ISSN 1757- 5958

amc technical briefs

Editor: Michael Thompson Analytical Methods Committee AMCTB No 36 January 2009

Experimental Design and Optimisation (3): Some Fractional Factorial Designs

In the first paper of this series (AMCTB24) we showed that the main factors affecting the outcomes of an experiment could be studied using *factorial designs*, in which the *levels* (values) of two or more factors are changed *together*, and the experimental response measured for each factor combination. If all the possible combinations of levels are studied, giving a *complete factorial design*, the number of experiments required grows rapidly with the number of factors. Even if each factor is studied at only two levels (so-called *screening* designs) and there are *k* factors, the number of experiments needed is 2^k. If we need to distinguish possible *interactions* between the factors from random measurement errors, the number of experiments rises to 2^{k+1}. So it is not surprising that in practice frequent use is made of *fractional factorial designs*, in which the number of experiments needed is lower, with some cost in the information available.

A Simple Fractional Factorial Design

In AMCTB24 we considered an example in which three factors – pH, ionic strength (I), and the choice of organic modifier – were studied in terms of their effects on the resolution of a reversed-phase high performance liquid chromatography experiment. Eight experiments would be necessary in a complete factorial design for such a system. We could assume that the three-fold interaction between all the factors was zero, so its apparent value could be used as an estimate of the random measurement error, and duplication of the eight experiments would not be necessary. Since there are three factors we can show the eight experiments graphically in a 3-dimensional way (Fig. 1).

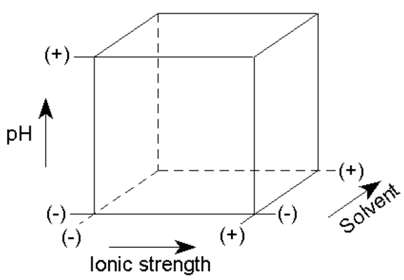

Figure 1. Complete Factorial Design for 3 Factors

Each of the vertices of the cube corresponds to one of the eight experiments – for example the top right-hand vertex represents the experiment in which all three factors are at their high (+) level. Using the same representation, we can easily describe a *half-fractional* design, in which the trial experiments are performed only at four of the vertices of the cube marked by the ● symbol, as shown in Figure 2. These four points lie at the vertices of a tetrahedron (shown by the dotted lines). Each of the three factors is studied twice at the high level and twice at the low level, giving a *balanced* design. Obviously it would be just as valid to do the four experiments at the other four vertices of the tetrahedron: in each case this minimal number of experiments maps the factor space as well as possible. These experiments, for both the full and the fractional factorial designs, with the factor levels given as + and – symbols, are summarised in Table 1.

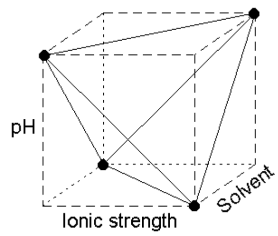

Figure 2:. Half-Factorial Design for 3 Factors

Table 1. Factorial Designs with Two Levels and Three Factors

Expt	Solv	pH	I	Response
1	-	-	-	y_1
2*	+	-	-	y_2
3	+	+	-	y_3
4*	-	+	-	y_4
5*	-	-	+	y_5
6	+	-	+	y_6
7*	+	+	+	y_7
8	-	+	+	y_8

If all eight experiments are used in the complete factorial design, we can calculate the effects of the three factors separately, the three possible two-fold interactions (i.e. pH interacting with ionic strength and with solvent, and ionic strength interacting with solvent) and the three-fold interaction. Such calculations are described in many standard texts and performed readily by experimental design programs. If we do only the four experiments marked with an asterisk we can find only the effects of the main factors, and *assume* that no significant interactions occur.

For the fractional design using experiments 2, 4, 5 and 7 the calculations are very easy. The effect of changing from high pH to low pH is clearly given by $0.5\{(y_4 + y_7) - (y_2 + y_5)\}$. Very often it is immediately apparent from such results whether or not a factor is an important one, but significance tests are also available to compare the effects with the random measurement error. The design is known as a "half-factorial" design, for obvious reasons. When more factors are under study, so that a full factorial design would involve too many experiments, other designs such as quarter-factorial ones may be feasible. In general, the more experiments we can do, the more information we shall get, including information on interactions as well as on the main effects. But as the number of factors and experiments grows a further complication may arise.

Confound It! He's got an alias!

In a system involving four experimental factors, A, B, C, and D, each studied at two levels, the complete factorial design would evidently involve 16 experiments, and a half-factorial design would require 8, again chosen to map the "factor space" as efficiently as possible. A suitable half-factorial design is shown in Table 2. As expected, each of the factors is studied four times at the high level (+) and four times at the low level (-). (It would be important to make the measurements in a random order, to minimise the effects of *uncontrolled factors*, e.g. instrument or temperature drifts).

Table 2. Half Factorial Design for Four Factors at Two Levels

Expt.	A	B	C	D	Response
1	+	+	+	+	y_1
2	+	+	-	-	y_2
3	+	-	+	-	y_3
4	+	-	-	+	y_4
5	-	+	+	-	y_5
6	-	+	-	+	y_6
7	-	-	+	+	y_7
8	-	-	-	-	y_8

Clearly the effect of A, for example, is given by the expression $(y_1 + y_2 + y_3 + y_4 - y_5 - y_6 - y_7 - y_8)/4$. But the results y_1 to y_8 also give information on some of the interactions between the factors, and it can be shown (see references) that the *same* expression is also a measure of the three-fold interaction between the factors B, C and D (called BCD for short). So when its value is computed we actually obtain the *sum* of the effects A and BCD. This problem is called *confounding*, and the pairs of effects involved are called *aliases* of each other. Sometimes the problem may not be significant (in this example it would be normal to neglect the possible 3-fold interaction BCD entirely), but in others it needs careful study. This is not at all surprising. With four factors we have four main effects, six two-factor interactions, four three-factor interactions, and one four-factor interaction. We cannot expect to resolve them all in only eight experiments! In a 2^{k-1} design every effect is confounded with another one, that is, each effect is one of a pair of aliases. The extent to which a fractional factorial design gives confounding problems is expressed by its *resolution*, R, written in Roman numerals. A design with a resolution R avoids confounding between a *p*-factor effect and an effect containing $<(R - p)$ factors. In our example R = IV (four!) so there is no confounding between main effects ($p = 1$) and *two-fold* (i.e. less than 3-fold) interactions. Many of these problems are avoided (again at the cost of some information loss) by using even simpler designs.

Available Software

Minitab® provides good experimental design facilities, and guidance on how to use them. Design-Ease® also offers much guidance, and a free trial version can be downloaded from www.statease.com.

Further reading

D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics*, Part A, Elsevier, Amsterdam, 1997.

James Miller

This Technical Brief was prepared for the Analytical Methods Committee by the Statistical Subcommittee (Chair M Thompson).

CPD Certification I certify that I have studied this document as a contribution to Continuing Professional Development.

Name…………………...

Signature…………………………….…….Date………….

Name of supervisor…………………………………..

Signature……………………….……….Date…………

