LINEAR MODELS

The Theory and Application of Analysis of Variance

Brenton R. Clarke

A JOHN WILEY & SONS, INC., PUBLICATION

This Page Intentionally Left Blank

LINEAR MODELS

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

Editors: *David J. Balding, Noel A. C. Cressie, Garrett M. Fitzmaurice, Iain M. Johnstone, Geert Molenberghs, David W. Scott, Adrian F. M. Smith, Ruey S. Tsay, Sanford Weisberg*
Editors Emeriti: *Vic Barnett, J. Stuart Hunter, Jozef L. Teugels*

A complete list of the titles in this series appears at the end of this volume.

LINEAR MODELS

The Theory and Application of Analysis of Variance

Brenton R. Clarke

WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Clarke, Brenton R.
Linear models : the theory and application of analysis of variance / Brenton R. Clarke.
p. cm. — (Wiley series in probability and statistics)
Includes bibliographical references and index.
ISBN 978-0-470-02566-6 (cloth)
1. Linear models (Statistics) 2. Analysis of variance. I. Title.
QA279.C55 2008
519.5'38—dc22 2008004930

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

To my wife, Erica
and
my sons, Andrew and Stephen

This Page Intentionally Left Blank

CONTENTS

Preface

In this book we provide a vector approach to linear models, followed by specific examples of what is known as the *canonical form* (Scheffé). This connection provides a transparent path to the subject of analysis of variance (ANOVA), illustrated for both regression and a number of orthogonal experimental designs. The approach endeavors to eliminate some of the mystery in the development of ANOVA and various representations of orthogonal designs. Many books list many different types of ANOVA for use in a variety of situations. To the mathematically oriented statistician or indeed any student of statistics, these books do not ease the understanding of where such ANOVA comes from but can be useful references when seeking ANOVA for use in particular situations. By coming to understand some basic rudiments of mathematics and statistics, one can prepare oneself to relate to the statistical applications, of which there are many. This may be a process that takes years and some experience. This book can be a useful foundation on such a career path.

In the first chapter, eight well-known examples of statistical models that involve fixed-effects parameters are presented in the vector form of the linear model. Some preliminary objectives of such linear models are then given. This gentle introduction is complemented by a simple model involving a regression through the origin with one explanatory variable. Such a model is used to demonstrate that the direct approach to least squares theory soon becomes unwieldy. This, then, is the reason for embarking on the vector approach, which is underpinned by a fairly succinct account of vector

space theory and projections onto orthogonal subspaces. In this sense the book is like many others written in the general area of ANOVA, since least squares theory involves the geometry of orthogonal spaces and projections of observation vectors onto orthogonal subspaces that lie within a vector space. Least squares regression is then justified in the classical way through Gauss–Markov theory, followed by some basic distribution theory and hypothesis testing of fixed-effects parameters.

Where this book differs significantly from most books on ANOVA is in the discussion beginning in Chapter 5 regarding Helmert matrices and Kronecker products (combined). This allows succinct and explicit forms of contrasts that yield both the orthogonal components in ANOVA, including projection matrices, and distributions of component sums of squares with illustrations for a number of designs, including two-way ANOVA, Latin squares, and 2^k factorial designs. The general approach to orthogonal designs is then discussed by introducing *relationship algebra* and the triple classification. As with any linear model estimation and fitting, one also needs to consider residual analysis in the form of diagnostic checking or the possibility of robust fitting and identification of outliers. The classical approaches to diagnostic checking are then followed up with a brief discussion of robust methods. Here recent research highlighting robust adaptive methods which automatically identify the outliers and give least squares estimates with the outliers removed is discussed. The particular approach to ANOVA given in the earlier chapters is then generalized to include models with random effects, such as in mixed model analysis. Illustrations include a split plot experimental design. The representation of orthogonal components here is new albeit the ANOVA techniques themselves are well documented in the literature.

Many books on the theory of linear models begin with basic distribution theory and descriptions of density functions with consequent definitions of expectation and variance. In this book we assume such theory, presuming students and/or researchers who embark on reading it have a knowledge of these basic statistical ideas. Typically, then, competing books develop early the ideas of likelihood theory, since least squares estimates can be motivated by introducing likelihood, assuming the normal parametric density for the errors. We however, take a historical approach, — beginning with least squares theory. Nevertheless, the theory of likelihood estimation allows for a general umbrella that covers estimation more generally. There are, of course, several likelihood approaches, just as there is more than one choice for the parametric density function for the errors. Consequently, we pay attention to some of the different likelihood approaches, which are discussed in detail in Chapter 9.

In what may seem a digression, in Chapter 10 we return to the general theory of the choice of contrasts. Although this may appear to be a chapter that could follow Chapter 7, it involves somewhat complicated algebra and is not necessarily directed immediately at ANOVA. The explicit forms of the orthogonal contrasts for the linear model given earlier in the book can in fact be generalized. By restricting ourselves to the error contrasts and full rank models, we illustrate a general formulation for the error contrasts and highlight some classical representations of such. Discussion then proceeds to extensions involving less than full rank models.

The final chapter relates to further directions and a summary. It is not meant to be

exhaustive but to put forth additional approaches to estimation and testing, several of which I have described elsewhere.

This volume has evolved from my experience as a student, teacher, and researcher at several institutions around the world, but in particular at Murdoch University in Western Australia, where since 1984, I have taught a one–semester graduate degree unit on linear models and experimental design. The unit currently sits at the honours or fourth–year level of an Australian university degree, so the book should be valuable as a graduate text for students at the master's or Ph.D. level at an American university. My research interests in the area of ANOVA were inspired by my holding a postdoctoral position at the University of London and the Swiss Federal Institute of Technology and a position as visiting professor at the University of North Carolina at Chapel Hill, the latter in the fall of 1983, even though my research at that time focused principally on time–series analysis and robustness theory. Using robust techniques in regression is a valid alternative to the classical least squares methods, combined with diagnostic tools based on examination of residuals. Typically, however, we must learn to walk before we can run, and the beauty of discussing classical distribution theory for regression and ANOVA in experimental design transcends the claims that we should always use robust techniques, as contended by some"robustniks". There are often good reasons for using robust methods and comparing results to classical approaches, but when there are few observations per cell or treatment combination exploiting the structure in the data using classical techniques can be more worthwhile than employing the black box approach of implementing a robust method. [See, in particular, a recent discussion of Clarke and Monaco (2004).]

I have included a set of problems at the end of each chapter. There are a total of 46 problems of varying degrees of difficulty. These have been developed over the years of teaching at Murdoch University. I encourage the reader to attempt as many problems as possible as these will reinforce knowledge learned in the chapter and in some cases, open up new areas of understanding.

There are some deliberate omissions in this book. The discussion of randomization in Section 6.7 is relatively brief. Randomization is important as a statistical concept, particularly in practice, and a full appreciation of it can be achieved by taking an applied statistics course. The book also has a relatively succinct discussion of variance components. The idea is to give a flavor of what can be done without elucidating each scenario that can be imagined.

Other topics, such as considering missing values and connectivity, although interesting, are not discussed, to keep the book to a reasonable length. Again, little space is given to distinguishing between the uppercase $\boldsymbol{Y}$ used as the random variable representation of a vector of observations and the lower case $\boldsymbol{y}$ used to denote actual measurements. As the distribution of components in ANOVA is discussed in depth we retain the use of upper case $\boldsymbol{Y}$ in representations of ANOVA.

The book assumes a knowledge of matrix theory and a course in probability and statistical inference where students are exposed to concepts such as expectation and variance and covariance, in particular of normal random variables. It is assumed that the classical results of asserting independence of jointly normal distributed variables when they have zero covariance were learned in a previous course. Also assumed are

the concepts of hypothesis testing and confidence intervals, in particular when the Student t-distribution is involved. Knowledge of chi-squared and Fisher's F-distributions is assumed, although discussion of these is given in more detail in Chapter 4. The vector space theory in Chapter 2 will hopefully build on concepts learned earlier in an appropriate mathematics course, but if this is not the case, several easy examples and illustrations of concepts are provided. These are important for an understanding of ANOVA as described later in the book.

In conclusion, I emphasize that this book is pitched at an advanced level of study and has been collated and inspired by the original research and teaching of the author. Although the inspiration for writing the book came from my own research, I have borrowed certain details from several authors, duly acknowledged in the text. To any authors who may, unwittingly, not have been mentioned, I apologize. Although much of the research in this area is historical, my presentation differs significantly from that of most books in this area.

BRENTON R. CLARKE

Perth, Western Australia

Acknowledgments

Many people have honed my skills and helped directly or indirectly in my creation of this book. I thank my earlier teachers earlier who have included John Darroch, Rob Muirhead, and Noel Cressie at Flinders University of South Australia, and Alan James while I was doing a unit on experimental design at the University of Adelaide as part of a shared honour's year curriculum between Flinders and Adelaide Universities. When working toward my Ph.D. at the Statistics Department in the Australian National University under the guidance of Chip Heathcote and Peter Hall, I explored the then newly emerging theory of robustness. This was so interesting a time for me that I finished my doctorate in three years, despite holding a full-time tutorship position in the Statistics Department. As a tutor in statistics units in analysis of variance (ANOVA), I have benefited from the guidance of David Manion at the Royal Holloway College, University of London, in 1980–1982 and Frank Hampel at the Swiss Federal Institute of Technology (ETH) in 1983. My research interests in the area of ANOVA were inspired by my holding postdoctoral positions at the University of London and ETH and a visiting professorship at the University of North Carolina (UNC) at Chapel Hill, the latter in the fall of 1983, even though my research was principally in time-series analysis and robustness theory. Although I have contributed widely in the area of robustness, I have kept my discussion of this area to Section 7.2, with some closing references in Chapter 11. I thank Toby Lewis for the references to Harter's history of least squares and its many alternatives.

I thank Edward J. Godolphin of the Royal Holloway College, University of London, for allowing me the opportunity to explore the methodology, in both time-series analysis and linear models and experimental design, of the structure associated with various types of uncorrelated residuals. This study was carried out while I was on a United Kingdom Social Science Research Council Grant in 1980–1982. Further work on this was carried out at ETH, UNC, Murdoch University, and the University of Edinburgh, the latter during a sabbatical in 1987, which led to a published paper (Clarke and Godolphin , 1992). I had the pleasure of supervising an honour's student, Rebecca Hogan, at Murdoch University in 2003, whose thesis set up the framework for discussing variance components along the lines of a paper of mine (Clarke, 2002). A summary paper was published in an issue of the ISI proceedings in (Clarke and Hogan, 2005). I thank Rebecca for her permission to give related details in Sections 8.1 and 8.2 of this book.

It goes without saying that I owe this presentation to the many students whom I have taught over the years, who have helped query, question, and correct the many earlier versions of this work. I owe much to the late Alex Robertson of Murdoch University and the subsequent senior academic in mathematics and statistics, Ian James, for careful checking and wording of a number of the problems, some of which found their way into mathematics and statistics examination papers. I also owe thanks to David Schibeci and Russell John, who helped to collate some of the early Latex versions of this work, which I have since updated. In addition, I would like to acknowledge the hospitality of the University of Western Australia Mathematics and Statistics Department, where in 1998–1999 and 2004 I spent a total of 18 months preparing this book.

B.R.C.

NOTATION

$\forall$	for every
$\Leftrightarrow$	equivalent
$\Rightarrow$	implies
$\| \cdot \|$	Euclidean norm
$\ll \cdot \gg$	linear space spanned by a vector or vectors
$\otimes$	Kronecker product
e	identity operator
$E[X]$	expectation of random variable X

$\dim\{U\}$	dimension of the space U
$\mathcal{K}(\boldsymbol{A})$	kernel of the linear transformation: $L_{\boldsymbol{A}}: \boldsymbol{x} \rightarrow \boldsymbol{A}\boldsymbol{x}$
$\mathcal{M}(\boldsymbol{A})$	space spanned by the columns of matrix $\boldsymbol{A}$
$N(\mu, \Sigma)$	multivariate normal distribution with mean μ and covariance matrix Σ
$\boldsymbol{p}$	projection operator
$\mathbb{R}^n$	Euclidean n-space
$\mathcal{R}(\boldsymbol{A})$	range of the linear transformation: $L_{\boldsymbol{A}}: \boldsymbol{x} \rightarrow \boldsymbol{A}\boldsymbol{x}$
S_{xy}	$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$
S^2	ANOVA sample estimate of residual or error variance
$\mathcal{S}^2$	true sum of squared unobserved residuals
$\mathcal{S}_\Omega$	error or residual sum of squares under the full model Ω
$\mathcal{S}_\omega$	error or residual sum of squares under a submodel ω
$U^\perp$	orthogonal complement of U
$U_1 \dot{+} U_2$	direct sum of spaces U_1, U_2 where $U_1 \cap U_2 = 0$
$U_1 \oplus U_2$	direct sum of two orthogonal spaces $U1, U_2$
$\eta = E[Y]$	vector whose elements are expectations of elements of Y
$\sum_{i=1}^{n} x_i$	$x_1 + x_2 + \ldots + x_n$

CHAPTER 1

INTRODUCTION

The objective of this chapter is to provide a formal definition of the linear model in its basic form and to illustrate, using examples that should be familiar to the interested reader of statistics, the motivation behind the use of this form of the linear model. Representation of the linear model in its vector matrix form is a unifying feature of these examples, which include both regression and factorial models usually associated with the analysis of variance. By considering the objectives of fitting and testing and writing down the confidence intervals for parameters for one of the simpler regression models, we show that the non-vector matrix approach soon becomes unwieldy and even quite complicated. On the other hand, the vector matrix approach requires some initial algebra, which is described in detail in Chapter 2. Along with some easy-to-follow theory dealing with distributions, this lends itself to a straightforward analysis of the regression model. All of this is then used to embark on a description of models commonly associated with the analysis of variance. Embracing the common approach to both types of models helps clear some of the mystery associated with analysis of variance, in particular by providing some understanding as to how the usual degrees of freedom and distributions for component sums of squares are actually derived. In the current chapter we introduce the notation and terminology

Linear Models: The Theory and Application of Analysis of Variance. By Brenton R. Clarke

used as a springboard for the rest of the book.

1.1 THE LINEAR MODEL AND EXAMPLES

The linear model embraces a large section of the statistical literature, having been studied seriously ever since the times of Gauss (1777–1855) and Legendre (1752–1833), and comes in a variety of forms. For our purposes up to Chapter 8, we consider the linear model to be expressible in the general form

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1.1}$$

where $\boldsymbol{Y}$ represents an $n \times 1$ vector of dependent observations, $\boldsymbol{X}$ is an $n \times k$ design matrix, $\boldsymbol{\beta}$ is a $k \times 1$ parameter vector, and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of unobserved residuals with the property that the expectation or mean value of each component in the vector is zero; this is expressed in vector notation as $E[\boldsymbol{\epsilon}] = 0$. That is, observations in the vector $\boldsymbol{Y}$ are scattered about their mean. Since the expectation operator is a linear operator, we can write in vector notation

$$E[\boldsymbol{Y}] = \boldsymbol{X}\boldsymbol{\beta} \quad (\equiv \boldsymbol{\eta}, \text{ say}). \tag{1.2}$$

In the following, the linear form (1.1), together with the assumption(s) made about the unobserved residuals $\boldsymbol{\epsilon}$, will be denoted by Ω. To demonstrate the variety of forms that this linear model (1.1) can include, we consider several examples. The model (1.1) is generalized in Chapter 8 to incorporate models with variance components, including *random effects*, but for the moment, we discuss *fixed-effects models*, which are numerous and of some importance. The first example is one of the simplest and introduces some basic terminology.

■ **EXAMPLE 1.1**

The simplest example of the linear model is provided by a sample of n independent observations from a univariate normal distribution with mean μ and variance σ^2, denoted from now on by $N(\mu, \sigma^2)$. The model for these data can be represented in the form (1.1) by setting the design matrix $\boldsymbol{X} = \mathbf{1}_n$, the $n \times 1$ column of 1's, and $\boldsymbol{\beta} = \mu$, so that

$$\begin{bmatrix} Y_1 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} \mu + \begin{bmatrix} \epsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_n \end{bmatrix}.$$

The vector $\boldsymbol{\epsilon}$ is $N(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$, where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix (i.e., $\boldsymbol{\epsilon}$ has a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\sigma^2 \boldsymbol{I}_n$). □

Note: A vector $\boldsymbol{Z} = [Z_1, \ldots, Z_n]'$ that follows a multivariate normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$ and variance-covariance matrix $\boldsymbol{\Sigma}$, with elements σ_{ij}, is such that each component $Z_i \sim N(\mu_i, \sigma_{ii})$ and, moreover, $\text{cov}(Z_i, Z_j) = \sigma_{ij}$. If $\sigma_{ij} = 0$ for all $i \neq j$ $(i = 1, \ldots, n;\ j = 1, \ldots, n)$, the Z_i's are independent. Further description of a multivariate normal distribution is given briefly in Chapter 4: in particular in definition 4.1. It is also noted here that the assumption of joint normality of the component variables $\{Z_i\}_{i=1}^n$ is required in order that one may presume independence of the Z_i's based on knowledge of zero covariances for the off-diagonal elements in the covariance matrix. For more discussion of this, see Broffitt (1986), for example.

The following example involves the modeling of a straight-line relationship between two variables restricted so that the line passes through the origin. This is easy to understand and quite common in practice. This example is used in Section 1.2 to discuss the objectives of fitting a linear model. The particular model here is simple enough to make "least squares estimation and testing" by the direct approach to fitting feasible, but is by no means elementary.

■ EXAMPLE 1.2

The period T for a swing of the pendulum of length ℓ is given by

$$T = 2\pi\sqrt{\frac{\ell}{g}}.$$

If $\beta = 2\pi/\sqrt{g}$ and $x = \sqrt{\ell}$, then

$$T = x\beta.$$

If one actually makes observations of T in the form of Y such that one can make an accurate measurement of ℓ, and hence of x, but the measurement of T is subject to error, due for example, to the reaction time of a person clicking a stopwatch, then one is only entitled to write

$$Y = x\beta + \epsilon,$$

where ϵ is the error of measurement of T (Figure 1.1). Suppose that measurements of Y are taken for different x, so that

$$Y_i = x_i\beta + \epsilon_i; \quad i = 1, \ldots, n.$$

Suppose that the ϵ_i can be described as independent normal errors $N(0, \sigma^2)$, and interest is in the pairs (Y_i, x_i). The model can be represented as

$$\Omega: \begin{bmatrix} Y_1 \\ . \\ . \\ . \\ Y_n \end{bmatrix} = \begin{bmatrix} x_1 \\ . \\ . \\ . \\ x_n \end{bmatrix} \beta + \begin{bmatrix} \epsilon_1 \\ . \\ . \\ . \\ \epsilon_n \end{bmatrix}$$

with $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \boldsymbol{I_n})$. □

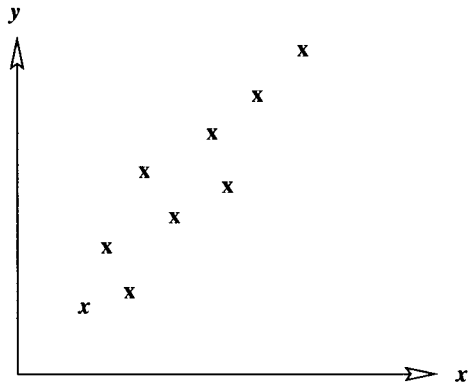

Figure 1.1 Possible scatterplot of periods of a pendulum versus the square root of a length of string.

The next example is, in fact, a simple extension of Example 1.2, again where one is contemplating the fit of a straight-line relationship between two variables, although there is no restriction that the line must pass through the origin. It is commonly referred to as a *simple linear regression.*

■ EXAMPLE 1.3

Linear Regression on One Variable. Suppose that one has a sample $Y_1, \ldots, Y_n$, where each Y_i is normal, with mean $\alpha+\beta x_i$, and variance σ^2, where x_i is known and α, β, and σ^2 are unknown parameters to be estimated (x_i nonrandom). Typical examples might be where Y is the weight of a baby at a certain age x_i, or where Y is a measure of chemical response in an experiment performed at temperature x_i. The model $\Omega : Y_i \sim N(\alpha + \beta x_i, \sigma^2)$ is represented in the form (1.1) by choosing the design matrix

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{bmatrix} \qquad (\equiv [\mathbf{1}_n, \boldsymbol{x}], \text{ say})$$

and the parameter vector $\boldsymbol{\beta} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$. Most frequently, one is interested in testing the hypothesis $H : \beta = 0$, and x is often referred to as a *concomitant variable.* In some examples, $x_1, \ldots, x_n$ are also random variables, but then one considers the conditional variation of Y_i given x_i. In this way the x's can be regarded as nonrandom. Examples include n married couples and

Y_i = age of wife
x_i = age of husband,

or n recordings of economic indices and

Y_i = price index in week i
x_i = wages index in week i.

□

A more general example introduces what is known as a *multiple linear regression model.* Here one has, for each observation of the response variable, several values that can be inputs to that response. This model generalizes the simple linear regression and has wide application in practice. See Problem 3-6 as a particular example involving two input variables.

■ EXAMPLE 1.4

Linear Regression on k Variables.; Suppose that one has n individuals and that for each a dependent variable Y and k concomitant variables, $x_1, \ldots, x_k$ are recorded. Let the value for the ith individual be

$$(Y_i, x_{1i}, \ldots, x_{ki}).$$

Assume that $E[Y_i] = \alpha + \beta_1 x_{1i} + \cdots + \beta_k x_{ki}$, so that the linear model is represented by

$$\begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ \vdots & \vdots & & \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix} \quad \text{and} \quad \beta = \begin{bmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}. \tag{1.3}$$

Frequently, there is interest in a hypothesis of the form

$$H_j : \beta_j = 0, \ 1 \leq j \leq k.$$ □

One is not always interested in fitting straight-line relationships between two variables. Indeed, there are many times when one wants to fit a quadratic or even a cubic relationship through a scatterplot of the two variables. The following example encompasses both suggestions and is even more general.

■ **EXAMPLE 1.5**

Polynomial Regression (on one variable). In many physical examples a nonlinear relationship may exist between the dependent variable Y and the concomitant variable x. Also, many nonlinear relationships can be approximated by a polynomial expression. This model is written

$$E[Y_i] = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_k x_i^k,$$

or equivalently (1.1), with each $x_{ij} = x_i^j$ in the design matrix of (1.3). Again the hypotheses of interest are of the form

$$H_j: \beta_j = 0, \qquad 1 \le j \le k.$$

For example, suppose that we are contemplating at most a cubic relationship where $k = 3$, but want to consider the hypothesis of a quadratic relationship only. This can be formulated as the hypothesis $H_3: \beta_3 = 0$. If, on the other hand, we were contemplating comparing the model with a cubic against a model with a straight line, this could be contemplated by considering a combination of hypotheses $H_2: \beta_2 = 0$ and $H_3: \beta_3 = 0$. For instance, combining these would lead to a *null hypothesis* of $H_0: \beta_2 = \beta_3 = 0$. See Example 4.1 for further discussion about combining tests for parameters and model selection.□

Remark 1.1. In each example the assumptions about the distribution of $\boldsymbol{Y}$ are of two types:

(a) The distribution of $\boldsymbol{Y}$ about its expected value $E[\boldsymbol{Y}] = \boldsymbol{X\beta}$ ($\equiv \boldsymbol{\eta}$, say), or equivalently, about the distribution of the errors $\boldsymbol{\epsilon}$

(b) The form for $\boldsymbol{\eta}$, for both the model Ω and the hypothesis $H \equiv (H_{i_1}, \ldots, H_{i_q})$, which represents a combination of hypotheses $\{H_j\}_{j=i_1}^{iq}$, where indices $(i_1, \ldots, i_q) \subset (1, \ldots, k)$

Remark 1.2. In the following we denote by Ω both the model (1.1) and any assumptions made about it, usually in the form of the distribution of errors $\boldsymbol{\epsilon}$. If we are considering any particular hypothesis H, under assumptions of the model, we denote

$$\omega = \Omega \cap H,$$

meaning the set of assumptions obtained by imposing the assumptions of hypothesis H in addition to assumptions of Ω. For the most part we will be interested in the assumption that errors $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I_n})$, and consider mainly the form of $\boldsymbol{\eta}$ as in Remark 1.1(b). In the latter part of the book, methods for examining the plausibility of this assumption on the error structure are described briefly. A typical departure from the assumption would be in cases of *heteroscedasticity*, where not all the errors in the vector $\boldsymbol{\epsilon}$ have common variance σ^2.

The approach taken to discuss or compare the model with hypothesis H_j will be to make use of vector algebra. Reconsider Example 1.3. Under the model Ω,

$$\boldsymbol{\eta} = \alpha \mathbf{1} + \beta \boldsymbol{x}.$$

That is, η is some linear combination of $\mathbf{1}$ and x. The hypothesis $H: \beta = 0$ is that $\eta = \alpha\mathbf{1}$, whence η is now restricted to the subspace spanned by the vector $\mathbf{1}$. Use of the words *subspace* and *spanned* here implies a knowledge of vector spaces that is introduced in Chapter 2, although they should easily be interpreted in this straightforward illustration as the set of vectors that are a multiple of the vector $\mathbf{1}$, as opposed to the full space of vectors which are from the linear combination given above. The hypothesis in question here is that one is fitting a line with zero slope through the data. Should we reject the hypothesis for this model, we are saying essentially that it is meaningful to fit a straight line through the data which has nonzero slope. This is often referred to as saying that the regression is useful since it describes a linear relationship between two variables.

Another example that comes up frequently in elementary statistics courses is where one wants to test a hypothesis of equal means, where, say, one has two independent samples of observations. Each sample could be as in Example 1.1, but the mean of each sample may be different; for instance, we may let μ_1 be the mean of the first sample and μ_2 be the mean of the second sample. Traditionally, the test of equality of means is discussed in the context of a *two-independent-sample Student t-test for equality of means*. However, we may couch such an example in the linear model framework by letting n_1 be the size of the first sample and n_2 be the size of the second sample, and describing the model as in the following example.

■ EXAMPLE 1.6

Let observations $\boldsymbol{Y} = [Y_{11}, \ldots, Y_{1n_1}, Y_{21}, \ldots, Y_{2n_2}]'$, such that under Ω the expectation vector

$$\eta = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} \quad \text{where} \quad \begin{cases} \eta_{11} = \eta_{12} = \cdots = \eta_{1n_1} \ (= \mu_1, \text{ say}) \\ \eta_{21} = \eta_{22} = \cdots = \eta_{2n_2} \ (= \mu_2, \text{ say}) \end{cases}$$

$$H: \mu_1 = \mu_2 \ (= \mu, \text{ say}).$$

Then under Ω :

$$\eta = \mu_1 \begin{bmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \mu_2 \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

and under ω :

$$\eta = \mu \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}. \qquad \square$$