OPTIMAL PROPERTIES OF EXPERIMENTAL DESIGNS

by

Chit Swe

Thesis submitted for M.Sc. degree of the University of Glasgow

in November, 1962.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

ACKNOWLEDGEMENT

Heartiest thanks are due to Dr. R. A. Robb for his general supervision, generous guidance and continued interest in the preparation of this thesis and also Mr. J. Aitchison for helpful and stimulating discussions.

Glasgow, November 1962. Chit Swe.

CONTENTS

Notations.

β = Power function.

$\underline{\beta}$ = vector of parameters i.e. $\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$

$\underline{\gamma}$ = A subvector of $\underline{\beta}$

i.e. a vector whose elements are β_i's

p = the number of parameters .

$\underline{p}$ = $\begin{pmatrix} p_1 \\ \vdots \\ p_r \end{pmatrix}$ = allocation vector , where $\sum p_i = 1$

δ = noncentrality parameter

δ_{ij} = kronecker delta

Chapter I

Introduction

A theory of efficient design of statistical investigations has been developed by R. A. Fisher, F. Yates and their followers mainly in connection with agricultural experimentation. Fisher's contributions to logic and scientific methods of experimentation of theory of experimental designs are outstanding. These methods, though developed mainly to be used in scientific investigations of agronomical problems are in fact general, to be applied to other fields also.

There are four broad categories into which problems of experimental design can be classified:

(1) The practical problem of deciding which experiments are relevant to the problems under consideration.

(2) The combinatorial aspects of construction of designs and the associated statistical analysis.

(3) The study of the optimum properties of well established designs.

(4) Construction of optimal designs under various criteria of optimality.

Most of the work done on design of experiments has concerned itself with the first two aspects while the remaining two have only recently been receiving attention. In early development of the subject, we have tried to devise sufficiently symmetrical plans,

so that the estimation of the relevant parameters and the associated test of significance are as comfortable as possible. There was little worrying about the best plans. However, most of the well established symmetrical designs are found to be highly efficient in terms of the recently established optimality criteria. Although much work has been done on the problem of optimal design in linear experiments in the past decade, all of them are found to be concentrated on establishing or denying optimum properties of well established designs, or on supplying sufficient conditions for designs to possess these optimum properties. Computational methods for the actual construction of optimal designs were investigated only recently.

A theory of design and analysis of experiment naturally has implications about how the experiment should be planned and performed and how the observations should be analysed so that we may have maximum information with minimum amount of risk and sometimes with minimum cost.

Formulation of the Problem.

By an experiment ξ we shall mean a set of random variables y_x the joint distribution of which depends on certain unknown parameters $\beta_1, \beta_2, \ldots \beta_p$, to be denoted by the vector $\underline{\beta}$. In all problems treated below the set (ξ) of possible experiments among which a choice is to be made is given. The purpose of the experimentsl investigation is

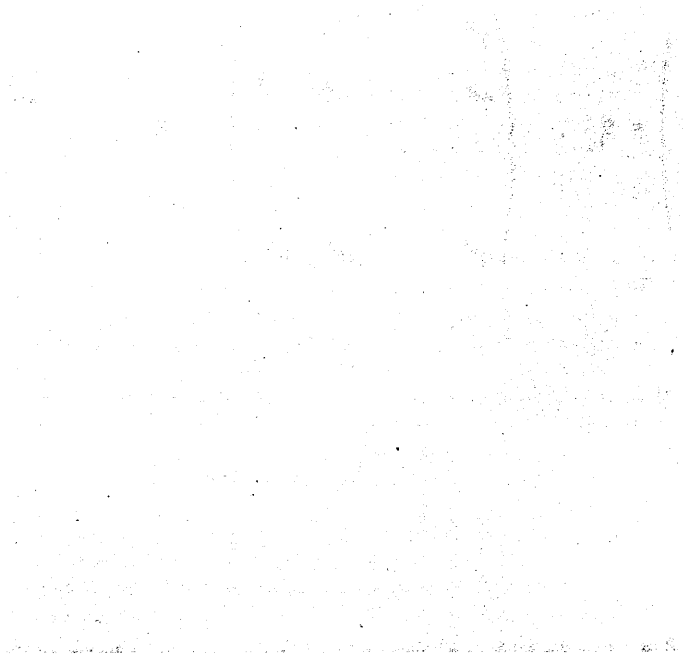
where the experiments are performed in such a way that $x_1, x_2, \cdots x_n$ are not random variables but have predetermined fixed values. It is known that the variance of the least square estimate (which is the same as the maximum likelihood estimate in this case because of normality) $\hat{\beta}$ of β is inversely proportional to $\sum_1^n (x_i - \bar{x})^2$ where $\bar{x} = \sum \frac{x_i}{n}$. Hence, if we can freely choose the values $x_1, \cdots x_n$, in a certain domain A, the greatest sensitivity of the test will be achieved by choosing so that $\left[\sum (x_i - \bar{x})^2\right]^{-1}$ becomes a minimum.

We shall later consider the various measures of the efficiency of the various designs of statistical investigations in connection with linear hypothesis.

It was A. Wald who first developed the modern theory of optimal statistical design by establishing a certain optimality criteria (now known as D-optimality) from the stand point of the power function of analysis of variance test. Then follows a thorough study of the subject by G. Elfving and S. Ehrenfeld of Scandinavian countries, P. G. Hoel, P. G. Guest, J. Kiefer, J. Wolfiwitz, E. J. Williams of United States, S. Moriguti and M. Masuyama of Japan and J. Aitcheson of Great Britain. These people have, on the basis of one or another goodness criteria, established some well known designs as optimal and/or developed methods of deriving best plans.

This thesis will survey various recent developments in the theory of optimal experimental designs. The primary aim of the thesis will be to stress certain ideas and principles which are easily set forth in

simple situations rather than to state things in the most possible setting and to try and apply the discussions to the more complicated problems. We shall be concerned with developments of optimality criteria based on the desirable properties of the design and methods for verifying whether or not given designs satisfy certain optimality criteria, the relation between them and some methods for computing optimal designs.



Chapter II

The General Linear Hypothesis

While, in some of the papers quoted below, more general set ups are considered, we shall in the following restrict entirely to linear experiments i.e. experiments in which the expected values are linear functions of the unknown parameters.

In matrix notation, the outcome of a linear experiment may be written

$$\underline{y} = \underline{X}\,\underline{\beta} + \underline{e}.$$

where

$$\underline{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \underline{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \quad \underline{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \underline{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

denote the observation vector, a known coefficient matrix, the elements of which the experimentor can freely choose, the unknown parameter vector and the error vector.

We may now list together, for easy reference, some points of

notation, definitions of various sets and functions which we may encounter in our theoretical development of the subject and some well established results of the general linear statistical model in connection with the estimation of the parameters and the tests of significance.

General Problems of a Linear Model.

In general, the parameters β_i's are not known and we shall have to estimate them using the prediction equation. In order to estimate the β_i a random sample of size n will be taken from the distribution $f(\underline{y}; \underline{X}, \underline{\beta})$. The sample will be denoted by $\underline{P}_j' = (y_j, x_{j1}, x_{j2}, \ldots\ldots\ldots, x_{jp})$ $j = 1, 2, 3.....m$ and the relations within the system of observations can be written

$$\underline{y} = \underline{X}\underline{\beta} + \underline{e} \qquad (2.1)$$

On examining $\underline{y} = \underline{X}\underline{\beta} + \underline{e}$, it can be seen that we must first select (either at random or by design) a set of x's, say $x_{j1}, x_{j2}, \ldots x_{jp}$, and then randomly select an observation y_j from the distribution $f(\underline{y}; \underline{x} = \underline{x}_j)$. We shall assume throughout that the model $\underline{y} = \underline{X}\underline{\beta} + \underline{e}$ was constructed by the sampling process above.

On the basis of the chosen matrix $\underline{X}$ and observed random variables $\underline{y}$ we shall now derive the estimators for $\underline{\beta}$ and for σ^2. We shall also study the distribution of pertinent statistics for estimation of certain parameters and functions of parameters and for testing of hypothesis about them.

2.1 General Linear Model of Full Rank.

Definition:- The model $\underset{\sim}{y}^{n\times 1} = \underset{\sim}{X}^{n\times p}\underset{\sim}{\beta}^{p\times 1} + \underset{\sim}{e}^{n\times 1}$ will be called the General Linear Hypothesis Model of full rank, if the rank of $\underset{\sim}{X}$ is equal to p where $p \leq n$.

Two cases concerning the distribution of $\underset{\sim}{e}$ will be considered.

Case A (Normal Theory Case) = $\underset{\sim}{e}$ is distributed $N(\underset{\sim}{0}, \sigma^2 \underset{\sim}{I})$, where is unknown.

Case B. (Distribution un specified). $\underset{\sim}{e}$ is a random vector such that $E(\underset{\sim}{e}) = \underset{\sim}{0}$ and $Cov(\underset{\sim}{e}) = \sigma^2 \underset{\sim}{I}$ where is unknown.

For point estimation we shall study both case A and case B, but for interval estimation and testing of hypothesis we shall restrict ourselves to normal theory situation only.

2.1.1 Point Estimation.

Under normal theory the parameters $\underset{\sim}{\beta}$ and σ^2 are estimated by Maximum Likelihood method resulting in the following theorem.

Theorem: 2.1 If $\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{e}$ is a general linear hypothesis model of full rank and if $\underset{\sim}{e}$ is distributed $N(\underset{\sim}{0}, \sigma^2 I)$ the estimators

$$\hat{\underset{\sim}{\beta}} = \underset{\sim}{M}^{-1}\underset{\sim}{X}'\underset{\sim}{y} \qquad \text{where} \quad \underset{\sim}{M} = \underset{\sim}{X}'\underset{\sim}{X}$$

$$\hat{\sigma^2} = \frac{\underset{\sim}{y}'(\underset{\sim}{I} - \underset{\sim}{X}\underset{\sim}{M}^{-1}\underset{\sim}{X}')\underset{\sim}{y}}{n-p}$$

have the following properties.

(1) Consistent.

(2) Efficient.

(3) Unbiased.

(4) Sufficient.

(5) $\hat{\underline{\beta}}$ is distributed $N(\underline{0}, \sigma^2 \underline{M}^{-1})$.

(6) Complete.

(7) Minimum variance unbiased.

(8) $(n-p)\frac{\hat{\sigma}^2}{\sigma^2}$ is distributed $\chi^2(n-p)$.

(9) $\hat{\underline{\beta}}$ and $\hat{\sigma}^2$ are independent.

Proof. The likelihood function of our sample, since the error vector $\underline{e}$ is distributed $N(\underline{0}, \sigma^2 \underline{I})$, is

$$L(\underline{e}; \underline{\beta}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}(\underline{e}'\underline{e})}$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}(\underline{y}-\underline{X}\underline{\beta})'(\underline{y}-\underline{X}\underline{\beta})}$$

$$\therefore \log L(\underline{e}; \underline{\beta}, \sigma^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(\underline{y}-\underline{X}\underline{\beta})'(\underline{y}-\underline{X}\underline{\beta}).$$

Maximum likelihood estimators of $\underline{\beta}$ and σ^2 are now given by

$$\frac{\partial}{\partial \underline{\beta}} \log L(\underline{e}; \underline{\beta}, \sigma^2) = \underline{0}$$

$$\frac{\partial}{\partial \sigma^2} \log L(\underline{e}; \underline{\beta}, \sigma^2) = 0$$

i.e. $$\frac{2}{2\sigma^2}(\underline{X}'\underline{y} - \underline{X}'\underline{X}\underline{\beta}) = \underline{0}$$

$$-\frac{n}{2\sigma^2} + \frac{(\underline{y}-\underline{X}\underline{\beta})'(\underline{y}-\underline{X}\underline{\beta})}{2\sigma^4} = 0.$$

The solutions $\tilde{\underline{\beta}}$ and $\tilde{\sigma}^2$ are given by

$$\underline{X}'\underline{X}\underline{\beta} = \underline{X}'\underline{y}$$, known as the Normal Equations,

and $\tilde{\sigma}^2 = \frac{1}{n}(\underline{y} - \underline{X}\tilde{\underline{\beta}})'(\underline{y} - \underline{X}\tilde{\underline{\beta}})$

Since rank $\underline{X}$ = rank $(\underline{X}'\ \underline{X})$ = p , $\underline{X}'\underline{X}$ is full rank and therefore has an inverse,

$$\tilde{\underline{\beta}} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y} = \underline{M}^{-1}\underline{X}'\underline{y}$$

where $\underline{M} = \underline{X}'\underline{X}$.

Large sample properties of consistency and efficiency of the estimators follow from the maximum likelihood method of estimation. For unbiasedness we have to proceed as follows.

$$E(\tilde{\underline{\beta}}) = E(\underline{M}^{-1}\underline{X}'\underline{y}) = \underline{M}^{-1}\underline{X}'E(\underline{y}) = \underline{M}^{-1}\underline{X}'\underline{X}\underline{\beta} = \underline{\beta}$$

$$E\left(\frac{n}{n-p}\tilde{\sigma}^2\right) = \frac{n}{n-p}E(\tilde{\sigma}^2) = \frac{n}{n-p}\frac{E(\underline{y} - \underline{X}\tilde{\underline{\beta}})'(\underline{y} - \underline{X}\tilde{\underline{\beta}})}{n}$$

$$= \frac{1}{n-p}E\left(\underline{y} - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y}\right)'\left(\underline{y} - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y}\right)$$

$$= \frac{1}{n-p}E\left[\underline{y}'(\underline{I} - \underline{X}\underline{M}^{-1}\underline{X}')'(\underline{I} - \underline{X}\underline{M}^{-1}\underline{X}')\underline{y}\right]$$

Since $\underline{I} - \underline{X}\underline{M}^{-1}\underline{X}'$ is idempotent and symmetric

$$E\left(\frac{n}{n-p}\tilde{\sigma}^2\right) = \frac{1}{n-p}E\left[\underline{y}'(\underline{I} - \underline{X}\underline{M}^{-1}\underline{X}')\underline{y}\right]$$

$$= \frac{1}{n-p}E\left[\underline{\varepsilon}'(\underline{I} - \underline{X}\underline{M}^{-1}\underline{X}')\underline{\varepsilon}\right]$$

$$= \frac{\sigma^2}{n-p}\operatorname{tr}(\underline{I} - \underline{X}\underline{M}^{-1}\underline{X}')$$

(The symbol ~ is used for Maximum Likelihood estimate and the symbol ^ for unbiased estimator.)

$$E\left(\frac{n}{n-p}\tilde{\sigma}^2\right) = \frac{\sigma^2}{n-p} tr\,(\underline{I} - \underline{X}\underline{M}^{-1}\underline{X}')$$

$$= \frac{\sigma^2}{n-p}\left[tr\,\underline{I}_n - tr\,\underline{X}\underline{M}^{-1}\underline{X}'\right]$$

$$= \frac{\sigma^2}{n-p}\left[tr\,\underline{I}_n - tr\,\underline{I}_p\right]$$

$$= \frac{\sigma^2}{n-p}\,(n-p)$$

$$= \sigma^2.$$

Sufficiency and completeness. Sufficiency of the estimator can be proved, using the identity

$$(\underline{y} - \underline{X}\underline{\beta})'(\underline{y} - \underline{X}\underline{\beta}) = (\underline{y} - \underline{X}\hat{\underline{\beta}})'(\underline{y} - \underline{X}\hat{\underline{\beta}}) + (\hat{\underline{\beta}} - \underline{\beta})\,\underline{M}\,(\hat{\underline{\beta}} - \underline{\beta})$$

in the joint density function of $\underline{e}$,

$$f(\underline{e}) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}\, e^{-\frac{1}{2\sigma^2}(\underline{y} - \underline{X}\underline{\beta})'(\underline{y} - \underline{X}\underline{\beta})}$$

Now, $$f(\underline{e}) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \cdot \exp\left[\frac{-(n-p)\hat{\sigma}^2 + (\hat{\underline{\beta}} - \underline{\beta})\,\underline{M}\,(\hat{\underline{\beta}} - \underline{\beta})}{2\sigma^2}\right]$$

Koopman-Pitman theorem establishes the joint sufficiency for $\hat{\sigma}^2$ and $\hat{\underline{\beta}}$. It can be shown that the estimators $\hat{\sigma}^2$, $\hat{\underline{\beta}}$ are complete. An immediate consequence of the completeness of a statistic is that only one function of that statistic can have a given expected value. Thus if one function of a statistic $\underline{t}$ is an unbiased estimator of a certain function of the parameter $\underline{\beta}$, no other function of $\underline{t}$ will be. Completeness confers a uniqueness property upon an estimator.

In the setting of the general linear hypothesis, suppose it is desired to find an estimator of the function $g(\sigma^2, \underline{\beta})$. Since $\hat{\sigma}^2$ and $\hat{\underline{\beta}}$ form a set of jointly sufficient statistics for the parameters σ^2, $\underline{\beta}$ and

since the estimators are complete, it follows that, if a function $h(\hat{\sigma}^2, \hat{\underline{\beta}})$ can be found such that

$$E[h(\hat{\sigma}^2, \hat{\underline{\beta}})] = g(\sigma^2, \underline{\beta})$$

then $h(\hat{\sigma}^2, \hat{\underline{\beta}})$ is the unique minimum variance unbiased estimate of $g(\sigma^2, \underline{\beta})$ for every sample size.

Normality of $\hat{\underline{\beta}}$. Since $\hat{\underline{\beta}} = \underline{M}^{-1}\underline{X}'\underline{y}$ where $\underline{y}$ is distributed Normal and rank $(\underline{M}^{-1}\underline{X}') = p$,

$\hat{\underline{\beta}}$ is a p-variate normal.

Already proved is that $E(\hat{\underline{\beta}}) = \underline{\beta}$

Now

$$\text{var}(\hat{\underline{\beta}}) = \underline{M}^{-1}\underline{X}'[\text{var}(\underline{y})]\underline{X}\underline{M}^{-1}$$
$$= \sigma^2 \underline{M}^{-1}\underline{X}'\underline{X}\underline{M}^{-1}$$
$$= \sigma^2 \underline{M}^{-1}.$$

$\therefore$ $\hat{\underline{\beta}}$ is distributed $N(\underline{\beta}, \sigma^2\underline{M}^{-1})$

Distribution of $(n-p)\frac{\hat{\sigma}^2}{\sigma^2}$

Since $(n-p)\frac{\hat{\sigma}^2}{\sigma^2} = \underline{y}'(\underline{I} - \underline{X}\underline{M}^{-1}\underline{X}')\underline{y}$ and since $\underline{I} - \underline{X}\underline{M}^{-1}\underline{X}'$ is idempotent of rank n - p, $(n-p)\frac{\hat{\sigma}^2}{\sigma^2}$ is distributed as noncentral $\chi^2(n-p, \delta)$ where noncentrality $\delta = \frac{1}{2\sigma^2}\underline{\beta}'\underline{X}'(\underline{I} - \underline{X}\underline{M}^{-1}\underline{X}')\underline{X}\underline{\beta}$.

But $(\underline{I} - \underline{X}\underline{M}^{-1}\underline{X}')\underline{X} = 0$, giving $\delta = 0$ and $\chi^2(n-p, \delta=0)$ reduces to central χ^2 distribution with n - p degrees of freedom.

In order to establish the last property that $\hat{\sigma}^2$ is independently distributed of the vector $\hat{\underline{\beta}}$, we may use the theorem which states that the p linear forms $\hat{\underline{\beta}} = (\underline{M}^{-1}\underline{X}')\underline{y}$ are independent of the quadratic form $\hat{\sigma}^2 = \left[\frac{1}{n-p}(\underline{y} - \underline{X}\underline{\beta})'(\underline{y} - \underline{X}\underline{\beta})\right] = \frac{1}{n-p}\underline{y}'(\underline{I} - \underline{X}\underline{M}^{-1}\underline{X}')\underline{y}$, if the

product of the matrix of the linear forms and that of the quadratic form is a null matrix.

i.e. $\hat{\beta}$ and $\hat{\sigma}^2$ are independent if $(M^{-1}X')(I - XM^{-1}X') = 0$, which is found to be true on simplification.

Estimation of parameters under case B and Gauss-Markoff Theorem.

Under the assumption that the random vector e has zero means and covariance matrix $\sigma^2 I$, without any specification of the form of the frequency function, the principle of maximum likelihood cannot be used to obtain the estimators of the unknown parameters. Instead we shall use the method of least squares; that is to say, we shall find the value of β which minimizes the error sum of squares

$$\sum_{i=1}^{n} e_i^2 = e'e = (y - X\beta)'(y - X\beta)$$

The value of β that minimizes $e'e$ is given by

$$\frac{\partial}{\partial \beta}(e'e) = 0 = 2X'y - 2X'X\hat{\beta}$$

the least square estimate of β is, therefore,

$$\hat{\beta} = M^{-1}X'y$$,which is the same as the maximum likelihood estimate, if the distribution of e is normal.

But minimizing the sum of squares $e'e$ does not provide us with an estimate of σ^2. However, the unbiased estimate of σ^2 based on the least square estimate of β is given by

$$\hat{\sigma}^2 = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{n - p}$$

Properties of Least square Estimators;

The frequency function of the random error vector $\underline{e}$ being unspecified, in general it is not possible to examine the goodness of the estimator $\hat{\underline{\beta}}$ relative to all functions. Instead, we shall have to limit ourselves to a subset of functions; for example, (since is a linear function of the y_i) to the set of all linear functions of the y_i. In the case of normality of $\underline{e}$, $\hat{\underline{\beta}} = \underline{M}^{-1}\underline{X}'\underline{y}$ has smaller variance than any other unbiased estimator of $\underline{\beta}$. For the least square estimator we cannot have so broad a statement, but we can compare the goodness of the least square estimator with other linear estimators, (i.e. estimators in the form of linear functions of observations.)

Gauss-Markoff theorem asserts the goodness of the least squares estimators in the model

$$\underline{y} = \underline{X}\underline{\beta} + \underline{e} \ .$$

The following theorem (which is the special case of General Gauss-Markoff theorem) is mentioned for continuity in the development of the theory. More generalised theorem will be proved later for the general linear hypothesis model of less than full rank.

Gauss-Markoff Theorem for the Linear Model of Full Rank.

Theorem: (2.2) Under the assumption

$$\Omega : \quad \underset{n\times 1}{\underline{y}} = \underset{n\times p}{\underline{X}}\ \underset{p\times 1}{\underline{\beta}} + \underset{n\times 1}{\underline{e}}, \ \text{rank}\ (\underline{X}) = p$$

$$E(\underline{e}) = \underline{0} \ , \quad E(\underline{e}\,\underline{e}') = \sigma^2 \underline{I}$$

the best (minimum variance) linear (linear function of the y_i) unbiased estimate of $\underline{\beta}$ is given by least

squares;

i.e. $\hat{\underline{\beta}} = \underline{M}^{-1}\underline{X}'\underline{y}$ is the best linear unbiased estimate of $\underline{\beta}$.

Further, the best linear unbiased estimate of any linear combination of the β_i is the same linear combination of the best linear estimates of the β_i, i.e. the best linear unbiased estimator of $\underline{\lambda}'\underline{\beta}$ ($\underline{\lambda}$ being a vector of constants) is $\underline{\lambda}'\hat{\underline{\beta}}$ where $\hat{\underline{\beta}}$ is the b.l.u.estimate of $\underline{\beta}$.

2.1.2 Choice of the Design Matrix $\underline{X}$.

We know that the elements x_{ij} must be known before the values are selected at random. In some cases, x_{ij} values may be picked or chosen by the experimenter in any way he wishes; in other cases, they cannot be so controlled. If the experimenter can choose the x_{ij} values, the question of how to select them arises. In general it would seem that the best way to pick them is so that the variance of certain estimators will be as small as possible.

For example, in the case of the full rank model, the variance of the estimator of $\underline{\lambda}'\underline{\beta}$ is $\sigma^2\underline{\lambda}'\underline{M}^{-1}\underline{\lambda}$ and we might want to choose the x_{ij} that minimize $\underline{\lambda}'\underline{M}^{-1}\underline{\lambda}$. We cannot do this in general for all vectors $\underline{\lambda}$, but might be able for some. We shall be discussing about the optimal choice of these elements x_{ij} or the design matrix $\underline{X}$.

2.1.3 Tests of Hypothesis

Likelihood Ratio tests and the General Linear Hypothesis of full rank:

We shall now consider some useful results associated with testing of hypothesis in the general linear model of full rank under normal theory. Through out this thesis, the symbols Ω and H will be used to denote a set of fundamental or underlying assumptions and the statement of the null hypothesis respectively. It is also convenient to introduce the symbol

$$\omega = H \wedge \Omega$$

meaning the set of assumptions obtained by imposing the assumptions of the hypothesis H on the assumptions Ω .

If $\underline{y} = (y_1, y_2, \cdots y_n)'$ stands for the observations or sample then $L_n(\underline{y}; \underline{\theta})$ where θ is an r-dimensional parameter vector is the probability density of $\underline{y}$ (also known as the likelihood function of $\underline{y}$)

Now consider the hypothesis

$$H(\omega; \Omega)$$

and let

$$(L_n)_\omega = \sup_{\theta \in \omega} L_n(\underline{y}; \underline{\theta})$$

and $$(L_n)_\Omega = \sup_{\theta \in \Omega} L_n(\underline{y}; \underline{\theta})$$

The likelihood ratio test statistic λ_H for testing $H(\omega; \Omega)$ or more briefly H , is defined to be

$$\lambda_H = (L_n)_\omega / (L_n)_\Omega .$$

The values of λ_H lie in the interval $[0, 1]$. The likelihood ratio test consists in rejecting H if $\lambda_H < c_\alpha$, the critical value; which is a constant so chosen that

$$\int_{W_\alpha} dF_\eta \leq \alpha \quad , \quad \theta \in \omega$$

W_α being the critical set in $E^{(n)}$.

In order to evaluate the power of the test, the distribution must also be known when the alternative hypothesis H_1, is true.

2.1.4 Testing the hypothesis $\beta_i = \beta_i^* \; , i = 1, 2, \ldots r,$ with the remaining β_i's unspecified in general linear model of full rank under normal theory.

Ω : $\underline{y} = \underline{X}\underline{\beta} + \underline{e}$ with $\underline{e}$ distributed $N(\underline{0}, \sigma^2 \underline{I})$ rank $\underline{X} = p \; (p < n)$

H : $\beta_i = \beta_i^* \quad , \quad i = 1, 2, \ldots r \quad (r \leq p)$

Now in order to apply the likelihood ratio criteria we may reformulate the model and redefine Ω and H as follows:

Ω : the half $(p + 1)$ -dimensional Euclidian space for which

$$-\infty < \beta_i < +\infty \quad , \quad i = 1, 2, \ldots p.$$

H : $\beta_i = \beta_i^* \quad , \quad i = 1, 2, \ldots r \quad (r \leq p)$

ω : the subset of Ω for which $\beta_i = \beta_i^* \; , i = 1, 2, \ldots r \; (r \leq p)$

In the model $\underline{y} = \underline{X}\underline{\beta} + \underline{e}$

partition the matrix $\underline{X}$ and the vector $\underline{\beta}$, so that

$$\underline{X} = (\underset{n \times r}{\underline{X}_1} \; , \; \underline{X}_2) \quad , \quad \underline{\beta} = \begin{pmatrix} \underline{\gamma}_1 \\ \underline{\gamma}_2 \end{pmatrix}$$

then $\underline{y} = \underline{X}_1 \underline{\gamma}_1 + \underline{X}_2 \underline{\gamma}_2 + \underline{e}$

We would like to test the hypothesis $\underline{\gamma}_1 = \underline{\gamma}_1^*$ ($\underline{\gamma}_1^*$ known)

with no stipulation on $\underline{\gamma}_2$.

Since $\underline{\gamma}_1 = (\beta_1, \beta_2, \cdots \beta_r)'$ and $\underline{\gamma}_1^* = (\beta_1^*, \beta_2^*, \cdots \beta_r^*)$

this test is equivalent to testing the hypothesis $\beta_i = \beta_i^*$, $i = 1, 2, \cdots r$

in the model $\underline{y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$.

Our sample consists of what may be regarded at the conditional random variable $(y_i \mid x_{i1}, x_{i2}, \cdots x_{ip};\ i = 1, 2 \cdots n)$, and the likelihood of the parameters $\underline{\beta}, \sigma^2$ is given by $L_n = L_n(\underline{y}; \underline{\beta}, \sigma^2) = L_n(\underline{y}; \underline{\gamma}_1, \underline{\gamma}_2, \sigma^2)$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}(\underline{y} - \underline{X}\underline{\beta})'(\underline{y} - \underline{X}\underline{\beta})\right]$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}(\underline{y} - \underline{X}_1\underline{\gamma}_1 - \underline{X}_2\underline{\gamma}_2)'(\underline{y} - \underline{X}_1\underline{\gamma}_1 - \underline{X}_2\underline{\gamma}_2)\right]$$

To find $(L_n)_\Omega$ and $(L_n)_\omega$ we shall work with $\log L_n$ the logarithm of the likelihood of the parameters:

$$\text{Log } L_n = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(\underline{y} - \underline{X}\underline{\beta})'(\underline{y} - \underline{X}\underline{\beta})$$

The value of $(\sigma^2, \underline{\beta})$ that maximize $\log L_n$ is given by

$$\frac{\partial}{\partial\sigma^2}(\log L_n) = 0$$

$$\frac{\partial}{\partial\underline{\beta}}(\log L_n) = \underline{0} .$$

Under the assumption Ω these equations reduce to

$$-\frac{n}{2\sigma^2} + (\underline{y} - \underline{X}\underline{\beta})'(\underline{y} - \underline{X}\underline{\beta}) = 0$$

$$\frac{\underline{X}'(\underline{y} - \underline{X}\underline{\beta})}{\sigma^2} = \underline{0}$$

giving $\underline{X}'\underline{X}\,\hat{\underline{\beta}} = \underline{X}'\underline{y}$ or $\hat{\underline{\beta}} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y} = \underline{M}^{-1}\underline{X}'\underline{y}$

and $\sigma^2 = \dfrac{(\underline{y} - \underline{X}\underline{\beta})'(\underline{y} - \underline{X}\underline{\beta})}{n}$

Substituting these values in the expression for L_n we have

$$(L_n)_\Omega = \frac{n^{n/2} e^{-n/2}}{(2\pi)^{n/2} [(y - X\hat{\beta})'(y - X\hat{\beta})]^{n/2}} \quad \text{where} \quad \hat{\beta} = M^{-1} X' y$$

$$= \frac{n^{n/2} e^{-n/2}}{(2\pi)^{n/2} \quad Q_0^{n/2}} \quad \text{where} \quad Q_0 = (y - X\beta)'(y - X\beta)$$

the minimum value of $e'e$ with respect to the unknown parameter β in the model $y = X\beta + e$. To obtain $(L_n)_\omega$, we may proceed in the similar manner with the restriction $\gamma_1 = \gamma_1^*$. Now if we put

$$T = y - X_1 \gamma_1^*$$

$$L_n = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}(T - X_2\gamma_2)'(T - X_2\gamma_2)\right]$$

The value of γ_2 and σ^2 that miaximize L_n are thus given by

$$\hat{\gamma}_2(\gamma_1^*) = (X_2' X_2)^{-1} X_2' T = (X_2' X_2)^{-1} X_2' (y - X_1 \gamma_1^*)$$

$$\hat{\sigma}^2(\gamma_1^*) = \frac{1}{n}\left[T - X_2 \hat{\gamma}_2(\gamma_1^*)\right]\left[T - X_2 \hat{\gamma}_2(\gamma_1)\right]$$

Substituting these values in L_n, we get

$$(L_n)_\omega = \frac{n^{n/2} e^{-n/2}}{(2\pi)^{n/2} \{[T - X_2 \hat{\gamma}_2(\gamma_1^*)]'[T - X_2 \hat{\gamma}_2(\gamma_1^*)]\}}$$

$$= \frac{n^{n/2} e^{-n/2}}{(2\pi)^{n/2} \quad (Q_0 + Q_1)^{n/2}}$$

where $Q_0 + Q_1$ = the maximum of $e'e$ with respect to the unknown parameter γ_2 on the model $y = X\beta + e$ restricted by $\gamma_1 = \gamma_1^*$. Taking the ratio of these two maxima, we get the likelihood ratio test statistic for H

$$\lambda_H = (L_n)_\omega / (L_n)_\Omega$$

$$= \left\{ \frac{[T - X_2 \hat{\gamma}_2(\gamma_1^*)]'[T - X_2 \hat{\gamma}_2(\gamma_1^*)]}{(y - X\beta)'(y - X\beta)} \right\}^{-n/2}$$

$$= \left(\frac{Q_0 + Q_1}{Q_0}\right)^{-n/2}$$

To find the distribution of Likelihood Ratio.

$X_2'X_2$ being a principal minor of the positive matrix $X'X$ is also positive definite i.e. it has an inverse.

From $$X'(I - X M^{-1} X') = 0$$

we obtain
$$\left.\begin{aligned} X_1'(I - X M^{-1} X') &= 0 \\ X_2'(I - X M^{-1} X') &= 0 \end{aligned}\right\} \quad \ldots \quad (i)$$

Substituting for I and $\hat{\gamma}_2(\gamma_1^*)$

$$Q_0 + Q_1 = (y - X_1\gamma_1^*)'[I - X_2(X_2'X_2)^{-1}X_2'](y - X_1\gamma_1^*)$$

From (i) follows

$$Q_0 = (y - X_1\gamma_1^*)'[I - X M^{-1} X'](y - X_1\gamma_1^*)$$

Using the notation

$$I = y - X_1\gamma_1^*$$
$$A = I - X M^{-1} X'$$
$$A_2 = I - X_2(X_2'X_2)^{-1}X_2' .$$

Now we shall study the identity

$$I'I = I'AI + I'(A_2 - A)I + I'(I - A_2)I = Q_0 + Q_1 + Q_2 \quad \ldots \quad (ii)$$

It follows that I is distributed $N[X(\beta - \beta^*), \sigma^2 I]$ where $\beta^* = \begin{pmatrix}\gamma_1^* \\ 0\end{pmatrix}$ since

$$E(I) = E(y - X_1\gamma_1^*) = X\beta - X_1\gamma_1^* = X\beta - X\begin{pmatrix}\gamma_1^* \\ 0\end{pmatrix}$$

In (ii) the identity

$$I = A + (A_2 - A) + (I - A_2)$$ is obvious.

Using (i) it can be easily verified that A , $(A_2 - A)$, $(I - A_2)$ are idempotent .

Hence 1. $\frac{I'AI}{\sigma^2} = \frac{Q_0}{\sigma^2}$ is distributed as $\chi^2(n-p)$, since by (i) the noncentrality is zero .

2. $\frac{T'(A_2 - A)T}{\sigma^2} = \frac{Q_1}{\sigma^2}$ is distributed as $\chi^2(r, \delta)$ where

$r = \text{rank}\ (A_2 - A)$ and the noncentrality

$$\delta = \frac{1}{2\sigma^2}\left[E(T')(A_2 - A)E(T)\right]$$

3. $\frac{Q_0}{\sigma^2}$ and $\frac{Q_1}{\sigma^2}$ are independent.

Therefore, the quantity $u = \left(\frac{Q_1}{Q_0}\right)\left(\frac{n-p}{r}\right)$ is distributed as $F'(r, n-p, \delta)$ This quantity u, being a monotonic function of λ_H may be used as a test statistic.

To show that rank $(A_2 - A) = r$

Using the fact that $A_2 - A$ is idempotent

$$\text{rank}\ (A_2 - A) = tr(A_2 - A)$$
$$= tr(A_2) - tr(A)$$
$$= n - (p-r) - (n-p) = r$$

To examine δ in more detail.

Substituting $X(\beta - \beta^*)$ for $E(T)$ we get

$$2\sigma^2\delta = (\beta - \beta^*)' X' (A_2 - A) X (\beta - \beta^*)$$
$$= \left[(\gamma_1 - \gamma_1^*)' X_1' + \gamma_2' X_2'\right]\left[X M^{-1} X' - X_2 (X_2' X_2)^{-1} X_2'\right]\left[X_1(\gamma_1 - \gamma_1^*) + X_2\gamma_2\right]$$

or $$\delta = \frac{1}{2\sigma^2}(\gamma_1 - \gamma_1^*)'\left[X_1' X_1 - X_1' X_2 (X_2' X_2)^{-1} X_2' X_1\right](\gamma_1 - \gamma_1^*).$$

$X_1' X_1 - X_1' X_2 (X_2' X_2)^{-1} X_2' X_1$ being positive definite, $\delta = 0$ if and only if $\gamma_1 = \gamma_1^*$ i.e. if and only if H_0 is true.

We therefore have the following basic result concerning the test of sub-hypothesis of the **general** linear hypothesis of full rank

Theorem: (2.3) Under

$$\Omega: \quad y = X\beta + \varepsilon \qquad \text{rank } X = p\ (< n) \quad \text{and}$$
$$\varepsilon \text{ is } N(0, \sigma^2 I)$$

to test the hypothesis $H_0 : \beta_i = \beta_i^*$, $i = 1, 2, \ldots r$.

we may partition $\underline{X}$ and $\underline{\gamma}$ so that

$$\underline{y} = \underline{X}_1 \underline{\gamma}_1 + \underline{X}_2 \underline{\gamma}_2 + \underline{e}$$

where $\underline{\gamma}_1$ is of dimension $r \times 1$ and

$$\underline{\gamma}_1 = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_r \end{pmatrix}$$

Then the likelihood ratio procedure to test the null hypothesis $H_0 : \underline{\gamma}_1 = \underline{\gamma}_1^*$ is

(1) obtain Q_0 the minimum value of $\underline{e}'\underline{e}$, the error sum of squares with respect to the unknown parameter $\underline{\beta}$ in the model $\underline{y} = \underline{X}\underline{\beta} + \underline{e}$.

(ii) obtain the minimum value of the same quantity with respect to $\underline{\gamma}_2$ in the restricted model $\underline{y} = \underline{X}_1 \underline{\gamma}_1^* + \underline{X}_2 \underline{\gamma}_2 + \underline{e}$ (i.e. the model restricted by the hypothesis) , $Q_0 + Q_1$.

Then the quantity

$$u = \frac{(n-p)}{r} \frac{Q_1}{Q_0}$$

is distributed as $F'(r, n-p, \delta)$, where

$$\delta = \frac{(\underline{\gamma}_1 - \underline{\gamma}_1^*)' \underline{B} (\underline{\gamma}_1 - \underline{\gamma}_1^*)}{2\sigma^2}$$

and where $\underline{B} = \underline{X}_1' \underline{X}_1 - \underline{X}_1' \underline{X}_2 (\underline{X}_2' \underline{X}_2)^{-1} \underline{X}_2 \underline{X}_1$

Since $\underline{B}$ is positive definite, u is distributed as $F(r, n-p)$ if and only if H_0 is true i.e. if and only if $\underline{\gamma}_1 = \underline{\gamma}_1^*$.

2.1.5 Test of the Hypothesis $\Psi = \underset{\sim}{\Lambda}_1^{r\times p}\,\underset{\sim}{\beta}^{p\times 1} = \Psi^*$, where $\underset{\sim}{\Lambda}_1 = \begin{pmatrix} \underset{\sim}{\lambda}_1' \\ \underset{\sim}{\lambda}_2' \\ \vdots \\ \vdots \\ \underset{\sim}{\lambda}_r' \end{pmatrix}$

Consider an augmented matrix $\underset{\sim}{\Lambda} = \begin{pmatrix} \underset{\sim}{\Lambda}_1 \\ \underset{\sim}{\Lambda}_2 \end{pmatrix}$ of rank p such that

$\underset{\sim}{\Lambda}^{-1} = \underset{\sim}{\Delta} = (\underset{\sim}{\Delta}_1, \underset{\sim}{\Delta}_2)$ where $\underset{\sim}{\Delta}_1$ has dimension $p\times r$.

Let $\underset{\sim}{x}\,\underset{\sim}{\Lambda}^{-1} = \underset{\sim}{X}\,\underset{\sim}{\Delta} = (\underset{\sim}{X}\underset{\sim}{\Delta}_1, \underset{\sim}{X}\underset{\sim}{\Delta}_2) = (\underset{\sim}{Z}_1, \underset{\sim}{Z}_2) = \underset{\sim}{Z}$

and $\underset{\sim}{\Lambda}\,\underset{\sim}{\beta} = \begin{pmatrix} \underset{\sim}{\Lambda}_1\underset{\sim}{\beta} \\ \underset{\sim}{\Lambda}_2\underset{\sim}{\beta} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{\alpha}_1 \\ \underset{\sim}{\alpha}_2 \end{pmatrix}$

Now our model becomes

$$\begin{aligned} \underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{e} &= \underset{\sim}{X}\,\underset{\sim}{\Lambda}^{-1}\underset{\sim}{\Lambda}\,\underset{\sim}{\beta} + \underset{\sim}{e} \\ &= (\underset{\sim}{Z}_1, \underset{\sim}{Z}_2)\begin{pmatrix} \underset{\sim}{\alpha}_1 \\ \underset{\sim}{\alpha}_2 \end{pmatrix} + \underset{\sim}{e} \\ &= \underset{\sim}{Z}_1\underset{\sim}{\alpha}_1 + \underset{\sim}{Z}_2\underset{\sim}{\alpha}_2 + \underset{\sim}{e}\,. \end{aligned}$$

Thus we have reduced our problem to the case of testing sub-hypothesis, for which the solution is known. This reduction is known as parametrization which we shall be discussing when we consider the model of less than full rank.

2.2. General Linear Hypothesis of less than full rank.

Ω: $\underset{\sim}{y}^{n\times 1} = \underset{\sim}{X}^{n\times p}\,\underset{\sim}{\beta}^{p\times 1} + \underset{\sim}{e}$ $(p \leq n)$

where rank $(\underset{\sim}{X}) = k < p$.

Most of the models associated with various designs are of this nature and by means of a process known as reparametrization we can always transform this model to a model of full rank for which the methods of analysis has been discussed already.

As in the case of the full rank model we shall consider two cases:-

Case A. $\underset{\sim}{e}$ is $N(\underset{\sim}{0}, \sigma^2\underset{\sim}{I})$: i.e. errors are independent normal variables with mean o and variances $\sigma^2\underset{\sim}{I}$.

Case B. $\underset{\sim}{e}$ is such that $E(\underset{\sim}{e}) = \underset{\sim}{0}$ and $Cov(\underset{\sim}{e}) = \sigma^2\underset{\sim}{I}$; i.e. errors are uncorrelated.

2.2.1. Point estimate of a linear combination of $\underset{\sim}{\beta}$ under case B.

Least square estimates of $\underset{\sim}{\beta}$ are given by

$$\frac{\partial}{\partial \underset{\sim}{\beta}} (\underset{\sim}{e}'\underset{\sim}{e}) = 2\underset{\sim}{X}'\underset{\sim}{X}\underset{\sim}{\beta} - 2\underset{\sim}{X}'\underset{\sim}{y} = \underset{\sim}{0}$$

i.e. as the solution of the normal equations

Since $\underset{\sim}{X}$ is of less than full rank, $\underset{\sim}{X}'\underset{\sim}{X}$ has no inverse and it is necessary to examine the system to see whether a solution exists at all. Now consider the coefficient matrix $\underset{\sim}{X}'\underset{\sim}{X}$ and the augmented matrix $(\underset{\sim}{X}'\underset{\sim}{X}, \underset{\sim}{X}'\underset{\sim}{y}) = \underset{\sim}{X}'(\underset{\sim}{X}, \underset{\sim}{y})$.

rank $(\underset{\sim}{X}'\underset{\sim}{X}, \underset{\sim}{X}'\underset{\sim}{y})$ = rank $\underset{\sim}{X}'(\underset{\sim}{X}, \underset{\sim}{y})$ $\leq$ rank $\underset{\sim}{X}'$ (= rank $\underset{\sim}{X}'\underset{\sim}{X}$)

Also rank $(X'X , X'y) \geq$ rank $X'X$.

rank $(X'X , X'y)$ = rank $X'X$.

Thus the system of equations is consistant.

It can be easily shown that there exist no linear functions of whatsoever that give rise to unbiased estimates of β i.e. there is no l.u.e of β . If possible consider a matrix $C_{(p \times n)}$ of constants such that $E(Cy) = \beta$ identically in β .

If so $E[C(X\beta + \varepsilon)] = CX\beta = \beta$ identically.

i.e $CX = I$.

which is impossible since CX is at most of rank k and rank $I = p > k$

It is thus natural to investigate next whether there exists an unbiased linear estimate of any linear combination of the β , say $\lambda'\beta$ where λ is a known vector of constants.

Definition. A function of parameters is said to be linearly estimable if there exists an unbiased linear estimate.

Theorem. 2.4 The linear combination $\lambda'\beta = \psi$ is estimable if and only if there exists a solution for r in the equations

$$X'X\, r = \lambda .$$

Proof. If $\lambda'\beta$ is estimable, there exist a vector c such that

$E(c'y) = c'Xy = \lambda'\beta$ identically in β .

i.e. $c'X = \lambda'$ or $X'c = \lambda$.

i.e. rank (X') = rank (X', λ)

Then rank $(X'X)$ = rank $(X'X, \lambda)$

Hence $\underline{X}'\underline{X}\,\underline{r} = \underline{\lambda}$ has a solution for $\underline{r}$.

On the other hand if $\underline{X}'\underline{X}\,\underline{r} = \underline{\lambda}$ has a solution for $\underline{r}$, we can put $\underline{X}\,\underline{r} = \underline{c}$ in the equation to get

$$\underline{X}'\underline{c} = \underline{\lambda}.$$

Theorem 2.5 . Generalized Gauss-Markoff Theorem .

The unique b. l. u. estimate for any estimable function $\underline{\lambda}'\underline{\beta}$ is $\underline{r}'\underline{X}'\underline{y}$ where $\underline{r}$ satisfies $\underline{X}'\underline{X}\,\underline{r} = \underline{\lambda}$.

Proof. Let the b. l. u. estimate be $\underline{c}'\underline{y}$, where $\underline{c}' = \underline{r}'\underline{X}' + \underline{b}'$ where $\underline{b}$ is any vector.

Then we have to find the vector $\underline{b}$ such that

i. $E(\underline{c}'\underline{y}) = \underline{\lambda}'\underline{\beta}$.

ii. $\mathrm{Var}(\underline{c}'\underline{y})$ is less than any other linear function of $\underline{y}$ that satisfies (i).

For unbiasedness, $E(\underline{c}'\underline{y}) = \underline{c}'\underline{X}\underline{\beta} = (\underline{r}'\underline{X}'\underline{X} + \underline{b}'\underline{X})\underline{\beta} = \underline{\lambda}'\underline{\beta}$

Since $\underline{X}'\underline{X}\,\underline{r} = \underline{\lambda}$, we must have $\underline{b}'\underline{X} = \underline{0}$.

Next consider $\mathrm{var}(\underline{c}'\underline{y}) = E(\underline{c}'\underline{y} - \underline{\lambda}'\underline{\beta})^2$

$$= E(\underline{c}'\underline{X}\underline{\beta} + \underline{c}'\underline{e} - \underline{\lambda}'\underline{\beta})^2$$

$$= E(\underline{r}'\underline{X}'\underline{X}\underline{\beta} + \underline{b}'\underline{X}\underline{\beta} + \underline{c}'\underline{e} - \underline{\lambda}'\underline{\beta})^2$$

$$= E(\underline{c}'\underline{e})^2$$

$$= E(\underline{c}'\underline{e}\,\underline{e}'\underline{c}) = \sigma^2 E(\underline{c}'\underline{c}) = \sigma^2\underline{c}'\underline{c}.$$

$$= \sigma^2(\underline{r}'\underline{X}' + \underline{b}')(\underline{X}\,\underline{r} + \underline{b})$$

$$= \sigma^2\underline{r}'\underline{X}'\underline{X}\,\underline{r} + \sigma^2\underline{b}'\underline{b}.$$

$\therefore$ $\mathrm{var}(\underline{c}'\underline{y})$ is minimum when $\underline{b}'\underline{b} = 0$ or $\underline{b} = \underline{0}$ which is consistent with $\underline{b}'\underline{X} = \underline{0}$.

Hence $\underline{c}' = \underline{r}' \underline{X}'$ and the b.l.u. estimate of the estimable function is $\underline{r}' \underline{X} \underline{y}$.

Though there are infinitely many solutions to the system $\underline{X}' \underline{X} \underline{r} = \underline{\lambda}$ when $\underline{\lambda}' \underline{\beta}$ is an estimable function for $\underline{X}'\underline{X}$ is not of full rank, any solution of the system gives the same estimate of $\underline{\lambda}' \underline{\beta}$.

Proof for uniqueness.

Consider any solution $\hat{\underline{\beta}}$ to the system $\underline{X}' \underline{X} \underline{\beta} = \underline{X}' \underline{y}$. Using $\underline{r}_1$ and $\underline{r}_2$, any two solutions of $\underline{X}' \underline{X} \underline{r} = \underline{\lambda}$.

We have $$\underline{r}_1' \underline{X}' \underline{X} \underline{\beta} = \underline{r}_1' \underline{X}' \underline{y}$$

and $$\underline{r}_2' \underline{X}' \underline{X} \underline{\beta} = \underline{r}_2' \underline{X}' \underline{y}$$

But $$\underline{r}_1' \underline{X}' \underline{X} = \underline{\lambda}' = \underline{r}_2' \underline{X}' \underline{X}.$$

Hence $$\underline{\lambda}' \hat{\underline{\beta}} = \underline{r}_1' \underline{X}' \underline{y} = \underline{\lambda}' \hat{\underline{\beta}}$$

$$\underline{\lambda}' \hat{\underline{\beta}} = \underline{r}_2' \underline{X}' \underline{y} = \underline{\lambda}' \hat{\underline{\beta}}.$$

It will be important to know how many independent linear functions are estimable. It will help us very much to have the following definition of linearly independent estimable function.

Definition. The estimable functions $\underline{\lambda}_1' \underline{\beta}, \underline{\lambda}_2' \underline{\beta}, \cdots \underline{\lambda}_r' \underline{\beta}$ are said to be linearly independent estimable functionsif there exists vectors $\underline{r}_1, \underline{r}_2, \cdots, \underline{r}_r$, such that

$$\underline{X}' \underline{X} \underline{r}_1 = \underline{\lambda}_1$$

$$\underline{X}' \underline{X} \underline{r}_2 = \underline{\lambda}_2$$

$$\vdots \qquad \vdots$$

$$\underline{X}' \underline{X} \underline{r}_r = \underline{\lambda}_r$$

and if the vectors $\underline{\lambda}_1, \underline{\lambda}_2, \cdots \underline{\lambda}_r$ are linearly independent.

If the $\underline{\lambda}_i$'s are not linearly independent but if the matrix

$\underline{\Lambda} = (\underline{\lambda}_1, \cdots \underline{\lambda}_r)$ has rank t, then the set contains t linearly independent estimable functions.

Theorem. 2.6 There are exactly k linearly independant estimable functions associated with the model $\underline{y} = \underline{X}\underline{\beta} + \underline{e}$ where k is the rank of $\underline{X}$.

Proof. The theorem will be proved if we can show that there are exactly k linearly independent vectors $\underline{\lambda}$ for each of which the system $\underline{X}'\underline{X}\ \underline{r} = \underline{\lambda}$ is consistent, where k is the rank of $\underline{X}$.

If there are q vectors $\underline{\lambda}_1, \underline{\lambda}_2, \cdots \underline{\lambda}_q$ $(q > k)$ satisfying $\underline{X}'\underline{X}\,\underline{r} = \underline{\lambda}$ then, since the parametric functions are estimable, there exist q vectors $\underline{r}_1, \underline{r}_2, \cdots \underline{r}_q$ such that

$$\underline{X}'\underline{X}\ \underline{r}_i = \underline{\lambda}_i \qquad (i = 1, 2, \cdots q)$$

$$\therefore \quad \underline{X}'\underline{X}\ (\underline{r}_1, \underline{r}_2, \cdots \underline{r}_q) = (\underline{\lambda}_1, \underline{\lambda}_2, \cdots \underline{\lambda}_q)$$

or $$\underline{X}'\underline{X}\ \underline{R} = \underline{\Lambda}$$

where $\underline{R}$ = the $p \times q$ matrix $(\underline{r}_1, \underline{r}_2, \cdots, \underline{r}_q)$

and $\underline{\Lambda}$ = " $(\underline{\lambda}_1, \underline{\lambda}_2, \cdots, \underline{\lambda}_q)$

But $\underline{X}'\underline{X}$ is of rank k ; so $\underline{\Lambda}$ must be at most of rank k ; hence there can be at most k linearly independent estimable functions.

Now for any $\underline{x}_i'$ the i^{th} row of $\underline{X}$ we can see that $\underline{x}_i'\underline{\beta}$ is estimable for rank $(\underline{X}'\underline{X}) =$ rank $(\underline{X}'\underline{X} \mid \underline{x}_i)$, $\underline{X}'\underline{X}\ \underline{r} = \underline{x}_i$ admits a solution $\underline{r}$.

Hence $\underline{x}_i'\underline{\beta}$ is estimable for $i = 1, 2, \cdots n$.

But these $\underline{x}_i$ form a set of k linearly independent vectors, and the theorem is thus established.

A natural consequence of the above theorem is that

$\underline{X}\underline{\beta}$ and $\underline{X}'\underline{X}\underline{\beta}$ are estimable, meaning that each element of these matrices are estimable.

Theorem. 2.7 If $\underline{\lambda}_i'\underline{\beta}$ for i=1, 2, $\ldots$, q are estimable, any linear combination of these is estimable.

Proof. Consider a linear combination $\sum_{i=1}^{q} a_i \underline{\lambda}_i'\underline{\beta}$ of estimable functions $\underline{\lambda}_i'\underline{\beta}$. This function may be estimable if there exist an $\underline{r}$, such that $\underline{M}\,\underline{r} = \sum a_i \underline{\lambda}_i$ $(\underline{M} = \underline{X}'\underline{X})$

But estimability of every $\underline{\lambda}_i'\underline{\beta}$ justifies the existence of an $\underline{r}_i$ such that $\underline{M}\,\underline{r}_i = \underline{\lambda}_i$ for i=I, 2, $\ldots$ q.

$$\therefore \sum_{i=1}^{q} \underline{M} a_i \underline{r}_i = \sum_{i=1}^{q} a_i \underline{\lambda}_i$$

or $\underline{M}\,\underline{r} = \sum a_i \underline{\lambda}_i$ where $\underline{r} = \sum a_i \underline{r}_i$.

and the proof is complete.

The following two simple results are useful in the study of estimable functions.

Corollary: 2.4.1. If $\underline{\lambda}'\underline{\beta}$ is estimable, then $\underline{\lambda}'$ must be a linear combination of the rows of $\underline{X}$.

i.e. $\underline{\lambda}' = \underline{a}'\underline{X}$

This lemma is useful in determining what functions are estimable by inspection of the model.

Corollary: 2.4.2. If $\underline{\lambda}'\underline{\beta}$ is estimable, then there exist a vector $\underline{r}$ such that $\underline{\lambda}' = \underline{r}'\underline{X}'\underline{X}$ and the b.l.u. estimate of $\underline{\lambda}'\underline{\beta}$ is $\underline{r}'\underline{X}'\underline{y}$.

Theorem: 2.8. The best linear unbiased estimate of a linear

combination of estimable functions is given by the same linear combination of the best unbiased estimates of the estimable functions.

Proof: Consider $\lambda'\beta = \sum a_i \lambda_i' \beta$ in which each $\lambda_i'\beta$ is estimable. Then by virtue of Theorem 2.7, $\lambda'\beta$ is estimable. B.l.u. estimates of $\lambda'\beta$ and $\lambda_i'\beta$ (i=1, 2, ... n.) are respectively $r'X'y$ and $r_i'X'y$ where r and r_i are any solutions of,

$$M r = \lambda$$

and $$M r_i = \lambda_i \text{ for } i=1 \ldots\ldots q.$$

Thus we have $M \sum a_i r_i = \sum a_i \lambda_i = \lambda$

i.e. $\sum a_i r_i$ is a solution of $M r = \lambda$

The b.l.u. estimate of $\lambda'\beta$ will then be

$$r'X'y = (\sum a_i r_i)' X'y$$

$$= (\sum a_i r_i') X'y = \sum a_i r_i' X'y .$$

2.2.2 Reparametrization.

In order to utilize the theorems on the model of full rank, we shall have to use the linear transformation known as reparametrization, i.e. we shall have to reparametrize the model of less than full rank to a model of full rank.

Formal Definition of Reparametrization.

By a reparametrization of the model $y = X\beta + e$ we shall mean a transformation from the vector β to the vector α by $\alpha = U\beta$, where each element of $\alpha = U\beta$ is an estimable function.

Since $\underline{M} = \underline{X}'\underline{X}$ is positive semidefinite of rank k, there exist a nonsingular matrix $\underline{W}^*(p \times p)$ such that

$$(\underline{W}^*)' \underline{X}'\underline{X}\, \underline{W}^* = \begin{pmatrix} \underline{B} & \underline{O} \\ \underline{O} & \underline{O} \end{pmatrix} \quad \text{where} \quad \underline{B}\,(k \times k) \quad \text{is of rank k.}$$

If $\underline{W}^*$ is partitioned into $\underline{W}^* = (\underline{W}, \underline{W}_1)$ where W is $p \times k$, we get

$$\begin{pmatrix} \underline{W}' \\ \underline{W}_1' \end{pmatrix} \underline{X}'\underline{X}\, (\underline{W}, \underline{W}_1) = \begin{pmatrix} \underline{B} & \underline{O} \\ \underline{O} & \underline{O} \end{pmatrix}$$

giving $\underline{W}'\underline{X}'\underline{X}\,\underline{W} = \underline{B}$ and $\underline{W}_1'\underline{X}'\underline{X}\,\underline{W}_1 = \underline{O}$.

This implies that $\underline{W}'\underline{X}'$ is of rank k and $\underline{W}_1'\underline{X}' = \underline{O}$.

Since $\underline{W}^*$ is nonsingular we can write the model $\underline{y} = \underline{X}\underline{\beta} + \underline{e}$.

or $\underline{y} = \underline{X}\underline{W}^*(\underline{W}^*)^{-1}\underline{\beta} + \underline{e}$

Partitioning $(\underline{W}^*)^{-1} = \underline{U}^*$ into $\underline{U}^* = \begin{pmatrix} \underline{U} \\ \underline{U}_1 \end{pmatrix}$ we get

$$\begin{aligned} \underline{y} &= \underline{X}\,(\underline{W}, \underline{W}_1)\begin{pmatrix} \underline{U} \\ \underline{U}_1 \end{pmatrix}\underline{\beta} + \underline{e} \\ &= (\underline{X}\underline{W})(\underline{U}\underline{\beta}) + (\underline{X}\underline{W}_1)(\underline{U}_1\underline{\beta}) + \underline{e} \\ &= (\underline{X}\underline{W})(\underline{U}\underline{\beta}) + \underline{e} \qquad \text{since } \underline{X}\underline{W}_1 = \underline{O} \end{aligned}$$

Now putting $\underline{X}\underline{W} = \underline{Z}$ and $\underline{U}\underline{\beta} = \underline{\alpha}$, the model finally assumes the form $\underline{y} = \underline{Z}\underline{\alpha} + \underline{e}$ where $\underset{n \times k}{\underline{Z}}$ is of rank k; hence a full rank reparametrization.

$\underline{y} = \underline{Z}\underline{\alpha} + \underline{e}$ being a full rank model, all the theorems of model of full rank apply. Obviously $\underline{\alpha}$ and any other linear combination are estimable and the unique estimates of $\underline{\alpha}$ are given by the normal equations $(\underline{Z}'\underline{Z})\underline{\alpha} = \underline{Z}'\underline{y}$

We know that there are exactly k linearly independent estimable functions for a model $\underline{y} = \underline{X}\underline{\beta} + \underline{e}$ with rank $\underline{X} = k$. By the process of reparametrization we obtain new parameters $\underline{\alpha}^{k \times 1}$ which

are independent linear functions of $\underset{\sim}{\beta}$ and estimable . Thus we arrive at the following theorem.

Theorem. 2.9. If $\underset{\sim}{\alpha}_{k\times 1}$ is a vector of k linearly independent estimable functions of the parameters $\underset{\sim}{\beta}$ in the model $\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{e}$ where rank($\underset{\sim}{X}$) = k, then there exists a reparametrization to the full rank model

$$\underset{\sim}{y} = \underset{\sim}{Z}\underset{\sim}{\alpha} + \underset{\sim}{e} \quad .$$

Since the diagonalizing matrix $\underset{\sim}{W}^{*}$ of $\underset{\sim}{X}'\underset{\sim}{X}$ is not unique, we may have different full rank reparametrizations.

Now consider two full rank reparametrizations

$$\underset{\sim}{y} = \underset{\sim}{Z}\underset{\sim}{\alpha} + \underset{\sim}{e}$$

and $$\underset{\sim}{y} = \underset{\sim}{T}\underset{\sim}{\delta} + \underset{\sim}{e}$$

of the model $$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{e} \quad .$$

Then, by definition, this means that there exist matrices $\underset{\sim}{V}$ and $\underset{\sim}{U}$ each of rank k and dimension k x p such that,

$$\underset{\sim}{\alpha} = \underset{\sim}{U}\underset{\sim}{\beta} \quad \text{and} \quad \underset{\sim}{\delta} = \underset{\sim}{V}\underset{\sim}{\beta} \quad .$$

But, there exist a relation between $\underset{\sim}{U}$ and $\underset{\sim}{V}$ of the form $\underset{\sim}{U} = \underset{\sim}{A}\underset{\sim}{V}$ with $\underset{\sim}{A}$ nonsingular . This means that there is a linear relationship between full rank reparametrizations. Thus the two estimates of an estimable function $\underset{\sim}{\lambda}'\underset{\sim}{\beta}$ using two different full rank reparametrized models of $\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{e}$ will be the same giving rise to the following corrolary on the uniqueness of the estimates.

Corrolary .2.9.1. Any full rank reparametrization gives the same estimate of the estimable function $\underset{\sim}{\lambda}'\underset{\sim}{\beta}$.

2.2.3. Estimation of Linear Combination of β_i under case A.

If $\underline{e}$ is $N(\underline{0}, \sigma^2 \underline{I})$, maximizing the LF or the method of maximum likelihood leads to the same normal equations as the Least-square method under case A.

If the model $\underline{y} = \underline{X}\underline{\beta} + \underline{e}$ be reparametrized to the model of full rank $\underline{y} = \underline{Z}\underline{\alpha} + \underline{e}$, than, by the invariance property of Maximum Likelihood, the ML estimate of $\underline{\alpha}$ is $\hat{\underline{\alpha}} = (\underline{Z}'\underline{Z})^{-1}\underline{Z}'\underline{y}$ and all the properties of the corresponding theorems of full rank model apply to $\hat{\underline{\alpha}}$.

Theorem. 2.10 In the model of less than full rank, under case A, the maximum likelihood estimate of any estimable function of $\underline{\beta}$ has all the properties mentioned in the corresponding full rank theorem

where $\hat{\sigma}^2 = \frac{1}{n-k} (\underline{y} - \underline{Z}\hat{\alpha})'(\underline{y} - \underline{Z}\hat{\alpha})$

and $\underline{y} = \underline{Z}\underline{\alpha} + \underline{e}$ is the reparametrized model of rank k = rank ($\underline{X}$).

Proof. If $\underline{y} = \underline{Z}\underline{\alpha} + \underline{e}$ be a full rank $\underline{y} = \underline{X}\underline{\beta} + \underline{e}$, then by the reparametrizing the model $\underline{y} = \underline{X}\underline{\beta} + \underline{e}$, then by the invariance property of maximum-likelihood estimate , the maximum-likelihood estimate of $\underline{\alpha}$ is $\hat{\underline{\alpha}} = (\underline{Z}'\underline{Z})^{-1}\underline{Z}'\underline{y}$ and all the properties listed in Theorem apply to $\underline{y} = \underline{Z}\underline{\alpha} + \underline{e}$.

If $\underline{\lambda}'\underline{\beta}$ is estimable, the b.l.u. estimate can be shown to be $\underline{\lambda}'\hat{\underline{\beta}}$ where $\hat{\underline{\beta}}$ is any solution to the normal equations.

Since $\underline{\lambda}'\underline{\beta}$ is estimable its b.l.u. estimate is $\underline{r}'\underline{X}'\underline{y}$ where $\underline{r}$ is any solution of

$$\underline{X}'\underline{X}\,\underline{r} = \underline{\lambda} \qquad \ldots \qquad (1)$$

$\hat{\underline{\beta}}$ being, any solution of the normal equations $(\underline{X}'\underline{X})\hat{\underline{\beta}} = \underline{X}'\underline{y}$, the best linear estimate $\underline{r}'\underline{X}'\underline{y} = \underline{r}'(\underline{X}'\underline{X})\hat{\underline{\beta}}$

By (1) it follows that $\underline{r}'\underline{X}'\underline{y} = \underline{\lambda}'\hat{\underline{\beta}}$

We shall now consider a similar invariance property of the estimation of σ^2.

Theorem 2.11 Under A $\underline{y} = \underline{X}\underline{\beta} + \underline{e}$, rank $(\underline{X}) = k \leq p$,

(1) $\hat{\sigma}^2 = \frac{1}{n-k}(\underline{y} - \underline{X}\hat{\underline{\beta}})'(\underline{y} - \underline{X}\hat{\underline{\beta}})$

is invariant for any solution $\hat{\underline{\beta}}$ of the normal equations.

(2) $\frac{\hat{\sigma}^2}{\sigma^2}(n-k)$ is distributed $\chi^2(n-k)$ under Case B.

Proof. Using the reparametrized model of rank k,

$$\hat{\sigma}^2 = \frac{1}{n-k}(\underline{y} - \underline{Z}\hat{\underline{\alpha}})'(\underline{y} - \underline{Z}\hat{\underline{\alpha}})$$

where the transformations used are $\underline{XW} = \underline{Z}$ and $\underline{U}\underline{\beta} = \underline{\alpha}$

$\underline{W}$ and $\underline{U}$ being defined as in Theorem 2.9 .

Now, $\hat{\sigma}^2 = \frac{1}{n-k}(\underline{y} - \underline{Z}\hat{\underline{\alpha}})'(\underline{y} - \underline{Z}\hat{\underline{\alpha}})$

$= \frac{1}{n-k}(\underline{y} - \underline{Z}\underline{W}.\underline{U}\hat{\underline{\beta}})'(\underline{y} - \underline{Z}\underline{W}.\underline{U}\hat{\underline{\beta}})$

$= \frac{1}{n-k}(\underline{y} - \underline{X}\hat{\underline{\beta}})'(\underline{y} - \underline{X}\hat{\underline{\beta}})$.

Since $\underline{\alpha}$ is estimable, it follows that the best unbiased linear

estimate $\hat{\underline{\alpha}} = \underline{U}\hat{\underline{\beta}} = \underline{U}\hat{\underline{\beta}}$, $\hat{\underline{\beta}}$ being any solution of the normal equation i.e. whatever solution we may use for $\hat{\underline{\beta}}$, we always get the same $\hat{\sigma}^2$.

(2) $\frac{\hat{\sigma}^2}{\sigma^2}(n-k) = \frac{(\underline{y} - \underline{Z}\underline{\alpha})'(\underline{y} - \underline{Z}\underline{\alpha})}{\sigma^2} = \frac{(\underline{y} - \underline{X}\hat{\underline{\beta}})'(\underline{y} - \underline{X}\hat{\underline{\beta}})}{\sigma^2}$

Since $\underline{y} = \underline{Z}\underline{\alpha} + \underline{e}$ is a model of full rank, $\frac{(\underline{y} - \underline{Z}\underline{\alpha})'(\underline{y} - \underline{Z}\underline{\alpha})}{\sigma^2}$ is distributed $\chi^2(n-k)$ under case B.

Thus the theorem is proved.

2.2.4 Testing of Hypothesis .

One of the useful test of the general linear hypothesis of less than full rank is of the form:

H: $\beta_1 = \beta_2 = \cdots = \beta_h \quad (h \leq k)$

We shall discuss the testing of the estimable hypothesis, which is defined below, only.

Definition: The hypothesis H : $\beta_1 = \beta_2 = \cdots = \beta_h$ is defined to be estimable if there exists a set of linearly independent estimable functions, $\underline{\lambda}_1'\underline{\beta}$, $\underline{\lambda}_2'\underline{\beta}$, $\cdots$ $\underline{\lambda}_s'\underline{\beta}$ such that H is true if and only if

$$\underline{\lambda}_1'\underline{\beta} = \underline{\lambda}_2'\underline{\beta} = \cdots = \underline{\lambda}_s'\underline{\beta} = 0 .$$

In general, if we want to test the hypothesis

H : $\beta_1 = \beta_2 = \cdots = \beta_h \quad (h \leq k)$

in the model of less than full rank, we have to find whether there exists a set of linearly independent estimable functions

$$\underline{\lambda}_1'\underline{\beta}, \ \underline{\lambda}_2'\underline{\beta}, \ \cdots \ \underline{\lambda}_s'\underline{\beta},$$

such that H is true if and only if

$$\begin{pmatrix} \underline{\lambda}_1'\underline{\beta} \\ \vdots \\ \underline{\lambda}_s'\underline{\beta} \end{pmatrix} = \underline{0}$$

Consider $(k-s)$ more estimable functions $\underline{\lambda}_{s+1}'\underline{\beta}, \ldots, \underline{\lambda}_k'\underline{\beta}$ such that $\underline{\lambda}_1'\underline{\beta}, \underline{\lambda}_2'\underline{\beta}, \ldots, \underline{\lambda}_k'\underline{\beta}$ form a set of k linearly independent estimable functions.

Now, if we let

$$\underline{\delta} = \begin{pmatrix} \underline{\alpha} \\ \underline{\gamma} \end{pmatrix} = \begin{pmatrix} \underline{\lambda}_1'\underline{\beta} \\ \vdots \\ \underline{\lambda}_s'\underline{\beta} \\ \underline{\lambda}_{s+1}'\underline{\beta} \\ \vdots \\ \underline{\lambda}_k'\underline{\beta} \end{pmatrix} \qquad \text{where } \underline{\alpha} = \begin{pmatrix} \underline{\lambda}_1'\underline{\beta} \\ \vdots \\ \underline{\lambda}_s'\underline{\beta} \end{pmatrix}$$

$$\text{and } \underline{\gamma} = \begin{pmatrix} \underline{\lambda}_{s+1}'\underline{\beta} \\ \vdots \\ \underline{\lambda}_k'\underline{\beta} \end{pmatrix}$$

we can reparametrize from the less than full rank model $\underline{y} = \underline{X}\underline{\beta} + \underline{e}$ to the full rank model

$$\underline{y} = \underline{Z}\,\underline{\delta} + \underline{e}$$
$$= \underline{Z}_1\,\underline{\alpha} + \underline{Z}_2\,\underline{\gamma} + \underline{e}$$

Now the hypothesis H is true if and only if $\underline{\alpha} = \underline{0}$; to test H we may use theorem 2.3 .

In order to use theorem 2.3 , we have to know

(1) Q_0 = Sum of squares due to error , $(\underline{y} - \underline{Z}\,\hat{\underline{\delta}})'(\underline{y} - \underline{Z}\,\hat{\underline{\delta}})$

where $\hat{\underline{\delta}}$ is the solution of the normal equations

(2) Q_1 = Sum of squares due to $\underline{\alpha}$ adjusted for $\underline{\gamma}$ = $R(\underline{\alpha} \mid \underline{\gamma})$

$$= \underline{\delta}' \underline{Z}' \underline{y} - \hat{\underline{\gamma}}' \underline{Z}_2' \underline{y}$$

where $\hat{\underline{\gamma}}$ is the solution of the normal equations

$$\underline{Z}_2' \underline{Z}_2 = \underline{Z}_2' \underline{y}$$

But from the theorem on the invariance property of $\hat{\sigma}^2$ it follows that

$$(\underline{y} - \underline{Z} \hat{\underline{\delta}})'(\underline{y} - \underline{Z} \hat{\underline{\delta}}) = (\underline{y} - \underline{X} \hat{\underline{\beta}})'(\underline{y} - \underline{X} \hat{\underline{\beta}})$$

By the similar argument the term $(\underline{y} - \underline{Z}_2 \hat{\underline{\gamma}})'(\underline{y} - \underline{Z}_2 \hat{\underline{\gamma}})$ can be obtained from the normal equations that are derived from the model $\underline{y} = \underline{X}' \underline{\beta} + \underline{e}$ with the condition $\beta_1 = \beta_2 = \cdots = \beta_h$

Thus we have the following theorem:

Theorem: 2.11 In less than full rank model $\underline{y} = \underline{X} \underline{\beta} + \underline{e}$, to test the hypothesis $\beta_1 = \beta_2 = \cdots = \beta_h$ $(h \leq k)$, which we assume is equivalent to testing the linearly estimable functions $\underline{\lambda}_1' \underline{\beta} = \underline{\lambda}_2' \underline{\beta} = \cdots = \underline{\lambda}_s' \underline{\beta}$ the statistic

$$\lambda_H = \frac{n-k}{s} \frac{Q_1}{Q_0} = \frac{\hat{\underline{\beta}}' \underline{X}' \underline{y} - \hat{\underline{\gamma}}' \underline{Z}_2' \underline{y}}{\underline{y}' \underline{y} - \hat{\underline{\beta}}' \underline{X}' \underline{y}} \cdot \frac{n-k}{s}$$

is used and is found to be distributed as $F(s, n-k, \delta)$ where

$Q_0 = (\underline{y} - \underline{X} \hat{\underline{\beta}})'(\underline{y} - \underline{X} \hat{\underline{\beta}})$ with any solution from the normal equations $\underline{X}' \underline{X} \hat{\underline{\beta}} = \underline{X}' \underline{y}$.

and

$Q_0 + Q_1 = (\underline{y} - \underline{Z}_2 \hat{\underline{\gamma}})'(\underline{y} - \underline{Z}_2 \hat{\underline{\gamma}})$ with any solution from the normal equations $\underline{Z}_2' \underline{Z}_2 = \underline{Z}_2' \underline{y}$ i.e. the

normal equations of the reduced model derived from the model

$\underline{y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$ under the conditions, $\beta_1 = \beta_2 = \cdots = \beta_h$.

2.3 General form of analysis of the model of less than full rank.

In our formulation of the linear model of full rank, we assumed that the number of observations n was larger than p, the number of parameters to be estimated. Now in our model we allow any value of n and suppose $\underline{X}$ to have rank $k < p$.

Our assumption is now generalised into the form

Ω : $\underline{y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$, rank $\underline{X} = k < p$ with any n.

We have shown already that we cannot find a linear estimator for $\hat{\underline{\beta}}$. In order to resolve this difficulty we transform the model by reparametrization to a full rank model of the form

$$\underline{y} = \underline{Z}\underline{\alpha} + \underline{\varepsilon} .$$

Now we shall consider the method of augmentation (due to Plackett) by which we found an unbiased estimator of $\underline{\beta}$ after introducing a set of linear constraints ,

$\underline{c} = \underline{B}\underline{\beta}$ where $\underline{c}$ is a $(p-k) \times 1$ vector of constants, $\underline{B}$ is a $(p-k) \times p$ matrix of rank $(p-k)$.

If we seek a linear estimator of the form

$$\hat{\underline{\beta}} = \underline{L}\underline{y} + \underline{N}\underline{c} .$$

the unbiasedness condition will be

$E(\hat{\underline{\beta}}) = E(\underline{Ly} + \underline{Nc})$ identically in

$\underline{\beta} = \underline{L}\underline{X}\underline{\beta} + \underline{NB}\underline{\beta}$

or $\underline{I} = \underline{L}\underline{X} + \underline{N}\underline{\beta}$.

Remembering that $\underline{X}$ is of rank k , we partition $\underline{X}$ into

$$\underline{X} = \begin{pmatrix} \underline{X}_{k,k} & \underline{X}_{k,p-k} \\ \underline{X}_{n-k,k} & \underline{X}_{n-k,p-k} \end{pmatrix}$$

the suffixes of the matrix elements indicating the number of rows and columns. Without any loss of generality, we may assume that $\underline{X}_{k,k}$ is nonsingular.

Define a new matrix of order $p \times (p-k)$,

$$\underline{D} = \begin{pmatrix} \underline{X}^{-1}_{k,k} \cdot \underline{X}_{k,p-k} \\ -\underline{I}_{p-k} \end{pmatrix}$$ where $\underline{I}_{p-k}$ is the identity matrix of that order.

Evidently, $\underline{D}$ is of rank $p-k$.

Then the product $\underline{X}\,\underline{D}$ will have its first k rows of zeroes.

Since rank $\underline{X} = k$, rank $\underline{X}\,\underline{D} \leq k$ and therefore all its rows are linearly dependent upon its first k rows.

Hence $\underline{X}\,\underline{D}$ consists entirely of zeroes, i.e.

$$\underline{X}\,\underline{D} = \underline{0}. \qquad (2.3.1)$$

Now $\underline{D} = \underline{I}\,\underline{D} = (\underline{L}\,\underline{X} + \underline{N}\,\underline{B})\,\underline{D} = \underline{L}\,\underline{X}\,\underline{D} + \underline{N}\,\underline{B}\,\underline{D} = \underline{N}\,\underline{B}\,\underline{D}$.

In order to avoid contradiction we shall assume that $\underline{B}\,\underline{D}$ is nonsingular.

<u>To Find a L.S. Solution Using The Augmented Model.</u>

$\underline{B}$, of rank $(p-k)$, makes up the deficiency in rank of $\underline{X}$. Regarding $\underline{c}$ as a vector of dummy random variables in the augmented model ,

$$\begin{pmatrix} \underline{Y} \\ \underline{c} \end{pmatrix} = \begin{pmatrix} \underline{X} \\ \underline{B} \end{pmatrix} \underline{\beta} + \begin{pmatrix} \underline{e} \\ \underline{0} \end{pmatrix}$$

In order to use the methods of the model of full rank; we have to see

whether $\begin{pmatrix} \underline{X} \\ \underline{B} \end{pmatrix}' \begin{pmatrix} \underline{X} \\ \underline{B} \end{pmatrix} = (\underline{X}'\underline{X} + \underline{B}'\underline{B})$ may be inverted.

This is a matrix of a non-negative quadratic form.

Consider $(\underline{X}'\underline{X} + \underline{B}'\underline{B})\,\underline{D} = \underline{X}'\underline{X}\,\underline{D} + \underline{B}'\underline{B}\,\underline{D}$

$= \underline{0} + \underline{B}'\underline{B}\,\underline{D}$, by virtue of (2.3.1)

Since $\underline{B}\,\underline{D}$ is nonsingular, rank $(\underline{B}'\underline{B}\,\underline{D})$ = rank $\underline{B}$

is rank $(\underline{X}'\underline{X} + \underline{B}'\underline{B})\,\underline{D}$ = rank B' = p-k.

= rank $\underline{D}$

i.e. $\underline{X}'\underline{X} + \underline{B}'\underline{B}$ is nonsingular.

From the theorem of full rank follows,

$$\hat{\underline{\beta}} = (\underline{X}'\underline{X} + \underline{B}'\underline{B})^{-1}(\underline{X}'\underline{y} + \underline{B}'\underline{c})$$

and var $(\hat{\underline{\beta}}) = \sigma^2 (\underline{X}'\underline{X} + \underline{B}'\underline{B})^{-1}\,\underline{X}'\underline{X}\,(\underline{X}'\underline{X} + \underline{B}'\underline{B})^{-1}$

The matrix $\underline{B}$ is arbitrary, subject to the condition $|\underline{B}\,\underline{D}| \neq 0$. In fact, if for $\underline{B}$ we use $\underline{U}\,\underline{B}$, where $\underline{U}$ is nonsingular (p-k) x (p - k) matrix, the expressions for $\hat{\underline{\beta}}$ and var $(\hat{\underline{\beta}})$ are unaltered in value.

2.4 Reduction of the case where the observations have correlations and known ratios of variances:-

We shall now consider the case when the covariance matrix $V_{\underline{y}}$ of the observations $\underline{y}$ is not of the form $\sigma^2 \underline{I}$ but $V_{\underline{y}}$ is known except for a scalar factor, i.e. $V_{\underline{y}} = \theta\,\underline{B}$, where θ is an unknown positive constant and $\underline{B}$ is a known constant matrix; $\underline{B}$ is necessarily symmetric and positive indefinite, and we shall assume furthermore that it is nonsingular. This is equivalent to knowing the correlation coefficient of all pairs of observations y_i and the ratios of their variances.

Our underlying assumptions are now

$$\Omega \quad : \quad E(\underline{y}) = \underline{X}\underline{\beta} \quad , \; V_{\underline{y}} = \sigma^2 \underline{B} \quad , \; |\underline{B}| \neq 0 \quad \text{and rank } \underline{X} = k \leq p$$

This case may now be reduced to the previously considered case, where

$V_{\underline{y}} = \sigma^2 \underline{I}$.

Since $\underline{B}$ is nonsingular, there exists a nonsingular $\underline{P}^{n \times n}$ such that

$$\underline{P}' \underline{B} \underline{P} = \underline{I}$$

Now put $\underline{y}^* = \underline{P}' \underline{y}$. Then

$$E(\underline{y}^*) = \underline{P}' E(\underline{y}) = \underline{P}' \underline{X} \underline{\beta} = \underline{X}^* \underline{\beta} \quad \text{where } \underline{X}^* = \underline{P}' \underline{X}$$

and

$$V_{\underline{y}^*} = \underline{P}' V_{\underline{y}} \underline{P} = \sigma^2 \underline{P}' \underline{B} \underline{P} = \sigma^2 \underline{I} \quad , \text{ where } \sigma^2 = \sigma^2 .$$

We may thus arrive at the required form of assumption

$$\Omega \quad : \quad E(\underline{y}^*) = \underline{X}^* \underline{\beta} \quad , \; V_{\underline{y}^*} = \sigma^2 \underline{I} \quad , \text{rank } \underline{X}^* = k$$

gives every justification for the assumption

$$\Omega \quad : \quad E(\underline{y}) = \underline{X}\underline{\beta} \quad , \; V_{\underline{y}} = \sigma^2 \underline{I} \quad , \text{ rank } \underline{X} = k \; ,$$

and there is obviously no loss of generality in this assumption of our theoretical consideration.

Chapter III

3.1 The General Linear Hypothesis and its Cannonical Form.

We shall now consider the problem of testing hypothesis in the general linear model, using the cannonical form which will be used in the study of the optimum properties of A O V test.

Ω : $\underline{y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$, rank $\underline{X} = p$, $E(\underline{\varepsilon}) = \underline{0}$, $E(\underline{\varepsilon}\underline{\varepsilon}') = \sigma^2 \underline{I}$.

H : $\underline{\Lambda}^{r \times p}\, \underline{\beta}^{p \times 1} = \underline{\varepsilon}^{r \times 1}$.

The hypothesis imposes $r\ (\leq p)$ constraints, which we shall take to be functionally independent so that $\underline{\Lambda}$ is of rank r .

If $\underline{A}$ is the first r rows of the (nxp) matrix $\underline{X}$, our hypothesis is concerned with the means of the first r y_i's .

We shall develop the theory with the model of full rank. If not, we can either use the method of augmentation or the method of reparametrization to get a model of full rank, under the assumption that $\underline{\Lambda}\,\underline{\beta}$ is a vector of estimable functions.

Now consider (nx1) vector,

$$\underline{z} = \underline{C}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y} = \underline{C}\underline{M}^{-1}\underline{X}'\underline{y} = \underline{C}\hat{\underline{\beta}},$$

where $\underline{C}$ is a (nxp) matrix and $\hat{\underline{\beta}}$ is the L S estimator of $\underline{\beta}$.

Then $\underline{z} = (\underline{C}\underline{M}^{-1}\underline{X}')(\underline{X}\underline{\beta} + \underline{\varepsilon})$

$= \underline{C}\underline{\beta} + \underline{C}\underline{M}^{-1}\underline{X}'\underline{\varepsilon}$

so that $\underline{\mu} = E(\underline{z}) = \underline{C}\underline{\beta}$

and var $(\underline{z}) = E[(\underline{z} - \underline{\mu})(\underline{z} - \underline{\mu})']$

$= E[(\underline{C}\underline{M}^{-1}\underline{X}'\underline{\varepsilon})(\underline{C}\underline{M}^{-1}\underline{X}'\underline{\varepsilon})']$

$= \underline{C}\,\underline{M}^{-1}\underline{X}'\,E(\underline{\varepsilon}\underline{\varepsilon}')\,\underline{X}\,\underline{M}^{-1}\underline{C}'$.

$$= \sigma^2 C M^{-1} X' X M^{-1} C' = \sigma^2 C M^{-1} C'$$

Now we may choose C so that the components of z are uncorrelated i.e. so that,

$$\text{var}(z) = \sigma^2 I$$

$$\therefore C M^{-1} C' = I$$

or if $C' C$ is nonsingular $C' C = M = X' X$.

which is the condition that the z_i be uncorrelated.

This transformation implies that

$$\begin{aligned}
(1)\ (z-\mu)'(z-\mu) &= (C M^{-1} X' e)'(C M^{-1} X' e) \\
&= e' X M^{-1} C' C M^{-1} X' e \\
&= e' X M^{-1} X' e = e' X (X' X)^{-1} X' e \\
(11)\ (z-\mu)'(z-\mu) &= \{C(\hat{\beta}-\beta)\}' \{C(\hat{\beta}-\beta)\} \\
&= (\hat{\beta}-\beta)' C' C (\hat{\beta}-\beta) \\
&= (\hat{\beta}-\beta)' X' X (\hat{\beta}-\beta) \\
&= \{X(\hat{\beta}-\beta)\}' \{X(\hat{\beta}-\beta)\}
\end{aligned} \qquad (3.0)$$

Now consider the matrix

$$C = \begin{pmatrix} A \\ D \\ F \end{pmatrix}$$

where $A^{r \times p}$ is the coefficient matrix of the hypothesis, D is $(p - r)$ xp matrix and F is a $(n - p)$xp matrix satisfying

$$F\beta = 0 \qquad (3.1)$$

Since A is of rank r, we can choose D so that the pxp matrix

is nonsingular, and thus $\underline{C}$ is of rank p. Then $\underline{C}'\underline{C}$ will also be of rank p, and hence nonsingular as required above.

Now we have, $\underline{\mu} = E(\underline{z}) = \underline{C}\underline{\beta} = \begin{pmatrix} \underline{\Lambda} \\ \underline{D} \\ \underline{F} \end{pmatrix} \underline{\beta}$

in which the first r z_i are precisely the L.H.S. of the statement of the hypothesis H .

Thus H is equivalent to testing

$$H : \mu_i = E(z_i) = \underline{\lambda}_i' \underline{\beta} = c_i, \quad i = 1, 2, \ldots r$$

a composite hypothesis imposing r constraints upon the parameters. From equation (3.1) follows that the last n-p of the μ_i are zero, so that there are p non-zero parameters μ_i, which together with σ^2 make up a total of (p+1) parameters.

We have thus reduced our problem to the following cannonical form:

"A set of n mutually uncorrelated variates z_i with equal variances σ^2, of which (n-p) have zero means and the others non-zero means".

The hypothesis to be tested is that r of p variates have specified means. In order to test hypothesis we need to make assumptions about the distribution of errors or $\underline{y}$. $\underline{y}$ is as assumed $N(\underline{X}\underline{\beta}, \sigma^2\underline{I})$ and hence since they are uncorrelated, independent. The $\underline{z}$, being linear function of $\underline{y}$, will be normally distributed and, being uncorrelated, independently normally distributed.

3.1.1 Connonical form of the assumptions and the Null Hypothesis.

Ω : $\underline{z}$ is $N(\underline{\mu}, \sigma^2 \underline{I})$ $\qquad \underline{\mu} = \begin{pmatrix} \underline{\mu}_r \\ \underline{\mu}_{p-r} \\ \underline{\mu}_{n-p} \end{pmatrix}$

with $\underline{\mu}_{n-p} = \underline{0}$.

(the suffixes denoting the number of components)

When we have more than one constraint, i.e. when r > 1 we know that there is no common uniformly most powerful unbiased critical region. Since there is no "optimum" test, we are tempted to use the L.R. method to give an intutively reasonable test.

Nevertheless the likelihood ratio test of the general linear hypothesis have certain optimum properties which we shall discuss briefly below.

3.1.2 The Derivation of L R statistic using the cannonical form.

$\underline{z}$ is $N(\underline{\mu}, \sigma^2 \underline{I})$

with $\underline{\mu} = \begin{pmatrix} \underline{\mu}_r \\ \underline{\mu}_{p-r} \\ \underline{\mu}_{n-p} \end{pmatrix}$

where $\underline{\mu}_{n-p} = \underline{0}$ and $\begin{pmatrix} \underline{\mu}_r \\ \underline{\mu}_{p-r} \end{pmatrix}$ are unspecified

H_o: $\underline{\mu}_r = \underline{\mu}^*_r$

In order to find the likelihood ratio statistic, we may redefine and H as follows:

$\underline{z}$ is distributed $N(\underline{\mu}, \sigma^2 \underline{I})$ with $\mu_i = 0$ for $i = p+1, \ldots, n$.

Ω : the half $(p + 1)$ - dimensional Euclidian Space

for which $-\infty < \mu_i < +\infty$, $i = 1, 2, \ldots p$, $\sigma^2 > 0$.

ω : the subset of Ω for which $\mu_i = \mu_i^*$, $i = 1, 2, \ldots r < p$.

The likelihood function of the sample

$$L = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\underline{z} - \underline{\mu})'(\underline{z} - \underline{\mu})\right\}$$

Therefore $\log L = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(\underline{z} - \underline{\mu})'(\underline{z} - \underline{\mu})$

$$= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(\underline{z}_p - \underline{\mu}_p)'(\underline{z}_p - \underline{\mu}_p) - \frac{1}{2\sigma^2}\underline{z}'_{n-p}\,\underline{z}_{n-p}$$

The unconditional maximum of L is obtained by solving

$$\left.\begin{aligned}\frac{\partial \log L}{\partial \underline{\mu}_p} &= \underline{0}\\ \frac{\partial \log L}{\partial \sigma^2} &= 0\end{aligned}\right\} \Rightarrow \qquad -\frac{1}{2\sigma^2}(\underline{z}_p - \underline{\mu}_p) = 0$$

$$-\frac{n}{2\sigma^2} + \frac{(\underline{z}_p - \underline{\mu}_p)'(\underline{z}_p - \underline{\mu}_p) + \underline{z}'_{n-p}\,\underline{z}_{n-p}}{2\sigma^4} = 0$$

giving $\hat{\underline{\mu}}_p = \underline{z}_p$

$$\hat{\sigma}^2 = \frac{1}{n}\,\underline{z}'_{n-p}\,\underline{z}_{n-p}$$

whence, $(L)_\Omega = (2\pi\hat{\sigma}^2 e)^{-n/2}$

$$= \left(\frac{2\pi}{n} e\ \underline{z}'_{n-p}\,\underline{z}_{n-p}\right)^{-\frac{n}{2}}$$

Now under H, the likelihood function

$$L = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[(\underline{z}_r - \underline{\mu}_r^*)'(\underline{z}_r - \underline{\mu}_r^*) + (\underline{z}_{p-r} - \underline{\mu}_{p-r})'(\underline{z}_{p-r} - \underline{\mu}_{p-r}) + \underline{z}'_{n-p}\,\underline{z}_{n-p}\right]\right\}$$

Therefore $\log L = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(\underline{z}_r - \underline{\mu}_r^*)'(\underline{z}_r - \underline{\mu}_r^*)$

$$- \frac{1}{2\sigma^2}(\underline{z}_{p-r} - \underline{\mu}_{p-r})'(\underline{z}_{p-r} - \underline{\mu}_{p-r}) - \frac{1}{2\sigma^2}\underline{z}'_{n-p}\,\underline{z}_{n-p}$$

Therefore the values of σ^2 and $\underline{\mu}_{p-r}$ that minimize log L afe given by

$$\left.\begin{aligned}\frac{\partial \log L}{\partial \underline{\mu}_{p-r}} &= \underline{0}\\ \frac{\partial \log L}{\partial \sigma^2} &= 0\end{aligned}\right\} \Rightarrow \qquad -\frac{1}{2\sigma^2}(\underline{z}_{p-r} - \underline{\mu}_{p-r}) = \underline{0}$$

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\left\{(\underline{z}_r - \underline{\mu}_r^*)'(\underline{z}_r - \underline{\mu}_r^*) + (\underline{z}_{p-r} - \underline{\mu}_{p-r})'(\underline{z}_{p-r} - \underline{\mu}_{p-r}) + \underline{z}'_{n-p}\,\underline{z}_{n-p}\right\} = 0$$

giving $\underline{z}_{p-r} = \hat{\underline{\mu}}_{p-r}$

and $\hat{\hat{\sigma}}^2 = \frac{1}{n}\left[(\underline{z}_r - \underline{\mu}^*_r)'(\underline{z}_r - \underline{\mu}^*_r) + \underline{z}'_{n-p}\,\underline{z}_{n-p} \right]$

therefore $(L)_\omega = (2\pi\hat{\hat{\sigma}}^2 e)^{-\frac{n}{2}} = \left[2\pi\frac{n}{n} e \left\{ \underline{z}'_{n-p}\,\underline{z}_{n-p} + (\underline{z}_r - \underline{\mu}^*_r)'(\underline{z}_r - \underline{\mu}^*_r) \right\} \right]^{-\frac{n}{2}}$

The L. R. statistic

$$\lambda_H = \frac{(L)_\omega}{(L)_\Omega} = \frac{\hat{\sigma}^2}{\hat{\hat{\sigma}}^2} = \frac{1}{1+u}$$

where $u = (\underline{z}_r - \underline{\mu}^*_r)'(\underline{z}_r - \underline{\mu}_r) \Big/ \underline{z}'_{n-p}\,\underline{z}_{n-p}$

Now $u = \dfrac{(\underline{z}_r - \underline{\mu}^*_r)'(\underline{z}_r - \underline{\mu}_r)/\sigma^2}{\underline{z}'_{n-p}\,\underline{z}_{n-p}/\sigma^2}$ of which the numerator and denominator, being the sum of squares of independent normal variates are distributed as χ^2 distributions with r and n-p d.f. respectively, when H_o holds. Thus, when H_o is true, $F = \frac{(n-p)}{r} u$ is distributed as F distribution with (r, n-p) d.f.

3.1.3 Computational aspect of the above theorem.

In the above discussion we encounter that $n\hat{\hat{\sigma}}^2$ and $n\hat{\sigma}^2$ are respectively the minima of $(\underline{z} - \underline{\mu})'(\underline{z} - \underline{\mu})$ with respect to $\underline{\mu}$ under H_o and H_1.

But we know from (3.0)

$$(\underline{z} - \underline{\mu})'(\underline{z} - \underline{\mu}) = \left\{ \underline{X}(\hat{\underline{\beta}} - \underline{\beta}) \right\}' \left\{ \underline{X}(\hat{\underline{\beta}} - \underline{\beta}) \right\}$$

Therefore there are the same as the minima of

$$S = \left\{ \underline{X}(\hat{\underline{\beta}} - \underline{\beta}) \right\}' \left\{ \underline{X}(\hat{\underline{\beta}} - \underline{\beta}) \right\}$$

with respect to $\underline{\beta}$.

Now consider the identity

$$S = \underline{e}'\underline{e} = (\underline{y} - \underline{X}\hat{\underline{\beta}})'(\underline{y} - \underline{X}\hat{\underline{\beta}}) + (\underline{y} - \underline{X}\underline{\beta})'(\underline{y} - \underline{X}\underline{\beta}) + S.$$

in which the form $(\underline{y}-\underline{X}\hat{\underline{\beta}})'(\underline{y}-\underline{X}\hat{\underline{\beta}})$ is independent of $\underline{\beta}$.

Therefore when S is minimized with respect to $\underline{\beta}$, so is $\mathcal{S}$.

The minimum values of $\mathcal{S}$ and S occur together. To obtain $\hat{\sigma}^2$ and $\hat{\hat{\sigma}}^2$ for the test statistic, we minimize $\underline{e}'\underline{e}$ in the original model.

Example.

As a special case of particular importance, we shall consider the hypothesis.

$$H : \underline{\beta}_r = \underline{0}$$

where $\underline{\beta}_r$ is a subvector of $\underline{\beta}$.

(i.e. a particular case of theorem)

We may now partition $\underline{X}$ and $\underline{\beta}$ conformably to get

$$\underline{Y} = (\underset{n\times r}{\underline{X}_1}, \underset{n\times p-r}{\underline{X}_2}) \begin{pmatrix} \underline{\beta}_r \\ \underline{\beta}_{p-r} \end{pmatrix} + \underline{e}$$

Thus when H_0 is true we have

$$\underline{Y} = \underline{X}_2 \underline{\beta}_{p-r} + \underline{e}$$

As explained above, we shall now consider the minima of $\mathcal{S} = \underline{e}'\underline{e}$.

As in theorem we can get the minimum under H_0 and H as follows:

$$n\hat{\hat{\sigma}}^2 = \underline{y}'\{\underline{I} - \underline{X}_2(\underline{X}_2'\underline{X}_2)^{-1}\underline{X}_2'\}\underline{y}$$

and $$n\hat{\sigma}^2 = \underline{y}'\{\underline{I} - \underline{X}(\underline{X}\underline{X})^{-1}\underline{X}'\}\underline{y}$$

The statistic

$$F = \frac{n-p}{r}\left(\frac{\hat{\hat{\sigma}}^2 - \hat{\sigma}^2}{\hat{\sigma}^2}\right) \quad \text{is distributed as } F(r, n-p).$$

3.1.4 Power Function of the L.R. test of the linear hypothesis.

We have seen that the L.R. test is based on the statistic

$$u = \frac{(\underset{\sim}{z}_{r} - \underset{\sim}{\mu}^{*}_{r})'(\underset{\sim}{z}_{r} - \underset{\sim}{\mu}^{*}_{r})}{\sigma^2} \Big/ \frac{\underset{\sim}{z}'_{n-p}\,\underset{\sim}{z}_{n-p}}{\sigma^2}$$

The denominator $\left(\frac{\underset{\sim}{z}'_{n-p}\,\underset{\sim}{z}_{n-p}}{\sigma^2}\right)$ is distributed as χ^2 with (n-p) d.f. no matter whether H_o holds or not. But the distribution of the numerator $\frac{(\underset{\sim}{z}_{r} - \underset{\sim}{\mu}^{*}_{r})'(\underset{\sim}{z}_{r} - \underset{\sim}{\mu}^{*}_{r})}{\sigma^2}$ depends on the H_o, and is distributed as a $\chi^2(r)$ variate only when H_o is true. But it will always be distributed as a non-central χ^2 variate with r d.f. and noncentrality

non-central χ^2 variate with r d.f. and noncentrality

$$\delta = \frac{1}{\sigma^2}(\underset{\sim}{\mu}_{r} - \underset{\sim}{\mu}^{*}_{r})'(\underset{\sim}{\mu}_{r} - \underset{\sim}{\mu}^{*}_{r})$$

where $\underset{\sim}{\mu}_{r}$ is the true mean of $\underset{\sim}{z}$.

Only when H_o is true, δ equals to zero, giving a χ^2 distribution. Now in order to evaluate the power of the L.R. test, we must know the distribution of u (or equivalently F) when the null hypothesis is not true. We know that the statistic F is always distributed as the non-central F distribution with d.f. n-p,r and noncentrality δ , the same as that of the non-central χ^2 distribution of the numerator.

3.2 Optimum Properties of the Likelihood Ratio Test of the General Linear Hypothesis:

We shall confine our discussion to the cannonical form:

Ω : $\{z_i\}$ are statistically independent .

$\underline{z}$ is $N(\underline{\mu}, \sigma^2 \underline{I})$

with $\underline{\mu} = \begin{pmatrix} \underline{\mu}_r \\ \underline{\mu}_{p-r} \\ \underline{\mu}_{n-p} \end{pmatrix}$ where $\underline{\mu}_{n-p} = \underline{0}$

H : $\underline{\mu}_r = \underline{\mu}_r^*$

In connection with the cannonical form we shall briefly review some concepts of the general theory of testing hypothesis (Neyman-Pearson Theory). We shall consider (p + 1) -dimensional space of the parameter $(\mu_1, \cdots \mu_p; \sigma)$ and denote any point in it by the vector $\underline{\theta} = (\mu_1, \cdots \mu_p; \sigma)$. Then the joint probability density $p_{\underline{\theta}}(\underline{z})$ of $\underline{z}$ is completely specified by the value of $\underline{\theta}$. Without any danger of confusion we may use the symbols Ω and ω for the sets of $\underline{\theta}$ corresponding to the assumptions Ω and ω , where $\omega = H \cap \Omega$ the set of assumptions obtained by imposing the conditions of H on the initial assumptions Ω . That is Ω denotes the (p + 1) -dimensional Euclidian half space,

$$\Omega = \{ \underline{\theta} \mid -\infty < \mu_i < +\infty \ , \ i = 1, 2, \cdots p ; \ \sigma > 0 \} .$$

known as the parameter space, while ω is a subset of the parameter space and is of dimension p + 1 - r.

$$\omega = \{ \underline{\theta} \mid \underline{\mu}_r = \underline{\mu}_r^* ; \ -\infty < \mu_r < +\infty \ , \ i = r+1, \cdots, p ; \ \sigma > 0 \}$$

Besides the parameter space, we also need to consider the n-dimensional sample space $E^{(n)}$ of the vector of observations $\underline{z} = (z_1, z_2, \cdots z_n)'$. The choice of a test of the null hypothesis consists in the choice of a (Borel) set W in $E^{(n)}$. The test procedure is that if $\underline{z} \in W$, we reject H, otherwise we accept H. W is called the critical set or critical region of the test.

Two types of errors are recognised here:

A Type I error is committed if H is rejected when it is true, and a type II error is committed if H is accepted when it is false. There will be no ambiguity if, for the sake of brevity, we refer to W as the test of H. The power of the test, or the power of W, defined to be the probability of rejecting H when $\underline{\theta}$ is the true parameter point, is a function of θ and W.

$$P(\underline{\theta}, W) = \int_W p_{\underline{\theta}}(\underline{z})\, d\underline{z} ,$$

where the integral is taken with respect to n-dimensional Libesgue measure.

In this expression for the power of the test ranges over

Definition: The size of the critical region W (also is known as the level of significance) is the quantity $\sup_{\underline{\theta} \in \omega} P(\underline{\theta}, W)$.

Desirable properties of the test W:

1. W is an unbiased critical region of size α if it has size α and $P(\underline{\theta}, W)$ for all $\underline{\theta} \in \Omega - \omega$ where $\Omega - \omega$ denotes the set of all points not in ω and corresponds to the alternative hypothesis

permitted alternative to H.

$$i.e. \quad P(W|\underline{\theta}) \leq \alpha \quad \forall\, \underline{\theta} \in \omega$$

$$\text{and} \quad P(W|\underline{\theta}) \geq \alpha \quad \forall\, \underline{\theta} \in \Omega - \omega .$$

2. W is a similar region of size α if $P(\underline{\theta}, W) = \alpha$ for all $\underline{\theta} \in \omega$.

3. If W and W^* are two critical regions of size α for testing H such that

$$P(W^*|\underline{\theta}) \leq P(W|\underline{\theta}) \quad , \forall\, \underline{\theta} \in \omega$$

$$P(W^*|\underline{\theta}) > P(W|\underline{\theta}) \quad , \forall\, \underline{\theta} \in \Omega - \omega .$$

then the critical region W is said to be uniformly more powerful than the critical region W.

More generally W is uniformly most powerful (UMP) critical region of a given class $\mathcal{C}$ if $W \in \mathcal{C}$ and if, for any $W^* \in \mathcal{C}$ and

$$(i) \quad P(W|\underline{\theta}) < P(W^*|\underline{\theta}) \quad , \quad \underline{\theta} \in \omega .$$

$$(ii) \quad P(W|\underline{\theta}) \geq P(W^*|\underline{\theta}) \quad , \quad \underline{\theta} \in \Omega - \omega .$$

It is a well known fact that for hypothesis about the value of two or more parameters (such as the above with $r \geq 2$) there never exists a UMP unbiassed critical region in cases of practical interest. Nevertheless the L.R. test of the general linear hypothesis has certain desirable properties which we shall now discuss.

Two possibilities are open:

1. We might be able to limit to a class $\mathcal{C}$ of critical regions smaller than the class of all unbiassed regions, in the hope of finding a UMP one in $\mathcal{C}$.

2. We would demand less than maximum power at every point of $\Omega - \omega$ (As UMP-ness requires), and settle for a critical region having some optimum

properties in the large, as for instance maximum average power for each of certain categories of alternative. Let us now consider the L.R. test of general linear hypothesis in the light of the above discussion.

The critical region of the test of H, is then given by

$$W_0 = \left\{ \underline{z} \;\middle|\; \frac{(\underline{z}_r - \underline{\mu}_r^*)'(\underline{z}_r - \underline{\mu}_r^*)/r}{\underline{z}'_{n-p}\,\underline{z}_{n-p}/n-p} > F_{\alpha;n,n-p} \right\} \qquad (3.2)$$

being the size of the test.

We know that the power $P(\underline{\theta}|W)$ of this test depends on $\underline{\theta}$ only through the intermediary

$$\delta' = \frac{(\underline{\mu}_r - \underline{\mu}_r^*)'(\underline{\mu}_r - \underline{\mu}_r^*)}{\sigma^2} .$$

In other words, $P(\underline{\theta}|W)$ is constant on the surface in the parameter space satisfying $(\underline{\mu}_r - \underline{\mu}_r^*)'(\underline{\mu}_r - \underline{\mu}_r^*) = \sigma^2\delta'$ where δ' is any non-negative constant.

This means that the power is constant on the surface

$$D_{\delta'} = \left\{ \underline{\theta} \;\middle|\; \underline{\theta} \in \Omega \;,\; (\underline{\mu}_r - \underline{\mu}_r^*)'(\underline{\mu}_r - \underline{\mu}_r^*) = \sigma^2\delta' \right\} .$$

The surface $D_{\delta'}$ is a cylinder in the $(p + 1)$-dimensional parameter space, whose base is a spherical cone of one nappe ($\sigma > 0$) in the ($r+1$) -dimensional space of $\mu_1, \mu_2, \cdots \mu_r, \sigma$.

The implication is that the critical region is a similar region, since under ω, $\underline{\theta}$ is restricted to the surface D_0.

Hsu has shown that

"Among all critical regions W of size α with property that the power depends on $\underline{\theta}$ only through the intermediary of

$$\delta' = \frac{(\underline{\mu}_r - \underline{\mu}_r^*)'(\underline{\mu}_r - \underline{\mu}_r^*)}{\sigma^2}$$

W_0 is U.M.P.

In order to introduce the concept of average power, we shall define in the parameter space the $(r-1)$-dimensional sphere (surfaces of r-dimensional sphere),

$$S(\mu'_{r+1}, \cdots, \mu'_p, \sigma', \delta') = \left\{ \underline{\theta} \;\middle|\; \sum_{i=1}^{r} (\mu_i - \mu_i^*)^2 = \sigma^2 \delta'^2 ; \mu_i = \mu'_i, \; i = r+1, \cdots p ; \; \sigma = \sigma' \right\}$$

where the constants $\mu'_{r+1}, \cdots, \mu'_p, \sigma', \delta'$ are chosen at liberty with only the restriction that $\sigma' > 0$, $\delta' \geq 0$.

If $\beta_\phi(\mu_1, \cdots \mu_r, \sigma')$ denotes the power of a test ϕ of H, then the average power is defined as

$$\int_S \beta_\phi(\mu_1, \cdots, \mu_r, \sigma')\, dS \Big/ \int_S dS$$

in which dS denotes the differential area of the surface of the sphere.

Wald's Theorem:- AOV test maximizes the average power

$$\int_S \beta_\phi(\mu_1, \cdots, \mu_r, \sigma')\, dS \Big/ \int_S dS$$

for every $\mu'_{r+1}, \cdots \mu'_p, \sigma'$ among all unbiassed (or similar) tests.

If $U = \sum_{i=1}^{r} (z_i - \mu_i)^2$ and $V = \sum_{i=p+1}^{n} z_i^2$, unbiassedness (or similarity) implies that the conditional probability of rejection given $z_{r+1}, \cdots, z_p$ and U + V equals α almost everywhere.

Hence for any given $\mu'_{r+1}, \cdots \mu'_p, \sigma'$ and δ' the average power is maximized by rejecting when the ratio of the average density to the density under H is larger than a suitable constant

$C(z_{r+1}, \cdots z_p, u+v)$ and hence when

$$g(z_1, z_2; \cdots z_r; \mu_1, \cdots \mu_r) = \int_S \exp \frac{1}{\sigma'^2}(\underline{z}_r - \underline{\mu}_r^*)'(\underline{z}_r - \underline{\mu}_r^*)\, dS$$

$$> C(z_{r+1}, \cdots z_p, u+v),$$

for average density / density under H

$$= \frac{\left(\frac{1}{\sqrt{2\pi}\sigma'}\right)^n \int_S \exp -\frac{1}{2\sigma'^2}\left[\sum_{i=1}^{p}(z_i - \mu_i^*)^2 + \sum_{p+1}^{n} z_i^2\right] dS}{\left(\frac{1}{\sqrt{2\pi}\sigma'}\right)^n \exp -\frac{1}{2\sigma'^2}\left[\sum_{1}^{r}(z_i - \mu_i^*)^2 + \sum_{r+1}^{p}(z_i - \mu_i)^2 + \sum_{p+1}^{n} z_i^2\right]}$$

$$= \int_S \exp -\frac{1}{2\sigma'^2}\sum_{i=1}^{r}\left[-2(z_i - \mu_i^*)(\mu_i - \mu_i^*) + (\mu_i - \mu_i^*)^2\right] dS$$

$$= \int_S \exp \cdot \left[\frac{1}{2\sigma'^2}\sum_{i=1}^{r}(z_i - \mu_i^*)(\mu_i - \mu_i^*) + \delta'\right] dS$$

$$= \text{constant} \int_S \exp \cdot \frac{1}{\sigma'^2}(\underline{z}_r - \underline{\mu}_r^*)'(\underline{z}_r - \underline{\mu}_r^*)\, dS$$

As will be indicated below, the function g depends on $\underline{z}$ only through u and is an increasing function of u.

Since under the hypothesis $\frac{U}{U+V}$ is independent of $z_{r+1}, \cdots, z_p$ and U + V it follows that the test is given by AOV test.

In the definition of g, $\frac{1}{\sigma'^2}(\underline{z}_r - \underline{\mu}_r)'(\underline{z}_r - \underline{\mu}_r)$ may be written as $(\sqrt{u\delta'}\cos\alpha)/\sigma$, where α is the angle ($0 \leq \alpha \leq \pi$) between ($\underline{z}_r - \underline{\mu}_r^*$) and ($\underline{\mu}_r - \underline{\mu}_r^*$). Because of the symmetry of the sphere this is unchanged if α is replaced by the angle γ between ($\underline{\mu}_r - \underline{\mu}_r^*$) and an arbitrary fixed vector. i.e. g depends on the z's only through u. For fixed $\underline{\mu}_r, \sigma$ we shall denote g by h (u).

Consider S', the subset of S in which $0 \leq \gamma \leq \frac{\pi}{2}$.

Then $h(u) = \int_{S'} \left[\exp.(\sqrt{u}s' \cos\gamma)/\sigma + \exp.(-\sqrt{u}s' \cos\gamma)/\sigma \right] dS$

showing that it is an increasing function of u.

Hsu's Theorem is a corollary to Wald's Theorem: If W is any critical region of size α, the power which depends on θ only through δ, then for any $\delta = \delta^*$, W and W_o each has constant power, say c and c_o on D_{δ^*} and hence on each S in D_{δ^*}. Because D_{δ^*}, can be expressed as the union of spheres S for $-\infty < \mu_i^* < +\infty$, $i = r+1, \cdots, p, \sigma^* > 0$. By Wald's theorem $c \leq c_o$ which proves Hsu's theorem.

Invariance property of F-test.

Consider a random variable $\underline{Z}^{n\times 1}$ which takes on values $\underline{z}^{n\times 1}$ in a sample space $\mathcal{Z}$ (n-dimensional Euclidean). We shall think now of Ω as a family of distributions P of $\underline{Z}$ (instead of considering Ω as a set in the parameter space), and assume that no two P in Ω are the same. Similarly ω reprsents a subfamily of distributions. Let g be any 1 to 1 Borel-measurable transformation of $\mathcal{Z}$ on to $\mathcal{Z}$. The random variable $\underline{Z}' = g\underline{Z}$ has a distribution P' and we write

$$P' = \bar{g} P \text{ , where } \bar{g} \text{ depends on } g .$$

In general, $\bar{g}P$ will not belong to Ω. We shall only consider transformations g for which the set of transformed distributions $\{\bar{g}P \mid P \in \Omega\}$ is the same as the oroginal set Ω ; we may write this $\bar{g}\Omega = \Omega$. Supposing G as a group of such transformations, we say that

(1) the problem (i.e. Ω, ω) is invariant under the group G, if, for all $g \in G$.

(i) $\bar{g}(\omega) = \omega$

(ii) $\bar{g}(\Omega - \omega) = \Omega - \omega$

(2) The critical region W is invariant under G if, for all $g \in G$

(iii) $g(W) = W$

(iv) $g(\bar{z} - W) = \bar{z} - W$

Then the following four groups of transformations obviously leave the canonical form of the General Linear Hypothesis model invariant.

(i) $z_i' = c z_i \qquad (i = 1, 2, \ldots n, c > 0)$

(ii) $z_i' = z_i + b_i \qquad (i = r+1, \ldots, p; -\infty < b_i < +\infty)$

(iii) Orthoginal transformation on $\{z_{p+1}, \ldots, z_n\}$

(iv) Orthoginal transformation on $\{z_1, \ldots, z_r\}$

Let G* be the smallest group containing these four groups as sub groups. The following two theorems (1 and 2) are due to Hunt and Stein.

Theorem 1. Of all critical regions W of size which are invariant under G*, W_0 defined by (3.) is UMP.

Theorem 2. W, is a most stringent test of size α .

i.e. it minimizes $\sup_{\underline{\theta} \in \Omega - \omega} \left[\tilde{\beta}_\alpha(\underline{\theta}) - \beta(W, \underline{\theta}) \right]$

where $\tilde{\beta}_\alpha(\underline{\theta})$ is defined as the maximum power obtainable by any critical region of size α against a particular alternative $\underline{\theta}$,

i.e. $\tilde{\beta}_\alpha(\underline{\theta}) = \sup_{W \text{ of size } \alpha} \beta(\underline{\theta}, W)$

Interpreation of Theorem 2: Given a particular alternative $\underline{\theta}$, a measure of how good a critical region

of size α is against that alternative is how close the power reaches to $\tilde{\beta}_{\alpha}(\underline{\theta})$ and thus the smaller $\tilde{\beta}_{\alpha} - \beta(W, \underline{\theta})$, the better the test is against $\underline{\theta}$. Accordingly, as a measure of how good a test is for the alternative, the quantity

$$\sup_{\underline{\theta} \in \Omega - \omega} \left[\tilde{\beta}_{\alpha}(\underline{\theta}) - \beta(W, \underline{\theta}) \right]$$ is a satisforty one.

Two more properties of AOV test.

(i) $F_{d,\alpha}$ is of type D in the class of tests of size α

(to be proved in Chapter VIII)

(ii) Among tests of size α, $F_{d,\alpha}$ is minimax for a variety of weight functions.

Chapter IV

Some useful results in Analysis of one-way, two-way and three-way layouts.

4.1 As we have discussed in Chapter (II) on the model of less than full rank we are interested in the estimable hypothesis. In practical applications, we often come across the hypothesis of the form.

$$H_o : \beta_1 = \beta_2 = \cdots = \beta_r = 0$$

where $\beta_1, \beta_2, \cdots, \beta_r$ represent a subset of the elements of in the model

$$\underline{y} = \underline{X}\underline{\beta} + \underline{e}$$

The following theorem is often found useful.

Theorem: 4.1 If $\beta_1, \cdots \beta_r$ represent a subset of the elements of of the model

$$\underline{y} = \underline{X}\underline{\beta} + \underline{e}$$

such that $\sum_{i=1}^{r} c_i \beta_i$ is estimable for every set of constants c_i where $\sum_{i=1}^{r} c_i = 0$ then the hypothesis

$$H_o: \beta_1 = \beta_2 = \cdots = \beta_r$$

is an estimable hypothesis.

Proof: Since $\sum c_i \beta_i$ is estimable for all sets of c_i such that $\sum c_i = 0$, the following in particular are estimable:

$$\underline{\lambda}_1' \underline{\beta} = \beta_1 - \beta_2$$

$$\underline{\lambda}_2' \underline{\beta} = \beta_1 + \beta_2 - 2\beta_3$$

$$\cdots\cdots\cdots$$

$$\underline{\lambda}_{r-1}' \underline{\beta} = \beta_1 + \beta_2 + \cdots + \beta_{r-1} - (r-1)\beta_r .$$

Obviously $\underline{\lambda}_1'\underline{\beta}$, $\cdots$ $\underline{\lambda}_{r-1}'\underline{\beta}$ constitute a set of linearly independent estimable functions that are all zero if and only if $\beta_1 = \beta_2 = \cdots = \beta_r$.

4.2 Computation of Normal Equations.

In most of the models, used in experimental investigation, the design matrix $\underline{X}$ consists entirely of 0's and 1's and a method is therefore desirable that may save some computational labour.

The normal equations are

$$\underline{X}'\underline{X}\,\underline{\beta} = \underline{X}'\underline{y} \quad , \quad \text{a system of p equations in p parameters.}$$

The procedure is to find the right-hand side of the normal equations and take the expected value to get the left-hand side.

i.e. $E(\underline{X}'\underline{y}) = E(\underline{X}'(\underline{X}\,\underline{\beta} + e)) = \underline{X}'\underline{X}\,\underline{\beta}$.

Consider $\underline{X} = (\underline{\xi}_1, \underline{\xi}_2, \cdots \underline{\xi}_p)$, $\underline{\xi}_i$ being the ith column of $\underline{X}$

The model can then be written in the form

$$\underline{y} = \sum \underline{\xi}_i \beta_i + \underline{e}.$$

On examining $\underline{\xi}_i \beta_i$, in which $\underline{\xi}_i$ is a vector of only o's and 1's, it is found that the parameter β_i occurs in the nth observation of the model if and only if the nth element of $\underline{\xi}_i$ is equal to one.

But $\underline{\xi}_i'\underline{y}$ represents the sum of the elements of $\underline{y}$ for those elements of $\underline{\xi}_i$ that are equal to 1. Therefore $\underline{\xi}_i'\underline{y}$ can be found by summing the elements of $\underline{y}$ over those elements in the model that contain β_i.

Thus the normal equations can be obtained by finding $\underline{\xi}_i'\underline{y}$ for each β_i, taking the expectation of each $\underline{\xi}_i'\underline{y}$ to obtain $\underline{\xi}_i'\underline{X}\,\underline{\beta}$, then putting the sign $\wedge$ over each parameter. Thus we get

$$\beta_i : \qquad \underline{\xi}_i'\underline{y} = \underline{\xi}_i'\underline{X}\,\hat{\underline{\beta}} \qquad i = 1, 2, \cdots, p.$$

Now we shall discuss the application of the results we have obtained so far on the analysis of one-way, two-way and higher-way layouts.

4.3 The One-way Layout.

The one-way layout (or one-way classification) refers to the comparison of the m4ans of several (univariate) population which we shall denote by $\beta_1, \cdots \beta_b$. We shall assume that we have populations normally distributed with equal variances σ^2 and that we have independent random samples of sizes $n_1, n_2, \cdots n_b$ from the respective populations.

Then the model associated with our underlying assumptions will be

$$\Omega: \quad y_{ij} = \mu + \beta_i + e_{ij} \quad (i = 1, 2, \cdots b; \; j = 1, 2, \cdots, n_i), \quad \sum n_i = n$$

$\{e_{ij}\}$ are independently $N(0, \sigma^2)$.

4.3.1 Point estimation of $\{\mu, \beta_1, \cdots \beta_b\}$.

Normal equations take the form

$$\begin{aligned}
\mu: &\quad n\hat{\mu} + n_1\hat{\beta}_1 + n_2\hat{\beta}_2 + \cdots + n_b\hat{\beta}_b = y_{..} \\
\beta_1: &\quad n_1\hat{\mu} + n_1\hat{\beta}_1 \qquad\qquad\qquad\qquad\quad = y_{1.} \\
&\vdots \qquad\qquad\qquad\qquad\qquad\qquad\quad \vdots \\
\beta_b: &\quad n_b\hat{\mu} + \cdots + n_b\hat{\beta}_b = y_{b.}
\end{aligned}$$

or, more compactly,

$$\begin{aligned}
\mu: &\quad n\hat{\mu} + \sum_{i=1}^{b} n_i\hat{\beta}_i = y_{..} \\
\beta_i: &\quad n_i\hat{\mu} + n_i\hat{\beta}_i = y_{i.} \qquad i = 1, 2, \cdots, b.
\end{aligned}$$

(Replacement of a dot for a subscript letter indicates a total over the possible values of the subscript and a bar over the letter means the corresponding means).

The sum of the last b equations being equal to the first, there is at least one linear dependence. It can be easily seen that the last b equations are linearly independent. Thus

$$\text{rank} \quad (\underline{X}' \underline{X}) = b$$

and consequently there are b linearly independent estimable functions.

Also we know that $E(y_{\cdot j})$ and any linear function of $E(y_{ij})$ are estimable. Also it is known that the best linear unbiased estimate of $\mu + \beta_i$ is given by

$$\widehat{E(y_{ij})} = \widehat{\mu + \beta_i} = \frac{y_{i\cdot}}{n_i} = \bar{y}_{i\cdot}$$

If $\sum c_i = 0$, then $\sum c_i \beta_i$ is estimable and the estimate is $\sum c_i \bar{y}_{i\cdot}$.

Suppose that $\sum c_i \beta_i$ is an estimable function. Then, some linear combination of the left-hand side of the normal equations must yield $\sum c_i \hat{\beta}_i$. This linear combination is obtained by multiplying the $(i+1)$th equation by $\frac{c_i}{n_i}$ and adding.

From the resulting equation

$$\mu \sum c_i + \sum c_i \beta_i = \sum \frac{c_i}{n_i} \bar{y}_{i\cdot} \quad ,$$

we see that if $\sum c_i = 0$ then $\sum c_i \beta_i$ is an estimable function.

From Theorem 4.1 it follows that the hypothesis

$$H_0: \quad \beta_1 = \beta_2 = \cdots = \beta_b$$

is an estimable hypothesis, since $\sum c_i \beta_i$ is estimable for all $\sum c_i = 0$.

4.4 Two-way layout with unequal numbers in subclasses and no intraction.

Suppose we are interested in the b levels of the factor B and t levels of the factor T on the outcome of an experiment. In such a

two-factor experiment, the observations can be arranged in a two-way layout or $b \times t$ table by letting the rows of the table correspond to the levels of factor B and the columns to the levels of T.

Let the number of observations in the (i, j) th cell be n_{ij} where not all $n_{ij} = 0$.

If y_{ijk} denotes the kth observation in the (i,j)th cell, then the model (with no interaction) associated with the design is given by

$$\Omega: \quad y_{ijk} = \mu + \beta_i + \tau_j + e_{ijk}$$

$\{e_{ijk}\}$ are independently $N(0, \sigma^2)$

$i = 1, 2, \ldots, b$

$j = 1, 2, \ldots, t$

$k = 0, 1, 2, \ldots, n_{ij}$

where k=o corresponds to no observation.

The normal equations of the model are

$$\mu: \quad n_{..}\hat{\mu} + \sum n_{i.}\hat{\beta}_i + \sum n_{.j}\hat{\tau}_j = y_{...}$$

$$\beta_i: \quad n_{i.}\hat{\mu} + n_{i.}\hat{\beta}_i + \sum_j n_{ij}\hat{\tau}_j = y_{i..}, \quad i = 1, 2, \ldots b.$$

$$\tau_j: \quad n_{.j}\hat{\mu} + \sum_i n_{ij}\hat{\beta}_i + n_{.j}\hat{\tau}_j = y_{.j.}, \quad j = 1, 2, \ldots t$$

Theorem: 4.2 Under the assumption that the n_{ij} in the model are such that $\beta_i - \beta_{i'}$, $\tau_j - \tau_{j'}$ are estimable for all $i \neq i'$ and $j \neq j'$, then

(i) there are exactly b + t − 1 linearly independent estimable functions.

(ii) $\sum c_i \beta_i$ and $\sum d_j \tau_j$ are estimable if $\sum c_i = \sum d_j = 0$.

Proof: (i) b + t − 1 estimable functions

$$\beta_1 - \beta_2, \ \beta_1 - \beta_2, \ \ldots, \ \beta_1 - \beta_b, \ \tau_1 - \tau_2, \ \ldots, \ \tau_1 - \tau_t,$$

and $n_{..}\hat{\mu} + \sum n_{i.}\hat{\beta}_i + \sum n_{.j}\hat{\tau}_j$ are obviously linearly independent.

Of the b + t + 1 equations in the normal equation, the first equation represented by μ equal to the sum of the b equations represented by β_i, and also equal to the sum of the t equations represented by τ_j.

Hence, there are at most b + t - 1 linearly independent estimable functions.

This, together with the fact that there are at least b + t - 1 linearly independent estimable functions ensure that there are excactly b + t - 1 linearly independent estimable functions.

(ii) Since every $\beta_i - \beta_{i'}$ is estimable, every linear combination of them is estimable.

Consider $\frac{1}{b}\sum_{i=1}^{b}(\beta_i - \beta_{i'}) \qquad i \neq i'$

$= \beta_i - \bar{\beta}_.$ where $\bar{\beta}_. = \frac{1}{b}\sum\beta_i$

This implies that $\beta_i - \bar{\beta}_.$ is estimable for all i; so is $\sum c_i(\beta_i - \bar{\beta}_.)$.

If $\sum c_i = 0$, this expression reduces to $\sum c_i\beta_i$.

The same argument applies to $\sum d_j\tau_j$ where $\sum d_j = 0$.

4.4.1 Solution of the normal equations for contrasts of the β_i's.

From the τ_j equations we obtain

$$\hat{\mu} + \hat{\tau}_j = \bar{y}_{.j.} - \frac{1}{n_{.j}}\sum_s n_{sj}\hat{\beta}_s \qquad (j = 1, 2, ..t)$$

From the β_i equations we obtain

$$\sum_j n_{ij}(\hat{\mu} + \hat{\tau}_j) + n_{i.}\hat{\beta}_i = y_{i..} \qquad (i = 1, 2, ..b)$$

From these two equations, we have

$$\sum_j n_{ij} \left(\bar{y}_{.j} - \frac{1}{n_{.j}} \sum n_{sj} \hat{\beta}_s \right) + n_{i.} \hat{\beta}_i = y_{i..} \quad (i = 1, 2, .. b)$$

On simplification, we get

$$n_{i.} \hat{\beta}_i - \sum_{j=1}^{t} \sum_{s=1}^{b} \frac{n_{ij} n_{sj}}{n_{.j}} \hat{\beta}_s = y_{i..} - \sum_{j=1}^{t} n_{ij} \bar{y}_{.j}.$$

$$(i = 1, 2, .., b).$$

$$\text{i.e.} \quad n_{i.} \hat{\beta}_i - \hat{\beta}_i \sum_{j=1}^{t} \frac{n_{ij}^2}{n_{.j}} - \sum_{j} \sum_{i \neq s}^{s} \frac{n_{sj} n_{ij}}{n_{.j}} \hat{\beta}_s = q_i$$

$$(i = 1, 2, .., b).$$

where $q_i = y_{i..} - \sum n_{ij} \bar{y}_{.j}.$

$$\text{i.e.} \quad \left(n_{i.} - \sum_{j=1}^{t} \frac{n_{ij}^2}{n_{.j}} \right) \hat{\beta}_i - \sum_{i \neq s} \sum_{s} \frac{n_{sj} n_{ij}}{n_{.j}} \hat{\beta}_s = q_i$$

This system of equations representing b equations in b unknowns may be written in matrix form as

$$\underline{C} \hat{\beta} = q$$

where b x b matrix $\underline{C}$ has the elements

$$c_{ii} = n_{i.} - \sum_{j=1}^{t} \frac{n_{ij}^2}{n_{.j}} \qquad (i = 1, 2, .., b)$$

$$c_{is} = -\sum_{s}^{b} \sum_{j=1}^{t} \frac{n_{ij} n_{sj}}{n_{.j}} \qquad (i \neq s, \; i = 1, 2, .., b)$$

Theorem: 4.3 rank $(\underline{C}) = b - 1$

Proof: By theorem 4.2 and the assumption of the theorem concerning the estimability of $\beta_i - \beta_{i'}$, there are $(b - 1)$ linearly independent estimable functions of the β_i

These functions must come from

$$\underline{C} \hat{\beta} = q$$

so that rank $(\underline{C})$ is at least $b - 1$.

These equations are not independent, it being easily verified that the sum of left-hand and of right-hand side are both

identically zero.

Therefore, $\underline{C}$ has one linearly dependent row, and its rank must be at most $b - 1$.

Thus rank $(\underline{C}) = b - 1$.

Now in order to obtain a unique solution of the equation

$$\underline{C}\ \hat{\underline{\beta}} = \underline{q}$$

we may impose any nonestimable condition, the simplest one (generally) being $\underline{1}'\hat{\underline{\beta}} = 0$, where $\underline{1}$ is a vector with each element equal to unity.

These two equations together may be written in the form

$$\begin{vmatrix} \underline{C} & \underline{1} \\ \underline{1}' & 0 \end{vmatrix} \begin{vmatrix} \hat{\underline{\beta}} \\ 0 \end{vmatrix} = \begin{vmatrix} \underline{q} \\ 0 \end{vmatrix} \qquad (4.3)$$

Theorem: 4.4 Under the assumptions of theorem

$$\text{rank } (\underline{C}^*) = b + 1$$

$$\text{where } \underline{C}^* = \begin{vmatrix} \underline{C} & \underline{1} \\ \underline{1}' & 0 \end{vmatrix}$$

Proof: We will show that the unique solution exist for

$$\begin{aligned} \underline{C}\ \hat{\underline{\beta}} &= \underline{q} \\ \underline{1}'\ \hat{\underline{\beta}} &= 0 \end{aligned} \qquad (4.4)$$

Consider the equations $\underline{C}_1\ \hat{\underline{\beta}} = \underline{q}_1$ $\qquad (4.4)^*$

$$\underline{1}'\ \hat{\underline{\beta}} = 0$$

where $\underline{C}_1\ \hat{\underline{\beta}} = \underline{q}_1$, represent any $b - 1$ equations of $\underline{C}\ \hat{\underline{\beta}} = \underline{q}$

The determinent of the coefficient, $\begin{vmatrix} \underline{C}_1 \\ \underline{1}' \end{vmatrix}$

is found to be nonsingular by adding all the columns to the first column and using the fact that any $(b-1) \times (b-1)$ submatrix of $\underline{C}$ is nonsingular.

i.e. The unique solution exists for (4·4)*

Therefore the unique solution exists for (4·4)

or $\quad \text{rank } \underline{C}^* = b + 1$

The solution of (4·3) may now be written

$$\begin{pmatrix} \hat{\underline{\beta}} \\ 0 \end{pmatrix} = \underline{C}^{*-1} \begin{pmatrix} \underline{q}_Y \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} \underline{B}_{11}^{b\times b} & \underline{B}_{12}^{b\times 1} \\ \underline{B}_{21} & \underline{B}_{22} \end{pmatrix} \begin{pmatrix} \underline{q}_Y \\ 0 \end{pmatrix}$$

or $\qquad \hat{\underline{\beta}} = \underline{B}_{11}\, \underline{q}_Y$

As a result of the imposition of the chosen condition, any $\hat{\beta}_i$ is the estimate in fact of ($\beta_i - \bar{\beta}_.$). This procedure is useful in that it produces as a by-product the variances and covariances of the $\hat{\underline{\beta}}$, for $\text{cov}(\hat{\underline{\beta}}) = \underline{B}_{11}\sigma^2$.

To understand these facts more in detail, we shall prove the following theorem.

Theorem: 4.5 If $\underline{C}^{*-1} = \begin{pmatrix} \underline{B}_{11} & \underline{B}_{12} \\ \underline{B}_{21} & \underline{B}_{22} \end{pmatrix}$

then (i) $\underline{B}_{21}$ (and $\underline{B}_{12}$) has all its elements equal to $\frac{1}{b}$.

(ii) $\underline{B}_{22} = 0$

(iii) $\underline{B}_{11}\ \underline{C}\ \underline{B}_{11} = \underline{B}_{11}$

(iv) $\underline{C}\ \underline{B}_{11}$ is idempotent of rank $b-1$, with diagonal elements each equal to $(b-1)/b$

and with off-diagonal elements each equal to $-\frac{1}{b}$.

(v) The rows of $\underline{B}_{11}$ add to zero.

Proof: From the relation $\underline{C}^* \underline{\bar{C}}^* = \underline{I}$

we get
$$\begin{pmatrix} \underline{C} & \underline{1} \\ \underline{1}' & 0 \end{pmatrix} \begin{pmatrix} \underline{B}_{11} & \underline{B}_{12} \\ \underline{B}_{21} & \underline{B}_{22} \end{pmatrix} = \begin{pmatrix} \underline{I} & \underline{0} \\ \underline{0} & 1 \end{pmatrix}$$

or
$$\underline{C}\ \underline{B}_{11} + \underline{1}\ \underline{B}_{21} = \underline{I} \qquad (a)$$
$$\underline{1}'\ \underline{B}_{11} = \underline{0} \qquad (b)$$
$$\underline{1}'\ \underline{B}_{12} = 1 \qquad (c)$$
$$\underline{C}\ \underline{B}_{12} + \underline{1}\ \underline{B}_{22} = \underline{0} \qquad (d)$$

It is also known that
$$\underline{1}'\ \underline{C} = \underline{0} \qquad (e)$$
$$\underline{1}'\ \underline{1} = b \qquad (f)$$

(1) Multiplying (a) by $\underline{1}'$, we have
$$\underline{1}'\ \underline{C}\ \underline{B}_{11} + \underline{1}'\underline{1}\ \underline{B}_{21} = \underline{1}'$$

Using (e) and (f), this gives
$$b\ \underline{B}_{21} = \underline{1}'$$
or
$$\underline{B}_{21} = \left(\frac{1}{b}\right)\ \underline{1}'$$

(2) Multiplying (d) by $\underline{1}'$ and using d) and (f) will give the required result.

(3) Multiply (a) by $\underline{B}_{11}$ to get
$$\underline{B}_{11}\ \underline{C}\ \underline{B}_{11} + \underline{B}_{11}\ \underline{1}\ \underline{B}_{21} = \underline{B}_{11}$$
in which $\underline{B}_{11}\ \underline{1} = \underline{0}$ by (b).
Thus $\underline{B}_{11}\ \underline{C}\ \underline{B}_{11} = \underline{B}_{11}$.

(4) From (a) we get $\underline{C}\,\underline{B}_{11} = \underline{I} - \underline{1}\,\underline{B}_{21}$

$= \underline{I} - \left(\frac{1}{b}\right)\,\underline{1}\,\underline{1}'$

Thus the diagonal elements are all equal to $\frac{b-1}{b}$, and

the off-diagonal elements are equal to $-\frac{1}{b}$.

(5) From (b) it follows that the rows of $\underline{B}$ add to zero.

4.5 To find the values of $E(\hat{\underline{\beta}})$ and $\mathrm{Cov}(\hat{\underline{\beta}})$.

We shall first consider $E(\hat{\underline{\beta}}) = E(\underline{B}_{11}\,\underline{q}_1)$

$= \underline{B}_{11}\,E(\underline{q}_1)$

Since $\underline{C}\,\hat{\underline{\beta}} = \underline{q}_1$ is obtained by linear combinations of the normal equations

$$E(\underline{q}_1) = E\,\underline{C}\,\hat{\underline{\beta}} = \underline{C}\,E(\hat{\underline{\beta}}) = \underline{C}\,\underline{\beta}$$

$$E(\hat{\underline{\beta}}) = \underline{B}_{11}\,\underline{C}\,\underline{\beta}$$

$$= \left(\underline{I} - \frac{1}{b}\,\underline{1}\,\underline{1}'\right)\underline{\beta}$$

$$= \underline{\beta} - \frac{1}{b}\,\underline{1}'\,\underline{\beta}.$$

i.e. ith element of $E(\hat{\underline{\beta}}) = E(\hat{\beta}_i) = \beta_i - \frac{1}{b}\Sigma\beta_i$

$= \beta_i - \bar{\beta}.$

We shall now consider $\mathrm{Cov}(\hat{\underline{\beta}}) = \mathrm{Cov}(\underline{B}_{11}\,\underline{q}_1)$

$= E\left[\hat{\underline{\beta}} - E(\hat{\underline{\beta}})\right]\left[\hat{\underline{\beta}} - E(\hat{\underline{\beta}})\right]'$

$= E\left[\underline{B}_{11}\,\underline{q}_1 - \underline{B}_{11}\,\underline{c}\,\underline{\beta}\right]\left[\underline{B}_{11}\,\underline{q}_1 - \underline{B}_{11}\,\underline{c}\,\underline{\beta}\right]'$

$= \underline{B}_{11}\,E\left[\underline{q}_1 - E(\underline{q}_1)\right]\left[\underline{q}_1 - E(\underline{q}_1)\right]\underline{B}_{11}$

Let $\underline{A} = \mathrm{Cov}(\underline{q}_1) = E\left[\underline{q}_1 - E(\underline{q}_1)\right]\left[\underline{q}_1 - E(\underline{q}_1)\right]$; then

$a_{ii} = \mathrm{var}(q_i)$

$$q_{ii} = \text{var}\left(y_{i..} - \sum n_{ij}\, \bar{y}_{.j.} \right)$$

$$= \text{var}\left(y_{i..} \right) + \text{var} \sum_{j=1}^{t} n_{ij}\, \bar{y}_{.j.} - 2\,\text{Cov}\left(\sum_{j=1}^{t} n_{ij}\, \bar{y}_{.j.}\, , y_{i..} \right)$$

$$= n_{i.}\sigma^2 + \sigma^2 \sum_{j=1}^{t} \frac{n_{ij}^2}{n_{.j}} - 2\sigma^2 \sum_{j=1}^{t} \frac{n_{ij}^2}{n_{.j}}$$

$$= \left(n_{i.} - \sum_{j=1}^{t} \frac{n_{ij}^2}{n_{.j}} \right) \sigma^2$$

Now, for the off diagonal element of $\underline{A}$,

$$q_{rs} = \text{Cov}\left(q_r, q_s \right) \qquad (r \neq s)$$

$$= \text{Cov}\left[\left(y_{r..} - \sum_{j=1}^{t} n_{rj}\, \bar{y}_{.j.} \right), \left(y_{s..} - \sum_{k=1}^{b} n_{sk}\, \bar{y}_{.k.} \right) \right]$$

$$= \text{Cov}\left(y_{r..}, y_{s..} \right) - \text{Cov}\left(y_{s..}, \sum_{j=1}^{t} n_{rj}\, \bar{y}_{.j.} \right)$$

$$- \text{Cov}\left(y_{r..}, \sum_{k=1}^{t} n_{sk}\, \bar{y}_{.k.} \right) + \text{Cov}\left(\sum_{j=1}^{t} n_{rj}\, \bar{y}_{.j.}, \sum_{k=1}^{t} n_{sk}\, \bar{y}_{.k.} \right)$$

$$= 0 - \sigma^2 \sum_{j=1}^{t} \frac{n_{sj}\, n_{rj}}{n_{.j}} - \sigma^2 \sum_{j=1}^{t} \frac{n_{sj}\, n_{rj}}{n_{.j}} + \sigma^2 \sum_{j=1}^{t} \frac{n_{sj}\, n_{rj}}{n_{.j}}$$

$$= -\sigma^2 \sum_{j=1}^{b} \frac{n_{sj}\, n_{rj}}{n_{.j}}$$

Thus, we have seen that $\underline{A} = \underline{C}\,\sigma^2$

so, $\text{cov}\left(\hat{\underline{\beta}} \right) = \underline{B}_{11}\, \underline{C}\, \underline{B}_{11}\, \sigma^2$

$= \underline{B}_{11}\, \sigma^2$

4.6 Bose's Information matrix:-

We have seen now the role of the matrix $\underline{C}$, known as Bose's information matrix, in the reduced normal equation

$$\underline{C}\ \hat{\underline{\beta}} = \underline{q}$$

in the one-way heterogeniety setting or two-way classification analysis of variance. In block designs, we have t treatments to be planted in b blocks, each block containing the same number k, of plots,

one "planting" to be allowed per plot. Then the total number of observations will be bk. Let n_{ij} be the number of appearances of treatment i in block j.

Then the elements of $\underline{C}$ will be

$$c_{ii} = n_{i.} - \sum_{s=1}^{t} \frac{n_{is}^2}{k} \qquad i=1,2,\ldots b.$$

$$c_{ij} = - \sum_{s=1}^{t} \frac{n_{is} n_{js}}{k}$$

and those of q will be

$$q_i = T_i - \sum n_{ij} B_j / k$$

where T_i = sum of all y's corresponding to treatment i and

B_j = sum of all y's arising from block j.

In the setting of two-way hetrogeniety, we have t treatments and $k_1 \times k_2$ arrays of plots, and the model for the observation corresponding to treatment i in row j and column h is

$$y_{ijhl} = \mu + \beta_j^{(1)} + \beta_h^{(2)} + \tau_i + \varepsilon_{ijhl}$$

Let $n_{ij}^{(1)}$ and $n_{ij}^{(2)}$ be the number of times treatment i appears in row j and in column h respectively. Let T be as before and $B_j^{(1)}$ and $B_h^{(2)}$ be the sum corresponding to the jth row and hth column respectively. Then following the same algebraic procedure as in the case of two-way classification, we finally arrive at the reduced normal equation

$$\underline{C}\ \underline{\tau} = \underline{q}$$

where $\underline{C}$ has elements

$$c_{ii} = \sum_s n_{is}^{(1)} - \sum_s \frac{n_{is}^{(1)2}}{k_2} - \sum \frac{n_{is}^{(2)2}}{k_1} + \frac{(\sum n_{is}^{(1)})(\sum n_{is}^{(2)})}{k_1 k_2}$$

$$c_{ij} = - \frac{\sum n_{is}^{(1)} n_{js}^{(1)}}{k_2} - \frac{\sum n_{is}^{(2)} n_{js}^{(2)}}{k_1} + \frac{\sum n_{is}^{(1)} \sum n_{js}^{(2)}}{k_1 k_2}$$

and

$$q_{v_i} = T_i - \sum_j \frac{n_{ij}^{(1)} B_j^{(1)}}{k_2} - \sum_h \frac{n_{ih}^{(2)} B_h^{(2)}}{k_1} + \sum_s n_{is}^{(1)} \sum_s \frac{T_s}{k_1 k_2}$$

N.B. $\sum_s n_{is}^{(1)} = \sum_s n_{is}^{(2)}$ = the number of replications of treatment i.

Chapter V

Admissibility and Complete Class Theorems.

When a set of possible designs is very extensive, we would like to have means for ruling out experiments (or designs) which are inferior for any purpose. If a design d is better than another design d^* for a collection of problems Π then we say that

d is better (π) than d^* or d dominates (π) d^*. If d is such that no d^* dominates (π) d, then we say that d is admissible .

A class of designs Δ is complete (π) (or essentially complete (π)) if for any d^* not in Δ there is a d in Δ which is better (π) than (or as good (π) as) d^*. If no proper subset of Δ has this property, Δ is said to be minimal complete (π). Similar definition may be given for minimal essentially complete (π).

Definition:- A class of design, Δ for the linear hypothesis

$$\underline{y} = \underline{X}\underline{\beta} + \underline{e}$$

with $\underline{\beta} \in B \subseteq E^{(p)}$ and $\underline{x}_i \in E^{(p)}$,

when the experimenter is restricted to choosing the vectors $\underline{x}_i \in A \subseteq E$ may be defined as follows.

$$\Delta_n(A) = \left\{ d \;\middle|\; \underline{X}_d = \begin{pmatrix} \underline{x}_1' \\ \vdots \\ \underline{x}_n' \end{pmatrix} \text{ with } \underline{x}_i \in A \subseteq E^{(p)} \quad i=1, 2, \ldots n. \text{ and } p \leq n. \right.$$

We shall now suppose the problem space Λ to be a set $\Lambda \subseteq E$. A point $(\lambda_1, \lambda_2, \ldots \lambda_p) = \underline{\lambda} \in \Lambda$ may be interpreted as the problem of estimating $\underline{\lambda}'\underline{\beta}$.

In the design matrix $\underline{X}_d$, some of the $\underline{x}_i$'s might be the same.

If $\underline{x}_1, \underline{x}_2, \ldots \underline{x}_r$ are r different vectors, the experiment may be interpreted as the one in which y corresponding $\underline{x}_i$ is observed n_i times. The design matrix may also be described in the form $(n_1, \underline{x}_1, n_2, \underline{x}_2, \ldots \ldots n_r, \underline{x}_r)$.

Thus if M (d) represents the information matrix associated with the design d,

$$M(d) = \sum_i n_i M_i \quad (d)$$

where $$M_i(d) = \underline{x}_i \underline{x}_i'$$

Let the variance of the least square estimate (which is the same as the maximum likelihood estimate under normal theory) of $\psi (= \underline{\lambda}' \underline{\beta})$ when the design d is used, be denoted by $V_d [\underline{\lambda}' \hat{\underline{\beta}}]$.

We know that $\psi = \underline{\lambda}' \underline{\beta}$ is estimable with respect to d, only when

$$\underline{M}(d) \underline{f} = \underline{\lambda} \qquad (5.1)$$

has solutions for $\underline{f}$.

The variance of the estimate, $V_d [\underline{\lambda}' \hat{\underline{\beta}}] = \sigma^2 \underline{f}' \underline{M}(d) \underline{f}$, where $\underline{f}$ is any solution of (5.1)

When M(d) is of full rank, ψ is always estimable and

$$V_d [\underline{\lambda}' \hat{\underline{\beta}}] = \sigma^2 \underline{\lambda}' M \quad (d)$$

Definition:- $\Delta_n(A_0)$ where $A_0 \subseteq A$ is said to be essentially complete with respect to $\Delta_n (A)$ if and only if for any $d \in \Delta_n(A)$ and any unknown $\underline{\beta} \in B$ there exists $d^* \in \Delta_n (A)$. such that

$$V_{d^*} [\underline{\lambda}' \hat{\underline{\beta}}] \leq V_d [\underline{\lambda}' \hat{\underline{\beta}}] \text{ for all } \underline{\lambda} \in \Lambda ,$$

a point $\underline{\lambda} \in \Lambda \subseteq E^{(p)}$ being interpreted as the problem of estimating $\underline{\lambda}'\underline{\beta}$.

In the above definition we use $V_d [\underline{\lambda}'\hat{\underline{\beta}}]$ as the measure of error. Other measures of error viz: $E[(\hat{\underline{\beta}} - \underline{\beta})' \underline{L} (\hat{\underline{\beta}} - \underline{\beta})]$ ($\underline{L}$ being same positive definite) and $\sup_{\underline{\lambda} \in \Lambda} V_d (\underline{\lambda}'\hat{\underline{\beta}})$ may also be used.

The following lemma, which we shall prove in order to simplify later proofs and to take care of estimability considerations may be true for these measures of errors. (Elfving).

Lemma 5.1 If M^* and M are pxp non-negative definite (symmetric) matrices such that

(5.1) $$\underline{\lambda}' \underline{M}^* \underline{\lambda} \geq \underline{\lambda}' \underline{M} \underline{\lambda} \text{ for all } \underline{\lambda} \in E^{(p)},$$

then

(5.2) (a) if, for any given $\underline{\lambda}$, there exists $\underline{\ell}$ such that

$$\underline{M}\, \underline{\ell} = \underline{\lambda}$$

then there exists $\underline{\ell}^*$ such that

$$\underline{M}^* \underline{\ell}^* = \underline{\lambda} \text{ and}$$

(5.3) (b) $$\underline{\ell}^{*\prime} \underline{M}^* \underline{\ell}^* \leq \underline{\ell}' \underline{M} \underline{\ell} .$$

The statistical significance of the lemma is as follows:

If $M(d^*) \geq M(d)$, d^* is at least as good (Π_L) as d, Π_L being the class of all problems of point estimation of any linear parametric function

i.e. under the conditions of the lemma, any linear parametric function which is estimable under d is estimable under d^* and its best

linear estimate under d^* has a variance no greater than the variance of the best linear estimator under d.

Proof. (a) Let υ and υ^* be the vector spaces spanned by the column vectors of $\underline{M}$ and $\underline{M}^*$, respectively. Since $\underline{M}$ and $\underline{M}^*$ are symmetric matrices, the spaces spanned by the column vectors are the same as the spaces spanned by the row vectors.

Then part (a) of the lemma states that $\upsilon^* \geq \upsilon$.

Let us suppose, on the contrary, that there exist $\bar{\lambda} \in \upsilon$ and $\bar{\lambda}$ orthogonal to υ^*.

Then $\underline{M}^* \bar{\lambda} = 0$, and therefore $\bar{\lambda}' \underline{M}^* \bar{\lambda} = 0$.

From (5.1) consequently, it follows that

$$-\bar{\lambda}' \underline{M} \bar{\lambda} \geq 0$$

and thus $\bar{\lambda}' \underline{M} \bar{\lambda} = 0$.

$\bar{\lambda}$ cannot be zero nor orthogonal to M, since it is a linear combination of the columns of υ.

Therefore $\underline{M} \bar{\lambda} = 0$.

i.e. $\bar{\lambda}$ is orthogonal to M, which is a contradiction.

(b) Since $\underline{M}^*$ and $\underline{M}$ are non-negative definite, there exist a nonsingular matrix $\underline{D}$, such that

$$\underline{D}' \underline{M} \underline{D} = \begin{vmatrix} \underline{A} & \underline{B} \\ \underline{B} & \underline{C} \end{vmatrix}, \quad \underline{D}' \underline{M}^* \underline{D} = \begin{vmatrix} \underline{I} & \underline{0} \\ \underline{0} & \underline{0} \end{vmatrix}.$$

It can be seen that $\underline{A}$ is positive definite since (5.1) becomes

$$\underline{\lambda}_1' \underline{A} \underline{\lambda}_1 \geq \underline{\lambda}_1' \underline{\lambda}_1$$

where $\underline{\lambda}' = (\underline{\lambda}_1', \underline{\lambda}_2')$

Furthermore there exist orthogonal matrices Q and P such that

$$\underline{Q}' \underline{A} \underline{Q} = \underline{\varkappa}, \quad \underline{P}' \underline{C} \underline{P} = \begin{pmatrix} \underline{I} & \underline{0} \\ \underline{0} & \underline{0} \end{pmatrix}$$

so that

$$\begin{pmatrix} \underline{Q} & \underline{0} \\ \underline{0} & \underline{P} \end{pmatrix} \begin{pmatrix} \underline{A} & \underline{B} \\ \underline{B} & \underline{C} \end{pmatrix} \begin{pmatrix} \underline{Q}' & \underline{0} \\ \underline{0} & \underline{P}' \end{pmatrix} = \begin{pmatrix} \underline{\varkappa} & \underline{E} & \underline{G} \\ \underline{E} & \underline{I} & \underline{0} \\ \underline{G} & \underline{0} & \underline{0} \end{pmatrix}$$

We can easily verify that $\underline{G} = \underline{0}$, for any submatrix of the form

$$\begin{pmatrix} \varkappa_i & q_{ij} \\ q_{ij} & 0 \end{pmatrix}$$

semi
is positive definite and thus

$$\begin{vmatrix} \varkappa_i & q_{ij} \\ q_{ij} & 0 \end{vmatrix} \geq -q_{ij}^2 \geq 0 \quad \text{i.e. } q_{ij} = 0.$$

Thus without any loss of generality, we may restrict our discussion to the matrices $\underline{M}^*$ and $\underline{M}$ of the form

$$\underline{M}^* = \begin{pmatrix} \underline{\varkappa} & \underline{E} & \underline{0} \\ \underline{E}' & \underline{I} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} \end{pmatrix} \quad \text{and} \quad \underline{M} = \begin{pmatrix} \underline{I} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} \end{pmatrix}$$

where

$$\underline{\varkappa} = \begin{pmatrix} \varkappa_1 & & \underline{0} \\ & \ddots & \\ \underline{0} & & \varkappa_n \end{pmatrix}$$

We can see that

$$\begin{pmatrix} \underline{\varkappa} & \underline{E} \\ \underline{E}' & \underline{I} \end{pmatrix}$$

is positive definite, since from (5.1) we can get

$$(\underline{x}_1', \underline{x}_2') \begin{pmatrix} \underline{\Lambda} & \underline{E} \\ \underline{E}' & \underline{I} \end{pmatrix} \begin{pmatrix} \underline{x}_1 \\ \underline{x}_2 \end{pmatrix} \geq (\underline{x}_1' \underline{x}_2') \begin{pmatrix} \underline{I} & \underline{0} \\ \underline{0} & \underline{0} \end{pmatrix} \begin{pmatrix} \underline{x}_1 \\ \underline{x}_2 \end{pmatrix} = \underline{x}_1' \underline{x}_1$$

and $(\underline{0}', \underline{x}_2') \begin{pmatrix} \underline{\Lambda} & \underline{E} \\ \underline{E}' & \underline{I} \end{pmatrix} \begin{pmatrix} \underline{x}_1 \\ \underline{x}_2 \end{pmatrix}$.

From the above consideration we can see that, only $\underline{\lambda}_1$ is relevant and we may derive

$$\underline{f}^{*'} \underline{M}^{*} \underline{f}^{*} = \underline{\lambda}_1' \underline{A} \underline{\lambda}_1$$

where $\underline{A} = (\underline{\Lambda} - \underline{E}\underline{E}')^{-1}$.

Thus $\underline{f}^{*'} \underline{M}^{*} \underline{f}^{*} = \underline{\lambda}_1' (\underline{\Lambda} - \underline{E}\underline{E}')^{-1} \underline{\lambda}_1$.

Relation (5.2) now becomes

$$\underline{\lambda}_1' (\underline{\Lambda} - \underline{E}\underline{E}')^{-1} \underline{\lambda}_1 \leq \underline{\lambda}_1' \underline{\lambda}_1 \qquad \text{for all } \underline{\lambda}_1 .$$

The last relation is equivalent to

$$\underline{\lambda}_1' (\underline{\Lambda} - \underline{E}\underline{E}')^{-1} \underline{\lambda}_1 - \underline{\lambda}_1 \underline{\lambda}_1' \geq 0 \qquad \text{for all } \underline{\lambda}_1 .$$

i.e. $(\underline{\Lambda} - \underline{I} - \underline{E}'\underline{E})$ is non-negative definite.

From the assumption of the lemma

$$\begin{pmatrix} \underline{\Lambda} & \underline{E} & \underline{Q} \\ \underline{E}' & \underline{I} & \underline{Q} \\ \underline{Q} & \underline{Q} & \underline{Q} \end{pmatrix} - \begin{pmatrix} \underline{I} & \underline{Q} & \underline{Q} \\ \underline{Q} & \underline{Q} & \underline{Q} \\ \underline{Q} & \underline{Q} & \underline{Q} \end{pmatrix} = \begin{pmatrix} \underline{\Lambda} - \underline{I} & \underline{E} & \underline{Q} \\ \underline{E}' & \underline{I} & \underline{Q} \\ \underline{Q} & \underline{Q} & \underline{Q} \end{pmatrix}$$

is non-negative definite and further

$$\begin{pmatrix} \underline{\Lambda} - \underline{I} & \underline{E} \\ \underline{E}' & \underline{I} \end{pmatrix}$$ is positive definite.

(because $\underline{\Lambda}$ is diagonal and $\underline{\lambda}' \underline{\Lambda} \underline{\lambda} \geq \underline{\lambda}' \underline{\lambda}$)

Thus $\underline{R}_{11}$ is positive definite, where $\underline{R}_{11}$ is defined by

$$\begin{pmatrix} \underline{R}_{11} & \underline{R}_{12} \\ \underline{R}_{21} & \underline{R}_{22} \end{pmatrix} = \begin{pmatrix} \underline{\Lambda} - \underline{I} & \underline{E} \\ \underline{E}' & \underline{I} \end{pmatrix}^{-1}$$

$$\underline{R}_{11} = \underline{\Lambda} - \underline{I} - \underline{E}\underline{E}' .$$

The positive definiteness of $\underline{R}_{11}$ yields the desired result.

In order to proceed to the next theorem, we shall define a new set R(A).

Definition: Given a compact set $A \subseteq E^{(p)}$. If $\underline{x} \in A$ and $\underline{x} \neq \underline{o}$, then there exists $\nu(\underline{x}) \geq 1$ such that $\nu(\underline{x})\underline{x} \in A$, and if $\nu_1 > \nu(\underline{x})$, then $\nu_1 \underline{x} \notin A$. With this notation we define $R(A) \subseteq A$ as $R(A) = \{\nu(\underline{x})\underline{x} \mid \underline{x} \neq o \; ; \; \underline{x} \in A.\}$.

Theorem: 5.1

If A is a compact set in $E^{(p)}$ then $\Delta_n(R(A))$ is essentially complete (Λ) with respect to $\Delta_n(A)$ for all n and all $\Lambda \subseteq E^{(p)}$

Proof:

Consider any design $d \in \Delta_n(A)$, with the associated information matrix

$$M(d) = \sum n_i M_i(d)$$

where $$M_i(d) = (\underline{x}_i \; \underline{x}'_i)$$

next consider $\underline{\bar{x}}_i = \nu(\underline{x}_i) \; \underline{x}_i \in R(A)$. in the definition of R(A). Then the information matrix of the design d^* with $\underline{\bar{x}}_i \in R(A)$ will be

$$M(d^*) = \sum n_i M_i(d)$$

where

$$\begin{aligned} M_i(d^*) &= (\underline{\bar{x}}_i \; \underline{\bar{x}}_i') \\ &= (\nu^2(\underline{x}_i) \; \underline{x}_i \; \underline{x}_i') \\ &= \nu^2(\underline{x}_i) \; (\underline{x}_i \; \underline{x}_i') \\ &= \nu^2(x_i) \; M_i(d) \end{aligned}$$

Then the condition of the lemma (5.1) that $\underline{\lambda}' M(d^*) \underline{\lambda} - \underline{\lambda}' M(d) \underline{\lambda} \geq o$ for all $\underline{\lambda} \in E^{(p)}$ is satisfied, for

$$\begin{aligned} \underline{\lambda}' M(d^*) \underline{\lambda} - \underline{\lambda}' M(d) \underline{\lambda} &= \underline{\lambda}' (M(d^*) - M(d)) \; \underline{\lambda} \\ &= \sum_i n_i \; (\nu^2(\underline{x}_i) - 1) \; \underline{\lambda}' M_i(d) \; \underline{\lambda} \\ &\geq 0 \end{aligned}$$

in consequence of

$$\nu(\underline{x}_i) \geq 1 \text{ and } \underline{\lambda}' M_i(d)\,\underline{\lambda} = \Big[\sum_j \lambda_j x_{ij}\Big]^2 \geq 0 .$$

Hence $\Delta_n(R(A))$ is essentially complete (Λ) with respect to $\Delta_n(A)$.

Theorem: 5.2

If A is convex body in $E^{(p)}$ with a total of m extreme points $\underline{\alpha}_1, \underline{\alpha}_2, \cdots \underline{\alpha}_m$ then $\Delta_{n+m}(\underline{\alpha}_1, \underline{\alpha}_2, \cdots \underline{\alpha}_m)$ is essentially complete (Λ) with respect to $\Delta_n(A)$ for all n and all $\Lambda \subseteq E^{(p)}$

Proof:

Let the design matrix $\underline{X}_d$ where $d \in \Delta_n(A)$ be

$(n_1, \underline{x}_1; n_2, \underline{x}_2; \ldots; n_r, \underline{x}_r)$ and X_d

where $d^* \in \Delta_{n+m}(\underline{\alpha}_1, \underline{\alpha}_2, \ldots, \underline{\alpha}_m)$ be

$(\bar{n}_1, \underline{\alpha}_1; \bar{n}_2, \underline{\alpha}_2; \ldots; \bar{n}_m, \underline{\alpha}_m)$

with $\sum \bar{n}_j = r \leq n+m$ and $\bar{n}_j \neq 0$ for $j=1,2,\ldots,m$.

Let the information matrices associated with designs d and d^* be $M(d)$ and $M(d^*)$ respectively.

We shall now show that there exist $\bar{n}_1, \ldots, \bar{n}_m$ such that $\bar{n}_j \neq 0$ and

$$\underline{\lambda}' M(d^*)\,\underline{\lambda} - \underline{\lambda}' M(d)\,\underline{\lambda} \geq 0 \text{ for all } \underline{\lambda} \in E^{(p)},$$

so that the proof of the theorem follows from 5.1.

Since $\underline{x}_i \in A$ and A is a convex set generated by $\underline{\alpha}_1, \ldots, \underline{\alpha}_m$ i.e.

$$A = \Big\{ \sum_{i=1}^{m} \lambda_i \underline{\alpha}_i \; ; \; \lambda_i \geq 0 \; , \; i = 1, 2, \ldots m \text{ and } \sum \lambda_i = 1 \Big\}$$

we have

$$\underline{x}_i = \sum_{j=1}^{m} \lambda_{ij} \underline{\alpha}_j \; , \; \lambda_{ij} \geq 0 \; , \; \sum_{j=1}^{m} \lambda_{ij} = 1 \; , \; \begin{matrix} j = 1, 2, \ldots m \\ i = 1, 2, \ldots r. \end{matrix}$$

Now

$$M(\bar{d}) = \sum_{i=1}^{m} \bar{n}_i M_i(d^*) = \sum_{i=1}^{m} \bar{n}_i \underline{\alpha}_i \underline{\alpha}_i'$$

$$= \underline{\alpha}\, \underline{P}\, \underline{\alpha}' \qquad \text{where } \underline{\alpha} = (\underline{\alpha}_1, \underline{\alpha}_2, \cdots \underline{\alpha}_m)$$

$$\text{and } \underline{P} = \begin{pmatrix} \bar{n}_1 & & \underline{0} \\ & \bar{n}_2 & \\ \underline{0} & & \ddots \\ & & \bar{n}_m \end{pmatrix}$$

$$\text{and } M(d) = \sum_{i=1}^{r} n_i M_i(d) = \sum_{i=1}^{r} n_i \underline{x}_i \underline{x}_i'$$

$$= \sum n_i (\underline{\alpha}\, \underline{\lambda}_i)(\underline{\alpha}\, \underline{\lambda}_i)'$$

$$= \sum n_i\, \underline{\alpha}\, (\underline{\lambda}_i \underline{\lambda}_i')\, \underline{\alpha}'$$

$$= \underline{\alpha}\, \underline{S}\, \underline{\alpha}' \qquad \text{where } S = (\sum_{i=1}^{r} n_i \lambda_{iu} \lambda_{iv})$$

It is known from matrix theory that there exists a nonsingular matrix Q such that

$$\underline{\alpha} = \underline{Q}\, \underline{\bar{\alpha}}$$

$$\underline{\text{with}} \quad \underline{Q}' \underline{S}\, \underline{Q} = \begin{pmatrix} \varkappa_1 & & \underline{0} \\ & \ddots & \\ \underline{0} & & \varkappa_m \end{pmatrix} \quad \underline{\text{and}} \quad \underline{Q}' \underline{P}\, \underline{Q} = \underline{I}_m ,$$

$\varkappa_1, \varkappa_2, \cdots \varkappa_m$ being the characteristic roots of

Therefore we have

$$M(d) = \underline{\bar{\alpha}}\, \underline{Q}' \underline{S}\, \underline{Q}\, \underline{\bar{\alpha}}' = \underline{\bar{\alpha}} \begin{pmatrix} \varkappa_1 & & \underline{0} \\ & \ddots & \\ \underline{0} & & \varkappa_m \end{pmatrix} \underline{\bar{\alpha}}'$$

$$= \sum \varkappa_i\, \underline{\bar{\alpha}}_i\, \underline{\bar{\alpha}}_i'$$

$$\text{and } M(\bar{d}) = \sum \underline{\bar{\alpha}}_i\, \underline{\bar{\alpha}}_i'$$

Obviously each $\underline{\bar{\alpha}}_i\, \underline{\bar{\alpha}}_i'$ being a Gram matrix is non-negative definite

$$\underline{\lambda}' \left[M(d^*) - M(d) \right] \underline{\lambda} = \sum_{i=1}^{m} (1 - \kappa_i)\, \underline{\alpha}_i \underline{\alpha}_i' \geq 0 .$$

if $\bar{n}_1, \bar{n}_2, \cdots, \bar{n}_m$ can be chosen so that $|\kappa_i| \leq 1$ for $i = 1, 2, \cdots m$. This can be done by choosing $\bar{n}_1, \cdots, \bar{n}_m$ so that the column sums of $\underline{S}\,\underline{\bar{P}}^{-1}$ are less than or equal to one, since a matrix that has all positive elements, and whose column sums are less than or equal to one, has all characteristic roots $|\kappa_i| \leq 1$.

Consider $\underline{S}\underline{\bar{P}}^{-1} = \left(\frac{1}{\bar{n}_v} \sum n_i \lambda_{iu} \lambda_{iv} \right)$ in which $\bar{n}_v > 0$, $n_i \geq 0$, $\lambda_{ij} \geq 0$. so that each and every element of $\underline{S}\underline{\bar{P}}^{-1}$ is positive.

Let $\Delta = (\Delta_{uv}) = \underline{S}\,\underline{\bar{P}}^{-1}$, then $\Delta_{uv} \geq 0$

and $\Delta_{uv} = \frac{1}{\bar{n}_v} \sum_i n_i \lambda_{iu} \lambda_{iv}$.

The column sums are

$$\sum_u \Delta_{uv} = \frac{1}{\bar{n}_v} \sum n_i \lambda_{iv} \Big]_{+1}$$

Now $\sum_u \Delta_{uv} \leq 1$ when $\bar{n}_v = \left[\sum n_i \lambda_{iv} \right] + 1$ where $[N]$ means the integral value of N.

Thus, $r = \sum_v \bar{n}_v \leq n + m$ and $\bar{n}_v \neq 0$.

This completes the proof.

Corollary. 5.2

Let A be a compact set in $E^{(p)}$ such that R(A) has the property that the convex closure of R(A) is generated by m vectors $\underline{\alpha}_1, \underline{\alpha}_2, \cdots \underline{\alpha}_m$ in R(A); then, $\Delta_{n+m}[\underline{\alpha}_1, \cdots \underline{\alpha}_m]$ is essentially complete (Λ) with respect to (A) for all n and all $\Lambda \leq E^{(p)}$.

Proof . By Theorem 5.2 $\Delta_{n+m}[x_1, \ldots x_m]$ is essentially complete with respect to $\Delta_n[R(A)]$ and by theorem 5.1 $\Delta_n[R(A)]$ is essentially complete (Λ) with respect to $\Delta_n(\Lambda)$.

The following theorem due to de la Garza about polynomial regression , which we shall prove in the chapter on polynomial regression, is mentioned here for continuity.

Theorem .5.3. Let A be the set of positive vectors

$$\underline{x} = \pm (1 , x , x^2 , \ldots\ldots , x^{p-1}) \quad (x_1 \leq x \leq x_2)$$

Then , to any design comprising more than p different observations there exists a p-observation design yielding the same information matrix.

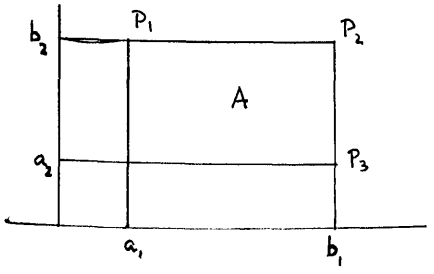
Thus , in any design associated with polynomial regression, for any specified estimation problem , the number of "levels" x needs never exceed the degree of the polynomial.

Since we are mainly concerned with designs in which the number of observations is small , we shall not study the asymtotic completeness in this thesis. Ehrenfeld has studied the subject quite thoroughly.

Example . Consider the model

$$E(y_i) = x_{i1}\beta_1 + x_{i2}\beta_2 .$$

with the set $A = \{(x_{i_1} , x_{i_2}) : 0 \leq a \leq x_i < b ; 0 \leq a \leq x_i \leq B\} \subseteq E$



The set R(A) is the boundary outlined in heavy lines in the figure viz . P_1P_2 and P_2P_3 .

Then theorem 5.1 states that for any n observations in the set A, there exist n observations in R(A) which do as well for estimating any linear combination $\lambda_1\beta_1 + \lambda_2\beta_2$.

Corollary 5.2 states that for any n observations in the set A, there exist r observations ($r \leq n+3$) at the points P_1, P_2, P_3 that are as efficient for estimating any linear combination $\lambda_1\beta_1 + \lambda_2\beta_2$.

It can sometimes happen that we are interested only in a subclass Π^* of problems in Π which are concerned with a fixed , proper subset of p parameters,say .

Let $\underline{M}(d) = \begin{vmatrix} \underline{M}_{11}(d) & \underline{M}_{12}(d) \\ \underline{M}_{21}(d) & \underline{M}_{22}(d) \end{vmatrix}$ and $\underline{M}^{-1}(d) = \begin{vmatrix} \underline{M}^{11}(d) & \underline{M}^{12}(d) \\ \underline{M}^{21}(d) & \underline{M}^{22}(d) \end{vmatrix}$

where $\underline{M}_{11}(d)$ is a (s x s) matrix.

Then the partial information matrix for the relevant parameters $\beta_1, \ldots \beta_s$ may be defined as

$$\underline{M}^*(d) = (\underline{M}^{11}(d))^{-1} = \underline{M}_{11}(d) - \underline{M}_{12}(d)\underline{M}_{22}^{-1}(d)\underline{M}_{21}(d) \qquad (5.4)$$

which is still semipositive definite.

When $\underline{M}(d)$ is singular, $\underline{M}^*(d)$ has to be redefined to serve the purpose purpose of the present theory.If $|\underline{M}_{22}(d)| \neq 0$,(5.4) may be used as a definition.

Otherwise, $\underline{M}_{22}^{-1}(d)\,\underline{M}_{21}(d)$ has to be replaced by any solution, always existing,of the matrix equation

$$\underline{M}_{22}(d)\,\underline{X} = \underline{M}_{21}(d);$$

it is known that the resulting $\underline{M}^*(d)$ is still uniquely determined and non-negative definite.

Uniqueness of $\underline{M}^{*}(d)$.

Let $\underline{X}_1$ and $\underline{X}_2$ be two different solutions of

$$\underline{M}_{22}(d)\,\underline{X} = \underline{M}_{21}(d)$$

i.e.

$$\underline{M}_{22}(d)\,\underline{X}_1 = \underline{M}_{21}(d)$$

$$\underline{M}_{22}(d)\,\underline{X}_2 = \underline{M}_{21}(d) \qquad (5.5)$$

Then the corresponding values of $\underline{M}^{*}(d)$ will be

$$\underline{M}_{11}(d) - \underline{M}_{12}(d)\,\underline{X}_1$$

and

$$\underline{M}_{11}(d) - \underline{M}_{12}(d)\,\underline{X}_2 .$$

$\underline{M}_{12}\underline{X}_1$ and $\underline{M}_{12}\underline{X}_2$ are both symmetric for their transposes are respectively equal to $\underline{X}_1'\underline{M}_{21} = \underline{X}_1'\underline{M}_{22}(d)\,\underline{X}_1$ and $\underline{X}_2'\underline{M}_{21} = \underline{X}_2'\underline{M}_{22}(d)\underline{X}_2$.

Now from (5.5) again

$$\underline{X}_1'\underline{M}_{21}(d) = \underline{X}_2'\underline{M}_{22}\underline{X}_1$$

and

$$\underline{X}_1'\underline{M}_{21}(d) = \underline{X}_1'\underline{M}_{22}\underline{X}_2 .$$

Since the left hand sides are symmetric, so are the right hand sides.

Therefore

$$\underline{X}_1'\underline{M}_{21}(d) = \underline{X}_2'\underline{M}_{21}(d)$$

or

$$\underline{M}_{12}(d)\underline{X}_1 = \underline{M}_{12}(d)\underline{X}_2 ,$$

establishing the uniqueness of the partial information matrix $\underline{M}^{*}(d)$.

Now Lemma 5.1 may be generalized as follows .

Lemma 5.1 . If $\underline{M}^{*}(\overset{*}{d}) \geq \underline{M}(\overset{*}{d})$, d is at least as good (π_L^{*}) as d ,

π_L^{*} being the class of all problems of point estimation of any linear parametric function of a subset of parameters.

i.e. Given $\underline{M}_{11}(\overset{*}{d}) - \underline{M}_{12}(\overset{*}{d})\,\underline{M}_{22}^{-1}(\overset{*}{d})\,\underline{M}_{21}(\overset{*}{d}) \geq \underline{M}_{11}(d) - \underline{M}_{12}(d)\,\underline{M}_{22}^{-1}(d)\,\underline{M}_{21}(d)$,

then we have to show

(a) if, for any given $\underline{\lambda}$, there exists $\underline{\rho}$ such that

$$\underline{M}^{*}(d)\,\underline{\rho}_{1} = \underline{\lambda}$$

then there exist $\underline{\rho}^{*}$ such that

$$\underline{M}^{*}(d)\,\underline{\rho}^{*}_{1} = \underline{\lambda} \qquad \text{and}$$

(b) $$\underline{\rho}^{*'}_{1}\,\underline{M}^{*}(d)\,\underline{\rho}^{*}_{1} \leq \underline{\rho}'_{1}\,\underline{M}^{*}(d)\,\underline{\rho}_{1} \quad .$$

Let the relevant parameters be $\beta_1, \ldots \beta_s$ and the linear parametric function be $\psi = \underline{\lambda}'\underline{\beta}$ where $\underline{\lambda} = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_s \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ and $\underline{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_s \\ \beta_{s+1} \\ \vdots \\ \beta_p \end{pmatrix}$

Then, if ψ is estimable under d,

$$\underline{M}(d)\,\underline{\rho} = \underline{\lambda} \text{ has a solution.}$$

i.e. $$\begin{pmatrix} \underline{M}_{11}(d) & \underline{M}_{12}(d) \\ \underline{M}_{21}(d) & \underline{M}_{22}(d) \end{pmatrix} \begin{pmatrix} \underline{\rho}_1 \\ \underline{\rho}_2 \end{pmatrix} = \begin{pmatrix} \underline{\lambda}_1 \\ \underline{0} \end{pmatrix}$$

i.e. $$\underline{M}_{11}(d)\underline{\rho}_1 + \underline{M}_{12}(d)\underline{\rho}_2 = \underline{\lambda}_1$$

$$\underline{M}_{21}(d)\underline{\rho}_1 + \underline{M}_{22}(d)\underline{\rho}_2 = \underline{0}$$

From these two equations, we have

$$\underline{\rho}_2 = -\underline{M}^{-1}_{22}(d)\underline{M}_{21}(d)\underline{\rho}_1$$

and $$(\underline{M}_{11}(d) - \underline{M}_{12}(d)\underline{M}^{-1}_{22}(d)\underline{M}_{21}(d))\underline{\rho}_1 = \underline{\lambda}_1$$

i.e. $$\underline{M}^{*}(d)\underline{\rho}_1 = \underline{\lambda}_1 \text{ has a solution.}$$

From lemma 5.1 we know that under the condition $\underline{M}(d) \geq \underline{M}(d)$

$$\underline{M}^{*}(d)\underline{\rho}_1 = \underline{\lambda}_1 \text{ has a solution}$$

implies that $$\underline{M}^{*}(d)\underline{\rho}_1 = \underline{\lambda}_1 \text{ has a solution.}$$

i.e. ψ is estimable under d^{*}.

Proof for the second part is the same as in lemma 5.1.

If we can show that $\text{var}(\psi) = \underline{\rho}_1 \underline{M}^{*}(d)\underline{\rho}_1 \sigma^2$ where $\underline{\rho}_1$ is a solution of $\underline{M}^{*}(d)\underline{\rho}_1 = \underline{\lambda}_1$, the lemma is proved,

Since ψ is estimable, $\underline{M}(d)\underline{\rho} = \underline{\lambda}$

$$\text{or} \qquad \underline{M}(d)\begin{pmatrix}\underline{\rho}_1\\ \underline{\rho}_2\end{pmatrix} = \begin{pmatrix}\underline{\lambda}_1\\ \underline{0}\end{pmatrix}$$

$$\text{or} \quad \begin{pmatrix}\underline{M}_{11}(d) & \underline{M}_{12}(d)\\ \underline{M}_{21}(d) & \underline{M}_{22}(d)\end{pmatrix}\begin{pmatrix}\underline{\rho}_1\\ \underline{\rho}_2\end{pmatrix} = \begin{pmatrix}\underline{\lambda}_1\\ \underline{0}\end{pmatrix}$$

$$\text{i.e.} \quad \begin{aligned}\underline{M}_{11}(d)\underline{\rho}_1 + \underline{M}_{12}(d)\underline{\rho}_2 &= \underline{\lambda}_1\\ \underline{M}_{21}(d)\underline{\rho}_1 + \underline{M}_{22}(d)\underline{\rho}_2 &= \underline{0}\end{aligned} .$$

Therefore $\text{var}\ \psi = \text{var}\,(\underline{\lambda}'\underline{\beta})$

$$= \sigma^2(\underline{\rho}_1', \underline{\rho}_2')\underline{M}(d)\begin{matrix}\underline{\rho}_1\\ \underline{\rho}_2\end{matrix}$$

$$= \sigma^2(\underline{\rho}_1', \underline{\rho}_2')\begin{pmatrix}\underline{M}_{11}(d) & \underline{M}_{12}(d)\\ \underline{M}_{21}(d) & \underline{M}_{22}(d)\end{pmatrix}\begin{pmatrix}\underline{\rho}_1\\ \underline{\rho}_2\end{pmatrix}$$

$$= \sigma^2(\underline{\rho}_1'\underline{M}_{11}(d)\underline{\rho}_1 + \underline{\rho}_1'\underline{M}_{12}(d)\underline{\rho}_2 + \underline{\rho}_2'\underline{M}_{22}(d)\underline{\rho}_2 + \underline{\rho}_2'\underline{M}_{21}(d)\underline{\rho}_1) .$$

$$= \sigma^2(\underline{\rho}_1'\underline{M}_{11}(d)\underline{\rho}_1 - \underline{\rho}_1'\underline{M}_{12}(d)\underline{M}_{22}^{-1}(d)\underline{M}_{21}(d)\underline{\rho}_1)$$

$$= \sigma^2\underline{\rho}_1'(\underline{M}_{11} - \underline{M}_{12}\underline{M}_{22}^{-1}\underline{M}_{21})\underline{\rho}_1.$$

$$= \sigma^2\,\underline{\rho}_1'\underline{M}^{*}\underline{\rho}_1 .$$

Chapter VI.

6.1 Optimum Allocation in Polynomial Regression:

In fitting a polynomial of degree p in a variable x to a dependent variable y, we are interested in estimating the coefficients β_i of x_j^i, setting confidence intervals on the β_i and testing its significance. If the n observations on x can be located at within given limits (which may be taken as ± 1), we may so allocate them to get maximum precision to the estimates of these coefficients. If the allocation is restricted for example,*to equally spaced points, we shall have to consider the restricted optimum i.e. optimum allocation under the restriction.

Our basic problem is how best to estimate the β_i's.

The model we are referring to can be written

$$y_j = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + \cdots + \beta_p x_j^p + e_j$$

$$j = 1, 2, \cdots, n.$$

or $\underline{y} = \underline{X}\,\underline{\beta} + \underline{e}$.

where

$$\underline{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \underline{X} = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^p \\ 1 & x_2 & x_2^2 & \cdots & x_2^p \\ \vdots & & & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^p \end{pmatrix}$$

$$\underline{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix} \quad \text{and} \quad \underline{e} = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}.$$

Now consider the determinant of any $(p+1) \times (p+1)$ subsmatrix of the coefficient matrix $\underline{X}$. Without any loss of generality we may consider the matrix of the first $(p + 1)$ rows, with determinant

$$\begin{vmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^p \\ 1 & x_2 & x_2^2 & \cdots & x_2^p \\ \vdots & & & & \vdots \\ & x_p & x_p^2 & \cdots & x_p^p \end{vmatrix} = \prod_{i<j} (x_j - x_i)$$

which is very well known as the alternant of Vandwmonde and has the value $\prod_{i<j} (x_j - x_i)$. Therefore the determinant vanishes when and only when $x_i = x_j$ (any i and j). We shall be considering at least p + 1 different values of x_j's. Consequently our square submatrix of order $(p + 1) \times (p + 1)$ is nonsingular and $\underline{X}$ is of full rank. Thus all the theorems of full rank model will apply to the polynomial regression problem.

Two aspects of the estimation problem. in polynomial region are (1) To determine the best method for using the information given by a set of observations $y_1, y_2, \cdots, y_n$.
(2) To determine the best method for choosing the x values at which to take observations.

The first aspect of the problem has been studied throroughly, though considerably very much less has been done on the second. We shall be concerned here with the latter aspect of the problem.

Very recently, De La Garza (1954) was able to show that just as much information is obtained from observations made at certain

$(p + 1)$ points in the interior of an interval as from n $(> p + 1)$ points on that interval; the measure of information being the variance-covariance matrix of the estimated polynomial coefficients.

6.2 Formulation of the problem:

Ω : $\underline{y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$ with rank $(\underline{X}) = p + 1$, $E(\underline{\varepsilon}) = 0$ and $E(\underline{\varepsilon}\underline{\varepsilon}') = V_{\underline{x}}$ i.e. there are at least $(p + 1)$ distinct observed variates x_i. By Gauss-Markoff theorem, the best linear unbiased minimum variance estimate is given by $(\underline{X}'V_{\underline{x}}^{-1}\underline{X})^{-1}\underline{X}'V_{\underline{x}}^{-1}\underline{y}$.

Thus the generalized variance of the estimates or $|\mathrm{var}(\hat{\underline{\beta}})|$

$$= |(\underline{X}'V_{\underline{x}}^{-1}\underline{X})^{-1}\underline{X}'V_{\underline{x}}^{-1}\ \mathrm{var}(\underline{y})\ V_{\underline{x}}^{-1}\underline{X}(\underline{X}'V_{\underline{x}}^{-1}\underline{X})^{-1}|$$

$$= |(\underline{X}'V_{\underline{x}}^{-1}\underline{X})^{-1}(\underline{X}'V_{\underline{x}}^{-1}\underline{X})(\underline{X}'V_{\underline{x}}^{-1}\underline{X})^{-1}|$$

$$= |(\underline{X}'V_{\underline{x}}^{-1}\underline{X})^{-1}|$$

$$= 1/|\underline{X}'V_{\underline{x}}^{-1}\underline{X}|$$

We shall therefore use the generalized variance $|\underline{X}'V_{\underline{x}}^{-1}\underline{X}|^{-1}$ as a measure of efficiency and an optimum solution will be based on this generalized variance.

Now we shall show (De La Garza) that the same information matrix $\underline{X}'V_{\underline{x}}^{-1}\underline{X}$, and hence the same value of the G.V. can be obtained by replacing a given set of n observations at the points $x_1, x_2, \ldots x_n$, by a total of n observations made at $(p + 1)$ properly selected points in the interval from x_1 to x_n.

The problem is to show that given a spacing of observations at x_i, $i = 1, 2, \ldots n$; $n \geq p+1$, these being at least $(p + 1)$ distinct x_i with the information matrix $\underline{M}$ it is always possible

to respace the observations at $(p + 1)$ distinct locations

in such a manner that $\text{minimum } x_i \leq r_j \leq \text{maximum } x_i$

$$i = 1, 2.....n$$

$$j = 1, 2.....p + 1$$

and $\underline{X}' \underline{V}_{\underline{x}}^{-1} \underline{X} = \underline{R} \underline{V}_{\underline{r}}^{-1} \underline{R}$ where $\underline{R}' \underline{V}_{\underline{r}}^{-1} \underline{R}$ is the information matrix of the respacing.

The problem is solved by prescribing a method for finding the required $\underline{V}_{\underline{r}}$ and $\underline{R}$ which determine the spacing of the observations.

Some useful relations:

Prior to investigating the problem as outlined above, several relations needed latter will be developed.

Consider the polynomial

$$P(\underline{x}) = \underline{x}' \underline{\beta} = (1, x, x^2, \ldots, x^p) \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}$$

$$= \beta_0 + \beta_1 x + \ldots + \beta_p x^p .$$

Choose $(p + 1)$ distinct numbers z_j

From Lagrange interpolation is follows that

$$P(x) = \frac{(x-z_2)(x-z_3) \cdots (x-z_{p+1})}{(z_1-z_2)(z_1-z_3) \cdots (z_1-z_{p+1})} P(z_1)$$

$$+ \frac{(x-z_1)(x-z_3) \cdots (x-z_{p+1})}{(z_2-z_1)(z_2-z_3) \cdots (z_2-z_{p+1})} P(z_2) + \ldots + \frac{(x-z_1) \cdots (x-z_p)}{(z_{p+1}-z_1) \cdots (z_{p+1}-z_p)} P(z_{p+1}).$$

With an obvious notation

$$P(x) = \sum_{j=1}^{p+1} F(x, z_j) P(z_j) .$$

Now consider

$$(P(z)) = \begin{pmatrix} P(z_1) \\ P(z_2) \\ \vdots \\ P(z_{p+1}) \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 z_1 + \beta_2 z_1^2 + \cdots + \beta_p z_1^p \\ \vdots \\ \beta_0 + \beta_1 z_{p+1} + \cdots + \beta_p z_{p+1}^p \end{pmatrix}$$

$$= \begin{pmatrix} 1 & z_1 & z_1^2 & \cdots & z_1^p \\ 1 & z_2 & z_2^2 & \cdots & z_2^p \\ \vdots & & & & \vdots \\ 1 & z_{p+1} & z_{p+1} & \cdots & z_{p+1}^p \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p+1} \end{pmatrix}$$

$$= \underline{Z}\,\underline{\beta}.$$

The matrix $\underline{Z}$ is nonsingular, since its determinant is a Vondermonde determinant not equal to zero due to the z_i being different. Thus

$$\underline{\beta} = \underline{Z}^{-1}(P(z)) \qquad \text{and for any } x$$

$$\underline{x}'\underline{\beta} = \underline{x}'\underline{Z}^{-1}(P(z)) \qquad \text{where } \underline{x}' = (1, x, x^2, \cdots x^p)$$

But $\mathbf{P}(x) = (F(x,z))'(P(z))$

where $$(F(x,z)) = \begin{pmatrix} F(x,z_1) \\ F(x,z_2) \\ \vdots \\ F(x,z_{p+1}) \end{pmatrix}$$

Thus $$\underline{x}'\underline{Z}^{-1}(P(z)) = (F(x,z))'(P(z))$$

Equality for any x implies

$$\underline{x}'\underline{Z}^{-1} = (F(x,z)) \qquad (6.1)$$

6.3 Investigation of the problem:

Using the appropriate transformations (i.e. $x = \frac{x - (x_1 + x_p)/2}{(x_p - x_1)/2}$) it may be shown that without any loss of generality the range of the variable x may be limited to minimum $x_i = -1$ and maximum $x_i = +1$. Suppose now that some of the x_i are not distinct; say $x_1 = x_2 = \cdots = x_k$, with the corresponding variances of $y_i = \sigma^2$ $(i = 1, 2, \cdots k)$

Therefore the variance of the mean y^*

$$= \text{Var}\left(\Sigma y_i/k\right) = \sigma^2/k .$$

Such a group, may be made for all x_i not distinct, thereby reducing the problem to considering only distinct x_i . Finally for $n = p+1$, there is no problem since the spacing is already at (p + 1) locations.

Now we have reduced the problem to a mathematically convenient form: Given a total of n observations at n distinct locations x_i, $i = 1, 2, \cdots n$, $n > (p+1)$ with minimum $x_i = -1$ and maximum $x_i = +1$, it is always possible to respace the observations at (p + 1) distinct locations r_j, $j = 1, 2, \cdots p+1$, in such a manner that $-1 \leq r_j \leq +1$, and $\underline{X}' V_{\underline{x}}^{-1} \underline{X} = \underline{R}' V_{\underline{r}}^{-1} \underline{R}$.

Suppose now that $\underline{R}$ exists, then

$$V_{\underline{r}}^{-1} = (\underline{X}\,\underline{R}^{-1})' V_{\underline{x}}^{-1} (\underline{X}\,\underline{R}^{-1})$$

From (6.1) it follows that the off diagonal elements of $(\underline{X}\,\underline{R}^{-1})' V_{\underline{x}}^{-1} (\underline{X}\,\underline{R}^{-1})$ are proportional to

$$c_{gh} = \sum_i \omega_i \left[\prod_j (x_i - r_j)\right]\left[\prod_k (x_i - r_k)\right] \quad (g \neq h)$$
$$= \sum_i \omega_i \phi_i \left[\prod_k (x_i - r_k)\right] \qquad (6.2)$$

Since $V_{\underline{r}}^{-1}$ is required to be diagonal, we must have all these equal to

zero for $g \neq h$.

This requirement is satisfied if the r_j are determined such that

$$\sum \omega_i \phi_i = 0 \,, \quad \sum \omega_i \phi_i x_i = 0 \,, \quad \cdots \,, \quad \sum \omega_i \phi_i x_i^{p-1} = 0 \quad \text{where} \quad \phi_i = \prod_j (x_i - r_j) \tag{6.3}$$

For reasons that will be discussed later, we shall further constrain the r_j by

$$\sum \omega_i \phi_i x_i^p = 0 \tag{6.4}$$

By direct expansion,

$$\phi_i = \alpha_1 + \alpha_2 x_i + \cdots + \alpha_{p+1} x_i^p + x_i^{p+1} \tag{6.5}$$

Hence r_j are the $p + 1$ roots of the polynomial

$$P(r) = \alpha_1 + \alpha_2 r + \cdots + \alpha_{p+1} r^p + r^{p+1} \tag{6.6}$$

Substituting (6.5) in (6.4) and (6.3) there results

$$\begin{vmatrix} f_0 & f_1 & \cdots & f_{p-1} & f_p \\ f_1 & f_2 & \cdots & f_p & f_{p+1} \\ \vdots & & & & \\ f_{p-1} & f_p & \cdots & f_{2p-2} & f_{2p-1} \\ f_p & f_{p+1} & \cdots & f_{2p-1} & f_{2p} \end{vmatrix} \begin{vmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{p+1} \end{vmatrix} + \begin{vmatrix} f_{p+1} \\ f_{p+2} \\ \vdots \\ f_{2p+1} \end{vmatrix} = \underline{0} \tag{6.7}$$

with $f_t = \sum \omega_i x_i^t$, $t = 0, 1, 2, \cdots 2p+1$, which is a system of $(p + 1)$ equations in $(p + 1)$ unknowns.

The square matrix is obviously $\underline{X}' \, V_x^{-1} \, \underline{X}$ which is nonsingular and hence (6.6) has a unique solution; $\underline{\alpha}$. This solution is not trivial, since it is readily seen that some f_t, $p+1 \leq t \leq 2p+1$, is not zero, because $f_t = \sum \omega_i x_i^t \neq 0$ for positive x_i. The corresponding r_j are then given by (6.6).

Thus, a method of determining the r_j that satisfy (6.3) and (6.4) has been prescribed. Accordingly these r_j make c_{gh} zero in (6.2) as

required. It will now be shown that they are real distinct and lie between -1 and +1.

To show that the r_j are real and distinct:

Since the roots of a polynomial occur in conjugate pairs, let a pair of complex r_j if possible, be $r_1 = b_1 + i b_2$ and $r_2 = b_1 - i b_2$ with $b_2 \neq 0$, the nature of the remaining roots being unspecified. Since in (6.2) c_{12} is then zero, it follows that

$$\sum \omega_i \left[(x_i - b_1)^2 + b_2^2\right] (x_i - r_3)^2 \cdots (x_i - r_p)^2 (x_i - r_{p+1})^2 = 0 \qquad (6.8)$$

All factors in each term being non-negative, equality to zero implies that all terms must be zero.

But $\left[(x_i - b_1)^2 + b_2^2\right]$ with $b_2 \neq 0$ never vanishes; and $(x_i - r_3)^2 \cdots (x_i - r_{p+1})^2$ can vanish for at most $(p - 1)$ distinct x_i. Since there are at least $(p + 2)$ distinct x_i it follows that (6.8) cannot be zero, and hence r_1 and r_2 are not complex. By the same argument for any other pair of roots, all the r_j are real;

To show that they are distinct, we may suppose $r_1 = r_2 = b_1$ and again from c_{12}, which is now given by (6.8) with $b_2 = 0$. Following the same line of argument, the terms can now vanish for at most p distinct x_i but since there are at least $p + 1$ distinct x_i, (6.8) cannot be zero with $b_2 = 0$. Hence, all the r_j are distinct.

Now, since the r_j are distinct, it follows that the matrix $\underline{R}$ is nonsingular and that c_{gh} being zero in (6.2) implies that $(\underline{X}\,\underline{R}^{-1})' \underline{V}_x^{-1} (\underline{X}\,\underline{R}^{-1}) = \underline{V}_r^{-1}$ is a diagonal matrix. Both $\underline{X}' \underline{V}_x^{-1} \underline{X}$ and $\underline{R}$ being nonsingular, it follows that no diagonal element of $\underline{V}_r^{-1}$ is zero.

Diagonal elements of V_{Σ}^{-1} are obtained by referring to (6.1)

$$u_h = \sum_i \omega_i \prod_{j \neq h} \frac{(x_i - r_j)^2}{(r_h - r_j)^2} \qquad (6.9)$$

and therefore all the u_h are positive. Thus, with V_{Σ}^{-1} given by (6.9)

$$X' V_x^{-1} X = R' V_{\Sigma}^{-1} R \quad .$$

Finally, we shall have to show that $-1 \leq r_j \leq +1$. Suppose that the r_j are such that two or more of the r_j are not in the closed interval (-1, 1). Say that r_1 and r_2 are such values. Since c_{12} in (6.2) must be zero, it follows that

$$\sum_i \omega_i (x_i - r_1)(x_i - r_2)(x_i - r_3)^2 \cdots (x_i - r_p)^2 (x_i - r_{p+1})^2 = 0 \qquad (6.10)$$

Consider that $(x_i - r_1)(x_i - r_2)$ never equals zero and always must have the same algebraic sign for $-1 \leq x_i \leq +1$. Furthermore

$$(x_i - r_3)^2 \cdots (x_i - r_p)^2 (x_i - r_{p+1})^2 \geq 0 \qquad (6.11)$$

Hence equality to zero in (6.10) implies that all terms must be zero. But

$$(x_i - r_1)(x_i - r_2)$$

can never vanish, and (6.11) can vanish for at most (p - 1) distinct

Since there are at least (p + 2) distinct x_i , all terms in (6.10) cannot vanish. Thus, two or more of the r_j cannot be excluded from the closed interval [-1, 1], and hence, it has been shown that p of the r_j are in the closed interval [- 1, 1].

Consider now the polynomial in (6.6) whose roots are the r_j . In determinant form by using Cramer's rule, this polynomial may be shown to be

$$P(r) = \frac{(-1)^{p+1}}{\Delta} \begin{vmatrix} 1 & r & r^2 & \cdots & r^p & r^{p+1} \\ f_0 & f_1 & & \cdots & f_p & f_{p+1} \\ \vdots & & & & & \\ f_p & f_{p+1} & & \cdots & f_{2p} & f_{2p+1} \end{vmatrix} \qquad (6.12)$$

where $\Delta = |\underline{X}' V_x^{-1} \underline{X}|$

Evaluate $P(r)$ at $r = -1$ and $r = +1$. It will be seen that, for $J = -1, 1$

$$P(J) = \frac{J^{p+1}}{\Delta} \begin{vmatrix} f_0 - \bar{J} f_1 & f_1 - \bar{J} f_2 & \cdots & f_p - \bar{J} f_{p+1} \\ f_1 - \bar{J} f_2 & f_2 - \bar{J} f_3 & \cdots & f_{p+1} - \bar{J} f_{p+1} \\ \vdots & & & \vdots \\ f_p - \bar{J} f_{p+1} & f_{p+1} - \bar{J} f_{p+2} & \cdots & f_{2p} - \bar{J} f_{2p+1} \end{vmatrix} \qquad (6.13)$$

Using the definition of f_L, the elements of the indicated determinant $|H_J|$ in the expression of $P(J)$ are of the form $\sum \omega_i (1 - \bar{J} x_i) x_i^L$,

$L = 0, 1, \ldots\ldots$ which shows that

$$\underline{H}_J = \underline{X}' \underline{\Omega}_J \underline{X} \qquad (6.14)$$

where Δ_J is the n x n diagonal matrix with diagonal elements $(1 - Jx_i)$. Since $\min x_i = -1$ and $\max x_i = 1$, it follows that Δ_J always has one diagonal element equal to zero, all others being positive.

Hence from (6.14), it follows that

$$\underline{H}_J = \underline{X}'\underline{V}_J\underline{X}_J ,$$

where $\underline{V}_J$ is the $(n-1) \times (n-1)$ matrix formed by striking out the row and column corresponding to $\min x_i$ for $J = -1$ and $\max x_i$ for $J = +1$, and $\underline{X}'_J$ is the $(n-1) \times (p+1)$ matrix formed from $\underline{X}$ by striking out the row corresponding to $\min x_i$ for $J = -1$ and $\max x_i$ for $J = 1$.

Since there are at least $(p+2)$ distinct x, $\underline{X}$ has rank $(p+1)$.

Also, since $\underline{V}_J$ is diagonal with nonzero diagonal elements, $\underline{V}_J$ is non-singular.

Obviously $$\underline{H}_J = \underline{X}'\underline{V}_J\underline{X}_J = \underline{X}'\underline{V}^*_J\underline{V}^*_J\underline{X}_J ,$$

where $\underline{V}^*_J$ = a diagonal matrix with elements equal to the square root of the corresponding elements of $\underline{V}_J$.

$\underline{H} = (\underline{V}^*_J\underline{X}_J)'(\underline{V}^*_J\underline{X}_J)$, which being a Gram matrix is always positive semi-definite.

Since $\underline{V}^*_J$ is full rank, $\underline{V}^*_J\underline{X}_J$ is of rank $p + 1$ equal to the number of its columns.

Therefore $(\underline{V}^*_J\underline{X})'(\underline{V}^*_J\underline{X})$ is positive definite.

It follows that $\underline{H}_J$ is positive definite and hence $|H_J| > 0$.

Now since $\Delta > 0$, (obtained on the same line of argument as above), the

following conclusion may now be made concerning $P(J)$.

$$\begin{aligned} P(+1) &> 0 \quad \text{and} \quad P(\infty) > 0 . \\ P(-1) &> 0 \quad \text{and} \quad P(-\infty) > 0 \quad \text{for odd } m . \\ P(-1) &< 0 \quad \text{and} \quad P(-\infty) < 0 \quad \text{for even } m . \end{aligned} \tag{6.15}$$

It was previously shown that p of the roots r_j are in closed interval $[-1 , 1]$. (6.15) shows that -1 and 1 cannot be roots , and hence , it may be stated that p of the roots are in the open interval $(-1,1)$. Furthermore, knowing the sign of $P(r)$ for $r = -1, 1$ and for sufficiently large values of $|r|$, it may be reasoned that all r_j are in the open interval $(-1, 1)$,for one exterior root would imply another.

In conclusion, it has been shown that for $n > (p + 1)$, $-1 < r_j \leq 1$. As obtained in the earlier discussion , for $n = p + 1$, $-1 \leq r_j \leq 1$. Hence $-1 \leq r_j \leq 1$ holds for all cases , and the solution of the problem is complete.

6.5 Some remarks about the constraints .

Returning to (6.2) , it may be noted that the constraints (6.3) are sufficient to make $\underline{V}_{\underline{r}}^{-1}$ a diagonal matrix . The added constraint (6.4) is sufficient to locate all r_j in the closed interval $[-1, 1]$ .To see this, we may consider the situation in which (6.4) is not imposed . This is equivalent to striking out the last row of $\underline{X}'\underline{V}_{\underline{x}}^{-1}\underline{X}$ in (6.7), leaving a system of p linear equations in $(p + 1)$ unknowns. The rank of the coefficient matrix of this system is p , and hence $p+1$ can be chosen at will. Now , the diagonality of $\underline{V}_{\underline{r}}^{-1}$ demands that p of the r_j be in the closed interval $[-1, 1]$. Accordingly, since

$\alpha_{p+1} = -\sum \gamma_i$ by choosing $|\beta p + 1|$ sufficiently large, a root can always be obtained exterior to $[-1, 1]$.

Application.

In interpolation problem, we have to estimate the coefficients in the model

$$\underline{Y} = \underline{X}\underline{\beta} + \underline{e} \quad \text{(the polynomial regression model).}$$

or $$y_i = \beta_1 + \beta_2 x_i + \cdots \beta_p x_i^{p-1} + e_i \,, \quad i = 1, 2, \ldots n$$

where $\underline{E}(\underline{e}) = \underline{0}$ and $E(\underline{e}\,\underline{e}') = V_{\underline{x}}$,

so that the fitted polynomial $P(x)$ does not differ very much from the theoretical model in the sense of a certain goodness criteria.

Using the results of the full rank model, the L.S. estimate of $\underline{\beta}$ and $P(x)$ are given respectively by

$$\hat{\underline{\beta}} = (\underline{X}' V_{\underline{x}}^{-1} \underline{X})^{-1} \underline{X}' V_{\underline{x}}^{-1} \underline{Y}$$

or $$\hat{P}(x) = \underline{x}' \hat{\underline{\beta}} \,.$$

Also we know

$$\text{var}\,(\hat{\underline{\beta}}) = (\underline{X}' V_{\underline{x}}^{-1} \underline{X})^{-1}$$

and $$\text{var}\,\hat{P}(x) = \text{var}\; \underline{x}' \hat{\underline{\beta}}$$

$$= \underline{x}' \,\text{var}\, \hat{\underline{\beta}}\; \underline{x}$$

$$= \underline{x}' (\underline{X}' V_{\underline{x}}^{-1} \underline{X}) \underline{x} \,.$$

For simplicity we shall consider an internal interpolation problem for the quadratic.

Problem. Given n independent observations on a model

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + e_i$$

to be taken in the specified interval $x_L \le x_i \le x_H$.

of equal variances σ^2 , to find the spacing of the n observations that will minimize the maximum variance $Y(\xi)$ for $x_L \leq \xi \leq x_H$, where $Y(\xi)$ is the L.S. estimator of $\mathcal{P}(\xi)$.

Solution.

The variance of $Y(\xi)$ have been found to be

$$\sigma^2_{Y(\xi)} = \underline{\xi}'(\underline{X}' \underline{V}_x^{-1} \underline{X})\, \underline{\xi}$$

Now, whatever be the optimum spacing, it will give rise to some matrix $\underline{X}' \underline{V}_x^{-1} \underline{X}$; let this be $(\underline{X}' \underline{V}_x^{-1} \underline{X})_0$.

From the above theory, it follows that there exists a matrix $\underline{R}$, where

$$\underline{R}' \underline{V}_x^{-1} \underline{R} = (\underline{X}' \underline{V}_x^{-1} \underline{X})_0 .$$

Since p = 2

$$\underline{R} = \begin{pmatrix} 1 & r_1 & r_1^2 \\ 1 & r_2 & r_2^2 \\ 1 & r_3 & r_3^2 \end{pmatrix} \qquad \text{and} \qquad \underline{V}_r^{-1} = \begin{pmatrix} n_1 & 0 & 0 \\ 0 & n_2 & 0 \\ 0 & 0 & n_3 \end{pmatrix} \frac{1}{\sigma^2}$$

implying n_j observations at r_j satisfying $x_L \leq r_j \leq x_H$, and

$\sum n_j = n, \quad j = 1, 2, 3.$

Hence, three locations suffix to establish the desired optimum spacing.

Let y_j^* be the average of n observations at r_j , let $\underline{\eta}$ be the column vector (y_1^* , y_2^* , y_3^*) .

Let $\underline{a}$ be the vector satisfying the normal equations

$$\underline{R}' \underline{V}_r^{-1} \underline{R}\ \underline{a} = \underline{R}' \underline{V}_r^{-1} \underline{\eta} .$$

Since $\underline{R}$ and $\underline{V}_r$ are nonsingular, $\underline{R}\ \underline{a} = \underline{\eta}$

But $\underline{R}\ \underline{a} = \begin{pmatrix} Y(r_1) \\ Y(r_2) \\ Y(r_3) \end{pmatrix}$

i.e. $Y(\xi)$ passes through y_j^* at $\xi = r_j$.

Hence $Y(\xi)$ may be written in the Lagrange form,

$$Y(\xi) = \frac{(\xi - r_2)(\xi - r_3)}{(r_1 - r_2)(r_1 - r_3)} y_1^* + \frac{(\xi - r_1)(\xi - r_3)}{(r_2 - r_1)(r_2 - r_3)} y_2^* + \frac{(\xi - r_1)(\xi - r_2)}{(r_3 - r_1)(r_3 - r_2)} y_3^* .$$

Now $\sigma^2_{y(r_j)} = \sigma^2_{y_j^*} = \sigma^2/n_j$

For any such spacing, let σ^2_{max} be the maximum variance of $Y(\xi)$ in the interval. Then

$$\frac{\sigma^2}{n_j} \leq \sigma^2_{max}$$

and thus,

$$\frac{\sigma^2}{3} \sum_1^3 \frac{1}{n_j} \leq \sigma^2_{max}$$

Therefore $\min \frac{\sigma^2}{3} \sum_1^3 \frac{1}{n_j} \leq \min \sigma^2_{max}$. (E.2)

But the minimum value of $\sum \frac{1}{n_j}$ is constrained by $\sum_1^3 n_j = n$.

The Lagrangian function

$$L = \sum_1^3 \frac{1}{n_j} + \lambda \left(\sum_1^3 n_j - n \right)$$

Therefore $\frac{\partial L}{\partial n_j} = -\frac{1}{n_j^2} + \lambda = 0$

or $\frac{1}{n_j} = +\sqrt{\lambda}$ (each n being positive)

i.e. $\frac{1}{n_1^2} = \frac{1}{n_2^2} = \frac{1}{n_3^2}$ or $n_1 = n_2 = n_3 = \frac{n}{3}$.

Thus the constrained minimum of $\sum_{j=1}^3 \frac{1}{n_j} = \frac{9}{n}$.

Hence from (E 2),

$$\frac{\sigma^2}{3} \cdot \frac{9}{n} \leq \min \sigma^2_{max} .$$

(E 3) i.e. $$3\frac{\sigma^2}{n} \leq \min \sigma^2_{max} .$$

From (E.1) we can see that $\sigma^2_{Y(\xi)}$ increases as ξ departs from the smallest

and the largest r_j in the direction of leaving the interval x_L, x_H. Now locate $\frac{n}{3}$ observations at x_L and x_H. Then $\sigma^2_{Y(\xi)}$ being quadratic in ξ, has one differential maximum occurring in the interior of the interval. From symmetry of the form of $\sigma^2_{Y(\xi)}$, this differential maximum then occurs at $\frac{x_L + x_H}{2}$, at which we may now locate the remaining $\frac{n}{3}$ observations.

Hence the maximum value of $\sigma^2_{Y(\xi)}$ for the spacing $r = x_L$, $r = (x_L + x_H)/2$, $r = x_H$, $n = \frac{n}{3}$ is obtained at $= \frac{x_L + x_H}{2}$ and its maximum value $\max \sigma^2_{Y(\xi)} = \frac{3\sigma^2}{n}$.

The inequality (6.3) assures that this particular spacing gives the desired minimum of the maximum variance, and this value of $\frac{3\sigma^2}{n}$ is attained with our spacing. Thus our spacing is optimal in the sense of minimax variance and also in the sense of the minimum possible number of allocations in the same or shorter interval.

This result is directly applicable for n divisible by 3. For large n, not divisible by 3, we may use the approximating integral values without much error. But for small n, not divisible by 3, a fine structure study, using $\frac{3\sigma^2}{n}$ as basis for comparison will indicate an acceptable spacing with little increase in variance.

6.6 Spacing of observations (Minimax variance)

For we have discussed only the possibility of a distribution of n observations on a polynomial regression of degree p at (p + 1) points for which the information matrix (determined by $\underline{X}' V_x^{-1} \underline{X}$)

were the same. In the example above we considered how these $(p + 1)$ points should be spaced and the observations allocated so that the maximum variance of the fitted value in the range of interpolation should be minimized in the particular case of quadratic regression. Now we shall consider the distribution of the points of observations in the minimax variance case for the general polynomial regression of degree p.

In this approach we shall use the simple regression polynomial curve to estimate the ordinate of the theoretical regression polynomial curve and find these values of $r_0, r_1, \cdots r_p$ that minimize the maximum variance of a single estimated ordinate.

Spacing for minimax variance.

The fitted value is given by $\mathbb{P}(r) = \sum_{j=1}^{p} F(r, r_j)\, \mathbb{P}(r_j)$ where $F(r, r_j)$ is the Lagrangean coefficient corresponding to the point of observation r_j and $\mathbb{P}(r_j) = y_j^*$, the mean of the observed values at this point. Then

$$\text{var}\, \mathbb{P}(r_j) = \sum_{j=0}^{p} F^2(r, r_j)\, \text{var}\,(y_j^*)$$

at a point of observation.

We know that the largest value of var (y_j^*) will be minimized when n observations are equally divided among the p + 1 points.

When this is done

$$\text{var}\, \mathbb{P}(r_j) = \text{var}\,(y_j^*) = (p + 1)\frac{\sigma^2}{n}$$

and

$$\text{var}\, \mathbb{P}(r) = \sum_{j=0}^{p} F^2(r, r_j)\,(p + 1)\frac{\sigma^2}{n}$$

Since this is a polynomial of degree , var $\mathbb{P}(r) \leq (p + 1)\frac{\sigma^2}{n}$ in the range -1 to +1 i.e. the minimax variance - conditions are obtained

when the maxima of var $\hat{P}(r)$ are at the $p-1$ internal points r_j, the end points x_0 and x_p being +1 and −1.

The minimax variance conditions are thus

$$F'(r_j) = 0 \quad , \quad j = 1, \ldots, p-1. \qquad (6.16)$$

Now if

$$L(r) = \prod_{j=0}^{p} (r - r_j)$$

then

$$F(r, r_j) = \frac{L(r)}{(r - r_j)} \frac{1}{L'(r_j)}$$

and so

$$(r - r_j) F(r, r_j) = L(r) / L'(r_j)$$

$$\therefore \quad (r - r_j) F''(r, r_j) + 2 F'(r, r_j) = L''(r) / L'(r_j)$$

$$L''(r) = \{ (r - r_j) F''(r, r_j) + 2 F'(r, r_j) \} L'(r_j)$$

The minimax variance condition (6.16) is thus equivalent to

$$L''(r_j) = 0 \quad , \quad j = 1, \ldots, p-1 \qquad (6.17)$$

The function $L(r)$ will be of the form $\alpha (r^2 - 1) \phi_{p-1}(r)$ where the polynomial $\phi_{p-1}(r)$ of degree $p-1$ is determined by the $(p-1)$ equations (6.17). The polynomial which satisfies these equations may be verified to be the derivative $P_p'(r)$ of the Legendre polynomial using the Legendre differential equation,

$$(r^2 - 1) \frac{d^2 y}{dr^2} + 2r \frac{dy}{dr} - p(p+1) y = 0$$

which is equivalent to

$$\frac{d}{dr} \left\{ (r^2 - 1) \frac{dy}{dr} \right\} = p(p+1) y$$

Now consider

$$L(r) = \alpha (r^2 - 1) P_p'(r)$$

then

$$L'(\tau) = \propto \frac{d}{d\tau}\left\{(\tau^2 - 1) \quad P_p'(\tau)\right\}$$

$$= \propto p(p+1) P_p(\tau)$$

and $$L''(\tau) = \propto p(p+1) P_p'(\tau)$$

and so $L''(\tau)$ vanishes at the interval points $L(\tau) = 0$. The points of observation for minimax variance are then to be located at +1, -1 and the roots of $P_p'(\tau)$.

6.7 Estimation Problem (minimum generalized variance).

When a number of parameters are to be estimated simultaneously the volume of the ellipsoid of concentration of the estimates is often used as a measure of the efficiency of the estimates. Since the square of the volume of the ellipsoid of concentration is proportional) to the generalized variance of the estimates one can see the justification of the use of G.V. as a measure of efficiency. G.V. of the L.S. estimates $\hat{\beta}$ has been found to be $\left|\underline{X}' \sum_{\epsilon}^{-1} \underline{X}\right|^{-1}$. The advantages of these L.S. estimates lies in the fact that among all linear estimates of the , the estimates given by this formula possess a minimum generalized variance. Thus if one restricts himself to linear estimates these are optimum estimates in the sense of minimum generalized variance. Now we have shown that the same information matrix, and hence the same value of G.V. can be obtained by replacing a given set of observations at the points $x_1, \ldots\ldots, x_n$ by a total of n observations made at p +1 properly selected points in the interval from X_1 to x_n. If the number of observations made at τ_i is n_i, then $\sum_{i=1}^{p+1} n_i = n$.

In terms of these substitute observations, the matrices in the G.V. are all square matrices and therefore the determinant of these products can be obtained by taking the product of these determinants. As a result ,

$$\frac{1}{G.V} = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ r_1 & r_2 & \cdots & r_{p+1} \\ \vdots & & & \\ r_1^p & r_2^p & \cdots & r_{p+1}^p \end{vmatrix} \begin{vmatrix} \frac{n_1}{\sigma^2} & & & \\ & \frac{n_2}{\sigma^2} & & \\ & & \ddots & \\ & & & \frac{n_{p+1}}{\sigma^2} \end{vmatrix} \begin{vmatrix} 1 & r_1 & \cdots & r_1^p \\ 1 & r_2 & \cdots & r_2^p \\ \vdots & & & \\ 1 & r_{p+1} & \cdots & r_{p+1}^p \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 1 & \cdots & 1 \\ r_1 & r_2 & \cdots & r_{p+1} \\ \vdots & & & \\ r_1^p & r_2^p & \cdots & r_{p+1}^p \end{vmatrix}^2 \prod_{i=1}^{p+1} \frac{n_i}{\sigma^{2p+2}}$$

$$= \frac{1}{\sigma^{2p+2}} \prod_{i<j}^{p+1} (r_i - r_j)^2 \prod_{i=1}^{p+1} n_i$$

Since $\prod_{i=1}^{p+1} n_i$, subject to the condition $\sum_{i=1}^{p+1} n_i = n$ is maximized when $n_1 = n = \cdots = n_{p+1} = \frac{n}{p+1}$, it follows that the generalized variance will be minimized for a fixed set of values when the same number of observations is taken at each of the values. This allocation assumes that n will be chosen to be a multiple of (p + 1).

Next we shall consider the maximization of $\prod_{i<j} (r_i - r_j)^2$, subject to the restriction that $x_1 \leq r_i \leq x_n$ (= 1, 2.....p + 1). We have shown already that x can be transformed linearly so that this

restriction assumes the form $-1 \leq t_i \leq 1$, $i = 1, 2.....p + 1$, then it can be shown that the set of t values that maximize $\pi (t_i - t_j)^2$ is given by the zeros of a polynomial $(1- t^2) P'_p (t)$ in which $P'_p (t)$ is the derivative of one of the Legendre polynomials as follows. These zeros may be obtained from the proper tables.

6.7.1 To show that the values of t's, $-1 \leq t \leq 1$, $i = 1, 2.....p + 1$ for which the function $\Delta = \prod_{i<j} (t_i - t_j)^2$ is a maximum are the zeros of the polynomial $P_{p+1} (t) = (1-t^2) P'_p (t)$, $P'_p (t)$ being the derivative of a Legendre polynomial of degree p.

Proof: Δ is a continuous function, and the region $-1 \leq t_i \leq +1$ defines a closed region G; hence there exist in G certain points at which Δ attains its maximum value.

At any such point the t_i $(i = 1,.....p.+ 1)$ are all distinct, and since Δ is symmetrical we can assume

$$t_1 < t_2 < \cdots < t_{p+1}$$

For $p = 1$, $\Delta \leq 4$ and equality is attained only when $x_1 = -1$, $x_2 = 1$.

For $p > 1$, we have for $1 < i < p + 1$

$$-1 < t_i < +1$$

Also

$$\frac{1}{\Delta} \frac{\partial \Delta}{\partial t_i} = 2 \sum^{p+1}_{*} \frac{1}{(t_i - t_j)} = 0 \quad (i \neq j)$$

Consider

$$f(t) = (t - t_1) (t - t_2) \cdots \cdots (t - t_{p+1})$$

$$f_i(t) = \frac{f(t)}{(t-t_i)}$$

then

$$\sum_{j=1}^{p+1} \frac{1}{(t-t_j)} = \frac{f_i'(t)}{f_i(t)} = 0 \quad \text{for } t = t_i \qquad \ldots \quad (6.18)$$

From $f(t) = (t-t_i) f_i(t)$

$$f''(t) = (t-t_i) f_i''(\ ) + 2 f_i'(t)$$

$$f''(t) = 2 f_i'(t) = 0 \quad \text{by virtue of } (6.18)$$

And since $f''(t)$ possesses exactly $(p-1)$ zeros, it follows that $t_1 = -1$, $t_{p+1} = 1$.

Hence $(t^2-1)\ f''(t)$ is devisible by $f(t)$, and (considering the degree of polynomials)

$$(t^2-1) f''(t) = C f(t), \qquad C, \text{ being a constant } C$$

may be determined by comparing leading coefficients on both sides. Thus

$$(t^2-1) f''(t) = (p+1)p\, f(t) \qquad (6.19)$$

We can see that $f(t) = \alpha (t^2-1) P_p'(t)$ satisfies the above equation (6.19), where $P_p(t)$ is the Legendre polynomial of degree p.

It can be clearly seen from inspecting the function $\prod (t_i - t_j)^2$ that the end points of the interval will always be chosen as two of the t values. It is also clear that the greater the range of n values the smaller will be the generalized variance.

In view of the results we have obtained so far; it follows that the optimum linear estimates of the coefficients of polynomial regression are obtained by using the L.S. estimates, choosing as a large range of n

values as possible, taking observations at the p + 1 points in this range given by means of the zeros of a polynomial (derivative of a Legendre polynomial) and the end points.

The preceding optimum manner of choosing n values assume that the G.V. of the estimates $\hat{\underline{\beta}}$ is the proper measure of the efficiency to use. Now shall study the situation when the sample regression curve is used for estimating the ordinates of the theoretical regression curve as we have done before in the minimax variance case and develop a measure of efficiency based on the variances and covariances of such estimated values.

From this point of view, let $r_1, \ldots r_{p+1}$ denote arbitrary points chosen in the given interval. Further, let y_i and $\hat{y}_i$ denote the ordinate and its estimate, of the polynomial regression curve at r_i.

Thus

$$y_i = \beta_0 + \beta_1 r_i + \cdots + \beta_p r_i^p, \quad i = 1, 2 \ldots\ldots p+1$$

or $$\underline{y} = \underline{R}\,\underline{\beta}$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 r_i + \cdots + \hat{\beta}_p r_i^p$$

or $$\hat{\underline{y}} = \underline{R}\,\hat{\underline{\beta}}$$

$$\text{var}\,(\underline{y}) = E\,(\hat{\underline{y}} - \underline{y})(\hat{\underline{y}} - \underline{y})' = E\,(\underline{R}\hat{\underline{\beta}} - \underline{R}\underline{\beta})\,(\underline{R}\hat{\underline{\beta}} - \underline{R}\underline{\beta})'$$

$$= \underline{R}\,[E\,(\hat{\underline{\beta}} - \underline{\beta})(\hat{\underline{\beta}} - \underline{\beta})]\,\underline{R}'$$

$$= \underline{R}\ \text{var}\ (\hat{\underline{\beta}})\ \underline{R}'$$

$$\text{G.V. of } \hat{\underline{y}} = |\text{var}\,(\hat{\underline{y}})| = |\underline{R}\ \text{var}\,(\hat{\underline{\beta}})\ \underline{R}'|$$

$$= |\underline{R}|\,|\text{var}\,(\hat{\underline{\beta}})|\,|\underline{R}'|$$

$$= \text{G.V.}\ (\hat{\underline{\beta}})\ \cdot\ |\underline{R}|^2.$$

$$= G.V.\ (\hat{\beta}) \prod_{i<j}^{p+1} (r_i - r_j)^2 .$$

This result shows that the generalized variance of the estimates of the ordinates of a polynomial regression curve at p + 1 arbitrary points will be minimized when the generalized variance of the estimates of the coefficients of the polynomial regression curve is minimized.

But we have seen that the set of values, $r_1, \ldots\ldots r_p$ which give the minimax variance estimates of a single ordinate is given by means of the zeros of the derivatives of the Legendre polynomial. It has been found that this set of values is the same set which minimizes the generalized variance. Thus, we have found that in polynomial regression, whether one is interested in efficient estimation of the regression coefficients, or in efficient ordinate estimation, either at one point or (p + 1) points the optimum choice is the same.

Chapter VII

Optimum Allocation Theory

Introduction: If for the estimation of the parameters $\{\beta_1, \beta_2, \cdots \beta_p\}$ or functions of them, different observations (sources) of form

$$y_i = \sum_{j=1}^{p} x_{ij} \beta_j + e_i \qquad (i = 1, 2, \cdots r)$$

are potentially available, each of them repeatable as many times as we please, the question arises which of them the experimenter should utilize and in what proportions. With appropriate optimality criteria we shall try and solve this problem of optimum allocation in the following section, first for the estimation of a single parametric function and second for the simultaneous estimation of all the parameters.

7.1 Estimation of a single parameter and a single parametric function of the form $\psi = \underline{\lambda}' \underline{\beta}$.

Suppose we want to determine each of the unknown parameters or a parametric form $\psi = \underline{\lambda}' \underline{\beta}$, based on a certain number r of different potential observations, the outcome of which are of the form

$$y_i = \sum_{j=1}^{p} x_{ij} \beta_j + e_i \qquad (i = 1, 2 \cdots r) \qquad (7.1)$$

where $\{x_{ij}\}$ are the known coefficients and

e_i is a random variable with mean zero and variance σ^2.

We shall assume furthermore, that the experimenter may perform each of the observations as many times as possible, or not at all, all actual observations being uncorrelated. If he has decided upon a certain total n of actual observations, he is faced with the problem which of the potential ones should be performed and in what proportion.

In order to distinguish between the potential and the actual observations, we shall refer to the former as sources (of information) and to the latter as observations. Since a source as well as an observation are described by the coefficient vector $(x_{i1}, x_{i2}, \cdots x_{ip})' = \underline{x}_i$ we will briefly speak of the source $\underline{x}_i'$ and the observation $\underline{x}_i'$ without any ambiguity.

The following normalization and idealization of our problem is mathematically convenient:

Let the experiment $\mathcal{E}$ consists of the observations $\underline{x}_1', \underline{x}_2', \cdots \underline{x}_r'$, repeated independently $n_1, n_2, \cdots n_r$, times respectively, with $\sum n_i = n$ i.e. the required number of observations on the source $\underline{x}_i'$ is n_i. If we put $n_i = n p_i$, then the p_i's are obviously multiples of $\frac{1}{n}$ fulfilling the conditions

$$p_i \geq 0 \quad , \quad \sum p_i = 1 \qquad (7.2)$$

We shall refer to $\underline{x} = (\underline{x}_1, \cdots \underline{x}_r)'$ as the spectrum and to $\underline{p} = (p_1, \cdots p_r)'$ as the allocation of $\mathcal{E}$.

Now the mean of the observations on the source $\underline{x}_i'$ will satisfy the model

$$y_i^* = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ip}\beta_p + \frac{e_i^*}{\sqrt{p_i}} \qquad (i = 1, 2 \cdots r) \qquad (7.3)$$

where e_i^* has variance σ^2/n. If a certain p_i is zero, the corresponding equation has to be left out of the system. For large n, the p_i's may be varied practically continuously over the range (7.2). Idealizing this feature we get a large-sample problem, which is essentially independent of σ and n. For simplicity, we may finally

assume $\frac{\sigma^2}{n} = 1$; in order to restore full generallity we have only to reintroduce this factor in all variance and covariance formula below. We now finally arrive at the normalized form of the problem.

Given a planned set of observations of the form

$$y_i^* = \underline{x}_i' \underline{\beta} + \frac{e_i^*}{\sqrt{p_i}} \quad \text{or} \quad \underline{y}^* = \underline{X}^{r \times p} \underline{\beta}^{p \times 1} + \frac{\underline{e}^*}{\sqrt{\underline{p}}}$$

$$\text{with } E(e_i^*) = 0 \text{ and } \operatorname{var}(e_i^*) = 1$$

and with the weights p_i at our disposal, subject to

$$\sum p_i = 1 \quad \text{and} \quad p_i \geq 0 .$$

What are the optimal p_i's ?

The solution of the problem obviously presupposes a specification of the word "optimal". Depending on the purpose of the experiment, various criteria for goodness may be used. We shall now discuss several such criteria all of them, however, based on the variance of one or more least-squares estimators, the desirable properties of which we have studied already (Gauss-Markoff Theorem). An allocation (or design) is considered optimal if it minimizes the variance of the least-squares estimators $\hat{\psi}$ of a parametric function ψ .

7.1.1 The Estimation Problem.

In this section we shall deal with the case where the interest of the experiment is centered upon a linear combination of the parameters.

$$\psi = \underline{\lambda}' \underline{\beta}$$

Particularly we may have $\psi = \beta_i$, $i = 1, 2, \ldots p$.

Now consider linear unbiased estimate $\underline{c}' \underline{y}^*$ of ψ ,

so that $E(\underline{c}' \underline{y}^*) = \psi = \underline{\lambda}' \underline{\beta}$ identically in $\underline{\beta}$.

i.e. $\underline{c}' E(\underline{y}^*) = \underline{\lambda}' \underline{\beta}$

$$\underline{c}' \underline{X} \underline{\beta} = \underline{\lambda}' \underline{\beta}$$

i.e. $\underline{c}' \underline{X} = \underline{\lambda}'$ or $\sum_{i=1}^{r} c_i \underline{x}_i' = \underline{\lambda}'$

Since $\underline{X}^{r \times p}$ is full rank, if $r > p$, there are infinitely many vectors $\underline{c}$ satisfying this condition. Among the corresponding estimates $\underline{c}' \underline{y}^*$, for any fixed set of p_i's, one with least variance is obtained using Gauss-Markoff Theorem, by substituting the least-squares estimate $\hat{\underline{\beta}}$ of the parameters in the expression for ψ.

The value of $\hat{\underline{\beta}}$ is the one which minimizes the weighted error sum squares

$$\sum p_i (y_i^* - x_{i1} \beta_1 - x_{i2} \beta_2 - \cdots - x_{ip} \beta_p)^2$$

The value of this estimate $\hat{\psi}$ is of course a function of $p_1, p_2, \cdots p_r$. We would like to find those weights p_i which yield the smallest minimum variance.

For this purpose we first notice that the smallest minimum variance by definition equals

$$\min_{p} \min_{c} \operatorname{Var} \{ \underline{c}' \underline{y}^* \}, \quad \underline{c} \text{ and } \underline{p} \text{ being subject to}$$

$$\underline{c}' \underline{X} = \underline{\lambda}'$$

and $$\sum p_i = 1.$$

Changing the order of minimization, we may calculate, to begin with, the minimum with respect to the p_i's for a fixed $\underline{c}$.

Consider $\min_{p} \operatorname{Var} \{ \underline{c}' \underline{y}^* \} = \min_{p} \underline{c}' \{ \operatorname{Var} \underline{y}^* \} \underline{c}$

$$= \min_{p} \sum_{i=1}^{r} \frac{c_i^2}{p_i} \qquad \text{subject to } \sum p_i = 1.$$

From the Lagrangian function,

$$L = \sum_{i=1}^{r} \frac{c_i^2}{p_i} + \ell (\textstyle\sum p_i - 1),$$

$$\frac{\partial L}{\partial p_i} = -\frac{c_i^2}{p_i^2} + l = 0 \qquad i = 1, 2, \ldots r.$$

$$p_i^2 = c_i^2 / l$$

$$p_i = |c_i| / \sqrt{l} \qquad (\because p_i \geq 0)$$

From $\sum p_i = 1$ follows that $\sum \frac{|c_i|}{\sqrt{l}} = 1$

or $\sqrt{l} = \sum |c_i|$

$$p = |c_i| / \sum |c_i|$$

Thus $\sum \frac{c_i^2}{p_i} = \left(\sum |c_i| \right)^2$

Therefore $\min_{p} \text{var} \left\{ \underline{c}' \underline{y}^* \right\} = \min_{p} \sum \frac{c_i^2}{p_i} = k_c^2$ where $k_c = \sum |c_i|$ (7.4)

The minimizing p_i values are $p_i = |c_i| / \sum |c_i|$.

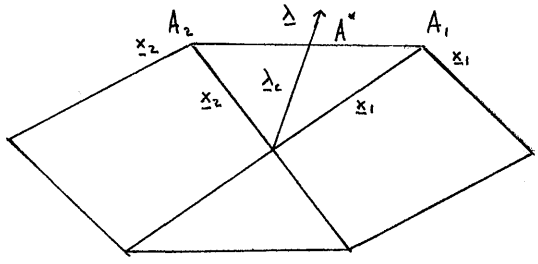
Now in order to minimize $\left\{ \min \text{var} \left(\underline{c}' \underline{y}^* \right) \right\}$ with respect to the c_i's,

Subject to the condition

$$\underline{c}' \underline{X} = \underline{\lambda}' \qquad \text{which we may write in the form}$$

$$\underline{\lambda}' = \sum_{i=1}^{r} c_i \underline{x}_i' = k_c \sum p_{c_i} \operatorname{sign} c_i \, \underline{x}_i' = k_c \underline{\lambda}_c' \text{ (say)} \qquad (7.5)$$

We may proceed to find the solution of the problem using convex sets.



The factor k, being a positive scalor, the vectors $\underline{\lambda}_c$ and $\underline{\lambda}$ are in the same direction. The weights p_{c_i} being nonnegative, with the sum one, it can be clearly seen that the end point of the vectors $\underline{\lambda}_c$ lies on or within the convex polyhedron spanned by the vectors $\pm \underline{x}_1$, $\pm \underline{x}_2$ $\ldots$ $\pm \underline{x}_r$. Since by (7.5) k_c is the ratio of the lengths of the vectors $\underline{\lambda}$ and $\underline{\lambda}_c$, it is obvious that (7.4) reaches its minimum when

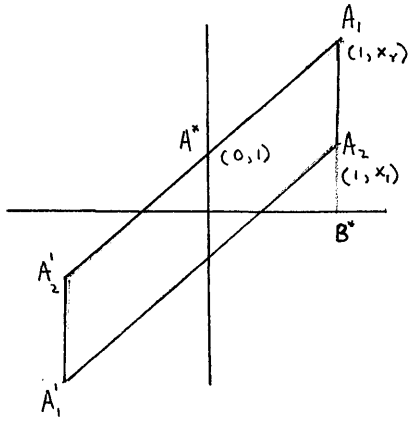
the endpoint $\underline{\lambda}_c$ coincides with the intersection A^* of the vector $\underline{\lambda}$ (or its extension) with the polyhedron in the p-dimensional space.

Illustrative example.

Consider the linear regression equation of y and x, where the potential observations are of the form

$$y = \alpha + x_i\beta + e_i, \qquad x_1 < x_2 < \cdots < x_r.$$

Here the polygon is a parallelogram spanned by the vectors $(1, x_1)$, ... $(1, x_r)$ and their opposite vectors. In this type of problem we are usually interested in the estimation of a single parameter.



Estimation of β alone.

Here $\underline{\lambda} = (0, 1)$ and $A^* = (0, 1)$ as shown in the figure and the ratio of the optimum weights is given by $\frac{A_1A^*}{A_2'A^*} = 1$.

This means that, if the experimentor is interested in the estimation of β alone, it is seen that he has to use only the extreme sources x_1 and x_r and observations of them have to be equal in number.

Estimation of α alone.

Here $\underline{\lambda} = (1, 0)$ and $A^* = B^*$ of the figure and the ratio of the optimum weights is given by $\frac{A_2B^*}{A_1'B^*} = \frac{x_1}{x_r}$. i.e. if α alone is to be estimated and if all x_i's have the same sign, the extreme one should again be used, this time in proportion $x_r : x_1$. If the x_i's include both positive and negative numbers, then the values of the p's are arbitrary

with the sole condition that the weighted average $\sum p_i x_i = 0$. Since in practice the p's have to be multiples of $1/n$, n being the number of observations, it is usually impossible to arrange the p's so that the condition mentioned is exactly fulfilled. In such circumstances, one can still make a useful choice between different approximations.

If there were three parameters in the model, we can still solve the problem geometrically. Since we can visualize the convex polyhedron with triangle side planes. For any estimation problem concerned with a single linear combination of the parameters we shall in general need three relevant sources.

In p parameters model, for any estimation problem concerned with a linear combination of the parameters, there will in general be p relevant sources. For more than two parameters we cannot solve the problem geometrically and algebraic procedure, which is in the form of programming must replace this geometrical method.

Now we may put the results of the last section in the form of a theorem.

The method of performing an experiment to estimate a parametric function $\psi = \underline{\lambda}'\underline{\beta}$ so as to minimize the variance $= \text{var}(\hat{\psi}) = \underline{\lambda}' \text{var}\,\hat{\underline{\beta}}\,\underline{\lambda}$ allowing, as possible experiments all convex combination of any finite number of sources $\underline{x}_1, \underline{x}_2, \ldots\ldots\ldots \underline{x}_r$ (which belongs to a symmetric closed, and bounded set R), depends on the following theorem.

Theorem: 7.1 In order that the allocation $(p_1, p_2, \ldots p_r)$ be optimal with regard to an experiment $(\underline{x}_1, \underline{x}_2, \ldots\ldots \underline{x}_r ; p_1, p_2, \ldots p_r)$, it is necessary and sufficient that the vector $\underline{\lambda}_i = \sum p_i \underline{x}_i$

be

(i) proportional to $\underline{\lambda}$

and (ii) a boundary vector of the convex hull $\bar{R}$ of R.

7.2. Estimation of Several parameters.

For a set of actual observations, i.e. for fixed $p_1, \cdots p_r$, the least-squares techniques yields minimum variance estimates of each of the parameters as well as all linear combinations of them; nothing is gained in the accuracy of some estimates by giving up accuracy in another. In the present setup, where the weights p_i are available, some information about the relative accuracy of different estimates is desirable. One natural way is to choose a nonnegative definite form in the estimation errors, say

$$\omega = (\hat{\underline{\beta}} - \underline{\beta})' \underline{L} (\hat{\underline{\beta}} - \underline{\beta}) = \sum_{i,j} l_{ij} (\hat{\beta}_i - \beta_i)(\hat{\beta}_j - \beta_j)$$

and minimize $E(\omega)$ with respect to the design (i.e. allocation).

If the rank of $\underline{L}$ is $s \leq p$, then $\underline{L}$ can be wrriten in the form

$$\underline{L} = \underline{l}_1 \underline{l}_1' + \cdots + \underline{l}_s \underline{l}_s'$$

where $\underline{l}_1, \underline{l}_2, \cdots \underline{l}_s$ are linearly independent p-vectors.

Using the linear transformation

$$\psi_j = \underline{l}_j' \underline{\beta} \qquad (j = 1, 2, \ldots s), \text{ we have}$$

$$E(\omega) = E \sum_{j=1}^{s} (\hat{\psi}_j - \psi_j)^2$$

$$= \sum_{j=1}^{s} \text{var}(\hat{\psi}_j) \qquad (7.6)$$

which is the sum of the variances of certain linear functions of the parameters. If s = 1 we are back in the situation of the previous section. The above consideration suggests that the min $\sum \text{var}(\hat{\psi}_j)$.

seems to be a reasonable optimality criteria for a design.

7.2.1 Two parameters case.

Under the assumption that the model is of full rank, the covariance matrix of $\hat{\beta}_1$, $\hat{\beta}_2$ in the model of the form (7.3) is the inverse of the information matrix

$$\underline{M} = \begin{vmatrix} \sum p_i x_{i1}^2 & \sum p_i x_{i1} x_{i2} \\ \sum p_i x_{i1} x_{i2} & \sum p_i x_{i2}^2 \end{vmatrix} = \sum p_i \underline{x}_i \underline{x}_i' .$$

To do the optimum allocation of observations in the sense of least variance sum, we need to minimize the quantity q_r, the trace of the inverse of this matrix $\underline{M}$ with respect to p 's satisfying (7.2)

Consider a point $\underline{p} = (p_1, \ldots, p_r)$ in (7.2) in which q_r reaches its minimum. If i and j are two relevant sources, that is, if $p_i > 0$, $p_j > 0$ any differential variation

$$dp_i = -\delta \qquad dp_j = \delta \qquad dp_h = 0 \quad (h \neq i,j)$$

of the coordinates leads to another point in (7.2). Now in order that $\underline{p}$ be a minimum point, we must have

$$\left(\frac{\partial q_r}{\partial p_j} - \frac{\partial q_r}{\partial p_i} \right) \delta \geq 0 \quad \text{for all } \delta .$$

i.e. we must have $\frac{\partial q_r}{\partial p_j} = \frac{\partial q_r}{\partial p_i}$ for any two relevant sources i and j.

If, on the other hand, i is relevant and j is irrevalent (i.e. $p_i > 0$, $p_j = 0$) than p_j can be varied in the positive direction, and we must, by the same argument as above, have

$$\left(\frac{\partial q_r}{\partial p_j} - \frac{\partial q_r}{\partial p_i} \right) \delta \geq 0 \quad \text{for any positive } \delta .$$

i.e. $\left(\frac{\partial q_r}{\partial p_j} \right)_{120} \geq \frac{\partial q_r}{\partial p_i}$

In conclusion, we come to know that corresponding to any solution of our minimization problem, there exists a constant $-\kappa^2$ such that

$$\frac{\partial q_r}{\partial p_i} = -\kappa^2 \qquad \text{for all relevant sources} \quad (7\cdot7)$$

$$\text{where as} \quad \frac{\partial q_r}{\partial p_i} \geq -\kappa^2 \qquad \text{for all irrelevant sources.}$$

As far as relevant sources are concerned, κ^2 is the ordinary Lagrange's multiplier in the minimization process.

But q_r is a homogeneous function of order -1, of $p_1, \ldots p_r$, for

$$\sum p_i \frac{\partial}{\partial p_i} \underline{M}^{-1} = -\sum p_i \underline{M}^{-1} \frac{\partial}{\partial p_i} \underline{M} \cdot \underline{M}^{-1} = -\underline{M}^{-1}$$

and therefore $\sum p_i \frac{\partial q_r}{\partial p_i} = -q_r$ (which is Euler's identity)

Also from (7·7) it follows that

$$\sum_{i=1}^{r} p_i \frac{\partial q_r}{\partial p_i} = -\kappa^2 \sum_{i=1}^{r} p_i = -\kappa^2.$$

We may now conclude from Euler's identity that κ^2 the minimum value of q_r. This also establishes the sign of κ^2 as positive, as already anticipated in the notation.

In order to find the value of $\frac{\partial q_r}{\partial p_i}$ we may use the relation

$$\frac{\partial}{\partial p_i} \underline{M}^{-1} = -\underline{M}^{-1} \frac{\partial \underline{M}}{\partial p_i} \underline{M}^{-1}$$

$$= -\underline{M}^{-1} \underline{x}_i \underline{x}_i' \underline{M}^{-1}$$

$$\therefore \quad \frac{\partial q_r}{\partial p_i} = \frac{\partial\, tr\, \underline{M}^{-1}}{\partial p_i} = -tr \{ (\underline{M}^{-1} \underline{x}_i \underline{x}_i' \underline{M}^{-1})$$

$$= -tr \{ (\underline{x}_i' \underline{M}^{-1} \underline{M}^{-1} \underline{x}_i) \}$$

$$= -tr \{ (\underline{M}^{-1} \underline{x}_i)' (\underline{M}^{-1} \underline{x}_i) \}$$

$$= -\| \underline{M}^{-1} \underline{x}_i \|^2 .$$

where $\|\underline{M}^{-1}\underline{x}_i\|$ denotes the length of the vector $\underline{M}^{-1}\underline{x}_i$.

Since $\|\underline{M}^{-1}\underline{x}\|^2$ is a positive definite quadratic form in the components of $\underline{x}$, the equation $\|\underline{M}^{-1}\underline{x}\|^2$ = constant represents an ellipse centred at the origin. Thus we finally arrive at the following result:

> "To any set $\{p_i\}$ that minimize the function q_v, the variance sum, there corresponds an ellipse, centred at the origin, such that all points $\underline{x}'_i$ representing relevant sources lie on the ellipse and none of the points representing irrelevant sources lie outside it".

Since, three points determine a conic centred at the origin, we can see that, in general, there are at most three relevant sources. Even in the case where four or more source points happen to lie on the same ellipse and the rest inside it, it may be shown that three relevant sources are enough for the minimization of q_v.

7.2.3 Generalization.

The preceeding arguments apply to an arbitrary s of parameters, the ellipse being replaced by an $(s - 1)$ - dimensional hyperellipsoid in R_s. Hence there will be at most $\frac{1}{2}s(s + 1)$ relevant sources.

7.3. Allocation problem of many observations experiment.

We shall next consider allocation problems associated with many observations experiments. As in the case of one observation experiment, each of the potential experiments (sources) may be repeated any number of times or not performed at all.

In the present situation each experiment consists in observing a vector variable $\underline{y}$ in which may be expressed as $\underline{X}\underline{\beta} + \underline{e}$. Thus the vector of observations in the ith experiment is

$$\underline{Y}_i = \underline{X}_i \underline{\beta} + \underline{e}_i \qquad (7.7)$$

or

$$\begin{pmatrix} y_i^{(1)} \\ y_i^{(2)} \\ \vdots \\ y_i^{(\nu_i)} \end{pmatrix} = \begin{pmatrix} x_{i1}^{(1)} & x_{i2}^{(1)} & \cdots & x_{ip}^{(1)} \\ x_{i1}^{(2)} & x_{i2}^{(2)} & \cdots & x_{ip}^{(2)} \\ \vdots & & & \vdots \\ x_{i1}^{(\nu_i)} & x_{i2}^{(\nu_i)} & \cdots & x_{ip}^{(\nu_i)} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} e_i^{(1)} \\ e_i^{(2)} \\ \vdots \\ e_i^{(\nu_i)} \end{pmatrix}$$

in which the coefficient $x_{ij}^{(\nu)}$ are known numbers.

The following table may help us to visualize the set-up pf potential as well as actual observations.

Experiment No.	Coefficient Matrix	Allocation	Replication No.	Actual Observations	Coefft. matrix
1	$\underline{X}_1$	n_1	1	(1) . . . (ν_1)	$\underline{X}$
			. .	. . .	
			n_1	(1) . . . (ν_1)	
. .	. . .	. .	. .	. . .	
r	$\underline{X}_r$	n_r	1	(1) . . . (ν_r)	
			. .	. . .	
			n_r	(1) . . . (ν_r)	

The horizontal matrices of the right-hand side of the table may be empty for one or more experiments viz. for those which are not performed at all. If an experiment consist of a single observation, the corresponding $\underline{X}$ matrix is a row vector. If, on the other hand, $n_i > 1$, we are concerned with the joint information. The matrix $\underline{X}$ is the coefficient of all observations actually carried out, it consists of n_1 matrices $\underline{X}_1$, n_2 matrices $\underline{X}_2$, etc.

For any given allocation, $\underline{X}$ is completely determined. $\underline{X}$ may be assumed to be nonsingular and thus all the theorems and results of full rank model apply.

If we denote

$$\underline{M}_i = \underline{X}_i' \underline{X}_i \quad , \quad \underline{M} = \frac{\underline{X}' \underline{X}}{n} \quad , \quad \text{then } \underline{M}^{-1} = (\underline{X}'\underline{X})^{-1} n.$$

Except for the factor $\frac{\sigma^2}{n}$, $\underline{M}^{-1}$ is the covariance matrix of $\hat{\beta}$.

Obviously

$$\underline{X}'\underline{X} = \sum n_i \underline{X}_i' \underline{X}_i$$

$$\underline{M} = \sum \frac{n_i}{n} \underline{X} \underline{X} = \sum p_i \underline{M}_i \quad \text{where } p_i = \frac{n_i}{n} \tag{7.8}$$

We shall consider the optimal allocation for the estimation of one or more of the parameters, say β_1, β_1 and β_2 etc., based on the criteria of minimum variance sum. i.e. allocation will be based on the quantities

$v_1 = m^{11}$, $v_{12} = m^{11} + m^{12}$ etc. where m^{ii} is the ith diagonal element of $\underline{M}^{-1}$.

Geometrical consideration.

As we have done before we shall drop the restriction that p_i must

vary through multiples of $\frac{1}{n}$. We may regard the φ's simply as functions of the elements m_{ij} of $\underline{M}$. This symmetric matrix is determined by $\frac{1}{2}p(p+1)$ elements; hence we may represent $\underline{M}$ by a vector in $\frac{1}{2}p(p+1)$ space, and similarly $\underline{M}_1$, $\underline{M}_2$, . . . , $\underline{M}_r$. According to (7.8) , the former vector varies over the convex polyhedron spanned by the latter vectors.

No matter whether we consider φ as a function of $p_1 \cdots p_r$ under the restriction $p_i \geq 0$, $\sum p_i = 1$ (which represents a barycentric simplex geometrically) or of m_{11} , m_{12} , . . . m_{pp} , the elements of $\underline{M}$, our aim is to minimize this function on a convex set.

As an illustration we shall consider the case $p = 2$, the situation being explained by fig. 7.1. Since the matrix $\underline{M}$ is always non-negative definite, and hence $m_{11} m_{22} - m_{12}^2 \geq 0$, $m_{11} \geq 0$, all M-vectors must lie on or within the rectangular cone V_0. Supposing there are five potential experiments, then the convex polyhedron C is spanned by five points $\underline{M}_i$, each representing one experiment.

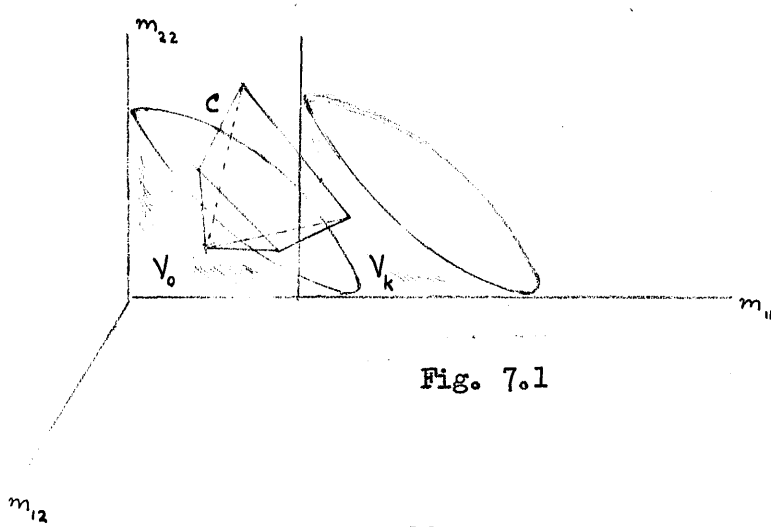


Fig. 7.1

If we are interested only in one parameter say β_1, we want to minimize q_{r_1}, (= var $\hat{\beta}_1$) i.e. to maximize

$$\frac{1}{q_{r_1}} = \frac{1}{m^{11}} = \frac{m_{11} m_{22} - m_{12}^2}{m_{22}} = k \qquad (7.9)$$

The level surfaces of this function have equations

$$(m_{11} - k)\, m_{22} - m_{12}^2 = 0$$

Thus, these surfaces are cones V_k obtained from V_0 by translation along the m_{11} axis. Hence it follows that (7.9) attains its maximum on C in the point T where the cone V_k, when moving towards the origin, has its first contact with C. This will clearly happen either in a corner or an edge of C (possibly at the same time in interior points of a side plane). Hence, we see that, in the case k = 2, only one or two different experiments will be required for an optimal design.

When we are interested in the estimation of both parameters the level surfaces of $q_{r_{12}}$ (which is relevant to the problem! turn out to be hyproboloids of two sheets, only one of which matters. In this case the minimum point may be either a corner, an edge point, or an interior point of a side plane. Accordingly, there will be required one, two or three experiments for an optimal design.

The above results are special cases of Elfving-Chirmoff's theorem (7.2)

"When s out of p parameters are to be estimated, the optimum allocation need comprise at most

$$p + (p - 1) + \cdots\cdots + (p - s + 1)$$

experiments".

The general result we infer in (7.2.3) is in fact the special

case of this theorem for p = s and s = s when all experiments consist of a single observation.

To throw more light on Chernoff-Elfving theorem, we shall consider the following simple proof, which does not cover the rather intricate case where $\underline{M}$ is singular. (Charnoff studied this case thoroughly). It may be noted that, if all $\underline{M}_i$ are nonsingular, then ϕ_r is continuous on C, and there is certainly a regular minimum point.

Consider an arbitrary differntiable function $\phi_r = \phi_r(M)$ which attains its minimum on C in a non-singular point M_0. We regard M as a function of $p_1, \ldots, p_s$ and denote by $\underline{p}^0 = (p_1^0, \ldots, p_s^0)$ aset of barycentric coordinates of the point M_0. As in the previous case one or more of the p_i^0 will be positive, the rest zero, the corresponding experiments being referred to as relevant or irrelevant respectively.

Then we know, at the minimum point, $\frac{\partial \phi_r}{\partial p_i}$ have a common value for all relevant experiments, and equal or larger value for the irrelevant ones.

By differentiating the matrix identity $\underline{M}^{-1}\underline{M} = \underline{I}$

$$\frac{\partial \underline{M}^{-1}}{\partial p_i} = -\underline{M}^{-1}\frac{\partial \underline{M}}{\partial p_i}\underline{M}^{-1} = -\underline{M}^{-1}\underline{M}_i\underline{M}^{-1}$$

In the one parameter case, the derivatives of $-\phi_1$, are

$$\frac{\partial \phi_1}{\partial p_i} = (\underline{M}^{-1}\underline{M}_i\underline{M}^{-1})_{11} = \sum_\alpha \sum_\beta m_{\alpha\beta}^{(i)} m^{1\alpha} m^{1\beta}.$$

$$= H_1^{(i)}(\underline{m}') \quad \text{(say)}.$$

where $m_{\alpha\beta}^{(i)}$ are the elements of $\underline{M}_i$ and $\underline{m}$ denotes the vector $(m^{11}, m^{12}, \ldots, m^{1p})'$.

Now, if $p_i^0 > 0$ for $i = 1, \ldots, x$, then we have by the preceeding arguments, $H_1^{(1)}(\underline{m}') = \ldots\ldots = H_1^{(x)}(\underline{m}')$.

These are (x - 1) relations between p - 1 variable, viz., the ratios of the m's and cannot in general be fulfilled unless x p. There are exceptional sets{Mi} admitting miminum points with more than p relevant experiments (the cone, in the case p = 2, being tangent to a side plane of the polyhedron C); but in such casesit is seen by a continuity argument that there also exists a minimum point with p (or fewer) positive p_i's.

Similarly, when two parameters are estimated we have

$$-\frac{\partial q_{12}}{\partial p_i} = (\bar{M}^{-1} M_i \bar{M}^{-1}) + (\bar{M}^{-1} M_i \bar{M}^{-1})_{22}$$
$$= H_{12}^{(i)} (m'', \ldots\ldots, m^{2p})$$

Since $m^{12} = m^{21}$, the number of variables in this quadratic form is p + (p - 1) and the theorem is thus found to hold for s = 2.

The conclusion may easily be extended to s = 2, ..., s = p.

7.4 Optimal selection problem in non-repeatable observations.

Another aspect of allocation problem is concerned with non-repeatable observations (very common in psychological experimentations). The problem may be formulated as follows:

"A finite set of potential observations is available each of them of the form".

$$y_i = \underline{x}_i' \underline{\beta} + e_i , \qquad i = 1, 2 \cdots n \quad (7.10)$$

when, the e's are uncorrelated error terms with zero expectation and (for convenience) variance 1 and $\underline{x}_i$'s are known coefficient vectors. The experimentor is allowed to perform m < n of the observations (7.10) each of them only once. The design problem, then, is to select a set

$\omega = (i_1, \quad i_m)$ from out of the subscript $i = 1, 2 \ldots, n$ so as to minimize the variance of the least-square estimator $\hat{\psi} = \underline{\lambda}' \hat{\underline{\beta}}$.

For any particular selection ω of subscripts, the estimatory properties of the corresponding set of observations are essentially described by the information matrix

$$\underline{M} = \sum_{\omega} \underline{x}_i \, \underline{x}_i' \qquad (7.11)$$

We shall assume that for any set ω of n observations, the matrix M is nonsingular. This will in general be the case when the coefficient vectors $\underline{x}$ are empirically determined. The variance of the least-square estimator $\hat{\psi}$ corresponding to is then, known to be

$$V_{\hat{\psi}} = \text{var} \quad (\hat{\psi}) \quad = \underline{\lambda}' \underline{M}^{-1} \underline{\lambda} \quad .$$

We shall now introduce the continuization device which greatly facilitates the solution of our selection problem, without, as we shall see later, impairing its practical applicability. For this purpose we shall consider, instead of the matrix (7.11) the generalized information matrix

$$\underline{M} = \sum_{i=1}^{n} p_i \, \underline{x}_i \, \underline{x}_i' \qquad (7.12)$$

where the allocation vector $\underline{p} = (p_1, \ldots, p_n)$ is subject to the restriction

$$0 \leq p_i \leq 1 \quad , \quad \sum_{i=1}^{n} p_i = m \qquad (7.13)$$

Obviously, the set of matrices (7.11) is a subset of the set of matrices (7.12) For a given p , we shall say that a source is totally selected, partially selected or excluded according as $p_i = 1, < 1$ or $= 0$

We shall see soon that the optimal allocation (when properly chosen) will contain at most p fractionally selected sources. Since the corresponding p's can total at most p-1, there will be from $m-p+1$ to m totally selected sources. If in practice, one may round off the fractional p_i's to 1's, i.e. select the corresponding sources on equal basis with the rest. Then at the expense of making at most $(p-1)$ more observations than originally planned, and will be sure to have a variance $V_{\hat{\psi}}$ not exceeding the smallest one that would be obtained for any selection of n,

Interpretation of the fractional p's:- We may imagine that the observation (7.11) be independently repeated each of them at most r times. The information matrix of the resulting experiment may then be written as $\underline{M} = r \sum p_i \underline{x}_i \underline{x}_i'$ when p_i varies from 0 to 1, through the multiple of $\frac{1}{r}$. The factor r in $\underline{M}$ is obviously irrelevant for the minimization problem. If, in particular, we take $m = 1$ and r very large, we will be concerned with the allocation problem of the repeatable observations we have discussed earlier; this problem then appears as a special case of the present one.

The following theorem is due to Elfving:

Theorem: 7.3 The variance of $\underline{\lambda}'\hat{\underline{\beta}} = V_{\hat{\psi}} = \underline{\lambda}'\underline{M}^{-1}\underline{\lambda}$ is minimum on the set

$$0 \leq p_i \leq 1, \qquad \sum_{i=1}^{n} p_i = m$$

if and only if the allocation satisfies the conditions

$$p_i \begin{cases} = 0 \\ \in [0,1] \\ = 1 \end{cases} \quad \text{when} \quad \begin{matrix} < \\ = \\ > \end{matrix} \quad h$$

where h is a certain non-negative constant.

The geometrical implication of this theorem is roughly that the "selection region" consists of that part of the p-space which lies outside the twin-hyperplans

$$\lambda' \underline{M}^{-1} \underline{x} = \pm h.$$

Proof: Since $M = \sum_{\omega} \underline{x}_i \underline{x}_i'$ is non-singular for any choice ω of m sources, under the assumption, it follows that $\underline{M} = \sum p_i \underline{x}_i \underline{x}_i'$ is nonsingular on (7·13).

Now if p_o denotes the smallest positive p_i, we may write

$$\underline{M} = p_o \sum_{p_i>0} \underline{x}_i \underline{x}_i' + \sum_{p_i>0} (p_i - p_o) \underline{x}_i \underline{x}_i'$$

The first form consists of at least m terms and, hence, is non-singular. This property is not destroyed by adjoining the second sum whose forms are non-negative definite. It follows that $V = \lambda' \underline{M}^{-1} \lambda$ is continuous on set (7·13) and hence attains its greatest lower bound .

Neccesity. Let $\underline{p}$ be the allocation vector for which the minimum of is attained and let i and j be two sources such that $p_i > 0$, $p_j < 1$, respectively.

Then, for any small enough δp , the variations

$dp_i = \delta p$ and $\delta p_j = \delta p$, $dp_k = 0$ $(k \neq i, j)$ is admissible with regard to (7·13). Since $\underline{p}$ was assumed to make V a minimum, the corresponding differential,

$$dV = \frac{\partial V}{\partial p_i} dp_i + \frac{\partial V}{\partial p_j} dp_j = \left(-\frac{\partial V}{\partial p_i} + \frac{\partial V}{\partial p_j} \right) \delta p$$

must be non-negative, and hence

$$-\frac{\partial V}{\partial p_i} \geq -\frac{\partial V}{\partial p_j} \qquad (p_i > 0,\ p_j < 1) \qquad (7\cdot14)$$

If $0 < p_i , p_j < 1$,i and j are interchangeable in (7·14) and it follows that $-\frac{\partial V}{\partial p_i} = h^2$ (say) for all fractionally selected items, if any (the non-negativity of $-\frac{\partial V}{\partial p_i}$ to be proved below) Applying (7·14) to a fractionally selected item, combined with a totally selected or excluded one, we have

$$-\frac{\partial V}{\partial p_i} \geq h^2 \quad \text{for the former}$$

$$\text{and} \qquad \leq h^2 \quad \text{for the latter type .}$$

Finally, if $\underline{p}$ contains no p_i with $0 < p_i < 1$, there will according to (7·14) exist a number h^2 (not uniquely determined) which is smaller than every $-\frac{\partial V}{\partial p_i}$ with $p_i = 1$ and larger than every $-\frac{\partial V}{\partial p_j}$ with $p_j = 0$. Thus in any case

$$p_i = \begin{cases} 0 \\ 1 \end{cases} \quad \text{whenever} \quad -\frac{\partial V}{\partial p_i} \lesseqgtr h^2 \qquad (7.15)$$

To find the value of $-\frac{\partial V}{\partial p_i}$:-

For a variation dp_i of p_i alone,

$$-dV = -\underline{\lambda}' d\underline{M}^{-1} \underline{\lambda} = \underline{\lambda}' \underline{M}^{-1} d\underline{M}\, \underline{M}^{-1} \underline{\lambda}$$

$$= \underline{\lambda}' \underline{M}^{-1} \underline{x}_i \underline{x}_i' \, dp_i \, \underline{M}^{-1} \underline{\lambda}$$

$$= \left| \underline{\lambda}' \underline{M}^{-1} \underline{x}_i \right|^2 dp_i$$

and therefore

$$-\frac{\partial V}{\partial p_i} = \left| \underline{\lambda}' \underline{M}^{-1} \underline{x}_i \right|^2 \geq 0 \qquad (7.16)$$

(7·15) and (7·16) together gives the necessary condition.

Sufficiency: Consider the allocation vector $\underline{p}^0$, together with a number $h \geq 0$ satisfying the conditions of the theorem. Let $\underline{p}'$ be another vector in (7·15) and consider the convex combination $\underline{p}(\theta) = (1-\theta)\,\underline{p}^0 + \theta\,\underline{p}'$ (also in (7·13)). Corresponding information matrix will then be

$$\underline{M}_\theta = \Sigma p_i(\theta)\, \underline{x}_i \underline{x}_i' = (1-\theta)\underline{M}_0 + \theta \underline{M}_1$$

and variances $\underline{V}(\theta) = \underline{\lambda}' \underline{M}_\theta^{-1} \underline{\lambda}$.

In order to prove that $\underline{V}(0)$, the variance corresponding to $\underline{p}^0$ is less than or equal to $\underline{V}(1)$, the variance corresponding to $\underline{p}^1$, we shall show that $\frac{\partial \underline{V}}{\partial \theta}\Big|_{\theta=0} \geq 0$, and $\frac{\partial^2 \underline{V}}{\partial \theta^2} \geq 0$ in $0 \leq \theta \leq 1$.

Now consider

$$-\frac{\partial \underline{V}}{\partial \theta} = \underline{\lambda}' \underline{M}_\theta^{-1} (\underline{M}_1 - \underline{M}_0) \underline{M}_\theta^{-1} \underline{\lambda} \qquad (7.17)$$

Thus

$$= \underline{\lambda}' \underline{M}_0^{-1} \Sigma (p_i^1 - p_i^0) \underline{x}_i \underline{x}_i' \underline{M}_0^{-1} \underline{\lambda}$$

$$= \Sigma (p_i^1 - p_i^0) \left| \underline{\lambda}' \underline{M}_0^{-1} \underline{x}_i \right|^2$$

Replacing the second factor in this sum by h^2 will, according to the conditions of the theorem , increase the sum ; because in those terms where $\left|\underline{\lambda}' \underline{M}_0^{-1} \underline{x}_i\right| > h$, the first factor will be $p_i^1 - 1 \leq 0$, and those terms where $\left|\underline{\lambda}' \underline{M}_0^{-1} \underline{x}_i\right| \leq h$, the first factor will be $p_i^0 - 0 \geq 0$.

Hence $-\frac{\partial \underline{V}}{\partial \theta}\Big|_{\theta=0} \leq h^2 \Sigma (p_i^1 - p_i^0) = h^2 (m - m) = 0$.

or $\frac{\partial \underline{V}}{\partial \theta}\Big|_{\theta=0} \geq 0$.

Differentiating (7.17) once more , we have

$$\frac{\partial^2 \underline{V}}{\partial \theta^2} = 2\, \underline{\lambda}' \underline{M}_\theta^{-1} (\underline{M}_1 - \underline{M}_0) \underline{M}_\theta^{-1} (\underline{M}_1 - \underline{M}_0) \underline{M}_\theta^{-1} \underline{\lambda}$$

$$= 2 \left[(\underline{M}_1 - \underline{M}_0) \underline{M}_\theta^{-1} \underline{\lambda} \right]' \underline{M}_\theta^{-1} \left[(\underline{M}_1 - \underline{M}_0) \underline{M}_\theta^{-1} \underline{\lambda} \right] .$$

Since $\underline{M}_\theta$, and hence $\underline{M}_\theta^{-1}$, is positive definite for $0 \leq \theta \leq 1$, it follows that $\frac{\partial^2 \underline{V}}{\partial \theta^2} \geq 0$.

7.5 Number of fractionally selected items .

Since in general the (p - 1)-dimensional boundary plane $\underline{\lambda}' \underline{M}^{-1} \underline{x} = h$ will contain at most p source points (opposite points $\pm \underline{x}_i$ being always counted as equivalent), and hence in general there will always be at most p

fractionally selected items. But it can happen that the boundary plane contains more than p source points, with the implication that there may well be optimal allocations with more than p fractional p_i's. However, the following theorem ensures that one can always find an equivalent allocation with at most p weights p_i between 0 and 1.

Theorem: 7.4 The function $V = \underline{\lambda}' \underline{M}^{-1} \underline{\lambda}$

when $\underline{M} = \sum_{i=1}^{n} p_i \underline{x}_i \underline{x}_i'$

and $0 \leq p_i \leq 1 \quad ; \quad \sum_{i=1}^{n} p_i = m$.

can always be minimized by means of an allocation vector $\underline{p}$ with at most p fractional p_i's.

Proof: Corresponding to the given set $\{\underline{x}_i\}$ of source points consider a sequence of sets $\{\underline{x}_i^j\}$ $(i = 1, \ldots, n)$ $(j = 1, \cdots)$ such that

(i) $\underline{x}_i^j \rightarrow \underline{x}_i$ as $j \rightarrow \infty$

and (ii) for each j, no p + 1 points $\underline{x}_i^j$ lie in the same (p - 1) dimensional hyperplane.

The sequence $\{\underline{x}_i^j\}$ may be constructed by choosing an arbitrary set $\{\underline{x}_i^*\}$ $(i = 1, \ldots, n)$ with property (ii), forming $\underline{x}_i(\lambda) = (1-\lambda)\underline{x}_i + \lambda \underline{x}_i^*$ and taking $\underline{x}_i^j = \underline{x}_i(\lambda_j)$ for a sequence $\lambda_1, \lambda_2, \cdots \rightarrow 0$.

In order for (i) to hold, we must choose λ_j so that all (p + 1) x (p + 1) determinants of type

$$\begin{vmatrix} x_{i_1 1}(\lambda) & \cdots & x_{i_1 p}(\lambda) & 1 \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ x_{i_{p+1} 1}(\lambda) & \cdots & x_{i_{p+1} p}(\lambda) & 1 \end{vmatrix} \qquad (7.18)$$

do not vanish for $\lambda = \lambda_j$. Now, each of these determinants is a polynomial in λ_j and cannot be identically zero, since in that case the set $\{ \underline{x}_i^* \}$ to which $\{ \underline{x}_i^j \}$ reduces for $\lambda = 1$ would not have property (ii) . To make sure that (ii) holds for all $\{ \underline{x}_i^j \}$, we only have to choose λ_j different from finitely many roots of all polynomial (7.18)

Now with each set $\{ \underline{x}_i^j \}$ $(j = 1, 2 \cdots)$ (λ and m being fixed once and for all) there exists an allocation vector $\underline{p} = (p_1^j , \quad p_n^j)$ in (7.13) and a corresponding information matrix $\underline{M}_j = \sum_i p_i^j \underline{x}_i^j \underline{x}_i^{j\prime}$ such that $\underline{\lambda}' \underline{M}_j^{-1} \underline{\lambda}$ is a minimum; moreover, there exists a number $h_j \geq 0$ such that the hyperplane $\underline{\lambda}' \underline{M}_j^{-1} \underline{\lambda} = \pm h_j$ separate the totally selected and the excluded source points, and contain all $\underline{x}_i^j$ for which $0 < p_i^j < 1$. But, by construction of the sequence $\{ \underline{x}_i^j \}$, there are at most p points $\underline{x}_i^j$ in any hyperplane for each j and hence at most p fractionally elected sources.

Since there are only finitely many ways to subdivide elements into three classes i.e. ω_+(totally selected), ω_0(fractionally selected) and ω_- (excluded), at least one such partition must repeat itself for infinitely many j's. We may now exclude all other j's, thus obtaining fixed sets ω_+ , ω_0 , ω_- of selected, excluded, and fractionally selected i's, the last group containing at most p elements.

Thus we have constructed a new subsequence (j) such that on this sequence $\underline{p}^j$ converge to a limiting allocation vector $\underline{p}$. This is still in (7.13), and the subscript of its fractional components, if any, still

belongs to ω_0 . If $\underline{M}$ denote the information matrix corresponding to $\underline{p}$, we have $\underline{M}_j \to \underline{M}$, also $\underline{M}^{-1}$ exists and $\underline{M}_j^{-1}\underline{\lambda} \to \underline{M}^{-1}\underline{\lambda}$,which implies that $\|\underline{M}_j^{-1}\underline{\lambda}\|$ is bounded.

Since $|\underline{x}_i^j|$ have a common bound and since $h_j \leq |\underline{\lambda}'\underline{M}_j\underline{x}_i|$ for some i , we conclude that the h_j are bounded . One can easily see that $h_j \to h$ follows from the selection procedure we have adopted.

For the sets $\{h_i^j\}$ constructed in this way we have, by the necessity part of the theorem

$$p_i^j = \begin{cases} 0 \\ 1 \end{cases} \quad \text{whenever } |\underline{\lambda}'\underline{M}\,\underline{x}_i^j| \lesssim h_j \qquad (7.19)$$

Meaning of the limiting p, h :

If for some subscript i , $|\underline{\lambda}'\underline{M}\,\underline{x}_i| > h$,then for some j on $|\underline{\lambda}'\underline{M}^{-1}\underline{x}_i^j| > h_j$, hence by (7.19) $p_i^j = 1$ and $p_i = \lim p_i^j = 1$; similarly, if for some i , $|\underline{\lambda}'\underline{M}\,\underline{x}_i| < h$,we conclude that $p_i = 0$.

The sufficiency part of the theorem ensures that p minimizes $\underline{\lambda}'\underline{M}^{-1}\underline{\lambda}$ on (7.13) .Since p has at most n fractional components , the proof of the theorem is complete.

Chapter VIII

Optimality Criteria & Optimality of Basic Designs.

Introduction:

Various classes of designs are now in use in experimental sciences of which very well known are randomized block, latin square, incomplete block designs of different type. In comparing the efficiencies of these designs the usual procedure is to relate some sort of an average of the variances associated with different estimates obtained from any particular design to a similar quantity for the corresponding randomized block design, the latter being considered as the most efficient (Yates, 1936). It seems to be, however, more desirable to start with a reasonable definition of efficiency of any experimental arrangement and to deduce the optimum character of randomized block, latin square and other design in terms of the optimality criteria so defined. It is from the latter point of view that Wald (1943) first proved the optimality of the latin square arrangement. We shall now study various optimality criteria and the optimum character of basic experimental designs on the basis of these criteria. After studying the desirable properties of the designs, we shall also derive sufficient conditions for designs to possess these optimum properties. Computation methods for the actual construction of optimal designs are considered in chapter VI & VII.

8.1 Definition of optimality

Wald's development of D-optimality;-

It is well known that the analysis of a design involve either tests

of linear hypothesis or estimation problems. So we shall first review some results we have obtained earlier in general linear hypothesis. Consider the general linear model of full rank. (without any loss of generality)

$$\underline{Y} = \underline{X}\,\underline{\beta} + \underline{e}$$

with $\underline{e}$ distributed $N(\underline{0}, \sigma^2 \underline{I})$

The p column vectors in $\underline{X}$ will be denoted by $\underline{\xi}_1, \underline{\xi}_2, \ldots \underline{\xi}_p$. (and n row vectors by $\underline{x}_1', \underline{x}_2', \ldots\ldots \underline{x}_n'$). In some cases the experimentor has some amount of freedom in the choice of p vectors in the n dimensional Euclidian space.

We shall suppose further that, for fixed α, $F_{d,\alpha}$ represents the critical value of the F test whatever design d (which correspond to the design matrix $\underline{X}$) is chosen.

Consider the partitioned model of full rank

$$\underline{Y} = \underline{X}_1 \underline{\gamma}_1 + \underline{X}_2 \underline{\gamma}_2 + \underline{e}$$

where $\underline{\gamma}_1$ is of dimension (rx1)

$\underline{e}$ may be assumed to be distributed $(\underline{0}, \sigma^2 \underline{I})$

Null hypothesis H: $\underline{\gamma}_1 = \underline{0}$

Then, as we have obtained earlier, the test statistic.

$$F = \frac{n-p}{r} \cdot \frac{\underline{y}' [\underline{X} \underline{M}^{-1} \underline{X}' - \underline{X}_2 (\underline{X}_2' \underline{X}_2)^{-1} \underline{X}_2'] \underline{y}}{\underline{y}' (\underline{I} - \underline{X}' \underline{M}^{-1} \underline{X}) \underline{y}}$$

is distributed as $F'(r, n-p, \delta)$, where

$$\delta = \frac{\underline{\gamma}_1' \underline{B} \underline{\gamma}_1}{2\sigma^2}$$

with $\underline{B} = \underline{X}_1' \underline{X}_1 - \underline{X}_1' \underline{X}_2 (\underline{X}_2' \underline{X}_2)^{-1} \underline{X}_2' \underline{X}_1$

$\underline{B}$ being positive definite, F is distributed as F(r,n-p) if and only

if H is true i.e. if and only if $\underline{\gamma}_1 = \underline{0}$.

It is known that the power function of the critical region of the design d of size α given by $F \geq F_{\alpha,d}$ is a monotonic increasing function of δ, which depends on $\underline{B}$, the elements of which are functions of of the vectors $\underline{\xi}_i$ of the design matrix $\underline{X}$.

Generally the conditions of the problem are such that the only factors under control are the values of $\underline{\xi}_i$'s and hence arises the problem of efficient designing i.e. choosing the $\underline{\xi}_i$'s so as to render the test procedure most sensitive to deviations from the null hypothesis H. Since the power of the F-test involved here is a monotonic increasing function of δ, it would be enough so to choose $\underline{\xi}_i$'s as to maximize δ for all values β_i $(i = 1,2, \ldots, r.)$.

To facilitate the argument simpler, we may write

$$\delta = \frac{1}{2\sigma^2} \sum_{i=1}^{r} \sum_{j=1}^{r} b_{ij} \beta_i \beta_j$$

where b_{ij} is the element of $\underline{B}$ (which is actually the inverse of the matrix $\underline{M}''$) where $\underline{M}''$ is made up of the elements of the first r rows and r columns of $\underline{M}^{-1}$.

When $r = 1$, the expression reduces to $\frac{1}{2\sigma^2} b_{11} \beta_1^2$.

Hence if $r = 1$, we may **maximize** δ by maximizing b_{11} or minimizing m'', the pivotal element of $\underline{M}^{-1}$.

From $\text{var}(\hat{\underline{\beta}}) = \underline{M}^{-1}\sigma^2$, we obtain $\text{var}(\hat{\beta}_0)$ is $\text{var}(\hat{\beta}_1)$. being the $(1,1)$th element of $\underline{M}^{-1}$. Thus when m'' is minimized so is $\text{var}(\hat{\beta}_1)$.

Now if $r = 1$, we can say that we obtain the powerfull test and the best estimate by minimizing m'' or we would equally well say that minimum

variance of $\hat{\beta}_1$ ensures the most powerful test.

But, when $r > 1$, the difficulty arises that no set of values $\underline{\xi}_i$ can be found for which δ becomes a maximum irrespective of the values of the unknown parameters.

Wald then suggested a compromise solution which is now known as E-optimality which Ehrenfield studied later thoroughly and developed a satisfactory working rule to find the sufficient conditions for this optimality. However Wald succeeded in developing a goodness criteria which is now known as D-optimality.

Wald first considered a unit hypersphere

$$\underline{\gamma}_1' \underline{\gamma}_1 = 1, \text{i.e.} \quad \beta_1^2 + \beta_2^2 + \cdots + \beta_r^2 = 1 \qquad (8.1)$$

in the space of the parameters $\beta_1, \cdots, \beta_r$ and mentioned the following well known result:

The smallest and the greatest characteristic roots of the equation

$$|\underline{B} - \lambda \underline{I}| = 0$$

are respectively the minimum and maximum values of $\sigma^2 \delta$ on the unit sphere (8.1). Though he mentioned the compromise solution of maximizing the smallest root of

$$|\underline{B} - \lambda \underline{I}| = 0$$

as a reasonable goodness criteria (which Erhenfeld exploited thoroughly later on), he approached the problem from a different angle. He proposed to maximize the product of r roots of the equation. Since the product of r roots of our equation is equal to the determinant $|\underline{B}|$ we have to maximize $|\underline{B}|$.

But $$|\underline{B}| = \frac{1}{|M_{11}''|}$$

Hence $\underline{B}$ will be maximized when $\underline{M}_{11}''$ is minimized. The generalized

variance of the set of variates $\hat{\beta}_1, \ldots, \hat{\beta}_r$ is equal to the product of σ^{2r} and the determinant $|M^{-1}|$.

Thus, the optimum choice of the values of ξ_i is that for which the generalized variance of $\hat{\beta}_i$ (i = 1, 2,.....r.) becomes a minimum.

Now we have the following definition of D-optimality.

Definition: - In a possible class of designs Δ, let $\Delta' \subset \Delta$ represents a subclass of designs for which ψ_i are estimable. Let V_d be the covariance matrix of best linear estimators of the ψ_i for a d in Δ'.

Then a design d is said to be D-optimum in Δ if $d^* \in \Delta'$ and

$$\det V_{d^*} = \min_{d \in \Delta'} \det V_d$$

8.1.2 Ehrenfeld's development of E-optimality.

We shall consider first an inequality which will lead to the proper understanding of E-optimality and later a theorem about quadratic forms which, the the aid of other considerations, will motivate a criteria for the optimality of a design.

A useful inequality. Suppose we have a real pxp symmetric semi-positive matrix $M = (X' X)$ of rank $r \leq p$. Let X be a column vector of p components, not all zero, such that $M\rho = \lambda$ has solutions for ρ. Then we have the following theorem.

Theorem 8.1 If ρ is any solution of $M\rho = \lambda$, then we have

$$\frac{1}{\kappa_{max}} \leq \frac{\rho' M \rho}{\lambda' \lambda} \leq \frac{1}{\kappa_{min}}$$

where κ_{max} and κ_{min} ar e the maximum and the minimum of

the nonzero characteristic roots of $\underline{M}$.

Proof: Since $\underline{M}$ is semipositive definite and of rank r there exists an orthogonal matrix $\underline{P}$ such that $\underline{P}' \underline{M} \underline{P} = (\kappa_i \delta_{ij})$, where $\kappa_i (i = 1, 2, \ldots r)$ are the nonzero characteristic roots of $\underline{M}$.

Therefore
$$\begin{aligned} \underline{\ell}' \underline{M} \underline{\ell} &= \underline{\ell}' \underline{P} \underline{P}' \underline{M} \underline{P} \underline{P}' \underline{\ell} \\ &= (\underline{P}' \underline{\ell})' (\underline{P}' \underline{M} \underline{P}) (\underline{P}' \underline{\ell}) \\ &= \underline{\ell}^{*\prime} (\underline{P}' \underline{M} \underline{P}) \underline{\ell}^{*} \\ &= \sum_{i=1}^{r} \kappa_i \ell_i^{*2} \end{aligned}$$

Since $\underline{M} \underline{\ell} = \underline{\lambda}$, we also have $\underline{\lambda}' \underline{\lambda} = \underline{\ell}' \underline{M}' \underline{M} \underline{\ell}$

$$\begin{aligned} &= \underline{\ell}' \underline{P} \underline{P}' \underline{M} \underline{P} \underline{P}' \underline{M} \underline{P} \underline{P}' \underline{\ell} \\ &= (\underline{P}' \underline{\ell})' (\underline{P}' \underline{M} \underline{P})^2 (\underline{P}' \underline{\ell}) \\ &= \underline{\ell}^{*\prime} (\underline{P}' \underline{M} \underline{P})^2 \underline{\ell}^{*} \\ &= \sum_{i=1}^{r} \kappa_i^2 \ell_i^{*2} \end{aligned}$$

Thus, putting $z_i = \sqrt{\kappa_i}\, \ell_i^{*}$, we have

$$\frac{\underline{\ell}' \underline{M} \underline{\ell}}{\underline{\lambda}' \underline{\lambda}} = \frac{\sum \kappa_i \ell_i^{*2}}{\sum \kappa_i^2 \ell_i^{*2}} = \frac{\sum z_i^2}{\sum \kappa_i z_i^2}$$

Since

$$\frac{1}{\kappa_{max}} \leqslant \frac{\sum z_i^2}{\sum \kappa_i z_i^2} \leqslant \frac{1}{\kappa_{min}}$$

we get the required relation

i.e.
$$\frac{1}{\kappa_{max}} \leqslant \frac{\underline{\ell}' \underline{M}' \underline{\ell}}{\underline{\lambda}' \underline{\lambda}} \leqslant \frac{1}{\kappa_{min}}$$

When $\underline{M}$ is of full rank, the above inequality becomes

$$\frac{1}{\kappa_{max}} \leqslant \frac{\underline{\lambda}' \underline{M}^{-1} \underline{\lambda}}{\underline{\lambda}' \underline{\lambda}} \leqslant \frac{1}{\kappa_{min}}$$

using $\underline{\ell} = \underline{M}^{-1} \underline{\lambda}$.

For the most part we shall restrict ourselves to the case where $\underline{M}$ has full rank i.e. $\underline{M}$ has an inverse and $\underline{\ell} = \underline{M}^{-1}\underline{\lambda}$. When $\underline{M}$ is not of full rank, we can either reparametrized the model so that the corresponding $\underline{M}$ is of full rank or proceed in a manner very similar to the case of full rank. Before we apply the above inequality, we shall reconsider the estimation problem and the power of the F-test of the hypothesis

H: $\beta_1 = \beta_2 = \cdots = \beta_r = 0$.

The best unbiased estimate of $\psi = \underline{\lambda}'\underline{\gamma}_1$, in the sense of minimum variance is provided by Gauss-Markoff theorem as $\hat{\psi} = \sum_{i=1}^{r} \lambda_i \hat{\beta}_i = \underline{\lambda}'\hat{\underline{\gamma}}_1$, where $\hat{\underline{\gamma}}_1$, is the Least Square estimate of $\underline{\gamma}_1$. Also it is known that the variance of $\hat{\psi}$ is $\sigma^2\underline{\lambda}'\underline{M}^{-1}\underline{\lambda}$. If $\underline{\lambda}$ is such that $\underline{M}\,\underline{\ell} = \underline{\lambda}$ has solution for $\underline{\ell}$, we know that $\psi = \underline{\lambda}'\underline{\gamma}_1$ is estimable. When $\underline{M}$ is of full rank, the variance of $\hat{\psi}$ is $\sigma^2\underline{\ell}'\underline{M}\,\underline{\ell}$ where $\underline{\ell}$ is the solution of $\underline{M}\,\underline{\ell} = \underline{\lambda}$. (In fact var $\hat{\psi} = \underline{\ell}'\underline{M}\,\underline{\ell}\,\sigma^2$ even when $\underline{M}$ is not full rank. Ref:Chapter IV.).

Next we shall consider the power function of the F-test associated with the null hypothesis H.

Consider now the similarly partitioned matrices of $\underline{M}$ and $\underline{M}^{-1}$,

$$\underline{M} = \begin{pmatrix} \underline{M}_{11} & \underline{M}_{12} \\ \underline{M}_{21} & \underline{M}_{22} \end{pmatrix} \qquad \text{with } \underline{M}_{12}' = \underline{M}_{21}$$

and

$$\underline{M}^{-1} = \begin{pmatrix} \underline{M}^{11} & \underline{M}^{12} \\ \underline{M}^{21} & \underline{M}^{22} \end{pmatrix} \qquad \text{with } \underline{M}^{12'} = \underline{M}^{21}$$

where $\underline{M}_{11}$ and $\underline{M}^{11}$ are rxr matrices.

The usual F-test for the null hypothesis $\beta_1 = \beta_2 = \cdots = \beta_r = 0$ or $\underline{\gamma}_1 = 0$ has a power function depending monotonically on δ, given by

$$\delta = \frac{\underline{\gamma}_1' \underline{B} \underline{\gamma}_1}{2\sigma^2} = \frac{\underline{\gamma}_1' (\underline{M}_1'')^{-1} \underline{\gamma}_1}{2\sigma^2} \quad \text{in our notation above.}$$

Since $(\underline{M}'')^{-1} = \underline{M}_{11} - \underline{M}_{12} \underline{M}_{22}^{-1} \underline{M}_{21} = (\underline{X}_1' \underline{X}_1 - \underline{X}_1' \underline{X}_2 (\underline{X}_2' \underline{X}_2)^{-1} \underline{X}_2' \underline{X}_1)$

$$\delta = \frac{\underline{\gamma}_1' \underline{M}_{11} \underline{\gamma}_1 - (\underline{M}_{21} \underline{\gamma}_1)' \underline{M}_{22}^{-1} (\underline{M}_{21} \underline{\gamma}_1)}{2\sigma^2}$$

Confidence interval for ψ with confidence coefficient $1-\alpha$ is

$$\hat{\psi} - k_\alpha \hat{\sigma}_\psi \leq \psi \leq \hat{\psi} + k_\alpha \hat{\sigma}_\psi$$

where k_α is an appropriate constant, being estimated by

$\hat{\sigma}_\psi^2 = \hat{\sigma}^2 \underline{\lambda}' \underline{M}^{-1} \underline{\lambda}$, where $\hat{\sigma}^2$ is the usual estimate of σ^2.

If we let L equal to the length of the confidence interval, we have $(E(L))^2 = A \sigma^2 \underline{\lambda}' \underline{M}^{-1} \underline{\lambda}$, where A is independent of σ^2.

Using the above inequality, we have the bounds for var $(\hat{\psi})$

Since $$1/\kappa_{max} \leq \underline{\lambda}' \underline{M}^{-1} \underline{\lambda} / \underline{\lambda}' \underline{\lambda} \leq 1/\kappa_{min}$$

$$1/\kappa_{max} \leq \sigma^2 \underline{\lambda}' \underline{M}^{-1} \underline{\lambda} / \sigma^2 \underline{\lambda}' \underline{\lambda} \leq 1/\kappa_{min}$$

or $$\sigma^2 \underline{\lambda}' \underline{\lambda} / \kappa_{max} \leq \text{var}(\hat{\psi}) \leq \sigma^2 \underline{\lambda}' \underline{\lambda} / \kappa_{min} \qquad (8.3)$$

Similarly we get the bounds for the expectation of the confidence interval

$$A \sigma^2 \underline{\lambda}' \underline{\lambda} / \kappa_{max} \leq (E(L))^2 \leq A \sigma^2 \underline{\lambda}' \underline{\lambda} / \kappa_{min} \qquad (8.4)$$

The noncentrality parameter of the power function of the F-test of the null hypothesis H is

$$\delta = \frac{\underline{\gamma}_1' \underline{B} \underline{\gamma}_1}{2\sigma^2}$$

Now consider $$\frac{\underline{\gamma}_1' \underline{B} \underline{\gamma}_1}{\underline{\gamma}_1' \underline{\gamma}_1} = \frac{\underline{\gamma}_1' \underline{P}' \underline{P} \underline{B} \underline{P}' \underline{P} \underline{\gamma}_1}{\underline{\gamma}_1' \underline{P}' \underline{P} \underline{\gamma}_1}$$ ($\underline{P}$ being an orthogonal matrix)

$$= \frac{\underline{\gamma}_1^{*\prime} (\kappa_i \delta_{ij}) \underline{\gamma}_1^*}{\underline{\gamma}_1^{*\prime} \underline{\gamma}_1^*} = \frac{\sum_{i=1}^{r} \kappa_i \beta_i^{*2}}{\sum_{i=1}^{r} \beta_i^{*2}}$$

such that κ_i are the characteristic roots of $\underline{B}$.

therefore $\min \varkappa_i \leq \dfrac{\sum_{i=1}^{r} \varkappa_i \beta_i^{*2}}{\sum \beta_i^{*2}} \leq \max \varkappa_i$

or $\min \varkappa_i \leq \dfrac{\underline{\gamma}_1' \underline{B} \underline{\gamma}_1}{\underline{\gamma}_1' \underline{\gamma}_1} \leq \max \varkappa_i$

Therefore $\dfrac{\underline{\gamma}_1' \underline{\gamma}_1}{2\sigma^2} \varkappa_{max} \geqslant \delta \geqslant \varkappa_{min} \dfrac{\underline{\gamma}_1' \underline{\gamma}_1}{2\sigma^2}$ (8.5)

It is to be noted that all these bounds can be attained.

In equalities (8.1), (8.2) and (8.3) suggest that it would be desirable to make $\varkappa_{min}$ as large as possible.

But $\underline{B} = V_d^{-1}$ where V_d is the covariance matrix of $\underline{\gamma}_1$.

Since $|\underline{B} - \varkappa \underline{I}| = 0$ is equivalent to

$$\left| V_d - \frac{1}{\varkappa} \underline{I} \right| = 0$$

the characterstic roots of V_d are the reciprocals of theose of $\underline{B}$.

Consequently maximizing $\varkappa_{min}(\underline{B})$ is the same as minimizing $\varkappa_{max}(V_d)$.

Now we shall formally define E optimality.

<u>Definition</u>:- Let $\Delta' \subset \Delta$ represents a subclass of designs for which are estimable in a class of possible designs and let V_d be the covariance matrix of the best linear estimates of the ψ_i for a design of d in Δ' .

Then a design d* is said to be E optimum in Δ if $d^* \in \Delta'$

and $\varkappa_{max}(V_{d^*}) = \min_{d \in \Delta'} \varkappa_{max}(V_d)$

where $\varkappa_{max}(V_d)$ = the maximum eigen value of V_d.

8.1.3 <u>The meaning of D- and E-optimality from the power function standpoint.</u>

In order to see the interpretation of D-optimality in terms of power

properties for hypothesis testing, we shall consider the notion of type E (or type D) test: Let the distribution of y depend on the parameters $(\underline{\gamma}_1 , \underline{\gamma}_2 ; \sigma^2)$. Then a test ϕ of the hypothesis H: $\underline{\gamma}_1 = \underline{\gamma}_1^*$ is said to be locally strictly unbiased if for each $(\underline{\gamma}_2 ; \sigma^2)$.

(i) $\beta_\phi (\underline{\gamma}_1^* , \underline{\gamma}_2 ; \sigma^2) = \alpha$

(ii) there exists a $\underline{\gamma}_1$-neighbourhood of $\underline{\gamma}_1^*$ in which

$$\beta_\phi (\underline{\gamma}_1 , \underline{\gamma}_2 ; \sigma^2) > \alpha \quad \text{for } \underline{\gamma}_1 = \underline{\gamma}_1^* .$$

supposing that the first and second derivative

$$\beta_\phi^i (\underline{\gamma}_2 , \sigma^2) = \frac{\partial}{\partial \beta_i} \beta_\phi (\underline{\gamma}_1 , \underline{\gamma}_2 ; \sigma^2) \Big|_{\underline{\gamma}_1^*} \text{ and } \beta_\phi^{ij} = \frac{\partial^2 \beta_\phi}{\partial \beta_i \partial \beta_j} \Big|_{\underline{\gamma}_1^*}$$

exist for all tests ϕ and all $(\underline{\gamma}_2 ; \sigma^2)$, then one can easily see that a necessary and sufficient condition for ϕ to be locally strictly unbiased is that $\beta_\phi^i (\underline{\gamma}_2 ; \sigma^2) = 0$ for all i and $(\underline{\gamma}_2 ; \sigma^2)$ and that the matrix $(\beta_\phi^{ij} (\underline{\gamma}_2 , \sigma^2))$ is positive definite for all $(\underline{\gamma}_2 ; \sigma^2)$.

<u>Definition:</u> A test of H is said to be of type E (type D if there is no nuisance parameters i.e. $\underline{\gamma}_1 = \underline{\beta}$ or no β_2). if it is locally strictly unbiased and among all tests with this property maximizes the determinant $|(\beta_\phi^{ij})|$. (This determinant under the stated conditions turns out to be the Gaussian curvature of the power surface at $\underline{\gamma}_1 *$).

Then the F test for testing the general linear univariate hypothesis is of type E.

Wald's theorem can now be stated as

"Analysis of variance test has the property of maximizing the surface integral

$$\int_S [\beta_\phi (\underline{\gamma}_1 , \underline{\gamma}_2 ; \sigma^2) - \alpha] \, dS .$$

among all similar (and hence locally unbiased) tests where S is the surface $\left\{ \underline{Y}_1 : (\underline{Y}_1 - \underline{Y}_1^*)'(\underline{Y}_1 - \underline{Y}_1^*) = \rho^2\sigma^2 \right\}$

letting ρ^2 tend to zero and utilizing the conditions

$$\beta_\phi^i(\underline{Y}_2; \sigma^2) = 0$$

$$\int_S (\beta_i - \beta_i^*)(\beta_j - \beta_j^*)\, dS = 0 \text{ for } i \neq j,$$

$$\int_S (\beta_i - \beta_i^*)^2 \cdot dS = k(\rho\sigma)$$

one can show that F maximizes $\sum_{i=1}^{r} \beta_\phi^{ii}(\underline{Y}_2, \sigma^2)$ among all locally unbiased tests as follows:

$$\int_S [\beta_\phi(\underline{Y}_1, \underline{Y}_2; \sigma^2) - \alpha] \cdot dS = \int_S [\beta_\phi(\underline{Y}_1, \underline{Y}_2; \sigma^2) - \beta_\phi(\underline{Y}_1^*, \underline{Y}_2; \sigma^2)] \cdot dS.$$

$$= \int_S [(\underline{Y}_1 - \underline{Y}_1^*) \mathcal{D}\beta_\phi(\underline{Y}_1, \underline{Y}_2; \sigma^2) + (\underline{Y}_1 - \underline{Y}_1^*) \mathcal{D}^2\beta_\phi(\underline{Y}_1, \underline{Y}_2; \sigma^2)(\underline{Y}_1 - \underline{Y}_1^*) + \varepsilon] \cdot dS$$

where $\mathcal{D}\beta_\phi = (\beta_\phi^i)$

$\mathcal{D}^2\beta_\phi = (\beta_\phi^{ij})$

and $\varepsilon \to 0$ when $\rho \to 0$.

$$= \int_S (\underline{Y}_1 - \underline{Y}_1^*)'(\beta_\phi^{ij})(\underline{Y}_1 - \underline{Y}_1^*)\, dS + \varepsilon \int_S dS.$$

$$= k(\rho\sigma) + \varepsilon \text{ (Surface Area)}$$

Thus Wald's theorem ensures that F test maximizes $\sum_{i=1}^{r} \beta_\phi^{ii}(\underline{Y}_2; \sigma^2)$ among all locally unbiased tests.

Now, since $(\beta_\phi^{ij}(\underline{Y}_2; \sigma^2))$ is positive definite

$$|(\beta_\phi^{ij})| \leq \Pi \beta_\phi^{ii}$$

(i.e. the determinant of a positive definite matrix is no greater than the product of its diagonal elements).

Using the inequality of the geometric and arithmetic means, we obtain

$$\left|\left(\beta_{\phi}^{ij}\right)\right| \leq \prod_{i=1}^{r} \beta_{\phi}^{ii} \leq \left[\sum_{i=1}^{r} \frac{\beta_{\phi}^{ii}}{r}\right]^{r}$$

$$\leq \left[\sum_{i=1}^{r} \frac{\beta_{F}^{ii}}{r}\right]^{r}$$

Since F test is invariant under nonsingular linear transformation we may assume that the components of Υ_1 are such that $\hat{\Upsilon}_1$ (the b.l.e. of Υ_1) have σ^2 times the identity for their covariance matrix.

Thus $$\left[\sum_{i=1}^{r} \frac{\beta_{F}^{ii}}{r}\right]^{r} = \prod_{i=1}^{r} \beta_{F}^{ii}$$

$$= \left|\left(\beta_{F}^{ij}\right)\right|$$

Thus the AOV test of the General Linear Hypothesis is of type E.

Obviously under F test $\left|\mathrm{Cov}\left(\hat{\Upsilon}_1\right)\right| = \left|\beta^{ij}\left(\Upsilon_2 ; \sigma\right)\right|^{-1}$ so that maximizing $\left|\beta^{ij}\left(\Upsilon_2 ; \sigma\right)\right|$ amounts to minimizing $\left|\mathrm{Cov}\left(\hat{\Upsilon}_1\right)\right|$.

Now we can see that for a D-optimum design, F-test of the general linear hypothesis is of type E, if it is suitable.

In the derivation of E-optimality, we have discussed, it is assumed that F-test is used in the testing of hypothesis. We know also that the power function of F test is a monotonic increasing function of the noncentrality parameter δ. Then a design which maximizes the minimum power on $\frac{\left(\Upsilon_1 - \Upsilon_1^*\right)' \left(\Upsilon_1 - \Upsilon_1^*\right)}{\sigma^2} = \rho^2$ simultaneously for all ρ is precisely one which maximizes the minimum of δ subject to $\frac{\left(\Upsilon_1 - \Upsilon_1^*\right)' \left(\Upsilon_1 - \Upsilon_1^*\right)}{\sigma^2} = \rho^2$

From (8.5) it follows that maximizing minimum power means

$$\text{maximizing} \min_{\xi'\xi = 1} \xi' V_{\Upsilon_1}^{-1} \xi = \frac{1}{\kappa\left(V_{\Upsilon_1}\right)}$$

i.e. minimizing $\max \kappa\left(V_{\Upsilon_1}\right)$ which is precisely E-optimality.

To summarize, D-optimality, although local properties seems more reasonable criteria than E-optimality, which is tied to the ad hoc

assumption that F test should always be used.

8.2. Before we study other optimality criteria and their meaning we would review some results we have obtained in the analysis of less than full rank model. The prediction equation based on the design d may now be written in the form

$$\underline{y} = \underline{X}_d \underline{\beta} + \underline{e} \qquad (8.6)$$

If there is an additional assumption of the form

$$\underline{\Gamma}\,\underline{\beta} = \underline{c}$$

where $\underline{\Gamma}$ and $\underline{c}$ are known g x p and g x 1 matrices, then it can be absorbed in (8.6). We shall suppose that this has been done, without any loss of generality.

Also a hypothesis H_o will be of the form $\underline{\Lambda}\,\underline{\beta} = \underline{0}$ where $\underline{\Lambda}$ is a specified rxp matrix ($r \leq p$) which we can take to be of rank r. We may think of the class H of alternatives as being all $\underline{\beta}$ for which $\underline{\Lambda}\,\underline{\beta} \neq 0$.

A hypothesis of the form $\underline{\Lambda}\,\underline{\beta} = \underline{\ell}$ can be transformed to the above form by using any solution $\underline{b}$ of $\underline{\Lambda}\,\underline{\beta} = \underline{\ell}$ and replace $\underline{y}$ by $\underline{y}^* = \underline{y} - \underline{X}_d\underline{b}$ and $\underline{\beta}^* = \underline{\beta} - \underline{b}$. It may or may not be that r elements of $\underline{\Lambda}\,\underline{\beta}$ are estimable when a given design d is used.

Suppose that there are s linearly independent linear combinations of the elements of $\underline{\Lambda}\,\underline{\beta}$ which have unbiased estimators when d is used, but not s + 1 such combinations. Then there exist an s x r matrix $\underline{Q}$ such that all components of $\underline{Q}\,\underline{\Lambda}\,\underline{\beta}$ are estimable. Let $\hat{\underline{\psi}}$ be the s-vector of such estimators with minimum variance. We shall use $\sigma^2\underline{V}_d$ to denote the

covariance of $\hat{\underline{\Psi}}$. We know that, if rank $\underline{X}_d$ is b, then there are b linearly independent estimable functions of the components of $\underline{\beta}$. Of this, we can take s of them to be the elements of $\underline{Q}\,\underline{\Lambda}\,\underline{\beta}$; thus there exists a (b - s) x p matrix R of rank (b - s) whose rows are orthogonal to those of $\underline{Q}\,\underline{\Lambda}$ and such that all components of $\underline{R}\,\underline{\beta}$ have unbiased estimates.

For any test Φ_d associated with d, let $\beta_{\Phi_d}(\underline{\beta}, \sigma^2)$ be the power function of Φ_d . For $0 < \alpha < 1$ we shall denote by $H_d(\alpha)$ the class of all Φ_d of size α i.e. all Φ_d for which

$$\beta_{\Phi_d}(\underline{\beta}, \sigma^2) \leq \alpha \text{ wherever } \underline{A}\,\underline{\beta} = \underline{0} \qquad (8.6)$$

and by $H_d^*(\alpha)$, the class of similar tests off size α i.e. those for which (8.6) holds with the inequality sign replaced by equality. Finally, let $F_{d,\alpha}$ denote the usual F test of H_o of size α with s and n - b degrees of freedom, based on $\hat{\underline{\Psi}}'\,\underline{V}_d^{-1}\,\hat{\underline{\Psi}} / s\hat{\sigma}^2$ where $\hat{\sigma}^2$ is the unbiased estimator of σ^2 . (If σ^2 is know, this is to be replaced by the appropriate χ^2 test).

In practical applications, we have to consider two cases of null hypothesis:

Case I $\quad \beta_1 = \beta_2 = \cdots = \beta_{r-1} = \beta_r = 0$

Case II $\quad \beta_1 = \beta_2 = \cdots = \beta_r$

The corresponding $\underline{A}$ matrices in $\underline{A}\,\underline{\beta} = \underline{0}$ will be $\underline{A}_1$ for Case I where $\underline{A}_1$ is the (r x r) identity matrix followed by p - r columns of zeros and R_{11} for Case II where $\underline{A}_{11}$ is a (r - 1) x r matrix $\underline{P}$ followed by p - r columns of zeros, $\underline{P}$ being made up of the last (r - 1) rows of a r x r orthogonal

matrix $\underline{O}$ whose first row elements are all $\frac{1}{\sqrt{r}}$. In Case II, as discussed in Chapter IV, it is often convenient to write the normal equations in the form

$$\underline{C}\,\hat{\underline{\gamma}}_1 = \underline{q}$$

where $\underline{C}$ is an (r x r) matrix of rank (r - 1), $\underline{q}$ a r-vector of linear forms in y with covariance matrix $\underline{C}\,\sigma^2$. For any solution $\hat{\underline{\gamma}}_1$ of these equations one obtains the best linear estimator of any contrast $\sum c_i \beta_i$ with $\sum c_i = 0$ by forming $\sum_{i=1}^{r} c_i \hat{\beta}_i$. Clearly, $\underline{P}\,\hat{\underline{\gamma}}_1$, is the best linear estimate of $\underline{A}_{11}\,\underline{\beta}$ ($\underline{\gamma}_1$ is the interested components of $\underline{\beta}$). We shall assume that $\underline{A}$ is full rank.

The last (r - 1) rows of the equation $\underline{O}\,\underline{C}\,\underline{O}'\,\underline{O}\,\hat{\underline{\gamma}}_1 = \underline{O}\,\underline{q}$ are thus

$$\underline{P}\,\underline{C}\,\underline{P}\,(\underline{P}\,\hat{\underline{\gamma}}_1) = \underline{P}\,\underline{q}$$ (the first row and column of $\underline{O}\,\underline{C}\,\underline{O}'$ are zero).

so that $\underline{P}\,\hat{\underline{\gamma}}_1 = (\underline{P}\,\underline{C}\,\underline{P}')^{-1}\,\underline{P}\,\underline{q}$

i.e. $$(\underline{A}_{11}\hat{\underline{\beta}}) = (\underline{P}\,\underline{C}\,\underline{P}')^{-1}\,\underline{P}\,\underline{q}$$

$$\mathrm{Cov}\,(\underline{A}_{11}\hat{\underline{\beta}}) = (\underline{P}\,\underline{C}\,\underline{P}')^{-1}\,\underline{P}\,\mathrm{Cov}\,\underline{q}\,\underline{P}\,(\underline{P}\underline{C}\underline{P}')^{-1}$$

$$= (\underline{P}\,\underline{C}\,\underline{P}')^{-1}\sigma^2$$

That the matrix $(\underline{P}\,\underline{C}\,\underline{P}')^{-1}$ is a multiple of the identity follows from the following lemma.

Lemma 8.1. If $\underline{C}$ is a r x r matrix with diagonal elements a and off diagonal elements c, then

$$\underline{O}\,\underline{C}\,\underline{O} = \begin{vmatrix} a + (r-1)c & 0 \\ 0 & (a-c)\,\underline{I}_{r-1} \end{vmatrix}$$

Proof: (i,j)ith element of $\bar{O}\,\underline{C}\,\bar{O}'$ $= \sum_t \sum_s o_{it}\, c_{ts}\, o_{js}$

For $i = j = 1$, we have $\sum_t \sum_s o_{1t}\, c_{ts}\, o_{1s} = \sum_t \sum_s \frac{1}{\sqrt{r}}\, c_{ts}\, \frac{1}{\sqrt{r}}$

$$= \frac{1}{r} \sum_t \sum_s c_{ts}$$

$$= \frac{1}{r} \left(ra + 2\left(\frac{r-1}{2}\right)\left[(r-1)c + c\right]\right)$$

$$= a + (r-1)\,c$$

This pivotal element vanishes when the row and column sums of $\underline{C}$ are zero.

For $i = j\ (\neq 1)$, we have

$$\sum_t \sum_s o_{it}\, c_{ts}\, o_{is} = (a-c) + c \sum_t \sum_s o_{it}\, o_{is}$$

$$= (a-c) + c \left(\sum_t o_{it}\right)^2$$

$$= (a-c) \qquad \left(\because \textstyle\sum o_{it} = 0\right)$$

For $i \neq j$,

$$\sum_t \sum_s o_{it}\, c_{ts}\, o_{js} = (a-c) \sum_t \sum_s o_{it}\, o_{is} + c \sum_t \sum_s o_{it}\, o_{js}$$

$$= (a-c) \left(\textstyle\sum o_{it}\right)^2 + c \left(\sum o_{it}\right)\left(\sum o_{js}\right)$$

$$= 0$$

Now the properties of the matrix C_d (Bose's information matrix) of the reduced normal equations $\underline{C}_d\, \hat{\underline{\gamma}} = \underline{q}_d$ for the varieties in a block design setting $(r < k)$ suggests the different optimality criteria of various designs. We know that $\underline{P}\,\underline{C}_d\,\underline{P}'$ is proportional to the inverse of the covariance matrix $\underline{V}_d$ of the best linear estimators of contrast $\psi = \sum_{i=1}^{v} c_i \beta_i$ with $\sum c_i = 0$ of the treatment effects $\underline{\gamma} = (\beta_1, \cdots \beta_v)$

Suppose a design d* is such that $\underline{C}_{d*}$ has the properties discussed above (i.e. $\underline{C}_{d*}$ has all diagonal elements equal, row and column sums equal to zero, and is of rank $(r-1)$)and has maximum possible trace. Then by Lemma (8.1) $\bar{Q}\,\underline{C}_{d*}\,\bar{Q}'$ has the same positive constant for each of the last $(r-1)$ diagonal elements, and is zero elsewhere. Since the lower

right-hand $(r-1) \times (r-1)$ submatrix of $\underline{Q}\, \underline{C}_{d*}\, \underline{Q}' = \underline{P}\, \underline{C}_d\, \underline{P}'$ is just the reciprocal of the covariance matrix $\underline{V}_{\underline{\psi}}$ for the b.l.e's of the contrasts $\psi_1, \cdots \psi_{r-1}$, we have the following theorem.

Theorem: 8.2 If $\underline{C}_{d*}$ has maximum possible trace, all diagonal elements equal and all off-diagonal elements equal, then $\underline{V}_{d*}^{-1}$ has maximum possible trace and is a multiple of the identity.

Proof: $\underline{Q}\, \underline{C}_d \underline{Q}'$ has the upper left-hand element equal to zero.

Since $\mathrm{tr}\left(\underline{V}_d^{-1}\right) = \mathrm{tr}\left(\underline{Q}\, \underline{C}_d \underline{Q}'\right) = \mathrm{tr}\left(\underline{C}_d\right)$,

$\mathrm{tr}\left(\underline{V}_{d*}^{-1}\right)$ is a maximum.

Now consider $\det \underline{V}_d^{-1}$

$$\leq \prod_{i=1}^{r-1} v_{d_{ii}}$$

$$\leq \left[\sum_{i=1}^{r-1} \frac{v_{d_{ii}}}{r-1}\right]^{r-1}$$

$$\leq \left[\sum_{i=1}^{r-1} \frac{v_{d*_{ii}}}{r-1}\right]^{r-1}$$

$$= \prod_{i=1}^{r-1} v_{d*_{ii}}$$

$$= \det \underline{V}_{d*}^{-1}$$

From this theorem we can deduce that any design d* with the above properties is optimum according to a wide variety of optimality criteria considered separately by various authors.

8.21. Various optimality criteria:-

Suppose we are interested in s given linearly independent parametric functions $\psi_j = \sum_{i=1}^{p} \lambda_{ji} \beta_i$, $1 \leq j \leq s$. Let Δ' be the class of designs in Δ for which all ψ_j are estimable, and let $\underline{V}_d \sigma^2$ be the covariance matrix of the b.l.e's of the ψ_j for a d in Δ'. Then a design d* is said to be

(1) D-optimum in Δ if $d* \in \Delta'$ and

$$\det \underline{V}_{d^*} = \min_{d \in \Delta'} \det \underline{V}_d$$

i.e. minimizing the generalized variance (Wald)

(ii) E-optimum in Δ if $d^* \in \Delta$ and

$$\varkappa(\underline{V}_{d^*}) = \min_{d \in \Delta'} \varkappa(\underline{V}_d)$$

where $\varkappa$ is the maximum eigen value of $\underline{V}_d$.

Also called the minimaxity with respect to all standard parametric forms: minimizing the largest eigen value of $\underline{V}_d$ (Wald & Ehrenfeld)

(iii) A -optimum in Δ if $d^* \in \Delta'$ and

$$\text{trace } \underline{V}_{d^*} = \min_{d \in \Delta'} \text{trace } \underline{V}_d$$

i.e. minimizing the average variance (Elfving & Charnoff)

(iv) Minimax with respect to single parameters

i.e. minimizing the maximum diagonal element of V_d (Elfving)

(v) Minimum average efficiency i.e. minimizing the average of the variances of the best linear estimators of $\beta_i - \beta_j$ this average being proved easily to be proportional to tr V_d.

(vi) L-optimum : maximizing under the assumption of normality the minimum power on spheres $\frac{\underline{\psi}'\underline{\psi}}{\sigma^2} = c^2$ as $c \to 0$ for testing hypothesis

i.e. for $d^* \in \Delta'$

$$\lim_{c \to 0} \left[\bar{\beta}(d^*, c, \alpha) - \alpha \right] \Big/ \left[\sup \bar{\beta}(d, c, \alpha) - \alpha \right] = 1$$

where $\bar{\beta}(d, c, \alpha)$ is the sup of $\bar{\beta}_\phi(c)$ over all ϕ of size α, $\bar{\beta}_\phi(c)$ being the informum of the power function

of the test ϕ over all alternatives for which $\frac{\psi'\psi}{\sigma^2} = c^2$.

Another optimality criterion, first considered by K. Smith and studied thoroughly in the most general setting by P. G. Guest is

(vii) G-optimality = minimizing the maximum variance of the estimated regression function

i.e. $\min_{d \in \Delta} \max \text{ var } \left[\sum (\hat{\beta}_i - \beta_i) f_i(x_i) \right]$

We have seen that G-optimum allocation and D-optimum allocation in polynomial regression are the same. In fact this result represents a special case of the following theorem.

Keifer-Wolfowitz Theorem: If $\mathcal{X}$ is a compact space on which the f_i are continuous and linearly independent, then the allocation $\underline{p}$ is D-optimym for $\underline{\beta}$ if and only if it is G-optimum.

8.2.2 Relations between various optimality criteria.

Having stated a relationship between G- and D-optimality, we may now turn to the question of relationships among the other criteria. In general, the remaining optimality criteria are unrelated. However, in certain situations such as those where balanced block designs, Latin Squares and Youlden Squares are employed, it happens that their criteria are related, due to the symmetry of the design.

Lemma: 8.2 Suppose b_d is constant for d in Δ'. If d^* is D-optimum and V_{d^*} is a multiple of the identity then d* is E-optimum, A-optimum, minimax with respect to a single parameter and L-optimum.

Proof: All optimality criteria except L-optimality is obvious from the nature of $\underline{V}_d$.

If d^* were not L-optimum, since $F_{d,\alpha}$ has property of maximizing average power (Wald), for some other design there would be an associated test ϕ_d in $H_d^*(\alpha)$ with

$$\inf_{\tau_2, \sigma^2} \left| \beta^{ij}_{\phi_d}(\tau_2, \sigma^2) \right| > \left| \beta^{ij}_{F_{d^*,\alpha}}(\tau_2, \sigma^2) \right|$$

If ϕ_d is replaced by $F_{d,\alpha}$ we have the equality.

This yields the contradiction that $\det \underline{V}_d < \det \underline{V}_{d^*}$.

Next we shall consider some theorems useful in many applications.

Theorem: 8.3 If $\prod_{i=1}^{r} c_{ii}$, the product of the diagonal elements of $\underline{C}_d$ is a maximum for $d = d^*$ and if $\underline{C}_d$ has all diagonal elements equal and all off-diagonal elements equal, row and column sums equal to zero, then $\det \underline{V}_d^{-1}$ is a maximum for $d = d^*$.

The proof of the above theorem depends on the following Lemma.

Lemma: 8.3 For $r > 1$ if $\underline{Q}$ is orthogonal rxr, $\underline{D}$ is diagonal rxr, $\underline{C}$ is symmetric non-negative definite rxr with row and column sums zero, and $\underline{Q}\,\underline{D}\,\underline{Q}' = \underline{C}$, then

$$\left(\frac{r-1}{r}\right)^r \left(\prod_{i=1}^{r-1} d_{ii}\right)^{\frac{r}{r-1}} \leq \prod_{i=1}^{r} c_{ii} \qquad (8.7)$$

Proof: Assume that $d_{rr} = 0 < d_{ii}$ for $i<r$, otherwise the result is trivial, since the row and column sums of $\underline{C}$ are zero,

$$0 = \sum_{i=1}^{r}\sum_{j=1}^{r} c_{ij} = \sum_{i=1}^{r}\sum_{j=1}^{r}\sum_{s=1}^{r-1} o_{is}\, o_{js}\, d_{ss} = \sum_{s=1}^{r-1} d_{ss} \left(\sum_{i=1}^{r} o_{is}\right)^2 \qquad (8.8)$$

Thus $\sum_{i=1}^{r} o_{is} = 0$ for $s = 1, 2, \ldots\ldots, r-1$.

i.e. the first $r-1$ columns of are orthogonal to the vector of ones . We may thus take $o_{jr} = \frac{1}{\sqrt{r}}$ $(j = 1, 2, \ldots\ldots, r)$.

Let the coordinates of a point ϵ in $r(r-1)$-dimensional Euclidean space be denoted by ϵ_{ij} $(i = 1, 2, \ldots\ldots, r ; j = 1, 2, \ldots, r-1)$.

and let R be the set of points ϵ in this space for which all $\epsilon_{ij} \geq 0$, for which $\sum_{j} \epsilon_{ij} = (r-1)/r$ for all i ,and for which $\sum_{i} \epsilon_{ij} = 1$ for all j .Then we shall prove below that ϵ in R implies

$$\prod_{i=1}^{r} \Big(\sum_{j=1}^{r-1} \epsilon_{ij} d_{jj} \Big) \geq \Big(\tfrac{r-1}{r} \Big)^{r} \Big(\prod_{i=1}^{r-1} d_{ii} \Big)^{\frac{r}{r-1}} \qquad (8.9)$$

Since the left hand side of (8.9) with $\epsilon_{ij} = o_{ij}^2$ gives the right hand side of (8.7) and since the restriction on the ϵ_{ij} in R must be satisfied by the o_{ij}^2 , (8.9) implies (8.7) .

Now consider the left hand side of (8.9) , $f(\epsilon) = \prod_{i=1}^{r} \Big(\sum_{j=1}^{r-1} \epsilon_{ij} d_{jj} \Big)$.

We can verify that $-\log f(\epsilon)$ is convex in ϵ on $r(r-1)$-space, and hence on R . Moreover , R is a convex body in $r(r-1)$-space ,and any extreme point of R is either

$$\begin{vmatrix} \epsilon_{11} & \cdots & \epsilon_{1,r-1} \\ \vdots & & \vdots \\ \epsilon_{r1} & \cdots & \epsilon_{r,r-1} \end{vmatrix} = \begin{vmatrix} \frac{r-1}{r} & 0 & \cdots & 0 \\ 0 & \frac{r-1}{r} & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & \frac{r-1}{r} \\ \frac{1}{r} & \frac{1}{r} & \cdots & \frac{1}{r} \end{vmatrix} \qquad (8.10)$$

or is obtained by permuting the rows of the matrix on the right side of (8.10) . Since a, convex function on a convex set attains its maximum at an extreme point, we conclude that the minimum of f is attained

at one of these extreme points . But f has the same value at any of these extreme points, namely

$$\min_{R} f(\epsilon) = \Big(\sum_{i=1}^{r-1} d_{ii}\Big)\prod_{i=1}^{r-1}\Big(\frac{r-1}{r}\, d_{ii}\Big). \qquad (8.11)$$

But from the well known inequality between geometric and arithmetic means, we have

$$\Big(\prod_{i=1}^{r-1} d_{ii}\Big)^{\frac{1}{r-1}} \le \frac{1}{r-1}\sum_{i=1}^{r-1} d_{ii}$$

$$\frac{r-1}{r}\Big(\prod_{i=1}^{r-1} d_{ii}\Big)^{\frac{1}{r-1}} \le \frac{1}{r}\sum_{i=1}^{r-1} d_{ii}$$

$$\Big(\frac{r-1}{r}\Big)^{r}\Big(\prod_{i=1}^{r-1} d_{ii}\Big)^{\frac{1}{r-1}}\Big(\prod_{i=1}^{r-1}\frac{r-1}{r}\, d_{ii}\Big) \le \frac{1}{r}\Big(\sum_{i=1}^{r-1} d_{ii}\Big)\Big(\prod_{i=1}^{r-1}\frac{r-1}{r}\, d_{ii}\Big)$$

$$\Big(\frac{r-1}{r}\Big)^{r}\Big(\prod_{i=1}^{r-1} d_{ii}\Big)^{\frac{1}{r-1}+1} \le \frac{1}{r}\Big(\sum_{i=1}^{r-1} d_{ii}\Big)\Big(\prod_{i=1}^{r-1}\frac{r-1}{r}\, d_{ii}\Big)$$

$$\Big(\frac{r-1}{r}\Big)^{r}\Big(\prod_{i=1}^{r-1} d_{ii}\Big)^{\frac{r}{r-1}} \le \min f(\epsilon) \le \prod_{i=1}^{r} c_{ii} \quad .$$

Proof of theorem 8.3.

In order to conform to the previous notation we may use lemma 8.3 with product on the left hand side of (8.7) going from 2 to r .In this form, with $\bar{\underline{Q}} = \underline{Q}$,it follows from lemma 8.1 that the left and right sides of (8.7) are equal for $\underline{C}_d = \underline{C}_{d^*}$.

Hence, by lemma 8.3 $\prod_{i=1}^{r-1} v_{ii}$ is a maximum for $d = d^*$. Since $\det \underline{V}_d^{-1}$ is non-negative definite $\prod_{i=1}^{r-1} v_{ii} \ge \det \underline{V}_d^{-1}$ with equality only for the diagonal matrix $\underline{V}_{d^*}^{-1}$.The proof of the theorem is thus complete.

Theorem 8.4. For $r > 1$, if $\varkappa(\underline{V}_d^{-1})$ is the minimum eigen value of $\underline{V}_d^{-1}$, then

$$\varkappa(\underline{V}_d^{-1}) \le \frac{r}{r-1}\min_i c_{ii}^{-1} \quad ,$$

the equality holding only if all diagonal elements of C_d are equal and all off-diagonal elements are equal.

Proof . Let e_i be the r-vector with ith element one and all other elements zero . Let $\xi_i = P\ e_i$.

Clearly $\sqrt{\frac{r}{r-1}}\ \xi_i$ has unit length .

Hence $c_{ii} = e_i'\ C_d\ e_i$.

$= (\ O\ e_i\ )'(\ O\ C_d\ O\ )\ (\ O\ e_i\ )$.

$= \xi_i'\ V_d\ \xi_i$.

$= \frac{r-1}{r}\ \min_{a'a=1}\ a'V_d^{-1}a$.

$= \frac{r-1}{r}\ \kappa\ (V_d)$.

8.3 Optimality of certain basic designs in common use .

8.3.1. Optimality of Balanced Block Designs . In the setting of one way heterogeneity, suppose we have b blocks of size k and t varieties without the assumption that $k \leq t$ (which is true in incomplete block designs). Generalizing the notion of a balanced incomplete block design, we define a design d in the above setting to be a balanced block design (B B D) if (a) the number of times n_{ij} that variety i appears in block j is k/t if this is an integer, and is one of the two closest integers otherwise ; (b) the number $\sum_j n_{ij}$ of replications of variety i is the same for all i , and (c) for each pair i_1 , i_2 with $i_1 \neq i_2$, $\Sigma\ n_{i_1 j} n_{i_2 j}$ is the same.

Theorem 8.5 . If a B B D d exists, then it is D-optimum, E-optimum, A-optimum, minimax with respect to single parameters and L-optimum .

Proof . In the setting of one way heterogeneity , Bose's information matrix , $\underline{C}_d$ is of the form ,

$$\left(c_{d_{ij}} \right) = \left(\delta_{ij} \left(\sum_j n_{ij} \right) - \sum_s \frac{n_{is} n_{js}}{k} \right)$$

so that $tr\left(\underline{C}_d \right) = bk - \sum_i \sum_s \frac{n_{is}^2}{k}$.

The expression for $tr(\underline{C}_d)$ is a maximum for $d = d^*$.

The result follows from theorem 8.2 and lemma 8.2 .

8.3.2. Optimality of Youdon Squares .

In the setting of two way heterogeneity with k rows , k columns and t treatments , suppose there exist a design d such that (a) $\sum_s n_{is}^{(1)}$ $\left(= \sum_s n_{is}^{(2)} \right)$, the number of replications of treatment i is the same for all i (b) $\sum_s n_{i_1 s}^{(1)} n_{i_2 s}^{(2)}$ is the same for each pair i_1 , i_2 with $i_1 \neq i_2$ (c) $\sum_s n_{j_1 s}^{(1)} n_{j_2 s}^{(2)}$ is the same for each pair j_1 , j_2 with $j_1 \neq j_2$ (d) all $n_{ij}^{(q)}$ are equal to k_q/t if k_q/t is an integer and are either of the two integers closest to k_q/t otherwise (q = 1, 2) . Thus , d^* is a B B D when the rows or columns are considered to be the blocks . Such a design d^* is known as a Y S if $k < t$ (and k/t is an integer). Without this condition, such a design d^* shall be called a Generalized Youdon Square (G Y S) .

Theorem 8.6. If k_1/t or k_2/t is an integer and if a G Y S d^* exist, then d is D-optimum, E-optomum, A-optimum, L-optimum and minimax with respect to single parameters.

Proof . If we can show that $\sum_i c_{ii}$ is a maximum for $d = d^*$,then theorem 8.2 yields the desired result.

We shall use the notation [x] for the greatest integer $\leq x$.

The expression $\sum_i^k m_j^2$ is minimized ,

subject to the restrictions that

(1) all m_j's are integers ,

(2) $\sum_{1}^{k} m_j = n$,

by taking $k - n + k[n/k]$ of the m_i to be $[n/k]$, and $n - k[n/k]$ of them to be $[n/k] + 1$, the corresponding minimum of Σm_j^2 being $n + (2n - k)[n/k] - k[n/k] = h(n,k)$ (say).

Assuming that k_1/t is an integer , we have for any design d , from Bose's information matrix ,

$$\Sigma c_{ii} = \sum_i \sum_s n_{is}^{(1)} - \sum_i \sum_s \frac{n_{is}^{(1)2}}{k_2} - \sum_i \sum_s \frac{n_{is}^{(2)2}}{k_1} + \sum_i \left(\frac{\sum_s n_{is}^{(1)} \sum_s n_{is}^{(2)}}{k_1 k_2} \right)$$

$$= k_1 k_2 - \sum_i \sum_s \frac{n_{is}^{(1)2}}{k_2} - \sum_i \sum_s \frac{n_{is}^{(2)2}}{k_1} + \frac{\sum_i \left(\sum_s n_{is}^{(1)} \right)^2}{k_1 k_2}$$

$$\therefore k_1 k_2 \left(k_1 k_2 - \sum_i c_{ii} \right) = k_1 \sum_i \sum_s n_{is}^{(1)2} + k_2 \sum_i \sum_s n_{is}^{(2)2} - \sum_i \left(\sum_s n_{is}^{(1)} \right)^2$$

$$= \sum_i \left\{ k_1 \sum_s n_{is}^{(1)2} + k_2 \sum_s n_{is}^{(2)2} - \left(\sum_s n_{is}^{(1)} \right)^2 \right\}$$

$$\geq \sum_i \left\{ k_1 h \left(\sum_s n_{is}^{(1)}, k_1 \right) + k_2 h \left(\sum_s n_{is}^{(2)}, k_2 \right) - \left(\sum_s n_{is}^{(1)} \right)^2 \right\} \qquad (8.12)$$

with equality in the case of a G Y S .

Our theorem will be proved if we can show that the right hand side of the above inequality attains its minimum for $d = d^*$. Now $h(n, k) \geq n^2/k$, the latter being the minimum of Σm_j^2 subject to $\Sigma m_j = n$ without the restriction that the m_j be integers . Hence ,the first and the third sums of the right hand side togather is at least zero. Moreover, this lower bound is achieved by this sum when $d = d^*$, since $\Sigma n_{is}^{(i)}/k = k_i/t$ is an integer . (For $k_1 k_2 / \Sigma n_{is}^{(i)} = t$).

We shall show that, subject to $\Sigma z_i = c$, the expression

$$q(z_1, z_2, \ldots\ldots, z_m) = \sum_{i=1}^{m} (2z_1 - 1)[z_i] - [z_i]^2 \} \quad (8.13)$$

is a minimum when all z_i are equal; putting $z_i = \left[\sum_s n_{is}^{(2)} / k_1 \right]$, we can see that this will yield the desired conclusion regarding the last sum of (8.12).

Proof regarding (8.13) is by induction.

Assuming the conclusion to be true for $m = M$ in proving the case $m = M + 1$, we may put $z_1 = z_2 = \ldots\ldots = z_M = s$ and $z_{M+1} = c - Ms$. in (8.13). The resulting expression is continuous in s and, except on a discrete set, has a derivative with respect to s which is equal to $2M([s] - [c - Ms])$. The latter is ≤ 0 if $s \leq c/(M + 1)$ and is $\geq 2M([s])$ if $s > c/(M + 1)$ sothat $s = c/(M + 1)$ yields a minimum. The proof we have considered depends on the assumption that k_1/t or k_2/t is an integer, otherwise the proof fails.

Analysis of other designs.

Many other design settings may be analysed in a similar manner. Though many authors have suggested different methods of establishing various optimum properties of well known designs, only J. Keifer has succeeded in presenting a unified theory.

Dicussion .

Most of the papers on optimum design of experiments are mainly concerned with establishing well known designs optimal or non-optimal in the sense of several optimality criteria we have discussed so far. Many more optimality criteria are suggested by different authors. Until very recently , few computational methods , mostly geometrical (main contribution by G.Elfving), for the actual construction of optimal designs are considered. The methods of construction of designs, optimal in some acceptable sense, are in the process of developement. Important aims in experimental designs are to estimate the relevant parameters with maximum precision for a given total number of experimental units, or total cost and to perform a test of the null hypothesis with maximum sensitivity. With these considerations in mind, we should construct a design which satisfies our specific need, i,e.which criteria should be adopted in the construction of the design depends on our aim in conducting the experiment .

For polynomial regression, P.G.Guest , P.G.Hoel , J.Keifer and J.Wolfowitz have constructed optimal designs under various criteria of optimality. Very recently , game-theorectic nature of the construction of optimum designs has been discovered , and recent papers by Keifer and Wolfowitz and J.Aitcheson show the possibility of constructing optimum designs by appealing to games theory.

Though a thorough study has been made on admissibility in the case of partial estimation using minimum variance goodness criteria in this

thesis mainly on the line of J.Keifer, H.Erhenfeld and G.Elfving , a systematic exploration is still needed in that direction.

BIBLIOGRAPHY

1. AITCHESON, J., The construction of optimal designs for the one-way classification Analysis of Variance. J.R.S.S., 23 (1961), 352-367.

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

3. EHRENFELD, S., On the efficiency of Experimental Designs. Ann. Math. Stat. 26 (1955), 247-255.

4. EHRENFELD, S., Complete class theorems in experimental designs. Proc. Third Berkley Symp. Math. Stat. and Prob. 1 (1955), 57-68.

5. ELFVING, G., Optimum allocation in linear regression theory. Ann. Math. Stat., 23 (1952), 255-262.

6. ELFVING, G., Convex sets in statistics. XII Congr. Math. Scand. (1953), 34-39.

7. ELFVING, G., Geometric Allocation Theory. Skand. Aktuarietidskr 1954, 170-190.

8. ELFVING, G., A selection problem in experimental design. Comm. Phys. Math. Soc. Sei. Fenn XX(2) (1957).

9. ELFVING, G., Minimax characters of balanced experimental designs. XIII Congr. Math. Scand. (1957) 69-76.

10. ELFVING, G., Design of linear experiments in Probability and statistics, The Harald Cramer Volume, editor, U Granander, New York. Wiley (1959).

11. DE LA GARZA, A., Spacing of information in polynomial regression. Ann. Math. Stat. 25 (1954), 123-130.

12. GRAYBILL, F.A., An Introduction of Linear Statistical Models. McGraw Hill. New York (1961).

13. GUEST, P.G., The spacing of observations in polynomial regression. Ann. Math. Stat. 29 (1958). 294-299.

14. HOEL, P.G., Efficiency problems in polymomial estimation Ann. Math. Stat. 29 (1958) 1134-1145.

15. HOTELLING. H., Some improvements in weighing and other experimental techniques. Ann. Math. Stat. 15 (1944), 297-306.

16. KEMPTHORNE, O., The Design and Analysis of Experiments. J. Wiley. New York (1952).

17. KENDALL, M.G., & STUART, A. The Advanced Theory of Statistics. Vol. 2. Charles Griffin, London, 1961.

18. KIEFER, J., On the non-randomized optimality and randomized non-optimality of Experimental Designs. Ann. Math. Stat. 29 (1958) 675-699.

19. KIEFER, J., Optimum Experimental Designs. J. R. S. S. (1959) , Vol. 21, 272-302 .

20. KSHIRSAGAR, A.M., A note on Incomplete Block Designs. Ann. Math. Stat. 29 (1958) 907-910.

21. LEHMANN, E.L. Testing Statistical Hypothesis. J. Wiley, New York. (1959).

22. MANN, H., Design and Analysis of Experiments. Dover, New York. (1949).

23. MOTE, V.L., On a minimax property of Balanced Incomplete Block Designs. 29 (1958), 910-914.

24. PLACKETT, R.L., Principles of Regression Analysis. Oxford (1960).

25. SCHEFFE, H., The Analysis of Variance. J. Wiley (1959).

26. WALD, A., On the efficient design of statistical investigations. Ann. Math. Stat. 14 (1943) 134-140.

27. WILKS, S.S., Mathematical Statistics. J. Wiley (1962).