Chapter 1
Introduction

This book is about linear models. Linear models are models that are linear in their parameters. A typical model considered is

$$Y = X\beta + e,$$

where Y is an $n \times 1$ vector of random observations, X is an $n \times p$ matrix of known constants called the *model* (or *design*) matrix, β is a $p \times 1$ vector of unobservable fixed parameters, and e is an $n \times 1$ vector of unobservable random errors. Both Y and e are random vectors. We assume that the errors have mean zero, a common variance, and are uncorrelated. In particular, $\mathrm{E}(e) = 0$ and $\mathrm{Cov}(e) = \sigma^2 I$, where σ^2 is some unknown parameter. (The operations $\mathrm{E}(\cdot)$ and $\mathrm{Cov}(\cdot)$ will be defined formally a bit later.) Our object is to explore models that can be used to predict future observable events. Much of our effort will be devoted to drawing inferences, in the form of point estimates, tests, and confidence regions, about the parameters β and σ^2. In order to get tests and confidence regions, we will assume that e has an n-dimensional normal distribution with mean vector $(0,0,\ldots,0)'$ and covariance matrix $\sigma^2 I$, i.e., $e \sim N(0, \sigma^2 I)$.

Applications often fall into two special cases: Regression Analysis and Analysis of Variance. Regression Analysis refers to models in which the matrix $X'X$ is nonsingular. Analysis of Variance (ANOVA) models are models in which the model matrix consists entirely of zeros and ones. ANOVA models are sometimes called classification models.

EXAMPLE 1.0.1. *Simple Linear Regression.*
Consider the model

$$y_i = \beta_0 + \beta_1 x_i + e_i,$$

$i = 1,\ldots,6$, $(x_1,x_2,x_3,x_4,x_5,x_6) = (1,2,3,4,5,6)$, where the e_is are independent $N(0,\sigma^2)$. In matrix notation we can write this as

R. Christensen, *Plane Answers to Complex Questions: The Theory of Linear Models*, Springer Texts in Statistics, DOI 10.1007/978-1-4419-9816-3_1,

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{bmatrix}$$

$$Y = X \quad \beta + e.$$

EXAMPLE 1.0.2 *One-Way Analysis of Variance.*
The model

$$y_{ij} = \mu + \alpha_i + e_{ij},$$

$i = 1, \ldots, 3,\ j = 1, \ldots, N_i,\ (N_1, N_2, N_3) = (3, 1, 2)$, where the e_{ij}s are independent $N\left(0, \sigma^2\right)$, can be written as

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{31} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{31} \\ e_{32} \end{bmatrix}$$

$$Y = X \quad \beta + e.$$

Examples 1.0.1 and 1.0.2 will be used to illustrate concepts in Chapters 2 and 3.

With any good statistical procedure, it is necessary to investigate whether the assumptions that have been made are reasonable. Methods for evaluating the validity of the assumptions will be considered. These consist of both formal statistical tests and the informal examination of residuals. We will also consider the issue of how to select a model when several alternative models seem plausible.

The approach taken here emphasizes the use of vector spaces, subspaces, orthogonality, and projections. These and other topics in linear algebra are reviewed in Appendices A and B. It is absolutely vital that the reader be familiar with the material presented in the first two appendices. Appendix C contains the definitions of some commonly used distributions. Much of the notation used in the book is set in Appendices A, B, and C. To develop the distribution theory necessary for tests and confidence regions, it is necessary to study properties of the multivariate normal distribution and properties of quadratic forms. We begin with a discussion of random vectors and matrices.

Exercise 1.1 Write the following models in matrix notation:

(a) Multiple regression

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i,$$

$i = 1, \ldots, 6$.

(b) Two-way ANOVA with interaction

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk},$$

$i = 1,2,3,\ j = 1,2,\ k = 1,2$.

(c) Two-way analysis of covariance (ACOVA) with no interaction

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma x_{ijk} + e_{ijk},$$

$i = 1,2,3,\ j = 1,2,\ k = 1,2$.

(d) Multiple polynomial regression

$$y_i = \beta_{00} + \beta_{10}x_{i1} + \beta_{01}x_{i2} + \beta_{20}x_{i1}^2 + \beta_{02}x_{i2}^2 + \beta_{11}x_{i1}x_{i2} + e_i,$$

$i = 1, \ldots, 6$.

1.1 Random Vectors and Matrices

Let $y_1, \ldots, y_n$ be random variables with $\mathrm{E}(y_i) = \mu_i$, $\mathrm{Var}(y_i) = \sigma_{ii}$, and $\mathrm{Cov}(y_i, y_j) = \sigma_{ij} \equiv \sigma_{ji}$.

Writing the random variables as an n-dimensional vector Y, we can define the expected value of Y elementwise as

$$\mathrm{E}(Y) = \mathrm{E}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \mathrm{E}y_1 \\ \mathrm{E}y_2 \\ \vdots \\ \mathrm{E}y_n \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} = \mu.$$

In general, we can define a random matrix $W = [w_{ij}]$, where each w_{ij}, $i = 1, \ldots, r$, $j = 1, \ldots, s$, is a random variable. The expected value of W is taken elementwise, i.e., $\mathrm{E}(W) = [\mathrm{E}(w_{ij})]$. This leads to the definition of the *covariance matrix* of Y as

$$\mathrm{Cov}(Y) = \mathrm{E}\big[(Y-\mu)(Y-\mu)'\big] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}.$$

A random vector is referred to as singular or nonsingular depending on whether its covariance matrix is singular or nonsingular. Sometimes the covariance matrix is called the *variance-covariance matrix* or the *dispersion matrix*.

It is easy to see that if Y is an n-dimensional random vector, A is a fixed $r \times n$ matrix, and b is a fixed vector in $\mathbf{R}^r$, then

$$\mathrm{E}(AY+b) = A\mathrm{E}(Y)+b$$

and

$$\mathrm{Cov}(AY+b) = A\mathrm{Cov}(Y)A'.$$

This last equality can be used to show that for any random vector Y, $\mathrm{Cov}(Y)$ is nonnegative definite. It follows that Y is nonsingular if and only if $\mathrm{Cov}(Y)$ is positive definite.

Exercise 1.2 Let W be an $r \times s$ random matrix, and let A and C be $n \times r$ and $n \times s$ matrices of constants, respectively. Show that $\mathrm{E}(AW+C) = A\mathrm{E}(W)+C$. If B is an $s \times t$ matrix of constants, show that $\mathrm{E}(AWB) = A\mathrm{E}(W)B$. If $s = 1$, show that $\mathrm{Cov}(AW+C) = A\mathrm{Cov}(W)A'$.

Exercise 1.3 Show that $\mathrm{Cov}(Y)$ is nonnegative definite for any random vector Y.

The covariance of two random vectors with possibly different dimensions can be defined. If $W_{r\times 1}$ and $Y_{s\times 1}$ are random vectors with $\mathrm{E}W = \gamma$ and $\mathrm{E}Y = \mu$, then the covariance of W and Y is the $r \times s$ matrix

$$\mathrm{Cov}(W,Y) = \mathrm{E}[(W-\gamma)(Y-\mu)'].$$

In particular, $\mathrm{Cov}(Y,Y) = \mathrm{Cov}(Y)$. If A and B are fixed matrices, the results of Exercise 1.2 quickly yield

$$\mathrm{Cov}(AW,BY) = A\mathrm{Cov}(W,Y)B'.$$

Another simple consequence of the definition is:

Theorem 1.1.1. If A and B are fixed matrices and W and Y are random vectors, and if AW and BY are both vectors in $\mathbf{R}^n$, then, assuming that the expectations exist,

$$\mathrm{Cov}(AW+BY) = A\mathrm{Cov}(W)A' + B\mathrm{Cov}(Y)B' + A\mathrm{Cov}(W,Y)B' + B\mathrm{Cov}(Y,W)A'.$$

PROOF. Without loss of generality we can assume that $\mathrm{E}(W) = 0$ and $\mathrm{E}(Y) = 0$:

$$\begin{aligned}
\mathrm{Cov}(AW+BY) &= \mathrm{E}[(AW+BY)(AW+BY)'] \\
&= A\mathrm{E}[WW']A' + B\mathrm{E}[YY']B' + A\mathrm{E}[WY']B' + B\mathrm{E}[YW']A' \\
&= A\mathrm{Cov}(W)A' + B\mathrm{Cov}(Y)B' + A\mathrm{Cov}(W,Y)B' + B\mathrm{Cov}(Y,W)A'.
\end{aligned}$$

$\square$

1.2 Multivariate Normal Distributions

It is assumed that the reader is familiar with the basic ideas of multivariate distributions. A summary of these ideas is contained in Appendix D.

Let $Z = [z_1, \ldots, z_n]'$ be a random vector with $z_1, \ldots, z_n$ *independent identically distributed (i.i.d.)* $N(0,1)$ random variables. Note that $\mathrm{E}(Z) = 0$ and $\mathrm{Cov}(Z) = I$.

Definition 1.2.1. Y has an r-dimensional *multivariate normal distribution* if Y has the same distribution as $AZ + b$, i.e., $Y \sim AZ + b$, for some n, some fixed $r \times n$ matrix A, and some fixed r vector b. We indicate the multivariate normal distribution of Y by writing $Y \sim N(b, AA')$.

Since A and b are fixed, and since $\mathrm{E}(Z) = 0$, $\mathrm{Cov}(Z) = I$, we have $\mathrm{E}(Y) = b$ and $\mathrm{Cov}(Y) = AA'$.

It is not clear that the notation $Y \sim N(b, AA')$ is well defined, i.e., that a multivariate normal distribution depends only on its mean vector and covariance matrix. Clearly, if we have $Y \sim AZ + b$, then the notation $Y \sim N(b, AA')$ makes sense. However, if we write, say, $Y \sim N(\mu, V)$, we may be able to write both $V = AA'$ and $V = BB'$, where $A \neq B$. In that case, we do not know whether to take $Y \sim AZ + \mu$ or $Y \sim BZ + \mu$. In fact, the number of columns in A and B need not even be the same, so the length of the vector Z could change between $Y \sim AZ + \mu$ and $Y \sim BZ + \mu$. We need to show that it does not matter which characterization is used. We now give such an argument based on characteristic functions. The argument is based on the fact that any two random vectors with the same characteristic function have the same distribution. Appendix D contains the definition of the characteristic function of a random vector.

Theorem 1.2.2. If $Y \sim N(\mu, V)$ and $W \sim N(\mu, V)$, then Y and W have the same distribution.

PROOF. Observe that

$$\varphi_Z(t) = \mathrm{E}[\exp(it'Z)] = \prod_{j=1}^{n} \mathrm{E}[\exp(it_j z_j)] = \prod_{j=1}^{n} \exp(-t_j^2/2) = \exp(-t't/2).$$

Define $Y \sim AZ + \mu$, where $AA' = V$. The characteristic function of Y is

$$\begin{aligned}
\varphi_Y(t) &= \mathrm{E}[\exp(it'Y)] = \mathrm{E}[\exp(it'[AZ + \mu])] \\
&= \exp(it'\mu)\varphi_Z(A't) \\
&= \exp(it'\mu)\exp(-t'AA't/2) \\
&= \exp(it'\mu - t'Vt/2).
\end{aligned}$$

Similarly,

$$\varphi_W(t) = \exp(it'\mu - t'Vt/2).$$

Since the characteristic functions are the same, $Y \sim W$. $\square$

Suppose that Y is nonsingular and that $Y \sim N(\mu, V)$; then Y has a density. By definition, Y nonsingular means precisely that V is positive definite. By Corollary B.23, we can write $V = AA'$, with A nonsingular. Since $Y \sim AZ + \mu$ involves a nonsingular transformation of the random vector Z, which has a known density, it is quite easy to find the density of Y. The density is

$$f(y) = (2\pi)^{-n/2}[\det(V)]^{-1/2}\exp[-(y-\mu)'V^{-1}(y-\mu)/2],$$

where $\det(V)$ is the determinant of V.

Exercise 1.4 Show that the function $f(y)$ given above is the density of Y when $Y \sim N(\mu, V)$ and V is nonsingular.

Hint: If Z has density $f_Z(z)$ and $Y = G(Z)$, the density of Y is

$$f_Y(y) = f_Z(G^{-1}(y))|\det(dG^{-1})|,$$

where dG^{-1} is the derivative (matrix of partial derivatives) of G^{-1} evaluated at y.

An important and useful result is that for random vectors having a joint multivariate normal distribution, the condition of having zero covariance is equivalent to the condition of independence.

Theorem 1.2.3. If $Y \sim N(\mu, V)$ and $Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$, then $\mathrm{Cov}(Y_1, Y_2) = 0$ if and only if Y_1 and Y_2 are independent.

PROOF. Partition V and μ to conform with Y, giving $V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$ and $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$. Note that $V_{12} = V_{21}' = \mathrm{Cov}(Y_1, Y_2)$.

$\Leftarrow$ If Y_1 and Y_2 are independent,

$$V_{12} = \mathrm{E}[(Y_1 - \mu_1)(Y_2 - \mu_2)'] = \mathrm{E}(Y_1 - \mu_1)\mathrm{E}(Y_2 - \mu_2)' = 0.$$

$\Rightarrow$ Suppose $\mathrm{Cov}(Y_1, Y_2) = 0$, so that $V_{12} = V_{21}' = 0$. Using the definition of multivariate normality, we will generate a version of Y in which it is clear that Y_1 and Y_2 are independent. Given the uniqueness established in Theorem 1.2.2, this is sufficient to establish independence of Y_1 and Y_2.

Since Y is multivariate normal, by definition we can write $Y \sim AZ + \mu$, where A is an $r \times n$ matrix. Partition A in conformance with $\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$ as $A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$ so that

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} = \begin{bmatrix} A_1A_1' & A_1A_2' \\ A_2A_1' & A_2A_2' \end{bmatrix}.$$

Because $V_{12} = 0$, we have $A_1A_2' = 0$ and

$$V = \begin{bmatrix} A_1A_1' & 0 \\ 0 & A_2A_2' \end{bmatrix}.$$

Now let $z_1, z_2, \ldots, z_{2n}$ be i.i.d. $N(0,1)$. Define the random vectors $Z_1 = [z_1, \ldots, z_n]'$, $Z_2 = [z_{n+1}, \ldots, z_{2n}]'$, and

$$Z_0 = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}.$$

Note that Z_1 and Z_2 are independent. Now consider the random vector

$$W = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} Z_0 + \mu.$$

By definition, W is multivariate normal with $\mathrm{E}(W) = \mu$ and

$$\begin{aligned} \mathrm{Cov}(W) &= \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}' \\ &= \begin{bmatrix} A_1A_1' & 0 \\ 0 & A_2A_2' \end{bmatrix} \\ &= V. \end{aligned}$$

We have shown that $W \sim N(\mu, V)$ and by assumption $Y \sim N(\mu, V)$. By Theorem 1.2.2, W and Y have exactly the same distribution; thus

$$Y \sim \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} Z_0 + \mu.$$

It follows that $Y_1 \sim [A_1, 0]Z_0 + \mu_1 = A_1Z_1 + \mu_1$ and $Y_2 \sim [0, A_2]Z_0 + \mu_2 = A_2Z_2 + \mu_2$. The joint distribution of (Y_1, Y_2) is the same as the joint distribution of $(A_1Z_1 + \mu_1, A_2Z_2 + \mu_2)$. However, Z_1 and Z_2 are independent; thus $A_1Z_1 + \mu_1$ and $A_2Z_2 + \mu_2$ are independent, and it follows that Y_1 and Y_2 are independent. $\square$

Exercise 1.5 Show that if Y is an r-dimensional random vector with $Y \sim N(\mu, V)$ and if B is a fixed $n \times r$ matrix, then $BY \sim N(B\mu, BVB')$.

In linear model theory, Theorem 1.2.3 is often applied to establish independence of two linear transformations of the data vector Y.

Corollary 1.2.4. If $Y \sim N(\mu, \sigma^2 I)$ and if $AB' = 0$, then AY and BY are independent.

PROOF. Consider the distribution of $\begin{bmatrix} A \\ B \end{bmatrix} Y$. By Exercise 1.5, the joint distribution of AY and BY is multivariate normal. Since $\text{Cov}(AY, BY) = \sigma^2 AIB' = \sigma^2 AB' = 0$, Theorem 1.2.3 implies that AY and BY are independent. □

1.3 Distributions of Quadratic Forms

In this section, quadratic forms are defined, the expectation of a quadratic form is found, and a series of results on independence and chi-squared distributions are given.

Definition 1.3.1. Let Y be an n-dimensional random vector and let A be an $n \times n$ matrix. A *quadratic form* is a random variable defined by $Y'AY$ for some Y and A.

Note that since $Y'AY$ is a scalar, $Y'AY = Y'A'Y = Y'(A+A')Y/2$. Since $(A+A')/2$ is always a symmetric matrix, we can, without loss of generality, restrict ourselves to quadratic forms where A is symmetric.

Theorem 1.3.2. If $\text{E}(Y) = \mu$ and $\text{Cov}(Y) = V$, then $\text{E}(Y'AY) = \text{tr}(AV) + \mu'A\mu$.

PROOF.

$$\begin{aligned}(Y-\mu)'A(Y-\mu) &= Y'AY - \mu'AY - Y'A\mu + \mu'A\mu, \\ \text{E}[(Y-\mu)'A(Y-\mu)] &= \text{E}[Y'AY] - \mu'A\mu - \mu'A\mu + \mu'A\mu,\end{aligned}$$

so $\text{E}[Y'AY] = \text{E}[(Y-\mu)'A(Y-\mu)] + \mu'A\mu$.

It is easily seen that for any random square matrix W, $\text{E}(\text{tr}(W)) = \text{tr}(\text{E}(W))$. Thus

$$\begin{aligned}\text{E}[(Y-\mu)'A(Y-\mu)] &= \text{E}(\text{tr}[(Y-\mu)'A(Y-\mu)]) \\ &= \text{E}(\text{tr}[A(Y-\mu)(Y-\mu)']) \\ &= \text{tr}(\text{E}[A(Y-\mu)(Y-\mu)']) \\ &= \text{tr}(A\text{E}[(Y-\mu)(Y-\mu)']) \\ &= \text{tr}(AV).\end{aligned}$$

Substitution gives

$$\text{E}(Y'AY) = \text{tr}(AV) + \mu'A\mu.$$ □

We now proceed to give results on chi-squared distributions and independence of quadratic forms. Note that by Definition C.1 and Theorem 1.2.3, if Z is an n-dimensional random vector and $Z \sim N(\mu, I)$, then $Z'Z \sim \chi^2(n, \mu'\mu/2)$.

Theorem 1.3.3. If Y is a random vector with $Y \sim N(\mu, I)$ and if M is any perpendicular projection matrix, then $Y'MY \sim \chi^2(r(M), \mu'M\mu/2)$.

PROOF. Let $r(M) = r$ and let $o_1, \ldots, o_r$ be an orthonormal basis for $C(M)$. Let $O = [o_1, \ldots, o_r]$ so that $M = OO'$. We now have $Y'MY = Y'OO'Y = (O'Y)'(O'Y)$, where $O'Y \sim N(O'\mu, O'IO)$. The columns of O are orthonormal, so $O'O$ is an $r \times r$ identity matrix, and by definition $(O'Y)'(O'Y) \sim \chi^2(r, \mu'OO'\mu/2)$ where $\mu'OO'\mu = \mu'M\mu$. □

Observe that if $Y \sim N\left(\mu, \sigma^2 I\right)$, then $[1/\sigma]Y \sim N([1/\sigma]\mu, I)$ and $Y'MY/\sigma^2 \sim \chi^2\left(r(M), \mu'M\mu/2\sigma^2\right)$.

Theorem 1.3.6 provides a generalization of Theorem 1.3.3 that is valid for an arbitrary covariance matrix. The next two lemmas are used in the proof of Theorem 1.3.6.

Lemma 1.3.4. If $Y \sim N(\mu, M)$, where $\mu \in C(M)$ and if M is a perpendicular projection matrix, then $Y'Y \sim \chi^2(r(M), \mu'\mu/2)$.

PROOF. Let O have r orthonormal columns with $M = OO'$. Since $\mu \in C(M)$, $\mu = Ob$. Let $W \sim N(b, I)$, then $Y \sim OW$. Since $O'O = I_r$ is also a perpendicular projection matrix, the previous theorem gives $Y'Y \sim W'O'OW \sim \chi^2(r, b'O'Ob/2)$. The proof is completed by observing that $r = r(M)$ and $b'O'Ob = \mu'\mu$. □

The following lemma establishes that, if Y is a singular random variable, then there exists a proper subset of $\mathbf{R}^n$ that contains Y with probability 1.

Lemma 1.3.5. If $\mathrm{E}(Y) = \mu$ and $\mathrm{Cov}(Y) = V$, then $\Pr[(Y - \mu) \in C(V)] = 1$.

PROOF. Without loss of generality, assume $\mu = 0$. Let M_V be the perpendicular projection operator onto $C(V)$; then $Y = M_V Y + (I - M_V)Y$. Clearly, $\mathrm{E}[(I - M_V)Y] = 0$ and $\mathrm{Cov}[(I - M_V)Y] = (I - M_V)V(I - M_V) = 0$. Thus, $\Pr[(I - M_V)Y = 0] = 1$ and $\Pr[Y = M_V Y] = 1$. Since $M_V Y \in C(V)$, we are done. □

Exercise 1.6 Show that if Y is a random vector and if $\mathrm{E}(Y) = 0$ and $\mathrm{Cov}(Y) = 0$, then $\Pr[Y = 0] = 1$.

Hint: For a random variable w with $\Pr[w \geq 0] = 1$ and $k > 0$, show that $\Pr[w \geq k] \leq \mathrm{E}(w)/k$. Apply this result to $Y'Y$.

Theorem 1.3.6. If $Y \sim N(\mu, V)$, then $Y'AY \sim \chi^2(\mathrm{tr}(AV), \mu'A\mu/2)$ provided that (1) $VAVAV = VAV$, (2) $\mu'AVA\mu = \mu'A\mu$, and (3) $VAVA\mu = VA\mu$.

PROOF. By Lemma 1.3.5, for the purpose of finding the distribution of $Y'AY$, we can assume that $Y = \mu + e$, where $e \in C(V)$. Using conditions (1), (2), and (3) of the theorem and the fact that $e = Vb$ for some b,

$$\begin{aligned} Y'AY &= \mu'A\mu + \mu'Ae + e'A\mu + e'Ae \\ &= \mu'AVA\mu + \mu'AVAe + e'AVA\mu + e'AVAe \\ &= Y'(AVA)Y. \end{aligned}$$

Write $V = QQ'$ so that $Y'AY = (Q'AY)'(Q'AY)$, where $Q'AY \sim N(Q'A\mu, Q'AVAQ)$. If we can show that $Q'AVAQ$ is a perpendicular projection matrix and that $Q'A\mu \in C(Q'AVAQ)$, then $Y'AY$ will have a chi-squared distribution by Lemma 1.3.4.

Since V is nonnegative definite, we can write $Q = Q_1Q_2$, where Q_1 has orthonormal columns and Q_2 is nonsingular. It follows that

$$Q_2^{-1}Q_1'V = Q_2^{-1}Q_1'[Q_1Q_2Q'] = Q'.$$

Applying this result, $VAVAV = VAV$ implies that $Q'AVAQ = Q'AQ$. Now $Q'AVAQ = (Q'AQ)(Q'AQ)$, so $Q'AQ$ is idempotent and symmetric and $Q'AQ = Q'AVAQ$ so $Q'AVAQ$ is a perpendicular projection operator.

From the preceding paragraph, to see that $Q'A\mu \in C(Q'AVAQ)$ it suffices to show that $Q'AQQ'A\mu = Q'A\mu$. Note that $VAVA\mu = VA\mu$ implies that $Q'AVA\mu = Q'A\mu$. However, since $Q'AVA\mu = Q'AQQ'A\mu$, we are done.

The noncentrality parameter is one-half of

$$(Q'A\mu)'(Q'A\mu) = \mu'AVA\mu = \mu'A\mu.$$

The degrees of freedom are

$$r(Q'AVAQ) = r(Q'AQ) = \mathrm{tr}(Q'AQ) = \mathrm{tr}(AQQ') = \mathrm{tr}(AV). \qquad \square$$

Exercise 1.7 (a) Show that if V is nonsingular, then the three conditions in Theorem 1.3.6 reduce to $AVA = A$. (b) Show that $Y'V^{-}Y$ has a chi-squared distribution with $r(V)$ degrees of freedom when $\mu \in C(V)$.

The next three theorems establish conditions under which quadratic forms are independent. Theorem 1.3.7 examines the important special case in which the covariance matrix is a multiple of the identity matrix. In addition to considering independence of quadratic forms, the theorem also examines independence between quadratic forms and linear transformations of the random vector.

Theorem 1.3.7. If $Y \sim N(\mu, \sigma^2 I)$ and $BA = 0$, then

(1) $Y'AY$ and BY are independent,
(2) $Y'AY$ and $Y'BY$ are independent,

where A is symmetric and in (2) B is symmetric.

PROOF. By Corollary 1.2.4, if $BA = 0$, BY and AY are independent. In addition, as discussed near the end of Appendix D, any function of AY is independent of any function of BY. Since $Y'AY = Y'AA^{-}AY$ and $Y'BY = Y'BB^{-}BY$ are functions of AY and BY, the theorem holds. $\square$

The final two theorems provide conditions for independence of quadratic forms under general covariance matrices.

Theorem 1.3.8. If $Y \sim N(\mu, V)$, A and B are nonnegative definite, and $VAVBV = 0$, then $Y'AY$ and $Y'BY$ are independent.

PROOF. Since A and B are nonnegative definite, we can write $A = RR'$ and $B = SS'$. We can also write $V = QQ'$.

$Y'AY = (R'Y)'(R'Y)$ and $Y'BY = (S'Y)'(S'Y)$ are independent

$$\begin{aligned} &\text{if} \quad R'Y \text{ and } S'Y \text{ are independent} \\ &\text{iff} \quad \mathrm{Cov}(R'Y, S'Y) = 0 \\ &\text{iff} \quad R'VS = 0 \\ &\text{iff} \quad R'QQ'S = 0 \\ &\text{iff} \quad C(Q'S) \perp C(Q'R). \end{aligned}$$

Since $C(AA') = C(A)$ for any A, we have

$$\begin{aligned} C(Q'S) \perp C(Q'R) \quad &\text{iff} \quad C(Q'SS'Q) \perp C(Q'RR'Q) \\ &\text{iff} \quad [Q'SS'Q][Q'RR'Q] = 0 \\ &\text{iff} \quad Q'BVAQ = 0 \\ &\text{iff} \quad C(Q) \perp C(BVAQ) \\ &\text{iff} \quad C(QQ') \perp C(BVAQ) \\ &\text{iff} \quad QQ'BVAQ = 0 \\ &\text{iff} \quad VBVAQ = 0. \end{aligned}$$

Repeating similar arguments for the right side gives $VBVAQ = 0$ iff $VBVAV = 0$. $\square$

Theorem 1.3.9. If $Y \sim N(\mu, V)$ and (1) $VAVBV = 0$, (2) $VAVB\mu = 0$, (3) $VBVA\mu = 0$, (4) $\mu'AVB\mu = 0$, and conditions (1), (2), and (3) from Theorem 1.3.6 hold for both $Y'AY$ and $Y'BY$, then $Y'AY$ and $Y'BY$ are independent.

Exercise 1.8 Prove Theorem 1.3.9.

Hints: Let $V = QQ'$ and write $Y = \mu + QZ$, where $Z \sim N(0, I)$. Using $\perp\!\!\!\perp$ to indicate independence, show that

$$\begin{bmatrix} Q'AQZ \\ \mu'AQZ \end{bmatrix} \perp\!\!\!\perp \begin{bmatrix} Q'BQZ \\ \mu'BQZ \end{bmatrix}$$

and that, say, $Y'AY$ is a function $Q'AQZ$ and $\mu'AQZ$.

Note that Theorem 1.3.8 applies immediately if AY and BY are independent, i.e., if $AVB = 0$. In something of a converse, if V is nonsingular, the condition $VAVBV = 0$ is equivalent to $AVB = 0$; so the theorem applies only when AY and BY are independent. However, if V is singular, the conditions of Theorems 1.3.8 and 1.3.9 can be satisfied even when AY and BY are not independent.

Exercise 1.9 Let M be the perpendicular projection operator onto $C(X)$. Show that $(I-M)$ is the perpendicular projection operator onto $C(X)^{\perp}$. Find $\mathrm{tr}(I-M)$ in terms of $r(X)$.

Exercise 1.10 For a linear model $Y = X\beta + e$, $\mathrm{E}(e) = 0$, $\mathrm{Cov}(e) = \sigma^2 I$, show that $\mathrm{E}(Y) = X\beta$ and $\mathrm{Cov}(Y) = \sigma^2 I$.

Exercise 1.11 For a linear model $Y = X\beta + e$, $\mathrm{E}(e) = 0$, $\mathrm{Cov}(e) = \sigma^2 I$, the residuals are

$$\hat{e} = Y - X\hat{\beta} = (I-M)Y,$$

where M is the perpendicular projection operator onto $C(X)$. Find

(a) $\mathrm{E}(\hat{e})$.
(b) $\mathrm{Cov}(\hat{e})$.
(c) $\mathrm{Cov}(\hat{e}, MY)$.
(d) $\mathrm{E}(\hat{e}'\hat{e})$.
(e) Show that $\hat{e}'\hat{e} = Y'Y - (Y'M)Y$.

[Note: In Chapter 2 we will show that for a least squares estimate of β, say $\hat{\beta}$, we have $MY = X\hat{\beta}$.]

(f) Rewrite (c) and (e) in terms of $\hat{\beta}$.

1.4 Generalized Linear Models

We now give a brief introduction to *generalized linear models*. On occasion through the rest of the book, reference will be made to various properties of linear models that extend easily to generalized linear models. See McCullagh and Nelder (1989) or Christensen (1997) for more extensive discussions of generalized linear models and their applications. First it must be noted that a *general linear model* is a linear model but a *generalized* linear model is a generalization of the concept of a linear model. Generalized linear models include linear models as a special case but also include logistic regression, exponential regression, and gamma regression as special cases. Additionally, log-linear models for multinomial data are closely related to generalized linear models.

Consider a random vector Y with $\mathrm{E}(Y) = \mu$. Let h be an arbitrary function on the real numbers and, for a vector $v = (v_1, \ldots, v_n)'$, define the vector function

$$h(v) \equiv \begin{bmatrix} h(v_1) \\ \vdots \\ h(v_n) \end{bmatrix}.$$

The primary idea of a generalized linear model is specifying that

$$\mu = h(X\beta),$$

where h is a known invertible function and X and β are defined as for linear models. The inverse of h is called the *link* function. In particular, linear models use the identity function $h(v) = v$, logistic regression uses the logistic transform $h(v) = e^v/(1+e^v)$, and both exponential regression and log-linear models use the exponential transform $h(v) = e^v$. Their link functions are, respectively, the identity, logit, and log transforms. Because the linear structure $X\beta$ is used in generalized linear models, many of the analysis techniques used for linear models can be easily extended to generalized linear models.

Typically, in a generalized linear model it is assumed that the y_is are independent and each follows a distribution having density or mass function of the form

$$f(y_i|\theta_i, \phi; w_i) = \exp\left\{\frac{w_i}{\phi}[\theta_i y_i - r(\theta_i)]\right\} g(y_i, \phi, w_i), \tag{1}$$

where $r(\cdot)$ and $g(\cdot,\cdot,\cdot)$ are known functions and θ_i, ϕ, and w_i are scalars. By assumption, w_i is a fixed known number. Typically, it is a known weight that indicates knowledge about a pattern in the variabilities of the y_is. ϕ is either known or is an unknown parameter, but for some purposes is always treated like it is known. It is related to the variance of y_i. The parameter θ_i is related to the mean of y_i. For linear models, the standard assumption is that the y_is are independent $N(\theta_i, \phi/w_i)$, with $\phi \equiv \sigma^2$ and $w_i \equiv 1$. The standard assumption of logistic regression is that the $N_i y_i$s are distributed as independent binomials with N_i trials, success probability

$$\mathrm{E}(y_i) \equiv \mu_i \equiv p_i = e^{\theta_i}/[1+e^{\theta_i}],$$

$w_i = N_i$, and $\phi = 1$. Log-linear models fit into this framework when one assumes that the y_is are independent Poisson with mean $\mu_i = e^{\theta_i}$, $w_i = 1$, and $\phi = 1$. Note that in these cases the mean is some function of θ_i and that ϕ is merely related to the variance. Note also that in the three examples, the h function has already appeared, even though these distributions have not yet incorporated the linear structure of the generalized linear model.

To investigate the relationship between the θ_i parameters and the linear structure $x_i'\beta$, where x_i' is the ith row of X, let $\dot{r}(\theta_i)$ be the derivative $dr(\theta_i)/d\theta_i$. It can be shown that

$$\mathrm{E}(y_i) \equiv \mu_i = \dot{r}(\theta_i).$$

Thus, another way to think about the modeling process is that

$$\mu_i = h(x_i'\beta) = \dot{r}(\theta_i),$$

where both h and $\dot{r}$ are invertible. In matrix form, write $\theta = (\theta_1, \ldots, \theta_n)'$ so that

$$X\beta = h^{-1}(\mu) = h^{-1}[\dot{r}(\theta)] \quad \text{and} \quad \dot{r}^{-1}[h(X\beta)] = \dot{r}^{-1}(\mu) = \theta.$$

The special case of $h(\cdot) = \dot{r}(\cdot)$ gives $X\beta = \theta$. This is known as a *canonical generalized linear model*, or as using a *canonical link function*. The three examples given earlier are all examples of canonical generalized linear models. Linear models with normally distributed data are canonical generalized linear models. Logistic regression is the canonical model having $N_i y_i$ distributed Binomial(N_i, μ_i) for known N_i with $h^{-1}(\mu_i) \equiv \log(\mu_i/[1-\mu_i])$. Another canonical generalized linear model has y_i distributed Poisson(μ_i) with $h^{-1}(\mu_i) \equiv \log(\mu_i)$.

1.5 Additional Exercises

Exercise 1.5.1 Let $Y = (y_1, y_2, y_3)'$ be a random vector. Suppose that $\mathrm{E}(Y) \in \mathscr{M}$, where $\mathscr{M}$ is defined by

$$\mathscr{M} = \{(a, a-b, 2b)' | a, b \in \mathbf{R}\}.$$

(a) Show that $\mathscr{M}$ is a vector space.
(b) Find a basis for $\mathscr{M}$.
(c) Write a linear model for this problem (i.e., find X such that $Y = X\beta + e$, $\mathrm{E}(e) = 0$).
(d) If $\beta = (\beta_1, \beta_2)'$ in part (c), find two vectors $r = (r_1, r_2, r_3)'$ and $s = (s_1, s_2, s_3)'$ such that $\mathrm{E}(r'Y) = r'X\beta = \beta_1$ and $\mathrm{E}(s'Y) = \beta_2$. Find another vector $t = (t_1, t_2, t_3)'$ with $r \neq t$ but $\mathrm{E}(t'Y) = \beta_1$.

Exercise 1.5.2 Let $Y = (y_1, y_2, y_3)'$ with $Y \sim N(\mu, V)$, where

$$\mu = (5, 6, 7)'$$

and

$$V = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 3 & 2 \\ 1 & 2 & 4 \end{bmatrix}.$$

Find
(a) the marginal distribution of y_1,
(b) the joint distribution of y_1 and y_2,

(c) the conditional distribution of y_3 given $y_1 = u_1$ and $y_2 = u_2$,
(d) the conditional distribution of y_3 given $y_1 = u_1$,
(e) the conditional distribution of y_1 and y_2 given $y_3 = u_3$,
(f) the correlations ρ_{12}, ρ_{13}, ρ_{23},
(g) the distribution of

$$Z = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} Y + \begin{bmatrix} -15 \\ -18 \end{bmatrix},$$

(h) the characteristic functions of Y and Z.

Exercise 1.5.3 The density of $Y = (y_1, y_2, y_3)'$ is

$$(2\pi)^{-3/2} |V|^{-1/2} e^{-Q/2},$$

where

$$Q = 2y_1^2 + y_2^2 + y_3^2 + 2y_1y_2 - 8y_1 - 4y_2 + 8.$$

Find V^{-1} and μ.

Exercise 1.5.4 Let $Y \sim N(J\mu, \sigma^2 I)$ and let $O = \left[n^{-1/2}J, O_1\right]$ be an orthogonal matrix.
(a) Find the distribution of $O'Y$.
(b) Show that $\bar{y}_{\cdot} = (1/n)J'Y$ and that $s^2 = Y'O_1O_1'Y/(n-1)$.
(c) Show that $\bar{y}_{\cdot}$ and s^2 are independent.
Hint: Show that $Y'Y = Y'OO'Y = Y'(1/n)JJ'Y + Y'O_1O_1'Y$.

Exercise 1.5.5 Let $Y = (y_1, y_2)'$ have a $N(0, I)$ distribution. Show that if

$$A = \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & b \\ b & 1 \end{bmatrix},$$

then the conditions of Theorem 1.3.7 implying independence of $Y'AY$ and $Y'BY$ are satisfied only if $|a| = 1/|b|$ and $a = -b$. What are the possible choices for a and b?

Exercise 1.5.6 Let $Y = (y_1, y_2, y_3)'$ have a $N(\mu, \sigma^2 I)$ distribution. Consider the quadratic forms defined by the matrices M_1, M_2, and M_3 given below.
(a) Find the distribution of each $Y'M_iY$.
(b) Show that the quadratic forms are pairwise independent.
(c) Show that the quadratic forms are mutually independent.

$$M_1 = \frac{1}{3}J_3^3, \qquad M_2 = \frac{1}{14}\begin{bmatrix} 9 & -3 & -6 \\ -3 & 1 & 2 \\ -6 & 2 & 4 \end{bmatrix},$$

$$M_3 = \frac{1}{42}\begin{bmatrix} 1 & -5 & 4 \\ -5 & 25 & -20 \\ 4 & -20 & 16 \end{bmatrix}.$$

Exercise 1.5.7 Let A be symmetric, $Y \sim N(0,V)$, and $w_1,\ldots,w_s$ be independent $\chi^2(1)$ random variables. Show that for some value of s and some numbers λ_i, $Y'AY \sim \sum_{i=1}^{s} \lambda_i w_i$.

Hint: $Y \sim QZ$ so $Y'AY \sim Z'Q'AQZ$. Write $Q'AQ = PD(\lambda_i)P'$.

Exercise 1.5.8. Show that

(a) for Example 1.0.1 the perpendicular projection operator onto $C(X)$ is

$$M = \frac{1}{6}J_6^6 + \frac{1}{70}\begin{bmatrix} 25 & 15 & 5 & -5 & -15 & -25 \\ 15 & 9 & 3 & -3 & -9 & -15 \\ 5 & 3 & 1 & -1 & -3 & -5 \\ -5 & -3 & -1 & 1 & 3 & 5 \\ -15 & -9 & -3 & 3 & 9 & 15 \\ -25 & -15 & -5 & 5 & 15 & 25 \end{bmatrix};$$

(b) for Example 1.0.2 the perpendicular projection operator onto $C(X)$ is

$$M = \begin{bmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{bmatrix}.$$