# amc technical brief

# Experimental design and optimisation (1): an introduction to some basic concepts

**Analytical scientists all too frequently think that the use of statistics and chemometrics is confined to the treatment of data obtained in completed experiments. In reality the proper planning of experiment**

**Table 2: Effects of two factors on reversed phase-HPLC resolution**

|        | MeOH | MeCN |
|--------|------|------|
| **pH 7** | 1.20 | 1.30 |
| **pH 8** | 1.30 | *    |

If the effect of altering both factor levels is to give a resolution of 1.40 for the table entry marked *, then the two factors are said to be **additive**. But in practice the resolution obtained when both alterations are made simultaneously may be significantly greater or less than 1.40, in which case the two factors are described as being **interactive**. In this case the interaction means that the effect of changing the pH *depends upon* the choice of organic modifier, and *vice-versa*. Obviously such effects could not be detected by studying just one factor at a time. Table 2 also highlights a further problem encountered in practice. Suppose that when both the factor levels are changed the resolution is found to be 1.43. Has a (slight) positive interaction between the two factors occurred, or is the value 1.43 simply the result of random measurement errors, i.e. is it not significantly different from 1.40? It seems that to answer that question it would be necessary to make replicate measurements so that the effects of random errors can be separated from interaction effects, for example by using analysis of variance (ANOVA). In summary, efficient EDs involve varying two or more factors at once, to minimise the effort involved and to study interactions; but separating the interactions from random errors may require extra work. Compromises may thus be necessary, and the chemist must choose the design best suited to the identified aims of the overall experiment.

**Factorial Designs**

Table 2 provides the simplest example of what is known as a **complete factorial design**. We have two factors, each studied at two levels, so $2^2 = 4$ measurements are needed to study all possible combinations of the factors and levels. If each experiment is duplicated, to obtain an estimate of the random measurement error, $2^{2+1} = 8$ measurements are necessary. In general, if $k$ factors are studied at 2 levels, the number of measurements in a complete factorial design is $2^k$ if the experiments are not duplicated, $2^{k+1}$ if they are. With 5 factors, therefore, either 32 or 64 measurements would be needed: in most cases such a protracted effort would be impossible. Complete factorial designs are evidently tedious in many cases, so