

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE



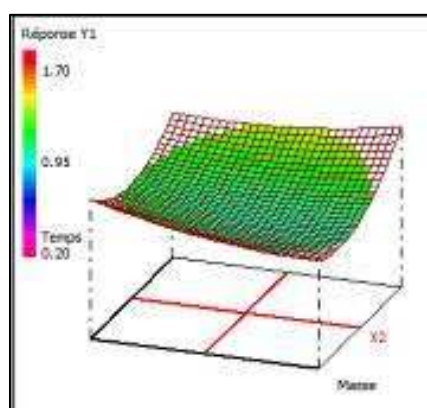
UNIVERSITE MOULOU D MAMMERI, TIZI-OUZOU

FACULTE DE SCIENCES

DEPARTEMENT DE CHIMIE



Initiation à la Modélisation par la Méthode des Plans d'expérience



Dr Hamida IBOUKHOULEF Epse BEKDA

2023-2024

Préface

Ce document de cours est destiné aux étudiants en deuxième année Master Chimie et à l'ensemble des doctorants et chercheurs intéressés par la modélisation avec la méthode des plans d'expérience.

La méthodologie de la recherche expérimentale (MRE), appelée communément, plans d'expérience est une suite d'expérience qui permettent d'organiser les essais qui accompagnent tout travail expérimental. Ils sont appliqués dans diverses disciplines où on cherche un lien entre une grandeur d'intérêt (réponse) et des variables (facteurs). Cette méthode a été mise au point en 1925, par Ronald Fisher, dans le cadre des études agronomiques, puis développée par Génichi Taguchi. Ce dernier a fortement contribué à l'expansion des plans d'expériences en milieu industriel dans les années 1960.

Au cours de ces dernières années, avec le progrès des logiciels, cette méthode a connu un essor spectaculaire et est aujourd'hui un outil indispensable à l'amélioration de plusieurs procédés. Cependant, cette méthodologie est basée sur des règles mathématiques strictes qui exigent une démarche rigoureuse de la part de l'expérimentateur. Une terminologie adaptée et précise doit être utilisée dans l'organisation des essais et dans la modélisation des résultats.

Les plans décrits dans ce document sont ceux du criblage par la matrice d'Hadamard, les plans factoriels complets, les plans factoriels fractionnaires et les plans pour surface de réponse.

L'identification des variables significatives est réalisée par les méthodes statistiques en faisant appel au test de STUDENT et aux outils d'aide à l'interprétation lorsque le degré de liberté (ddl) est élevé.

Concernant l'optimisation, les plans composites centrés, de Box Behnken et de Doehlert sont abordés. La validation des modèles élaborés a été faite par une analyse de variance et le calcul du coefficient de détermination.

Pour faciliter les calculs et la lecture des résultats, la méthodologie des plans d'expérience nécessitent des logiciels scientifiques (JMP, Nemrodw, Modde 6...). Ces derniers sont obtenus sous licence, pour la recherche scientifique et un seul poste de travail ! Pour contourner cette situation, nous avons proposé de travailler sur Excel. C'est pour cela, que ce cours est accompagné par des applications sur Excel.

Ce cours est présenté en cinq chapitres. Dans le premier chapitre, nous avons introduit les notions statistiques. Le deuxième chapitre présente des notions générales sur les plans d'expérience, en se concentrant d'abord sur le vocabulaire requis, la terminologie et l'élaboration d'un modèle. En second lieu, l'importance des PE est démontrée par des exemples de pesées et la représentation d'une droite.

Le troisième chapitre développe le criblage de facteurs par la matrice d'HADAMARD, au calcul des coefficients d'un modèle élaboré et l'estimation des facteurs influents par le test de STUDENT. L'identification des facteurs influents en utilisant les outils d'aide à l'interprétation, à savoir, les méthodes de Pareto, de Lenth et le Normal/Half plot a été aussi développée.

Le quatrième chapitre présente l'élaboration et la validation des modèles des plans factoriels complets et fractionnaires par le test statistique de Fischer et l'analyse de variance. Le dernier chapitre est consacré aux surfaces de réponse et à l'étude des modèles polynomiaux du second degré.

Nous espérons que cet ouvrage, fruit des travaux d'encadrement et de formation que nous avons menés depuis plusieurs années à la faculté des sciences de l'université Mouloud Mammeri de Tizi Ouzou, sera d'une grande utilité pour les étudiants. C'est aussi une manière de leur permettre de dissiper l'idée que les PE sont réservés uniquement aux personnes ayant un bagage important en mathématiques.

Les exemples traités dans ce cours ont été tirés des travaux de recherche publiés, ils ont été cités dans les références.

Liste des abréviations

\bar{x} : Moyenne,
 σ : Ecart type,
 σ^2 : Variance,
IC : Intervall de confiance,
Abs : Valeur absolue,
(-1) : Niveau bas du facteur,
(+1) : Niveau haut du facteur,
 X_i : Variable codée,
 X^T : Matrice transposée,
N : Nombre d'expérience,
 I_d : Matrice identité,
N : Nombre d'expériences à réaliser,
p : Nombre de coefficients du modèle,
k : Nombre de facteurs,
 e_i : Résidus,
 t_{exp} : Valeur calculée du test de Student,
ddl : Degré de liberté,
 t_{crit} : Valeur critique donnée par la table de Student,
 S^2 : Estimateur de variance,
IP : Indice de Pareto,
IPc : Indice de Pareto cumulé,
PSE : Pseudo écart-type,
P : Probabilité,
J : Rang,
 b_j : Coefficient du modèle,
 E_i : Effet principal du facteur X_i .
 E_{ij} (b_{ij}) : Effet de l'interaction entre les variables X_i et X_j ,
SCEL : Somme des carrés des écarts dues à la liaison,
SCER : Somme des carrés des écarts dues aux résidus,
STCE : Somme totale des carrés des écarts,
 F_{obs} : Valeur de F observée,
 F_{crit} : Valeur de F critique donnée par la table de Fisher-Snedecor,
 R^2 : Coefficient de détermination,
 L_i : contraste,
RSM : Response surface methodology,
 d_A : Fonction de variance au point A,
CC : Matrice composite centrée,
 α : Bras de l'étoile,
BB : Plans Box Behnken,

TABLE DES MATIERES

Chapitre I : Notions statistiques

I. Introduction	5
II. Apport des méthodes statistiques	5
1. Moyenne	5
2. Variance.....	5
3. Intervalle de confiance.....	6
Application	6

Chapitre II : Généralités sur les plans d'expériences

I. Introduction.....	7
II. Objectif des plans d'expérience (PE)	7
III. Terminologie.....	7
1. Réponse	8
2. Facteur	8
3. Domaine expérimental	9
4. Variables codées	9
5. Points expérimentaux.....	11
6. Matrice d'expérience	12
7. Plan d'expérimentation	12
IV. Démarche méthodologique	12
V. Modélisation par les plans d'expérience	13
VI. Importance des plans d'expérience	15
1. Exemple de pesée.....	15
2. Qualité de l'information.....	22
Application : Mon premier plan d'expérience	23

Chapitre III : Criblage de facteurs

I. Introduction	28
II. Criblage de facteurs- matrice d'Hadamard.....	28
III. Construction de la matrice d'Hadamard	29
IV. Estimation des facteurs influents	31
V. Identification des facteurs influents en utilisant les outils d'aide à l'interprétation	37
1. Approche de Pareto.....	37
2. Approche de Lenth.....	37
3. Normal Plot ou Half Normal Plot	38
Application	39

Chapitre IV : Etude quantitative des facteurs

Première partie : Plan factoriel complet.....	47
I. Définition	47
II. Construction de la matrice	47
III. Modèle interactif.....	48
1 ^{er} cas : Modèle sans interaction	49
2 ^{eme} cas : Modèle avec interaction	53
IV. Analyse des résultats	56
V. Intervalle de confiance.....	65
VI. Validation du modèle linéaire –ANOVA	66
VII. Coefficient de détermination (R^2).....	67
Application	68
Deuxième partie : Les plans fractionnaires	75
I. Introduction.....	75
II. Définition	75
III. Calcul de BOX.....	76

IV. Générateur et relation de définition.....	78
V. Notion d'alias et de contraste.....	80
VI. Interprétation des contrastes.....	81
VII. Résolution d'un plan fractionnaire.....	82
Application	84

Chapitre VI : Initiation aux surfaces de réponse (RSM)

I. Introduction.....	91
II. Critère d'optimalité	93
1. Qualité de prédiction.....	93
2. Isovariance par rotation.....	93
3. Orthogonalité	94
III. Matrices d'expérience pour surface de réponse	95
1. Matrice composite centrée (PCC).....	95
2. Matrice Box Behnken	96
3. Matrice de Doehlert	102
IV. Recherche des conditions optimales	104
1. Surface d'isoréponse.....	104
2. Fonction de désirabilité.....	105
3. Analyse canonique (AC).....	105
Références bibliographiques	107

Chapitre I

Notions statistiques

Chapitre I : Notions statistiques

I. Introduction

La chimie est une science expérimentale. Pour accéder aux grandeurs qui caractérisent une matière, le chimiste doit faire des mesures. En dépit de tous les soins que l'on peut prendre pour les réaliser, celles-ci ne conduisent pas à la même valeur. Chaque mesure est entachée d'une incertitude et l'expérimentateur n'obtient qu'une valeur approchée du résultat.

L'incertitude ou l'erreur expérimentale est due à des variations qui sont le résultat du hasard (facteur non contrôlable). Si on refait plusieurs fois la même expérience, on n'obtient pas le même résultat à chaque fois, cette grandeur qui peut fluctuer est appelée variable aléatoire.

II. Apport des méthodes statistiques

L'analyse chimique peut se définir comme une suite d'opérations élémentaires statistiquement indépendantes les unes des autres. On désire étudier une population P, on tire un échantillon E de taille N issu de la population P, On analyse les caractéristiques de E puis on généralise à P.

Lorsqu'on répète plusieurs fois la même expérience, la grandeur mesurée (variable aléatoire) prend plusieurs valeurs. On doit estimer *la moyenne, la variance, l'écart-type et l'intervalle de confiance* de ces mesures.

Soit x_1, x_2, \dots, x_N les résultats d'une même expérience.

- **Moyenne**

Quand on répète une mesure faite sur un même échantillon N fois, on obtient des valeurs individuelles légèrement différentes. Dans ce cas, pour calculer le résultat final, il est préférable de se baser sur la moyenne arithmétique de ces N mesures.

La moyenne (\bar{x}) d'une série statistique est donnée par la relation :

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

- **Variance** (var (x) ou σ^2)

La variance (var x) de la même série statistique de N échantillon est donnée par la relation :

$$\text{var}(x) = \sigma^2 = \frac{1}{N-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

L'écart type σ représente la racine carré de la variance, il est donné par la relation :

$$\sigma = \sqrt{\text{var}(x)}$$

• **Intervalle de confiance**

Si on suppose une distribution normale d'une série de résultats, les bornes de l'intervalle de confiance sur une moyenne calculée sont données par l'équation :

$$X = \bar{x} \pm \frac{t \cdot \sigma}{\sqrt{N}}$$

t est le coefficient de Fisher-Student dépendant du nombre de mesures N qui a servi au calcul de s. Ce coefficient se trouve dans des tables pour différents niveaux de probabilité. On choisit le plus généralement la valeur t_α correspondant à un intervalle de confiance de 95 %.

Application

L'analyse du plomb dans une eau potable, par absorption atomique, donne les résultats suivants :

N°	1	2	3	4	5	6
[Pb] (ppm)	19.4	19.5	19.6	19.8	20.1	20.3

- Calculer la moyenne, la variance et l'écart type de ces mesures,
- Quelles sont les limites de l'intervalle de confiance du plomb à 99% ? ($t_{99\%}=4,03$).

Corrigé

N°	x	$(x - \bar{x})$	$(x - \bar{x})^2$
1	19,4	-0,38	0,1444
2	19,5	-0,28	0,0784
3	19,6	-0,18	0,0324
4	19,8	0,02	0,0004
5	20,1	0,32	0,1024
6	20,3	0,52	0,2704
$\overline{(x)}$	19,78		
		somme	0,63
		Var (x)	0,13
		Ecart type	0,35
		racine N	2,45
		IC	0,58

$$X = \bar{x} \pm \frac{t \cdot \sigma}{\sqrt{N}} = 19.78 \pm 0.57 \text{ ppm}$$

Chapitre II

Généralités sur les plans d'expérience

Chapitre II : Généralités sur les plans d'expériences

I. Introduction

L'approche des plans d'expérience (PE) est un outil robuste pour la compréhension et l'optimisation des paramètres expérimentaux. Elle permet une étude rationnelle de leurs influences sur les réponses choisies avec une diminution considérable du nombre d'essais.

Les plans d'expérience est une suite d'expérience qui permettent d'organiser au mieux les essais qui accompagnent la recherche scientifique ou des études industrielles. Ils sont appliqués dans diverses disciplines où on cherche un lien entre une grandeur d'intérêt (Y) et des variables (Xi). Ces plans d'expérience appartiennent à une branche des statistiques appliquées, ils sont basés sur l'utilisation de méthodes scientifiques pour la planification, la réalisation, l'analyse et l'interprétation des données. Ces méthodes permettent de modifier simultanément les variables d'entrée (appelées facteurs) sur les variables de sortie (appelées réponses).

En combinant plusieurs variables dans une étude au lieu de créer une étude pour chacune, le nombre d'essai requis sera considérablement réduit. Ce qui permet d'obtenir un maximum de renseignement avec un minimum d'expérience et par conséquent un coût réduit.

II. Objectif des plans d'expérience (PE)

L'objectif des plans d'expériences est :

- De planifier l'étude d'un très grand nombre de paramètres : de nombreuses expériences maintiennent certains facteurs constants et modifient le niveau d'une seule variable. L'utilisation de cette approche, appelée un facteur à la fois, pour réaliser puis traiter des données est souvent inefficace et ne reflètent pas l'évolution simultanée des niveaux des facteurs. En effet, les interactions entre ces facteurs seront omises. Dans la méthode des PE, l'ensemble des paramètres pouvant influencer ces expériences sera modifié puis quantifié au même temps.

- De minimiser le coût : En maîtrisant les PE, le nombre d'expérience à réaliser sera considérablement réduit. Par conséquent, un PE correct peut faire gagner du temps et du coût lors d'une étude expérimentale.

- De modéliser et d'interpréter les résultats obtenus dans le domaine expérimental choisi.

III. Terminologie

Au cours de ces dernières années, avec le progrès des logiciels, la méthode des plans d'expérience a connu un essor spectaculaire et est aujourd'hui un outil indispensable pour l'amélioration de plusieurs procédés. Cependant, cette méthodologie est basée sur des règles

mathématiques strictes qui exigent une démarche rigoureuse de la part de l'expérimentateur. Une terminologie et un vocabulaire adapté et précis doivent être définis :

1. Réponse

On appelle réponse, la grandeur physique étudiée dans l'examen d'un phénomène. Cette étude peut se traduire par une ou plusieurs réponses (rendement d'une réaction, abattement de la couleur d'un polluant..). La valeur d'une réponse est le résultat d'une expérience ou une simulation. Elle ne peut être modifiée que de manière indirecte en faisant varier les facteurs (figure 1).

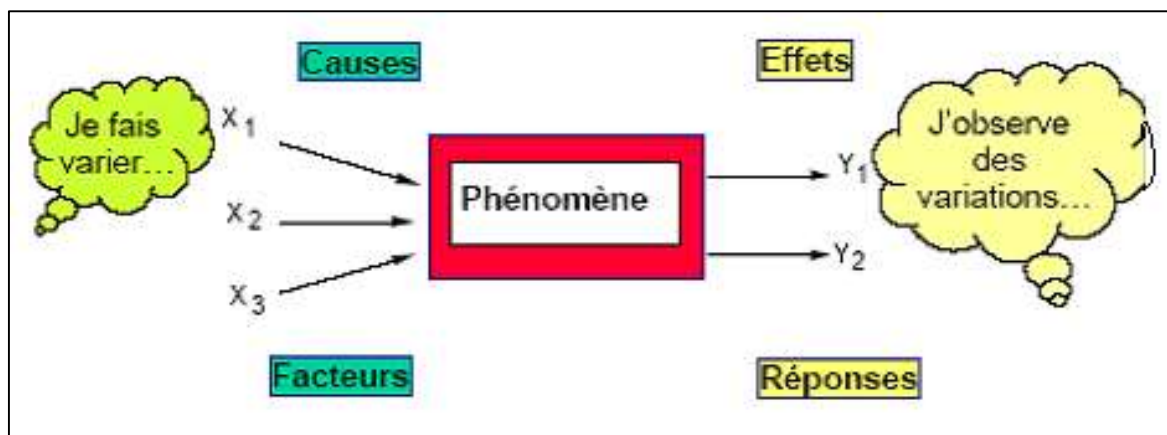


Figure 1 : plan d'expérience

2. Facteur

Un facteur est une grandeur (ou une variable) physique ou chimique, modifiable ou contrôlable par l'expérimentateur, censée avoir une influence sur les réponses considérées. Les facteurs **sont indépendants** peuvent être qualitatifs ou quantitatifs, continues ou discontinus.

On fait souvent varier un facteur entre un niveau inférieur (ou bas) et un niveau supérieur (ou haut). Lorsqu'un facteur varie on dit qu'il change de niveau (figure 2).

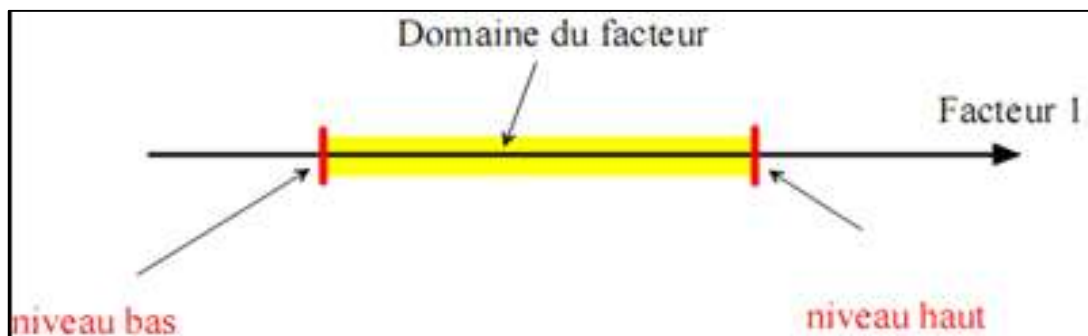


Figure 2 : Domaine de variation d'un facteur

3. Domaine expérimental

Il présente l'espace dans lequel peuvent varier les facteurs. Les informations tirées des résultats expérimentaux ne seront valables que dans ce domaine (figure 3). Ce domaine est défini entre le niveau bas et le niveau haut du facteur.

S'il y a un second facteur, il est représenté, lui aussi, par un axe gradué et orienté. On note, comme pour le premier facteur, son niveau haut, son niveau bas et son domaine de variation. Ce second axe est disposé orthogonalement au premier. On obtient ainsi un repère cartésien qui définit un espace euclidien à deux dimensions. Cet espace est appelé l'espace expérimental.

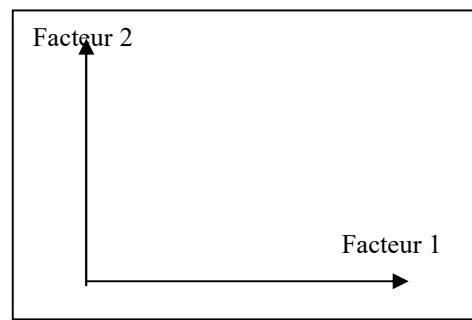


Figure 3 : Domaine d'étude.

4. Variables codées

Les variables naturelles sont des valeurs algébriques qui sont définies pour quantifier les différents facteurs. Généralement, elles ne sont pas exprimées dans le même système d'unités, ce qui rend difficile la comparaison de leurs effets. Pour cela, dans les plans d'expériences, toutes les variables naturelles sont transformées en valeurs codées sans dimension.

L'utilisation des variables codées présente l'intérêt de pouvoir généraliser la méthode des plans d'expérience quel que soit le domaine de l'expérimentateur.

Soit A , la variable naturelle qui passe d'un niveau bas (A_-) à un niveau haut (A_+),

On appelle A_0 , la valeur au centre du domaine donnée par :

$$A_0 = \frac{(A_+ + A_-)}{2}$$

On appelle : Pas de variation :

$$Pas = \frac{(A_+ - A_-)}{2}$$

Pour passer d'une variable naturelle A, à une variable centrée réduite (ou codée) notée X :

$$X = \frac{A - A_0}{Pas}$$

Exemple

Une extraction solide-liquide réalisée à deux températures 25 et 55°C.

Quelle sera la variable codée correspondante à une température de 32.5°C.

$$A_- = 25^\circ C \rightarrow X = -1$$

$$A_+ = 55^\circ C \rightarrow X = +1$$

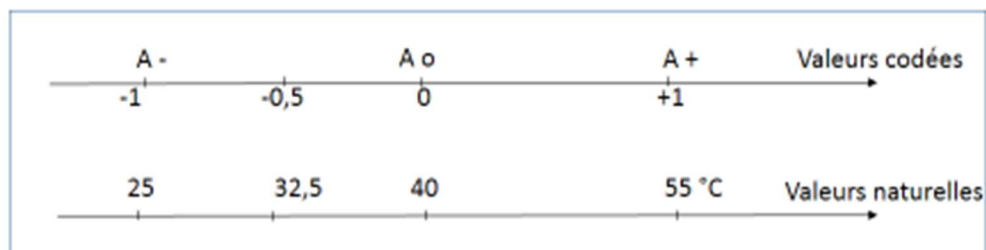
$$A_0 = \frac{(A_+ + A_-)}{2} = \frac{25 + 55}{2} = 40^\circ C$$

Le pas de variation est donné par :

$$Pas = \frac{(A_+ - A_-)}{2} = \frac{55 - 25}{2} = 15^\circ C$$

La variable codée correspondante à une température de 32.5°C est donnée par :

$$X = \frac{32.5 - 40}{15} = -0.5$$



5. Points expérimentaux

Le niveau X1 du facteur 1 et le niveau X2 du facteur 2 peuvent être considéré comme les coordonnées d'un point de l'espace expérimental. Une expérience donnée est alors représentée par un point dans ce système d'axe. Un plan d'expérience est représenté par un ensemble de points expérimentaux.

Le regroupement des domaines des facteurs définit le « domaine d'étude ». Ce domaine d'étude est la zone de l'espace expérimental choisie par l'expérimentateur pour faire ses essais. Une étude, c'est-à-dire plusieurs expériences bien définies, est représentée par des points répartis dans le domaine d'étude (Figure 4).

Cette façon de représenter une expérimentation par des points dans un espace cartésien est une représentation géométrique de l'étude. Une autre représentation d'une étude sera introduite ultérieurement.

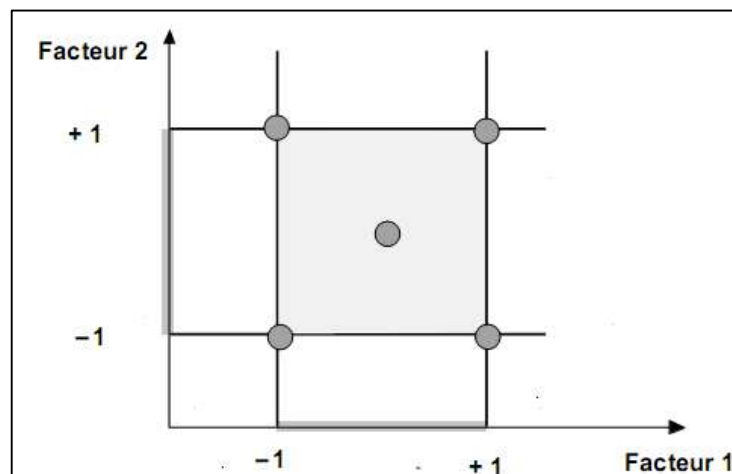


Figure 4 : Disposition des points expérimentaux dans un domaine d'étude.

Jusqu'à trois facteurs, il est possible de dessiner le domaine d'étude. Au-delà de trois facteurs, on utilise une représentation en tableau, dite matricielle, plus générale puisqu'elle permet de représenter les points d'expériences dans un hypervolume à un nombre quelconque de dimensions.

6. Matrice d'expérience

Une matrice d'expérience est un objet mathématique qui représente l'ensemble des expériences à réaliser. Elle est toujours écrite sous forme **codée**. Elle est constituée de :

- N lignes correspondant au nombre d'expérience à réaliser,
- k colonnes correspondant au nombre de facteurs à faire varier.

7. Plan d'expérimentation

Un plan d'expérimentation correspond à la « traduction » de la matrice d'expériences en une matrice directement utilisable par l'expérimentateur car les facteurs seront exprimés en variables naturelles.

Le plan d'expérimentation doit faire l'objet d'une analyse minutieuse pour voir si toutes les expériences sont réalisables et si elles ne présentent aucun risque.

IV. Démarche méthodologique

La démarche à respecter est la suivante :

- 1- Définition de l'objectif :
 - Recherche exploratoire,
 - Criblage des facteurs,
 - Etude quantitative des facteurs,
 - Etude quantitatives des surfaces de réponse,
- 2- Choix de (s) la (les) réponse (s),
- 3- Choix des facteurs et du domaine expérimental d'intérêt,
- 4- Etablissement de la stratégie expérimentale :
 - Construction de la matrice d'expérience,
 - Construction du plan d'expérimentation,
 - Expérimentation,
- 5- Calcul des coefficients du modèle proposé,
- 6- Interprétation des résultats.

V. Modélisation par les plans d'expérience

La modélisation est l'objectif des plans d'expérience. Elle permet, une fois que les facteurs influents sont identifiés, d'écrire l'équation qui relie la réponse (y) aux variables codées (X_i).

L'intérêt de modéliser la réponse par un polynôme est de pouvoir calculer ensuite toutes les réponses du domaine d'étude sans faire les expériences.

1. Modèle théorique

On choisit une fonction mathématique qui relie la réponse aux facteurs. On prend un développement limité de la série de Taylor-Mac Laurin qui prend la forme d'un polynôme de degré plus ou moins élevé.

$$y = b_0 + \sum_{i=1}^k b_i X_i + \sum_{i,j=1, i \neq j}^k b_{ij} X_i X_j + \sum_{i=1}^k b_{ii} X_i^2$$

Où

y est la réponse. Elle est mesurée au cours de l'expérimentation et elle est obtenue avec une précision donnée.

X représente le niveau attribué au facteur i par l'expérimentateur pour réaliser un essai.

b_0 , b_i , b_{ij} , b_{ii} sont les coefficients du modèle mathématique adopté. Ils ne sont pas connus et doivent être calculés à partir des résultats des expériences.

2. Modèle de l'expérimental

Deux compléments doivent être apportés au modèle précédemment décrit :

- *Le premier complément* est le "manque d'ajustement". Cette expression traduit le fait que le modèle théorique est fort probablement différent du modèle réel qui régit le phénomène étudié. Il y a un écart entre ces deux modèles. Cet écart est le manque d'ajustement (lack of fit).

- *Le second complément* est la prise en compte de la nature aléatoire de la réponse. En effet, si l'on mesure plusieurs fois une réponse en un même point expérimental, on n'obtient pas exactement le même résultat. Les résultats sont dispersés. Les dispersions ainsi constatées sont appelées erreurs expérimentales (pure error). Ces deux écarts, manque d'ajustement et erreur expérimentale, sont souvent réunis dans un seul écart, notée e . Le modèle utilisé par l'expérimentateur s'écrit alors :

$$y = b_0 + \sum_{i=1}^k b_i X_i + \sum_{i,j=1, i \neq j}^k b_{ij} X_i X_j + \sum_{i=1}^k b_{ii} X_i^2 + e$$

e : l'écart entre les deux modèles

3. Système d'équation

Chaque point expérimental permet d'obtenir une valeur de la réponse. Cette réponse est modélisée par un polynôme dont les coefficients sont les inconnues qu'il faut déterminer. A la fin du plan d'expériences, on a un système de n équations (s'il y a n essais) à p inconnues. Ce système s'écrit d'une manière simple en notation matricielle :

$$y = bX + e$$

y est le vecteur des réponses,

X est la matrice de calcul, ou matrice du modèle, qui dépend des points expérimentaux,

b est le vecteur des coefficients ou des effets.

e est le vecteur des écarts.

Pour résoudre ce système, on utilise une méthode de régression basée sur le critère des moindres carrés. On obtient ainsi les estimations des coefficients que l'on note :

$$b = (X^T X)^{-1} X^T Y$$

X^T : la matrice transposée de la matrice X .

VI. Importance des plans d'expérience

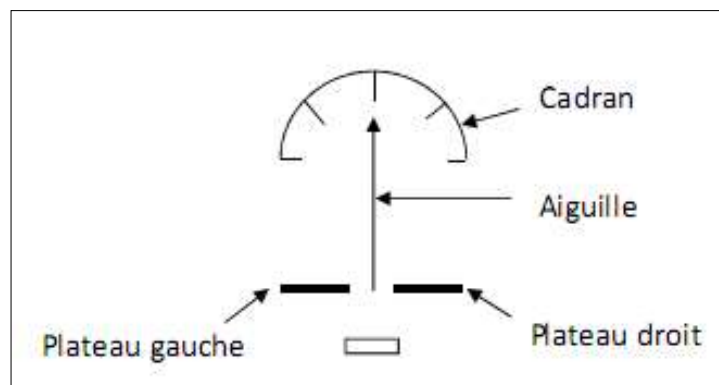
La méthodologie de la recherche expérimentale nous permet de choisir de façon judicieuse des essais pouvant répondre aux questions posées avec une précision maximum.

✚ Exemple de pesée

❖ **Problématique** : Déterminer la masse de trois objets à l'aide d'une balance à deux plateaux en quatre pesées avec un coût minimum.

❖ Conditions

- Peser aucun, un ou plusieurs objets à la fois,
- Pas d'objet sur le plateau : 0
- Objet sur le plateau droite : +1,
- Objet sur le plateau gauche : -1,
- Payer 100 DA à chaque pesée.



❖ Hypothèses

- L'erreur est une variable aléatoire différente systématique, la distribution des erreurs suit une loi normale,
- On ne connaît pas la position de l'aiguille quand il n'y a aucun objet sur la balance (erreur systématique),
- L'ordre de grandeur de l'erreur est donné par la variance expérimentale, elle est constante quel que soit l'objet pesé,
- Les pesées sont indépendantes entre elles : $cov(y_1, y_2) = 0$.

❖ **Rappels**

1. La variance de l'erreur expérimentale est constante dans tout le domaine expérimental.

$$y_i = n_i + e_i \rightarrow \text{var}(y_i) = \sigma^2$$

y_i : résultat de la pesée,

n_i : valeur théorique,

e_i : erreur.

2. Rappelons que :

$$\text{var}(y_2 + y_1) = \text{var}(y_2) + \text{var}(y_1) + 2 \text{cov}(y_1, y_2) = 2\sigma^2$$

$$\text{var}(y_2 - y_1) = \text{var}(y_2) + \text{var}(y_1) - 2 \text{cov}(y_1, y_2) = 2\sigma^2$$

$$\text{var}(ky) = k^2 \text{var}(y) = k^2 \sigma^2$$

Pour déterminer la masse d'un objet, nous allons proposer plusieurs stratégies expérimentales et montrer que selon la méthode adoptée, les résultats peuvent être différents.

Méthode 1 : Elle consiste à travailler comme le font la plupart des expérimentations, c'est-à-dire peser **un objet à la fois**.

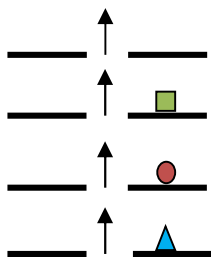
1^{ère} pesée : rien sur la balance : y_1

2^e pesée : objet 1 sur la balance : y_2

3^e pesée : objet 2 sur la balance : y_3

4^e pesée : objet 3 sur la balance : y_4

Plan s'expérience



N° pesée	■	●	▲	Résultats
1	0	0	0	y_1
2	1	0	0	Y_2
3	0	1	0	Y_3
4	0	0	1	Y_4

On appelle b_1 , b_2 et b_3 les estimations des masses des trois objets.

$$\begin{cases} y_1 = b_0 \\ y_2 = b_0 + b_1 \\ y_3 = b_0 + b_2 \\ y_4 = b_0 + b_3 \end{cases} \rightarrow \begin{cases} b_1 = y_2 - y_1 \\ b_2 = y_3 - y_1 \\ b_3 = y_4 - y_1 \end{cases}$$

$$\text{var}(b_1) = \text{var}(b_2) = \text{var}(b_3) = \text{var}(y_i - y_1) = 2\sigma^2 \rightarrow \mathbf{var(b_j) = 2\sigma^2}$$

Coût : Nous avons fait quatre pesées : **400DA**

Comment réduire la variance ?

Faire chaque pesée deux fois puis calculer la moyenne des résultats. En gardant la même stratégie, que faut-il faire pour obtenir une meilleure précision. Il suffit de répéter les pesées plusieurs fois. Si on répète deux fois les mêmes pesées, on obtiendra :

$$b_0 = \frac{y_1 + y'_1}{2} = \frac{1}{2}(y_1 + y'_1)$$

$$b_1 = \frac{y_2 + y'_2}{2} - \frac{y_1 + y'_1}{2} = \frac{1}{2}(y_2 + y'_2 - y_1 + y'_1)$$

$$b_2 = \frac{y_3 + y'_3}{2} - \frac{y_1 + y'_1}{2} = \frac{1}{2}(y_3 + y'_3 - y_1 + y'_1)$$

$$b_3 = \frac{y_4 + y'_4}{2} - \frac{y_1 + y'_1}{2} = \frac{1}{2}(y_4 + y'_4 - y_1 + y'_1)$$

$$\text{On obtient : } \text{var } b_1 = \text{var } \frac{1}{2}(y_2 + y'_2 - y_1 + y'_1) = \frac{1}{4} \cdot 4\sigma^2 = \sigma^2$$

$$\text{var } b_1 = \text{var } b_2 = \text{var } b_3 = \sigma^2 \quad \text{ou} \quad \mathbf{var(b_j) = \sigma^2} \implies \mathbf{Coût 800 DA}$$

Ceci montre que doubler le nombre d'essais permet de **diminuer la variance** des estimations par deux mais en **doublant le prix** de l'expérimentation.

Si on désire augmenter la précision des résultats, on doit continuer à répéter les mêmes essais plusieurs fois, on obtient :

Nombre de répétitions	1	2	4	8
Nombre d'essais	4	8	16	32
Var (b _i)	2 σ ²	σ ²	$\frac{\sigma^2}{2}$	$\frac{\sigma^2}{4}$
Cout (DA)	400	800	1600	3200

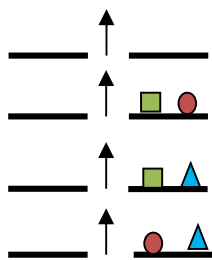
Conclusion : plan doublé variance réduite et coût élevé !

Méthode 2 : Peser deux objets à la fois

Cette méthode ne permet pas de visualiser directement le résultat désiré. Nous allons placer successivement sur le plateau gauche de la balance toutes les combinaisons des 3 objets pris 2 à 2.

- 1° pesée : Rien sur la balance.
- 2° pesée : objet 1 et 2 sur la balance.
- 3° pesée : objet 1 et 3 sur la balance.
- 4° pesée : objet 2 et 3 sur la balance.

Plan d'expérimentation



N° pesée	■	●	▲	Résultats
1	0	0	0	y ₁
2	1	1	0	Y ₂
3	1	0	1	Y ₃
4	0	1	1	Y ₄

Obtient :

$$\left\{ \begin{array}{l} y_1 = b_0 \\ y_2 = b_0 + b_1 + b_2 \\ y_3 = b_0 + b_1 + b_3 \\ y_4 = b_0 + b_2 + b_3 \end{array} \right. \rightarrow \left\{ \begin{array}{l} b_1 = \frac{(y_2 + y_3 - y_1 - y_4)}{2} \\ b_2 = \frac{(y_2 + y_4 - y_1 - y_3)}{2} \\ b_3 = \frac{(y_3 + y_4 - y_1 - y_2)}{2} \end{array} \right.$$

$$Var b_1 = var \frac{(y_2 + y_3 - y_1 - y_4)}{2} = \frac{1}{4} [var (y_2) + var (y_3) + var (y_1) + var (y_4)]$$

$$var(b_1) = var(b_2) = var(b_3) = var(b_4) = \sigma^2$$

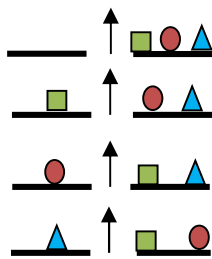
$$var(b_j) = \sigma^2 \implies \text{cout} = 400 DA$$

On remarque que cette méthode est meilleure que la première car elle permet l'estimation des masses des objets avec plus de précision sans pour autant augmenter le coût.

Méthode 3 : Peser trois objets à la fois

- 1° pesée : objet 1, 2 et 3 sur plateau de droite.
- 2° pesée : objet 1 sur le plateau de gauche.
- 3° pesée ; objet 2 sur le plateau de gauche
- 4° pesée : objet 3 sur le plateau de gauche.

Plan d'expérimentation



N° pesée	■	●	▲	Résultats
1	+1	+1	+1	y ₁
2	-1	+1	+1	Y ₂
3	+1	-1	+1	Y ₃
4	+1	+1	-1	Y ₄

$$\begin{cases} y_1 = b_0 + b_1 + b_2 + b_3 \\ y_2 = b_0 - b_1 + b_2 + b_3 \\ y_3 = b_0 + b_1 - b_2 + b_3 \\ y_4 = b_0 + b_1 + b_2 - b_3 \end{cases} \rightarrow \begin{cases} b_1 = \frac{(y_2 - y_1)}{2} \\ b_2 = \frac{(y_1 - y_3)}{2} \\ b_3 = \frac{(y_1 - y_4)}{2} \end{cases}$$

$$var b_1 = var\left(\frac{1}{2}(y_2 - y_1)\right) = \frac{1}{4} [var (y_2) + var (y_1)] = \frac{1}{4} (\sigma^2 + \sigma^2) = \frac{\sigma^2}{2}$$

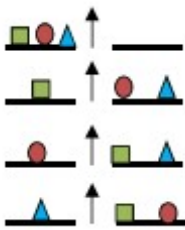
$$Var(b_j) = \frac{\sigma^2}{2} \text{ et le coût} = 400DA$$

Les méthodes 2 et 3 ont montré qu'il est conseillé de peser plusieurs objets à la fois pour obtenir une meilleure précision.

L'examen de la dernière stratégie montre qu'il existe une certaine asymétrie dans le nombre de fois qu'apparaissent les objets sur chacun des plateaux : chaque objet apparaît 3 fois sur le plateau de droite, et une seule fois sur le plateau de gauche. La 4^e méthode consiste à réécrire cette symétrie en plaçant tous les objets sur le plateau de gauche. Ainsi, chaque objet apparaîtra deux fois sur chaque plateau.

Méthode 4 : inverser la stratégie 3 dans la première pesée

Plan d'expérimentation



N° pesée	■	●	▲	Résultats
1	-1	-1	-1	y_1
2	-1	+1	+1	Y_2
3	+1	-1	+1	Y_3
4	+1	+1	-1	Y_4

$$\begin{cases} y_1 = b_0 - b_1 - b_2 - b_3 \\ y_2 = b_0 - b_1 + b_2 + b_3 \\ y_3 = b_0 + b_1 - b_2 + b_3 \\ y_4 = b_0 + b_1 + b_2 - b_3 \end{cases} \rightarrow \begin{cases} b_1 = \frac{(-y_1 - y_2 + y_3 + y_4)}{4} \\ b_2 = \frac{(-y_1 + y_2 - y_3 + y_4)}{4} \\ b_3 = \frac{(-y_1 + y_2 + y_3 - y_4)}{4} \end{cases}$$

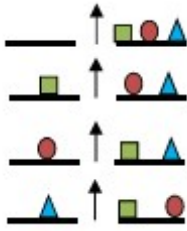
$$\text{var}(b_1) = \frac{1}{16} [\text{var}(y_1) + \text{var}(y_2) + \text{var}(y_3) + \text{var}(y_4)] = \frac{1}{4} (\sigma^2 + \sigma^2) = \frac{4\sigma^2}{16}$$

$$\text{var}(b_1) = \text{var}(b_2) = \text{var}(b_3) = \text{var}(b_4) = \frac{\sigma^2}{4}$$

$$\text{var}(b_j) = \frac{\sigma^2}{4} \text{ et le coût} = 400 \text{ DA}$$

En permutant les objets entre le plateau gauche et le plateau droit, on obtient les mêmes résultats. C'est la même stratégie où la symétrie est respectée.

Plan d'expérimentation



N° pesée	■	●	▲	Résultats
1	+1	+1	+1	y_1
2	+1	-1	-1	Y_2
3	-1	+1	-1	Y_3
4	-1	-1	+1	Y_4

$$\begin{cases} y_1 = b_0 + b_1 + b_2 + b_3 \\ y_2 = b_0 + b_1 - b_2 - b_3 \\ y_3 = b_0 - b_1 + b_2 - b_3 \\ y_4 = b_0 - b_1 - b_2 + b_3 \end{cases} \rightarrow \begin{cases} b_1 = \frac{(y_1 + y_2 - y_3 - y_4)}{4} \\ b_2 = \frac{(y_1 - y_2 + y_3 - y_4)}{4} \\ b_3 = \frac{(y_1 - y_2 - y_3 + y_4)}{4} \end{cases}$$

Le calcul de la variance nous donne :

$$\text{var}(b_j) = \frac{\sigma^2}{4} \text{ et le coût est de } 400 \text{ DA}$$

Théorème : Pour un plan de N pesée utilisant une balance à deux plateaux, la variance de l'estimateur b_j est telle que :

$$\text{var}(b_j) \geq \frac{\sigma^2}{N}$$

k : nombre d'objets

N : nombre d'expériences : $N = k + 1$

Conclusion

Cette expérience montre l'intérêt de la planification des expériences sur la précision, la qualité et le coût.

✚ Qualité de l'information

Problématique : trouver les coefficients de la droite qui a pour équation :

$$\eta = \beta_0 + \beta_1 X$$

η : La réponse à mesurer,

β : Coefficients de la droite.

❖ *Cas idéal* : Pas d'erreurs expérimentales

Toutes les droites sont tracées avec deux points au minimum :

$$\eta_1 = \beta_0 + \beta_1 X_1$$

$$\eta_2 = \beta_0 + \beta_1 X_2$$

Le coefficient directeur de la droite est donné par :

$$\beta_1 = \frac{\eta_2 - \eta_1}{X_2 - X_1}$$

❖ *Cas réel* : toute expérience est entachée d'erreurs expérimentales

$$y_i = \eta_i + e_i$$

Lorsqu'on fait une expérience, on ne trouve pas η_i mais y_i .

y_i est l'estimateur de η_i ,

e_i : erreur.

$$y_2 = \eta_2 + e_2 = \beta_0 + \beta_1 X_2 + e_2 = b_0 + b_1 X_2$$

$$y_1 = \eta_1 + e_1 = \beta_0 + \beta_1 X_1 + e_1 = b_0 + b_1 X_1$$

b_0 et b_1 sont les estimations de β_0 et β_1 .

A partir des expressions précédentes, on déduit que :

$$\beta_1 = \frac{y_2 - y_1}{X_2 - X_1} - \frac{e_2 - e_1}{X_2 - X_1}$$

Et

$$b_1 = \frac{y_2 - y_1}{X_2 - X_1}$$

En comparant, on conclut que :

$$b_1 = \beta_1 + \frac{e_2 - e_1}{X_2 - X_1}$$

$$b_1 = \beta_1 \text{ si } e_1 - e_2 = 0 \text{ (non) et } X_2 - X_1 \rightarrow \infty$$

Conclusion

La qualité de l'information ne dépend pas des résultats de l'expérience mais de la position des points expérimentaux donc des conditions expérimentales !

✚ Mon premier plan d'expérience

Exemple : Rendement d'une réaction chimique

Un chimiste veut améliorer le rendement d'une réaction chimique, il décide d'étudier deux facteurs pouvant, à son avis, influencer l'évolution de sa réaction.

Objectif : déterminer la masse du produit de la réaction (g),

Facteurs : les facteurs choisis sont la température et la pression, on suppose qu'ils sont dépendants. Les niveaux de variation de ces facteurs sont donnés par :

Facteurs	Symbole	Niveaux	
		(-1)	(+1)
Température (°C)	X1	60	80
Pression (bar)	X2	1	2

Domaine expérimental : un plan factoriel complet a été choisi, le nombre d'expérience à réaliser est de 4.

Matrice d'expérience :

N°	X1	X2
1	-1	-1
2	+1	-1
3	-1	+1
4	+1	+1

Le modèle postulé : un polynôme du premier degré sans interaction donné par

$$Y = b_0 + b_1X_1 + b_2X_2 + b_{12}X_1X_2$$

Y est la réponse (la masse du produit),

X représente le niveau attribué au facteur i par l'expérimentateur pour réaliser un essai,

b_0 la valeur de la réponse au centre du domaine (la moyenne des réponses),

b_i, b_{ij} sont les coefficients du modèle mathématique adopté.

Expérimentation : Les expériences doivent être faites selon l'ordre donné par la matrice, le plan d'expérience correspondant est le suivant :

N°	T (°C)	P(bar)
1	60	1
2	80	1
3	60	2
4	80	2

Après expérimentation, on trouve les résultats expérimentaux suivants :

$$y = \{60, 70, 80, 95\}.$$

L'objectif étant de calculer les coefficients du modèle appelés « effet », pour cela, on écrit la matrice des effets en insérant la colonne X0 qui nous permettra de calculer la moyenne des réponses et les résultats de l'expérimentation.

Matrice des effets :

N°	X0	X1	X2	X1X2	Y(g)
1	+1	-1	-1	+1	60
2	+1	+1	-1	-1	70
3	+1	-1	+1	-1	80
4	+1	+1	+1	+1	95
b _j	b ₀	b ₁	b ₂	b ₁₂	

Le calcul des coefficients du modèle se fait de deux manières :

1. Méthode analytique

A partir de la matrice des effets, on peut écrire :

$$\begin{cases} y_1 = b_0 - b_1X_1 - b_2X_2 + b_{12}X_1X_2 \\ y_2 = b_0 + b_1X_1 - b_2X_2 - b_{12}X_1X_2 \\ y_3 = b_0 - b_1X_1 + b_2X_2 - b_{12}X_1X_2 \\ y_4 = b_0 + b_1X_1 + b_2X_2 + b_{12}X_1X_2 \end{cases}$$

La combinaison de ces équations nous permet de déduire la valeur des coefficients :

$$b_0 = \frac{(+1)(60) + (+1)(70) + ((+1)(80) + (+1)(95))}{4} = 76.25$$

$$b_1 = \frac{(-1)(60) + (+1)(70) + ((-1)(80) + (+1)(95))}{4} = 6.25$$

$$b_2 = \frac{(-1)(60) + (-1)(70) + ((+1)(80) + (+1)(95))}{4} = 11.25$$

$$b_{12} = \frac{(+1)(60) + (-1)(70) + ((-1)(80) + (+1)(95))}{4} = 1.25$$

La réponse en tout point du domaine expérimental est donnée par :

$$Y = 76.25 + 6.25X_1 + 11.25X_2 + 1.25 X_1X_2$$

2. Méthode matricielle

Cette méthode permet le calcul des coefficients simultanément. Le modèle choisi peut être écrit sous la forme matricielle :

$$Y = X \cdot b$$

Y : Vecteur réponse,

X : Matrice des effets,

b : vecteur effet.

Pour résoudre ce système, on utilise une méthode des moindres carrés. On obtient ainsi les estimations des coefficients que l'on note :

$$Y = X \cdot b \rightarrow X^{-1}Y = X^{-1}X \cdot b = b$$

$$\mathbf{b} = \mathbf{X}^{-1}\mathbf{Y}$$

X^{-1} : matrice inverse de la matrice X, si le déterminant de la matrice X est différent de zéro :

$$X^{-1} = \frac{1}{n}X^T$$

X^T : la matrice transposée de la matrice X.

n : ordre de la matrice (n=4).

En remplaçant X^{-1} dans l'équation de b, on trouvera :

$$\mathbf{b} = \mathbf{X}^{-1}\mathbf{Y} = \frac{1}{n}\mathbf{X}^T \cdot \mathbf{Y}$$

Donc : $\mathbf{b} = \frac{1}{4}\mathbf{X}^T \cdot \mathbf{Y}$

$$b = \frac{1}{4} \cdot \begin{vmatrix} +1 & +1 & +1 & +1 \\ -1 & +1 & +1 & +1 \\ -1 & -1 & +1 & +1 \\ +1 & -1 & -1 & +1 \end{vmatrix} \cdot \begin{vmatrix} 60 \\ 70 \\ 80 \\ 95 \end{vmatrix} = \frac{1}{4} \cdot \begin{vmatrix} 305 \\ 25 \\ 45 \\ 5 \end{vmatrix} = \begin{vmatrix} 76.25 \\ 6.25 \\ 11.25 \\ 1.25 \end{vmatrix}$$

La réponse en tout point du domaine expérimental est donnée par :

$$Y = 76.25 + 6.25X_1 + 11.25X_2 + 1.25 X_1X_2$$

Chapitre III

Criblage de facteurs

Chapitre III : Criblage de facteurs

I. Introduction

Un criblage de facteurs est une méthode qui nous permet de sélectionner parmi l'ensemble des facteurs jugés influents, ceux qui le sont réellement, pour cela, on utilise les matrices d'Hadamard. L'étape de criblage est une étape importante, elle nous permet de tirer des conclusions rapides.

❖ Hypothèses

- Les facteurs sont indépendants entre eux,

L'hypothèse d'additivité est une hypothèse forte. Donc nous posons implicitement que le modèle est un polynôme du 1^{er} degré. Tous les autres coefficients de degré supérieur sont nuls :

$$y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k$$

b_j sont les coefficients du modèle calculés avec la méthode linéaire ou la méthode matricielle.

❖ Avantages

- Un minimum d'expérience pour l'étude d'un grand nombre de facteurs,
- Les effets des facteurs sont comparables,
- Les facteurs peuvent être qualitatifs ou quantitatifs.

❖ Inconvénients

- Les interactions sont supposées négligeables avec risque,
- Chaque facteur ne prend que deux niveaux.

II. Criblage de facteurs- matrice d'Hadamard

Les plans de criblage sont souvent une première étape très efficace pour une étude complète. Les matrices d'Hadamard sont les matrices les plus utilisées pour un criblage des facteurs. Dans ces matrices, les facteurs sont indépendants et ne prennent généralement que deux niveaux distincts notés (-1) et (+1) en variables codées.

Dans les plans de criblage, il faut estimer "le poids" de chaque facteur. La variance de l'estimation de chaque facteur est donnée par la relation :

$$\text{Var}(b_j) = \frac{\sigma^2}{N}$$

Où :

σ^2 est la variance des réponses,

N le nombre d'expériences à réaliser, b_j l'estimation du poids du facteur j.

Pour k facteurs à deux niveaux, le nombre d'expériences N à réaliser doit satisfaire la condition :

$$\begin{cases} N \geq k + 1 \\ N \text{ est un multiple de } 4 \end{cases}$$

III. Construction de la matrice d'Hadamard

L'algorithme de la construction de la matrice de Hadamard à k facteurs et N expériences est le suivant :

- 1- Ecrire la première ligne (voir tableau III.1.),
- 2- Les autres lignes ou colonnes sont générées par permutation circulaire à droite ou à gauche ou vers le haut ou vers le bas,
- 3- S'arrête à la (N-1) i^{ème} ligne, la dernière ligne ne comporte que des (-1),
- 5- Supprimer les colonnes superflues.

Les premières lignes des matrices de Hadamard sont données par le tableau suivant :

Tableau III.1. Première ligne de la matrice de Hadamard

Nombre de facteurs	Nombre d'expériences	Ligne de départ
$K \leq 3$	4	++-
$4 \leq k \leq 7$	8	+++--
$8 \leq k \leq 11$	12	++-+++--
$12 \leq k \leq 15$	16	++++-+-+---
$16 \leq k \leq 19$	20	++--++++-+-+---
$20 \leq k \leq 23$	24	+++++-+---+---+---

Exemple

Pour trois facteurs ($k=3$), le nombre d'expérience à réaliser est de quatre. En adoptant la permutation de droite à gauche, on aura :

	X1	X2	X3
1	+	+	-
2	-	+	+
3	+	-	+
4	-	-	-

Remarques

1. Une matrice d'Hadamard est orthogonale, une matrice d'expérience X est dite orthogonale si le produit $(X.X^T) = NI$, (X^T étant la matrice transposée et I est la matrice identité).
2. La matrice d'Hadamard est une matrice optimale car la variance est minimale,
3. Dans les matrices d'Hadamard, Le poids d'un facteur est le double de la valeur de son coefficient,
4. Avant de calculer les coefficients du modèle, on compare toujours les valeurs minimale et maximale de la réponse. Si l'écart est important, il y aura au moins un facteur influent.
5. Dans les plans de criblage, la variance expérimentale peut être calculée de différentes manières :
 - La méthode du manque d'ajustement,
 - En faisant des expériences N_0 au centre du domaine expérimentale ($N_0 > k+1$),
 - Doubler le plan d'expérience,

IV. Estimation des facteurs influents

Les calculs statistiques qui permettent de savoir si les effets sont significatifs et de valider un modèle font intervenir d'une part les résidus e_i , c'est-à-dire la différence entre la valeur expérimentale et la valeur prédite par le modèle et, d'autre part un estimateur sans biais de la variance commune des résidus. Cet estimateur est donné par la relation :

$$S^2 = \frac{1}{N-p} \sum e_i^2$$

Où N est le nombre d'expériences à réaliser,

p le nombre de coefficients du modèle incluant la constante du modèle.

Dans ces conditions, on peut montrer que tous les effets ont la même variance donnée par :

$$\sigma_i^2 = \frac{S^2}{N}$$

Pour identifier les effets principaux impactant la/les réponse(s) observée(s), il faut calculer l'écart type σ et le t de Student (t_{exp}) pour chaque coefficient.

Le t_{exp} est donné par la relation :

$$t_{exp} = \frac{|b_j|}{\sigma}$$

Connaissant le nombre de degré de liberté (ddl = N-p), on détermine le $t_{critique}$ en utilisant la table de Student en choisissant un risque α le plus souvent de 5%.

Nous considérons que si

$$t_{exp} \geq t_{critique} \rightarrow \text{l'effet est significatif}$$

$$t_{exp} < t_{critique} \rightarrow \text{l'effet est non significatif}$$

Remarques

➤ Un seuil de signification α de 5% indique que 5% de risque à conclure à tort que le modèle explique la variation de la réponse,

➤ α est seuil de signification et $(1-\alpha)$ est appelé niveau de confiance,

➤ On peut aussi utiliser la probabilité (p value) pour mesurer le degré de certitude.

Pour déterminer les effets influents, la valeur de P est comparée au seuil de signification α :

Si $P \leq$ seuil de signification α : Le modèle explique la variation de la réponse,

Si $P >$ seuil de signification α : Le modèle n'explique pas la variation de la réponse, il faut le réajuster.

Exercice d'application

Un procédé pharmaceutique d'extrusion –sphéronisation se propose de fabriquer des billes dont la granulométrie doit être comprise entre 900 et 1100 μm . Ce procédé contient trois parties :

- Granulométrie,
- Extrusion,
- Sphéronisation.

Les responsables de l'unité décident de réaliser un criblage. Les facteurs étudiés et leur domaine de variation sont donnés dans le tableau suivant :

Facteur	Niveau bas	Niveau haut
Quantité de liant (%)	0.5	1
Quantité d'eau (%)	40	50
Durée de granulation (min)	1	2
Charge à sphéroniser (Kg)	1	4
Vitesse de sphéronisation (tr/mn)	700	1100
Durée de sphéronisation (min)	2	5

1- Quel est le nombre d'expérience à réaliser, Construire la matrice d'Hadamard en variables codées,

2- Après expérimentation, on trouve :

Y (%) : (55.9 ; 51.7 ; 78.1 ; 61.9 ; 76.1 ; 59.1 ; 50.8 ; 62.1).

Calculer les coefficients b_j , déduire le modèle mathématique puis retrouver les facteurs influents ($t_{\text{critique}} = 1.00$).

3- Tracer qualitativement le diagramme des effets.

Corrigé

- 1- Pour k facteurs à deux niveaux, le nombre d'expériences N à réaliser doit satisfaire la condition :

$$\begin{cases} N \geq k + 1 \\ N \text{ est un multiple de } 4 \end{cases}$$

$$k = 6 \rightarrow N = 8$$

Après avoir éliminé les colonnes superflues, la matrice d'Hadamard écrite par permutation de droite à gauche est donnée par :

N	X0	X1	X2	X3	X4	X5	X6
1	1	1	1	1	-1	1	-1
2	1	-1	1	1	1	-1	1
3	1	-1	-1	1	1	1	-1
4	1	1	-1	-1	1	1	1
5	1	-1	1	-1	-1	1	1
6	1	1	-1	1	-1	-1	1
7	1	1	1	-1	1	-1	-1
8	1	-1	-1	-1	-1	-1	-1

- 2- Pour calculer les coefficients du modèle postulé, nous devons d'abord écrire la matrice des effets. Le calcul a été fait avec la méthode analytique, les résultats obtenus seront notés en bas de la matrice.

N°	X0	X1	X2	X3	X4	X5	X6	Yexp
1	1	1	1	1	-1	1	-1	55,900
2	1	-1	1	1	1	-1	1	51,700
3	1	-1	-1	1	1	1	-1	78,100
4	1	1	-1	-1	1	1	1	61,900
5	1	-1	1	-1	-1	1	1	76,100
6	1	1	-1	1	-1	-1	1	59,100
7	1	1	1	-1	1	-1	-1	50,800
8	1	-1	-1	-1	-1	-1	-1	62,100
bj	61,963	-5,038	-3,338	-0,763	-1,338	6,038	0,238	

Le modèle est un polynôme du 1^{er} degré donné par :

$$y = 61.96 - 5.038 X1 - 3.338 X2 - 0.763X3 - 1.338X4 + 6.038X5 + 0.238X6$$

3- Les calculs statistiques qui permettent de savoir les effets significatifs font intervenir d'une part les résidus e_i , la variance expérimentale et l'écart type. Nous avons utilisé la méthode du manque d'ajustement. L'ensemble des résultats est regroupé dans le tableau suivant :

4-

N	X0	X1	X2	X3	X4	X5	X6	Yexp	Ythe	e^2
1	1	1	1	1	-1	1	-1	55,90	59,962	16,500
2	1	-1	1	1	1	-1	1	51,70	55,762	16,500
3	1	-1	-1	1	1	1	-1	78,10	74,038	16,500
4	1	1	-1	-1	1	1	1	61,90	65,964	16,516
5	1	-1	1	-1	-1	1	1	76,10	72,040	16,484
6	1	1	-1	1	-1	-1	1	59,10	55,038	16,500
7	1	1	1	-1	1	-1	-1	50,80	46,736	16,516
8	1	-1	-1	-1	-1	-1	-1	62,10	66,164	16,516
bj	61,963	-5,038	-3,338	-0,763	-1,338	6,038	0,238		Somme	132,031
									S^2	132,031
									σ^2	16,504
									σ	4.063

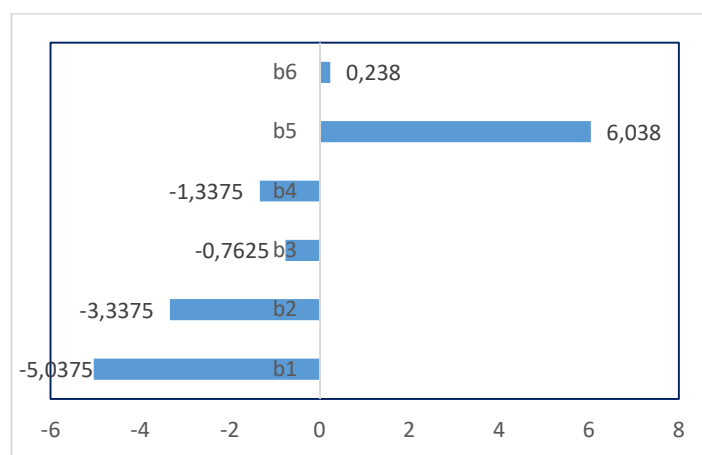


Diagramme des effets

Pour identifier les effets principaux impactant la réponse, il faut calculer le t de Student (t_{exp}) pour chaque coefficient, Le t_{exp} est donné par la relation :

$$t_{exp} = \frac{|b_j|}{\sigma}$$

Connaissant le nombre de degré de liberté (ddl = N-p =1), on détermine le $t_{critique}$ en utilisant la table de Student en choisissant un risque α de 1%. Les résultats sont résumés dans le tableau suivant :

Facteur	b_j	écart type	t_{exp}	t_{crit}	Conclusion
Quantité de liant (%)	5,038	4,06	1,24	1,000	Influent
Quantité d'eau (%)	3,338	4,06	0,82	1,000	Non influent
Durée de granulation (min)	0,763	4,06	0,19	1,000	Non influent
Charge à sphéroniser (Kg)	1,338	4,06	0,33	1,000	Non influent
Vitesse de sphéronisation (tr/mn)	6,038	4,06	1,49	1,000	Influent
Durée de sphéronisation (min)	0,238	4,06	0,06	1,000	Non influent

V. Identification des facteurs influents en utilisant les outils d'aide à l'interprétation

Pour identifier les effets principaux impactant la réponse observée lorsque nombre de degré de liberté est élevé, nous pouvons utiliser les outils d'aide à l'interprétation. Dans ces outils, on considère que l'erreur expérimentale est une variable aléatoire distribuée selon une loi normale.

1. Approche de Pareto

Pareto a montré que dans la majorité des situations, un petit nombre de facteurs a une influence majeure sur les résultats. C'est la loi dite de Pareto des 80-20, où, 20% des facteurs expliquent 80% des résultats (20 % des étudiants demandent 80% de l'effort d'un enseignant, 20% de la population mondiale a accès à 80% de l'eau potable, 20% de la population paie 80% des impôts, 20% des problèmes représentent 80% des préoccupations, 20% des pays émettent 80% des gaz à effet de serre...etc).

Pour identifier les effets principaux et les interactions entre les facteurs, nous devons :

- Ecrire la matrice d'Hadamard en adoptant la permutation à droite,
- Calculer les coefficients et le carré des coefficients b_j et b_j^2 ,
- Classer les valeurs par ordre décroissant (Z à A),
- Calculer $\sum b_j^2$,
- Calculer l'indice de Pareto (PI) donné par la relation :

$$PI = 100 * \frac{b_i^2}{\sum b_j^2}$$

- Calculer l'indice de Pareto cumulé (PIc).
- Tracer les histogrammes des b_j en fonction de PIc.
- Déduire les facteurs influents.

2. Approche de Lenth

Lenth a mis au point une méthode qui consiste à estimer le pseudo écart-type des estimations des coefficients calculés. Sous cette hypothèse Lenth démontre qu'il y a un lien entre la médiane des valeurs absolues des b_j et l'écart-type :

$$S_0 = 1.5 * \text{médiane} |b_j|$$

S_0 est appelé « pseudo écart-type » noté PSE,

$|b_j|$ représente la valeur absolue des b_j .

Le PSE correspond à la valeur de S_0 trouvée à la dernière itération.

Pour identifier les effets principaux par l'approche de Lenth, nous devons :

- Ecrire la matrice d'Hadamard en adoptant la permutation à droite,
- Calculer les coefficients b_j ,
- Calculer $Abs|b_j|$ puis classer les valeurs par ordre croissant (A à Z),
- Calculer la médiane,
- Calculer l'écart type S_0 ,
- Calculer $2.5*S_0$,
- Eliminer les b_j supérieurs à $(2.5*S_0)$ puis reprendre les calculs.
- Le PSE correspond à la valeur de S_0 trouvée à la dernière itération,

Calculer les limites : Les limites de signification sont calculées à l'aide de la relation :

$$limites = \pm t_{\alpha,d} * PSE$$

α = seuil de signification qui est généralement de 5%,

d = nombre d'effets restants/3.

3. Normal Plot ou Half Normal Plot

Comme dans l'approche de Lenth, on considère que tous les b_j sont distribués selon une loi normale autour de zéro. Les graphes Normal Plot et Half normal Plot permettent de tester la répartition des valeurs de ces effets.

La construction du Normal Plot se fait de la manière suivante :

- Ecrire la matrice d'Hadamard en adoptant la permutation à droite,
- On ordonne les valeurs des effets dans un ordre croissant (A à Z) en affectant à chaque effet son rang j ($j=1,2,\dots,k$). La valeur des effets constitue l'abscisse du graphe,
- On calcule la probabilité correspondant au rang j à l'aide de la relation :

$$P = \frac{j - 0.5}{k}$$

- On trace la probabilité en fonction de b_j . Les points qui s'écartent de la droite correspondent aux effets influents.

Pour la construction du Half Normal Plot (HNP), on adopte la même démarche mais le classement se fait avec la valeur absolue des valeurs des effets.

Application

L'étude de la robustesse d'un procédé de préparation de sulfate amide des acides gras de l'huile de grignons d'olive a été étudiée. Ce procédé est réalisé en plusieurs étapes : la saponification, l'hydrolyse des savons obtenus, l'estérification des acides gras, l'amidation et la sulfatation des diéthanolamines des acides gras et enfin une neutralisation avec une solution de KOH.

Nous avons appliqué un plan de criblage du type Plackett et Burman pour distinguer parmi onze facteurs jugés potentiellement influents ceux qui le sont réellement. Ces facteurs ainsi que leurs niveaux bas et haut sont donnés par le tableau II :

	Facteurs	Niveau (-1)	Niveau (+1)
X1	Durée d'hydrolyse (h)	1	1,5
X2	Température d'estérification (°C)	115	125
X3	Durée d'estérification (h)	3	5
X4	Température d'amidation (°C)	120	140
X5	Durée d'amidation (h)	10	12
X6	Méthanoate de sodium (%)	0,25	0,5
X7	Amine /Ester	1,2	1,6
X8	Réactif sulfatation	oléum	acide chlorosulfonique
X9	SO3/Ester	1,5	2
X10	Température de sulfatation (°C)	5	15
X11	Durée de sulfatation (h)	2	3

Les rendements de la réaction globale de préparation des sulfate-amides sont :

$$y = \{76, 80, 106, 113, 140, 86, 92, 134, 96, 88, 91, 73\}.$$

En utilisant les trois méthodes précédentes, quels sont les facteurs influents ?

Corrigé

Le nombre d'expérience à réaliser est de 12, la matrice d'Hadamard en variables codées et le Calcul des coefficients b_j sont regroupés dans le tableau suivant :

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	Y
1	1	1	1	-1	1	1	1	-1	-1	-1	1	-1	76
2	1	-1	1	1	-1	1	1	1	-1	-1	-1	1	80
3	1	1	-1	1	1	-1	1	1	1	-1	-1	-1	106
4	1	-1	1	-1	1	1	-1	1	1	1	-1	-1	113
5	1	-1	-1	1	-1	1	1	-1	1	1	1	-1	140
6	1	-1	-1	-1	1	-1	1	1	-1	1	1	1	86
7	1	1	-1	-1	-1	1	-1	1	1	-1	1	1	92
8	1	1	1	-1	-1	-1	1	-1	1	1	-1	1	134
9	1	1	1	1	-1	-1	-1	1	-1	1	1	-1	96
10	1	-1	1	1	1	-1	-1	-1	1	-1	1	1	88
11	1	1	-1	1	1	1	-1	-1	-1	1	-1	1	91
12	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	73
b_j	97,92	1,25	-0,08	2,25	-4,58	0,75	5,75	-2,42	14,25	12,08	-1,58	-2,75	

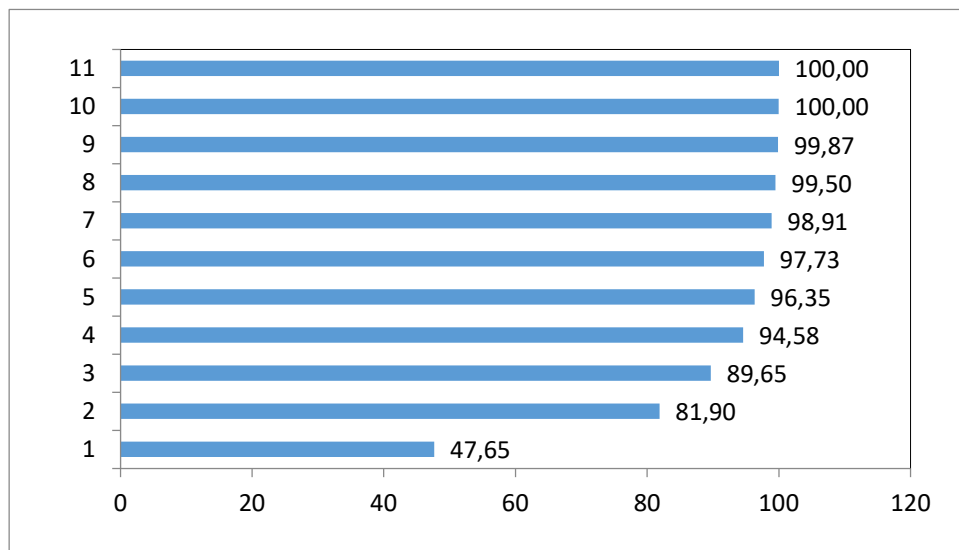
Les facteurs influents en utilisant les outils d'aide à l'interprétation :

➤ **Approche de Pareto**

En respectant la démarche donnée précédemment, on trouve :

Nom	b_j (tri)	b_j^2	PI	PI _{Cumulé}
b8	14,25	203,0625	47,652	47,652
b9	12,08	145,9264	34,244	81,896
b6	5,75	33,0625	7,759	89,655
b4	-4,58	20,9764	4,922	94,577
b11	-2,75	7,5625	1,775	96,352
b7	-2,42	5,8564	1,374	97,726
b3	2,25	5,0625	1,188	98,914
b10	-1,58	2,4964	0,586	99,500
b1	1,25	1,5625	0,367	99,867
b5	0,75	0,5625	0,132	99,999
b2	-0,08	0,0064	0,002	100,000
	Somme b_j^2	426,137		

Selon Pareto, seul le facteur b8 est considéré influent.



Représentation de Pareto

➤ Approche de Lenth

Nom	b_j	Abs b _j + (tri A à Z)	
b2	-0,08	0,08	0,08
b5	0,75	0,75	0,75
b1	1,25	1,25	1,25
b10	-1,58	1,58	1,58
b3	2,25	2,25	2,25
b7	-2,42	2,42	2,42
b11	-2,75	2,75	2,75
b4	-4,58	4,58	4,58
b6	5,75	5,75	5,75
b9	12,08	12,08	
b8	14,25	14,25	
	Médiane	2,42	2,25
	S₀	3,63	3,375
	2.5*S₀	9,075	8,4375

A partir des itérations faites, on conclut :

PSE=3.38

ddl=3

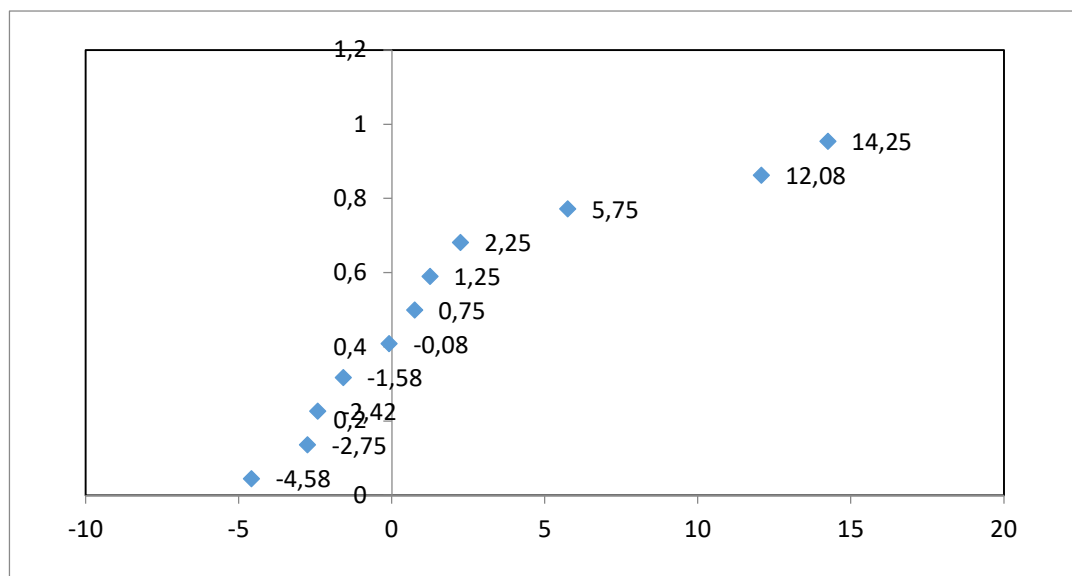
t=3.18

$$\text{Limites} = \pm PSE \cdot t = \pm 10.76$$

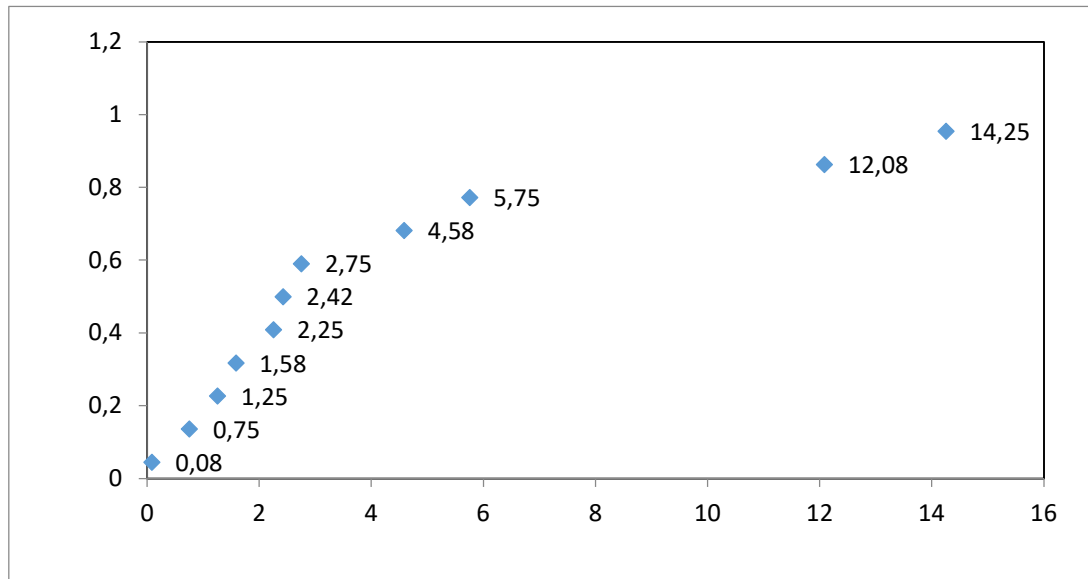
Les facteurs jugés influents sont : b8 et b9

➤ **Normal Plot (NP) et Half Normal Plot (HNP)**

Nom	bj (tri)	Rang	P
b4	-4,58	1	0,045
b11	-2,75	2	0,136
b7	-2,42	3	0,227
b10	-1,58	4	0,318
b2	-0,08	5	0,409
b5	0,75	6	0,500
b1	1,25	7	0,591
b3	2,25	8	0,682
b6	5,75	9	0,773
b9	12,08	10	0,864
b8	14,25	11	0,955



Normal Plot



Half Normal Plot

Les points b8 et b9 s'écartent de la droite, ils sont jugés influents.

Chapitre IV

Etude quantitative des facteurs

Chapitre IV : Etude quantitative des facteurs

L'expérimentation coûte cher, il faut avoir pour objectif d'obtenir les informations les plus fiables et les plus efficaces possibles en un minimum d'essais : l'expérimentation doit être optimisée d'où la nécessité d'une recherche expérimentale bien planifiée.

Un plan d'expériences efficace propose de faire varier plusieurs facteurs à la fois selon des règles d'organisation précises et rigoureuses : les plans que nous allons étudier sont multifactoriels et sont appelés habituellement plans factoriels.

Lors de l'étude d'un phénomène, plusieurs questions se posent, auxquelles répondent différents types de plans. On peut distinguer 3 grandes étapes dans l'acquisition des connaissances

- *Recherche des facteurs influents*,
- *Modélisations* : quand les facteurs influents ont été identifiés et leur importance quantifiée, on recherche ensuite l'équation permettant de décrire les variations de la réponse étudiée en fonction de celles des facteurs influents,
- *Optimisation* : déterminer quelles conditions expérimentales (les valeurs prises par les facteurs influents) permettent d'obtenir le meilleur résultat pour la réponse.

Parmi tous les facteurs susceptibles d'influer sur le phénomène (c'est à dire sur la réponse mesurée du phénomène),

- Lequel des facteurs a une influence significative ?
- Que vaut quantitativement cette influence ?
- Existe-t-il des interactions entre ces facteurs ?

Les plans qui permettent de rechercher les facteurs influents sont :

- Les plans factoriels complets,
- Les plans fractionnaires.

Première partie : Plan factoriel complet

I. Définition

Un plan est dit factoriel complet, noté (2^k) , si tous les niveaux de tous les facteurs sont programmés pour varier simultanément et de manière équilibrée dans l'ensemble de l'expérimentation. Ces plans sont à utiliser pour rechercher les facteurs agissant sur une réponse mesurée. Ce sont les plus simples à interpréter et ils présentent le meilleur rapport coût/efficacité.

- Ces plans peuvent être utilisés pour les variables continues et les variables discrètes.
- Dans ce cas, les facteurs ne sont pas indépendants, l'effet de chaque facteur dépend du niveau de l'autre.
- Au-delà de 5 facteurs, il faut passer aux plans fractionnaires.

Les plans factoriels permettent de mettre en évidence parmi les facteurs étudiés, ceux qui ont une influence sur la réponse, ainsi que d'éventuelles interactions entre les facteurs. La grandeur de cette influence est appelée **effet du facteur** ou de l'interaction.

Un effet positif signifie que la valeur de la réponse augmente quand le facteur passe du niveau bas au niveau haut ; un effet négatif signifie que la réponse diminue quand on passe du niveau bas au niveau haut.

Un facteur A est d'autant plus influent que la valeur absolue de son effet est plus grande, il est plus influent qu'un facteur B si la valeur absolue de son effet est plus grande.

II. Construction de la matrice

Il est commode de symboliser par (-1) le niveau bas et par (+1) le niveau haut de chaque facteur. Ce qui permet de rassembler les éléments relatifs à chaque facteur dans une matrice d'expérience que l'on présente en correspondance avec une colonne donnant les résultats expérimentaux de la réponse y.

En général, la matrice d'expérience comporte **k** colonnes pour les facteurs principaux et 2^k lignes correspondantes à **k essais**. La construction se fait avec l'algorithme de YATES.

Colonne 1 du premier facteur : alternance -1 et +1,

Colonne 2 du deuxième facteur : alternance -1 et +1 de 2 en 2,

Colonne 3 du troisième facteur : alternance -1 et +1 de 4 en 4,

Colonne 4 du quatrième facteur : alternance -1 et +1 de 8 en 8.

III. Modèle interactif

L'équation du modèle dans le cas des plans complet est donnée par :

$$y = b_0 + \sum_{i=1}^k b_i X_i + \sum_{i,j=1, i \neq j}^k b_{ij} X_i X_j$$

y est la réponse, X représente le niveau attribué au facteur i ,

b_0 : moyenne des réponses.

b_i : Effet principal du facteur X_i .

b_{ij} : Effet d'interaction entre les variables X_i et X_j .

Les effets principaux et les interactions ne sont pas connus. Ils doivent être calculés à partir des résultats des expériences.

• Matrice des effets ou matrice du modèle

Pour calculer les coefficients du modèle, on utilisera la matrice des effets donnée dans le tableau suivant (pour $k=2$) :

	X0	X1	X2	X1X2	Y
1	+1	-1	-1	+1	Y1
2	+1	+1	-1	-1	Y2
3	+1	-1	+1	-1	Y3
4	+1	+1	+1	+1	Y4

Comme nous l'avons déjà vu (chapitre 2), les réponses y_i permettent de calculer les effets comme suit :

$$E_0 = b_0 = \frac{1}{4}(y_1 + y_2 + y_3 + y_4)$$

$$E_A = E_1 = b_1 = \frac{1}{4}(-y_1 + y_2 - y_3 + y_4)$$

$$E_B = E_2 = b_2 = \frac{1}{4}(-y_1 - y_2 + y_3 + y_4)$$

$$E_{AB} = E_{12} = b_{12} = \frac{1}{4}(+y_1 - y_2 - y_3 + y_4)$$

1^{er} cas : Modèle sans interaction

Nous avons choisi l'exemple de l'étude de l'influence de la température et de la concentration d'un réactif sur le rendement d'une réaction chimique. Le domaine de variation de la température (θ) est entre 60°C et 80°C et celui de la concentration est entre 10 g/L et 15 g/L.

Le nombre d'expérience à réaliser est $2^k = 2^2 = 4$ expériences.

Les résultats expérimentaux (rendement de la réaction (%)) obtenus sont :

$$y = \{60; 70; 80; 90\}$$

- **Matrice d'expérience**

La matrice d'expérience et les résultats obtenus sont :

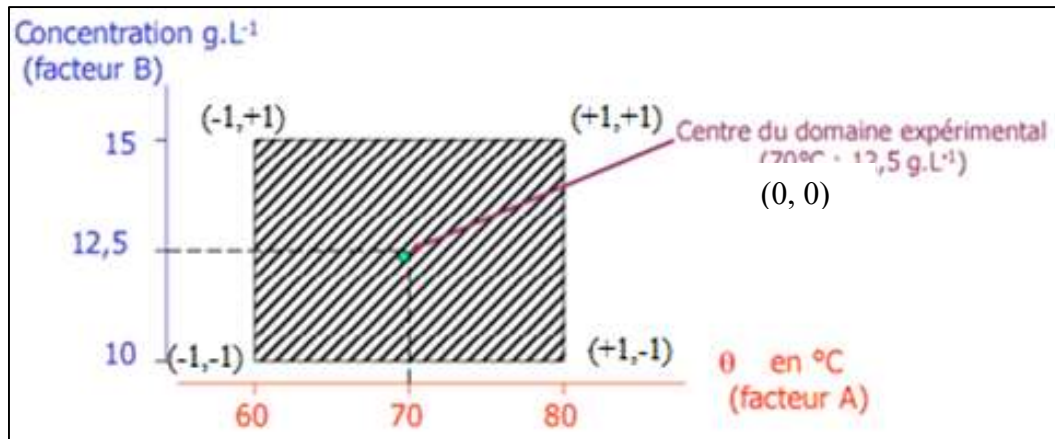
N°	X1	X2
1	-1	-1
2	+1	-1
3	-1	+1
4	+1	+1

- **Matrice des effets**

Dans la matrice des effets, il faut faire apparaître la colonne X0 et la réponse Y,

N°	X0	X1	X2	Y
1	+1	-1	-1	60
2	+1	+1	-1	70
3	+1	-1	+1	80
4	+1	+1	+1	90

- **Domaine expérimental**



Nous avons deux facteurs, le domaine expérimental est un carré. Les quatre expériences sont disposées sur le sommet dans un système d'axe à variables codées.

Les facteurs codés sont dits **facteurs centrés réduits** et notés X :

- Facteur A : dans le cas de la température (θ) :

$X_A = -1$ correspond à une température de 60°C ,

$X_B = +1$ correspond à une température de 80°C .

Le centre du domaine expérimental : $X=0$ est donné par $\frac{60+80}{2} = 70^\circ\text{C}$

- Facteur B : dans le cas de la concentration (conc):

$X_A = -1$ correspond à une concentration de 10 g.L^{-1} ,

$X_B = +1$ correspond à une concentration de 15 g.L^{-1} .

Le centre du domaine expérimental : $X=0$ est donné par $\frac{10+15}{2} = 12.5 \text{ g.L}^{-1}$

- **Effet global et effet moyen d'un facteur**

- Effet global

Il correspond à la variation d'un facteur lorsque celui-ci passe d'un niveau bas (-1) au niveau haut (+1). L'effet peut être évalué de deux façons :

θ re	Conc B	Réponse %
A		
-1	-1	$y_1 = 60$
+1	-1	$y_2 = 70$
-1	+1	$y_3 = 80$
+1	+1	$y_4 = 90$

- température (A)		
à 10 g.L ⁻¹	effet global = $y_2 - y_1 = 70 - 60 = 10\%$	
à 15 g.L ⁻¹	effet global = $y_4 - y_3 = 90 - 80 = 10\%$	
- concentration (B)		
à 60°C	effet global = $y_3 - y_1 = 80 - 60 = 20\%$	
à 80°C	effet global = $y_4 - y_2 = 90 - 70 = 20\%$	

Dans ce cas, l'effet d'un facteur est exactement le même quel que soit le niveau de l'autre, les deux facteurs agissent indépendamment. On dit qu'il n'y a pas d'interaction entre la température et la concentration.

$$\text{L'effet global de la température} = \frac{1}{2}(-y_1 + y_2 - y_3 + y_4) = 10$$

$$\text{L'effet global de la concentration} = \frac{1}{2}(-y_1 - y_2 + y_3 + y_4) = 20$$

- Effet moyen

- L'effet global correspond à une variation du facteur centré réduit de 2 unités (de -1 à +1),
 - L'effet moyen correspond à une variation de 1 unité du facteur centré réduit (de 0 à 1 par exemple), il correspond à la moitié de l'effet global.

$$\text{L'effet moyen de la température : } E_A = E_1 = b_1 = \frac{1}{4}(-y_1 + y_2 - y_3 + y_4) = 5$$

$$\text{L'effet moyen de la concentration : } E_B = E_2 = b_2 = \frac{1}{4}(-y_1 - y_2 + y_3 + y_4) = 10$$

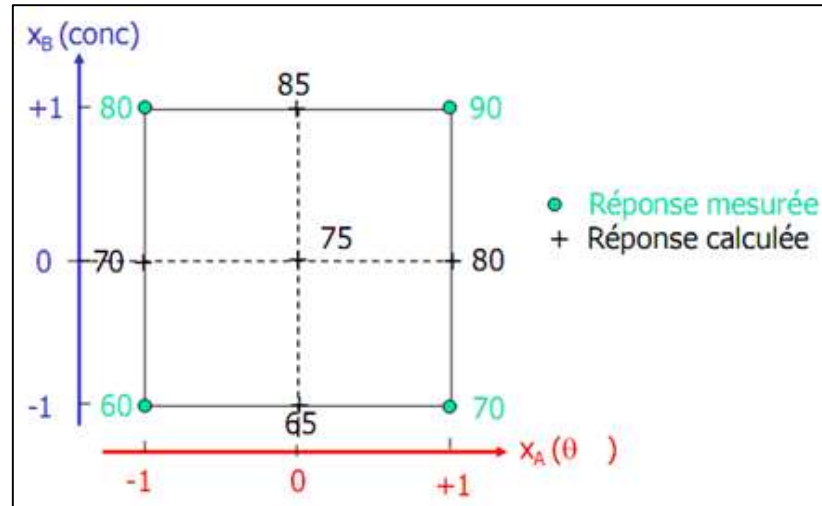
$$\text{La moyenne des essais s'écrit : } \bar{y} = b_0 = \frac{1}{4}(y_1 + y_2 + y_3 + y_4) = 75$$

b_0 peut être considérée comme la valeur calculée de la réponse quand les facteurs centrés réduits sont au niveau 0, c'est à dire au centre du domaine expérimental.

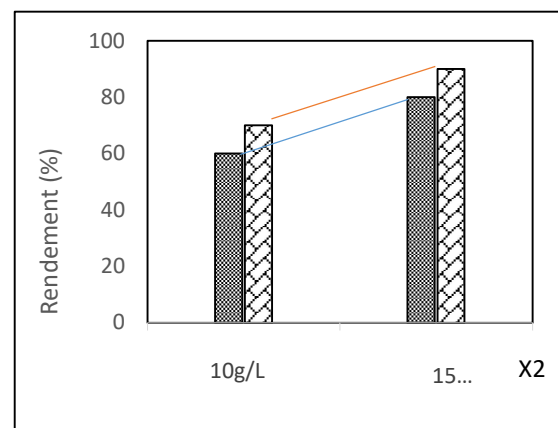
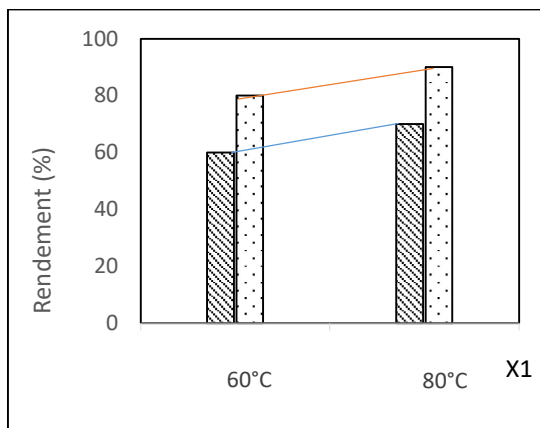
- **Représentations graphiques**

Pour visualiser les résultats on peut :

1. Reporter les réponses mesurées et calculées sur le domaine expérimental :



2. Reporter les réponses moyennes calculées en fonction du facteur :



Les graphiques présentés montrent les effets des facteurs A et B, ayant chacun deux niveaux. On peut observer que l'effet du facteur A ne dépend pas des niveaux de B ; en effet les droites correspondantes sont parallèles. De la même façon l'effet du facteur B ne dépend pas des niveaux de A.

Cela signifie qu'il n'y a aucune interaction entre les facteurs A et B. On dit alors que les facteurs A et B sont indépendants.

2^{ème} cas : Modèle avec interaction

On reprend le même exercice mais en présence d'un catalyseur, les conditions expérimentales sont identiques mais les résultats diffèrent.

- **Matrice d'expérience**

La matrice d'expérience et les résultats obtenus sont :

	X1	X2	Y
1	-1	-1	60
2	+1	-1	70
3	-1	+1	80
4	+1	+1	95

- **Matrice des effets**

La colonne X1X2 s'obtient en multipliant les colonnes X1 et X2 :

	X0	X1	X2	X1X2	Y
1	+1	-1	-1	+1	60
2	+1	+1	-1	-1	70
3	+1	-1	+1	-1	80
4	+1	+1	+1	+1	95
Coefficient	76,25	6,25	11,25	1,25	

La présence du catalyseur est favorable pour cette réaction.

Qu'a-t-il changé dans les effets ?

- Effet de la température

$$\text{A } 10\text{g. L}^{-1}: E_{\theta} = \frac{1}{2}(70 - 60) = 5$$

$$\text{A } 15\text{g. L}^{-1}: E_{\theta} = \frac{1}{2}(95 - 80) = 7.5$$

- Effet de la concentration

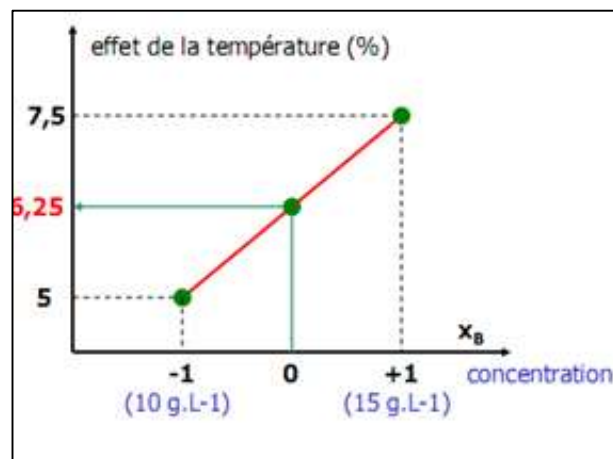
$$\text{A } 60^{\circ}\text{C} : E_{\text{Conc}} = \frac{1}{2}(80 - 60) = 10$$

$$\text{A } 80^{\circ}\text{C} : E_{\text{Conc}} = \frac{1}{2}(95 - 70) = 12.5$$

Dans ce cas, l'effet d'un facteur est différent et dépend du niveau de l'autre, les deux facteurs sont dépendants. On dit qu'il y a interaction entre la température et la pression.

- **Signification de l'effet d'un facteur dans le cas des interactions**

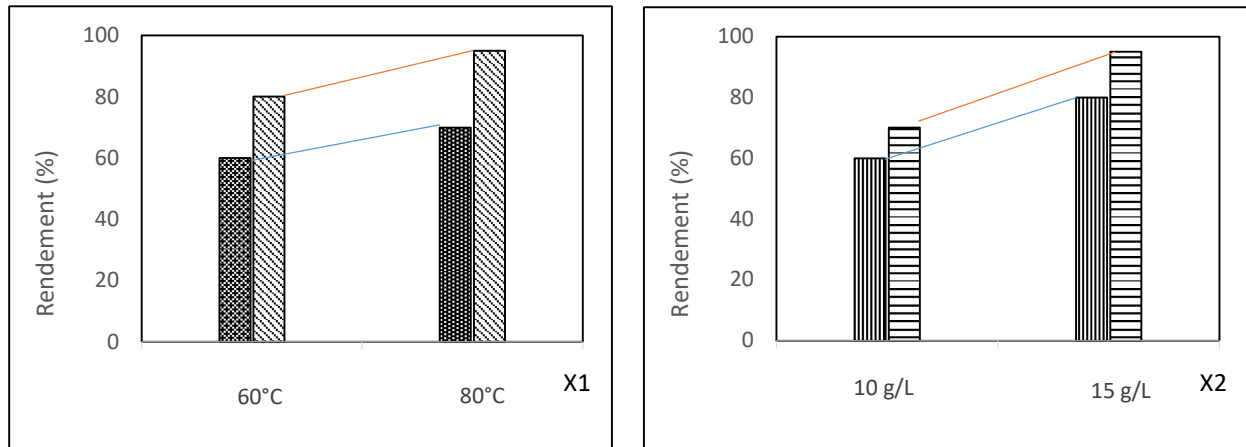
La signification de l'effet d'un facteur, appelé effet principal, apparait quand on représente la variation de l'effet en fonction du niveau de l'autre facteur. L'effet principal d'un facteur est l'effet de ce facteur lorsque l'autre facteur est au niveau 0.



- **Evaluation de l'interaction ($E_{X_1X_2}$)**

Elle se définit comme la moitié de la différence entre l'effet moyen d'un facteur au niveau haut de l'autre et l'effet moyen de ce facteur au niveau bas de l'autre.

$$E_{X_1X_2} = \frac{1}{2} (7.5 - 5) = 1.25 \text{ ou bien } \frac{1}{2} (12.5 - 10) = 1.25$$



Ces graphes illustrent l'existence d'une interaction entre les facteurs A et B, l'effet de A au niveau 2 n'est pas égal à l'effet de A au niveau 1. Une interaction est une action mutuelle : L'effet de B dépend réciproquement du niveau de A.

En observant ces histogrammes, on remarque immédiatement que les droites traduisant l'effet des facteurs ne sont pas parallèles. L'effet du facteur A change suivant que le facteur B est aux niveaux 1 ou 2. Cela signifie qu'une interaction entre les facteurs A et B s'est manifestée. On dit alors que les facteurs A et B **ne sont pas indépendants**.

V. Analyse des résultats : quels effets (principaux, interactions) sont significatifs ?

Un coefficient (ou effet) est dit significatif s'il est différent de zéro. Un effet, positif ou négatif, est d'autant plus crédible que sa valeur absolue est grande.

Dans les chapitres précédents, nous avons évalué l'écart type σ à partir de la variance, il est donné par :

$$\sigma^2 = \frac{S^2}{N} \rightarrow \sigma = \frac{S}{\sqrt{N}}$$

Avec :

S^2 : l'estimateur de la variance,

σ^2 : la variance expérimentale,

σ : l'écart-type,

N : le nombre d'expérience.

Cette formule est donc valable pour les effets principaux, les interactions et la réponse au centre du domaine. La connaissance de l'estimateur S^2 permet donc de calculer la variance et déduire l'écart-type σ .

Dans la pratique expérimentale, plusieurs cas peuvent se présenter :

1. **L'écart-type est connu** par des expériences antérieures de même type que celles du plan et suffisamment nombreuses pour être fiables. C'est le meilleur des cas.

Pour estimer les effets influents, on calcule le t de Student (t_{exp}) pour chaque coefficient, donné par :

$$t_{exp} = \frac{|b_j|}{\sigma}$$

b_j : effet des facteurs principaux ou des interactions,

Connaissant le nombre de degré de liberté (ddl), on détermine le $t_{critique}$ en utilisant la table de Student en choisissant un risque α le plus souvent de 5%. Nous considérons que :

si $t_{exp} > t_{critique}$, l'effet est significatif.

si $t_{exp} < t_{critique}$, l'effet est non significatif.

2. **L'écart-type n'est pas connu** : pour l'estimer, l'expérimentateur a prévu :

- Soit faire des expériences au centre du domaine,
- Soit de doubler le plan.

a. Des expériences au centre du domaine expérimental

Pour un plan d'expérience qui contient N_0 essais, On prévoit des répétitions au centre de domaine (N_0 fois $\geq k+1$). On calcule d'abord la valeur moyenne des réponses qui permettent de calculer l'écart type d'un effet (voir chapitre 1).

$$\bar{y} = \frac{1}{n_0} \sum_{i=1}^{n_0} y_i \rightarrow S^2 = \frac{1}{n_0 - 1} \sum_{i=1}^{n_0} (y_i - \bar{y})^2$$

$$\sigma = \frac{S}{\sqrt{N}} \rightarrow t_{exp} = \frac{|b_j|}{\sigma}$$

Chaque effet (principal et interaction) est alors testé par un test t de Student comme précédemment.

Exemple

Reprenons l'exemple de l'étude de l'influence de la température et de la concentration d'un réactif sur le rendement d'une réaction chimique. Le domaine de variation de la température (θ) est entre 60°C et 80°C et la concentration entre 10 g/L et 15 g/L ($t_{critique} = 2,57$).

Six expériences au centre du domaine expérimental ont été faites, les résultats trouvés (rendement de la réaction (%)) obtenus sont :

$$y = \{77,3; 79,1; 77,8; 77,0; 77,7; 79,1\}$$

Essai	y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1	77,30	-0,70	0,49
2	79,10	1,10	1,21
3	77,80	-0,20	0,04
4	77,00	-1,00	1,00
5	77,70	-0,30	0,09
6	79,10	1,10	1,21
moyenne	78,00	Somme $(y_i - \bar{y})^2$	4,04
		ddl	6 - 1 = 5
		S^2	$(4,04/5) = \mathbf{0,81}$
		var (bj) = σ^2	$(0,81/4) = \mathbf{0,20}$
		σ	$\sqrt{\mathbf{0,20}} = \mathbf{0,45}$

b. Double le plan d'expérience

Cela veut dire que chaque expérience doit être réalisée deux fois au minimum. Le calcul des effets principaux et des interactions se fait de deux manières :

1^{er} cas : on calcule la moyenne des réponses pour chaque condition expérimentale puis on effectue le calcul classique des effets avec les méthodes analytique ou matricielle.

	X1	X2	X1X2	Réponse 1	Réponse 2	Moyenne
1	-1	-1	+1	y_1	y'_1	\bar{y}_1
2	+1	-1	-1	y_2	y'_2	\bar{y}_2
3	-1	+1	-1	y_3	y'_3	\bar{y}_3
4	+1	+1	+1	y_4	y'_4	\bar{y}_4

$$E_1 = b_1 = \frac{1}{4}(-\bar{y}_1 + \bar{y}_2 - \bar{y}_3 + \bar{y}_4)$$

$$E_2 = b_2 = \frac{1}{4}(-\bar{y}_1 - \bar{y}_2 + \bar{y}_3 + \bar{y}_4)$$

$$E_{12} = b_{12} = \frac{1}{4}(+\bar{y}_1 - \bar{y}_2 - \bar{y}_3 + \bar{y}_4)$$

2^{ème} cas : on peut copier la matrice des effets autant de fois qu'il y a répétition et on effectue le calcul des effets à partir des réponses individuelles en divisant par le nombre d'expériences réalisées. Par exemple, pour un plan 2^2 dupliqué :

Essai 1	X1	X2	X1X2	Réponse
1	-1	-1	+1	y_1
2	+1	-1	-1	y_2
3	-1	+1	-1	y_3
4	+1	+1	+1	y_4
Essai 2	X1	X2	X1X2	Réponse
1	-1	-1	+1	y_1'
2	+1	-1	-1	y_2'
3	-1	+1	-1	y_3'
4	+1	+1	+1	y_4'

Le calcul des coefficients sera :

$$E_A = E_1 = b_1 = \frac{1}{8}(-y_1 + y_2 - y_3 + y_4 - y_1' + y_2' - y_3' + y_4')$$

$$E_B = E_2 = b_2 = \frac{1}{8}(-y_1 - y_2 + y_3 + y_4 - y_1' - y_2' + y_3' + y_4')$$

$$E_{12} = b_{12} = \frac{1}{8}(+y_1 - y_2 - y_3 + y_4 + y_1' - y_2' - y_3' + y_4')$$

Application 1

Lors de l'optimisation des conditions expérimentales d'une réaction de précipitation, on a retenu comme facteurs à étudier susceptibles d'agir sur le poids du précipité pesé les facteurs suivants:

Facteurs		Niveau bas	Niveau haut
Température de la réaction	X1	60°C	70°C
Concentration d'un réactif	X2	1 g/L	2 g/L
Temps de contact	X3	30mn	45mn
Débit de lavage du précipité	X4	1 L/mn	0.5 L/mn

1- Quel est le nombre d'expérience à réaliser ? En ne tenant compte que les effets principaux et les interactions du 1^{er} ordre, construire la matrice d'expérience et la matrice des effets,

2- La réponse étudiée est la masse du précipité obtenue :

Y= {60.6 ; 61.0 ; 60.3 ; 61.7 ; 62.0 ; 61.5 ; 61.7 ; 62.4 ; 59.6 ; 61.1 ; 60.7 ; 61.3 ; 61.6 ; 61.9 ; 62.3 ; 62.8}.

Calculer les coefficients b_j et déduire le modèle mathématique,

3- Retrouver les effets influents ($t_{\text{critique}} = 0.727$), tracer qualitativement le diagramme des effets.

4- Etudier les interactions influentes.

Corrigé

- Le nombre d'expérience à réaliser est de $2^4=16$ expériences.
- La matrice d'expérience est la suivante :

N°	X0	X1	X2	X3	X4
1	1	-1	-1	-1	-1
2	1	1	-1	-1	-1
3	1	-1	1	-1	-1
4	1	1	1	-1	-1
5	1	-1	-1	1	-1
6	1	1	-1	1	-1
7	1	-1	1	1	-1
8	1	1	1	1	-1
9	1	-1	-1	-1	1
10	1	1	-1	-1	1
11	1	-1	1	-1	1
12	1	1	1	-1	1
13	1	-1	-1	1	1
14	1	1	-1	1	1
15	1	-1	1	1	1
16	1	1	1	1	1

En ne tenant compte que les effets principaux et les interactions du 1^{er} ordre, la matrice des effets sera :

N°	X0	X1	X2	X3	X4	X1X2	X1X3	X1X4	X2X3	X2X4	X3X4	C(g/l)
1	1	-1	-1	-1	-1	1	1	1	1	1	1	60,6
2	1	1	-1	-1	-1	-1	-1	-1	1	1	1	61
3	1	-1	1	-1	-1	-1	1	1	-1	-1	1	60,3
4	1	1	1	-1	-1	1	-1	-1	-1	-1	1	61,7
5	1	-1	-1	1	-1	1	-1	1	-1	1	-1	62
6	1	1	-1	1	-1	-1	1	-1	-1	1	-1	61,5
7	1	-1	1	1	-1	-1	-1	1	1	-1	-1	61,7
8	1	1	1	1	-1	1	1	-1	1	-1	-1	62,4
9	1	-1	-1	-1	1	1	1	-1	1	-1	-1	59,6
10	1	1	-1	-1	1	-1	-1	1	1	-1	-1	61,1
11	1	-1	1	-1	1	-1	1	-1	-1	1	-1	60,7
12	1	1	1	-1	1	1	-1	1	-1	1	-1	61,3
13	1	-1	-1	1	1	1	-1	-1	-1	-1	1	61,6
14	1	1	-1	1	1	-1	1	1	-1	-1	1	61,9
15	1	-1	1	1	1	-1	-1	-1	1	1	1	62,3
16	1	1	1	1	1	1	1	1	1	1	1	62,8
	b0	b1	b2	b3	b4	b12	b13	b14	b23	b24	b34	
	61,406	0,306	0,244	0,619	0,006	0,094	-0,181	0,056	0,031	0,119	0,119	

Après avoir calculer les coefficients du modèle, il faut chercher les facteurs et les interactions influentes.

La variance expérimentale est inconnue, pour interpréter nos résultats, On calcule l'estimateur :

$$var(y) = S^2 = \frac{1}{N-p} \sum e_i^2 = \frac{1}{N-p} \sum (y_{exp} - y_{theo})^2$$

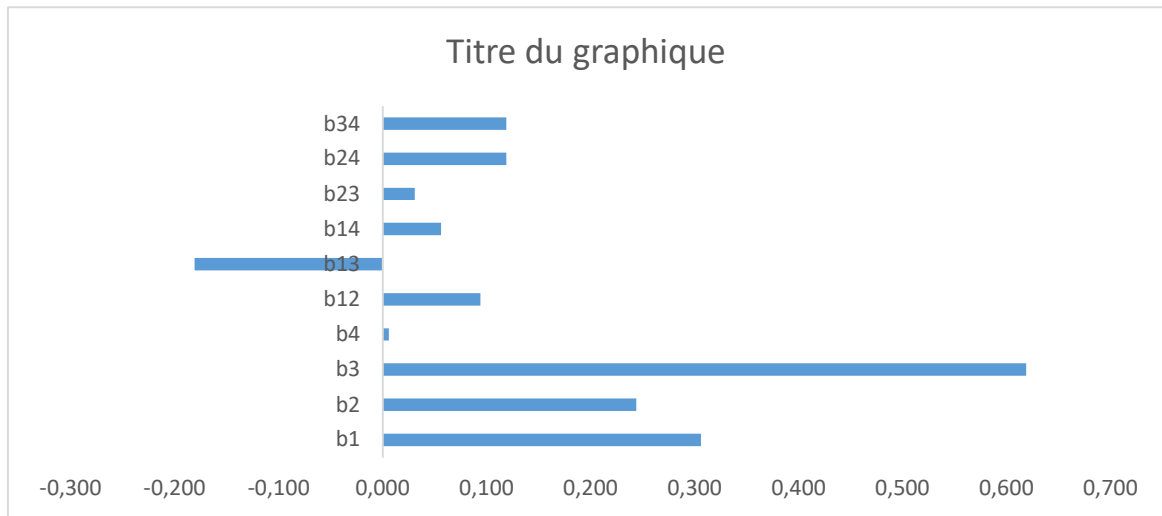
La variance des coefficients est donnée par :

$$var(b_j) = \sigma^2 = \frac{var(y)}{N} \rightarrow \text{l'écart - type: } \sigma = \sqrt{var(b_j)}$$

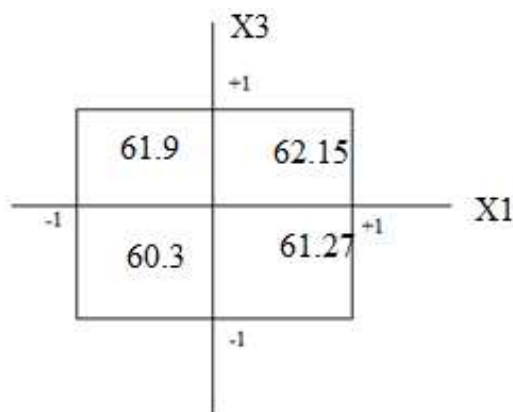
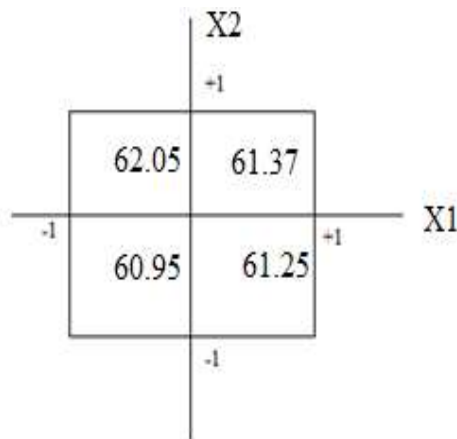
Les résultats trouvés sont regroupés dans le tableau suivant : (tc = 0727)

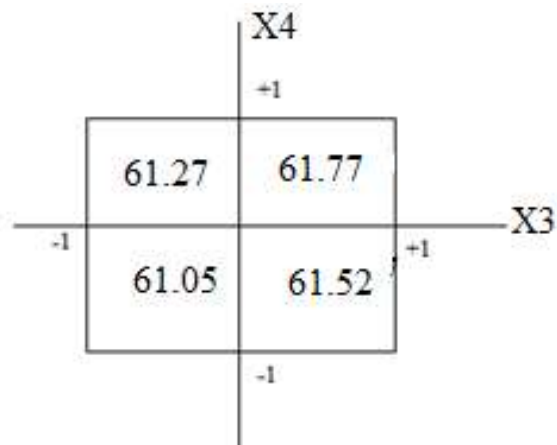
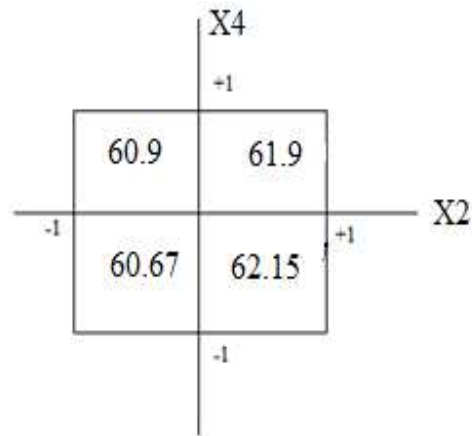
Coefficient	abs(bj)	Ecart type	texp	Conclusion
b1	0,306	0,096	3,188	influent
b2	0,244	0,096	2,542	influent
b3	0,619	0,096	6,448	influent
b4	0,006	0,096	0,063	non influent
b12	0,094	0,096	0,979	influent
b13	0,181	0,096	1,885	influent
b14	0,056	0,096	0,583	non influent
b23	0,031	0,096	0,323	non influent
b24	0,119	0,096	1,240	influent
b34	0,119	0,096	1,240	influent

Le diagramme des effets est donné par :



Etude des interactions





Application 2 : Procédé de préparation des sulfates amide des acides gras (voir chapitre III)

La détermination de la variance expérimentale a été faite en reprenant quelques expériences du plan d'expérimentation. Les expériences (1, 5, 8 et 12) ont été faites deux fois.

	ESSAI 1	ESSAI 2	moyenne	Variance	ddl
1	74	78	76	$(76-74)^2 + (76-78)^2 = 8$	1
5	141	139	140	$(140-141)^2 + (140-139)^2 = 2$	1
8	123	145	134	$(134-123)^2 + (134-145)^2 = 242$	1
12	69	76	72.5	$(72.5-69)^2 + (72.5-76)^2 = 24,5$	1
		Var(Y) = S ²		$\frac{1}{4}(8+2+242+24.5) = 69,13$	4
		Var(bj) = σ^2		$69,13/12 = 5,76$	
		$\sigma(bj)$		2,40	

3. **L'écart-type n'est pas connu** et il n'y a eu ni répétitions du plan, ni essais complémentaires. On fait appel aux outils d'aide à l'interprétation (approche de Lenth ou Normal/Half Plot).

VI. Intervalle de confiance

Connaissant l'écart-type, l'intervalle de confiance d'un effet est donné, par :

$$I_c = \left[b_i - t \cdot \frac{\sigma}{\sqrt{N}}; b_i + t \cdot \frac{\sigma}{\sqrt{N}} \right]$$

Pour $\alpha = 5\%$: t (Student) = 1,96.

Pour $\alpha = 1\%$: t (Student) = 2,58.

VII. Validation du modèle linéaire –analyse de variance-ANOVA

L'analyse de la variance consiste à comparer à l'aide d'un test de Fisher (test F), la somme des carrés des écarts due uniquement à la régression (ou modèle), avec la somme des carrés des résidus. Précisons ces notions en introduisant un vocabulaire spécifique à l'analyse de variance.

On notera par la suite Y_{exp} les réponses observées lors de la réalisation des expériences et Y_{cal} la réponse estimée à l'aide du modèle linéaire. On notera, de même, Y_{moy} la moyenne des réponses ($Y_{moy} = b_0$).

On définit alors trois types de "variations"

- 1- La variation due à la liaison linéaire

$$SCEL = \sum (y_{calc} - y_{moy})^2 = \sum (y_{calc} - b_0)^2$$

SCEL se lit : "somme des carrés des écarts dues à la liaison".

- 2- La variation résiduelle

$$SCER = \sum (y_{exp} - y_{calc})^2 = \sum (e_i)^2$$

SCER se lit : "somme des carrés des écarts des résidus".

- 3- La variation totale

$$STCE = SCEL + SCER$$

STCE se lit : " somme totale des carrés des écarts".

On définit de plus un "carré moyen" qui est le quotient d'une somme de carrés par son degré de liberté.

SCEL aura $(p - 1)$ degrés de liberté (p est le nombre de coefficients estimé à partir du modèle).

SCER aura $(N - p)$ degrés de libertés (N est le nombre d'expériences réalisées).

SCET aura $(N - 1)$ degrés de liberté.

Le tableau d'ANOVA se présente comme suit :

Variation due à	Somme des carrés	ddl	Carré moyen	F
Liaison	SCEL	p-1	$CML = \frac{SCEL}{p-1}$	$F_{obs} = \frac{CML}{s^2}$
Résidus	SCER	N-p	$s^2 = \frac{SCER}{N-p}$	
Totale	SCET	N-1		

Le test F permet alors de comparer pour un risque fixé à l'avance le F_{obs} que l'on a calculé dans le tableau précédent avec la valeur de F (critique) lue dans la table de Fisher-Snedecor avec (p-1) et (N - p) degrés de liberté.

Le test est le suivant pour un risque α :

Si $F_{obs} < F$ (critique), on rejette l'hypothèse de linéarité du modèle.

Si $F_{obs} > F$ (critique), on accepte l'hypothèse de linéarité du modèle.

VII. Coefficient de détermination (R^2)

Un modèle théorique est d'autant proche des réponses expérimentales lorsque le coefficient de détermination est proche de 1. Généralement, un modèle est validé lorsque $R^2 > 0,90$.

Il est donné par :

$$R^2 = 1 - \frac{\sum(y_{exp} - y_{calc})^2}{\sum(y_{exp} - y_{moy})^2} = 1 - \frac{\sum(e_i)^2}{\sum(y_{exp} - b_0)^2} = 1 - \frac{SCER}{SCEL}$$

Le coefficient de détermination ajusté est donné par :

$$R_A^2 = 1 - \frac{\frac{\sum(y_{exp} - y_{calc})^2}{N-p}}{\frac{\sum(y_{exp} - y_{moy})^2}{N-1}}$$

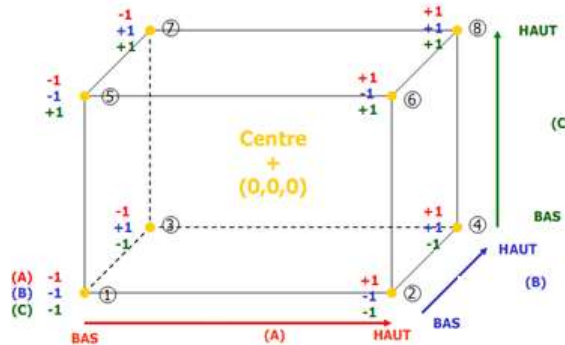
Application : Etude d'un plan 2³

Soit un plan factoriel complet à trois facteurs. Le nombre d'expérience à réaliser est $2^k = 2^3 = 8$ expériences.

Le domaine expérimental peut être représenté par un cube ; les points expérimentés sont aux 8 sommets.

Les réponses sont :

$$y = \{5,2; 4,7; 5,1; 5,5; 4,9; 4,6; 4,8; 5,3\}$$



L'équation du modèle est :

$$y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3 + b_{123}X_1X_2X_3$$

On néglige l'interaction d'ordre 3.

• **Matrice des effets**

	X0	X1	X2	X3	X1X2	X1X3	X2X3	Y (%)
1	+1	-1	-1	-1	+1	+1	+1	5,2
2	+1	+1	-1	-1	-1	-1	+1	4,7
3	+1	-1	+1	-1	-1	+1	-1	5,1
4	+1	+1	+1	-1	+1	-1	-1	5,5
5	+1	-1	-1	+1	+1	-1	-1	4,9
6	+1	+1	-1	+1	-1	+1	-1	4,6
7	+1	-1	+1	+1	-1	-1	+1	4,8
8	+1	+1	+1	+1	+1	+1	+1	5,3
b _j	5,0125	0.0125	0.1625	-0.1125	0.2125	0.0375	-0.0125	

Après avoir calculer les coefficients du modèle, il faut chercher les facteurs et les interactions influentes.

La variance expérimentale est inconnue, pour interpréter nos résultats, On calcule l'estimateur :

$$\text{var}(y) = S^2 = \frac{1}{N-p} \sum e_i^2 = \frac{1}{N-p} \sum (y_{\text{exp}} - y_{\text{theo}})^2$$

La variance des coefficients est donnée par :

$$\text{var}(b_j) = \sigma^2 = \frac{\text{var}(y)}{N} \rightarrow \text{l'écart - type: } \sigma = \sqrt{\text{var}(b_j)}$$

N étant le nombre d'expérience = 8. (voir tableau)

Le modèle postulé est donné par :

$$y = 5,0125 + 0,1625 X2 + 0.2125 X1X2$$

N°	X0	X1	X2	X3	X1X2	X1X3	X2X3	y	y _{cal}	e _i	e _i ²
1	1	-1	-1	-1	1	1	1	5,2	5,1875	0,0125	0,0001563
2	1	1	-1	-1	-1	-1	1	4,7	4,7125	-0,0125	0,0001563
3	1	-1	1	-1	-1	1	-1	5,1	5,1125	-0,0125	0,0001562
4	1	1	1	-1	1	-1	-1	5,5	5,4875	0,0125	0,0001562
5	1	-1	-1	1	1	-1	-1	4,9	4,9125	-0,0125	0,0001563
6	1	1	-1	1	-1	1	-1	4,6	4,5875	0,0125	0,0001562
7	1	-1	1	1	-1	-1	1	4,8	4,7875	0,0125	0,0001563
8	1	1	1	1	1	1	1	5,3	5,3125	-0,0125	0,0001563
b_j	5,0125	0,0125	0,1625	-	0,2125	0,0375	0,0125			Somme	0,00125
										Estimateur	0,00125
										Variance	0,0001562
										Ecart type	0,0125

Le test de Student appliqué à ces résultats est donné par :

Coefficient	b _j	abs (b _j)	Ecart type	t (Student)	tcritique	Conclusion
b1	0,0125	0,0125	0,0125	1	12,70	Non influent
b2	0,1625	0,1625	0,0125	13	12,70	influent
b3	-0,1125	0,1125	0,0125	9	12,70	Non influent
b12	0,2125	0,2125	0,0125	17	12,70	influent
b13	0,0375	0,0375	0,0125	3	12,70	Non influent
b23	-0,0125	0,0125	0,0125	1	12,70	Non influent

Le modèle postulé est donné par :

$$y = 5,0125 + 0,1625 X_2 + 0,2125 X_1 X_2$$

Faisons l'analyse de variance avec le modèle complet puis avec le modèle postulé :

- **Analyse de variance (ANOVA)**

Variation due à	Somme des carrés	Ddl	Carré moyen	F
Liaison	0,6875	7-1 =6	$CML = 0,1146$	$F_{obs} = 91,66$
Résidus	0,00125	8-7 =1	$s^2 = 0,00125$	
Totale	0.6887	8-1=7		

La table de Fischer-Snédecor donne pour $v_1 = 6$ et $v_2 = 1$, $F(\text{crit}) = 233,98$ pour un risque de 5%.

On a : ($F_{obs} = 91,667$) < ($F_{crit} = 233,98$) donc on rejette l'hypothèse de linéarité du modèle.

En revanche, effectuons une nouvelle analyse de variance avec un modèle linéaire ne contenant que les coefficients significatifs, b_2 et b_{12} . Le nouveau tableau d'analyse de variance est alors :

On a cette fois $N= 8$, mais $p = 3$ (3 coefficients estimés).

Variation due à	Somme des carrés	Ddl	Carré moyen	F
Liaison	$SCEL=0.5725$	3-1 =2	$CML = 0,2863$	$F_{obs} = 12,340$
Résidus	$SCER=0.1162$	8-3 =5	$s^2 = 0,0232$	
Totale	SCET	8-1=7		

La table de Fischer-Snédecor donne pour $v_1 = 2$ et $v_2 = 5$, $F(\text{crit}) = 5,79$, pour un risque de 5%. On a :

$$F_{obs} = 12,3118 > F_{crit} = 5,79$$

Donc on accepte l'hypothèse H_1 de linéarité du modèle.

- **Intervalle de confiance**

Intervalle de confiance des effets significatifs à 5% :

$b_2 = 0,1625$:

$$I_c = \left[b_i - t \cdot \frac{\sigma}{\sqrt{N}} ; b_i + t \cdot \frac{\sigma}{\sqrt{N}} \right]$$

$$I_c = \left[0,1625 - 1,96 \cdot \frac{0,0125}{\sqrt{8}} ; 0,1625 + 1,96 \cdot \frac{0,0125}{\sqrt{8}} \right]$$

$b_{12} = 0,2125$:

$$I_c = \left[0,2125 - 1,96 \cdot \frac{0,0125}{\sqrt{8}} ; 0,2125 + 1,96 \cdot \frac{0,0125}{\sqrt{8}} \right]$$

TABLE 3.1: Table de student

$\alpha\%$	0.5	0.4	0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.002	0.001
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.080	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.950
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.335	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.071
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.220	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.699	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.681	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.675	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.201

TABLE 3.3: Table de Fischer - Snedecor (Risque $\alpha = 0.05$)

n_1	1	2	3	4	5	6	7	8	9	10	12	15	20
n_2													
1	161	199,5	215,7	224,6	230,2	234	236,8	239	240,5	241,9	243,9	245,9	248
2	18,5	19	19,16	19,25	19,3	19,33	19,35	19,4	19,38	19,4	19,41	19,43	19,45
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,7	8,66
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6	5,96	5,91	5,86	5,8
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,1	4,06	4	3,94	3,87
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44
8	5,32	4,46	4,07	3,84	3,69	3,58	3,5	3,44	3,39	3,35	3,28	3,22	3,15
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94
10	4,96	4,1	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77
11	4,84	3,98	3,59	3,36	3,2	3,09	3,01	2,95	2,9	2,85	2,79	2,72	2,65
12	4,75	3,89	3,49	3,26	3,11	3	2,91	2,85	2,8	2,75	2,69	2,62	2,54
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,6	2,53	2,46
14	4,6	3,74	3,34	3,11	2,96	2,85	2,76	2,7	2,65	2,6	2,53	2,46	2,39
15	4,54	3,68	3,29	3,06	2,9	2,79	2,71	2,64	2,59	2,54	2,48	2,4	2,33
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28
17	4,45	3,59	3,2	2,96	2,81	2,7	2,61	2,55	2,49	2,45	2,38	2,31	2,23
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19
19	4,38	3,52	3,13	2,9	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16
20	4,35	3,49	3,1	2,87	2,71	2,6	2,51	2,45	2,39	2,35	2,28	2,2	2,12
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,1
22	4,3	3,44	3,05	2,82	2,66	2,55	2,46	2,4	2,34	2,3	2,23	2,15	2,07
23	4,28	3,42	3,03	2,8	2,64	2,53	2,44	2,37	2,32	2,27	2,2	2,13	2,05
24	4,26	3,4	3,01	2,78	2,62	2,51	2,42	2,36	2,3	2,25	2,18	2,11	2,03
25	4,24	3,39	2,99	2,76	2,6	2,49	2,4	2,34	2,28	2,24	2,16	2,09	2,01
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07	1,99
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,2	2,13	2,06	1,97
28	4,2	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04	1,96
29	4,18	3,33	2,93	2,7	2,55	2,43	2,35	2,28	2,22	2,18	2,1	2,03	1,94
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2	1,92	1,84
60	4	3,15	2,76	2,53	2,37	2,25	2,17	2,1	2,04	1,99	1,92	1,84	1,75
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,83	1,75	1,66
infini	3,84	3	2,6	2,37	2,21	2,1	2,01	1,94	1,88	1,83	1,75	1,67	1,57

Deuxième partie : Les plans fractionnaires

I. Introduction

Les plans factoriels complets 2^k présentent le désavantage de nécessiter rapidement de nombreux essais quand le nombre de facteurs à étudier augmente ; l'expérimentation devient « encombrante », consommatrice de temps et d'argent (matériels, personnels, matières ...).

C'est ainsi qu'un plan 2^5 nécessite 32 essais, qu'un plan 2^6 en comporte 64, pour un plan 2^7 il en faut 128 ...etc.

Ces plans permettent, certes, le calcul de tous les effets principaux et de toutes les interactions mais la plupart de ces interactions (particulièrement celles qui mettent en jeu plus de 2 facteurs) ne présentent aucun intérêt pour l'expérimentateur.

Dans un plan 2^7 par exemple, on sera intéressé par les 7 effets principaux et par les interactions de 2^{ème} ordre mettant en jeu 2 facteurs actifs (il ya 21 interactions de ce type) : il est raisonnable de penser qu'il n'est pas nécessaire de faire 128 essais pour obtenir ces renseignements.

Un plan fractionnaire permet de réduire le nombre d'essais à effectuer ; on n'effectue dans l'expérimentation qu'une fraction des essais du plan complet. Pour étudier les effets éventuels de 7 facteurs (chacun à 2 niveaux), on peut effectuer :

Seulement 64 essais (la moitié du plan complet) : le plan est dit 2^{7-1} ,

Seulement 32 essais (le quart du plan complet) : le plan est dit 2^{7-2} ,

Seulement 16 essais (le huitième du plan complet) : le plan est dit 2^{7-3} ,

Seulement 8 essais (le seizième du plan complet) : le plan est dit 2^{7-4} .

Attention, on ne peut pas choisir n'importe quels essais du plan complet : le fractionnement d'un plan doit obéir à des organisations qui font précisément l'objet de ce chapitre.

II. Définition

Les plans fractionnaires ont été conçus pour remédier à l'inflation rapide du nombre d'essais dans les plans complets. L'objectif des plans fractionnaires va consister à réduire le nombre d'expériences à réaliser par rapport au nombre maximum donné par le plan complet. Les plans fractionnaires utilisent les matrices des effets des plans complets.

On parlera de plan 2^{k-p} (p entier) pour indiquer un plan fractionnaire issu du plan complet 2^k avec k facteurs à 2 niveaux. Par exemple le plan 2^{4-1} est un plan fractionnaire

permettant l'étude de 4 facteurs en utilisant la matrice des effets du plan 2^3 . 2^3 expériences sont à réaliser au lieu des 2^4 expériences du plan complet. Le nombre d'expériences est divisé par 2, il correspond à la réalisation d'un demi plan complet 2^4 .

Néanmoins les plans fractionnaires nécessitent une phase de conception plus longue car l'interprétation des résultats dépend essentiellement du choix de p. Plus le nombre p augmente, plus la charge expérimentale va diminuer mais au détriment d'un risque de plus en plus grand sur la qualité des informations tirées du plan. Il faudra donc évaluer les risques avant de démarrer l'expérimentation et les minimiser en construisant le plan fractionnaire adéquat.

III. Calcul de BOX

Le calcul de Box permet facilement de retrouver les effets aliasés pour les plans fractionnaires à deux niveaux (-1) et (+1). On note :

La colonne du facteur 1 par 1,

La colonne du facteur 2 par 2,

La colonne du facteur 3 par 3, etc.....

La colonne I, la colonne qui possède des (+1),

La colonne (-I) la colonne qui possède des (-1)

$$\diamond \quad \mathbf{1x1 = 2x2 = 3x3 = I}$$

$$\mathbf{1x1 = \begin{matrix} & -1 & & -1 & & 1 \\ & 1 & & 1 & & 1 \\ -1 & & x & -1 & = & 1 \\ & 1 & & 1 & & 1 \end{matrix} = \mathbf{I}}$$

$$\diamond \quad \mathbf{1xI = 1}$$

$$\diamond \quad \mathbf{2xI = 2}$$

$$\begin{array}{ccccccc}
 & & -1 & & 1 & & -1 \\
 2xI = & & -1 & \times & 1 & = & -1 = 2 \\
 & & 1 & & 1 & & 1 \\
 & & 1 & & 1 & & 1
 \end{array}$$

❖ $1x(-I) = -1$

❖ $2x(-I) = -2$

$$\begin{array}{ccccccc}
 & & -1 & & -1 & & 1 \\
 & & 1 & & -1 & & -1 \\
 1x(-I) = & & -1 & \times & -1 & = & 1 = -I \\
 & & 1 & & -1 & & -1
 \end{array}$$

❖ $IxI = I$

$$\begin{array}{ccccccc}
 & & 1 & & 1 & & 1 \\
 IxI = & & 1 & \times & 1 & = & 1 = I \\
 & & 1 & & 1 & & 1 \\
 & & 1 & & 1 & & 1
 \end{array}$$

❖ *Commutativité:*

❖ $1x2 = 2x1$

$$\begin{array}{ccccccc}
 & -1 & -1 & 1 & -1 & -1 & 1 \\
 \\
 1x2 = & 1 & x & -1 & = & -1 & = 2x1 = & -1 & x & 1 & = & -1 \\
 \\
 & -1 & 1 & -1 & 1 & -1 & -1 \\
 \\
 & 1 & 1 & 1 & 1 & 1 & 1
 \end{array}$$

❖ *Associativité :*

❖ $1x2x3 = (1x2)x3 = 1x(2x3)$

IV. Générateur et relation de définition

Considérons un plan factoriel complet 2^3 . La matrice des effets sera :

	X0	X1	X2	X3	X1X2	X1X3	X2X3	X1X2X3
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

Selon le signe de la colonne X1X2X3, divisons la matrice en deux parties : les lignes N° 5, 2, 3 et 8 et les lignes 1, 6, 7 et 4. Nous aurons :

	X0	X1	X2	X3	X1X2	X1X3	X2X3	X1X2X3
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

On appelle générateur d'alias, la colonne identique à la colonne X0 (la colonne identité).

Dans le demi-plan supérieur (N° 5, 2, 3 et 8) :

$I = 123$ est la relation de définition, en utilisant le calcul de Box, on retrouve les autres alias :

$$1.I = 1.123 = 23$$

$$2.I = 2.123 = 13$$

$$3.I = 3.123 = 12$$

Dans le demi-plan inférieur (N° 1, 6, 7 et 4) :

$I = -123$ est la relation de définition, en utilisant le calcul de Box, on retrouve les autres alias :

$$1.I = 1.(-123) = -23$$

$$2.I = 2.(-123) = -13$$

$$3.I = 3.(-123) = -12$$

V. Notion d'alias et de contraste

Pour l'étude d'un plan fractionnaire 2^{4-1} à 4 facteurs, 8 expériences sont réalisées. Le modèle de plan factoriel avec 4 facteurs est le suivant :

$$Y = b_0 + b_1A + b_2B + b_3C + b_4D + b_{12}AB + b_{13}AC + b_{14}AD + b_{23}BC + b_{24}BD \\ + b_{34}CD + b_{123}ABC + b_{124}ABD + b_{134}ACD + b_{234}BCD + b_{1234}ABCD$$

Les matrices des effets des plans factoriels doivent comporter autant de lignes que de colonnes. Chaque colonne correspond à un coefficient du modèle donc il faudrait 16 colonnes : pour utiliser une matrice à 2^{4-1} colonnes, il faut donc faire en sorte que les termes de $(16 - 8)$ colonnes se retrouvent dans les 8 autres colonnes.

Les termes étant tous égaux à -1 ou $+1$, une solution consiste à écrire que la variable D aura la même valeur dans chacune des 8 expériences que la valeur du produit de variables A.B.C. Le choix des valeurs de D dans la matrice d'expérience sera donc fixé.

On peut donc écrire pour les 8 expériences : $D = A.B.C$. Comme cette égalité est vraie pour les 8 expériences, on peut écrire l'égalité de colonne suivante :

$$D = ABC$$

Le choix de $D = A.B.C$, permet alors d'écrire les égalités suivantes sur les variables en utilisant le fait que le produit d'une variable quelconque par elle-même est toujours égal à $+1$.

$$A.D = B.C$$

$$B.D = A.C$$

$$C.D = A.B$$

$$A.B.D = C$$

$$A.C.D = B$$

$$B.C.D = A$$

$$A.B.C.D = 1$$

On dira que D est l'**alias** initiale (choisie par l'expérimentateur).

D est aliasé avec A.B.C,

B est aliasé avec A.C.D,

A.D est aliasé avec B.C ...

Le modèle s'écrit alors sous la forme :

$$Y = (b_0 + b_{1234}) + (b_1 + b_{234})A + (b_2 + b_{134})B + (b_3 + b_{124})C + (b_{12} + b_{34})AB \\ + (b_{13} + b_{24})AC + (b_{14} + b_{23})BC + (b_4 + b_{123})ABC$$

On reconnaît bien le modèle d'un plan factoriel complet 2^3 dont on pourra utiliser la matrice des effets. Les nouveaux coefficients L_i calculés avec la matrice des effets par la méthode habituelle seront nommés **contrastes**.

Un contraste L_i est une somme d'effets et d'interactions. On les écrit ainsi :

$$L_0 = (b_0 + b_{1234})$$

$$L_1 = (b_1 + b_{234})$$

$$L_2 = (b_2 + b_{134})$$

$$L_3 = (b_3 + b_{124})$$

$$L_5 = (b_{12} + b_{34})$$

$$L_6 = (b_{13} + b_{24})$$

$$L_7 = (b_{14} + b_{23})$$

$$L_4 = (b_4 + b_{123})$$

Remarque : un contraste apparaît ici comme une somme mais dans le cas général des signes négatifs peuvent apparaître. Il suffirait par exemple de choisir comme aliase initiale

$$D = -A.B.C.$$

VI. Interprétation des contrastes

Les difficultés d'interprétation apparaissent dans l'expression des contrastes. Les 16 coefficients initiaux b_j du modèle sont a priori impossibles à calculer car ils sont groupés par 2.

La détermination des effets et des interactions est néanmoins possible moyennant certaines hypothèses couramment admises. Bien entendu il faut valider ces hypothèses avant de

conclure une étude, par exemple avec la vérification de points expérimentaux particuliers comme au centre du domaine d'étude. Les hypothèses sont les suivantes :

Hypothèse 1 : les interactions d'ordre égal ou supérieur à 3 peuvent être négligées ($b_{1234}=0$).

Hypothèse 2 : on peut considérer les interactions d'ordre 2 négligeables avec un risque acceptable ($b_{123}=0$).

Hypothèse 3 : Un effet d'interaction b_{ij} peut être significatif **si** l'un des effets principaux b_i **ou** b_j est significatif (hypothèse simple hérédité),

Hypothèse 4 : Un effet d'interaction b_{ij} est probablement significatif **si** les deux effets b_i **et** b_j sont significatifs (hypothèse double hérédité).

Il faut néanmoins être attentif à toute anomalie dans les résultats même si on estime que ces hypothèses sont vérifiées dans 95 % des cas.

VII. Résolution d'un plan fractionnaire

Le choix de l'alias initiale est fondamental pour la conception d'un plan fractionnaire.

Dans l'exemple précédent, le choix $D = A.B.C$ a permis d'obtenir des contrastes où les effets des facteurs A,B,C et D sont déterminés directement par le calcul des contrastes L_1 , L_2 , L_3 et L_4 .

L'hypothèse 1 permet d'annuler les interactions d'ordre 3, on écrit :

$$L_0 = (b_0 + b_{\overline{1234}}) = b_0$$

$$L_1 = (b_1 + b_{\overline{234}}) = b_1$$

$$L_2 = (b_2 + b_{\overline{134}}) = b_2$$

$$L_3 = (b_3 + b_{\overline{124}}) = b_3$$

$$L_4 = (b_4 + b_{\overline{123}}) = b_4$$

Par contre on dispose uniquement de la somme des interactions AD et BC, BD et AC, CD et AB. Seule l'application des hypothèses sur les contrastes calculés permettra peut-être de pouvoir connaître toutes les interactions.

On définit la notion de résolution d'un plan fractionnaire selon les termes de l'expression des contrastes :

Plan de résolution III: L'effet d'un facteur principal est aliasé avec l'effet de l'interaction d'ordre 1 (par exemple $A = BC$).

Plan de résolution IV : Les effets principaux sont aliasés avec les effets des interactions d'ordre 2 (par exemple $A = BCD$), les effets des interactions d'ordre 1 sont aliasés entre elles (par exemple $AB = CD$).

Plan de résolution V : Les effets principaux sont aliasés avec les effets des interactions d'ordre 3 (par exemple $A = BCDE$), les effets des interactions d'ordre 1 sont aliasés avec les effets des interactions d'ordre 2 (par exemple $AB = CDE$).

Il apparaît que les plans d'ordre III sont risqués car si une interaction n'est pas supposée négligeable, elle s'ajoute à l'effet d'un facteur. Ce type de plan est à entreprendre pour réduire énormément le nombre d'expériences dans le cas d'un grand nombre de facteurs. Dans ce cas il est parfois nécessaire de réaliser un plan complémentaire pour accéder à toutes les interactions.

Les plans d'ordre V permettent de lever tous les doutes sur la détermination des effets des facteurs et des interactions. Néanmoins ils sont peu fractionnaires et sont ainsi coûteux en nombre d'essais.

Les plans d'ordre IV présentent un bon compromis : les effets des facteurs sont faciles à déterminer, il reste par contre des doutes sur les interactions d'ordre 2. Le plan 2^3 de l'exemple précédent est un plan d'ordre IV.

Exercice d'application

On étudie le durcissement d'une pâte. Les facteurs choisis sont la température, le taux d'humidité, la durée de malaxage et la durée de stockage (voir tableau).

Le but est de rechercher le temps mis par la pâte pour qu'elle soit molle.

	Facteurs	Niveau (-1)	Niveau (+1)
A	Température	15	25
B	Taux d'humidité	0.1	0.5
C	Durée de malaxage	Court	Long
D	Durée de stockage	1 semaine	6 mois

On a vu qu'un plan factoriel complet 2^4 nécessite 16 essais pour déterminer 16 coefficients (moyenne, 4 effets de facteurs, 6 interactions du second ordre, 4 interactions du troisième ordre et l'interaction du quatrième ordre). On souhaite diviser par 2 le nombre d'essais soit réaliser un plan fractionnaire 2^{4-1} dont la matrice des effets est celle du plan complet 2^3 .

Corrigé**Conception du plan fractionnaire 2^{4-1}**

On écrit la matrice des effets du plan 2^3 . Cette matrice dispose de 8 lignes et 8 colonnes : on peut donc placer 8 termes du modèle au maximum.

On choisit comme aliase initiale : **D = A.B.C**

Ce choix s'impose : en effet les interactions d'ordre égal ou supérieur à 3 étant négligeables, il sera donc possible de déterminer l'effet des facteurs sans ambiguïté.

Le générateur d'alias est donc : **I = ABCD**.

Les autres alias sont obtenues en multipliant chaque colonne par le générateur d'alias. On obtient :

$$I = ABCD$$

$$A = BCD$$

$$B = ACD$$

$$C = ABD$$

$$AB = CD$$

$$AC = BD$$

$$BC = AD$$

$$ABC = D$$

Par exemple

$$AC = ACI = AC \mathbf{ABCD} = AA \ CC \ BD = \mathbf{II} \ BD = BD$$

La multiplication des colonnes avec le générateur d'aliases permet ainsi de "retrouver" tous les termes du modèle avec le choix de l'alias originale ($D = A.B.C$).

La matrice d'expérience s'écrit ci-dessous avec la même colonne pour D que ABC dans la matrice des effets,

N°	A	B	C	D	Y
1	-1	-1	-1	-1	61,00
2	1	-1	-1	1	78,00
3	-1	1	-1	1	72,00
4	1	1	-1	-1	29,00
5	-1	-1	1	1	81,00
6	1	-1	1	-1	61,00
7	-1	1	1	-1	33,00
8	1	1	1	1	68,00

La matrice des effets s'écrit avec la matrice d'expériences contenue. Les réponses sont données dans la matrice des effets :

	I=ABCD	A = BCD	B = ACD	C = ABD	AB=CD	AC=BD	BC=AD	ABC=D	y
1	1	-1	-1	-1	1	1	1	-1	61,00
2	1	1	-1	-1	-1	-1	1	1	78,00
3	1	-1	1	-1	-1	1	-1	1	72,00
4	1	1	1	-1	1	-1	-1	-1	29,00
5	1	-1	-1	1	1	-1	-1	1	81,00
6	1	1	-1	1	-1	1	-1	-1	61,00
7	1	-1	1	1	-1	-1	1	-1	33,00
8	1	1	1	1	1	1	1	1	68,00
Contraste	60,38	-1,38	-9,88	0,38	-0,63	5,13	-0,38	14,38	
	<i>L0</i>	<i>L1</i>	<i>L2</i>	<i>L3</i>	<i>L4</i>	<i>L5</i>	<i>L6</i>	<i>L7</i>	
	moy+ABCD	A+BCD	B+ACD	C+ABD	AB+CD	AC+BD	BC+AD	ABC+D	

D'après la relation entre les contrastes et les colonnes, on peut écrire les relations suivantes :

$$L_0 = (b_0 + b_{1234})$$

$$L_1 = (b_1 + b_{234})$$

$$L_2 = (b_2 + b_{134})$$

$$L_3 = (b_3 + b_{124})$$

$$L_4 = (b_{12} + b_{34})$$

$$L_5 = (b_{13} + b_{24})$$

$$L_6 = (b_{14} + b_{23})$$

$$L_7 = (b_4 + b_{123})$$

On peut alors commencer à déterminer les facteurs et interactions avec les hypothèses. D'après ce qui a été dit auparavant, parmi les interactions seul le calcul de celles d'ordre 2 présentent un intérêt.

$$L_0 = (b_0 + b_{1234}) = b_0$$

$$L_1 = (b_1 + b_{234}) = b_1$$

$$L_2 = (b_2 + \cancel{b_{134}}) = b_2$$

$$L_3 = (b_3 + \cancel{b_{124}}) = b_3$$

$$L_4 = (b_{12} + b_{34})$$

$$L_5 = (b_{13} + b_{24})$$

$$L_6 = (b_{14} + b_{23})$$

$$L_7 = (b_4 + b_{123})$$

❖ **Détermination des contrastes influents**

La signification statistique des contrastes est à envisager comme celle des coefficients. Une valeur de contraste peut être élevée sans être pour autant significative et inversement. Dans le cas étudié, nous avons choisi la méthode de Lenth pour déterminer les contrastes influents.

Nom	Li	Abs Li+ tri		
L6	-0,38	0,38	0,3800	0,38
L3	0,38	0,38	0,3800	0,38
L4	-0,63	0,63	0,6300	0,63
L1	-1,38	1,38	1,3800	1,38
L5	5,13	5,13	5,1300	
L2	-9,88	9,88		
L7	14,38	14,38		
	Mediane	1,38	0,6300	0,505
	S0	2,07	0,9450	0,7575
	2.5*S0	5,175	2,3625	1,89375
		PSE	0,7575	
		ddl	1,0000	
		t _{critique}	12,7062	
		limites	±9,6250	

On voit clairement que les contrastes influents sont :

$$L_2 = (b_2 + b_{134}) = b_2 \quad , \quad L_5 = (b_{13} + b_{24}) \quad \text{et} \quad L_7 = (b_4 + b_{123}) = b_4$$

Les contrastes L_2 et L_7 sont influents. Compte tenu des hypothèses, b_2 et b_4 sont jugés influents.

Le contraste L_5 est influent, il est constitué des des interactions b_{13} et b_{24} .

b_1 et b_3 sont non significatifs donc l'interaction b_{13} est non significative.

$$L_5 = (b_{13} + b_{24}) = b_{24} = 5.13$$

Il faut étudier l'interaction 24 ou BD :

$$b_2^+ = \frac{70 - 79.5}{2} = -4.72$$

$$b_2^- = \frac{31 - 61}{2} = -15$$

Chapitre V

Initiation aux surfaces de réponse

Chapitre V : Initiation aux surfaces de réponse (RSM)

I. Introduction

Les plans examinés précédemment n'avaient que deux niveaux par facteur (-1 et +1) et les modèles mathématiques utilisés étaient du premier degré (avec ou sans interactions) par rapport à chaque facteur. Ces plans sont les plus employés car ils permettent le criblage des facteurs et conduisent parfois à des modélisations simples et suffisantes. Pourtant, il existe de nombreux cas où il est nécessaire d'avoir une bonne modélisation des phénomènes étudiés et où il faut passer à des modèles mathématiques du second degré. Dans ce cas, on fait alors appel aux plans pour surfaces de réponse (RSM : Response Surface Methodology).

L'optimisation par la méthode de surface de réponse d'un processus quelconque consiste à trouver l'optimum d'une fonction donnée, qui correspond aux conditions optimales de conduite de ce processus.

Quand un modèle mathématique est construit pour décrire un phénomène, il peut être utilisé pour prédire les réponses dans le domaine d'étude ou pour trouver les conditions expérimentales qui conduisent à la réponse optimale. Ce n'est qu'après une analyse minutieuse des phénomènes que l'on peut définir la ou les bonnes réponses.

Le modèle mathématique utilisé avec les plans pour surfaces de réponse est un modèle polynomiale du second degré :

$$y = \underbrace{b_0}_{\text{Moyenne des réponses}} + \underbrace{\sum_{i=1}^k b_i X_i}_{\text{Modèle du criblage}} + \underbrace{\sum_{i,j=1, i \neq j}^k b_{ij} X_i X_j}_{\text{Modèle plan factoriel}} + \underbrace{\sum_{i=1}^k b_{ii} X_i^2}_{\text{Terme quadratique}}$$

Modèle RSM

y est la réponse,

X_i représente le niveau attribué au facteur i ,

b_0 , b_i ($i = 1, 2, \dots, k$), b_{ii} ($i = 1, 2, \dots, k$) et b_{ij} ($i = 1, 2, \dots, k-1; j = 2, 3, \dots, k$) sont respectivement les coefficients de régression pour les termes d'ordonnée à l'origine, linéaire, quadratique et d'interaction.

La matrice du modèle obtenue à partir de la matrice d'expérience et du modèle postulé permet d'écrire sous forme matricielle l'équation :

$$y = Xb + e$$

L'objectif étant de trouver un ensemble de p coefficients qui résout le système d'équation, en minimisant la somme des carrés des résidus (SCER).

La résolution nous permettra d'accéder aux coefficients du modèle postulé, ils sont donnés par :

$$b = (X^t X)^{-1} \cdot X^t \cdot y$$

X^t est la matrice transposée de X ,

$X^t X$ est la matrice d'information,

$(X^t X)^{-1}$ est la matrice de dispersion.

L'examen du modèle postulé va nous permettre :

- D'établir une relation entre la réponse y et les facteurs X_1, X_2, \dots, X_k ,
- De déterminer à l'aide des tests statistiques la signification des facteurs,
- De calculer les niveaux optimaux (minimum ou maximum) pour chaque facteur.

Le domaine d'étude peut être sphérique ou cubique et les variables sont toujours **quantitatives**.

Les plans de surfaces de réponse les plus utilisés en chimie sont : les plans composites centrés (1951), les plans de Box-Behnken (1960) et les plans de Doehlert (1970).

II. Critère d'optimalité

Pour un modèle donné, on cherchera le placement « optimal » des points d'expériences pour lesquels l'erreur sur les réponses prédites est la plus faible possible.

1. Qualité de prédiction

Considérons un point A du domaine expérimental, soit X_A , la matrice du modèle au point A.

La fonction de variance au point A (d_A) dépend seulement des coordonnées du point A, de la matrice d'expérience et du modèle postulé. Cette fonction de variance est donnée par :

$$d_A = X_A^t \cdot (X_A^t X_A)^{-1} \cdot X_A$$

X_A : matrice du modèle au point A.

Une matrice d'expérience est dite de prévision acceptable si elle conduit à une fonction de variance maximale de la réponse $d_{\max} \leq 1$ dans tout le domaine expérimental. Il est très utile de connaître la valeur de d_{\max} pour juger si on peut passer à l'expérimentation.

1. Isovariance par rotation

La fonction de variance dépend des coordonnées du point pour lequel la valeur de la réponse est calculée. Nous ne pouvons pas comparer les valeurs calculées au point A et au point B si les variances en ces points sont différentes.

Une première approche consiste à rendre l'erreur de prédiction identique pour tous les points situés à égale distance du centre du domaine, c'est -à-dire rendre la fonction de variance indépendante de l'orientation. Cette propriété est appelée isovariance par rotation. On parle d'une isovariance par rotation (ou rotatabilité) lorsque les réponses calculées avec le modèle aient une erreur de prévision identique pour des points situés à la même distance du centre du domaine d'étude (figure 1).

Le modèle mathématique permet de calculer la valeur de la réponse (Y_{calc}) en tout point du domaine expérimental.

$$Y_{\text{calc}} = b \cdot X_A$$

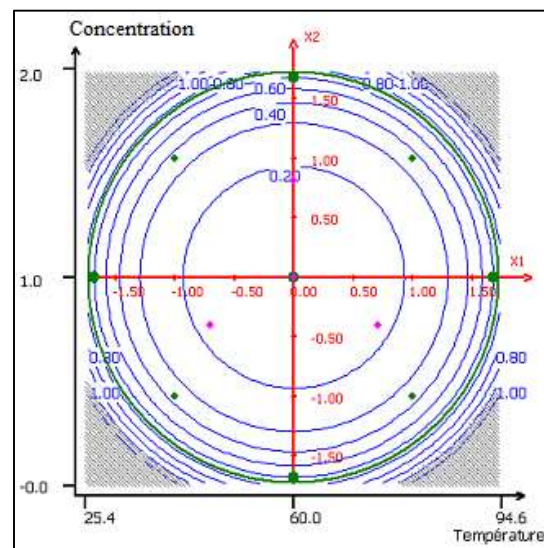


Figure 1 : Variation de la fonction de variance dans le plan : température, concentration

La fonction de la variance au point A est donnée par :

$$\text{Var}(Y_A) = d_A \sigma^2$$

Y_A : valeur de la réponse au point A,

σ^2 : variance de l'erreur expérimentale,

d_A : fonction de la variance au point A,

2. Orthogonalité

Une matrice X est *orthogonale* si sa matrice de dispersion est *diagonale*. La matrice de dispersion est donnée par $(X_A^t X_A)^{-1}$. Cette orthogonalité est respectée dans le cas du modèle polynomiale du premier degré, elle n'est pas respectée dans le modèle polynomiale du second degré. Dans ce cas, on exclut la première ligne et la première colonne de la matrice de dispersion. On obtiendra une sous matrice et parlera d'une *presque orthogonalité*.

Remarques

Pour toutes les matrices dédiées aux surfaces de réponse, il est toujours nécessaire de réaliser des points au centre du domaine. Ces points permettront :

- Obtenir des informations sur l'orthogonalité et l'isovariance par rotation,
- Obtenir des informations sur la variabilité de la réponse (estimer la variance expérimentale),
- Tester la validité du modèle.
-

III. Matrices d'expérience pour surface de réponse

1- Matrice composite centrée (PCC)

Les plans composites centrés permettent une modélisation du second degré. La démarche séquentielle est constituée de trois parties : plan factoriel, plan en étoile et points au centre.

Pour k facteurs, une matrice d'expérience composite centrée contient :

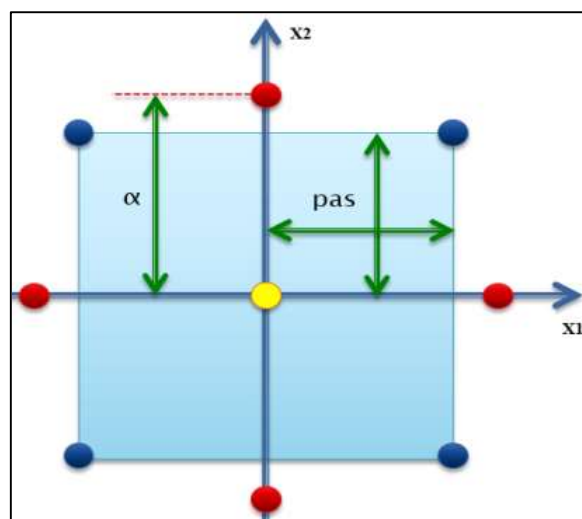
$N_f = 2^k$ expérience correspondante à une matrice d'expérience complète à deux niveaux,

$N_\alpha = 2k$ expérience placée sur les axes représentant les facteurs à une distance α du centre du domaine (points en étoile),

N_0 expériences au centre du domaine (figure 2).

Chaque facteur peut prendre cinq niveaux $(-\alpha, -1, 0, +1, +\alpha)$.

α qui est la distance du centre du plan à un point en étoile est appelé « bras de l'étoile ».



Le nombre total des expériences à faire est de :

$$N = N_f + N_\alpha + N_0 = 2^k + 2k + N_0$$

La condition d'isovariance par rotation nous a permis de déterminer la valeur de α , elle est donnée par :

$$\alpha = (n_f)^{\frac{1}{4}} = (2^k)^{\frac{1}{4}}$$

Par exemple pour : $k = 2 \rightarrow \alpha = (2^2)^{\frac{1}{4}} = 1.41$

Remarque

Si les points en étoiles (N_α) sont sur les faces du cube, on parle d'un plan composite à face centré (le domaine est cubique), on aura trois niveaux par facteurs (-1, 0 et +1).

Si le domaine est sphérique ($\alpha > 1$), le facteur prendra cinq niveaux (- α , -1, 0, +1, + α) et on aura un plan composite centré.

Les valeurs de α en fonction du nombre de facteur k sont données dans le tableau suivant :

Tableau I : Matrice composite centrée – Nombre d'expérience

K	2	3	4	5	6
Matrice factorielle	2^2	2^3	2^4	2^{5-1} (I=12345)	2^{6-1} (I=123456)
Nf	4	8	16	16	32
N_α	4	6	8	10	12
α	1,41	1,68	2,00	2,00	2,378
N_0 (théorie)	5	6	7	6	9
N	13	20	31	32	53

Exemple :

Pour $k = 2 \rightarrow N = 2^2 + 2.2 + 3$

Ainsi pour deux facteurs ($k = 2$), l'équation du modèle s'écrit :

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

La position des points expérimentaux est illustrée dans les figures suivantes :

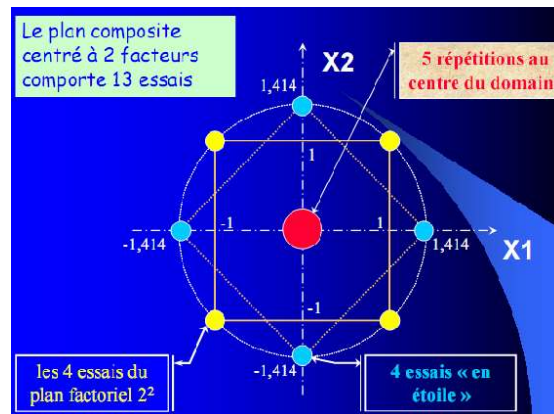


Figure 2 : Plan composite centré à deux facteurs

La matrice d'expérience est donnée par :

Tableau II : Matrice composite centrée à deux facteurs

N°		X1	X2
1	$2^2 = 4$	-1	-1
2		1	-1
3		-1	1
4		1	1
5	$2.2=4$	-1,414	0
6		1,414	0
7		0	-1,414
8		0	1,414
9	N_0	0	0
10		0	0
11		0	0
12		0	0
13		0	0

Pour $k = 3 \rightarrow N = 2^3 + 2.3 + 6 = 20$ expériences,

Le modèle proposé est donné par :

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 + \varepsilon$$

La position des points expérimentaux est illustrée dans les figures suivantes

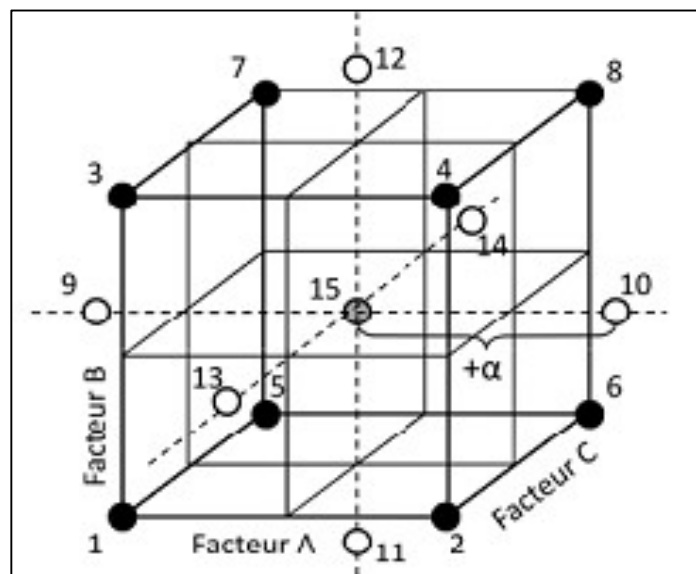
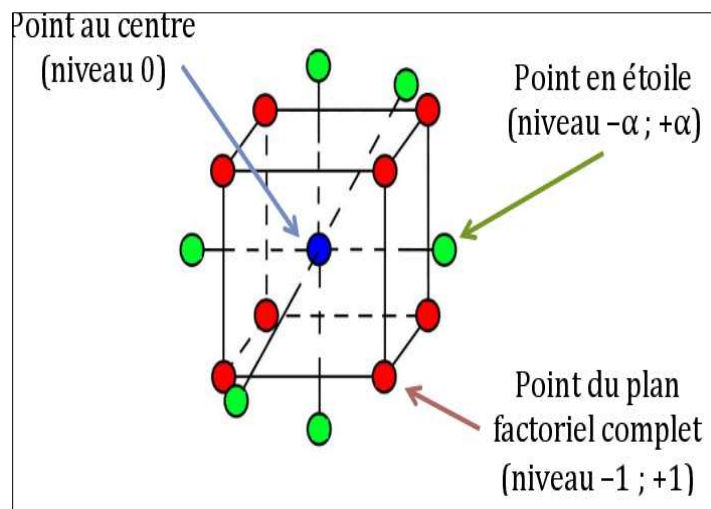


Figure 3 : Plan composite centré à trois facteurs

La matrice d'expérience est la suivante :

Tableau III : Matrice composite centrée à trois facteurs

N°		A	B	C
1	$2^3 = 8$	-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
9	$2 \cdot 3 = 6$	-1.682	0	0
10		1.682	0	0
11		0	-1.682	0
12		0	1.682	0
13		0	0	-1.682
14		0	0	1.682
15	Point au centre	0	0	0
16		0	0	0
17		0	0	0
18		0	0	0
19		0	0	0
20		0	0	0

2- Matrice Box Behnken

Les plans de Box-Behnken sont des plans économiques comportant moins d'essai que le plan composite centré correspondant au même nombre de facteurs pour localiser un optimum. Par exemple pour 3 facteurs, il n'y a que 15 expériences à faire au lieu de 20.

La matrice de Box Behnken est une matrice facile à mettre en œuvre, symétrique dont toutes les variables ne prennent que trois niveaux (-1, 0 et 1). Les points expérimentaux sont situés au milieu des arêtes (figure 2).

Le plan de Box-Behnken à trois facteurs est construit sur un cube.

Pour $k = 3 \rightarrow 15$ expériences,

Le modèle proposé est donné par :

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 + \varepsilon$$

La matrice d'expérience pour un modèle polynomiale du second degré de Box Behnken pour trois facteurs est la suivante :

Tableau IV : Matrice Box Behnken à trois facteurs

N	A	B	C
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	0	0	0
14	0	0	0
15	0	0	0

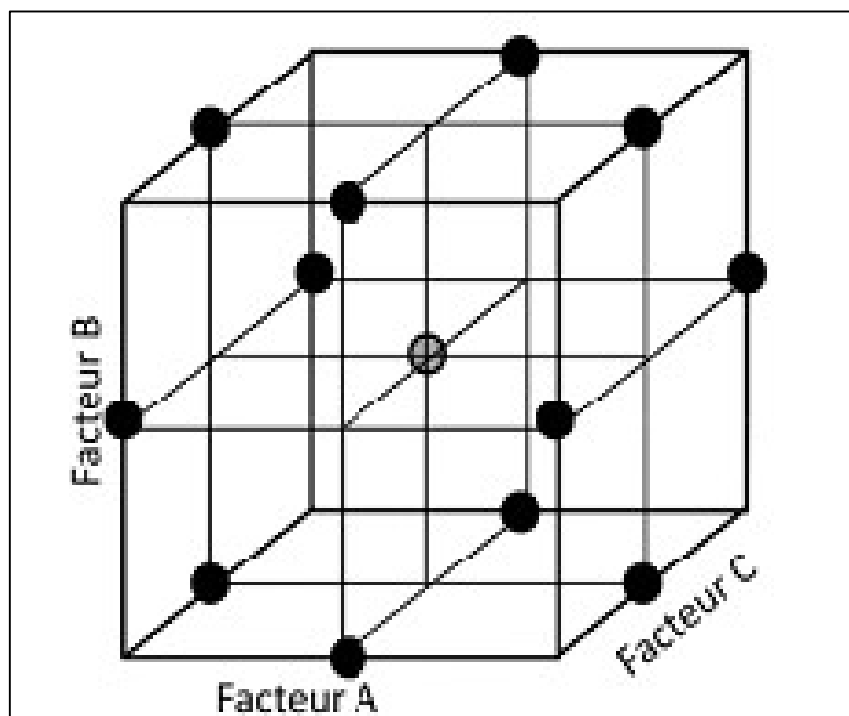
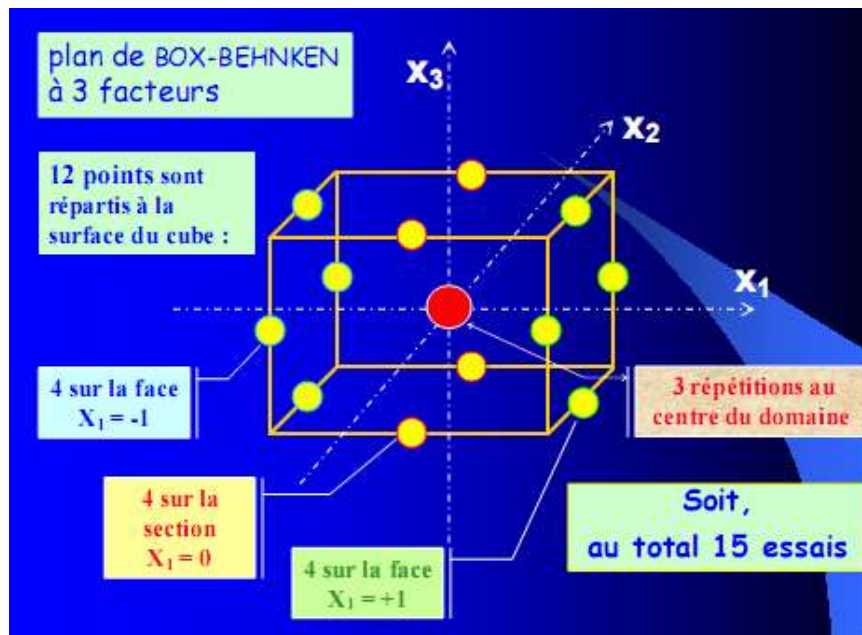


Figure 2 : Plan Box Behnken pour trois facteurs

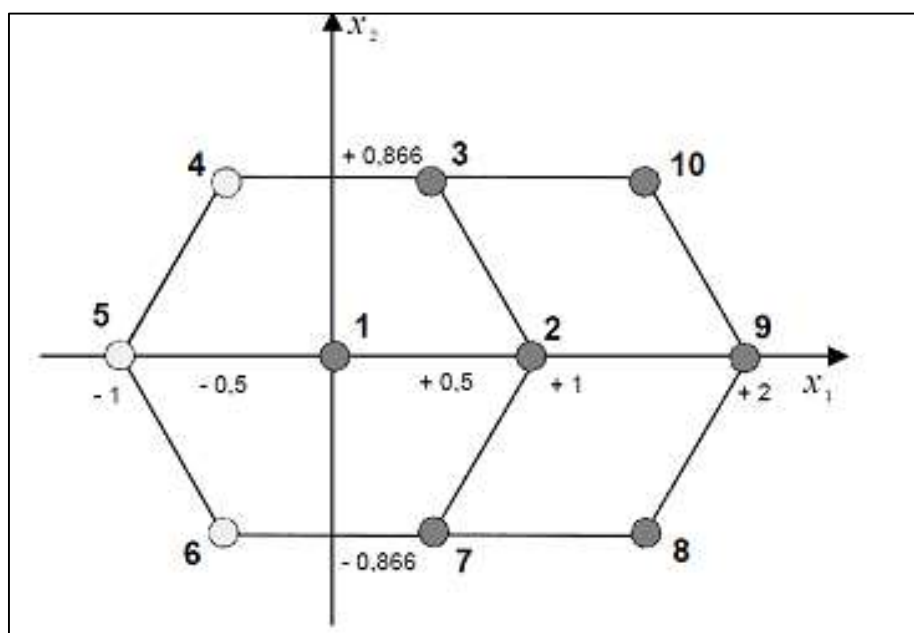
3- Matrice de Doehlert

Doehlert a proposé une distribution uniforme des points expérimentaux dans l'espace des facteurs codés. Le domaine défini par les plans de Doehlert est un domaine sphérique, un cercle dans un espace à deux dimensions, une sphère dans un espace à trois dimensions, une hyper sphère dans un espace à plus de trois dimensions. Le nombre d'expérience à réaliser pour k facteur est donné par la relation :

$$N = k^2 + k + 1$$

Le nombre de niveaux des facteurs est différent dans ce cas, 5 niveaux pour la première variable (-1, -0.5, 0, +0.5 et +1), trois niveaux pour la dernière variable et 7 niveaux pour les autres.

La caractéristique principale des plans de Doehlert est d'avoir une répartition uniforme des points expérimentaux dans l'espace expérimental. La Figure ci-dessous donne la disposition de ces points pour un plan à deux facteurs (essais 1 à 7). Tous les points sont à la même distance du centre du domaine d'étude et sont situés sur le cercle trigonométrique. Ils forment un hexagone régulier.



Les points 1 à 7 illustrent un premier plan de Doehlert. Les trois points 8, 9 et 10 illustrent les expériences supplémentaires. Les points 2, 7, 8, 9, 10, 3 et 1 illustrent un deuxième plan de Doehlert.

Si l'expérimentateur désire explorer le domaine expérimental, il peut facilement ajouter des points d'expériences supplémentaires et retrouver une disposition identique à celle de départ.

La Figure montre qu'avec trois points d'expériences supplémentaires (essais 8, 9 et 10), on peut obtenir un nouveau plan de Doehlert (essais 2, 7, 8, 9, 10, 3 et 1). Ce type de plans existe pour un nombre quelconque de facteurs.

Matrice de Doehlert pour modèle polynomiale du second degré

Tableau V : Matrice de Doehlert à trois facteurs

N°	A	B	C
1	0	0	0
2	1	0	0
3	0,5	0,866	0
4	-0,5	0,866	0
5	-1	0	0
6	-0,5	-0,866	0
7	0,5	-0,866	0
8	0	0	0
9	-0,5	-0,289	0,816
10	0	-0,577	0,816
11	0,5	0,289	0,816
12	-0,5	-0,289	-0,816
13	0	0,577	-0,816
14	0,5	-0,289	-0,816
15	0	0	0

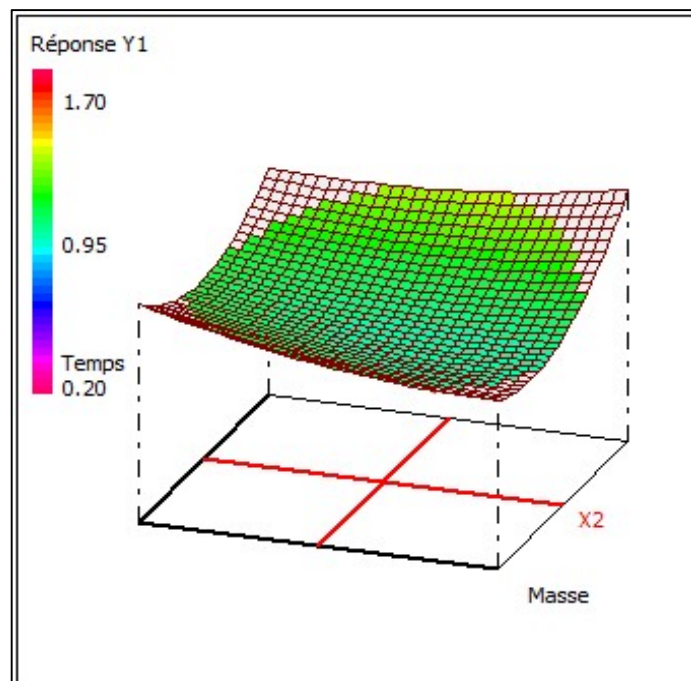
$$\frac{1}{2\sqrt{3}} = 0.289; \quad \frac{1}{\sqrt{3}} = 0.577 \quad ; \quad \frac{\sqrt{2}}{3} = 0.816; \quad \frac{\sqrt{3}}{2} = 0.866$$

IV. Recherche des conditions optimales

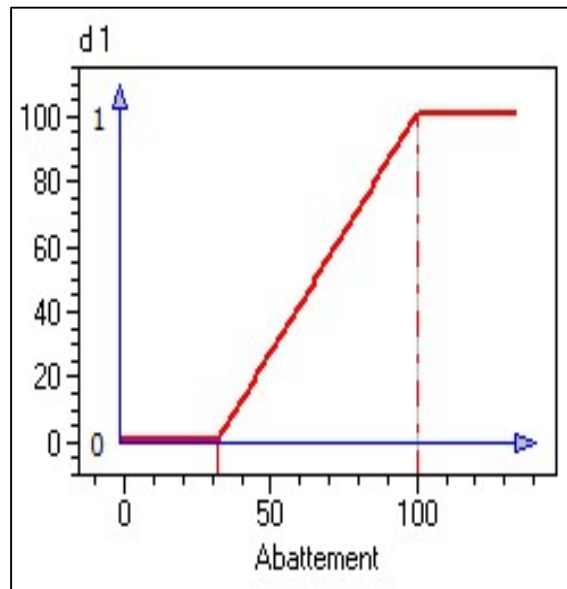
La recherche des conditions optimales se fait avec plusieurs méthodes. Elle peut se faire manuellement si le nombre des facteurs à optimiser est de deux mais au-delà, il faut utiliser des logiciels tel que modde6, statistica, JMP, Nemrodw...etc.

La détermination de l'optimum peut se faire avec :

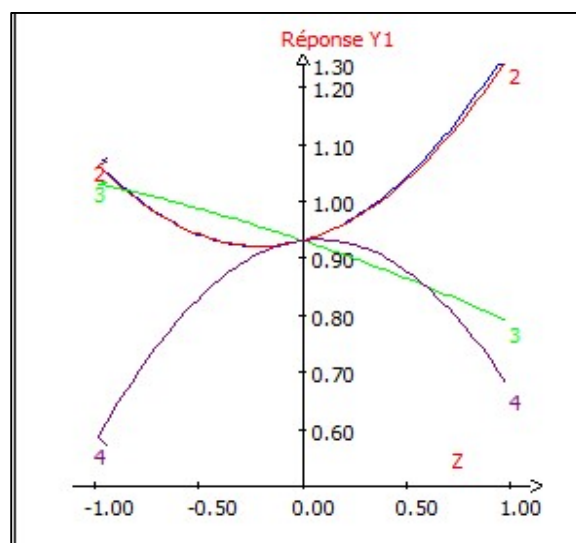
- **Surface d'isoréponse** : Un graphique de contour affiche une vue en deux ou trois dimensions de la surface, où les points ayant la même réponse sont reliés pour produire des lignes de contour de réponses constantes. Les graphiques de contour permettent d'établir les valeurs de réponses et les conditions d'exploitation souhaitables.



- **Fonction de désirabilité** : c'est l'une des approches les plus utilisées lors de l'optimisation simultanée de plusieurs réponses. Les valeurs des fonctions de désirabilité (d_i) sont comprises entre 0 et 1. Une désirabilité de 0 appelée aussi « désirabilité élémentaire » prenant la valeur nulle, représente une solution inacceptable pour l'objectif choisi, alors qu'une désirabilité prenant la valeur 1 indique la performance maximale souhaitée donc aucune amélioration n'est plus possible.



- **Analyse canonique (AC)** : c'est une méthode de statistique descriptive multidimensionnelle qui permet la construction et l'interprétation de graphiques. L'objectif général de l'analyse canonique est d'explorer les relations pouvant exister entre des variables quantitatives observées sur la réponse mesurée.



Références

J. Goupy, Introduction aux plans d'expériences, Dunod- Paris, 336 pages, 2001,

J. Goupy, Plans d'expériences pour surfaces de réponse, Dunod. Paris, 414 pages, 1999,

J. Goupy, Les plans d'expériences, Revue modulad, 34, 2006, pp74-116,

J. Goupy, étude comparative des divers plans d'expérience, revue statistique appliquée, tome 38 N°4, 1990, p5-44.

Walter Tinsson, Plan d'expérience : constructions et analyses statistiques, Mathématiques et applications, Springer, 535pages, 2010.

R. Bchitou, Chimiométrie universite Mohammed V, Agdal cours de master chimie faculté des sciences département de chimie, Rabat. <https://www.studocu.com/row/document/universite-abdelmalek-essaadi/chimie/chimiometrie/46576532>

Claude Hoinard, LES PLANS FACTORIELS COMPLETS, Laboratoire de Biophysique et Mathématiques, Faculté de Pharmacie de Tours, 2009

http://unt-ori2.crihan.fr/unspf/2010_Tours_Hoinard_PlansExperimentaux/res/1PFCdiapo.pdf

Claude Hoinard, LES PLANS FACTORIELS fractionnaires, Laboratoire de Biophysique et Mathématiques, Faculté de Pharmacie de Tours, 2009

<https://www.etudier.com/dissertations/Plan-Factoriel-Et-Fractionnaire/65405377.html>

Claude Hoinard, LES méthodes d'optimisation, Laboratoire de Biophysique et Mathématiques, Faculté de Pharmacie de Tours, 2009

http://unt-ori2.crihan.fr/unspf/2010_Tours_Hoinard_PlansExperimentaux/res/3OPTpolyp.pdf

Philippe TRIBOULET, NOTIONS DE BASES SUR LES PLANS D'EXPERIENCES, 2008, <https://pdfslide.tips/documents/notions-de-bases-sur-les-plans-dexperiences-i-2014-01-22-philippe-triboulet.html>

Jean-Pierre Gauchi. Introduction à la méthode des plans d'expériences. ISUP (Institut de Statistique de l'Université Pierre et Marie Curie) - édition 2016 (Introduction à la planification expérimentale), 2016, 403 pages, hal-02795365.

Thomas Silou, Marc Malanda, Laurent Loubaki, Optimisation de l'extraction de l'huile essentielle de *Cymbopogon citratus* grace à un plan factoriel complet 2^3 , Journal of Food Engineering 65(2):219-223,

Torbjorn Lundstedt ,Elisabeth Seifert ,Lisbeth Abramo , Bernt Thelin , Asa Nystrom, Jarle Pettersen, Rolf Bergman, Experimental design and optimization, Chemometrics and Intelligent Laboratory Systems 42, (1998)pp3- 40.

F. Rais, A. Kamoun, M. Chaabouni, M. Claeys-Bruno, R. Luu, M. Sergent, Etude par plans d'expériences de la robustesse d'un procédé de préparation de sulfate-amide des acides gras de l'huile de grignons d'olives, Journal de la Société Chimique de Tunisie, 2011, 13, 191-202.