2
MODEL FITTING

2.1 Introduction

The transmission and reception of information involves a message, or *signal*, which is distorted by *noise*. It is sometimes useful to think of scientific data as measurements composed of signal and noise and to construct mathematical models incorporating both of these components. Often the signal is regarded as *deterministic* (i.e. non-random) and the noise as random. Therefore, a mathematical model of the data combining both signal and noise is probabilistic and it is called a statistical model.

Another way of thinking of a statistical model is to consider the signal component as a mathematical description of the main features of the data and the noise component as all those characteristics not 'explained' by the model (i.e. by its signal component).

Our goal is to extract from the data as much information as possible about the signal as it is defined by the model. Typically the mathematical description of the signal involves several unknown constants, termed *parameters*. The first step is to estimate values for these parameters from the data.

Once the signal component has been quantified we can partition the total variability observed in the data into a portion attributable to the signal and the remainder attributable to the noise. A criterion for a good model is one which 'explains' a large proportion of this variability, i.e. one in which the part attributable to signal is large relative to the part attributable to noise. In practice, this has to be balanced against other criteria such as simplicity. Occam's Razor suggests that a parsimonious model which describes the data adequately may be preferable to a complicated one which leaves little of the variability 'unexplained'.

In many situations we wish to test hypotheses about the parameters. This can be performed in the context of model fitting by defining a series of different models corresponding to different hypotheses. Then the question about whether the data support a particular hypothesis can be formulated in terms of the adequacy of fit of the corresponding model (i.e. the amount of variability it explains) relative to other models.

These ideas are now illustrated by two detailed examples.

A. J. Dobson, *Introduction to Statistical Modelling*

2.2 Plant growth example

Suppose that genetically similar seeds are randomly assigned to be raised either in a nutritionally enriched environment (treatment) or under standard conditions (control) using a *completely randomized experimental design.* After a predetermined period all plants are harvested, dried and weighed. The results, expressed as dried weight in grams, for samples of 10 plants from each environment are given in Table 2.1. Fig. 2.1 shows the distributions of these weights.

Table 2.1 Plant weights from two different growing conditions.

Control (1)	4.17	5.58	5.18	6.11	4.50	4.61	5.17	4.53	5.33	5.14
Treatment (2)	4.81	4.17	4.41	3.59	5.87	3.83	6.03	4.89	4.32	4.69

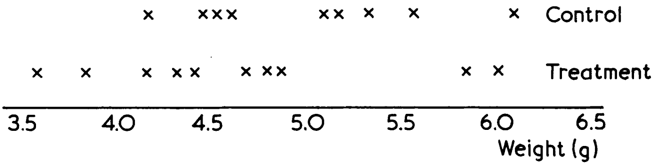

Figure 2.1 Plant growth data from Table 2.1.

The first step is to formulate models to describe these data, for example

$$Y_{jk} = \mu_j + e_{jk}, \tag{2.1}$$

where

(i) Y_{jk} is the weight of the kth plant ($k = 1, \ldots, K$ with $K = 10$ in this case) from the jth sample ($j = 1$ for control and $j = 2$ for treatment);

(ii) μ_j is a parameter, the signal component of weight, determined by the growth environment. It represents a common characteristic of all plants grown under the conditions experienced by sample j;

(iii) e_{jk} is the noise component. It is a random variable (although by convention it is usually written using the lower case). It is sometimes called the *random error term.* It represents that element of weight unique to the kth observation from sample j.

From the design of the experiment we assume that the e_{jk}s are independent and identically distributed with the Normal distribution with mean zero and variance σ^2, i.e. $e_{jk} \sim \text{NID}(0, \sigma^2)$ and therefore $Y_{jk} \sim \text{NID}(\mu_j, \sigma^2)$ for all j and k.

We would like to know if the enriched environment made a difference to the weight of the plants so we need to estimate the difference between μ_1 and μ_2 and test whether it differs significantly from some pre-specified value (such as zero).

An alternative specification of the model which is more suitable for comparative use is

$$Y_{jk} = \mu + \alpha_j + e_{jk} \tag{2.2}$$

where

(i) Y_{jk} and e_{jk} are defined as before;

(ii) μ is a parameter representing that aspect of growth common to both environments; and

(iii) α_1 and α_2 are parameters representing the differential effects due to the control or treatment conditions; formally $\alpha_j = \mu_j - \mu$ for $j = 1, 2$.

If the nutritionally enriched conditions do not enhance (or inhibit) plant growth, then the terms α_j will be negligible and so the model (2.2) will be equivalent to

$$Y_{jk} = \mu + e_{jk}. \tag{2.3}$$

Therefore, testing the hypothesis that there is no difference in weight due to the different environments (i.e. $\mu_1 = \mu_2$ or equivalently $\alpha_1 = \alpha_2 = 0$) is equivalent to comparing the adequacy of (2.1) and (2.3) for describing the data.

The next step is to estimate the model parameters. We will do this using the *likelihood function* which is the same as the joint probability density function of the response variables Y_{jk} but viewed primarily as a function of the parameters, conditional on the observations. *Maximum likelihood estimators* are the estimators which correspond to the maximum value of the likelihood function or, equivalently, its logarithm which is called the *log-likelihood function.*

We begin by estimating parameters μ_1 and μ_2 in (2.1) treating σ^2 as a known constant (in this context σ^2 is often referred to as a *nuisance parameter*). Since the Y_{jk}s are independent, the likelihood function is

$$\prod_{j=1}^{2}\prod_{k=1}^{K}\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}}\exp\left\{-\frac{1}{2\sigma^2}(y_{jk}-\mu_j)^2\right\}$$

and the log-likelihood function is

$$l_1 = -K\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=1}^{2}\sum_{k=1}^{K}(y_{jk}-\mu_j)^2,$$

so the maximum likelihood estimators of μ_1 and μ_2 are given by the solutions of

$$\frac{\partial l_1}{\partial \mu_j} = \frac{1}{\sigma^2}\sum_{k=1}^{K}(y_{jk}-\mu_j) = 0, \quad j = 1, 2$$

i.e. $$\hat{\mu}_j = \frac{1}{K}\sum_{k=1}^{K} y_{jk} = \frac{1}{K}y_{j\cdot} = \bar{y}_j \quad \text{for } j = 1, 2.$$

By considering the second derivatives it can be verified that $\hat{\mu}_1$ and $\hat{\mu}_2$ correspond to the maximum of l_1. Let

$$\hat{l}_1 = -K\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\hat{S}_1$$

denote the maximum of l_1 where $\hat{S}_1 = \sum_{j=1}^{2}\sum_{k=1}^{K}(y_{jk}-\bar{y}_j)^2$.

For the model given by (2.3) the likelihood function is

$$\prod_{j=1}^{2}\prod_{k=1}^{K}\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}}\exp\left\{-\frac{1}{2\sigma^2}(y_{jk}-\mu)^2\right\}$$

since $Y_{jk} \sim \mathrm{NID}(\mu, \sigma^2)$ for $j = 1, 2$ and $k = 1, \ldots, K$. Therefore the log-likelihood function is

$$l_0 = -K\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=1}^{2}\sum_{k=1}^{K}(y_{jk}-\mu)^2,$$

and so the estimator $\hat{\mu}$ obtained from the solution of $\partial l_0/\partial\mu = 0$ is

$$\hat{\mu} = \frac{1}{2K}\sum_{j=1}^{2}\sum_{k=1}^{K}y_{jk} = \frac{1}{2K}y_{..} = \bar{y} = \tfrac{1}{2}(\bar{y}_1+\bar{y}_2).$$

Hence the maximum of l_0 is

$$\hat{l}_0 = -K\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\hat{S}_0$$

where

$$\hat{S}_0 = \sum_{j=1}^{2}\sum_{k=1}^{K}(y_{jk}-\bar{y})^2.$$

For the plant data the values of the maximum likelihood estimates and the statistics $\hat{S}_1$ and $\hat{S}_0$ are shown in Table 2.2.

Table 2.2 Analysis of plant growth data in Table 2.1.

Model (2.1):	$\hat{\mu}_1 = 5.032, \hat{\mu}_2 = 4.661$ and $\hat{S}_1 = 8.729$
Model (2.3):	$\hat{\mu} = 4.8465$ and $\hat{S}_0 = 9.417$

The third step in the model fitting procedure involves testing hypotheses. If the null hypothesis $\mathrm{H}_0: \mu_1 = \mu_2$ is correct then the models (2.1) and (2.3) are the same so the maximum values $\hat{l}_1$ and $\hat{l}_0$ of the log-likelihood functions should be nearly equal, or equivalently, $\hat{S}_1$ and $\hat{S}_0$ should be nearly equal. If the data support this hypothesis, we would feel justified in using the simpler model (2.3) to describe the data. On the other hand, if the more general hypothesis $\mathrm{H}_1: \mu_1$ and μ_2 are not necessarily equal, is true then $\hat{S}_0$ should be larger than $\hat{S}_1$ (corresponding to $\hat{l}_0$ smaller than $\hat{l}_1$) and the model given by (2.1) would be preferable.

To assess the relative magnitude of $\hat{S}_1$ and $\hat{S}_0$ we need to consider the sampling distributions of the corresponding random variables

$$S_1 = \sum_{j=1}^{2}\sum_{k=1}^{K}(Y_{jk}-\bar{Y}_j)^2 \quad\text{and}\quad S_0 = \sum_{j=1}^{2}\sum_{k=1}^{K}(Y_{jk}-\bar{Y})^2.$$

It can be shown that

$$\frac{1}{\sigma^2}S_1 = \frac{1}{\sigma^2}\sum_{j=1}^{2}\sum_{k=1}^{K}(Y_{jk}-\bar{Y}_j)^2 = \frac{1}{\sigma^2}\sum_{j=1}^{2}\sum_{k=1}^{K}(Y_{jk}-\mu_j)^2 - \frac{K}{\sigma^2}\sum_{j=1}^{2}(\bar{Y}_j-\mu_j)^2.$$

For the more general model (2.1) we assume that $Y_{jk} \sim \text{NID}(\mu_j, \sigma^2)$ and so $\bar{Y}_j \sim \text{NID}(\mu_j, \sigma^2/K)$. Therefore (S_1/σ^2) is the difference between the sum of the squares of $2K$ independent random variables $(Y_{jk}-\mu_j)/\sigma$ which each has the distribution N(0, 1) and the sum of two independent random variables $(\bar{Y}_j-\mu_j)/(\sigma^2/K)^{\frac{1}{2}}$ which also have the N(0, 1) distribution. Hence, from definition (1.1),

$$\frac{1}{\sigma^2}S_1 \sim \chi^2_{2K-2}.$$

Similarly for the simpler model (2.3), let $\bar{\mu} = \frac{1}{2}(\mu_1+\mu_2)$ then

$$\begin{aligned}\frac{1}{\sigma^2}S_0 &= \frac{1}{\sigma^2}\sum_{j=1}^{2}\sum_{k=1}^{K}(Y_{jk}-\bar{Y})^2\\ &= \frac{1}{\sigma^2}\sum_{j=1}^{2}\sum_{k=1}^{K}(Y_{jk}-\bar{\mu}) - \frac{2K}{\sigma^2}(\bar{Y}-\bar{\mu})^2.\end{aligned}$$

If $Y_{jk} \sim \text{NID}(\mu_j, \sigma^2)$ then $\bar{Y} \sim \text{N}(\bar{\mu}, \sigma^2/2K)$. Also if $\mu_1 = \mu_2 = \bar{\mu}$ (corresponding to H_0) then the first term of (S_0/σ^2) is the sum of the squares of $2K$ independent random variables $(Y_{jk}-\bar{\mu})/\sigma \sim \text{N}(0, 1)$ and therefore

$$\frac{1}{\sigma^2}S_0 \sim \chi^2_{2K-1}.$$

However, if μ_1 and μ_2 are not necessarily equal (corresponding to H_1) then $(Y_{jk}-\bar{\mu})/\sigma \sim \text{N}(\mu_j-\bar{\mu}, 1)$ so that (S_0/σ^2) has a non-central chi-squared distribution with $2K-1$ degrees of freedom.

The statistic S_0-S_1 represents the difference in fit between the two models. If $\text{H}_0\colon \mu_1 = \mu_2$, is correct then

$$\frac{1}{\sigma^2}(S_0-S_1) \sim \chi^2_1;$$

otherwise it has a non-central chi-squared distribution. However, since σ^2 is unknown we cannot compare S_0-S_1 directly with the χ^2_1 distribution. Instead we eliminate σ^2 by using the ratio of $(S_0-S_1)/\sigma^2$ and the central chi-squared random variable (S_1/σ^2), each divided by its degrees of freedom, i.e.

$$f=\frac{1}{\sigma^2}\frac{(S_0-S_1)}{1}\Big/\frac{1}{\sigma^2}\frac{S_1}{2K-2}=\frac{S_0-S_1}{S_1/(2K-2)}.$$

If H_0 is correct, by definition (1.4), f has the central F-distribution with 1 and $(2K-2)$ degrees of freedom; otherwise f has a non-central F-distribution and so it is likely to be larger than predicted by $F_{1,2K-2}$.

For the plant weight data,

$$f=\frac{9.417-8.729}{8.729/18}=1.42$$

which is not statistically significant when compared with the $F_{1,18}$ distribution. Thus the data provide no evidence against H_0 so we conclude that there is probably no difference in weight due to the different environmental conditions and we can use the simpler model (2.3) to describe the data.

The more conventional approach to testing H_0 against H_1 is to use a t-test, i.e. to calculate

$$T=\frac{\bar{Y}_1-\bar{Y}_2}{s\left(\frac{1}{K}+\frac{1}{K}\right)^{\frac{1}{2}}}$$

where s^2, the pooled variance, is

$$s^2=\frac{1}{2K-2}\sum_{j=1}^{2}\sum_{k=1}^{K}(Y_{jk}-\bar{Y}_j)^2=\frac{1}{2K-2}S_1.$$

If H_0 is correct the statistic T has the distribution t_{2K-2}. The relationship between the test statistics T and f is obtained as follows:

$$T^2=\frac{(\bar{Y}_1-\bar{Y}_2)^2}{2s^2/K}=\frac{K(\bar{Y}_1-\bar{Y}_2)^2}{2S_1/(2K-2)},$$

but

$$S_0-S_1=\sum_{j=1}^{2}\sum_{k=1}^{K}[(Y_{jk}-\bar{Y})^2-(Y_{jk}-\bar{Y}_j)^2]$$

$$=\tfrac{1}{2}K(\bar{Y}_1-\bar{Y}_2)^2$$

so that

$$T^2=\frac{S_0-S_1}{S_1/(2K-2)}=f$$

corresponding to the distributional relationship that if $T\sim t_n$ then $T^2\sim F_{1,n}$ (see (1.5)).

The advantages of using an F-test instead of a t-test are:

(i) it can be generalized to test the equality of more than two means;

(ii) it is more closely related to the general methods considered in this book which involve comparing statistics that measure the 'goodness of fit' of competing models.

2.3 Birthweight example

The data in Table 2.3 are the birthweights (g) and estimated gestational ages (weeks) of 12 male and female babies born in a certain hospital. The mean ages are almost the same for both sexes but the mean birthweight for males is higher than for females. The data are plotted in Fig. 2.2; they suggest a linear trend of birthweight increasing with gestational age. The question of interest is whether the rate of increase is the same for males and females.

Table 2.3 Birthweight and gestational age for male and female babies

	Male		Female	
	Age (weeks)	Birthweight (g)	Age (weeks)	Birthweight (g)
	40	2968	40	3317
	38	2795	36	2729
	40	3163	40	2935
	35	2925	38	2754
	36	2625	42	3210
	37	2847	39	2817
	41	3292	40	3126
	40	3473	37	2539
	37	2628	36	2412
	38	3176	38	2991
	40	3421	39	2875
	38	2975	40	3231
Means	38.33	3024.00	38.75	2911.33

A fairly general statistical model for these data is

$$Y_{jk} = \alpha_j + \beta_j x_{jk} + e_{jk}, \tag{2.4}$$

where

(i) the response Y_{jk} is the birthweight for the kth baby of sex j where $j = 1$ for males, $j = 2$ for females and $k = 1, \ldots, K = 12$;

(ii) the parameters α_1 and α_2 represent the intercepts of the lines for the two sexes;

(iii) the parameters β_1 and β_2 represent the slopes or rates of increase for the two sexes;

(iv) the independent variable x_{jk} is the age of the (j, k)th baby (it is not a random variable);

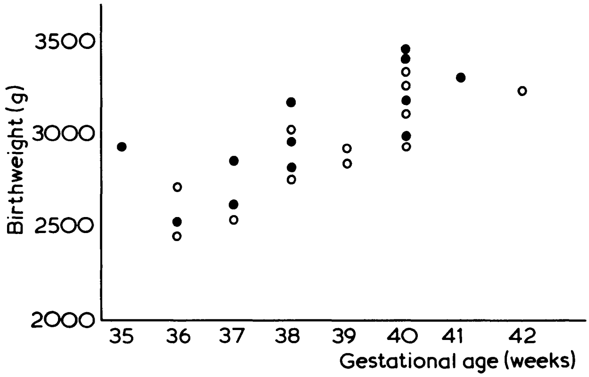

Figure 2.2 Birthweight and gestational age for male and female babies. ○, male; ●, female.

(v) the random error term is e_{jk}; we assume that $e_{jk} \sim \text{NID}(0, \sigma^2)$ for all j and k.

If the rate of increase is the same for males and females then the simpler model

$$Y_{jk} = \alpha_j + \beta x_{jk} + e_{jk} \qquad (2.5)$$

is appropriate, where the one parameter β in (2.5) corresponds to the two parameters β_1 and β_2 in (2.4). Thus we can test the null hypothesis

$$\text{H}_0: \beta_1 = \beta_2 \; (= \beta)$$

against the more general hypothesis

$$\text{H}_1: \beta_1 \text{ and } \beta_2 \text{ not necessarily equal,}$$

by comparing how well the models (2.4) and (2.5) fit the data.

The next step in the modelling process is to estimate the parameters. For this example we will use the *method of least squares* instead of the method of maximum likelihood. It consists of minimizing the sum of squares of the differences between the responses and their expected values. For the model (2.4) $E(Y_{jk}) = \alpha_j + \beta_j x_{jk}$ because we assumed that $E(e_{jk}) = 0$ so

$$S = \sum_j \sum_k (Y_{jk} - \alpha_j - \beta_j x_{jk})^2.$$

Geometrically, S is the sum of squares of the vertical distances from the points (x_{jk}, y_{jk}) to the line $y = \alpha_j + \beta_j x$ (Fig. 2.3). Algebraically it is the sum of squares of the error terms,

$$S = \sum_j \sum_k e_{jk}^2.$$

Estimators derived by minimizing S are called *least squares estimators* and the minimum value of S is a measure of the fit of the model. An advantage

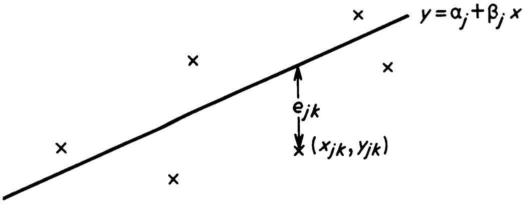

Figure 2.3 Distance from a point (x_{jk}, y_{jk}) to the line $y = \alpha_j + \beta_j x$.

of this method is that it does not require detailed assumptions about the distribution of the error terms (for example, that they are Normally distributed). However, such assumptions are required in order to compare minimum values of S obtained from different models.

Firstly, for (2.4)

$$S_1 = \sum_{j=1}^{2} \sum_{k=1}^{K} (Y_{jk} - \alpha_j - \beta_j x_{jk})^2,$$

so the least squares estimators for the parameters are the solutions of

$$\frac{\partial S_1}{\partial \alpha_j} = -2 \sum_{k=1}^{K} (Y_{jk} - \alpha_j - \beta_j x_{jk}) = 0,$$

$$\frac{\partial S_1}{\partial \beta_j} = -2 \sum_{k=1}^{K} x_{jk}(Y_{jk} - \alpha_j - \beta_j x_{jk}) = 0, \quad \text{for } j = 1, 2.$$

These equations can be simplified to the form

$$\left.\begin{aligned} \sum_{k=1}^{K} Y_{jk} - K\alpha_j - \beta_j \sum_{k=1}^{K} x_{jk} &= 0 \\ \sum_{k=1}^{K} x_{jk} Y_{jk} - \alpha_j \sum_{k=1}^{K} x_{jk} - \beta_j \sum_{k=1}^{K} x_{jk}^2 &= 0 \end{aligned}\right\} j = 1, 2$$

when they are called the *normal equations*. The solutions are

$$b_j = \frac{K \sum_k x_{jk} y_{jk} - \left(\sum_k x_{jk}\right)\left(\sum_k y_{jk}\right)}{K \sum_k x_{jk}^2 - \left(\sum_k x_{jk}\right)^2}$$

$$a_j = \bar{y}_j - b_j \bar{x}_j$$

for $j = 1, 2$. Then the value for S_1 is

$$\hat{S}_1 = \sum_{j=1}^{2} \sum_{k=1}^{K} (y_{jk} - a_j - b_j x_{jk})^2.$$

Secondly, for (2.5)

$$S_0 = \sum_{j=1}^{2} \sum_{k=1}^{K} (Y_{jk} - \alpha_j - \beta x_{jk})^2,$$

so the least squares estimators are the solutions of

$$\frac{\partial S_0}{\partial \alpha_j} = -2\sum_k (Y_{jk} - \alpha_j - \beta x_{jk}) = 0, \qquad j = 1, 2,$$

and

$$\frac{\partial S_0}{\partial \beta} = -2\sum_j\sum_k x_{jk}(Y_{jk} - \alpha_j - \beta x_{jk}) = 0.$$

Hence

$$b = \frac{K\sum_j\sum_k x_{jk}y_{jk} - \sum_j\left(\sum_k x_{jk}\sum_k y_{jk}\right)}{K\sum_j\sum_k x_{jk}^2 - \sum_j\left(\sum_k x_{jk}\right)^2}$$

and $a_j = \bar{y}_j - b\bar{x}_j$.

For the birthweight example, the data are summarized in Table 2.4 (summation is over $k = 1, \ldots, K$ with $K = 12$). The least squares estimates for both models are given in Table 2.5.

Table 2.4 Summary of birthweight data in Table 2.3.

	Male, $j = 1$	Female, $j = 2$
Σx	460	465
Σy	36288	34936
Σx^2	17672	18055
Σy^2	110623496	102575468
Σxy	1395370	1358497

Table 2.5 Analysis of birthweight data in Table 2.3.

Model (2.4):	$b_1 = 111.983,$	$a_1 = -1268.672,$	
	$b_2 = 130.400,$	$a_2 = -2141.667,$	$\hat{S}_1 = 652424.5$
Model (2.5):	$b = 120.894,$	$a_1 = -1610.283,$	
		$a_2 = -1773.322,$	$\hat{S}_0 = 658770.8$

To test the hypothesis $\mathrm{H}_0: \beta_1 = \beta_2$, i.e. to compare the models given by (2.4) and (2.5), we need to know the sampling distribution of the minimum of the sum of squares, S. By analogous arguments to those used in the previous example, it can be shown that $(S_1/\sigma^2) \sim \chi^2_{20}$ and if H_0 is correct then $(S_0/\sigma^2) \sim \chi^2_{21}$. In each case the number of degrees of freedom is the number of observations minus the number of parameters estimated. The improvement in fit for (2.4) compared with (2.5) is

$$\frac{1}{\sigma^2}(S_0 - S_1),$$

which can be compared with fit of the more detailed model (2.4), i.e. with (S_1/σ^2) using the test statistic

$$f = \frac{(S_0 - S_1)/1}{S_1/(2K-4)}.$$

If the hypothesis H_0 is correct, $f \sim F_{1, 2K-4}$. For these data the value of f is 0.2 which is certainly not statistically significant, so the data provide no evidence against the hypothesis $\beta_1 = \beta_2$ and we have reason for preferring the simpler model given by (2.5).

2.4 Notation for linear models

The models considered in the above examples can be written in matrix notation in the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \qquad (2.6)$$

where

(i) $\mathbf{y}$ is a vector of responses,
(ii) $\boldsymbol{\beta}$ is a vector of parameters,
(iii) $\mathbf{X}$ is a matrix whose elements are zeros or ones or values of 'independent' variables, and
(iv) $\mathbf{e}$ is a vector of random error terms.

For quantitative explanatory variables (e.g. age in the birthweight example) the model contains terms of the form βx where the parameter β represents the rate of change in the response corresponding to changes in the independent variable x.

For qualitative explanatory variables there is a parameter to represent each level of a factor (e.g. the effects due to environmental conditions in the plant growth example). The corresponding elements of $\mathbf{X}$ are chosen to exclude or include the appropriate parameters for each observation; they are called *dummy variables* (if only zeros and ones are used for $\mathbf{X}$ the term *indicator variable* is used).

Example 2.1

For the plant growth example the more general model was

$$Y_{jk} = \mu_j + e_{jk}; \qquad j = 1, 2 \text{ and } k = 1, \ldots, K.$$

The corresponding elements of (2.6) are

$$\mathbf{y}=\begin{bmatrix} Y_{11}\\ Y_{12}\\ \vdots\\ Y_{1K}\\ Y_{21}\\ \vdots\\ Y_{2K}\end{bmatrix},\qquad \boldsymbol{\beta}=\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix},\qquad \mathbf{X}=\begin{bmatrix}1 & 0\\ 1 & 0\\ \vdots & \vdots\\ 1 & 0\\ 0 & 1\\ \vdots & \vdots\\ 0 & 1\end{bmatrix}\qquad\text{and}\qquad \mathbf{e}=\begin{bmatrix} e_{11}\\ e_{12}\\ \vdots\\ e_{1K}\\ e_{21}\\ \vdots\\ e_{2K}\end{bmatrix}.$$

Example 2.2

For the simpler plant growth model

$$Y_{jk}=\mu+e_{jk},\qquad j=1,2 \text{ and } k=1,\ldots,K$$

so

$$\mathbf{y}=\begin{bmatrix} Y_{11}\\ Y_{12}\\ \vdots\\ Y_{1K}\\ Y_{21}\\ \vdots\\ Y_{2K}\end{bmatrix},\qquad \boldsymbol{\beta}=[\mu],\qquad \mathbf{X}=\begin{bmatrix}1\\ 1\\ \\ \vdots\\ \\ 1\end{bmatrix}\qquad\text{and}\qquad \mathbf{e}=\begin{bmatrix} e_{11}\\ e_{12}\\ \vdots\\ e_{1K}\\ e_{21}\\ \vdots\\ e_{2K}\end{bmatrix}.$$

Example 2.3

For the model

$$Y_{jk}=\alpha_j+\beta_j x_{jk}+e_{jk};\quad j=1,2 \text{ and } k=1,\ldots,K$$

for birthweight the corresponding matrix and vector terms are

$$\mathbf{y}=\begin{bmatrix} Y_{11}\\ Y_{12}\\ \vdots\\ Y_{1K}\\ Y_{21}\\ \vdots\\ Y_{2K}\end{bmatrix},\qquad \boldsymbol{\beta}=\begin{bmatrix}\alpha_1\\ \alpha_2\\ \beta_1\\ \beta_2\end{bmatrix},\qquad \mathbf{X}=\begin{bmatrix}1 & 0 & x_{11} & 0\\ 1 & 0 & x_{12} & 0\\ \vdots & \vdots & \vdots & \vdots\\ 1 & 0 & x_{1K} & 0\\ 0 & 1 & 0 & x_{21}\\ \vdots & \vdots & \vdots & \vdots\\ 0 & 1 & 0 & x_{2K}\end{bmatrix}\qquad\text{and}\qquad \mathbf{e}=\begin{bmatrix} e_{11}\\ e_{12}\\ \vdots\\ e_{1K}\\ e_{21}\\ \vdots\\ e_{2K}\end{bmatrix}.$$

Models of the form $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{e}$ are called *linear models* because the signal part of the model, $\mathbf{X}\boldsymbol{\beta}$, is a linear combination of the parameters and the noise part, $\mathbf{e}$, is also additive. If there are p parameters in the model and N observations, then $\mathbf{y}$ and $\mathbf{e}$ are $N\times 1$ random vectors, $\boldsymbol{\beta}$ is a $p\times 1$ vector of parameters (usually to be estimated) and $\mathbf{X}$ is an $N\times p$ matrix of known constants.

2.5 Exercises

2.1 For the plant growth example verify that the least squares estimators for the parameters in (2.1) and (2.3) are the same as the maximum likelihood estimators.

2.2 Write the equations

$$Y_{jkl} = \mu + \alpha_j + \beta_j x_{jk} + \gamma x_{jk}^2 + e_{jkl},$$

where $j = 1, \ldots, J$, $k = 1, \ldots, K$ and $l = 1, \ldots, L$ in matrix notation. [Hint: form a new independent variable $t_{jk} = x_{jk}^2$.]

2.3 The weights (kg) of 10 people before and after going on a high carbohydrate diet for 3 months are shown in Table 2.6. You want to know if, overall, there was any significant change in weight.

Table 2.6 Weights (kg) of people before and after a diet.

Before	64	71	64	69	76	53	52	72	79	68
After	61	72	63	67	72	49	54	72	74	66

Let Y_{jk} denote the weight of the kth person at time j where $j = 1$ before the diet, $j = 2$ afterwards and $k = 1, \ldots, 10$.

(i) Use the models (2.1) and (2.3) from the plant growth example to test the hypothesis H_0 that there was no change in weight, i.e. $E(Y_{1k}) = E(Y_{2k})$ for all k.

(ii) Let $D_k = Y_{1k} - Y_{2k}$ for $k = 1, \ldots, 10$. If H_0 is true then $E(D_k) = 0$, so another test of H_0 is to compare the models

$$D_k = \mu + e_k \tag{2.7}$$

and

$$D_k = e_k, \tag{2.8}$$

assuming that $e_k \sim \text{NID}(0, \sigma^2)$ for $k = 1, \ldots, 10$ (where μ in (2.7) is not necessarily zero). Use the method of maximum likelihood to estimate μ and compare the values of the likelihood function under (2.7) and (2.8) to test H_0.

(iii) List all the assumptions you made for the analyses in (i) and (ii). Which analysis was more appropriate?