

General Introduction to Design of Experiments (DOE)

Ahmed Badr Eldin
Sigma Pharmaceutical Corp.,
Egypt

1. Introduction

Experimental design and optimization are tools that are used to systematically examine different types of problems that arise within, e.g., research, development and production. It is obvious that if experiments are performed randomly the result obtained will also be random. Therefore, it is a necessity to plan the experiments in such a way that the interesting information will be obtained.

2. Terminology

Experimental domain: the experimental 'area' that is investigated (defined by the variation of the experimental variables).

Factors: experimental variables that can be changed independently of each other

Independent Variables: same as factors

Continuous Variables: independent variables that can be changed continuously

Discrete Variables: independent variables that are changed step-wise, e.g., type of solvent.

Responses: the measured value of the result(s). from experiments

Residual: the difference between the calculated and the experimental result

3. Empirical models

It is reasonable to assume that the outcome of an experiment is dependent on the experimental conditions. This means that the result can be described as a function based on the experimental variables^[2],

$Y = f(x)$. The function $f(x)$ is approximated by a polynomial function and represents a good description of the relationship between the experimental variables and the responses within a limited experimental domain. Three types of polynomial models will be discussed and exemplified with two variables, x_1 and x_2 .

The simplest polynomial model contains only linear terms and describes only the linear relationship between the experimental variables and the responses. In a *linear model*, the two variables x_1 and x_2 are expressed as:

$$y = b_0 + b_1x_1 + b_2x_2 + residual.$$

The next level of polynomial models contains additional terms that describe the interaction between different experimental variables. Thus, a *second order interaction model* contains the following terms:

$$y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + \text{residual}.$$

The two models above are mainly used to investigate the experimental system, i.e., with screening studies, robustness tests or similar.

To be able to determine an optimum (maximum or minimum), quadratic terms have to be introduced in the model. By introducing these terms in the model, it is possible to determine non-linear relationships between the experimental variables and responses. The polynomial function below describes a *quadratic model* with two variables:

$$y = b_0 + b_1x_1 + b_2x_2 + b_{11}x_1^2 + b_{22}x_2^2 + b_{12}x_1x_2 + \text{residual}.$$

The polynomial functions described above contain a number of unknown parameters ($b_0, b_1, b_2, \text{etc.}$) that are to be determined. For the different models different types of experimental designs are needed.

4. Screening experiments

In any experimental procedure, several experimental variables or factors may influence the result. A screening experiment is performed in order to determine the experimental variables and interactions that have significant influence on the result, measured in one or several responses.^[3]

5. Factorial design^[4]

In a factorial design the influences of all experimental variables, factors, and interaction effects on the response or responses are investigated. If the combinations of k factors are investigated at two levels, a factorial design will consist of $2k$ experiments. In Table 1, the factorial designs for 2, 3 and 4 experimental variables are shown. To continue the example with higher numbers, six variables would give $2^6 = 64$ experiments, seven variables would render $2^7 = 128$ experiments, etc. The levels of the factors are given by - (minus) for low level and + (plus) for high level. A zero-level is also included, a centre, in which all variables are set at their mid value. Three or four centre experiments should always be included in factorial designs, for the following reasons:

- The risk of missing non-linear relationships in the middle of the intervals is minimised, and
- Repetition allows for determination of confidence intervals.

What - and + should correspond to for each variable is defined from what is assumed to be a reasonable variation to investigate. In this way the size of the experimental domain has been settled. For two and three variables the experimental domain and design can be illustrated in a simple way. For two variables the experiments will describe the corners in a quadrangle (Fig. 1), while in a design with three variables they are the corners in a cube (Fig. 2).

Two variables			Three variables				Four variables				
Exp. no.	Variables		Exp. no.	Variables			Exp. no.	Variables			
	x_1	x_2		x_1	x_2	x_3		x_1	x_2	x_3	x_4
1	-	-	1	-	-	-	1	-	-	-	-
2	+	-	2	+	-	-	2	+	-	-	-
3	-	+	3	-	+	-	3	-	+	-	-
4	+	+	4	+	+	-	4	+	+	-	-
			5	-	-	+	5	-	-	+	-
			6	+	-	+	6	+	-	+	-
			7	-	+	+	7	-	+	+	-
			8	+	+	+	8	+	+	+	-
							9	-	-	-	+
							10	+	-	-	+
							11	-	+	-	+
							12	+	+	-	+
							13	-	-	+	+
							14	+	-	+	+
							15	-	+	+	+
							16	+	+	+	+

Note that all variables are changed simultaneously in a controlled way, to ensure that every experiment in each design is a unique combination of variable levels.

Table 1. Factorial designs

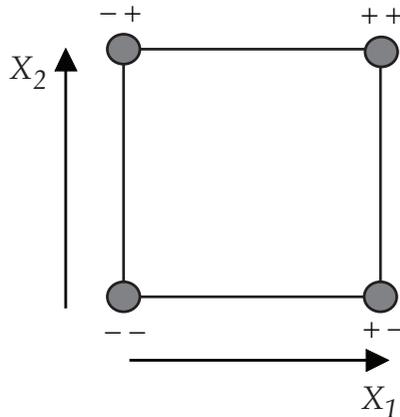


Fig. 1. The experiment in a design with two variables

6. Signs of interaction effects^[5]

The sign for the interaction effect between variable 1 and variable 2 is defined as the sign for the product of variable 1 and variable 2 (Table 2). The signs are obtained according to normal multiplication rules. By using these rules it is possible to construct sign columns for all the interactions in factorial designs.

Example 1: A 'work-through' example with three variables

This example illustrates how the sign tables are used to calculate the main effects and the interaction effects from a factorial design. The example is from an investigation of the influence from three experimental variables.

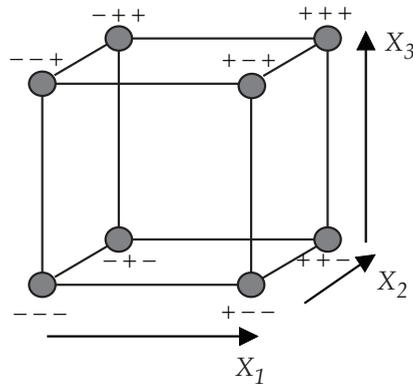


Fig. 2. The experiment in a design with three variables

7. Fractional factorial design

To investigate the effects of k variables in a full factorial design, 2^k experiments are needed. Then, the main effects as well as all interaction effects can be estimated. To investigate seven experimental variables, 128 experiments will be needed; for 10 variables, 1024 experiments have to be performed; with 15 variables, 32,768

experiments will be necessary. It is obvious that the limit for the number of experiments it is possible to perform will easily be exceeded, when the number of variables increases. In most investigations it is reasonable to assume that the influence of the interactions of third order or higher are very small or negligible and can then be excluded from the polynomial model. This means that 128 experiments

are too many to estimate the mean value, seven main effects and 21 second order interaction effects, all together 29 parameters. To achieve this, exactly 29 experiments are enough. On the following pages it is shown how the fractions $(1/2, 1/4, 1/8, 1/16 \dots 1/2^p)$ of a factorial design with 2^k experiments are defined, where

k is the number of variables and p the size of the fraction. The size of the fraction will influence the possible number of effects to estimate and, of course, the number of experiments needed. If only the main effects are to be determined it is sufficient to perform only 4 experiments to investigate 3 variables, 8 experiments for 7 variables, 16 experiments for 15 variables, etc. This corresponds to the following response function:

$$v = \beta_n + \sum \beta_i x_i + \varepsilon$$

It is always possible to add experiments in order to separate and estimate interaction effects, if it is reasonable to assume that they influence the result. This corresponds to the following second order response function:

$$y = \beta_0 + \sum \beta_i x_i + \sum \sum \beta_{ij} x_i x_j + \varepsilon$$

In most cases, it is not necessary to investigate the interactions between all of the variables included from the beginning. In the first screening it is recommended to evaluate the result

and estimate the main effects according to a linear model (if it is possible to calculate additional effects they should of course be estimated as well.).

After this evaluation the variables that have the largest influence on the result are selected for new studies. Thus, a large number of experimental variables can be investigated without having to increase the number of experiments to the extreme.

8. Optimization

In this part, two different strategies for optimization will be introduced; simplex optimization and response surface methodology. An exact optimum can only be determined by response surface methodology, while the simplex method will encircle the optimum. simplex is a geometric figure with $(k+1)$ corners where k is equal to the number of variables in a k -dimensional experimental domain. When the number of variables is equal to two the simplex is a triangle (Fig. 16.).

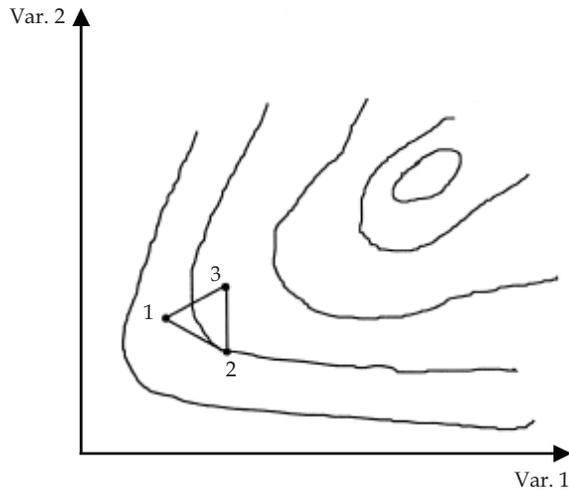


Fig. 3. A simplex in two variables

Simplex optimization is a stepwise strategy. This means that the experiments are performed one by one. The exception is the starting simplex in which all experiments can be run in parallel. The principles for a simplex optimization are illustrated in Fig. 17. To maximize the yield in a chemical synthesis, for example, the first step is to run $k+1$ experiments to obtain the starting simplex. The yield in each corner of the simplex is analyzed and the corner showing the least desirable result is mirrored through the geometrical midpoint of the other corners. In this way, a new simplex is obtained. The co-ordinates (i.e., the experimental settings) for the new corner are calculated and the experiment is performed. When the yield is determined,

the worst of the three corners is mirrored in the same way as earlier and another new simplex is obtained, etc. In this way, the optimization continues until the simplex has rotated and the optimum is encircled. A fully rotated simplex can be used to calculate a response surface. The type of design described by a rotated simplex is called a Doehlert design.

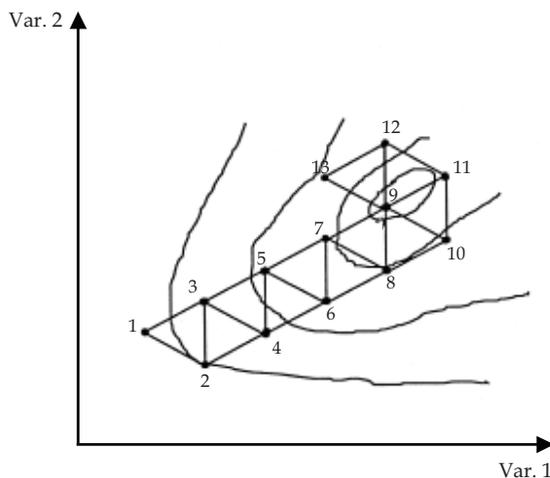


Fig. 4. Illustration of a simplex optimization with two variables

9. Rules for a simplex optimization

With k variables $k+1$ experiments are performed with the variable settings determined by the co-ordinates in the simplex. For two variables the simplex forms a triangle. For three variables it is recommended to use a 2^{3-1} fractional factorial design as a start simplex.

10. References

- [1] Experimental design and optimization, Chemometrics and Intelligent Laboratory Systems 42_1998. 3-40
- [2] R. Sundberg, Interpretation of unreplicated two-level factorial experiments, Chemometrics and intelligent laboratory system, 24_1994. 1-17.
- [3] Atkinson, A. C. and Donev, A. N. Optimum Experimental Designs Clarendon Press, Oxford p.148.
- [4] Kowalski, S.M., Cornell, J.A., and Vining, G.G. (2002) "Split Plot Designs and Estimation Methods for Mixture Experiments with Process Variables," Technometrics 44: 72-79.
- [5] Goos, P. (2002) The Optimal Design of Blocked and Split-Plot Experiments, New York: Springer



Wide Spectra of Quality Control

Edited by Dr. Isin Akyar

ISBN 978-953-307-683-6

Hard cover, 532 pages

Publisher InTech

Published online 07, July, 2011

Published in print edition July, 2011

Quality control is a standard which certainly has become a style of living. With the improvement of technology every day, we meet new and complicated devices and methods in different fields. Quality control explains the directed use of testing to measure the achievement of a specific standard. It is the process, procedures and authority used to accept or reject all components, drug product containers, closures, in-process materials, packaging material, labeling and drug products, and the authority to review production records to assure that no errors have occurred. The quality which is supposed to be achieved is not a concept which can be controlled by easy, numerical or other means, but it is the control over the intrinsic quality of a test facility and its studies. The aim of this book is to share useful and practical knowledge about quality control in several fields with the people who want to improve their knowledge.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Ahmed Badr Eldin (2011). General Introduction to Design of Experiments (DOE), Wide Spectra of Quality Control, Dr. Isin Akyar (Ed.), ISBN: 978-953-307-683-6, InTech, Available from:

<http://www.intechopen.com/books/wide-spectra-of-quality-control/general-introduction-to-design-of-experiments-doe->

INTECH

open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.