Chapter 16

Theory of Orthogonal Designs

Chapter Table of Contents

Chapter 16
Theory of Orthogonal Designs

Overview

This chapter provides the mathematical and statistical background for designs constructed by the FACTEX procedure; it also outlines the search algorithm that is used to find suitable construction rules. Note that the material in this chapter is general and theoretical; you do not need to read this chapter to use the procedure for constructing most common experimental designs. On the other hand, you should read this chapter

- to understand the general structure of designs that can be constructed with the FACTEX procedure
- to construct designs for factors with more than two levels, especially if interactions are involved
- to improve the search used by the procedure when constructing complicated designs for many factors

Structure of General Factorial Designs

The FACTEX procedure constructs a fractional design for q-level factors using the *Galois field* (or *finite field*) of size q. This is a system with q elements and two operations $+$ and $\times$, which satisfy the usual mathematical axioms for addition and multiplication. When q is a prime number, finite field arithmetic is equivalent to regular integer arithmetic modulo q. When $q = 2$, addition of the two elements of the finite field is equivalent to multiplication of the integers $+1$ and -1. Since designs for factors with levels $+1$ and -1 are the factorial designs most commonly covered in textbooks, the arithmetic for fractional factorial designs is usually shown in multiplicative form. However, throughout this section a more general notation is used.

A design for q-level factors in q^m runs constructed by the FACTEX procedure has the following general form. The first m factors are taken to index the runs in the design, with one run for each different combination of the levels of these factors, where the levels run from 0 to $q-1$. These factors are called *run-indexing factors*. For a particular run, the value F of any other factor in the design is derived from the levels $P_1, P_2, \ldots, P_m$ of the run-indexing factors by means of *confounding rules*. These rules are of the general form

$$F = r_1 P_1 + r_2 P_2 + \ldots + r_m P_m$$

where all the arithmetic is performed in the finite field of size q.

The linear combination on the right-hand side of the preceding equation is called a *generalized interaction* between the run-indexing factors. A generalized interaction is part of the statistical interaction between the factors with nonzero coefficients in the linear combination. The factor F is said to be *confounded* or *aliased* with this generalized interaction; two terms are confounded when the levels they take in the design yield identical partitions of the runs, so that their effects cannot be distinguished. The confounding rules characterize the design, and the problem of constructing the design reduces to finding suitable confounding rules.

Suitable Confounding Rules

Design Factors

This section explains how the criteria for a design can be reduced to prescribing that certain generalized interactions are *not* to be "confounded with zero."

Suitable confounding rules depend on the effects you want to estimate with the design. For example, if you want to estimate the main effects of both A and B, the following rule is inappropriate:

$$A = B$$

With this rule, the levels of A and B are the same in every run of the design, and the main effects of the two factors cannot be estimated independently of one another. Thus, the first criterion for a suitable confounding rule is that no two effects you want to estimate should be confounded with each other.

Furthermore, an effect you want to estimate should not be confounded with an effect that is nonnegligible. For example, if the interaction between C and D is nonnegligible and you want to estimate the main effect of A, the following confounding rule is inappropriate:

$$A = C + D$$

(Recall that this section uses a general linear form for confounding rules instead of the usual multiplicative form. For factors with levels $+1$ and -1, the preceding rule is equivalent to $A = C * D$.)

Another kind of confounding involves *confounding with zero*. If a factor or a generalized interaction F has the same value in every run of the design, then F is *confounded with zero*. Such confounding is denoted as

$$0 = F$$

Interactions are estimable with the design if and only if they are not confounded with zero. Consequently, another criterion for a suitable confounding rule is that no effect that you want to estimate can be confounded with zero. The confounding rule for two main effects

$$A = B$$

can be written as a generalized interaction confounded with zero.

$$0 \;=\; -A + B$$

The right-hand side of the preceding equation is part of the interaction between A and B. Thus, for any two effects to be unconfounded, it is equivalent to prescribe that no part of their generalized interaction be confounded with zero.

Note that it is not enough to make sure that only the confounding rules themselves satisfy these restrictions. The consequences of the confounding rules must also satisfy the restrictions. For example, suppose you want to make sure that main effects are not confounded with two-factor interactions, and suppose that the confounding rule for factor E is

$$E \;=\; A + B + C + D$$

Then the following rule cannot be used for factor F:

$$F \;=\; A + B + C$$

Even though the rule for F does not confound F with a two-factor interaction, this rule forces a generalized interaction between E and F to be aliased with the main effect of D, since

$$E - F \;=\; (A + B + C + D) - (A + B + C) \;=\; D$$

Block Factors

If your design involves blocks, additional confounding criteria need to be considered. Blocks are introduced into designs by means of *block pseudo-factors*. (See "Types of Factors" on page 488 for details.) A design for q-level factors in q^s blocks contains s block pseudo-factors. Denoting the levels of these factors for any given run by $B_1, B_2, \ldots, B_s$, the index of the block in which the run occurs is given by

$$B_1 + qB_2 + q^2B_3 + \ldots + q^{s-1}B_s$$

For each block to occur in the design, every possible combination of block pseudo-factors must occur. This can happen only if all main effects and interactions between the block factors are estimable, which leads to yet another criterion for the confounding rules. Moreover, the effects you want to estimate cannot be confounded with blocks. In general,

- no generalized block pseudo-factors can be confounded with zero
- no generalized interactions between block pseudo-factors and effects you want to estimate can be confounded with zero

General Criteria

The criteria for an orthogonally confounded q^k design reduce to requiring that no generalized interactions in a certain set $\mathcal{M}$ can be confounded with zero. (See "Structure of General Factorial Designs" on page 503 for a definition of *generalized interaction*.) This section presents the general definition of $\mathcal{M}$. First, define three sets, as follows:

$\mathcal{E}$ the set of effects that you want to estimate

$\mathcal{N}$ the set of effects you do not want to estimate but that have unknown nonzero magnitudes (referred to as *nonnegligible* effects)

$\mathcal{B}$ the set of all generalized interactions between block pseudo-factors

Furthermore, for any two sets of effects $\mathcal{A}$ and $\mathcal{B}$, denote by $\mathcal{A} \times \mathcal{B}$ the set of all generalized interactions between the effects in $\mathcal{A}$ and the effects in $\mathcal{B}$.

Then the general rules for creating the set of effects $\mathcal{M}$ that are not to be confounded with zero are as follows:

- Put $\mathcal{E}$ in $\mathcal{M}$. This ensures that all effects in $\mathcal{E}$ are estimable.
- Put $\mathcal{E} \times \mathcal{E}$ in $\mathcal{M}$. This ensures that all pairs of effects in $\mathcal{E}$ are unconfounded with each other.
- Put $\mathcal{E} \times \mathcal{N}$ in $\mathcal{M}$. This ensures that effects in $\mathcal{E}$ are unconfounded with effects in $\mathcal{N}$.
- Put $\mathcal{B}$ in $\mathcal{M}$. This ensures that all q^s blocks occur in the design.
- Put $\mathcal{E} \times \mathcal{B}$ in $\mathcal{M}$. This ensures that effects in $\mathcal{E}$ are unconfounded with blocks.

Searching for Confounding Rules

The goal in constructing a design, then, is to find confounding rules that do not confound with zero any of the effects in the set $\mathcal{M}$ defined previously. This section describes the sequential search performed by the FACTEX procedure to accomplish this goal.

First, construct the set C_1 of candidates for the first confounding rule, taking into account the set $\mathcal{M}$ of effects not to be confounded with zero. If C_1 is empty, then no design is possible; otherwise, choose one of the candidates $r_1 \in C_1$ for the first confounding rule and construct the set C_2 of candidates for the second confounding rule, taking both $\mathcal{M}$ and r_1 into account. If C_2 is empty, choose another candidate from C_1; otherwise, choose one of the candidates rules $r_2 \in C_2$ and go on to the third rule. The search continues either until it succeeds in finding a rule for every non-run-indexing factor or the search fails because the set C_1 is exhausted.

The algorithm used by the FACTEX procedure to select confounding rules is essentially a depth-first tree search. Imagine a tree structure in which the branches connected to the root node correspond to the candidates C_1. Traversing one of these branches corresponds to choosing the corresponding rule r_1 from C_1. The branches attached to the node at the next level correspond to the candidates for the second rule given r_1. In general, each node at level i of the tree corresponds to a set of feasible choices for rules $r_1, \ldots, r_i$, and the rest of the tree above this node corresponds to the set of all possible feasible choices for the rest of the rules.

Speeding up the Search

For designs with many factors or blocks, the tree of candidate confounding rules can be very large and the search can take a very long time. In these cases, the FACTEX procedure spends a lot of time exploring sets of rules that are essentially the same and that all result in failure. A technique for pruning the search tree (see Figure 16.1) is as follows. Suppose that for some selection r_i for rule i, all the branches for the next rule eventually result in failure. Then any other selection r_i' is immediately declared a failure if the resulting number of candidates is the same as for the failed rule r_i. The search goes on to the next selection for rule i.

This method of pruning is not perfect; it may prune a branch of the search tree that would have resulted in a success. In mathematical terms, candidate sets C_i are not necessarily isomorphic because they have the same size. You can use the NOCHECK option in the PROC FACTEX statement to turn off the pruning. With the NOCHECK option, the FACTEX procedure searches the entire tree of feasible confounding rules; and if given enough time, will find a design if one exists. The default argument for the TIME= option on the PROC FACTEX statement limits the search time to one minute.

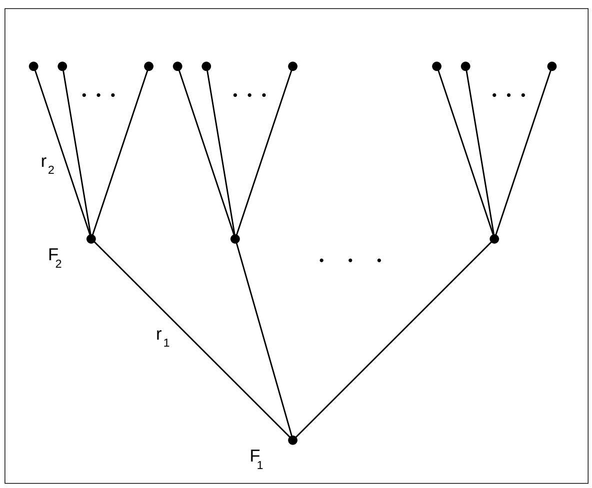

Figure 16.1. Search Tree

On the other hand, you should recognize how rarely the NOCHECK option is needed to produce a design with a given resolution. For example, consider all possible blocked and unblocked two-level designs with minimum resolution for 50 or fewer factors and 128 or fewer runs. Of the 849 different designs, the NOCHECK option is required in only five cases. The five designs for which the NOCHECK option is required are listed in Table 16.1. Note that all of these are block designs, most for many factors and relatively small blocks.

Table 16.1. Designs Requiring the NOCHECK Option

Number of Factors	Number of Runs	Block Size	Resolution
5	16	2	4
21	32	4	3
22	32	4	3
23	32	4	3
39	64	4	3

General Recommendations

Choosing appropriate confounding rules can be difficult, especially if the set $\mathcal{M}$ is at all complicated. Even if a design is found that satisfies the model specification, it is a good idea to examine the alias structure to make sure that you understand the alias structure generated by the confounding rules. To do so, use the ALIAS option in the EXAMINE statement.

For more details on the general mathematical theory of orthogonal factorial designs, refer to Bose (1947).

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *SAS/QC® User's Guide, Version 8*, Cary, NC: SAS Institute Inc., 1999. 1994 pp.

SAS/QC® User's Guide, Version 8

ISBN 1–58025–493–4

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, October 1999