

of freedom in this case is  $(n - 1)$  where  $n$  is the number of pairs of data.

Comparisons between means (and standard deviations) can be extended to the study of three or more sets of data. Such comparisons require the very important statistical method called analysis of variance (ANOVA).

crucial issue is provided by the resources listed in the Further Reading section.

**See also:** **Chemometrics and Statistics:** Experimental Design; Optimization Strategies; Multivariate Classification Techniques; Multivariate Calibration Techniques; Expert Systems; Multicriteria Decision Making; Signal Processing; Spectral Deconvolution and Filtering.

## Application of Statistical Techniques

Many of the significance tests and other procedures summarized in this article (and many others) are very readily performed with the aid of Microsoft Excel®, Minitab®, and other widely available programs. Such software also gives instant access to the most important descriptive statistics (mean, median, standard deviation, s.e.m., confidence limits, etc.). In practice, the major problems are therefore (1) accurate entry of the experimental data into the program (this problem may not arise if an analytical instrument is directly interfaced to a PC); and (2) choice of the appropriate test once data entry has been successfully completed. Guidance on the latter

## Further Reading

- De Levie R (2001) *How to Use Excel® in Analytical Chemistry*. Cambridge: Cambridge University Press.
- ISO (1993) *Guide to the Expression of Uncertainty in Measurement*. Geneva: International Organisation for Standardisation.
- Meier PC and Zund RE (2000) *Statistical Methods in Analytical Chemistry*, 2nd edn. New York: Wiley.
- Miller JN and Miller JC (2000) *Statistics and Chemometrics for Analytical Chemistry*, 4th edn. Harlow: Pearson Education.
- Neave HR (1981) *Elementary Statistics Tables*. Routledge: London.

## Experimental Design

**G Hanrahan, J Zhu, S Gibani, and D G Patil,**  
California State University, Los Angeles, CA, USA

© 2005, Elsevier Ltd. All Rights Reserved.

This article is a revision of the previous-edition article by J Goupy, pp. 659–666, © 1995, Elsevier Ltd.

## Introduction

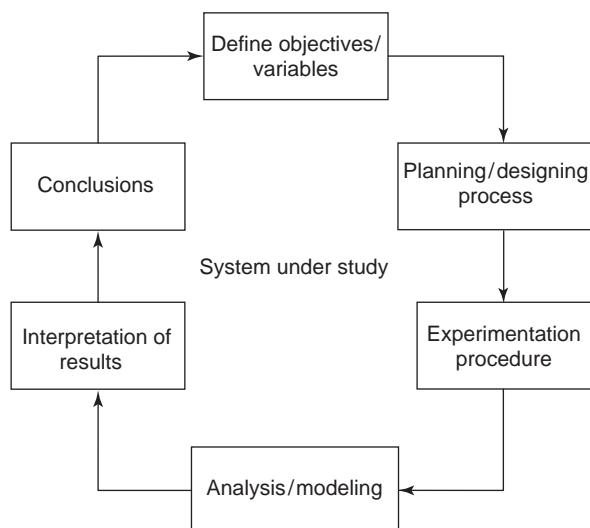
Experimental design methods allow the experimenter to understand better and evaluate the factors that influence a particular system by means of statistical approaches. Such approaches combine theoretical knowledge of experimental designs and a working knowledge of the particular factors to be studied. Although the choice of an experimental design ultimately depends on the objectives of the experiment and the number of factors to be investigated, initial experimental planning (as shown in Figure 1) is essential.

The relationship between the various factors and response within a given system can be shown

mathematically as follows:

$$y = f(x_1, x_2, x_3, \dots, x_k) \quad [1]$$

where  $y$  is the response of interest in the system and  $x$  are the factors that affect the response when their values change. In general, the following types of factors can be distinguished: (1) continuous, e.g., temperature; and (2) discrete, e.g., experimenters. Factors are considered to be independent if there is no relationship between them and dependent if a relationship exists. The values or settings attributed for each factor are called levels. Each experimental run in an experimental design study requires that one or more treatments (stimulus applied to one or more factors) be applied to the system and the response measured. The experimenter then employs statistical design methods to determine if the treatment of interest or combination of treatments was significant in influencing the response of the system under study. Calculation of the treatment effects can then be used to identify which variables lead to an optimal response.



**Figure 1** Essential criteria during early experimental planning.

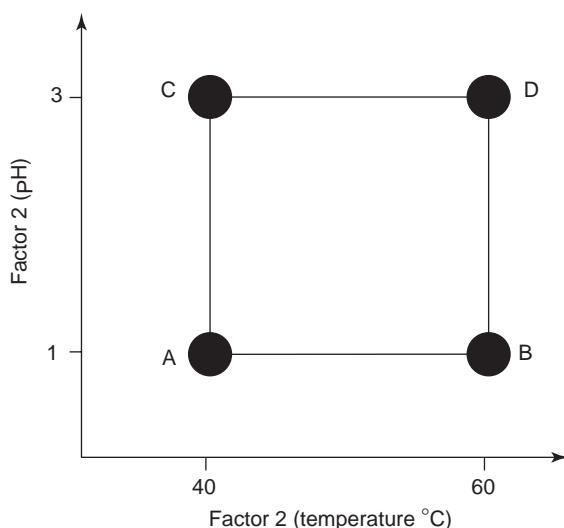
The two main applications of experimental design are screening, in which the factors that influence the experiment are identified, and optimization, in which the optimal settings or conditions for an experiment are found. A screening experiment is a systematic approach to identifying the key input parameters of a process or product that affect the output performance. The usual approach is to start with a screening design including all controllable factors that may possibly influence the experiment, identify the most important ones, and then proceed with an experimental optimization design.

## Experimental Designs

### Full Factorial Designs (Two Levels per Factor)

The most general two-level designs is a full factorial design and described as  $2^k$ -designs where the base 2 stands for the number of factor levels and  $k$  the number of factors each with a high and low value. In a full factorial design, the levels of the factors are chosen in such a way that they span the complete factor space. Often, only a lower and upper level is chosen. With two factors, this defines a square in the factor space, and with three factors, this defines a cube. The lower level is usually indicated with a ‘–’ sign; the higher level with a ‘+’ sign.

The method can be graphically illustrated in a simplified example: the effects of reaction temperature and pH in determining the spectrophotometric response (absorbance) of a standard analyte solution. Figure 2 shows a graphical definition of the experimental domain, with the reaction temperature varying from 40°C (low level) to 60°C (high level)



**Figure 2** Graphical definition of the effects of reaction temperature and pH in determining the spectrophotometric response of a standard analyte solution.

**Table 1** Experimental matrix: spectrophotometric response of a standard analyte solution

Experiment number	Temperature	pH	Response
A	–1	–1	$y_1$
B	+1	–1	$y_2$
C	–1	+1	$y_3$
D	+1	+1	$y_4$
Factor levels			
(–)	40°C	pH 1	
(+)	60°C	pH 3	

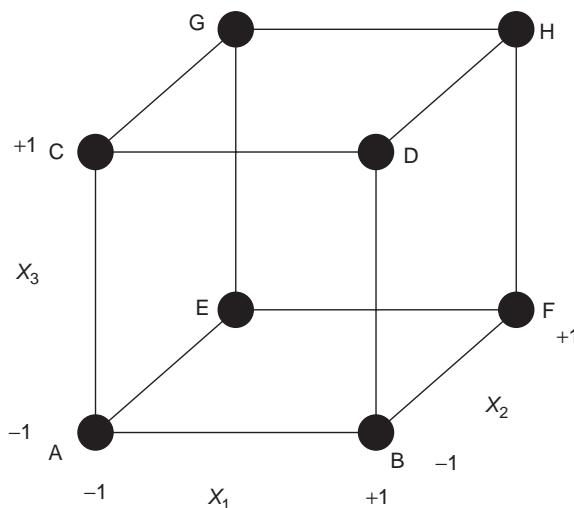
and the reaction pH varying from 1 (low level) to 3 (high level). The best experimental points in the domain are located in the corners A, B, C, and D as follows: A (40°C, pH 1); B (60°C, pH 1); C (40°C, pH 3); D (60°C, pH 3).

The four trials of experimental matrix used in this experiment are shown in Table 1, with the results of each experiment indicated in the response column and the factor levels in the rows below the experimental matrix. Note that –1 is used for the low level of each factor and +1 for the high level.

If we introduce another variable (e.g., reagent concentration) in the experiment, it is then possible to represent the factors as faces on one or more cubes with the responses at the points. The distribution of experimental points within this type of experimental domain ( $2^3$  design) is shown schematically in Figure 3.

### Fractional Factorial Design

Fractional factorial designs are arguably the most widely used designs in experimental investigations,



**Figure 3** Full factorial design at two levels,  $2^3$  design.

and mainly used for the screening portion of experiments. Such designs are good alternatives to a full factorial design, especially in the initial stage of a project, and considered a carefully prescribed and representative subset of a full factorial design. In fractional factorial designs, the number of experiments is reduced by a number  $p$  according to a  $2^{k-p}$  design. In the most commonly employed fractional design, the half-fraction design ( $p=1$ ), exactly one-half of the experiments of a full design are performed.

Suppose a situation occurs in which three factors, each at two levels are of interest, but the experimenter does not want to run all eight treatment combinations ( $2^3=8$ ). A design with four treatment combinations can then be performed when considering the one-half fraction of the  $2^3$  design ( $2^{3-1}=4$ ).

The fractional factorial design is based on an algebraic method of calculating the contributions of factors to the total variance with less than a full factorial number of experiments. Such designs are useful when the numbers of potential factors are relatively large because they reduce the total number of runs required for the overall experiment. However, by reducing the number of runs, a fractional factorial design will not be able to evaluate the impact of some of the factors independently.

### Latin Squares

A Latin square is a block design with the arrangement of  $v$  Latin letters into a  $v \times v$  array (a table with  $v$  rows and  $v$  columns). Latin square designs are often used in experiments where subjects are allocated treatments over a given time period where time is

thought to have a major effect on the experimental response. Suppose the treatments are labeled A, B, and C. In this particular situation, the design would be

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

This type of design allows the separation of an additional factor from an equal number of blocks and treatments. If there are more than three blocks and treatments, then a number of Latin square designs are possible. It can be noted that Latin square designs are equivalent to specific fractional factorial designs (e.g., the  $4 \times 4$  Latin square design is equivalent to a  $4^{3-1}$  fractional factorial design).

### Greco-Latin Squares

The Greco-Latin square design involves two Latin squares that are superimposed on each other. It contains two treatment factors instead of one and contains four factors overall instead of three. An example design would look as follows:

	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>
B <sub>1</sub>	C <sub>1</sub> D <sub>3</sub>	C <sub>2</sub> D <sub>4</sub>	C <sub>3</sub> D <sub>1</sub>	C <sub>4</sub> D <sub>2</sub>
B <sub>2</sub>	C <sub>4</sub> D <sub>2</sub>	C <sub>1</sub> D <sub>1</sub>	C <sub>2</sub> D <sub>3</sub>	C <sub>3</sub> D <sub>4</sub>
B <sub>3</sub>	C <sub>3</sub> D <sub>1</sub>	C <sub>4</sub> D <sub>2</sub>	C <sub>1</sub> D <sub>3</sub>	C <sub>2</sub> D <sub>4</sub>
B <sub>4</sub>	C <sub>2</sub> D <sub>4</sub>	C <sub>1</sub> D <sub>3</sub>	C <sub>3</sub> D <sub>2</sub>	C <sub>4</sub> D <sub>1</sub>

The analysis for the Greco-Latin square design is similar to that of a Latin square design. However, one noticeable difference is that two treatment sum of squares have to be computed (factors C and D) by listing two sets of means outside the design table. As an additional note, Greco-Latin squares are most effective if replicated and are subject to the same randomization rules as for the Latin squares.

### Response Surface Designs (More than Two Levels for One or More Factors)

Response surface methodology is designed to allow experimenters to estimate interactions, therefore giving them an idea of the shape of the response surface they are investigating. This approach is often used when simple linear and interaction models are not adequate, e.g., experimentation far from the region of optimum conditions. Here, the experimenter can expect curvature to be more prevalent and will need a mathematical model, which can represent the curvature. The simplest such model has the quadratic

form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 \quad [2]$$

which contains linear terms for all factors, squared terms for all factors, and products of all pairs of factors. The two most common designs generally used in response surface modeling are central composite designs and Box–Behnken designs. In these designs the inputs take on three or five distinct levels, but not all combinations of these values appear in the design.

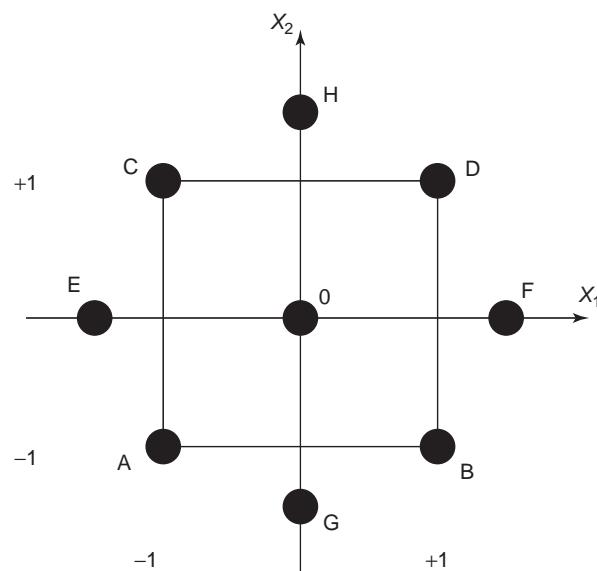
Central composite designs contain imbedded factorial or fractional factorial design with center points that is augmented with a group of axial (star) points that allow estimation of curvature (Figure 4). One central composite design consists of cube points at the corners of a unit cube that is the product of the intervals  $[-1, 1]$ , star points along the axes at or outside the cube, and center points at the origin. Points A, B, C, and D are the points of the initial factorial design with points E, F, G, and H being the star points at the central 0.

A central composite design always contains twice as many star points as there are factors in the design. The star points represent new extreme values (low and high) for each factor in the design. There are three types of central composite designs, ultimately depending on where the star points are placed. Circumcised central composite designs are the original forms of the central composite design. The star points are at some distance  $\alpha$  from the center based on the properties desired for the design and the number of factors in the design. The  $\alpha$  values depend on the number of factors in the factorial part of the design.

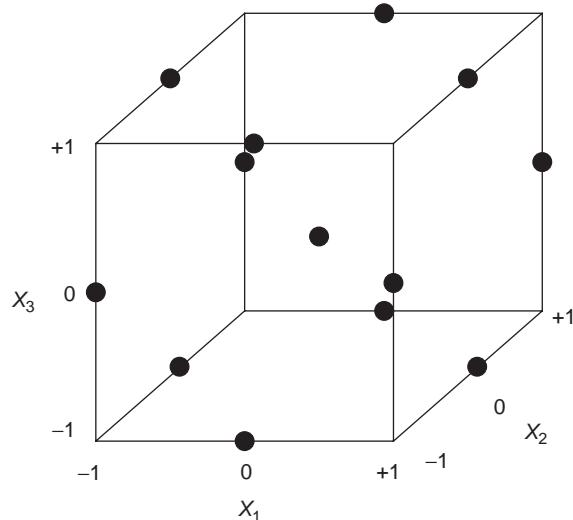
Inscribed central composite designs are a scaled down version of the circumcised version in that they use the factor settings as the star points and create factorial or fractional factorial design within those limits. In other words, inscribed designs are those where each factor of the circumcised version is divided by  $\alpha$  for their generation. Face centered central composite designs occur when the star points are at the center of each face of the factorial space, where  $\alpha = \pm 1$ .

### Box–Behnken Designs

The Box–Behnken is considered an efficient option in response surface methodology and an ideal alternative to central composite designs. It has three levels per factor, but avoids the corners of the space, and fills in the combinations of center and extreme levels (Figure 5). It combines a fractional factorial with incomplete block designs in such a way as to avoid the



**Figure 4** Central composite design consisting of a full factorial two-level and star design.



**Figure 5** Box–Behnken design with three levels per factor.

extreme vertices and to present an approximately rotatable design with only three levels per factor. A design is rotatable if the variance of the predicted response at any point  $x$  depends only on the distance of  $x$  from the design center point. It must be noted, however, that Box–Behnken designs should be confined to uses where the experimenter is not interested in predicting response at extremes (corners of the cube).

### Mixture Designs

In a mixture design experiment, the independent factors are proportions of different components of a

blend and often measured by their portions, which sum to 100% or normalized to 1, i.e.,

$$\sum_{i=1}^N X_i = 1 \quad \text{for } x_i \geq 0 \quad [3]$$

As shown, mixture components are subject to the constraint that they must equal to the sum of one. In this case, standard mixture designs for fitting standard models such as simplex-lattice and simplex-centroid designs are employed. When mixtures are subject to additional constraints, constrained mixture designs (extreme-vertices) are then appropriate. Like the factorial experiments discussed above, mixture experimental errors are independent and identically distributed with zero mean and common variance. In addition, the true response surface is considered continuous over the region being studied. Overall, the measured response is assumed to depend only on the relative proportions of the components in the mixture and not on the amount.

*See also: Chemometrics and Statistics:* Optimization Strategies; Multivariate Calibration Techniques.

## Further Reading

- Box GEP, Hunter WG, and Hunter JS (1987) *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building*. New York: Wiley.
- Cornell JA (1990) *Experiments with Mixtures*. New York: Wiley.
- Dean A and Voss D (1999) *Design and Analysis of Experiments*. New York: Springer.
- Goupy J (1993) *Methods for Experimental Design. Principles and Applications for Physicists and Chemists*. Amsterdam: Elsevier.
- Kempthorne O (1976) *Design and Analysis of Experiments*. New York: Wiley.
- Martens H and Martens M (2001) *Multivariate Analysis of Quality*. Chichester: Wiley.
- Massart DL, Vandeginste BGM, Buydens LMC, et al. (1997) *Handbook of Chemometrics and Qualimetrics, Part A*. Amsterdam: Elsevier.
- Miller JN and Miller JC (2000) *Statistics and Chemometrics for Analytical Chemistry*, 4th edn. New York: Prentice-Hall.
- Morgan E (1991) *Chemometrics: Experimental Design*. Chichester: Wiley.
- Myers RH and Montgomery DC (2002) *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. New York: Wiley.
- Otto M (1999) *Chemometrics: Statistics and Computer Application in Analytical Chemistry*. Chichester: Wiley-VCH.
- Sharaf MA, Illman DL, Kowalski BR (1986) *Chemometrics*. Chemical Analysis Series, vol. 82. New York: Wiley.

Walters FH, Parker LR, Morgan SL, and Deming SN (1991) *Sequential Simplex Optimization*. Boca Raton, FL: CRC Press.

## Glossary

Blocking	procedure by which experimental units are grouped into homogeneous clusters in an attempt to improve the comparison of treatments by randomly allocating the treatments within each cluster or 'block'.
Control	A control is a treatment, which is included to provide a reference set of data which can be compared with data obtained from the experimental treatments.
Experiment	An investigation in which the investigator applies some treatment(s) to experimental units to be observed and evaluated by measuring one or more response variables.
Experimental unit	A physical entity or subject subjected to the treatment independently of other units.
Factor	A categorical explanatory variable studied in an experiment, e.g., pH, flow rate.
Factorial designs	A factorial design is used to evaluate two or more factors simultaneously. The treatments are combinations of levels of the factors. The advantages of factorial designs over one-factor-at-a-time experiments are that they are more efficient and allow interactions to be detected.
Levels	The different values assigned to a factor.
Randomization	A random assignment of experimental material to treatments prior to the start of the experiment. Randomization is vital in the experimental design process and provides: (1) the basis for a valid interpretation of the experimental outcomes in terms of a test of statistical significance, and (2) the basis for computing a valid estimate of experimental error by justifying the assumption of independence of responses over experimental units.
Replication	When a given combination of factors is present in a system, replication can be used to: (1) demonstrate that results are reproducible, (2) provide a degree of assurance against erroneous results due to unforeseen reasons, (3) provide the means to estimate experimental error, and (4) provide the capacity to increase

Response variable	A characteristic of an experimental unit measured after treatment and analyzed	Treatments	The set of circumstances created for an experiment.

## Optimization Strategies

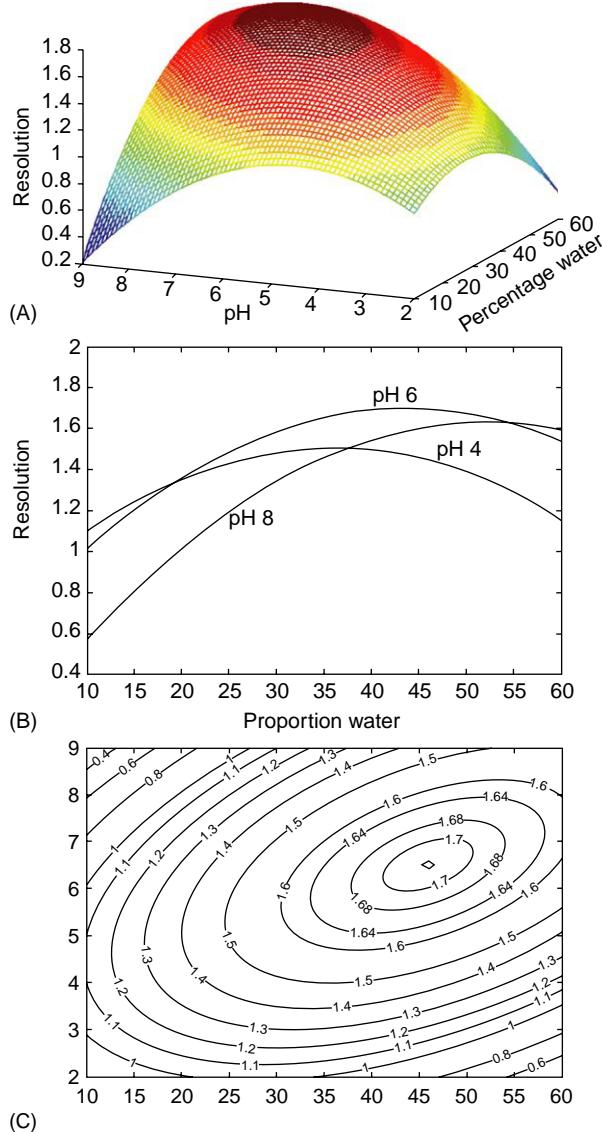
**R G Brereton**, University of Bristol, Bristol, UK

© 2005, Elsevier Ltd. All Rights Reserved.

### Introduction

Optimization has a significant role in analytical science. There are many reasons for finding an optimum. For example, it may be important to maximize the extraction efficiency of a compound from a matrix; there may be a large number of factors involved in the extraction procedure. Other examples involve improving chromatographic separations and optimizing the factors that influence signal intensity in atomic spectroscopy.

Traditional methods involve studying the influence of one factor at a time. For example, we might want to look at chromatographic resolution of two compounds using isocratic conditions, as a function of proportion of water (using a mobile phase of water and acetonitrile) and acidity (as controlled by the nature and amount of buffer). We do not know how the resolution varies as a function of proportion of water and pH prior to the experiment. Figure 1A illustrates a possible underlying response surface. We might try to set pH at a constant level, then vary the percentage water, until we reach an optimum, then set this percentage constant, varying the pH until a fresh optimum is chosen, which we use as the best conditions. The problem with this strategy is illustrated in Figure 1B. The change in resolution as a function of proportion of water differs according to pH, so a different optimum proportion of water is found according to the pH that is used. Hence, if we performed an optimization in which the initial experiments were performed at pH 4, we would obtain a different result for the best conditions to an optimization at pH 8 and so, in both cases a false optimum unless we happen by accident to have hit on the correct pH in the original experiments. The reason for this is that the influence of pH and mobile phase composition are said to interact and so cannot be considered independent factors. If there are several different factors involved in an optimization



**Figure 1** (A) Response surface for a typical chromatographic resolution as pH and proportion of water (in a water-acetonitrile mixture) is altered. (B) Resolution as a function of proportion of water at three different values of pH. (C) Representation as a contour plot.

it can be extremely hard to find true optima using traditional methods.

In order to overcome these difficulties, it is necessary to employ systematic approaches for determination