathbf{AZB} + \mathbf{C}] = \mathbf{A}E[\mathbf{Z}]\mathbf{B} + \mathbf{C}.$$

Proof. Let $\mathbf{W} = \mathbf{AZB} + \mathbf{C}$; then $W_{ij} = \sum_{r=1}^{m} \sum_{s=1}^{n} a_{ir} Z_{rs} b_{sj} + c_{ij}$ and

$$\begin{aligned} E[\mathbf{AZB} + \mathbf{C}] = (E[W_{ij}]) &= \left(\sum_r \sum_s a_{ir} E[Z_{rs}] b_{sj} + c_{ij} \right) \\ &= ((\mathbf{A}E[\mathbf{Z}]\mathbf{B})_{ij}) + (c_{ij}) \\ &= \mathbf{A}E[\mathbf{Z}]\mathbf{B} + \mathbf{C}. \end{aligned}$$ □

In this proof we note that l, m, n, and p are any positive integers, and the matrices of constants can take any values. For example, if $\mathbf{X}$ is an $m \times 1$ vector, then $E[\mathbf{AX}] = \mathbf{A}E[\mathbf{X}]$. Using similar algebra, we can prove that if $\mathbf{A}$ and $\mathbf{B}$ are $m \times n$ matrices of constants, and $\mathbf{X}$ and $\mathbf{Y}$ are $n \times 1$ vectors of random variables, then

$$E[\mathbf{AX} + \mathbf{BY}] = \mathbf{A}E[\mathbf{X}] + \mathbf{B}E[\mathbf{Y}].$$

In a similar manner we can generalize the notions of covariance and variance for vectors. If $\mathbf{X}$ and $\mathbf{Y}$ are $m \times 1$ and $n \times 1$ vectors of random variables, then we define the generalized covariance operator Cov as follows:

***Definition* 1.2**

$$\mathrm{Cov}[\mathbf{X},\mathbf{Y}] = (\,\mathrm{cov}[X_i, Y_j])\,.$$

THEOREM 1.2 *If* $E[\mathbf{X}] = \boldsymbol{\alpha}$ *and* $E[\mathbf{Y}] = \boldsymbol{\beta}$, *then*

$$\mathrm{Cov}[\mathbf{X},\mathbf{Y}] = E\left[(\mathbf{X}-\boldsymbol{\alpha})(\mathbf{Y}-\boldsymbol{\beta})'\right].$$

Proof.

$$\begin{aligned}\mathrm{Cov}[\mathbf{X},\mathbf{Y}] &= (\,\mathrm{cov}[X_i, Y_j])\\ &= \{E[(X_i-\alpha_i)(Y_j-\beta_j)]\}\\ &= E\,\{[(X_i-\alpha_i)(Y_j-\beta_j)]\}\\ &= E\left[(\mathbf{X}-\boldsymbol{\alpha})(\mathbf{Y}-\boldsymbol{\beta})'\right].\end{aligned}$$

□

***Definition* 1.3** *When* $\mathbf{Y} = \mathbf{X}$, $\mathrm{Cov}[\mathbf{X},\mathbf{X}]$, *written as* $\mathrm{Var}[\mathbf{X}]$, *is called the variance (variance–covariance or dispersion) matrix of* $\mathbf{X}$. *Thus*

$$\begin{aligned}\mathrm{Var}[\mathbf{X}] &= (\,\mathrm{cov}[X_i, X_j])\\ &= \begin{pmatrix}\mathrm{var}[X_1] & \mathrm{cov}[X_1,X_2] & \cdots & \mathrm{cov}[X_1,X_n]\\ \mathrm{cov}[X_2,X_1] & \mathrm{var}[X_2] & \cdots & \mathrm{cov}[X_2,X_n]\\ \cdots & \cdots & \cdots & \cdots\\ \mathrm{cov}[X_n,X_1] & \mathrm{cov}[X_n,X_2] & \cdots & \mathrm{var}[X_n]\end{pmatrix}.\end{aligned} \tag{1.3}$$

Since $\mathrm{cov}[X_i, X_j] = \mathrm{cov}[X_j, X_i]$, *the matrix above is symmetric. We note that when* $\mathbf{X} = X_1$ *we write* $\mathrm{Var}[\mathbf{X}] = \mathrm{var}[X_1]$.

From Theorem 1.2 with $\mathbf{Y} = \mathbf{X}$ we have

$$\mathrm{Var}[\mathbf{X}] = E\left[(\mathbf{X}-\boldsymbol{\alpha})(\mathbf{X}-\boldsymbol{\alpha})'\right], \tag{1.4}$$

which, on expanding, leads to

$$\mathrm{Var}[\mathbf{X}] = E[\mathbf{X}\mathbf{X}'] - \boldsymbol{\alpha}\boldsymbol{\alpha}'. \tag{1.5}$$

These last two equations are natural generalizations of univariate results.

EXAMPLE 1.4 If $\mathbf{a}$ is any $n \times 1$ vector of constants, then

$$\mathrm{Var}[\mathbf{X}-\mathbf{a}] = \mathrm{Var}[\mathbf{X}].$$

This follows from the fact that $X_i - a_i - E[X_i - a_i] = X_i - E[X_i]$, so that

$$\mathrm{cov}[X_i - a_i, X_j - a_j] = \mathrm{cov}[X_i, X_j].$$

□

THEOREM 1.3 *If* $\mathbf{X}$ *and* $\mathbf{Y}$ *are* $m \times 1$ *and* $n \times 1$ *vectors of random variables, and* $\mathbf{A}$ *and* $\mathbf{B}$ *are* $l \times m$ *and* $p \times n$ *matrices of constants, respectively, then*

$$\mathrm{Cov}[\mathbf{AX},\mathbf{BY}] = \mathbf{A}\,\mathrm{Cov}[\mathbf{X},\mathbf{Y}]\mathbf{B}'. \tag{1.6}$$

Proof. Let $\mathbf{U} = \mathbf{AX}$ and $\mathbf{V} = \mathbf{BY}$. Then, by Theorems 1.2 and 1.1,

$$\begin{aligned}\mathrm{Cov}[\mathbf{AX},\mathbf{BY}] &= \mathrm{Cov}[\mathbf{U},\mathbf{V}]\\ &= E\left[(\mathbf{U}-E[\mathbf{U}])(\mathbf{V}-E[\mathbf{V}])'\right]\\ &= E\left[(\mathbf{AX}-\mathbf{A}\boldsymbol{\alpha})(\mathbf{BY}-\mathbf{B}\boldsymbol{\beta})'\right]\\ &= E\left[\mathbf{A}(\mathbf{X}-\boldsymbol{\alpha})(\mathbf{Y}-\boldsymbol{\beta})'\mathbf{B}'\right]\\ &= \mathbf{A}E\left[(\mathbf{X}-\boldsymbol{\alpha})(\mathbf{Y}-\boldsymbol{\beta})'\right]\mathbf{B}'\\ &= \mathbf{A}\,\mathrm{Cov}[\mathbf{X},\mathbf{Y}]\mathbf{B}'.\end{aligned}$$

□

From the theorem above we have the special cases

$$\mathrm{Cov}[\mathbf{AX},\mathbf{Y}] = \mathbf{A}\,\mathrm{Cov}[\mathbf{X},\mathbf{Y}] \quad \text{and} \quad \mathrm{Cov}[\mathbf{X},\mathbf{BY}] = \mathrm{Cov}[\mathbf{X},\mathbf{Y}]\mathbf{B}'.$$

Of particular importance is the following result, obtained by setting $\mathbf{B} = \mathbf{A}$ and $\mathbf{Y} = \mathbf{X}$:

$$\mathrm{Var}[\mathbf{AX}] = \mathrm{Cov}[\mathbf{AX},\mathbf{AX}] = \mathbf{A}\,\mathrm{Cov}[\mathbf{X},\mathbf{X}]\mathbf{A}' = \mathbf{A}\,\mathrm{Var}[\mathbf{X}]\mathbf{A}'. \tag{1.7}$$

EXAMPLE 1.5 If $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{U}$, and $\mathbf{V}$ are any (not necessarily distinct) $n \times 1$ vectors of random variables, then for all real numbers a, b, c, and d (including zero),

$$\begin{aligned}&\mathrm{Cov}[a\mathbf{X}+b\mathbf{Y}, c\mathbf{U}+d\mathbf{V}]\\ &\quad= ac\,\mathrm{Cov}[\mathbf{X},\mathbf{U}] + ad\,\mathrm{Cov}[\mathbf{X},\mathbf{V}] + bc\,\mathrm{Cov}[\mathbf{Y},\mathbf{U}] + bd\,\mathrm{Cov}[\mathbf{Y},\mathbf{V}].\end{aligned} \tag{1.8}$$

To prove this result, we simply multiply out

$$\begin{aligned}&E\left[(a\mathbf{X}+b\mathbf{Y}-aE[\mathbf{X}]-bE[\mathbf{Y}])(c\mathbf{U}+d\mathbf{V}-cE[\mathbf{U}]-dE[\mathbf{V}])'\right]\\ &\quad= E\left[(a(\mathbf{X}-E[\mathbf{X}])+b(\mathbf{Y}-E[\mathbf{Y}]))\,(c(\mathbf{U}-E[\mathbf{U}])+d(\mathbf{V}-E[\mathbf{V}]))'\right].\end{aligned}$$

If we set $\mathbf{U} = \mathbf{X}$ and $\mathbf{V} = \mathbf{Y}$, $c = a$ and $d = b$, we get

$$\begin{aligned}\mathrm{Var}[a\mathbf{X}+b\mathbf{Y}] &= \mathrm{Cov}[a\mathbf{X}+b\mathbf{Y}, a\mathbf{X}+b\mathbf{Y}]\\ &= a^2\,\mathrm{Var}[\mathbf{X}] + ab(\mathrm{Cov}[\mathbf{X},\mathbf{Y}] + \mathrm{Cov}[\mathbf{Y},\mathbf{X}])\\ &\quad + b^2\,\mathrm{Var}[\mathbf{Y}].\end{aligned} \tag{1.9}$$

□

In Chapter 2 we make frequent use of the following theorem.

THEOREM 1.4 *If $\mathbf{X}$ is a vector of random variables such that no element of $\mathbf{X}$ is a linear combination of the remaining elements [i.e., there do not exist $\mathbf{a}\,(\neq \mathbf{0})$ and b such that $\mathbf{a}'\mathbf{X} = b$ for all values of $\mathbf{X} = \mathbf{x}$], then* $\text{Var}[\mathbf{X}]$ *is a positive-definite matrix (see A.4).*

Proof. For any vector $\mathbf{c}$, we have

$$\begin{aligned} 0 &\leq \text{var}[\mathbf{c}'\mathbf{X}] \\ &= \mathbf{c}'\,\text{Var}[\mathbf{X}]\mathbf{c} \qquad \text{[by equation (1.7)]}. \end{aligned}$$

Now equality holds if and only if $\mathbf{c}'\mathbf{X}$ is a constant, that is, if and only if $\mathbf{c}'\mathbf{X} = d$ ($\mathbf{c} \neq \mathbf{0}$) or $\mathbf{c} = \mathbf{0}$. Because the former possibility is ruled out, $\mathbf{c} = \mathbf{0}$ and $\text{Var}[\mathbf{X}]$ is positive-definite. □

EXAMPLE 1.6 If $\mathbf{X}$ and $\mathbf{Y}$ are $m \times 1$ and $n \times 1$ vectors of random variables such that no element of $\mathbf{X}$ is a linear combination of the remaining elements, then there exists an $n \times m$ matrix $\mathbf{M}$ such that $\text{Cov}[\mathbf{X}, \mathbf{Y} - \mathbf{M}\mathbf{X}] = \mathbf{0}$. To find $\mathbf{M}$, we use the previous results to get

$$\begin{aligned} \text{Cov}[\mathbf{X}, \mathbf{Y} - \mathbf{M}\mathbf{X}] &= \text{Cov}[\mathbf{X}, \mathbf{Y}] - \text{Cov}[\mathbf{X}, \mathbf{M}\mathbf{X}] \\ &= \text{Cov}[\mathbf{X}, \mathbf{Y}] - \text{Cov}[\mathbf{X}, \mathbf{X}]\mathbf{M}' \\ &= \text{Cov}[\mathbf{X}, \mathbf{Y}] - \text{Var}[\mathbf{X}]\mathbf{M}'. \end{aligned} \tag{1.10}$$

By Theorem 1.4, $\text{Var}[\mathbf{X}]$ is positive-definite and therefore nonsingular (A.4.1). Hence (1.10) is zero for

$$\mathbf{M}' = (\,\text{Var}[\mathbf{X}])^{-1}\,\text{Cov}[\mathbf{X}, \mathbf{Y}].$$ □

EXAMPLE 1.7 We now give an example of a singular variance matrix by using the two-cell multinomial distribution to represent a binomial distribution as follows:

$$\text{pr}(X_1 = x_1, X_2 = x_2) = \frac{n!}{x_1!\,x_2!} p_1^{x_1} p_2^{x_2}, \quad p_1 + p_2 = 1,\ x_1 + x_2 = n.$$

If $\mathbf{X} = (X_1, X_2)'$, then

$$\text{Var}[\mathbf{X}] = \begin{pmatrix} np_1(1-p_1) & -np_1p_2 \\ -np_1p_2 & np_2(1-p_2) \end{pmatrix},$$

which has rank 1 as $p_2 = 1 - p_1$. □

EXERCISES 1a

1. Prove that if $\mathbf{a}$ is a vector of constants with the same dimension as the random vector $\mathbf{X}$, then

$$E[(\mathbf{X} - \mathbf{a})(\mathbf{X} - \mathbf{a})'] = \text{Var}[\mathbf{X}] + (E[\mathbf{X}] - \mathbf{a})(E[\mathbf{X}] - \mathbf{a})'.$$

If $\text{Var}[\mathbf{X}] = \boldsymbol{\Sigma} = (\sigma_{ij})$, deduce that

$$E[||\mathbf{X} - \mathbf{a}||^2] = \sum_i \sigma_{ii} + ||E[\mathbf{X}] - \mathbf{a}||^2.$$

2. If $\mathbf{X}$ and $\mathbf{Y}$ are $m \times 1$ and $n \times 1$ vectors of random variables, and $\mathbf{a}$ and $\mathbf{b}$ are $m \times 1$ and $n \times 1$ vectors of constants, prove that

$$\text{Cov}[\mathbf{X} - \mathbf{a}, \mathbf{Y} - \mathbf{b}] = \text{Cov}[\mathbf{X}, \mathbf{Y}].$$

3. Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$ be a vector of random variables, and let $Y_1 = X_1$, $Y_i = X_i - X_{i-1}$ ($i = 2, 3, \ldots, n$). If the Y_i are mutually independent random variables, each with unit variance, find $\text{Var}[\mathbf{X}]$.

4. If $X_1, X_2, \ldots, X_n$ are random variables satisfying $X_{i+1} = \rho X_i$ ($i = 1, 2, \ldots, n-1$), where ρ is a constant, and $\text{var}[X_1] = \sigma^2$, find $\text{Var}[\mathbf{X}]$.

1.5 MEAN AND VARIANCE OF QUADRATIC FORMS

Quadratic forms play a major role in this book. In particular, we will frequently need to find the expected value of a quadratic form using the following theorem.

THEOREM 1.5 *Let $\mathbf{X} = (X_i)$ be an $n \times 1$ vector of random variables, and let $\mathbf{A}$ be an $n \times n$ symmetric matrix. If $E[\mathbf{X}] = \boldsymbol{\mu}$ and $\text{Var}[\mathbf{X}] = \boldsymbol{\Sigma} = (\sigma_{ij})$, then*

$$E[\mathbf{X}'\mathbf{A}\mathbf{X}] = \underbrace{\text{tr}(\mathbf{A}\boldsymbol{\Sigma})}_{\text{variance}} + \underbrace{\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}}_{\text{bias}}.$$

Proof.

$$\begin{aligned} E[\mathbf{X}'\mathbf{A}\mathbf{X}] &= \text{tr}(E[\mathbf{X}'\mathbf{A}\mathbf{X}]) \\ &= E[\text{tr}(\mathbf{X}'\mathbf{A}\mathbf{X})] \\ &= E[\text{tr}(\mathbf{A}\mathbf{X}\mathbf{X}')] \quad \text{[by A.1.2]} \\ &= \text{tr}(E[\mathbf{A}\mathbf{X}\mathbf{X}']) \\ &= \text{tr}(\mathbf{A}E[\mathbf{X}\mathbf{X}']) \\ &= \text{tr}\,[\mathbf{A}(\,\text{Var}[\mathbf{X}] + \boldsymbol{\mu}\boldsymbol{\mu}')] \quad \text{[by (1.5)]} \\ &= \text{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \text{tr}(\mathbf{A}\boldsymbol{\mu}\boldsymbol{\mu}') \\ &= \text{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} \quad \text{[by A.1.2]}. \end{aligned}$$ □

We can deduce two special cases. First, by setting $\mathbf{Y} = \mathbf{X} - \mathbf{b}$ and noting that $\text{Var}[\mathbf{Y}] = \text{Var}[\mathbf{X}]$ (by Example 1.4), we have

$$E[(\mathbf{X} - \mathbf{b})'\mathbf{A}(\mathbf{X} - \mathbf{b})] = \text{tr}(\mathbf{A}\boldsymbol{\Sigma}) + (\boldsymbol{\mu} - \mathbf{b})'\mathbf{A}(\boldsymbol{\mu} - \mathbf{b}). \tag{1.11}$$

Linear Regression Analysis

Second Edition

GEORGE A. F. SEBER
ALAN J. LEE
Department of Statistics
University of Auckland
Auckland, New Zealand

WILEY-INTERSCIENCE
A JOHN WILEY & SONS PUBLICATION