BU-813-M

PRINCIPLES OF STATISTICAL DESIGN WITH SPECIAL REFERENCE TO EXPERIMENT AND TREATMENT DESIGN

by

Walter T. Federer, Cornell University

Abstract

Statistical design has many facets, two of which are experiment design and treatment design. The former refers to the arrangement of treatments in an experiment, while the latter refers to the selection of treatments for the experiment. The frequently used term "experimental design" is not used here because it has been used in many different contexts and besides the "design" should not be "experimental". Sir Ronald A. Fisher laid down the three basic principles of experiment design as replication, randomization, and blocking (local control) in the 1920's. Frank Yates added orthogonality and confounding in 1933. Since then several other principles have been discussed, *viz.* sensitivity, efficiency, balancedness of various types, variance optimality, connectedness, resolvability, and sufficiency. Many of the above principles have been used for treatment design and others have been added, e.g., unbiasedness, rotatability, mean squared error optimality, alias balance, saturation, and symmetry. To many writers of statistical literature, "experimental design" is combinatorics, computing, and/or hypothesis testing. While the first is useful in the construction of plans and in attaining certain properties for the plans, the last two are *not* statistical design. Also, many statisticians confuse *plans* for statistical designs with *planning* investigations. Plans are useful but are only one aspect of planning. Five axioms are given to aid in planning experiments.

PRINCIPLES OF STATISTICAL DESIGN WITH SPECIAL REFERENCE TO EXPERIMENT AND TREATMENT DESIGN

by

Walter T. Federer, Cornell University

BU-813-M

May, 1983
Revised July 1983

Introduction

Statistical design for an investigation encompasses many items; these are:

(i) Complete description of population involved; of the sampling unit (s.u.), the elements making up the population; unit of the experiment (e.u.), or experiment unit, the smallest unit to which one treatment is applied; of the variables to be measured, as well as their distributions; the response equations prior to applying treatments; and the observational unit (o.u.), the smallest unit on which one observation is made.

(ii) Complete description, validation, and standardization of measurements and measuring instruments for the investigation.

(iii) Treatment design, the selection of the treatments to be used in the investigation as well as selection of adequate points of reference (controls, standards, placebos).

(iv) Experiment design, the arrangement of the treatment in the experiment. Randomized designs are preferred over non-randomized ones.

(v) Survey design. If the investigation is a survey of what is in a population, a probability sample survey with proper stratification into relatively homogeneous subgroups is usually preferred.

(vi) Sequential design. In some sequentially conducted experiments, previous results determine which treatment(s) is(are) applied to the next e.u.(s) entering the experiment in a sequential manner.

Paper No. BU-813-M in the Biometrics Unit, Cornell University.

(vii) Modeling and model design. A mathematical expression of results is formulated to explain a phenomenon in an investigation. Selection of an appropriate model or response equation is often a difficult and arduous task involving many experiments. The selection and arrangement of the e.u.'s represents the model design.

(viii) Sample size and replication number, including allocation of s.u.'s and/or e.u.'s to various categories. Statistical criteria, as well as available resources, determine number of samples and replicates.

(ix) Size, shape, and handling of e.u.'s in an experiment (often referred to as experiment technique or plot technique). Too often the technique used is that popular in the field of investigation. Often more efficient techniques can be used, and the experimenter should be aware of this.

(x) Principles and properties of statistical design. Several principles have been formulated for designing experiments. Some of these are the three basic principles enunciated by Fisher (1935), i.e., randomization, blocking, and local control (blocking) and several others such as orthogonality, confounding, sensitivity, balance, efficiency, connectedness, resolvability, variance optimality, and sufficiency. Several of the above and others have been used in treatment designs. The statistical designs obtained by using a certain principle have the property associated with the principle. For example, using the principle of orthogonality, experiment and treatment designs have the property of being orthogonal designs.

Some basic axioms (rules) to follow in designing an experiment are given below. Failure to follow any one of them can lead to meaningless results for the experimenter, with the study ending up as an experience rather than an experiment. One crucial rule which appears to be universally ignored by statisticians and experimenters is:

<u>Axiom I: A complete, precise, and rigorous description of the population to which inferences are made is essential if inferences are to have any meaning.</u>

It is insufficient to say "the results of this experiment apply to the population of which it is a sample". What is the population precisely? Textbooks and literature, statistical or otherwise, ignore this fact for the most part. Any reference to the population is more of a "hand-waving" description. If one does not know what the population is, any inferences are meaningless.

<u>Axiom II: Design for the experiment; do not experiment for the design.</u>

All too often an experimenter and statistician feel their choice of an experiment and/or treatment design is limited to those appearing in tables or in the literature. The experiment should be considered as it is to be conducted rather than being changed to fit a tabled design. Treatments need not be added or dropped to make the experiment fit a tabled design. Following Axiom II means that new designs may have to be constructed to meet the requirements of the experiment. One can easily construct many types of designs that are not in tables, using simple construction procedures (see, e.g., Patterson and Williams, 1976, and Khare and Federer, 1981).

<u>Axiom III: Use the minimum amount of blocking possible to control heterogeneous variation among the e.u.'s.</u>

The main idea in blocking is to group the e.u.'s in such a manner that the variation among e.u.'s within a group is a minimum and that among groups is a maximum, and this should be done with as few groups as is possible. There is no need to allocate extra degrees of freedom to blocking when they should be allocated to error variance. For example, if one has relatively uniform groups of seven e.u.'s and has four treatments, the block size should be seven, not four. Each of the treatments but one can be included twice in each block, and a different treatment is omitted from each block. For more than four blocks, the procedure can be repeated until the desired number of blocks are obtained. (An experiment design of this type does not appear in tables of designs.) Also, if several sources of experimental variation can be grouped together, this should be done in order to minimize blocking. For example, a block in the greenhouse, the day of transplanting material from greenhouse to field, and a block in a field can all be called a block rather than blocking on three separate sources of variation and using a complex experiment design. If there is no grouping of e.u.'s into subpopulations, then there should be no blocking. Complex designs should be avoided <u>if</u> simpler designs will suffice. Ease of statistical analyses, especially when data are lost, and interpretations are additional reasons for using a design with minimum blocking.

<u>Axiom IV</u>: <u>Treatments with different numbers of randomizations will have different numbers of replicates, will have different e.u.'s, and will have different error variances.</u>

Consider the two sets of data, 100 measurements on one plant and one measurement from each of 100 plants. Means, variance, and interval estimates are computed using exactly the same formulae for both sets of data. The individual interested in the computing aspects would fail to distinguish between the two populations to which the data refer. The first set refers only to the population of measurements on the <u>one plant</u>. The second set refers to the population which is a mixture of two populations, plants and measurements. Thus, if the one plant was randomly selected from the population of plants, the sample size is one, no matter if a million measurements on that one plant are made. In the second case, 100 randomizations were used and the sample size of plants is 100. The differences in this example should be obvious to almost everyone, but in many situations the distinction may be rather subtle. This is especially true when the e.u.'s for different treatments are different sizes and involve different blocking patterns.

Axiom V: A valid error variance for the difference between two treatment effects must contain all sources of variation in the e.u.'s except that due to the treatments themselves.

This is Sir R. A. Fisher's (1935) definition of a valid estimate of an error variance for differences of treatment effects. The error variance is not necessarily the last line in an analysis of variance table or some term designated as "error" on computer output. Every single contrast in an experiment could have a different error variance because of differences in variances for each treatment, differences in size of e.u.'s for each treatment or groups of treatments, use of an incorrect response model, or different units of measurement for different treatments.

Many, many experiment designs and classes of designs have been devised by statisticians and mathematicians, but they have been constructed without any consideration of a particular experiment. For these designs, a statistical analysis has been provided for a (sometimes erroneously called the) linear model. It should be noted that the linear model provided may be incorrect and inappropriate, leading to incorrect statistical analyses and inferences. In using any experiment design, the above five axioms should be kept in mind at all times.

Principles of Experiment Design

Fisher (1935) first presented the three principles of experiment design known as randomization, replication, and blocking (see Figure 1). Figure 2 is an extended version of Figure 1 as given in Federer and Balaam (1972). The principles of orthogonality and confounding were discussed by Yates (1933, 1937). Blocking has been discussed to some extent in the previous section. Replication refers to the number of e.u.'s on each treatment, not necessarily the number of observations. Perhaps the most frequently occurring

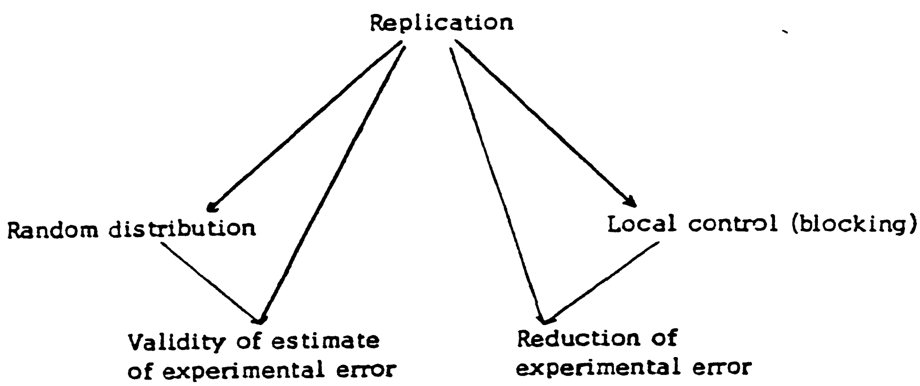

Figure 1. Fisher's principle of design of experiments.

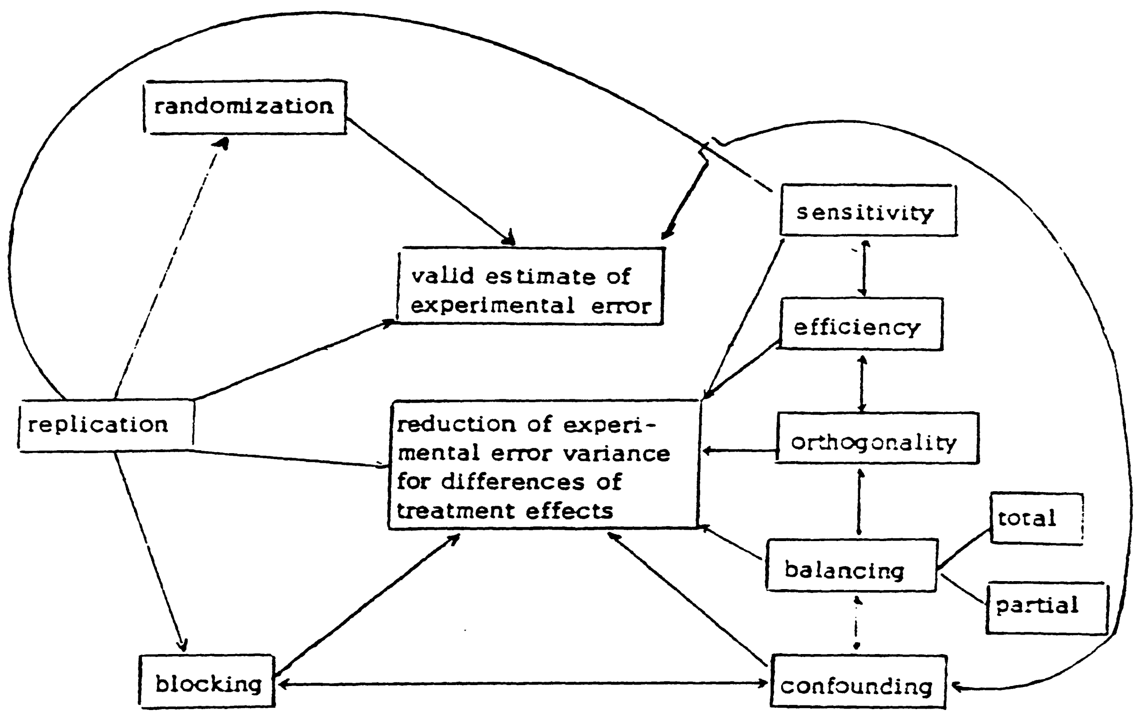

Figure 2. An expanded version of Fisher's diagram.

mistake in scientific and textbook literature is to have one block of a randomized complete block design with multiple observations on each treatment and then to consider this as a completely randomized design. The reason for this mistake is that the concepts of replication and valid estimate of error variance are not understood. Unless the e.u. is precisely defined, used, and understood, the idea of replication remains fuzzy at best. Additional principles have been added to the above five and no doubt more will be added in the future. Some of the additional ones discussed below are balance, sensitivity, efficiency, connectedness, optimality, resolvability, and sufficiency.

Replication

The replication principle is a necessary part of scientific investigation. Experiment units are replicated to study treatment variation from e.u. to e.u. The number of e.u.'s, not the number of s.u.'s or o.u.'s, determine the number of replications in an experiment. Entire experiments may need to be replicated (repeated) by other experimenters to validate the results. The idea of repeatability of results by other investigators is a basic tenet of scientific inquiry. Experiment designs which have the property of being replicated designs are an essential tool in the hands of competent scientists. The person who said "Damn the duplicate plot, give me one and I know where I am" was not much of a scientist. He wanted to ignore the universal fact that e.u.'s differ and that repeatability of results of scientific investigations is a necessity.

Randomization

Randomization is the basis for obtaining a valid estimate of a treatment effect and valid estimate of error variation. The concept of randomization has been unclear for many, in that they do not understand how blocking and randomization can both be used. The confinement of randomization within blocks bothers some. This should not be a point of confusion, because the restriction on all possible permutations to the subset of permutations that can occur within blocks does not vitiate obtaining valid estimates of treatment effects and error variances. However, the complete restriction to a single permutation results in a systematically arranged experiment design, which usually results in biased estimates of treatment effects and of error variances. Hence, such designs are to be avoided.

The randomization principle is difficult for certain scientists because they feel they know how to place treatments in an experiment. Very few, if any, investigators can judiciously place treatments in an experiment and obtain unbiased results. The fundamental nature of a human being implies biasedness. In order to be fair to the treatments under investigation, the experimenter should block the experimental material as best as possible. Then, within each of the blocks, treatments are randomly allocated to the e.u.'s. Random allocation provides equal fairness to treatments in an experiment.

Blocking

Basically, there are four methods of reducing the error variance of a difference between estimated treatment effects. Blocking as discussed above is one method. The second procedure is to refine experiment techniques. The third procedure is to measure related extraneous variables and use covariance techniques; this last method has often been overlooked and/or misused. Continuous variates classified into groups and called blocks, e.g., weight of plants, is an inefficient method. Instead, covariance should be used as fewer degrees of freedom are usually used for covariates than for blocking. If none of the above three methods suffice, the alternative is to increase replication number (sample size) until the degree of precision desired is attained.

The blocking principle implies that extraneous, non-treatment variation in the experimental material is controlled and that this variation can be removed from estimates of treatment variation. Blocking can be used to reduce experiment error, thereby increasing the precision of estimated treatment effects. Blocking can reduce the amount of experimental material required for a specified degree of precision. In general, blocking is not well understood by experimenters, whether in agriculture, medicine, or some other field. The same holds for several types of statisticians. Theorems on and procedures for blocking are relatively non-existent. One exception is Vithayasai and Robson (1974) who have given a blocking strategy for an experiment.

Orthogonality

The principle of orthogonality is a desirable one, as orthogonal experiment designs have the property that the various sets of effects are orthogonal to each other, that simple statistical analyses result, that valid estimates of effects and their variances are obtained, and the greatest precision usually results. Simple arithmetic averages may be used to obtain the differences between estimated effects, and the variances of these differences are easily computed when a linear model holds. Since there is no mixing up of effects (confounding) in orthogonal experiment designs, there is no loss in information due to this source, and maximum precision results.

As Preece (1977) has amply demonstrated, the term "orthogonal designs" has been used in at least three distinct ways. In addition, the adjective "orthogonal" has several conflicting meanings as does the noun "design". The reader is referred to Preece's excellent paper for a complete and up-to-date discussion on orthogonality. An early paper on the subject was written by Yates (1933). To illustrate some of the difficulties, suppose that we wish a combinatorial definition of orthogonality. Let r_{ij} equal the number of times treatment i occurs in block j. Now consider the relative proportions of times that treatments 1, 2, $\cdots$, v occur in a given block j, i.e., $r_{1j} : r_{2j} : r_{3j} : \cdots : r_{vj}$. If these relative proportions are the same for every block j, we say that treatment effects are orthogonal to block effects (Federer, 1973). An additive linear model formulation for this situation would be given the following definition of orthogonality (Federer, 1973). If differences between arithmetic means for any and all pairs of means, contains only treatment and random error terms, treatments are orthogonal to all other effects in the linear model. If one had a multiplicative model, the definition would be altered as follows: the ratios of pairs of any two geometric means, ignoring random error terms, contains only the ratio of treatment parameters. For example, the ratio of treatment geometric means for treatments one and two, ignoring error terms, is $\mu\tau_1\left(\prod_{j=1}^{r}\beta_j\right)^{1/r}/\mu\tau_2\left(\prod_{j=1}^{r}\beta_j\right)^{1/r} = \tau_1/\tau_2$. The combinatoric definition is not model dependent as are the other two.

Use of the orthogonality principle has several advantages. However, the user must first decide what type of orthogonality is being considered for his situation. Other principles may determine that a non-orthogonal design will be the most precise in terms of cost and utilization of material. One such example is discussed under balance.

Confounding

Orthogonal experiment designs do not exist for all situations, and hence it is necessary to use non-orthogonal designs. Whenever non-orthogonality is present there is a mixing up (confounding) of the effects in an experiment. If two or more effects are not separable, the effects are said to be completely confounded. Otherwise, when the effects can be separately estimated and they are not orthogonal, they are said to be partially confounded. For example, consider the two design plans for $v = 3$ treatments, A, B, C, and for $b = 4$ blocks:

Plan I Block 1	2	3	4	Plan II Block 1	2	3	4
A	A	B	C	A	A	A	B
A	A	B	C	A	A	A	B
A	A	B	C	B	B	A	C
A	A	B	C	C	C	A	C

In the first design plan treatment A effects and blocks 1 and 2 are completely confounded. No separate estimates of the A effects or of the blocks 1 and 2 effects can be obtained. In design plan II, the effects are partially confounded, and they can be estimated separately. If treatment A had been in twice in every block and if treatments B and C had been included once in each of the four blocks, the plan would have been an orthogonal design plan with blocks and treatment effects being completely unconfounded, i.e., orthogonal. The principle of confounding is useful for many design situations. If there are more treatments than can be included in a block, one must use an incomplete block design with partial confounding of treatment and block effects. When less information is required on some effects than others, or when it is necessary by the nature of the treatment, the principle of confounding may be invoked to obtain smaller blocks than can accommodate all the treatments in each block.

When there is no confounding of various types of effects in a statistical design, orthogonality results. When there is complete confounding, effects cannot be estimated in the experiment. Confounding results in loss of precision and statistical information. In order to have confounding of block and treatments, blocking must be present, illustrating the connection between the principles of blocking and confounding. If treatment effects are of equal importance, confounding should be avoided whenever possible and desirable.

Balancedness

Balancedness of effects is another useful principle in selecting and constructing design plans for an experiment. Balancedness and orthogonality are sometimes both obtainable, but in general different classes of experiment designs are obtained. Balancedness refers to a relationship between blocking and treatment effects. Hence, this illustrates the connection between the two principles. There are many types of balance (see Hedayat and Federer, 1974, and Preece, 1982), but attention here is confined mostly to variance balance; i.e., when the errors in the experiment come from a distribution with a single variance parameter (homoscedastic), all differences between treatments will have the same variance. When orthogonality cannot be obtained, then balancedness is a desirable second choice. It is relatively easy computationally to obtain solutions for the effects in most cases. There are situations where a balanced design has a lower error variance for differences of means than a competing orthogonal experiment design (see Shafiq and Federer, 1979). For example, consider the following three designs for treatments A, B, and C.

I. Complete Block

1	2	3
A	A	A
A	A	A
A	A	B
B	B	B
B	B	B
B	C	C
C	C	C
C	C	C

II. Complete Block

1	2	3	4
A	A	A	A
A	A	A	A
B	B	B	B
B	B	B	B
C	C	C	C
C	C	C	C

III. Complete Block

1	2	3	4	5	6	7	8
A	A	A	A	A	A	A	A
B	B	B	B	B	B	B	B
C	C	C	C	C	C	C	C

Suppose that groups of k = 8 relatively homogeneous e.u.'s were available. One could form blocks of k = 8 as in plan I, blocks of k = 6 as in plan II, or blocks of k = 3 as in plan III. Plan I would require 3 groups, II 4 groups, and III 8 groups. II would cost 1/3 more than I as one more group would be required, and III would cost $8/3 - 1 = 5/3$ more than I and twice as much as II. If each e.u. costs the same regardless of which group it is in, then the three plans would cost the same. However, there are many situations where the e.u.'s do not cost the same and some e.u.'s would need to be discarded to obtain plans II and III. For example, suppose that there were 8 positions in an oven for baking pies. There would only be 3 bakings for I but 8 for III, resulting in considerably more expense.

Plan I is variance balanced, and II and III are both variance balanced and orthogonal. This assumes a linear model for effects. The three designs are also pairwise balanced in that every pair of treatments occurs together in the blocks an equal number of times λ; note that $\lambda = 21$ for I, 16 for II, and 8 for III.

As stated above, there are many types of balance. Although it took Preece (1977) 23 pages to discuss the orthogonality muddle, it took him (1982) 102 pages to discuss the balance tangle. His papers should be read by users of these two terms.

Efficiency

The principle of efficiency is useful in deciding which one or ones of a class of experiment designs is optimal (best) with respect to a selected criterion. Making use of this principle, designs can be ranked with respect to their relative efficiencies using the selected criterion. Orthogonal designs usually have 100% relative efficiency with respect to variance (perhaps not if cost and variance are both considered). Partially confounded designs have efficiency less than 100%, as some information is lost due to the confounding. The efficiency principle says to utilize the most efficient design at one's disposal and suited for the experiment. In almost all situations discussed in statistical literature, efficiency is related to variances of treatment contrasts from two different designs. For example, the three experiment designs given under balancedness have average variances of a difference between two treatment means of $16\sigma_\epsilon^2/63$, $\sigma_\epsilon^2/4$, and $\sigma_\epsilon^2/4$. The relative intrablock efficiencies, when corrected for the differences in degrees of freedom for the remainder sums of squares (i.e., 19, 18, and 14, respectively), are 0.989, 1.000, and 0.975. This example shows that even ignoring cost of the additional groups required for designs II and III, a balanced design is more efficient than a competing orthogonal design, III, and is almost as efficient, 0.989, as the best competing orthogonal design. If one considers the additional cost, this balanced incomplete block design is more efficient than either of the competing orthogonal designs. Hence, it is incorrect to say that an orthogonal design is always more efficient than a balanced block design.

Another fallacious statement which frequently appears in statistical literature is that a balanced incomplete block design is variance optimal among all incomplete block designs. That this statement is false is illustrated in the examples below for treatments A, B, C, and D:

Plan IV. Balanced Incomplete Block

Block

1	2	3	4	5	6
A	A	A	B	B	C
A	A	A	B	B	C
B	C	D	C	D	D
B	C	D	C	D	D

Plan V. Partially Balanced Incomplete Block

Block

1	2	3	4	5	6
A	A	A	A	A	C
B	B	B	B	A	C
C	C	C	C	B	D
D	D	D	D	B	D

The average variance of a difference for plan IV is $\sigma_\epsilon^2/2$ and that for plan V is $7\sigma_\epsilon^2/18$. Since plan V has a smaller variance than IV, it demonstrates that IV is not variance optimal. The statement is only true if occurrences of treatments is limited to the case where treatment i occurs in block j either zero or once, i.e., $n_{ij} = 0, 1$. It is not true in general for $n_{ij} = m_0, m_1$.

Another erroneous statement is that a binary design is always more efficient than a ternary design, a ternary than a quaterary, etc. It is only true for basic designs where $n_{ij} = 0, 1, 2, \cdots$ (see Shafiq and Federer, 1979). It is not true for $n_{ij} = m_0, m_1, m_2, \cdots$. The above example demonstrates this. In plan IV, $n_{ij} = 0, 2$, a binary design. Plan V is a ternary design, $n_{ij} = 0, 1, 2$, and has a smaller variance.

Other criteria are often necessary in experimentation. For example, in a nutrition experiment, a variance minimal design was selected for one-half of the experiment because it was nutritionally optimal in the sense of having other nutritionists believe the results of the experiment. The other half was variance optimal, thus compromising nutritional and statistical optimality. Other criteria could be cost, ability of technicians to conduct the experiment, etc.

Optimality

The optimal principle states that the most optimal experiment design be utilized in conducting experiments. When variance optimality is used, this ensures the greatest precision with minimum expenditure of experimental material. There are many criteria which could be utilized in setting up optimality measures. A great majority of the statistical literature on optimality deals with variance optimality, with the adjective variance frequently being omitted. One could set up measures of combinatorial optimality or optimality criteria for a multiplicative response model. Variance optimality is very much related to efficiency.

Connectedness

The connectedness principle is related to the blocking and confounding principles. Totally confounded designs are not connected. Unconfounded (orthogonal) and partially confounded designs are connected, i.e., all estimable treatment contrasts can be estimated from the experiment design. This principle should be applied to all proposed experimental procedures to determine confounding structures and connectedness. The first plan given under confounding is not connected but the remaining plans given are.

Some relatively recent papers on this principle and on designs which have or have not the property of being connected are by Vartak (1963), Eccleston and Hedayat (1974), Eccleston and Russell (1975) and by Raghavarao and Federer (1975).

Resolvability

The principle of resolvability is to have all treatments in a complete block regardless of how many incomplete blocks are used. Resolvable designs are useful when alternate analyses may be used, such as, e.g., a randomized complete blocks design analysis for an experiment actually laid out as an incomplete block design. Occasionally it may be desirable to eliminate one or more complete blocks in the analysis.

Sensitivity

The sensitivity of a statistical design first reached attention of researchers when Fisher (1935) used the example of a lady tasting tea where the two kinds of tea were milk first and then tea versus tea with milk infused. Some literature on sensitivity since that time include papers by Bradley and Schumann (1957), Dar (1962, 1964), Lashof et al. (1956), Mandel and Stiehler (1954), and Schumann and Bradley (1957). In the first citation above, several examples are given illustrating a need for comparing sensitivities of various experimental procedures in several different fields. Experimenters need to select procedures which will be the most sensitive for attaining the goals of the experiment. For example, measurement of time to recurrence of cancer after treatment is a much more sensitive measurement than is measuring whether or not the disease recurred. Measuring the effect of applying nitrogen on the amount found in the plant can be done much more effectively in samples from the 8 - 10 node area than in the leaves. The 8 - 10 node area samples were able to detect a 25 lb. increase in nitrogen fertilizer with the same precision as leaf samples could detect a 150 lb. increase. Thus the 8 - 10 node samples were six times as sensitive as the leaf samples in detecting additional applications of nitrogen fertilizer. Samples taken at the base of the sugar cane stalk would have been even more sensitive in measuring changes in nitrogen levels.

Sufficiency

The sufficiency principle states that an experiment design and a given response model should provide statistics that contain all the information in the experiment. Experiment designs yielding a minimal set of sufficient statistics which is complete with respect to a given response model would be desirable. For example, Graybill and Weeks (1959) show that Yates' combined interblock and intrablock estimator is based on a set of minimal sufficient statistics but the set is not complete. The sufficiency principle needs much more amplification in statistical literature with regard to experiment design.

Principles of Treatment Design

Treatment design (the selection of treatment for an experiment) can be categorized into the following types:

(i) controls, standards, placebos, or other points of reference,

(ii) discrete levels of one or more factors (factorial designs),

(iii) continuous level of one or more factors (regression or response surface designs),

(iv) single level of two or more factors in mixtures (diallel crossing, tournaments, matched pairs, intercropping, etc.), and

(v) combinations of the above.

The principles of treatment design discussed below apply to each of the above types. Principles found useful in treatment design are efficiency, variance optimality, replication, orthogonality, balance, symmetry, unbiasedness, connectedness, saturation, rotatability, and mean squared error.

Replication

Replication of treatments (or combinations) in an experiment should be related to their importance. Otherwise replication should be equal. For fractionally replicated experiments, a certain class of treatment effects needs to be zero or at least small in comparison with other classes of treatment effects. In special cases, a fractional replicate is all that is possible, e.g., diallel cross, tournaments, etc. In some cases, missing subclass combinations leads to difficulties in statistical analyses and in interpretation of effects. The replication principle needs to be considered seriously for any fractional replicate of a treatment design. Likewise, the replication principle can be applied to repetitions of a fractional replicate, i.e., a fractional replicate should not be repeated unless one is certain that the assumed class of effects is actually zero. Instead another fraction and different combinations should be selected. This procedure tends toward completion of a full replicate of a treatment design. After a full replicate is achieved then replication of a combination should be considered.

Orthogonality

The orthogonality principle for a treatment design ensures ease of statistical analysis and variance optimality problems. Use of this principle has led to construction of such fractional replicates of factorial designs as orthogonal main effect plans, orthogonal resolution V plans, etc. Use of the principle to obtain orthogonal plans leads to more precise interpretation of effects, i.e., they are unentangled with other effects.

Confounding (aliasing)

When fractions of a full treatment design are less than one, some effects will necessarily be confounded with other effects. The principle used here is to confound effects of most importance with those that are small or negligible. This would allow zero or small bias in the estimates of important effects. When the magnitude of treatment effects is unknown, confounding of effects should be avoided. The less important effects which are confounded, totally or partially, are denoted as aliases of the more important effects. Following Raktoe et al. (1981), one could use the principle that the type of confounding for all important effects should be similar in nature, i.e., a balanced arrangement of aliases.

Symmetry

The symmetry principle is to have the same number and type of levels for each of the factors under study when the factors are of equal importance. One may know something about the response function for certain factors, one may be interested in the response for one factor over a smaller range than others, or one may wish less detailed information on one or more of the factors. If any of these were true, non-symmetrical or asymmetrical levels would be selected.

Saturation

Saturated designs are parsimonious designs in that the number of observations taken is equal to the number of parameters to be estimated. In many situations, e.g., some quality control investigations, observations are very costly. Hence, a minimal number of observations are required to estimate a parameter set, say $\underline{\beta}_1$. The saturation principle leads to construction of saturated treatment designs. Of course, in the class of saturated designs one would use the one(s) which have variance optimality properties.

Efficiency

The efficiency principle for treatment design applies in various ways. First, a treatment design should be selected to meet the goals of the experiment. Unless this is done, more than one experiment may be necessary to meet the stated goals, resulting in wasteful expenditure of resources. Secondly, the levels of factors or variables must be selected with care. Otherwise, the goals of the experiment may not be achieved or achieved with low precision, resulting in low efficiency. This principle is closely tied to variance optimality, which follows, when the goal is to optimize a procedure with respect to a stated criterion.

Balance

As noted for experiment designs and by Preece (1982), there are many types of balance for treatment designs and many uses of the term. Many phrases have been utilized as is amply demonstrated in Preece's (1982) discussion of balance. Others could be added from the literature on linear models (e.g., Searle, 1971). In utilizing the balance principle, one can say that in general most uses of the principle have some desirable properties, and it is essential that the writer precisely defines the context in which the term balance is being used. Otherwise, considerable confusion can arise.

Yates (1936, 1937) introduced the idea of balance in constructing confounded arrangements for factorial treatment design. A balanced arrangement is one for which all interactions of a specified order are confounded equally. Thus, suppose one confounds all three-factor interactions once with blocks and all four-factor interactions three times with blocks; this would be a balanced confounding arrangement. The idea is that if interactions of a specified degree are of equal importance, equal intrablock information should be obtained on each one. Unbalanced confounding would be used when unequal amounts of information are desired. The degree of confounding is related to the degree of importance of effects.

The same concept can be carried over to regression designs. All trends of a specified degree can be confounded equally, assuming that they are of equal importance to the experimenter.

Raktoe et al. (1981) set up measures of alias optimality and alias balance. If A is the aliasing matrix, then when the absolute value of the determinant of s is minimal, this is denoted as alias optimal. Let a_{hj} be the element in the h'th row and j'th column of A. Then a measure of alias balance for a design is given below:

$$Q_8 = \sum_{h=1}^{s} \left[\left(\sum_{j=1}^{s} a_{hj} \right)^{\frac{1}{2}} - \frac{\sum_j (\Sigma a_{hj})^{\frac{1}{2}}}{s} \right]^2 .$$

When Q_8 is a minimum, zero, the design is alias balanced.

Variance Optimality

In statistical literature considerable emphasis is placed upon variance optimality or precision of estimating treatment effects. Many variance optimality criteria have been established. Some of these are (e.g., see Kiefer, 1959, and Raktoe et al., 1981):

(i) D-optimal or determinant optimal – minimum value of determinant of inverse of the information matrix.

(ii) A-optimal or average variance optimal – average variance of linear contrasts of treatment effects is minimal.

(iii) E-optimal or eigenvalue optimal – smallest eigenvalue is a maximum over the class of designs possible.

(iv) G-optimal or global optimal – largest variance for any treatment effect is minimal.

The variance optimality principle is closely tied to the efficiency principle. Use of either principle often leads to the same treatment designs.

Unbiasedness

In the absence of knowledge about the size and nature of the bias, the unbiasedness principle would indicate that unbiased treatment designs would be utilized. These designs yield unbiased estimates of all treatment effects. Suppose that one has a full or complete treatment design with parametric vector $\underline{\beta}$, with an observation for every treatment, say $\underline{Y}$, and with a design matrix X. Suppose the partitioning is as follows:

$$E(\underline{Y}) = E\begin{bmatrix}\underline{Y}_1\\ \underline{Y}_2\\ \underline{Y}_3\end{bmatrix} = X\underline{\beta} = \begin{bmatrix}X_{11} & X_{12} & X_{13}\\ X_{21} & X_{22} & X_{23}\\ X_{31} & X_{32} & X_{33}\end{bmatrix}\begin{bmatrix}\underline{\beta}_1\\ \underline{\beta}_2\\ \underline{\beta}_3\end{bmatrix}.$$

If one takes a fraction $\underline{Y}_1$, say, then

$$E(\underline{Y}_1) = [X_{11}\ \ X_{12}\ \ X_{13}]\begin{bmatrix}\underline{\beta}_1\\ \underline{\beta}_2\\ \underline{\beta}_3\end{bmatrix} = X_{11}\underline{\beta}_1 + X_{12}\underline{\beta}_2 + X_{13}\underline{\beta}_3 .$$

If one wishes to estimate $\underline{\beta}_1$, say, then

$$\hat{\underline{\beta}}_{11} = (X_{11}'X_{11})^{-1}X_{11}'\underline{Y}_1 - (X_{11}'X_{11})^{-1}X_{11}'X_{12}\underline{\beta}_2 - (X_{11}'X_{11})^{-1}X_{11}'X_{13}\underline{\beta}_3 .$$

If either, or both, $\underline{\beta}_2$ or $\underline{\beta}_3$ are not identically equal to zero, the above is a biased estimate of $\underline{\beta}_1$. If $\underline{\beta}_2$ and $\underline{\beta}_3$ are small compared to $\underline{\beta}_1$, then the bias is small.

Let $A = (X_{11}'X_{11})^{-1}X_{11}'X_{12}$ and let $\underset{\sim}{\beta}_3 \equiv 0$, then A is called the aliasing matrix of $\underset{\sim}{\beta}_2$ with respect to $\underset{\sim}{\beta}_1$. If $\underset{\sim}{\beta}_3 \neq 0$, and if $A^* = (X_{11}'X_{11})^{-1}X_{11}'X_{13}$, then the aliasing matrices would be

$$[A \quad A^*] \quad \text{for the vector} \quad \begin{bmatrix} \underset{\sim}{\beta}_2 \\ \underset{\sim}{\beta}_3 \end{bmatrix}$$

with respect to $\underset{\sim}{\beta}_1$. Raktoe <u>et</u> <u>al</u>. (1981) discuss four situations regarding whether or not $\underset{\sim}{\beta}_2$ and/or $\underset{\sim}{\beta}_3$ are zero.

To simplify the discussion in later sections, let $\underset{\sim}{\beta}_3 \equiv 0$.

Connectedness

As in experiment design, it is desirable that all non-zero effects be estimable. For the situation in the previous section, one needs to have a $\underset{\sim}{Y}_1$ and a $\underset{\sim}{Y}_2$ that allow estimation of $\underset{\sim}{\beta}_1$ and $\underset{\sim}{\beta}_2$. In the event that $\underset{\sim}{\beta}_2 \equiv 0$, then $(X_{11}'X_{11})^{-1}$ must exist in order to estimate $\underset{\sim}{\beta}_1$. When all estimable effects can be estimated from a design, the design is said to be connected. During the course of statistical consulting and in reading scientific literature, experiment and/or treatment designs that are not connected are utilized. This is a wasteful use of resources and should be discontinued.

Mean Squared Error

Accuracy includes precision (variance) plus bias. The mean squared error principle states that the mean squared error (MSE), which is the variance plus the square of the bias, should be minimal in order to obtain the most accurate measurement of a parametric vector, say $\underset{\sim}{\beta}_1$, i.e.,

$$\text{MSE}(\hat{\underset{\sim}{\beta}}_1) = \text{cov}(\hat{\underset{\sim}{\beta}}_1) + A\underset{\sim}{\beta}_2\underset{\sim}{\beta}_2'A' ,$$

where cov($\hat{\beta}_1$) means the covariance matrix for $\hat{\beta}_1$. Raktoe *et al.* (1981) define a design to be alias optimal if the absolute value of the determinant of A is *minimal* and alias balanced if the sum of squares of square roots of the sums of squares of coefficients of the rows of A is zero. They discuss some of the problems in implementing the mean squared error criterion. Further discussion also appears in a recent paper by Welch (1983) on response surface designs.

Rotatability

Rotatability is the principle that states that all points equidistant from the center of the region in regression designs have equal variance (see Box and Hunter, 1957). If levels of factors from the center for all factors are of equal importance, this is a desirable criterion to use. If they are not, then some sort of ellipsoidal criterion should be used in the definition of rotatability.

Discussion

It seems odd that Sir Ronald A. Fisher did not relate the principles of statistical design given in his design book (1935), *viz.* replication, randomization, and local control with the principles of estimation given in his statistical methods book (1925), *viz.* efficiency, sufficiency, and consistency. It would appear that principles of statistical design and principles of statistical estimation should be considered together rather than as separate entities. Should the estimation principle of consistency be used as a statistical design principle? Perhaps not. Should only mathematical and statistical considerations be used in establishing a principle, or should applications of the procedure in actual practice dictate the principles to be used? Are there nonstatistical principles that should be used in statistical design? What is the relative importance of the above discussed principles in the real world where these designs are used?

An overview of the principles of design and estimation indicates that there are many unanswered questions and that the philosophical aspects need further study.

Literature Cited

Box, G. E. P. and J. S. Hunter (1957). Multi-factor experimental designs for exploring response surfaces. Annals of Mathematical Statistics 28, 195-241.

Bradley, R. A. and D. E. W. Schumann (1957). The comparison of sensitivities of similar experiments: Applications. Biometrics 13, 496-510.

Dar, S. N. (1962). On the comparison of the sensitivities of experiments. Journal of the Royal Statistical Society B 24, 447-453.

Dar, S. N. (1964). Comparison of sensitivities of dependent experiments. Biometrics 20, 209-212.

Eccleston, J. A. and A. Hedayat (1974). On the theory of connected designs: Characterization and optimality. Annals of Statistics 2, 1238-1255.

Eccleston, J. A. and K. Russell (1975). Connectedness and orthogonality in multi-factor designs. Biometrika 62, 341-345.

Federer, W. T. (1973). Statistics and Society. Marcel Dekker, Inc., New York.

Federer, W. T. and L. N. Balaam (1972). Bibliography on Experiment and Treatment Design Pre 1968. Published for the International Statistical Institute by Oliver and Boyd, Edinburgh.

Fisher, R. A. (1925). Statistical Methods for Research Workers (first of several editions). Oliver and Boyd, Edinburgh.

Fisher, R. A. (1935). The Design of Experiments (first of several editions). Oliver and Boyd, Edinburgh.

Graybill, F. and D. L. Weeks (1959). Combining inter-block and intra-block information in balanced incomplete blocks. Annals of Mathematical Statistics 30, 799-805.

Hedayat, A. and W. T. Federer (1974). Pairwise and variance balanced incomplete block designs. Annals of the Institute of Statistical Mathematics 26, 331-338.

Khare, M. and W. T. Federer (1981). A simple construction procedure for resolvable incomplete block designs for any number of treatments. Biometrical Journal 23(2), 121-132.

Kiefer, J. (1959). Optimum experimental designs. Journal of the Royal Statistical Society B 21, 272-304.

Lashof, T. W., J. Mandel, and V. Worthington (1956). Use of the sensitivity criterion for the comparison of the Bekk and Sheffield smoothness testers. Tappi 39, 532-543.

Mandel, J. and R. D. Stiehler (1954). Sensitivity - A criterion for the comparison of methods of the test. Journal of Research, National Bureau of Standards 53, 155-159.

Patterson, H. D. and E. R. Williams (1976). A new class of resolvable incomplete block designs. Biometrika 63, 83-92.

Preece, D. A. (1977). Orthogonality and designs: A terminological muddle. Utilitas Mathematica 12, 201-223.

Preece, D. A. (1982). Balance and designs: Another terminological tangle. Utilitas Mathematica 21C, 85-186.

Raghavarao, D. and W. T. Federer (1975). On connectedness in two-way elimination of heterogeneity designs. Annals of Statistics 3, 730-735.

Raktoe, B. L., A. Hedayat, and W. T. Federer (1981). Factorial Designs. John Wiley and Sons, Inc., New York.

Schumann, D. E. W. and R. A. Bradley (1957). A comparison of the sensitivities of similar experiments: Theory. Annals of Mathematical Statistics 28, 902-920.

Searle, S. R. (1971). Linear Models. John Wiley and Sons, Inc., New York.

Shafiq, M. and W. T. Federer (1979). Generalized N-ary balanced block designs. Biometrika 66, 115-123.

Vartak, M. N. (1963). Connectedness of Kronecker product designs. Journal of India Statistical Association 1, 215-218.

Vithayasai, C. and D. S. Robson (1974). A blocking strategy. Communications in Statistics 3(11), 1041-1052.

Welch, W. J. (1983). A mean squared error criterion for the design of experiments. Biometrika 70, 205-213.

Yates, F. (1933). The principles of orthogonality and confounding in replicated experiments. Journal of Agricultural Science XXIII, part I, 108-145.

Yates, F. (1936). Complex experiments (with discussion). Supplement, Journal of the Royal Statistical Society 2, 181-247.

Yates, F. (1937). The design and analysis of factorial experiments. Technical Communication No. 35, Imperial Bureau of Soil Science, Harpenden, England.