ON SOME ASPECTS OF SECOND ORDER RESPONSE SURFACE METHODOLOGY.

A thesis presented in partial fulfilment of the requirements for the degree of M.Sc. in Mathematics at Massey University.

Vernon John Thomas

1972

Abstract

A unified development of the theoretical basis of response surface methodology, particularly as it applies to second order response surfaces, is presented. A rigorous justification of the various tests of hypothesis usually used is given, as well as a convenient means of making tests on whole factors, rather than on terms of a given degree, as is customary at present. Finally, the super--imposition of some elementary classification designs on a response surface design is considered.

Acknowledgement

I would like to take this opportunity of expressing my gratitude to Dr B.S.Weir for his comments and encouragement in the preparation of this thesis.

Table of contents

1. Response Surface Methodology.

Introduction

Response surface methodology seeks to estimate, by regression methods, that linear combination of previously specified graduating functions of a number of independent variables which provides, in some sense, the best fit to an observed response.

While the techniques of fitting are identical with, or closely related to, those of multiple linear regression, the emphasis is slightly different, in that considerable stress is laid on the design aspect of the problem. It is assumed that the levels of the independent variables may be pre-specified at will, within broad limits. The space defined on the independent variables, and within these limits, is termed the region of operability. The sub-space of this region, in which estimates of response are of interest to the experimenter, is termed the region of interest.

Typically, a number of experiments are carried out, according to some previously decided experimental plan. Each experiment consists of the measurement of an observed response at a point defined by some combination of the independent variables. In some cases, sequential designs are used - that is, the curve fitted to date is used as information to assist in the specification of the combination of independent variables to be used in the next experiment.

The basic variability of the observed response is measured by replication of experimental points, or by the residual error of the observed response from the fitted surface. This latter error can arise from a true observational error or from inadequate specification of the model, whereas the error based on point replication estimates true experimental error only. For this reason, when point replication is used, the residual error may be used to test model adequacy.

<u>The model</u>

The model is developed by assuming p independent variables, given by

$$\underset{\sim}{\xi} = (\xi_1 \ \dots \ \xi_p)^T$$

and k pre-specified vector graduating functions of these variables, given by

$$\underset{\sim}{x} = \underset{\sim}{x}(\underset{\sim}{\xi}) \text{ where } \underset{\sim}{x} \text{ is kx1}$$

The observational response is assumed (or known) to be

$$y = \eta + \epsilon$$

where ϵ is a random variate with zero mean, and the so-called "true response" η is given by the exact relationship

$$\eta = \underset{\sim}{x}^T \underset{\sim}{\beta} \qquad (1.1)$$

where $\underset{\sim}{\beta}$ is a vector of unknown coefficients. The measurement of an observed y, for some known $\underset{\sim}{\xi}$, is termed an experiment. The values of ϵ arising from different experiments are assumed to be statistically independent, with constant, unknown, variance σ^2.

The aim of the sequence of experiments is to estimate $\boldsymbol{\beta}$ by $\underset{\sim}{b}$, and, from this, to estimate the response at any point of the region of interest by

$$\hat{y} = \underset{\sim}{x}^T \underset{\sim}{b}$$

To achieve this, n experiments are conducted, at the points $\underset{\sim}{\xi}_u$, u=1, ... n, yielding n observed responses

$$\underset{\sim}{y} = (y_1 \ldots y_n)^T$$

Now let

$$\Xi = (\underset{\sim}{\xi}_1 \ldots \underset{\sim}{\xi}_n)^T \text{ of dimension nxp}$$

$$\underset{\sim}{x}_u = \underset{\sim}{x}(\underset{\sim}{\xi}_u)$$

and $$X = (\underset{\sim}{x}_1 \ldots \underset{\sim}{x}_n)^T \text{ of dimension nxk}$$

so that $\underset{\sim}{y}$ is the observed value of the true response $X\boldsymbol{\beta}$.

Properly speaking, Ξ is the design matrix, since Ξ determines X. However, once Ξ is chosen, according to some design criterion, it is convenient to refer to X as the design matrix, since all operations are in terms of X.

In the vast majority of applications, $\underset{\sim}{x}$ consists of all powers of the ξ, separately or together, up to some maximum degree d. The design is then referred to as a dth order design. Thus, for a second order design

$$\underset{\sim}{x}(\underset{\sim}{\xi}) = (1;\ \xi_1 \ldots \xi_p;\ \xi_1^2 \ldots \xi_p^2;\ \xi_1\xi_2 \ldots \xi_{p-1}\xi_p)^T$$

For this type of design, it is convenient to use the subscripts occurring in the corresponding element of $\underset{\sim}{x}$ to identify the elements of $\underset{\sim}{\beta}$, thus, for second order designs,

$$\underset{\sim}{\beta} = (\beta_0; \beta_1 \cdots \beta_p; \beta_{11} \cdots \beta_{pp}; \beta_{12} \cdots \beta_{(p-1)p})^T$$

In general, for a dth order design, there will be $\binom{p+d}{d}$ coefficients.

Within this framework, $\underset{\sim}{x}^T\underset{\sim}{\beta}$ is a general dth order polynomial in p variables.

The exceptions to this kind of polynomial are of two types. In the first type, the elements of $\underset{\sim}{x}$ are not powers of the elements of $\underset{\sim}{\xi}$. For example, M.J.Box (1968) considerdd the functions given by

$$x_i = \exp(\xi_i)$$

as well as other non-polynomial functions.

The second type occurs when certain of the terms of the polynomial $\underset{\sim}{x}^T\underset{\sim}{\beta}$ cannot be estimated, and must, therefore, be omitted. For example, in the bivariate case (p=2), if Ξ specified the points of a 3x5 factorial design, necessarily the polynomial elements of $\underset{\sim}{x}$ must be a subset of

$$1, \xi_1, \xi_2, \xi_1^2, \xi_2^2, \xi_1\xi_2, \xi_2^3, \xi_1^2\xi_2, \xi_1\xi_2^2, \xi_2^4, \xi_1^2\xi_2^2, \xi_1\xi_2^3, \xi_1\xi_2^4, \xi_1^2\xi_2^4$$

which omits the combinations ξ_1^3, ξ_1^4, and $\xi_1^3\xi_2$, whose coefficients cannot be estimated because an insufficient number of levels of ξ_1 was used. Similarly only two coefficients of degree five or higher may be estimated from this design. In practice it is unlikely that an attempt would be made to estimate the coefficients of $\xi_1\xi_2^4$ or $\xi_1^2\xi_2^4$. If it were, and if the factorial were unreplicated, an exact fit would be obtained.

Estimation

Methods, culminating in the estimates $\underset{\sim}{b}$ and $\hat{y}$, may be divided into design procedures and estimation procedures. Design procedures are those used to specify Ξ and hence X. Discussion on methods of selecting the design is outside the scope of this thesis. Estimation procedures are those which, given X and $\underset{\sim}{y}$, seek to estimate $\underset{\sim}{b}$. In general, $\underset{\sim}{b}$ is assumed to be a linear combination of the observed responses, of the form

$$\underset{\sim}{b} = T\underset{\sim}{y}$$

where T depends only on X (not, for example, on $\underset{\sim}{\beta}$).

The commonest estimator arises from minimization of the sum of squares of the errors $y_u - \hat{y}_u$. This is known as the least squares estimator, and is, in fact, identical to that obtained when ϵ is assumed to have a normal distribution, and maximum likelihood estimation used.

The quantity to be minimized is

$$(\underset{\sim}{y}-X\underset{\sim}{b})^T(\underset{\sim}{y}-X\underset{\sim}{b}) \qquad (1.2)$$

Differentiation with respect to $\underset{\sim}{b}$ and equation to zero yields

$$-2X^T(\underset{\sim}{y}-X\underset{\sim}{b}) = \underset{\sim}{0}$$

from which

$$\underset{\sim}{b} = (X^TX)^{-1}X^T\underset{\sim}{y}$$

or

$$T = (X^TX)^{-1}X^T$$

provided that X^TX is non-singular. If X^TX is singular, a generalized inverse may be used, but is unnecessary in the present case.

This straight-forward estimator has many desirable properties. In particular,

$$E(\underset{\sim}{b}) = (X^TX)^{-1}X^TX\underset{\sim}{\beta} = \underset{\sim}{\beta} \qquad (1.3)$$

$$\begin{aligned} Var(\underset{\sim}{b}) &= (X^TX)^{-1}X^T \; Var(\underset{\sim}{y}) \; X(X^TX)^{-1} \qquad (1.4) \\ &= \sigma^2 \quad (X^TX)^{-1} \end{aligned}$$

from the assumptions about ϵ. Hence the estimator is unbiased from (1.3). It can also be shown that (1.4) gives the minimum variance arising from an unbiased linear estimator.

Finally

$$\begin{aligned} Var(\hat{y}) &= Var(\underset{\sim}{x}^T\underset{\sim}{b}) \\ &= \sigma^2 \; \underset{\sim}{x}^T(X^TX)^{-1}\underset{\sim}{x} \end{aligned}$$

for arbitrary $\underset{\sim}{x}$, not necessarily one of the $\underset{\sim}{x}_u$. Thus the various variances can easily be derived from $(X^TX)^{-1}$.

This leads naturally to the concept of a rotatable design, which is a polynomial design for which $Var(\hat{y})$ depends only on σ^2 and $\underset{\sim}{\xi}^T\underset{\sim}{\xi}$. That is, $Var(\hat{y})$ is invariant under orthogonal rotation of the ξ-axes.

It should be emphasized that the estimator defined above is not the only linear estimator possible. In particular, in conditions where the specified model (1.1) is inadequate, that is, where y contains other terms than those in a linear combination of the specified $\underset{\sim}{x}$, a different estimator may assist in compensating, in some degree, for this inadequacy, at the expense of greater variance.

Hypothesis testing

From this point hypothesis testing will be considered, and the additional assumption that the ϵ are normally distributed will be required.

Now let

$$M = I - X(X^TX)^{-1}X^T$$
$$N = X(X^TX)^{-1}X^T$$

Note that both M and N are idempotent matrices, nxn, and that MN=0, MX=0. Also

$$\text{tr } N = \text{tr}\{X(X^TX)^{-1}X^T\}_{nxn}$$
$$= \text{tr}\{(X^TX)^{-1}X^TX\}_{pxp}$$

since compatible matrices commute under the trace operator. Hence

$$\text{tr } N = \text{tr } I_{pxp} = p$$
$$\text{tr } M = \text{tr } I_{nxn} - \text{tr } N = n-p$$

The residual sum of squares (1.2) is equal, on expansion, to

$$\begin{aligned} SSE &= \underset{\sim}{y}^T\underset{\sim}{y} - \underset{\sim}{y}^T X\underset{\sim}{b} \\ &= \underset{\sim}{y}^T\underset{\sim}{y} - \underset{\sim}{y}^T X(X^TX)^{-1}X^T\underset{\sim}{y} \qquad (1.5) \\ &= \underset{\sim}{y}^T M\underset{\sim}{y} \end{aligned}$$

It is necessary to recall a theorem on the distribution of quadratic forms (see, for example, Graybill (1961)).

Theorem: If $\underset{\sim}{y} \sim N(\underset{\sim}{\mu}, \sigma^2 I)$, then $\underset{\sim}{y}^T A \underset{\sim}{y}/\sigma^2$ is distributed as $\chi'^2(k,\lambda)$, where χ'^2 represents the non-central chi-squared distribution, and $\lambda = \frac{1}{2\sigma^2}\underset{\sim}{\mu}^T A \underset{\sim}{\mu}$, if, and only if, A is an idempotent matrix and tr A = k.

In the present situation, $\underset{\sim}{y} \sim N(X\underset{\sim}{\beta}, \sigma^2 I)$ and hence

$$\frac{SSE}{\sigma^2} = \frac{\underset{\sim}{y}^T M \underset{\sim}{y}}{\sigma^2} \sim \chi'^2(n-p, \lambda)$$

where $\lambda = \frac{1}{2\sigma^2}\underset{\sim}{\beta}^T X^T M X \underset{\sim}{\beta} = 0$. Thus SSE/σ^2 has a central χ^2 distribution with n-p degrees of freedom.

The second term in (1.5) is the sum of squares accounted for by the regression, and is

$$SSR = \underset{\sim}{y}^T N \underset{\sim}{y}$$

By a process of reasoning similar to that for SSE, it is an easy matter to establish that

$$\frac{SSR}{\sigma^2} \sim \chi'^2(p, \lambda)$$

where $\lambda = \frac{1}{2\sigma^2}\underset{\sim}{\beta}^T X^T N X \underset{\sim}{\beta} = \frac{1}{2\sigma^2}\underset{\sim}{\beta}^T X^T X \underset{\sim}{\beta}$

Again from the theory of quadratic forms, a necessary and sufficient condition for $\underset{\sim}{y}^T A \underset{\sim}{y}$ and $\underset{\sim}{y}^T B \underset{\sim}{y}$ to be independent is that AB=O.

Hence, since MN=O, SSE and SSR are independent and

$$F = \frac{SSR}{p} \Big/ \frac{SSE}{n-p}$$

has a non-central F-distribution with p and n-p degrees of freedom and non-centrality parameter $\frac{1}{2\sigma^2}\underset{\sim}{\beta}^T X^T X \underset{\sim}{\beta}$. Thus F may be used to test the hypothesis that $\underset{\sim}{\beta} = \underset{\sim}{0}$.

In response surface design it is usual to further subdivide SSE by taking advantage of point replication.

As a preliminary, suppose that the model specification (1.1) is incorrect and that while the model

$$\eta = \underset{\sim}{x}_1^T \underset{\sim}{\beta}_1$$

has been assumed, the true model is

$$\eta = \underset{\sim}{x}_1^T \underset{\sim}{\beta}_1 + \underset{\sim}{x}_2^T \underset{\sim}{\beta}_2$$

Using X_1 and X_2 in an obvious way,

$$\underset{\sim}{b} = (X_1^T X_1)^{-1} X_1^T \underset{\sim}{y}$$

In these circumstances

$$E(\underset{\sim}{b}) = (X_1^T X_1)^{-1} X_1^T (X_1 \underset{\sim}{\beta}_1 + X_2 \underset{\sim}{\beta}_2)$$
$$= \underset{\sim}{\beta}_1 + (X_1^T X_1)^{-1} X_1^T X_2 \underset{\sim}{\beta}_2$$

and $\underset{\sim}{b}$ is a biased estimator of $\underset{\sim}{\beta}_1$. The matrix $A = (X_1^T X_1)^{-1} X_1^T X_2$ is known as the alias matrix (Box and Wilson (1951)) and measures the extent of the bias.

Putting

$$M_1 = I - X_1 (X_1^T X_1)^{-1} X_1^T$$

and using the same expansion as before,

$$SSE = \underset{\sim}{y}^T M_1 \underset{\sim}{y}$$

and $\frac{SSE}{\sigma^2} \sim \chi'^2(n-p, \lambda)$.

However, in the present case

$$\underset{\sim}{y} \sim N(X_1 \underset{\sim}{\beta}_1 + X_2 \underset{\sim}{\beta}_2,\ \sigma^2 I)$$

and thus

$$\lambda = \frac{1}{2\sigma^2} (\underset{\sim}{\beta}_2^T X_2^T + \underset{\sim}{\beta}_1^T X_1^T) M_1 (X_1 \underset{\sim}{\beta}_1 + X_2 \underset{\sim}{\beta}_2)$$
$$= \frac{1}{2\sigma^2} \underset{\sim}{\beta}_2^T X_2^T M_1 X_2 \underset{\sim}{\beta}_2 \neq 0$$

in general, and the F-test described above is no longer available.

Suppose, however, that point replication has been used, and that r distinct points have been included in the design, the sth of them n_s times, s=1, ... r, and $\Sigma_{s=1}^{r} n_s = n$. Also, let $\bar{y}_s$ be the group mean of the y-values measured at the sth distinct point.

Without loss of generality, the points may be arranged in such a way that the n_s points in the sth group are together in the $\underset{\sim}{y}$ and X matrices.

Now define

$$K = I - \begin{pmatrix} \frac{1}{n_1} J_{n_1} & & & 0 \\ & \cdot & & \\ & & \cdot & \\ 0 & & & \frac{1}{n_r} J_{n_r} \end{pmatrix} = I - J$$

where J_{n_s} is the $n_s x n_s$ matrix with all unit elements, so that, without point replication, $n_s=1$, and K=O.

Now $K^2=K$, hence K is idempotent, and tr K = n-r. Also, since Ξ, and hence X_1 and X_2, consist of r groups of n_1, ... n_r identical rows, $JX_1=X_1$ and $JX_2=X_2$, from which $KX_1=KX_2=O$. Hence $KM_1=K$ and $KN_1=O$.

If the y-values are standardized by $z_u=y_u-\bar{y}_s$, where $\bar{y}_s$ is the group mean containing y_u, then

$$\begin{aligned} \underset{\sim}{z} &= K\underset{\sim}{y} \\ &= K(X_1\underset{\sim}{\beta}_1+X_2\underset{\sim}{\beta}_2+\underset{\sim}{\epsilon}) \\ &= K\underset{\sim}{\epsilon} \end{aligned}$$

Now SSW, the sum of squares within groups of observations at the same point, is given by

$$SSW = \underset{\sim}{z}^T\underset{\sim}{z} = \underset{\sim}{y}^T K\underset{\sim}{y} = \underset{\sim}{\epsilon}^T K\underset{\sim}{\epsilon}$$

and since $\underset{\sim}{\epsilon} \sim N(\underset{\sim}{0},\sigma^2 I)$, and tr K = n-r,

$$\frac{SSW}{\sigma^2} \sim \chi^2(n-r)$$

by the theorem quoted for SSE. Also, SSW and SSR are independent, since

$$KN_1 = KX_1(X_1^TX_1)^{-1}X_1^T = 0$$

Now consider SSF (for sum of squares due to lack of fit), defined by

$$\begin{aligned} SSF &= SSE - SSW \\ &= \underset{\sim}{y}^T(M_1-K)\underset{\sim}{y} \end{aligned}$$

$$\begin{aligned} \text{Now} \qquad (M_1-K)^2 &= M_1^2 - M_1K - KM_1 + K^2 \\ &= M_1-K \\ \text{tr } (M_1-K) &= \text{tr } M_1 - \text{tr } K \\ &= r - p \end{aligned}$$

Hence, from the theorem,

$$\frac{SSF}{\sigma^2} \sim \chi'^2(r-p,\lambda)$$

$$\begin{aligned} \text{where } \lambda &= \frac{1}{2\sigma^2}(\underset{\sim}{\beta}_2^TX_2^T+\underset{\sim}{\beta}_1^TX_1^T)(M_1-K)(X_1\underset{\sim}{\beta}_1+X_2\underset{\sim}{\beta}_2) \\ &= \frac{1}{2\sigma^2}\underset{\sim}{\beta}_2^TX_2^TM_1X_2\underset{\sim}{\beta}_2 \end{aligned}$$

This requires, reasonably enough, $r>p$.

Finally, SSR= $\underset{\sim}{y}^TN_1\underset{\sim}{y}$ where $N_1=X_1(X_1^TX_1)^{-1}X_1$, and SSR has a $\chi'^2(p,\lambda)$ distribution where

$$\lambda = \frac{1}{2\sigma^2}(\underset{\sim}{\beta}_2^TX_2^T+\underset{\sim}{\beta}_1^TX_1^T)N_1(X_1\underset{\sim}{\beta}_1+X_2\underset{\sim}{\beta}_2)$$

$$= \frac{1}{2\sigma^2}[\underset{\sim}{\beta}_1^TX_1^TX_1\underset{\sim}{\beta}_1+2\underset{\sim}{\beta}_1^TX_1^TX_2\underset{\sim}{\beta}_2+\underset{\sim}{\beta}_2^TX_2^TX_1(X_1^TX_1)^{-1}X_1^TX_2\underset{\sim}{\beta}_2]$$

Now note that, where L is an arbitrary matrix,

$$\begin{aligned} E(\underset{\sim}{y}^TL\underset{\sim}{y}) &= E(\text{tr } \underset{\sim}{y}^TL\underset{\sim}{y}) = E(\text{tr } L\underset{\sim}{y}\underset{\sim}{y}^T) \\ &= \text{tr } [LE(X_1\underset{\sim}{\beta}_1+X_2\underset{\sim}{\beta}_2+\underset{\sim}{\epsilon})(X_1\underset{\sim}{\beta}_1+X_2\underset{\sim}{\beta}_2+\underset{\sim}{\epsilon})^T] \\ &= \text{tr } [L(X_1\underset{\sim}{\beta}_1+X_2\underset{\sim}{\beta}_2)(X_1\underset{\sim}{\beta}_1+X_2\underset{\sim}{\beta}_2)^T+\sigma^2L] \\ &= \text{tr } \underset{\sim}{\beta}^TX^TLX\underset{\sim}{\beta} + \sigma^2\text{tr } L \end{aligned}$$

where

$$\underset{\sim}{\beta} = \begin{pmatrix} \underset{\sim}{\beta}_1 \\ \underset{\sim}{\beta}_2 \end{pmatrix} \qquad X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

From this, the expected values of the various sums of squares are readily derived. The analysis is given in table 1.

Table 1

Basic response surface ANOV

Source	Sum of squares	DF	E(MS)
Regression	$SSR = \underset{\sim}{y}^T X_1 \underset{\sim}{b}$	p	$\sigma^2 + \frac{1}{p}\underset{\sim}{\beta}^T X^T M_1 X \underset{\sim}{\beta}$
Lack of fit	SSF by subtraction	r-p	$\sigma^2 + \frac{1}{r-p}\underset{\sim}{\beta}_2^T X_2^T M_1 X_2 \underset{\sim}{\beta}_2$
Error within replicated points	$SSW = \underset{\sim}{y}^T K \underset{\sim}{y}$	n-r	σ^2
Total	$SST = \underset{\sim}{y}^T \underset{\sim}{y}$	n	

Note that without replication n=r, and if $\underset{\sim}{\beta}_2=0$, this table reduces to the simpler form derived earlier.

While the above argument establishes the theoretical justification for the use of the F-tests, the test of the whole regression is, in practice, of little use. However, it is perfectly general, and not dependent on a polynomial specification of $\underset{\sim}{x}$. In the event that a polynomial is used, the SSR is ordinarily broken down into the classification shown in table 2.

Table 2

Conventional ANOV for regression coefficients in polynomial model

Source	DF
Mean, β_0	1
Linear terms	p
Second order terms	$\frac{1}{2}p(p+1)$
Third order terms	$\frac{1}{6}p(p+1)(p+2)$
...	
dth order terms	$\sum_{i=1}^{p} i^{d-1}$

While this is suitable for establishing the true degree of the polynomial, it is inadequate for establishing the importance, in the final response, of a particular ξ. Section 3 of this thesis considers the structure of $X^T X$, for the second order polynomial model, in some detail, in order to facilitate tests aimed at establishing the importance of particular elements of $\underset{\sim}{\xi}$.

Further topics

In field experiments, each experiment usually consists of a plot of ground. In most circumstances, the number of such plots which can be assumed to represent essentially the same external conditions is quite limited.

In order to control this type of environmental variation, a block structure may be superimposed on the response surface design, yielding a model of the form

$$\eta = \alpha_w + \underset{\sim}{x}^T \underset{\sim}{\beta} \qquad (1.6)$$

where α_w is the block effect associated with the wth block, with $\sum_w \alpha_w = 0$.

Designs including such block structures were introduced by DeBaun (1956) and elaborated by Box and Hunter (1957) in the case of rotatable designs. These designs allow adequate control of environmental variation.

A natural extension of this type of design is to consider the possibility of superimposing a further treatment effect, which, in practice, could represent something like a species effect. The model would be

$$\eta = \alpha_w + \tau_v + \underset{\sim}{x}^T \underset{\sim}{\beta}$$

where now τ_v is the vth treatment effect. As far as treatments are concerned, such a model is identical to the analysis of covariance model, which uses the regression variables $\underset{\sim}{x}$ to reduce variation in the response, major interest being focussed on the superimposed treatment effects. A response surface approach would have equal interest in both parts of the fitted model.

Pursuing this line of enquiry further, section 4 of this thesis considers the implications of combining various classification designs with a response surface design.

One obvious extension of the model described by (1.6) is to allow $\boldsymbol{\beta}$ to vary with the block, giving a model of the form

$$\eta = \alpha_w + \mathbf{x}^T \boldsymbol{\beta}_w$$

In many applications the question of the degree of correspondence between the individual regressions $\boldsymbol{\beta}_w$ and the overall regression $\boldsymbol{\beta}$ is of considerable importance. Section 4 also considers, briefly, this aspect of response surface methodology.

2 Historical Development

After a small number of related papers in the nineteen-forties, Box and Wilson (1951) laid the foundation for later work on response surface analysis. They were primarily concerned with a sequential series of experiments to determine the maximum or minimum point of a quadratic response surface. Their approach was to fit a linear model over the region of interest and make additional experiments in the direction of increasing response until first order effects, over some small sub-region, were negligible, and then fit a second order model. They also introduced the concept of the alias matrix to measure the bias arising from the use of an inadequate model. The designs they considered are known as central composite designs, consisting as they do of a superimposition of two or more centrally symmetric designs, usually a cuboidal (or factorial) design and a simplex design (a type of design which varies each variable in turn, setting all others to the central level). An example of a three-way design (that is, one involving ξ_1, ξ_2, and ξ_3) of this type is

$(\pm 1, \pm 1, \pm 1)$ $(\pm 2, 0, 0)$ $(0, \pm 2, 0)$ and $(0, 0, \pm 2)$

together with replicated central points $(0,0,0)$.

These ideas were further developed by Box (1952) who used rotations to minimize quadratic bias in linear models.

Elfving (1952) considered the two-variable model $y_u = \beta_1 x_{1u} + \beta_2 x_{2u} + \epsilon_u$ with no constant term, and showed that a particular design minimized the sum of the variances of the coefficients.

Elfving's paper, and that of Box and Wilson (1951), were reveiwed by Anderson (1953).

Chernoff (1953) generalized Elfving's work to more than two dimensions, and used Fisher's maximum likelihood information matrix to minimize the sum of the diagonal elements of this matrix.

Box and Hunter (1954) developed methods for establishing confidence regions for the solution of a set of simultaneous equations, and applied this to the problem of the confidence region for the stationary point on a fitted second order response surface.

Box (1954a) has a comment on a "confidence cone" of an estimated vector which, in this case, is the vector of steepest ascent of a response surface, as used by Box and Wilson (1951).

Hunter (1954, 1956) discussed, in general terms, the problem of finding a stationary point on a response surface, and pointed out that a general second order response surface could be transformed to a canonical form

$$y-\beta_0 = \beta_{11}\xi_1^2+\beta_{22}\xi_2^2$$

Box (1954a) and Davies (1954) gave general surveys of the then current state of response surface methodology.

De la Garza (1954), discussing a dth degree polynomial regression, with one independent variable, showed that, for any arbitrary spacing of experimental points, it is always possible to obtain the same X^TX matrix, using not more than d+1 distinct experimental levels of ξ. He then considered how these points may be selected in such a way as to minimize the variance of that coefficient which has the maximum variance. This criterion is known as the minimax variance criterion. Guest (1958) obtained general formulae for minimax variance spacing and compared this spacing with a uniform spacing.

Box and Youle (1955) considered the application of response surfaces in the field of chemistry.

DeBaun (1956) was the first to apply methods of blocking to central composite designs, with a rather cursory survey of the possibilities. His ideas were extended by Box and Hunter (1957), who considered rotatable designs in general, and made an extensive study of central composite designs in particular. Box and Hunter's paper gives what is probably the best summary of the classical approach to response surface experimental design.

The first discussions on response surface methods to appear in textbooks were Davies, mentioned above, DeBaun and Schneider (1958), who described particular applications, and Plackett (1960), who summarized the early optimum oriented work, in his book on regression analysis.

Many of the papers that appeared in the late 1950's and early to middle 1960's merely list particular designs or classes of designs. Hartley (1959) considered the smallest composite designs for fitting quadratic response surfaces, based on fractional factorials, plus simplex designs and centre points. Bose and Draper (1959) used a transformation group to generate point sets leading to quadratic response surfaces in three dimensions. Box and Behnken (1960a) used superimposed simplex designs to derive second order designs from first order designs. The resulting designs they called simplex-sum designs. Das (1963) and Das and Narasimhan (1962) developed quadratic designs from balanced incomplete block designs. Draper (1960a) and Herzberg (1967a) gave rather similar methods for generating second order designs based on permuting point sets and building up designs in p dimensions from designs in p-1 dimensions. Then, together, (Draper and Herzberg (1968)) they developed methods based on composite designs with more than one fractional factorial. Das (1961) considered second order and third order designs derived from factorials.

Third order designs are not, properly speaking, within the scope of this thesis, however, third order designs have been developed by Gardiner, Grandage, and Hader (1959), Draper (1960b, 1960c, 1961b, 1962), and Herzberg (1964).

DeBaun (1959) and Box and Behnken (1960b) considered designs in which, for reasons depending on the context of the experiment, each factor is limited to only three levels. Draper and Stoneman (1968) extended this work to the case where some factors are restricted to two levels and others to three or four levels. Herzberg (1966, 1967b) developed cylindrical designs, in which one factor was set at a predetermined number of levels, but the design was rotatable in the remaining factors.

A more complicated three-factor design, using the properties of dodecahedrons, was developed by Hermanson et al. (1964).

Bose and Carter (1959) used complex number properties to examine some of the characteristics of two-factor designs.

Missing values were considered by Draper (1961a) and the effects of point replication by Box (1959) and Dykstra (1959, 1960).

Kitagawa (1959) extended the early work on sequential experiments, Umland and Smith (1959) gave an interesting example of the use of LaGrange multipliers in fitting second order response surfaces under a second order constraint, and Box and Tidwell (1962) gave a useful summary of the effect of transformations of the independent variables.

A literature survey was compiled by Hill and Hunter (1966). However, their emphasis was on applications, and many important theoretical papers were omitted.

In the related field of multiple regression, a number of papers which considered the effect of model inadequacy appeared in the late nineteen-fifties. Since they did not pertain directly to response surfaces, bare mention will be made of them. They were Ehrenfield (1955), Griliches (1957), Hoel (1958), Kiefer (1958, 1959), Kiefer and Wolfowitz (1959), and David and Arens (1959). In a PhD thesis Folks (1958) compared various optimality criteria, with response surfaces in mind.

Their work was extended and related more directly to response surface methods in an important paper by Box and Draper (1959), who considered the problem of estimating a response by

$$y = \underset{\sim}{x}_1^T \underset{\sim}{b}$$

where $\underset{\sim}{x}_1$ includes all terms up to degree d_1, when in fact the true model is

$$\eta = \underset{\sim}{x}^T \underset{\sim}{\beta} = \underset{\sim}{x}_1^T \underset{\sim}{\beta}_1 + \underset{\sim}{x}_2^T \underset{\sim}{\beta}_2$$

where $\underset{\sim}{x}$ includes all terms up to degree d_2, with $d_2 > d_1$.

They discussed design criteria based on minimization of bias and of variance, integrated over the region of interest. Their main conclusion was that bias considerations were likely to have a much greater effect on design optimality than was variance minimization.

Since Box and Draper's paper, some forty papers have appeared considering response surface design from the point of view of various optimality criteria. It is not proposed to pursue this aspect of response surface methodology further in this thesis.

3. Form of X^TX for second order designs

Non-quadratic effects

In the analysis of response surface results, the only difficult calculation step is the inversion of the matrix $S=X^TX$. For this reason, some attention will be paid to the form of this matrix.

The general second order model is given by

$$\eta = \beta_0 + \sum_{i=1}^{p} \beta_i \xi_i + \sum_{i=1}^{p} \beta_{ii} \xi_i^2 + \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \beta_{ij} \xi_i \xi_j \qquad (3.1)$$

which suggests a natural partition of S into elements related to the constant term, those related to the linear terms, those related to the quadratic terms (that is, terms involving ξ_i^2) and those related to interaction terms (that is, those involving $\xi_i \xi_j$, $i \neq j$).

The introduction of zeros, or blocks of zeros, into S will not only simplify the inversion of S, but will enable tests of hypotheses on orthogonal groups of coefficients to be made without refitting the parameters.

Some progress in this direction may be made by measuring the variables $\xi_1, \ldots \xi_p$ from their respective means over the design. Using the subscript u to range over the actual sample points, and making this transformation,

$$\sum_u \xi_{iu} = 0 \qquad i = 1, \ldots p$$

This has the immediate effect of making the linear effects orthogonal to the constant term.

Within this framework, attention will be confined to designs for which all odd moments, that is, moments which include at least one odd power, about the mean of the independent variables ξ_i, up to the fourth moments, are zero. Such designs include all symmetric designs, that is, designs for which the inclusion of a point containing a co-ordinate ξ_i implies the inclusion of an otherwise similar point with co-ordinate $-\xi_i$, which may be the same point if $\xi_i=0$.

This restriction still includes all rotatable designs, and, in particular, all central composite designs. This latter type of design has received by far the greatest attention in the literature. As noted in section 2 of this thesis, they are formed by the superimposition of cuboidal designs, simplex or "star" designs, and centre points.

For symmetric designs, for every term $\xi_{iu}\xi_{ju}$, either ξ_{iu} or $\xi_{ju} = 0$, or there exists another term $\xi_{iv}\xi_{jv} = -\xi_{iu}\xi_{ju}$. Hence $\sum_u \xi_{iu}\xi_{ju}=0$. Similarly $\sum_u \xi_{iu}^3=\sum_u \xi_{iu}^2\xi_{ju}=0$.

This requirement that odd moments be zero is necessary for rotatability and convenient for orthogonality, and in most cases does not restrict the choice of design to any significant degree.

The effect of the restriction is to make the linear and interaction terms orthogonal to each other and to all other terms. Hence the parts of S corresponding to these terms are diagonal. Thus it is possible to give the estimates for β_i and β_{ij} ($i \neq j$) immediately as

$$b_i = \sum_u \xi_{iu} y_u / \sum_u \xi_{iu}^2$$

$$b_{ij} = \sum_u \xi_{iu} \xi_{ju} y_u / \sum_u \xi_{iu}^2 \xi_{ju}^2 \qquad i \neq j$$

However, the quadratic terms are still non-orthogonal to each other and to the constant term. Suppose now that the quadratic functions ξ_i^2 are replaced by a quadratic polynomial

$$\zeta_{iu} = \xi_{iu}^2 + \theta_i \xi_{iu} + \mu_i$$

The values of θ_i and μ_i may be selected to improve orthogonality. The orthogonality conditions are given in table 3.

Table 3

Orthogonality conditions on θ_i, μ_i

To achieve orthogonality of ζ_{iu} with	Requirement
1. Constant term	$\sum_u \zeta_{iu} = 0$
2. Linear terms	$\sum_u \zeta_{iu} \xi_{ju} = 0$
3. Interaction terms	$\sum_u \zeta_{iu} \xi_{ju} \xi_{ku} = 0 \quad j \neq k$
4. Other quadratic terms	$\sum_u \zeta_{iu} \xi_{ju}^2 = 0$

Using the restriction on odd moments, and expanding ζ_{iu}, condition 3 is automatically satisfied.

Condition 1 gives

$$\sum_u \xi_{iu}^2 + n\mu_i = 0$$

whence
$$\mu_i = -\frac{1}{n}\sum_u \xi_{iu}^2$$

Condition 2 is automatically satisfied for $i \neq j$.

When $i=j$,

$$\theta_i \sum_u \xi_{iu}^2 = 0$$

or
$$\theta_i = 0$$

Hence
$$\zeta_{iu} = \xi_{iu}^2 - \frac{1}{n}\sum_u \xi_{iu}^2 \qquad (3.2)$$

and
$$\sum_u \zeta_{iu} = 0$$

Thus ζ_{iu} is orthogonal to all terms except those in ζ_{ju}. To ensure this orthogonality for $i \neq j$, condition 4 must be satisfied, or

$$\sum_u \xi_{iu}^2 \xi_{ju}^2 = \frac{1}{n}\sum_u \xi_{iu}^2 \sum_u \xi_{ju}^2 \qquad i \neq j \qquad (3.3)$$

This is equivalent to the requirement that ξ_{iu}^2 and ξ_{ju}^2 have zero covariance, in any of the $\frac{1}{2}p(p-1)$ combinations. Since n may be adjusted, by the addition of centre points, which do not affect any of the summations, if

$$\sum_u \xi_{iu}^2 \sum_u \xi_{ju}^2 / \sum_u \xi_{iu}^2 \xi_{ju}^2$$

is an integer greater than n, the design may be made orthogonal by the addition of centre points, thus increasing n to satisfy (3.3).

Consider, for example, the three-way central composite design mentioned earlier, with the size of the simplex part of the design made general.

The design is

$$(\pm 1, \pm 1, \pm 1),\ (\pm\delta, 0, 0),\ (0, \pm\delta, 0),\ (0, 0, \pm\delta)$$

and

$$\sum_u \xi_{iu}^2 = 8 + 2\delta^2 \qquad i=1,2,3$$

$$\sum_u \xi_{iu}^2 \xi_{ju}^2 = 8 \qquad i \neq j$$

$$n = 14 + n_c$$

where n_c is the number of central points (0,0,0). Condition 4 then requires

$$\sum_u \xi_{iu}^2 \sum_u \xi_{ju}^2 / \sum_u \xi_{iu}^2 \xi_{ju}^2 = \tfrac{1}{2}(4 + \delta^2)^2 = 14 + n_c$$

if the design is to be made orthogonal. This requires that $(4 + \delta^2)^2$ be an even integer. Practical possibilities for this integer are 30, 32, 34, and so forth. If $n_c = 4$, $(4 + \delta^2)^2 = 36$, $\delta = \sqrt{2}$. Hence the levels and numbers of points must both be taken into consideration.

When condition 4 is satisfied,

$$b_{ii} = \sum_u \zeta_{iu} y_u / \sum_u \zeta_{iu}^2$$

The only estimate of the coefficients in (3.1) that is altered by the transformation to ζ_{iu}, is that for β_0. Using the transformation, the element of S^{-1} corresponding to β_0 is 1/n. Hence, in the transformed model, from the formula $\underset{\sim}{b} = S^{-1}X^T\underline{y}$,

$$b_0 = \frac{1}{n} \sum_u y_u = \bar{y}$$

Thus, tests of hypotheses on the transformed b_0 are, in fact, tests on the sample mean.

The estimates and regression sums of squares for the situation in which condition 4 is met are given in table 4. The only part of the table that does not apply in the general case is that for β_{ii}.

Table 4

Estimates and regression sums of squares when individual coefficients are to be tested

Coefficient	Estimate	S.S.	D.F.
β_0 (mean)	$b_0 = \bar{y}$	$b_0 \sum_u y_u$	1
β_i	$b_i = \sum_u \xi_{iu} y_u / \sum_u \xi_{iu}^2$	$b_i \sum_u \xi_{iu} y_u$	1
β_{ii} (orthogonal case only)	$b_{ii} = \sum_u \zeta_{iu} y_u / \sum_u \zeta_{iu}^2$	$b_{ii} \sum_u \zeta_{iu} y_u$	1
β_{ij} $(i \neq j)$	$b_{ij} = \dfrac{\sum_u \xi_{iu} \xi_{ju} y_u}{\sum_u \xi_{iu}^2 \xi_{ju}^2}$	$b_{ij} \sum_u \xi_{iu} \xi_{ju} y_u$	1

Quadratic terms when orthogonality does not hold

In the event that the condition leading to orthogonality between different quadratic terms does not hold, it will be necessary to invert that submatrix of S that pertains to the quadratic terms. Denote this submatrix by Q. That is, Q is the submatrix whose (i,j)th element is $\sum_u \zeta_{iu} \zeta_{ju}$. The corresponding elements of b will be called $\underset{\sim}{b}_Q$, or

$$\underset{\sim}{b}_Q = (b_{11} \ldots b_{pp})^T$$

and those of the appropriate part of the transformed X matrix will be called Z. That is, the (u,i)th element of Z will be ζ_{iu}.

Now

$$Q = Z^T Z$$

$$\underset{\sim}{b}_Q = Q^{-1} Z^T \underset{\sim}{y}$$

In order to test a subset of the β_{ii}, together with their associated effects on the constant term, it is necessary to refit the model. To do this, one must specify which coefficients, of the β_{ii}, are to be tested, which are taken to be already fitted, and which are to be ignored. These latter are accounted for by deleting the corresponding columns of Z and thereafter ignoring them. Thus, without loss of generality, those to be ignored may be disregarded entirely, assuming that $\beta_{11}, \dots \beta_{pp}$ consist only of those to be tested and those considered already fitted. Assume that the first p_1 elements of $\underset{\sim}{b}_Q$ have been fitted, and that the last $p-p_1$ are to be tested. Now assume that Q, Z and $\underset{\sim}{b}_Q$ are appropriately partitioned. That is,

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \qquad Z = (Z_1 \ \ Z_2) \qquad \underset{\sim}{b}_Q = \begin{pmatrix} \underset{\sim}{b}_{Q1} \\ \underset{\sim}{b}_{Q2} \end{pmatrix}$$

The reduction in the residual sum of squares arising from fitting $\underset{\sim}{b}_Q$ is

$$\underset{\sim}{b}_Q^T Z^T \underset{\sim}{y} = \underset{\sim}{y}^T Z Q^{-1} Z^T \underset{\sim}{y}$$

and that from fitting $\underset{\sim}{b}_{Q1}$ alone is

$$\underset{\sim}{b}_{Q1}^T Z_1^T \underset{\sim}{y} = \underset{\sim}{y}^T Z_1 Q_{11}^{-1} Z_1^T \underset{\sim}{y}$$

Hence the improvement from fitting $\underset{\sim}{b}_{Q2}$ is

$$\underset{\sim}{y}^T (Z Q^{-1} Z^T - Z_1 Q_{11}^{-1} Z_1^T) \underset{\sim}{y} \qquad (3.4)$$

Representing Q^{-1} by P, with suitable partitioning, and using the formula for the inverse of a partitioned matrix,

$$Q_{11}^{-1} = P_{11}-P_{12}P_{22}^{-1}P_{21}$$

Also

$$ZPZ^T = Z_1P_{11}Z_1^T+2Z_1P_{12}Z_2^T+Z_2P_{22}Z_2^T$$

since $P_{12}=P_{21}^T$.

The bracketed expression in (3.4) now becomes

$$\begin{aligned}Z_1P_{11}Z_1^T+2Z_1P_{12}Z_2^T+Z_2P_{22}Z_2^T-Z_1(P_{11}-P_{12}P_{22}^{-1}P_{21})Z_1^T \\ = (P_{21}Z_1+P_{22}Z_2)^TP_{22}^{-1}(P_{21}Z_1+P_{22}Z_2)\end{aligned}$$

Hence the improvement in the residual sum of squares from fitting b_{Q2} is

$$\begin{aligned}\underset{\sim}{y}^T(P_{21}Z_1+P_{22}Z_2)^TP_{22}^{-1}(P_{21}Z_1+P_{22}Z_2)\underset{\sim}{y} &= \underset{\sim}{b}_{Q2}^TP_{22}^{-1}\underset{\sim}{b}_{Q2} \\ &= \underset{\sim}{b}_{Q2}^TQ_{22}\underset{\sim}{b}_{Q2}-\underset{\sim}{b}_{Q2}^TQ_{21}Q_{11}^{-1}Q_{12}\underset{\sim}{b}_{Q2} \quad (3.5)\end{aligned}$$

When $\underset{\sim}{b}_{Q2}$ consists of a single element, b_{ii}, say, this reduces to

$$b_{ii}^2/p_{ii}$$

where p_{ii} is the ith diagonal element of P. This enables a test to be made of the hypothesis $b_{ii}=0$, in the presence of the other quadratic coefficients.

Special forms for Q

The above analysis covers the case of general Q. However, in many cases it will be found that Q can be put into the form

$$Q = \Delta + \underset{\sim}{\gamma}\underset{\sim}{\gamma}^T$$

where Δ is easily inverted (usually diagonal) and

$\underset{\sim}{\gamma}$ is some vector. This pattern arises particularly in the case of permutation designs (which include central composite designs based on full factorials). These designs are such that if, for each point each co-ordinate ξ_i is divided by the scale factor $(\sum_u \xi_{iu}^2)^{\frac{1}{2}}$, then for any particular point, every permutation of these standardized co-ordinates exists in the design. Thus, if, using ξ_{iu}^* for the standardized co-ordinates, there exists a point

$$(\xi_{iu}^* \ldots \xi_{pu}^*)$$

then for every permutation of these values, there exists a point (which may be the same point if the permuted co-ordinates are equal) whose co-ordinates are these permuted values. This arrangement has the effect that $\sum_u \zeta_{iu}\zeta_{ju}$, $i \neq j$, has the form $c\sum_u \xi_{iu}^2 \sum_u \xi_{ju}^2$ where c is a constant, independent of i or j. Thus $\underset{\sim}{\gamma}$ is proportional to $(\sum_u \xi_{1u}^2 \ldots \sum_u \xi_{pu}^2)^T$. Then the ith diagonal element of the diagonal matrix is

$$\sum_u \xi_{iu}^4 - (c+\frac{1}{n})(\sum_u \xi_{iu}^2)^2 \qquad i=1, \ldots p$$

The inverse of this special form of Q is readily calculated as

$$Q^{-1} = \Delta^{-1} - \frac{1}{\mu}\Delta^{-1}\underset{\sim}{\gamma}\,{}^T\underset{\sim}{\gamma}\Delta^{-1}$$

where μ is the scalar $1+\underset{\sim}{\gamma}^T\Delta^{-1}\underset{\sim}{\gamma}$, thus Q^{-1} has the same form as Q.

If Δ is block diagonal, with blocks Δ_s, $s=1, \ldots r$ and the corresponding blocks of Q^{-1} are P_s, then

$$P_s = \Delta_s^{-1} - \frac{1}{\mu}\Delta_s^{-1}\underset{\sim}{\gamma}_s\underset{\sim}{\gamma}_s^T\Delta_s^{-1}$$

which has inverse

$$P_s^{-1} = \Delta_s + \frac{1}{\mu-\mu_s}\underset{\sim}{\gamma}_s\underset{\sim}{\gamma}_s^T$$

where $\mu_s = \underset{\sim}{\gamma}_s^T \Delta_s^{-1} \underset{\sim}{\gamma}_s$.

Hence, using (3.5), the improvement from fitting the sth block, in the presence of the other coefficients is

$$\underset{\sim}{b}_{Qs}^T \Delta_s \underset{\sim}{b}_{Qs} + (\underset{\sim}{b}_{Qs}^T \underset{\sim}{\gamma}_s)^2/(1+\underset{\sim}{\gamma}^T \Delta^{-1} \underset{\sim}{\gamma} - \underset{\sim}{\gamma}_s^T \Delta_s^{-1} \underset{\sim}{\gamma}_s)$$

and if Δ is diagonal, the improvemant from fitting b_{ii} in the presence of the remaining quadratic coefficients is

$$b_{ii}^2 [d_{ii} + \underset{\sim}{\gamma}_i^2/(1+\sum_{j\neq i} \underset{\sim}{\gamma}_j^2/d_{jj})]$$

using d_{ii} for the diagonal elements of Δ.

Rotatable designs

The conditions for a second order design to be rotatable are (Box and Hunter (1957)) that all the moments containing an odd power be zero, and that the two kinds of standardized fourth moment each be constant. Also, the relationship

$$\lambda_4 = \frac{n\sum_u \xi_{iu}^2 \xi_{ju}^2}{\sum_u \xi_{iu}^2 \sum_u \xi_{ju}^2} = \frac{1}{3} \frac{n\sum_u \xi_{lu}^4}{(\sum_u \xi_{lu}^2)^2}$$

must hold for all i, j, and l. Thus λ_4 is the basic parameter for the design.

In the present notation these conditions become, using (3.2)

$$q_{ii} = \sum_u \zeta_{iu}^2 = (3\lambda_4 - 1)(\Sigma_u \xi_{iu}^2)^2/n$$

$$q_{ij} = \sum_u \zeta_{iu}\zeta_{ju} = (\lambda_4 - 1)\sum_u \xi_{iu}^2 \sum_u \xi_{ju}^2/n$$

Now, using the notation of the previous subsection,

$$\underset{\sim}{\gamma} = \left(\frac{\lambda_4 - 1}{n}\right)^{\frac{1}{2}} \left(\sum_u \xi_{1u}^2 \ \cdots \ \sum_u \xi_{pu}^2\right)^T$$

and

$$\Delta = \frac{2\lambda_4}{n} \begin{pmatrix} (\sum_u \xi_{1u}^2)^2 & & 0 \\ & \ddots & \\ 0 & & (\sum_u \xi_{pu}^2)^2 \end{pmatrix}$$

from which

$$\frac{1}{\mu} = \frac{1}{1+\underset{\sim}{\gamma}^T \Delta^{-1} \underset{\sim}{\gamma}} = \frac{2\lambda_4}{\lambda_4(p+2)-p}$$

Evidently, from the definition of λ_4 and equation (3.3), orthogonality is achieved if $\lambda_4=1$, which would imply $\underset{\sim}{\gamma}=0$.

The (i,j)th element of Q^{-1} is

$$\frac{n}{2\lambda_4}\left\{\frac{\delta_{ij}}{(\sum_u \xi_{iu}^2)^2} - \frac{\lambda_4 - 1}{\lambda_4(p+2)-p}\,\frac{1}{\sum_u \xi_{iu}^2 \sum_u \xi_{ju}^2}\right\} \qquad (3.6)$$

where δ_{ij} is the Kronecker delta and has the value 1 if i=j, and otherwise zero. From (3.6) the diagonal elements of Q^{-1} are

$$p_{ii} = \frac{n[\lambda_4(p+1)-p+1]}{2\lambda_4[\lambda_4(p+2)-p]}\,\frac{1}{(\sum_u \xi_{iu}^2)^2}$$

and

$$b_{ii} = \frac{n}{2\lambda_4 \sum_u \xi_{iu}^2}\left\{\frac{\sum_u \xi_{iu}^2 y_u}{\sum_u \xi_{iu}^2} - \frac{\lambda_4 - 1}{\lambda_4(p+2)-p} \sum_j \frac{\sum_u \xi_{ju}^2 y_u}{\sum_u \xi_{ju}^2}\right\}$$

from which the effect of fitting b_{ii} is easily calculated by b_{ii}^2/p_{ii}.

4. Overlaid Experimental Designs

Introduction

As noted in section 1, the control of error by blocking has been considered by a number of authors. The design requirements in this case are recapitulated below.

It is natural to extend this to the two-way classification situation, both with and without interaction. Further extension, to multiple classification models is likely to make the scheme unwieldy in practice, but is conceptually straightforward.

Another natural development is to assume that more accurate information may be wanted on the classification part of the design than on the regression part. In this situation, a split-plot arrangement might be used, with closely related sub-plots containing representatives of each of the classification treatments, and each whole plot concerned bearing only one combination of the regression treatments. Alternatively, the emphasis may be placed on the regression part of the model.

All these designs are generalizations of the analysis of covariance model, except that the regression aspect is fully analysed.

General experimental design model

In the succeeding discussion, the overall mean will be assumed to be part of the regression model rather than the experimental design or qualitative model.

The design matrix for these compound designs will be represented by the partitioned matrix

$$W = (D\ X)$$

where X is the regression design matrix discussed earlier, and D is the superimposed experimental design matrix.

Suppose now that D, (mxr), is an arbitrary design matrix, with the imposed constraint that the sum of the r effects is zero (in order to include the overall mean in the regression model). This constraint can now be used, as in normal experimental design, to reparameterize the qualitative part of the model in order to make all the effects orthogonal to the mean. This means that, where $\underset{\sim}{j}_m$ is an mx1 vector, all of whose elements are unity,

$$\underset{\sim}{j}_m^T D = 0$$

where D is the reparameterized design matrix.

Now generate a design in which the whole design matrix D is repeated n times, each repetition corresponding to one point of some regression design with matrix X. If the rows of X are represented by $\underset{\sim}{x}_1^T, \ldots \underset{\sim}{x}_n^T$, the overall design now has the form

$$W = \left(\begin{array}{c|c} D & \begin{matrix}\underset{\sim}{x}_1^T\\ \vdots \\ \underset{\sim}{x}_1^T\end{matrix} \\ \hline & \cdots \\ \hline D & \begin{matrix}\underset{\sim}{x}_n^T\\ \vdots \\ \underset{\sim}{x}_n^T\end{matrix}\end{array}\right)$$

Consider now the submatrix

$$W_i = \left(\begin{array}{c|c} D & \begin{matrix}\underset{\sim}{x}_i^T\\ \vdots \\ \underset{\sim}{x}_i^T\end{matrix}\end{array}\right) = (D \quad \underset{\sim}{j}_m \underset{\sim}{x}_i^T)$$

Now

$$\begin{aligned} W_i^T W_i &= \begin{pmatrix} D^T \\ \underset{\sim}{x}_i \underset{\sim}{j}_m^T \end{pmatrix} (D \quad \underset{\sim}{j}_m \underset{\sim}{x}_i^T) \\ &= \begin{pmatrix} D^T D & D^T \underset{\sim}{j}_m \underset{\sim}{x}_i^T \\ \underset{\sim}{x}_i \underset{\sim}{j}_m^T D & \underset{\sim}{x}_i \underset{\sim}{j}_m^T \underset{\sim}{j}_m \underset{\sim}{x}_i^T \end{pmatrix} \\ &= \begin{pmatrix} D^T D & 0 \\ 0 & m\underset{\sim}{x}_i \underset{\sim}{x}_i^T \end{pmatrix} \end{aligned}$$

and

$$W^T W = \sum_i W_i^T W_i = \begin{pmatrix} nD^T D & 0 \\ 0 & mX^T X \end{pmatrix}$$

hence, with this arrangement, qualitative and quantitative effects are orthogonal. Thus, any experimental design in which the effects may be made orthogonal to the mean may be combined with an arbitrary rsponse surface model in such a way that qualitative and quantitative effects may be estimated separately.

The simplest way to analyse any of these "full replication" designs, is to analyse the qualitative model, regarding the quantitative points as a further classification effect, akin to replications, then sum over the qualitative model before fitting the regression. An interaction between elements of the qualitative design, and the replications arising from the quantitative model would indicate that the regression model was dependent on the classification, and varied according to the classification model effects involved.

However, the above strategy of repeating the entire design could well be extravagant in experimental points. In practice, more compact designs are possible.

In the material that follows, the X designs are assumed to satisfy the constraint requiring zero odd moments.

<u>One-way classification</u>

The simplest design is the one-way classification, with design matrix, before reparameterization,

$$D = \begin{pmatrix} \underset{\sim}{j}_{n_1} & & 0 \\ & \ddots & \\ 0 & & \underset{\sim}{j}_{n_r} \end{pmatrix} \qquad (4.1)$$

with $n_1 + \dots + n_r = n$. The design matrix, after reparameterization, becomes

$$D = \begin{pmatrix} \underset{\sim}{j}_{n_1} & & \\ & \ddots & \\ & & \underset{\sim}{j}_{n_r} \end{pmatrix} - \frac{1}{n} \underset{\sim}{j}_n \underset{\sim}{n}^T$$

where $\underset{\sim}{n}^T = (n_1 \dots n_r)$. This has the required property that $\underset{\sim}{j}_n^T D = 0$.

Now introduce the subscript w, to range over the classification effects, giving independent regression variables ξ_{iwu}, $i=1, \dots p$, $w=1, \dots r$, $u=1, \dots n_w$. The full design matrix, using D defined above, is $W=(D\ X)$. Under these conditions, the requirement for orthogonality between the linear term and the ith regression variable, and the wth classification effect is

$$\sum_{u=1}^{n_w} \xi_{iwu} - \frac{n_w}{n} \sum_{w=1}^{r} \sum_{u=1}^{n_w} \xi_{iwu} = 0$$

However, by the requirement of zero odd moments, the latter term must be zero. Hence the classification effects are orthogonal to the linear effects if

$$\sum_{u=1}^{n_w} \xi_{iwu} = 0 \text{ for all } i,\ w$$

The equivalent for the regression interaction effects is derived in an identical way, and gives

$$\sum_{u=1}^{n_w} \xi_{iwu} \xi_{jwu} = 0 \text{ for all } i \neq j, \text{ all } w$$

For quadratic effects, the requirement is

$$\sum_{u=1}^{n_w} \xi_{iwu}^2 = \frac{n_w}{n} \sum_{w=1}^{r} \sum_{u=1}^{n_w} \xi_{iwu}^2$$

which is identical to the proportionate variance requirement derived in a more intuitive manner by Box and Hunter (1957).

Two-way classification designs

In the two-way classification, without classification interaction, the one-way conditions must be satisfied for each of the classifications. In addition, if the two sets of classification effects are to be orthogonal (using v as the subscript for the second classification)

$$n_{wv} = n_w n_v / n$$

is required.

If a classification interaction term is included in the model, a further condition is required to ensure that this interaction is orthogonal to the quadratic term in the regression.

In this case the model needs to be reparameterized in such a way that the main effects and classification interaction are orthogonal to the mean. The reparameterization is more complex, and is not readily expressible in matrix notation. Table 5 summarizes the distinct elements of the design matrix for a particular combination (w,v).

In this form the columns of the matrix may be added readily, to show that each column sum is zero, and hence that each effect is orthogonal to the overall mean. The condition for orthogonality between main effects is conceptually straightforward and is obtained from:

Sum over rows of table (Number of similar rows x w-treatment effect x v-treatment effect) = 0

which, noting the common factor $(n_w+n_{wv})(n_v+n_{wv})$ is

$$(n_w+n_{wv})(n_v+n_{wv})\left(\frac{n_{wv}}{n_w n_v} - \frac{1}{n}\right) = 0$$

Table 5

Summary of reparameterized design matrix for the two-way classification with interaction case.

From	Number of similar rows	w effect	v effect	(w,v) effect
		$(n_w+n_{wv})x$	$(n_v+n_{wv})x$	$n_{wv}x$
Both w & v	n_{wv}	$\frac{1}{n_w}-\frac{1}{n}$	$\frac{1}{n_v}-\frac{1}{n}$	$\frac{1}{n_{wv}}-\frac{1}{n_w}-\frac{1}{n_v}+\frac{1}{n}$
w but not v	n_w-n_{wv}	$\frac{1}{n_w}-\frac{1}{n}$	$-\frac{1}{n}$	$-(\frac{1}{n_w}-\frac{1}{n})$
v but not w	n_v-n_{wv}	$-\frac{1}{n}$	$\frac{1}{n_v}-\frac{1}{n}$	$-(\frac{1}{n_v}-\frac{1}{n})$
Other terms	$n-n_w-n_v+n_{wv}$	$-\frac{1}{n}$	$-\frac{1}{n}$	$\frac{1}{n}$

Hence, as before, the orthogonality requirement, for main effects, is

$$n_{wv} = n_w n_v / n$$

The condition for orthogonality of the w-effect and the interaction, obtained in the same way, is

$$n_{wv}(n_w+n_{wv})(\frac{1}{n} - \frac{n_{wv}}{n_w n_v}) = 0$$

which leads to the same condition as before.

In discussing orthogonality between classification effects and regression effects, it is convenient to define

$$\phi_{wvu} = \xi_{iwvu},\ \xi_{iwvu}\xi_{jwvu},\ \text{or } \xi^2_{iwvu}$$

as required. Also, use the dot notation to denote summation, thus for example $\Phi_{...} = \sum_{wvu}\phi_{wvu}$.

Now the orthogonality requirements can be derived quite quickly from table 5, and summarized as:

For orthogonality of regression effects with:

w - effects: $\frac{1}{n_w}\Phi_{w..} - \frac{1}{n}\Phi_{...} = 0$

v - effects: $\frac{1}{n_v}\Phi_{.v.} - \frac{1}{n}\Phi_{...} = 0$

w, v interaction: $\frac{1}{n_{wv}}\Phi_{wv.} - \frac{1}{n_w}\Phi_{w..} - \frac{1}{n_v}\Phi_{.v.} + \frac{1}{n}\Phi_{...} = 0$

or, using the main effects conditions in the interaction condition,

$$\frac{1}{n_{wv}}\Phi_{wv.} = \frac{1}{n}\Phi_{...}$$

The main effects conditions are implied by this condition, which is thus the requirement for orthogonality of regression effects and classification effects. In terms of the original regression variables, this condition becomes

$$\sum_u \xi_{iwvu} = \sum_u \xi_{iwvu}\xi_{jwvu} = 0$$

$$\sum_u \xi^2_{iwvu} = \frac{n_{wv}}{n} \sum\sum\sum_{wvu} \xi^2_{iwvu}$$

As an illustration, consider the case in which the whole of a response surface design is repeated for each w, v combination. In this case the first two conditions above are automatically satisfied. Also

$$\sum_u \xi^2_{iwvu} = \kappa, \text{ say}$$

is constant, and since n_{wv} is constant (equal to the size of the regression design) the last condition reduces to

$$\kappa = \frac{n_{wv}}{n} \left(\frac{n}{n_{wv}} \kappa\right)$$

which is also satisfied. Thus the design satisfies the conditions, in agreement with the general result on the full replication model derived earlier.

Regression model dependent on classification

A natural extension of the classification model is to allow the regression coefficients to vary with the classification effect. Thus $\underset{\sim}{\beta}$ is replaced by a series $\underset{\sim}{\beta}_w$, w=1, ... r. To achieve orthogonality with this scheme, the full replication type of design is required.

The simplest and most natural way to proceed, is to fit the regressions separately, and combine the results for an overall analysis of variance afterward.

5. Summary

After a general development of the theory behind response surface methodology, with particular reference to polynomial models and rotatable designs, section 1 of this thesis gives a rigorous development of the justification of the standard F-tests used. In particular, the lack of fit test, using an error estimate based on point replication, is justified.

Section 2 surveys the literature relating to response surfaces, with the exception of recent bias-oriented work. The emphasis is on theoretical development, rather than on applications.

In section 3 the analysis of second order designs is considered in some detail, the aim being to provide a means of testing hypotheses about individual coefficients. This separation is acheived for the slightly restricted case in which the dependent variables have zero odd moments about their means. Conditions for orthogonality between different quadratic terms are developed. Methods are derived for testing subsets of the quadratic terms when orthogonality does not hold, and, in particular, the case in which the part of the sums of squares and cross-products matrix relating to the quadratic terms has a particular form is considered. Finally, the ideas developed are

applied to rotatable designs.

Section 4 considers the combination of response surface designs with ordinary qualitative experimental designs. The first design considered is a general one enabling the combination of any response surface design with any experimental design in which the mean is orthogonal to the treatment effects. Since the design described is likely to be extravagant in experimental points, consideration is given to orthogonality conditions for general one- and two-way classification designs. In particular, the two-way classification with interaction is studied in some detail. Finally, brief mention is made of the possibility of regression coefficients differing with treatments.

6. References

ANDERSON, R.L. (1953) "Recent advances in finding best operating conditions." J.Am.Statist.Assoc 48 789-798

BOSE, R.C. & CARTER, R.L. (1959) "Complex representation in the construction of rotatable designs." Ann.Math.Statist. 30 771-780

BOSE, R.C. & DRAPER, N.R. (1959) "Second order rotatable designs in three dimensions." Ann.Math.Statist. 30 1097-1112

BOX, G.E.P. (1952) "Multifactor designs of first order." Biometrika 39 49-57

- (1954a) Discussion on the symposium on interval estimation. J.R.Statist.Soc. B 16 211-212

- (1954b) "The exploration and exploitation of response surfaces: some general considerations and examples." Biometrics 10 16-60

- (1959) Answer to query: "Replication of non-centre points in the rotatable and near-rotatable central composite design." Biometrics 15 133

BOX, G.E.P. & BEHNKEN, D.W. (1960a) "Simplex-sum designs: a class of second order rotatable designs derivable from those of first order." Ann.Math.Statist. 31 838-864

- (1960b) "Some new three level designs for the study of quantitative variables." Technometrics 2 455-475, erratum 3 576

BOX, G.E.P. & DRAPER, N.R. (1959) "A basis for the selection of a response surface design." J.Am.Statist.Assoc. 54 622-654

BOX, G.E.P. & HUNTER, J.S. (1954) "A confidence region for the solution of a set of simultaneous equations with an application to experimental design." Biometrika 41 190-199

- (1957) "Multifactor experimental designs for exploring response surfaces." Ann.Math.Statist 28 195-241

BOX, G.E.P. & TIDWELL, P.W. (1962) "Transformations of the independent variable." Technometrics 4 531-550

BOX, G.E.P. & WILSON, K.B. (1951) "On the experimental attainment of optimum conditions." J.R.Statist.Soc. B 13 1-45

BOX, G.E.P. & YOULE, P.V. (1955) "The exploration and exploitation of response surfaces: an example of the link between the fitted surface and the basic mechanism of the system." Biometrics 11 287-323

BOX, M.J. (1968) "The occurrence of replications in optimal designs of experiments to estimate parameters in non-linear models." J.R.Statist.Soc. 30 290-302

CHERNOFF, H. (1953) "Locally optimal designs for estimating parameters." Ann.Math.Statist. 24 586-602

DAS, M.N. (1961) "Construction of rotatable designs from factorial designs."

J.Indian Soc. Agric. Statist. 13 169-194

- (1963) "On the construction of second order rotatable designs through balanced incomplete block designs with blocks of unequal size."

Bull.Calcutta Stat. Assoc. 12 31-46

DAS, M.N. & NARASIMHAN, V.L. (1962) "Construction of rotatable designs through balanced incomplete block designs." Ann.Math.Statist. 33 1421-1439

DAVID, H.A. & ARENS, B.E. (1959) "Optimal spacing in regression analysis."

Ann.Math.Statist. 30 1072-1081

DAVIES, O.L. (1954) The design and analysis of industrial experiments. Hafner, New York, Ch 11, 495

DEBAUN, R.M. (1956) "Block effects in the determination of optimum conditions." Biometrics 12 20-22

- (1959) "Response surface designs for three factors at three levels." Technometrics 1 1-8

DEBAUN, R.M. & SCHNEIDER, A.M. (1958) "Experiences with response surface designs." In V.Chew (ed.) Experimental designs in industry

John Wiley & Sons, New York, p235

DE LA GARZA, A. (1954) "Spacing of information in polynomial regression."

Ann.Math.Statist. 25 123-130

DRAPER, N.R. (1960a) "Second order rotatable design in four or more dimensions."
Ann.Math.Statist. 31 23-33

\- (1960b) "Third order rotatable design in four dimensions." Ann.Math.Statist. 31 865-874

\- (1960c) "Third order rotatable design in three dimensions." Ann.Math.Statist. 31 875-877

\- (1961a) "Missing values in response surface designs." Technometrics 3 389

\- (1961b) "Third order rotatable designs in three dimensions: some specific designs."
Ann.Math.Statist. 32 910-913

\- (1962) "Third order rotatable designs in three dimensions: analysis." Technometrics 4 219-234

DRAPER, N.R. & HERZBERG, A.M. (1968) "Further second order rotatable designs."
Ann.Math.Statist. 39 1995-2001

DRAPER, N.R. & STONEMAN, D.M. (1968) "Response surface designs for factors at two and three levels and at two and four levels." Technometrics 10 177-192

DYKSTRA, O. (1950) "Partial duplication of factorial experiments." Technometrics 1 63-75

\- (1960) "Partial duplication of response surface designs." Technometrics 2 185-195

EHRENFELD, S. (1955) "On the efficiency of experimental designs." Ann. Math. Statist. 26 247-255

ELFVING, G. (1952) "Optimum allocation in linear regression theory." Ann.Math.Statist. <u>23</u> 255-262

FOLKS, J.L. (1958) "Comparison of designs for exploration of response relationships." Ph.D. Thesis, Iowa State College

GARDINER, D.A., GRANDAGE, A.H.E. & HADER, R.J. (1959) "Third order rotatable designs for exploring response surfaces." Ann.Math.Statist. <u>30</u> 1082-1096

GRAYBILL, F.A. (1961) <u>Linear statistical models</u>, Vol 1 McGraw-Hill

GRILICHES, Z. (1957) "Specification bias in estimates of production functions." Farm Economics <u>39</u> 8-20

GUEST, P.G. (1958) "The spacing of observations in polynomial regression." Ann.Math.Statist. <u>29</u> 294-299

HARTLEY, H.O. (1959) "Smallest composite design for quadratic response surface." Biometrics <u>15</u> 611-624

HERMANSON, H.P., GATES, C.E., CHAPMAN, J.W. & FARNHAM, R.S. (1964) "An agronomically useful three-factor response surface design based on dodecahedron symmetry." Ann.Math.Statist. <u>35</u> 445-446

HERZBERG, A.M. (1964) "Two third order rotatable designs in four dimensions." Ann.Math.Statist. <u>35</u> 445-446

\- (1966) "Cylindrically rotatable designs." Ann.Math.Statist. <u>37</u> 242-247

HERZBERG, A.M. (1967a) "A method for the construction of second order rotatable designs in k dimensions." Ann.Math.Statist. 38 177-180

- (1967b) "Cylindrically rotatable designs of types 1, 2, and 3." Ann.Math.Statist. 38 167-176

HILL, W.J. & HUNTER, W.G. (1966) "A review of response surface methodology: a literature survey." Technometrics 8 571-590

HOEL, P.G. (1958) "Efficiency problems in polynomial regression." Ann.Math.Statist. 29 1134-1145

HUNTER, J.S. (1954) "Searching for optimum conditions." Trans.N.Y.Acad.Sci., Series 11 17

- (1956) "Statistical methods for determining optimum conditions." Trans.Tenth Annual Convention A.S.Q.C. p415

KIEFER, J. (1958) "On the nonrandomized optimality and randomized non-optimality of symmetric designs." Ann.Math.Statist. 29 675-699

- (1959) "Optimum experimental designs." J.R.Statist.Soc. B 21 272-319

KIEFER, J. & WOLFOWITZ, J. (1959) "Optimum designs in regression problems." Ann.Math.Statist. 30 271-294

KITAGAWA, T. (1959) "Successive process of statistical inference applied to linear regression analysis and its specialization to response surface analysis." Bull.Math.Statist. 8 80-114

PLACKETT, R.L. (1960) Principles of regression analysis. Oxford U.P.

UMLAND, A.W. & SMITH, W.N. (1959) "The use of LaGrange multipliers with response surfaces." Technometrics 1 289