Computer assisted design of experiments

B. Jones

A Thesis submitted to the University of Kent at Canterbury for the degree of Doctor of Philosophy.

September 1980.

Abstract

It is now generally true that the computer is widely used for analysing data from comparative experiments but relatively little attention has been paid to how the computer can assist at the design stage of an experiment. This thesis attempts to remedy this situation.

In Chapter 1 an attempt is made to formalize the conversation that takes place between an experimenter and a statistical consultant with a view to transferring the consultant's skills to the computer. To determine the difficulties involved experiments to compare two treatment means are considered.

Where there are more than two treatments it is essential to determine which treatment comparisons are important. Existing computerised algorithms for constructing suitable experimental designs deal solely with the case where there is an equal interest in all treatment comparisons and where all blocks are of the same size. Chapter 2 describes an algorithm without these limitations. It searches for treatment interchanges that improve a given design.

In Chapter 3 this algorithm is extended to cover row-and-column designs. A further algorithm is required to combine component designs.

One disadvantage of these algorithms is the invariance of the treatment replications once chosen. In Chapter 4 completely different algorithms without this disadvantage are developed. They are based on an exchange procedure and an interchange procedure.

CONTENTS

Introduction

Computers, nowadays, provide experimenters and statisticians with valuable assistance in the analysis of data from designed experiments. It is natural therefore, to enquire if the computer is able to assist at the design stage of an experiment. The work described in this thesis was begun in order to provide an answer to this question.

When executing a computer program it is usual to do it either interactively or in batch mode. With the former, information is fed into the computer and received back from it via a terminal device of some sort, e.g. a teletype or visual display unit. With the latter, cards or their equivalent are fed in and output is written to a line printer or equivalent medium. The difference between these two forms of use is that when executing interactively it is possible to get an immediate response from the computer. In batch mode there is usually some delay in the output getting printed as the computer jobs are executed sequentially.

Using the computer interactively means that it is possible to write a program that can carry out a dialogue with the user. That is, information can be requested by the program, acted upon and results printed or further information requested. The ability to program in this conversational mode suggests that a program could be written to provide experimenters with useful statistical advice on the design of their experiment. The computer would output questions on various aspects of the design and then, if possible, advise the experimenter on the best course of action.

Interactive computer programs for data analysis are, of course, available [the program GLIM described by Baker and

Nelder (1978), for example] but there does not appear to be a program that provides advice on the design of an experiment. To discover why this might be, an attempt was made to write such a program. The results of doing this are described in Chapter 1. There, to illustrate some of the difficulties involved, an experiment to compare two treatment means is considered in detail. Also noted are those aspects of designing an experiment that cannot be incorporated into a computer program but require the advice of the experienced statistical consultant. Eventually, either during the conversation with a statistical consultant or with an interactive computer program, the point will be reached when it is necessary to provide the experimenter with a design that is ideally suited to his needs. There will only be a limited amount of experimental material available and this may have to be arranged in blocks or rows and columns to eliminate systematic variation from the treatment comparisons. Further, the experimenter may only be interested in certain treatment comparisons and may not have an equal interest in each one. The human consultant may find that a design, ideally suited to the experimenter's needs, is not available in the literature and may have to devise one somehow. If the interactive computer program described above is to provide designs optimal for the experimenter's needs then computerised algorithms for constructing such designs are required. The rest of this thesis is therefore concerned with developing computerised algorithms that search for designs that are optimal for the experimenter's requirements. Such algorithms are another way the computer can assist at the design stage of an experiment.

In Chapter 2 an algorithm that searches for optimal

block designs is described. This algorithm is based on interchanging treatments between blocks in an attempt to improve an appropriate optimality criterion. Two forms of the criterion are compared and the performance of the algorithm is evaluated.

In Chapter 3 the algorithm described in Chapter 2 is modified to search for the component designs that make up an optimal row-and-column design. Again an appropriate optimality criterion is used. The algorithm only provides the component designs and a second algorithm is required to arrange them as a row-and-column design. Three alternative versions of this algorithm are proposed and compared. Again the performance of the optimality algorithm is evaluated.

Some of the disadvantages of the optimality algorithms described in Chapters 2 and 3 are overcome in Chapter 4 by using a different approach in the search for optimal designs. An exchange procedure and an interchange procedure are developed which can be used to search for optimal block designs and optimal row-and-column designs. The performance of these procedures are evaluated and compared with that of the algorithms described in the previous chapters.

Chapter 1. The problem of writing an interactive computer program that gives advice on the design of an experiment.

1.1 Introduction

In recent years the computer has played an important role in the analysis of results from comparative experiments. The data from large experiments can be analysed quickly and complicated experimental designs can now be dealt with as a matter of routine. It is reasonable therefore for the experimenter to expect the computer also to provide assistance at the design stage of an experiment. Statistically informed experimenters can of course use a number of computerised algorithms to construct designs that have certain optimal properties but there does not appear to be a computer program that can be used by the layman. There is no program that can be used to help the experimenter decide on a suitable experimental design. Expressed somewhat differently, there is no computer program that attempts to replace the conversation that takes place between the experimenter and the statistical consultant.

In this chapter the problems of writing such a program are considered. The role of the statistical consultant is discussed and the views of a number of statisticians on the problems of consultancy are noted. Those aspects of consultancy that can be usefully transferred to the computer are discussed with a view to seeing how much can be undertaken by the computer.

An attempt is made to write a program that will provide guidance on the design of an experiment for comparing two treatment means. Following this the additional difficulties

that arise when more than two treatments are to be compared are noted.

1.2 The nature of the interactive computer program

The program outputs questions onto a computer terminal and invites the user to give suitable replies. This question and answer procedure continues until sufficient information has been accumulated for an appropriate experimental design to be suggested. The output from the program consists of the type of design and the number of experimental units to use and a randomised plan of the design.

1.3 The computer as a statistical consultant

If the experimenter is going to be guided by a computer program on how to design his experiment he must be able to describe the experimental conditions and the aim of the experiment to the computer. Consequently, if the advice given by the computer is to be useful the computer must in some way perform the functions of the statistical consultant whose role therefore needs to be considered. Those aspects that can be usefully transferred to the computer can then be identified and perhaps more importantly those features that cannot be transferred can be noted. In the following the terms client and experimenter will be used interchangeably and so with the terms consultant and statistician.

Consider first some definitions of the statistical consultant.

Eisenhart (1947): "The function of a statistical consultant in a research organization is to furnish advice and guidance in the collection and use of data to provide quantitative foundations for decisions."

Kastenbaum (1969): "The resident bookie ... [who] can quote

odds on most quantitative studies before they are performed ... he can tell how many observations are required to detect differences of specific magnitude for pre-assigned levels of risk."

C.P. Cox (1968): "The general aim of statistical consulting [is] the application of statistical principles and methods to the furtherance of research."

Cameron (1969): quotes Eisenhart as defining the role of the statistical consultant "to use the Mathematical Theory of Probability to determine what conclusions can be drawn and with what confidence, and in determining the amount and type of data needed as evidence."

It seems therefore that the computer could perform some of the functions mentioned above. The computer can perform complicated mathematical calculations in seconds and so is ideal for determining levels of risk, numbers of observations needed, analysing data and any calculations that can be programmed in advance. The computer then is likely to bear most similarity to Kastenbaum's definition of a consultant, rather than Cox's or Eisenhart's. The difficult problem for the computer is that of extracting the information it needs from the experimenter. This is difficult even for the human consultant:

Daniel (1969): "Every successful client-consultant relation seems to ... require <u>three</u> client conditions and <u>one</u> consultant condition. There must be a <u>good</u> problem, a <u>ready</u> client and a <u>favourable</u> organizational situation. The consultant must be <u>well prepared</u>."

Hyams (1971): "The Ideal Consultation is <u>not</u> a consultation. It is a working together, a voluntary meeting of minds and the union of energies whose prime aim is to seek 'truth'."

C.P. Cox (1968): "... consultancy is a craft an art scarcely a science"

It is clear therefore that the attitudes and personalities of the client and consultant are very important if their collaboration is to be successful. The computer program if it is to be wholly general and able to satisfy all comers must be prepared to advise the unready or difficult client as well as the well prepared. The program must at the same time possess as many qualities of the 'good' consultant as possible.

Hyams (1971) has discussed some negative stereotypes of both client and consultant and whilst he admits these stereotypes are extreme they do help focus attention on the bad features that can enter into any consultation.

Consider the negative stereotypes of the consultant:

(1) The Model Builder fits any and every data set to a model he is presently interested in or knows something about.

(2) The Hunter will subject every data set to an exhaustive and extreme computer analysis (even for a relatively simple problem).

(3) The Gong who starts every conference by drawing a bell shaped curve.

(4) The Traditionalist who is convinced nothing really new has happened in statistics since R.A. Fisher ... He views computers as the devil's work.

(5) The Randomophiliac who believes it doesn't matter what else you do as long as you have randomised.

(6) The Quantophiliac who believes that it doesn't matter if you observe what you want as long as you get a hard measurement.

(7) The More Data Yeller.

(8) The Nit Picker will always focus his attention on the inconsequential but debatable.

In fact if not carefully programmed it would be very easy for the computer to suffer from all the above faults. Some analogous faults the computer could suffer from are

(1a) The Model Builder will suggest one of the designs programmed into its memory no matter what.

(2a) The Hunter will ask an enormous amount of questions without checking whether they are relevant to the problem posed or not.

(3a) The Gong assumes the data that are to be collected will always be of a particular form.

(4a) The Traditionalist is likely to be outdated and less effective unless the program is regularly updated to include new developments in the subject of the design of experiments.

(5a) The Randomophiliac does not persevere and determine the real nature of the client's problem.

(6a) The Quantophiliac will suggest any design as long as it is a 'respectable' one without considering if it is wholly appropriate.

(7a) The More Data Yeller will suggest the most obvious and probably the largest design without determining if the client is prepared to sacrifice some of the less important points he was hoping to explore in order to reduce the size of the design. For example, confounding certain unimportant treatment comparisons in order to obtain smaller blocks.

(8a) The Nit Picker will ask an endless list of questions on a relatively unimportant aspect of the design structure. This stereotype is perhaps indistinguishable from (2a) above.

The computer program should lack all the above faults whilst possessing as many features of the good consultant as possible. Before discussing these good features it is useful first to examine some negative stereotypes of the client.

Hyams (1971) has given the following list:

(a) The Probabilist's only interest is in a significant 'p' value.

(b) The Numbers Collector has performed ... N^N experiments. Settled in his chair he has the greatest difficulty framing his questions.

(c) The Sporadic Leech is the casual aquaintance who makes it clear that the consultant's expertise is nice but unnecessary.

(d) The Amateur Statistician who calls the type of analysis to apply.

(e) The Long Distance Runner who has no time for the consultant. He asks you to find something to say about his data, anything!

Before discussing these it is useful to note Sprent's (1970) list of client types:

(i) The timid apologist who has little statistical knowledge and expresses reluctance at wasting the consultant's time.

(ii) Significant difference and least significant difference experts who think all numerical results can be made respectable by quoting significant differences and significance levels.

(iii) The one technique amateur statistician.

(iv) The believer in sacred texts or computers.

(v) The experimenter with addled statistical ideas.

(vi) The expert data handler who is not a statistician.

(vii) The statistically informed experimenter.

Pridmore (1970) added to Sprent's list the experimenter who has unlimited faith in statistics. He suggests that such questions as "How big a difference do you want to detect?" are liable to be met by the answer "Surely that is a statistical matter."

One bad characteristic that the program is likely to possess is that it might get used frequently by Sprent's types (iii) and (iv) and Hyam's type (d). The believer in computers is unlikely to question the advice given by the program, and not appreciate that it will only give reasonable advice if it can obtain the full details of the client's problem. He may be tempted to make his problem fit the program rather than vice-versa. The program could also end up as the amateur statistician's one technique for solving each and every problem, even though some problems are really outside the program's scope. This type of client is likely to find a way round some of the traps laid by the program to stop it getting misused and misapplied. Sprent's type (v) could misinterpret some of the questions or advice put out by the program. He may be tempted to reject offers by the program to give additional help with some questions and so eventually end up receiving wrong advice having misunderstood some of the questions. Sprent's type (ii) and Hyam's type (a) could abuse the program by finding ways around some of the traps set to stop the program being misused. They are only interested in getting a design that has been 'suggested by the computer'. The analogue of Hyam's type (b) is the client who has spent many hours studying the details of his proposed experiment without seriously considering what fundamental problems he hopes the experiment will solve. He may find he cannot answer the questions posed by the computer program. The computer is programmed in advance and so will not be able to 'talk around the problem' in an attempt to elucidate from the client his true aims. Sprent's type (i) is likely to find the computer program a convenient way of avoiding contact with the statistician. He probably won't mind wasting his own time

sitting at a computer terminal as much as he minds wasting the statistician's. However, if the client is timid and apologetic he may not be prepared to question the computer's advice and will accept it blindly. Sprent's types (vi) and (vii) are likely to use the program intellegently, if they need to use it at all. They will soon realise if the program is designed to help solve their particular problem or not. Hyam's types (c) and (e) are not likely to have the inclination to discuss their problems with a computer anyway and so we can probably ignore them in this discussion. Pridmore's type of client, like some of the others discussed so far, is likely to accept all the computer tells him without question.

The program if it is to serve as a consultant to a wide variety of users must be able to determine which type of client is using the program. This is likely to be difficult and even if it were possible would probably require a very large computer program. The two types of client that can most easily be prepared for in advance are the clients with no statistical expertise whatsoever and the statistically informed experimenters. The former it can be assumed requires help with everything whilst the latter is likely to require no help at all. Therefore, when the form of a particular type of program is discussed in detail below it will be assumed it is a program for the layman.

However, before moving on to discuss the positive features of consultancy attention should be drawn to two sorts of relationship that can take place between a client and a consultant. Boen (1972), for example, has mentioned the problems that can arise when the client insists on a parent-to-child relationship, i.e. when the client insists on being the boss and deciding what will be done. Boen has also mentioned

the child-to-parent relationship, i.e. when the client expects the consultant to make all the decisions even though some of them relate to the client's own subject area. If a parent-to-child relationship exists then the client is likely to find ways of forcing the computer program to follow the route he thinks it should follow. If a child-to-parent relationship exists then the client is likely to accept the computer's advice without question and not be prepared to consider that the computer could make a mistake.

So far attention has concentrated only on the negative aspects of consultancy. It is of course important to identify what makes a good consultant. However, before doing this it should be noted that there are many differences between statistical consultants. In fact Healy (1973) has likened the population of statisticians to "a single linkage cluster in which a good deal of chaining has taken place. Any two statisticians can be related by a series of quite close similarity links passing through other statisticians; but this is quite consistent with many pairs being formed by people having relatively little in common." Therefore, when trying to identify what characteristics a good consultant will possess one is really trying to identify the main things they have in common. Some characteristics possessed by the consultant in an agricultural research station, for example, may not be entirely useful to the consultant in an industrial laboratory.

A number of articles in the literature discuss the training the new statistical consultant should be given before he embarks on his career. For example, see C.P. Cox (1968), Boen (1972), Watts (1970), Schucany (1972) and Greenfield (1979). One of the important things they stress is the ability to communicate with the client. Other writers have made the same

point:

Tarter and Berger (1972): "Clarity of communication is also the hallmark of the consultation process."

Perry (1970): "A major problem of consultancy is communication."

Marquardt (1979): "learn the client's subject matter and lingo."

Obviously good communication means that both client and consultant should understand each others language, and a number of writers have mentioned this. For example, C.P. Cox (1968) suggests that consultancy problems in biology would be greatly reduced if the experimenter were fully trained in both the biological and statistical disciplines. Finney (1956) has said that the experimenter "and his statistical colleagues should have sufficient understanding of one another's disciplines to avoid the errors that can result from imperfect communication of ideas."

It is therefore essential that the computer be programmed so that it outputs questions and advice that can be easily understood by the client. Communication between colleagues in the same profession or business is often speeded up by using technical terms or jargon. To the outsider such terms are meaningless and seem to make communication more difficult. However to those familiar with it jargon can be a short-hand way of expressing lengthy phrases or complex ideas. If the computer program is used say, in a particular agricultural research station or industrial laboratory the experimenters there will probably find the program more convenient to use if it uses their own particular jargon rather than expressing things in generalities. For example, the agriculturalist will probably refer to his experimental units as "plots", the engineer to "units" and the chemist to "determinations".

Some other characteristics that a good consultant should

possess are the following:

Daniel (1969): "A good consultant has a good scientific background, enjoys cooperative enterprises, a few years experience and some statistical training."

C.P. Cox (1968): "ability to work in a team, understanding of inductive scientific method."

The computer will only possess those features that can be programmed in advance and certainly a scientific background and an understanding of inductive scientific method are probably going to be difficult to program. The computer is unlikely to 'enjoy' cooperative enterprises but will possess a good deal of "statistical training" in the form of numerous computer subroutines for performing various statistical calculations. The computer can also learn from experience by keeping records of all previous encounters with clients. For example, the most frequently suggested design could be noted. The transformation of the data that is usually required after the experiment could be noted. Sizes of variances or coefficients of variation could be stored. Whether the experimental material was usually divided into blocks or rows and columns could be recorded.

Therefore, to a limited extent some of the good features of a consultant can be programmed and clients should be able to receive useful advice. However, being programmed in advance the computer will be less likely to have the ability to "talk around the problem" in order to get the client to explain fully all the possible complications of his problem.

What is meant by talking around the problem can be best explained by quoting Salsburg (1973) who says:

"I spend my time asking stupid questions. I know when the experiment is finished I will have to analyse the data...

I must protect myself from impending chaos.

... I ask such questions as whether it is possible to observe something every fifteen minutes ... or whether the thing they've given a name to can in fact be measured at all. I ask them what can go wrong ... I think of the in between outcomes, the two correlated variables that happen to go different ways, the test tube someone is bound to drop, the patient who revives from death's door on placebo. Practical experience has taught me that almost every observable outcome is in between."

Being programmed in advance the computer will also lack the versatility of the consultant. A quality Greenfield (1979), for example, stressed was important.

It should also be noted from the above remarks that the computer will have an obligation to analyse the client's data or be able to suggest where advice on the analysis can be obtained.

What is apparent from all the above remarks is that a computer program is only likely to be able to give sound advice on a limited number of problems. Safeguards must be introduced so that the program and the client can quickly realize if the problem to be solved is within the computer's capability. Therefore it is required that

(1) the problem be clearly defined so that the computer has no difficulty deciding if it can solve it or not.

(2) the language used by the program and by the experimenter must be understood by both.

Such requirements are likely to be satisfied for problems which are

(a) fairly commonplace so far as the client's particular organization is concerned

(b) relatively straightforward so that an unambiguous solution is often possible.

Of course, if the problem was too simple there would be no need for a computer. A handbook of prepared solutions is all that would be required. The sort of problem best suited to the computer is one for which there are quite a few possible answers depending on the client's particular requirements and where lengthy calculations are necessary. Such a problem is considered below.

However, before going on to consider this problem in detail it must be borne in mind that the computer is much more likely than the human consultant to make, in the words of Kimball (1957), an "error of the third kind". That is, where "the right answer is given to the wrong problem." This happens when due to a breakdown in communications between the client and the consultant the problem that appears to require a solution is not the one the client brought to the consultant. Such errors can be avoided Kimball says "with proper preparation, sufficient patience and persistent questioning of the experimenter." It is the computer's inability to talk around the problem which is likely to increase its error rate. It is therefore essential to try and reduce such errors by including in the program adequate checks that the problem presented by the client is able to be solved by the computer and that the client has not unknowingly perhaps been forced to modify his problem to fit in with the program's pre-set repertoire of solutions. Sprent's (1970) "timid apologist" and experimenters with "addled statistical ideas" are likely to fall into the trap of believing the computer knows best as is Sprent's "believer in computers". The "expert data handler" and the "statistically informed experimenter" are perhaps less likely to be led astray.

It is clear that the program must by printing out a suitable message make it obvious what type of problem it is able to advise on. If the computer can advise on the problem it should check that any assumptions made to achieve a solution are satisfied by the client's particular problem.

One difficulty already mentioned is that if the program is going to be totally successful it must be able to advise all client types. Even if this were possible it would probably require a large and unmanageable computer program. If the program is written to deal with clients who have different amounts of statistical expertise the computer might by mistake pitch its questions at the wrong level and an unsatisfactory consultation would result. As mentioned previously, the clients who can be most easily catered for in advance are clients who either know nothing about statistical techniques or know all they need to solve their particular problems. The client who knows nothing can be offered additional explanation on almost every point without fear that he will find such offers tedious. The expert will require no additional help and will regard the computer as something that will do his calculations and confirm his ideas.

To highlight some of the particular problems that have to be overcome the rest of this chapter is concerned with writing a computer program that will give advice on the design of an experiment for comparing two treatment means. This to the statistician is a relatively straightforward problem that has in most cases an unambiguous solution. This problem does require certain calculations to be performed and has sufficient variations to make it worthwhile programming a computer to solve it.

1.4 The information needed to design an experiment

The information needed to design a comparative experiment will depend on the peculiarities of the proposed experiment. However, certain basic questions have to be answered before an experiment can be designed and Finney (1956), for example, has suggested that for field experiments these questions can be regarded as being of five types:

1) Choice of treatments for comparison.
2) Specification of the units to which the treatments are to be applied.
3) Specification of the measurements to be made on each unit.
4) Rules by which units are allocated to the various treatments.
5) Number of replications of any one factor or of the whole experiment.

Nelder (1956) produced a similar list of questions, again for field experiments:

(a) Has anyone already answered the questions?
(b) What treatments are needed in the experiment?
(c) Where should the experiment be done?
(d) How should the experiment be laid out?
(e) What records should be made?

D.R. Cox (1958, p.4) speaking more generally suggested the following should be considered:

(i) The choice of treatments to be compared, of observations to be made and of the experimental units to be used.
(ii) The method of assigning treatments to experimental units and the decision about how many units should be used.

Jeffers (1978) considered in more detail the information required and produced a check list of 71 questions that need to be answered.

The above lists emphasize the limitations of the computer.

It is in the making of the important decisions regarding the choice of treatments, choice of experimental units, where the experiment should be carried out, and what records should be made, that the computer is likely to be of least assistance. Such questions are of too general a nature to make it easy for a prepared set of questions to be stored in the computer. These questions really need discussion and "talking around" before decisions can be made. Therefore, in this chapter, it will be assumed such questions have been answered before the experimenter approaches the computer for help. With only two treatments there is probably little that the statistician can advise except to bring to the attention of the experimenter that he should consider including a control as one of his treatments. With more than two treatments the statistician can offer more useful advice, particularly if the response to be observed is measured on a quantitative scale and the treatments are to be made up of combinations of levels of certain factors. The choice of the number of levels of each factor will depend, for example, on whether a linear or quadratic relationship is assumed to exist between the levels of a factor and the response. The choice of experimental units is also something the statistician could advise upon in particular experiments. For example, in a field experiment he may be able to decide whether a few large experimental plots are desirable or a large number of smaller plots would be preferred.

In order to decide what information the computer program must obtain before it can design an experiment to compare two treatment means it is necessary to first define the problem.

In this section and the following it will be assumed that data are to be sampled from two populations:

x_{1i}, $i = 1, 2, \ldots, n_1$ will be the sample of size n_1 from

population 1, and

x_{2i}, $i = 1, 2, \ldots, n_2$ will be the sample of size n_2 from population 2.

Further, it will be assumed that all x_{1i} and x_{2i} can be considered mutually independent and that

x_{1i} is distributed normally with mean μ_1 and variance σ^2 and x_{2i} is distributed normally with mean μ_2 and variance σ^2. The situation where some of these assumptions are not valid will be considered in Section 1.6.

The null hypothesis to be tested is

$$H_o : \mu_1 = \mu_2.$$

The alternative hypothesis will be one of the following depending on the experimenter's requirements

$$H_1 : \mu_1 \neq \mu_2$$

$$H_2 : \mu_1 > \mu_2$$

$$H_3 : \mu_2 > \mu_1.$$

To test the null hypothesis the following statistic will be used.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_{\bar{x}_1 - \bar{x}_2}},$$

where $\bar{x}_1$ is the mean of the x_{1i} values

$\bar{x}_2$ is the mean of the x_{2i} values

and $S_{\bar{x}_1 - \bar{x}_2}$ is an estimate of the standard deviation of the difference $(\bar{x}_1 - \bar{x}_2)$.

On the null hypothesis, and under the assumptions stated above, this statistic will follow the t-distribution on v degrees of freedom, where v depends on n_1 and n_2 and on the design of the experiment.

More generally it will be necessary to consider the null

hypothesis that $|\mu_1 - \mu_2| \leq \delta$ where δ is the size of difference between the treatment means it is useful to detect in the experiment. To test this null hypothesis the statistic t is redefined as

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{S_{\bar{x}_1 - \bar{x}_2}}.$$

The experimenter will expect the computer to design an experiment to test his null hypothesis and one thing he will want to know is how many experimental units to use. Cochran and Cox (1957, p.18) proposed a method of determining the approximate number of experimental units to be used for each treatment. Their argument is as follows.

If the observed difference between the treatment means is d then d is judged to be significantly different from zero at the 100α% significance level if

$$d \geq \sqrt{\frac{2}{n}}\, s t_1, \qquad (1.4.1)$$

where n is the number of observations on each treatment, s is the standard error per unit of the observations and t_1 is the tabulated value of t at the 100α% level, and will depend on whether a one-sided or two-sided test is required. The probability of this happening is required.

Now,

$$d \geq \sqrt{\frac{2}{n}}\, s t_1$$

if and only if

$$t_2 = \frac{d - \delta}{\sqrt{\frac{2}{n}}\, s} \geq \frac{\sqrt{\frac{2}{n}}\, s t_1 - \delta}{\sqrt{\frac{2}{n}}\, s} = -\ell, \text{ say.}$$

Hence the probability that (1.4.1) happens is the probability that the value of t should exceed $-\ell$.

The flaw in the argument is the assumption that the

value of s can be assumed equal to σ, whereas s is actually a random variable. If the argument is carried through algebraically, it gives the following rule for calculating the approximate value of the power P.

1) Find t_2 from the relation

$$\delta = \sqrt{\frac{2}{n}}\,\sigma(t_1 + t_2) \qquad (1.4.2)$$

where n is the number of observations to be taken on each treatment

t_1 is the significant value of t in the test of significance (with care to distinguish between one-sided and two-sided tests).

2) Set $P = 1 - (\frac{1}{2})p_2$

where p_2 is the probability corresponding to t_2 in the ordinary (two-sided) t-table. The degrees of freedom in t_1 and t_2 are those available for the estimate of σ.

A rule is required for determining the value of n and this can be obtained by inverting the above rule. Given the value of P

$$n \geq \frac{2\sigma^2}{\delta^2}(t_1 + t_2)^2 \qquad (1.4.3)$$

where t_2 is the $2(1 - P)\%$ point of the t-distribution on v degrees of freedom. The value of v will depend on the design of the experiment as well as n.

In order to be able to use the above formula, the following information must be obtained.

1) Whether a one- or two-sided hypothesis test is required.

2) A value for δ.

3) A value for σ.

4) Values for the significance level and power.

5) Whether the experimental units can be paired or not.

1.5 Obtaining information using the computer

The information is probably easy for the human consultant to obtain in a conversation with the client. The consultant can quickly see if the client does not understand the questions he is asked and can explain or illustrate certain points to clear up any difficulties. The computer, however, can only proceed on the assumption that the client has understood each question as printed. There is no way of telling if the client had difficulty answering a question. What the computer can do however, in an attempt to help the client answer difficult questions, is to offer help. This can be done as follows. The computer program first outputs a direct question on some aspect of 1) to 5) as listed at the end of the previous section. Whenever possible the valid replies to the question are listed at the end of the question. These answers will usually be yes (Y), no (N), help (H), or an integer (1, 2, ...,) indicating a choice from a list of alternatives. If the H reply is given the program outputs some additional text to help the client give his or her reply. When in the following, some examples of questions are given, the above convention will be used.

The computer program, it will be noted, cannot assume any prior information about the client's experiment unless that information has already been stored in the computer's memory. Consequently, the questions output by the program must be such that they cover every eventuality so that a valid experimental design is produced at the end of the computing session. There will therefore be more questions asked by the computer than by the human consultant. The wording of the questions and the order of asking them is therefore extremely important. The answer from one question must lead naturally

onto the next and unlike a human consultation there can be no skipping back if it is realized that some particular piece of information is required but was not obtained earlier.

To illustrate how difficult this aspect of writing the program can be even for a relatively straightforward program, each of the five points listed at the end of the previous section will be considered in turn. Then the order of asking questions on these points will be considered. It should be noted that in this section all the assumptions made in the previous section are taken to hold. The additional difficulties that arise when these assumptions have to be checked in the program are considered in the next section.

1.5.1 One or two sided hypothesis test

This information is relatively straightforward to obtain. The client could be asked by the computer:

DO YOU WANT TO TEST IF THE AVERAGE RESPONSE OF ONE TREATMENT IS

1) HIGHER THAN THE OTHER?

2) LOWER THAN THE OTHER?

3) HIGHER OR LOWER? (I.E. A DIFFERENCE IN EITHER DIRECTION IS OF INTEREST)

(1, 2, 3 OR H)

In reply to H the following could be given

IF ANSWER 1) IS CHOSEN THEN IT MUST BE ASSUMED THAT THE AVERAGE RESPONSE OF THE TREATMENT CONCERNED COULD NOT, UNDER ANY CIRCUMSTANCES, BE LOWER THAN THE OTHER'S.

IF ANSWER 2) IS CHOSEN THEN IT MUST BE ASSUMED THAT THE AVERAGE RESPONSE OF THE TREATMENT CONCERNED COULD NOT, UNDER ANY CIRCUMSTANCES, BE HIGHER THAN THE OTHER'S.

IF ANSWER 3) IS CHOSEN THEN EITHER SORT OF DIFFERENCE

COULD OCCUR AND THERE IS AN EQUAL INTEREST IN BOTH.

THE CHOICE BETWEEN 1), 2) AND 3) IS IMPORTANT AS IT AFFECTS HOW MANY EXPERIMENTAL UNITS TO USE.

WHICH REPLY TO THE ABOVE QUESTION DO YOU WISH TO GIVE?

(1, 2 OR 3)

1.5.2 The value of δ

Questions to determine the value of δ are more difficult to phrase because the client may be familiar with expressing δ as a percentage of the mean of one of the treatments, or in the units used to measure the response, or in some other way. However, as can be seen from formula (1.4.3), the absolute value of δ is not needed, only its value relative to σ, the standard deviation. If the client is likely to express δ as a percentage of the mean he is quite likely to express σ in a similar fashion. The computer program therefore needs to check that δ and σ have been expressed in similar ways so that a value of (σ/δ) can be obtained.

Here, however, it is assumed that the client will give a value for δ. Pridmore (1970), as noted earlier, pointed out that the question "How big a difference do you want to detect?" is likely to be met by the answer "Surely that is a statistical matter?"

This problem could be approached as follows.

IT IS ESSENTIAL TO KNOW WHAT SIZE OF DIFFERENCE BETWEEN THE TREATMENT MEANS THE EXPERIMENT IS REQUIRED TO DETECT.

BEFORE BEING ASKED TO SPECIFY A VALUE DO YOU WANT FURTHER EXPLANATION?

(Y OR N)

In reply to Y the following could be printed.

AN EXPERIMENT CAN BE DESIGNED TO DETERMINE IF ANY DIFFERENCE EXISTS BETWEEN THE TWO TREATMENT MEANS. HOWEVER, YOU MAY DECIDE THAT DIFFERENCES SMALLER THAN A CERTAIN VALUE ARE OF NO IMPORTANCE.
THEREFORE IF A USEFUL EXPERIMENT IS TO BE DESIGNED IT IS ESSENTIAL THAT YOU SPECIFY A VALUE FOR THE SIZE OF THE SMALLEST DIFFERENCE IT IS USEFUL TO DETECT.

IF YOU DO NOT SPECIFY A VALUE THE PROGRAM CANNOT CONTINUE.

The client would then be asked if a value is to be specified. If the answer is N then the program cannot continue. Perhaps the best advice that could be output by the program, if the answer N is given, is that suggested by Finney (1970) when discussing an interactive data analysis program:

GO CONSULT A STATISTICIAN.

However, even though the statistician can help explain at greater length why a value of δ is needed, it is really up to the client to reconsider his aims. If he cannot specify a value it could be that his own ideas are muddled.

If a value of δ can be specified the program could print

IN WHICH WAY WOULD YOU PREFER TO INPUT THIS DIFFERENCE?

1) AS A PERCENTAGE OF THE MEAN OF ONE OF THE TREATMENTS.
2) IN THE UNITS OF MEASUREMENT USED TO MEASURE THE RESPONSE.
3) SOME OTHER WAY.

(1, 2 OR 3)

If reply 1) is given then the program would need to know which treatment the percentage is referred to, as this will be needed when a value of σ is asked for. Reply 3) has been

included in case, at the time of writing the program, some other way of specifying δ, peculiar to the clients who will use the program, is to be used.

1.5.3 A value for σ

D.R. Cox (1958, p.165) suggested five ways of obtaining an estimate of σ^2. Three of them refer to estimates obtained after the data has been collected and the other two refer to estimates obtained before the experiment takes place. These last two are (i) from the results of previous similar experiments and (ii) from theoretical considerations. The second of these is no help in this section as the fact that the observations are normally distributed does not enable an estimate of σ^2 to be obtained. If the data were sampled from a Poisson distribution for example, then knowledge of the treatment mean gives information on the treatment variance. Consider, therefore, suggestion (i).

An important idea to put over to the client, who is assumed to have little or no knowledge of statistics, is that his data will show some degree of variability and that it is essential to have some quantitative estimate of the variability.

The computer program could therefore lead into the set of questions designed to obtain a value for σ by printing

YOU ARE AWARE NO DOUBT THAT DIFFERENT MEASUREMENTS ON THE SAME TREATMENT WILL SHOW SOME VARIABILITY. IT IS ESSENTIAL TO OBTAIN A NUMERICAL ESTIMATE OF THE SIZE OF THIS VARIABILITY.

The program would then ascertain if past data are available by printing

HAVE YOU IN THE PAST PERFORMED SIMILAR EXPERIMENTS TO THE ONE YOU ARE PLANNING NOW?

(Y OR N)

If the reply was Y the program could follow up by asking

IS THE VARIABILITY OF THE DATA ON EACH TREATMENT COLLECTED IN THE PAST EXPERIMENTS SIMILAR TO THE VARIABILITY OF THE DATA TO BE COLLECTED IN THE NEW EXPERIMENT?

(Y OR N)

If the reply to this question was Y there is the possibility that an estimate of σ can be obtained. However, it is desirable to know how many experiments have been done in the past. If only one or two have been done then the estimate of the error variance could be unreliable as an unusually high or low variance could have been obtained. If only one or two experiments have been done it is probably a good idea to ask the client to consider these estimates in the light of his own knowledge to decide if a representative value of the variance was obtained in the past.

If past data are available it is also necessary to consider what sort of experimental design was used, if any, in the past experiments. This brings up the point of whether the experimental units were paired or not. This is considered in sub-section 1.5.5.

If no estimate of σ is forthcoming, either from past data or other considerations a different approach must be adopted or the computer program cannot continue.

One reason why an estimate of σ is unavailable could be that the data from the past experiments were not kept. If, however, a record of treatment means or ranges or coefficients of variation have been kept these might be used to obtain a rough estimate of σ.

If the range of a treatment is available along with

the number of data on which the range is based, then the range could be compared with tables of (σ/range) which have been tabulated by Pearson and Hartley (1954), for example. If the number of data is not available a rough estimate of σ could be obtained by arguing that the size of the range is likely to be approximately equal to a 95% confidence interval for the mean and to take the estimate of σ as range/4.

If the values of the treatment means are available, along with a coefficient of variation, then a rough estimate of σ is (coefficient of variation $\times$ mean).

If no past information is available that will lead to an estimate of σ one further approach is to ask the client to express an opinion on the value of σ. The client is assumed to have little knowledge of statistics and so is unlikely to understand the concepts of variance and standard deviation. The idea of variability is therefore best put over by considering the treatment ranges.

A possible way for the program to do this could be as follows.

THE VARIABILITY OF THE MEASUREMENTS ON A TREATMENT CAN BE THOUGHT OF IN TERMS OF THE RANGE. THE RANGE IS THE DIFFERENCE BETWEEN THE LARGEST MEASUREMENT AND THE SMALLEST.

OBVIOUSLY THE MORE VARIABLE ARE THE DATA THE LARGER THE RANGE.

FOR ONE OF THE TREATMENTS CAN YOU GIVE A VALUE FOR THE LIKELY SIZE OF THE RANGE?

(Y OR N)

If a range is specified σ can be estimated as described above.

If a range cannot be given the program could try yet

another approach.

MAYBE YOU CANNOT SUGGEST A VALUE FOR THE RANGE BUT COULD YOU SAY THE RANGE IS LIKELY TO BE ± X% OF THE VALUE OF THE MEAN OF A TREATMENT?

(Y OR N)

If the reply is Y then a value for X would be asked for. If the reply is N then the program cannot continue and the client would be advised to consult a statistician.

In the above, when discussing if past data were available, the possibility of having to combine estimates of the variance from more than one experiment was not mentioned. One way of combining the estimates would be to take a weighted mean of the estimates. The weights being inversely proportional to the number of data used to calculate each estimate. Doubtless other ways could be devised. A more difficult problem arises when different designs were used in the past experiments. It is essential to elucidate from the client the exact nature of the variance he has estimated. More will be said on this in sub-section 1.5.5 when pairing of experimental units is considered.

If no estimate of σ can be obtained from the client using the approaches given above then one alternative is to suggest that a pilot experiment is performed to obtain an estimate of σ. A two-stage procedure, similar to that suggested by Stein (1945) could be suggested. From the first stage experiment the total size of the experiment needed to achieve the desired significance level and power to detect a difference of a given size could then be calculated.

It might also be useful before asking about past data and attempting to explain the meaning of variability to ask a direct question first, e.g.

HAVE YOU AN ESTIMATE OF THE VARIANCE OR STANDARD DEVIATION OF THE DATA TO BE COLLECTED IN THE EXPERIMENT?

(Y OR N)

The possibility arises here of the reply N being given because an estimate is available but the client does not have it with him. The option to stop the program and to return with an estimate must therefore be included.

1.5.4 The significance level and power

To the client without any statistical training the ideas of significance and power are probably the most difficult to understand. The statistican can visualize the shape of the frequency curve of the normal distribution for example, and is able to picture $2\frac{1}{2}\%$ the area under the curve being shaded at either end to represent the 5% significance level of a two-sided test. The layman however, has to have such ideas explained and this must be done by a short piece of text output by the computer program. Once the client has grasped the idea of a significance level, for example, he will not want a full explanation each time. One approach is therefore to first ask a direct question:

WHAT VALUE OF THE SIGNIFICANCE LEVEL IS APPROPRIATE?

1) 5%

2) 1%

3) ONE SUGGESTED BY YOU.

(1, 2, 3 OR H)

In reply to H the program could print

AS YOU WILL KNOW THE DATA FROM A TREATMENT WILL SHOW SOME DEGREE OF VARIABILITY. CONSEQUENTLY THE DATA MAY INDICATE A DIFFERENCE BETWEEN THE MEANS WHEN THERE IS

NONE. THE OBSERVED DIFFERENCE BEING A CHANCE RESULT.

THE EXPERIMENT MUST HAVE ENOUGH OBSERVATIONS TO ENSURE THAT THE CHANCES OF MAKING THE ABOVE MISTAKE ARE QUITE SMALL. STATISTICIANS REFER TO THE PROBABILITY OF FINDING A SPURIOUS DIFFERENCE AS THE "SIGNIFICANCE LEVEL", IT IS OFTEN EXPRESSED AS A PERCENTAGE, FOR EXAMPLE, 5% REPRESENTS A 1 IN 20 CHANCE AND 1% REPRESENTS A 1 IN 100 CHANCE. HOW HIGH A PROBABILITY CAN YOU TOLERATE OF FINDING A DIFFERENCE WHERE NONE EXISTS?

USUALLY 5% IS ADEQUATE BUT IF YOU WANT TO REDUCE THE RISK EVEN FURTHER CHOOSE 1%. IT SHOULD BE REMEMBERED THAT THE SMALLER THE PROBABILITY THE GREATER THE NUMBER OF OBSERVATIONS THAT WILL HAVE TO BE TAKEN.

WHICH REPLY DO YOU WISH TO GIVE?

(1, 2 OR 3)

One of the client's requirements is that the experiment should have a good chance of detecting a difference between the treatment means of size δ. That is, the power of the hypothesis test must be considered. Suppose earlier in the program a value of δ had been decided upon. The program taking the same approach as for the significance level could ask

WHAT SIZE OF POWER IS APPROPRIATE?

1) 80%

2) 90%

3) A SIZE SUGGESTED BY YOU.

(1, 2, 3 OR H)

In reply to H the program could print

IN REPLY TO THE LAST QUESTION YOU CHOSE A SIGNIFICANCE LEVEL OF 5%. WHAT YOU HAVE SAID THEREFORE IS THAT YOU ARE PREPARED TO TAKE A 1 IN 20 CHANCE OF DECIDING A

DIFFERENCE EXISTS WHEN THERE IS NONE.

HOWEVER, IF A DIFFERENCE OF SIZE δ DOES EXIST YOU WANT A GOOD CHANCE OF DETECTING IT. STATISTICIANS REFER TO THE SIZE OF THIS CHANCE AS THE POWER.

AGAIN, THE HIGHER THE POWER THE GREATER THE NUMBER OF OBSERVATIONS THAT WILL BE NEEDED. A VALUE OF 80% IS QUITE REASONABLE.

WHICH REPLY TO THE ABOVE QUESTION DO YOU WISH TO GIVE?

(1, 2 OR 3)

In the above replies to H the values suggested could be put equal to values particularly appropriate to the sort of experiments the computer program will be required to design.

1.5.5 Pairing the experimental units

When comparing two treatment means it is often a good idea to pair the experimental units so that the units in a pair are more similar than they are to the units as a whole. In some situations it may also be useful to take account of the ordering of the units in a pair. To simplify matters a little in this chapter, such ordering of units will not be considered. The computer program then, must somehow determine if the experimental units should be paired or not. The aim of pairing is to eliminate from the treatment comparison any systematic variation between units. It is essential therefore that the client realizes that variability needs to be reduced. This can be done in stages as follows.

IT IS IMPORTANT TO CONSIDER THE VARIABILITY OF THE MEASUREMENTS THAT WILL BE TAKEN IN THE EXPERIMENT. DO YOU WANT FURTHER EXPLANATION OF THE TERM VARIABILITY?

(Y OR N)

In reply to Y the program could print

SUPPOSE ONLY ONE TREATMENT WAS USED IN THE EXPERIMENT. IF THERE WAS ABSOLUTELY NO DIFFERNCE BETWEEN THE UNITS AND IN THE WAY THE MEASUREMENTS WERE TAKEN, AND IN THE CONDITIONS THAT PREVAILED DURING THE COURSE OF THE EXPERIMENT THEN ALL MEASUREMENTS WOULD BE IDENTICAL.

HOWEVER IN PRACTICE IT IS IMPOSSIBLE TO ENSURE COMPLETE UNIFORMITY OF CONDITIONS, UNITS AND METHODS OF MEASUREMENT AND CONSEQUENTLY THE MEASUREMENTS WILL EXHIBIT SOME DEGREE OF VARIABILITY.

This could then be followed by

SOME VARIABILITY CAN USUALLY BE ATTRIBUTED TO DIFFERENCES AMONG THE UNITS. IS PAIRING OF UNITS A POSSIBILITY?

(Y, N OR H)

In reply to H the program could print

IT IS SOMETIMES POSSIBLE TO GROUP THE UNITS INTO PAIRS IN SUCH A WAY THAT THE UNITS IN A PAIR ARE MUCH MORE SIMILAR TO ONE ANOTHER THAN THEY ARE TO THE GROUP OF UNITS AS A WHOLE. FOR EXAMPLE,

1) IF THE UNITS WERE THE KIDNEYS OF AN ANIMAL IT IS LIKELY THAT THE TWO KIDNEYS FROM THE SAME ANIMAL ARE MORE SIMILAR TO EACH OTHER THAN THEY ARE TO THE KIDNEYS OF THE OTHER ANIMALS IN THE EXPERIMENT.

2) THE EXPERIMENTAL CONDITIONS COULD CHANGE PROGRESSIVELY AS TIME GOES ON. THE LATER MEASUREMENTS BEING SYSTEMATICALLY DIFFERENT FROM THE EARLIER ONES. HERE TWO MEASUREMENTS TAKEN CLOSE TOGETHER IN TIME WILL BE MORE SIMILAR THAN TWO TAKEN FAR APART.

WHICH REPLY TO THE ABOVE QUESTION DO YOU WISH TO

GIVE?

(Y OR N)

The program would then note if pairing was possible for possible use later.

It will be recalled that when discussing how to obtain an estimate of σ in sub-section 1.5.3, one method was to use past data. When discussing these past data with the client it is important to discover if pairing was used in some or all of the past experiments. It is important to know if the value of the variance or standard deviation that is quoted by the client refers to the variance of the paired differences or the variance of the data on each treatment. The appropriate variance for the present experiment is the one that must be obtained from the client. These remarks have implications on the order the program asks the questions. When trying to obtain an estimate of the variance the possibility of pairing will have to be explained. If pairing is possible an estimate of the variance of the paired differences will have to be somehow obtained from the client. It is necessary therefore to ensure the client understands the need for pairing before moving to questions on the variance.

If pairing is possible in the present experiment it will be necessary to decide if pairing is worthwhile or not. That is, the likely increase in precision of the treatment comparison will have to be weighted against the loss of error degrees of freedom if pairing is used. At this point therefore, it is convenient to digress from the structure of the program to consider how to decide if pairing is worthwhile or not.

1.5.6 Pairing versus no pairing

The degrees of freedom for the error sum of squares in

the analysis of variance will depend on whether pairing has been used or not. If pairing has been used then n pairs will provide (n - 1) degrees of freedom for error. If no pairing has been used the 2n experimental units will provide 2(n - 1) degrees of freedom for error. Whether pairing has been used or not will affect the standard error of the difference between the treatment means when using formula (1.4.3). In fact, if pairing is used the formula becomes

$$n \geq \frac{\sigma_D^2}{\delta^2} (t_1' + t_2')^2, \qquad (1.5.1)$$

where σ_D^2 is the variance of the difference between the two observations in each pair, and t_1' and t_2' are analogous values to t_1 and t_2 in formula (1.4.3) except now the degrees of freedom are (n - 1) instead of 2(n - 1).

If x_1 is one observation in a pair and x_2 the other then

$$V(x_1 - x_2) = \sigma_D^2 = 2\sigma^2 - 2\mathrm{Cov}(x_1, x_2).$$

If the pairing is at all useful then $\mathrm{Cov}(x_1, x_2) > 0$. Therefore,

$$\sigma_D^2 < 2\sigma^2.$$

However, even though $2\sigma^2$ has been reduced to σ_D^2, the paired experiment is not necessarily more precise because the degrees of freedom are reduced.

Suppose $2\sigma^2$ and σ_D^2 are known. Then given δ, the significance level and the power, the unpaired experiment would have size $2n_1$, where

$$n_1 \geq \frac{2\sigma^2}{\delta^2} (t_1 + t_2)^2,$$

and t_1 and t_2 are on $2(n_1 - 1)$ degrees of freedom.

The paired experiment, on the other hand, would have

size $2n_2$, where

$$n_2 \geq \frac{\sigma_D^2}{\delta^2} (t_1' + t_2')^2,$$

and t_1' and t_2' are on $(n_2 - 1)$ degrees of freedom.

The paired experiment would be preferred if $n_2 < n_1$, and no pairing would be preferred if $n_1 < n_2$.

The difficulty is obtaining estimates of σ^2 and σ_D^2 so that the above calculations may be done.

If past data are available the human consultant could examine these with a view to deciding if pairing is likely to be useful. The computer program would have difficulty doing this because it is hard to plan a general set of questions in advance when the past experiments could be of quite different types. One approach that could be adopted that does not specifically ask for information on past experiments is the following. Of course if pairing is possible in the present experiment and was always used in the past an estimate of σ_D^2 might be readily available.

In the earlier part of the program a value for the range of one of the treatments could be obtained. The following hypothetical situation could then be put to the client. Suppose the units are paired and one treatment is applied to all units, i.e. both units in a pair receive the same treatment. The client is then asked to consider the largest difference between the measurements in a pair. This is considered as a range for the paired differences. This range is likely to be smaller than the treatment range if pairing is effective. The client is then asked to give a figure which indicates how much the range of the difference is likely to be smaller than the treatment range. For example, is it 10% smaller? 20% smaller? etc.

However, in practice this might be too much to ask the client. A better approach before writing the program might be to first consider the type of experiments the program is to advise on. If pairing is thought likely to be useful in these experiments then the program could be set to always advise a paired experiment if the client indicated that pairing was possible.

In the above sub-sections some details of the difficulties that have to be overcome have been considered. One further point to consider, once the questions and their wording have been decided, is the order in which the questions have to be asked. As mentioned earlier there should be a natural flow of questions with one answer leading naturally on to the next. One possible ordering is given in Table 1.5.1.

Once the size of the experiment has been calculated this can be put to the client. It may be that the ideal size is too large for the client's resources because the client has been too ambitious. This in itself would be a useful outcome as the client, if not advised by the program, would have wasted valuable experimental units. A dialogue between the client and the program could then ensue with the aim of reaching a compromise. Perhaps a significance level of 5% could be used instead of 1%, or the power reduced from 90% to 80%, or the difference to be detected made a little larger. For given values of σ^2 (or σ_D^2), significance level and power, the smallest value of δ that could be detected using all the available units could be calculated.

If the experiment has been accepted by the client then a detailed randomized plan of it could be printed. The importance of the randomization could be emphasized and details of where the analysis of the data may be done could also be

pointed out.

Even when the assumptions can be assumed to hold it is apparent that a large number of questions and associated text must be printed by the computer program. If the assumptions must be checked by the program then it has to be enlarged. Some examples of the extra difficulties involved if the assumptions have to be checked are given in the next section.

Table 1.5.1 Order of asking questions

1) Explain the meaning of variability.
2) In particular, discuss the variability of the experimental units with a view to deciding if pairing is possible.
3) If pairing is possible, and pairing was always used in the past experiments, attempt to get a direct estimate of σ_D^2. If pairing is not possible and was never used in the past attempt to get a direct estimate of σ^2. If neither of these applies then attempt to get an estimate of σ^2 in some form or other, e.g. as a function of the range. If pairing is possible attempt to determine the relative sizes of σ_D^2 and σ^2.
4) Determine δ, the size of the smallest difference to be detected.
5) Decide if a one- or two-sided hypothesis test is required.
6) Determine the significance level.
7) Determine the power.
8) Decide, if necessary, if pairing is worthwhile or not.
9) Determine if the size of the design is appropriate.

1.6 Checking on the assumptions

In the last section the advice printed by the computer program was made on the assumption that the observations to be taken in the experiment would be

(i) normally distributed

(ii) such that each treatment had the same variance

(iii) independent.

If these fail to hold the program is likely to give wrong advice. However, as Scheffé (1959, Chapter 10) concludes, the effects of departures from the above are likely to have little affect on inferences about means if the number of observations on each treatment is large. Further, he concludes that even if the number of observations on each treatment is not large the effects of departures from the above are not serious if each treatment is replicated the same number of times. However, if the reason for the failure of the assumptions can be determined then, in some circumstances, it is possible to remedy the situation, e.g. by a suitable transformation of the data.

It is important then, that the program, whenever possible, attempts to check the assumptions. The experienced human consultant quite often has no difficulty deciding on the matter. This may be because of previous encounters with similar data or because of an ability to sense that things are not what they seem. This skill is more a part of the consultant's art than craft. However, the computer program, lacking any such skill, must decide the matter by asking in as systematic an order as possible, a set of suitably framed questions. In the following, some circumstances where (i) and (ii) fail to hold are considered in order to illustrate some of the difficulties encountered when devising such a list of questions. Obviously

not every possible circumstance is covered, only enough to make the point. In this chapter, questions to decide on assumption (iii) are not considered. If a properly randomized experiment is used then quite often the observations may be considered as independent for the purpose of hypothesis testing, even if there is some degree of correlation between them.

If the data are not normally distributed then there is often a relationship between a treatment's mean and its variance. Therefore, the failure of (i) may be accompanied by the failure of (ii). If the data to be collected in the experiment are on what is essentially a discrete variate then the data will not be normal but their distribution may be close enough for practical purposes. If the variate is not continuous then this may be an indication that the data will not be normally distributed, and this should be one of the first things the program enquires about. One way of doing this is to use the approach taken in the previous section where a direct question is asked first, followed, if necessary, by some help. For example,

IS THE UNDERLYING NATURE OF THE VARIATE YOU ARE ABOUT
TO MEASURE IN THE EXPERIMENT

1) DISCRETE

OR

2) CONTINUOUS?

(1, 2 OR H)

In reply to H the following could be printed

IN STATISTICAL TERMINOLOGY THE MEASUREMENTS ARE MADE
ON A 'VARIATE' OF INTEREST. THE VALUES OF THE VARIATE
INDICATE THE EFFECT OF THE TREATMENT.

IN EXPERIMENTS ON DIETS, FOR EXAMPLE, THE VARIATE

MIGHT BE 'CHANGE IN WEIGHT'. THIS MAY BE RECORDED TO ANY DEGREE OF ACCURACY THE WEIGHING SCALE OR THE EXPERIMENTER PERMITS. IN A CERTAIN RANGE ALL VALUES OF THE VARIATE ARE POSSIBLE. 'CHANGE IN WEIGHT' IS THEREFORE A CONTINUOUS VARIATE. LENGTH IS ANOTHER CONTINUOUS VARIATE.

ON THE OTHER HAND, IF IN A DIFFERENT EXPERIMENT, THE VARIATE WAS 'THE NUMBER OF PETALS ON A FLOWER' THEN NOT ALL POSSIBLE VALUES IN A CERTAIN RANGE ARE POSSIBLE. THE ONLY VALUES THAT CAN OCCUR ARE 0, 1, 2, ..., ETC. VALUES SUCH AS 3.126, FOR EXAMPLE, CANNOT OCCUR. SUCH VARIATES ARE CALLED DISCRETE. THE 'AVERAGE NUMBER OF DOTS' APPEARING ON TWO DICE IS ALSO A DISRETE VARIATE BECAUSE THE ONLY VALUES THAT CAN OCCUR ARE 1, 1.5, 2, ..., 5.5, 6.

WHICH REPLY TO THE ABOVE QUESTION DO YOU WISH TO GIVE?

(1 OR 2)

However, even the above question is not totally satisfactory. For example, even though the variate is strictly continuous it may be measured in large steps and could be better considered as discrete for the purposes of the analysis. A better approach might therefore be to first ask what range of values the observations on a treatment are likely to take. Then to ask to what degree of accuracy the observations will be recorded, e.g. to the nearest 10 units of measurement, 1 unit, $\frac{1}{10}$ unit. If only a relatively small number of different values are possible the variate could be treated as discrete even if it was continuous. If a large number of different values are possible then the variate could be treated as continuous even if it was discrete. The difficulty with this approach, however, is

how best to decide what is "relatively small" and what is "relatively large". A reasonable way of approaching this is to consider the "correction" for grouped data suggested by Sheppard (1898). If data are grouped using a grouping interval of size h units, and the variance of the grouped data is calculated then $\sigma^2 g = \sigma^2 + \frac{h^2}{12}$. Here $\sigma^2 g$ is the variance of the grouped data and σ^2 is the variance of the ungrouped data. If the ratio $\frac{h^2}{12\sigma^2}$ is small then little information on σ^2 is lost by grouping the data. By analogy, h and σ^2 could be determined from the experimenter and if $\frac{h^2}{12\sigma^2}$ was small the data could be considered as continuous. Otherwise, they would be considered as discrete. However, an estimate of σ^2 may be difficult to obtain. Therefore, as done earlier the value $\frac{R}{4}$, where R is the range of the data, could be taken as an estimate of σ. Then if $h^2/[12(R/4)^2]$ was small the data could be considered as continuous. The difficulty left to resolve is how small $\frac{h^2}{12\sigma^2}$ should be in order that the data can be considered as continuous. Snedecor and Cochran (1967, p.81) considered the number of classes into which data should be grouped. They give an example where 12 classes are used and the range of the data is taken as 6σ. That is, h is $\sigma/2$. The relative increase in σ^2 due to grouping is then $\frac{1}{48}$ or 2%. If the range is taken as 4σ then h will be $\sigma/3$ and the increase due to grouping is $\frac{1}{108}$ or approximately 1%. If a 10% increase in variance is tolerated then, if the range is taken as 4σ, the number of classes should be at least 4.

By considering the different sorts of data the program

is likely to encounter in any particular experimental set-up, the value that $\frac{h^2}{12\sigma^2}$ must take before the data are considered as continuous could, perhaps, be decided upon.

Suppose it can be determined if the data are to be treated as discrete or continuous. The program must then decide if the method of analysis used in the previous section, i.e. Student's t-test can be applied to the data from the experiment. To simplify the following the two types of data are considered separately, discrete data being considered first.

1.6.1 Discrete data

Before going on to consider discrete data per se, it is important to be aware that an essentially discrete variate is sometimes recorded as a proportion or a percentage. The client may have been misled by the explanation of discrete and continuous variates and indicated that his proportion or percentage is a continuous variate. Clark and Leonard (1939) discussed the analysis of data expressed as percentages and distinguished three types:

(a) Continuous data may be expressed as percentages when each variate is divided by an arbitrary constant value, whereby each variate becomes a percentage of some standard value.

(b) Continuous data expressed in percentages to show concentrations, because a comparison of concentrations is the principle objective.

(c) Discrete data expressed as percentages, for example where the variate is the percentage of items out of a pre-determined number that possess a characteristic of interest after treatment.

An example of type (c) is an insecticide trial where the number of diseased plants on each plot after treatment is

counted and expressed as a percentage of the total number of plants on the plot.

Clark and Leonard suggest that percentages of type (a) should be treated as if the data were in their untransformed state; percentages of type (b) should not as a rule be subjected to transformation; percentages of type (c) should be transformed before analysis.

Attention to percentages of type (c) will be drawn a little later on in this section. For the moment however, attention is centered on simpler forms of discrete data. Of great importance in the following is the fact that unless otherwise stated an unpaired experiment is considered. The paired experiment will be considered in due course.

The simplest discrete variate is the one that takes only two values, e.g. 0 and 1, Yes and No, Success and Failure, etc. If this is the case the properties of the Binomial distribution may be used. The computer program could first ask

DOES THIS DISCRETE VARIATE TAKE ONLY TWO VALUES?
E.G. 0 AND 1, YES AND NO, SUCCESS AND FAILURE, ETC.
(Y OR N)

If the variate is of the Binomial type then the client is probably interested in comparing the proportion or percentage of experimental units that possess some attribute after treatment. The program would therefore ask

IT IS USUAL WHEN THE VARIATE IS AS YOU DESCRIBE TO
COMPARE THE PROPORTIONS OR PERCENTAGES POSSESSING
ONE OF THE TWO VALUES AFTER APPLYING EACH TREATMENT.
DO YOU INTEND COMPARING PROPORTIONS OR
PERCENTAGES?
(Y OR N)

Suppose that proportions are to be compared and this has been determined by the program. That is, 2n experimental units are used and n of them receive one treatment and n the other. Further, all observations taken in the experiment can be assumed independent. If the true probability of a 'Success' with one treatment is p_1 for treatment 1 and p_2 for treatment 2 then

Prob(r successes out of n are observed for treatment 1)

$$= {}^{n}C_{r} p_1^{r} (1 - p_1)^{n-r}$$

Prob(r successes out of n are observed for treatment 2)

$$= {}^{n}C_{r} p_2^{r} (1 - p_2)^{n-r}.$$

If $\hat{p}_i$ is the observed proportion of successes for treatment i, i = 1 or 2, then

$$E(\hat{p}_i) = p_i,$$

and

$$V(\hat{p}_i) = \frac{p_i(1 - p_i)}{n}.$$

The null hypothesis is now

$$H_o : p_1 = p_2$$

and the alternative hypothesis is one of

$$H_1 : p_1 \neq p_2$$

$$H_2 : p_1 > p_2$$

$$H_3 : p_2 > p_1.$$

It can be assumed that as n increases the distribution of $\hat{p}_i$ approaches the normal distribution with mean p_i and variance $\frac{p_i(1 - p_i)}{n}$. To test the null hypothesis

$$z = \frac{\hat{p}_1 - \hat{p}_2}{(\hat{p}_1(1 - \hat{p}_1)2/n)^{\frac{1}{2}}}$$

would be calculated, where

$$V(\hat{p}_1 - \hat{p}_2) = V(\hat{p}_1) + V(\hat{p}_2) = \frac{p_1(1 - p_1)}{n} + \frac{p_2(1 - p_2)}{n}$$

$$= \frac{2p_1(1 - p_1)}{n} \text{ on the null hypothesis.}$$

On the null hypothesis the distribution of z is approximately normal.

However, as in the case where two treatment means were compared, the client may require that the experiment be capable of detecting a difference as small as δ between the two proportions.

If $\delta = p_2 - p_1$ and $\sigma^2 = \{p_1(1 - p_1) + p_2(1 - p_2)\}$, and a value α for the significance level and a value P for the power are specified, as done in the previous section, the number of observations that must be made on each treatment is n where

$$n \geq \frac{\sigma^2}{\delta^2}(z_1 + z_2)^2. \qquad (1.6.1)$$

Here the variance of the difference between one observation on treatment 1 and one observation on treatment 2 is

$$V(x_1 - x_2) = V(x_1) + V(x_2) = p_1(1 - p_1) + p_2(1 - p_2).$$

The value z_1 is the $100\alpha\%$ point of the normal distribution and z_2 is the $2(1 - P)\%$ point of the normal distribution. Here the P value is considered as a two-sided probability.

To use the formula estimates of p_1 and p_2 are needed. At this stage it is easier to talk about percentages rather than proportions and so for the purpose of illustration it will be assumed in the following that percentages are being compared. An example of the situation that can be dealt with by the program using the above formula is the following. Suppose one treatment is a standard and known to give a 50% 'success

rate'. A treatment is available and seems likely to be superior. In comparing the new treatment with the standard the experimenter wants a probability of 0.9 of finding a significant difference in a one-tailed test at the 5% level if the new treatment gives a success rate of 80% or more. The appropriate value of $(z_1 + z_2)^2$ is 8.6 and so the number of observations to be taken on each treatment is n, where

$$n \geq \{(50)(50) + (80)(20)\}/(30)^2 \times 8.6$$
$$= 39.2.$$

That is 40 experimental units should be used for each treatment.

The sort of questions needed to determine the significance level, the power, and whether a one-sided or two-sided test is to be used have already been discussed in the previous section. The new questions are those needed to determine an estimate of p_1 (or p_2) and $\delta = (p_1 - p_2)$.

This could be done as follows.

TO DESIGN THE EXPERIMENT IT IS ESSENTIAL TO SPECIFY THE LIKELY VALUE OF THE PERCENTAGE (OR PROPORTION) OF SUCCESSES TO BE OBTAINED USING ONE OF THE TREATMENTS.

CAN SUCH A VALUE BE SPECIFIED?

(Y OR N)

If the answer is N the program can either stop or advise the client to take a number of 'pilot' observations so that estimates of p_1 and or p_2 can be determined.

If the answer is Y the estimate is input. Following this the program would ask:

WHAT SORT OF DIFFERENCE BETWEEN THE PERCENTAGES IS IT IMPORTANT FOR YOU TO DETECT IN THE EXPERIMENT?

This is then a similar question to that considered in the previous section when discussing treatment means.

Paulson and Wallis (1947) pointed out that the problem

with using the normal approximation discussed above is that (1) the proportions are not normally distributed when the true proportion is small and (2) the standard deviation depends on the true proportion. They also suggested that these problems can be overcome by using

$$\theta = 2\text{arc sin}\sqrt{p}$$

instead of the proportion p, as long as n is not too small. Then θ may be regarded as a normally distributed random variable with standard deviation $1/\sqrt{n}$ when θ is measured in radians and $57.3/\sqrt{n}$ when θ is measured in degrees. The sample size n is then given by

$$n \geq \tfrac{1}{2}\left(\frac{k_\alpha + k_\beta}{\text{arc sin}\sqrt{p_1} - \text{arc sin}\sqrt{p_2}}\right)^2$$

if angles are measured in radians, or by

$$n \geq 1641.6\left(\frac{k_\alpha + k_\beta}{\text{arc sin}\sqrt{p_1} - \text{arc sin}\sqrt{p_2}}\right)^2$$

if angles are measured in degrees.

Here, k_ϵ is defined by

$$\frac{1}{2\pi}\int_{k_\epsilon}^{\infty} e^{-\frac{x^2}{2}}\,dx = \epsilon,$$

α is the significance level and $1 - \beta$ is the power of the test. The value β is the probability of rejecting the alternative hypothesis when the true probabilities are p_1 for treatment 1 and p_2 for treatment 2.

Similar expressions for percentages could be given.

Suppose now that the data are of Clark and Leonard's type (c), i.e. discrete data expressed as a percentage or proportion. If each experimental unit is made up of n items and the measurements are a_{ij}, where a_{ij} is the number of successes out of n obtained on the j^{th} replicate of treatment

i, then the variate that is analysed is usually $p_{ij} = a_{ij}/n$. This variate has variance $p_{ij}(1 - p_{ij})/n$, i.e. depends on n and the treatment mean.

To eliminate this dependency the transformed variate $\theta_{ij} = \text{arc sin } (\sqrt{p_{ij}})$ should be analysed.

Bliss (1938) seems to have been the first to publish this transformation, following a suggestion by R.A. Fisher. With n less than 50, Bartlett (1947) suggested that a zero proportion should be counted as $\frac{1}{4n}$ before transforming to angles, and a proportion of one to $\frac{n - \frac{1}{4}}{n}$. Anscombe (1948) suggested using the transformation

$$\theta_{ij} = \text{arc sin } \left(\frac{a_{ij} + \frac{3}{8}}{n + \frac{3}{4}}\right)^{\frac{1}{2}}$$

for improving the equality of the variance in the angles.

However Cochran (1938) stated that "For percentages between 0 and 20 the variance increases roughly as the mean so that a square root transformation can be used. For percentages between 30 and 70, on the other hand, the variance is more or less independent of the mean, so that the analysis can be performed on the original data without transformation. For percentages [p] between 80 and 100, the square root of (100 - p) may be used."

Therefore, if the data are continuous the program would check with the experimenter if he really means they are percentages or proportions. If they are the program would check if they are of Clark and Leonard's type (c) or not.

Apart from explaining the above sorts of percentages, proportions or counts to the client the program could tackle it another way.

The client would be asked if his data are such that

they cannot take values below a fixed number or above a fixed number. Most of the data the program is likely to be concerned with will be bounded below by zero, e.g. yields or counts. If, however, the data are bounded above in the same way as percentages and proportions, for example, then the data will not be normally distributed. A difficulty with this approach however, is explaining the nature of the upper bound to the client. An upper bound to most types of data could be stated. For example, no crop of wheat has yielded more than 100 tons per acre. However, such crop yields are often analysed as if they were normally distributed. The type of upper bound that causes difficulty is the sharp cut-off as in the case of a percentage or proportion.

Consider now the situation where a paired design is used. If the experimental units are paired then the analysis of the data is not as described above. Now the data are expressed as a two-way table as follows,

		Y_2		
		1	0	Total
Y_1	1	Z_{11}	Z_{10}	Z_1
	0	Z_{01}	Z_{00}	$N - Z_1$
Total		Z_2	$N - Z_2$	N

where 1 represents success and 0 failure, N is the total number of pairs in the experiment and Z_{ij} is the number of pairs with response i for treatment 1 and j for treatment 2.

Miettinen (1968) considered the power and sample size of such experiments in terms of asymptotic theory. In his approach it is assumed the pairing is based on some matching

variate M which may be multidimensional. The k^{th} pair show some realization m_k for M. The probability of a success is then $P_i = p_i(M)$ for treatment i. The values observed in the k^{th} pair are y_{1k} and y_{2k} and these are taken as observations on the response vector (Y_1, Y_2), where $Y_i = 0$ or 1, i = 1, 2. The object of the statistical inference is

$$\delta = \theta_1 - \theta_2,$$

where $$\theta_i = E(Y_i) = E_M(Y_i/M) = E_M[p_i(M)] = E(P_i), \quad i = 1, 2.$$

The null hypothesis is $H_o : \delta = 0$. A large sample test can be based on

$$T = \frac{(Z_{10} - Z_{01})}{(Z_{10} + Z_{01})^{\frac{1}{2}}} \quad \text{and referring}$$

its value to a table of the standard normal distribution. This statistic is the square root of the customary McNemar (1947) statistic, without the correction for non-continuity.

The way this test is set up implies the effective sample size is $s = z_{01} + z_{10}$, the observed number of pairs with discordant responses. It is a realization of the random variable $S = Z_{01} + Z_{10}$. The asymptotic distribution of (T/s) is normal with expectation

$$\frac{s^{\frac{1}{2}}(\lambda_{10} - \lambda_{01})}{(\lambda_{10} + \lambda_{01})}$$

and variance

$$\frac{4\lambda_{01}\lambda_{10}}{(\lambda_{10} + \lambda_{01})^2},$$

where the λ's are the parameters of the quadrinomial distribution of (Y_1, Y_2). The quantity $\lambda_{10} - \lambda_{01}$ can be shown to equal δ. Furthermore, letting $\Psi = \lambda_{01} + \lambda_{10}$, it can be shown that the asymptotic power function for a one-sided test is

$$\Pi(\delta/s) = \Phi\left[\frac{-u_\alpha \Psi + s^{\frac{1}{2}}|\delta|}{(\Psi^2 - \delta^2)^{\frac{1}{2}}}\right]$$

where Φ denotes the distribution function of a standard normal variate and u_α is the $100(1 - \alpha)$ percentile of the standard normal distribution. A similar expression can be given for the two-sided test.

However, the power unconditional on s is required in order that the sample size can be determined before the experiment and Miettinen (1968) gives approximations to this unconditional power function. Given α, δ and an estimate of Ψ, the sample size required to achieve a power of $1 - \beta$ in the case of the one-sided test is

$$N \doteq [u_\alpha \Psi + u_\beta (\Psi^2 - \delta^2)^{\frac{1}{2}}]^2 / \Psi \delta^2.$$

Miettinen adds that, as far as the two-sided test is concerned, in most practical situations α and β are small enough to make the size of any second terms in his approximation negligible, and the two-sided case can be treated using the above formula with u_α replaced by $u_{\frac{\alpha}{2}}$.

A difficulty with using Miettinen's results is that a value of Ψ needs to be specified. However, Miettinen shows that

$$\Psi = \theta_1 + \theta_2 - 2\theta_1\theta_2 - 2\mathrm{Cov}(P_1, P_2).$$

In practice, it is usual to assume $\mathrm{Cov}(P_1, P_2) \geq 0$ i.e.

$$\Psi \leq \theta_1 + \theta_2 - 2\theta_1\theta_2.$$

Further, it can be shown that

$$|\delta| \leq \Psi \leq \theta_1 + \theta_2 - 2\theta_1\theta_2.$$

Therefore, to use this approach some estimates of θ_1 and θ_2 are needed.

Tamhane (1980) took a different approach and considered

the situation where the client wants an experiment which guarantees that the probability of choosing the treatment which has the higher probability of success is P*, whenever $P_{[2]} - P_{[1]} = \delta \geq \delta^*$, and $\pi_{10} + \pi_{01} = \pi \leq \pi^*$. Here $P_{[2]} \geq P_{[1]}$ denote the ordered values of P_1 and P_2, the probabilities of a success, π_{ij} is the true probability that a paired observation on treatment 1 and treatment 2, respectively, has outcome (i, j), and π^* is an upper limit on π; $0 < \pi^* \leq 1$, $0 < \delta^* \leq \pi^*$ and $\frac{1}{2} < P^* < 1$. The client must specify values of π^*, δ^* and P* in order that the size of the experiment can be calculated.

The corresponding treatment of this problem where independent samples are compared (i.e. pairing not used) was given by Sobel and Huyett (1957).

Therefore depending on the purpose of the experiment, suitable questions will have to be devised to obtain the information required to use the above formulae.

Having given some insight into the problems that present themselves when binary data are to be collected in the experiment, the situation when another form of discrete data is to be collected is now considered.

If the variate to be measured is discrete but not binary it may be that it is a count, i.e. a whole number which can take any of the integers 0, 1, 2, ..., etc. If this is the case the Poisson distribution may be applicable. However, it should be noted that although the following remarks are made in terms of the Poisson distribution there may be instances in practice where a treatment's mean is proportional to or approximately proportional to its variance and the data are not Poisson distributed.

Cochran (1938) suggested that if the whole numbers are

likely to be 100 or less then the square root of the numbers should be analysed. If, however, the numbers are less than 10 a half should be added to the numbers before taking square roots. If the square root transformation is used then the variance of the transformed data is a $\frac{1}{4}$. The transformation assumes the data are from a Poisson distribution. In this case the variance equals the mean. Cochran has suggested that in practice variates with large means rarely follow a Poisson distribution and consequently a transformation is not usually needed. However, if the whole numbers are likely to vary over a wide range e.g. 0 to 72,000 then a logarithmic transformation is appropriate.

The transformations suggested above can be derived if it is known what function relates the variance of the data to their mean. Bartlett (1947) showed that if the variance, σ_x^2, of the data is related to the mean, μ, by a function

$$\sigma_x^2 = f(\mu),$$

then for any function of the data, $g(x)$, it is approximately true that

$$\sigma_{g(x)}^2 = \left(\frac{\partial g}{\partial x}\right)^2_{x=\mu} f(\mu).$$

If $\sigma_{g(x)}^2$ is to be a constant c^2, say, then

$$g(\mu) = \int \frac{c\,\partial\mu}{\sqrt{f(\mu)}}.$$

For example, in the Poisson distribution,

$$\sigma_x^2 = \mu$$

and so

$$g(\mu) = c\int \frac{\partial\mu}{\sqrt{\mu}} = \frac{c}{2}\sqrt{\mu}.$$

That is, a square root transformation is needed.

If $\sigma_x^2 = \mu^2$, then a logarithmic transformation is needed, and if $\sigma_x^2 = \mu^4$, then a reciprocal transformation is needed.

Hoyle (1973) reviewed the reasons for using transformations and has described those that have been found useful.

Therefore, if the program can detect that a particular transformation is likely to be needed the client could be warned about it. The difficulty that occurs when a transformation is required is that to calculate the size of the experiment using the formulae given in the previous section the variance of the transformed variate and the size of difference to be detected on the transformed scale are required. These are likely to be difficult to obtain unless the client has used a similar transformation in previous analyses. Of course, in cases like the Poisson where the variance of the transformed variate is known for theoretical reasons some of the difficulties can be overcome. Cochran (1938), for example, has noted the variance of the transformed data for some of the more common transformations. This problem will be considered again when continuous data are discussed in the next sub-section.

One way out of these difficulties is to suggest that the client perform some pilot experiments to discover the variance of the transformed variate and to estimate the means of the treatments on the transformed scale. The pilot experiment could be planned to provide, as D.R. Cox (1958, p.167) suggests, at least six degrees of freedom for error.

1.6.2 Continuous data

When the client states his data are on a continuous variate the assumptions (i) and (ii) may be more difficult to check. For example, there were some clear instances when the variance of a discrete variate was related to its mean, and

direct questions on the relationship were not needed. When the variate is continuous it will again be necessary to determine if the mean and variance are related and suitable questions will have to be devised. One approach is to ask the client if he has analysed similar data before and then to question him on these past data. He may be able to state that he suspects the variance is related to the mean and might be able to specify the approximate or exact form of the relationship. However, with only two treatments it may be difficult for the client to detect the exact relationship, if any, between the mean and variance. The sort of question on his past data that he might be able to answer is the following

IS THERE ANY REASON TO SUSPECT THAT THE VARIABILITY OF THE DATA ON ONE TREATMENT WAS A LOT LARGER (OR SMALLER) THAN THE VARIABILITY OF THE DATA ON THE OTHER TREATMENT?

(Y, N OR H)

In reply to H the program could print

FOR EXAMPLE DID THE SIZE OF VARIABILITY DEPEND ON THE SIZE OF THE MEAN OF THE TREATMENT? SOMETIMES AS THE SIZE OF THE MEAN INCREASES THE VARIABILITY INCREASES.

WHICH REPLY TO THE ABOVE QUESTION DO YOU WISH TO GIVE?

(Y OR N)

If the variance did depend on the mean the client might then be asked about the particular form of the relationship. For example, he could be asked if the range of the data on a treatment varied in proportion to the mean. If it did then as Bartlett's (1947) procedure indicates a logarithmic transformation is appropriate. If the range was proportional to the square of the mean then a reciprocal transformation would be appropriate. However, even if the nature of the transformation could be

determined it is still necessary to check that such a transformation could be applied. For example, although the client's replies indicate a square root transformation is necessary it may be that some of the data can take negative values. Or a logarithmic transformation seems called for but zero data values are possible. Therefore, at some point in the program the client should be asked if it is possible for his data to take negative or zero values.

The distributions which are such that the mean is related to the variance are in general skewed distributions. That is, the frequency distribution of the observations is not symmetrical about the mean, but has a long tail either to the right or left. As well as questioning the client about the probable size of the variance of each treatment it is also wise to question him on the likely shape of the frequency distribution of the measurements on each treatment. Again if past data are available he could be questioned on these. If they are not available the question could be put to the client, as for example, in the following.

CONSIDER THE MEASUREMENTS ON ONLY ONE OF THE TREATMENTS FOR A MOMENT. AS YOU KNOW THEY WILL EXHIBIT SOME DEGREE OF VARIABILITY ABOUT THEIR MEAN VALUE. IS THERE ANY REASON TO EXPECT THAT THE MEASUREMENTS WILL NOT BE DISTRIBUTED SYMMETRICALLY ABOUT THEIR MEAN?

(Y, N OR H)

In reply to H the program could print

FOR EXAMPLE, THE MEASUREMENTS COULD BE SUCH THAT THERE WERE A LOT OF SMALL VALUES BELOW THE MEAN BALANCED BY A FEW LARGE VALUES ABOVE THE MEAN, OR VICE VERSA.

WHICH REPLY DO YOU WISH TO GIVE TO THE ABOVE QUESTION?

(Y OR N)

If the reply indicated the observations were likely to be skewed and it had previously been established that the variances were unequal then a transformation of the data, as described above will most likely be required. If, however, the client replies that the distribution is skewed but variances are equal or vice versa then there is some ambiguity as to what to suggest. In this situation it is probably best to suggest that the client consult a statistician.

The above consideration of what to do if the treatment variances are unequal is necessary only if a combined estimate of the variance of a treatment is to be used. This would be the case if two independent sets of observations were to be compared. If, however, a paired comparison design is to be used then the variance of the differences between pairs will be used in the test of hypothesis. If pairing is used the above transformations may not be required. However, it is necessary to check that the values of the paired differences satisfy the assumptions.

If the client knows enough about the treatments to be able to say the variances are likely to be different he may have some idea of the size of these variances. If he can give these values they can be used to give an upper bound on the variance of a paired difference. For example, suppose the variance of treatment 1 is σ_1^2 and treatment 2 is σ_2^2 then the variance of a paired difference, $(x_1 - x_2)$, say, is

$$V(x_1 - x_2) = \sigma_1^2 + \sigma_2^2 - 2\mathrm{Cov}(x_1, x_2).$$

The difficulty remaining is to determine $\mathrm{Cov}(x_1, x_2)$. If the client can also give a value for the correlation between x_1 and x_2 then

$$\text{Correlation } (x_1, x_2) = \frac{\mathrm{Cov}(x_1, x_2)}{\sigma_1 \sigma_2},$$

i.e.

$$\text{Cov}(x_1, x_2) = \sigma_1\sigma_2 \text{ Correlation } (x_1x_2).$$

If such information is likely to be difficult to obtain then the program could take $V(x_1 - x_2) = \sigma_1^2 + \sigma_2^2$ by default or suggest a pilot experiment.

Consider, once again, the problem of calculating the size of an unpaired experiment assuming a transformation has been suggested. To use formula 1.4.3 given in section 1.4 the variance of the transformed data and a value for the size of difference to be detected on the transformed scale are required. As an example of how this problem may be dealt with consider the logarithmic transformation. Here it is assumed that after taking logarithms to base e the data on each treatment are normally distributed with equal variance σ^2. Assume further that on the transformed scale the mean of treatment 1 is μ_1 and the mean of treatment 2 is μ_2. The untransformed variate will be denoted by x and the transformed variate by y. That is $y = \log_e(x_i)$ is normally distributed as $N(\mu_i, \sigma^2)$, if x_i is an observation on treatment i, i = 1, 2. The distribution of the x variate is log normal. This distribution has been described in detail by Aitchison and Brown (1966), for example. Assuming this distribution

$$E(x_i) = e^{\mu_i + \frac{1}{2}\sigma^2} \qquad (1.6.1)$$

and

$$V(x_i) = e^{2\mu_i + \sigma^2}(e^{\sigma^2} - 1). \qquad (1.6.2)$$

The quantiles ξ_q of the log normal distribution are related to the quantiles v_q of the normal distribution by the relationship

$$\xi_q = e^{\mu + v_q\sigma}.$$

Therefore, if the quantiles of the normal distribution of the

y's corresponding to a 95% confidence interval are such that $-v_q = v_{1-q}$ then the corresponding quantiles of the distribution of the x_i's are $e^{\mu i - 2\sigma}$ and $e^{\mu i + 2\sigma}$.

If the range of treatment i can be specified as the difference between two values, i.e. range = $x_U - x_L$, then taking σ^2 as a quarter of the range, the estimate of σ is

$$\hat{\sigma} = (\log_e x_U - \log_e x_L)/4.$$

Further, if the smallest difference to be detected in the experiment can also be specified as a difference between two mean values, i.e. $\delta = m_2 - m_1$ then

$$m_1 \doteq e^{\mu 1 + \frac{1}{2}\sigma^2}$$

and

$$m_2 \doteq e^{\mu 2 + \frac{1}{2}\sigma^2}.$$

Also as

$$\log_e(m_1) = \mu_1 + \tfrac{1}{2}\sigma^2$$
$$\log_e(m_2) = \mu_2 + \tfrac{1}{2}\sigma^2$$

then

$$\mu_2 - \mu_1 = \log_e(m_2) - \log_e(m_1).$$

That is, the smallest difference to be detected on the transformed scale is

$$\delta' = \log_e(m_2) - \log_e(m_1).$$

In the last resort, it is not unreasonable to base the calculation of the size of the experiment on estimates of σ and δ, if available, that are given in terms of the untransformed data. The aim of the transformation is to enable a more precise experiment to be performed. Therefore, the worse that can happen is that a larger experiment than necessary will be suggested by the program.

From the points considered in this and the previous

section it is clear that the program must obtain information from the client on eight main areas.

(1) Whether or not the observations to be taken in the experiment can be considered to be normally distributed. (Implicit here is the assumption that the observations are on a continuous variate.)

(2) If the variance of the observations is the same for each treatment.

(3) An estimate of the variance of each treatment.

(4) The size of difference between the means the experiment must be capable of detecting.

(5) Whether a one-sided or two-sided hypothesis is to be tested.

(6) The significance level of the test.

(7) The power of the test.

(8) Whether pairing of the units is possible.

Often in a real consultation the human consultant is able to take for granted the answers to some of the above questions because he is familiar with the client's work. The computer however can assume nothing and must ask questions on every point before it can make an unambiguous decision. It is clear from what has been said previously that the fact the sort of questions that must be asked have been identified does not mean it is necessarily a simple matter to incorporate them into an interactive computer program. If the user of the program is a statistically informed experimenter then most probably a set of direct questions to determine (1) to (8) above are all that will be required. However, the statistical layman will not understand such direct questions. He will require additional explanation of unfamiliar words or phrases and illustrative examples to help illuminate difficult points. It is in the writing of a program for the layman that most problems

arise. A strategy for obtaining the information in an indirect manner must be devised. At the time of writing the program all the necessary questions must be thought of and be such that they cover all eventualities. The program must be written in a way that ensures that the answer from one question leads naturally on to the next. The order of asking the questions is very important. A sensible order of obtaining the information specified in (1) to (8) above is as follows.

It is first explained that the experimental units will exhibit some degree of variability and this will be reflected in the observations on each treatment. Having introduced the idea of variability the client is asked if it is possible to pair the units so that the units in a pair are more similar to one another than they are to the units as a whole. If pairing is possible this is noted. Whether or not the observations can be considered as continuous or discrete for the purposes of the analysis is decided next. One way of doing this is to ask what range of values is expected for one of the treatments and then to ask to what degree of accuracy the data will be recorded, e.g. to the nearest 10 units of measurement, 1 unit, $\frac{1}{10}$ unit. If only a relatively small number of values is possible the data would be considered as discrete and the program may have to state that it cannot advise on such an experiment. For the purposes of the analysis discrete data like counts may be considered as continuous. If the data are counts the Poisson distribution may apply and the square roots of the observations should be analysed. If the observations are from a Binomial distribution a normal approximation may be appropriate. Whether or not the data can take zero or negative values is determined next. Such information may prove useful later in the program.

Still on the subject of variability the client is asked if the degree of variability is the same for each treatment. If it is not the program takes a special route described below. If it is then it is explained to the client that it is essential to have a numerical value for the degree of variability. If similar experiments have been done in the past there is the possibility that the variance of the past data is likely to be a good estimate of the variance of the data about to be obtained. Asking for such information will of course involve explaining the meaning of a variance to the layman. If the client is unlikely to understand such an explanation he could at least be given instructions on how to calculate it and to return to the program with his calculated value. If pairing was used in the past experiment and it is to be used in the new experiment then an estimate of the difference of the paired observations may be available. If an estimate of any of these variances is not forthcoming the value of the range of each treatment, if available, can be used to provide an estimate of the standard deviation. Information on the range may have already been obtained in the earlier questions to determine if the observations are discrete or continuous. The client may be more familiar with expressing variability as a coefficient of variation. If he is then questions to obtain an estimate of it must be included. If all else fails the client can be advised to perform a pilot experiment to estimate the required variance.

If the data on each treatment are normally distributed the frequency distribution of the data should not be markedly skew. The client is asked if he thinks extremely small or extremely large data values are equally likely to occur. If they are not then this could be taken as evidence of skewness.

If the data are such that the treatment variances are likely to be equal, skewness seems unlikely and the data can be considered as continuous then the program is unlikely to be far wrong if it assumes the data are normally distributed.

If the treatment variances are unequal and extreme skewness seems likely it could be that the variance is related to the mean of each treatment. If it is and the relationship can be determined then a transformation may eliminate the dependency of the variance on the mean. For example, if the variance is proportional to the mean a square root transformation is called for. An estimate of the variance of the square rooted data would then be required. If however the client had stated in an earlier reply that the data can take negative values a square root transformation cannot be applied and the computer is unable to continue. The advice of the human consultant must be sought to clear up this apparent contradiction. Similarly if a logarithmic transformation seemed appropriate and many zero data values are possible then this transformation is inappropriate. If pairing of experimental units is possible then the heterogeneity of variance can be overcome by analysing the differences between the paired observations. An estimate of the variance of these paired differences would then be required, and this may be difficult to obtain without performing a pilot experiment.

Having obtained an estimate of the variance in some form the client is asked what size of difference between the treatment means it is essential for the experiment to detect. Again the client would be advised that the program cannot continue without such a value. If the data are to be transformed an estimate of the difference on the transformed scale is required. The way of expressing the difference must be the same as that

used to express the variance. For example, if a coefficient of variation has been given the size of difference must be expressed as a proportion or percentage of the mean. Once the size of difference to be detected is decided upon the client is asked if he wants to test a one-sided or two-sided hypothesis, with suitable explanations of these terms being provided.

Next the client is asked what significance level is required. The significance level will probably be an unfamiliar concept and so must be explained to the client. This will involve explaining the type I error. If the client is unable to decide on a significance level then default values of 5% or 1% can be taken.

Following on from this is a question on the size of power of the hypothesis test. It is highly likely that the client will not understand this concept and so a full explanation must be given. Again default values can be taken if all else fails.

Having obtained the necessary information the computer calculates the size of the experiment. If this is too large then, if the experiment is to be undertaken at all, the client will have to limit his objectives. A dialogue can then take place in which various modifications can be tried. For example, planning to detect a larger difference than was first suggested, reducing the significance level from 1% to 5%, reducing the power and considering whether the estimates of the variance are realistic. A useful outcome of the consultation could be that the client is being too ambitious and needs to limit his aims.

1.7 Computer programming difficulties

The only way the suggestions made in the previous sections can be put to a realistic test is if a computer program is written and used by experimenters. With this aim in mind a program based on some of the more easily programmable suggestions made in the previous sections was written for the ICL 4130 computer. This was the computer possessed by the University of Kent at the time the work described in this chapter was begun. At that time the only interactive computing language was BASIC, and this only existed in a limited form suitable for teaching purposes. With such limited resources the writing of a suitable program was made more difficult. The main problem was the limited resources of the computer which permitted only small BASIC programs to be held in core at one time. The interactive program that was written was too large and so two ways were used to enable it to run. The first way was to divide the program into small sections with information obtained in one section being passed onto the next. One section is held in core at a time and this is overwritten by the next section as the client moves through the program. The second way was to keep all the text that is printed by the program stored on a separate disc file. Each separate question, help statement and other pieces of text were each given a label and stored on the disc in the order they were to be printed. Stored directly beneath each piece of text was a list of labels which indicated the next piece of text to print depending on the client's reply. For example, the following piece of text might have label N on the disc file. Beneath it are three labels (P, Q, R).

N) IS THE UNDERLYING NATURE OF THE VARIATE YOU ARE
ABOUT TO MEASURE IN THE EXPERIMENT

1) DISCRETE

OR

2) CONTINUOUS?

(1, 2 OR H)

P, Q, R

If the user's reply is 1 then the text labelled as P is printed next, if the reply is 2 then the text labelled as Q is printed next, etc. The program moves to the appropriate place on the disc file and prints the text.

A brief summary of this and the previous sections was given by Jones (1980a).

1.8 Some extra difficulties when more than two treatments are to be compared

In this section some of the difficulties peculiar to the situation where more than two treatments are to be compared are noted. Such things as checking the assumptions of normality, equality of variance and independence will be essentially the same and therefore will not be considered further. The major difference comes from the possibility of structure among the treatments, i.e. there may be a choice of treatment contrasts to be tested and it may be that not all contrasts are of equal importance. For example, in an experiment to compare four treatments, one might be a standard and the other three new varieties. The aim of the experiment being to decide if any or all of the new varieties are superior to the standard. The contrasts among the new varieties would be of no interest. In a different experiment the four treatments might be the four treatment combinations of a 2 × 2 factorial experiment. The contrasts of interest will then be the main effects and interactions. It could be that the client requires more precise estimates of the main effects than of the interaction.

An experimental design that satisfies this objective will then be required. It is also possible that the client knows about the main effects and only wants information on the interaction.

Therefore, it is essential that the computer program enquires about the treatment structure, the treatment contrasts of interest, and the relative importance of these contrasts.

If treatments are to be compared in a factorial set it may be necessary to enquire about the likely size and importance of high order interactions, in case confounding and or fractional replication are needed.

The problem of getting the computer to provide experimental designs, optimal for the client's requirements must also be considered. The client's experimental material may not fit neatly into the "text-book" set of block sizes, nor will the treatments necessarily be equally replicated. At the inception of this research programme, computer algorithms for constructing optimal designs, that were of a sufficiently general nature for use in an interactive computer program such as that envisaged here, did not exist. Therefore, the rest of the chapters in this thesis are devoted to the development of such algorithms. Of course, such algorithms are useful in their own right, and not only as a useful addition to an interactive program. They are useful aids to the experienced statistician as well as the statistically informed experimenter. They represent yet another way in which computers can assist in the design of comparative experiments.

Therefore, before leaving the study of the difficulties involved in writing an interactive computer program that gives advice on the design of a comparative experiment, some general conclusions are drawn in the next section.

1.9 Some general conclusions

1) There are certain qualities possessed by the human consultant that make it difficult to transfer his skills to a computer program. Not the least of these are his experience, versatility and ability to "talk around" the proposed experiment.

2) The good consultant is able to modify his approach to satisfy the needs of clients with quite different levels of knowledge about statistics. A computer program intended to be used by a wide variety of clients is likely, however, to be large and unmanageable. The clients that can be most easily catered for in a computer program are those with no knowledge of statistics and those who are expert in statistics.

3) The types of experiment that can be most easily advised upon by a computer program are those that are so familiar and well defined that it is not difficult to decide in advance what information must be obtained by the program.

4) A major aspect of writing the program is to divide up the total amount of information required to design the experiment into smaller parts that can be structured in an ordered sequence.

5) The way the questions are phrased is very important. It is essential that they be easily understood, unambiguous, and if possible, short and to the point. It will be necessary for the client to have the option of obtaining extra text to help answer some of the more difficult questions.

6) Having broken down the total amount of information required as described in 4) and devised suitable questions as described in 5), it is important that the order of asking the questions be such that there is a natural flow through

the program. The answer from one question should lead naturally on to the next.

7) The most important questions such as those on the choice of treatments, experimental units, type of measurement to take, and choice of experimental site are likely to be too difficult to incorporate into a computer program. It will have to be assumed that these questions have been answered before the client uses the computer program.

8) Even for relatively simple experiments, quite a large number of questions and related pieces of text may have to be printed by the program. Efficient computer programming techniques will therefore be required to store and print this text.

9) The size of the program is likely to increase considerably when it is enlarged to cater for more complicated experimental situations.

10) When more than two treatments are to be compared, computer algorithms that search for designs that are optimal for the client's requirements will be required.

The remainder of this thesis, as mentioned earlier, is devoted to the development of computer algorithms that search for optimal designs. These algorithms are developed for use in their own right, and not specifically for use in an interactive computer program such as that considered in this chapter.

Chapter 2. Construction of optimal block designs

2.1 Introduction

The final aim of the interactive computer program described in the previous chapter is to present the user with a design that is best suited to his requirements. The concept of an optimal design is therefore a natural one to consider in this context, but this presents the problem of what is meant by optimality. In this chapter the different criteria in the literature are reviewed and a new approach is described. This new approach is particularly suited to the needs of an experimenter who requires a block design for studying certain contrasts among the treatments.

The review begins by looking at the theory of optimal designs for the general linear model and then concentrates on the theory for block designs.

2.2 Optimality criteria for the general linear model

The general linear model may be defined as

$$E(\underset{\sim}{Y}) = \underset{\sim\sim}{X}\underset{\sim}{\beta}, \qquad (2.2.1)$$

where $\underset{\sim}{Y}'$ is the vector of n data $(y_1, y_2, \ldots, y_n)$, assumed to be normally distributed random variables with variance σ^2, $\underset{\sim}{X}$ is a $n \times p$ design matrix of elements (x_{ij}), and $\underset{\sim}{\beta}$ is a vector of P unknown parameters $(\beta_1, \beta_2, \ldots, \beta_p)$.

The normal equations for the least squares estimate of $\underset{\sim}{\beta}$ are then

$$\underset{\sim}{X}'\underset{\sim\sim}{X}\hat{\underset{\sim}{\beta}} = \underset{\sim}{X}'\underset{\sim}{Y} \ . \qquad (2.2.2)$$

The power of the F-test for the analysis of variance of the hypothesis that all elements of $\underset{\sim}{\beta}$ are zero depends only on the non-centrality parameter μ, where

$$\underset{\sim}{\mu} = (\underset{\sim}{\beta}'\underset{\sim}{X}'\underset{\sim\sim}{X}\underset{\sim}{\beta})/2\sigma^2 \ . \qquad (2.2.3)$$

As μ increases the power of the F-test increases and therefore by an appropriate choice of the values of x_{ij} the power can be

maximized. Also, if $\underset{\sim}{X}$ has full rank, the same x_{ij} minimize the variance of $\hat{\underset{\sim}{\beta}}$ which is $\sigma^2(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$.

However, for $p > 1$ it is not possible to maximize μ for all values of $\underset{\sim}{\beta}$ and Wald (1943) suggested as a compromise to maximize the minimum value of μ subject to $\underset{\sim}{\beta}'\underset{\sim}{\beta} = 1$. This is equivalent to maximizing the minimum eigenvalue of $\underset{\sim}{X}'\underset{\sim}{X}$. However, for certain mathematical simplifications Wald further proposed that the product of the non-zero eigenvalues of $\underset{\sim}{X}'\underset{\sim}{X}$ should be maximized. If $\underset{\sim}{X}$ has full rank this is equivalent to the maximization of the determinant of $\underset{\sim}{X}'\underset{\sim}{X}$. This criterion, which is known as D (for determinant) - optimality has been investigated in detail in regression situations. The former criterion is known as E-optimality and has been investigated further by Ehrenfeld (1955). He showed that when $\underset{\sim}{X}$ has full rank the E-optimal design minimizes the variance of any linear combination of the parameters $\beta_1, \beta_2, \ldots, \beta_p$, scaled so that the sum of squared coefficients is unity.

Kiefer (1958) considered the properties of a number of optimality criteria, including D- and E-optimality, in terms of the power function of a test of hypothesis.

Moriguti (1954) considered the design that minimizes the maximum variance of the unbiased linear estimates of $\underset{\sim}{\beta}$ as optimal.

Elfving (1952) and Chernoff (1953) minimized the trace of $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$, where $\underset{\sim}{X}$ has full rank, to obtain optimal regression designs. That is, they minimized the average variance of the parameters $\beta_1, \beta_2, \ldots, \beta_p$. Box and Draper (1971) considered the D-optimality of factorial designs.

Kiefer (1959, 1961a, 1961b, 1962) and Kiefer and Wolfowitz (1959, 1960) considered a design as a mass distribution of points over a general space equivalent to a probability measure. In this

approach the linear model (2.2.1) is redefined as

$$E(y_i) = \underset{\sim}{\beta}'\underset{\sim}{f}(x_i), \tag{2.2.4}$$

where the variables y_i, $i = 1, 2, \ldots, n$, are as before, each y_i corresponding to a value x_i of the independent variable, and $\underset{\sim}{f}'(x) = [f_1(x), f_2(x), \ldots, f_p(x)]$ is a vector of functions f_i evaluated at x.

It is further assumed that the functions $\underset{\sim}{f}(x)$ are linearly independent on the space χ of x, which is continuous in a topology in which χ is compact, and usually χ will be a closed compact set in a Euclidean space of a particular dimension.

An exact or discrete design d is then a choice of n points $x_1, x_2, \ldots, x_n$ in χ.

Using the model (2.2.4), let $\xi_d(x)$ be the proportion of points x_i, i=1, 2, ..., n, which are equal to x in the design d. Then $\xi_d(x)$ may be regarded as a probability measure on χ. For any probability measure ξ on χ define

$$m_{ij} = \int_{\chi} f_i(x) f_j(x) \xi(dx), \tag{2.2.5}$$

and let $\underset{\sim}{M}(\xi)$ be the matrix of elements m_{ij}. The information matrix of the design is then

$$\underset{\sim}{A}_d = n\underset{\sim}{M}(\xi_d). \tag{2.2.6}$$

In this setting a discrete design is a probability measure ξ_d which takes only integer multiples of $1/n$.

The main difficulties associated with discrete designs are that they often exhibit a fine structural dependence on n, necessitating a lengthy table of optimum designs for a given problem, and that their construction often presents a difficult combinatorial problem admitting no simple method of solution.

Therefore, to overcome these problems the concept of a continuous or approximate design has been developed. This development is again contained in the works of Kiefer and of Kiefer and Wolfowitz, referred to earlier.

To obtain continuous designs it is assumed that there is an optimality criterion which maximizes some function Q of the information matrix A_d and that there exists a probability measure ξ^* which maximizes $Q[\underset{\sim}{M}(\xi)]$ over all probability measures ξ on χ. It will be recalled that in the discrete theory ξ only takes on integer multiples of $1/n$. The main advantages of continuous designs are that they are independent of n and are optimal to within order $1/n$. Also they can yield designs that are discrete for many n.

In this setting a design ξ^* is D-optimal if

$$\det[\underset{\sim}{M}(\xi^*)] = \sup_{\xi} \det[\underset{\sim}{M}(\xi)], \tag{2.2.7}$$

where $\det[\underset{\sim}{M}]$ is the determinant of $\underset{\sim}{M}$.

Kiefer and Wolfowitz investigated the equivalence of several types of optimality. Defining

$$d(x,\xi) = \underset{\sim}{f}'(x)\underset{\sim}{M}^{-1}(\xi)\underset{\sim}{f}(x), \tag{2.2.8}$$

i.e. $d(x,\xi)$ is the variance function obtained using the model (2.2.4), Kiefer and Wolfowitz (1960), showed that the D-optimality of ξ^* as defined in (2.2.7) is equivalent to

$$\inf_{\xi} \sup_{x} d(x,\xi) = \sup_{x} d(x,\xi^*) \tag{2.2.9}$$

and also to

$$\sup_{x} d(x,\xi^*) = p \tag{2.2.10}$$

where p is the dimension of $\underset{\sim}{f}(x)$.

The equivalence of (2.2.7), (2.2.9) and (2.2.10) is referred to as the General Equivalence Theorem. The criterion (2.2.9) is known as G-optimality, and if χ is a compact space on which the functions f_i are linearly independent, then for continuous ξ^* a design is D-optimal for $\underset{\sim}{\beta}$ if and only if it is G-optimal.

Kiefer (1961a) gave a generalization of the General Equivalence Theorem for deriving D-optimal designs for a subset s(<p) of the parameters $(\beta_1, \beta_2, \ldots, \beta_p)$.

White (1973) considered the D-optimality of non-linear models and gave a result analogous to the General Equivalence Theorem.

Wynn (1970) considered the sequential generation of of D-optimal designs and later (Wynn,1972) showed that designs that are D-optimal for a given subset s of the parameters $(\beta_1, \beta_2, \ldots, \beta_p)$ can be generated sequentially. Dykstra (1971) also considered the sequential generation of D-optimal designs.

Kiefer (1974) gave further equivalence results between D-optimality and other criteria.

A geometric approach to D-optimality was given by Silvey and Titterington (1973). In this approach a unified treatment of the General Equivalence Theorem is given in the form of duality theorems established by means of Strong Lagrangian Theory.

Consideration of the optimal properties of regression experiments can be traced back to Smith (1918), who determined the G-optimal designs for polynomials of order $p \leq 6$. Guest (1958) characterized the G-optimum design for polynomials of arbitrary order p in terms of the zeros of

the derivative of the Legendre polynomial. Hoel (1958) computed the D-optimal design for the polynomial and noted that the D-optimal design coincided with Guest's G-optimal design.

The book by Fedorov (1972) contains much of the work done by the author on continuous and discrete designs.

A review of the theory of D-optimality when applied to regression situations was given by St. John and Draper (1975a), and St. John and Draper (1975b) reviewed the application of D-optimality to regression experiments with mixtures.

More recently, Herzberg and Andrews (1975) have considered the effect of missing observations on D-optimal and G-optimal designs for simple regression situations, and Tsay (1976) has given a general procedure for the construction of D-optimal designs, which includes Wynn's procedure as a special case.

Published algorithms are mainly for the construction of continuous D-optimal designs for regression experiments. These designs are of interest primarily because the D-optimal continuous design provides a reference against which the discrete designs can be evaluated and often the continuous design corresponds to a discrete design.

Kiefer and Wolfowitz (1959) suggested a game theoretic approach to constructing designs and Kiefer (1961a, 1961b) gave examples of constructing D-optimal designs, but his methods did not lead to a general algorithm. Fedorov (1969a, 1969b, 1972) and Fedorov and Dubova (1968) seem to have been the first to develop a general algorithm for obtaining D-optimal designs.

Wynn (1970) proved the convergence of an algorithm

similar to that of Fedorov (1972) and Dykstra (1971). Hebble and Mitchell (1972) and Covey-Crump and Silvey (1970) suggested algorithms similar to that of Wynn but did not give convergence proofs.

Atwood (1973) offered some improvements to Fedorov's algorithm and St. John (1973) modified one of Atwood's suggestions. Silvey and Titterington (1973) outlined an algorithm for obtaining a D-optimal design measure on the space spanned by the response function specified by (2.2.4).

Wynn extended his algorithm for constructing D-optimal designs for a subset s<p parameters. Further refinements of this algorithm were given by Wynn (1973).

Despite the algorithms given above, D-optimality is not always an easy criterion to implement and Nalimov et al. (1970) and Box and Draper (1971) pointed out that creating a discrete design from the probability measure may require an extremely large number of points.

One approach to this difficulty is to consider the size n, of the experiment, as fixed beforehand, and to maximize the determinant of $n\underset{\sim}{M}[\xi(n)]$, where this matrix is expressed as $\underset{\sim}{X}'\underset{\sim}{X}$, and $\xi(n)$ denotes the discrete design with n points. The design that maximizes the determinant of $\underset{\sim}{X}'\underset{\sim}{X}$ ($\underset{\sim}{X}'\underset{\sim}{X}$ non-singular) is referred to as D_n-optimal.

Several authors obtained D_n-optimal designs using mathematical programming methods. Designs using this approach were given by Box and Draper (1971), Atkinson (1969), Bloom, Pfaffenberger and Kochenberger (1972), Box (1966), Hartley and Rudd (1969), Mahoney (1970) and Neuhardt and Bradley (1971).

As no General Equivalence Theorem exists for discrete designs, the algorithms for obtaining D_n-optimal designs are

somewhat different from those for obtaining D-optimal designs. The first algorithm for constructing D_n-optimal designs was given by Fedorov (1969b, 1972) and Van Schalkwyk (1971). The second was given by Wynn (1972) and Mitchell and Miller (1970). Mitchell (1974) proposed a variation on this latter algorithm.

The main disadvantage of these algorithms for obtaining D_n-optimal designs is that there is no guarantee that they will converge to the true optimum. Often they converge to local optima and to overcome this Mitchell (1973) suggested using a number of different starting points, sometimes as many as fifty.

2.3 Optimality of block designs

In order to consider the optimality of block designs it is necessary to introduce some notation. It will be assumed that the data from a block experiment can be expressed as

$$\underset{\sim}{y} = \underset{\sim}{1}\alpha + \underset{\sim}{D}'\underset{\sim}{\beta} + \underset{\sim}{\Delta}'\underset{\sim}{\gamma} + \underset{\sim}{\eta}, \qquad (2.3.1)$$

where $\underset{\sim}{y}$ is a column vector of n data; $\underset{\sim}{1}$ is a column vector with unit elements; $\underset{\sim}{D}'$ is an n×b design matrix for blocks, b being the number of blocks, and $\underset{\sim}{\Delta}'$ is an n×v design matrix for treatments, v being the number of treatments; also α is a general parameter, $\underset{\sim}{\beta}$ a vector of block parameters, $\underset{\sim}{\gamma}$ a vector of treatment parameters and $\underset{\sim}{\eta}$ a vector of n uncorrelated residuals distributed about zero with variance σ^2.

If G, $\underset{\sim}{B}$, and $\underset{\sim}{T}$ represent the grand total of the data, the vector of block totals and the vector of treatment totals respectively, the least squares normal equations for estimating the parameters are

$$G = \underset{\sim}{1}'\underset{\sim}{y} = n\hat{\alpha} + \underset{\sim}{r}'\hat{\underset{\sim}{\gamma}} \qquad (2.3.2)$$

$$\underset{\sim}{B} = \underset{\sim\sim}{D}\underset{\sim}{y} = \underset{\sim}{k}\hat{\alpha} + \underset{\sim}{k}^{\delta}\hat{\underset{\sim}{\beta}} + \underset{\sim}{N}'\hat{\underset{\sim}{\gamma}} \tag{2.3.3}$$

$$\underset{\sim}{T} = \underset{\sim\sim}{\Delta}\underset{\sim}{y} = \underset{\sim}{r}\hat{\alpha} + \underset{\sim\sim}{N}\hat{\underset{\sim}{\beta}} + \underset{\sim}{r}^{\delta}\hat{\underset{\sim}{\gamma}} \tag{2.3.4}$$

where $\hat{\alpha}$, $\hat{\underset{\sim}{\beta}}$ and $\hat{\underset{\sim}{\gamma}}$ are the least squares estimates of α, $\underset{\sim}{\beta}$ and $\underset{\sim}{\gamma}$ respectively, and $\underset{\sim}{N} = \underset{\sim\sim}{\Delta}\underset{\sim}{D}'$ is the incidence matrix.

By writing

$$\underset{\sim}{Q} = \underset{\sim}{T} - \underset{\sim\sim}{N}\underset{\sim}{k}^{-\delta}\underset{\sim}{B} \tag{2.3.5}$$

we have

$$\underset{\sim}{Q} = (\underset{\sim}{r}^{\delta} - \underset{\sim\sim}{N}\underset{\sim}{k}^{-\delta}\underset{\sim}{N}')\hat{\underset{\sim}{\gamma}}, \tag{2.3.6}$$

that is

$$\underset{\sim}{Q} = \underset{\sim\sim}{C}\hat{\underset{\sim}{\gamma}} \tag{2.3.7}$$

The matrix $\underset{\sim}{C}$ is singular however, and so either a generalized inverse $\underset{\sim}{C}^{-}$ is used to give

$$\hat{\underset{\sim}{\gamma}} = \underset{\sim}{C}^{-}\underset{\sim}{Q},$$

or constraints on the treatment parameters are imposed.

On the assumption that $\underset{\sim}{C}$ had only one zero eigenvalue, Tocher (1952) applied the constraint $\underset{\sim}{r}'\hat{\underset{\sim}{\gamma}} = 0$ and wrote equation (2.3.6) as

$$\underset{\sim}{Q} = (\underset{\sim}{r}^{\delta} - \underset{\sim\sim}{N}\underset{\sim}{k}^{-\delta}\underset{\sim}{N}' + \underset{\sim\sim}{r}\underset{\sim}{r}'/n)\hat{\underset{\sim}{\gamma}}.$$

He wrote $$\underset{\sim}{\Omega}^{-1} = \underset{\sim}{r}^{\delta} - \underset{\sim\sim}{N}\underset{\sim}{k}^{-\delta}\underset{\sim}{N}' + \underset{\sim\sim}{r}\underset{\sim}{r}'/n \tag{2.3.8}$$

and $$\hat{\underset{\sim}{\gamma}} = \underset{\sim\sim}{\Omega}\underset{\sim}{Q}.$$

The matrix $\underset{\sim}{\Omega}$ is one possible generalized inverse. Another can be obtained from the eigensystem of $\underset{\sim}{C}$. Writing the non-zero eigenvalues of $\underset{\sim}{C}$ as $\lambda_1, \lambda_2, \ldots, \lambda_{v-1}$ and the corresponding eigenvectors as $\underset{\sim}{u}_1, \underset{\sim}{u}_2, \ldots, \underset{\sim}{u}_{v-1}$ a second generalized inverse is

$$\underset{\sim}{C}^{+} = \sum_{i=1}^{v-1} \frac{1}{\lambda_i}\underset{\sim}{u}_i\underset{\sim}{u}_i' \quad .$$

The coefficients matrix $\underset{\sim}{C}$ has played an important part in the development of criteria for block designs, particularly incomplete block designs. That is where all $k_j = k<v$, $j = 1, 2, \ldots, b$.

Kempthorne (1956), showed that when all $r_i = r$, $i = 1, 2, \ldots, v$, the efficiency factor of a design is r times the harmonic mean of the non-zero eigenvalues of $\underset{\sim}{C}$. Here the efficiency factor is the ratio of (a) the mean variance of simple treatment differences in a complete block design to (b) the mean variance of the intrablock estimates of simple treatment differences in the incomplete block design. The optimum design suggested by this approach would be the one that maximizes the harmonic mean just defined.

Masuyama (1957, 1958) suggested minimizing trace $(\underset{\sim}{C}^2)$, subject to trace $(\underset{\sim}{C})$ being constant, and showed that the balanced incomplete block design was optimal.

Kshirsagar (1958) considered the same problem as Kempthorne (1956) but expressed his results in terms of the average variance of v-1 orthogonal normalized treatment contrasts. He showed that the generalised variance of these contrasts is proportional to the geometric mean of the non-zero eigenvalues of $\underset{\sim}{C}$. He further proved that the balanced incomplete block design, if it exists, optimizes both of these means.

Mote (1958) showed that the balanced incomplete block design maximizes the minimum non-zero eigenvalue of $\underset{\sim}{C}$.

Kiefer (1958) showed that the design which maximizes trace $(\underset{\sim}{C})$ has all diagonal elements of $\underset{\sim}{C}$ equal and is A-, D- and E-optimal.

Shah (1960) reviewed the above criteria. The three criteria are (a) minimize $\sum_{i=1}^{v-1} \lambda_i^{-1}$, (b) minimize $\prod_{i=1}^{v-1} \lambda_i^{-1}$, (c) maximize (minimum λ_i). Shah noted that the balanced incomplete block design, if it exists, is optimal on all these criteria, and that Kempthorne's and Kshirsagar's

results were valid for the case of unequal replications. As Shah pointed out, if a balanced incomplete block design does not exist, the criteria given above need not agree on the optimal design. Further, as all the non-zero eigenvalues are equal for a balanced incomplete block design he suggested the criterion (d) minimize $\sum_{i=1}^{v-1} (\lambda_i - \bar{\lambda})^2/v-1$, with $\sum_{i=1}^{v-1} \lambda_i$ constant. That is, among designs of a given size, minimize $\sum_{i=1}^{v-1} \lambda_i^2$.

Eccleston and Hedayat (1974) following on from the work of Shah (1960) suggested choosing the design as optimal by (i) forming a class of designs which maximize trace ($\underset{\sim}{C}$) and then (ii) selecting from this class the design which minimizes trace ($\underset{\sim}{C}^2$). They refer to this criterion as (M,S)-optimality. Although they allow different treatment replications they remark that the (M,S)-optimality of designs with unequal block sizes remains unsolved.

More recently, Mitchell and John (1976) considered optimality criteria for incomplete block designs, based on the eigenvalues of

$$\underset{\sim}{\Omega}^{-1} = \underset{\sim}{C} + \frac{r^2}{bk}\underset{\sim}{J},$$

where $r_i = r$ is the replication of treatment i; i=1, 2, ..., v; $k_j = k$ is the size of block j; j = 1, 2, ..., b; and $\underset{\sim}{J}$ is a $v \times v$ matrix of ones.

The three criteria considered by Mitchell and John (1976) were

(i) Maximize the product of the eigenvalues of $\underset{\sim}{\Omega}^{-1}$, i.e. D-optimality.

(ii) Maximize the harmonic mean of the eigenvalues of $\underset{\sim}{\Omega}^{-1}$, i.e. A-optimality.

(iii) Maximize the smallest eigenvalue of $\underset{\sim}{\Omega}^{-1}$, i.e. E-optimality.

It is important to note that in all the above considerations of optimality the block sizes are kept equal, the replications are sometimes kept equal and, most importantly of all, the criteria assume an equal interest in all possible comparisons among the treatments.

Few algorithms for constructing optimal block designs have appeared in the literature. Mitchell (1974) used his algorithm DETMAX to search for small D-optimal incomplete block designs. Mitchell and John (1976) mention a computer program they used to search for "regular graph" designs that optimize the criteria (i) - (iii) given above. Regular graph designs are either in balanced incomplete blocks or are such that every pair of treatments occurs in a_1 or a_2 blocks, where $a_2 = a_1 + 1$. John and Turner (1977) describe an application of a computer program that searches for the incidence matrix of an incomplete block design, given the matrix of treatment concurrences.

As already mentioned, all the criteria considered so far assume an equal interest in all treatment contrasts. Pearce (1975a) remarked that it is more usual for an experimenter to be interested in a particular set of treatment contrasts and to require a design that estimates these contrasts as efficiently as possible. He therefore considered certain "basic" contrasts, previously put forward by Caliński (1971), and suggested choosing as optimal the design that maximizes particular weighted functions of the efficiency factors of these contrasts. It is assumed that the i^{th} basic contrast has been assigned a weight ω_i $(0 \leq \omega_i \leq 1)$ and has an efficiency factor $\varepsilon_i (0 \leq \varepsilon_i \leq 1)$. He proposed three criteria:

(a) Maximize the arithmetic mean of the efficiency factors.

(b) Maximize the harmonic mean of the efficiency factors.

(c) Maximize the geometric mean of the efficiency factors.

Both (a) and (b) are easily generalized to deal with weighted efficiency factors but Pearce (1975a) preferred the weighted arithmetic mean because then an efficiency factor could take the value zero without calling for a modification of the criterion.

Although weighting of contrasts was considered by Pearce (1975a), the contrasts of interest had to be "basic".

If weighting is required then criterion (c) cannot be used because the weights get merged. The choice between (a) and (b) depends on how contrasts that get confounded are to be treated. If a contrast is totally confounded then its efficiency factor is zero. If criterion (a) is used and a contrast of interest is confounded then the value of the criterion can still be calculated. The only effect is that its value is reduced by the confounding. If criterion (b) is used however, and a contrast of interest is confounded the criterion takes an infinite value and the corresponding design must be withdrawn from consideration. A design which confounds a contrast with a low weighting might still be preferred under criterion (a) because the efficiencies of the other contrasts are high. Criterion (b) however does not admit this possibility.

It is clear therefore that before the work described in this chapter was begun there was no criterion available that could be used to choose a design that satisfied the following requirements:

(a) block sizes chosen by the experimenter (not necessarily all equal)

(b) certain contrasts to be estimated as efficiently as possible (not necessarily an orthogonal set).

(c) contrast weights can be specified that indicate which contrasts are the more important

Further, an algorithm that searched for designs that satisfy the above was needed.

A new criterion was therefore proposed by the present author. While it was being evaluated however, Freeman (1976) independently described the same one. However, Freeman did not describe any algorithm to obtain optimal designs, but only illustrated how the criterion could be used to decide among a set of competing designs.

The criterion and an algorithm that searches for designs that optimize it are described in the next section.

2.4 An algorithm for deriving optimal block designs

In Section 2.3 it was noted that one possible generalized inverse of the matrix $\underset{\sim}{C}$ was $\underset{\sim}{\Omega}$ as defined in equation (2.3.8). For calculating the variances of unconfounded contrasts it can be shown that the variance-covariance matrix of the estimates $\hat{\underset{\sim}{\gamma}}$ is

$$V(\hat{\underset{\sim}{\gamma}}) = \underset{\sim}{\Omega}\sigma^2.$$

If a set of treatment contrasts are specified and held columnwise in a matrix $\underset{\sim}{L}$ then

$$V(\underset{\sim}{L}'\hat{\underset{\sim}{\gamma}}) = \underset{\sim}{L}'\underset{\sim}{\Omega}\underset{\sim}{L}\sigma^2.$$

A sensible criterion for discriminating between designs is to choose the design that minimizes the sum of the weighted variances of the contrasts. That is, if $\underset{\sim}{\omega}^{\delta}$ is a diagonal matrix of contrast weights, the criterion is

$$\text{Minimize trace } (\underset{\sim}{L}'\underset{\sim}{\Omega}\underset{\sim}{L}\underset{\sim}{\omega}^{\delta}). \qquad (2.4.1)$$

The aim here is to obtain a design in which the most important contrasts are estimated with the highest precision.

Weighting is necessary because, as Pearce (1975a) pointed

out, any criterion that regards all treatment contrasts as equally important must give equal status to all treatments and therefore will evolve a design in which their replications are as nearly equal as possible and their concurrences also.

The differences between this criterion and that suggested by Pearce (1975a) are that Pearce's contrasts are what he termed 'basic'. Here the contrasts $\underset{\sim}{L}$, are not necessarily 'basic' or orthogonal, although orthogonal contrasts are usually specified as they lead to a full partitioning of the treatments sum of squares. Further, working with the weighted mean of the efficiencies as suggested by Pearce means that if a contrast of interest is confounded with blocks then the efficiency of the contrast is zero and Pearce's criterion can still be calculated. The variance of the confounded contrast however would be infinite and the criteria also.

It could be argued that it would be simpler to consider criteria based on $\underset{\sim}{\Omega}^{-1}$ and to avoid the matrix inversion. However, as the direct inversion of $\underset{\sim}{\Omega}^{-1}$ can be avoided and only connected designs are considered in this chapter, only criteria based on $\underset{\sim}{\Omega}$ will be considered in the following.

To use the criterion given in equation (2.4.1) however, the matrix $\underset{\sim}{\Omega}^{-1}$ must be inverted. Fortunately, the direct inversion of this matrix can be avoided by using an iterative formula given by Caliński (1971) and developed by Pearce, Caliński and Marshall (1974) as

$$\underset{\sim}{\Omega} = \underset{\sim}{r}^{-\delta} - \sum_{h=1}^{\infty} (\underset{\sim\sim}{1}\underset{\sim}{1}'/n - (\underset{\sim}{r}^{-\delta}\underset{\sim\sim}{N}\underset{\sim\sim}{k}^{-\delta}\underset{\sim}{N}')^{h}\underset{\sim}{r}^{-\delta}), \quad (2.4.2)$$

where $\underset{\sim}{\Omega}^{-1}$ is assumed to be non-singular.

If only connected designs are considered, $\underset{\sim}{\Omega}^{-1}$ will be as defined in equation (2.3.8) and equation (2.4.2) can be applied. The definition of a connected design adopted here is the one given by John (1971, p.233). He considered a block and a treatment

to be associated if the treatment occurred in the block. Two treatments A and B are then said to be connected if a chain of treatments and blocks can be formed as treatment-block-treatment-block-····treatment beginning with A and ending with B, such that every block is associated with both treatments adjacent to it. A design is connected if every pair of treatments is connected.

The formula (2.4.2) allows $\underset{\sim}{\Omega}$ to be approximated by

$$\underset{\sim}{\Omega}_H = \underset{\sim}{r}^{-\delta} - \sum_{h=1}^{H} (\underset{\sim\sim}{1}\underset{\sim}{1}'/n - (\underset{\sim}{r}^{-\delta}\underset{\sim\sim}{N}\underset{\sim\sim}{k}^{-\delta}\underset{\sim}{N}')^h \underset{\sim}{r}^{-\delta}), \quad (2.4.3)$$

where H is some finite positive integer.

The criterion (2.4.1) can then be re-defined as

$$\text{Minimize trace } (\underset{\sim}{L}'\underset{\sim}{\Omega}_H\underset{\sim\sim}{L}\underset{\sim\sim}{\omega}^{\delta}), \quad (2.4.4)$$

for some pre-determined value of H.

Now that the direct inversion of $\underset{\sim}{\Omega}^{-1}$ can be avoided, and if necessary an approximate inverse $\underset{\sim}{\Omega}_H$ can be calculated using a small value for H, it is convenient to consider using the criterion given in (2.4.4) in an algorithm to search for optimal designs. Such an algorithm is given below. The algorithm requires an initial design and the matrices $\underset{\sim}{L}$ and $\underset{\sim}{\omega}^{\delta}$ to be specified. Improvements to this design are then searched for by interchanging treatments between blocks. After the description of the algorithm a method of obtaining a starting design is given.

The Algorithm

Step 1. Evaluation of the starting design.

Calculate T = trace $(\underset{\sim}{L}'\underset{\sim}{\Omega}_H\underset{\sim\sim}{L}\underset{\sim\sim}{\omega}^{\delta})$ = trace $(\underset{\sim\sim}{L}\underset{\sim\sim}{\omega}^{\delta}\underset{\sim}{L}'\underset{\sim}{\Omega}_H)$ for some pre-determined value of H. The value of H will be considered below.

Step 2. Order the blocks on their degree of non-orthogonality.

Let N_{ji} be the number of times treatment j occurs in block i,

k_i the size of block i, and r_j the replication of treatment j. In an orthogonal design N_{ji}/k_i is proportional to r_j/n. If the allocation of treatments to blocks in an arbitrary design is to be improved by considering blocks individually those that should be amended first are those which, on aggregate, differ most from the ideal treatment allocation given by an orthogonal design. In the present algorithm the deviation of a block i from the ideal is measured by

$$M = \sum_{j=1}^{v} |N_{ji}/k_i - r_j/n|.$$

The blocks with the largest values of M are considered first.

Suppose that when ordered in decreasing magnitude of M the blocks are $b_1, b_2, \ldots, b_b$.

Step 3. Evaluate the effect of interchanging treatments between blocks.

For blocks b_1 and b_2 calculate the change in T that would occur if treatment ℓ in block b_1 was interchanged with treatment m in block b_2. Repeat this for all possible different interchanges of treatments between these blocks. Any interchange that would lead to a disconnected design is excluded.

Select that pair of treatments, if any, that when interchanged leads to a design that gives the largest reduction in T and proceed to Step 4. If there is a tie the first pair found are interchanged.

If no interchange causes a reduction in T consider blocks b_1 and b_3 and perform the interchanges as above. If after all blocks have been paired with b_1 and no interchange that reduces T has been found then blocks b_2 and b_3 are considered and so on. If when all possible pairs of blocks have been considered no reduction in T has been obtained by an interchange then stop and regard the current design as

optimal.

The alteration in $\underset{\sim}{\Omega}_H$ caused by interchanging say, treatment ℓ in block b_p with treatment m in block b_q is most easily made by noting that only certain elements of $\underset{\sim\sim}{N}\underset{\sim\sim}{k}^{-\delta}\underset{\sim}{N}'$ are altered by the interchange. If A_{ij} represents the i,j^{th} element of $\underset{\sim\sim}{N}\underset{\sim\sim}{k}^{-\delta}\underset{\sim}{N}'$ and N_{ij} the i,j^{th} element of $\underset{\sim}{N}$, the only elements altered by the interchange are those given below (where ΔA_{ij} represents the value of an element A_{ij} before the interchange less its value after).

$$\Delta A_{\ell\ell} = (2N_{\ell p} - 1)/k_p - (2N_{\ell q} + 1)/k_q,$$

$$\Delta A_{mm} = -(2N_{mp} + 1)/k_p + (2N_{mq} - 1)/k_q,$$

$$\Delta A_{\ell m} = (-N_{\ell p} + N_{mp} + 1)/k_p + (N_{\ell q} - N_{mq} + 1)/k_q,$$

$$\Delta A_{\ell j} = -\Delta A_{mj} = N_{jp}/k_p - N_{jq}/k_q, \quad (j \neq \ell \neq m)$$

Step 4. Update the design.

Interchange the treatment pair found in Step 3 to give a new design with its associated $\underset{\sim}{\Omega}_H$ and T. Return to Step 2.

2.5 Choice of starting design

The number of interchanges required is obviously minimized by a good choice of starting design. Two considerations apply. The first concerns the relative replications of the treatments and the following rule derived from the paper of Pearce (1975a) has proved satisfactory. This rule which assumes that the optimal design has high efficiency requires choosing the replications to minimize the sum of weighted treatment variances. In particular the replications are taken in the ratio of the quantities $(\sum_j \omega_j \ell_{ij}^2)^{\frac{1}{2}}$, where ℓ_{ij} is the element of the j^{th} contrast vector corresponding to the i^{th} treatment, all contrasts having been scaled to make $\sum_i \ell_{ij}^2 = 1$, and ω_j is the weight

assigned to that contrast. If as a result of using this rule a treatment is given zero replication then that treatment should be dropped from the design. The vector $\underset{\sim}{r}$ of replications is invariant over the interchanges and therefore if the rule gives non-integral replications various different replications may be used to give different starting designs.

The other consideration is to disperse each treatment as much as possible over blocks. A good method of doing this, which can be programmed onto a computer, is the following which constructs the incidence matrix $\underset{\sim}{N}$.

The treatments are allocated to one block at a time. Suppose the block to be considered first has size k_1 and the second has size k_2 and so on.

The elements N_{j1}, $j = 1, 2, \ldots, v$ are assigned integer values that are as closely as possible equal to

$$z_j = r_j \times k_1/n, \qquad j = 1, 2, \ldots, v,$$

subject of course to the constraint that $\sum_j N_{j1} = k_1$. When each plot in this block has been allocated a treatment the values r_j are reduced to $r_j - N_{j1}$ and the value of n is reduced to $n - k_1$. The values of z_j are then recalculated using k_2 and the adjusted values of r_j and n.

The elements N_{j2} are then assigned integer values that are as nearly equal as possible to the new z_j values, with $\sum_j N_{j2} = k_2$. The values of r_j and n are then reduced further by N_{j2} and k_2 respectively.

This procedure is repeated for the remaining blocks using the appropriate block size and the values of r_j and n that result from allocating treatments to the previous block.

If, however, the block sizes are all equal and the number of treatments is a multiple of the block size the above procedure will produce a disconnected design. Accordingly

the procedure should be modified as follows.

Allocate the treatments to one block at a time including as many different treatments in a block as possible. Having done this for the first block repeat this for the second block ensuring that a treatment from the first block occurs in the second. Repeat this for each successive block ensuring that at least one treatment in the current block occurs in the previous block until all treatments have occurred at least once. The remaining treatment replications are then allocated using the dispersal procedure described above beginning with the remaining replications in place of the r_j and the remaining number of plots in place of n.

2.6 An alternative form of the optimality criterion

The optimality criterion trace $(\underset{\sim}{L}'\underset{\sim}{\Omega}_H\underset{\sim}{L}\omega^{\delta})$ is only an approximation to trace $(\underset{\sim}{L}'\underset{\sim}{\Omega}\underset{\sim}{L}\omega^{\delta})$. To be useful in practice the value of H should be small to avoid wasting valuable computing time. It is important therefore to consider if an alternative approximation can be found that is more sensitive to the effect of interchanging treatments between blocks. In this section such an alternative is advanced and in the remaining sections the two approximations are compared. The performance of the algorithm is also discussed.

Consider now the alternative approximation. It is clear that in a comparative experiment the number of times treatments concur, i.e. occur together in the same block, is very important. The more often two treatments concur the more precisely will the difference between them be estimated. Pearce (1976) therefore proposed judging the extent to which a treatment participates in an experiment, not by its replication, but by the sum of its weighted concurrences. He defined these weighted concurrences as the off-diagonal elements of $\underset{\sim}{N}\underset{\sim}{k}^{-\delta}\underset{\sim}{N}'$,

the weights being the reciprocals of the block sizes. In particular he wrote

$$\underset{\sim}{N}\underset{\sim}{k}^{-\delta}\underset{\sim}{N}' = \underset{\sim}{p}^{\delta} + \underset{\sim}{W},$$

where $\underset{\sim}{W}$ is a matrix with all diagonal elements equal to zero and $\underset{\sim}{p}^{\delta}$ is a diagonal matrix. The matrix $\underset{\sim}{W}$ is called the matrix of weighted concurrences.

Further, we have

$$\underset{\sim}{r} = \underset{\sim}{N}\underset{\sim}{k}^{-\delta}\underset{\sim}{N}'\underset{\sim}{1} = \underset{\sim}{p} + \underset{\sim}{W}\underset{\sim}{1}.$$

Pearce (1976) also defined the vector $\underset{\sim}{q}$ of quasi-replications, where

$$\underset{\sim}{q} = \underset{\sim}{r} - \underset{\sim}{p} = \underset{\sim}{W}\underset{\sim}{1},$$

and wrote

$$u = \underset{\sim}{q}'\underset{\sim}{1}.$$

Using these he obtained an alternative generalized inverse of $\underset{\sim}{C}$ by writing equation (2.3.6) as

$$\underset{\sim}{Q} = (\underset{\sim}{q}^{\delta} - \underset{\sim}{W})\hat{\underset{\sim}{\gamma}}. \tag{2.6.1}$$

Then

$$\underset{\sim}{\Xi} = (\underset{\sim}{q}^{\delta} - \underset{\sim}{W} + \underset{\sim}{q}\underset{\sim}{q}'/u)^{-1} \tag{2.6.2}$$

is as legitimate a generalized inverse of $\underset{\sim}{C}$ as $\underset{\sim}{\Omega}$.

Therefore as $V(\hat{\underset{\sim}{\gamma}}) = \underset{\sim}{\Xi}\sigma^2$ we have that

$$\underset{\sim}{L}'\underset{\sim}{\Xi}\underset{\sim}{L}\underset{\sim}{\omega}^{\delta} = \underset{\sim}{L}'\underset{\sim}{\Omega}\underset{\sim}{L}\underset{\sim}{\omega}^{\delta}$$

If a formula for $\underset{\sim}{\Xi}$ that is analogous to that given in (2.4.2) for $\underset{\sim}{\Omega}$ can be derived then an alternative way of approximating the optimality criterion given in the previous section can be obtained. Let us consider deriving such a formula. Using the same approach as that adopted by Pearce, Caliński and Marshall (1974) to obtain a formula for $\underset{\sim}{\Omega}$, we write equation (2.6.1) as

$$\underset{\sim}{Q} = \underset{\sim}{q}^{\frac{1}{2}\delta}\underset{\sim}{F}\underset{\sim}{q}^{\frac{1}{2}\delta}\hat{\underset{\sim}{\gamma}}, \tag{2.6.3}$$

where

$$\underset{\sim}{q}^{\frac{1}{2}\delta}\underset{\sim}{q}^{\frac{1}{2}\delta} = \underset{\sim}{q}^{\delta}$$

and

$$\underset{\sim}{F} = \underset{\sim}{I} - \underset{\sim}{q}^{-\frac{1}{2}\delta}\underset{\sim\sim}{W}\underset{\sim}{q}^{-\frac{1}{2}\delta}, \tag{2.6.4}$$

with $\underset{\sim}{I}$, the identity matrix.

The equation (2.6.3) does not, however, provide a solution for $\hat{\underset{\sim}{\gamma}}$, because $\underset{\sim}{F}$ is singular as can be seen from the relation

$$\underset{\sim}{F}(\underset{\sim}{q}^{\frac{1}{2}\delta}\underset{\sim}{1}) = \underset{\sim}{q}^{-\frac{1}{2}\delta}(\underset{\sim}{q}^{\delta} - \underset{\sim}{W})\underset{\sim}{1} = \underset{\sim}{0}. \tag{2.6.5}$$

Since $\underset{\sim}{F}$ is symmetric it has a complete set of orthogonal eigenvectors, though they will not be uniquely determined if there are multiplicities among the eigenvalues. It will be convenient to order the eigenvectors so that those with zero eigenvalues come at the end, the eigenvector given by (2.6.5) coming last of all. It is now possible to write

$$\underset{\sim}{F} = \sum_{i=1}^{h} \epsilon_i \underset{\sim}{p}_i \underset{\sim}{p}_i', \tag{2.6.6}$$

where ϵ_i is the eigenvalue corresponding to the eigenvector $\underset{\sim}{p}_i$, and $h<v$. It will also be convenient to order so that those, g of them, with unit eigenvalues, $0 \le g \le h$, come at the beginning.

Writing

$$\underset{\sim}{X} = \begin{cases} \sum_{i=1}^{g} \underset{\sim}{p}_i \underset{\sim}{p}_i' & (g>0) \\ \underset{\sim}{0} & (g=0) \end{cases} \tag{2.6.7}$$

$$\underset{\sim}{Y} = \begin{cases} \sum_{i=g+1}^{h} \epsilon_i \underset{\sim}{p}_i \underset{\sim}{p}_i' & (g<h) \\ \underset{\sim}{0} & (g=h) \end{cases} \tag{2.6.8}$$

$$\underset{\sim}{Y}^{+} = \begin{cases} \sum_{i=g+1}^{h} \underset{\sim}{p}_i \underset{\sim}{p}_i'/\epsilon_i & (g<h) \\ \underset{\sim}{0} & (g=h) \end{cases} \tag{2.6.9}$$

$$\underset{\sim}{Z} = \sum_{i=h+1}^{v} \underset{\sim}{p}_i \underset{\sim}{p}_i', \tag{2.6.10}$$

and noting from the orthonormality of the eigenvectors that

$$\underset{\sim}{p}_i'\underset{\sim}{p}_i = 1, \quad \underset{\sim}{p}_i'\underset{\sim}{p}_j = 0 \qquad (i \neq j), \tag{2.6.11}$$

we have that

$$\left.\begin{aligned} &\underset{\sim\sim}{X}\underset{\sim\sim}{X} = \underset{\sim}{X}, \quad \underset{\sim}{Y}^+\underset{\sim}{Y} = \underset{\sim\sim}{Y}\underset{\sim\sim}{Y}^+ = \sum_{i=g+1}^{h} \underset{\sim}{p}_i\underset{\sim}{p}_i' \qquad (g<h) \\ &\text{and } \underset{\sim\sim}{Z}\underset{\sim\sim}{Z} = \underset{\sim}{Z}. \end{aligned}\right\} \tag{2.6.12}$$

Apart from those in (2.6.12) and $\underset{\sim\sim}{Y}\underset{\sim\sim}{Y}$ and $\underset{\sim\sim}{Y}\underset{\sim\sim}{Y}^+$ all pairwise products of $\underset{\sim}{X}$, $\underset{\sim}{Y}$, $\underset{\sim}{Y}^+$ and $\underset{\sim}{Z}$ are zero matrices. Also

$$\underset{\sim}{I} = \sum_{i=1}^{v} \underset{\sim}{p}_i\underset{\sim}{p}_i' = \underset{\sim}{X} + \underset{\sim\sim}{Y}\underset{\sim\sim}{Y}^+ + \underset{\sim}{Z}, \tag{2.6.13}$$

and (2.6.6) may be written as

$$\underset{\sim}{F} = \underset{\sim}{X} + \underset{\sim}{Y}, \tag{2.6.14}$$

so that

$$\underset{\sim}{Q} = \underset{\sim}{q}^{\frac{1}{2}\delta}(\underset{\sim}{X} + \underset{\sim}{Y})\underset{\sim}{q}^{\frac{1}{2}\delta}\hat{\underset{\sim}{\gamma}}. \tag{2.6.15}$$

If now $\underset{\sim}{\Xi}^{-1} = \underset{\sim}{q}^{\frac{1}{2}\delta}(\underset{\sim}{X} + \underset{\sim}{Y} + \underset{\sim}{Z})\underset{\sim}{q}^{\frac{1}{2}\delta}$, then $\underset{\sim}{\Xi}^{-1}$ is a generalization of the matrix proposed by Pearce (1976) and equals it when there are no zero eigenvalues apart from that in (2.6.5).

From (2.6.12) and (2.6.13)

$$\underset{\sim}{\Xi} = \underset{\sim}{q}^{-\frac{1}{2}\delta}(\underset{\sim}{X} + \underset{\sim}{Y}^+ + \underset{\sim}{Z})\underset{\sim}{q}^{-\frac{1}{2}\delta}. \tag{2.6.16}$$

The importance of $\underset{\sim}{\Xi}$ is that it is a generalized inverse of $\underset{\sim}{q}^{\frac{1}{2}\delta}\underset{\sim\sim}{F}\underset{\sim}{q}^{\frac{1}{2}\delta}$ and so although $\underset{\sim}{\gamma}$ cannot be estimated it is possible to estimate

$$\begin{aligned} \tilde{\underset{\sim}{\gamma}} &= \underset{\sim\sim}{\Xi}\underset{\sim}{q}^{\frac{1}{2}\delta}\underset{\sim\sim}{F}\underset{\sim}{q}^{\frac{1}{2}\delta}\underset{\sim}{\gamma} \\ &= \underset{\sim}{\gamma} - \tilde{\underset{\sim}{q}}^{\frac{1}{2}\delta}\underset{\sim\sim}{Z}\underset{\sim}{q}^{\frac{1}{2}\delta}\underset{\sim}{\gamma}. \end{aligned} \tag{2.6.17}$$

To estimate $\underset{\sim}{\gamma}$ constraints may be imposed as follows. Put

$$a_i = \underset{\sim}{p}_i'\underset{\sim}{q}^{\frac{1}{2}\delta}\underset{\sim}{\gamma}, \tag{2.6.18}$$

and it is required that

$$a_i = 0 \qquad (i>h), \tag{2.6.19}$$

which may be written alternatively as

$$\underset{\sim\sim}{Z}\underset{\sim}{q}^{\frac{1}{2}\delta}\underset{\sim}{\gamma} = 0. \tag{2.6.20}$$

One such constraint arises from (2.6.5),

$$\underset{\sim}{p}_v = \underset{\sim}{q}^{\frac{1}{2}\delta}\underset{\sim}{1}/u^{\frac{1}{2}}, \tag{2.6.21}$$

hence

$$a_v = \underset{\sim}{l}'\underset{\sim}{q}^{\delta}\underset{\sim}{\gamma}/u^{\frac{1}{2}} = 0$$

i.e.

$$\underset{\sim}{q}'\underset{\sim}{\gamma} = 0. \qquad (2.6.22)$$

When Pearce, Caliński and Marshall (1974) employed this method of estimating $\underset{\sim}{\gamma}$ using $\underset{\sim}{\Omega}$, the constraint they found that is analogous to (2.6.22) was

$$\underset{\sim}{r}'\underset{\sim}{\gamma} = 0,$$

the more usual constraint.

From (2.6.17), because $\underset{\sim}{\Xi}$ is a generalized inverse of $\underset{\sim}{q}^{\frac{1}{2}\delta}\underset{\sim}{F}\underset{\sim}{q}^{\frac{1}{2}\delta}$, it is permissible to write

$$\hat{\underset{\sim}{\gamma}} = \underset{\sim}{\Xi}\underset{\sim}{Q}, \qquad (2.6.23)$$

and it can be shown that the variance-covariance matrix of $\hat{\underset{\sim}{\gamma}}$ may be written as

$$V(\hat{\underset{\sim}{\gamma}}) = (\underset{\sim}{\Xi} - \underset{\sim}{q}^{-\frac{1}{2}\delta}\underset{\sim}{Z}\underset{\sim}{q}^{-\frac{1}{2}\delta})\sigma^2. \qquad (2.6.24)$$

An iterative formula for $\underset{\sim}{\Xi}$ may be obtained as follows.

Let

$$\begin{aligned}\underset{\sim}{t}_1 &= \underset{\sim}{q}^{-\frac{1}{2}\delta}(\underset{\sim}{I} - \underset{\sim}{Z})\underset{\sim}{q}^{-\frac{1}{2}\delta} \\ &= \underset{\sim}{q}^{-\frac{1}{2}\delta}(\underset{\sim}{X} + \underset{\sim}{Y}\underset{\sim}{Y}^{+})\underset{\sim}{q}^{-\frac{1}{2}\delta},\end{aligned}$$

and

$$\underset{\sim}{t}_j = \underset{\sim}{q}^{-\delta}\underset{\sim}{W}\underset{\sim}{t}_{j-1}, \qquad j \geq 2.$$

Consider first

$$\begin{aligned}\underset{\sim}{q}^{-\delta}\underset{\sim}{W} &= \underset{\sim}{q}^{-\frac{1}{2}\delta}(\underset{\sim}{q}^{-\frac{1}{2}\delta}\underset{\sim}{W}\underset{\sim}{q}^{-\frac{1}{2}\delta})\underset{\sim}{q}^{\frac{1}{2}\delta} \\ &= \underset{\sim}{q}^{-\frac{1}{2}\delta}(\underset{\sim}{I} - \underset{\sim}{F})\underset{\sim}{q}^{\frac{1}{2}\delta} \\ &= \underset{\sim}{q}^{-\frac{1}{2}\delta}(\underset{\sim}{Y}\underset{\sim}{Y}^{+} - \underset{\sim}{Y} + \underset{\sim}{Z})\underset{\sim}{q}^{\frac{1}{2}\delta} \\ &= \underset{\sim}{q}^{-\frac{1}{2}\delta}(\sum_{i=g+1}^{h}(1 - \epsilon_i)\underset{\sim}{p}_i\underset{\sim}{p}_i' + \underset{\sim}{Z})\underset{\sim}{q}^{\frac{1}{2}\delta} \\ &= \underset{\sim}{q}^{-\frac{1}{2}\delta}(\underset{\sim}{Y}_o + \underset{\sim}{Z})\underset{\sim}{q}^{\frac{1}{2}\delta},\end{aligned}$$

where

$$\underset{\sim}{Y}_o = \sum_{i=g+1}^{h}(1 - \epsilon_i)\underset{\sim}{p}_i\underset{\sim}{p}_i'.$$

Then

$$\begin{aligned} \underset{\sim}{t}_2 &= \underset{\sim}{q}^{-\frac{1}{2}\delta}(\underset{\sim}{Y}_o + \underset{\sim}{Z})\underset{\sim}{q}^{\frac{1}{2}\delta}\underset{\sim}{q}^{-\frac{1}{2}\delta}(\underset{\sim}{X} + \underset{\sim\sim}{YY}^+)\underset{\sim}{q}^{-\frac{1}{2}\delta} \\ &= \underset{\sim}{q}^{-\frac{1}{2}\delta}\underset{\sim o\sim\sim}{Y\,YY}^+\underset{\sim}{q}^{-\frac{1}{2}\delta} \\ &= \underset{\sim}{q}^{-\frac{1}{2}\delta}\underset{\sim o\sim}{Y}\underset{\sim}{q}^{-\frac{1}{2}\delta}. \end{aligned}$$

Therefore, if it is assumed that

$$\underset{\sim}{t}_j = \underset{\sim}{q}^{-\frac{1}{2}\delta}\underset{\sim o}{Y}^{j-1}\underset{\sim}{q}^{-\frac{1}{2}\delta}$$

then

$$\begin{aligned} \underset{\sim}{t}_{j+1} &= \underset{\sim}{q}^{-\frac{1}{2}\delta}(\underset{\sim}{Y}_o + \underset{\sim}{Z})\underset{\sim}{q}^{\frac{1}{2}\delta}\underset{\sim}{q}^{-\frac{1}{2}\delta}\underset{\sim o}{Y}^{j-1}\underset{\sim}{q}^{-\frac{1}{2}\delta} \\ &= \underset{\sim}{q}^{-\frac{1}{2}\delta}\underset{\sim o\sim}{Y}^{j}\underset{\sim}{q}^{-\frac{1}{2}\delta}. \end{aligned}$$

Therefore by induction, it must be true that

$$\underset{\sim}{t}_j = \underset{\sim}{q}^{-\frac{1}{2}\delta}\underset{\sim o}{Y}^{j-1}\underset{\sim}{q}^{-\frac{1}{2}\delta} \qquad \text{for all } j \geq 2.$$

Writing

$$\underset{\sim\sim}{YY}^+ = \underset{\sim o}{Y}^0 = \sum_{i=g+1}^{h} \underset{\sim i}{p}_i \underset{\sim i}{p}_i'$$

$$\underset{\sim}{t}_1 = \underset{\sim}{q}^{-\frac{1}{2}\delta}(\underset{\sim}{X} + \underset{\sim o}{Y}^0)\underset{\sim}{q}^{\frac{1}{2}\delta}$$

and then

$$\sum_{j=1}^{\infty} \underset{\sim}{t}_j = \underset{\sim}{q}^{-\frac{1}{2}\delta}(\underset{\sim}{X} + \sum_{j=1}^{\infty} \underset{\sim o}{Y}^{j-1})\underset{\sim}{q}^{-\frac{1}{2}\delta}.$$

But

$$\underset{\sim o}{Y}^j = \sum_{i=g+1}^{h} \{(1 - \epsilon_i)^j \underset{\sim i}{p}_i \underset{\sim i}{p}_i'\},$$

therefore

$$\begin{aligned} \sum_{j=1}^{\infty} \underset{\sim o}{Y}^{j-1} &= \sum_{i=g+1}^{h} \sum_{j=1}^{\infty} \{(1 - \epsilon_i)^{j-1} \underset{\sim i}{p}_i \underset{\sim i}{p}_i'\} \\ &= \sum_{i=g+1}^{h} \underset{\sim i}{p}_i \underset{\sim i}{p}_i' / \epsilon_i = \underset{\sim}{Y}^+ \end{aligned}$$

if

$$| (1 - \epsilon_i) | < 1 \text{ for all } i, \qquad g+1 \leq i \leq h.$$

For some designs this does not hold as will be seen in Section 2.8. However, where it does hold

$$\begin{aligned} \sum_{j=1}^{\infty} \underset{\sim}{t}_j &= \underset{\sim}{q}^{-\frac{1}{2}\delta}(\underset{\sim}{X} + \underset{\sim}{Y}^+)\underset{\sim}{q}^{-\frac{1}{2}\delta} \\ &= \underset{\sim}{\Xi} - \underset{\sim}{q}^{-\frac{1}{2}\delta}\underset{\sim\sim}{Z}\underset{\sim}{q}^{-\frac{1}{2}\delta} \quad . \end{aligned}$$

Provided no effects are totally confounded in the design the

matrix $\underset{\sim}{\Xi}$ can be calculated as follows.

Let

$$\underset{\sim}{t}_1 = \underset{\sim}{q}^{-\delta} - \underset{\sim\sim}{1}\underset{\sim}{1}'/u$$

and

$$\underset{\sim}{t}_j = \underset{\sim}{q}^{-\delta}\underset{\sim\sim}{W}\underset{\sim}{t}_{j-1}, \qquad j \geq 2$$

then

$$\underset{\sim}{\Xi} = \sum_{j=1}^{\infty} \underset{\sim}{t}_j + \underset{\sim\sim}{1}\underset{\sim}{1}'/u,$$

i.e.

$$\underset{\sim}{\Xi} = \underset{\sim}{q}^{-\delta} - \sum_{h=1}^{\infty} (\underset{\sim\sim}{1}\underset{\sim}{1}'/u - (\underset{\sim}{q}^{-\delta}\underset{\sim}{W})^h \underset{\sim}{q}^{-\delta}). \tag{2.6.25}$$

This enables $\underset{\sim}{\Xi}$ to be approximated by

$$\underset{\sim H}{\Xi} = \underset{\sim}{q}^{-\delta} - \sum_{h=1}^{H} (\underset{\sim\sim}{1}\underset{\sim}{1}'/u - (\underset{\sim}{q}^{-\delta}\underset{\sim}{W})^h \underset{\sim}{q}^{-\delta}), \tag{2.6.26}$$

for some finite positive integer H.

Therefore, a different approximation to the optimality criterion defined in Section 2.4 is

$$T = \text{trace } (\underset{\sim}{L}'\underset{\sim H}{\Xi}\underset{\sim\sim}{L}\omega^{\delta}). \tag{2.6.27}$$

2.7 Testing the algorithm and comparing the criteria

In this section both approximations to the criterion will be used to test the algorithm and the results compared. It will be recalled that the two approximations are $T = \text{trace } (\underset{\sim}{L}'\underset{\sim H}{\Omega}\underset{\sim\sim}{L}\omega^{\delta})$ and $T = \text{trace } (\underset{\sim}{L}'\underset{\sim H}{\Xi}\underset{\sim\sim}{L}\omega^{\delta})$.

Test 1

A design is required in three blocks of sizes 6, 12 and 18, respectively, where treatment 1 is to be compared as precisely as possible with the other four treatments.

The problem was presented to the algorithm in the following form.

$$\underset{\sim}{r}' = (12, 6, 6, 6, 6), \qquad \underset{\sim}{L}' = (\tfrac{1}{2})^{\frac{1}{2}} \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix}$$

and $\underset{\sim}{\omega}' = (1, 1, 1, 1)$.

The procedure described in Section 2.4 for generating a starting design did in this instance give the optimal design and so no interchanges were needed. Therefore, to test the algorithm the following deliberately perverse starting design was used:

Block

1 233333

2 344444455555

3 111111111111222225.

With H set equal to 4 the algorithm gave the following design using each of the approximations.

Block

1 112345

2 111122334455

3 111111222333444555.

This design is orthogonal and is the best possible with the given replications. A computer program that executes the algorithm took 0.336 CP seconds on the University of London CDC 7600 computer.

Test 2

Pearce (1975a) gives an example where a design is required in four blocks of five plots each and there are three treatments representing equally spaced doses of some chemical. The treatments are labelled here as 1, 2, and 3. Interest centres on the linear and curvilinear effects which are allocated weights of $1 - w_1$ and w_1, respectively.

The values of $\underset{\sim}{L}'$ and $\underset{\sim}{\omega}$ for this example are

$$\underset{\sim}{L}' = \begin{bmatrix} (\frac{1}{2})^{\frac{1}{2}} & 1 & 0 & -1 \\ (\frac{1}{6})^{\frac{1}{2}} & 1 & -2 & 1 \end{bmatrix} \quad \text{and} \quad \underset{\sim}{\omega} = \begin{bmatrix} 1 - w_1 \\ w_1 \end{bmatrix}.$$

The optimal design will depend on the value of w_1 and Freeman (1976) using his criterion, showed analytically that the designs given in Table 2.7.1 are optimal for the ranges of w_1 indicated.

Table 2.7.1. Freeman's optimal designs for Test 2.

Design		Replication	w_1 in the range
A)	11223 11223 12233 12233	(6, 8, 6)	(0.531, 1)
B)	11223 12233 11233 11233	(7, 6, 7)	(0.346, 0.531)
C)	11233 11233 11233 11233	(8, 4, 8)	(0.052, 0.346)
D)	11233 11233 11133 11333	(9, 2, 9)	(0, 0.052)

If the dispersal procedure given in Section 2.5 is used with treatment replications (6, 8, 6) then depending on how the non-integral values of z_j are dealt with alternative starting designs are possible. One of these is design A as given in Table 2.7.1. Similarly, using the dispersal rule with treatment replications of (7, 6, 7), (8, 4, 8) and (9, 2, 9) the designs B, C, D respectively, as given in Table 2.7.1 are obtained. Therefore, to test the algorithm the slightly perverse starting designs given in Table 2.7.2 were used for all values of w_1 in the range 0, (0.01), 1.0.

Table 2.7.2. Starting designs for Test 2.

W)	11233	X)	12223
	11233		11233
	12223		11233
	12223		11233
Y)	11133	Z)	13333
	11233		11233
	11223		11233
	12333		11113

Using the starting designs given in Table 2.7.2 the algorithm, with H equal to 4, gave the designs A, B, C and D, respectively, as optimal for the ranges of w_1 given in Table 2.7.3, when both approximations of the criterion were used.

Table 2.7.3. Optimal designs obtained by the present algorithm for Test 2.

Starting Design	Optimal Design	w_1 in the range
W	A	(0.34, 1.0)
X	B	(0.13, 1.0)
Y	C	(0, 1.0)
Z	D	(0, 1.0)

If the values of T are compared for the designs A, B, C and D, respectively, throughout the range of w_1 it is found that these designs are optimal for the ranges of w_1 given in Table 2.7.1.

This example illustrates the point that when there is some doubt about which replications to use with a particular weighting, the alternative optimal designs produced by the algorithm using the different replications should be compared using their respective T values. The design with the

smallest value of T for a particular weighting is optimal.

Test 3

To determine how well the algorithm performs, the designs considered by Mitchell (1973) were used. He used his computer algorithm DETMAX (1974) to obtain incomplete block designs with $n \leq 44$ and $v + b - 1 \leq 30$. He excluded designs where all block sizes and all treatment replications equalled two.

The major differences between DETMAX and the present algorithm are:

(i) DETMAX is based on the maximization of a determinant, i.e. D-optimality is sought,

(ii) equal interest in all treatment contrasts is assumed, i.e. weighting is not possible,

(iii) Mitchell recommends using different starting designs, sometimes as many as fifty, in his search for the optimum design,

(iv) "candidate" points are successively exchanged for points already in the design in an attempt to improve the value of the criterion.

Before comparing the designs found by DETMAX with those obtained by the algorithm described in Section 2.4, the designs obtained by the algorithm using each approximation of the trace criterion will be compared.

It will be recalled that one reason for using the approximation based on Ξ_H was that it might lead to a better approximation of the true value of the criterion. Further, it was thought that Ξ_H might be more sensitive to interchanges, as it is based explicity on treatment concurrences, and so result in the algorithm requiring fewer interchanges to reach the optimum design.

Therefore, in the following a comparison will be made not only of the properties of the final designs but also the number of interchanges attempted during the course of the algorithm. The contrasts and weights used were chosen to specify an equal interest in all treatment contrasts. The results of this comparison are summarized below.

For 31 sets of values of (v, k, r) the algorithm found the same design using $\underset{\sim}{\Omega}_H$ and $\underset{\sim}{\Xi}_H$ and attempted the same number of interchanges.

For 4 sets of values of (v, k, r) the algorithm found the same design using $\underset{\sim}{\Omega}_H$ and $\underset{\sim}{\Xi}_H$ but attempted different numbers of interchanges. These sets are given in Table 2.7.4.

Table 2.7.4. Designs where $\underset{\sim}{\Omega}_H$ and $\underset{\sim}{\Xi}_H$ gave the same design but different numbers of interchanges were attempted.

(v, k, r)	No. of ints. ($\underset{\sim}{\Omega}_H$)	No. of ints. ($\underset{\sim}{\Xi}_H$)
(5, 2, 6)	499	510
(6, 2, 6)	1417	1432
(6, 2, 7)	1909	1836
(8, 2, 5)	1178	1691

It can be seen that for only one of the designs in Table 2.7.4 the algorithm took a shorter route using $\underset{\sim}{\Xi}_H$.

Therefore, there is some preference for the $\underset{\sim}{\Omega}_H$ approximation. Further, for six of the sets of values, the $\underset{\sim}{\Omega}_H$ criterion led to a better design. These sets are given in Table 2.7.5 along with the efficiencies of the final designs found by each approximation. The efficiency is defined as the ratio of (a) the variance of any treatment difference in a randomized complete block design to (b) the

average variance of all treatment differences for the given design.

Table 2.7.5. Designs where $\underset{\sim}{\Omega}_H$ and $\underset{\sim}{\Xi}_H$ gave different final designs.

	$\underset{\sim}{\Omega}_H$		$\underset{\sim}{\Xi}_H$	
(v, k, r)	No. of ints.	Efficiency	No. of ints.	Efficiency
(6, 2, 3)	170	0.555556	138	0.531915
(7, 2, 4)	538	0.544987	510	0.541667
(8, 2, 4)	890	0.538462	599	0.516944
(9, 2, 4)	885	0.511140	854	0.506061
(10, 2, 4)	1540	0.500000	1178	0.489400

For all of the sets of values of (v, k, r) given in Table 2.7.5 the $\underset{\sim}{\Xi}_H$ took a shorter route but a less efficient design was obtained. One reason for this was that the series formula (2.6.25) did not converge. This difficulty will be examined in more detail later.

A description of the algorithm and a summary of the results obtained using $\underset{\sim}{\Xi}_H$ were given by Jones (1976). At that time the performance of the algorithm using $\underset{\sim}{\Omega}_H$ had not been assessed. Some examples of using the algorithm were given by Jones (1978).

It is clear from the above tests that the $\underset{\sim}{\Omega}_H$ approximation to the criterion is to be preferred, although it will be later shown that the difficulties associated with $\underset{\sim}{\Xi}_H$ can be overcome.

It is important now to compare the final designs obtained using the $\underset{\sim}{\Omega}_H$ approximation to the criterion with the final designs obtained by Mitchell. Mitchell compared his final designs with the best he could find in the literature. These best designs will be referred to as "assumed optima".

The results of the comparison of the final designs found by Mitchell and the present algorithm are summarized below. To aid the comparison values of T, D and E are given for some (v, k, r) in Tables 2.7.6 to 2.7.9. Here T is the trace criterion based on $\underset{\sim}{\Omega}_H$ with H equal to 4, D is Mitchell's determinant and E is the efficiency. The subscripts p, m and a are used to distinguish between the values of T, D and E appropriate to the present algorithm, Mitchell's algorithm and the assumed optima, respectively.

For twenty sets of values of (v, k, r) the present algorithm found the same designs as Mitchell's, which were the assumed optima. These sets of values were (4, 2, 4), (4, 2, 5), (4, 2, 7), (4, 2, 8), (4, 2, 10), (5, 2, 6), (5, 3, 3), (6, 2, 3), (6, 2, 4), (6, 2, 6), (6, 3, 3), (6, 3, 4), (6, 3, 6), (6, 3, 7), (6, 4, 4), (7, 5, 5), (8, 2, 3), (8, 2, 4), (8, 4, 4) and (8, 5, 5).

For five sets of values of (v, k, r) the present algorithm found the assumed optima where Mitchell's failed. These were (6, 4, 6), (8, 3, 3), (9, 3, 3), (10, 2, 3) and (10, 2, 4).

For two sets of values of (v, k, r) the present algorithm failed to match Mitchell's which found the assumed optima. These sets are given in Table 2.7.6.

Table 2.7.6. Designs where the present algorithm failed to find the assumed optima where Mitchell's succeeded.

(v, k, r)	$T_m = T_a$	T_p	$D_m = D_a$	D_p	$E_m = E_a$	E_p
(8, 4, 5)	1.6386	1.6441	0.34399E10	0.339739E10	0.8542	0.8514
(10, 4, 4)	2.7317	2.7336	0.50388E10	0.503733E10	0.8232	0.8231

For three sets of values of (v, k, r) the present

algorithm and Mitchell's failed to find the assumed optima. These sets are given in Table 2.7.7.

Table 2.7.7. Designs where both the present algorithm and Mitchell's failed to find the assumed optima.

(v, k, r)	T_a	T_p	D_a	D_p	E_a	E_p
(9, 4, 4)	2.3970	2.3976	0.46760E09	0.467251E09	0.8340	0.8338
(11, 4, 4)	3.0624	3.0661	0.55479E11	0.548561E11	0.8165	0.8147
(12, 3, 3)	5.3292	5.3360	0.16200E09	0.160161E09	0.6801	0.6777

For seven sets of values Mitchell's improved on the assumed optima and the present algorithm matched Mitchell's. These sets were (4, 2, 11), (6, 2, 7), (7, 2, 4), (8, 2, 5), (9, 2, 4), (10, 3, 3) and (11, 3, 3).

For one set of values of (v, k, r) Mitchell's failed to improve on the assumed optimum but the present algorithm improved on it. These values are given in Table 2.7.8.

Table 2.7.8. The design where Mitchell's failed to improve on the optimum but the present algorithm improved on it.

(v, k, r)	T_a	T_p	D_a	D_p	E_a	E_p
(13, 3, 3)	5.9043	5.8895	0.81202E09	0.825044E09	0.6667	0.6688

For two sets of values of (v, k, r) Mitchell's improved on the assumed optimum and the present algorithm improved on Mitchell's. These values are given in Table 2.7.9.

The above comparison reveals that for thirteen sets of values of (v, k, r) the present algorithm found a different design from Mitchell's. As Mitchell did not give details of his failures it is impossible to know if the present algorithm matched the three sets given in Table 2.7.7. Of the remaining

Table 2.7.9. Designs where Mitchell's improved on the assumed optima and the present algorithm improved on Mitchell's.

(v, k, r)	T_m	T_p	D_m	D_p	E_m	E_p
(12, 2, 3)	-	7.4630	0.12247E07	0.125440E07	0.4205	0.4241
(14, 3, 3)	-	6.4371	0.41998E10	0.431594E10	0.6587	0.6630

ten sets however, there were eight where Mitchell was bettered and two where he was not. Therefore, overall, the present algorithm based on the $\underset{\sim}{\Omega}_H$ approximation was better than Mitchell's, and as already mentioned, Mitchell's algorithm does not permit contrasts to be weighted.

It is apparent that for any given design the larger the value of H in equations (2.4.4) and (2.6.27) the better will be the approximation to the true value of the trace criterion. Therefore, to determine if a value of H larger than four would have led to any better designs Test 3 as described above was repeated using a value of H equal to ten. The result for the $\underset{\sim}{\Omega}_H$ approximation of the criterion was that one improved design was found. This was for (10, 4, 4) and the design found matched the assumed optimum which Mitchell's also found. For the $\underset{\sim}{\Xi}_H$ approximation improved designs for five sets of values of (v, k, r) were obtained. These sets are given in Table 2.7.10.

Of these five the design found for (7, 2, 4) matched Mitchell's which was an improvement on the assumed optimum, the designs found for (8, 2, 4) and (9, 2, 4) were not as good as those found by Mitchell's, the design found for (10, 2, 4) matched the assumed optimum which Mitchell's failed to obtain, and finally the design found for (10, 4, 4)

Table 2.7.10. Designs improved upon using the $\underset{\sim}{\Xi}_H$ approximation with H = 10 instead of 4.

(v, k, r)	D_p	E_p
(7, 2, 4)	0.30720E06	0.5450
(8, 2, 4)	0.19661E07	0.5250
(9, 2, 4)	0.12617E08	0.5092
(10, 2, 4)	0.83886E08	0.5000
(10, 4, 4)	0.50388E10	0.8232

matched the assumed optimum which was also found by Mitchell's. Further tests showed that a value of H equal to six was sufficient to obtain the design found when H equalled ten for (8, 2, 4), (9, 2, 4), (10, 2, 4) and (10, 4, 4), using the $\underset{\sim}{\Xi}_H$ approximation.

Therefore even when an increased value of H is used with the $\underset{\sim}{\Xi}_H$ approximation it still fails to match the $\underset{\sim}{\Omega}_H$ approximation for (6, 2, 3), (8, 2, 4) and (9, 2, 4). The $\underset{\sim}{\Omega}_H$ approximation is therefore to be preferred. It is important however, to discover why the $\underset{\sim}{\Xi}_H$ approximation does worse. One factor is that the series (2.6.25) fails to converge for some designs and so the approximation to the trace criterion is not a good one. In fact the series failed to converge for all the starting designs obtained when k equalled two, except for (5, 2, 6).

The reasons why the series formula for $\underset{\sim}{\Xi}$ fails to converge are considered in the next section.

2.8 Some results concerning $\underset{\sim}{\Xi}_H$

The $\underset{\sim}{\Xi}_H$ approximation to the criterion was developed because it was based directly on treatment concurrences and so might possibly be more sensitive to treatment interchanges.

It was therefore surprising that it fared so badly when compared with the $\underset{\sim}{\Omega}_H$ approximation. One reason why $\underset{\sim}{\Xi}_H$ did so badly was that the series formula (2.6.25) did not converge for some designs. Therefore some consideration was given to determine why the series did not converge.

In this section some results are given for the case where all quasi replications are equal, i.e. $\underset{\sim}{q} = q\underset{\sim}{1}$. This is an important case to consider because it corresponds to an equal interest in all treatment comparisons. Then $\underset{\sim}{q}^{-\delta}\underset{\sim}{W} = \frac{1}{q}\underset{\sim}{W}$, a real symmetric matrix, has a complete set of orthogonal eigenvectors. Let $\lambda_1, \lambda_2, \ldots, \lambda_v$ be the eigenvalues of $\frac{1}{q}\underset{\sim}{W}$, with corresponding orthogonal eigenvectors $\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_v$, which have been scaled so that $\underset{\sim}{x}_i'\underset{\sim}{x}_i = 1$, $i = 1, 2, \ldots, v$.

If $\underset{\sim}{q} = q\underset{\sim}{1}$ then $u = \underset{\sim}{q}'\underset{\sim}{1} = q\underset{\sim}{1}'\underset{\sim}{1} = qv$, and therefore $\underset{\sim}{1}'\underset{\sim}{1} = \frac{u}{q}$.

One eigenvector of $\frac{1}{q}\underset{\sim}{W}$ is always $\underset{\sim}{1}$ as can be seen from

$$\frac{1}{q}\underset{\sim}{W}\underset{\sim}{1} = \frac{1}{q}\underset{\sim}{q} = \frac{1}{q}q\underset{\sim}{1} = 1\underset{\sim}{1}.$$

In order that this eigenvector satisfies the scaling requirement that $\underset{\sim}{x}_i'\underset{\sim}{x}_i = 1$, $\underset{\sim}{1}$ must be divided by $v^{\frac{1}{2}} = (\frac{u}{q})^{\frac{1}{2}}$. Therefore $\frac{1}{q}\underset{\sim}{W}$ always has one eigenvector equal to $\underset{\sim}{1}(\frac{q}{u})^{\frac{1}{2}}$, with eigenvalue equal to 1.

In what follows it will be assumed that

$$\left.\begin{aligned} \underset{\sim}{x}_1 &= \underset{\sim}{1}(\tfrac{q}{u})^{\frac{1}{2}} \\ \text{and} \qquad & \\ \lambda_1 &= 1 \end{aligned}\right\} \qquad (2.8.1)$$

Experience has shown that the series fails to converge when one other eigenvalue takes the value -1. In the following if -1 is an eigenvalue it will be labelled as λ_2.

Further, it will be noted that because

$$\frac{1}{q}\underset{\sim}{W} = \sum_{\ell=1}^{v} \lambda_\ell \underset{\sim}{x}_\ell \underset{\sim}{x}_\ell', \tag{2.8.2}$$

then

$$(\frac{1}{q}\underset{\sim}{W})^h = \sum_{\ell=1}^{v} \lambda_\ell^h \underset{\sim}{x}_\ell \underset{\sim}{x}_\ell'. \tag{2.8.3}$$

Theorem 2.8.1 If $\underset{\sim}{q} = q\underset{\sim}{1}$ and the eigenvalues of $(\frac{1}{q}\underset{\sim}{W})$ are such that $\lambda_1 = 1$, $\lambda_2 = -1$ and $0 \leq \lambda_j < 1$, for $j \neq 1$, $j \neq 2$, then for H even

$$\underset{\sim}{\Xi} - \underset{\sim}{\Xi}_H = -\frac{1}{2q}\underset{\sim}{x}_2 \underset{\sim}{x}_2'.$$

Proof

$$\underset{\sim}{\Xi}_H = \frac{1}{q}\underset{\sim}{I} - \sum_{h=1}^{H} [\underset{\sim}{1}\underset{\sim}{1}'/u - (\frac{1}{q}\underset{\sim}{W})^h \frac{1}{q}],$$

from equation (2.6.26).

That is,

$$q\underset{\sim}{\Xi}_H = \underset{\sim}{I} - \sum_{h=1}^{H} [(\frac{q}{u})\underset{\sim}{1}\underset{\sim}{1}' - (\frac{1}{q}\underset{\sim}{W})^h],$$

which may be written as

$$q\underset{\sim}{\Xi}_H = \sum_{\ell=1}^{v} \underset{\sim}{x}_\ell \underset{\sim}{x}_\ell' - \sum_{h=1}^{H} (\frac{q}{u})\underset{\sim}{1}\underset{\sim}{1}' + \sum_{h=1}^{H} [\sum_{\ell=1}^{v} \lambda_\ell^h \underset{\sim}{x}_\ell \underset{\sim}{x}_\ell'],$$

where $\underset{\sim}{I}$ has been expressed as $\sum_{\ell=1}^{v} \underset{\sim}{x}_\ell \underset{\sim}{x}_\ell'$.

Now

$$(\frac{q}{u})\underset{\sim}{1}\underset{\sim}{1}' = \underset{\sim}{x}_1 \underset{\sim}{x}_1',$$

and

$$\sum_{h=1}^{H} \sum_{\ell=1}^{v} \lambda_\ell^h \underset{\sim}{x}_\ell \underset{\sim}{x}_\ell' = H\underset{\sim}{x}_1 \underset{\sim}{x}_1' + \sum_{h=1}^{H} (-1)^h \underset{\sim}{x}_2 \underset{\sim}{x}_2' + \sum_{h=1}^{H} \sum_{\ell=3}^{v} \lambda_\ell^h \underset{\sim}{x}_\ell \underset{\sim}{x}_\ell',$$

therefore if H is even and chosen so large that, given any $\varepsilon > 0$,

$$|\lambda_\ell|^H < \varepsilon, \quad \ell \geq 3,$$

then

$$\sum_{h=1}^{H} \sum_{\ell=1}^{v} \lambda_\ell^h \underset{\sim}{x}_\ell \underset{\sim}{x}_\ell' = H\underset{\sim}{x}_1 \underset{\sim}{x}_1' + \sum_{\ell=3}^{v} (\lambda_\ell/1 - \lambda_\ell) \underset{\sim}{x}_\ell \underset{\sim}{x}_\ell'.$$

Therefore,

$$q\underset{\sim}{\Xi}_H = \sum_{\ell=1}^{v} \underset{\sim}{x}_\ell \underset{\sim}{x}_\ell' - H\underset{\sim}{x}_1\underset{\sim}{x}_1' + H\underset{\sim}{x}_1\underset{\sim}{x}_1' + \sum_{\ell=3}^{v} (\lambda_\ell/1-\lambda_\ell)\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell'. \tag{2.8.4}$$

Consider now

$$\underset{\sim}{\Xi}^{-1} = q\underset{\sim}{I} - \underset{\sim}{W} + q^2\underset{\sim\sim}{1}\underset{\sim\sim}{1}'/u,$$

from equation (2.6.2).

That is,

$$\begin{aligned}
\frac{1}{q}\underset{\sim}{\Xi}^{-1} &= \underset{\sim}{I} - \frac{1}{q}\underset{\sim}{W} + \left(\frac{q}{u}\right)\underset{\sim\sim}{1}\underset{\sim\sim}{1}' \\
&= \sum_{\ell=1}^{v} \underset{\sim}{x}_\ell\underset{\sim}{x}_\ell' - \sum_{\ell=1}^{v} \lambda_\ell \underset{\sim}{x}_\ell\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell' + \underset{\sim}{x}_1\underset{\sim}{x}_1' \\
&= \sum_{\ell=1}^{v} (1 - \lambda_\ell)\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell' + \underset{\sim}{x}_1\underset{\sim}{x}_1' \\
&= \underset{\sim}{x}_1\underset{\sim}{x}_1' + 2\underset{\sim}{x}_2\underset{\sim}{x}_2' + \sum_{\ell=3}^{v} (1 - \lambda_\ell)\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell'. \qquad (2.8.5)
\end{aligned}$$

Therefore,

$$q\underset{\sim}{\Xi} = \underset{\sim}{x}_1\underset{\sim}{x}_1' + \frac{1}{2}\underset{\sim}{x}_2\underset{\sim}{x}_2' + \sum_{\ell=3}^{v} (1/1 - \lambda_\ell)\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell',$$

and consequently, using (2.8.4),

$$\begin{aligned}
q\underset{\sim}{\Xi} - q\underset{\sim}{\Xi}_H &= \sum_{\ell=3}^{v} (1/1 - \lambda_\ell)\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell' + \underset{\sim}{x}_1\underset{\sim}{x}_1' + \frac{1}{2}\underset{\sim}{x}_2\underset{\sim}{x}_2' \\
&\quad - \sum_{\ell=1}^{v} \underset{\sim}{x}_\ell\underset{\sim}{x}_\ell' - \sum_{\ell=3}^{v} (\lambda_\ell/1 - \lambda_\ell)\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell' \\
&= \sum_{\ell=3}^{v} [(1/1 - \lambda_\ell) - 1 - (\lambda_\ell/1 - \lambda_\ell)]\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell' \\
&\quad - \underset{\sim}{x}_1\underset{\sim}{x}_1' - \underset{\sim}{x}_2\underset{\sim}{x}_2' + \underset{\sim}{x}_1\underset{\sim}{x}_1' + \frac{1}{2}\underset{\sim}{x}_2\underset{\sim}{x}_2' \\
&= -\frac{1}{2}\underset{\sim}{x}_2\underset{\sim}{x}_2'.
\end{aligned}$$

That is,

$$\underset{\sim}{\Xi} - \underset{\sim}{\Xi}_H = -\frac{1}{2q}\underset{\sim}{x}_2\underset{\sim}{x}_2'.$$

Therefore, when $\underset{\sim}{q} = q\underset{\sim}{1}$ and the conditions of Theorem 2.8.1 hold, it is possible to calculate $\underset{\sim}{\Xi}$ using the series (2.6.26) by adjusting $\underset{\sim}{\Xi}_H$ by a factor of $-\frac{1}{2q}\underset{\sim}{x}_2\underset{\sim}{x}_2'$, where H is even and large enough for $|\lambda_\ell|^H < \varepsilon$.

Theorem 2.8.2 If the conditions of Theorem 2.8.1 hold, then

$$\underset{\sim}{\Xi}_H\underset{\sim}{\Xi}^{-1} = \underset{\sim}{I} + \underset{\sim}{x}_2\underset{\sim}{x}_2'.$$

Proof

$$\underset{\sim}{\Xi}_H\underset{\sim}{\Xi}^{-1} = \frac{1}{q}\Big[\sum_{\ell=1}^{v}\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell' + \sum_{\ell=3}^{v}(\lambda_\ell/1-\lambda_\ell)\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell'\Big] \times q\Big[\sum_{\ell=3}^{v}(1-\lambda_\ell)\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell' + \underset{\sim}{x}_1\underset{\sim}{x}_1' + 2\underset{\sim}{x}_2\underset{\sim}{x}_2'\Big],$$

using equation (2.8.4) and (2.8.5).

That is,

$$\underset{\sim}{\Xi}_H\underset{\sim}{\Xi}^{-1} = \sum_{\ell=1}^{v}\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell'\Big[\sum_{\ell=3}^{v}(1-\lambda_\ell)\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell' + \underset{\sim}{x}_1\underset{\sim}{x}_1' + 2\underset{\sim}{x}_2\underset{\sim}{x}_2'\Big] + \sum_{\ell=3}^{v}(\lambda_\ell/1-\lambda_\ell)\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell'\Big[\sum_{\ell=3}^{v}(1-\lambda_\ell)\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell' + \underset{\sim}{x}_1\underset{\sim}{x}_1' + 2\underset{\sim}{x}_2\underset{\sim}{x}_2'\Big].$$

Recalling that $\underset{\sim}{x}_\ell'\underset{\sim}{x}_\ell = 1$ and $\underset{\sim}{x}_\ell'\underset{\sim}{x}_j = 0$, $\ell \neq j$,

$$\begin{aligned}\underset{\sim}{\Xi}_H\underset{\sim}{\Xi}^{-1} &= \sum_{\ell=3}^{v}(1-\lambda_\ell)\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell' + \underset{\sim}{x}_1\underset{\sim}{x}_1' + 2\underset{\sim}{x}_2\underset{\sim}{x}_2 + \sum_{\ell=3}^{v}\lambda_\ell\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell'\\ &= \sum_{\ell=3}^{v}\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell' + \underset{\sim}{x}_1\underset{\sim}{x}_1' + 2\underset{\sim}{x}_2\underset{\sim}{x}_2'\\ &= \sum_{\ell=1}^{v}\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell' + \underset{\sim}{x}_2\underset{\sim}{x}_2'\\ &= \underset{\sim}{I} + \underset{\sim}{x}_2\underset{\sim}{x}_2'\end{aligned}$$

Theorem 2.8.3 If $\underset{\sim}{q} = q\underset{\sim}{1}$ and $\lambda_1 = 1$, $\lambda_2 = -1$, $0 \leq \lambda_j < 1$, $j \neq 1$, $j \neq 2$, then

$$\underset{\sim}{\Xi}_H = \frac{1}{q}\underset{\sim}{I} - \sum_{\ell=3}^{v}\Big[(\lambda_\ell/1-\lambda_\ell)\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell'\Big] + \sum_{h=1}^{H}(-1)^h\underset{\sim}{x}_2\underset{\sim}{x}_2'/q,$$

where H is such that, given $\varepsilon > 0$, $|\lambda_\ell|^H < \varepsilon$ for all ℓ.

Proof

$$\begin{aligned}\underset{\sim}{\Xi}_H &= \frac{1}{q}\underset{\sim}{I} - \sum_{h=1}^{H}\Big[\underset{\sim}{1}\underset{\sim}{1}'/u - \Big(\frac{1}{q}\underset{\sim}{W}\Big)^h\frac{1}{q}\Big]\\ &= \frac{1}{q}\underset{\sim}{I} - \sum_{h=1}^{H}\underset{\sim}{1}\underset{\sim}{1}'/u + \sum_{h=1}^{H}\sum_{\ell=1}^{v}\Big[\lambda_\ell^h\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell'\Big]\frac{1}{q}\\ &= \frac{1}{q}\underset{\sim}{I} - H\underset{\sim}{1}\underset{\sim}{1}'/u + H\underset{\sim}{x}_1\underset{\sim}{x}_1'/q + \sum_{h=1}^{H}(-1)^h\underset{\sim}{x}_2\underset{\sim}{x}_2'\frac{1}{q}\\ &\quad + \sum_{h=1}^{H}\Big[\sum_{\ell=3}^{v}\lambda_\ell^h\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell'\Big]\frac{1}{q}.\end{aligned}$$

Recalling that

$$\underset{\sim}{x}_1\underset{\sim}{x}_1' = \left(\frac{q}{u}\right)\underset{\sim}{1}\underset{\sim}{1}'$$

and

$$\sum_{h=1}^{H}\left[\sum_{\ell=3}^{v}\lambda_\ell^h \underset{\sim}{x}_\ell\underset{\sim}{x}_\ell'\right] = \sum_{\ell=3}^{v}(\lambda_\ell/1-\lambda_\ell)\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell' + H\underset{\sim}{x}_1\underset{\sim}{x}_1',$$

it can be seen that

$$\begin{aligned}\underset{\sim}{\Xi}_H &= \frac{1}{q}\underset{\sim}{I} - H\underset{\sim}{1}\underset{\sim}{1}'/u + H\underset{\sim}{x}_1\underset{\sim}{x}_1'/q + \sum_{h=1}^{H}(-1)^h\underset{\sim}{x}_2\underset{\sim}{x}_2'/q \\ &\quad + H\underset{\sim}{1}\underset{\sim}{1}'/u + \sum_{h=1}^{H}(-1)^h\underset{\sim}{x}_2\underset{\sim}{x}_2'/q + \sum_{\ell=3}^{v}(\lambda_\ell/1-\lambda_\ell)\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell'\frac{1}{q} \\ &= \frac{1}{q}\underset{\sim}{I} + \sum_{\ell=3}^{v}\left[(\lambda_\ell/1-\lambda_\ell)\frac{1}{q}\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell'\right] + \sum_{h=1}^{H}(-1)^h\underset{\sim}{x}_2\underset{\sim}{x}_2'/q.\end{aligned}$$

Corollary 2.8.3 For H even

$$\underset{\sim}{\Xi}_H = \frac{1}{q}\underset{\sim}{I} + \sum_{\ell=3}^{v}(\lambda_\ell/1-\lambda_\ell)\frac{1}{q}\underset{\sim}{x}_\ell\underset{\sim}{x}_\ell'.$$

Theorem 2.8.4 If $\underset{\sim}{q} = q\underset{\sim}{1}$ and $v = 2$, then

$$\lambda_2 = -1 \text{ and } \underset{\sim}{x}_2 = \begin{bmatrix}1\\-1\end{bmatrix}\left(\frac{q}{u}\right)^{\frac{1}{2}}.$$

Proof

If $v = 2$ then $\underset{\sim}{W}$ is of the form

$$\underset{\sim}{W} = \begin{bmatrix}0 & \beta\\ \beta & 0\end{bmatrix},$$

where β is the weighted concurrence of the two treatments.

Therefore,

$$\underset{\sim}{q} = \underset{\sim}{W}\underset{\sim}{1} = \begin{bmatrix}\beta\\ \beta\end{bmatrix} = \beta\underset{\sim}{1}$$

and

$$\underset{\sim}{q}^{-\delta}\underset{\sim}{W}\begin{bmatrix}1\\-1\end{bmatrix} = \frac{1}{\beta}\begin{bmatrix}0 & \beta\\ \beta & 0\end{bmatrix}\begin{bmatrix}1\\-1\end{bmatrix} = \begin{bmatrix}-1\\1\end{bmatrix} = (-1)\begin{bmatrix}1\\-1\end{bmatrix}.$$

That is, $\lambda_2 = -1$ and scaling $\underset{\sim}{x}_2$ so that $\underset{\sim}{x}_2'\underset{\sim}{x}_2 = 1$ implies that

$$\underset{\sim}{x}_2 = \begin{bmatrix}1\\-1\end{bmatrix}\left(\frac{q}{u}\right)^{\frac{1}{2}}.$$

This last Theorem therefore states that for any block design with two treatments the series (2.6.26) for $\underset{\sim}{\Xi}_H$

will not converge.

To illustrate some of the above Theorems consider the following design which has 6 treatments, each equally replicated, in 9 blocks of size 2.

Block	
I	1 2
II	2 3
III	3 4
IV	4 5
V	5 6
VI	6 1
VII	1 4
VIII	2 5
IX	3 6

For this design

$$q = 1.5$$

and

$$\frac{1}{q}\underset{\sim}{W} = \frac{1}{3}\begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}.$$

The eigenvalues of $\frac{1}{q}\underset{\sim}{W}$ are

$$\lambda_1 = 1,\ \lambda_2 = -1,\ \lambda_j = 0 \quad j \neq 1,\ j \neq 2,$$

and the eigenvectors $\underset{\sim}{x}_1$ and $\underset{\sim}{x}_2$ are

$$\underset{\sim}{x}_1' = 0.408248\ \underset{\sim}{1}'$$

and

$$\underset{\sim}{x}_2' = 0.408248\ [1, -1, 1, -1, 1, -1].$$

Then,

$$\underset{\sim}{\Xi}^{-1} = \begin{bmatrix} 1.75 & -0.25 & 0.25 & -0.25 & 0.25 & -0.25 \\ -0.25 & 1.75 & -0.25 & 0.25 & -0.25 & 0.25 \\ 0.25 & -0.25 & 1.75 & -0.25 & 0.25 & -0.25 \\ -0.25 & 0.25 & -0.25 & 1.75 & -0.25 & 0.25 \\ 0.25 & -0.25 & 0.25 & -0.25 & 1.75 & -0.25 \\ -0.25 & 0.25 & -0.25 & 0.25 & -0.25 & 1.75 \end{bmatrix},$$

$$\underset{\sim}{\Xi} = \begin{bmatrix} 0.6111 & 0.0556 & -0.0556 & 0.0556 & -0.0556 & 0.0556 \\ 0.0556 & 0.6111 & 0.0556 & -0.0556 & 0.0556 & -0.0556 \\ -0.0556 & 0.0556 & 0.6111 & 0.0556 & -0.0556 & 0.0556 \\ 0.0556 & -0.0556 & 0.0556 & 0.6111 & 0.0556 & -0.0556 \\ -0.0556 & 0.0556 & -0.0556 & 0.0556 & 0.6111 & 0.0556 \\ 0.0556 & -0.0556 & 0.0556 & -0.0556 & 0.0556 & 0.6111 \end{bmatrix}$$

and

$$\underset{\sim}{\Xi}_H = \begin{bmatrix} 0.6667 & & & & & \\ & 0.6667 & & & & \\ & & 0.6667 & & & \\ & & & 0.6667 & & \\ & & & & 0.6667 & \\ & & & & & 0.6667 \end{bmatrix}$$

for H even.

To illustrate Theorem 2.8.1, it can be seen that

$$\underset{\sim}{\Xi} - \underset{\sim}{\Xi}_H = \begin{bmatrix} -0.0556 & 0.0556 & -0.0556 & 0.0556 & -0.0556 & 0.0556 \\ 0.0556 & -0.0556 & 0.0556 & -0.0556 & 0.0556 & -0.0556 \\ -0.0556 & 0.0556 & -0.0556 & 0.0556 & -0.0556 & 0.0556 \\ 0.0556 & -0.0556 & 0.0556 & -0.0556 & 0.0556 & -0.0556 \\ -0.0556 & 0.0556 & -0.0556 & 0.0556 & -0.0556 & 0.0556 \\ 0.0556 & -0.0556 & 0.0556 & -0.0556 & 0.0556 & -0.0556 \end{bmatrix}$$

and

$$-\frac{1}{2q}\underset{\sim}{x}_2\underset{\sim}{x}_2' = -\frac{1}{2\times 1.5}(0.408248)^2 \begin{bmatrix} 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 \end{bmatrix},$$

where

$$\frac{1}{2\times 1.5}(0.408248)^2 = 0.0556.$$

To illustrate Theorem 2.8.2 it can be seen that

$$\underset{\sim}{\Xi}_H\underset{\sim}{\Xi}^{-1} = \begin{bmatrix} 1.1667 & -0.1667 & 0.1667 & -0.1667 & 0.1667 & -0.1667 \\ -0.1667 & 1.1667 & -0.1667 & 0.1667 & -0.1667 & 0.1667 \\ 0.1667 & -0.1667 & 1.1667 & -0.1667 & 0.1667 & -0.1667 \\ -0.1667 & 0.1667 & -0.1667 & 1.1667 & -0.1667 & 0.1667 \\ 0.1667 & -0.1667 & 0.1667 & -0.1667 & 1.1667 & -0.1667 \\ -0.1667 & 0.1667 & -0.1667 & 0.1667 & -0.1667 & 1.1667 \end{bmatrix}$$

and a typical element of $\underset{\sim}{x}_2\underset{\sim}{x}_2'$ is

$$\pm (0.408248)^2 = \pm (0.1667).$$

To illustrate Theorem 2.8.3 it can be seen that for H even

$$\underset{\sim}{\Xi}_H = \frac{1}{q}\underset{\sim}{I} = \frac{1}{1.5}\underset{\sim}{I} = 0.6667\ \underset{\sim}{I}.$$

Therefore whenever the series (2.6.25) does not converge and $\lambda_1 = 1$, $\lambda_2 = -1$ and $0 \leq \lambda_j < 1$, $j \neq 1$, $j \neq 2$, $\underset{\sim}{\Xi}$ can still be calculated by ensuring H is even and then calculating $\underset{\sim}{x}_2\underset{\sim}{x}_2'$ from the relationship

$$\underset{\sim}{\Xi}_H \underset{\sim}{\Xi}^{-1} = \underset{\sim}{I} + \underset{\sim}{x}_2 \underset{\sim}{x}_2'$$

and then putting

$$\underset{\sim}{\Xi} = \underset{\sim}{\Xi}_H - \frac{1}{2q} \underset{\sim}{x}_2 \underset{\sim}{x}_2'.$$

In the above example,

$$\underset{\sim}{\Xi}_H \underset{\sim}{\Xi}^{-1} = \begin{bmatrix} 1.1667 & -0.1667 & 0.1667 & -0.1667 & 0.1667 & -0.1667 \\ -0.1667 & 1.1667 & -0.1667 & 0.1667 & -0.1667 & 0.1667 \\ 0.1667 & -0.1667 & 1.1667 & -0.1667 & 0.1667 & -0.1667 \\ -0.1667 & 0.1667 & -0.1667 & 1.1667 & -0.1667 & 0.1667 \\ 0.1667 & -0.1667 & 0.1667 & -0.1667 & 1.1667 & -0.1667 \\ -0.1667 & 0.1667 & -0.1667 & 0.1667 & -0.1667 & 1.1667 \end{bmatrix}$$

which implies that

$$\underset{\sim}{x}_2 \underset{\sim}{x}_2' = 0.1667 \begin{bmatrix} 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 \end{bmatrix}$$

and so $\underset{\sim}{\Xi}$ can be found from $\underset{\sim}{\Xi}_H$.

When deriving the expression for $\underset{\sim}{\Xi}_H$ as given in (2.6.26) use was made of the eigensystem of $\underset{\sim}{F}$ as defined in (2.6.6). It was noted that $\underset{\sim}{\Xi}_H$ will not converge to $\underset{\sim}{\Xi}$ if $\underset{\sim}{F}$ has an eigenvalue ϵ_i such that $|1 - \epsilon_i| \geq 1$ for $g + 1 \leq i \leq h$. In fact, in the above example the eigenvalues are $\epsilon_i = 1$, $i = 1, 2, 3$ and 4, $\epsilon_5 = 2$ and $\epsilon_6 = 0$. Therefore because $|1 - \epsilon_5| = 1$ the series (2.6.25) fails to converge.

After seeing the above Theorems, Dr. E. A. Catchpole in a private communication and later (Catchpole, 1980) confirmed that in the general case (i.e. $\underset{\sim}{q}$ not necessarily equal to $q\underset{\sim}{1}$) at most one eigenvalue of $\underset{\sim}{q}^{-\delta}\underset{\sim}{W}$ could take the value -1, and derived a necessary and sufficient condition for the series formula for $\underset{\sim}{\Xi}$ to fail to converge. This condition is that the treatments fall into two groups, such that no two treatments from the same group ever occur together in the same block. In other words each block must contain only two different treatments, one from each group.

(There may also be an arbitrary disconnected part of the design.) Catchpole further showed that the series can be made to always converge by using

$$c\underset{\sim}{q} \text{ in place of } \underset{\sim}{q},$$

and

$$\underset{\sim}{W} + (c - 1)\underset{\sim}{q}^{\delta} \text{ in place of } \underset{\sim}{W},$$

for some constant c such that $c > 1$.

Catchpole also showed that $c = 3/2$ was an optimal choice. Obviously with $c = 1$ the formula is as given in (2.6.25).

The algorithm was therefore used to search for the designs where the criteria based on $\underset{\sim}{\Xi}_H$ and $\underset{\sim}{\Omega}_H$ differed in the tests described in the previous section, but this time the series for $\underset{\sim}{\Xi}_H$ was modified as above. These designs were for the sets (6, 2, 3), (7, 2, 4), (8, 2, 4), (9, 2, 4) and (10, 2, 4). The algorithm with $H = 4$ gave the same designs and attempted the same number of interchanges as when the $\underset{\sim}{\Omega}_H$ approximation with $H = 4$ was used.

The major difference between the two approximations $\underset{\sim}{L}'\underset{\sim}{\Xi}_H\underset{\sim\sim}{L\omega}^{\delta}$ and $\underset{\sim}{L}'\underset{\sim}{\Omega}_H\underset{\sim\sim}{L\omega}^{\delta}$ seems therefore to have been eliminated. However, both approximations can lead to a local optimum. Further, both are only approximations, all be it good ones. It might be possible to obtain better approximations but really the objective should be to derive an algorithm which can work directly with the criterion $\underset{\sim}{L}'\underset{\sim\sim}{\Omega L\omega}^{\delta} = \underset{\sim}{L}'\underset{\sim\sim}{\Xi L\omega}^{\delta}$ without resorting to an approximation. This is done in Chapter 4. However, the performance of the algorithm using the approximations was good enough to warrant considering if it could be extended to search for optimal row-and-column designs. This is done in the next chapter.

Chapter 3. Construction of optimal row-and-column designs

3.1 Introduction

In many experiments the plots form a rectangular grid, for example, agricultural plots on a rectangular field or potted plants on a shelf. In this situation the experimenter may wish to use two blocking systems, one based on the rows of the grid and the other on its columns, to eliminate heterogeneity of plots in both directions. A formally similar situation arises when a block experiment has been concluded, each treatment having been applied once in each block, and a further set of treatments is applied, the assumption being made that the effects of the second set and the residual effects of the first do not interact.

In Chapter 2 an algorithm that searched for optimal connected block designs was described. In this chapter it is shown that this algorithm can be modified to search for the component designs of an optimal row-and-column design. A row-and-column design is a design for v treatments in b_1 rows and b_2 columns. The rows of the design can be considered as a block design with b_1 blocks of size b_2 and is known as the row component, and the columns can be considered as a block design with b_2 blocks of size b_1 and is known as the column component. The optimality criterion used is similar to that given in the previous chapter.

The algorithm however, only searches for the component designs and these have to be combined to form a row-and-column design. By "combining" the components is meant the arranging of the v treatments as a $b_1 \times b_2$ grid in such a way that the rows of the grid are equal to the row component and the columns of the grid are equal to the column component. Three

alternative algorithms for combining the components are described and compared.

To simplify the presentation of the material considered in this chapter the problem of searching for the optimal components and the problem of deriving ways of combining the components are considered separately. Sections 3.2 to 3.7 are concerned with the development of an algorithm to search for optimal components and sections 3.8 to 3.11 are concerned with the development of an algorithm to combine the components. The remainder of the chapter is concerned with further evaluations of the performance of the algorithm to search for optimal components and an improvement to the algorithm is described.

3.2 Optimality criteria for row-and-column designs

The observations from a row-and-column design for v treatments in b_1 rows and b_2 columns may be formally expressed as

$$\underset{\sim}{y} = \underset{\sim}{1}\alpha + \underset{\sim}{D}_1'\underset{\sim}{\beta}_1 + \underset{\sim}{D}_2'\underset{\sim}{\beta}_2 + \underset{\sim}{\Delta}'\underset{\sim}{\gamma} + \underset{\sim}{\eta}, \qquad (3.2.1)$$

where $\underset{\sim}{y}$ is a column vector of $n = b_1 \times b_2$ observations, α a general parameter, $\underset{\sim}{1}$ a vector with unit elements, $\underset{\sim}{\beta}_1$ and $\underset{\sim}{\beta}_2$ are vectors respectively, of the parameters of the two blocking systems and $\underset{\sim}{\gamma}$ is a vector of treatment parameters. The design matrices for the two blocking systems and the allocation of treatments are $\underset{\sim}{D}_1$, $\underset{\sim}{D}_2$ and $\underset{\sim}{\Delta}$, respectively. The vector $\underset{\sim}{\eta}$ contains the n uncorrelated residuals which are distributed about zero with variance σ^2.

Minimizing $\underset{\sim}{\eta}'\underset{\sim}{\eta}$ in (3.2.1) and assuming that $\underset{\sim}{\beta}_1'\underset{\sim}{1} = \underset{\sim}{\beta}_2'\underset{\sim}{1} = 0$, the reduced normal equations for the treatment parameters $\underset{\sim}{\gamma}$ are

$$\underset{\sim}{Q} = \underset{\sim\sim}{C}\hat{\underset{\sim}{\gamma}}, \qquad (3.2.2)$$

where

$$\underset{\sim}{Q} = \underset{\sim}{T} - \underset{\sim}{N}_1\underset{\sim}{B}_1/b_2 - \underset{\sim}{N}_2\underset{\sim}{B}_2/b_1 + \underset{\sim\sim}{r}\underset{\sim}{1}'\underset{\sim}{y}/n. \qquad (3.2.3)$$

$$\underset{\sim}{C} = \underset{\sim}{r}^{\delta} - \underset{\sim}{N}_1\underset{\sim}{N}_1'/b_2 - \underset{\sim}{N}_2\underset{\sim}{N}_2'/b_1 + \underset{\sim}{r}\underset{\sim}{r}'/n, \qquad (3.2.4)$$

$\underset{\sim}{r}$ denotes the vector of treatment replications, $\underset{\sim}{r}^{\delta}$ is the vector $\underset{\sim}{r}$ written as a diagonal matrix, $\underset{\sim}{N}_1$ is the $v \times b_1$ incidence matrix for rows, $\underset{\sim}{N}_2$ is the $v \times b_2$ incidence matrix for columns, $\underset{\sim}{T}$ is the vector of treatment totals, $\underset{\sim}{B}_1$ is the vector of row totals and $\underset{\sim}{B}_2$ is the vector of column totals.

The matrix $\underset{\sim}{C}$ however, is singular and so equation (3.2.2) cannot be solved directly.

A unique solution to equation (3.2.2) can be obtained however, by applying the constraint $\underset{\sim}{r}'\underset{\sim}{\gamma} = 0$. Then

$$\underset{\sim}{Q} = \underset{\sim}{\Omega}^{-1}\hat{\underset{\sim}{\gamma}}, \qquad (3.2.5)$$

where

$$\underset{\sim}{\Omega}^{-1} = \underset{\sim}{C} + \underset{\sim}{r}\underset{\sim}{r}'/n$$

and

$$\hat{\underset{\sim}{\gamma}} = \underset{\sim}{\Omega}\underset{\sim}{Q}. \qquad (3.2.6)$$

The variance-covariance matrix of $\hat{\underset{\sim}{\gamma}}$, for the purposes of calculating the variances of unconfounded contrasts, is as Pearce (1975b) noted

$$V(\hat{\underset{\sim}{\gamma}}) = \underset{\sim}{\Omega}\sigma^2. \qquad (3.2.7)$$

If $\underset{\sim}{L}$ is a set of treatment contrasts, held columnwise then

$$V(\underset{\sim}{L}'\hat{\underset{\sim}{\gamma}}) = \underset{\sim}{L}'\underset{\sim}{\Omega}\underset{\sim}{L}\sigma^2. \qquad (3.2.8)$$

As was the case with block designs the matrix $\underset{\sim}{C}$ played an important part in the early suggestions for optimality criteria for row-and-column designs. Before mentioning these however, it should be noted that the approximate and continuous theory for the general linear model described in Section 2.2 of Chapter 2 can be used to search for optimal designs. Although compared to the search for optimal block designs the row-and-column design has received very little attention.

Wald (1943) and Nandi (1950) showed that the Latin Square design was D-optimal and Ehrenfeld (1955) showed it

was E-optimal. Masuyama and Okuno (1957) showed that Latin Squares and Youden Squares were optimal in the sense that they minimized trace ($\underset{\sim}{C}^2$) subject to trace ($\underset{\sim}{C}$) being constant. Kshirsagar (1958) showed that the Youden Square was optimal in the sense that it maximized the harmonic mean and geometric mean of the non-zero eigenvalues of $\underset{\sim}{C}$. Kiefer (1958) also showed that Latin Squares were D- and E-optimal and also considered the D- and E-optimality of what he termed Generalized Youden Designs (GYD's). A GYD is a row-and-column design such that the row component is a balanced block design (BBD) and the column component is a BBD. In a BBD all replications are equal to r, all block sizes are equal to k, all off-diagonal elements of $\underset{\sim\sim}{NN}'$, where $\underset{\sim}{N}$ is the incidence matrix of the block design, are equal, and $|N_{ij} - k/r| \leq 1$ for all i and j, where N_{ij} is $(i,j)^{th}$ element of $\underset{\sim}{N}$. Kiefer (1975) considered the optimality of GYD's further and indicated ways of constructing them. Kurotchka and Dwyer (1974) considered the G-, D- and A-optimality of three-way layouts without interactions and Shah, Raghavarao and Khatri (1976) considered the A-, D- and E-optimality of three factor designs.

Therefore, before the work described in this chapter was begun there was no algorithm to search for general row-and-column designs.

As done in the previous chapter it will be assumed that the experimenter is interested in treatment contrasts held column-wise in a matrix $\underset{\sim}{L}$, which have been assigned weights held in a diagonal matrix $\underset{\sim}{\omega}^{\delta}$. For the rest of this chapter the optimal row-and-column design will be considered to be the one that minimizes

$$T = \text{trace } (\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}\underset{\sim}{\omega}^{\delta}). \qquad (3.2.9)$$

However, if this criterion is to be used in a computerised

algorithm it is desirable to avoid the direct inversion of $\underset{\sim}{\Omega}^{-1}$. Therefore in the next section an iterative formula for $\underset{\sim}{\Omega}$ is derived.

3.3 An iterative formula for $\underset{\sim}{\Omega}$

Following the methods of Pearce, Caliński and Marshall (1974), as done in Chapter 2, an iterative formula for $\underset{\sim}{\Omega}$ can be obtained as follows.

Equation (3.2.2) can be re-written as

$$\underset{\sim}{Q} = \underset{\sim}{r}^{\frac{1}{2}\delta}\underset{\sim}{F}\underset{\sim}{r}^{\frac{1}{2}\delta}\hat{\underset{\sim}{\gamma}}, \tag{3.3.1}$$

where

$$\underset{\sim}{F} = \underset{\sim}{I} - \underset{\sim}{r}^{-\frac{1}{2}\delta}\frac{\underset{\sim}{N}_1\underset{\sim}{N}_1'}{b_2}\underset{\sim}{r}^{-\frac{1}{2}\delta} - \underset{\sim}{r}^{-\frac{1}{2}\delta}\frac{\underset{\sim}{N}_2\underset{\sim}{N}_2'}{b_1}\underset{\sim}{r}^{-\frac{1}{2}\delta} + \underset{\sim}{r}^{\frac{1}{2}\delta}\underset{\sim}{1}\underset{\sim}{1}'\underset{\sim}{r}^{\frac{1}{2}\delta}/n, \tag{3.3.2}$$

$\underset{\sim}{r}^{\delta} = \underset{\sim}{r}^{\frac{1}{2}\delta}\underset{\sim}{r}^{\frac{1}{2}\delta}$, and $\underset{\sim}{r}^{-\frac{1}{2}\delta}$ is the inverse of $\underset{\sim}{r}^{\frac{1}{2}\delta}$.

However, (3.3.1) does not provide a unique solution for $\hat{\underset{\sim}{\gamma}}$ because $\underset{\sim}{F}$ is singular as can be seen from the relation

$$\underset{\sim}{F}(\underset{\sim}{r}^{\frac{1}{2}\delta}\underset{\sim}{1}) = \underset{\sim}{r}^{\frac{1}{2}\delta}\underset{\sim}{1} - \underset{\sim}{r}^{-\frac{1}{2}\delta}\frac{\underset{\sim}{N}_1\underset{\sim}{N}_1'\underset{\sim}{1}}{b_2} - \underset{\sim}{r}^{-\frac{1}{2}\delta}\frac{\underset{\sim}{N}_2\underset{\sim}{N}_2'\underset{\sim}{1}}{b_1} + \underset{\sim}{r}^{\frac{1}{2}\delta}\underset{\sim}{1}\underset{\sim}{1}'\underset{\sim}{r}/n$$

$$= \underset{\sim}{r}^{\frac{1}{2}\delta}\underset{\sim}{1} - \underset{\sim}{r}^{-\frac{1}{2}\delta}\underset{\sim}{r} - \underset{\sim}{r}^{-\frac{1}{2}\delta}\underset{\sim}{r} + \underset{\sim}{r}^{\frac{1}{2}\delta}\underset{\sim}{1} = \underset{\sim}{0}, \tag{3.3.3}$$

recalling that $\underset{\sim}{N}_1'\underset{\sim}{1} = \underset{\sim}{1}b_2$, $\underset{\sim}{N}_2'\underset{\sim}{1} = \underset{\sim}{1}b_1$, and $\underset{\sim}{N}_1\underset{\sim}{1} = \underset{\sim}{r} = \underset{\sim}{N}_2\underset{\sim}{1}$.

However, since $\underset{\sim}{F}$ is symmetric it has a complete set of orthogonal eigenvectors, though they will not be uniquely determined if there are multiplicities among the eigenvalues. It will be convenient to order the eigenvectors so that those with zero eigenvalues come at the end, the eigenvector given by (3.3.3) coming last of all. It is now possible to write

$$\underset{\sim}{F} = \sum_{i=1}^{h} \epsilon_i \underset{\sim}{p}_i\underset{\sim}{p}_i', \tag{3.3.4}$$

where ϵ_i is the eigenvalue corresponding to the eigenvector $\underset{\sim}{p}_i$, and $h < v$. It will also be convenient to order so that any eigenvectors with unit eigenvalues, should any exist,

g in number, where $0 \leq g \leq h$, come at the beginning.

In exactly the same way as done in Section 2.6 of Chapter 2, matrices $\underset{\sim}{X}$, $\underset{\sim}{Y}$, $\underset{\sim}{Y}^{+}$ and $\underset{\sim}{Z}$ may be defined. Then (3.3.4) may be written as

$$\underset{\sim}{F} = \underset{\sim}{X} + \underset{\sim}{Y}, \tag{3.3.5}$$

and

$$\underset{\sim}{Q} = \underset{\sim}{r}^{\frac{1}{2}\delta}(\underset{\sim}{X} + \underset{\sim}{Y})\underset{\sim}{r}^{\frac{1}{2}\delta}\hat{\underset{\sim}{\gamma}}. \tag{3.3.6}$$

If now $\underset{\sim}{\Omega}^{-1} = \underset{\sim}{r}^{\frac{1}{2}\delta}(\underset{\sim}{X} + \underset{\sim}{Y} + \underset{\sim}{Z})\underset{\sim}{r}^{\frac{1}{2}\delta}$, then $\underset{\sim}{\Omega}^{-1}$ is a generalization of the matrix proposed by Pearce (1975b) and equals it when there are no zero eigenvalues apart from that in (3.3.3). Further,

$$\underset{\sim}{\Omega} = \underset{\sim}{r}^{-\frac{1}{2}\delta}(\underset{\sim}{X} + \underset{\sim}{Y}^{+} + \underset{\sim}{Z})\underset{\sim}{r}^{-\frac{1}{2}\delta}. \tag{3.3.7}$$

The importance of $\underset{\sim}{\Omega}$ is that it is a generalized inverse of $\underset{\sim}{r}^{\frac{1}{2}\delta}\underset{\sim\sim}{F}\underset{\sim}{r}^{\frac{1}{2}\delta}$ and so although it is not possible to estimate $\underset{\sim}{\gamma}$ it is possible to estimate

$$\begin{aligned}\tilde{\underset{\sim}{\gamma}} &= \underset{\sim\sim}{\Omega}\underset{\sim}{r}^{\frac{1}{2}\delta}\underset{\sim\sim}{F}\underset{\sim}{r}^{\frac{1}{2}\delta}\underset{\sim}{\gamma} \\ &= \underset{\sim}{\gamma} - \underset{\sim}{r}^{-\frac{1}{2}\delta}\underset{\sim\sim}{Z}\underset{\sim}{r}^{\frac{1}{2}\delta}\underset{\sim}{\gamma}.\end{aligned} \tag{3.3.8}$$

To estimate $\underset{\sim}{\gamma}$ constraints may be imposed as follows. Put

$$a_i = \underset{\sim}{p}_i'\underset{\sim}{r}^{\frac{1}{2}\delta}\underset{\sim}{\gamma} \tag{3.3.9}$$

and it is required that

$$a_i = 0 \qquad (i > h), \tag{3.3.10}$$

which may be written alternatively as

$$\underset{\sim\sim}{Z}\underset{\sim}{r}^{\frac{1}{2}\delta}\underset{\sim}{\gamma} = 0. \tag{3.3.11}$$

One such constraint arises from (3.3.3),

$$\underset{\sim}{p}_v = \underset{\sim}{r}^{\frac{1}{2}\delta}\underset{\sim}{1}/n^{\frac{1}{2}}, \tag{3.3.12}$$

hence

$$a_v = \underset{\sim}{1}'\underset{\sim}{r}^{\delta}\underset{\sim}{\gamma}/n^{\frac{1}{2}} = 0,$$

i.e.

$$\underset{\sim}{r}'\underset{\sim}{\gamma} = 0.$$

From (3.3.6), because $\underset{\sim}{\Omega}$ is a generalized inverse of $\underset{\sim}{r}^{\frac{1}{2}\delta}\underset{\sim\sim}{F}\underset{\sim}{r}^{\frac{1}{2}\delta}$, it is permissible to write

$$\hat{\underset{\sim}{\gamma}} = \underset{\sim\sim}{\Omega}\underset{\sim}{Q}, \tag{3.3.13}$$

and it can be shown that the covariance matrix of $\hat{\underset{\sim}{\gamma}}$ may be written as

$$\mathrm{Cov}(\hat{\underset{\sim}{\gamma}}) = (\underset{\sim}{\Omega} - \underset{\sim}{r}^{-\frac{1}{2}\delta}\underset{\sim\sim}{Z}\underset{\sim}{r}^{-\frac{1}{2}\delta})\sigma^2.$$

An iterative formula for $\underset{\sim}{\Omega}$ may be obtained as follows. Let

$$\begin{aligned}\underset{\sim}{t}_1 &= \underset{\sim}{r}^{-\frac{1}{2}\delta}(\underset{\sim}{I} - \underset{\sim}{Z})\underset{\sim}{r}^{-\frac{1}{2}\delta} \\ &= \underset{\sim}{r}^{-\frac{1}{2}\delta}(\underset{\sim}{X} + \underset{\sim\sim}{YY}^{+})\underset{\sim}{r}^{-\frac{1}{2}\delta}\end{aligned} \tag{3.3.14}$$

and

$$\underset{\sim}{t}_j = \underset{\sim}{r}^{-\frac{1}{2}\delta}(\underset{\sim}{Y}_o + \underset{\sim}{Z})\underset{\sim}{r}^{\frac{1}{2}\delta}\underset{\sim}{t}_{j-1} \qquad j \geq 2 \tag{3.3.15}$$

where

$$\underset{\sim}{Y}_o = \underset{\sim\sim}{YY}^{+} - \underset{\sim}{Y} = \sum_{i=g+1}^{h}(1 - \epsilon_i)\underset{\sim}{p}_i\underset{\sim}{p}_i'. \tag{3.3.16}$$

It may be noted that

$$\underset{\sim}{r}^{-\frac{1}{2}\delta}(\underset{\sim}{Y}_o + \underset{\sim}{Z})\underset{\sim}{r}^{\frac{1}{2}\delta} = \underset{\sim}{r}^{-\delta}\underset{\sim}{N}_1\underset{\sim}{N}_1'/b_2 + \underset{\sim}{r}^{-\delta}\underset{\sim}{N}_2\underset{\sim}{N}_2'/b_1 - \underset{\sim\sim}{1r'}/n,$$

and that $\underset{\sim}{t}_1$ and $\underset{\sim}{t}_j$ have been defined in an analogous manner to that of Section 2.6.

Then

$$\begin{aligned}\underset{\sim}{t}_2 &= \underset{\sim}{r}^{-\frac{1}{2}\delta}(\underset{\sim}{Y}_o + \underset{\sim}{Z})\underset{\sim}{r}^{\frac{1}{2}\delta}\underset{\sim}{r}^{-\frac{1}{2}\delta}(\underset{\sim}{X} + \underset{\sim\sim}{YY}^{+})\underset{\sim}{r}^{-\frac{1}{2}\delta} \\ &= \underset{\sim}{r}^{-\frac{1}{2}\delta}(\underset{\sim}{Y}_o\underset{\sim\sim}{YY}^{+})\underset{\sim}{r}^{-\frac{1}{2}\delta} \\ &= \underset{\sim}{r}^{-\frac{1}{2}\delta}\underset{\sim}{Y}_o\underset{\sim}{r}^{-\frac{1}{2}\delta}.\end{aligned}$$

Therefore, if it is assumed that

$$\underset{\sim}{t}_j = \underset{\sim}{r}^{-\frac{1}{2}\delta}\underset{\sim}{Y}_o^{j-1}\underset{\sim}{r}^{-\frac{1}{2}\delta},$$

then

$$\begin{aligned}\underset{\sim}{t}_{j+1} &= \underset{\sim}{r}^{-\frac{1}{2}\delta}(\underset{\sim}{Y}_o + \underset{\sim}{Z})\underset{\sim}{r}^{\frac{1}{2}\delta}\underset{\sim}{r}^{-\frac{1}{2}\delta}\underset{\sim}{Y}_o^{j-1}\underset{\sim}{r}^{-\frac{1}{2}\delta} \\ &= \underset{\sim}{r}^{-\frac{1}{2}\delta}\underset{\sim}{Y}_o^{j}\underset{\sim}{r}^{-\frac{1}{2}\delta}.\end{aligned}$$

Therefore, by induction, it must be true that

$$\underset{\sim}{t}_j = \underset{\sim}{r}^{-\frac{1}{2}\delta} \underset{\sim}{Y}_o^{j-1} \underset{\sim}{r}^{-\frac{1}{2}\delta} \quad \text{for all } j \geq 2.$$

Writing

$$\underset{\sim}{Y}\underset{\sim}{Y}^{+} = \underset{\sim}{Y}_o^{0} = \sum_{i=g+1}^{h} \underset{\sim}{p}_i \underset{\sim}{p}_i',$$

and

$$\underset{\sim}{t}_1 = \underset{\sim}{r}^{-\frac{1}{2}\delta}(\underset{\sim}{X} + \underset{\sim}{Y}_o^{0})\underset{\sim}{r}^{-\frac{1}{2}\delta},$$

then

$$\sum_{j=1}^{\infty} \underset{\sim}{t}_j = \underset{\sim}{r}^{-\frac{1}{2}\delta}(\underset{\sim}{X} + \sum_{j=1}^{\infty} \underset{\sim}{Y}_o^{j-1})\underset{\sim}{r}^{-\frac{1}{2}\delta}.$$

But

$$\sum_{j=1}^{\infty} \underset{\sim}{Y}_o^{j-1} = \sum_{i=g+1}^{h} \sum_{j=1}^{\infty} \{(1 - \epsilon_i)^{j-1} \underset{\sim}{p}_i \underset{\sim}{p}_i'\}$$

$$= \sum_{i=g+1}^{h} \underset{\sim}{p}_i \underset{\sim}{p}_i' / \epsilon_i = \underset{\sim}{Y}^{+},$$

if $|1 - \epsilon_i| < 1$ for all i, $g + 1 \leq i \leq h$.

It will be shown that $|1 - \epsilon_i| < 1$ for all $g + 1 \leq i \leq h$,

therefore,

$$\sum_{j=1}^{\infty} \underset{\sim}{t}_j = \underset{\sim}{r}^{-\frac{1}{2}\delta}(\underset{\sim}{X} + \underset{\sim}{Y}^{+})\underset{\sim}{r}^{-\frac{1}{2}\delta},$$

i.e.

$$\sum_{j=1}^{\infty} \underset{\sim}{t}_j = \underset{\sim}{\Omega} - \underset{\sim}{r}^{-\frac{1}{2}\delta} \underset{\sim}{Z} \underset{\sim}{r}^{-\frac{1}{2}\delta}. \tag{3.3.17}$$

To see that $|1 - \epsilon_i| < 1$ for all $g + 1 \leq i \leq h$ consider

$$\underset{\sim}{s}_i = \underset{\sim}{r}^{\frac{1}{2}\delta} \underset{\sim}{p}_i. \tag{3.3.18}$$

Then if $\underset{\sim}{S}$ is the matrix such that the i^{th} column is $\underset{\sim}{s}_i$ and $\underset{\sim}{P}$ is the matrix such that the i^{th} column is $\underset{\sim}{p}_i$

$$\underset{\sim}{S}'\underset{\sim}{r}^{-\delta}\underset{\sim}{S} = \underset{\sim}{P}'\underset{\sim}{P} = \underset{\sim}{I}$$

Since from (3.3.12), $\underset{\sim}{p}_v = \underset{\sim}{r}^{\frac{1}{2}\delta}\underset{\sim}{1}/n^{\frac{1}{2}}$ we have that

$$\underset{\sim}{s}_i'\underset{\sim}{1} = 0 \quad (i \neq v)$$

For this reason $\underset{\sim}{s}_i$ $(i \neq v)$ are contrasts and the estimates of these contrasts are given by $\underset{\sim}{S}'\hat{\underset{\sim}{\gamma}}$ which has variance-covariance matrix

$$\underset{\sim}{S}'(\underset{\sim}{\Omega} - \underset{\sim}{r}^{-\frac{1}{2}\delta}\underset{\sim\sim}{Z}\underset{\sim}{r}^{-\frac{1}{2}\delta})\underset{\sim}{S} = \underset{\sim}{P}'(\underset{\sim}{X} + \underset{\sim}{Y}^{+})\underset{\sim}{P}\sigma^2 = \underset{\sim}{\epsilon}_*^{-\delta}\sigma^2 \qquad (3.3.19)$$

Here $\underset{\sim}{\epsilon}_*^{-\delta}$ is a diagonal matrix with elements $\frac{1}{\epsilon_i}$ for $1 \le i \le h$.

If the design had been orthogonal (3.3.19) would have been $\underset{\sim}{S}'\underset{\sim}{r}^{-\delta}\underset{\sim}{S}\sigma^2 = \underset{\sim}{I}\sigma^2$ and so ϵ_i, $1 \le i \le h$ are efficiency factors, i.e. $0 < \epsilon_i \le 1$ for $1 \le i \le h$. Therefore,

$$|1 - \epsilon_i| < 1 \text{ for } g + 1 \le i \le h.$$

Provided no effects are totally confounded in the design $\underset{\sim}{\Omega}$ can be found as follows.

Set

$$\underset{\sim}{t}_1 = \underset{\sim}{r}^{-\frac{1}{2}\delta}(\underset{\sim}{I} - \underset{\sim}{Z})\underset{\sim}{r}^{-\frac{1}{2}\delta}$$

$$= \underset{\sim}{r}^{-\frac{1}{2}\delta}(\underset{\sim}{I} - \underset{\sim}{r}^{\frac{1}{2}\delta}\frac{\underset{\sim\sim}{1}\underset{\sim}{1}'\underset{\sim}{r}^{\frac{1}{2}\delta}}{n})\underset{\sim}{r}^{-\frac{1}{2}\delta}$$

$$= \underset{\sim}{r}^{-\delta} - \underset{\sim\sim}{1}\underset{\sim}{1}'/n$$

and

$$\underset{\sim}{t}_j = [\underset{\sim}{r}^{-\delta}\underset{\sim}{N}_1\underset{\sim}{N}_1'/b_2 + \underset{\sim}{r}^{-\delta}\underset{\sim}{N}_2\underset{\sim}{N}_2'/b_1 - \underset{\sim\sim}{1}\underset{\sim}{r}'/n]\underset{\sim}{t}_{j-1}, \quad j \ge 2.$$

Using equation (3.3.17)

$$\underset{\sim}{\Omega} = \sum_{j=1}^{\infty}\underset{\sim}{t}_j + \underset{\sim\sim}{1}\underset{\sim}{1}'/n$$

$$= \underset{\sim}{r}^{-\delta} + \sum_{j=1}^{\infty}[\underset{\sim}{r}^{-\delta}\underset{\sim}{N}_1\underset{\sim}{N}_1'/b_2 + \underset{\sim}{r}^{-\delta}\underset{\sim}{N}_2\underset{\sim}{N}_2'/b_1 - \underset{\sim\sim}{1}\underset{\sim}{r}'/n]^j$$

$$\times [\underset{\sim}{r}^{-\delta} - \underset{\sim\sim}{1}\underset{\sim}{1}'/n].$$

However,

$$[\underset{\sim}{r}^{-\delta}\underset{\sim}{N}_1\underset{\sim}{N}_1'/b_2 + \underset{\sim}{r}^{-\delta}\underset{\sim}{N}_2\underset{\sim}{N}_2'/b_1 - \underset{\sim\sim}{1}\underset{\sim}{r}'/n][\underset{\sim\sim}{1}\underset{\sim}{1}'/n]$$

$$= \underset{\sim\sim}{1}\underset{\sim}{1}'/n + \underset{\sim\sim}{1}\underset{\sim}{1}'/n - \underset{\sim\sim}{1}\underset{\sim}{1}'/n = \underset{\sim\sim}{1}\underset{\sim}{1}'/n.$$

Therefore,

$$\underset{\sim}{\Omega} = \underset{\sim}{r}^{-\delta} - \sum_{h=1}^{\infty}[\underset{\sim\sim}{1}\underset{\sim}{1}'/n - (\underset{\sim}{r}^{-\delta}\underset{\sim}{N}_1\underset{\sim}{N}_1'/b_2 + \underset{\sim}{r}^{-\delta}\underset{\sim}{N}_2\underset{\sim}{N}_2'/b_1 - \underset{\sim\sim}{1}\underset{\sim}{r}'/n)^h\underset{\sim}{r}^{-\delta}].$$

(3.3.20)

This formula enables $\underset{\sim}{\Omega}$ to be approximated by

$$\underset{\sim}{\Omega}_H = \underset{\sim}{r}^{-\delta} - \sum_{h=1}^{H} [\underset{\sim\sim}{11}'/n - (\underset{\sim}{r}^{-\delta}\underset{\sim}{N}_1\underset{\sim}{N}_1'/b_2 + \underset{\sim}{r}^{-\delta}\underset{\sim}{N}_2\underset{\sim}{N}_2'/b_1 - \underset{\sim\sim}{1r}'/n)\underset{\sim}{r}^{h-\delta}],$$

(3.3.21)

for some finite positive integer H.

Therefore, the criterion (3.2.9) can be approximated by

$$T = \text{trace } (\underset{\sim\sim}{L}\omega^{\delta}\underset{\sim}{L}'\underset{\sim}{\Omega}_H), \qquad (3.3.22)$$

an analagous approximation to that used in Section 2.4 of Chapter 2.

As in Chapter 2 it will be seen that H does not need to be large.

3.4 An algorithm to search for the components of an optimal row-and-column design

The algorithm described in Section 2.5 of Chapter 2 searched for optimal block designs. This algorithm can be modified to search for the component designs of an optimal row-and-column design if the amalgamated design is considered. The amalgamated design consists of the two component designs taken together as a block design. The treatment replications are therefore doubled and blocks may be of two sizes. Pearce (1975b), for example, has noted the usefulness of the amalgamated design. Further, it is important to note that given the incidence matrices $\underset{\sim}{N}_1$ and $\underset{\sim}{N}_2$ for the row component and the column component, respectively, the matrix $\underset{\sim}{\Omega}$ can be obtained even though the components have not been combined to give a row-and-column design. In fact $\underset{\sim}{\Omega}$ can be obtained even though it may be impossible to combine the components.

Further Caliński (1971) considered the contrasts that satisfy the following matrix equation

$$\underset{\sim}{\Omega}^{-1}\underset{\sim}{r}^{-\delta}\underset{\sim}{c} = \varepsilon\underset{\sim}{c}$$

as worthy of special interest and called them basic contrasts.

Here Ω^{-1} is as defined in equation (2.3.8) of Chapter 2, and ε is the efficiency factor of the contrast.

Pearce (1975b) showed that if the efficiency factor of a basic contrast was ε_1 in the row component and ε_2 in the column component, then its efficiency factor in the row-and-column design is $\varepsilon_1 + \varepsilon_2 - 1$. That is, if $\varepsilon_1 + \varepsilon_2 = 1$ that contrast is totally confounded in the row-and-column design. Also if $\varepsilon_i = 0$, $(i \neq j)$ then $\varepsilon_j = 1$ and again $\varepsilon_1 + \varepsilon_2 - 1 = 0$.

The components found by the algorithm described below may therefore be such that there is a complete loss of information on a contrast of interest. However, by ensuring that each component forms a connected block design, as defined in Section 2.4 of Chapter 2 the extreme cases of $\varepsilon_1 = 0$ or $\varepsilon_2 = 0$ will be avoided.

The steps in the algorithm are as follows.

Step 1. Construct the initial row component and column component

Using the procedure described in Section 2.5 determine the optimal treatment replications and construct a connected starting design for the row component. Using the same treatment replications construct a starting design for the column component.

As in the algorithm described in Section 2.4 the treatment replications, once chosen, remain fixed throughout the remainder of the algorithm.

Step 2. Evaluation of the starting components

Using the components found in Step 1 calculate

$$T = \text{trace}(\underset{\sim}{L}\omega^{\delta}\underset{\sim}{L}'\underset{\sim}{\Omega}_H),$$

for some predetermined value of H, using equation (3.3.21).

Step 3. Order the blocks on their degree of non-orthogonality

Using the two component designs found in Step 1 form the amalgamated design, i.e. form a block design consisting of the b_1 blocks of the row component and the b_2 blocks of the column component. Order these blocks on their degree of non-orthogonality as done in Step 2 of the algorithm described in Section 2.4 of Chapter 2.

Step 4. Evaluate the effect of interchanging treatments between the blocks of the amalgamated design

In exactly the same way as done in Step 3 of the algorithm described in Section 2.4, evaluate the effect on the criterion T as defined in Step 1 above, by interchanging treatments between the blocks of the amalgamated design. However, ensuring that no interchange is made between a block of the row component and a block of the column component and that no interchange leads to a disconnected component.

Having found the interchange that leads to the largest reduction in T proceed to step 5.

Step 5. Update the design

Interchange the treatment pair to give a new component. Calculate T using the two components and return to Step 3.

In order to illustrate the algorithm consider the following example.

Example 3.4.1

A design is required for four treatments in six rows and six columns and there is an equal interest in all treatment differences.

Therefore,

$$\underset{\sim}{L}' = \begin{matrix} (\tfrac{1}{2})^{\frac{1}{2}} \\ (\tfrac{1}{6})^{\frac{1}{2}} \\ (\tfrac{1}{12})^{\frac{1}{2}} \end{matrix} \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 1 & -2 & 0 \\ 1 & 1 & 1 & -3 \end{bmatrix} \quad \text{and} \quad \underset{\sim}{\omega}^{\delta} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Using the procedure described in Section 2.5 the starting design for each component is found to be

Block						
I	2	2	3	3	4	4
II	1	1	1	2	3	4
III	2	2	3	3	4	4
IV	1	1	1	2	3	4
V	2	2	3	3	4	4
VI	1	1	1	2	3	4

which has a vector of treatment replications equal to

$$\underset{\sim}{r}' = (9, 9, 9, 9).$$

The starting design is the same for each component because they have the same block sizes.

The starting design for Step 3 of the algorithm therefore consists of the six blocks corresponding to the row component plus the six blocks corresponding to the column component.

The value of the criterion T, with H equal to 4, for this starting design is 0.5117.

Using this starting design the algorithm gave as optimal the amalgamated design in Table 3.4.1.

Table 3.4.1. Optimal amalgamated design

Block							
I	1	2	3	3	4	4	row component
II	2	1	1	2	3	4	
III	2	2	1	3	4	4	
IV	2	1	1	2	3	4	
V	1	2	3	3	4	4	
VI	1	3	1	2	3	4	
VII	1	2	3	3	2	4	column component
VIII	4	1	1	2	3	4	
IX	2	2	1	3	4	4	
X	2	3	1	2	3	4	
XI	1	2	1	3	4	4	
XII	1	3	1	2	3	4	

These components when combined give the following row-and-column design,

1 2 3 3 4 4
4 1 1 2 3 2
2 4 2 3 4 1
3 4 2 1 2 1
2 3 4 4 1 3
3 1 4 2 1 3,

for which

$$\underset{\sim}{\Omega} = \begin{bmatrix} 0.1178 & -0.0022 & -0.0022 & -0.0022 \\ -0.0022 & 0.1178 & -0.0022 & -0.0022 \\ -0.0022 & -0.0022 & 0.1178 & -0.0022 \\ -0.0022 & -0.0022 & -0.0022 & 0.1178 \end{bmatrix}.$$

In each row and column of this design there is a complete replicate of the four treatments 1, 2, 3 and 4 with two excesses. The excesses consist of the six possible pairs of the four treatments. It would have been neater if the six possible pairs of excesses had occurred in each component. However, although its components are not balanced, this design is totally balanced and has the same value of the criterion T as, for example, the totally balanced design

1 4 3 2 1 4
2 2 4 1 3 3
4 3 2 3 4 1
3 1 4 1 2 2
3 3 2 4 1 1
2 4 1 4 2 3

which is of type 0:TT in the notation of Pearce (1963). Therefore because a design of type 0:TT is an optimal row-and-column, so is the design obtained by the algorithm.

It should be noted that in this section and the next two the problem of how to combine the components is not considered. It will be assumed that the components can be combined somehow. The performance of the algorithm is considered in the next section.

3.5 The performance of the algorithm

To test the algorithm described in Section 3.4 the 37 designs given in Table 13.2 of Cochran and Cox (1957) for which $v \leq 11$ and $b_1 + b_2 \leq 22$ were used. These designs were all based on either incomplete Latin Squares or extended Youden Squares. For each design H was set equal to 4 and $\underset{\sim}{L}$ and $\underset{\sim}{\omega}^{\delta}$ were chosen to specify an equal interest in all treatment differences. The present algorithm failed to match three of the designs given by Cochran and Cox. These designs are given in Table 3.5.1, where T refers to the trace criterion with H equal to 4 and E is the efficiency factor of the design. This efficiency factor is defined as the ratio of (a) the variance of any treatment difference in a Latin Square design to (b) the average variance of all treatment differences for the given design.

Table 3.5.1. Designs where the present algorithm failed to match Cochran and Cox

v	b_1	b_2	Present Algorithm		Cochran and Cox	
			T	E	T	E
9	18	4	1.1867	0.8426	1.1851	0.8437
11	11	5	2.2820	0.8764	2.2727	0.8800
11	11	6	1.8208	0.9153	1.8182	0.9167

These tests reveal that the algorithm can get trapped at local optima. However, none of the local optima found were much worse than the designs given by Cochran and Cox. Increasing H to ten did not lead to any improvement.

In the next section a different approximation to the optimality criterion is defined in a way similar to that

used in Chapter 2. This is done in an attempt to obtain a more sensitive approximation to the true value of the criterion

3.6 A second approximation to the optimality criterion for row-and-column designs

In Chapter 2 when the optimality of block designs was considered the optimality criterion was expressed in terms of $\underset{\sim}{\Xi}_H$ rather than $\underset{\sim}{\Omega}_H$ in an attempt to obtain a more sensitive approximation. The reason for doing this was that it was thought that as $\underset{\sim}{\Xi}_H$ was constructed from the treatment concurrences, it would be more sensitive to treatment interchanges. Here a similar approach is adopted for row-and-column designs.

The normal equations for the estimation of the treatment parameters $\underset{\sim}{\gamma}$, for a row-and-column design were given in equation (3.2.2) as

$$\underset{\sim}{Q} = (\underset{\sim}{r}^{\delta} - \underset{\sim}{N}_1\underset{\sim}{N}_1'/b_2 - \underset{\sim}{N}_2\underset{\sim}{N}_2'/b_1 + \underset{\sim}{r}\underset{\sim}{r}'/n)\hat{\underset{\sim}{\gamma}}.$$

If $\underset{\sim}{r}^{\delta} - \underset{\sim}{N}_1\underset{\sim}{N}_1'/b_2 - \underset{\sim}{N}_2\underset{\sim}{N}_2'/b_1 + \underset{\sim}{r}\underset{\sim}{r}'/n$ is set equal to $\underset{\sim}{q}^{\delta} - \underset{\sim}{W}$, where $\underset{\sim}{q}^{\delta}$ is a diagonal matrix and $\underset{\sim}{W}$ is a matrix with zero diagonal elements then the normal equation for $\hat{\underset{\sim}{\gamma}}$, can be written as

$$\underset{\sim}{Q} = (\underset{\sim}{q}^{\delta} - \underset{\sim}{W})\hat{\underset{\sim}{\gamma}}, \qquad (3.6.1)$$

which is analogous to equation (2.6.1) in Chapter 2. However in (3.6.1) $\underset{\sim}{W}$ and $\underset{\sim}{q}^{\delta}$ do not depend solely on the treatment concurrences.

If $\underset{\sim}{q} = \underset{\sim}{q}^{\delta}\underset{\sim}{1}$ and $u = \underset{\sim}{q}'\underset{\sim}{1}$, then a matrix $\underset{\sim}{\Xi}^{-1}$ may be defined as

$$\underset{\sim}{\Xi}^{-1} = \underset{\sim}{q}^{\delta} - \underset{\sim}{W} + \underset{\sim}{q}\underset{\sim}{q}'/u. \qquad (3.6.2)$$

The matrix $\underset{\sim}{\Xi}^{-1}$ satisfies the relation

$$\underset{\sim}{\Xi}^{-1}\underset{\sim}{1} = \underset{\sim}{q} - \underset{\sim}{q} + \underset{\sim}{q} = \underset{\sim}{q}, \qquad (3.6.3)$$

because $\underset{\sim}{W}\underset{\sim}{1} = \underset{\sim}{q}$ as can be seen from

$$(\underset{\sim}{r}^{\delta} - \underset{\sim}{N}_1\underset{\sim}{N}_1'/b_2 - \underset{\sim}{N}_2\underset{\sim}{N}_2'/b_1 + \underset{\sim}{r}\underset{\sim}{r}'/n)\underset{\sim}{1} = \underset{\sim}{0},$$

i.e.

$$(\underset{\sim}{q}^{\delta} - \underset{\sim}{W})\underset{\sim}{1} = \underset{\sim}{0}.$$

Equation (3.6.1) can be written as

$$\underset{\sim}{Q} = \underset{\sim}{q}^{\frac{1}{2}\delta}\underset{\sim}{F}\underset{\sim}{q}^{\frac{1}{2}\delta}\underset{\sim}{\hat{\gamma}}$$

where

$$\underset{\sim}{F} = \underset{\sim}{I} - \underset{\sim}{q}^{-\frac{1}{2}\delta}\underset{\sim}{W}\underset{\sim}{q}^{-\frac{1}{2}\delta}.$$

The matrix $\underset{\sim}{F}$ is singular, as can be seen from the relation

$$\begin{aligned}\underset{\sim}{F}\underset{\sim}{q}^{\frac{1}{2}\delta}\underset{\sim}{1} &= \underset{\sim}{q}^{\frac{1}{2}\delta}\underset{\sim}{1} - \underset{\sim}{q}^{-\frac{1}{2}\delta}\underset{\sim}{W}\underset{\sim}{1} \\ &= \underset{\sim}{q}^{\frac{1}{2}\delta}\underset{\sim}{1} - \underset{\sim}{q}^{\frac{1}{2}\delta}\underset{\sim}{1} = \underset{\sim}{0}. \qquad (3.6.4)\end{aligned}$$

Having now defined matrices $\underset{\sim}{q}^{\delta}$, $\underset{\sim}{W}$ and $\underset{\sim}{F}$ in this way an iterative formula for $\underset{\sim}{\Xi}$ can be derived by following exactly the steps given in Section 2.6, where an iterative formula for the block design case was derived.

That is,

$$\underset{\sim}{\Xi} = \underset{\sim}{q}^{-\delta} - \sum_{h=1}^{\infty}(\underset{\sim}{1}\underset{\sim}{1}'/u - (\underset{\sim}{q}^{-\delta}\underset{\sim}{W})^h\,\underset{\sim}{q}^{-\delta}). \qquad (3.6.5)$$

This enables $\underset{\sim}{\Xi}$ to be approximated by

$$\underset{\sim}{\Xi}_H = \underset{\sim}{q}^{-\delta} - \sum_{h=1}^{H}(\underset{\sim}{1}\underset{\sim}{1}'/u - (\underset{\sim}{q}^{-\delta}\underset{\sim}{W})^h\underset{\sim}{q}^{-\delta}), \qquad (3.6.6)$$

for some finite positive integer H.

The criterion given in (3.2.9) can therefore be approximated by

$$T = \text{trace}(\underset{\sim}{L}'\underset{\sim}{\Xi}_H\underset{\sim}{L}\underset{\sim}{\omega}^{\delta}). \qquad (3.6.7)$$

It should be noted that the comments made in Section 2.6 concerning the iterative formula also apply here. That is, the constraint on the treatment parameters is $\underset{\sim}{q}'\underset{\sim}{\gamma} = 0$ and the series formula may not converge.

The formula (3.6.5) will fail to converge if any eigenvalue ϵ_i of $\underset{\sim}{F}$ fails to satisfy the condition $|1 - \epsilon_i| < 1$, $g + 1 \leq i \leq h$.

This condition will not hold, for example, for a Latin Square design for two treatments.

In this case

$$\underset{\sim}{C} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \qquad \underset{\sim}{W} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

$$\underset{\sim}{q}^{\delta} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \text{and } \underset{\sim}{F} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

One eigenvalue of $\underset{\sim}{F}$ is 1 and the other is -1.

However, this is the only design known to the author for which the series (3.6.5) fails to converge.

As the matrix $\underset{\sim}{\Xi}$ is a completely new generalized inverse for $\underset{\sim}{C}$ some examples of it are given in the next section. Also included in the next section is a comparison of $\underset{\sim}{\Omega}_H$ and $\underset{\sim}{\Xi}_H$, for small H, based on the examples used to illustrate $\underset{\sim}{\Xi}$.

3.7 Examples of the matrix $\underset{\sim}{\Xi}$

Example 3.7.1

A Youden Square design for seven treatments in seven rows and three columns.

7 1 3
1 2 4
2 3 5
3 4 6
4 5 7
5 6 1
6 7 2

For this design all diagonal elements of $\underset{\sim}{\Xi}$ are equal to 0.4388 and all off-diagonal elements are equal to 0.0102.

Example 3.7.2

Three treatments in six rows and five columns.

1 2 2 3 3
1 1 2 3 3
2 3 1 2 1
2 3 3 2 1
3 1 3 1 2
3 2 1 1 2

For this design all diagonal elements of $\underset{\sim}{\Xi}$ are equal to 0.1215 and all off-diagonal elements are equal to 0.0174.

For Example 3.7.1 the matrix $\underset{\sim}{\Omega}$ has all diagonal elements equal to 0.4150 and all off-diagonal elements equal to -0.0136. For Example 3.7.2 the matrix $\underset{\sim}{\Omega}$ has all diagonal elements equal to 0.1028 and all off-diagonal elements equal to -0.0014.

Of particular interest is whether $\underset{\sim H}{\Xi}$ and $\underset{\sim H}{\Omega}$, with H equal to 4, are good approximations to $\underset{\sim}{\Xi}$ and $\underset{\sim}{\Omega}$, respectively. For Example 3.7.1 all elements of $\underset{\sim}{\Xi}$ are equal to the elements of $\underset{\sim 4}{\Xi}$ to four decimal places, but the diagonal elements of $\underset{\sim}{\Omega}$ differ from the diagonal elements of $\underset{\sim 4}{\Omega}$ by 0.0002 to four decimal places, with all off-diagonal elements of $\underset{\sim}{\Omega}$ being equal to those of $\underset{\sim 4}{\Omega}$. So in this case $\underset{\sim 4}{\Xi}$ is better than $\underset{\sim 4}{\Omega}$.

However, for Example 3.7.2 the reverse is the case. Here all elements of $\underset{\sim}{\Omega}$ are equal to all elements of $\underset{\sim 4}{\Omega}$ to four decimal places, but the diagonal elements of $\underset{\sim}{\Xi}$ differ from those of $\underset{\sim 4}{\Xi}$ by 0.0022 and the off-diagonal elements of $\underset{\sim}{\Xi}$ differ from those of $\underset{\sim 4}{\Xi}$ by 0.0011.

It appears therefore, that for some designs $\underset{\sim}{L}'\underset{\sim 4}{\Xi}\underset{\sim}{L}$ may be a closer approximation to $\underset{\sim}{L}'\underset{\sim}{\Omega}\underset{\sim}{L}$ than $\underset{\sim}{L}'\underset{\sim 4}{\Omega}\underset{\sim}{L}$, but for other designs $\underset{\sim}{L}'\underset{\sim 4}{\Omega}\underset{\sim}{L}$ may be closer.

The relationship between $\underset{\sim H}{\Xi}$ and $\underset{\sim H}{\Omega}$ is not investigated further in this thesis, as the main aim is to obtain a successful algorithm to search for optimal row-and-column designs.

To check if using $\underset{\sim H}{\Xi}$ in place of $\underset{\sim H}{\Omega}$ in the algorithm described in Section 3.4 led to any improvement, the algorithm with $\underset{\sim H}{\Xi}$ was used to search for the 37 designs described in the previous section. The results of this

test are given below.

3.8 Performance of the algorithm using $\underset{\sim}{\Xi}_H$

The algorithm using $\underset{\sim}{\Xi}_4$ gave designs that were the same as those obtained when $\underset{\sim}{\Omega}_4$ was used. In particular, designs that were local optima were found for (9, 18, 4), (11, 11, 5) and (11, 11, 6) as before. Increasing H to ten did not lead to any improvement on these designs.

Therefore, for the rest of this chapter, only the algorithm based on $\underset{\sim}{\Omega}_H$ will be considered further. Additional tests of the performance of the algorithm are given in Sections 3.12 to 3.14. In the next three sections the problem of how to combine two component designs to give a row-and-column design will be considered.

3.9 An algorithm for combining two component designs

In the next three sections three different algorithms for combining two component designs are described. The reason three different algorithms are given is that they could be considered as of increasing complexity. The algorithms were developed in the order presented. To test each algorithm a large number of pairs of components were required and to obtain these the design problem given below was considered. This new problem was also introduced as a further test of the performance of the algorithm that searches for the components of an optimal design, the problem requiring contrasts to receive different weights.

The design problem is as follows. There are nine treatments to be arranged in six rows and six columns. The nine treatments are the nine treatment combinations corresponding to two factors each at three levels. The factors are labelled here as "letters", L and "numbers", N. The eight treatment contrasts were taken to be those

given in Table 3.9.1. For use in the algorithm that searches for the components and for calculating variances and covariances, each contrast was scaled so that the sum of squares of its coefficients was one.

Table 3.9.1. The eight treatment contrasts

Contrast No.	Treatment No.	1	2	3	4	5	6	7	8	9	Explanation
	Treatment Label	A_1	A_2	A_3	B_1	B_2	B_3	C_1	C_2	C_3	
1		1	1	1	-1	-1	-1	0	0	0	L_ℓ
2		1	1	1	1	1	1	-2	-2	-2	L_q
3		1	-1	0	1	-1	0	1	-1	0	N_ℓ
4		1	1	-2	1	1	-2	1	1	-2	N_q
5		1	-1	0	-1	1	0	0	0	0	$L_\ell \times N_\ell$
6		1	1	-2	-1	-1	2	0	0	0	$L_\ell \times N_q$
7		1	-1	0	1	-1	0	-2	2	0	$L_q \times N_\ell$
8		1	1	-2	1	1	-2	-2	-2	4	$L_q \times N_q$

Here the subscripts ℓ and q refer to the linear effect and quadratic effect, respectively, and $L_\ell \times N_\ell$ refers to the linear by linear interaction etc.

Pearce (1963) gave two examples of designs for nine treatments in six rows and six columns. The first was of type O:GG(F) and is given in Table 3.9.2. In this design, which is in factorial balance, the groups of treatments in the rows are (A_1, A_2, A_3), (B_1, B_2, B_3) and (C_1, C_2, C_3) and the groups in the columns are (A_1, B_1, C_1), (A_2, B_2, C_2) and (A_3, B_3, C_3). The second design was of type O:FF(T) and is given in Table 3.9.3. This design is in total balance. The variance-covariance matrix for the design in Table 3.9.2 is given in Table 3.9.4 and the variance-covariance matrix for the design in Table 3.9.3 is given in Table 3.9.5.

Table 3.9.2. Pearce's design of type O:(F)

A_1	A_2	A_3	B_1	B_2	B_3
B_1	C_3	B_3	B_2	C_2	C_1
A_2	C_2	C_1	A_1	C_3	A_3
C_1	A_3	C_3	C_2	A_2	A_1
C_2	B_2	B_1	C_1	B_3	C_3
B_2	B_3	A_1	A_2	A_3	B_1

Table 3.9.3. Pearce's design of type O:(T)

A_2	B_1	A_1	A_3	B_3	B_2
C_2	B_2	B_3	B_1	C_3	C_1
B_3	C_3	C_1	A_1	B_1	A_3
C_1	A_3	C_3	C_2	A_2	A_1
B_1	C_1	A_2	B_2	A_1	C_2
A_3	A_2	B_2	C_3	C_2	B_3

Table 3.9.4. Variance-covariance matrix for O:(F) design

$$\begin{bmatrix} 0.33 & & & & & & \\ & 0.33 & & & & & \\ & & 0.33 & & & & \\ & & & 0.33 & & & \\ & & & & 0.25 & & \\ & & & & & 0.25 & \\ & & & & & & 0.25 \\ & & & & & & & 0.25 \end{bmatrix}$$

Table 3.9.5. Variance-covariance matrix for O:(T) design

$$\begin{bmatrix} 0.2857 & & & & & & \\ & 0.2857 & & & & & \\ & & 0.2857 & & & & \\ & & & 0.2857 & & & \\ & & & & 0.2857 & & \\ & & & & & 0.2857 & \\ & & & & & & 0.2857 \\ & & & & & & & 0.2857 \end{bmatrix}$$

From the variance-covariance matrices it can be seen that

the 0:(F) design would be appropriate for the situation where the interactions were more important than the main effects, and the 0:(T) design would be appropriate for the situation where all treatment contrasts were of equal importance.

To obtain different sets of components three sets of weights were used. The main effects of L were each given a weight of W_1, the main effects of N were each given a weight of W_2 and the interactions were each given a weight of W_3. In all, nineteen different combinations of weights made up of the values 1, $\frac{1}{2}$ and 0 were used. Here a weight of 1 was intended to indicate that a contrast was to be estimated with as small a variance as possible and a weight of 0 was intended to indicate that a contrast was of no importance and its variance need not be kept small. A weight of a $\frac{1}{2}$ was intended to indicate that a contrast was not as important as a contrast with weight 1 and its variance need not be made as small as a contrast with weight 1, but even so should be made as small as possible. For all sets of weights the method of constructing the starting design gave the design given in Table 3.9.6. This design is of type 0:GG(G) and the groups of treatments in the rows (and columns) are (A_1, A_2, A_3), (B_1, B_2, B_3) and (C_1, C_2, C_3). The variance-covariance matrix of the contrasts, obtained using this design, is given in Table 3.9.7. This design would be appropriate for the situation where the main effects of L were of less importance than those of the other main effects and interactions.

Table 3.9.6. The starting design produced by the algorithm

$C_3\ C_1\ B_2\ B_3\ C_2\ B_1$
$C_1\ C_3\ A_2\ C_2\ A_3\ A_1$
$B_1\ A_1\ A_3\ B_2\ A_2\ B_3$
$B_3\ C_2\ B_1\ C_1\ C_3\ B_2$
$C_2\ A_2\ A_1\ C_3\ C_1\ A_3$
$B_2\ A_3\ B_3\ B_1\ A_1\ A_2$

Table 3.9.7. Variance-covariance matrix for the starting design

$$\begin{bmatrix} 0.5 & & & & & & & \\ & 0.5 & & & & & & \\ & & 0.25 & & & & & \\ & & & 0.25 & & & & \\ & & & & 0.25 & & & \\ & & & & & 0.25 & & \\ & & & & & & 0.25 & \\ & & & & & & & 0.25 \end{bmatrix}$$

To obtain different pairs of components three different starting designs were used. The first, labelled here as P1 (Pearce's first design), was the O:(F) design given in Table 3.9.2, the second, labelled here as P2 (Pearce's second design), was the O:(T) design given in Table 3.9.3, and the third, labelled here as C (Computer generated), was the O:(G) design given in Table 3.9.6. In each run of the algorithm to search for the components H was set equal to 4.

Consider now an algorithm for combining two component designs. To illustrate the steps in this algorithm the two components found in Example 3.4.1 will be used for simplicity.

For the purpose of writing a computer algorithm, the task of combining two components can be considered as finding some arrangement of the treatments in each block

of the row component that satisfies the following. The set of treatments that occupies the first position in each block of the row component satisfies the first column of the incidence matrix $\underset{\sim}{N}_2$ of the column component, the set of treatments in the second position of each block satisfies the second column of $\underset{\sim}{N}_2$, and so on.

For Example 3.4.1 the column component found by the algorithm had incidence matrix

$$\underset{\sim}{N}_2 = \begin{bmatrix} 1 & 2 & 1 & 1 & 2 & 2 \\ 2 & 1 & 2 & 2 & 1 & 1 \\ 2 & 1 & 1 & 2 & 1 & 2 \\ 1 & 2 & 2 & 1 & 2 & 1 \end{bmatrix}.$$

Any treatment in the first block of the row component found by the algorithm may occur in any column therefore a suitable first row would be

1 2 3 3 4 4.

The choice of arrangement for the first row may affect the choice of arrangement for the second row. The effect can be determined by subtracting 1 from the $(i,j)^{th}$ element of $\underset{\sim}{N}_2$ if treatment i occurred in position j of the first row. For Example 3.4.1 the adjusted incidence matrix after choosing the first row as above is

$$\text{adjusted } \underset{\sim}{N}_2 = \begin{bmatrix} 0 & 2 & 1 & 1 & 2 & 2 \\ 2 & 0 & 2 & 2 & 1 & 1 \\ 2 & 1 & 0 & 1 & 1 & 2 \\ 1 & 2 & 2 & 1 & 1 & 0 \end{bmatrix}.$$

From this adjusted incidence matrix, lists S_i, $i = 1, 2, \ldots, b_2$ can be formed that indicate which treatments may occur in column i. These lists will be useful in the computer algorithm described below. For Example 3.4.1, if the second block of the row component (2 1 1 2 3 4) is considered as the second row, the lists S_i are $S_1 = (2, 0, 0, 2, 3, 4)$, $S_2 = (0, 1, 1, 0, 3, 4)$,

$S_3 = (2, 1, 1, 2, 0, 4)$, $S_4 = (2, 1, 1, 2, 3, 4)$, $S_5 = (2, 1, 1, 2, 3, 4)$ and $S_6 = (2, 1, 1, 2, 3, 0)$. The lists are constructed by writing in the list S_i, $i = 1, 2, \ldots, 6$, the treatments that occur in the block and then replacing those treatments that cannot occur in column i by zero.

Therefore the arrangement

1 1 2 2 3 4

is not suitable because 1 cannot occur in column 1; a suitable arrangement is

4 1 1 2 3 2.

A procedure for combining the two components is therefore as follows. Construct the column incidence matrix $\underset{\sim}{N}_2$ and from it the lists S_i, $i = 1, 2, \ldots, b_2$, for the first block of the row component. Attempt to find an arrangement of the treatments in the first block to satisfy these lists. If a suitable arrangement can be found then take this as the first row. Adjust $\underset{\sim}{N}_2$ to take into account this choice and form new lists S_i for the treatments in the second block of the row component. Attempt to find a suitable arrangement of these treatments to satisfy the lists. If an arrangement can be found take this as the second row and adjust $\underset{\sim}{N}_2$ further to take account of the choice of second row. Form lists S_i for the third block of the row component and so on until all rows have been formed or no suitable arrangement for a particular row can be found.

The procedure as described leaves two difficulties to be resolved. The first is that of deciding if a suitable arrangement can be found and the second is that of deciding what to do if not.

Consider first the problem of choosing a suitable arrangement of the treatments in a block to satisfy the lists S_i. For each block the lists S_i are formed as

$$S_i = (E_1, E_2, \ldots, E_{b_2}),$$

where E_j is the treatment in position j of the block if that treatment can occur in column i, or $E_j = 0$ otherwise.

The problem of choosing a suitable arrangement is therefore that of choosing from the lists S_i an element E_j such that no E_j is zero and no E_j is chosen from more than one list. The elements E_j and E_k are considered distinct if $j \neq k$, e.g. E_1 and E_4 in list S_1, given above for the first row, are distinct even though $E_1 = 2$ and $E_4 = 2$.

When framed in this way the problem reduces to the familiar one of choosing from the lists $S_1, S_2, \ldots, S_{b_2}$ a set of distinct representatives. The lists are so constructed to take into account the choice of arrangements of any previous row. Hall (1935) gave a theorem which states that a necessary and sufficient condition for a set of distinct representatives to exist is that for every value of k, $1 \leq k \leq b_2$, any k lists have in their union at least k different members. Anderson (1974, p.26) gave a different formulation of this theorem and a constructive proof. This theorem and its proof are given below.

The Marriage Theorem (Anderson, 1974).

Given a set of n men and a set of n women, each man makes a list of women he is willing to marry. Then each man can be married off to a woman on his list if and only if,

(*) {for every value of k, any k lists contain in their union at least k names.}

Anderson's Proof of the Marriage Theorem.

It is shown how, on the assumption that $r < n$ men have been paired off with suitable women, to increase this to $(r + 1)$ men.

Suppose r men have been paired off. If there is a man left who has on his list a woman who is still unattached, an $(r + 1)^{th}$ pairing is immediate. So suppose that all women on remaining lists are already attached.

Figure 3.5.1.

Choose any unmarried man A_o (see Figure 3.5.1). By (*) with k = 1, there is a woman B_1 on his list. B_1 is married to A_1, say. By (*) with k = 2, the combined lists of A_o and A_1 contain the name of at least one more woman B_2. If B_2 is unmarried, stop. If B_2 is married to A_2, then by (*) with k = 3, the combined lists of A_o, A_1 and A_2 contain a third name, say B_3. If B_3 is unmarried, stop. If B_3 is married to A_3, repeat the process, and continue until an unmarried woman B_s is reached. (This must happen eventually since not all women are married, and no B_i occurs twice in the process.) Note that by construction, each B_i is on the list of at least one A_j with $j < i$. This is very important. Consider now B_s. Pair her off with an A_i on whose list she appears $(i < s)$. This frees B_i. Next pair off B_i with some A_j $(j < i)$ on whose list

she appears. This frees B_j. Repeat until some woman is freed and re-paired with A_o. This must eventually happen. Then take all the new pairings and all the original ones which have not been tampered with. Now (r + 1) pairs have been obtained. Repeat the process if $r + 1 < n$.

Anderson (1974) notes that this proof has the advantage that the condition (*) need not be checked before the construction is attempted. If (*) does not hold, this will become clear when the method breaks down. On the other hand, if (*) holds, the method will not break down.

The proof given by Anderson can be considered as an algorithm for finding a set of distinct representatives. In the following the references to finding a set of distinct representatives refer to the application of a computerised version of the above algorithm written by the present author.

Consider now the second problem of deciding what to do when a set of distinct representatives cannot be found for a particular block of the row component. If this is the first block of the row component a row-and-column design cannot be constructed. For the remaining blocks either a row-and-column design cannot be constructed or the choice of arrangements for the previous rows has made it impossible to find a set of distinct representatives. In this second possibility a choice of different sets of distinct representatives for the previous rows may make it possible to find a set for the block under consideration. A possible method is as follows. Label the blocks as 1, 2, ..., L with L the block under consideration and 1, 2, ..., L-1 the blocks that have already been arranged

as rows of the design. Then if a set of distinct representatives for block L cannot be found attempt to find a new set of distinct representatives for row L-1 that will allow row L to be formed. If such a set cannot be found for row L-1, attempt to find a new set for row L-2 that will allow row L-1 and row L to be formed, and so on, each time keeping a record of the rows that do not allow row L to be formed. The procedure is repeated until either a suitable arrangement of rows 1, 2, ..., L can be found or no arrangement is possible. If the second possibility occurs then a row-and-column design cannot be constructed from the components using the above algorithm.

When deciding if a new set of distinct representatives can be found for a block ℓ, $1 \leq \ell \leq L-1$, the lists S_i are constructed in the following way. The adjusted incidence matrix is restored to the state it had prior to fitting row ℓ, i.e. N_2 adjusted only for the present choice of rows 1, 2, ..., $\ell-1$. The lists are then formed in the usual way. These lists are then adjusted further to take into account all arrangements of block ℓ that have followed the present row $\ell-1$ and did not permit rows $\ell+1$, $\ell+2$, ..., L to be formed. This is done, for each list S_i by setting $E_j = 0$ if E_j is an element in the list i corresponding to a treatment that occurred in column i of any of the previous choices for row ℓ.

When this algorithm is applied to the two components found for Example 3.4.1 the row-and-column design is found to be

1 2 3 3 4 4
4 1 1 2 3 2
2 4 2 3 4 1
3 4 2 1 2 1
2 3 4 4 1 3
3 1 4 2 1 3.

Return now to the fifty-seven pairs of components found for the design for nine treatment combinations in six rows and six columns. It will be convenient to refer to the pairs of components using a combination of the letters P1, P2, and C, and a set of weights (W_1, W_2, W_3). So, for example, the components obtained using starting design C, as given in Table 3.9.6 and set of weights (0, 1, 1), will be identified by the code C(0, 1, 1).

Out of the fifty-seven pairs of components, the pairs found for C(0, 1, 1), C(0, 1, 0), C(0, 0, 1), C(0, 1, $\frac{1}{2}$) and C(0, $\frac{1}{2}$, 1) were equal to the starting design given in Table 3.9.6. The pairs found for P1(0, 0, 1) and P1($\frac{1}{2}$, $\frac{1}{2}$, 1) were equal to the starting design given in Table 3.9.2. The pair found for P2(1, 1, 1) was equal to the starting design given in Table 3.9.3. Therefore, of the forty-nine pairs found, seventeen could not be combined by the above algorithm. If the three starting designs, the components of which could be combined, are also considered the failure rate for the above algorithm is 17/52.

The reason the algorithm failed was that not all possible arrangements of the rows of the design were considered at each step. After a row had been rejected the procedure for choosing an alternative row eliminated a number of possible choices. The task of enumerating all possible sets of distinct representatives for a given row, storing them in the computer and successively trying each

one if necessary would have meant an unwieldy and time consuming computer program. Therefore, some alternative approaches were considered.

The failures of the algorithm were not all because it was impossible to combine the components, but because of the above flaw in the algorithm.

Consider therefore a second algorithm.

3.10 A second algorithm for combining the component designs

In the algorithm described in the previous section the components were combined by arranging a row of the design at a time. The arrangement of the treatments in each row was such that the incidence matrix for the columns was $\underset{\sim}{N}_2$. As noted above the algorithm sometimes failed to combine the components when in fact, as will be seen in the following, some of them could be combined. A second algorithm that arranges a row and a column at each step was therefore developed.

The idea for this second algorithm stemmed from the following theorems given by Freeman (1957).

Theorem 1. For two block designs to combine into a row-and-column design, a necessary condition is that the matrix formed by multiplying the matrix $\underset{\sim}{N}_1'$ into the matrix $\underset{\sim}{N}_2$ shall have no zeros.

Theorem 2. For two block designs to combine into a row-and-column design, a necessary condition is that, given any block of the row design, a matrix $\underset{\sim}{M}$ with v rows and b_2 columns can be obtained with the following properties: (i) the totals in successive rows of $\underset{\sim}{M}$ shall be a vector corresponding to the given block of the row design, (ii) in each column of $\underset{\sim}{M}$ there shall be one value of unity and the rest zeros, (iii) it shall be possible to subtract

$\underset{\sim}{M}$ from $\underset{\sim}{N}_2$ without the occurrence of any negative numbers.

Theorem 3. For two block designs to combine into a row-and-column design, a necessary condition is that if a matrix with the same first row as the matrix $\underset{\sim}{N}_1'$ and values in the other rows corresponding to positions in the first column of the matrix $\underset{\sim}{N}_2$ is subtracted from the matrix $\underset{\sim}{N}_1'$ to form a reduced row-treatment matrix $(\underset{\sim}{n}_1')$ and a reduced treatment-column matrix $(\underset{\sim}{n}_2)$ is obtained similarly, then there shall be no zeros, except in the first row and column, in the matrix formed by multiplying the matrix $\underset{\sim}{n}_1'$ into the matrix $\underset{\sim}{n}_2$.

He further states that the process of construction of row-and-column designs can be likened to fitting a nest of boxes of ever decreasing size into one another, any of the theorems corresponding to a condition that a particular box shall fit into the preceding one.

As Freeman pointed out in Theorem 1 there must be no zero values in the matrix $\underset{\sim}{N}_1'\underset{\sim}{N}_2$ and therefore before beginning the algorithm described below this is assumed.

Obviously the treatments in the blocks of the row component must be arranged as rows of the design so that the incidence matrix for columns is $\underset{\sim}{N}_2$, and the treatments in the blocks of the column component must be arranged as columns of the design so that the incidence matrix for rows is $\underset{\sim}{N}_1$. In the following a "suitable arrangement" will mean an arrangement that satisfies one of these conditions.

For convenience, the blocks of the row component are numbered as 1, 2, ..., b_1 and the blocks of the column component as 1, 2, ..., b_2. At each step in the

algorithm the treatments in block j, say, of the row component and block j of the column component are arranged as row and column j of the design. Obviously, the treatment assigned to plot (j, j) of the design $\{j = 1, 2, \ldots, \min(b_1, b_2)\}$ must be one of the treatments common to block j of each component. Therefore, before arranging row and column j a list L_j is compiled which gives the treatments common to block j of the row component and block j of the column component. However, not all treatments in block j of each component are considered, because unless j = 1 some of them will have been assigned to plots in rows 1, 2, ..., j-1 and columns 1, 2, ..., j-1. Therefore, in the following, L_j contains only those treatments common to block j of each component that are eligible to be assigned to plot (j, j). It is possible that list L_j will contain more than one element and so there may be a choice of treatment to assign to plot (j, j). Further, it may be that some of the treatments in list L_j when assigned to plot (j, j) make it impossible for the remaining plots in row and column j to be assigned a suitable arrangement of treatments. Therefore, in the algorithm the elements in list L_j are considered successively until one is found that enables a suitable arrangement of row and column j to be found.

If no element in list L_j (j > 1), when assigned to plot (j, j), enables row and column j to be suitably arranged then all record of L_j is deleted and a new arrangement sought for row and column j-1. If j = 1 (i.e. j-1 = 0) then the algorithm has failed to combine the components. In order to find a new arrangement of row and column j-1 that will enable row and column j to

be arranged, the next treatment in list L_{j-1} that has not yet been assigned to plot (j-1, j-1) is assigned to that plot. If no treatment in list L_{j-1} when assigned to plot (j-1, j-1) enables row and column j-1 to be suitably arranged then all record of list L_{j-1} is deleted and a new arrangement sought for row and column j-2, and so on. Eventually either a new arrangement of some or all of rows and columns 1, 2, ..., j-1 is found that enables row and column j to be suitably arranged, or all arrangements of the treatments common to each pair of row and column blocks will have been tried on plots (1, 1), (2, 2), ..., (j-1, j-1) without a suitable arrangement being found. Even if the latter happens it is not certain that it is impossible to combine the components because in the algorithm not all possible arrangements of rows and columns are considered, as will be explained below.

As in the algorithm described in Section 3.9 a suitable arrangement is found by compiling lists of treatments that may occur on a given set of plots and finding one set of distinct representatives for these lists. Consequently as not all sets of distinct representatives are obtained it is possible that the one chosen may make it impossible to arrange a row or column later in the algorithm. However, in this algorithm there is usually a choice of treatment to assign to plot (j, j), $\{j = 1, 2, \ldots, \min(b_1, b_2)\}$ and so a number of sets of distinct representatives for row or column j can be sought by assigning successively to plot (j, j) every treatment in list L_j that is eligible to be assigned to that plot.

Having outlined the main steps in the algorithm it is necessary to fill in the details. After a treatment from

list L_1 has been assigned to plot (1, 1) the remaining $b_2 - 1$ treatments (not all necessarily distinct) in block 1 of the row component need to be assigned to plots 2, 3, ..., b_2 of row 1 of the design. These treatments must be arranged in such a way that in the final design the incidence matrix for columns is $\underset{\sim}{N}_2$. Therefore, in a similar way to that described in Section 3.9, columns 2, 3, ..., b_2 of $\underset{\sim}{N}_2$ are used to compile lists S_2, S_3, ..., S_{b_2} which give respectively, the treatments that can be assigned to plots 2, 3, ..., b_2 of row 1. A set of distinct representatives, if one exists, is found using the Marriage Theorem algorithm given by Anderson (1974) and this defines the arrangement of row 1. The remaining $b_1 - 1$ treatments in block 1 of the column component are similarly assigned to plots 2, 3, ..., b_1 of column 1 by compiling lists based on columns 2, 3, ..., b_1 of $\underset{\sim}{N}_1$ and finding a set of distinct representatives for these lists.

A suitable arrangement of row and column 1 having been found the incidence matrices $\underset{\sim}{N}_1$ and $\underset{\sim}{N}_2$ are adjusted to take account of these arrangements. That is, if treatment ℓ, say, (ℓ = 1, 2, ..., v) was assigned to plot p, (p = 2, 3, ..., b_2) of row 1 then a 1 is subtracted from element (ℓ, p) of $\underset{\sim}{N}_2$. Similarly, a 1 is subtracted from element (m, q) of $\underset{\sim}{N}_1$ if treatment m, (m = 1, 2, ..., v) was assigned to plot q, (q = 2, 3, ..., b_1) of column 1.

Making these adjustments makes the compilation of list L_2 quite easy. The treatments common to block 2 of each component can be determined from column 2 of each of the adjusted matrices $\underset{\sim}{N}_1$ and $\underset{\sim}{N}_2$. If the product of element i, (i = 1, 2, ..., v) in column 2 of $\underset{\sim}{N}_1$ and element i in column 2 of $\underset{\sim}{N}_2$ is zero then treatment i is not common to

block 2 of each component. Further, if there is a zero element in row or column k, ($k = 2, 3, \ldots, b_1$ or b_2) of the matrix $\underset{\sim}{N}_1'\underset{\sim}{N}_2$ then as Freeman pointed out in Theorem 2, the remaining treatments in blocks 2, 3, ..., b_1 of the row component and blocks 2, 3, ..., b_2 of the column component cannot be arranged as rows and columns of the design. If this happens a new arrangement of row and column 1 must be sought.

After rows and columns 1, 2, ..., j-1 have been suitably arranged there will exist matrices $\underset{\sim}{N}_1$ and $\underset{\sim}{N}_2$ suitably adjusted for these arrangements. From $\underset{\sim}{N}_1$ and $\underset{\sim}{N}_2$ it can be determined if the remaining treatments in blocks j, j+1, ..., b_1 of the row component and blocks j, j+1, ..., b_2 of the column component can be arranged as rows and columns of the design. If they cannot, a new arrangement of rows and columns 1, 2, ..., j-1 must be sought. If they can, then L_j is compiled. After a treatment has been assigned to plot (j, j) lists $S_{j+1}, S_{j+2}, \ldots, S_{b_2}$ are compiled using columns j+1, j+2, ..., b_2 of $\underset{\sim}{N}_2$ and lists $S_{j+1}, S_{j+2}, \ldots, S_{b_1}$ using columns j+1, j+2, ..., b_1 of $\underset{\sim}{N}_1$.

If no arrangement for row and column j can be found then a new arrangement for row and column j-1 is sought. However, before looking for this new arrangement the matrices $\underset{\sim}{N}_1$ and $\underset{\sim}{N}_2$ are restored to the state they had immediately after being adjusted for the present arrangement of row and column j-2.

Obviously if no suitable arrangement for plots j+1, j+2, ..., b_2 of row j can be found there is no point in looking for a suitable arrangement for plots j+1, j+2, ..., b_1 of column j. Therefore, if this happens the next treatment, if any, in list L_j is assigned to plot (j, j)

and a new arrangement for row and column j is sought.

After rows and columns $\{1, 2, \ldots, \min(b_1, b_2)\}$ have been suitably arranged, two possibilities are left. If $b_1 = b_2$ then the design has been completed and no more arranging is required. If $b_1 \neq b_2$ then if $b_1 > b_2$ plots $b_2 + 1, b_2 + 2, \ldots, b_1$ of column b_2 still need to be assigned treatments and if $b_2 > b_1$ plots $b_1 + 1, b_1 + 2, \ldots, b_2$ need to be assigned treatments.

If $b_1 > b_2$, this is done by compiling lists S_{b_2+1}, $S_{b_2+2}, \ldots, S_{b_1}$ using columns $b_2 + 1, b_2 + 2, \ldots, b_1$ of N_1 and finding a set of distinct representatives for plots $b_2 + 1, b_2 + 2, \ldots, b_1$ of column b_2. If $b_2 > b_1$ lists $S_{b_1+1}, S_{b_1+2}, \ldots, S_{b_2}$ are compiled using columns $b_1 + 1, b_1 + 2, \ldots, b_2$ of N_2 and finding a set of distinct representatives for plots $b_1 + 1, b_1 + 2, \ldots, b_2$ of row b_1.

If no suitable arrangement can be found then a new arrangement for row and column $\{\min(b_1, b_2) - 1\}$ is sought and so on.

To illustrate the steps in the algorithm consider again Example 3.4.1 for four treatments in six rows and six columns. Here

$$N_1 = \begin{bmatrix} 1 & 2 & 1 & 2 & 1 & 2 \\ 1 & 2 & 2 & 2 & 1 & 1 \\ 2 & 1 & 1 & 1 & 2 & 2 \\ 2 & 1 & 2 & 1 & 2 & 1 \end{bmatrix} \quad \text{and} \quad N_2 = \begin{bmatrix} 1 & 2 & 1 & 1 & 2 & 2 \\ 2 & 1 & 2 & 2 & 1 & 1 \\ 2 & 1 & 1 & 2 & 1 & 2 \\ 1 & 2 & 2 & 1 & 2 & 1 \end{bmatrix}.$$

No element of $N_1 N_2'$ is zero so the algorithm can be applied.

List $L_1 = (1, 2, 3, 4)$ and so treatment 1 is initially assigned to plot (1, 1). An arrangement of treatments (2, 3, 3, 4, 4) is now required that will give a suitable first row. Using columns 2, 3, ..., 6 of N_2 each of the lists $S_1, S_2, \ldots, S_6$ is found to be equal to (2, 3, 3, 4, 4) and a set of distinct representatives is (2, 3, 3, 4, 4).

Therefore the arrangement of the first row is (1, 2, 3, 3, 4, 4). To find the arrangement of the first column we must find an arrangement of (2, 3, 3, 2, 4). Using columns 2, 3, ..., 6 of $\underset{\sim}{N}_1$ each of the lists S_2, S_3, ..., S_6 is found to be equal to (2, 3, 3, 2, 4) and a set of distinct representatives is (2, 2, 3, 3, 4). The arrangement of the first column is therefore (1, 2, 2, 3, 3, 4). At this stage the design has the form

$$\begin{array}{l} 1\ 2\ 3\ 3\ 4\ 4 \\ 2 \\ 2 \\ 3 \\ 3 \\ 4 \end{array} ,$$

$$\underset{\sim}{N}_1 = \begin{bmatrix} 0 & 2 & 1 & 2 & 1 & 2 \\ 0 & 1 & 1 & 2 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 2 \\ 0 & 1 & 2 & 1 & 2 & 0 \end{bmatrix} \quad \text{and} \quad \underset{\sim}{N}_2 = \begin{bmatrix} 0 & 2 & 1 & 1 & 2 & 2 \\ 0 & 0 & 2 & 2 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 2 \\ 0 & 2 & 2 & 1 & 1 & 0 \end{bmatrix}.$$

The matrix $\underset{\sim}{N}_1'\underset{\sim}{N}_2$ has no zero elements except in the first row and column and so the next step in the algorithm is considered.

The list L_2 = (1, 3, 4) and so initially treatment 1 is assigned to plot (2, 2). This leaves treatments (1, 2, 3, 4) to be assigned to plots 3, 4, 5, 6 of row 2 and treatments (1, 3, 4, 4) to be assigned to plots 3, 4, 5, 6 of column 2. Using $\underset{\sim}{N}_2$, the lists for row 2 are S_3 = (1, 2, 0, 4), S_4 = S_5 = (1, 2, 3, 4) and S_6 = (1, 2, 3, 0). A set of distinct representatives is (4, 2, 3, 1) and so a suitable second row would be (2, 1, 4, 2, 3, 1). Using $\underset{\sim}{N}_1$, the lists for column 2 are S_3 = (1, 3, 4, 4), S_4 = (1, 0, 4, 4), S_5 = (1, 3, 4, 4) and S_6 = (1, 3, 0, 0). A set of distinct representatives is (4, 4, 3, 1) and so a suitable second column would be (2, 1, 4, 4, 3, 1). At this stage the design has the form

1 2 3 3 4 4
2 1 4 2 3 1
2 4
3 4
3 3
4 1 .

Continuing in this way the algorithm gives the combined design as

1 2 3 3 4 4
2 1 4 2 3 1
2 4 1 2 4 3
3 4 2 1 1 2
3 3 4 4 2 1
4 1 2 3 1 3.

The above algorithm was applied to the three starting designs and the forty-nine different pairs of components. This new algorithm failed to combine twelve pairs of components, i.e. a failure rate of 12/52. This is slightly lower than the failure rate of 17/52 for the algorithm described in Section 3.9. So for the fifty-two pairs of components considered here the new algorithm has been more successful. However, as will be seen in the next section most of these pairs of components can be combined and so this second algorithm also suffers from the disadvantage of not being able to determine all possible arrangements of each row and column. A failure rate of 12/52 is still too high and so a more successful algorithm must be devised. This is done in the next section.

3.11 A third algorithm for combining components

In the previous algorithms for combining components not all possible arrangements of each row and column were considered. This meant some components were not combined when in fact it was possible to combine them. However, unless all possible arrangements are considered these failures are bound to occur. The previous algorithms could probably

be modified to search for all arrangements but this would require a very efficient procedure for generating them if the algorithm is to run quickly on the computer. Therefore, rather than modify the algorithms a new approach is adopted. In this approach a treatment is assigned to a plot at a time, and an algorithm to search for sets of distinct representatives is not needed.

Initially for each plot a list of possible treatments is prepared. Thus, if plot j is in row p and column q, it is a question of listing those treatments common to block p of the row component and block q of the column component. If any list is empty then as Freeman's first theorem points out, the two components cannot be combined. It will however be assumed that plot j has S_j possible treatments, $E_1, E_2, \ldots, E_{S_j}$, where $S_j \geq 1$ for all j, the list being called L_j.

A treatment is then allocated to each plot in turn, working along the rows. The treatment allocated to plot j is the first eligible treatment in list L_j; a treatment may have become ineligible because of an allocation already made. If treatment E_{ℓ_j} say, is allocated to plot j then E_{ℓ_j} is replaced by $-E_{\ell_j}$ in list L_j, the minus sign acting as an indicator for possible use later. Further, a one is subtracted from element (E_{ℓ_j}, p) of $\underset{\sim}{N}_1$, where p is the row in which plot j occurs, and a one is subtracted from element (E_{ℓ_j}, q) of $\underset{\sim}{N}_2$, where q is the column in which plot j occurs. These adjustments to $\underset{\sim}{N}_1$ and $\underset{\sim}{N}_2$ ensure that a treatment does not occur too many times in any row or column. Therefore, treatment E_{ℓ_k} from list L_k, $k > j$, may become ineligible for plot k because element (E_{ℓ_k}, p) of $\underset{\sim}{N}_1$ is zero,

where p is the row in which plot k occurs, or because element (E_{ℓ_k}, q) of N_2 is zero, where q is the column in which plot k occurs. If this happens, treatment E_{ℓ_k} is replaced by $-E_{\ell_k}$ in list L_k and the next treatment in list L_k is considered.

Of course, list L_k may contain no treatment eligible for plot k because of the way the treatments have been allocated to the previous plots. When this happens, a new arrangement of treatments must, if possible, be found that will allow a treatment from list L_k to be allocated to plot k. This is done as follows. Any minus sign is removed from list L_k and then plot k - 1 is reconsidered. The matrices N_1 and N_2 are readjusted to the state they had after the present assignment of treatments had been made to plots 1, 2, ..., k - 2. The first treatment in list L_{k-1} that does not have a minus sign is then considered for possible allocation to plot k - 1. If this treatment is ineligible for plot k - 1 then a minus sign is added to it and the next treatment in list L_k considered and so on. If a treatment in list L_{k-1} can be allocated to plot k - 1 then this is done, and a minus sign added to that treatment in list L_{k-1}, and the matrices N_1 and N_2 adjusted to take into account the allocation of this treatment to plot k - 1. Plot k is then reconsidered. If no treatment in list L_{k-1} is eligible for plot k - 1 then the procedure described for plots k and k - 1 is repeated for plots k - 1 and k - 2, and so on until either an arrangement can be found that allows a treatment from list L_k to be allocated to plot k or no such arrangement can be found. In the latter case the components cannot be combined to give a row-and-column design.

In the above algorithm the labelling of treatments with a minus sign ensures that no arrangement is considered more than once. Further, different arrangements are considered successively until a suitable one is found. This algorithm also has the advantage of not requiring sets of distinct representatives to be obtained at each step.

For Example 3.4.1 the lists for all thirty-six plots are $L_i = (1, 2, 3, 4)$, $i = 1, 2, \ldots, 36$, and the algorithm described above gives the following design

1 2 3 3 4 4
2 1 1 2 4 3
2 1 4 4 2 3
3 4 2 1 1 2
3 4 4 2 3 1
4 3 2 3 1 1.

The above algorithm failed to combine three of the fifty-two pairs of components. These were the final designs found for P1(1, 0, 0), C(1, 0, 0) and C(1, 0, $\frac{1}{2}$). As the algorithm considers every possible arrangement these failures occurred because the components could not be combined. Therefore, the success rate of the algorithm is obviously a 100% whenever the components can be combined, and it is usually very quick in its execution. For example, it took 0.411 CP seconds on the University of London CDC 7600 computer to combine the components of the design for 21 treatments in 21 rows and 5 columns given by Youden (1937). The failure rate of the algorithm that fits a row and a column at each step, and described in Section 3.10, was therefore 9/49 and the failure rate for the algorithm that fits a row at a time and described in Section 3.9 was therefore 14/49.

What is also apparent is that the algorithm that finds the components can find a pair that cannot be

combined. A failure rate of 3/52 does appear to be high but these three pairs are the only ones that have occurred in over one hundred runs of the algorithm.

Having now developed an efficient algorithm for combining the components the properties of the forty-nine different designs obtained by the algorithm that searches for the components can be considered. This is done in the next section.

A description of the algorithm that searches for the components and of the third algorithm for combining them was given by Jones (1979). A computerised version of the third algorithm for combining the components (Jones, 1980b) has been accepted for publication in Applied Statistics.

3.12 Properties of the designs found by the optimality algorithm

When considering the properties of the designs found by the algorithm that searches for components it should be borne in mind that the optimality criterion uses only the weighted variances of the estimates of the contrasts. The covariances between the estimates are not considered. Obviously, a desirable property of a design is that the estimates of the contrasts of interest are mutually orthogonal, if that is possible, and also that these estimates are orthogonal to the estimates of any other contrasts that taken together with the contrasts of interest account for all the degrees of freedom for treatments. The partitioning of the treatment sums of squares into orthogonal components in an analysis of variance is then made possible.

It will be convenient to make use of the following definition.

Definition 3.12.1.

Given a complete set of orthogonal contrasts, $\underset{\sim}{L}$, a design for which $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ has some non-zero off-diagonal elements will be referred to as being of type NOEC (i.e. Non-Orthogonal Estimates of Contrasts).

A design of type NOEC is not necessarily a bad one because the non-zero covariances could all be between estimates of contrasts for which no interest has been expressed. This point will become clear in the following.

Table 3.12.1 gives a summary of the properties of the fifty-seven designs obtained in the tests of the algorithm. It will be recalled that there were eight treatment contrasts. The two main effects of "letters" (L) were each given a weight of W_1, the two main effects of "numbers" (N) were each given a weight of W_2 and the remaining four interactions were each given a weight of W_3. The * in the table indicates that the two component designs found for that set of weights could not be combined. These designs will be considered again in the next section. Also indicated in the table are those designs that were equal to the starting design.

For convenience the designs will be considered in groups depending on the nature of their weights. As in each design the replication of each treatment is four the estimate of a contrast cannot have a variance smaller than 0.25.

Consider first the designs with weights (1, 1, 1). The P2(1, 1, 1) design was equal to the starting design and so was optimal, being of type O:(T). The other two designs, however, were of type NOEC and were not optimal. The variance-covariance matrix for the estimates of the

Table 3.12.1. Types of design

Weights W_1	W_2	W_3	Starting design	Type of final design
1	1	1	P1	NOEC
1	1	1	P2	O:(T)†
1	1	1	C	NOEC
1	1	0	P1	NOEC
1	1	0	P2	NOEC
1	1	0	C	NOEC
1	0	1	P1	O:(G)
1	0	1	P2	O:(G)
1	0	1	C	O:(G)
0	1	1	P1	O:(G)
0	1	1	P2	O:(G)
0	1	1	C	O:(G)†
1	0	0	P1	*
1	0	0	P2	NOEC
1	0	0	C	*
0	1	0	P1	NOEC
0	1	0	P2	NOEC
0	1	0	C	O:(G)†
0	0	1	P1	O:(F)†
0	0	1	P2	NOEC
0	0	1	C	O:(G)†
1	1	$\frac{1}{2}$	P1	O:(F)
1	1	$\frac{1}{2}$	P2	O:(F)
1	1	$\frac{1}{2}$	C	O:(F)
1	$\frac{1}{2}$	1	P1	O:(F)
1	$\frac{1}{2}$	1	P2	O:(F)
1	$\frac{1}{2}$	1	C	NOEC
$\frac{1}{2}$	1	1	P1	O:(F)
$\frac{1}{2}$	1	1	P2	O:(F)
$\frac{1}{2}$	1	1	C	NOEC
1	$\frac{1}{2}$	$\frac{1}{2}$	P1	NOEC
1	$\frac{1}{2}$	$\frac{1}{2}$	P2	NOEC
1	$\frac{1}{2}$	$\frac{1}{2}$	C	NOEC
$\frac{1}{2}$	1	$\frac{1}{2}$	P1	O:(F)
$\frac{1}{2}$	1	$\frac{1}{2}$	P2	NOEC
$\frac{1}{2}$	1	$\frac{1}{2}$	C	O:(F)
$\frac{1}{2}$	$\frac{1}{2}$	1	P1	O:(F)†
$\frac{1}{2}$	$\frac{1}{2}$	1	P2	O:(F)
$\frac{1}{2}$	$\frac{1}{2}$	1	C	O:(F)
1	$\frac{1}{2}$	0	P1	NOEC
1	$\frac{1}{2}$	0	P2	NOEC
1	$\frac{1}{2}$	0	C	NOEC
1	0	$\frac{1}{2}$	P1	O:(G)
1	0	$\frac{1}{2}$	P2	O:(G)
1	0	$\frac{1}{2}$	C	*
$\frac{1}{2}$	1	0	P1	NOEC
$\frac{1}{2}$	1	0	P2	NOEC
$\frac{1}{2}$	1	0	C	NOEC
$\frac{1}{2}$	0	1	P1	O:(F)
$\frac{1}{2}$	0	1	P2	O:(F)
$\frac{1}{2}$	0	1	C	O:(F)
0	1	$\frac{1}{2}$	P1	O:(G)
0	1	$\frac{1}{2}$	P2	O:(G)
0	1	$\frac{1}{2}$	C	O:(G)†
0	$\frac{1}{2}$	1	P1	O:(F)
0	$\frac{1}{2}$	1	P2	O:(F)
0	$\frac{1}{2}$	1	C	O:(G)†

† Here the final design equalled the starting design.

treatment contrasts, $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$, for design P1(1, 1, 1), is given in Table 3.12.2. The matrix $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ for C(1, 1, 1) was identical to this except some of the signs of the covariances were reversed. The variances are quite close to the optimal value of 0.2857 and all covariances are quite small. Therefore this design is near optimal. A method of improving these designs is described in the next section.

Table 3.12.2. $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ for the P1(1, 1, 1) design

0.3066		-0.0183			-0.0065	0.0065
	0.2857					
-0.0183		0.2742			-0.0023	0.0094
			0.2733	0.0088		-0.0088
			0.0088	0.2795		0.0062
-0.0065		-0.0023			0.2887	0.0121
0.0065		0.0094			0.0121	0.3087
			-0.0088	0.0062		0.2795

Consider now the next eighteen designs in Table 3.12.1 which have weights made up of both 1's and 0's, but no $\frac{1}{2}$'s. The designs labelled as 0:(G) or 0:(F) were such that all eight contrasts had mutually orthogonal estimates and the estimates of the contrasts with weight 1 had the minimum variance of 0.25. The designs labelled as NOEC were such that all contrasts with weight 1 had mutually orthogonal estimates and the estimates had the minimum variance of 0.25. In addition, all the estimates of the contrasts with weight 1 were orthogonal to the estimates of the contrasts with weight 0, the contrasts of no interest. All these NOEC designs were therefore optimal as they satisfied

the requirements set at the start of the algorithm and had the additional feature of orthogonality. To illustrate this point the $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ matrix for design P1(1, 1, 0) is given in Table 3.12.3. Here the non-zero covariances are labelled by an x.

Table 3.12.3. $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ for the P1(1, 1, 0) design

0.2500							
	0.2500						
		0.2500			(all zeros)		
			0.2500				
				0.8274	x	x	x
(all zeros)				x	0.8487	x	x
				x	x	0.4886	x
				x	x	x	0.5524

Consider now the designs with weights that include a $\frac{1}{2}$. Deriving designs for these sets of weights is more difficult because there is less opportunity to reduce the variance of one contrast at the expense of the other. The designs found for weights (1, 1, $\frac{1}{2}$), (1, $\frac{1}{2}$, 1) and ($\frac{1}{2}$, 1, 1) that were of type 0:(F) were such that all contrasts had estimates that were mutually orthogonal. For the designs with weights (1, 1, $\frac{1}{2}$) the estimates of the contrasts with weight 1 had variance 0.25 and all those with weight $\frac{1}{2}$ had variance 0.3333. The designs with weights (1, $\frac{1}{2}$, 1) and ($\frac{1}{2}$, 1, 1) were such that the estimates of the main effects with weight 1 had variance 0.2857, the estimates of the main effects with weight $\frac{1}{2}$ had variance 0.40, and the estimates of the interactions had variance 0.25. Overall, a most satisfactory outcome.

The designs $C(1, \frac{1}{2}, 1)$ and $C(\frac{1}{2}, 1, 1)$ were of type NOEC and so were less satisfactory. They were however, near optimal. The matrix $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ for each of these designs is given in Tables 3.12.4 and 3.12.5, respectively.

Table 3.12.4. $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ for design $C(1, \frac{1}{2}, 1)$

0.2540				-0.0097	-0.0056	
	0.2857					
		0.4000				
			0.4000			
-0.0097				0.2738	0.0137	
-0.0056				0.0137	0.2579	
						0.2500
						0.2500

Table 3.12.5. $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ for the design $C(\frac{1}{2}, 1, 1)$

0.4000						
	0.4000					
		0.2659		0.0049	0.0084	0.0028	-0.0146
			0.2500			
		0.0049		0.2560	0.0103	-0.0069	
		0.0084		0.0103	0.2679	-0.0119	
		0.0028		-0.0069	-0.0019	0.2639	-0.0103
		-0.0146				-0.0103	0.2679

The designs found for weights $(1, \frac{1}{2}, \frac{1}{2})$ were all NOEC and the matrix $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ for each of these designs is given in Tables 3.12.6 to 3.12.8, respectively. One good feature of these designs is that the contrasts with weight 1 have minimum variance estimates which are mutually orthogonal, and orthogonal to the estimates of the other contrasts.

Table 3.12.6. $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ for the P1(1, $\frac{1}{2}$, $\frac{1}{2}$) design

0.2500						
	0.2500					
		0.3182	0.0262			
		0.0262	0.2879			
				0.2932	-0.0130	
				-0.0130	0.3083	
						0.2932 -0.0130
						-0.0130 0.3083

Table 3.12.7. $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ for the P2(1, $\frac{1}{2}$, $\frac{1}{2}$) design

0.2500						
	0.2500					
		0.2960	0.0083	0.0063	0.0002	-0.0031 -0.0121
		0.0083	0.2864	0.0104	-0.0004	0.0054 -0.0002
		0.0063	0.0104	0.3087	0.0007	0.0090 0.0079
		0.0002	-0.0004	0.0007	0.3095	-0.0231 0.0007
		-0.0031	0.0054	-0.0090	-0.0231	0.2999 0.0090
		-0.0121	-0.0002	0.0079	0.0007	-0.0090 0.3087

Table 3.12.8. $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ for the C(1, $\frac{1}{2}$, $\frac{1}{2}$) design

0.2500						
	0.2500					
		0.2960	0.0005	-0.0126	0.0106	-0.0039 0.0003
		0.0005	0.3166	-0.0113		-0.0007 -0.0101
		-0.0126	-0.0113	0.3099		0.0179 -0.0068
		0.0106			0.3083	0.0075
		-0.0039	-0.0007	0.0179	0.0075	0.2988 -0.0004
		0.0003	-0.0101	-0.0068		-0.0004 0.2796

The designs for P1($\frac{1}{2}$, 1, $\frac{1}{2}$) and C($\frac{1}{2}$, 1, $\frac{1}{2}$) were such that all contrasts had mutually orthogonal estimates. The variance of the estimates of the main effects of L was 0.3333, the variance of the estimates of the main effects of N was 0.25 and the variance of the estimates of the interactions was 0.2857. The NOEC designs for P2($\frac{1}{2}$, 1, $\frac{1}{2}$) had the matrix $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ given in Table 3.12.9. A pleasing feature of this design is that the estimates of the contrasts with weight 1 have minimum variance and are mutually orthogonal to the estimates of the other contrasts.

Table 3.12.9. $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ for the P2($\frac{1}{2}$, 1, $\frac{1}{2}$) design

0.3277	0.0097			0.0029	0.0017	-0.0051	-0.0029
	0.3166			-0.0051	-0.0029	0.0088	0.0051
		0.2500					
			0.2500				
0.0029	-0.0051			0.2861	-0.0041	-0.0006	0.0072
0.0017	-0.0029			-0.0041	0.2908	0.0072	-0.0089
-0.0051	0.0088			-0.0006	0.0072	0.2867	-0.0124
-0.0029	0.0051			0.0072	-0.0089	-0.0124	0.3011

All three designs for weights ($\frac{1}{2}$, $\frac{1}{2}$, 1) were such that the estimates of all the contrasts were mutually orthogonal. The variance of the estimates of the contrasts with weight $\frac{1}{2}$ was 0.3333 and the variance of the estimates of the contrasts with weight 1 was 0.25.

Consider now the designs with weights that are made up of a 1, a $\frac{1}{2}$ and a 0. All three NOEC designs for weights (1, $\frac{1}{2}$, 0) had estimates of the contrasts with non-zero weights that were all mutually orthogonal and had minimum

variance. Further, these estimates were orthogonal to the estimates of the contrasts that had weight 0. That is, the design was optimal as it satisfied the requirements set at the start of the algorithm.

The designs P1(1, 0, $\frac{1}{2}$) and P2(1, 0, $\frac{1}{2}$) had estimates of the contrasts that were all mutually orthogonal. The variance of the estimates of the contrasts with non-zero weights was 0.25 and those with zero weights was 0.50.

All three NOEC designs for weights ($\frac{1}{2}$, 1, 0) were such that the estimates of the contrasts with non-zero weights had minimum variance and were mutually orthogonal. Further these estimates were orthogonal to the contrasts with weight 0. Therefore these designs are optimal for the requirements set.

All nine of the designs remaining in Table 3.12.1 were such that the estimates of all eight contrasts were mutually orthogonal. The three designs for weights ($\frac{1}{2}$, 0, 1) were such that the estimates of the contrasts with weight $\frac{1}{2}$ had variance 0.2857, the estimates of the contrasts with weight 0 had variance 0.40, and the estimates of the contrasts with weight 1 had variance 0.25.

The designs for weights (0, 1, $\frac{1}{2}$) were similar to those just described except the estimates of the contrasts with zero weight had variance 0.5 and all other contrasts had estimates with variance 0.25.

The designs P1(0, $\frac{1}{2}$, 1) and P2(0, $\frac{1}{2}$, 1) were also similar to those just described except the estimates of the contrasts with weight 0 had variance 0.4, the estimates of the contrasts with weight $\frac{1}{2}$ had variance 0.2857 and the estimates of the contrasts with weight 1 had variance 0.25. The design C(0, $\frac{1}{2}$, 1) was equal to the starting design

and as such the estimates of the contrasts with non-zero weights had variance 0.25 and those with weight zero had variance 0.50.

Whether or not the above designs are truely optimal for the requirements stated will be considered further in Section 3.14. What is apparent however, is that those designs which are of type NOEC and have no zero weights are not optimal as a design which is not of type NOEC exists that would satisfy the requirements. It would be an improvement to the algorithm that searches for optimal components if less designs of type NOEC occurred. A way of doing this is considered in the next section.

3.13 A modification of the optimality algorithm to improve designs of type NOEC

In the previous section it was noted that some designs found by the optimality algorithm for sets of non-zero weights were of type NOEC. These designs are less desirable because their analysis of variance is more difficult as the treatment sum of squares cannot be partitioned into a full set of orthogonal components. It would be useful therefore if such designs could be modified so that all off-diagonal elements of the matrix $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ were zero. The algorithm described in Section 3.4 can be easily altered for this purpose by changing the optimality criterion. Instead of searching for interchanges that minimize trace$(\underset{\sim}{L}'\underset{\sim H\sim}{\Omega}\underset{\sim\sim}{L}\omega^{\delta})$, the algorithm is altered to search for interchanges that minimize the sum of the absolute values of the off-diagonal elements of $\underset{\sim}{L}'\underset{\sim H\sim}{\Omega}\underset{\sim}{L}$, where $\underset{\sim H}{\Omega}$ is as defined in formula (3.3.21).

The modified algorithm with H equal to four was applied to all twenty-one NOEC designs described in the previous section and the three designs labelled as * in

Table 3.13.1. Types of design and values of the new criterion before and after the second minimization

Design	Criterion value before	Criterion value after	Type of design found
P1(1, 1, 1)	0.1574	0.0	OEC
C (1, 1, 1)	0.1574	0.0	OEC
P1(1, 1, 0)	1.4600	0.0	OEC
P2(1, 1, 0)	0.2488	0.2214	NOEC
C (1, 1, 0)	0.5520	0.0	OEC
P1(1, 0, 0)	1.6634	0.3066	NOEC
P2(1, 0, 0)	0.6188	0.1860	NOEC
C (1, 0, 0)	2.0168	0.0	OEC
P1(0, 1, 0)	0.6486	0.0916	NOEC
P2(0, 1, 0)	0.4800	0.1166	NOEC
P2(0, 0, 1)	0.1922	0.1922	NOEC
C (1, $\frac{1}{2}$, 1)	0.0582	0.0	OEC
C ($\frac{1}{2}$, 1, 1)	0.1400	0.0714	NOEC
P1(1, $\frac{1}{2}$, $\frac{1}{2}$)	0.1040	0.1040	NOEC
P2(1, $\frac{1}{2}$, $\frac{1}{2}$)	0.1928	0.0860	NOEC
C (1, $\frac{1}{2}$, $\frac{1}{2}$)	0.1644	0.0774	NOEC
P2($\frac{1}{2}$, 1, $\frac{1}{2}$)	0.1684	0.0	OEC
P1(1, $\frac{1}{2}$, 0)	1.4600	0.0	OEC
P2(1, $\frac{1}{2}$, 0)	0.2488	0.2214	NOEC
C (1, $\frac{1}{2}$, 0)	0.5520	0.0	OEC
C (1, 0, $\frac{1}{2}$)	0.4998	0.0	OEC
P1($\frac{1}{2}$, 1, 0)	1.4600	0.0	OEC
P2($\frac{1}{2}$, 1, 0)	0.3324	0.3258	NOEC
C ($\frac{1}{2}$, 1, 0)	0.5520	0.0	OEC

Table 3.12.1, which had components that could not be combined. It will be noted however, that not all of these twenty-one designs needs modification because the designs with zero weights, P1(1, 1, 0) for example, were optimal for the requirements set at the start of the algorithm. It is designs such as C(1, 1, 1) for example, that need modification. All twenty-four designs were included however, to test the performance of the modified algorithm. The results of this test are summarized in Table 3.13.1. There the new code OEC (Orthogonal Estimates of all Contrasts) is used to indicate the designs that were no longer of type NOEC after modification by the new algorithm.

Out of the twenty-one NOEC designs, ten were modified to be of type OEC and all except two of the remaining eleven had values of the non-zero covariances that were smaller. Also, and quite importantly, the three designs that had components that could not be combined were modified so that the combining of their components was possible. These designs it will be recalled were P1(1, 0, 0), C(1, 0, 0) and C(1, 0, $\frac{1}{2}$). Further, these designs were now optimal for the requirements set at the start of the algorithm as the estimates of the contrasts with non-zero weights had minimum variance, were mutually orthogonal and were orthogonal to the estimates of the contrasts with zero weights.

One unfortunate outcome was that the modified design for P1(1, 1, 0) was no longer optimal. The estimate of the linear main effect of N no longer had minimum variance, and this estimate was no longer orthogonal to the estimates of all the other contrasts. However, this design did not need modification as it was already optimal and so in practice this outcome would not have occurred. It does serve as a warning however, that it

is not always a good thing to tamper with a satisfactory design.

Before the second minimization procedure was employed there were eight NOEC designs that had no zero weights and were therefore the ones that actually needed modification. Of these eight the designs P1(1, 1, 1), C(1, 1, 1), C(1, $\frac{1}{2}$, 1) and P2($\frac{1}{2}$, 1, $\frac{1}{2}$) were now of type OEC. Of the remaining four, C($\frac{1}{2}$, 1, 1), P2(1, $\frac{1}{2}$, $\frac{1}{2}$) and C(1, $\frac{1}{2}$, $\frac{1}{2}$) had reduced values of the non-zero covariances, but P1(1, $\frac{1}{2}$, $\frac{1}{2}$) was not improved.

Therefore, overall, it appears that the modified version of the algorithm has been successful and is a useful addition to the procedure for searching for the components of the optimal design.

Those designs described above that had some zero weights and were such that the estimates of the contrasts with non-zero weights had minimum variance and were mutually orthogonal to all the other estimates of the contrasts were in fact optimal designs. These designs satisfied the requirements set at the start of the algorithm and had the additional feature of orthogonality. Most of the designs described above that had all non-zero weights looked as if they were optimal as the pattern of $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ satisfied the requirements set at the start of the algorithm. It is important, however, to try and determine if better designs exist so that it can be established whether or not these designs are in fact optimal. This is done in the next section.

3.14 Optimality of the designs described in the previous section

Here we consider the best designs found for each set of weights, having applied the modified version of the algorithm, as described in the previous section, where necessary. There were thirty-one designs which were certainly optimal as the

estimates of the contrasts with non-zero weights had minimum variance and were orthogonal to each other and the estimates of the contrasts with zero weights. These designs were P1(1, 1, 0), P2(1, 1, 0), C(1, 1, 0), P1(1, 0, 1), P2(1, 0, 1), C(1, 0, 1), P1(0, 1, 1), P2(0, 1, 1), C(0, 1, 1), P1(1, 0, 0), P2(1, 0, 0), C(1, 0, 0), P1(0, 1, 0), P2(0, 1, 0), C(0, 1, 0), P1(0, 0, 1), P2(0, 0, 1), C(0, 0, 1), P1(1, $\frac{1}{2}$, 0), P2(1, $\frac{1}{2}$, 0), C(1, $\frac{1}{2}$, 0), P1(1, 0, $\frac{1}{2}$), P2(1, 0, $\frac{1}{2}$), C(1, 0, $\frac{1}{2}$), P1($\frac{1}{2}$, 1, 0), P2($\frac{1}{2}$, 1, 0), C($\frac{1}{2}$, 1, 0), P1(0, 1, $\frac{1}{2}$), P2(0, 1, $\frac{1}{2}$), C(0, 1, $\frac{1}{2}$) and C(0, $\frac{1}{2}$, 1).

The remaining designs however may or may not be optimal. It depends if a better design can be found for a particular set of weights. As there is no theoretical method of deciding which design is certainly optimal for a given set of weights the following approach was adopted. An attempt was made to build up a design appropriate to a given set of weights using "bricks" made up of smaller block designs.

For the particular example of 9 treatments in 6 rows and 6 columns it is possible to suggest pairs of components for which the variances of the estimates have a pattern that reflects the pattern of the weights. Further, the designs so found are all of type OEC.

This is possible because the nine treatment combinations can be written as an array

$$A_1\ A_2\ A_3$$
$$B_1\ B_2\ B_3$$
$$C_1\ C_2\ C_3.$$

The contrasts between rows represent the main effects of "letters" and the contrasts between columns represent the main effects of "numbers". Of the eight degrees of freedom

four are left and correspond to the degrees of freedom for interactions. These can be divided into two sets of two. The first corresponding to the contrasts between (A_1, B_2, C_3), (A_2, B_3, C_1) and (A_3, B_1, C_2) and the second corresponding to the contrasts between (A_3, B_2, C_1), (A_2, B_1, C_3) and (A_1, B_3, C_2). As defined, the rows, columns and interactions contrasts are all mutually orthogonal.

Consider the following four block designs for 9 treatments in 3 blocks of 6.

1)

I	$A_1\ A_2\ A_3\ B_1\ B_2\ B_3$
II	$A_1\ A_2\ A_3\ C_1\ C_2\ C_3$
III	$B_1\ B_2\ B_3\ C_1\ C_2\ C_3$

2)

I	$A_1\ B_1\ C_1\ A_2\ B_2\ C_2$
II	$A_1\ B_1\ C_1\ A_3\ B_3\ C_3$
III	$A_2\ B_2\ C_2\ A_3\ B_3\ C_3$

3)

I	$A_1\ B_2\ C_3\ A_2\ B_3\ C_1$
II	$A_1\ B_2\ C_3\ A_3\ B_1\ C_2$
III	$A_2\ B_3\ C_1\ A_3\ B_1\ C_2$

4)

I	$A_3\ B_2\ C_1\ A_2\ B_1\ C_3$
II	$A_3\ B_2\ C_1\ A_1\ B_3\ C_2$
III	$A_2\ B_1\ C_3\ A_1\ B_3\ C_2$

In design 1) the linear and quadratic contrasts between the letters as defined in Table 3.9.1 are estimated with less than full efficiency. In design 2) the linear and quadratic effects between the numbers are estimated with less than full efficiency. When designs 3) and 4) are taken together the four interaction contrasts are estimated with less than full efficiency. Therefore, for convenience the four designs are labelled respectively, as L, N, I_1 and I_2, where the letters indicate which contrasts are estimated with less than full efficiency. The label L, for example, refers to the fact that the main effects of letters are estimated with less than full efficiency. The subscripts are used to differentiate between designs 3) and 4).

These four "bricks" can now be used to build up the

components of a row-and-column design. As the efficiency loss of a contrast in the row-and-column design can be calculated by adding the efficiency losses of the contrast over the bricks, it is relatively easy to decide which bricks to use to obtain a particular pattern of $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ in the row-and-column design. For example, to estimate the main effects of letters and all four interactions more efficiently than the main effects of numbers, a row-and-column design with each component made up of two type N bricks is required. In all, nine different pairs of components that lead to different patterns in $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ can be chosen. These are given in Table 3.14.1, along with the type of each design so formed. The codes L_ℓ, L_q etc. are as defined in Table 3.9.1.

Table 3.14.1. The nine different pairs of components and corresponding contrast variances

Components	Type	Contrast variances		
		L_ℓ and L_q	N_ℓ and N_q	All four interactions
NL/I_1I_2	O:FF(T)	0.2857	0.2857	0.2857
LL/LL	O:GG(G)	0.5000	0.2500	0.2500
NN/NN	O:GG(G)	0.2500	0.5000	0.2500
NL/NL	O:FF(F)	0.3333	0.3333	0.2500
I_1I_2/I_1I_2	O:FF(F)	0.2500	0.2500	0.3333
LL/I_1I_2	O:GF(F)	0.3333	0.2500	0.2857
NN/I_1I_2	O:GF(F)	0.2500	0.3333	0.2857
LL/LN	O:GF(F)	0.4000	0.2857	0.2500
NN/NL	O:GF(F)	0.2857	0.4000	0.2500

If each of the nineteen different sets of weights is considered in turn and the design in Table 3.14.1 that gives the minimum value of trace$(\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim\sim}{L}\omega^\delta)$ is determined, then the optimal design for that set of weights is obtained. The

results of doing this are summarized in Table 3.14.2, where for each set of weights the optimal design is given. For some sets of weights more than one of the designs given in Table 3.14.1 gives the same minimum value for the trace. For convenience only one of them has been included in Table 3.14.2.

Table 3.14.2. Optimal designs for each set of weights

Weights	Components	Contrast variances		
		L_ℓ and L_q	N_ℓ and N_q	All four interactions
1 1 1	NL/I_1I_2	0.2857	0.2857	0.2857
1 1 0	I_1I_2/I_1I_2	0.2500	0.2500	0.3333
1 0 1	NN/NN	0.2500	0.5000	0.2500
0 1 1	LL/LL	0.5000	0.2500	0.2500
1 0 0	NN/NN	0.2500	0.5000	0.2500
0 1 0	LL/LL	0.5000	0.2500	0.2500
0 0 1	LL/LL	0.5000	0.2500	0.2500
1 1 $\frac{1}{2}$	I_1I_2/I_1I_2	0.2500	0.2500	0.3333
1 $\frac{1}{2}$ 1	NN/NL	0.2857	0.4000	0.2500
$\frac{1}{2}$ 1 1	LL/LN	0.4000	0.2857	0.2500
1 $\frac{1}{2}$ $\frac{1}{2}$	NN/I_1I_2	0.2500	0.3333	0.2857
$\frac{1}{2}$ 1 $\frac{1}{2}$	LL/I_1I_2	0.3333	0.2500	0.2857
$\frac{1}{2}$ $\frac{1}{2}$ 1	NL/NL	0.3333	0.3333	0.2500
1 $\frac{1}{2}$ 0	I_1I_2/I_1I_2	0.2500	0.2500	0.3333
1 0 $\frac{1}{2}$	NN/NN	0.2500	0.5000	0.2500
$\frac{1}{2}$ 1 0	I_1I_2/I_1I_2	0.2500	0.2500	0.3333
$\frac{1}{2}$ 0 1	NN/NN	0.2500	0.5000	0.2500
0 1 $\frac{1}{2}$	LL/LL	0.5000	0.2500	0.2500
0 $\frac{1}{2}$ 1	LL/LL	0.5000	0.2500	0.2500

Consider now the designs obtained by the algorithm. There were twenty-six designs for which it was not certain

Table 3.14.3. Variances of the designs obtained by the algorithm

Starting Design	Contrast variances L_ℓ and L_q	N_ℓ and N_q	All four interactions
P1(1, 1, 1)	0.2857	0.2857	0.2857
P2(1, 1, 1)			
C (1, 1, 1)			
P1(1, 1, $\frac{1}{2}$)	0.2500	0.2500	0.3333
P2(1, 1, $\frac{1}{2}$)			
C (1, 1, $\frac{1}{2}$)			
P1(1, $\frac{1}{2}$, 1)	0.2857	0.4000	0.2500
P2(1, $\frac{1}{2}$, 1)			
C (1, $\frac{1}{2}$, 1)			
P1($\frac{1}{2}$, 1, 1)	0.4000	0.2857	0.2500
P2($\frac{1}{2}$, 1, 1)			
P1($\frac{1}{2}$, 1, $\frac{1}{2}$)	0.3333	0.2500	0.2857
P2($\frac{1}{2}$, 1, $\frac{1}{2}$)			
C ($\frac{1}{2}$, 1, $\frac{1}{2}$)			
P1($\frac{1}{2}$, $\frac{1}{2}$, 1)	0.3333	0.3333	0.2500
P2($\frac{1}{2}$, $\frac{1}{2}$, 1)			
C ($\frac{1}{2}$, $\frac{1}{2}$, 1)			
P1($\frac{1}{2}$, 0, 1)	0.2857	0.4000	0.2500
P2($\frac{1}{2}$, 0, 1)			
C ($\frac{1}{2}$, 0, 1)			
P1(0, $\frac{1}{2}$, 1)	0.4000	0.2857	0.2500
P2(0, $\frac{1}{2}$, 1)			

Table 3.14.4. $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ for C($\frac{1}{2}$, 1, 1)

$$\begin{bmatrix}
0.4000 & & & & & & & & \\
 & 0.4000 & & & & & & & \\
 & & 0.2500 & & & & & & \\
 & & & 0.2500 & & & & & \\
 & & & & 0.2679 & & & & 0.0179 \\
 & & & & & 0.2679 & -0.0179 & & \\
 & & & & & -0.0179 & 0.2679 & & \\
 & & & & 0.0179 & & & & 0.2679
\end{bmatrix}$$

Table 3.14.5. $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ for P2(1, $\frac{1}{2}$, $\frac{1}{2}$)

$$\begin{bmatrix}
0.2500 & & & & & & & & \\
 & 0.2500 & & & & & & & \\
 & & 0.3529 & & & & & & \\
 & & & 0.2857 & & & & & \\
 & & & & 0.2806 & & & & 0.0216 \\
 & & & & & 0.3094 & -0.0216 & & \\
 & & & & & -0.0216 & 0.2806 & & \\
 & & & & 0.0216 & & & & 0.3094
\end{bmatrix}$$

Table 3.14.6. $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ for C(1, $\frac{1}{2}$, $\frac{1}{2}$)

$$\begin{bmatrix}
0.2500 & & & & & & & & \\
 & 0.2500 & & & & & & & \\
 & & 0.3014 & & & & -0.0194 & & \\
 & & & 0.3014 & & & & & -0.0194 \\
 & & & & 0.3333 & & & & \\
 & & & & & 0.3333 & & & \\
 & & -0.0194 & & & & 0.2740 & & \\
 & & & -0.0194 & & & & & 0.2740
\end{bmatrix}$$

if they were optimal. The values of the estimates of the eight treatment contrasts obtained using each of the twenty-two OEC designs out of these twenty-six are given in Table 3.14.3. The remaining four designs $C(\frac{1}{2}, 1, 1)$, $P1(1, \frac{1}{2}, \frac{1}{2})$, $P2(1, \frac{1}{2}, \frac{1}{2})$ and $C(1, \frac{1}{2}, \frac{1}{2})$ were not optimal because the matrix $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ for each of these designs had some non-zero off-diagonal elements. However, these designs were near optimal. The matrix $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ for each of designs $C(\frac{1}{2}, 1, 1)$, $P2(1, \frac{1}{2}, \frac{1}{2})$ and $C(1, \frac{1}{2}, \frac{1}{2})$ is given in Tables 3.14.4 to 3.14.6, respectively. The matrix $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ for $P1(1, \frac{1}{2}, \frac{1}{2})$ was given in Table 3.12.6.

Comparing Tables 3.14.2 and 3.14.3 indicates that only five designs were sub-optimal. These designs, $P1(\frac{1}{2}, 0, 1)$, $P2(\frac{1}{2}, 0, 1)$, $C(\frac{1}{2}, 0, 1)$, $P1(0, \frac{1}{2}, 1)$ and $P2(0, \frac{1}{2}, 1)$, were however only slightly sub-optimal. The value of $\text{trace}(\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim\sim}{L}\omega^{\delta})$ for all five designs was 1.2857 as compared with the optimum value of 1.2500.

The final conclusion of this chapter is therefore that the algorithm described in Section 3.4 that searches for the components, supplemented when necessary by the modified algorithm described in Section 3.13, usually leads to optimum designs. Those local optima that are found are only slightly worse than the true optima.

It is clear that the algorithms that search for optimal designs, that were described in this chapter and the previous one, give very satisfactory results. On the relatively few occasions that the algorithms get trapped at a local optimum, the design obtained is only slightly worse than the true optimum. One slight disadvantage of these algorithms, however, is that the treatment replications have to be chosen before the algorithms can be applied and remain fixed throughout the

course of the algorithms. Two new algorithms that permit the replications to vary are described in the next chapter.

Chapter 4. Exchange and interchange procedures to search for optimal designs.

4.1 Introduction

Two disadvantages of the algorithms described in Chapters 2 and 3 are that the treatment replications have to be chosen first and remain fixed, and an approximation to $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ is used. A further disadvantage of the algorithm that searches for row-and-column designs is the need to combine the row and column components. In this chapter a more general algorithm that can be applied to both block designs and row-and-column designs is described. The treatment replications are allowed to vary, the true value of $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$ is used and for row-and-column designs the need to combine components is eliminated.

The initial idea for this new algorithm stems from "exchange" algorithms such as Mitchell's DETMAX (1974) that search for D_n-optimal designs. In DETMAX, for example, the design points eligible for inclusion in the design are specified in advance as a set of "candidates". An initial design of the required size is then chosen as a starting point. Mitchell (1974), for example, suggested choosing a random selection of candidate points for inclusion in the starting design. Once an initial design has been chosen attempts are made to improve it by removing one point from the design and replacing it by a point from the set of candidates. When searching for D_n-optimal designs the best point to remove and the best point to replace it can be determined quite easily. This exchanging of points continues until no further improvement in the design can be made.

Part of the work done in this chapter was done in collaboration with Dr. J.A. Eccleston of the University of

New South Wales who was visiting the University of Kent during the academic year 1978-1979. The basic structure of the new algorithm was conceived independently by each of us. The main contribution of Dr. Eccleston is a set of theorems (Eccleston, 1979), which he proved before his visit, with a different purpose in mind. These theorems are necessary for the "exchange procedure" described below. After testing this procedure, joint discussion led to the "interchange procedure" also described below.

Before describing the algorithm it is necessary to describe some recursive formulae. This is done in the next section.

4.2 Recursive formulae

The recursive formulae given here will first be described for the general linear model and then for the block design and the row-and-column design.

The linear model (2.2.1) as defined in Chapter 2 is written as

$$y_i = \underset{\sim}{x}_i'\underset{\sim}{\beta} + \varepsilon_i, \quad i = 1, 2, \ldots, \tag{4.2.1}$$

where y_i is the i^{th} observation, $\underset{\sim}{x}_i$ is a $p \times 1$ vector of known values and ε_i is an independent random variable with mean zero and variance σ^2. Further, $\underset{\sim}{Y}_i' = (y_1, y_2, \ldots, y_i)$, $\underset{\sim}{\varepsilon}_i = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_i)$ and $\underset{\sim}{X}_i' = (\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_i)$. It is assumed that there are $q < p$ estimable functions between the elements of $\underset{\sim}{\beta}$, that is the model does not have full rank. For convenience let $\underset{\sim}{x}_j'$, $j = 1, 2, \ldots, q$ be chosen so that $\underset{\sim}{X}_q$ has rank q and hence the remaining $\underset{\sim}{x}_i'$, $i = q+1, q+2, \ldots$ are in the row space of $\underset{\sim}{X}_q$. Eccleston (1979) gave the following theorems, where $(\underset{\sim}{X}_i'\underset{\sim}{X}_i)^-$ is a generalized inverse of $(\underset{\sim}{X}_i'\underset{\sim}{X}_i)$. That is, $(\underset{\sim}{X}_i'\underset{\sim}{X}_i)(\underset{\sim}{X}_i'\underset{\sim}{X}_i)^-(\underset{\sim}{X}_i'\underset{\sim}{X}_i) = (\underset{\sim}{X}_i'\underset{\sim}{X}_i)$. These theorems relate values obtained after the $(i + 1)^{th}$ observation has been added to the design to the values obtained from the design containing

i observations.

Theorem 4.2.1

$$(\underset{\sim}{X}_{i+1}'\underset{\sim}{X}_{i+1})^{-} = (\underset{\sim}{X}_i'\underset{\sim}{X}_i)^{-} - \frac{(\underset{\sim}{X}_i'\underset{\sim}{X}_i)^{-}\underset{\sim}{x}_{i+1}\underset{\sim}{x}_{i+1}'(\underset{\sim}{X}_i'\underset{\sim}{X}_i)^{-}}{1 + \underset{\sim}{x}_{i+1}'(\underset{\sim}{X}_i'\underset{\sim}{X}_i)^{-}\underset{\sim}{x}_{i+1}}.$$

If interest only centres on a subset of the parameters the model (4.2.1) can be partitioned as follows.

$$\underset{\sim}{\beta} = \begin{bmatrix} \underset{\sim}{\alpha} \\ \hline \underset{\sim}{\theta} \end{bmatrix}, \quad \underset{\sim}{x}_i' = (\underset{\sim}{a}_i' \,\vdots\, \underset{\sim}{b}_i') \text{ and } \underset{\sim}{X}_i = (\underset{\sim}{A}_i \,\vdots\, \underset{\sim}{B}_i).$$

Here $\underset{\sim}{\alpha}$ is an $s \times 1$ vector of parameters of interest, and not all of which are estimable. The vector $\underset{\sim}{a}_i'$ is $s \times 1$ and $\underset{\sim}{A}_i' = (\underset{\sim}{a}_1, \underset{\sim}{a}_2, \ldots, \underset{\sim}{a}_i)$, $\underset{\sim}{b}_i'$ and $\underset{\sim}{B}_i$ are therefore also defined. Let the partition be such that $\underset{\sim}{a}_{i+1}'$ and $\underset{\sim}{b}_{i+1}'$ are in the row spaces of $\underset{\sim}{A}_i$ and $\underset{\sim}{B}_i$, $i = q, q+1, \ldots$, respectively. This will be true for block designs and row-and-column designs.

For the first i observations the reduced normal equations for the effect of $\underset{\sim}{\alpha}$ adjusted for the effect of $\underset{\sim}{\theta}$ are

$$\underset{\sim}{C}_i\hat{\underset{\sim}{\alpha}}_i = \underset{\sim}{Q}_i, \quad i = q, q+1, \ldots, \tag{4.2.2}$$

where

$$\underset{\sim}{C}_i = \underset{\sim}{A}_i'\underset{\sim}{A}_i - \underset{\sim}{A}_i'\underset{\sim}{B}_i(\underset{\sim}{B}_i'\underset{\sim}{B}_i)^{-}\underset{\sim}{B}_i'\underset{\sim}{A}_i$$

and

$$\underset{\sim}{Q}_i = \underset{\sim}{A}_i'\underset{\sim}{y}_i - \underset{\sim}{A}_i'\underset{\sim}{B}_i(\underset{\sim}{B}_i'\underset{\sim}{B}_i)^{-}\underset{\sim}{B}_i'\underset{\sim}{y}_i.$$

A solution to (4.2.2) is $\hat{\underset{\sim}{\alpha}}_i = \underset{\sim}{C}_i^{-}\underset{\sim}{Q}_i$, where $\underset{\sim}{C}_i^{-}$ is a generalized inverse of $\underset{\sim}{C}_i$.

Theorem 4.2.2

$$\underset{\sim}{C}_{i+1} = \underset{\sim}{C}_i + \underset{\sim}{c}_{i+1}\underset{\sim}{c}_{i+1}',$$

where

$$\underset{\sim}{c}_{i+1} = \frac{\underset{\sim}{a}_{i+1} - \underset{\sim}{A}_i'\underset{\sim}{B}_i(\underset{\sim}{B}_i'\underset{\sim}{B}_i)^{-}\underset{\sim}{b}_{i+1}}{(1 + \underset{\sim}{b}_{i+1}'(\underset{\sim}{B}_i'\underset{\sim}{B}_i)^{-}\underset{\sim}{b}_{i+1})^{\frac{1}{2}}}, \quad i = q, q+1, \ldots$$

Theorem 4.2.3

$$\underset{\sim}{C}^{-}_{i+1} = \underset{\sim}{C}^{-}_{i} - \frac{\underset{\sim}{C}^{-}_{i}\underset{\sim}{c}_{i+1}\underset{\sim}{c}'_{i+1}\underset{\sim}{C}^{-}_{i}}{(1 + \underset{\sim}{c}'_{i+1}\underset{\sim}{C}^{-}_{i}\underset{\sim}{c}_{i+1})}, \quad i = q, q+1, \ldots$$

Similar theorems to the above can be given for the case where an observation is removed from a design.

In the next section an algorithm to search for optimal block designs will be described. The results quoted in the above theorems will be used in the algorithm.

4.3 Exchange and interchange procedures to search for optimal block designs

As defined in Chapter 2 the block design has size $n = \sum_{j=1}^{b} k_j = \sum_{i=1}^{v} r_i$. The incidence matrix for this design is $\underset{\sim}{N}_n$, the diagonal matrix of treatment replications is $\underset{\sim}{r}^{\delta}_n$ and the diagonal matrix of block sizes is $\underset{\sim}{k}^{\delta}_n$. The subscript (n) refers to the size of the design. Then,

$$\underset{\sim}{C}_n = \underset{\sim}{r}^{\delta}_n - \underset{\sim}{N}_n\underset{\sim}{k}^{-\delta}_n\underset{\sim}{N}'_n.$$

If an observation is removed from the design then, using a slightly simplified notation to that used in the previous section,

$$\underset{\sim}{C}_{n-1} = \underset{\sim}{C}_n - \underset{\sim}{c}_1\underset{\sim}{c}'_1,$$

where

$$\underset{\sim}{c}_1 = \frac{\underset{\sim}{a}_1 - \underset{\sim}{N}_{n-1}\underset{\sim}{k}^{-\delta}_{n-1}\underset{\sim}{b}_1}{(1 + \underset{\sim}{b}'_1\underset{\sim}{k}^{-\delta}_{n-1}\underset{\sim}{b}_1)^{\frac{1}{2}}}.$$

The $\underset{\sim}{a}_1$ is a $v \times 1$ vector of zeros except for a one in the position corresponding to the deleted treatment and $\underset{\sim}{b}_1$ is a $b \times 1$ vector of zeros except for a one in the position corresponding to the block from which the treatment has been deleted. The depleted block design of n - 1 observations has new incidence matrix $\underset{\sim}{N}_{n-1}$ and a new matrix of block sizes $\underset{\sim}{k}^{-\delta}_{n-1}$.

If an observation is added to the depleted design a new $\underset{\sim}{C}_n$ is obtained as

$$\text{new } \underset{\sim}{C}_n = \underset{\sim}{C}_{n-1} + \underset{\sim}{c}_2\underset{\sim}{c}_2',$$

where

$$\underset{\sim}{c}_2 = \frac{\underset{\sim}{a}_2 - \underset{\sim}{N}_{n-1}\underset{\sim}{k}_{n-1}^{-\delta}\underset{\sim}{b}_2}{(1 + \underset{\sim}{b}_2'\underset{\sim}{k}_{n-1}^{-\delta}\underset{\sim}{b}_2)^{\frac{1}{2}}}$$

and $\underset{\sim}{a}_2$ and $\underset{\sim}{b}_2$ characterize the new observation in the same way that $\underset{\sim}{a}_1$ and $\underset{\sim}{b}_1$ did the deleted observation. As the number of blocks and their sizes are kept fixed the vectors $\underset{\sim}{b}_2$ and $\underset{\sim}{b}_1$ are the same.

The generalized inverses of the above matrices can also be specified as

$$\underset{\sim}{C}_{n-1}^{-} = \underset{\sim}{C}_n^{-} + \frac{\underset{\sim}{C}_n^{-}\underset{\sim}{c}_1\underset{\sim}{c}_1'\underset{\sim}{C}_n^{-}}{1 - \underset{\sim}{c}_1'\underset{\sim}{C}_n^{-}\underset{\sim}{c}_1}$$

and

$$\text{new } \underset{\sim}{C}_n^{-} = \underset{\sim}{C}_{n-1}^{-} - \frac{\underset{\sim}{C}_{n-1}^{-}\underset{\sim}{c}_2\underset{\sim}{c}_2'\underset{\sim}{C}_{n-1}^{-}}{1 + \underset{\sim}{c}_2'\underset{\sim}{C}_{n-1}^{-}\underset{\sim}{c}_2}.$$

Consider now an algorithm to search for optimal block designs. The criterion used to decide if a design is optimal will be that used in Chapter 2, except that rather than specify any particular generalized inverse of $\underset{\sim}{C}_n$ the criterion is written as

$$\text{minimize trace } (\underset{\sim}{L}'\underset{\sim}{C}_n^{-}\underset{\sim}{L}\underset{\sim}{\omega}^{\delta}).$$

Taking up the ideas of the exchange algorithms that search for D_n-optimal designs mentioned in Section 4.1, the weakest observation in a block design is defined to be that one for which

$$\text{trace } (\underset{\sim}{L}\underset{\sim}{\omega}^{\delta}\underset{\sim}{L}'\underset{\sim}{C}_{n-1}^{-}) - \text{trace } (\underset{\sim}{L}\underset{\sim}{\omega}^{\delta}\underset{\sim}{L}'\underset{\sim}{C}_n^{-})$$

is a minimum for all possible $\underset{\sim}{C}_{n-1}^{-}$ (i.e. all $\underset{\sim}{c}_1$). This means that the observation which is of "least importance" is termed

the weakest. This observation is deleted and replaced by the strongest which is defined to be the one for which

$$\text{trace } (L\underset{\sim}{\omega}^{\delta}L'\underset{\sim}{C}_{n-1}^{-}) - \text{trace } \{L\underset{\sim}{\omega}^{\delta}L'(\text{new } \underset{\sim}{C}_{n})^{-}\}$$

is positive and a maximum for all possible $\underset{\sim}{c}_2$.

These operations are relatively straightforward for a block design. For instance, if treatment i in block j is exchanged for treatment ℓ then

$$\underset{\sim}{c}_1 = \frac{1}{[k_j(k_j - 1)]^{\frac{1}{2}}} \begin{bmatrix} -n_{1j} \\ \vdots \\ -n_{ij} + k_j \\ \vdots \\ -n_{vj} \end{bmatrix}$$

and

$$\underset{\sim}{c}_2 = \frac{1}{[k_j(k_j - 1)]^{\frac{1}{2}}} \begin{bmatrix} -n_{1j} \\ \vdots \\ -n_{\ell j} + k_j - 1 \\ \vdots \\ -n_{ij} + 1 \\ \vdots \\ -n_{vj} \end{bmatrix},$$

where n_{ij} is the i,j^{th} element of $\underset{\sim}{N}_n$.

In the following, the exchanging of treatments in the above manner will be referred to as the exchange procedure (EP). A simple algorithm based on the exchange procedure is at each step to exchange the weakest observation for the strongest. If at any step the exchange procedure does not lead to an improved design, that is a smaller trace $(L\underset{\sim}{\omega}^{\delta}L'\underset{\sim}{C}_{n}^{-})$, then the next weakest observation is considered and so on. The procedure continues until no further improvement in the design can be obtained. At this point any alteration in the current replication scheme does not improve the design. The EP has

produced what might be termed the "correct" replications.

However, for some tests of the EP where the optimal design was known it was found that although the correct replications were found the final design could still be improved upon. Therefore, it was necessary to rearrange the treatments between blocks to obtain the optimal set of treatment concurrences. A good way of doing this is the interchange procedure (IP) described below.

The IP interchanges a pair of observations by first exchanging the weakest observation, a replicate of treatment ℓ say, with the strongest, a replicate of treatment m say. Secondly, to maintain the "correct" replications, the weakest observation involving an m is replaced by an ℓ and so on. When an improvement is achieved that interchange is made and the IP restarts. If there is no improvement for all m then the procedure restarts with the next weakest observation other than the particular ℓ and so on. The procedure terminates when no interchange can be found that improves the design.

The recursive formulae given above are used to determine the change in $\underset{\sim}{C}_n^-$ effected by an exchange or interchange. It is worth noting that $\underset{\sim}{C}_n^-$ is calculated directly only once and all other $\underset{\sim}{C}_n^-$'s and $\underset{\sim}{C}_{n-1}^-$'s are obtained by using the recursive formulae.

The initial design can be determined using the methods described in Section 2.5 of Chapter 2, and the formula (2.4.2) for $\underset{\sim}{\Omega}$ can be used to give the initial generalized inverse $\underset{\sim}{C}_n^-$.

A new algorithm to search for optimal block designs is therefore to first apply the EP to a given starting design to search for the "correct" replications. Then apply the IP to search for optimal treatment concurrences.

Some tests of the performance of the above algorithms

are given below. When running a computer program, rounding errors may make it difficult to decide precisely if one trace is smaller than another. Therefore, in the computerised version of the EP and IP, a pre-set tolerance value, ε, is used. An improved design is deemed to have been obtained if T_{old}, the value of the trace before an exchange or interchange is made is such that $(T_{old}-T_{new}) \geq \varepsilon$, where T_{new} is the value of the trace after the exchange or interchange. Experience has shown that the choice of ε affects the value of the trace of the final design.

Test 1

Here $v = 5$, $b = 6$, $k_1 = k_2 = 4$ and $k_3 = k_4 = k_5 = k_6 = 2$, and there is an equal interest in all treatment differences. The value of ε was 0.00001.

The procedure for generating the starting design gave

Block	
I	2 3 4 5
II	1 3 4 5
III	2 5
IV	1 5 ,
V	3 4
VI	1 2

which has $\underset{\sim}{r}' = (3, 3, 3, 3, 4)$ and $T = \text{trace}\,(L\underset{\sim\sim}{\omega}^{\delta}L'\underset{\sim}{C}_n^{-}) = 1.7278$.

The EP gave the following design

Block	
I	2 3 4 5
II	1 3 4 5
III	2 5
IV	1 3 ,
V	1 4
VI	1 2

which has $\underset{\sim}{r}' = (4, 3, 3, 3, 3)$ and $T = 1.6602$.

The IP gave the following design

Block

I	2 3 4 5
II	2 3 4 5
III	1 2
IV	1 3 ,
V	1 5
VI	1 4

which has T = 1.6000.

This design is totally balanced and therefore optimal.

To test if the correct replications would be determined by the EP, the following deliberately perverse starting design was used.

Block

I	5 5 5 5
II	5 5 5 5
III	5 4
IV	5 3 .
V	5 2
VI	5 1

This design has replications $\underset{\sim}{r}' = (1, 1, 1, 1, 12)$ and T = 6.4000.

The EP gave the optimal design given above with T = 1.6000.

Test 2

This was a repetition of a test used in Chapter 2. Here $v = 5$, $b = 3$, $k_1 = 6$, $k_2 = 12$, $k_3 = 18$,

$$\underset{\sim}{L}' = (\tfrac{1}{2})^{\frac{1}{2}} \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix} \quad \text{and} \quad \underset{\sim}{\omega}^{\delta} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The value of ε was 0.00001, and the starting design generated by the procedure described in Chapter 2 was

Block

I	1 1 2 3 4 5
II	1 1 1 1 2 2 3 3 4 4 5 5
III	1 1 1 1 1 1 2 2 2 3 3 3 4 4 4 5 5 5.

This design is optimal as noted in Chapter 2.

Therefore, as done in Chapter 2, the deliberately perverse starting design given below was used.

Block

I	2 3 3 3 3 3
II	3 4 4 4 4 4 4 5 5 5 5 5
III	1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 5.

The EP followed by the IP gave the optimal design noted above.

To test again the ability of the EP to determine the correct replications, the following perverse starting design was used, which has $\underset{\sim}{r}' = (1, 1, 1, 1, 32)$.

Block

I	1 4 5 5 5 5
II	2 5 5 5 5 5 5 5 5 5 5 5
III	3 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5.

The EP gave the correct replications, i.e. $\underset{\sim}{r}' = (12, 6, 6, 6, 6)$ and the optimal design.

Test 3

This was also a repetition of a test used in Chapter 2. The designs considered were those described by Freeman (1976). Here $v = 3$, $b = 4$, $k_i = 5$, $i = 1, 2, 3$ and 4,

$$\underset{\sim}{L}' = \begin{matrix} (\frac{1}{2})^{\frac{1}{2}} \\ (\frac{1}{6})^{\frac{1}{2}} \end{matrix} \begin{bmatrix} 1 & 0 & -1 \\ -1 & 2 & -1 \end{bmatrix} \quad \text{and} \quad \underset{\sim}{\omega}^{\delta} = \begin{bmatrix} 1-w_1 & 0 \\ 0 & w_1 \end{bmatrix}.$$

In the test described in Chapter 2 the replications were fixed at either (6, 8, 6), (7, 6, 7), (8, 4, 8) or (9, 2, 9) and slightly perverse starting designs, labelled as W, X, Y and Z, respectively, were used to search for designs for all values of w_1 in the range 0, (0.01), 1. The final designs found there agreed with those obtained analytically by Freeman. The design labelled as A in Chapter 2 was optimal for w_1 in the range (0.531, 1), the design B for w_1 in the range (0.346, 0.531),

the design C for w_1 in the range (0.052, 0.346) and the design D for w_1 in the range (0, 0.052).

To test the EP and IP the five different starting designs were used with all weights in the range 0, (0.1), 1.0. The starting designs were the ones obtained using the procedure described in Chapter 2, and designs W, X, Y and Z respectively, as defined in Test 2 in Chapter 2. For each value of the weight w_1 all five different starting designs gave equivalent final designs, although the treatment replications of the final design sometimes differed for $w_1 = 1.0$. Here two different sets of treatment replications, (5, 10, 5) and (6, 8, 6), gave equivalent designs. When $w_1 = 0$ the EP indicated that treatment 2 should have zero replication. A summary of the final designs obtained using all five different starting designs and all weights in the range 0.1, (0.1), 1.0 are given in Table 4.3.1. When $w_1 = 0$ the computer program stopped after the EP because at the start of the EP there were three treatments but at the end there were only two. As the program still assumed there were three treatments it had to be stopped. When restarted with only two treatments the optimal design was obtained. Here again ε was set equal to 0.00001.

Table 4.3.1. Final designs found using the five different starting designs

w_1	Final replications	Final trace
1.0	(5, 10, 5) or (6, 8, 6)	0.1389
0.9	(6, 8, 6)	0.1429
0.8	(6, 8, 6)	0.1468
0.7	(6, 8, 6)	0.1508
0.6	(6, 8, 6)	0.1548
0.5	(7, 6, 7) or (6, 7, 7)	0.1569
0.4	(7, 6, 7)	0.1549
0.3	(8, 4, 8)	0.1500
0.2	(8, 4, 8)	0.1417
0.1	(8, 4, 8)	0.1333

All the designs were optimal and it is interesting to note that a design with replication (5, 10, 5) was obtained that was equivalent to Freeman's design A which had replication (6, 8, 6).

To determine if the EP and IP would determine the optimal designs close to the boundary values of the weights, i.e. w_1 = 0.531, 0.346 and 0.052, the above tests were repeated for the following values of w_1: 0.04, 0.05, 0.06, 0.33, 0.34, 0.35, 0.52, 0.53 and 0.54. The final designs obtained are summarized in Table 4.3.2, where all five starting designs led to equivalent designs for each value of w_1.

Table 4.3.2. Final designs found for weights near the boundaries

w_1	Final replications	Final trace	Freeman's optimal trace
0.54	(7, 7, 6) or (6, 7, 7)	0.1565	0.1572
0.53	(7, 7, 6) or (6, 7, 7)	0.1566	0.1575
0.52	(7, 7, 6) or (6, 7, 7)	0.1567	0.1573
0.35	(8, 5, 7) or (7, 5, 8)	0.1531	0.1539
0.34	(8, 5, 7) or (7, 5, 8)	0.1526	0.1533
0.33	(8, 5, 7) or (7, 5, 8)	0.1521	0.1525
0.06	(9, 3, 8)	0.1286	0.1300
0.05	(9, 3, 8) or (8, 3, 9)	0.1270	0.1288
0.04	(9, 3, 8) or (8, 3, 9)	0.1254	0.1258

It is apparent that near the boundary values of w_1 the design obtained is usually a mixture of the two designs either side of the boundary. Perhaps this is not surprising. It would be very unusual for an experimenter to give the value of a weight correct to two decimal places. It will be noted that all the designs found by the EP and IP have a smaller trace than the 'optimal' designs given by Freeman, and so are to be preferred on the trace criterion. The designs considered by Freeman, however, were constrained so that the replications of

treatments 1 and 3 were equal. It is seen from the above that if this condition is relaxed better designs in terms of the trace criterion are obtained. However, most of the designs obtained by the EP and IP were such that the off-diagonal elements of $\underset{\sim}{L}'\bar{C}\underset{\sim}{L}$ were equal to -0.01 or 0.01.

Test 4

Here an attempt was made to generate the 209 optimal block designs given by Mitchell and John (1976), (henceforth referred to as M. and J.), which have equal treatment replications and equal block sizes. In these designs the number of treatments varies from two to twelve, and the block size varies from two to ten. When ε was set equal to 0.00001 161 of the 209 optimal designs were found. When ε was decreased to 0.00000001 21 further optimal designs were found, and when ε was decreased to zero a further 10 optimal designs were found. In fact the efficiencies differed by 0.0015 or less, except for (v, k, r) = (11, 2, 6) which had an efficiency 0.003 less than the M. and J. design. The efficiency is the average variance of all pairs of treatment differences, divided into the minimum which would be achieved by a randomized block design if it existed.

The efficiency can be expressed as the ratio

$$\text{Efficiency} = \frac{\frac{2\sigma^2}{r}}{\frac{2\sigma^2}{(v-1)}\left[\frac{1}{c} - \frac{1}{\lambda_{max}}\right]}$$

$$= \frac{v-1}{r\left(\frac{1}{c} - \frac{1}{\lambda_{max}}\right)},$$

where

$$c = \left(\sum_{i=1}^{v-1} \frac{1}{\lambda_i} \right)^{-1},$$

λ_i are the non-zero eigenvalues of $\underset{\sim}{\Omega}^{-1}$, as defined in equation 2.3.8, in Chapter 2, and λ_{max} is the maximum non-zero eigenvalue of $\underset{\sim}{\Omega}^{-1}$.

When this definition is adopted it is found that the EP and IP yielded designs with higher efficiencies than those of M. and J., but with unequal replications. Table 4.3.3 gives the replications and efficiencies of the M. and J. designs and those generated by the EP and IP. These designs are identified by their treatment number v = 10, 11 or 12.

Table 4.3.3. Designs with unequal replications

	M. and J.		EP and IP	
v	replication	efficiency		efficiency
10	2	0.273	(8,2,2,2,1,1,1,1,1,1)	0.291
11	2	0.250	(9,2,2,2,1,1,1,1,1,1,1)	0.285
12	2	0.231	(10,2,2,2,1,1,1,1,1,1,1,1)	0.281

Also for these three particular cases the M. and J. design is better with respect to D-optimality, but the reverse is true for E- and A-optimality.

When the replications were fixed at 2 and only the IP employed the optimal designs given by M. and J. were obtained.

There were seven parameter sets (v, k, r) = (12, 2, 5), (12, 2, 6), (12, 3, 3), (12, 3, 8), (12, 4, 9), (12, 6, 10) and (12, 9, 9) which M. and J. were unable to handle. For these (v, k, r) the EP and IP generated designs which had equal or greater efficiency than those given by Clatworthy (1973). Except for (12, 6, 10) for which the design found by the EP and IP had an efficiency 0.00005 less than that of the

most efficient found by Clatworthy.

A description of the EP and IP and some of the above tests was given by Jones and Eccleston (1980).

The 209 designs considered in the above test included all but four of the forty designs considered by Mitchell (1973) that were used to test the algorithm described in Chapter 2. The designs not included were for (v, k, r) = (4, 2, 11), (12, 3, 3), (13, 3, 3) and (14, 3, 3). When the EP and IP were used to obtain designs for these sets of values of (v, k, r) the overall result was that optimal designs were found for thirty-nine sets. The only local optimum obtained was for (12, 3, 3). This local optimum equalled the design found for (12, 3, 3) in Chapter 2. The algorithm based on $\underset{\sim}{\Omega}_H$ and described in Chapter 2 only managed to obtain thirty-six optimal designs out of the forty. Therefore, the EP and IP has done better on this test than either Mitchell or the algorithm described in Chapter 2.

4.4 Exchange and interchange procedures to search for optimal row-and-column designs

As noted in Chapter 3, the expected values of the n observations from an experiment in b_1 rows and b_2 columns can be expressed as

$$E(\underset{\sim}{y}_n) = \underset{\sim}{\Delta}'\underset{\sim}{\gamma} + \underset{\sim}{D}_1'\underset{\sim}{\beta}_1 + \underset{\sim}{D}_2'\underset{\sim}{\beta}_2 + \alpha\underset{\sim}{1}_n.$$

If the parameters of the above model are partitioned as $[\underset{\sim}{\gamma}\,;\,\underset{\sim}{\beta}_1, \underset{\sim}{\beta}_2, \alpha]$, then the notation introduced in Section 4.2 after Theorem 4.2.1 may be used. That is,

$$\underset{\sim}{A}_n = \underset{\sim}{\Delta}'$$

and

$$\underset{\sim}{B}_n = [\underset{\sim}{D}_1', \underset{\sim}{D}_2', \underset{\sim}{1}_n].$$

Then $\underset{\sim}{A}_n'\underset{\sim}{A}_n = \underset{\sim}{r}_n^{\delta}$ is the v × v diagonal matrix of treatment

replications and

$$\underset{\sim}{A}'\underset{\sim}{B}_n = [\underset{\sim}{M}_n, \underset{\sim}{N}_n, \underset{\sim}{r}_n],$$

where $\underset{\sim}{M}_n$ is the $v \times b_1$ incidence matrix for the row component, $\underset{\sim}{N}_n$ is the $v \times b_2$ incidence matrix for the column component and $\underset{\sim}{r}_n$ is the $v \times 1$ vector of treatment replications. It will be recalled that $\underset{\sim}{M}_n$ was referred to as $\underset{\sim}{N}_1$ in Chapter 3, and $\underset{\sim}{N}_n$ as $\underset{\sim}{N}_2$.

The matrix $\underset{\sim}{C}_n$ of the coefficients of the equations for estimating the treatment effects adjusted for rows and columns is therefore

$$\underset{\sim}{C}_n = \underset{\sim}{r}_n^{\delta} - \frac{1}{b_2}\underset{\sim}{M}_n\underset{\sim}{M}_n' - \frac{1}{b_1}\underset{\sim}{N}_n\underset{\sim}{N}_n' + \underset{\sim}{r}_n\underset{\sim}{r}_n'/n. \qquad (4.4.1)$$

As done in the previous section, a generalized inverse of $\underset{\sim}{C}_n$ will be denoted by $\underset{\sim}{C}_n^-$. For a connected design a convenient generalized inverse of $\underset{\sim}{C}_n$ is $\underset{\sim}{\Omega}$ as defined in Chapter 3.

Suppose in the experimental design treatment i, $i = 1, 2, \ldots, v$, is removed from the plot in the p^{th} row and the q^{th} column. Then

$$\underset{\sim}{C}_{n-1} = \underset{\sim}{C}_n - \underset{\sim}{c}_1\underset{\sim}{c}_1'$$

and

$$\underset{\sim}{C}_{n-1}^- = \underset{\sim}{C}_n^- + \frac{\underset{\sim}{C}_n^-\underset{\sim}{c}_1\underset{\sim}{c}_1'\underset{\sim}{C}_n^-}{1 - \underset{\sim}{c}_1'\underset{\sim}{C}_n^-\underset{\sim}{c}_1},$$

where

$$\underset{\sim}{c}_1 = \frac{\underset{\sim}{a}_1 - (\underset{\sim}{A}'\underset{\sim}{B}_n)(\underset{\sim}{B}'\underset{\sim}{B}_n)^-\underset{\sim}{b}_1}{[1 - \underset{\sim}{b}_1'(\underset{\sim}{B}_n'\underset{\sim}{B}_n)^-\underset{\sim}{b}_1]^{\frac{1}{2}}}.$$

Here $\underset{\sim}{a}_1$ is a $v \times 1$ vector of zeros except for a one in position i, $\underset{\sim}{b}_1$ is a vector of zeros except for a one in positions p, $b_1 + q$, and $(b_1 + b_2 + 1)$,

$$(B'B)_{\sim n\sim n} = \begin{bmatrix} b_2 I_{\sim b_1} & 1_{\sim b_1} 1'_{\sim b_2} & b_2 1_{\sim b_1} \\ 1_{\sim b_2} 1'_{\sim b_1} & b_1 I_{\sim b_2} & b_1 1_{\sim b_2} \\ b_2 1'_{\sim b_1} & b_1 1'_{\sim b_1} & n \end{bmatrix},$$

$$(B'B)^{-}_{\sim n\sim n} = \begin{bmatrix} \frac{1}{b_2} I_{\sim b_1} & \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0} & \frac{1}{b_1} I_{\sim b_2} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} & -\frac{1}{n} \end{bmatrix},$$

$I_{\sim b_1}$ is the identity matrix of order b_1 and $\underset{\sim}{0}$ is a matrix of appropriate size with all elements equal to zero.

In fact $c_{\sim 1}$ reduces to

$$c_{\sim 1} = \frac{1}{(1 - \frac{1}{b_1} - \frac{1}{b_2} + \frac{1}{n})^{\frac{1}{2}}} \begin{bmatrix} -\frac{M_{1p}}{b_2} - \frac{N_{1q}}{b_1} + \frac{r_1}{n} \\ \vdots \\ 1 - \frac{M_{ip}}{b_2} - \frac{N_{iq}}{b_1} + \frac{r_i}{n} \\ \vdots \\ -\frac{M_{vp}}{b_2} - \frac{N_{vq}}{b_1} + \frac{r_v}{n} \end{bmatrix},$$

where M_{ij} is the $(i, j)^{th}$ element of $M_{\sim n}$, N_{ij} is the $(i, j)^{th}$ element of $N_{\sim n}$ and r_i is the i^{th} element of $r_{\sim n}$.

If treatment j is now placed on the $(p, q)^{th}$ plot in place of i, then

$$\text{new } C_{\sim n} = C_{\sim n-1} + c_{\sim 2} c'_{\sim 2}$$

and

$$(\text{new } C_{\sim n})^{-} = C^{-}_{\sim n-1} - \frac{C^{-}_{\sim n-1} c_{\sim 2} c'_{\sim 2} C^{-}_{\sim n-1}}{1 + c'_{\sim 2} C^{-}_{\sim n-1} c_{\sim 2}},$$

where

$$\underset{\sim}{c}_2 = \frac{\underset{\sim}{a}_2 - (\underset{\sim}{A}'_{n-1}\underset{\sim}{B}_{n-1})(\underset{\sim}{B}'_{n-1}\underset{\sim}{B}_{n-1})^{-}\underset{\sim}{b}_2}{(1 + \underset{\sim}{b}'_2(\underset{\sim}{B}'_{n-1}\underset{\sim}{B}_{n-1})^{-}\underset{\sim}{b}_2)^{\frac{1}{2}}}$$

and $\underset{\sim}{a}_2$ and $\underset{\sim}{b}_2$ characterize the new observation in the same way that $\underset{\sim}{a}_1$ and $\underset{\sim}{b}_1$ did the deleted observation.

Further,

$$(\underset{\sim}{A}'_{n-1}\underset{\sim}{B}_{n-1}) = (\underset{\sim}{A}'\underset{\sim}{B}_n) - \underset{\sim}{a}_1\underset{\sim}{b}'_1 \qquad (4.4.2)$$

and

$$(\underset{\sim}{B}'_{n-1}\underset{\sim}{B}_{n-1})^{-} = (\underset{\sim}{B}'\underset{\sim}{B}_n)^{-} + \frac{(\underset{\sim}{B}'\underset{\sim}{B}_n)^{-}\underset{\sim}{b}_1\underset{\sim}{b}'_1(\underset{\sim}{B}'\underset{\sim}{B}_n)^{-}}{1 - \underset{\sim}{b}'_1(\underset{\sim}{B}'_n\underset{\sim}{B}_n)^{-}\underset{\sim}{b}_1}. \qquad (4.4.3)$$

It is worth noting that the second matrix on the right hand side of (4.4.2) has only three non-zero elements and the second matrix on the right hand side of (4.4.3) has only nine non-zero elements. All these non-zero elements can be calculated directly without the need for numerous matrix multiplications.

The criterion used to search for optimal designs is, as before,

$$\text{minimize trace } (\underset{\sim}{L}'\underset{\sim}{C}_n^{-}\underset{\sim}{L}\omega^{\delta}),$$

where $\underset{\sim}{C}_n^{-}$ is a generalized inverse of $\underset{\sim}{C}_n$ as defined in equation (4.4.1).

An algorithm to search for optimal row-and-column designs is then similar to that described in the previous section. Given a connected starting design the EP is employed to obtain the "correct" replications. That is, at each step the weakest observation is exchanged for the strongest.

In the IP the weakest observation, a replicate of treatment ℓ say, is exchanged for the strongest, a replicate of treatment m say. Secondly, to maintain the correct replications, the weakest observation involving an m is replaced by an ℓ. If this does not improve the design the last

mentioned observation is replaced and the next weakest observation involving an m is replaced by an ℓ and so on. If there is no improvement for all m then the procedure restarts with the next weakest observation instead of the particular one involving ℓ and so on. The procedure terminates when no interchange can be found that improves the design.

The recursive formulae given above in this section are used to determine the change in $\underset{\sim}{C}_n^-$ effected by an exchange or interchange. It should be noted that $\underset{\sim}{C}_n^-$ and $(\underset{\sim}{B}_n'\underset{\sim}{B}_n)^-$ are only calculated directly once and all other generalized inverses are obtained using the recursive formulae.

Some tests of the performance of the above algorithm are given below.

A description of the EP and IP applied to row-and-column designs was given by Eccleston and Jones (1980).

Test 1

Here an attempt was made to obtain the 37 designs given in Table 13.2 of Cochran and Cox (1957) for which $v \leq 11$ and $b_1 + b_2 \leq 22$. It will be recalled that these designs were used in Chapter 3 to test the performance of the algorithms described there.

The starting designs were generated using the procedure described in Section 3.4 of Chapter 3. With ε set equal to 0.00001 the EP and IP did not find the optimal designs for six sets of values of (v, b_1, b_2). These sets were (4, 8, 5), (7, 7, 4), (7, 14, 4), (9, 18, 4), (11, 11, 5) and (11, 11, 6). The algorithm described in Chapter 3 did better in that it found all but three of the optimal designs, and these three were for sets (9, 18, 4), (11, 11, 5) and (11, 11, 6). The results were therefore disappointing in view of the performance of the EP and IP for block designs.

The most surprising result was that the design obtained by the EP for (4, 8, 5) had unequal treatment replications. The reason why this happened, however, was clearly seen. The incidence matrices $\underset{\sim}{M}_n$ and $\underset{\sim}{N}_n$ for rows and columns, respectively, of the optimal design are

$$\underset{\sim}{M}_n = \begin{bmatrix} 2 & 1 & 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 & 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 & 1 & 2 \end{bmatrix} \quad \text{and} \quad \underset{\sim}{N}_n = \begin{bmatrix} 2 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 \end{bmatrix}.$$

The incidence matrices for the starting design were

$$\underset{\sim}{M}_n = \begin{bmatrix} 2 & 1 & 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 1 & 2 & 0 & 2 \\ 1 & 1 & 2 & 1 & 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 & 2 & 1 \end{bmatrix} \quad \text{and} \quad \underset{\sim}{N}_n = \begin{bmatrix} 2 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 \end{bmatrix}.$$

The first obvious improvement to be made to the starting design is to increase the zero value in column seven of $\underset{\sim}{M}_n$. The EP does this by removing treatment four from plot (7, 1) and replacing it by treatment two. This results in column seven of $\underset{\sim}{M}_n$ having one entry equal to two and the rest equal to one. That is, a pattern more similar to that of the optimal design. Once the EP has produced the wrong treatment replications it is impossible for the IP to find the optimal design. When the IP only was applied to the starting design the optimal design was found. Therefore, a more selective use of the EP and IP has produced the optimum.

The optimal design for (7, 7, 4) was found by applying the EP and IP to the starting design with $\varepsilon = 0.0$ instead of $\varepsilon = 0.00001$. The disadvantage of always using $\varepsilon = 0.0$ is that rounding errors in the calculation of the trace sometimes cause the design to make an interchange and then interchange back the same pair of treatments. The consequence of this is that valuable computer time can be lost.

The optimal design found for (7, 14, 4) could not be

improved upon using $\varepsilon = 0.0$ and so the possibility of rounding errors affecting the calculations was considered more seriously. For the larger designs, the number of times the original $\underset{\sim}{C}_n^-$ is updated when determining the effect of an exchange or interchange is quite large.

The first way of overcoming this difficulty was to take advantage of the fact that if the original $\underset{\sim}{C}^-$ is $\underset{\sim}{\Omega}$ as defined in equation (3. 3. 20) then it must satisfy the constraint

$$\underset{\sim}{\Omega}\underset{\sim}{r}_n = \underset{\sim}{1}$$

where $\underset{\sim}{r}_n$ is the vector of treatment replications and $\underset{\sim}{1}$ is an $n \times 1$ vector of ones.

In the updating formula

$$\underset{\sim}{C}_{n-1}^- = \underset{\sim}{\Omega} + \frac{\underset{\sim}{\Omega}\underset{\sim}{c}_1\underset{\sim}{c}_1'\underset{\sim}{\Omega}}{1 - \underset{\sim}{c}_1'\underset{\sim}{\Omega}\underset{\sim}{c}_1},$$

and so

$$\underset{\sim}{C}_{n-1}^-\underset{\sim}{r}_n = \underset{\sim}{1}$$

as

$$\underset{\sim}{c}_1'\underset{\sim}{1} = 0.$$

Further as,

$$(\text{new } \underset{\sim}{C}_n)^- = \underset{\sim}{C}_{n-1}^- - \frac{\underset{\sim}{C}_{n-1}^-\underset{\sim}{c}_2\underset{\sim}{c}_2'\underset{\sim}{C}_{n-1}^-}{1 + \underset{\sim}{c}_2'\underset{\sim}{C}_{n-1}^-\underset{\sim}{c}_2}$$

then

$$(\text{new } \underset{\sim}{C}_n)^-\underset{\sim}{r}_n = \underset{\sim}{1}$$

as

$$\underset{\sim}{c}_2'\underset{\sim}{1} = 0.$$

The EP and IP were therefore modified so that whenever a plot was removed the resulting $\underset{\sim}{C}_{n-1}^-$ was altered, if necessary, so that $\underset{\sim}{C}_{n-1}^-\underset{\sim}{r}_n = \underset{\sim}{1}$, and whenever a plot was added the resulting $(\text{new } \underset{\sim}{C}_n)^-$ was altered, if necessary, so that

(new $\underset{\sim}{C}_n$)$\bar{}$ $\underset{\sim}{r}_n = \underset{\sim}{1}$.

For example, suppose that when a plot is removed, the first row of $\underset{\sim}{C}^{-}_{n-1}$ is such that its elements $c_1, c_2, \ldots, c_v$, say, satisfy

$$\sum_{i=1}^{v} c_i r_i = 1 + a,$$

where r_i, $i = 1, 2, \ldots, v$ are the elements of $\underset{\sim}{r}_n$ and a is a non-zero constant. To restore this row so that $\underset{\sim}{C}^{-}_{n-1}\underset{\sim}{r}_n = \underset{\sim}{1}$ the elements c_i are altered to $(c_i - \frac{a}{n})$. Then,

$$\sum_{i=1}^{v} (c_i - \frac{a}{n}) r_i = 1 + a - a = 1.$$

The fact that a constant may have been added to the elements of each row of $\underset{\sim}{C}^{-}_{n-1}$ does not affect any of the other calculations needed in the computer program that executes the EP and IP.

The second way was to recalculate $\underset{\sim}{C}^{-}_{n}$ directly after the EP so that the IP begins with an accurate matrix $\underset{\sim}{C}^{-}_{n}$.

With these modifications the optimal design was obtained for (7, 14, 4) when the EP and IP were applied to the starting design and ε was set equal to 0.00001.

Even with these improvements the optimal designs for (9, 18, 4), (11, 11, 5) and (11, 11, 6) were not found by the EP and IP. The efficiency factors of these designs and those of the optimal design are given in Table 4.4.1. It can be seen that the designs found are not much worse than

Table 4.4.1 Efficiencies of designs

(v, b_1, b_2)	Design found by EP and IP	Optimal design
(9, 18, 4)	0.8426	0.8437
(11, 11, 5)	0.8757	0.8800
(11, 11, 6)	0.9137	0.9167

the optimal designs. However, it would be useful to determine if they can be improved upon. The method adopted in Chapter 3 was to repeat the interchanging of treatments in an attempt to minimize the sum of the absolute values of the off-diagonal elements of $\underset{\sim}{L}'\underset{\sim\sim}{\Omega}\underset{\sim}{L}$. Therefore, to see if this would lead to any improvement the EP and IP were applied to the three final designs found above for (9, 18, 4), (11, 11, 5) and (11, 11, 6), respectively, using this time the criterion

minimize the sum of the absolute values of the off-diagonal elements of $\underset{\sim}{L}'\underset{\sim n\sim}{\bar{C}}\underset{\sim}{L}$.

The results were disappointing in that although the value of the above criterion was reduced in all three cases, the value of trace $(\underset{\sim}{L}'\underset{\sim n\sim\sim}{\bar{C}}\underset{\sim\sim}{L\omega}^{\delta})$ was increased. Further, none of these final designs had all off-diagonal elements of $\underset{\sim}{L}'\underset{\sim n\sim}{\bar{C}}\underset{\sim}{L}$ equal to zero, and the design found for (11, 11, 5) had unequal treatment replications. When only the IP was applied to the final design found for (11, 11, 5), i.e. the design noted in Table 4.4.1, using the new criterion, a design with a larger value of trace $(\underset{\sim}{L}'\underset{\sim n\sim\sim}{\bar{C}}\underset{\sim\sim}{L\omega}^{\delta})$ was obtained and not all off-diagonal elements of $\underset{\sim}{L}'\underset{\sim n\sim}{\bar{C}}\underset{\sim}{L}$ were zero.

To determine if changing the value of ε might result in better final designs the design for (11, 11, 5) was considered further. Using only the trace $(\underset{\sim}{L}'\underset{\sim n\sim\sim}{\bar{C}}\underset{\sim\sim}{L\omega}^{\delta})$ criterion and the original starting design, the EP and IP were applied using values of $\varepsilon = 0.001$ and $\varepsilon = 0.0$ respectively. The design found when $\varepsilon = 0.001$ was worse than that obtained previously and the design found for $\varepsilon = 0.0$ was the same as that found previously. When the IP was applied to each of these two final designs, in an attempt to minimize the covariance criterion defined earlier, the optimal design was still not found. In fact the designs found were worse on the original trace criterion.

In one last attempt to overcome this problem of obtaining local optima the IP was extended slightly. In the original version, treatment ℓ, say was removed and replaced by treatment m, say. Then the weakest m was removed and replaced by ℓ. If this led to a smaller trace the interchange was kept and the design updated. If not, the next weakest m was considered for replacement by ℓ, and so on. In the extended version all the m's in the second part of the interchange are considered in turn, in order of the weakest, and replaced by an ℓ. After all the m's have been considered the one that when replaced by ℓ gives the largest reduction, if any, in the trace is chosen and the interchange completed. However, this did not lead to the optimal design for (11, 11, 5) being found and so this extended IP was not considered further.

Test 2

The aim of this test was to check if the EP and IP would give the correct design, when applied to a very perverse starting design. The example chosen was a design for four treatments in six rows and six columns and for which there was an equal interest in all treatment contrasts. The starting design was

```
1 2 3 4 1 1
2 1 1 1 1 1
3 1 1 1 1 1
4 1 1 1 1 1
1 1 1 1 1 1
1 1 1 1 1 1
```

The EP produced the optimal design, that is, an equi-replicated design in total balance.

In fact, when the EP and IP were applied to the design generated by the method described in Section 3.4 of Chapter 3, the optimal design was also obtained.

Test 3

Here the aim was to determine how well the EP and IP

would be at finding designs for estimating contrasts with unequal weights. As in the previous tests an example used in Chapter 3 is used. There are nine treatment combinations of a 3 × 3 factorial experiment to be assigned to six rows and six columns. The contrasts of interest are those given in Table 3.9.1 of Chapter 3. The weights are defined, as in Chapter 3, by a vector (w_1, w_2, w_3). Here w_1 is the weight of each main effect of the first factor, w_2 is the weight of each main effect of the second factor, and w_3 is the weight of each of the four interactions. As computer time was limited, only four different sets of weights were used. These were $(1, 1, 1)$, $(1, 1, \frac{1}{2})$, $(\frac{1}{2}, 1, 1)$ and $(1, 0, 0)$. With ε set equal to 0.00001 and using the starting design generated by the method described in Section 3.4 of Chapter 3, the EP and IP were employed using each set of weights. All four final designs did not have all off-diagonal elements of $\underset{\sim}{L}'\bar{\underset{\sim}{C}}_n\underset{\sim}{L}$ equal to zero, although the design found for weights $(1, 0, 0)$ was optimal for the requirements set. It will be recalled that when the algorithm described in Chapter 3 was applied using these four sets of weights the optimal design was found for weights $(1, 1, \frac{1}{2})$ and $(1, 0, 0)$. As done previously the IP was applied to the final designs found for weights $(1, 1, 1)$, $(1, 1, \frac{1}{2})$ and $(\frac{1}{2}, 1, 1)$ in an attempt to minimize the sum of the absolute values of the off-diagonal elements of $\underset{\sim}{L}'\bar{\underset{\sim}{C}}_n\underset{\sim}{L}$, using $\varepsilon = 0.00001$. However, the final designs, although improved in terms of the second criterion, still had some non zero off-diagonal elements of $\underset{\sim}{L}'\bar{\underset{\sim}{C}}_n\underset{\sim}{L}$.

The algorithm described in Chapter 3 was therefore more successful in obtaining designs for these sets of weights.

Overall, the conclusions to be drawn from the above tests are that the EP and IP when applied to row-and-column designs

are likely to get trapped more often at local optima than when they are applied to block designs. The EP and IP are more sensitive to changes in the value of ε and to the effect of rounding errors. The interchange algorithm described in Chapter 3 was more successful at locating optimal row-and-column designs.

However, none of the local optima found were much worse than the true optima. A general conclusion is that the local optima are so near the true optima that to improve them a radical change is needed that cannot be achieved by simple interchanges.

For example, the local optimum found for (11, 11, 5) was the following row-and-column design.

1	10	3	7	5
6	7	1	2	11
2	8	5	6	10
4	6	7	8	9
5	2	4	3	7
7	9	8	11	3
9	5	6	1	4
8	3	2	9	1
10	11	9	5	2
3	4	11	10	6
11	1	10	4	8.

Here the column component is optimal but not the row component. Here no sequence of simple interchanges can improve the design unless the design is allowed to get worse first.

However, one major advantage of the EP and IP is that they are quicker to execute on the computer than the algorithms described in Chapters 2 and 3. For example, the algorithm described in Chapter 3 took 684 CP seconds on the University of London CDC 7600 computer to search for the design for (11, 11, 6), whereas the EP and IP took approximately 100 CP seconds. The algorithm described in Chapter 2 took just

over 31 CP seconds to search for the design for 14 treatments in 14 blocks of 3, whereas the EP and IP took just over 11 CP seconds.

Summary and conclusions

In this thesis I have considered how the computer can be used to assist the experimenter at the design stage of an experiment. The computer can provide advice on the design of an experiment through an interactive computer program. It can also be programmed to search for optimal designs.

In Chapter 1 the problem was that of formalizing the conversation that takes place between an experimenter and a statistical consultant, so as to write an interactive computer program to advise on the design. Experiments to compare two treatment means were considered by way of illustration. Although some aspects of the consultant's role could be programmed, e.g. asking a set of questions to obtain the necessary information and performing calculations on this, it was found that some aspects were difficult to incorporate into a computer program. It appeared that the consultant's versatility, experience, ability to talk around a problem and his ability to modify his approach to suit his client were very difficult to simulate. If a program is to be written, however, one difficulty that must be overcome is that of deciding in advance all the information that will be required to design the experiment. If that can be done, then the total amount of information must be divided up into smaller parts so that an ordered set of simple questions can be planned. It was apparent that it was not always an easy matter to devise questions that could be easily understood by different experimenters. The fact that it was desirable that the replies should be simple 'Yes', 'No' type answers also added to the difficulty. The way a question was phrased was very important. It was also important to plan carefully the order so that one reply led naturally on to the next. It was apparent that a

program to suit all client types would be large and difficult to write. The two types of client that could most easily be catered for were found to be those without any statistical expertise and those who were statistical experts. Some aspects of the design that were difficult to program concerned the choice of treatments, choice of experimental units and choice of experimental site. The types of experiment about which the computer could most easily advise were those that were familiar and required some numerical calculations. The number of questions and related pieces of text that could be printed by the program was very large and efficient ways of organising the printing were required. One way was suggested in Chapter 1. If advice is to be given about more complicated experiments the size of the program is likely to increase considerably. For example, when more than two treatments are to be compared, questions to determine the treatment structure, if any, will be required, as well as questions on the treatment contrasts of interest and their relative importance. If optimal experimental designs are to be provided by the program then suitable computer algorithms that search for them would be needed. Such algorithms were considered in the remainder of this thesis.

In Chapter 2 a review was made of the literature on methods of constructing optimal designs for the general linear model and for the case where the experimental units are arranged as blocks. It was discovered that the available algorithms were not general enough to deal with the situation when there is special interest in a certain set of treatment contrasts. Further, blocks may be of unequal size. To remedy this deficiency an algorithm was devised based on interchanging treatments between blocks, with the aim of improving an appropriate optimality criterion. The criterion was to minimize

the sum of the weighted variances of the contrasts of interest. The calculation of this criterion involved inverting a matrix $\underset{\sim}{\Omega}^{-1}$. To avoid that an approximation to $\underset{\sim}{\Omega}$, denoted by $\underset{\sim}{\Omega}_H$ was used. The approximation was iterative, H being the number of cycles used. To decide if this approximation to the true value of the criterion could be improved, an alternative matrix to $\underset{\sim}{\Omega}_H$ was devised, namely $\underset{\sim}{\Xi}_H$. It was based solely on treatment concurrences. Tests of the performance of the interchange algorithm resulted in a preference for $\underset{\sim}{\Omega}_H$, although it was noted that some of the disadvantages of $\underset{\sim}{\Xi}_H$ could be overcome. Some theoretical results concerning $\underset{\sim}{\Xi}_H$ were given. A value of H was found to be sufficient for use in the interchange algorithm. A starting design was required and a way of obtaining one was described. When the performance of the interchange algorithm was compared with another which was considered to be among the best in the literature, it was found to be superior. It was noted that the other algorithm did not permit treatment contrasts to be weighted. One disadvantage of the interchange algorithm, in common with others that have been suggested, was its getting trapped at a local optimum. However, in tests none of the local optima found were much inferior to the true optima. The performance of the interchange algorithm was indeed good.

Chapter 3 began by giving a review of the literature on methods of constructing optimal row-and-column designs. It was noted that the row-and-column case had received less attention than the block case. Since no suitable algorithm was available in the literature, the interchange algorithm described in Chapter 2 was modified to search for the component designs that make up an optimal row-and-column design. This modification was made possible by making use of the amalgamated

design. The interchange algorithm used a criterion similar to that in Chapter 2, and two approximations to the true value of the criterion, based on new matrices $\underset{\sim}{\Omega}_H$ and $\underset{\sim}{\Xi}_H$, were devised. The roles of $\underset{\sim}{\Omega}_H$ and $\underset{\sim}{\Xi}_H$ were similar to those of the corresponding matrices used in Chapter 2. Tests of the interchange algorithm resulted again, in a preference for the approximation based on $\underset{\sim}{\Omega}_H$. As before, a value of H equal to four was found to be sufficient. The interchange algorithm, however, only provided the component designs and a further algorithm was required to arrange them as a row-and-column design. Three alternative algorithms were devised. In the first of these a row of the row-and-column design was arranged at each step, in the second a row and a column were arranged at each step and in the third a plot was arranged at each step. The first two of these algorithms, however, suffered from the disadvantage of not being able to consider every possible arrangement and so sometimes failed to combine a pair of components when in fact it was possible. The failure rate of the first algorithm was a little higher than that of the second. The third algorithm, however, did consider every possible arrangement and so was guaranteed to combine a pair if it was at all possible. To test these three algorithms and to test further the performance of the interchange algorithm a particular set of designs was used. It was noted that the interchange algorithm could provide a pair of components that could not be combined, although in tests that happened on only three occasions out of fifty-two. With a few of the designs obtained in the tests there were some that had orthogonal contrasts that were not estimated orthogonally. To overcome that a modified form of the interchange algorithm was applied to the offending designs. The optimality criterion was the minimization of the sum of

the absolute values of the off-diagonal elements of an appropriate approximation to the variance-covariance matrix for contrasts. This modification led to improvements in most cases and had the additional advantage of altering the three pairs of components that could not be combined so that they could be. It was found that in nearly all cases the design found was optimal. However, even though the interchange algorithm did sometimes get trapped at a local optimum, the local optimum was nevertheless nearly as good as the true optimum.

One disadvantage of the interchange algorithms described in Chapters 2 and 3 was that the treatment replications once chosen, remain fixed. To overcome that a different approach was described in Chapter 4. There given a suitable starting design, an exchange procedure (EP) and an interchange procedure (IP) were used to search for optimal block designs and optimal row-and-column designs. The EP and IP made use of recursive formulae that enabled the true value of the optimality criteria used in Chapters 2 and 3 to be calculated for the design obtained by deleting a treatment and for the design obtained by adding a different treatment to the depleted design. During the EP the treatment replications were modified, if necessary, as the design was improved. The IP was then used, if necessary, to search for an optimal set of treatment concurrences. As the EP and IP worked directly with the row-and-column design, a further algorithm to combine components was not needed. The performance of the EP and IP when applied to block designs was very good and was better than that of the interchange algorithm described in Chapter 2. As before, however, the procedures could get trapped at a local optimum, although the local optimum found was nearly always as good as the true optimum. The

performance of the EP and IP when applied to row-and-column designs, however, was worse than that of the interchange algorithm described in Chapter 3, more local optima being located. No modification was found to improve the procedures. However, as before, the local optima were nearly as good as the true optima. An important advantage of the EP and IP was that they were quicker to execute on the computer than the interchange algorithms described in Chapters 2 and 3.

Acknowledgements

I am much indebted to my supervisor Professor S.C. Pearce for providing advice and encouragement. Financial support for the early part of the work was provided by the Science Research Council, in the form of a CASE award in conjunction with I.C.I., Mond Division, Runcorn. For this award I am most grateful. Finally I am much indebted to my wife for her support and encouragement, and for typing this thesis.

References

AITCHISON, J. and BROWN, J.A.C. (1966). The lognormal distribution. Cambridge University Press.

ANDERSON, I. (1974). A first course in combinatorial mathematics. Clarenden Press: Oxford.

ANSCOMBE, F.J. (1948). The transformation of Poisson, Binomial and Negative-Binomial data. Biometrika, 35, 246 - 254.

ATKINSON, A.C. (1969). Constrained maximization and the design of experiments. Technometrics, 11, 616 - 618.

ATWOOD, C.L. (1973). Sequences converging to D-optimal designs of experiments. Ann. Statist., 1, 342 - 352.

BAKER, R.J. and NELDER, J.A. (1978). General Linear Interactive Modelling (GLIM, Release 3). Numerical Algorithms Group: Oxford.

BARTLETT, M.S. (1947). The use of transformations. Biometrics, 3, 39 - 52.

BLISS, C.I. (1938). The transformation of percentages for use in the analysis of variance. Ohio J. Science, 38, 9 - 12.

BLOOM, T., PFAFFENBERGER, R. and KOCHENBERGER, G. (1972). The efficient construction of optimal response surface designs based on the $|\underset{\sim}{X}'\underset{\sim}{X}|$ criterion. Presented at 1972 Amer. Statist. Assn. Meeting, Montreal, Canada, August 14 - 18.

BOEN, J.R. (1972). The teaching of personal interaction in statistical consultancy. American Statistician, 26, 30 - 31.

BOX, M.J. (1966). A comparison of several current optimization methods, and the use of transformations in constrained problems. Computer J., 9, 67 - 77.

BOX, M.J. and DRAPER, N.R. (1971). Factorial designs, the $|X'X|$ criterion, and some related matters. Technometrics, 13, 731 - 742.

CALIŃSKI, T. (1971). On some desirable patterns in block designs (With Discussion). Biometrics, 27, 275 - 292.

CAMERON, J.M. (1969). The statistical consultant in a scientific laboratory. Technometrics, 11, 247 - 254.

CATCHPOLE, E.A. (1980). On the convergence of Jones' iteration. To appear in Biom. Zeit.

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

CLARK, A. and LEONARD, W.H. (1939). The analysis of variance with special reference to data expressed as percentages. J. Amer. Soc. Agron., 31, 55 - 66.

CLATWORTHY, W.H. (1973). Tables of two-associate-class partially balanced designs. Nat. Bur. Stand. (U.S.), Appl. Math. Ser. 63.

COCHRAN, W.G. (1938). Some difficulties in the statistical analysis of replicated experiments. Emp. J. Expt. Agric., 6, 157 - 175.

COCHRAN, W.G. and COX, G.M. (1957). Experimental designs, 2nd ed. Wiley: New York.

COVEY-CRUMP, P.A.K. and SILVEY, S.D. (1970). Optimal regression designs with previous observations. Biometrika, 57, 551 - 566.

COX, C.P. (1968). Some observations on the teaching of statistical consulting. Biometrics, 24, 789 - 801.

COX, D.R. (1958). Planning of experiments. Wiley: New York.

DANIEL, C. (1969). Some general remarks on consulting in statistics. Technometrics, 11, 241 - 245.

DYKSTRA, O. Jnr. (1971). The augmentation of experimental data to maximize $|X'X|$. Technometrics, 13, 682 -688.

ECCLESTON, J.A. (1979). Recursive techniques for the construction of experimental designs. To appear in J. Statist. Plan. Inf.

ECCLESTON, J.A. and HEDAYAT, A. (1974). On the theory of connected designs: characterization and optimality. Ann. Statist., 2, 1238 - 1255.

ECCLESTON, J.A. and JONES, B. (1980). Exchange and interchange procedures to search for optimal row-and-column designs. To appear in J. R. Statist. Soc., B, 42.

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

EISENHART, C. (1947). The role of the statistical consultant in a research organisation. Proc. Inter. Stat. Conf. III, 308 - 313.

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

FEDOROV, V.V. (1969a). Properties and methods for construction of point optimal designs in regression experiments. Preprint No 5. LSM, Izd-vo Moscow State University, Moscow, USSR.

FEDOROV, V.V. (1969b). Theory of optimal experiments. Preprint No 7. LSM, Izd-vo Moscow State University, Moscow, USSR.

FEDOROV, V.V. (1972). The theory of optimal experiments. Academic Press: New York.

FEDOROV, V.V. and DUBOVA, I.S. (1968). Methods for constructing optimal designs in regression experiments. Preprint No 4. LSM, Izd-vo Moscow State University, Moscow, USSR.

FINNEY, D.J. (1956). The statistician and the planning of field experiments (With Discussion). J. R. Statist. Soc., B, 119, 1 - 27.

FINNEY, D.J. (1970). Contribution to the discussion of "Some problems of statistical consultancy" by P. Sprent. J. R. Statist. Soc., A, 133, 139 - 164.

FREEMAN, G.H. (1957). Some experimental designs for use in changing from one set of treatments to another. Part II. Existence of designs. J. R. Statist. Soc., B, 19, 163 - 165.

FREEMAN, G.H. (1976). On the selection of designs for comparative experiments. Biometrics, 32, 195 - 199.

GREENFIELD, A.A. (1979). Statisticians in industrial research: the role and training of an industrial consultant. The Statistician, 28, 71 - 82.

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

HALL, P. (1935). On representatives of subsets. Bull. London Math. Soc., 10, 26 - 30.

HARTLEY, H.O. and RUDD, P.G. (1969). Computer optimization of second order surface designs. In Statistical computing. Ed. R.C. Milton and J.A. Nelder, Academic Press: New York. 441 - 462.

HEALY, M.J.R. (1973). The varieties of statistician. J. R. Statist. Soc., A, 136, 71 -74.

HEBBLE, T.L. and MITCHELL, T.J. (1972). 'Repairing' response surface designs. Technometrics, 14, 767 - 779.

HERZBERG, A.M. and ANDREWS, D.F. (1975). The robust design of experiments. Proc. 40th Session ISI, 370 - 374.

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

HOYLE, M.H. (1973). Transformations - an introduction and a bibliography. Int. Stat. Rev., 41, 203 - 223.

HYAMS, L. (1971). The practical psychology of biostatistical consultation. Biometrics, 27, 201 - 211.

JEFFERS, J.N.R. (1978). Design of experiments. Statistical checklist 1. Institute of Terrestial Ecology.

JOHN, J.A. and TURNER, G. (1977). Some new group divisable designs. J. Statist. Plan. Inf., 1, 103 - 107.

JOHN, P.W.M. (1971). Statistical design and analysis of experiments. Macmillan: New York.

JONES, B. (1976). An algorithm for deriving optimal block designs. Technometrics, 18, 451 - 458.

JONES, B. (1978). Some examples of the use of an algorithm for constructing optimal block designs. B.I.A.S., 5, 82 - 90.

JONES, B. (1979). Algorithms to search for optimal row-and-column designs. J. R. Statist. Soc., B, 41, 210 - 216.

JONES, B. (1980a). The computer as a statistical consultant. To appear in B.I.A.S., 7.

JONES, B. (1980b). Combining two component designs to form a row-and-column design. To appear in the Statistical Algorithm section of Appl. Statist.

JONES, B. and ECCLESTON, J.A. (1980). Exchange and interchange procedures to search for optimal designs. J. R. Statist. Soc., B, 42, 238 - 243.

KASTENBAUM, M.A. (1969). The consulting statistician: who needs him? Oak Ridge National Laboratory Review, 9 - 11.

KEMPTHORNE, O. (1956). The efficiency factor of an incomplete block design. Ann. Math. Statist., 27, 846 - 849.

KIEFER, J. (1958). On non-randomized optimality and randomized non-optimality of symmetrical designs. Ann. Math. Statist., 20, 675 - 699.

KIEFER, J. (1959). Optimum experimental designs. J. R. Statist. Soc., B, 21, 272 -304.

KIEFER, J. (1961a). Optimum designs in regression problems, II. Ann. Math. Statist., 32, 298 - 325.

KIEFER, J. (1961b). Optimum experimental designs V, with applications to rotatable designs. Proc. 4th Berkeley Symp. on Math. Statist. and Prob., 1, 381 - 405.

KIEFER, J. (1962). Two more criteria equivalent to the D-optimality of design. Ann. Math. Statist., 33, 792 - 796.

KIEFER, J. (1974). General equivalence theory for optimum designs (approximate theory). Ann. Math. Statist., 2, 849 - 879.

KIEFER, J. (1975). Construction and optimality of generalized Youden designs. In A survey of statistical design and linear models. Ed. J.N. Srivastava. North Holland: New York.

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

KIEFER, J. and WOLFOWITZ, J. (1960). The equivalence of two extremum problems. Canad. J. Math., 12, 363 - 366.

KIMBALL, A.W. (1957). Errors of the third kind in statsitical consulting. J. Amer. Statist. Assoc., 52, 133 - 142.

KSHIRSAGAR, A.M. (1958). A note on incomplete block designs. Ann. Math. Statist., 29, 907 - 910.

KUROTCHKA, V.G. and DWYER, P.S. (1974). Optimal design of three way layouts without interactions. Math. Operationsforch. Statist., 5, 131 - 145.

MAHONEY, F. (1970). The design of experiments for regression analysis by computer search procedures. Ph. D. Thesis, Rensselaer Polytechnic Institute.

MARQUARDT, D.W. (1979). Statistical consulting in industry. American Statistician, 33, 102 - 107.

MASUYAMA, M. (1957). On the optimality of balanced incomplete block designs. Rep. Statist. Appl. Res., J.U.S.E., 5, 4 - 8.

MASUYAMA, M. (1958). Correction to "On the optimality of balanced incomplete block designs." Rep. Statist. Appl. Res., J.U.S.E., 5, 37 - 38.

MASUYAMA, M. and OKUNO, T. (1957). On the optimality of Latin- Youden- and Shrikhande Square designs. Rep. Statist. Appl. Res., J.U.S.E., 5, 17 - 19.

Mc NEMAR, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika, 12, 153 - 157.

MIETTINEN, O.S. (1968). The matched pairs design in the case of all-or-none response. Biometrics, 24, 339 - 352.

MITCHELL, T.J. (1973). Computer construction of small "D-optimal" incomplete block designs. Proc. 39th Session ISI, 199 - 205.

MITCHELL, T.J. (1974). An algorithm for the construction of "D-optimal" experimental designs. Technometrics, 16, 203 - 211.

MITCHELL, T.J. and JOHN, J.A. (1976). Optimal incomplete block designs. ORNL/CDS-8. Available from the National Technical Information Service, U.S. Dept. of Commerce.

MITCHELL, T.J. and MILLER, F.L. Jnr. (1970). Use of design repair to construct designs for special linear models. Math. Div. Ann. Progr. Rept. (ORNL-4661), Oak Ridge National Laboratory, Oak Ridge, Tennessee, 130 - 131.

MORIGUTI, S. (1954). Optimality of orthogonal designs. Rep. Stat. Appl. Res., J.U.S.E., 3, 74 - 98.

MOTE, V.L. (1958). On a minimax property of a balanced incomplete block design. Ann. Math. Statist., 29, 910 - 913.

NALIMOV, V.V., GOLIKOVA, T.I. and MIKESHINA, N.G. (1970). On practical use of the concept of D-optimality. Technometrics, 12, 799 - 812.

NANDI, H.K. (1950). On the efficiency of experimental designs. Bull. Calcutta Statist. Assoc. Bull., 3, 167 - 171.

NELDER, J.A. (1956). Notes on the planning and execution of field experiments. N.A.A.S. Quart. Res., 34, 139 - 146.

NEUHARDT, J.B. and BRADLEY, H.E. (1971). Selection of multi-factor experimental arrangements with resource constraints. J. Amer. Statist. Assoc., 66, 618 - 621.

PAULSON, E. and WALLIS, W.A. (1947). Techniques of statistical analysis. Ed. C. Eisenhart, M.W. Hastay and W.A. Wallis, McGraw-Hill.

PEARCE, S.C. (1963). The use and classification of non-orthogonal designs (With Discussion). J. R. Statist. Soc., A, 126, 353 - 377.

PEARCE, S.C. (1975a). Optimality of design in plot experiments. Proc. 8th International Biometric Conference, August 25 - 30, 1974, Constanta, Romania. Editura Academiei Republicii Socialiste Romania Bucharest.

PEARCE, S.C. (1975b). Row-and-column designs. Appl. Statist., 24, 60 - 74.

PEARCE, S.C. (1976). Concurrence and quasi-replication: an alternative approach to precision in designed experiments. Biom. Zeit., 18, 105 - 116.

PEARCE, S.C., CALIŃSKI, T. and MARSHALL, T.F. de C. (1974). The basic contrasts of an experimental design with special reference to the analysis of data. Biometrika, 61, 449 - 460.

PEARSON, E.S. and HARTLEY, H.O. (1954). Biometrika tables for statisticians Vol. 1. Cambridge University Press.

PERRY, K.R. (1970). Contribution to the discussion of "Some problems of statistical consultancy" by P. Sprent. J. R. Statist. Soc., A, 133, 139 - 164.

PRIDMORE, W.A. (1970). Contribution to the discussion of "Some problems of statistical consultancy" by P. Sprent. J. R. Statist. Soc., A, 133, 139 - 164.

SALSBURG, D.S. (1973). Sufficiency and the waste of information. American Statistician, 27, 152 - 154.

SCHEFFÉ, H. (1959). The analysis of variance. Wiley: New York.

SCHUCANY, W.R. (1972). Some remarks on educating problem solvers. Proc. Comp. Sci. and Statist. 6th Annual Symp. on the Interface, 33 - 36.

SHAH, K.R. (1960). Optimality criteria for incomplete block designs. Ann. Math. Statist., 31, 791 - 794.

SHAH, K.R., RAGHAVARAO, D.R. and KHATRI, C.K. (1976). Optimality of two and three factor designs. Ann. Statist., 4, 419 - 422.

SHEPPARD, W.F. (1898). On the calculation of the most probable values of frequency contrasts, for data arranged according to equidistant divisions of a scale. Proc. London Math. Soc., 29, 353 - 381.

SILVEY, S.D. and TITTERINGTON, D.M. (1973). A geometric approach to optimal design theory. Biometrika, 60, 21 - 32.

SMITH, K. (1918). On the standard deviations and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. Biometrika, 12, 1 - 85.

SNEDECOR, G.W. and COCHRAN, W.G. (1967). Statistical Methods, 6th ed. Iowa State.

SOBEL, M. and HUYETT, M.J. (1957). Selecting the best one of several binomial populations. Bell System Tech. J., 36, 537 - 576.

SPRENT, P. (1970). Some problems of statistical consultancy (With Discussion). J. R. Statist. Soc., A, 133, 139 - 164.

ST. JOHN, R.C. (1973). Models and designs for experiments with mixtures. Ph. D. Thesis, Department of Statist., University of Wisconsin, Madison, Wisconsin.

ST. JOHN, R.C. and DRAPER, N.R. (1975a) D-optimality for regression designs: a review. Technometrics, 17, 15 - 23.

ST. JOHN, R.C. and DRAPER, N.R. (1975b). D-optimality and mixtures. Presented at the I.M.S. Eastern Regional.

STEIN, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. Ann. Math. Statist., 16, 243 - 258.

TAMHANE, A.C. (1980). Selecting the better Bernoulli treatment using a matched samples design. J. R. Statist. Soc., B, 42, 26 - 30.

TARTER, M. and BERGER, B. (1972). On the training and practice of computer science and statistical consultants. Proc. Comp. Sci. and Statist. 6th Annual Symp. on the Interface, 16 - 23.

TOCHER, K.D. (1952). The design and analysis of block experiments. J. R. Statist. Soc., B, 14, 45 - 91.

TSAY, JIA-YEONG. (1976). On the sequential construction of D-optimal designs. J. Amer. Statist. Soc., 71, 671 - 674.

VAN SCHALKWYK, D.J. (1971). On the design of mixture experiments. Ph. D. Thesis. University of London, London, England.

WALD, A. (1943). On the efficient design of statistical investigations. Ann. Math. Statist., 14, 134 - 140.

WATTS, D.G. (1970). A program for training statistical consultants. Technometrics, 4, 737 - 740.

WHITE, L.V. (1973). An extension of the general equivalence theorem to non-linear models. Biometrika, 60, 345 - 348.

WYNN, H.P. (1970). The sequential generation of D-optimum experimental designs. Ann. Math. Statist., 41, 1655 - 1664.

WYNN, H.P. (1972). Results in the theory and construction of D-optimum experimental designs. J. R. Statist. Soc., B, 34, 133 - 147.

WYNN, H.P. (1973). Simple conditions for optimum experimental design algorithms. Presented at the Symp. on Linear Models, Colorado Springs, Colorado, March 23.

YOUDEN, W.J. (1937). Use of incomplete block replications in estimating tobacco-mosaic virus. Contr. Boyce Thompson Inst., 9, 41 - 48.

GLOSSARY OF NOTATION

Chapter 1

α	significance level of hypothesis test
β	probability of accepting the null hypothesis when it is false, i.e. $(1-\beta)$ is the power of the hypothesis test.
k_α	$100(1-\alpha)$ percentage point of the standard normal distribution (Paulson and Wallis's notation, used on page 49)
k_β	as above but with α replaced by β.
u_α, u_β	as for k_α and k_β, respectively (Miettinen's notation, used on page 53).
z_1	value of the standard normal deviate corresponding to a two-tailed significance level of α (Snedecor and Cochran's notation, used on page 47).
z_2	value of the standard normal deviate corresponding to a two-tailed significance level of $2(1-P)$, where P is the power of the hypothesis test (Snedecor and Cochran's notation, used on page 47).
p_2	equal to $2(1-P)$ where P is the power of the hypothesis test (as used on page 22).
p_2	probability of success with treatment 2 (as used on pages 46 to 48).
S	random variable whose values equal the number of pairs with discordant responses in a paired Binomial trial (as used on page 52).
s	observed value of S as defined above (as used on pages 52 and 53).
s	observed standard deviation of the data (as used on page 21).

Chapter 2

n	number of plots in the experiment.
p	the number of parameters in linear model.
$\underset{\sim}{X}$	$n \times p$ design matrix.
A-optimality	maximize trace of $\underset{\sim}{X}'\underset{\sim}{X}$.
D-optimality	maximize determinant of $\underset{\sim}{X}'\underset{\sim}{X}$.
E-optimality	maximize minimum eigenvalue of $\underset{\sim}{X}'\underset{\sim}{X}$.
G-optimality	minimize the maximum variance of a predicted observation.
$\underset{\sim}{x}^{t\delta}$	a diagonal matrix whose diagonal elements equal the elements of the vector $\underset{\sim}{x}$ raised to the power t.
v	number of treatments in the design.
$\underset{\sim}{\gamma}$	$v \times 1$ vector of treatment parameters.
$\underset{\sim}{r}$	$v \times 1$ vector of treatment replications.
$\underset{\sim}{\Delta}$	$v \times n$ design matrix for treatments.
b	number of blocks in the design.
$\underset{\sim}{k}$	$b \times 1$ vector of block sizes.
$\underset{\sim}{D}$	$b \times n$ design matrix for blocks .
$\underset{\sim}{N}$	$v \times b$ incidence matrix for the design.
$\underset{\sim}{C}$	equals $\underset{\sim}{r}^{\delta} - \underset{\sim\sim}{N}\underset{\sim\sim}{k}^{-\delta}\underset{\sim}{N}'$.
$\underset{\sim}{C}^-$	a generalized inverse of $\underset{\sim}{C}$, i.e. $\underset{\sim\sim}{C}\underset{\sim}{C}^-\underset{\sim}{C} = \underset{\sim}{C}$.
$\underset{\sim}{\Omega}$, $\underset{\sim}{\Xi}$ and $\underset{\sim}{C}^+$	alternative generalized inverses of $\underset{\sim}{C}$.
$\underset{\sim H}{\Omega}$ and $\underset{\sim H}{\Xi}$	approximations to $\underset{\sim}{\Omega}$ and $\underset{\sim}{\Xi}$, respectively.
$\underset{\sim}{W}$	a $v \times v$ matrix with zero diagonal elements and off-diagonal elements equal to those of $\underset{\sim\sim}{N}\underset{\sim\sim}{k}^{-\delta}\underset{\sim}{N}'$.
$\underset{\sim}{q}$	a $v \times 1$ vector of quasi-replications; $\underset{\sim}{q} = \underset{\sim\sim}{W}\underset{\sim}{1}$.
u	equals $\underset{\sim}{q}'\underset{\sim}{1}$.

Chapter 3

b_1	number of rows in row-and-column design.
b_2	number of columns in row-and-column design.
n	number of plots in design.
$\underset{\sim}{D}_1$	$b_1 \times n$ design matrix for rows
$\underset{\sim}{D}_2$	$b_2 \times n$ design matrix for columns
$\underset{\sim}{N}_1$	$v \times b_1$ incidence matrix for rows
$\underset{\sim}{N}_2$	$v \times b_2$ incidence matrix for columns
$\underset{\sim}{C}$	equals $\underset{\sim}{r}\delta - \underset{\sim}{N}_1\underset{\sim}{N}_1'/b_2 - \underset{\sim}{N}_2\underset{\sim}{N}_2'/b_1 + \underset{\sim}{r}\underset{\sim}{r}'/n$, where $\underset{\sim}{r}$ is a $v \times 1$ vector of treatment replications.
$\underset{\sim}{\Omega}$ and $\underset{\sim}{\Xi}$	alternative generalized inverses of $\underset{\sim}{C}$.

Chapter 4.

$\underset{\sim}{Y}_i$	$i \times 1$ vector of data.
$\underset{\sim}{\beta}$	$p \times 1$ vector of parameters in linear model.
$\underset{\sim}{X}_i$	$i \times p$ design matrix that relates $\underset{\sim}{\beta}$ to $\underset{\sim}{Y}_i$.

For pages 183 - 188

$\underset{\sim}{A}_i$	$i \times v$ design matrix for v treatments in a block design with i plots.
$\underset{\sim}{B}_i$	$i \times b$ design matrix for b blocks in a block design with i plots.
$\underset{\sim}{r}_n$, $\underset{\sim}{k}_n$, $\underset{\sim}{N}_n$, $\underset{\sim}{C}_n$ and $\underset{\sim}{C}_n^-$	are for a block design and are as defined in Chapter 2 except that now there is a subscript to indicate the size of the design.

For pages 195-207

$\underset{\sim}{A}_n$	$n \times v$ design matrix for v treatments in a row-and-column design.
$\underset{\sim}{B}_n$	$n \times (b_1+b_2+1)$ matrix equal to $[\underset{\sim}{D}_1', \underset{\sim}{D}_2', \underset{\sim}{1}_n]$, where $\underset{\sim}{D}_1$ and $\underset{\sim}{D}_2$ are as defined in Chapter 3 and $\underset{\sim}{1}_n$ is a $n \times 1$ vector of ones.
$\underset{\sim}{M}_n$	$v \times b_1$ incidence matrix for the rows of a row-and-column design with n plots.
$\underset{\sim}{N}_n$	$v \times b_2$ incidence matrix for the columns of a row-and-column design with n plots.
$\underset{\sim}{r}_n$ and $\underset{\sim}{C}_n$	are for a row-and-column design and are as defined in Chapter 3 except that now there is a subscript to indicate the size of the design.
$\underset{\sim}{C}_n^-$	a generalized inverse of $\underset{\sim}{C}_n$.

Chapter 1 - Corrigenda

Page 30 line 13 — remove the word 'inversely'.

Page 49 line 17 — replace '2π' by '$\sqrt{2\pi}$' .

Chapter 2 - Corrigenda

Page 72 line 28 — replace '$\underset{\sim}{\mu}$' by 'μ'.

Page 102 line 23 — replace 'six' by 'five'.

Page 111 line 19 — replace first '-' sign by '+' .

Page 112 line 7 — remove ' $+\underset{\sim\sim}{H}\underset{\sim}{1}'/u + \sum_{h=1}^{H}(-1)^h \underset{\sim}{x}_2\underset{\sim}{x}_2'/q$ ' .

Chapter 3 - Corrigenda

Page 118 line 6 — replace '3.7' by '3.8'.

Page 118 line 8 — replace '3.8' by '3.9'.

Page 123 line 20 — replace 't_2' by '$\underset{\sim}{t}_2$' .

Page 127 line 1 — replace 'Ω^{-1}' by '$\underset{\sim}{\Omega}^{-1}$'

Page 134 line 6 — replace '1' by 'zero' and '-1' by '2'.

Chapter 4 - Corrigenda

Page 182 line 22 — replace '$\underset{\sim}{\varepsilon}_i$' by '$\underset{\sim}{\varepsilon}_i'$'.

Page 193 lines 23 and 24 — replace '$\frac{1}{\lambda_{max}}$' by '$\underset{\sim}{1}'\underset{\sim\sim}{\Omega}\underset{\sim}{1}'\frac{1}{v}$' .

Page 194 line 2 — replace 'v-1' by 'v'.

Page 194 lines 4 & 5 — delete all reference to '$\frac{1}{\lambda_{max}}$' from these lines.

Page 194 lines 15,16 and 17 — replace '0.291, 0.285 and 0.281' by '0.304, 0.297 and 0.292', respectively.

Page 197 line 3 — replace '$b_1\underset{\sim}{1}'b_1$' by '$b_1\underset{\sim}{1}_{b_2}$' .