The Theory of Designed Experiments

7. The Factorial Treatment Structure

UNESP-Botucatu, August 2010

Example

A bioreactor is to be designed for the enzymatic synthesis of derivatives of sucrose. The physical design of the bioreactor and the conditions applied to the substrate have to be optimized. A number of questions can be asked: which factors and how many should be studied? Which levels of the factors should be used? Should all combinations of all levels be used? How many replicates of each combination should be run?

Choice of factors

In the example, the experimenters quickly identified five factors: pH, total sugar, donor:acceptor ratio, temperature and mix speed. Should these all be varied in one experiment, or should a few be studied first?

It is good practice to get experimenters to list all factors which *might possibly* have an effect before deciding which ones to actually vary in the experiment.

Choice of factors

The number of factors actually varied will depend on the resources available for the experiment, the amount of blocking needed and the number of levels used for each factor. The only restriction on the number of factors used is usually the total number of units available. Sufficient degrees of freedom should be left for estimating error.

A useful concept is the *resource equation*:

$$n = \frac{n}{n_b} + p + n_r,$$

where n is the number of units, n_b is the number of units in each block, p is the number of parameters for the treatment model and n_r, the residual d.f., should usually be about 10-20.

Choice of levels

For qualitative factors when fixed effects are of interest, all categories have to be used, e.g. if new and standard drugs and a placebo are of interest, all must be used in the experiment.

If random treatment effects are of interest, a large enough sample from the population must be chosen to get a reliable estimate of the variance component. How large is this? In practice, relatively small samples are often used.

For continuous factors, the number of levels chosen should usually be the fewest which will meet the objectives, as when a single continuous factor is under study.

Choice of levels

- Two levels allow a simple change in response to be detected and might be appropriate at early, exploratory, stages, *if* the response can be assumed to be monotonic, i.e. there will not be a turning point between the two levels.
- Three levels allow curvature to be checked and optimal levels to be estimated, but only assuming a symmetrical response about the optimum. This is often appropriate at early stages in research and sometimes later, if the location of the optimum can be checked using confirmatory runs.

Choice of levels

- Four levels allow symmetry of the response about the optimum to be checked. They are often appropriate at later, modelling, stages in research.
- Five levels do the same as four, but slightly less efficiently. Sometimes they are useful because they allow centre points (e.g. standard conditions) to be included in design.

Choice of combinations

The factors used and their levels define the set of factorial combinations. However, sometimes not all combinations are run because:

- some combinations are impossible, dangerous or unethical;
- the required information can be obtained using a fractional replicate.

Choice of replication

As usual the number of replicates should be chosen to ensure the effects of interest (main effects and interactions) are estimated with sufficient precision and there are a reasonable number of residual degrees of freedom. Usually, (as near as possible) equal replication is appropriate.

Incomplete Factorial Structures

Example: An experiment is to be carried out to test the robustness of a bioassay to changing a number of factors from their usual levels. Five factors were identified: percentage of acetonitrile, flow rate, pH of the mobile phase, octylamine concentration and temperature. Each was to be used at two levels, target and off-target. The experimenters do not want to perform more than 16 assays. What should our advice be?

Incomplete Factorial Structures

If high order interactions can be assumed to be zero or interest is only in very large effects, even a single replicate might be considered excessive, e.g. if three-factor and higher order interactions are negligible, the 2^5 design gives:

Source	df
Main effects	5
2-factor interactions	10
Error	16
Total	31

Incomplete Factorial Structures

Is it possible to choose half of the 32 combinations to get the following analysis of variance?

Source	df
Main effects	5
2-factor interactions	10
Error	0
Total	15

Yes, but the 16 combinations must be chosen with care.

Incomplete Factorial Structures

When thinking about hidden replication, we noted that to estimate main effects we only needed to look at the one-way table of means, to estimate two-factor interactions we needed to look at two-way tables of means, etc. To get good estimates we need to consider:

- Efficiency: estimate the effects with low variance.
- Near orthogonality: obtain low correlations between pairs of effects, because:
  - ensures simplicity of interpretation;
  - needed for efficiency.

Incomplete Factorial Structures

- To ensure efficiency of estimation of the main effect of a factor, we need equal replication of the levels of that factor.
- To ensure orthogonality of the estimated main effects, we need equal replication of every combination of every pair of factors.
- To ensure efficiency of estimation of the interaction of two factors, we need equal replication of each combination of levels of these two factors.
- To ensure orthogonality of estimated main effects and estimated two-factor interactions, we need equal replication of every combination of every set of three factors.
- To ensure orthogonality of estimated two factor interactions, we need equal replication of every combination of levels of every set of four factors.
- and so on ...

Incomplete Factorial Structures

In general, try to get (as near as possible) equal replication for every combination of levels of all sets of factors for as large sets as possible.

In 16 runs it is possible to get all projections onto four factors with equal replication.

A	B	C	D	E
0	0	0	0	0
0	0	0	1	1
0	0	1	0	1
0	0	1	1	0
0	1	0	0	1
0	1	0	1	0
0	1	1	0	0
0	1	1	1	1
1	0	0	0	1
1	0	0	1	0
1	0	1	0	0
1	0	1	1	1
1	1	0	0	0
1	1	0	1	1
1	1	1	0	1
1	1	1	1	0

Incomplete Factorial Structures

Note that fractional replication only works because we make assumptions about terms in the full model being negligible. If the full model were appropriate, our estimates of effects would be biased as follows:

$$
\begin{aligned}
E(\hat{\mu}) &= \mu + (abcde)_{11111} \\
E(\hat{a}_1) &= a_1 + (bcde)_{1111} \\
&\vdots \\
E(\widehat{(ab)}_{11}) &= (ab)_{11} + (cde)_{111} \\
&\vdots
\end{aligned}
$$

Incomplete Factorial Structures

Regular fractional replicates are those which are $1/p$ fractions, where p is a power of the lowest common multiple of the number of levels of each factor. For two-level designs regular fractions are half replicates, quarter replicates, etc.

ABCDE is known as the *defining contrast*, denoted $I \equiv ABCDE$. A and BCDE are said to be *aliases*, or A is said to be *aliased* with BCDE. The aliasing pattern for regular two-level fractional replicates is particularly simple and can be obtained from the defining contrast. Simply multiply both sides of the defining contrast, using the rule that for any effect P, $P^2 = I$.

Incomplete Factorial Structures

Note that the columns in the design matrix corresponding to, for example, A and BCDE have identical patterns. This allows us to obtain regular fractional replicates simply by choosing the defining contrast.

Many published lists of regular fractions are available and many packages have routines for constructing regular fractions.

Regular fractions are not always suitable, especially with more than two levels. Then, there are various ways of searching for *irregular* fractions. If at least two-factor interactions are of interest, these usually involve using search routines.

Factor screening experiments

In most experiments fractions will be chosen so that at least two-factor interactions can be estimated. Designs which only allow main effects to be estimated are used when the objective is *factor screening* - finding a few important factors among a large number of unimportant factors.

Example: Find a design to screen the main effects of 7 two-level factors in 8 runs.

A	B	C	D	E	F	G
0	0	0	0	0	0	0
0	0	1	1	1	1	0
0	1	0	1	1	0	1
0	1	1	0	0	1	1
1	0	0	1	0	1	1
1	0	1	0	1	0	1
1	1	0	0	1	1	0
1	1	1	1	0	0	0

Factor screening experiments

For two-level factors, regular fractions exist only for the number of experimental units equal to a power of 2. Other designs, based on *orthogonal arrays* of strength 2, are very popular in practice.

An orthogonal array of strength t is an $n \times q$ array of symbols such that any $n \times t$ submatrix contains each possible row vector of t symbols an equal number of times.

Orthogonal arrays can be used as screening designs. For example, for two-level factors they exist for numbers of units being (almost) any multiple of 4 (*Plackett-Burman designs*).

Factor screening experiments

Plackett-Burman design for 11 factors in 12 runs.

A	B	C	D	E	F	G	H	I	J	K
+	+	-	+	+	+	-	-	-	+	-
+	-	+	+	+	-	-	-	+	-	+
-	+	+	+	-	-	-	+	-	+	+
+	+	+	-	-	-	+	-	+	+	-
+	+	-	-	-	+	-	+	+	-	+
+	-	-	-	+	-	+	+	-	+	+
-	-	-	+	-	+	+	-	+	+	+
-	-	+	-	+	+	-	+	+	+	-
-	+	-	+	+	-	+	+	+	-	-
+	-	+	+	-	+	+	+	-	-	-
-	+	+	-	+	+	+	-	-	-	+
-	-	-	-	-	-	-	-	-	-	-

Experiments with Quantitative Factors

Example: In the example on the enzymatic synthesis of derivatives of sucrose, at the beginning of the lecture, it was decided to run an experiment to optimize the reaction conditions. Three factors were to be varied: pH, total sugar and donor:acceptor ratio. Because of the cost of the substrate, only 18 runs could be afforded. Which combinations of which levels should be used?

If we are going to fit a polynomial model to our data, the standard arguments about projection no longer hold. e.g. our *primary model* is often the second order polynomial,

$$E(Y) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=i}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j.$$

Experiments with Quantitative Factors

For each main effect, we are primarily interested in 2 df, corresponding to linear and quadratic components. For each interaction, we are primarily interested in 1 df, corresponding to the linear$\times$linear component of the interaction. We require only a subset of the projections onto two factors.

The choice of how many and which levels of the factors to use is usually the same as for a single quantitative factor. In practice, three, four or five levels are usually enough. At the early stages of experimentation, if the effect of a factor can be assumed to be monotonic, two levels may be used. For two levels the design requirements are exactly the same as for two level qualitative factors.

Three level designs

Code the factors so that the three levels are -1, 0 and 1. For each factor, as with a single factor, some compromise between linear and quadratic effects is necessary and near-equal replication of each level is often sensible. For interactions, we require high and (as near as possible) equal replication of the $(\pm 1, \pm 1)$ combinations for each pair of factors.

The *central composite design* is the most widely used response surface design. It is made up from:

- the two-level factorial points, $(\pm 1, \ldots, \pm 1)$, or a fraction of them;
- the *axial points*, $(\pm 1, 0, \ldots, 0)$, etc.;
- some centre points, $(0, \ldots, 0)$.

Four level designs

Good designs can be obtained from two level designs by replacing each four level factor with a pair of two level *pseudo-factors*, as follows.

X_1	A	B
-1	-1	-1
$-1/3$	-1	$+1$
$+1/3$	$+1$	-1
$+1$	$+1$	$+1$

Four level designs

The two pseudo-factors give three contrasts, A, B and AB. However, the contrasts we are interested in are those which correspond to the linear and quadratic effects of the four level factor. These are:

$$\begin{aligned} X_{1L} &= (2A + B)/3 \\ X_{1Q} &= AB. \end{aligned}$$

An interaction contrast comes from direct multiplication of the main effect contrasts, e.g. if X_2 has pseudo-factors C and D,

$$X_{1L} \times X_{2L} = (4AC + 2BC + 2AD + CD)/9.$$

We then choose a good four level design by choosing a good two level design for the pseudo-factors. Note that it is not necessary to estimate all two-factor interactions for the pseudo-factors.

Five level designs

The central composite design can be modified by bringing in the corner points to $(\pm\gamma, \ldots, \pm\gamma)$ for $\gamma < 1$.

It is then made up from:

- the two-level factorial points, $(\pm\gamma, \ldots, \pm\gamma)$, or a fraction of them;
- the *axial points*, $(\pm 1, 0, \ldots, 0)$, etc.;
- some centre points, $(0, \ldots, 0)$.

The levels are often recoded, so that the factorial points are at $(\pm 1, \ldots, \pm 1)$ and the axial points at $(\pm 1/\gamma, 0, \ldots, 0) = (\pm\alpha, 0, \ldots, 0)$, etc.

Search Algorithms

In many experiments, especially with quantitative factors, regular fractions and other standard constructions are not available, or not appropriate.

Then an optimal design approach is useful.

Criteria can be the same as for a single factor, e.g. weighted-A-efficiency or D-efficiency, but criteria which allow for model uncertainty are more useful, e.g. a weighted average D-efficiency over a suitable set of candidate models.

Direct optimization is rarely possible. Instead a search routine is usually used.

Search Algorithms

Start by defining a set of *candidate points*, i.e. possible treatments, from which we are going to try to find an optimal subset of size n.

Often the candidate points will be a full set of factorial points, e.g. 3^q or 5^q.

Many algorithms are available for searching through the candidate points. The most basic, of which many modifications are possible, is the Fedorov exchange algorithm.

Search Algorithms

1. Generate a randomly chosen set of n points.
2. If the design is singular (i.e. doesn't allow the model to be fitted), go to 1.
3. Evaluate the criterion for the starting design, which becomes the current design.
4. For every possible exchange of a point in the current design with a point in the candidate set, evaluate the criterion.
5. If any exchange improves the current design, accept the exchange which gives the largest improvement to create the new current design and go to 4.
6. The current design is returned.

It is usual to run the design with several restarts, or *tries*.

Search Algorithms

Many modifications are possible, e.g.

- Instead of finding the optimal exchange, find the optimal point(s) to add, then the optimal point(s) to delete.
- Allow excursion, i.e. make several exchanges before evaluating the new design.
- Accept the best exchange even if it makes the worse by a small amount, the acceptable amount decreasing as the algorithm continues.
- *Simulated annealing* accepts any exchange with small probability even if it makes the design worse, this probability decreasing as the algorithm continues.
- Other stochastic search algorithms, such as *genetic algorithms*, have been proposed.

Search Algorithms

For most fairly standard design problems the simplest algorithms are to be recommended because:

- They run more quickly, so more tries can be made.
- They do not require the choice of tuning constants.
- They can be used interactively, for example to modify the criterion to eliminate undesirable features of the designs produced.

The most important point is that none of these methods is perfect. They should be considered to be just ways of generating reasonable designs for discussion with the experimenters.