الجمهورية الجزائرية الديمقراطية الشعبية

People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

University Blida 1

FACULTY OF SCIENCE

DEPARTMENT OF MATHEMATICS

Master degrees

Option: Stochastic Modeling and Statistics

TITLE

Methodology of Experimental Design: Synthesis and Development of an Application in Python

---

Presented by: Zitoun Kaouther

Examing boord:

Dr MESSAOUDI A.N.	Associate Professor B	President
Mr BOUKHARI M.	Assistant Professor A	Examinator
Dr. ELMOSSAOUI H.	Associate Professor A	Supervisor

2023/2024

الملخص

تتمتع منهجية البحث التجريبي بمكانة بارزة في علم الرياضيات كجزء من علم الإحصاء. تُستخدم هذه المنهجية على نطاق واسع من قبل الباحثين والصناعيين لتخطيط تجاربهم، ونمذجة النتائج، وإجراء التحليلات.

يهدف هذا العمل إلى تحقيق هدفين رئيسيين: أولاً، تقديم خلاصة لمنهجية تصميم التجارب حيث يتم وصف فئتين من التصاميم التجريبية الكلاسيكية؛ وثانياً، تطوير تطبيق بايثون لتلبية احتياجات كل مستخدم ومجرب. تكمن فائدة وجود برنامج خاص في تسهيل عملية تخطيط التجارب، ونمذجة النتائج، وتحليلها بشكل كبير. بالإضافة إلى ذلك، يمكن تحسين هذا البرنامج في المستقبل عن طريق إضافة خيارات أخرى.

الكلمات الرئيسية : تصاميم التجارب، الانحدار الخطي المتعدد، تصاميم التجارب الإستكشافية، تصاميم سطح الاستجابة، معايير الأمثلية، التصاميم العاملية الكاملة.

Abstract

The methodology of experimental research is a relatively recent discipline in mathematics, forming part of statistics. It is widely used by both researchers and industrialists to plan their experiments, model the results, and perform their analysis.

This work has two main objectives: first, to provide a synthesis of the methodology of experimental design, where two categories of classical experimental designs are described, and second, to use Python to elaborate an application to meet the needs of each user and experimenter. The utility of having one's own software is evident. The planning of experiments, modeling, and analysis of results will be considerably facilitated. Additionally, this software can be further improved in the future by adding other options.

Key Words : Experimental designs, multiple linear regression, screening experimental designs, response surface designs, optimality criteria, full factorial designs.

ACKNOWLEDGEMENT

First and foremost, I would like to express my deepest gratitude to God almighty for granting me the courage,patience and determination to complete this humble work.

I would like to thank my supervisor, Dr. ELMOSSAOUI HICHEM, who helped me with his expertise and understanding. Without his invaluable guidance and support, this thesis would not have been possible.

I would also like to thank the assistance and support provided by the administrative of the Mathematics Department, in particular MR. O.Tami, Mr. H.Hadj-Allah, Ms. N.Djenas and Ms. S.Takarli. Their help and support were crucial in the completion of this thesis.

I take this opportunity to express my profound gratitude and sincere thanks to all the teachers who have take part to our education.

I am profoundly grateful to the members Dr.Messaoudi A.N, Mr.Boukhari M, for their insightful comments, constructive criticism, and encouragement, which have greatly enhanced the quality of this work.

DEDICATION

To my dearest parents,Ali and Faiza, Your belief in me, your patience, and your sacrifices have been my greatest source of strength. I am especially grateful to you for instilling in me the value of education and hard work. Thank you for always being there, through the challenges and the triumphs, and for inspiring me to strive for excellence.

To my lovely brothers Adel and Oussama, my gorgeous sister Wafaa, forever grateful and forever in my heart.

To my big family, your companionship, understanding, and motivation have made this journey not only bearable but also enjoyable.

To my friends, Thank you for lifting me up when I was down and for celebrating every success, big or small, with me, in particular Maleg.H.

To my colleagues at the Mathematics Department, Fatiha, Lydia, Manelle Soumia, Serine, Abdelkader, Abderraouf, Yassine, Noureddine, Hamza and specially Oud feul Adel, whose camaraderie and intellectual discussions have made my time here both productive and enjoyable. Thank you all.

Finally, I would like to thank all those who have contributed to this thesis in one way or another, who have stood by my side through thick and thin. Your assistance and support have been invaluable and are deeply appreciated.

This achievement is as much yours as it is mine.
With all my gratitude.

KAOUTHER

Contents

List of Figures

List of Tables

Meanings

$F_{1-\alpha,p,n-p-1}$	:	Tabulated value.
$(X^tX)^{-1}$	:	Dispersion matrix.
(X^tX)	:	Information matrix.
$Var(\hat{y})$	:	Variance of the error on the measured responses.
$Var(\hat{A})$	:	Variance of the coefficients estimated by the least squares method.
SST	:	Total sum of squares.
SSR	:	Sum of squares due to regression.
SSE	:	Sum of squares due to error.
$Im(X)$	:	Vector subspace spanned by the columns of X.
$Im(X)^{\perp}$	:	Orthogonal subspace.
$\sigma(\hat{a}_i)$	:	Standard deviation of the coefficients determined by the least squares method.
σ_r^2	:	Variance of the residuals.
σ_y	:	Standard deviation of the error on the measured responses.
$d(\hat{y}_p)$	:	Prediction error function.
$d^2(\hat{y}_p)$	:	Prediction variance function.
$\bar{y}_-$	:	Mean of the responses at the low level of a factor.
$\bar{y}_+$	:	Mean of the responses at the high level of a factor.
$\hat{y}$	:	Vector of predicted responses.
$\hat{y}_i$	:	Predicted response of trial i.
Y	:	Vector of observed responses.
$\bar{Y}$	:	Observed mean response.
Y^*	:	Vector of centered observed responses.
y	:	Response or response variable.
y_i	:	Response of trial i.
a_0	:	Constant of a model.
a_i	:	Coefficients of the first-degree terms.
$a_{i,i}$	:	Coefficients of the squared terms.
$a_{i,j}$	:	Coefficients of rectangular terms.
A	:	Vector of coefficients of the model.
$\hat{A}$	:	Vector of coefficients estimated by the least squares method.
α	:	Distance to the center of the star points in a composite design.
X^t	:	Transposed matrix of X.
X	:	Design matrix.
x_p	:	Modeled vector of the point p.

x_i	:	Level assigned to the different factors and expressed in centered reduced values,variable i.
$\hat{a}_i$	:	The coefficient of the variable x_i in a model determined by the least squares method.
e_i	:	The residual gap.
ε_i	:	The true value of the gap e_i.
E	:	Vector of deviations.
ξ	:	Mathematical expectation.
$E(.)$	:	Mathematical expectation.
f	:	Function.
F	:	Fisher-Snedecor.
I	:	Unit matrix, alias generator, mean in an effects table.
k	:	Number of factors.
n_0	:	Number of trials at the center of the study domain.
n_f	:	Number of trials in the factorial design.
n_α	:	Number of trials in the star design.
N	:	Number of trials in an experimental design.
p	:	Number of additional factors for a fractional factorial design.
q	:	Number of coefficients in a mathematical model.
z_0	:	The central value of the interval [-1, +1] expressed in current units.
z	:	Level assigned to the factor expressed in current units.
Q	:	Change Proposal Transition.
$\partial(.)$	:	Neighborhood.
ddl	:	Degrees of freedom.
Δ	:	The lack of fit.
R^2	:	Correlation coefficient.

Introduction

The Experimental Research Methodology (Design of Experiments) is valuable for anyone engaged in scientific research or industrial studies. The use of experimental designs for the empirical study of a response law poses particular challenges for statisticians or researchers. While they have little information about this law, they generally only have a very limited sample of observations relative to the number of parameters of the models they can consider for their analyses. Before any observation of the response, they must specify not only which models to use but also how to organize the experiments. Indeed, the quality of statistical analysis depends closely on the experimental design used to observe the response. Additionally, combinatorial analysis is generally required to construct the proposed experimental designs.

To propose a solution that meets industrial objectives, it is sometimes necessary to seek missing information by conducting a series of experiments. Important decisions based on experimental results and the significant cost of experimentation prohibit relying solely on the experimenter's intuition to find the solution. A methodological approach is needed to not only reduce the cost of experimentation but also to establish an optimal organization of experiments.

The goal of the design of experiments methodology is to propose one or more strategies to solve a particular problem posed by experimental research.

In our work, the general principles of constructing experimental designs are presented based on the notion of experimental space. The geometric representation of experimental points is very illustrative, but it quickly becomes limited as the space's dimension increases. Therefore, a matrix representation is used. With the help of both geometric and matrix representations, the main experimental designs are described: full factorial designs at two levels, fractional designs, Mozzo designs, full factorial designs at three levels, composite designs, Box-Behnken designs, Doehlert designs, and Roqumore designs. A Python application has been developed. It allows the construction of the eight plans mentioned above, provides the model, representative graphs, as well as the analysis and numerical results necessary to conduct a successful experimental study. The main concepts developed in the Experimental Research Methodology have been addressed in the first chapter.

In the second chapter, we explored the method of multilinear regression. This method was utilized in our application to compute the model coefficients. To validate this model, we provided a review of statistical tools such as Analysis of Variance, Residual Analysis, etc. Additionally, we discussed important concepts such as error propagation to model coefficients, and consequently, to predicted responses, as well as optimality criteria.

The third chapter focuses on describing classical screening designs, while in the fourth chapter, we delve into detailed descriptions of the primary optimization designs

used to establish a response surface.

Finally, the description of the developed software was addressed in the fifth chapter.

Chapter 1

Generalities, Method And Models For Experimental Design

In this chapter, various hypotheses involved in the use of the experimental design method are synthesized and summarized. This method is useful for any experimenter conducting scientific research or industrial studies. It is applicable to all disciplines as long as one is seeking the relationship between a variable of interest, y, and other variables that can influence its values. To achieve this, it is necessary to follow mathematical rules and adopt a rigorous approach.

1.1 History

The method of experimental design is not new; it is indeed extensive and can be applied across many fields [1]. It has been around for a long time. These designs are valuable to anyone undertaking scientific research or industrial studies. The earliest users of these methods were agronomists and statisticians. After 1945, experimental designs sparked numerous publications and research in the Anglo-Saxon world. Many researchers have continued to develop this branch of statistics in various ways and directions.

Statisticians like Yates, Youden, Cochran, Plackett, and Burman, among others, subsequently followed in the footsteps of Fisher [2], enriching and disseminating the use of experimental design techniques in other fields. In the 1950s, Box and Hunter [3], drawing notably on the work of Yates, developed specific methods for constructing fractional factorial designs at two levels. Taguchi and Masuyama elaborated tables for constructing orthogonal experimental designs suitable for the majority of industrial problems [4]. These tables were published in 1959 and 1961 [3]. Subsequently, Taguchi, with pragmatism, simplified and clarified the use of these designs.

Once understood, the method becomes an irreversible step in the technician's career, as they can no longer consider conducting experiments without using an experimental design. Once confidential, it is now experiencing a spectacular development. With experimental designs, we obtain the maximum amount of information with the minimum number of experiments. To achieve this, we must follow mathematical rules and adopt a rigorous approach [5].

1.2 Limitation of Traditional Testing Methods

1.2.1 Study Of A Phenomenon

The analysis of a phenomena may be schematized as follows: the researcher concentrates on a quantity, such as the amount of wheat produced on a plot of land or the price of a chemical product used in an engine for an automobile. There are several factors that will affect this amount, including the kind of ground, the quality of fertilizer applied, the environment, exposure to sunshine, and more [5].

Mathematically, we can write that the variable of interest y as a multi-variable function x_i.

$$y = f(x_1, x_2, \ldots, x_k) \tag{1.1}$$

The examination of a phenomenon is then reduced to measuring the quantity based on the various values that we can give to the variables. The traditional approach to studying the function will be briefly described(1.1).

1.2.2 The Classical Method

In traditional practice, trials are performed sequentially [6] by varying the variables one after the other without prior planning for the complete set of them. We fix the level of all variables except one (variable $x1$) and we measure the quantity y corresponding to different levels of this last.

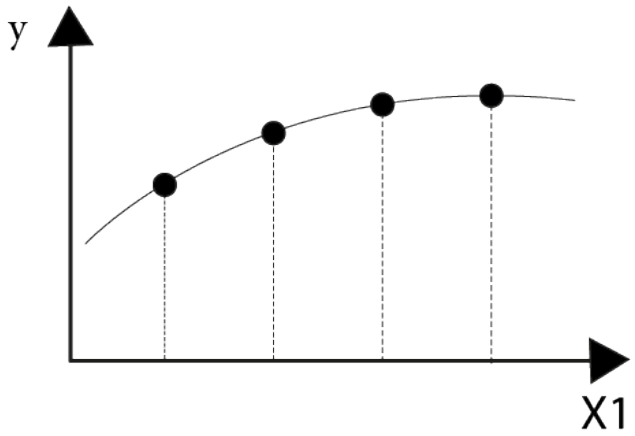

Figure 1.1: All variables are fixing at well-defined value except x_1 who takes different levels

At the end of the experiment on this first variable,we can plot a curve represanting $y = f(x_1)$ (fig 1.1). We can repeat the same experience with all different variables. If we want to study six variables and we decide to take five experimental points for each variable , we must run $5^6 = 15625$ experiences.
We only act in two ways when the tester must reduce down on the number of trials:

Decrease experimental points: for three points instead of five it still $3^6 = 729$ experiences to be made. And for two points, we have $2^6 = 64$ experiences .

Decrease number of variables: for four variables we take each three value, and we must run $3^4 = 81$ experiences.

Hence, the drawbacks of this approach become evident when dealing with security risks or significant financial amounts. We will then discuss the Experimental Design Method.

1.3 The Experimental Design Method

For any experimenter doing industrial studies or scientific research, the experimental design approach is a useful tool.While it does so in a planned and logical way, it applies to all fields. It also links between a magnitude of interest y and variables x_i. for this purpose we should follow and take strict mathematical rules such as [7] :

-Decrease the number of trials.

-The number of factors studied can be quite large.

-Detection of interactions among factors.

-Better precision on the results.

-Modeling the results.

Understanding Two essential ideas"Experimental space" and "Mathematical modeling" are the foundation of the experimental design technique.

1.4 Concept of Experimental Space

1.4.1 Responses, Factors, Levels

- The response is the quantified quantity of interest in every experiment. A number of things influence this response's value. We exclusively use the expression "factors" in place of the term "variables."
- The initial factor can be represented by a graduated and oriented axis (fig 1.2). The value assigned to a factor for a specific trial is called a level. When we're studying the influence of a factor, we typically limited its variation between two bounds (the lower bound is the low level, and the upper bound is the high level).

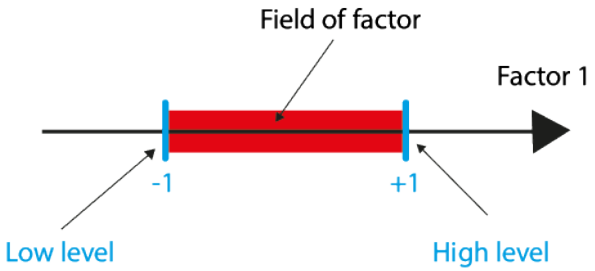

Figure 1.2: Range of variation of the factor

- The factor's field of variation is the collection of all values that it can take between the low and high levels. Typically, the low level is denoted as -1, while the high level is represented as +1.
- An axis is used to indicate any additional factors that may exist. Its range of fluctuation, high level, and low level can alternatively be defined as the first factor.This second axis is positioned orthogonally to the first one. Thus, a Cartesian coordinate system is obtained defining a two-dimensional Euclidean space known as the experimental space (fig 1.3).

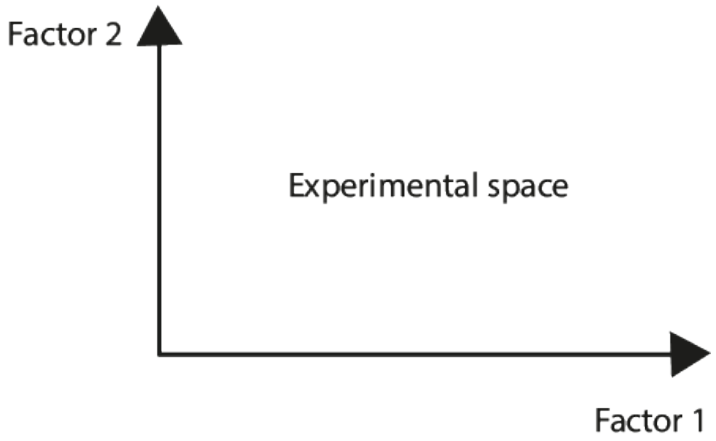

Figure 1.3: Definition of the experimental design

- The levels x_1 and x_2 of factors 1 and 2 respectively may be seen as the coordinates of a point in the experimental space(fig 1.4). In this axis system, a point corresponds to each individual experiment.An experimental design is shown by a set of data points.

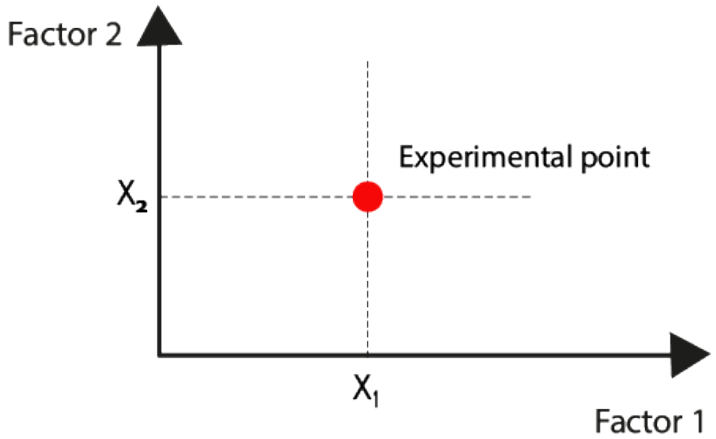

Figure 1.4: Levels of factors defining experimental points in the experimental space

1.4.2 Field of Study, Response Surface

- The factors' fields are grouped to form the "study field." The region of the experimenter's chosen experimental space where the trials will be conducted is known

as the field of study. Points dispersed over the study's region indicate an experiment, which consists of many well-defined experiments in (fig1.5). This method of using points in a Cartesian space to illustrate an experiment is a geometric representation of the research [7].

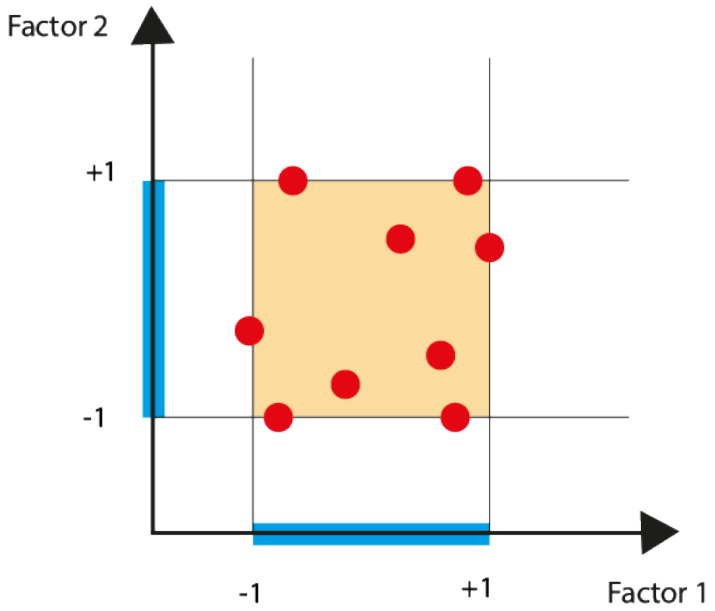

Figure 1.5: defenition of field of study

- The geometric representation of the experimental design and the response requires a space with one more dimension than the experimental space. A three-dimensional space is used to represent a two-factor design, with one dimension for the response and two for the factors. Each point within the study domain corresponds to a set of responses which are located on a surface known as the response surface (fig 1.6) [8].

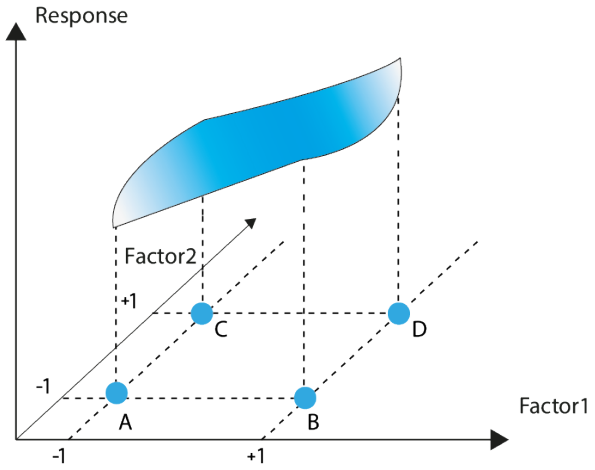

Figure 1.6: Defenition of the response surface

1.4.3 Centered Reduced Coordinates

When we assigned the value of -1 to the low level of a factor and +1 to the high level, two significant modifications are performed:

-The origin of measurements is shifted.

-The unit of measurements is altered.

New variables known as standardized centered variables are created as a result of these two changes, the terms "standardized" and "centered" relate to the new unit and the origin shift, respectively. The transformation from the original variables z to the standardized centered variables x, and vice versa, is given by the following formula:

$$x = \frac{z - z_0}{step}$$

with:

$$z_0 = \frac{highlevel + lowlevel}{2}$$
$$step = \frac{highlevel - lowlevel}{2}$$

Standardized variables are used to present experimental designs uniformly, independent of the factors and study domains selected. This gives the theory of experimental designs a great degree of generality [7].

1.4.4 Experimental Design

Every point in the research area represents a potential operational situation, which gives the operator the opportunity to perform out an experiment.

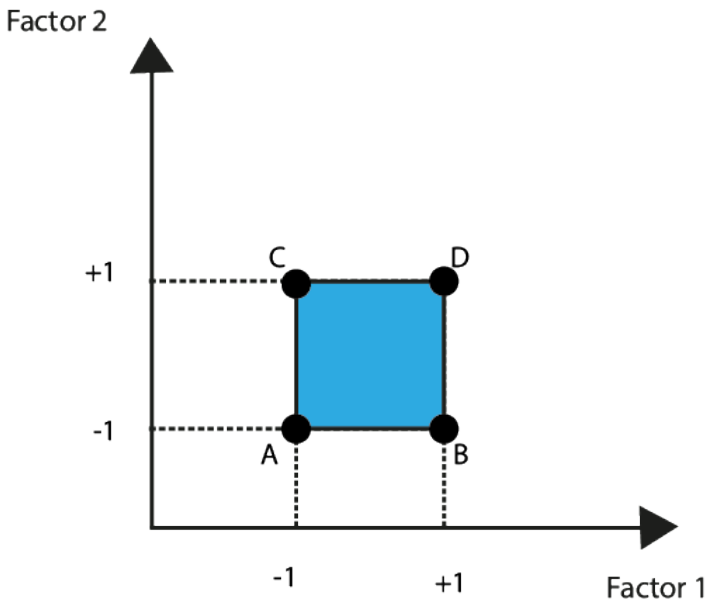

Figure 1.7: The vertices A, B, C, and D of the study domain are the ideal positions, as shown by the theory of experimental designs

The primary issue in experimental designs is the selection of the quantity and placement of experimental points. Typically, experimental designs are collections of experimental points that follow certain criteria. These are the traditional experimental designs, which are well-known and frequently reported. Non-conventional designs refer to experimental point arrangements that differ from those of traditional experimental designs. Compared to classical designs, their qualities are frequently less ideal. It is not always feasible to follow the specifications of traditional experimental designs, which leads to the occurrence of non-conventional designs [8].

1.4.5 Matrix of Experiments

For the task of visualizing the positions of experimental points within the research domain, an experimental design's geometric representation is useful. When there are more than three components, though, it becomes unfeasible. We use a table or experiment matrix as our representation for multidimensional spaces. The trials to be performed are specified in the experiment matrix (Tab1.1). The idea of an experimental point is synonymous with the phrase "trial" [9].

Table 1.1: Experiences matrix

N°.Trials	Factor 1	Factor 2
1(A)	-1	-1
2(A)	1	-1
3(C)	-1	1
4(D)	1	1

1.5 Concept Of Mathematical Modeling

We choose a mathematical function a priori that relates the response to the factors. The Taylor-Mac Laurin series is expanded to a limited expansion by us. The derivatives are assumed to be constant, and the expansion takes the form of a polynomial of more or less high degree [10]:

$$y = a_0 + \sum a_i x_i + \sum a_{ij} x_i x_j + \sum a_{ii} x_i^2 + \dots$$

- y is the response, or the quantity of interest, is measured during the experimentation, and it is obtained with a given precision,

- x_i represents the level assigned to factor i by the experimenter to conduct a trial. This value is perfectly known. It is even assumed that this level is determined without error (classic regression hypothesis),

- $a_0, a_i, a_{ij}, a_{ii}, \dots$ are coefficients of the model.they aren't familiar and they should be calculated starting from the results of experiments.

The term "a priori model" or "postulated model" refers to this concept.

1.5.1 Interest Of Polynomial Representation

Compared to other modeling techniques, the polynomial representation of the response makes matrix processes possible, thus explaining why it is so valuable. Additionally, it

allows the interactions and effects of elements to be included in the expression directly of y something that would not be feasible if we used other mathematical functions like exponentials, algorithms, or more. Polynomials are the favored choice for experimenters since experience shows them to be a successful solution to a wide range of issues.

1.5.2 Mathematical Model For Experimental Design

During an experiment, the measured responses are randomized quantities that require to be handled randomly.It is necessary to take reality into account.The identical experiment will not produce precisely the same results in every determination, no matter how hard one tries. An uncertainty known as experimental error, measurable error, or pure error taints every measurement. This uncertainty arises from variations in specific parameters that the investigator is unaware of. Those models fall into three types [11].

Mathematician's model

The lack of experimental error is a characteristic of a mathematician's mathematical model. For example, with two factors:

$$y = a_0 + a_1x_1 + a_2x_2 + a_{12}x_1x_2 + a_{11}x^2 + a_{22}x^2$$

This model simplifies analysis but doesn't reflect real-world experiments where measurements have inherent uncertainty.

Experimentator's model

There are three parts to this concept. A mathematician's model makes up the first component, adjustment error makes up the second, and random error makes up the third:

$$y = a_0 + a_1x_1 + a_2x_2 + a_{12}x_1x_2 + a_{11}x^2 + a_{22}x^2 + \Delta + \sigma_y$$

This model provides a more realistic picture but can be complex to analyze statistically.

Statistician's model

The prior model is too complicated to be solved:

$$y = a_0 + a_1x_1 + a_2x_2 + a_{12}x_1x_2 + a_{11}x^2 + a_{22}x^2 + \sigma_y$$

This model allows for easier statistical analysis and hypothesis testing but may not capture all sources of error.

1.6 System Of Equations

A response value is brought to each experimental location. The unknowns that must be found are represented by the coefficients of a polynomial in this form. When the experimental design is complete, we have a system of N equations with q unknowns (as long as the model chosen a priori has q coefficients) if there are N trials. This system can be expressed using a basic matrix notion [12]:

$$Y = XA + e \tag{1.2}$$

with:

- Y: responses vector,

- X: calculate matrix, who depends on experimental points chosen for run the design and the applied model,

- A: coefficients vector,

- e: residual vector.

This problem is usually difficult to solve since there are fewer equations than unknowns. as a matter of fact, there are N equations and $q+N$ unknowns; this resolution can only conducted in a good way if a regression method is used, introducing q additional equations. This approach often relies on the least squares optimization criteria.

Chapter 2

MLR Regression, Error Transmission and Optimality Criteria

In this chapter, we explore three key concepts essential for understanding and applying experimental design: multiple linear regression (MLR), error transmission, and optimality criteria. Multiple linear regression allows for modeling the relationship between a dependent variable and several independent variables, providing a powerful approach for modeling data from an experimental design. Error transmission plays a crucial role in assessing the precision of measurements and estimates derived from statistical models. Finally, optimality criteria provide guidelines for designing experiments to achieve reliable and significant results. This chapter aims to examine these concepts in detail, illustrating their importance in optimizing experimental processes.

2.1 Multilinear Regression (MLR)

The aim is to determine which set of q coefficients best solves the system of equations.

2.1.1 Principle

The multiple linear regression model is the most commonly used statistical tool for the study of multidimensional data. As a specific case of the linear model, it represents the natural generalization of simple regression.

Table 2.1: Presentation of the results of trials for a simple regression

x	y
x_1	y_1
x_2	y_2
$\vdots$	$\vdots$
x_i	y_i
$\vdots$	$\vdots$
x_N	y_N

The term "multiple" refers to the fact that there are several explanatory variables x_j to explain y.

Table 2.2: Presentation of the results of trials for a multilinear regression

x_1	...	x_i	x_i	x_k	y
x_{11}	...	x_{1i}	...	x_{1k}	y_1
⋮	⋮	⋮	⋮	⋮	⋮
x_{i1}	...	x_{ii}	...	x_{ik}	y_i
⋮	⋮	⋮	⋮	⋮	⋮
x_{N1}	...	x_{Ni}	...	x_{Nk}	y_N

For N observations, the model is expressed in this form:

$$y = a_0 + a_1x_1 + a_2x_2 + \cdots + a_jx_j + \cdots + a_qx_q + e$$

with:
y: Endogenous variable.
$x_1, x_2, \ldots, x_j, \ldots, x_q$: Exogenous variables,

The least squares approach is used to estimate all of the model's coefficients. And here's the plot of the least squares regression line:

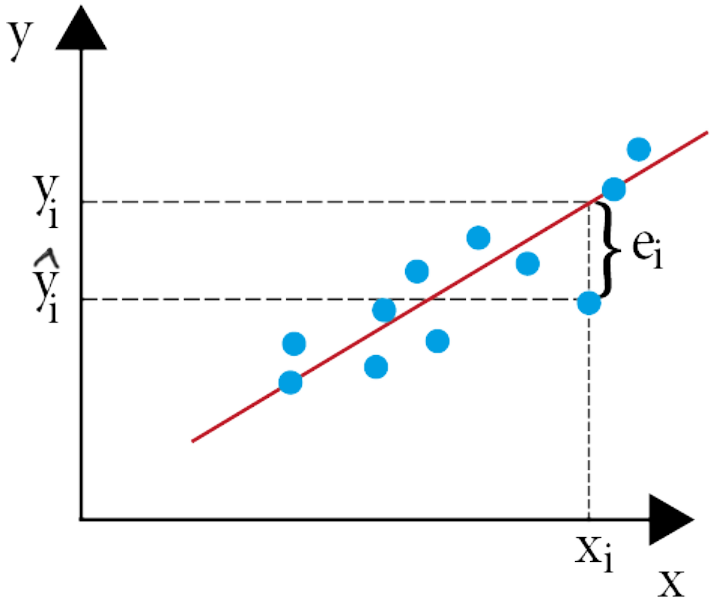

Figure 2.1: The least squares line

2.1.2 Estimation of Coefficients by the Least Squares Method

Selecting the best possible estimator $\hat{A}$ of A is the problem now that the model has been established.One traditional method of searching for $\hat{A}$ is to align the vector of observed responses Y with the vector of predicted means : $\hat{Y} = X\hat{A}$ as closely as feasible :

Definition :
We say that $\hat{A}$ is the least squares estimator of A if and only if $\hat{A}$ minimizes the

function:

$$Q(A) = ||Y - X\hat{A}||^2$$

The function Q is minimized by the least squares estimator of A, and this minimum is thus:

$$Q(\hat{A}) = ||Y - X\hat{A}||^2 = ||Y - \hat{Y}||^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

This demonstrates that there is a relationship between this quantity and the (quadratic) error between the mean responses predicted by the model $\hat{Y}_i$ and the observed responses Y_i. As for how this estimate may be determined practically, we demonstrate that:

Proposition 1.1 : The statistical model $Y = XA + e$ with X being a full rank matrix. The least squares estimator of A is given by:

$$\hat{A} = (X^tX)^{-1}X^tY$$

The proof : We are looking for $\hat{A}$ that minimizes the quantity : $||Y - \hat{Y}||^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$, Let's represent $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$ as a function of $\hat{A}$ in order to achieve this :

$$\begin{aligned}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 &= (Y - X\hat{A})^t(Y - X\hat{A}) \\ &= (Y^t - \hat{A}^tX^t)(Y - X\hat{A}) \\ &= Y^tY - \hat{A}^tX^tY - Y^tX\hat{A} - \hat{A}^tX^tX\hat{A}\end{aligned}$$

Note that $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$ is a scalar, and it is easy to verify that all terms in the sum are also scalars. Therefore, we have:

$$Y^tX\hat{A} = (\hat{A}^tX^tY)^t = \hat{A}^tX^tY$$

Is required to :

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = Y^tY - 2\hat{A}^tX^tY + \hat{A}^tX^tX\hat{A}$$

Let us get the derivative of $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$ concerning the variable $\hat{A}$:

$$\frac{\partial\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\partial\hat{A}} = \frac{\partial(Y^tY)}{\partial\hat{A}} - 2\frac{\partial(\hat{A}^tX^tY)}{\partial\hat{A}} + \frac{\partial(\hat{A}^tX^tX\hat{A})}{\partial\hat{A}}$$

Where :

- $\frac{\partial(Y^tY)}{\partial\hat{A}} = 0$ (Because Y^tY doesn't depend on A).
- $\frac{\partial(\hat{A}^tX^tY)}{\partial\hat{A}} = X^tY$ (Because $\hat{A}^tX^tY$ is a linear form in $\hat{A}$).
- $\frac{\partial(\hat{A}^tX^tX\hat{A})}{\partial\hat{A}} = 2X^tX\hat{A}$ (Because $\hat{A}^tX^tX\hat{A}$ is a quadratic form in $\hat{A}$).

Thus, it comes:

$$\frac{\partial\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\partial\hat{A}} = -2X^tY + 2X^tX\hat{A}$$

The value of $\hat{A}$ that minimizes $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$ must satisfy:

$$\frac{\partial \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\partial \hat{A}} = 0$$
$$\implies -2X^tY + 2X^tX\hat{A} = 0$$
$$\implies X^tX\hat{A} = X^tY$$

$$\implies \hat{A} = (X^tX)^{-1}X^tY \tag{2.1}$$

2.1.3 Hypotheses of Multilinear Regression

We separate between stochastic hypotheses and structural hypotheses by applying the normal distribution as we see here [13]:

- The values $X_{1i}, X_{2i}, \ldots, X_{pi}$ are observed without errors,
- Residuals have a mean of zero $E(\epsilon_i) = 0, \forall i = 1, \ldots, n$,
- The variance is equal to the unbiased estimator $var(\epsilon_i) = \sigma_\epsilon^2; \forall i = 1, \ldots, n$.
- The errors are uncorrelated ; $E(\epsilon_i\epsilon_j) = 0; \forall i \neq j$,
- The error ϵ_t is independent of the observations $X_{1t}, X_{2t}, \ldots, X_{pt}$, $E(X_{it}\epsilon_t) = 0, \forall i = 1, \ldots, p$ and $\forall t = 1, \ldots, n$,
- The exogenous variables $X_1, X_2, \ldots, X_p$ are not collinear; therefore, the matrix X^TX is invertible,
- This matrix $(X^TX)/n$ is a finite matrix,
- $n >> p + 1$ the number of observations is substantially higher than the number of exogenous variables.

2.1.4 Properties of Coefficients

If $\hat{A}$ is the least squares estimator of A and the residuals' hypotheses are satisfied, then:

1. $\hat{A}$ is an unbiased estimator of A,
2. $\hat{A}$ accepts as its matrix of covariance : $var(\hat{A}) = \sigma^2(X^TX)^{-1}$.

Mathematical expectation of the coefficients: $E(\hat{A}) = E[(X^tX)^{-1}X^tY] = (X^tX)^{-1}X^tE(y)$, because the elements of X are considered fixed. Referring to A as the vector of p true coefficients and ε as the vector of n deviations between experimental results and theoretical responses, we have then :

$$Y = XA + \varepsilon$$
$$\implies E(Y) = E(XA + \varepsilon) = E(XA) + E(\varepsilon) = XE(A)$$

From our hypotheses ($E(\varepsilon) = 0$), so we have:

$$E(\hat{A}) = (X^tX)^{-1}X^tXA = A$$

Variance of the coefficients : Giving $var(\hat{A}) = E[(\hat{A} - A)(\hat{A} - A)^t]$, then we replace $\hat{A}$ by $(X^tX)^{-1}X^tY$ and Y by $XA + \varepsilon$, We get :

$$(\hat{A} - A) = (X^tX)^{-1}X^t(XA + \varepsilon) - A = A + (X^tX)^{-1}X^t\varepsilon - A = (X^tX)^{-1}X^t\varepsilon$$

As $(\hat{A} - A)^t = \varepsilon^t X(x^tX)^{-1}$,so :

$$var(\hat{A}) = E[(X^tX)^{-1}X^t\varepsilon\varepsilon^tX(X^tX)^{-1}] = (X^tX)^{-1}X^tE(\varepsilon\varepsilon^t)X(X^tX)^{-1}$$

Changes $E(\varepsilon\varepsilon^t)$ by $E[(\varepsilon - 0)(\varepsilon - 0)^t] = var(\varepsilon) = \sigma^2$, we can write :

$$\begin{aligned} var(\hat{A}) &= (X^tX)^{-1}X^t\sigma^2X(X^tX)^{-1} \\ &= \sigma^2(X^tX)^{-1}X^tX(X^tX)^{-1} \end{aligned}$$

$$\implies var(\hat{A}) = \sigma^2(X^tX)^{-1} \tag{2.2}$$

2.2 Evaluation of Model Quality

The model's ability to effectively summarise the experiment results within the experimental design is determined largely by the model's quality evaluation. With statistical techniques, the quality may be evaluated. There have been four methods explained for carrying out this evaluation.

2.2.1 Graphical Examination of Results

It is always enriching and frequently aids in preventing significant misunderstandings when this assessment is possible.

2.2.2 Analysis of Variance

Once the model is fitted, the issue of the quality of the obtained fit arises. Using what are known as analysis of variance techniques, it is practical to produce numerical indicators that enable this to be quantified. These methods rely on judicious decompositions into sums of squares. Let Y^* be the centered observed response vector and $\bar{Y}$ the observed mean response. Observe that in the event when 1_n represents the order n indication, meaning the vector of length n with all of its components equal to 1, then [11]:

$$\begin{aligned} \bar{Y} &= \frac{1}{n}1_nY^t \\ Y^* &= Y - \bar{Y}1_n \end{aligned}$$

At this point, we define the three classical sums "Sum of Squares (SS)" that follow :

Total Sum of Squares :	$\sum_{i=1}^{n}(Y_i - \bar{Y})^2$ (noted SST).
Sum of Regression Squares :	$\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$ (noted SSR).
Sum of Error Squares :	$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$ (noted SSE).

Proposition 1.2 : For the least squares model, if $P = X(X^tX)^{-1}X^t$ is the orthogonal projector onto $\mathbb{R}^n$ over the subspace $Im(X)$, and if $I_n \subset Im(X)$, then the above sum of squares can be written as:

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = Y^tY - n\bar{Y}^2.$$
$$\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = Y^tPY - n\bar{Y}^2.$$
$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = Y^t(I_n - P)Y.$$

As a result, the following essential relationship results:

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2.$$

Proof : Using a matrix form, what follows can get written as:

- For $\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = (Y - \bar{Y}1_n)(Y - \bar{Y}1_n)^t = YY^t - \bar{Y}Y^t1_n - \bar{Y}1_nY^t + \bar{Y}^21_n1_n^t$, thus $1_nY^t = Y^t1_n = n\bar{Y}$ and $1_n1_n^t = n$, so $\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = Y^tY - n\bar{Y}^2$.

- For $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = (Y - \hat{Y})(Y - \hat{Y})^t$ with $\hat{Y} = X(X^tX)^{-1}XY^t = PY$, so $Y - \hat{Y} = Y - PY = Y(I_n - P)$, and we get $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = (Y - \hat{Y})(Y - \hat{Y})^t = Y^t(I_n - P)(I_n - P)Y = Y^t(I_n - P)Y$

- For $\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = Y^tPY - n\bar{Y}^2$, we have $\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = \hat{Y}^t\hat{Y} - 2\bar{Y}\hat{Y}^tI_n + \bar{Y}^2I_n^tI_n$. Alternatively, for the least squares model: $X^tY = X^tX\hat{A}$. Consequently: $X^tY = X^tX\hat{A} = X^t\hat{Y}$, then we multiply the two parts by $\frac{1}{n}1_n$, we get : $X^t\frac{1}{n}YI_n = X^t\frac{1}{n}\hat{Y}I_n \implies X^t\bar{Y} = X^t\bar{\hat{Y}}$. Therefore, $\bar{Y} = \bar{\hat{Y}}$. Where : $\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = \hat{Y}^t\hat{Y} - 2\bar{Y}\hat{Y}^t1_n + \bar{Y}^21_n^t1_n = \hat{Y}^t\hat{Y} - n\bar{Y}^2 = Y^tPY - n\bar{Y}^2$. This effectively develops as:

  $$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

  Cause, $\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = Y^tY - n\bar{Y}^2 + Y^t(I_n - P)Y = Y^tPY - n\bar{Y}^2$.

Calling the degrees of freedom of Y^tMY the rank of the matrix M, where Y is a random vector of $\mathbb{R}^n$ and $M(n,n)$ is a non-random matrix. Typically, a chi-square distribution is used to represent degrees of freedom. Indeed, it's shown that if y is a random vector with Gaussian distribution $N(\mu, \delta^2I_n)$ and if M is the matrix of a projector, then Y^tMY follows a non-central chi-square distribution, with non-centrality parameter $1/2\mu^tA\mu$ and degrees of freedom equal to the rank of M [12]. Below are the various sums of squares' corresponding degrees of freedom.

Proposition 1.3: The sums of squares SST, SSE, and SSR are associated, respectively, with degrees of freedom (n-1), (n-p), and (p-1). This allows us to define the

following mean square sums:

$$MST = \frac{\sum_{i=1}^{n} YY_i^2 - n\bar{Y}^2}{n-1}.$$
$$MSE = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2}{n-p}.$$
$$MSR = \frac{\sum_{i=1}^{n} \hat{Y}_i^2 - n\bar{Y}^2}{p-1}.$$

(With the notation MS denoting the mean square sum).

Proof : According to the previous proposition, we have:
$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = Y^tY - n\bar{Y}^2 = Y^t(I_n - \frac{1}{n}1_n1_n^t)Y$ and since $\frac{1}{n}1_n1_n^t = 1_n(1_n^t1_n)^{-1}1_n^t$ is the orthogonal projector onto ImX of $\mathbb{R}^n$, so $(I_n - \frac{1}{n}1_n1_n^t)$ is the orthogonal projector onto the subspace $(ImX)^{\perp}$ of $\mathbb{R}^n$. Hence the matrix $(I_n - \frac{1}{n}1_n1_n^t)$ for the renk $n-1$ cause the dimension of $(ImX)^{\perp}$ equal to $n-1$.

Similarly to preceding: $\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = Y^tPY - n\bar{Y}^2 = Y^t(P - \frac{1}{n}1_n1_n^t)Y$, if $1_n \subset ImX$ So, it is also found in the kernel of $P - \frac{1}{n}1_n1_n^t$ and the rank of $P - \frac{1}{n}1_n1_n^t$ is equal to $p-1$ [11].

2.2.3 Statistical Tests

The outcomes of the analysis of variances are used in these tests. Under specific conditions, they permit the computation of three often used statistics:

The multiple correlation coefficient R^2:
The ratio of the part of the results that the model can explain to the full quantity that it should explain is known as R^2 :

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

According to this formula, we can see that the R2 ratio varies between 0 and 1. The adjusted model is considered to be "close" to the observed responses as R^2 approaches 1. A classical approach consists of verifying the model according to $R^2 \geq 0.95$.

The Fisher test F :
The real Fisher coefficient is called:

$$F = \frac{\frac{SSR}{p-1}}{\frac{SSE}{n-p}}$$

A high Fisher's F-value suggests that the computed answers, or regression variance, is greater than the residual variance. In that case, there is an extremely little chance that any of the model's coefficients are zero. A low probability (p-value) and a high Fisher's F-value are necessary for significant coefficients. If $F > F_{1-\alpha,p,n-p-1}$, the model will be termed globally significant. It is also possible to state that a model has higher quality when the adjustment variance is smaller than the experimental variation [13].

2.2.4 Residual Analysis

It is the residuals represented graphically. Plotting the "$X-$" and "$Y-$" axes respectively shows the values of the responses that the model $\hat{Y}_i$ anticipated and the residuals

e_i. The appearance of random distribution in the residuals is investigated. If the points ($\hat{Y}_i$, e_i) seem to be arranged randomly, the model is becomes acceptable.

2.3 Error Transmission

Based on experimental error, the answers that the researcher measures are unpredictable variables. This error may be calculated from a sample and estimated by the population standard deviation. The factor levels and the mathematical model selected a priori aren't considered as random variables since it's expected that they don't add any errors. Consequently, there's no mistake introduced by the matrix X and it's not a random variable. Only the matrix Y is a matrix whose elements are random variables. Because of a connection, changes in any one of the matrix's members might cause variances or errors in the mathematical model's coefficients.

$$\hat{A} = (X^t X)^{-1} X^t Y$$

Random variables make up the coefficients that the least squares technique determines. Consequently, when included into a mathematical relationship, they produce additional random variables. For example, because of the relation, the expected replies are random variables:

$$\hat{Y} = X\hat{A}$$

According to this brief analysis, the position of experimental points inside the research domain, the degree of errors in the measured responses, and the model selected a priori will all influence the errors in the model coefficients and predicted responses. The connections that exist between these various faults will now be thoroughly examined.

2.3.1 Precision on Coefficients

The primary diagonal of $V(\hat{A})$ in equation (2.2) indicates where the coefficient variances are placed. For our proposed use, the above formula becomes:

$$DiagV(\hat{A}) = \sigma_r^2 Diag(X^t X)^{-1} \tag{2.3}$$

This variance connection, which highlights the existence of three separate components causing mistakes in the coefficients, is quite important. These elements consist of :

- The inaccuracy in the replies that resulted from using the initial σ_r^2,
- Where the experimental points are located,
- The mathematical model that was selected beforehand(a priori).

2.3.2 Precision on Predicted Responses

Computed coefficients make it easier to compute replies for the whole research domain when using the least squares approach. When no experiments have been done in the research domain, this suggests the capacity to estimate responses for all points in it.

In the research domain, the expected response at a point p, which is described by its coordinates, is provided by [9]:

$$\hat{Y}_p = X_p \hat{A}$$

Assuming that p represents an experimental point in the experimental design, a row of the matrix X corresponds to the matrix-vector X_p. The row vector X_p is the modeled vector of point p, and the matrix of coefficients obtained by the least squares approach is A. The estimated response is impacted by the coefficients' uncertainty:

$$v(\hat{Y}_p) = var(X_p \hat{A})$$

Assumed by hypothesis, the coordinates of the experimental points are precisely known and don't lead to errors. The row matrix X_p in this equation depends on the coordinates of a point in the research domain. Consequently, we are able to remove the brackets from the modeled vector of point p :

$$var(\hat{Y}_p) = X_p^t var(\hat{A}) X_p^t$$

The variance of $\hat{A}$ in this formula is known, and it is determined by:

$$V(\hat{A}) = \sigma_r^2 (X^t X)^{-1}$$

The variance of the calculating response for p points is :

$$var(\hat{Y}_p) = X_p^t \sigma^2 (X^t X)^{-1} X_p$$

This relationship allows for the calculation of the uncertainty in the calculated response at point p. It is observed that this error in the calculated response (or predicted response) depends on four quantities :

- The experimental error in the measured responses,
- Point p's location inside the research region,
- The set of points used to establish the coefficients of the model that is, the actual experimental design,
- The hypothesis that was selected to clarify the results (through the use of the residual variance and the matrix for calculating coefficients).

2.3.3 Prediction Variance Function

The type of experimentation, the precision of the technology used, the experimenter's expertise and care, and several other factors are all under the experimenter's control and determine the error produced on the measured answers. These variables are dependent on actual experimental practice rather than theory on experimental designs. In order to separate this experimental portion from the theoretical portion, the prediction variance function $d^2(\hat{Y}_p)$ is given by [9] :

$$d^2(\hat{Y}_p) = X_p^t (X^t X)^{-1} X_p$$

The prediction error function may be obtained by taking the variance function's square root:

$$d(\hat{y}_p) = [X_p^t (X^t X)^{-1} X_p]^{\frac{1}{2}}$$

In the research region, we can plot curves with identical prediction errors. Curiously, the prediction error function is independent of the experiment results, that is, the measured response values. Fundamentally, it is dependent on the postulated model and the locations of the experimental points inside the research region. As a result, we can determine, before starting the experiment, how the model and experimental point locations will impact the precision of the predicted answers. The link between them may be shown as follows by presenting the prediction variance function and standard deviations:

$$d(\hat{Y}_p) = [\sigma_r^2 X_p^t (X^t X)^{-1} X_p]^{\frac{1}{2}}$$

The anticipated replies are more accurate than the measured responses if the prediction error function is less than 1. If the prediction error function is greater than 1, it is the opposite. In general, we make sure that the experimental error is not more than the forecast error. As a result, predictions in areas where the prediction error function is larger than unity are avoided.

2.4 Optimality Criteria

Quantitative criteria must be used to assess the quality of the structure of a set of points extracted from a database or an experimental design. The quality of an experimental design may be assessed using a variety of metrics, among other things.

According to the chosen criteria, the placement of experimental points may vary from one design to another. There are several optimality criteria, some of which focus on the distribution of variance within the study domain.One such criteria is the criterion of isovariance by rotation [14]. The goal of optimality criteria is to produce a mathematical model of superior quality. The precision of the model coefficients is the main emphasis of these requirements.

2.4.1 Quality of Coefficient Representation

According to several authors, the limit of the confidence domain of the coefficients is given by the relationship [14]:

$$(A - \hat{A})(X^t X)(A - \hat{A}) = qs^2 F_\alpha(p, v) \tag{2.4}$$

With :

- s^2 it's an estimation of the experimental variance,
- F is the statistic corresponding to the Fisher's test,
- v is the chosen confidence level.

The calculated vector $\hat{A}$ serves as the center of this domain, which is a hyperellipsoid in the coefficient space. We may conclude that this hyperellipsoid contains the real values of the coefficients with a risk alpha. The hyperellipsoid is graphically shown for two coefficients as:

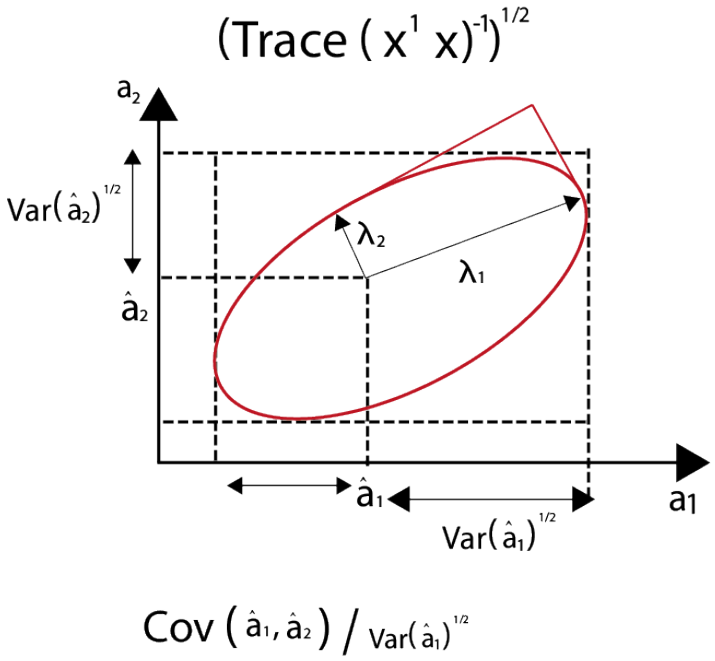

Figure 2.2: confidence ellipsoid

The two coefficients identified via multilinear regression are $(\hat{a}_1, \hat{a}_2)$ and the two ellipsoid eigenvalues are (λ_1, λ_2). The performance of all observations for a specific model is represented by this hyperellipsoid. Three descriptions apply to it:

- **Volume :** $\pi(\lambda_1\lambda_2)^{\frac{1}{2}} = \pi(\det(X^tX)^{-1})^{\frac{1}{2}}$ is the ellipsoid's volume. The determinant of the dispersion matrix $(X^tX)^{-1}$ is linked to this volume. We approach the real solution of A by decreasing the volume, which causes the hyperellipsoid to gravitate more towards a point.

- **Form :** When the shape is very elongated, there is a large disparity in the precision of the coefficients.
  The ellipsoid becomes a hypersphere when $(\lambda_1 = \lambda_2)$, implying that all coefficients are determined with the same precision.

- **Orientation :** A coefficient's calculated value is independent of the other coefficients' calculated values if the hyperellipsoid's primary axes are parallel to the coefficients' axes.

  According to the formula (2.4), the hyperellipsoid depends on the dispersion matrix $(X^tX)^{-1}$, so the criteria are based on the matrices (X^tX) and $(X^tX)^{-1}$.

2.4.2 Criterion A

Minimize the sum, or average, of the variances of the parameter estimates.

$$min \sum_{i=1}^{p} \frac{1}{\lambda_i};$$

2.4.3 Criterion D

A design is D-optimum if it maximizes the value of $|X^T X|$, that is the generalized variance of the parameter estimates is minimized.

$$min \prod_{i=1}^{p} \frac{1}{\lambda_i};$$

2.4.4 Criterion E

Minimize the variance of the least well-estimated linear combination $A^T \hat{\beta}$ with $A^T A = 1$

$$\min \max_i \frac{1}{\lambda_i};$$

2.4.5 Criterion G

A G-optimum design minimizes the maximum over the design region X of the standardized variance $d(x, \xi)$. For some designs this maximum value equals:

$$\bar{d}(\xi) = \max_{x \in X} d(x, \xi).$$

2.4.6 Criterion M

Using the M-criterion allows consideration of the experiment's quality of results.The quantity of trials in the experimental design has impact on this criteria. The moment matrix, is defined by :

$$M = \frac{(X^t X)}{N}$$

Two experimental design matrices, which might not contain the same number of experiments, can be compared. In terms of the M-criterion, we will agree that the first plan is more effective than the second if: $|M_1| > |M_2|$ and we have two moment matrices associated with two experimental design matrices composed respectively of N_1 and N_2 experiments:

$$M_1 = \frac{(X_1^t X_1)}{N_1}$$
$$M_2 = \frac{(X_2^t X_2)}{N_2}$$

2.4.7 Criterion of Orthogonality

Once estimates of independent coefficients can be obtained, an experimental design matrix is considered orthogonal. Parallel to the coefficients' axes are ellipsoid axes that define this. A diagonal $X^t X$, or one in which the coefficient covariances are zero as well [15], is required to satisfy this condition.

2.4.8 Criterion of Near Orthogonality

The nearly orthogonality criteria is obtained if the submatrix created by eliminating the first row and first column of the matrix $(X^tX)^{-1}$ is diagonal.

2.4.9 Criterion of Isovariation by Rotation

It is desired that, for sites situated at equal distances from the research domain's center, the responses computed using the model obtained from the experimental design have the same prediction error. This type of design is known as a rotationally invariant (rotable) design.

Chapter 3

Screening Designs

Screening experimental designs are essential tools in research and development for identifying the most influential factors among a large number of variables. They allow experimental efforts to be concentrated on the most significant variables, thereby reducing the time and costs associated with testing. This chapter explores different screening methods, focusing on their design, application, and effectiveness. We will present the theoretical foundations of these designs, followed by practical examples, and discuss the advantages and disadvantages of each experimental design.

3.1 Full Factorial Designs at Two Levels

The most simple factorial design consists of two levels. There are usually only two levels (a lower and an upper value) for each element in these designs. Both discrete and continuous variables can be employed randomly with them [16].

3.1.1 Construction of Full Factorial Designs at Two Levels

Trials matrix

A factor has to be given at least two levels in order to be studied in relation to a response. To study two factors 1 and 2, each with two levels, we need to perform $2^2 = 4$ trials. The operating conditions for each trial are described in Table 3.1:

Table 3.1: Experiences matrix of 2^2 design

Trials N°	Factor 1	Factor 2
1	-1	-1
2	1	-1
3	-1	1
4	1	1

The number of trials that need to be done for three factors is $2^3 = 8$ trials. Table 3.2 specifies the trial or experiment matrix.

Table 3.2: Experiences matrix of 2^3 design

Trial N°	Factor 1	Factor 2	Factor 3
1	-1	-1	-1
2	1	-1	-1
3	-1	1	-1
4	1	1	-1
5	-1	-1	1
6	1	-1	1
7	-1	1	1
8	1	1	1

This design family is called "2^k design" as for k factors, the number of trials is 2^k. The trial matrix involves 2^k rows and k columns. It's made simply by:

- The 1^{st} factor's column switching between -1 and +1.
- The 2^{nd} factor's column every 2 rows,switching between -1 and +1.
- The 3^{rd} factor's column every 4 rows,switching between -1 and +1.
- The 4^{th} factor's column every 8 rows,switching between -1 and +1.

And so on for a higher number of factors.

Experimental field

The experimental space of factors with k dimensions can be used to describe the study domain. A square is obtained when $k = 2$. The four segments of the square represent the position of the experimental points(fig 3.1).

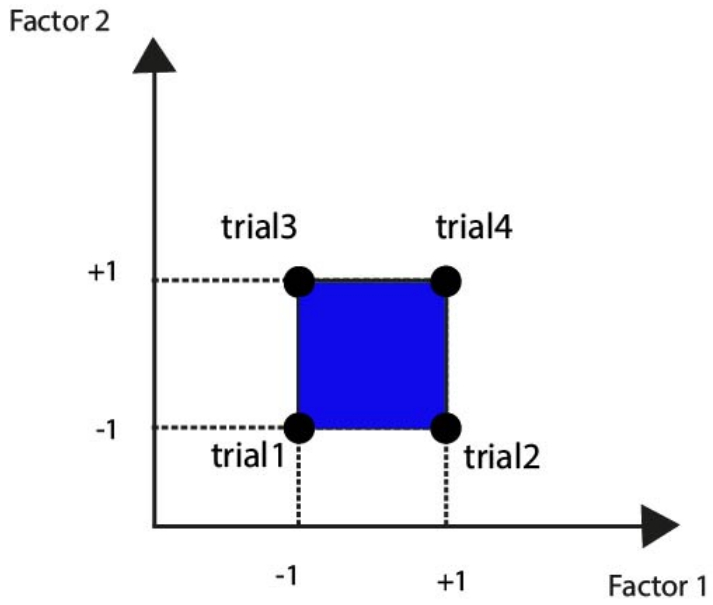

Figure 3.1: 2^2 design's study field

With eight segments representing the eight design experiments, a cube shows the experimental field for $k = 3$ (fig 3.2).

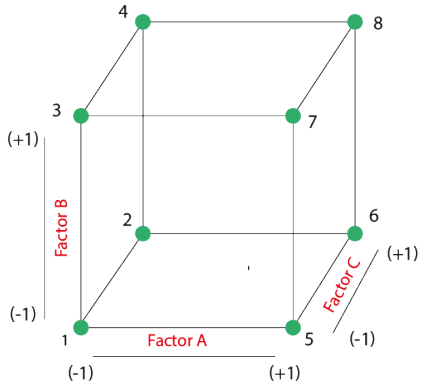

Figure 3.2: 2^3 factorial design's experimental field

The experimental points remain at the segments of a k-dimensional hypercube, even if the geometric representation of the field is no longer possible when $k > 3$.

Analysis and interpretation

We found it easier to demonstrate the concepts with a 2^2 factorial design as the most basic example, and to progressively increase the number of factors until we reached the 2^k design.

3.1.2 Two-Factor Factorial Design

The field of study is a square for two factors. An example of a fully realized two-factor factorial design is shown in Figure (3.1). In relation to each factor, the proposed mathematical model is a first-degree model:

$$Y = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + \hat{a}_{12} x_1 x_2 \tag{3.1}$$

with:

- Y is the response.
- x_i represents the level assigned to factor i.
- $\hat{a}_1$(resp $\hat{a}_2$)the principal effect of factor 1 (resp) 2.
- $\hat{a}_{12}$ interaction between factors 1 and 2.

Factor effect : The tester who conducted the trials has four values for the response: y_1,y_2,y_3 and y_4. Therefore,we have a system of four equations with four unknowns. The unknowns are the coefficients of the model: $\hat{a}_0$,$\hat{a}_1$,$\hat{a}_2$,$\hat{a}_{12}$. By replacing the values of x_i into the ralation we had in (3.1), we obtain:

$$\begin{aligned}
\hat{a}_0 &= y_1 + y_2 + y_3 + y_4 \\
\hat{a}_1 &= -y_1 + y_2 - y_3 + y_4 \\
\hat{a}_2 &= -y_1 - y_2 + y_3 + y_4 \\
\hat{a}_{12} &= y_1 - y_2 - y_3 + y_4
\end{aligned}$$

The resolution of that system gives us:

$$\hat{a}_0 = \frac{1}{4}(y_1 + y_2 + y_3 + y_4) \tag{3.2}$$

$$\hat{a}_1 = \frac{1}{4}(-y_1 + y_2 - y_3 + y_4) \tag{3.3}$$

$$\hat{a}_2 = \frac{1}{4}(-y_1 - y_2 + y_3 + y_4) \tag{3.4}$$

$$\hat{a}_{12} = \frac{1}{4}(y_1 - y_2 - y_3 + y_4) \tag{3.5}$$

Meaning of a_0
The center of the study field is defined if we set both x_1 and x_2 to zero. After then, the relation (3.1) becomes:

$$y = \hat{a}_0$$

The response value in the center of the research domain is represented by the coefficient a_0. Equation (3.2) also demonstrates that the mean of the four replies could be considered as a_1.

Meaning of a_1
Let us now situate ourselves at the midpoint of factor 2. Let's set x_2's value to zero in order to do this. The relation (3.3) turns into:

$$y = \hat{a}_0 + \hat{a}_1 x_1$$

Plotting the development of the anticipated response on a slice plane with $x_2 = 0$ is made possible by this relationship (Figure 3.3). When factor 1 increases from a low to a high level, the response varies, showing the influence of factor 1.

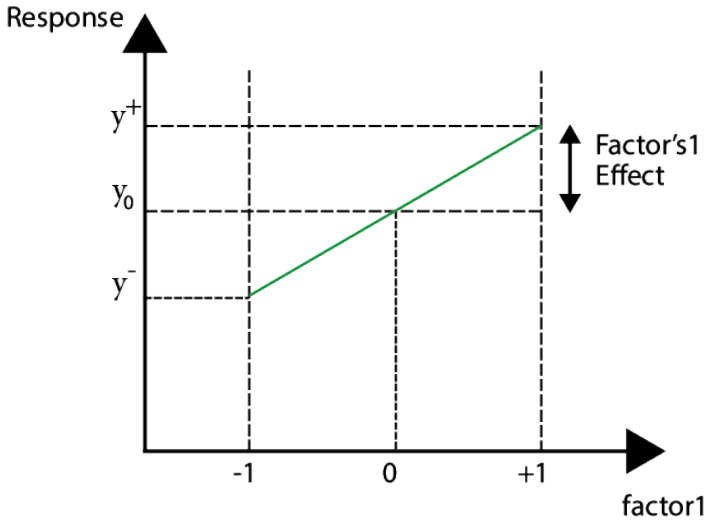

Figure 3.3: Factor's 1 effect

The mean response value at the high (or low) level of factor 1 is denoted by y_+ (or y_-).

Meaning of a_{12}
The relationship (3.5) can be written as:

$$y = \frac{1}{2}[\frac{1}{2}(y_4 - y_3) - \frac{1}{2}(y_2 - y_1)] = \frac{1}{2}(\grave{y}_+ - \grave{y}_-)$$

Between the impact of factor 1 at the high level of factor 2 (effect indicated $\grave{y}_+$) and the effect of factor 1 at the low level of factor 2 (effect noted $\grave{y}_-$), the interaction appears to correspond to the half of the difference. It shows how the degree of one factor affects the other, reflecting variations in the effects of both.

Calcul of factor effect: Let's go back to the formula (3.3) that gave the effect of factor 1 :

$$\hat{a}_1 = \frac{1}{4}(-y_1 + y_2 - y_3 + y_4)$$

We not that:

-All answers are involved to the calculation of the effect.

-Each response is preceded by a sign, and that the following one are the same as the column for Factor 1 in the experimental matrix (table 3.1).

-There's a coefficient here $\frac{1}{4}$,which the denominator is equal to the number of trials made.

-We multiply each response by the corresponding sign from the factor's column,we add the products, and then divide the sum by the number of trials.

Effect matrix: As we just saw, the effects may be calculated via to the signs in the experimental matrix. But you would also need to figure out the mean and the interaction.

Mean calcul : The calculation process adopte for the effects can be applied using a column of positive signs (+) because it's the only sign in formula (3.2).

Interaction calcul : The following signs of relation (3.5) are $+ - -+$. Each sign comes from the product of $x_1.x_2$ showing in relation (3.1). This sequence of signs can be found as follows : write in column the signs belongs to x_1 and x_2 then we product the scalar of the corresponding elements in the columns of the factors.

$$\begin{pmatrix} -1 & -1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix}$$

With the experience matrix in hand, it is easy to construct the effects matrix (Table 3.3) by adding a column with a plus sign for the mean and calculating that of the interaction as previously done.

Table 3.3: Matrix effects for a 2^2 design

Trials N°	Mean	Factor 1	Factor 2	Interaction 12
1	1	-1	-1	1
2	1	1	-1	-1
3	1	-1	1	-1
4	1	1	1	1

Box notation: Factor 1 (resp 2) signals will be represented by the column that is defined 1 (resp 2).The sequence of signs for the interaction between components 1 and 2 may be obtained by multiplying these two columns using the method described in the preceding paragraph. Therefore, we define multiplication for an algebra of sign columns, and we introduce it by expressing the column of signs for this interaction as 12. If we multiply a column of signs by itself, we obtain a column containing only plus (+) signs. This column will be marked as I. Leading to :

$$1.1=\text{I and for the same } 2.2=\text{I}.$$

3.1.3 Three-Factor Factorial Design

An example:

For road network maintenance, oil companies are required to prepare bitumen emulsions. These emulsions must remain stable from their production to their placement. The study we present aims to investigate the stability conditions of bitumen emulsion based on its composition.The manager has identified three factors [16]:

Factor 1 : Low and high concentration of fatty acid.

Factor 2 : Highly and slightly diluted hydrochloric acid.

Factor 3 : Nature of bitumen A and B.

On the stability of a bitumen emulsion. The implemented plan is a 2^3, 3 factors with 2 levels per factor. This plan totals $2^3 = 8$ experiments. Figure (3.2) provides the geometric representation of the experimental design. To calculate the effects of each factor, we construct the effect calculation matrix (Table 3.4).

Table 3.4: effect matrix of 2^3 design

Trial N°	Mean	1	2	3	12	13	23	123	responses
1	1	-1	-1	-1	1	1	1	-1	38
2	1	1	-1	-1	-1	-1	1	1	37
3	1	-1	1	-1	-1	1	-1	1	26
4	1	1	1	-1	1	-1	-1	-1	24
5	1	-1	-1	1	1	-1	-1	1	30
6	1	1	-1	1	-1	1	-1	-1	28
7	1	-1	1	1	-1	-1	1	-1	19
8	1	1	1	1	1	1	1	1	16
Effects	27,25	-1	-6	-4	0,25	0,25	0,25	0	

Only effects 2 and 3 are significant.

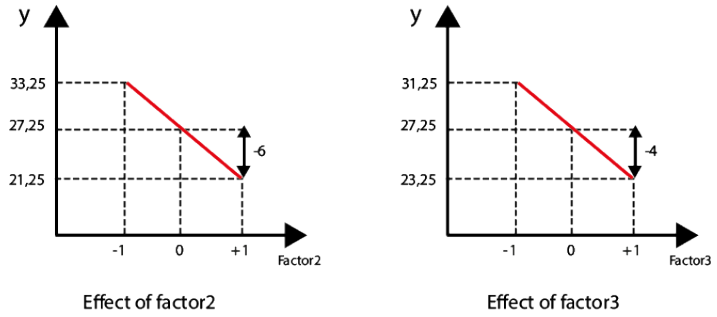

Figure 3.4: effect of factor 2, factor 3 on the stability of a bitumen emulsion

Study conclusions:

The concentration of fatty acid has practically no influence on the stability of the emulsion. On the other hand, the dilution of hydrochloric acid is an important factor with a negative effect. The nature of the bitumen is also important; the best stability will be obtained with bitumen B, with no significant interaction.

Note :

When the response decreases as the matching factor moves from level -1 to level +1, the effect is negative.

3.1.4 Four-Factor Full Factorial Design

The prior method may be applied more broadly when there are four, five, or more elements. The full factorial design for four elements has $2^4 = 16$ trials. Sixteen pieces of information are included in the matrix of effects in terms of:

-the effects of each of the 4 factors.

-the interactions of the second order 12, 13,. . . , etc., the number that is marked by $C_4^2 = 6$.

-the third-order interactions,123,124,. . . ,etc.the number that is marked by $C_4^3 = 4$.

-The fourth-order interaction 1234, which is unique.

-The mean, which is unique.

3.1.5 2^k Full Factorial Design

According to these designs, there are k factors, two levels for each. The polynomial used as the mathematical model accounts for the mean, the effects of each factor, and the interactions between factors are taken two by two.

$$Y = \hat{a}_0 + \sum \hat{a}_i x_i + \sum \hat{a}_{ij} x_i x_j \tag{3.6}$$

In matrix form, we replace the x_i by their values in centered reduced coordinates :

$$Y = X\hat{A} \tag{3.7}$$

Reviewing the 2^2 plan's system, equation (3.7) looks like this:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \end{pmatrix}$$

The Hadamard matrix, or matrix X, is orthogonal when it relates to two-level factorial designs. One of its essential features is that $X^tX = NI$, where N is the number of trials and I is the identity matrix. By replacing this equation in formula (2.1), we get:

$$\hat{A} = (X^tX)^{-1}X^tY = (NI)^{-1}X^tY$$

With :

$$\hat{A} = \frac{1}{N}X^tY$$

3.1.6 Optimality of 2^k Designs

A minimum variance $var(Y)/N$, is estimated for each effect and interaction independently.

3.1.7 Advantages and Disadvantages of Full Factorial Designs at Two Levels

- There are lots of advantages to factorial designs [17]; we will focus on the essential ones here:

  - Constructing full factorial designs is simple.

  - There are less risks of error and the trials are simple to manage because each factor only requires two levels.

  - Effects and interactions are fairly easy to calculate.

  - Any experimenter may interpret the results; an advanced knowledge of statistics isn't necessary.

  - Partially or fully, the results from an original design might be applied for studying a different section of the experimental field or to develop a higher-order mathematical model.

- Having to run a large number of trials rapidly becomes a disadvantage of these systems. It is also not possible to create models of second degree and higher while only primary effects and interactions can be calculated.

3.2 Fractional Factorial Designs at Two Levels

As we just saw, the number of experiments rises quickly as the number of factors evaluated grows when using complete factorial designs [18]. This is the reason why fractional factorial designs are preferred by the testers. Less trials are carried out at the risk of information loss, which is often not frightening.

3.2.1 Definition of Fractional Factorial Designs

It is usually observed in practice that interactions with orders of three or higher are frequently negligible. The full factorial design therefore proposes $2^5 = 32$ tests for five factors, while we are only interested in 16 information points.

- The unique mean.
- The effects of the factors.
- The second-order interactions, of which there are $C_5^2 = 10$.

It is unfortunate that 32 trials must be conducted in order to obtain just 16 pieces of information. Fractional factorial designs suggest using the effect matrix of full factorial designs 2^{k-1} , 2^{k-2} or 2^{k-p} for the analysis of k factors. The advantage of these designs is clear: the experimental load is divided by 2^p, as $2^{k-p} = \frac{2^k}{2^p}$ [6].

3.2.2 Theory of Aliases

Turning back to the previous example. The experimenter conducted eight trials. Suppose that the experimenter does $2^{3-1} = 2^2 = 4$ experiments, or just half of the total. Of course, the responses are the same as in the preceding case, and Table 3.5 shows the experimental design matrix.

Table 3.5: experiences matrix (Study of a bitumen emulsion)

Trial N°	1	2	12	Y
1	-1	-1	1	$y_1 = 30$
2	1	-1	-1	$y_2 = 37$
3	-1	1	-1	$y_3 = 26$
4	1	1	1	$y_4 = 16$

We can compare the effects calculated from this fractional design with those from the full design (table 3.6).

Table 3.6: Comparison of the calculated effects

Effects	Full design	Fractional design
Mean	27,25	27,25
1	-1	0,75
2	-6	-6,25
3	-4	-5,25

The results obtained are identical to the ones that were attained using the whole eight-experiment design. Seems like we can get the same results with less work. Let's check the effect 3 value and interaction 12 that were calculated using the full design.
3^{rd} factor effect :

$$\hat{a}_3 = \frac{1}{8}(-y_1 - y_2 - y_3 - y_4 + y_5 + y_6 + y_7 + y_8)$$

Interaction between the first factor and the second:

$$\hat{a}_{12} = \frac{1}{8}(+y_1 - y_2 - y_3 + y_4 + y_5 - y_6 - y_7 + y_8)$$

Adding them up:

$$\hat{a}_3 + \hat{a}_{12} = \frac{1}{4}(-y_2 - y_3 - y_5 + y_8)$$

Then, we found the quantity :

$$c_3 = \hat{a}_3 + \hat{a}_{12} = -4,25$$

This means that c_3 is equal to the main effect a_3 increased by the interaction $\hat{a}_{12}$. We say that $\hat{a}_1$ and $\hat{a}_{12}$ are aliased. The quantity c_3 can be called an alias, contrast, or simply effect.

We also see that :

$$c_1 = \hat{a}_1 + \hat{a}_{23} = -0,75$$
$$c_2 = \hat{a}_2 + \hat{a}_{13} = -6,25$$

The model of the 2^{3-1} fractional design can be written as:

$$Y = c_0 + c_1 x_1 + c_2 x_2 + c_3 x_3$$

3.2.3 Hypotheses of Interpretation

1^{st} **Hypothesis**

Interactions of the third order or higher order are considered as negligible.

2^{nd} **Hypothesis**

A neutral contrast can indicate what follows:

- We will focus on the most probable case, which is that all of the aliased effects are null.

- That the aliased effects cancel each other out. This hypothesis is unlikely and is not retained.

3^{rd} **Hypothesis**

We shall be careful of an interaction between two strong impacts, as it may also be strong.

4^{th} **Hypothesis**

If two effects are small, we will assume that their interaction is also small.

3.2.4 Calculation of Contrasts

Consider the effects matrix of a 2^3 design, where trials were ordered to highlight two 2^2 designs for factors 1 and 2. The 2^3 design was divided into two half-plans (Table 3.7).

Table 3.7: Matrix for calculating contrasts for the two fractional half-plans

Trial N°	I	1	2	3	12	13	23	123
5	1	-1	-1	1	1	-1	-1	1
2	1	1	-1	-1	-1	1	1	1
3	1	-1	1	-1	-1	-1	-1	1
8	1	1	1	1	1	1	1	1
1	1	-1	-1	-1	1	1	1	-1
6	1	-1	1	1	-1	1	-1	-1
7	1	1	1	1	-1	-1	1	-1
4	1	1	-1	-1	1	-1	-1	-1

Consider the higher half plan. In Box notation, we can see that 3 and 12 are equivalent as they have the same sign sequence $+--+$. We could write:

$$3 = 12$$

However, we had previously demonstrated :

$$c_3 = \hat{a}_3 + \hat{a}_{12}$$

So, $3 = 12$ is equivalent to $c_3 = \hat{a}_3 + \hat{a}_{12}$. This equivalence relationship holds in both directions and forms the basis of aliasing theory. Similarly, we can show that:

$$1 = 23 \text{ equal to } c_1 = \hat{a}_1 + \hat{a}_{23}$$
$$2 = 13 \text{ equal to } c_2 = \hat{a}_2 + \hat{a}_{13}$$

In the higher half plan, we can find these relations. The two columns with a positive $(+)$ sign allow us to write:

$$I = 123$$

By successively multiplying this connection, known as the alias generator, by 1, 2, and 3:

$$1.I = 1.123 = 12231 = 23$$
$$2.I = 2.123 = 1.2232 = 13$$
$$3.I = 3.123 = 1.2.323 = 12$$

3.2.5 Practical Construction of a Fractional Design

Choosing a complete factorial design and writing its calculation matrix without including the plus $(+)$ sign column is the foundation of the practical contraction. The term "basic design" refers to this new grid. In this base, we designate an extra component to a column of signs that corresponds to an interaction. The indicators of the chosen interaction become the high and low research levels for this extra component.

We can generalize this method and use all the columns of a basic design. For example, to illustrate this construction of fractional designs, let's take a basic design of 2^3, the effects matrix includes 3 second-order interactions and one third-order interaction. The full factorial design allows studying three factors on columns 1, 2, and 3.

Table 3.8: Fractional Factorial Design for 3 factors

Trial N°	I	1	2	3	12	13	23	123
1	1	-1	-1	-1	1	1	1	-1
2	1	1	-1	-1	1	-1	1	1
3	1	-1	1	-1	-1	1	-1	1
4	1	1	1	-1	1	-1	-1	-1
5	1	-1	-1	1	1	-1	-1	1
6	1	1	-1	1	-1	1	-1	-1
7	1	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1	1

In order to examine four factors, we take the first three sign columns for the first three factors and select a column for the fourth component that represents an interaction. For clarification, let's take the interaction 12 column and write that the levels described by this interaction's signals are given to the extra factor 4:

$$4=12$$

Hence the alias generator:

$$4.4=4.12$$
$$I=124$$

Successively ,multiplying this generator by 1, 2, 3, and 4, we get:

$$1=24 \text{ is equal to } c_1 = \hat{a}_1 + \hat{a}_{24}$$
$$2=14 \text{ is equal to } c_2 = \hat{a}_2 + \hat{a}_{14}$$
$$4=12 \text{ is equal to } c_3 = \hat{a}_3 + \hat{a}_{12}$$
$$3=1234 \text{ is equal to } c_4 = \hat{a}_4 + \hat{a}_{1234}$$

We could aliased the 4^{th} factor in an other interaction and obtained other contrast values. It is absolutely possible to study two additional factors, 2^{5-2} design. We can selected as aliases :

$$4=12$$
$$5=13$$

hence, the independent aliase generators :

$$I=124$$
$$I=135$$

If we multuply those two generator, one by one .We obtain:

$$I.I=124.135$$
$$I=2345$$

Two additional factors introduce an alias group with four terms:

$$I=124=135=2345$$

Using this group, we are able to see how the fractionner design aliases the components and interactions in the calculat contract. For example to identify the constract c_1 , we multiply all the terms of groupe by the first column :

$$1.\text{I}=1.124=1.135=1.2345$$
$$1=24=135 \text{ is equal to } c_1 = \hat{a}_1 + \hat{a}_{24} + \hat{a}_{135} + \hat{a}_{2345}$$

In a based 2^4 design, the effects matrix has six second-order interactions, four third-order interactions, and one fourth-order interaction. The full design allows for the examination of four factors: 1, 2, 3, and 4. The full design has 256 effects and interactions to evaluate 8 factors, while the $2^{8-4} = 2^4$ design only allows for 16 contrasts. To investigate the other four factors, four interactions are required :

$$5 = 123$$
$$6 = 124$$
$$7 = 134$$
$$8 = 1234$$

The four independent alias generators are:

$$I = 1235 = 1246 = 1347 = 12348$$

The dependent generators are calculated from the independent generators by multiplying them pairwise, three at a time, and four at a time.

Multiplying pairwise :

$$1235.1246 = 3456$$
$$1235.1347 = 2457$$
$$1235.12348 = 458$$
$$1246.1347 = 2367$$
$$1246.12348 = 368$$
$$2347.12348 = 278$$

Multiplying three at a time :

$$1235.1246.1347 = 1567$$
$$1235.1246.12348 = 12568$$
$$1235.12347.12348 = 13576$$
$$1246.12348.1347 = 14678$$

Multiplying four at a time :

$$1235.1246.1347.12348 = 2345678$$

The group of alias generators will thus be:

$$I = 278 = 368 = 458 = 1235 = 1246 = 1347 = 3456 = 2457 = 2367$$
$$= 12348 = 1567 = 2568 = 13578 = 14678 = 234578.$$

Here is the complete calculation for factor 1:

$$c_1 = \hat{a}_1 + \hat{a}_{235} + \hat{a}_{246} + \hat{a}_{347} + \hat{a}_{567} + \hat{a}_{1278} + \hat{a}_{1368} + \hat{a}_{1458} + \hat{a}_{13456} + \hat{a}_{12457} + \hat{a}_{12367} + \hat{a}_{12568} + \hat{a}_{2348} + \hat{a}_{3578} + \hat{a}_{4678} + \hat{a}_{12345678}$$

Furthermore, if we neglect interactions of order more than three, we may write:

$$\begin{array}{llll} c_1 = \hat{a}_1, & c_5 = \hat{a}_5 + \hat{a}_{48}, & c_{12} = \hat{a}_{12} + \hat{a}_{35} + \hat{a}_{46}, & c_{24} = \hat{a}_{24} + \hat{a}_{16} + \hat{a}_{57}, \\ c_2 = \hat{a}_2 + \hat{a}_{78}, & c_6 = \hat{a}_6 + \hat{a}_{38}, & c_{13} = \hat{a}_{13} + \hat{a}_{25} + \hat{a}_{47}, & c_{34} = \hat{a}_{34} + \hat{a}_{17} + \hat{a}_{56}, \\ c_3 = \hat{a}_3 + \hat{a}_{68}, & c_7 = \hat{a}_7 + \hat{a}_{28}, & c_{14} = \hat{a}_{14} + \hat{a}_{26} + \hat{a}_{37}, & c_{234} = \hat{a}_{234} + \hat{a}_{18}, \\ c_4 = \hat{a}_4 + \hat{a}_{58}, & c_8 = \hat{a}_8 + \hat{a}_{45} + \hat{a}_{36} + \hat{a}_{27}, & c_{23} = \hat{a}_{23} + \hat{a}_{15} + \hat{a}_{67}, & \end{array}$$

3.2.6 Concept of Resolution

We discovered that in a fractional factorial design, primary effects are aliased by interactions with varying orders. A design's resolution $[X]$ equals to 1 plus the lowest-order interaction's order value. For example, if a major effect is aliased by interactions with orders 2, 3 and 4,the resolution is $1 + 2 = 3$. The resolution is frequently expressed in Roman numerals.

3.2.7 Adventages And Disadvantages

On two levels, the advantages of fractional factorial designs are identical to those of full factorial designs. But they also make it possible to get around the latter's primary disadvantage, which is a huge number of trials. The capacity to do the task in a sequential manner is a significant advantage. As an example, one can begin with a 2^{6-3} design and carry out additional tests by conducting a second 2^{6-3} design if more trials are required to settle ambiguities. Adding this to the initial design results in two 2^{6-3} designs, or $2 \cdot 2^{6-3}$, which can be more simply expressed as 2^{6-2}.

Given all of these advantages, there are just a few disadvantages. The most important consideration is that the experimenter make an effort to acquire the use of fractional factorial designs.

3.3 MOZZO's Designs

Two advantages of Mozzo's designs [8] are their sequentiality and the limited number of levels that can be studied. Twelve trials allow us to study four factors, and we can begin by examining two factors in three trials. To examine a third factor, we need only realize three additional trials (the 4^{th},5^{th} and 6^{th} trials from the table 3.9).

Not every design created by Mozzo permits the creation of a second-degree model. It is vital to use Mozzo's nested quadratic designs in these situations.

Table 3.9: Mozzo's design for 2, 3 and 4 factors

Trials N°	Factor 1	Factor 2	Factor 3	Factor 4
1	0,268	1	-1	-1
2	0,732	-0,732	-1	-1
3	-1	-0,268	-1	-1
4	-0,268	-1	1	-1
5	-0,732	0,732	1	-1
6	1	0,268	1	-1
7	-0,268	-1	-1	1
8	-0,732	0,732	-1	1
9	1	0,268	-1	1
10	0,268	1	1	1
11	0,732	-0,732	1	1
12	-1	-0,268	1	1

3.3.1 MOZZO's Design for Two Factors

It is a design for evaluating two factors in three trials. The experimental points build a triangle. Figure 3.5 shows one of the alternative arrangements for this triangle :

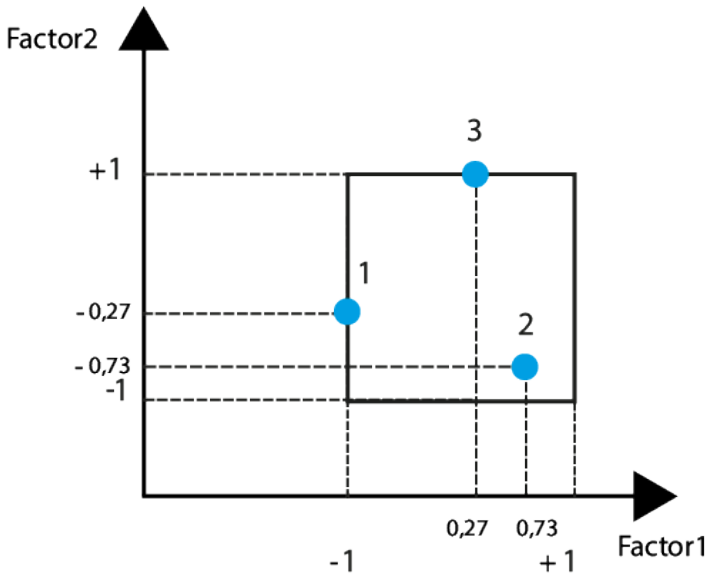

Figure 3.5: Study field of Mozzo's design for two factors

Given that there are just a few experimental points, the mathematical model is quite basic. It is a first-degree model that is without interactions.

$$Y = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2$$

By writting the X matrix

$$X = \begin{pmatrix} 1 & 0.268 & 1 \\ 1 & 0.732 & -0.732 \\ 1 & -1 & -0.268 \end{pmatrix}$$

Matrix X^tX is calculated immediately as

$$X^tX = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1.608 & 0 \\ 0 & 0 & 1.608 \end{pmatrix}$$

The matrix X is orthogonal. It can be observed that the elements of the first-degree terms are equal. As a result, this plan also satisfies the criterion of rotation isovariance.

3.3.2 MOZZO's Design for Three Factors

These are the first six experiments from Table 4.8. Figure 3.6 shows the arrangement of the experimental points in the experimental space.

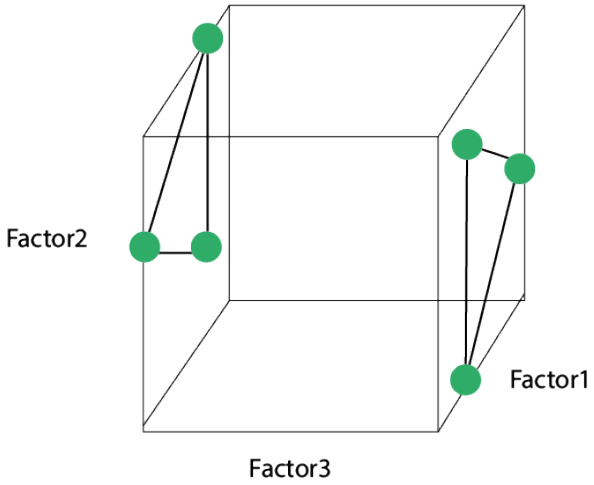

Figure 3.6: Study field of Mozzo's design for three factors

Considering there are six experimental points, we could theoretically determine six unknowns. As a result, we can construct a first-degree model with interactions. But because of the points' arrangement, we are unable to get interaction with the third element. We can only discuss the interaction between factors 1 and 2. Thus, the model is:

$$Y = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + \hat{a}_3 x_3 + \hat{a}_{12} x_1 x_2$$

Writting the X matrix

$$X = \begin{pmatrix} 1 & 0.268 & 1 & -1 & 0.268 \\ 1 & 0.732 & -0.732 & -1 & -0.536 \\ 1 & -1 & -0.268 & -1 & 0.268 \\ 1 & -0.268 & -1 & 1 & 0.268 \\ 1 & -0.732 & 0.732 & 1 & -0.536 \\ 1 & 1 & 0.268 & 1 & 0.268 \end{pmatrix}$$

We can calculate the matrix to see if orthogonality is still preserved.

$$X^tX = \begin{pmatrix} 6 & & & & \\ & 3.22 & & & \\ & & 3.22 & & \\ & & & 6 & \\ & & & & 0.86 \end{pmatrix}$$

3.3.3 Advantages and Disadvantages

- The main advantage of these designs is the very limited number of trials required. For two factors, only three trials need to be conducted, with each factor taking three levels.
- The disadvantages of Mozzo's design are: the design does not exist for any number of factors, and the selected model generally does not account for all interactions between factors.

Chapter 4

Exploration of Experimental design for response surface

This chapter will present the fundamental principles of response surface designs, illustrated with practical examples. We will discuss the advantages and disadvantages of these designs, as well as strategies for their effective implementation. By exploring these concepts, we will provide researchers and practitioners with powerful tools to design and optimize their experiments, ensuring robust and reliable results.

4.1 Full Factorial Designs at Three Levels

These designs allow the study of k factors with three levels for each factor. The designation for these plans is 3^k, where k represents the number of factors that we will study and three represents each factor's three levels. This notation also specifies the number of trials to be undertaken [19].

4.1.1 Construction of Full Factorial Designs at Three Levels

The three-level designs will all be defined as follows:

- The sequence of the first factor's numbers is -1, 0, +1;-1, 0, +1;...
- The second factor's numbers are -1,-1,-1, followed by 0, 0, 0 and +1, +1, +1,...,-1,-1,-1; 0, 0, 0; +1, +1, +1;...
- The third factor will have 9 levels equal to -1, followed by 9 equal to 0, then 9 equal to +1.
- The fourth factor will have 27 levels of -1, followed by 27 levels of 0, then 27 levels of +1.
- And so on ...

For two factors, the number of experiments needed is $3^2 = 9$ trials. The table 4.1includes details about the experience matrix.

Table 4.1: Experience Matrix for a 3^2 design

Trials N°	Factor 1	Factor 2
1	-1	-1
2	0	-1
3	1	-1
4	-1	0
5	0	0
6	1	0
7	-1	1
8	0	1
9	1	1

The figure 4.1 shows the study's field.

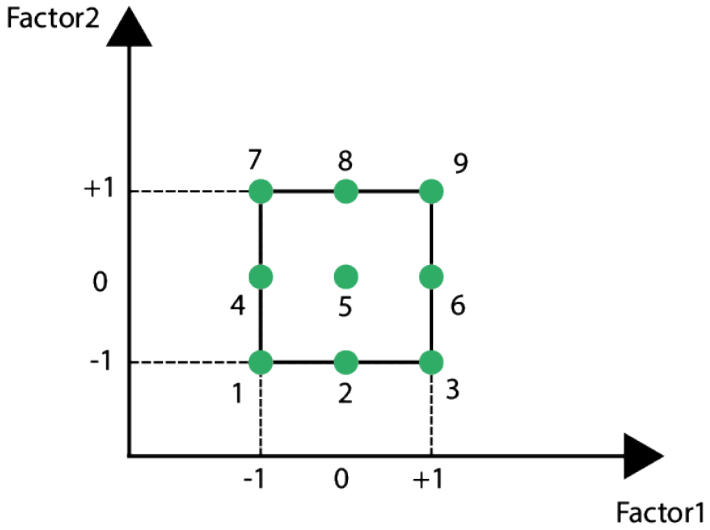

Figure 4.1: Study's field of 3^2 design

For 3 factors, we obtain a cube with twenty-seven trials as indicated in the figure 4.2.

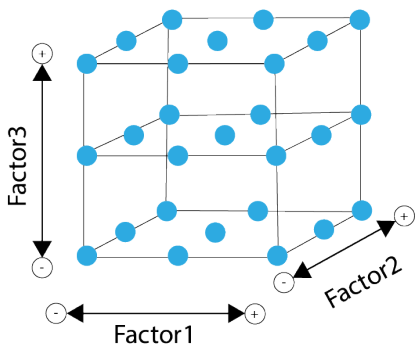

Figure 4.2: Study's field of 3^3 design

4.1.2 Advantages and Disadvantages

- Althought the model adopted by these designs is of second degree.The advantages of full factorial designs with three levels are the same as those of full factorial designs with two levels.
- The basic disadvantage of these designs is that as the number of factors rises, the design size might grow greatly.

4.2 Composite Designs

A composite design [20, 21] consists of three parts:

1. A factorial design where the factors take two levels.
2. At least, there should be an experimental point in the center of the study's regions.
3. Axial points. These experimental points are located on the axes of each of the factors.

using a composite design. The sum of the trials from the factorial design n_f, the star design n_α, and the center n_0 equals the total number of trials N that need to be done by :

$$N = n_f + n_\alpha + n_0$$

Figure 4.3 shows a composite design for two elements. Points A, B, C, and D are the experimental points in a 2^2-plan. Point E is the center point. This point may have been repeated one or more times. F, G, H, and I are the axial points. The remaining four points form what is termed as the star design.

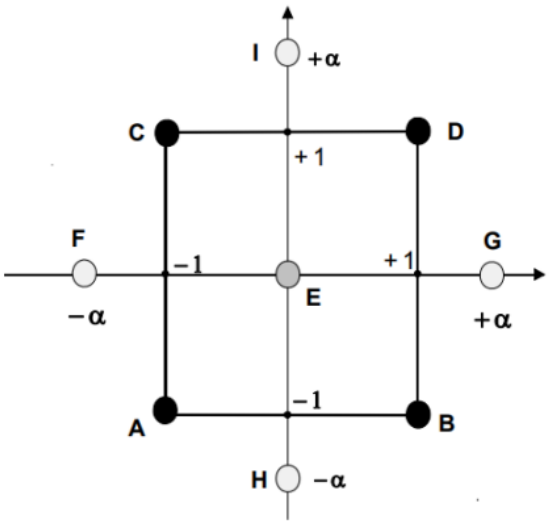

Figure 4.3: Composite design for two factors

We run nine trials, and six coefficients must be determined. Consequently, it is necessary to solve a system of nine equations with six unknowns.

Table 4.2 shows the experience matrix for this design. The letter α indicates the distance from the center of the field to the points in the star.

Table 4.2: Experiences matrix of a composite design for two factors

Trials N°	Factor 1	Factor 2
1	-1	-1
2	1	-1
3	-1	1
4	1	1
5	$-\alpha$	0
6	α	0
7	0	$-\alpha$
8	0	α
9	0	0

4.2.1 Properties of Composite Designs

Postulated mathematical model
A second-degree model with interactions is the mathematical model used with composite designs. Usually, only interactions of the second order are kept.

•For two factors :

$$Y = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + \hat{a}_{12} x_1 x_2 + \hat{a}_{11} x_1^2 + \hat{a}_{22} x_2^2$$

• For three factors :

$$Y = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + \hat{a}_3 x_3 + \hat{a}_{12} x_1 x_2 + \hat{a}_{13} x_1 x_3 + \hat{a}_{23} x_2 x_3 + \hat{a}_{11} x_1^2 + \hat{a}_{22} x_2^2 + \hat{a}_{33} x_3^2$$

The design matrix
The design matrix X for two factors is a matrix of dimension (11, 6), as the postulated model has 6 coefficients and 11 trials :

$$X = \begin{pmatrix} 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -\alpha & 0 & 0 & \alpha^2 & 0 \\ 1 & \alpha & 0 & 0 & \alpha^2 & 0 \\ 1 & 0 & -\alpha & 0 & 0 & \alpha^2 \\ 1 & 0 & \alpha & 0 & 0 & \alpha^2 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The information matrix
Using the formula $(X^tX)^{-1}$, the information matrix is calculated from the design matrix. For two factors, we have:

$$X^tX = \begin{pmatrix} N & 0 & 0 & 0 & 4+2\alpha^2 & 4+2\alpha^2 \\ 0 & 4+2\alpha^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4+2\alpha^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ 4+2\alpha^2 & 0 & 0 & 0 & 2+\alpha^2 & 4 \\ 4+2\alpha^2 & 0 & 0 & 0 & 4 & 4+2\alpha^2 \end{pmatrix}$$

The optimality criteria
The value of α will change depending on the chosen optimality criterion.

Isovariance by rotation
The information matrix's elements have to satisfy the relation [8]

$$\begin{aligned} & 3n_f = n_f + 2\alpha^4 \\ \Longrightarrow\ & 2n_f = 2\alpha^4 \\ \Longrightarrow\ & \alpha = (n_f)^{\frac{1}{4}} \end{aligned}$$

In the case of a composite design for two factors with star points located at a distance of 1.414 from the center, the isovariance by rotation requirement is in fact satisfied by this information matrix :

$$X^tX = \begin{pmatrix} 12 & & & & 8 & 8 \\ & 8 & & & & \\ & & 8 & & & \\ & & & 4 & & \\ 8 & & & & 12 & 4 \\ 8 & & & & 4 & 12 \end{pmatrix}$$

Nearly orthogonal
Instead of rotation isovariance, probably we should follow the orthogonality criterion. A diagonal X^t matrix would be produced by this point arrangement. We are unable to invalidate the elements that correspond to constant terms and squared terms, hence this is not achievable. Remember that the submatrix of $(X^tX)^{-1}$ that results from removing the first row and first column has to be diagonal. This is proven to be possible if [8]:

$$\alpha = \frac{n_f(\sqrt{N} - \sqrt{n_f})^2}{4}$$

For two factors, we have $\alpha = 1.21$. Let's write the corresponding information matrix :

$$X^tX = \begin{pmatrix} 12 & & & & 6.928 & 6.928 \\ & 6.928 & & & & \\ & & 6.928 & & & \\ & & & 4 & & \\ 6.928 & & & & 8.287 & 4 \\ 6.928 & & & & 4 & 8.287 \end{pmatrix}$$

The corresponding dispersion matrix is indeed nearly orthogonal since if we eliminate the first row and column, we obtain a diagonal matrix

$$(X^tX)^{-1} = \begin{pmatrix} 0.239 & & & & & -0.135 & -0.135 \\ & 0.144 & & & & & \\ & & 0.144 & & & & \\ & & & 0.25 & & & \\ -0.135 & & & & & 0.233 & \\ -0.135 & & & & & & 0.233 \end{pmatrix}$$

Standard deviation of the coefficients of the predictive model
The variance of each coefficient of the postulated model is given by (the formula 2.2). By taking the square roots of the diagonal elements of the variance-covariance matrix, we obtain the standard deviations of the coefficients.

$$\sigma(\hat{a}_i) = \sqrt{\sigma_r^2 Var(\hat{a}_i)} = \sigma_r \sqrt{Var(\hat{a}_i)}$$

4.2.2 Advantages and Disadvantages

A gradual and methodical approach becomes possible through composite designs. One can perform an initial factorial design. This design will indicate the impact of every factor and if the selected domain should be kept. The experimenter can conduct a supplemental design if, after the examination of this first factorial design, there are still difficulties. Verifying the validity of the first-degree model is important after defining the appropriate field and identifying the influential factors. The experimenter can stop if it is valid; if not, they have to move to a second-degree model. The additional experimental points from the composite design are useful in this case.

4.3 Doehlert Designs

Experimental space is consistently filled with the experimental points of designs suggested by David H. Doehlert in 1970 [22]. It's an alternative strategy to what we have discovered thus far [23]. Finding the ideal placement for the experimental points to suit the fitted model is not our goal. Rather, we just split the points in a regular manner, avoiding the optimality criterion.

4.3.1 Doehlert Designs for two factors

The experimental points are located at the borders of a regular hexagon, with an additional point at the center. (Figure 4.4) shows the arrangement of these points for a two-factor design (trials 1 to 7) :

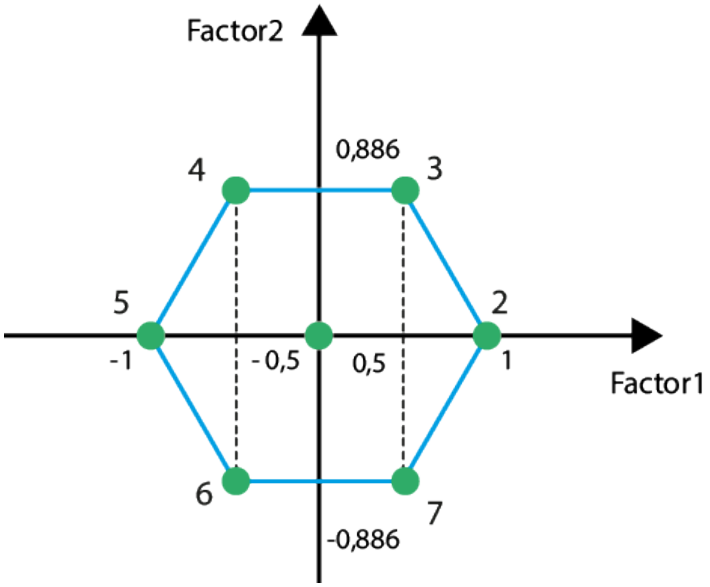

Figure 4.4: Doehlert Designs for two factors

The experimental design matrix is constructed by taking the coordinates of each experimental point (Table 4.3).

Table 4.3: Experiences matrix of Doehlert designs for two factors

Trial N°	Factor 1	Factor 2
1	0	0
2	1	0
3	0.5	0.866
4	-0.5	0.866
5	-1	0
6	-0.5	-0.866
7	0.5	-0.866

All points in the Doehlert design are on a circle with a unit radius. A circle for one factor, a sphere for three factors, and a hypersphere for more than three factors build up the spherical region that the Doehlert designs specify.

4.3.2 Postulated Mathematical Model

The postulated mathematical model for Doehlert designs is generally a second-degree model with second-order interactions. In terms of two factors, we have:

$$Y = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + \hat{a}_{12} x_1 x_2 + \hat{a}_{11} x_1^2 + \hat{a}_{22} x_2^2$$

4.3.3 Doehlert Designs for k factors

This type of design exists for any number of factors. The experimenter may simply add more experimental points to create a design that is exactly the same

as the first one, if they would like to investigate the experimental field longer. With three (or more) experimental points, a new Doehlert design can be created.

The Doehlert designs with a maximum of four factors are shown in a single (Table 4.4). Nothing stops us from realizing many central points, even though we have only pointed out one.

Table 4.4: Doehlert designs from two to four factors

Trial N°	Factor 1	Factor 2	Factor 3	Factor 4
1	0	0	0	0
2	1	0	0	0
3	0.5	0.866	0	0
4	-0.5	0.866	0	0
5	-1	0	0	0
6	-0.5	-0.866	0	0
7	0.5	-0.866	0	0
8	0.5	0.289	0.816	0
9	-0.5	0.289	0.816	0
10	0	-0.577	0.816	0
11	0.5	-0.289	-0.816	0
12	-0.5	-0.289	-0.816	0
13	0	0.577	-0.816	0
14	0.5	0.289	0.204	0.791
15	-0.5	0.289	0.204	0.791
16	0	-0.577	0.204	0.791
17	0	0	-0.612	0.791
18	0.5	-0.289	-0.204	-0.971
19	-0.5	-0.289	-0.204	-0.971
20	0	0.577	-0.204	-0.971
21	0	0	0.612	-0.971
22	0	0	0	0

4.3.4 Advantages and Disadvantages

Advantages

- Permit a limited number of trials to efficiently explore the response in a multidimensional space.
- Ideal for researching factor interactions and nonlinear responses.
- Offer great flexibility in choosing factor levels and the distribution of experimental points.
- Make it possible for second-order interactions in second-degree models to be formed.
- Easy to analyze and interpret.

Disadvantages

- Larger designs could need a large number of trials, which can be expensive both in terms of time and resources.
- The analysis may become more difficult if nonlinear equations must be solved in order to estimate model coefficients.
- Generated models can become complex and difficult to interpret as the number of factors increases.
- In light of experimental restrictions, some combinations of experimental sites might not be practical in reality.

4.4 Roquemore Designs

Roquemore designs proposed on 1976 [24] are a family of experimental designs designed for factorial experiments analysis. They are characterized by a uniform distribution of experimental points in the factor space. These plans allow for studying interactions between factors and constructing mathematical models to predict the behavior of the system under study. Roquemore designs are particularly useful for investigating a limited number of factors while minimizing the number of required trials. They aim to satisfy both the criterion of rotation isovariance and the criterion of near orthogonality.

- For three factors, there are two Roquemore designs, 311A and 311B. They allow the study of 3 factors in 11 trials.

- For four factors, there are two Roquemore designs, 411A and 411B.

- For six factors, there is one Roquemore design, 618A.

We will detail the 311A design and simply mention the other designs without analyzing them.

4.4.1 Roquemore's 311A Design

This design allows three factors to be studied in eleven trials.

Experiences matrix : The experimental points are as shown in (Table 4.5).

Table 4.5: Roquemore's 311A Design for three factor

Trial N°	Factor 1	Factor 2	Factor 3
1	0	0	2
2	0	0	-1
3	-1.414	-1.414	1
4	1.414	-1.414	1
5	-1.414	1.414	1
6	1.414	1.414	1
7	-2	0	-1
8	2	0	-1
9	0	-2	-1
10	0	2	-1
11	0	0	0

Location of experimental points : Evaluating the table makes it possible to identify the following elements:

- A 2^2 design consists of four points located at the corners of a square (trials 3, 4, 5, and 6). These trials are at level +1 for factor 3.

- Four points located at the corners of a square (trials 7, 8, 9, and 10) shifted by 45 degrees compared to the previous square. These points are at level -1 of factor 3.

- Level 0 of factor 3, the central point, is one of three locations on an axis that passes through the centers of the two preceding squares. For the third element, the other two spots are situated at levels +2 and -2, respectively (Figure 4.5) :

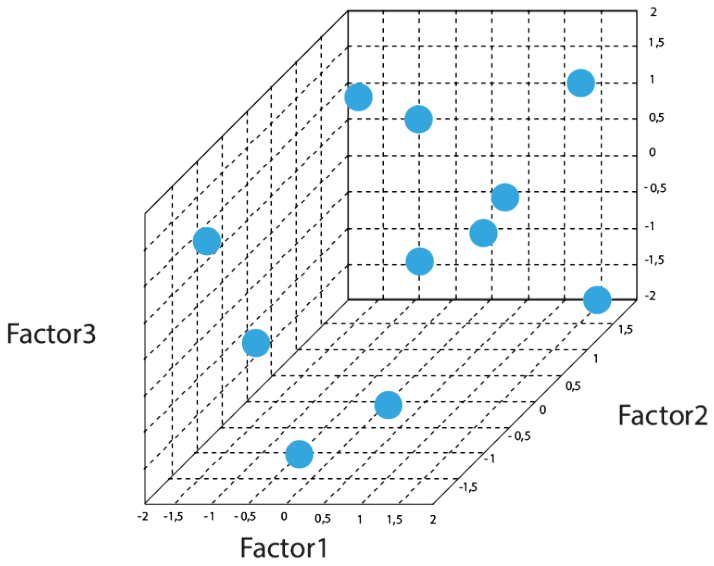

Figure 4.5: Location of experimental points in the Roquemore's 311A design

Mathematical model : The used model in Roquemore Designs is a second-degree model with second-order interactions :

$$Y = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + \hat{a}_3 x_3 + \hat{a}_{12} x_1 x_2 + \hat{a}_{13} x_1 x_3 + \hat{a}_{23} x_2 x_3 + \hat{a}_{11} x_1^2 + \hat{a}_{22} x_2^2 + \hat{a}_{33} x_3^2$$

The design matrix : In this regard, this strategy is almost saturated.So, we given that there are 10 coefficients in the postulated model and 11 trials are performed out, the calculation matrix X for the Roquemore 311A plan is a (10,10) matrix.

The information matrix : In this case, isovariance by rotation is respected for factors 1 and 2, and it's not respected for factor 3. So the information matrix is a (10,10) :

$$X^t X = \begin{pmatrix} 11 & & & & & & & 16 & 16 & 16 \\ & 16 & & & & & & & & \\ & & 16 & & & & & & & \\ & & & 16 & & & & & & \\ & & & & 16 & & & & & \\ & & & & & 16 & & & & \\ & & & & & & 16 & & & \\ 16 & & & & & & & 48 & 16 & 16 \\ 16 & & & & & & & 16 & 48 & 16 \\ 16 & & & & & & & 16 & 16 & 40 \end{pmatrix}$$

Variance-covariance matrix : This matrix is the inverse of the information matrix $(X^t X)^{-1}$. It will be used to calculate the variances of the coefficients and it's also a (10,10) matrix :

$$\begin{pmatrix} 1 & & & & & & & -0.188 & -0.188 & -0.188 \\ & 0.063 & & & & & & & & \\ & & 0.063 & & & & & & & \\ & & & 0.063 & & & & & & \\ & & & & 0.063 & & & & & \\ & & & & & 0.063 & & & & \\ & & & & & & 0.063 & & & \\ -0.188 & & & & & & & 0.061 & 0.029 & 0.029 \\ -0.188 & & & & & & & 0.029 & 0.06 & 0.029 \\ -0.188 & & & & & & & 0.029 & 0.039 & 0.094 \end{pmatrix}$$

Optimality cretiria : The Roquemore 311A plan for three factors satisfies the isovariance by rotation condition for two factors but not for the third, according to an analysis of the information matrix. The Roquemore 311A design does not satisfy the nearly orthogonality requirement, as can be shown from the dispersion matrix analysis. Even so, the values of the secondary diagonal elements are weak.

4.4.2 Other Roquemore Designs

The majority of hybrid Roquemore designs are three- or four-factor designs. They are also present for six factors. The next tables (Table 4.6 and Table 4.7), include

these different plans. The 311A plan's properties can be obtained in the same manner as these plans' properties.

Table 4.6: Roquemore's 311B Design for three factor

Trial N°	Factor 1	Factor 2	Factor 3
1	0	0	2.449
2	0	0	-2.449
3	-0.751	-2.106	-1
4	0.751	-2.106	1
5	-0.751	2.106	1
6	0.751	2.106	-1
7	-2.106	-0.751	1
8	2.106	-0.751	-1
9	-2.106	0.751	-1
10	2.106	0.751	1
11	0	0	0

Table 4.7: Roquemore's Design from two to four factors

Trial N°	Factor 1	Factor 2	Factor 3	Factor 4
1	0	0	0	1.732
2	0	0	0	-0.269
3	-1	-1	-1	0.605
4	1	-1	-1	0.605
5	-1	1	-1	0.605
6	1	1	-1	0.605
7	-1	-1	1	0.605
8	1	-1	1	0.605
9	-1	1	1	0.605
10	1	1	1	0.605
11	-1.518	0	0	-1.05
12	1.518	0	0	-1.05
13	0	-1.518	0	-1.05
14	0	1.518	0	-1.05
15	0	0	-1.518	-1.05
16	0	0	1.518	-1.05

Table 4.8: Roquemore's 416C Design for four factors

Trial N°	Factor 1	Factor 2	Factor 3	Factor 4
1	0	0	0	1.765
2	0	0	-1	-0.568
3	-1	-1	-1	0.568
4	1	-1	-1	0.568
5	-1	1	-1	0.568
6	1	1	1	0.568
7	-1	-1	1	0.568
8	1	-1	1	0.568
9	-1	1	1	0.568
10	1	0	0	-1.051
11	-1.470	0	0	-1.051
12	1.470	1.470	0	-1.051
13	0	1.470	0	-1.051
14	0	0	-1.470	-1.051
15	0	0	-1.470	-1.051
16	0	0	0	-1.051

4.4.3 Advantages and Disadvantages

Advantages

- Through applying these designs, a second-degree model may be created and the 4 factors can be studied in just 16 trials.
- Contrary to this, Doehlert designs need 21 trials.

Disadvantages

- These designs' disadvantage is that they are only available for factors 3, 4, and 6.

4.5 BOX-BEHNKEN Designs

These ideas for second-degree models, which enable the study of three levels -1, 0 and +1 were put out by Box and Behnken in 1960 [4]. These designs have the sequentiality trait, which makes them easier to implement and satisfies a certain optimization requirement.

The results of previously completed tests can be preserved while exploring the first k variables and keeping open the potential of adding more.

4.5.1 Construction of BOX-BEHNKEN Designs

For three factors: The experimental points are located at the midpoints of each edge of the cube (Figure 4.6). Twelve trials represent this design, and one or more central points may be added.

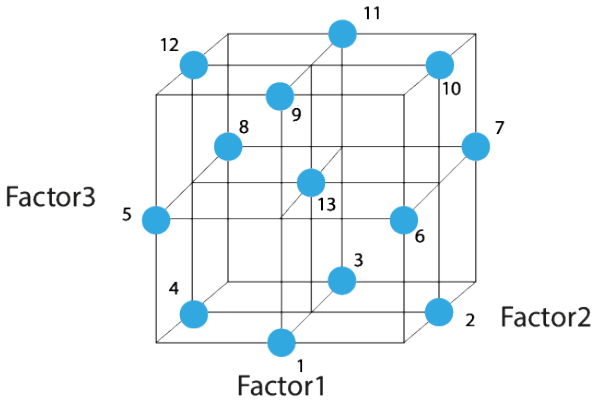

Figure 4.6: Box-Behnken design for three factor

The construction of the design follows these principle:

- Two factors describe a square (4 experiments of a 22 design).
- The corresponding coordinates of the third factor are set to zero.

(Table 4.9) shows these trials :

Table 4.9: Box-Behnken designs for three factor

Trial N°	Factor 1	Factor 2	Factor 3
1	-1	-1	0
2	1	-1	0
3	-1	1	0
4	1	1	0
5	-1	0	-1
6	1	0	-1
7	-1	0	1
8	1	0	1
9	0	-1	-1
10	0	1	-1
11	0	-1	1
12	0	1	1
13 to 15	0	0	0

For four factors: The design is built on a hypercube in 4 dimensions, the experimental points are located at the center of the 24 squares.(Figure 4.7) shows:

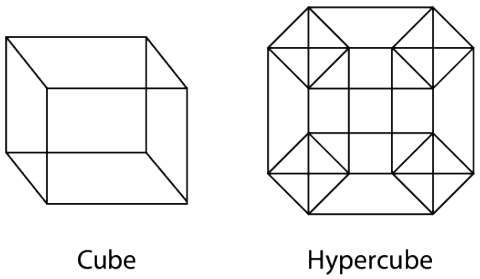

Figure 4.7: Difference between Box-Behnken designs for 3 and 4 factors

The construction of the design is as follows:

Table 4.10: Box-Behnken designs for four factor

Trial N°	Factor 1	Factor 2	Factor 3	Factor 4
1 to 4	±1	±1	0	0
5 to 8	±1	±1	±1	0
9 to 12	±1	0	0	0
13 to 16	0	±1	±1	0
17 to 20	0	±1	0	0
21 to 24	0	0	±1	0
25 to 27	0	0	0	0

In this table, the factor levels are indicated by ±1. This means that the factor alternately takes the levels +1 and -1, and that all combinations are realized.

4.5.2 Properties of BOX-BEHNKEN Designs

postulated mathematical model
The used model is a second-degree model with second-order interactions :

$$Y = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + \hat{a}_3 x_3 + \hat{a}_{12} x_1 x_2 + \hat{a}_{13} x_1 x_3 + \hat{a}_{23} x_2 x_3 + \hat{a}_{11} x_1^2 + \hat{a}_{22} x_2^2 + \hat{a}_{33} x_3^2$$

The design matrix
The design matrix is constructed based on the experimental design and the postulated model. For three factors the matrix is :

$$X = \begin{pmatrix} 1 & -1 & -1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & -1 & 0 & -1 & 0 & 0 & 1 & 1 & 0 \\ 1 & -1 & 1 & 0 & -1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 \\ 1 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & -1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & -1 & 0 & 0 & -1 & 0 & 1 & 1 \\ 1 & 0 & -1 & 1 & 0 & 0 & -1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The information matrix
Using the formula $(X^t X)$, the information matrix is calculated from the design matrix. We notice that for 3 factors, the plan does not respect the criterion of

isovariance by rotation,so we have:

$$X^tX = \begin{pmatrix} 15 & & & & & & & 8 & 8 & 8 \\ & 8 & & & & & & & & \\ & & 8 & & & & & & & \\ & & & 8 & & & & & & \\ & & & & 4 & & & & & \\ & & & & & 4 & & & & \\ & & & & & & 4 & & & \\ 8 & & & & & & & 8 & 4 & 4 \\ 8 & & & & & & & 4 & 8 & 4 \\ 8 & & & & & & & 4 & 4 & 8 \end{pmatrix}$$

Variance-covariance matrix
This matrix is the inverse of the information matrix :

$$(X^tX)^{-1} = \begin{pmatrix} 0.333 & & & & & & & -0.167 & -0.167 & -0.167 \\ & 0.125 & & & & & & & & \\ & & 0.125 & & & & & & & \\ & & & 0.125 & & & & & & \\ & & & & 0.25 & & & & & \\ & & & & & 0.25 & & & & \\ & & & & & & 0.25 & & & \\ -0.167 & & & & & & & 0.271 & 0.021 & 0.021 \\ -0.167 & & & & & & & 0.021 & 0.271 & 0.021 \\ -0.167 & & & & & & & 0.021 & 0.021 & 0.271 \end{pmatrix}$$

The dispersion matrix shows that this design does not respect the criterion of nearly orthogonality. But, if four points are added to the center instead of three, we will obtain a Box-Behnken design that respect this criterion.

4.5.3 Principal of Box-Behnken Designs

There are several other Box-Behnken designs available, but the one with three factors (Table 4.9) is the most popular. As released by the researchers [4], we additionally provide the designs for factors 6 and 7 (Tables 4.11 and 4.12).

Table 4.11: Box-Behnken designs for six factor

Trial N°	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
1 to 8	±1	±1	0	±1	0	0
9 to 16	0	±1	±1	0	±1	0
17 to 24	0	0	±1	±1	0	±1
25 to 32	±1	0	0	±1	±1	0
33 to 40	0	±1	0	0	±1	±1
41 to 48	±1	0	±1	0	0	±1
49 to 52	0	0	0	0	0	0

Table 4.12: Box-Behnken designs for seven factor

Trial N°	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
1 to 8	0	0	0	±1	±1	±1	0
9 to 16	±1	0	0	0	0	±1	±1
17 to 24	0	±1	0	0	±1	0	±1
25 to 32	±1	±1	0	±1	0	0	0
33 to 40	0	0	±1	±1	0	0	±1
41 to 48	±1	0	±1	0	±1	0	0
49 to 56	0	±1	±1	0	0	±1	0
57 to 62	0	0	0	0	0	0	0

4.5.4 Advantages and Disadvantages

Box-Behnken designs are easy to construct and allow the making of a second-degree model. Even so, these designs need a large number of trials when the number of studied factors rises.

Chapter 5

Software Description

The construction of experimental designs is facilitated by the use of specific software. Interpreting experimental designs requires numerous calculations and graphs. Here again, the software we developed simplifies the construction of the most common experimental designs. It performs the necessary calculations for interpreting results and provides the ability to plot numerous graphs that illustrate these results. This makes it possible to appreciate the capabilities offered by experimental design software. Interested readers can use it by selecting one of the programmed designs for various studies in a wide range of fields.

There are several experimental design software programs, each with its own advantages and disadvantages. This version of the software gives a good idea of the contribution of computing to experimental designs and the support that an experimenter can expect from it.

5.1 Startup

The software being installed on the hard drive, we perform to open it double click on his icon, we get the software opening window. After clicking on the button "Open", you access the home window consisting of the toolbar:

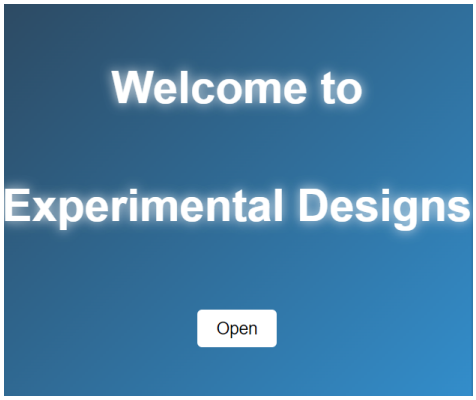

Figure 5.1: Open Window

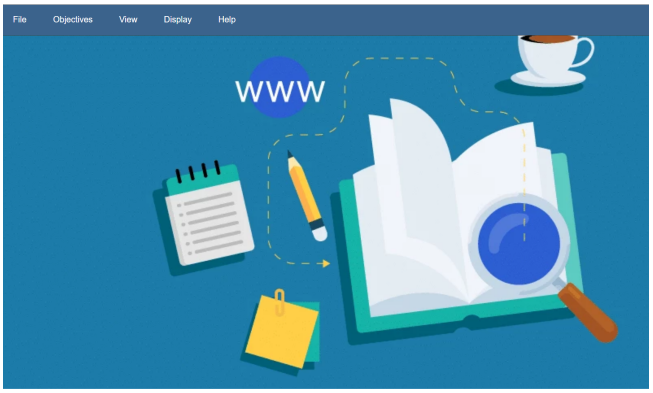

Figure 5.2: Home Window

This toolbar contains five menu choices, which are :

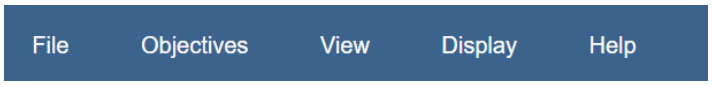

Figure 5.3: Toolbar

5.1.1 Example

Let's consider the process of baking a cake where we want to study the effect of temperature (T) and baking time (BT) on the thickness (E) of the cake. The experimenter decides to use **a full factorial design at 2 levels**. He defines the study domain of the two parameters as follows:

Table 5.1: Our example for 2 study field with 2 factors

	low level	hight level
Temperature	150°	200°
Baking time	15 mn	25 mn

The measured results of the thickness (E) for each experiment conducted are given in a table.

Table 5.2: Experimental designs with results

Trial N°	T	BT	Thickness (E)
1	-1	-1	26
2	1	-1	22
3	-1	1	52
4	1	1	24

5.2 Description Of The Menu Bar

5.2.1 File Menu

It has the following fields:

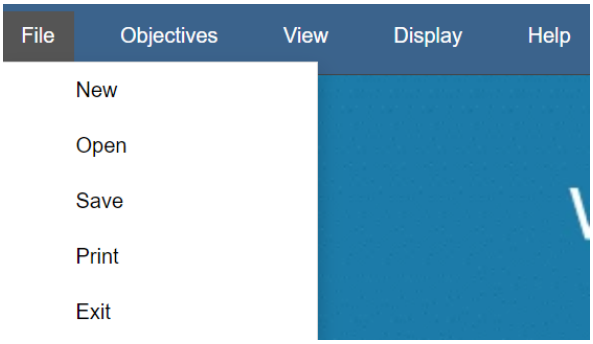

Figure 5.4: File Menu

New : When the "New" item is clicked, a window appears, and the user may follow its steps :

- A window specifying to define at first number of factors then the name, unit, high level, and low level.

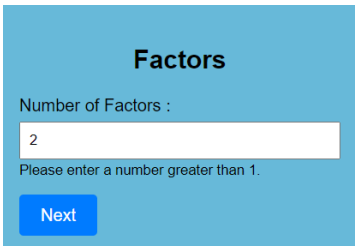

Figure 5.5: Enter Number of Factors

Choosing a number then we click on the "Next" button, we get a form to fill :

Figure 5.6: Introducing factors

- We keep clicking the "Next" button and fill in the factor's information. A window opens up to show and stock all what we had, after validating we introducing the response.

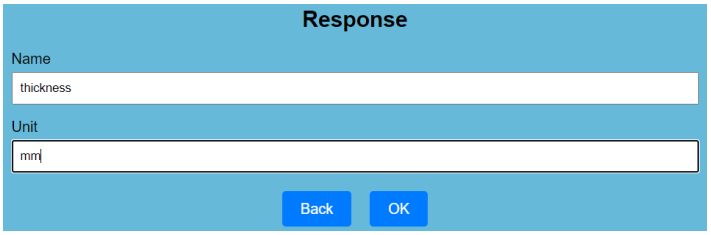

Figure 5.7: Introducing Responses

- Once we click on "OK", we get an other window it shows us a remark.

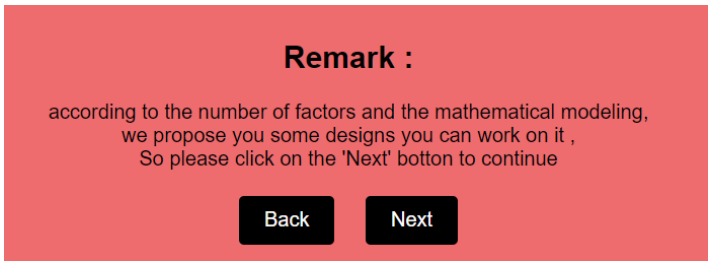

Figure 5.8: Remark

- By tapping the "Next" bottom, a window give us the opportunity to choose a design that we want to work on :

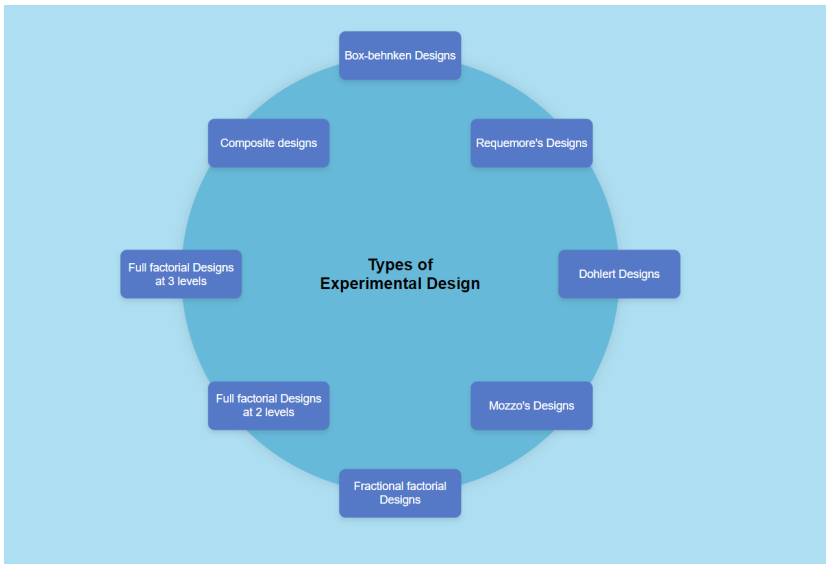

Figure 5.9: Types of Designs

- We choose the Full factorial designs at 2 levels, and we face an **input prompt** to enter responses values.

```
Response for Trial 1: 26
Response for Trial 2: 22
Response for Trial 3: 52
Response for Trial 4: 24
```

Figure 5.10: responses

- Now we obtain the following results :
  1. **Estimation by the least squares method :**
     It select the best possible estimator of $\hat{A}$ which is a vector and it has a relationship with the predicted $\hat{Y}$, which is also a vector. And here is the results :

```
a^ (Estimated Coefficients):
[31. -8.  7. -6.]
y^ (Adjusted Responses):
[26. 22. 52. 24.]
```

Figure 5.11: Estimated coefficients with $\hat{y}$ ajusted

Now, it shows the value of each coefficients that's refer to " the A coefficients vector" representer by :

```
Mean of each coefficient:
Mean of coefficient a0: 31.000
Mean of coefficient a1: -8.000
Mean of coefficient a2: 7.000
Mean of coefficient a3: -6.000
```

Figure 5.12: Coefficients

And here are the residuals :

```
Residuals:
e1: 0.000
e2: 0.000
e3: 0.000
e4: 0.000
```

Figure 5.13: Residuals

The reason why they are all equals to null, because the $\hat{Y}$ equals to Y.

2. **Variance analysis :**
   It is rely to judicious decompositions into sums of squares, obtained as follow :

```
Total Sum of Squares (SST): 596.000
Sum of Regression Squares (SSR): 596.000
Sum of Error Squares (SSE): 0.000
```

Figure 5.14: Sums of squares

We get the sums of error squares null, because we had null residuals in (Figure 5.13). And here are the degrees of freedom (Figure 5.15), and variances (Figure 5.16) as follows :

```
Degrees of freedom :
 SST: 3,
 SSR: 3,
 SSE: 0
```

Figure 5.15: Degrees of freedom

```
Variance :
 SST: 198.667,
 SSR: 198.667,
 SSE: nan
```

Figure 5.16: Variances

3. **Statistical tests :**
   The outcomes of the variance analysis are used to obtain these tests, which are :

```
Multiple correlation coefficient (R^2): 1.000
Adjusted R^2: nan
Fisher test (F): nan
```

Figure 5.17: R^2, R^2ajusted, and Fisher test

4. **Graphical designs :**
   Although this examination is possible, here are the graphical results :

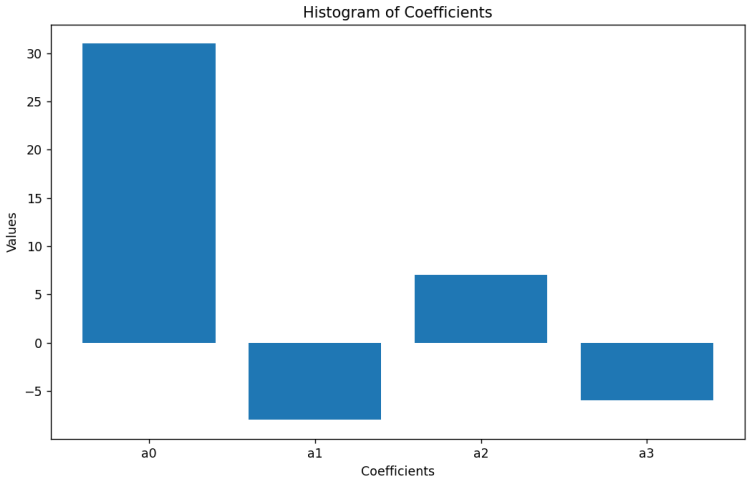

Figure 5.18: Histogram Coefficients

- Histogram showing the different value of each coefficients.

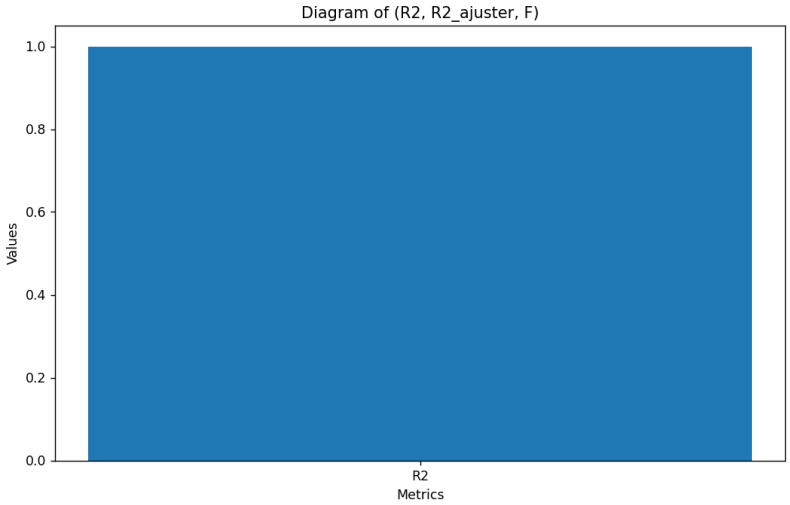

Figure 5.19: Diagram of R^2, R^2ajusted, and Fisher test

- We have just the R^2, because it is the most frequent value.

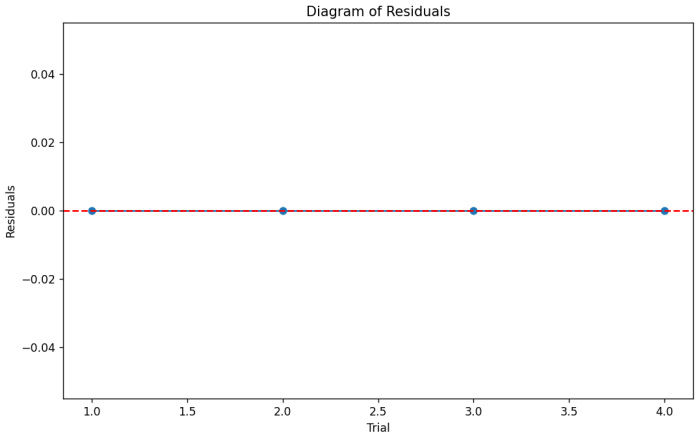

Figure 5.20: Residuals Diagram

• The residuals are defined in that way because they are null ($\hat{Y} = Y$).

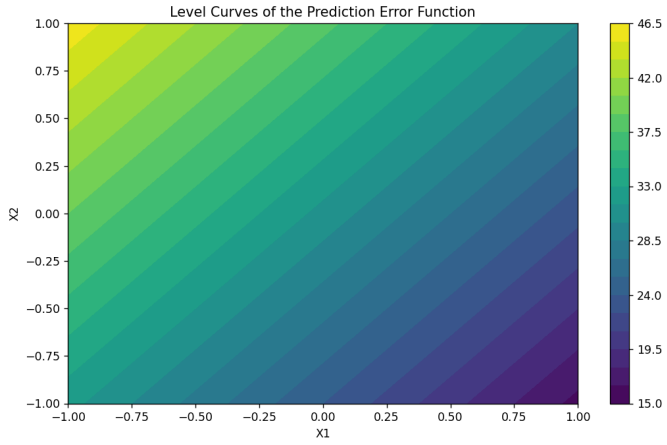

Figure 5.21: Prediction Error (X_1, X_2)

• When using the least squares approach, it suggests to estimate responses, and above we had the curve of the prediction error.

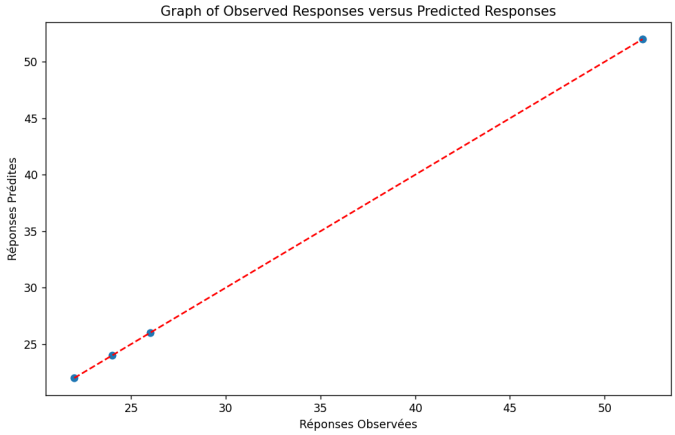

Figure 5.22: Responses Observed Compared to Responses Predicted

• As mentioned the Y and the $\hat{Y}$ have the same value.

Open : Open an application that the user has previously saved.

Save : Saves the problem data to a file or a folder.

Print : Prints the study results.

Leave : Permits you to stop using the program or not.

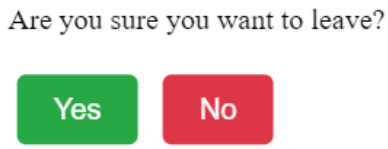

Figure 5.23: Verify Remark For Leaving The Application

5.2.2 Objectives Menu

It allows the choice between a screening designs and optimization ones.

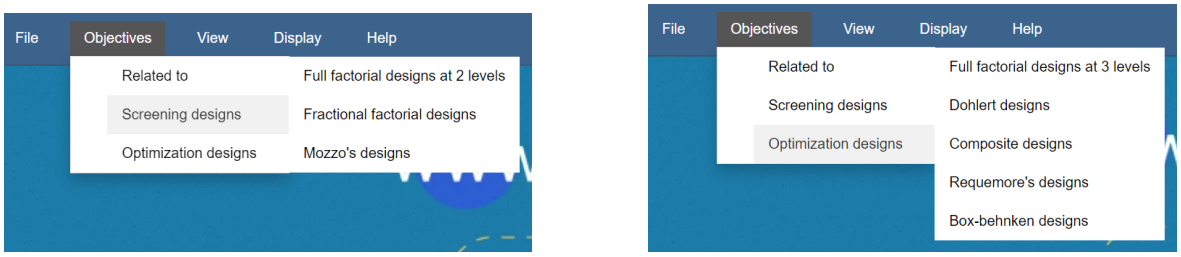

Figure 5.24: Types Designs

After choosing a specific design, the software ask to enter number of factors, then gives us how the trial matrix is formed.

For the menu item "Related to", it will show you what's about the types of designs :

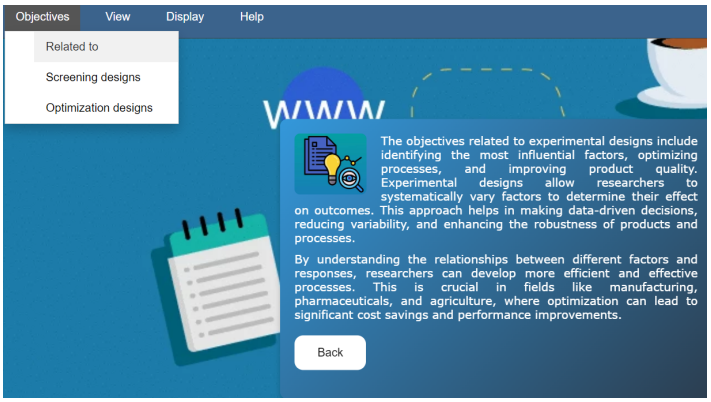

Figure 5.25: Related to Types

5.2.3 View Menu

In that menu, we will putting up the 4^{th} matrix we have in our designs, with the centered reduced coordinates.

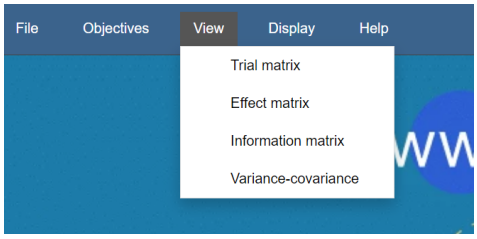

Figure 5.26: View bar

5.2.4 Display bar

It has the following traits :

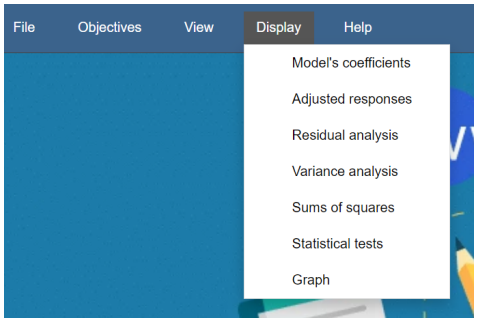

Figure 5.27: Display menu screen

Each menu item contains a specific results we can have it simply when we finish our study.

5.2.5 Help Menu

The help bar show you a screen that explain what the software is able to do, and here is what it contains :

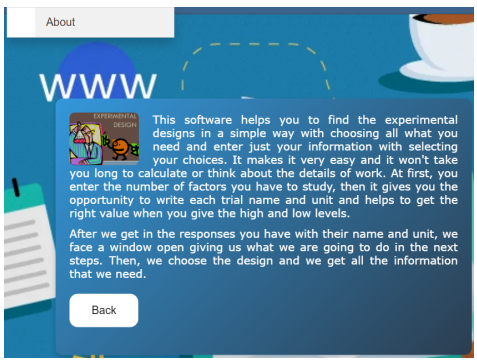

Figure 5.28: About our Software

Conclusion

Mathematicians have always been at the forefront of research, developing several scientific theories, both fundamental and practical. They have consistently improved and refined these theories. However, there has often been a gap between the discovery of these theories and their practical application in the industrial world. This is because these theories often require a considerable volume of calculations when put into practice. Fortunately, the advent of computer tools has addressed this issue by avoiding researchers the burden of tedious and often impossible calculations. For example, Data Analysis, whose initial theoretical developments date back to the 1930s, was only widely used in the industrial world in the 1960s. The same applies to the application of graph theory and many other mathematical theories.

Experimental Research Methodology, also known as experimental design, is a mathematical discipline within inferential statistics. The theory of experimental designs is constantly evolving, and its significance is such, thanks to the existence of specialized software, that it is legitimate to wonder if this discipline should not be taught from the early university years. The work proposed to us had two main objectives: first, to provide a state-of-the-art overview of the various categories of experimental designs existing in the literature, to study their conception, and to evaluate their advantages and disadvantages in order to provide users with a comprehensive document to guide them in choosing and implementing these designs according to their specific needs. Secondly, to develop our own software. This will not only free us from tedious calculations and allow us to process and analyze any practical experiment in chemical, pharmaceutical, agronomic, physical, etc., fields, but also to improve this software.

Bibliography

[1] Paul Schimmerling, Jean Claude Sisson, and Ali Zaïdi. *"Pratique des plans d'expériences"*. TEC & DOC, (1998).

[2] RA Fisher. "the design of experiments",. *2nd edn, Olivier and Boyd, Edinburgh*, (1935).

[3] G. Taguchi and Wu Y. "introduction to off-line quality control". *Central Japan Quality Control Association, Nagoya*, (1980).

[4] George EP Box and Donald W Behnken. "some new three level designs for the study of quantitative variables". *Technometrics*, 2(4):455–475, (1960).

[5] Jacques Goupy. "modélisation par les plans d'expériences". *Techniques de l'ingénieur. Mesures et contrôle*, (R275):1–23, (2000).

[6] GEP Box and J Stuart Hunter. "the 2 k—p fractional factorial designs". *Technometrics*, 42(1):28–47, (2000).

[7] J. Goupy. "plan d'expérience". *Techniques de l'Ingénieur, traité analyse et caractérisation, P 230*, pages 1–20, (1997).

[8] Jacques Goupy. *"Plans d'expériences pour surfaces de réponse"*. Dunod, (1999).

[9] D Mathieu and R Phan-Tan-Luu. "methodology for research using optimal design". *LPRAI Université d'Aix-Marseille*, (1998).

[10] Gilbert Saporta. *"Probabilités, analyse des données et statistique"*. Editions technip, (1990).

[11] Walter Tinsson. *"Plans d'expérience: constructions et analyses statistiques"*, volume 67. Springer Science & Business Media, (2010).

[12] SR Searle. "linear models". john wiley & sons, inc., new york-london-sydney-toronto, (1971).

[13] Yadolah Dodge and Valentin. Rousson. *"Analyse de régression appliquée"*. n° ISBN : 2-10-048659-4, Dunod, (2004).

[14] D Mathieu. "contribution de la méthodologie de la recherche expérimentale à la relation structure-activité". *Thèse de doctorat, Université d'Aix-Marseille*, (1981).

[15] Didier MATHIEU and PHAN-TAN-LUU Roger. *"Planification d'expériences en formulation: criblage"*. Ed. Techniques de l'Ingénieur, traité Génie des procédés, Doc.J242, (2000).

[16] Dominique C. *"Plans d'expériences factoriels"*. Spriger, Paris, (1996).

[17] Jacques L Goupy. ""étude comparative de divers plans d'expériences". *Revue de statistique appliquée*, 38(4):5–44, (1990).

[18] GEP Box and J Stuart Hunter. "the 2 k—p fractional factorial designs". *Technometrics*, 42(1):28–47, (2000).

[19] Max D Morris. "a class of three-level experimental designs for response surface modeling". *Technometrics*, 42(2):111–121, (2000).

[20] George EP Box and Norman R Draper. *"Empirical model-building and response surfaces"*. John Wiley & Sons, New York, 669 pages, (1987).

[21] Jean-Jacques Droesbeke, Jeanne Fine, and Gilbert Saporta. *"Plans d'expériences: applications à l'entreprise"*. Editions technip, 509 pages, (1997).

[22] DH DOHLERT. "uniform shell design". *Applied statistics*, (19):231–239, (1970).

[23] Hu Yuzhu. "program for doehlert matrix design of experiments". *Trends in Analytical Chemistry*, (1989).

[24] KG Roquemore. "hybrid designs for quadratic response surfaces". *Technometrics*, 18(4):419–423, (1976).