CHAPTER 12

Models Containing Functions of the Predictors, Including Polynomial Models

The general regression methods formulated for linear models in Chapters 5–11 are valid whatever (linear in the parameters) form the model function takes. Thus the model function can in theory be set up in any way one wishes. We discuss this first in a general way. Subsequently, we examine a set of data to which a second-order model is fitted. This example provides motivation for some suggested rules for testing and omitting terms from polynomial models.

12.1. MORE COMPLICATED MODEL FUNCTIONS

Many models fitted in practice involve just the observed predictor variables X_1, $X_2, \ldots, X_k$, say, in their original form; that is, we fit

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon. \tag{12.1.1}$$

Many more general forms are possible. We can write the most general type of linear model in variables $X_1, X_2, \ldots, X_k$ in the form

$$Y = \beta_0 Z_0 + \beta_1 Z_1 + \beta_2 Z_2 + \cdots + \beta_{p-1} Z_{p-1} + \epsilon. \tag{12.1.2}$$

($Z_0 = 1$ is a dummy variable that is always unity and will in general not be shown. However, it is sometimes mathematically convenient to have a Z_0 in the model. For example, if

$$(Z_{1i}, Z_{2i}, \ldots, Z_{p-1}) \qquad i = 1, 2, \ldots, n$$

are n settings of the variables Z_j, $j = 1, 2, \ldots, p-1$, corresponding to observations Y_i, $i = 1, 2, \ldots, n$, then when $j \neq 0$, and $Z_{0i} = 1$,

$$\sum_{i=1}^{n} Z_{ji} = \sum_{i=1}^{n} Z_{ji} Z_{0i}$$

and thus can be represented by the general cross-product expression

$$\sum_{i=1}^{n} Z_{ji} Z_{li}$$

if the normal equations are written out. Note that $\Sigma_{i=1}^{n} Z_{0i}^2 = n$.)

In (12.1.2), each Z_j, $j = 1, 2, \ldots, p-1$, is a general function of $X_1, X_2, \ldots, X_k$,

$$Z_j = Z_j(X_1, X_2, \ldots, X_k),$$

and can take any form. In some examples, each Z_j may involve only one X-variable.

Any model that can be written, perhaps after rearrangement or transformation, in the form of Eq. (12.1.2) can be analyzed by the general methods given in Chapters 5–11. We now provide some specific examples of models that can be treated by these methods and relate them to the general form of Eq. (12.1.2).

Polynomial Models of Various Orders in the X_j

First-Order Models

1. If $p = 2$ and $Z_1 = X$ in Eq. (12.1.2), we obtain the simple first-order model with one predictor variable:

$$Y = \beta_0 + \beta_1 X + \epsilon. \tag{12.1.3}$$

2. If $p = k + 1$ and $Z_j = X_j$, we obtain a first-order model with k predictor variables:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon. \tag{12.1.4}$$

Second-Order Models

1. If $p = 3$, $Z_1 = X$, $Z_2 = X^2$, and $\beta_2 = \beta_{11}$, we obtain a second-order (quadratic) model with one predictor variable:

$$Y = \beta_0 + \beta_1 X + \beta_{11} X^2 + \epsilon. \tag{12.1.5}$$

2. If $p = 6$, $Z_1 = X_1$, $Z_2 = X_2$, $Z_3 = X_1^2$, $Z_4 = X_2^2$, $Z_5 = X_1 X_2$, $\beta_3 = \beta_{11}$, $\beta_4 = \beta_{22}$, and $\beta_5 = \beta_{12}$, we obtain a second-order model with two predictor variables:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 + \epsilon. \tag{12.1.6}$$

A full second-order model in k variables can be obtained in similar fashion when $p = 1 + k + k + +\frac{1}{2}k(k-1) = \frac{1}{2}(k+1)(k+2)$. Second-order models are used particularly in response surface studies where it is desired to graduate, or approximate to, the characteristics of some unknown response surface by a polynomial of low order. Note that all possible second-order terms are in the model. This is sensible because omission of terms implies possession of definite knowledge that certain types of surface (those that cannot be represented *without* the omitted terms) cannot possibly occur. Knowledge of this sort is not often available. When it is, it would usually enable a more theoretically based study to be made.

An example of a second-order response surface analysis is given in Section 12.2.

Third-Order Models

1. If $p = 4$, $Z_1 = X$, $Z_2 = X^2$, $Z_3 = X^3$, $\beta_2 = \beta_{11}$, and $\beta_3 = \beta_{111}$, we obtain a third-order model with one predictor variable:

$$Y = \beta_0 + \beta_1 X + \beta_{11} X^2 + \beta_{111} X^3 + \epsilon. \tag{12.1.7}$$

2. If $p = 10$ and proper identification of the β_i and Z_i is made (we omit the details now since the examples above should have made the idea clear), the model (12.1.2) can represent a third-order model with two predictor variables given by

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{12} X_1 X_2 + \beta_{22} X_2^2 + \beta_{111} X_1^3 + \beta_{112} X_1^2 X_2 + \beta_{122} X_1 X_2^2 + \beta_{222} X_2^3 + \epsilon. \tag{12.1.8}$$

The general third-order model for k factors $X_1, X_2, \ldots, X_k$ can be obtained similarly. Third-order models are also used in response surface work though much less frequently than second-order models. Note the method of labeling the β's. This may seem confusing at first but it is done to enable the coefficients to be readily associated with their corresponding powers of the X's. For example, $X_1 X_2^2 = X_1 X_2 X_2$ has a coefficient β_{122}, and so on. A similar notation is used above for second-order models and is standard in response surface work. (Note that we write β_{122} with lower-valued subscripts first rather than β_{212} or β_{221}.)

Models of *any* desired order can be represented by Eq. (12.1.2) by continuing the process illustrated above.

Transformations

If a second-order model is not adequate, a third-order model may be. However, one should not routinely add higher-order terms. It is often more fruitful to investigate the effects produced by other transformations of the predictor variables, or by transformations of the response variable, or by both. The same remark also applies in the first-order versus second-order decision. For example, a straight-line fit of the response log Y versus X, if appropriate, would usually be preferred to a quadratic fit of Y versus X, assuming the behavior of the residuals showed that either fit was a workable choice.

Models Involving Transformations Other than Integer Powers

The polynomial models above involved powers, and cross-products of powers, of the predictor variables $X_1, X_2, \ldots, X_k$. Here we provide a few examples of other types of transformations that are often useful in forming the model function.

The Reciprocal Transformation. If in Eq. (12.1.2) we take $p = 2$, $Z_1 = 1/X_1$, and $Z_2 = 1/X_2$, we obtain the model

$$Y = \beta_0 + \beta_1 \left(\frac{1}{X_1}\right) + \beta_2 \left(\frac{1}{X_2}\right) + \epsilon. \tag{12.1.9}$$

The Logarithmic Transformation. By taking $p = 2$, $Z_1 = \ln X_1$, and $Z_2 = \ln X_2$, Eq. (12.1.2) can represent

$$Y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \epsilon. \tag{12.1.10}$$

The Square Root Transformation. For example,

$$Y = \beta_0 + \beta_1 X_1^{1/2} + \beta_2 X_2^{1/2} + \epsilon. \tag{12.1.11}$$

Clearly, there are many possible transformations, and models can be postulated that contain few or many such terms. *Several different transformations could occur in the same model, of course.* [The examples (12.1.9)–(12.1.11) use the same transformation throughout.] The choice of what, if any, transformation to make is often difficult to decide. The choice would often be made on the basis of previous knowledge of the variables under study. The purpose of making transformations of this type is to be able to use a regression model of simple form in the transformed variables, rather than a more complicated one in the original variables.

Transformations could also involve several X_j variables simultaneously, for example, $Z_1 = X_1^{1/2} \ln X_2$. Transformations of this type are sometimes suggested by the form of the fitted equation in untransformed variables. When the best power of an X to use is not known, a parameter can be substituted. In such cases, nonlinear estimation methods are usually needed.

Plots Can Be Useful

Suitable transformations of the predictor variables are also sometimes suggested by plotting the data in various ways. See, for example, Hoerl (1954). Other references are Dolby (1963) and Tukey (1957).

Our discussion above relates entirely to choosing the model function. (The response Y is untouched.) When, as we assume here, the predictor variables are not subject to error, there are no problems in transforming them.

Transformations on the response are discussed in Chapter 13. For those, one must be especially careful to check that the least squares assumptions [errors independent, $N(0, \sigma^2)$] are not violated by making the transformation. This is usually done by checking the residuals in the *transformed metric,* after the transformation has been chosen and the model fitted.

Using Ratios as Responses and/or Predictors

When ratios with a common component are used in regression analyses, there is a danger that a strong regression relationship will be introduced spuriously by the component. Warnings have appeared in the literature for about a century, but the problems still recur. A useful discussion of possible problems is given by Kronmal (1993). Kronmal's conclusions and recommendations section (his pp. 390–391) begins: "The message of this paper is that ratio variables should only be used in the context of a full linear model in which the variables that make up the ratio are included and the intercept term is also present. The common practice of using ratios for either the [response] or the [predictor] variables in regression analyses can lead to misleading inferences and rarely in any gain." Kronmal then provides some suggestions for responding to researchers who do not wish to "give up their most prized ratio or index."

12.2. WORKED EXAMPLES OF SECOND-ORDER SURFACE FITTING FOR $k = 3$ AND $k = 2$ PREDICTOR VARIABLES

Aia, Goldsmith, and Mooney (1961) reported a pilot plant investigation under the title "Predicting Stoichiometric $CaHPO_4{\cdot}2H_2O$." This section is adapted from that paper with the permission of the American Chemical Society. We omit the chemical details here and also make several minor changes to their original analysis.

In the problem studied, there were seven candidates for predictor variables but four of these were kept fixed throughout the experiment. The three selected for the response surface study and their chosen ranges were as follows:

Variable	Designation	Range Chosen
Mole ratio $NH_3/CaCl_2$ in the calcium chloride solution	r	0.70–1.00
Addition time in minutes of ammoniacal $CaCl_2$ to $NH_4H_2PO_4$	t	10–90
Starting pH of $NH_4H_2PO_4$ solution	pH	2–5

TABLE 12.1. A Worked Example: The X Matrix and Two Responses

		Design Matrix										
u	1	X_1	X_2	X_3	X_1^2	X_2^2	X_3^2	X_1X_2	X_1X_3	X_2X_3	Y	Y_4
1	1	−1	−1	−1	1	1	1	1	1	1	52.8	6.95
2	1	1	−1	−1	1	1	1	−1	−1	1	67.9	5.90
3	1	−1	1	−1	1	1	1	−1	1	−1	55.4	7.10
4	1	1	1	−1	1	1	1	1	−1	−1	64.2	7.08
5	1	−1	−1	1	1	1	1	1	−1	−1	75.1	5.64
6	1	1	−1	1	1	1	1	−1	1	−1	81.6	5.18
7	1	−1	1	1	1	1	1	−1	−1	1	73.8	6.84
8	1	1	1	1	1	1	1	1	1	1	79.5	5.67
9	1	$-\frac{5}{3}$	0	0	$\frac{25}{9}$	0	0	0	0	0	68.1	6.00
10	1	$\frac{5}{3}$	0	0	$\frac{25}{9}$	0	0	0	0	0	91.2	5.67
11	1	0	$-\frac{5}{3}$	0	0	$\frac{25}{9}$	0	0	0	0	80.6	5.52
12	1	0	$\frac{5}{3}$	0	0	$\frac{25}{9}$	0	0	0	0	77.5	6.47
13	1	0	0	$-\frac{5}{3}$	0	0	$\frac{25}{9}$	0	0	0	36.8	7.17
14	1	0	0	$\frac{5}{3}$	0	0	$\frac{25}{9}$	0	0	0	78.0	5.36
15	1	0	0	0	0	0	0	0	0	0	74.6	6.48
16	1	0	0	0	0	0	0	0	0	0	75.9	5.91
17	1	0	0	0	0	0	0	0	0	0	76.9	6.39
18	1	0	0	0	0	0	0	0	0	0	72.3	5.99
19	1	0	0	0	0	0	0	0	0	0	75.9	5.86
20	1	0	0	0	0	0	0	0	0	0	79.8	5.96

There were seven responses of interest. For each response, the idea was tentatively entertained that it could be graduated by a second-order function of r, t, and pH. We make use here only of the first response (which we call Y) and record the fourth (which we call Y_4) for use in an exercise.

The experimental design chosen was a "cube plus star plus six center points" composite design with $\alpha = \frac{5}{3} = 1.667$. (For "rotatability," α should have the value $2^{3/4} = 1.6818$ so that the selected design is nearly, but not quite, rotatable. The authors ignored this difference and in fact used $\alpha = 1.6818$ in their calculations. For this reason our calculations are slightly different from theirs.)

The design requires five levels of each variable. The experimental variables were coded by the transformations

$$X_1 = (r - 0.85)/0.09, \qquad X_2 = (t - 50)/24, \qquad X_3 = (\text{pH} - 3.5)/0.9. \quad (12.2.1)$$

Thus the equivalence of the design and actual levels can be expressed by the following table.

Coded Levels, X_1 or X_2 or X_3	Actual Levels		
	r	t	pH
$\frac{5}{3}$	1.00	90	5.0
1	0.94	74	4.4
0	0.85	50	3.5
−1	0.76	26	2.6
$-\frac{5}{3}$	0.70	10	2.0

The actual design in coded units is shown as the indicated part of the **X** matrix in Table 12.1. In coded units the postulated second-order model can be written

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{33} X_3^2 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 \tag{12.2.2}$$

for the response Y, and similarly for Y_4.

The runs were performed in random order and the observed values of the seven responses were recorded. Two of these,

$$Y = \text{yield as percentage of theoretical yield},$$
$$Y_4 = \text{bulk density in grams per cubic inch},$$

are given in Table 12.1. The appropriate $\mathbf{X'X}$ matrix is the same no matter which response is being fitted and has the form

$$\mathbf{X'X} = \begin{bmatrix} N & 0 & 0 & 0 & B & B & B & 0 & 0 & 0 \\ 0 & B & 0 & 0 & & & & & & \\ 0 & 0 & B & 0 & & \mathbf{0} & & & \mathbf{0} & \\ 0 & 0 & 0 & B & & & & & & \\ B & & & & C & D & D & & & \\ B & & \mathbf{0} & & D & C & D & & \mathbf{0} & \\ B & & & & D & D & C & & & \\ 0 & & & & & & & D & 0 & 0 \\ 0 & & \mathbf{0} & & & \mathbf{0} & & 0 & D & 0 \\ 0 & & & & & & & 0 & 0 & D \end{bmatrix}, \tag{12.2.3}$$

where, for our specific example,

$$\begin{aligned} N &= 20, \\ B &= 8 + 2\alpha^2 = \tfrac{122}{9}, \\ C &= 8 + 2\alpha^4 = \tfrac{1898}{81}, \\ D &= 8. \end{aligned} \tag{12.2.4}$$

This type of partitioned matrix occurs frequently in planned second-order response surface studies, and its inverse is easily obtained. In the general case, when there are k factors (rather than three) and the same symbols are used in a larger $\mathbf{X'X}$ matrix in the obvious way, the inverse can be written

$$
(\mathbf{X}'\mathbf{X})^{-1} =
\begin{array}{c|cccc|ccc|ccc|c}
 & 0 & 1 & 2 & k & 11 & 22 & kk & 12 & 13 & (k-1)k & \\
 & P & 0 & 0 & 0 & Q & Q & Q & 0 & 0 & 0 & 0 \\
\hline
0 & 1/B & 0 & 0 & 0 & & & & & & & 1 \\
0 & 0 & 1/B & 0 & 0 & & & & & & & 2 \\
\vdots & \vdots & & & & & \mathbf{0} & & & \mathbf{0} & & \vdots \\
0 & 0 & 0 & 0 & 1/B & & & & & & & k \\
\hline
Q & & & & & R & S & S & & & & 11 \\
Q & & & & & S & R & S & & & & 22 \\
\vdots & & \mathbf{0} & & & \vdots & & & & \mathbf{0} & & \vdots \\
Q & & & & & S & S & R & & & & kk \\
\hline
0 & & & & & & & & 1/D & 0 & 0 & 12 \\
0 & & & & & & & & 0 & 1/D & 0 & 13 \\
\vdots & & \mathbf{0} & & & & \mathbf{0} & & \vdots & & & \vdots \\
0 & & & & & & & & 0 & 0 & 1/D & (k-1)k
\end{array}
\tag{12.2.5}
$$

The values of P, Q, R, and S are shown in Table 12.2, in the second column marked $C \neq 3D$. [The values in the third column are the simplified forms when $C = 3D$, which happens when the design is "rotatable," that is, when the contours of $V\{\hat{Y}(X)\}$ are spherical ones. In such circumstances, the design can be rotated in the predictor (i.e., the X-) space without affecting the precision of the information obtained.]

In our case, $k = 3$ and $3D = \frac{1944}{81}$ so that $C - 3D$ is small but not zero as it would be for a rotatable design. Consequently, we obtain

$$
\begin{aligned}
P &= 1597/9614, & Q &= -549/9614, \\
R &= 685.3248/9614, & S &= 62.3376/9614.
\end{aligned}
\tag{12.2.6}
$$

Also, $1/B = \frac{9}{122}$, $1/D = \frac{1}{8}$. (The figures are exact; to avoid round-off error later, the final division by 9614 has been postponed until after the subsequent matrix multiplication.) We now need $\mathbf{X}'\mathbf{Y}$. These are given below for the responses Y and Y_4.

TABLE 12.2. Formulas for Obtaining Elements of $(\mathbf{X}'\mathbf{X})^{-1}$

Symbol	Value when $C \neq 3D$	Value when $C = 3D$ (Rotatable Design)
P	$(C - D)(C + (k-1)D)/A$	$2(k+2)D^2/A$
Q	$-(C - D)B/A$	$-2DB/A$
R	$\{N(C + (k-2)D) - (k-1)B^2)/A$	$\{N(k+1)D - (k-1)B^2)/A$
S	$(B^2 - ND)/A$	$(B^2 - ND)/A$
A	$(C - D)\{N(C + (k-1)D) - kB^2\}$	$2D\{N(k+2)D - kB^2\}$

Note that A occurs in the formulas for P, Q, R, and S.

$$\mathbf{X'Y} = \tfrac{1}{9}\begin{bmatrix} 12{,}941.1 \\ 671.4 \\ -87.0 \\ 1{,}245.3 \\ 8{,}935.2 \\ 8{,}905.2 \\ 7{,}822.7 \\ -63.9 \\ -105.3 \\ -20.7 \end{bmatrix}, \qquad \mathbf{X'Y}_4 = \tfrac{1}{9}\begin{bmatrix} 1{,}108.26 \\ -29.25 \\ 41.43 \\ -60.45 \\ 744.99 \\ 752.99 \\ 766.49 \\ 2.88 \\ -5.04 \\ 3.24 \end{bmatrix}. \tag{12.2.7}$$

Applying the usual formula $\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$ for the estimates of the regression parameters for the first response Y, we obtain the fitted equation

$$\begin{aligned} \hat{Y} = {} & 76.022 + 5.503X_1 - 0.713X_2 + 10.207X_3 \\ & + 0.712X_1^2 + 0.496X_2^2 - 7.298X_3^2 \\ & - 0.888X_1X_2 - 1.463X_1X_3 - 0.288X_2X_3. \end{aligned} \tag{12.2.8}$$

When the second-order design gives rise to an $\mathbf{X'X}$ matrix of the type shown

TABLE 12.3. Standard Analysis of Variance Table for Certain Types of Second-Order Designs

Source	df	SS
b_0 (mean)	1	$\left(\sum_{u=1}^{N} Y_u\right)^2/N$
b_i (first order)	k	$\sum_{i=1}^{k} b_i(iY)$
$b_{ii}\|b_0$ (pure second-order, given b_0)	k	$b_0(0Y) + \sum_{i=1}^{k} b_{ii}(iiY) - \left(\sum_{u=1}^{N} Y_u\right)^2/N$
b_{ij} (mixed second-order)	$\frac{1}{2}k(k-1)$	$\sum_{i=1}^{k}\sum_{j=1}^{k} b_{ij}(ijY)$
Lack of fit	$N - n_e - \frac{1}{2}(k+1)(k+2)$	By subtraction
Pure error	n_e	By usual calculation
Total	N	$\sum_{u=1}^{N} Y_u^2$

inverted in (12.2.5) the analysis of variance table is as shown in Table 12.3. Here,

$$\begin{aligned}(0y) &= \sum_{u=1}^{N} Y_u,\\ (iy) &= \sum_{u=1}^{N} X_{iu}Y_u,\\ (iiy) &= \sum_{u=1}^{N} X_{iu}^2 Y_u,\\ (ijy) &= \sum_{u=1}^{N} X_{iu}X_{ju}Y_u,\end{aligned} \tag{12.2.9}$$

all of these expressions being cross-products of columns of the $\mathbf{X}$ matrix with the column $\mathbf{Y}$ of observations and so all are elements of the $\mathbf{X'Y}$ vector. Usually we would combine the $\text{SS}(b_{ii}|b_0)$ and $\text{SS}(b_{ij})$ to give a

$$\text{SS(second order terms}|b_0) \text{ with } \tfrac{1}{2}k(k+1) \text{ degrees of freedom}$$

but we have displayed them separately in the table to emphasize that the only *extra* sum of squares that arises is $\text{SS}(b_{ii}|b_0)$ due to the orthogonality of many pairs of columns in $\mathbf{X}$, a feature that is true only for specific design choices. We can now proceed, in the usual way, to test lack of fit and the usefulness of the second-order and first-order terms.

It should be noted that many of the special features of this least squares estimation and analysis apply *only* to designs whose $\mathbf{X'X}$ matrices take the special form given, so that $(\mathbf{X'X})^{-1}$ can be found from the formulas above. Designs without this feature must be subjected to the usual least squares analysis without recourse to special formulas. Nevertheless, the "source" column of the analysis of variance table given above provides a framework to aim at. The pure error sum of squares is obtained as usual, and the successive entries for sums of squares for parameter estimates would all be obtained as extra sums of squares as described in Chapter 6.

For our example, the appropriate analysis of variance table is given as Table 12.4. Since $F(5, 5, 0.95) = 5.05 > 3.04$, no lack of fit is indicated. We can recombine the lack of fit and pure error sums of squares and estimate $V(Y_i) = \sigma^2$ by

$$s^2 = (93.91 + 30.86)/(5 + 5) = 12.477.$$

Dividing this into the first-order mean square gives a ratio $609.93/12.477 = 48.88$, which exceeds $F(3, 10, 0.999) = 12.55$, while from the second-order mean square we obtain the ratio $135.59/12.477 = 10.88$, which exceeds $F(6, 10, 0.999) = 9.93$. Thus both first- and second-order terms appear to be needed in the fitted model.

TABLE 12.4. Analysis of Variance Table for the Fitted Model

Source	df	SS	MS	F
Mean (b_0)	1	103,377.82		
First-order	3	1,829.80	609.93	
Second-order $\|b_0$	6	813.54	135.59	
Lack of fit	5	93.91	18.78	3.04
Pure error	5	30.86	6.17	
Total	20	106,145.93		

TABLE 12.5. Analysis of Variance for the Reduced Second-Order Model in X_1 and X_3

Source	df	SS	MS	F
First-order	2	1822.91	911.46	89.80
Second-order $\|b_0$	3	803.12	267.71	26.38
Lack of fit	9	111.22	12.36	
	14	142.08	10.15	2.00
Pure error	5	30.86	6.17	
Total (corrected)	19	2768.11		

Do We Need X_2?

In the original paper, the authors noted the small size of all estimated coefficients with a subscript 2 compared with their standard errors and concluded that their model should not contain X_2 at all. When the situation is this clear-cut—all the coefficients being small compared with their standard errors—such a conclusion is unlikely to be wrong. However, the extra sum of squares principle should be applied in such situations in general and we apply it here to illustrate.

Suppose we wish to test the null hypothesis H_0: $\beta_2 = \beta_{22} = \beta_{12} = \beta_{23} = 0$ against the alternative hypothesis H_1 that at least one of these β's is not zero. The regression sum of squares for the full second-order model in X_1, X_2, and X_3 given b_0 is, from the analysis of variance table,

$$\begin{aligned} S_1 &= \text{SS(first-order terms)} + \text{SS(second-order terms}|b_0) \\ &= 1829.80 + 813.54 \\ &= 2643.34 \qquad (\text{with } 3 + 6 = 9 \text{ df}). \end{aligned}$$

Application of the hypothesis H_0 to the original model implies use of the reduced model

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_{11} X_1^2 + \beta_{33} X_3^2 + \beta_{13} X_1 X_3.$$

The appropriate $\mathbf{X}$ matrix can be obtained from Table 12.1 by deleting the X_2, X_2^2, X_1X_2, and X_2X_3 columns. The $\mathbf{X'X}$ matrix can be obtained from the previous one by deleting rows and columns corresponding to X_2, X_2^2, X_1X_2, and X_2X_3. The $\mathbf{X'Y}$ vector is obtained from the previous one by a similar row deletion. The fitted equation is thus

$$\hat{Y} = 76.420 + 5.503X_1 + 10.207X_3 + 0.667X_1^2 - 7.343X_3^2 - 1.463X_1X_3.$$

The regression sum of squares given b_0 for this reduced model is now required. We find it to be $S_2 = 2626.025$ (with 5 df). The extra sum of squares due to b_2, b_{22}, b_{12}, and b_{23} is therefore

$$\begin{aligned} S_1 - S_2 &= 2643.34 - 2626.03 \\ &= 17.31 \qquad (\text{with } 9 - 5 = 4 \text{ df}). \end{aligned}$$

This leads to a mean square of $17.31/4 = 4.33$, which can be compared with the residual mean square estimate of σ^2 from the original three-factor regression. The null hypothesis that $\beta_2 = \beta_{22} = \beta_{12} = \beta_{23} = 0$ cannot be rejected. Thus it seems sensible to adopt the reduced model, which does not involve terms in X_2. The analysis of variance table appropriate to this reduced fitted model is shown in Table 12.5. No

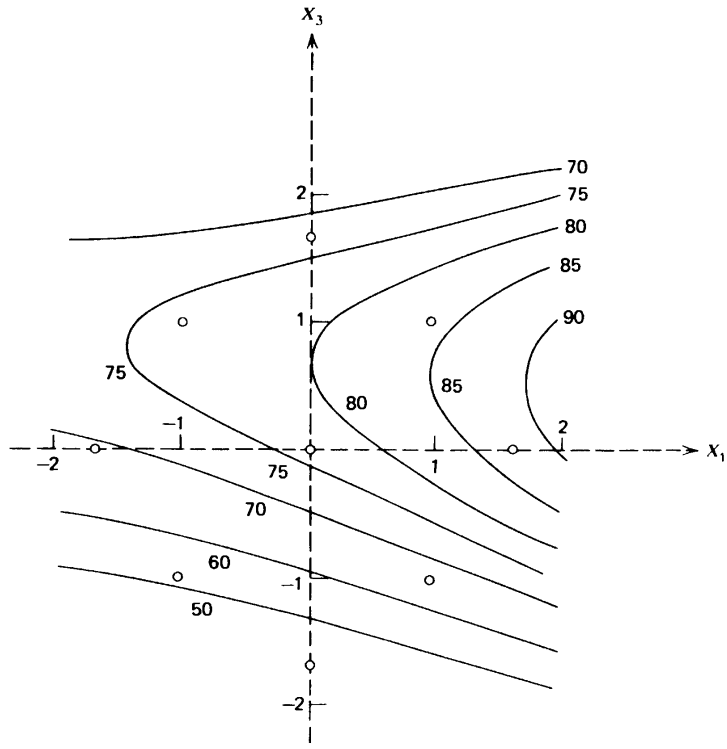

Figure 12.1. Contours of the fitted second-order equation relating response Y to variables X_1 and X_3.

lack of fit is shown and the regression is highly significant for both first- and second-order terms.

In order to examine a fitted second-order response surface, we would usually perform a "canonical analysis" in which the surface is described in terms of coordinates placed along its major axes. Such an analysis is extremely useful and enables the overall situation to be grasped even when many factors are involved. With only two factors, however, as here, we can plot the contours of $\hat{Y}$ directly by writing the fitted equation in the form

$$-7.343X_3^2 + (10.207 - 1.463X_1)X_3 + (0.667X_1^2 + 5.503X_1 + 76.420 - \hat{Y}) = 0.$$

If a value of $\hat{Y}$ is selected, the corresponding contour can be drawn by substituting values of X_1 and solving for X_3. Contours obtained in this way are shown in Figure 12.1. The experimental points are indicated on the diagram by dots. Repeat points are not distinguished, however, and must be obtained by looking at Table 12.6. The contours are those of a rising ridge. Examination of this contour system led the authors to hypothesize on the chemical reactions that could cause such contours. (Quite frequently, response surface investigations provide the initial step in a more fundamental, theoretical, investigation of the system under study.)

These contours may also be viewed in conjunction with the residuals, given in Table 12.6. A "pattern" plot of the residuals in which each residual is placed near its corresponding design point is shown in Figure 12.2. Of the twenty residuals, the six

TABLE 12.6. The Fitted Values and Residuals Obtained from Fitting a Second-Order Surface $\hat{Y} = f(X_1, X_3)$

u	X_1	X_3	Y	$\hat{Y}$	$e = Y - \hat{Y}$
1	−1	−1	52.8	52.57	0.23
2	1	−1	67.9	66.50	1.40
3	−1	−1	55.4	52.57	2.83
4	1	−1	64.2	66.50	−2.30
5	−1	1	75.1	75.91	−0.81
6	1	1	81.6	83.99	−2.39
7	−1	1	73.8	75.91	−2.11
8	1	1	79.5	83.99	−4.49
9	$-\frac{5}{3}$	0	68.1	69.10	−1.00
10	$\frac{5}{3}$	0	91.2	87.44	3.76
11	0	0	80.6	76.42	4.18
12	0	0	77.5	76.42	1.08
13	0	$-\frac{5}{3}$	36.8	39.01	−2.21
14	0	$\frac{5}{3}$	78.0	73.04	4.97
15	0	0	74.6	76.42	−1.82
16	0	0	75.9	76.42	−0.52
17	0	0	76.9	76.42	0.48
18	0	0	72.3	76.42	−4.12
19	0	0	75.9	76.42	−0.52
20	0	0	79.8	76.42	3.38

with the largest absolute values occur at points $(X_1, X_3) = (0, 0)$ (three); $(\frac{5}{3}, 0)$, $(0, \frac{5}{3})$, and (1, 1) (one each). Thus the model appears to fit least well in the first quadrant of the (X_1, X_3) plane and any conclusions that rely on the validity of the fitted surface in that region could be suspect. (What effect this might have on the authors' original conclusions is a matter for a chemical engineer, rather than a statistician to examine, and we avoid discussion of the point here.) Doubts of this kind can

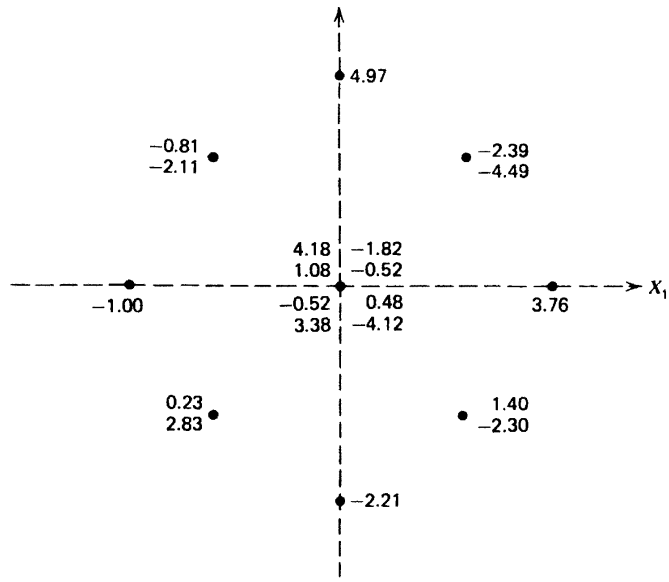

Figure 12.2. Pattern of residuals from the fitted second-order equation relating response Y to variables X_1 and X_3.

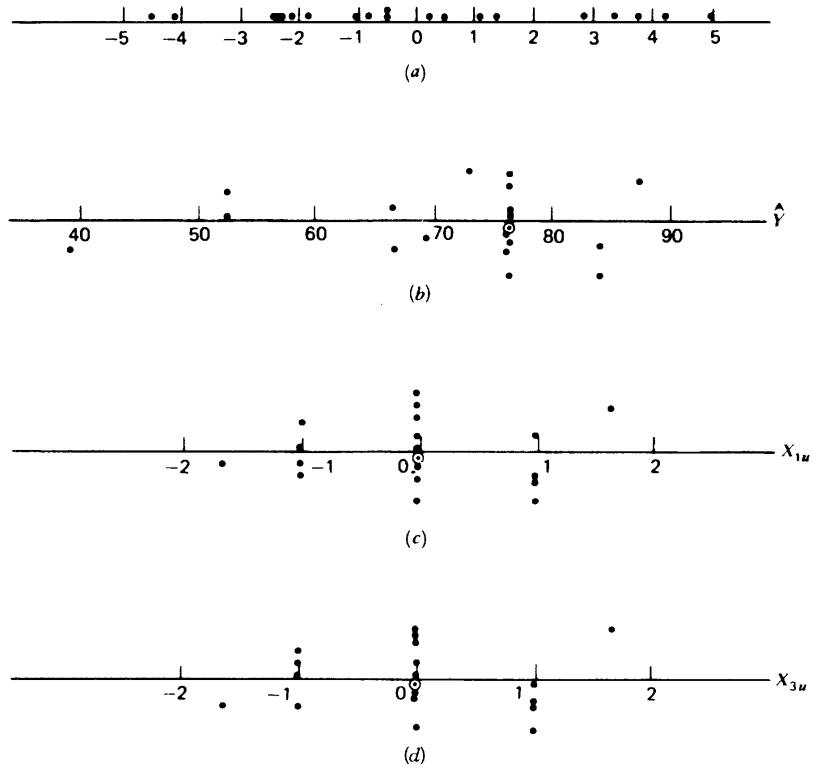

Figure 12.3. Standard plots of the residuals from the fitted second-order equation relating Y to X_1 and X_3: (*a*) overall, (*b*) against $\hat{Y}$, (*c*) against X_1, and (*d*) against X_3.

sometimes be resolved by doing further experimental runs in the region in which the shape of the fitted surface is suspect and refitting a suitable response function in that more limited region.

We can now examine the residuals in other ways to see if any abnormality is indicated. Figure 12.3 shows the following standard plots of residuals (*a*) overall, (*b*) against the fitted values $\hat{Y}_u$, (*c*) against X_{1u}, and (*d*) against X_{3u}.

The overall plot does not appear to deny the assumption of normality implicit in the testing of variance ratios in the analysis of variance. The plot against the $\hat{Y}_u$ exhibits "widening" tendencies at first sight but this is deceptive due to the fact that most of the residuals are large and the size of the residuals band is not well established at the lower end of the $\hat{Y}$ scale. Similar behavior occurs in the plots against X_{1u} and X_{3u}, where the size of the residuals band is not well defined at the extremes. Thus one cannot conclude in any of the plots that abnormality is indicated. It does not appear that the basic regression assumptions are unjustified therefore. (Note that, since we do not know the order in which the observations were taken, we are unable to check whether a time trend has affected the response.)

If this investigation were to be continued, additional efforts might involve attempts to account for the large first quadrant residuals by reexamining the original data and their relationships to other predictor variables whose variations have possibly not been considered. In this way, an improvement in the model might be possible. Also, or alternatively, the region in which the fit of the model is questionable could be examined in more detail, as previously suggested.

TABLE 12.6. The Fitted Values and Residuals Obtained from Fitting a Second-Order Surface $\hat{Y} = f(X_1, X_3)$

u	X_1	X_3	Y	$\hat{Y}$	$e = Y - \hat{Y}$
1	−1	−1	52.8	52.57	0.23
2	1	−1	67.9	66.50	1.40
3	−1	−1	55.4	52.57	2.83
4	1	−1	64.2	66.50	−2.30
5	−1	1	75.1	75.91	−0.81
6	1	1	81.6	83.99	−2.39
7	−1	1	73.8	75.91	−2.11
8	1	1	79.5	83.99	−4.49
9	$-\frac{5}{3}$	0	68.1	69.10	−1.00
10	$\frac{5}{3}$	0	91.2	87.44	3.76
11	0	0	80.6	76.42	4.18
12	0	0	77.5	76.42	1.08
13	0	$-\frac{5}{3}$	36.8	39.01	−2.21
14	0	$\frac{5}{3}$	78.0	73.04	4.97
15	0	0	74.6	76.42	−1.82
16	0	0	75.9	76.42	−0.52
17	0	0	76.9	76.42	0.48
18	0	0	72.3	76.42	−4.12
19	0	0	75.9	76.42	−0.52
20	0	0	79.8	76.42	3.38

Treatment of Pure Error When Factors Are Dropped

The foregoing analysis raises a question we have avoided until now. When a factor like X_2 is dropped from a model, should we reassess our treatment of pure error? In Table 12.1, runs 15–20 are the only repeats but, when X_2 is dropped so that the data become as in Table 12.6, the pairs of runs numbered (1, 3), (2, 4), (5, 7), and (6, 8) now apparently form four pairs of repeat runs in variables X_1 and X_3. Also, runs 11 and 12 are now apparently center points. Thus it could be argued that the design should now be considered as a replicated 2^2 factorial (eight points, 1–8) plus a two-factor star (four points, 9, 10, 13, 14) plus eight center points (11, 12, 15, . . . , 20). If this were done, Table 12.5 would have to be revised to show entries for

$$\text{Lack of fit SS} = 78.26\ (3\ \text{df}), \quad \text{MS} = 26.08;$$
$$\text{Pure error SS} = 63.82\ (11\ \text{df}), \quad \text{MS} = 5.80.$$

The consequent F-ratio is $4.50 > F(3, 11, 0.95) = 3.59$, leading to the rather surprising conclusion that there *is* lack of fit. Thus, in this revised analysis, while the variable X_2 appears to be unnecessary in the model, lack of fit is shown if we remove it! However, it is clear that X_2 does little to help explain the variation in the observations. In fact, the size of the pure error mean square is inflated if X_2 *is* used in the model, and, at the same time, the reduction in the degrees of freedom for pure error provides a less sensitive F-test for lack of fit.

Which analysis is correct? One could argue both sides of the issue. On the whole, however, we favor using the pure error as initially calculated before dropping of factors. Presumably, repeats in the original data *are* genuine repeats if so reported, but the same cannot be said of the runs, which look like repeats when a factor is dropped. Also, in many sets of data, the opposite problem may occur, that is, genuine lack of fit will be missed because the "new" repeats will show more variability than the genuine repeats.

A safe way to proceed would be to do the analysis both ways and see if they agree. For many sets of data, they will. If they do not, the data can be subjected to further scrutiny.

How should we proceed in the present example? The model must definitely be placed under suspicion. However, the first- and second-order terms account for a proportion $R^2 = (1822.91 + 803.12)/2768.11 = 0.949$ of the total (corrected for the mean) variation in the data, for the loss of only five degrees of freedom. (When the terms in X_2 are added, the figure rises only to 0.955.) In other words, the model is explaining 95% of the variation about the mean even though, technically, lack of fit is possible. By examining the fitted contours and the residuals together, as we have already done, we can discover where the lack of fit may exist. If the graduation of the true surface appears good over a large region of the X-space, conclusions obtained from the fitted model in that region may still be valid. Further examination of the residuals may also reveal if any of the basic regression assumptions (normality, constant variance, independence of observations) appear to be violated, or may suggest ways to revise the model.

Sometimes, in practical work, the pure error is "too small" simply because the pure error runs have not been randomized (or, at minimum, distributed) over the whole of an experiment. If several pure error runs are done consecutively or close together in time, there is a tendency for the responses to be more alike than they otherwise would be. In other words, the pure error in such a case would not be representative of the range of errors typically found throughout the experiment. This sometimes leads to false signals that lack of fit exists and needs to be investigated carefully. Repeated analyses of the same experimental run also rarely constitute true "repeat runs."

Comment. The example we have just discussed is somewhat unusual in one respect. When, for sound reasons, terms are dropped from a model, lack of fit does not usually appear in the reduced model unless peculiarities exist in the data. We have seen that these peculiarities do not appear to arise from violation of least squares assumptions. Their source remains a matter for speculation.

In a wider sense, this example is not unusual. While the experiment answered some questions, it left others unresolved. These questions become the subject of further conjectures for future work. In this respect, it is typical of much practical experimentation.

Treatment of Pure Error When a Design Is Blocked

The design of Table 12.1 was unblocked. Often, however, response surface designs are performed in blocks in such a way that the blocks are orthogonal to the model. Runs that would be repeat runs in an unblocked design are often divided among the blocks. In such a case, these runs are no longer repeat runs *unless they occur in the same block*, and the pure error must be calculated on that basis. Also, the analysis of variance must contain a sum of squares for blocks. For blocks orthogonal to the model, the appropriate sum of squares for blocks is usually

$$\text{SS(blocks)} = \sum_{w=1}^{m} \frac{B_w^2}{n_w} - \frac{G^2}{N}, \quad \text{with } (m-1) \text{ degrees of freedom,}$$

in the analysis of variance table, where B_w is the total of the n_w observations in the wth block (there are m blocks in all) and G is the grand total of all the observations

in all of the m blocks. When blocks are not orthogonal to the model, the extra sum of squares principle applies. (It can, of course, be applied in all cases, orthogonally blocked or not, and produces the answer given above in the former case.)

In some situations, even runs that occur in *different* blocks can be used to measure "almost pure" error, provided an appropriate estimate of the block difference(s) is (are) available. See, for example, Box and Draper (1987, p. 375).

On Dropping Terms

In our second-order polynomial example, we decided to remove all terms in X_2. This raises the more general question of which terms can be considered for removal without damaging the relationships between terms. (Might it have been reasonable to remove, for example, just the term in X_2? We answer this particular question with a resounding no.) For some further conversation and two suggested criteria, please see Section 12.3.

12.3. RETAINING TERMS IN POLYNOMIAL MODELS

We argue in this section that individual terms should not, in general, be dropped from a polynomial model of order two or more, unless the situation is carefully assessed. We also offer two criteria and two consequent rules, which make sense to us and which we recommend. An implication of all this is that to use an equation derived by allowing some mechanical selection procedure to pick a subset of terms from a polynomial model is risky.

Example 1. Quadratic Equation in X

Consider the fitted quadratic in a single variable X,

$$\hat{Y} = b_0 + b_1 X + b_{11} X^2. \tag{12.3.1}$$

The maximum or minimum of this quadratic occurs where the first derivative of $\hat{Y}$ with respect to X is zero. This is where $b_1 + 2b_{11}X = 0$, that is, at the location $X = -b_1/(2b_{11})$. If b_{11} is positive, we obtain a minimum; if negative, a maximum.

Let us consider the consequences of dropping just one term. (The equation must be refitted then, of course, to give new coefficients.)

(a) Drop b_0. The intercept at $X = 0$ was $\hat{Y} = b_0$. It will now be forced to be zero, in the refit.

(b) Drop b_1. The maximum or minimum was at $X = -b_1/(2b_{11})$. It will now be forced to be at zero in the refit.

(c) Drop b_{11}. We initially had a quadratic with intercept b_0. We now have a straight line in the refit.

Even if b_0 or b_1 is not statistically significant, actions (a) and (b) will usually produce distortion in the refit. If b_{11} is not significant, however, so that whatever curvature exists is small, the new straight line will tend to do a reasonable job in most cases.

We would thus recommend against dropping either b_0 or b_1 if b_{11} is retained, because this will force the fitting of a quadratic with a built-in restriction and bias. We would argue that dropping b_{11} is reasonable, if the quadratic slope is very slight, because the

resulting line will tend to fit well. (Otherwise b_{11} would have been large.) In higher dimensions (more X's), exactly the same sorts of considerations arise.

One way of thinking about the effects of the three different "drops" is to consider what happens to the reduced equations under a shift of origin. Suppose we let $Z = X - a$, corresponding to an origin shift from $X = 0$ to $X = a$. Then we get, by substituting $X = Z + a$,

(a) $b_1X + b_{11}X^2 = a(b_1 + ab_{11}) + a(b_1 + 2b_{11})Z + b_{11}Z^2$.

(b) $b_0' + b_{11}'X^2 = (b_0' + a^2b_{11}') + 2ab_{11}'Z + b_{11}'Z^2$.

(c) $b_0'' + b_1''X = (b_0'' + ab_{11}'') + b_{11}''Z$.

We see that only the third equation has retained its original form. The other two have reacquired the terms that were deleted! This gives a basis for the first criterion and rule. We also suggest a second criterion and rule, below. In proceeding, we now use the notation for the model function with β's rather than that for a fitted equation with b's, but that does not affect any of our points. So the general question is: Which reduced models are reasonable to adopt, and which are not?

Criterion 1. The Origin Shift Criterion

We shall consider a reduced model to be a sensible one (some say "well formulated") if a shift in the origin of the X-space produces a model of unchanged *form* in the new variables $Z_1, Z_2, \ldots, Z_k$.

Example 2. Second-Order Polynomial in Two X's

Consider the two-X second-order (quadratic) model

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_{11}X_1^2 + \beta_{22}X_2^2 + \beta_{12}X_1X_2 + \epsilon. \tag{12.3.2}$$

If we substitute $X_1 = Z_1 + a_1$, $X_2 = Z_2 + a_2$ into the full second-order model (12.3.2), corresponding to an origin shift to the point $(X_1, X_2) = (a_1, a_2)$, we get

$$\begin{aligned} Y &= \beta_0 + \beta_1(Z_1 + a_1) + \beta_2(Z_2 + a_2) \\ &\quad + \beta_{11}(Z_1 + a_1)^2 + \beta_{22}(Z_2 + a_2)^2 + \beta_{12}(Z_1 + a_1)(Z_2 + a_2) + \epsilon \qquad (12.3.3) \\ &= \alpha_0 + \alpha_1Z_1 + \alpha_2Z_2 + \alpha_{11}Z_1^2 + \alpha_{22}Z_2^2 + \alpha_{12}Z_1Z_2 + \epsilon, \qquad (12.3.4) \end{aligned}$$

where

$$\begin{aligned} \alpha_0 &= \beta_0 + \beta_1a_1 + \beta_2a_2 + \beta_{11}a_1^2 + \beta_{22}a_2^2 + \beta_{12}a_1a_2, \quad \text{a quadratic in } (a_1, a_2), \\ \alpha_1 &= \beta_1 + 2\beta_{11}a_1 + \beta_{12}a_2, \\ \alpha_2 &= \beta_2 + \beta_{12}a_1 + 2\beta_{22}a_2, \\ \alpha_{11} &= \beta_{11}, \\ \alpha_{22} &= \beta_{22}, \\ \alpha_{12} &= \beta_{12}. \end{aligned} \tag{12.3.5}$$

What terms can be omitted in (12.3.2) to produce a model of the same (reduced) form in (12.3.4)? Or, equivalently, what β's can be set to zero in (12.3.2) to give a model (12.3.4) of the same form? Obviously, only quadratic coefficients can be dropped. Setting $\beta_{11} = 0$, for example, forces $\alpha_{11} = 0$, but all other coefficients remain in both

models. (We ignore the relatively unlikely case where the a's and β's are such that one of the α's vanished identically. This would not occur for all origin moves, nor in most practical cases.)

The general truth, which can be seen via the specific example in (12.3.5), is that only the highest-order coefficients are unaffected by a shift in origin, while all lower-order coefficients become, after the shift, a combination of both lower- and higher-order ones. The consequent rule is the following.

Rule 1

If a model is to be consistent under a shift in origin, only the highest-order terms can be deleted at first and any chosen deletions must keep the model well formulated. Moreover, if any of the highest-order terms are retained, all terms of lower order affected by them in a shift of origin must also be retained, whether or not their estimates are significant in the regression fit.

Note: A model that lacks β_0 cannot be a well-formulated one under origin shift, in any circumstances.

Example 2. Continued

For example, suppose $\beta_{11} \neq 0$, $\beta_{22} = 0$, $\beta_{12} = 0$. Then [see (12.3.5)] $\beta_1 X_1$ must be retained, because α_1 depends on β_{11}; however, $\beta_2 X_2$ is a candidate for possible deletion because $\alpha_2 = \beta_2$ when $\beta_{22} = \beta_{12} = 0$.

Example 3. Third-Order Polynomial in Three Factors

To aid thinking about this in slightly wider contexts, we now give the equivalent of (12.3.5) for second-order (quadratic) and third-order (cubic) models in three factors (X_1, X_2, X_3), with an origin shift to (a_1, a_2, a_3). For the cubic,

$$\begin{aligned} Y = {} & \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \\ & + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{33} X_3^2 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 \\ & + \beta_{111} X_1^3 + \beta_{222} X_2^3 + \beta_{333} X_3^3 \\ & + \beta_{122} X_1 X_2^2 + \beta_{133} X_1 X_3^2 + \beta_{112} X_1^2 X_2 + \beta_{233} X_2 X_3^2 + \beta_{113} X_1^2 X_3 \\ & + \beta_{223} X_2^2 X_2 + \beta_{123} X_1 X_2 X_3 + \epsilon, \end{aligned} \tag{12.3.6}$$

we find that:

$$\begin{aligned} \alpha_0 &= \text{full cubic in } (a_1, a_2, a_3), \\ \alpha_1 &= \beta_1 + 2\beta_{11} a_1 + \beta_{12} a_2 + \beta_{13} a_3 + 3\beta_{111} a_1^2 + \beta_{122} a_2^2 + \beta_{133} a_3^2 \\ &\quad + 2\beta_{112} a_1 a_2 + 2\beta_{113} a_1 a_3 + \beta_{123} a_2 a_3, \\ \alpha_2 &= \beta_2 + \beta_{12} a_1 + 2\beta_{22} a_2 + \beta_{23} a_3 + \beta_{112} a_1^2 + 3\beta_{222} a_2^2 + \beta_{233} a_3^2 \\ &\quad + 2\beta_{122} a_1 a_2 + 2\beta_{223} a_2 a_3 + \beta_{123} a_1 a_3, \\ \alpha_3 &= \beta_3 + \beta_{13} a_1 + \beta_{23} a_2 + 2\beta_{33} a_3 + \beta_{113} a_1^2 + \beta_{223} a_2^2 + 3\beta_{333} a_3^2 \\ &\quad + 2\beta_{133} a_1 a_3 + 2\beta_{233} a_2 a_3 + \beta_{123} a_1 a_2, \\ \alpha_{11} &= \beta_{11} + 3\beta_{111} a_1 + \beta_{112} a_2 + \beta_{113} a_3, \end{aligned} \tag{12.3.7}$$

$$
\begin{aligned}
\alpha_{22} &= \beta_{22} + \beta_{122}a_1 + 3\beta_{222}a_2 + \beta_{223}a_3, \\
\alpha_{33} &= \beta_{33} + \beta_{133}a_1 + \beta_{233}a_2 + 3\beta_{333}a_3, \\
\alpha_{12} &= \beta_{12} + 2\beta_{112}a_1 + 2\beta_{122}a_2 + \beta_{123}a_3, \\
\alpha_{13} &= \beta_{13} + 2\beta_{113}a_1 + \beta_{123}a_2 + 2\beta_{133}a_3, \\
\alpha_{23} &= \beta_{23} + \beta_{123}a_1 + 2\beta_{223}a_2 + 2\beta_{233}a_3, \\
\alpha_{111} &= \beta_{111}, \\
\alpha_{222} &= \beta_{222}, \\
\alpha_{333} &= \beta_{333}, \\
\alpha_{122} &= \beta_{122}, \\
\alpha_{133} &= \beta_{133}, \\
\alpha_{112} &= \beta_{112}, \\
\alpha_{233} &= \beta_{233}, \\
\alpha_{113} &= \beta_{133}, \\
\alpha_{223} &= \beta_{223}, \\
\alpha_{123} &= \beta_{123},
\end{aligned}
\tag{12.3.7}
$$

continued

The origin-shift parameter situation for the quadratic model in three variables is found by setting all β's with three subscripts equal to zero in (12.3.6) and (12.3.7). In both cases, detailed conclusions developed by extending the ideas of Example 2 are left to the reader.

Example 4

Peixoto (1990) gives the more general polynomial example model

$$
\begin{aligned}
Y = \beta_0 &+ \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \beta_{22} X_2^2 + \beta_{122} X_1 X_2^2 \\
&+ \beta_{222} X_2^3 + \beta_{1222} X_1 X_2^3 + \epsilon
\end{aligned}
\tag{12.3.8}
$$

as a well-formulated one that satisfies Criterion 1.

[*Exercise:* Substitute $X_1 = Z_1 + a_1$, $X_2 = Z_2 + a_2$ into (12.3.8) and show that no new types of polynomial terms occur in the expansion.]

Criterion 2. The Axes Rotation Criterion

We shall consider a reduced model to be a sensible one if a rotation of the X-axes produces a model of unchanged *form* in the X's.

Example 5. Second-Order Polynomial in Two X's

We consider again the second-order model (12.3.2). An axial rotation from the X's to (say) the W's is such that

$$
\begin{aligned}
W_1 &= c_1 X_1 + c_2 X_2, \\
W_2 &= d_1 X_1 + d_2 X_2,
\end{aligned}
\tag{12.3.9}
$$

or $\mathbf{W} = \mathbf{MX}$, say, where $\mathbf{M}$ is an orthonormal matrix, such that $\mathbf{M'M} = \mathbf{MM'} = \mathbf{MM}^{-1} = \mathbf{I}$. These conditions, which preserve both length and orthogonality of the axes after rotation, imply that $\mathbf{M}$ must take the form

$$\mathbf{M} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} c & s \\ -s & c \end{bmatrix}, \tag{12.3.10}$$

say, or a similar form with changes of signs in rows and/or columns.

To substitute for the X's in (12.3.2), we invert via $\mathbf{X} = \mathbf{M}^{-1}\mathbf{W} = \mathbf{M'W}$ to give

$$\begin{aligned} X_1 &= cW_1 - sW_2, \\ X_2 &= sW_1 + cW_2. \end{aligned} \tag{12.3.11}$$

The surface after rotation is then

$$Y = \gamma_0 + \gamma_1 W_1 + \gamma_2 W_2 + \gamma_{11} W_1^2 + \gamma_{22} W_2^2 + \gamma_{12} W_1 W_2 + \epsilon,$$

where

$$\begin{aligned} \gamma_0 &= \beta_0, \\ \gamma_1 &= c\beta_1 + s\beta_2, \\ \gamma_2 &= -s\beta_1 + c\beta_2, \\ \gamma_{11} &= c^2\beta_{11} + s^2\beta_{22} + cs\beta_{12}, \\ \gamma_{22} &= s^2\beta_{11} + c^2\beta_{22} - cs\beta_{12}, \\ \gamma_{12} &= -2cs\beta_{11} + 2cs\beta_{22} + (c^2 - s^2)\beta_{12}. \end{aligned} \tag{12.3.12}$$

[The case $\theta = 45°$, $c = s = 2^{-1/2}$ was used by Box and Draper (1987), pp. 447–448.] We see that order is completely preserved, namely, zero-order, first-order, and second-order γ coefficients are linear combinations of only zero-order, first-order, and second-order coefficients, respectively. This example, which extends to more X's, illustrates the more general statement of Rule 2.

Rule 2

If a model is to be consistent under rotation of axes, all terms of a particular order must be considered as a unit. Subsets of terms of a given order cannot be removed.

(For example, it is senseless to consider dropping β_{12} alone, because this does not remove γ_{12}, which is still a combination of β_{11} and β_{22}.)

Application of Rules 1 and 2 Together

If a model is to be consistent both under a shift in origin and under rotation, terms can be dropped only in units of order (e.g., all second-order terms, all first-order terms). Moreover, lower-order terms cannot be deleted if higher-order terms are retained in the model.

In situations where the original axial directions are considered mandatory and untouchable, so that rotation to new axes is out of the question (e.g., from practical considerations in the interpretation of the effects shown by the fitted model), only Rule 1 would apply.

"Do We Need This *X*?"

Assume now that we do *not* wish to consider rotations of the surface, so that Rule 2 will not come into consideration. We can then ask questions about specific X-variables. Look again at (12.3.5) and consider the question: "Does X_2 contribute to second-order curvature?" This is equivalent to testing $H_0: \beta_{22} = \beta_{12} = 0$ versus H_1: not so. We see that this is equivalent to $\alpha_{22} = \alpha_{12} = 0$, so that Rule 1 is satisfied. We can then test this null hypothesis via an extra sum of squares test in the usual way. It is also possible to test whether X_2 is necessary *at all*, via $H_0: \beta_2 = \beta_{22} = \beta_{12} = 0$, because then $\alpha_2 = \alpha_{22} = \alpha_{12} = 0$. This implies the following:

If rotation of the axes is not an option because of the nature of the predictor variables, subsets of higher-order terms (or all terms) that depend on one (or more) specific predictors can be deleted, but only in combinations that keep the model well formulated, that is, satisfy Rule 1.

For example, in (12.3.7), some possibilities are:

(a) $\beta_{111} = \beta_{122} = \beta_{112} = \beta_{133} = \beta_{113} = \beta_{123} = 0$. (Does X_1 contribute to third-order curvature?)

(b) $\beta_{11} = \beta_{12} = \beta_{13} = 0$, in addition to (a). (Does X_1 contribute to second- and third-order curvature?)

(c) $\beta_1 = 0$, in addition to (b). (Do we need X_1 at all?)

Summary Advice

In deciding what terms can sensibly be deleted from a polynomial model, we suggest:

1. Always apply Rule 1.
2. Consider whether the original predictor variables are either:
   a. not absolute descriptors of the response surface, so that rotation of the axes to give the surface in terms of new variables that are rotational linear combinations of the original variables can be considered; apply Rule 2 in addition to Rule 1; or,
   b. always to be retained (so that descriptions of the response surface are always in these variables); apply only Rule 1, perhaps considering hypotheses involving specific X's, as in the examples above.

Using a Selection Procedure for a Polynomial Fit

Selection procedures (see Chapter 15) do not incorporate rules of the type we have given. Thus their use to get polynomial models is suspect. It might be reasonable to let a selection procedure offer its choice, but the equation should then be reviewed and refined with the criteria given above in mind.

References

Box and Draper (1987); Driscoll and Anderson (1980); Peixoto (1987, 1990).

EXERCISES FOR CHAPTER 12

A. Eighteen observations were obtained on four predictor variables and one response variable in a process. It is suggested that the model

$$Y = \beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_{11} X_1^2 + \beta_{12} X_1 X_2 + \beta_{22} X_2^2 + \beta_{13} X_1 X_3 + \beta_{14} X_1 X_4 + \epsilon$$

is a reasonable one. The data are shown in Table A.

Requirements

1. Examine the data and the model. Is it possible to fit the proposed model to the data? Why or why not?
2. Estimate $V(Y_i) = \sigma^2$.

TABLE A. Data

X_1	X_2	X_3	X_4	Y
20	50	75	15	27
27	55	60	20	23
22	62	68	16	18
27	55	60	20	26
24	75	72	8	23
30	62	73	18	27
32	79	71	11	30
24	75	72	8	23
22	62	68	16	22
27	55	60	20	24
40	90	78	32	16
32	79	71	11	28
50	84	72	12	31
40	90	78	32	22
20	50	75	15	24
50	84	72	12	31
30	62	73	18	29
27	55	60	20	22

B. (*Source:* "Variable shear rate viscosity of SBR-filler-plasticizer systems," by G. C. Derringer, *Rubber Chemistry and Technology,* **47,** September 1974, 825–836.) Fit the model

$$Y = (\alpha_0 + \alpha_1 Z + \alpha_2 Z^2) + (\beta_0 + \beta_1 Z + \beta_{11} Z^2) X_1 + (\gamma_0 + \gamma_1 Z + \gamma_{11} Z^2) X_2 + \epsilon,$$

where $Z = \ln(X_3 + 1)$, to the data below, and give a complete analysis. Note that there are six repeat runs.

X_1	X_2	X_3	Y	X_1	X_2	X_3	Y
47.1	33.9	7.5	11.97	47.1	33.9	750	8.46
72.9	33.9	750	8.63	60	21	75	10.65
47.1	8.1	750	8.80	60	21	3000	7.60
60	21	75	10.73	60	21	3	13.06
60	21	75	10.69	39	21	75	10.51
72.9	8.1	7.5	13.12	60	0	75	11.22
47.1	8.1	7.5	12.58	60	21	75	10.67
72.9	33.9	7.5	12.24	60	42	75	10.24
60	21	75	10.64	81	21	75	10.74
72.9	8.1	750	9.09	60	21	75	10.69

C. (*Source:* "A short life test for comparing a sample with previous accelerated test results," by Wayne Nelson, *Technometrics,* **14,** 1972, 175–185.) The data in the table below are accelerated life test results on 24 units of a type of sheathed tabular heater. T is the temperature in °F and Y is the life in hours at that temperature for a single unit. Six units are tested at each temperature. Plot the data and look at them. Fit the model

$$\log_{10} Y = \beta_0 + \beta_1\{1000/(T + 460)\} + \epsilon$$

and perform all the usual analyses. [*Note:* $T + 460$ is the absolute temperature in °F.]

T	Y
1520	1953, 2135, 2471, 4727, 6143, 6314
1620	1190, 1286, 1550, 2125, 2557, 2845
1660	651, 837, 848, 1038, 1361, 1543
1708	511, 651, 651, 652, 688, 729

D. The experiment summarized in the table below was run on a pilot plant to examine the effects of varying the percentage of a certain mix component (X_1), the temperature of the mix(X_2), and the flow-through rate (X_3), on three responses, Y_1, Y_2, and Y_3. The input variables have been coded, but the responses are in their original units. The experimental design shown is a central composite design consisting of eight cube points (X_1, X_2, X_3) = (±1, ±1, ±1), six axial points ($\pm\alpha$, 0, 0), (0, $\pm\alpha$, 0), (0, 0, $\pm\alpha$), where $\alpha = 1.2154$, and one center point (0, 0, 0). The run order shown is the randomized order in which the design was performed.

X_1	X_2	X_3	Y_1	Y_2	Y_3
−1	−1	1	85.3	72.7	97.1
1	1	−1	72.3	57.6	96.9
0	1.2154	0	71.4	56.5	96.4
0	−1.2154	0	72.0	64.6	96.8
−1	−1	−1	87.0	79.2	97.0
1	1	1	55.6	32.6	96.2
0	0	−1.2154	85.0	75.9	97.2
1.2154	0	0	70.9	53.4	97.9
0	0	0	75.9	59.3	97.4
1	−1	1	76.1	63.2	97.4
−1	1	−1	85.0	75.3	97.2
0	0	1.2154	68.0	57.2	95.5
−1.2154	0	0	89.6	83.6	97.2
−1	1	1	75.0	61.5	96.5
1	−1	−1	74.2	61.0	98.2

Requirements. Is the design rotatable? (See Section 12.2.) Using multiple regression techniques, formulate and fit suitable models of first or second order to Y_1, Y_2, and Y_3 separately. Perform a complete analysis and provide practical conclusions. If larger values of the Y's were more desirable, where in the X-space would it be better to operate?

E. A new product was being considered by the bakery goods research division of a large corporation. Of paramount concern was the maximum peak height obtained from a standard container of mixed dough just prior to baking. Four major ingredients were thought to be important in affecting peak height: percentage of fat, percentage of water, amount of flour in the brew, and the speed of the mixer in rpm. The experiments given in the table were performed, and the values of the maximum peak heights (shown in the body of the table) were recorded for each run. Note that there are four repeat runs at the conditions (12, 50, 20, 130) with responses 492, 523, 530, and 590.

Experimental Data Extensigraph Maximum Peak Height

Percent Fat	Flour in Brew	Percent Water 46 rpm 90	46 rpm 130	46 rpm 170	50 rpm 90	50 rpm 130	50 rpm 170	54 rpm 90	54 rpm 130	54 rpm 170
	10	833		540				673		493
8	20					537				
	30	577		547				660		512
	10					547				
						492				
12	20		653		650	523	553		487	
						530				
						590				
	30					595				
	10	802		477				710		520
16	20					575				
	30	568		401				572		483

Requirements

1. By suitable choice of central levels and scale divisors, code all four predictor variables so that their levels are $(-1, 0, 1)$. Write out the design in the coded variables, and confirm that it is a "cube plus star plus four center points" type design. Is it rotatable?
2. Using multiple regression techniques, construct a suitable model of first or second order for predicting maximum peak heights. In your conclusions, indicate the relative importance of the predictor variables, and make any other comments you find relevant.

F. Fit a full second-order model and perform a complete analysis, using the Y_4 data of Table 12.1.

G. If we believe in the "origin-shift" criterion," is the model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \beta_{22} X_2^2 + \beta_{122} X_1 X_2^2 + \epsilon$$

a "well-formulated" one?

H. A proposed model, based on theoretical considerations, is

$$Y = \alpha X_1^{\beta} X_2^{\gamma} X_3^{\delta} \epsilon.$$

Requirements. After transformation, fit the proposed model by least squares. State which predictor variable appears most important and check all coefficients for statistical significance (take $\alpha = 0.05$). Is the model a satisfactory one?

The data shown below, which relate to a study of the quantity of vitamin B_2 in turnip green, are taken from the "Annual progress report on the soils–weather project, 1948," by J. T. Wakeley, University of North Carolina (Raleigh) Institute of Statistics Mimeo Series 19 (1949). The variables are:

X_1 = radiation in relative gram calories per minute during the preceding half day of sunlight (coded by dividing by 100),

X_2 = average soil moisture tension (coded by dividing by 100),

X_3 = air temperature in degrees Fahrenheit (coded by dividing by 10),

Y = milligrams of vitamin B_2 per gram of turnip green.

These data were used by R. L. Anderson and T. A. Bancroft in *Statistical Theory in Research,* McGraw-Hill, New York, 1959, on p. 192, to fit the model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \epsilon.$$

Requirements. Develop a suitable fitted equation using these data and compare its form with the form of the one fitted by Anderson and Bancroft.

X_1	X_2	X_3	Y
1.76	0.070	7.8	110.4
1.55	0.070	8.9	102.8
2.73	0.070	8.9	101.0
2.73	0.070	7.2	108.4
2.56	0.070	8.4	100.7
2.80	0.070	8.7	100.3
2.80	0.070	7.4	102.0
1.84	0.070	8.7	93.7
2.16	0.070	8.8	98.9
1.98	0.020	7.6	96.6
0.59	0.020	6.5	99.4
0.80	0.020	6.7	96.2
0.80	0.020	6.2	99.0
1.05	0.020	7.0	88.4
1.80	0.020	7.3	75.3
1.80	0.020	6.5	92.0
1.77	0.020	7.6	82.4
2.30	0.020	8.2	77.1
2.03	0.474	7.6	74.0
1.91	0.474	8.3	65.7
1.91	0.474	8.2	56.8
1.91	0.474	6.9	62.1
0.76	0.474	7.4	61.0
2.13	0.474	7.6	53.2
2.13	0.474	6.9	59.4
1.51	0.474	7.5	58.7
2.05	0.474	7.6	58.0