REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

Université de Saïda – Dr. Moulay Tahar
Faculté des Sciences et de la Technologie
Département : Génie des Procédés

Polycopié Pédagogique

Intitulé :

Plans d'expériences

Présenté par :

Dr. ARDJANI Taki Eddine Ahmed

Maître de Conférences A (MCA)

Pour les étudiants de la 2ème année Master en Génie Pharmaceutique

Semestre S3

Année universitaire : 2025-2026

L'avant-propos

Ce polycopié de cours, conforme au programme enseigné, s'adresse aux étudiants de $2^{ème}$ année du Master Génie Pharmaceutique. Il est le fruit d'un travail personnel issu de ma propre expérience pédagogique dans le cadre de l'enseignement du module Plans d'expériences.

Ce support a pour objectif de fournir aux étudiants une base solide et méthodologique pour la conception, l'analyse et l'optimisation d'expériences, compétence essentielle dans les industries de recherche, de développement et de production pharmaceutique.

Ce module est présenté en cinq chapitres :

- Le premier chapitre pose les fondements généraux et introduit les principes des plans factoriels complets, avec des applications immédiates.
- Le deuxième chapitre approfondit l'analyse statistique en se concentrant sur la signification des effets et les méthodes rigoureuses de validation des modèles obtenus.
- Le troisième chapitre est consacré aux plans factoriels fractionnaires, une approche indispensable pour le criblage efficace d'un grand nombre de facteurs avec un minimum d'essais.
- Le quatrième chapitre traite des plans de surface de réponse, dédiés à la modélisation fine et à l'optimisation de procédés, en présentant les plans composites centrés et Box-Behnken.
- Le cinquième chapitre aborde la spécificité des plans de mélange, cruciaux pour les problématiques de formulation rencontrées en ingénierie pharmaceutique.

Tout au long de ces chapitres, une place centrale est accordée à la mise en pratique. Des études de cas, des exemples détaillés étape par étape et des exercices d'application, souvent illustrés, visent à concrétiser les concepts théoriques et à faciliter leur assimilation.

Nous clôturons ce polycopié par une bibliographie indicative, rassemblant les ouvrages et ressources de référence qui ont alimenté sa rédaction et qui pourront accompagner l'approfondissement des connaissances des étudiants.

Sommaire

Sommaire

Chapitre III: Plans Factoriels Fractionnaires

Sommaire

Sommaire

Liste des figures

Liste des tableaux

Introduction

Dans l'industrie et la recherche, nous sommes constamment confrontés à des questions:

- Comment améliorer le rendement de cette réaction chimique?
- Quels paramètres de réglage de cette machine permettent de réduire le nombre de défauts?
- Quelle est la combinaison d'ingrédients qui maximise le goût de ce produit alimentaire?

La méthode intuitive est de tester les facteurs un par un. Cette approche est non seulement longue et coûteuse, mais elle rate complètement un aspect crucial: les interactions entre les facteurs.

Les plans d'expériences (DoE (Design of Experiments))

Les plans d'expériences sont une méthodologie statistique qui nous permet de:

- Concevoir des expériences de manière rationnelle;
- Analyser les résultats pour identifier les facteurs influents et leurs interactions;
- Optimiser le processus pour atteindre l'objectif fixé (maximiser, minimiser, atteindre une cible).

1. Qu'est-ce qu'un plan d'expérience?

Définition: Un plan d'expérience est un protocole organisé qui définit la manière de recueillir des données de façon à obtenir le maximum d'informations pertinentes avec un minimum d'essais, tout en permettant une analyse statistique rigoureuse.

Analogie: Imaginez chercher une salle dans un bâtiment inconnu.

- **Sans plan:** Vous ouvrez les portes au hasard. C'est long et vous pouvez passer à côté de la bonne salle;
- **Avec un plan:** Vous avez un plan du bâtiment. Vous suivez un chemin logique (par étage, par couloir). C'est systématique, rapide et efficace. Un plan d'expérience, c'est le "plan du bâtiment" pour votre étude.

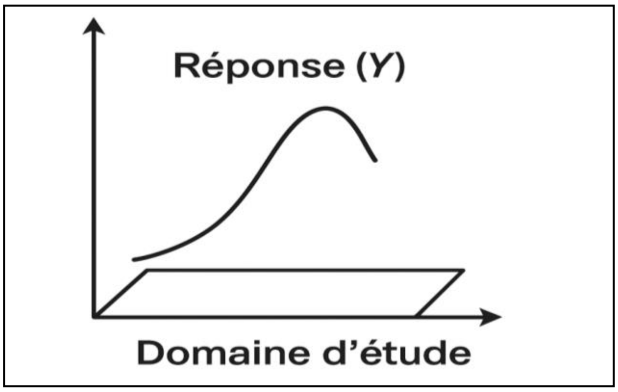
2. Domaine d'étude et surface de réponse

2.1. Domaine d'étude

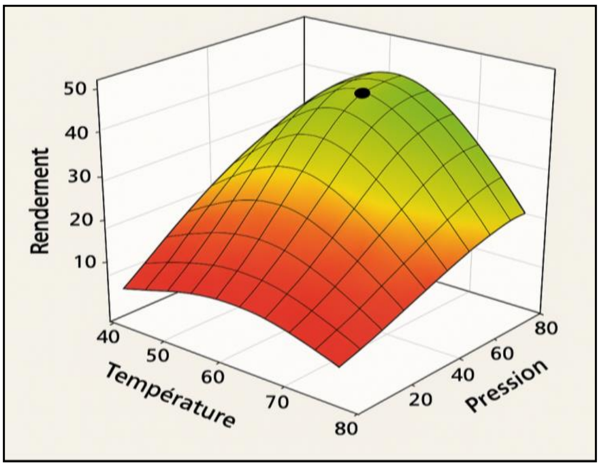
C'est l'espace défini par l'ensemble des valeurs que peuvent prendre tous les facteurs que vous avez décidé d'étudier. Si vous avez deux facteurs (Température, Pression), le domaine d'étude est un rectangle.

2.2. Surface de réponse

C'est la représentation graphique de la relation entre les facteurs et la réponse mesurée. C'est une "carte" de la performance du processus.



Exemple: Si votre réponse est le Rendement **(Y)** et vos facteurs sont la Température **(X_1)** et la Pression **(X_2)**, la surface de réponse est un paysage en 3D où l'altitude est le rendement.



L'objectif est de trouver les coordonnées (X_1, X_2) qui nous amènent au sommet (O) de cette surface.

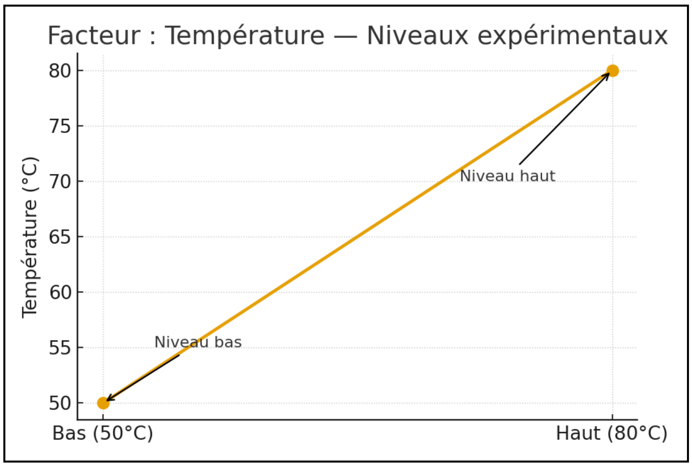
3. Les Facteurs

Définition: Un facteur est une variable indépendante dont on suspecte qu'elle peut influencer la variable réponse. On les classe ainsi:

- **Facteurs qualitatifs**: Nature non numérique (**ex:** Type de catalyseur A, B, C; Marque de raw material).
- **Facteurs quantitatifs**: Nature numérique (**ex:** Température 50°C, 70°C; Temps 10 min, 20 min).

Pour chaque facteur, on définit des niveaux: ce sont les valeurs spécifiques auxquelles le facteur sera testé lors de l'expérimentation.

Exemple: Pour un facteur "Température", les niveaux pourraient être: Niveau bas = 50°C, Niveau haut = 80°C.



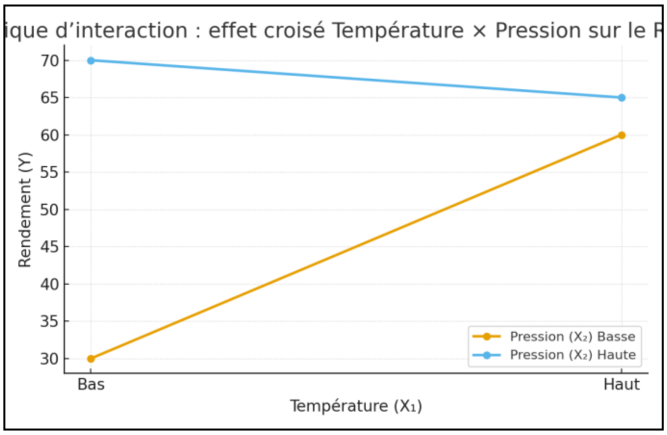
4. Notion d'interaction

C'est un concept fondamental et souvent négligé.

Définition: On dit qu'il y a une interaction entre deux facteurs lorsque l'effet d'un facteur sur la réponse dépend du niveau de l'autre facteur.

Exemple: L'effet de la "Température" (X_1) sur le "Rendement" (Y) n'est pas le même si la "Pression" (X_1) est basse ou haute.

<u>Représentation graphique: Graphique d'interaction</u>



- **Pas d'interaction:** Les deux lignes sont parallèles. L'effet de la température est le même, quelle que soit la pression.
- **Interaction présente:** Les deux lignes ne sont pas parallèles (elles se croisent ou divergent). L'effet de la Température sur le Rendement change selon le niveau de Pression.

5. Blocage et Réplication

Le blocage est une technique expérimentale visant à éviter toute variation indésirable du processus d'entrée ou du processus expérimental. Par exemple, une expérience peut être menée avec le même équipement pour éviter toute variation de l'équipement. Les praticiens répliquent également des expériences, en effectuant la même combinaison plusieurs fois, afin d'obtenir une estimation de la quantité d'erreur aléatoire pouvant faire partie du processus.

6. Matrice d'expériences

La matrice d'expériences montre toutes les combinaisons possibles des niveaux haut et bas pour chaque facteur d'entrée. Ces niveaux haut et bas peuvent être codés +1 et -1. Par exemple, une expérience à 2 facteurs nécessitera 4 essais.

Un plan pour lequel nous avons k facteurs est appelé un plan 2^k. le nombre d'expériences à réaliser sera donc 2^k expériences. Ce nombre devient rapidement très important. Par exemple pour seulement 7 facteurs, il faudrait 128 expériences. Pour diminuer le nombre des essais en conservant la possibilité d'étudier tous les facteurs, les plans factoriels fractionnaires à deux niveaux ont été proposés.

Une matrice d'expériences peut se construire à la main en suivant l'algorithme de Yates; pour une colonne *c* nous alternons: une série de 2^{c-1} (-1) et de 2^{c-1} (+1). La figure (2.2) montre les matrices d'expériences pour deux plans factoriels complets 2^2 et 2^3.

Exp	x_1	x_2
1	-1	-1
2	+1	-1
3	-1	+1
4	+1	+1

Exp	x_1	x_2	x_3
1	-1	-1	-1
2	+1	-1	-1
3	-1	+1	-1
4	+1	+1	-1
5	-1	-1	+1
6	+1	-1	+1
7	-1	+1	+1
8	+1	+1	+1

Figure I.1: Matrices d'expériences pour les plans factoriels complets 2^2 et 2^3.

7. Notion de modèle et de régression linéaire multiple

Pour décrire mathématiquement la surface de réponse, nous utilisons un modèle.

<u>Définition d'un modèle</u>: C'est une équation mathématique qui lie la variable réponse Y aux facteurs $X_1, X_2, ..., X_k$.

La forme la plus simple est le modèle linéaire multiple:

$$\mathbf{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \varepsilon}$$

Où:

- **Y**: Variable réponse (**ex:** Rendement);
- $\boldsymbol{\beta_0}$: Ordonnée à l'origine (la moyenne générale);
- $\boldsymbol{\beta_1, \beta_2}$: Coefficients des effets principaux (la pente de l'effet de chaque facteur);
- $\mathbf{X_1, X_2}$: Variables représentant les facteurs (souvent codées entre -1 et +1);
- $\boldsymbol{\varepsilon}$: Terme d'erreur (bruit, variabilité non expliquée).

Un modèle peut inclure des termes d'interaction (**ex:** $\beta_{ij}X_iX_j$) et des termes quadratiques (**ex:** $\beta_{ij}X_i^2$) pour capturer des courbures.

8. Variables Codées

l'utilisation des variables centrées réduites ou codées (-1,+1) présente l'intérêt de pouvoir généraliser la théorie des plans d'expériences quels soient les facteurs ou les domaines d'études retenus. Utiliser les variables codées va permettre d'avoir pour chaque facteur le même domaine de variation [-1,+1] et de pouvoir ainsi comparer entre eux les effets des facteurs, et ce sont ces variables codées (sans dimensions et entre -1 et +1) qui sont utilisées dans l'équation du modèle obtenu.

Exemple: nous choisissons pour le facteur température un domaine d'étude: [10°C, 40°C], nous affectons la valeur -1 à 10°C et +1 à 40°C et nous notons V_e la valeur codée et V_n la valeur naturelle. La figure (1.2) montre la correspondance proportionnelle entre les valeurs codées et naturelles.

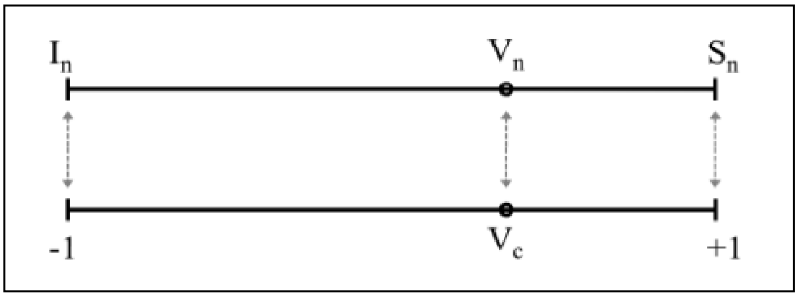


Figure I.2: Domaines de variation naturels/codés.

Ou S_n et I_n sont respectivement la limite supérieure et inférieure du domaine d'étude de la variable température. Les variables centrées réduites sont sans unité. Une température de 20°C correspond à une variable centrée réduite de -0.33. Les formules de conversion sont:

$$\mathbf{V_e = \frac{2V_n-(S_n+I_n)}{(S_n-I_n)}} \qquad \mathbf{(1.1)}$$

$$\mathbf{V_n = \frac{V_e(S_n-I_n)+(S_n+I_n)}{2}} \qquad \mathbf{(1.2)}$$

9. Objectifs d'utilisation des plans d'expériences

9.1. Objectif de Comparaison d'alternatives

Dans le cas de notre exemple de la cuisson du gâteau, nous pourrions vouloir les résultats deux différents types de farine. S'il s'avérait que la farine des différents fournisseurs avait un résultat non significatif, nous pourrions choisir le fournisseur le moins coûteux. Si le résultat était significatif, nous choisirions celui de la meilleure farine.

9.2. Objectif de Criblage

Il existe souvent de nombreux facteurs possibles, dont certains peuvent être critiques et d'autres qui ne peuvent avoir que peu ou pas d'effet sur la réponse. En tant que but en soi, il est peut être souhaitable de réduire le nombre de facteurs à un ensemble relativement petit (2 à 5), de manière à concentrer l'attention sur le contrôle de ces facteurs. Les expériences de criblage sont un moyen efficace, avec un nombre minimal de tests, pour déterminer les facteurs importants. Nous pourrions poser une question: Quels sont les facteurs importants qui ont un impact sur la couleur du gâteau parmi: la température, la durée de la cuisson, la qualité du sucre et de la farine, le nombre d'œufs et l'utilisation du zeste de citron?

9.3. Objectifs de Modélisation

La méthodologie de surface de réponse est un ensemble de techniques mathématiques et statistiques pour la construction de modèles mathématiques empiriques. L'exploitation de ces modèles peut répondre à plusieurs objectifs:

a. Objectif de Prédiction

Un modèle mathématique obtenu peut être utilisé pour prédire une réponse à un point quelconque dans les limites des domaines expérimentaux. La précision des réponses obtenues dépend du degré de capacité de prédiction du modèle.

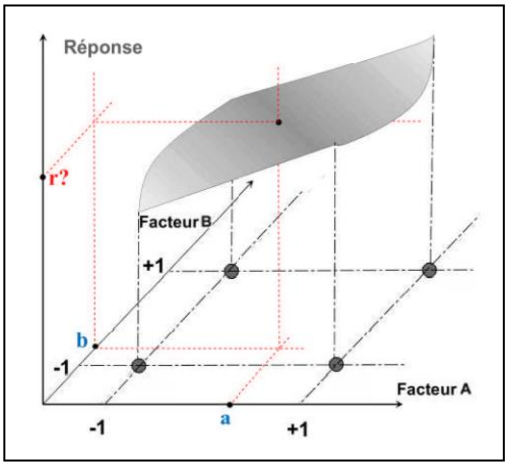


Figure I.3: Exemple graphique de prédiction de la réponse.

b. Objectif d'Optimisation

Une optimisation est effectuée pour déterminer les valeurs des entrées du processus à utiliser pour obtenir la sortie du processus souhaitée. Les objectifs d'optimisation habituels peuvent être de maximiser le rendement d'un processus, de minimiser le temps de traitement nécessaire à la fabrication d'un produit ou d'atteindre une spécification du produit cible.

Nous pourrions poser une question: Comment rendre le gateau aussi que possible sans le désintégrer?.

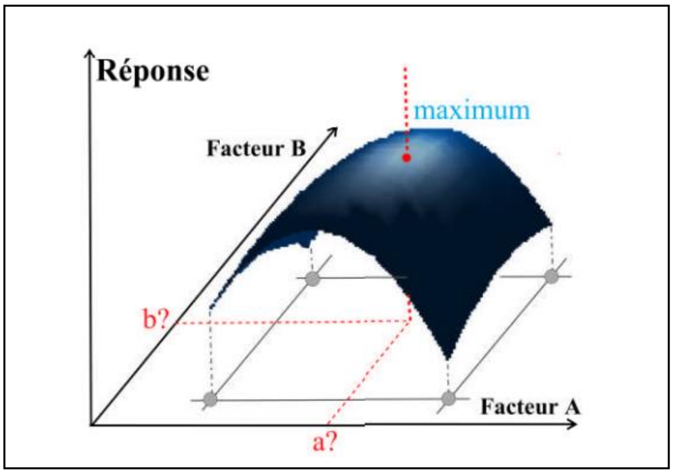


Figure I.4: Exemple graphique de l'optimisation de la réponse.

c. Objectifs d'optimisation multi-réponse (Equilibre les compromis)

L'objectif est de déterminer les paramètres des facteurs permettant d'optimiser simultanément (un compromis) plusieurs réponses.

10. Quelles sont les étapes des plans d'expériences?

Obtenir de bons résultats par un PEX implique les sept étapes suivantes:

a. Poser des objectifs

Les objectifs d'une expérience sont lieux déterminés par une discussion en équipe. Tous les objectifs doivent être écrits, même ceux qui ne semblent pas très intéressants. Le groupe devrait discuter des objectifs clés et des objectifs "Bons mais pas vraiment nécessaires". La priorisation des objectifs vous aide à choisir la direction à prendre en ce qui concerne la sélection des facteurs, les réponses et un plan particulier.

b. Sélectionner les variables du processus eu leurs plages

les variables d'un processus incluent à la fois les entrées et les sorties - c'est-à-dire les facteurs et les réponses. La sélection de ces variables s'effectue mieux en équipe. L'équipe devrait:

- Inclure tous les facteurs importants (Selon le jugement de l'ingénieur);

- Déterminer une plage d'étude pour chaque facteur;

- Etre prudent en choisissant les niveaux bas et haut des facteurs;

- vérifier les réglages des facteurs pour des combinaisons non pratiques ou impossibles;

- inclure toutes les réponses pertinentes.

c. Choisir un plan expérimental: Le choix d'un plan expérimental dépend de l'expérience, du nombre de facteurs à étudier et da la quantité de ressources disponibles. Le tableau (1.1) montre un simple guide exemple de sélection d'un plan.

Tableau I.1: Guide de sélection d'un plan.

Nombre de facteurs	Objectif comparatif	Objectif de dépistage	Surface de réponse
1	Plan totalement aléatoire à 1 facteur	-	-
2-4	Plan de bloc aléatoire	Factoriel complet ou fractionnaire	Central composite ou Box-Behnken
5 ou plus	Plan de bloc aléatoire	Factoriel fractionnaire po Plackett-Burman	Dépistage en premier pour réduire le nombre de facteurs

d. Exécuter le plan: Les expériences du plan sont réalisées et les résultats des essais (réponses) sont rassemblés.

e. Vérifier que les données sont cohérentes aves les hypothèses expérimentales

Dans tous les modèles, nous formulons des hypothèses et nous exigeons également que certaines conditions soient approximativement remplies à des fins d'estimation. Ceux-ci sont:

- Les systèmes de mesure sont-ils capables de répondre à toutes vos questions?.

- Vos réponses sont-elles susceptibles d'être bien approchées par de simples modèles polynomiaux?.

- Les résultats (la différence entre les prédictions du modèle et les observations réelles) se sont-ils bien comportés?.

f. Analyser et interpréter les résultats

En supposant l'existence d'un modèle de départ que nous souhaitons l'adapter à nos données expérimentales et que l'expérience ait été conçue correctement pour notre objectif. La plupart des progiciels PEX analyseront ces données et peuvent fournir plusieurs statistiques numériques ainsi que graphiques.

11. Classes des plans d'expériences

Il existe trois grandes familles de plans d'expériences:

a. Les plans de criblages: dont l'objectif est de découvrir les facteurs les plus influents sur une réponse donnée en un minimum d'expériences.

b. Les plans pour surface de réponse: dont l'objectif est de trouver une relation mathématique (modèle) qui lie les réponses mesurées aux variables associées aux facteurs soit via une démarche mathématique analytique ou purement matricielle. Ce modèle peut être aussi utilisé à des fins d'optimisation du processus étudié.

c. Les plans de mélange: dont l'objectif est le même que la deuxième famille mais ou les facteurs ne sont pas indépendants et sont contraints. Comme exemple de contrainte, la fractions molaires d'un mélange doit être égale à 1.

12. Plan factoriel 2^k complet

C'est le plan de base pour étudier k facteurs.

a. Définition: Un plan factoriel 2^k est un plan où chaque facteur n'a que **deux niveaux** (généralement notés "-" ou "-1" pour le niveau bas, et "+" ou "+1" pour le niveau haut), et où toutes les combinaisons possibles de ces niveaux sont réalisées.

- Pour 2 facteurs : $2^2 = 4$ essais.
- Pour 3 facteurs : $2^3 = 8$ essais.

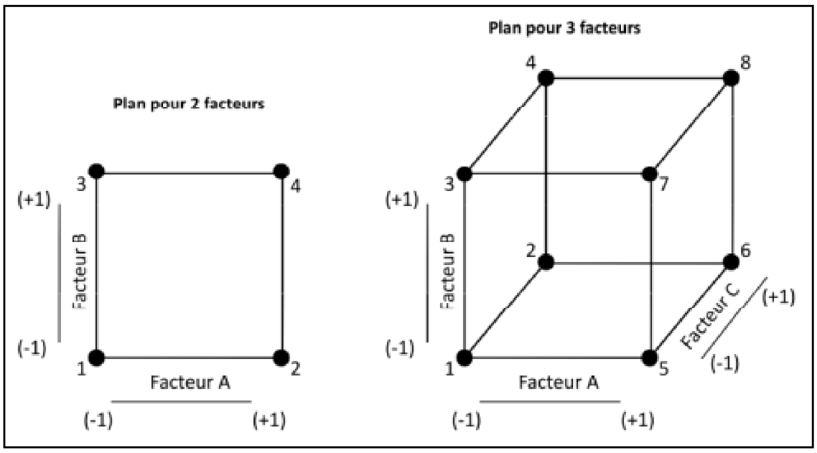


Figure I.5: Plans factoriels complets 2^2 et 2^3.

Ce plan permet d'estimer tous les effets principaux et toutes les interactions.

b. Exemple de calcul des effets

Scénario: On étudie le Rendement (Y) d'une réaction chimique avec 2 facteurs :

- **A** : Température (Bas: 80°C, Haut: 100°C);
- **B** : Concentration (Bas: 20%, Haut: 30%).

c. Matrice d'expériences et résultats:

Tableau I.2: Matrice d'expériences.

Essai	A (Temp)	B (Conc)	AB (Interaction)	Rendement Y
1	-1	-1	+1	60%
2	+1	-1	-1	72%
3	-1	+1	-1	54%
4	+1	+1	+1	80%

La colonne AB est le produit des colonnes A et B.

1. Calcul de l'effet principal d'un facteur:

L'effet est la différence moyenne de la réponse lorsque le facteur passe de son niveau bas à son niveau haut.

- **Effet de A (Température):**

Effet A = $[(Y_2 + Y_4) - (Y_1 + Y_3)] / 2 = [(72+80) - (60+54)] / 2 = (152 - 114) / 2 = 19$

Conclusion: En moyenne, augmenter la Température augmente le Rendement de 19%.

- **Effet de B (Concentration):**

Effet B = $[(Y3 + Y4) - (Y1 + Y2)] / 2 = [(54+80) - (60+72)] / 2 = (134 - 132) / 2 = 1$

Conclusion: L'effet moyen de la Concentration est faible (+1%).

- **Effet de l'interaction AB:**

Effet AB = [(Y1 + Y4) - (Y2 + Y3)] / 2 = [(60+80) - (72+54)] / 2 = (140 - 126) / 2 = 7

Conclusion: Il y a une interaction positive de 7%.

2. La représentation graphique des effets

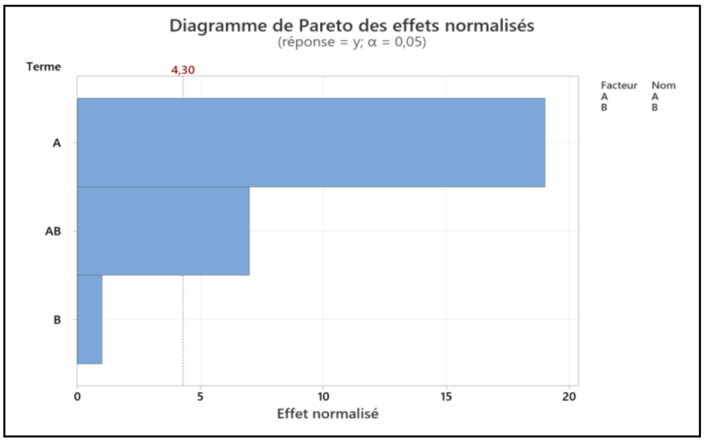


Figure I.6: Diagramme de Pareto des effets.

Ce graphique classe les effets par ordre d'importance. Seul l'effet A dépasse le seuil de significativité.

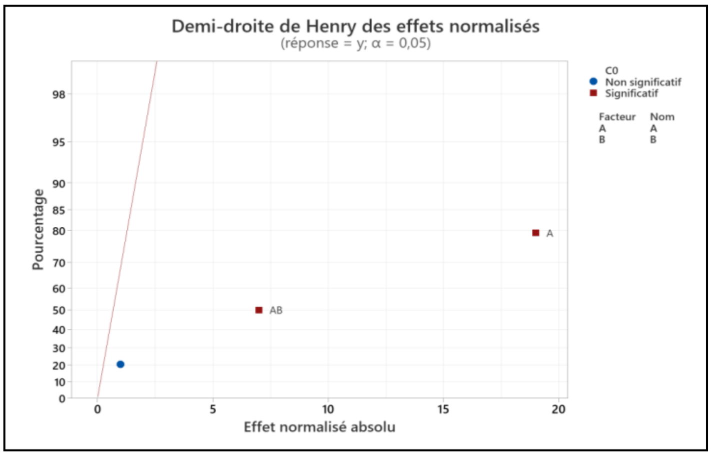


Figure I.7: Graphique des effets normaux.

On représente les effets sur un graphique de probabilité normale. Les effets non significatifs (proches de zéro, comme B) s'alignent sur la droite. Les effets significatifs (comme A et AB) s'écartent de la droite.

3. Forme matricielle - Régression multilinéaire

La même analyse peut être faite via un modèle de régression.

Notre modèle est: $\mathbf{Y = \beta_0 + \beta_1 A + \beta_2 B + \beta_{12} AB + \varepsilon}$

Sous forme matricielle: $\mathbf{Y = X\beta + \varepsilon}$

Avec:

- **Y** = Vecteur des réponses [60, 72, 54, 80];
- **X** = Matrice du modèle (avec une colonne de 1 pour l'ordonnée à l'origine);
- **β** = Vecteur des coefficients [β_0, β_1, β_2, β_{12}].

En utilisant l'algèbre matricielle, on trouve les coefficients : $\beta = (X'X)^{-1}X'Y$

Pour notre exemple, on trouverait :

- $\mathbf{\beta_0 = 66.5}$ (Moyenne générale);
- $\mathbf{\beta_1 = 9.5}$ (La moitié de l'effet A calculé précédemment);
- $\mathbf{\beta_2 = 0.5}$ (La moitié de l'effet B);
- $\mathbf{\beta_{12} = 3.5}$ (La moitié de l'effet AB).

L'équation du modèle prédictif est donc : $\mathbf{\hat{Y} = 66.5 + 9.5*A + 0.5*B + 3.5AB}$

Application 1: Optimisation d'un procédé de nettoyage

Énoncé:

Vous êtes ingénieur procédé dans une usine chimique. Vous souhaitez optimiser l'efficacité d'un bain de nettoyage. Après brainstorming, votre équipe identifie 2 facteurs influents:

Facteur A: Temps de nettoyage (10 min vs 20 min);

Facteur B: Concentration du détergent (2% vs 5%);

Variable réponse: Efficacité du nettoyage (mesurée par un score de 0 à 100%).

Données collectées:

Essai	Temps	Concentration	Efficacité
1	10 min	2%	60%
2	20 min	2%	72%
3	10 min	5%	54%
4	20 min	5%	80%

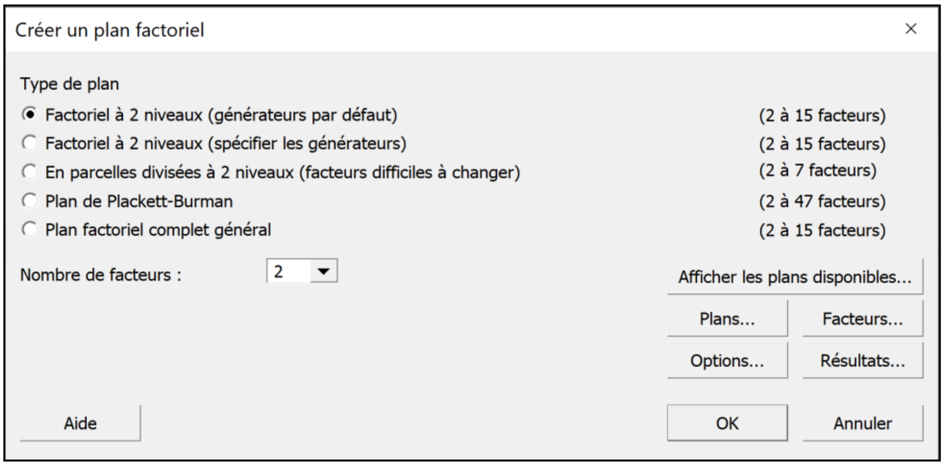
Consignes:

- Créer le plan dans Minitab;
- Analyser les effets principaux et interactions;
- Trouver les conditions optimales;
- Interpréter les résultats.

Solution étape par étape:

Étape 1: Création du plan dans Minitab

Ouvrir Minitab → Aller dans Stat > DOE > Factorial > Create Factorial Design

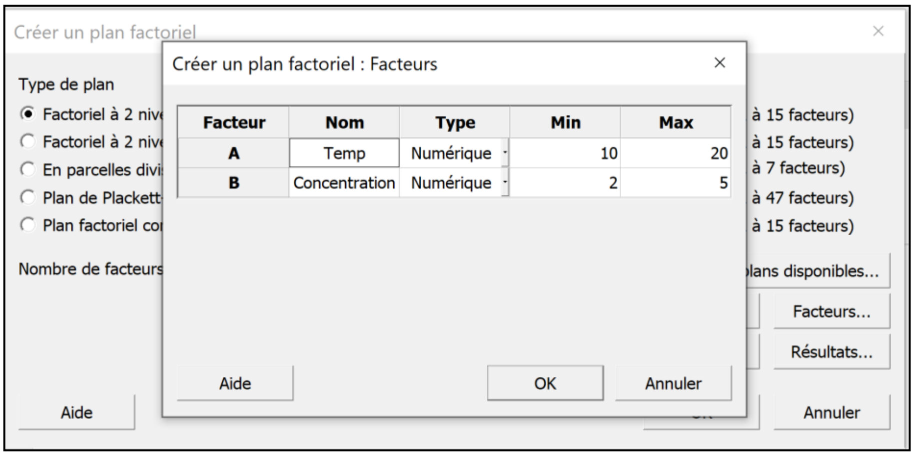


Type de plan: Choisir "2-level factorial (default generators)"

Nombre de facteurs: 2

Cliquer sur Designs: Vérifier que le nombre de runs est 4 → OK.

Cliquer sur Factors: Remplir comme suit:



Facteur A: Name = "Temps", Low = 10, High = 20

Facteur B: Name = "Concentration", Low = 2, High = 5

Cliquer sur Options : Décocher "Randomize runs" pour cet exercice → OK.

OK pour générer le plan.

Étape 2: Saisie des données

Dans la feuille de travail Minitab, entrer les données d'efficacité dans la colonne "Efficacité"

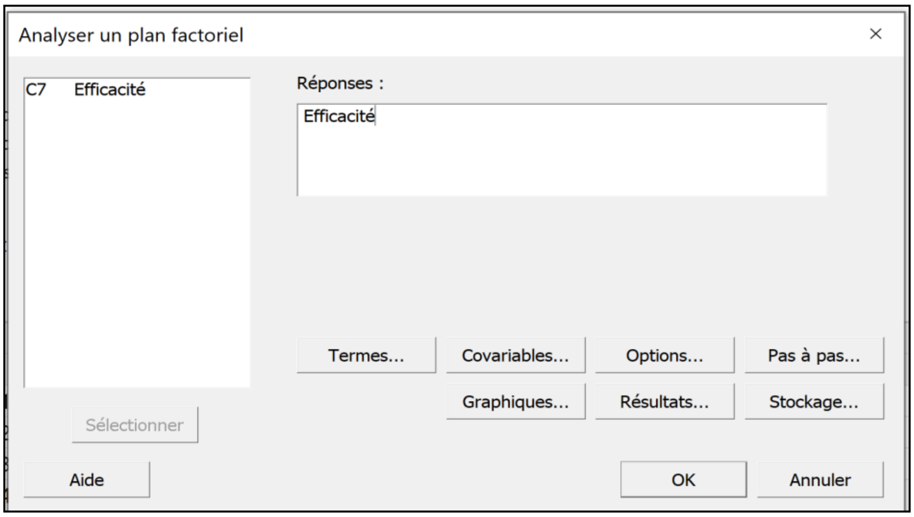
Votre feuille devrait ressembler à:

	C1	C2	C3	C4	C5	C6	C7	C8
	OrdreStd	OrdEssai	PtCentr	Blocs	Temp	Concentration	Efficacité	
1	1	1	1	1	-1	-1	60	
2	2	2	1	1	1	-1	72	
3	3	3	1	1	-1	1	54	
4	4	4	1	1	1	1	80	
5	5	5	0	1	0	0	72	
6	6	6	0	1	0	0	70	
7	7	7	0	1	0	0	71	

Étape 3: Analyse du plan

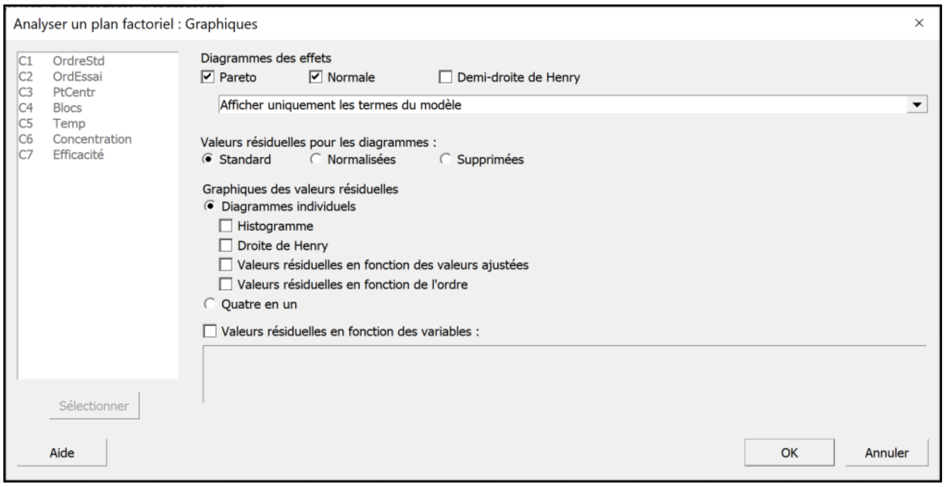
Aller dans: Stat > DOE > Factorial > Analyse Factorial Design.

Sélectionner "Efficacité" dans Responses.



Cliquer sur Terms: S'assurer que tous les termes (A, B, AB) sont inclus → OK.

Cliquer sur Graphs: Choisir "Pareto" et "Normal" pour les effets → OK.



OK pour lancer l'analyse.

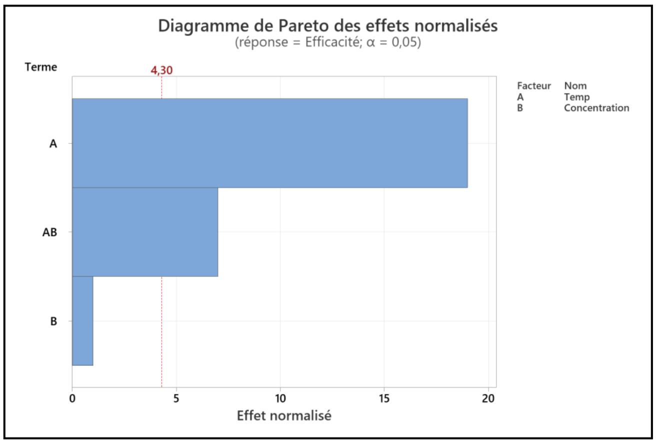
Étape 4: Interprétation des résultats

Résultats dans la fenêtre Session:

Coefficients codés

Terme	Effet	Coeff	Coef ErT	Valeur de T	Valeur de p	FIV
Constante		66,500	0,500	133,00	0,000	
Temp	19,000	9,500	0,500	19,00	0,003	1,00
Concentration	1,000	0,500	0,500	1,00	0,423	1,00
Temp*Concentration	7,000	3,500	0,500	7,00	0,020	1,00
Pt ctr		4,500	0,764	5,89	0,028	1,00

Graphique de Pareto des effets:

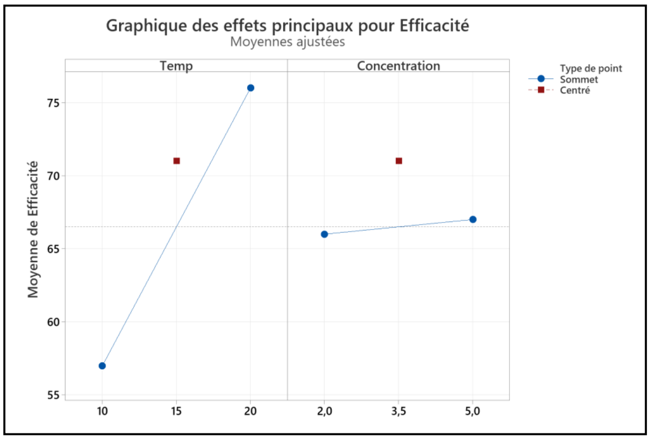


L'effet Temps (A) dépasse la ligne de référence → significatif

L'interaction Temps*Concentration (AB) est également significative

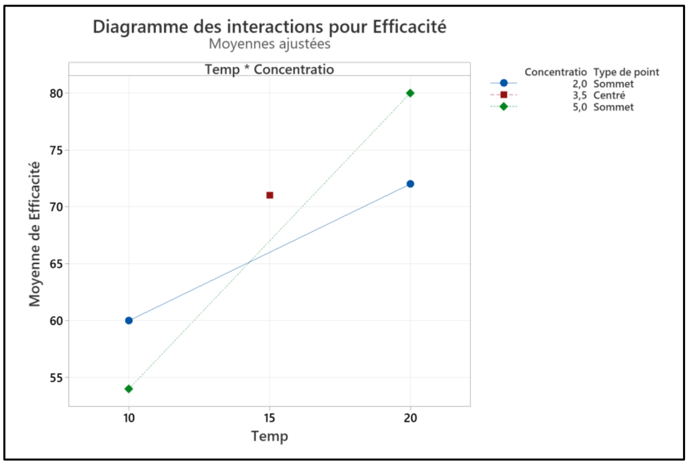
L'effet Concentration (B) seul n'est pas significatif

Graphique des effets principaux:



- La pente pour Temps est forte et positive → Temps élevé est meilleur
- La pente pour Concentration est presque plate → effet faible seul

Graphique d'interaction:



Les lignes ne sont pas parallèles → interaction présente.

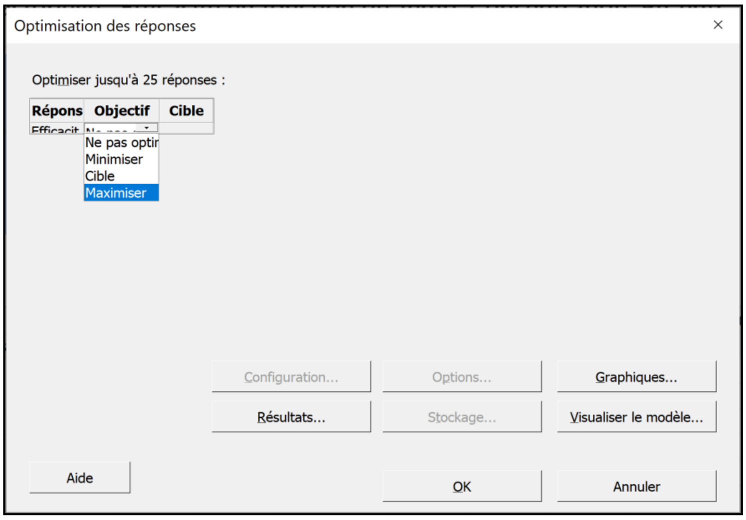
L'effet du Temps est plus fort quand la Concentration est haute.

<u>Étape 5: Optimisation</u>

Aller dans: Stat > DOE > Factorial > Response Optimiser.

Sélectionner "Efficacité".

Cliquer sur Setup : Choisir "Maximiser" → **OK.**



Solution proposée: Temps = 20 min, Concentration = 5%;

Efficacité prédite: 80%.

<u>**Conclusion:**</u> Pour maximiser l'efficacité du nettoyage, il faut utiliser le temps le plus long (20 min) et la concentration la plus élevée (5%). L'interaction significative indique que ces deux facteurs agissent en synergie.

Application 2: Réduction de la porosité en fonderie

Énoncé:

Dans une fonderie d'aluminium, vous cherchez à réduire les défauts de porosité. Trois facteurs sont étudiés :

A: Température du moule (500°C vs 600°C);

B: Temps de refroidissement (5 min vs 10 min);

C: Type de sable (Type1 vs Type2);

Variable réponse: Nombre de défauts de porosité par cm² (à minimiser).

Plan réalisé et résultats:

Essai	A	B	C	Défauts
1	-1	-1	-1	10
2	1	-1	-1	15
3	-1	1	-1	8
4	1	1	-1	12
5	-1	-1	1	9
6	1	-1	1	14
7	-1	1	1	7
8	1	1	1	11

Solution dans Minitab:

Étape 1: Création du plan 2^3

Stat > DOE > Factorial > Create Factorial Design;

Nombre de facteurs: 3;

Dans Facteurs:

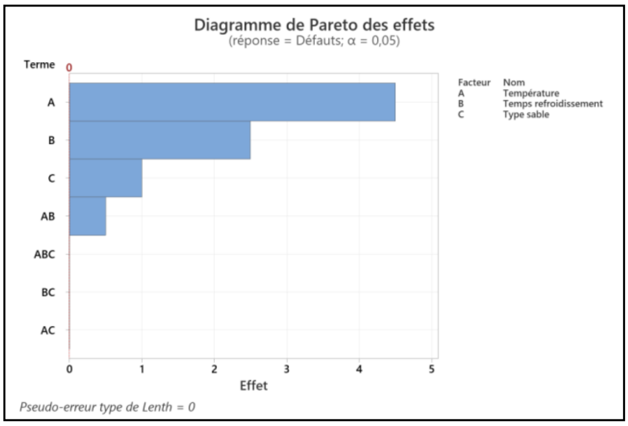
A: "Température" (500, 600);

B: "Temps refroidissement" (5, 10);

C: "Type sable" (1, 2) → Minitab gère les facteurs qualitatifs.

<u>Étape 2: Analyse et interprétation</u>

Graphique de Pareto :

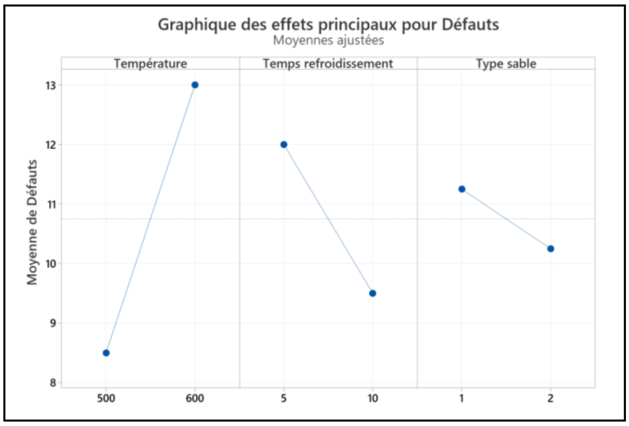


<u>Effets (du plus grand au plus petit):</u>

- A (Température): +4.5;
- B (Temps-refroidissement): -2.5;
- C (Type-sable): -1;
- AB (interaction): -0.5.

<u>Interprétation:</u>

- **B** a l'effet négatif → Temps de refroidissement long réduit les défauts.
- **C** a un effet négatif → Type2 est meilleur que Type1.
- **A** a un effet positif → Température basse est meilleure.



L'interaction AB est faible mais notable

<u>Étape 3: Optimisation</u>

<u>Conditions optimales:</u>

- ✓ Température: 500°C (niveau bas);
- ✓ Temps de refroidissement: 10 min (niveau haut);
- ✓ Type de sable: Type 2.

Application 3: Amélioration de la résistance d'un composite

Énoncé:

Vous développez un nouveau matériau composite. Vous étudiez l'influence de :

A: Pourcentage de fibres (10% vs 20%);

B: Pression de moulage (100 bar vs 150 bar);

Variable réponse: Résistance à la traction (MPa) - à maximiser.

Plan avec points centraux:

Essai	A	B	Résistance
1	-1	-1	50
2	1	-1	70
3	-1	1	60
4	1	1	80
5	0	0	75
6	0	0	76
7	0	0	74

Solution dans Minitab:

Étape 1: Création du plan avec points centraux

Stat > DOE > Factorial > Create Factorial Design;

2 facteurs;

Dans Designs: Number of center points per block = 3;

Dans Facteurs:

A: "Pourcentage_fibres" (10, 20);

B: "Pression" (100, 150).

Étape 2: Analyse avec test de courbure

Résultats clés:

Terme	Effet	Coeff	Coef ErT	Valeur de T	Valeur de p	FIV
Constante		65,000	0,500	130,00	0,000	
A	20,000	10,000	0,500	20,00	0,002	1,00
B	10,000	5,000	0,500	10,00	0,010	1,00
A*B	-0,000	-0,000	0,500	-0,00	1,000	1,00
Pt ctr		10,000	0,764	13,09	0,006	1,00

Effet A: +20 (très significatif)

Effet B: +10 (significatif)

Interaction AB: faible et non significative

Test de courbure : p-value = 0.006 < 0.05 → **Courbure significative !**

Analyse de la variance

Source	DL	SomCar ajust	CM ajust	Valeur F	Valeur de p
Modèle	4	671,429	167,857	167,86	0,006
Linéaires	2	500,000	250,000	250,00	0,004
A	1	400,000	400,000	400,00	0,002
B	1	100,000	100,000	100,00	0,010
Interactions à 2 facteur(s)	1	0,000	0,000	0,00	1,000
A*B	1	0,000	0,000	0,00	1,000
Courbure	1	171,429	171,429	171,43	0,006
Erreur	2	2,000	1,000		
Total	6	673,429			

Étape 3: Interprétation de la courbure

La courbure significative indique que la relation n'est pas linéaire. Le point central (15%, 125 bar) donne de meilleurs résultats que ce que prédit le modèle linéaire.

Conclusion importante

Nous sommes proches d'un optimum ! Un plan de surface de réponse (que nous verrons au prochain chapitre) serait nécessaire pour trouver le véritable maximum.

Recommandation

Utiliser des pourcentages de fibres entre 15% et 20% et des pressions autour de 125-150 bar pour des résistances maximales.

Conclusion

Les plans factoriels 2^k sont des outils extrêmement puissants et efficaces pour une première exploration d'un processus. Ils nous donnent une vue d'ensemble rapide des effets principaux et des interactions clés avec un nombre d'essais raisonnable. Maîtriser ce chapitre est essentiel avant d'aborder des plans plus complexes, parmes les points clés à retenir:

La démarche systématique: Créer le plan → Saisir données → Analyser → Interpréter → Optimiser;

L'importance des interactions: Ne jamais négliger les graphiques d'interaction;

L'utilité des points centraux: Détecter la courbure et éviter de passer à côté d'un optimum;

L'interprétation pratique: Toujours relier les résultats statistiques à la réalité physique du procédé.

Introduction

Au chapitre précédent, nous avons calculé des effets, mais une question essentielle se pose: ces effets sont-ils réels ou simplement dus au hasard des variations expérimentales?.

Objectif du chapitre: Apprendre à distinguer les effets significatifs du bruit expérimental et valider la qualité de notre modèle.

1. Erreurs expérimentales

<u>Définition:</u> L'erreur expérimente (ou bruit) est la variabilité inévitable des mesures due à:

- La précision des instruments;
- Les variations des matières premières;
- L'opérateur;
- Les conditions environnementales.

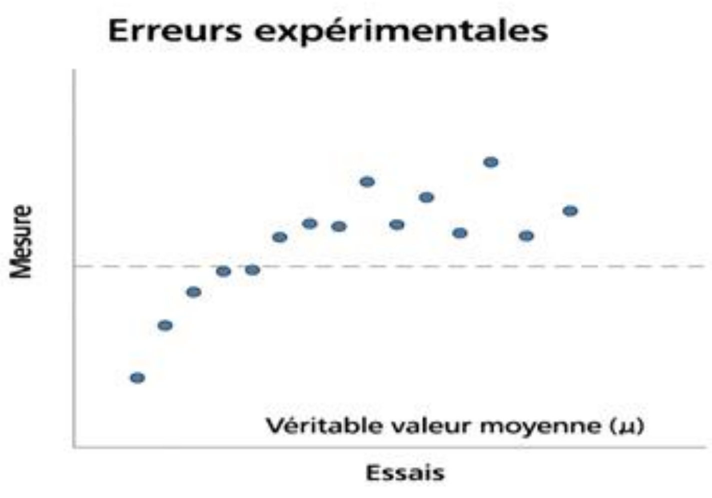


Figure II.1: Erreurs expérimentales.

<u>Types d'erreurs:</u>

- Aléatoires: Imprévisibles, suivent une distribution normale;
- Systématiques: Reproductibles, dues à un biais;
- Conséquence: Nous devons déterminer si un effet calculé est plus grand que l'erreur expérimentale naturelle.

2. Test de signification des effets

2.1. Hypothèses statistiques

- **H_0** (hypothèse nulle): L'effet = 0 (le facteur n'a pas d'influence);
- **H_1** (hypothèse alternative): L'effet ≠ 0 (le facteur a une influence).

2.2. Principe du test

L'influence des facteurs et de leurs interactions est interprétée par les coefficients du modèle postulé. Il faut donc trouver une valeur étalon (t_{crit}) pour la prise de décision si l'effet d'un facteur ou d'une interaction est important ou non (Un effet sera dit significatif s'il est, pour un risque donné, significativement différent de 0). Le test de Student a pour but de fournir une règle de décision. La valeur à tester t_i sera le rapport de la valeur du coefficient a_i sur la valeur de son écart-type S_i (équation 1):

$$\mathbf{t_i = \frac{a_i}{s_i}} \qquad \textbf{(1)}$$

a. Calcul de la variance des coefficients S_i^2

En statistique la formule qui détermine la variance des coefficients S_i^2 en fonction de la variance des écarts S^2 est donnée par l'équation (2).

$$\mathbf{S_i^2 = kS^2} \qquad \textbf{(2)}$$

Ou la constante k dépend du modèle postulé et de la matrice d'expériences. Généralement k très long à calculer (Des logiciels spécialisés possèdent les algorithmes pour faire ce calcul). Mais dans le cas des plans factoriels la relation est simple devient (équation 3):

$$\mathbf{S_i^2 = \frac{1}{n}S^2 \Rightarrow S_i = \sqrt{\frac{S^2}{n}}} \qquad \textbf{(3)}$$

Les calculs statistiques qui permettent de déterminer l'écart type S_i font intervenir la variance des écarts (Les différences entre les valeurs expérimentales y_i et les valeurs estimés (prédites) par le modèle $\hat{y}_i^2$) selon l'équation (4).

$$\mathbf{S^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-p}} \qquad \textbf{(4)}$$

Remarque:

Pour pouvoir conduire les calculs statistiques, il est clair que (n − p) doit être différent de 0. Pour cela, en pratique, il est nécessaire de l’estime l'erreur expérimentale:

- Via les réplicats (répétition des mêmes conditions);
- Via les points centraux;
- Via la suppression des interactions d'ordre élevé (supposées nulles).

Pour notre cas en va négliger un ou plusieurs termes (en général l'interaction (les interactions) d'ordre plus élevé pour que p sont différent de n. Si un plan complet 2^3est considéré, ça donne 8 expériences et 8 termes du modèle (a_0, 3 effets principaux, 3 effets d'interactions d'ordre deux et 1 effet d'interaction d'ordre trois). Pour permettre le test statistique, l'interaction d'ordre trois peut être négligée ce qui donne un modèle réduit avec 7 termes (p = 7).

b. Détermination de la valeur critique $\mathbf{t_{crit}}$

Afin de pouvoir tester la signification d'un effet avec un risque donné α, le test de Student est utilisé: le rapport t_i est comparé à une valeur critique t_{crit} pour un risque α et un degré de liberté ddl = n – p. Cette valeur critique peut être directement lue à partir de tableau de Student (II.1).

$$\mathbf{t_{crit} = v(a, ddl)} \qquad \textbf{(5)}$$

Par exemple la valeur critique de Student t_{crit} pour un modèle dont p = 7, n = 8 et pour un risque α = 0.05 correspond à la cellule de l'intersection de la première ligne (n – p = 1) et la sixième colonne (α = 0.05) dans le tableau de Student (II.1).

Tableau II.1: Lecture d'une valeur t_{crit} depuis la tableau de Student.

dll/α	**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~~	$\boxed{12.71}$	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

c. Réalisation du test et interprétation

L'hypothèse selon laquelle l'effet α_i est nul s'appelle l'hypothèse nulle et est H_0. N'importe quelle autre hypothèse qui diffère de l'hypothèse H_0 s'appelle l'hypothèse alternative et est notée H_1 (α_0 de confiance pour que l'effet soit significativement différent de zéro). Nous aurons:

- H_0: l'effet α_i est nul c-à-d α_i n'est pas significatif;
- H_1: l'effet α_i n'est pas nul c-à-d α_i est significatif.

Après avoir calculé les rapports t_i ces valeurs sont comparées avec la valeur t_{crit} pour déterminer la signification des effets ou leur non signification.

- Si ($|t_i| > t_{crit}$), l'hypothèse H_0 est rejetée (H_1 est acceptée) c-à-d l'effet α_i est significatif;
- Si non ($|t_i| > t_{crit}$), l'hypothèse H_0 est acceptée (H_1 est rejetée) c-à-d l'effet α_i n'est pas significatif.

A partir de t de Student, l'évaluation de la probabilité peut révéler que le coefficient est peu significatif (ou nul) ou non. Cette probabilité est la P_{valeur}.

- Si la P_{valeur} est proche de 0 (probabilité nulle), le coefficient est influent car il n'est pas nul;
- Si la P_{valeur} est proche de 1. Le coefficient est proche de 0 et donc négligeable;
- Si la P_{valeur} possède une valeur intermédiaire, le coefficient est peut être légèrement significatif ou non significatif.

3. Intervalle de confiance des effets du modèle

Définition:

Intervalle dans lequel on estime, avec un certain niveau de confiance (généralement 95%), que se trouve la vraie valeur de l'effet.

Formule:

$$\mathbf{IC(Effet) = Effet} \pm \mathbf{t}_{(\frac{\mathbf{a}}{\mathbf{2}},\mathbf{dd})} \times \mathbf{S}_{\mathbf{effet}}$$

où:

- $\mathbf{t}_{(\frac{\mathbf{a}}{\mathbf{2}},\mathbf{dd})}$ = valeur de la loi de Student;
- dd = degrés de liberté de l'estimation de l'erreur.

Interprétation:

- Effet A: 19.0 ± 3.2 → IC à 95% = [15.8; 22.2].
- Effet B: 1.0 ± 3.2 → IC à 95% = [-2.2; 4.2].

Conclusion: L'effet A est significatif (0 n'est pas dans l'intervalle), l'effet B ne l'est pas (0 est dans l'intervalle).

3.1. Exemple applicatif

En prenant l'exemple précèdent de la cuisson d'un gâteau ou nous considérons l'ajout ou le non ajout de la levure comme un facteur supplémentaire. Nous voulons déterminer quels sont les facteurs et les interactions qui ont un effet significatif et ceux qui ont un effet négligeable. L'expérimentateur décidé d'utiliser un plan d'expériences factoriel complet 2^3. Il définit le domaine d'étude des trois paramètres comme suit (Tableau II.2).

Tableau II.2: Plan factoriel fractionnaire 2^{3-1}.

	T	D	L
Niveau bas : -1	a_T **= 23**	a_D **= 38**	**Sans**
Niveau haut : +1	a_T^+ **= 39**	a_D^+ **= 24**	**Avec**

Les résultats mesurés de l'épaisseur (E_i) pour chaque expérience réalisée ainsi que les coefficients calculés sont donnés dans le tableau (II.3).

Tableau II.3: Matrice d'expériences avec effets de l'exemple étudié.

Exp(i)	Moy	T	D	L	TD	TL	DL	~~TDL~~	E_i^{exp}
1	1	-1	-1	-1	1	1	1	~~-1~~	66.82
2	1	1	-1	-1	-1	-1	1	~~1~~	45.22
3	1	-1	1	-1	-1	1	-1	~~1~~	69.22
4	1	1	1	-1	1	-1	-1	~~-1~~	38.48
5	1	-1	-1	1	1	-1	-1	~~1~~	66.6
6	1	1	-1	1	-1	1	-1	~~-1~~	74.82
7	1	-1	1	1	-1	-1	1	~~-1~~	74.2
8	1	1	1	1	1	1	1	~~1~~	74.28
	a_0=63.705	a_T=5.505	a_d=0.34	a_L=8.77	a_{TD}=2.16	a_{TL}=7.58	a_{DL}=1.425	~~a_{TDL}~~=0.125	

L'équation (6) montre le modèle mathématique complet:

$$\hat{E} = 63.705 - 5.505T + 0.34D + 8.77L - 2.16TD + 7.58TL + 1.425DL + 0.125TDL \qquad (6)$$

- Pour répondre à la question déterminant la signification des différents effets, nous utilisons le test de Student.

Afin de pouvoir conduire le test statistique de Student, nous choisissons de supprimer dernier terme du modèle complet (l'interaction d'ordre le plus élevé) pour éviter le cas ou le dénominateur n – p = 0. Le nouveau modèle à considérer dans l'équation (7):

$$\hat{E} = 63.705 - 5.505T + 0.34D + 8.77L - 2.16TD + 7.58TL + 1.425DL \qquad (7)$$

• Pour calculer S^2 nous calculons les réponses estimées $\hat{E}_i$ par le modèle (7) et les résidus dans le tableau (II.4).

Tableau II.4: Résultats des réponses estimées et des écarts au carré.

Exp (i)	**E_i**	$\hat{E}_i$	$(\mathbf{E_i} - \mathbf{\hat{E}_i})^2$
1	66.82	66.945	**0.015625**
2	45.22	45.095	**0.015625**
3	69.22	69.095	**0.015625**
4	38.48	38.605	**0.015625**
5	66.6	66.475	**0.015625**
6	74.82	74.945	**0.015625**
7	74.2	74.325	**0.015625**
8	74.28	74.155	**0.015625**
		$\sum(\mathbf{E_i} - \mathbf{\hat{E}_i})^2$	**0.125**

Nous aurons:

$$\mathbf{S^2} = \frac{\sum_{i=1}^{8}(\mathbf{y_i} - \mathbf{\hat{y}_i})^2}{\mathbf{n} - \mathbf{p}}$$

$$= \frac{0.125}{8-7} = \mathbf{0.125}$$

Les valeurs des S_i seront:

$$\mathbf{S_0} = \mathbf{S_T} = \mathbf{S_D} = \mathbf{S_L} = \mathbf{S_{TD}} = \mathbf{S_{TL}} = \mathbf{S_{DL}} = \sqrt{\frac{S^2}{n}} = \sqrt{\frac{0.125}{8}} = \mathbf{0.125}$$

- La valeur de t_{crit} pour risque α = 0.05 et un degré de liberté ddl = n –p = 8 -7 est prélevée de la tableau de Student (équation 2.1) $t_{crit} = \upsilon(0.05,1) = 12.71$
- Finalement nous calculons les valeurs $t_i = a_i/s_i$ et nous comparons leurs valeurs absolues à t_{crit} dans le tableau (II.5) pour établir une décision concernant la signification des effets des facteurs et de leurs interactions.

Tableau II.5: Résultats obtenu des t_i et a_i et leurs signification.

Effet	a_i	s_i	t_i	$\lvert t_i \rvert >= t_{crit}$?	a_i est Sinificatif?
a_0	63.705	0.125	509.64	Oui	**Sinificatif**
a_T	-5.505	0.125	44.04	Oui	**Sinificatif**
a_D	0.34	0.125	2.72	Non	Non Sinificatif
a_L	8.77	0.125	70.16	Oui	**Sinificatif**
a_{TD}	-2.16	0.125	17.28	Oui	**Sinificatif**
a_{TL}	7.58	0.125	60.64	Oui	**Sinificatif**
a_{DL}	1.425	0.125	11.4	Non	Non Sinificatif

- D'après Le tableau (II.3), la duré de cuisson D et l'interaction DL ont un effet négligeable (considérés comme = 0) c-à-d qu'il y a moins de 5% de risque que les effets ne soient pas nuls. Quant aux effets de la température, de la levure et les interactions TD et TL sont significatifs c.-à-d. qu'ils ont plus de 5% de risque que ces effets ne soient pas nuls (considérés comme ≠ 0).

3.2. Régression linéaire et analyse du modèle

3.2.1. Régression linéaire

Supposons que nous un modèle polynomial sous la forme générique suivante (8):

$$\boldsymbol{y = a_0 + \sum_{i=1}^{k} a_i x_i} \qquad \textbf{(8)}$$

Ce polynôme peut être utilisé pour représenter plusieurs formes de modèles (linéaires, quadratiques, etc). Par exemple, pour écrire le modèle: $y = 3 + 2.1z_1 + 4z_2 + 2z_1z_2 + z_2^2$ sous la forme (8) nous posons: $x_1 = z_1$, $x_2 = z_2$, $x_3 = z_1z_2$ et $x_4 = z_2^2$. Notez que, bien qu'un modèle ne soit pas linéaire, le modèle de régression est toujours linéaire car il est linéaire dans les paramètres.

A la fin du plan d'expériences, un système de n équations (s'il y a n essais) et à p inconnues (s'il y a p coefficients dans le modèle choisi) peut être écrit d'une manière simple en notation matricielle:

$$\mathbf{y - X_a + \epsilon} \qquad \textbf{(9)}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdot\cdot & x_{1p} \\ 1 & x_{21} & x_{22}. & \cdot\cdot & x_{2p} \\ . & . & . & \cdot\cdot & . \\ . & . & . & \cdot\cdot & . \\ . & . & . & \cdot\cdot & . \\ 1 & x_{n1} & x_{n2} & \cdot\cdot & x_{np} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ . \\ . \\ . \\ a_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ . \\ . \\ . \\ \epsilon_n \end{bmatrix} \quad \textbf{(10)}$$

- $\boldsymbol{y}$: est le vecteur des réponses;
- $\boldsymbol{x}$: est la matrice de calcul des coefficients ou matrice du modèle qui dépend des points expérimentaux choisis pour exécuter le plan et du modèle postulé;
- $\boldsymbol{a}$: est le vecteur des coefficients;
- $\boldsymbol{\epsilon}$: est le vecteur des écarts.

Ce système ne peut pas être résolu simplement car en général le nombre d'équations est inférieur au nombre d'inconnues (n équations et pn inconnues). Cette résolution peut être menée en utilisant une méthode de régression. Souvent cette méthode est basée sur la méthode d'optimisation des moindres carrés. Les estimations des coefficients obtenues sont les éléments du vecteur $\hat{a}$.

$$\hat{\mathbf{a}} = (\mathbf{X}\acute{\mathbf{X}})^{-1}\acute{\mathbf{X}}\mathbf{y} \quad \textbf{(11)}$$

Ou $\grave{X}$ est la transposée de la matrice X.

Pour calculer directement les valeurs des coefficients, il existe plusieurs logiciels qui exécutent ce calcul.

3.2.2. Test de validation de modèle

Dans les cas ou il y a plus de deux échantillons de test, une *ANOVA* est utilisée pour déterminer s'il existe des différences statistiquement significatives entre les moyennes des échantillons (Dans les cas de deux échantillons, le test t suffit à vérifier s'il existe des différences statistiquement significatives entre les moyennes des échantillons). La procédure de test implique une analyse de variance (*ANOVA*) et la réalisation du test F ($Fisher - Snedecor$) qui teste la signification de la régression dans sa globalité (il teste la nullité de tous les coefficients en même temps) Il ne permet donc pas de préjuger la signification particulière des coefficients pris isolément. C'est ce que fait le test de Student qui teste un à un la signification des coefficients. Il s'agit de tester les hypothèses:

H_0: $a_1 = a_2 = \ldots = a_p = 0$.

H_1: il existe au moins un $a_i \neq 0$.

La variation de la variable à expliquer (ou totale) se décompose en somme de la variance expliquée par le modèle et la variance résiduelle.

$$\mathbf{STCE = SCEL + SCER} \quad \textbf{(12)}$$

- **Somme totale des carrés des écarts ($STCE$):** est la somme totale des carrés corrigés de y.

$$\mathbf{STCE} = \sum_{i=1}^{n}(y_i - \bar{y})^2 \quad \textbf{(13)}$$

- **Somme des carrés des écarts dues à la liaison ($SCEL$):** est la somme des carrés expliquée par le modèle:

$$\mathbf{SCEL} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \quad \textbf{(14)}$$

- **Somme des carrés des écarts des résidus ($SCER$):** est la somme des carrés des résidus.

$$\mathbf{SCER} = \sum_{i=1} \hat{e}_i^2 = \sum_{i=1}(y_i - \hat{y}_i)^2 \quad \textbf{(15)}$$

Notons que: y_i sont les réponses observées lors de la réalisation des expériences, $\bar{y}$ est la moyenne des réponses et $\hat{y}_i$ sont les réponses estimées à l'aide du modèle. De plus des "carrés moyens" sont définis qui sont le quotient d'une somme de carrés par son degré de liberté (CML, CMR, CMT).

- *SCEL* aura *p*-1 degrés de liberté (*p* est le nombre de coefficients estimes du modèle);
- *SCER* aura *n* -*p* degrés de liberté (ou *n* est le nombre d'expériences réalisées);
- *STCE* aura *n* -1 degrés de liberté.

Le tableau dit analyse de la variance se présente sous la forme suivante:

Tableau II.6: Tableau d'analyse de la variance

Variation du $\dot{a}$	Somme des carrés	ddl	Carré moyen	F
Liaison	***SCEL***	***p-1***	$\frac{SCEL}{p-1}$ = ***CML***	$Fcal = \frac{CML}{CMR}$
Résidus	***SCER***	***n-p***	$\frac{SCER}{n-p}$ = ***CMR***	
Totale	***STCE***	***n-1***	$\frac{STCE}{n-1}$ = ***CMT***	

Le test F permet alors de comparer le $\mathrm{F_{cal}}$ que nous avons calculé dans le tableau précédent avec le $\mathrm{F_{crit}(ddl_1.ddl_2.a)}$ lu à partir de la tableau de *Fisher-Snedecor* (Tableau II.6) avec les degrés de liberté $\mathrm{ddl_1 = p - 1, ddl_2 = n - p}$ et avec un risque fixe à l'avance a. Le critère est rejeter $\mathrm{H_0}$ au seuil a si $\mathrm{F_{cal}} > \mathrm{F_{crit}}$.

- Accepter $\mathrm{H_0}$ implique que l'on conclut qu'il n'y a pas de relation globale entre $\mathrm{x_i, y}$. Ceci peut signifier que:

- Le modèle utilisé n'est pas adéquat.
- La variation des x_i influent peu ou pas sur la variation de y.

- Au contraire, rejeter $\mathrm{H_0}$ implique que nous concluons que la variation des $\mathrm{x_i}$ influe la variation de y.

Autrement dit:

- Si $F_{cal} > F_{crit}$ le modèle de la régression linéaire est considère comme valide;
- Si non, le modèle est considéré comme non valide.

3.2.3. Quelques mesures statistiques d'évaluation de modèle

En effectuant une analyse de la variance du modèle de régression, plusieurs mesures statistiques peuvent être obtenues (R^2, R^2_{ajuste}, R^2_{predit} $et\ la\ Precision - adequate$). Ces mesures permettent de tester la qualité de modèle.

a. Coefficient de détermination R^2

Le coefficient de détermination R^2 montre dans quelle mesure un modèle de fonction correspond aux données. Plus R^2 est proche de 1, meilleur est la correspondance. R^2 est donc une mesure de la qualité du modèle, il est toujours entre 0 et 1. S'il est égal à 1, le modèle permet de retrouver les va leurs des réponses mesurées.

$$\mathbf{R^2 = 1 - \frac{SCER}{STCE} = \frac{SCEL}{STCE}} \qquad \textbf{(16)}$$

Attention, le R^2 ne révèle pas tous sur la qualité du modèle. R^2 doit être considéré comme une donnée descriptive, intéressante en soi, et pratique pour comparer des modèles sur les mêmes données, mais il ne peut être considéré comme une absolue. Dans la pratique, il est difficile d'indiquer la valeur d'un bon R^2 car les valeurs varient beaucoup d'une discipline à l'autre et du processus étudié.

b. R au carré ajuste $\mathbf{R^2_{ajuste}}$

Le $\mathrm{R^2_{ajuste}}$ est une version modifiée de $\mathrm{R^2}$ qui a été ajustée en fonction du nombre de prédicteurs du modèle. Le $\mathrm{R^2_{ajuste}}$ peut être négatif, mais ce n'est généralement pas le cas. Il est toujours inférieur à $\mathrm{R^2}$.

$$\mathbf{R^2_{ajuste} = 1 - \frac{CMR}{CMT} = 1 - \frac{(n-1)(1 - R^2)}{n-p}} \qquad \textbf{(17)}$$

c. Somme des carrés des erreurs résiduelles prédites *PRESS*

PRESS (Predicted Residual Error Sum of Squares) est une mesure de l'adaptation de modèle à chaque point du plan. L *PRESS* est calculé en prédisant d'abord ou chaque point doit se trouver à partir d'un modèle contenant tous les autres points, à l'exception du point en question. Les résidus au carré (différence entre les valeurs réelles et prédites) sont ensuite additionnés. *PRESS* est utile pour déterminer la valeur de R^2_{predit}.

$$\textbf{PRESS} = \sum_{i=1}^{n}(y_i - \hat{y}_{-i})^2 \qquad \textbf{(18)}$$

d. R au carré prédit R^2_{predit}

Le R^2_{predit} indique dans quelle mesure un modèle de régression prédit les réponses à de nouvelles observations.

$$\mathbf{R^2_{predit} = 1 - \frac{PRESS}{STCE}} \qquad \textbf{(19)}$$

Le R^2_{predit} est calculé en suppriment systématiquement chaque observation de l'ensemble de données, en estimant l'équation de régression et en déterminant dans quelle mesure le modèle prédit l'observation supprimée. Comme R^2_{ajuste}, le R^2_{predit} peut être négatif et il est toujours inférieur à R^2.

Même si vous ne prévoyez pas d'utiliser le modèle pour les prévisions, le R^2_{predit} fournit néanmoins des informations cruciales.

e. Précision adéquate Pre_{adeq}

C'est un rapport signal sur bruit. Il compare la plage des valeurs prédites aux points du plan à l'erreur de prédiction moyenne. Des rations supérieurs à 4 indiquent une discrimination de modèle adéquate.

$$\mathbf{Pre_{adeq} = \frac{max(\hat{y}) - min(\hat{y})}{\sqrt{\frac{p * CMR}{n}}}} \qquad \textbf{(20)}$$

Tableau II.7: Tableau de Fisher-Snedecor (Risque α = 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

3.2.4. Exemple applicatif

En prenant l'exemple précédent de la cuisson de gâteau ou nous avons obtenu le modèle mathématique suivant:

$$\hat{E} = 63.705 - 5.505T + 0.34D + 8.77L - 2.16TD + 7.58TL + 1.425DL \quad (21)$$

Nous voulons étudier la validité du modèle obtenu avec un risque (α = 0.05) ainsi que sa qualité (sa correspondance aux données mesurées, sa capacité de prédiction et sa précision adéquate) en effectuant une analyse de la variance (*ANOVA*) et en calculant les différentes mesures statistiques.

Tableau II.8: Résultats des réponses estimées et des écarts et des écarts au carré.

Exp(i)	E_i	$\hat{E}_i$	$(E_i - \hat{E}_i)^2$	$(E_i - \bar{E})^2$
1	66.82	66.945	0.015625	10.4976
2	45.22	45.095	0.015625	346.3321
3	69.22	69.095	0.015625	29.0521
4	38.48	38.605	0.015625	630.01
5	66.6	66.475	0.015625	7.6729
6	74.82	74.945	0.015625	126.3376
7	74.2	74.325	0.015625	112.7844
8	74.28	74.155	0.015625	109.2025
/	**63.705**	/	**0.125**	**1371.8892**
/	$\bar{E}$	/	$\sum(E_i - \hat{E}_i)^2$	$\sum(E_i - \bar{E})^2$

SCEL = $\sum_{i=1}^{n}(\hat{E}_i - \bar{E})^2$ **= 1371.8892**

$$\text{CML} = \frac{SCEL}{p-1} = \frac{1371.8892}{7-1} = 228.6482$$

SCER = $\sum_{i=1}^{n}(E_i - \hat{E}_i)^2$ **= 0.125**

$$\text{CMR} = \frac{SCER}{n-p} = \frac{0.125}{8-7} = 0.125$$

STCE = SCEL + SCER = 1372.0142

$$\text{CMT} = \frac{STCE}{n-1} = \frac{1372.0142}{8-1} = 196.0020$$

Tableau II.9: Tableau d'analyse de la variance de l'exemple étudié.

Variation	Somme des carrés	ddl	Carré moyen	F
Liaison	**1371.8892**	**7-1**	**228.6482**	
Résidus	**0.125**	**8.7**	**0.125**	**Fcal =** $\frac{CML}{CMR}$ **= 1829.1856**
Totale	**1372.0142**	**8-1**	**196.0020**	

Pour tester la validité du modèle obtenu, F_{cal} est comparé à F_{crit}. F_{crit} est obtenu de la tableau de Fisher pour les paramètres $\alpha = 0.05$, $dll_1 = p$ et $dll_2 = n-p$ ($F_{crit}(0.05,6,1) = 234$. $F_{cal} > F_{crit}$ → le modèle de régression est valide.

$R^2 = \frac{2791.75}{2794.875} = \frac{SCEL}{STCE} = 0.99990889$ et $R^2 = 1 - \frac{CMR}{CMT} = 1 - \frac{0.125}{196.002028} = 0.999362$.

Nous avons les valeurs de R^2 et R^2_{ajuste} sont très proches de 1 ce qui prouve que la qualité du modèle est très bonne en ce qui concerne sa données observées.

$Pre_{adeq} = \frac{\max(\hat{y}) - \min(\hat{y})}{\sqrt{\frac{p*CMR}{n}}} = 109.882$. $Pre_{adeq} > 4$ → Le modèle a une précision adéquate.

Le calcul de la valeur PRESS est long; nous donnons directement sa valeur PRESS = 8 (En utilisant a logiciel nous pouvons directement obtenir toutes les statistiques).

$R^2_{predit} = 1 - \frac{PRESS}{STCE} = 1 - \frac{8}{1372.0142} = 0.9942$. La valeur de R^2_{predit} est très proche de 1 → le modèle a un grande capacité de prédire de nouvelles observations.

Dans l'étude du modèle ci-dessus nous avons négligé que le terme TDL. Nous avons trouvé que les effets de D et TDL sont négligeables. Si nous négligeons ces deux termes et nous étudions la qualité du nouveau modèle simplifié (équation 2.7), nous aurons des résultats qui reflètent la bonne qualité de ce modèle simplifié.

Application 01: Test de signification sur un plan 2^2 avec réplicats

Énoncé:

Reprenons l'exemple du procédé de nettoyage avec des **réplicats** pour estimer l'erreur expérimentale.

Données:

Essai	Temps	Concentration	Efficacité (%)
1	10 min	2%	60, 62, 61
2	20 min	2%	72, 70, 71
3	10 min	5%	54, 56, 55
4	20 min	5%	80, 78, 79

Objectif: Déterminer quels effets sont statistiquement significatifs.

Étapes Minitab:

Étape 1: Création du plan avec réplicats

1. Stat > DOE > Factorial > Create Factorial Design;
2. 2 facteurs, 2 niveaux;
3. Dans Designs: Number of replicates = 3;
4. Facteurs:

   - Temps (10, 20);
   - Concentration (2, 5).

Étape 2: Saisie des données

Entrer les 12 valeurs d'efficacité dans l'ordre.

Étape 3: Analyse statistique

1. Stat > DOE > Factorial > Analyze Factorial Design;
2. Response: Efficacité;
3. Terms: Inclure tous les termes (A, B, AB);

4. Graph: Sélectionner:

- Pareto standardisé;
- Graphique des effets normaux;
- Quatre-en-un des résidus.

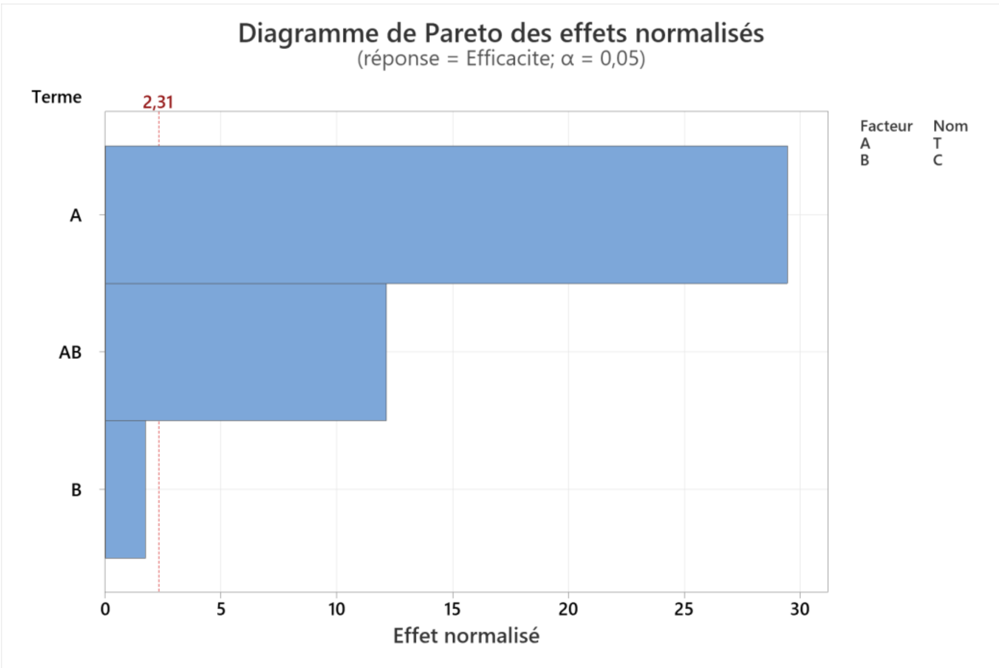
Étapes 4: Résultats et interprétation

Tableau ANOVA attendu:

Source	Ddl	SC	MC	F	P
Modèle	3	788.67	262.89	65.72	0.000
Linéaire	2	780.50	390.25	97.56	0.000
Temps	1	722.00	722.00	480.50	0.000
Concentration	1	58.50	58.50	14.63	0.005
Interaction	1	7.50	7.50	1.88	0.208
Erreur	8	32.00	4.00		
Total	11	820.67			

S = 2,000 R^2 = 96,10% R^2(adj) = 94,64%

Graphique de Pareto standardisé:



<u>Interprétation:</u>

- **Temps**: p = 0,000 < 0,05 → **très significatif;**
- **Concentration**: p = 0,005 < 0,05 → **significatif;**
- **Interaction**: p = 0,208 > 0,05 → **non significative;**
- **R^2** = 96,1% → modèle explique très bien les données.

Application 02: Validation avec points centraux et test de manque d'ajustement

Énoncé:

On étudie un procédé de fabrication avec 2 facteurs. Des points centraux ont été ajoutés pour détecter la courbure.

Facteurs:

- A: Température (100°C, 150°C);
- B: Pression (1 bar, 2 bar).

Données:

Essai	A	B	Rendement (%)
1	-1	-1	45
2	+1	-1	65
3	-1	+1	55
4	+1	+1	75
5, 6, 7	0	0	70, 71, 69

Étapes Minitab:

Étape 1: Création du plan avec points centraux

1. Create Factorial Design (2 factors);
2. Dans Designs : Number of center points = 3;
3. Saisir les données.

Étape 2: Analyse avec test de courbure

1. Analyse Factoriel Design;
2. Vérifier que "Include center points in model" est coché.

Étape 3: Résultats et interprétation

Sortie Minitab:

text

Analyse de variance pour Rendement (données codées):

Source	Ddl	SC	MC	F	P
Modle	3	1000.0	333.3	66.67	0.000
Linéaire	2	1000.0	500.0	100.00	0.000
A	1	800.0	800.0	160.00	0.000
B	1	200.0	200.0	40.00	0.001
Courbure	1	90.0	90.0	18.00	0.013
Erreur	4	20.0	5.0		
Total	7	1110.0			

S = 2,236 R^2 = 90,09% R^2(adj) = 85,59%

Test de manque d'ajustement détaillé:

Décomposition de l'erreur:

Source	Ddl	SC	MC	F	P
Erreur	4	20.0	5.0		
Erreur pure	2	2.0	1.0		
Manque d'ajust	2	18.0	9.0	9.00	0.100

Interprétation:

- **Courbure significative** (p = 0,013) → relation non linéaire;
- **Manque d'ajustement non significatif** (p = 0,100) → modèle acceptable;
- **Recommandation** : Un plan de surface de réponse serait plus approprié.

Application 03: Analyse complète avec décision stratégique

Énoncé:

Une entreprise pharmaceutique veut optimiser la synthèse d'un principe actif avec 3 facteurs, en considérant à la fois le rendement et le coût.

Facteurs:

- A: Température (30°C, 50°C) - coût énergétique différent;
- B: Temps de réaction (1h, 3h) - coût temporel;
- C: Catalyseur (Type A, Type B) - coût matériel.

Variables réponses:

- Y_1: Rendement (%) → à maximiser;
- Y_2: Coût total (€) → à minimiser.

Plan 2^3 avec 2 réplicats:

Essai	A	B	C	Rendement	Coût
1	-1	-1	-1	60, 62	100, 102
2	+1	-1	-1	75, 73	150, 148
3	-1	+1	-1	65, 67	180, 182
4	+1	+1	-1	80, 78	230, 228
5	-1	-1	+1	70, 72	120, 118
6	+1	-1	+1	85, 83	170, 172
7	-1	+1	+1	75, 77	200, 202
8	+1	+1	+1	90, 88	250, 248

Étapes Minitab:

Étape 1: Création et analyse multi-réponses

1. Créer le plan 2^3 avec 2 réplicats;
2. Analyze Factorial Design pour chaque réponse;
3. Pour le Rendement : Maximise;

4. Pour le Coût : Minimise.

<u>Étape 2: Analyse des résidus</u>

Pour chaque modèle, vérifier:

- Normalité des résidus (graphique normal);
- Homoscédasticité (résidus vs valeurs ajustées);
- Indépendance (résidus vs ordre).

<u>Étape 3: Optimisation multi-objectifs</u>

1. Stat > DOE > Factoriel > Réponse Optimiser
2. Paramétrer les objectifs:
   - Rendement: Cible = 90, Limite inf = 60;
   - Coût: Cible = 100, Limite sup = 250;
3. Définir l'importance relative (poids).

<u>Étape 4: Résultats et compromis</u>

<u>Modèle Rendement:</u>

Coefficients (données codées) :

Constante : 74,75

A: 8,75*** B: 3,75** C: 6,25***

AB : 1,25 AC : 0,75 BC: -0,25 ABC: 0,25

R^2 = 95,8% $R^2(adj)$ = 92,3%

<u>Modèle Coût:</u>

Coefficients (données codées) :

Constante : 177,5

A: 37,5*** B: 52,5*** C: 12,5**

AB: 2,5 AC : 1,5 BC: -0,5 ABC: 0,5

R^2 = 99,1% R^2(adj) = 98,2%

Solution d'optimisation:

Paramètres optimaux :

A (Température) : +1 (50°C)

B (Temps) : -1 (1h)

C (Catalyseur) : +1 (Type B)

Prédictions :

Rendement : 85,0% ± 2,1%

Coût : 170€ ± 5€

Désirabilité globale : 0,82

Compromis analysé:

- Solution idéale (rendement max, coût min) : non réalisable;
- Solution optimale : bon compromis qualité/prix;
- Alternative : A=+1, B=0, C=+1 → rendement 83%, coût 155€.

Exercice supplémentaire de validation

Exercice 1: Interprétation de R^2 et R^2 ajusté

Scénario: Comparaison de deux modèles pour les mêmes données :

Modèle 1 (termes principaux seulement) :

$R^2 = 85{,}2\%$

- R^2 ajusté = 83,8%

Modèle 2 (termes principaux + interactions) :

- $R^2 = 87{,}5\%$
- R^2 ajusté = 82,1%

Question: Quel modèle choisiriez-vous et pourquoi ?

Réponse: Modèle 1, car le R^2 ajusté est plus élevé, indiquant que les interactions n'apportent pas d'amélioration significative.

Grille d'évaluation de la validation

Check-list à compléter pour chaque analyse:

Critère de validation	TD1	TD2	TD3	Commentaires
Effets principaux significatifs ($p<0{,}05$)	✓	✓	✓	
Interactions significatives identifiées	✗	✗	✓	TD3 : interactions faibles
Modèle global significatif ($p<0{,}05$)	✓	✓	✓	
$R^2 > 0{,}80$	✓	✓	✓	
R^2 ajusté $> 0{,}75$	✓	✓	✓	
Résidus normaux	✓	✓	✓	Vérifié graphiquement
Résidus indépendants	✓	✓	✓	
Homoscédasticité	✓	✓	✓	
Pas de manque d'ajustement	✓	(courbure)	✓	TD2 : courbure détectée
Validation expérimentale	-	-	✓	Prédictions cohérentes

Conclusion:

1. **L'importance des réplicats**: Permettent d'estimer l'erreur expérimentale et de faire des tests de signification fiables
2. **L'utilité des points centraux**: Détectent les non-linéarités et aident à choisir le bon type de modèle
3. **La nécessité des diagnostics**: Un modèle statistiquement significatif n'est pas forcément valide si les résidus ne respectent pas les hypothèses
4. **L'optimisation multi-réponses**: Nécessite souvent des compromis entre objectifs contradictoires

Introduction

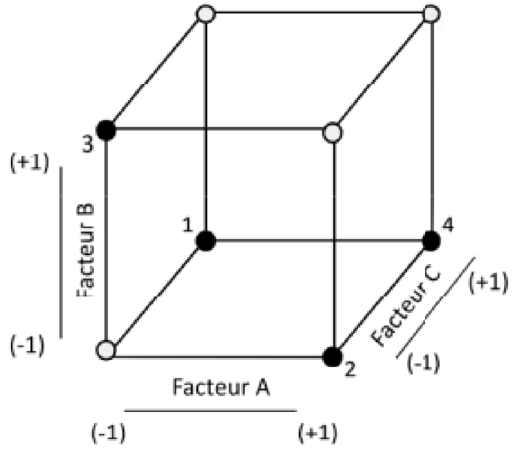
la visualisation d'une expérience comportant plus de trois facteurs est plus complexe et difficile à décrire, car les dimensions requis pour la visualisation sont égales au nombre de facteurs, mais le principe des expériences en angle est le même, quel que soit le nombre de facteurs. Pour les études de dépistage, il est assez courant d'utiliser des plans factoriels fractionnaires.

1. Plans factoriels fractionnaires

Comme son nom l'indique, le plan factoriel fractionnaire est une fraction d'un plan factoriel complet. Un plan factoriel fractionnaire est construit de manière à pouvoir encore identifier les principaux effets sans acquérir les informations détaillées fournies par un plan factoriel complet. Les expériences sont sélectionnées en utilisant une sélection symétrique des angles, des diagonales et des diagonales opposées.

La demi-fraction d'un plan à trois facteurs nécessite quatre essais. Le plan comprend deux ensembles de quatre essais chacun, exactement équivalents sur le plan mathématique, ou chaque ensemble peut être choisi.

Un plan factoriel complet est facilement complété à l'aide de l'ensemble des autres paramètres des facteurs. Le plan est équilibré, ce qui signifie que chaque facteur est exécuté avec le même nombre de fois pour chaque niveau.



Exp	A	B	C
1	-1	-1	+1
2	+1	-1	-1
3	-1	+1	-1
4	+1	-1	+1

Figure III.1: Plan factoriel 2^{3-1}.

En général, un plan factoriel fractionnaire peut être décrit par $N = 2^{k-p}$ ou :

- N est le nombre d'expériences.
- K est le nombre de facteurs à étudier.
- P est la taille de la fraction (1 →1/2,2 →1/4,3 →1/8, etc).

Par conséquent, $N = 2^{4-1}$ signifie que quatre facteurs seront examinés en $2^3 = 8$ exécutions.

La force d'un plan factoriel fractionnaire réside dans le fait qu'il permet de sélectionner de nombreux facteurs en utilisant relativement peu d'expériences. Cependant, l'inconvénient de ce type de plan est que les effets sont confondus. En fonction de la résolution (Res), le plan factoriel fractionnaire prend en charge les effets des facteurs principaux et d'interaction.

Les plans fractionnaires sont généralement utilisés en tant que plans de criblage pour déterminer quels sont les facteurs les plus influents sans forcément étudier les interactions d'ordre 2. C'est souvent le cas si le nombre de facteurs est très élevé.

2. Notion d'alias et de contraste

Un plan factoriel fractionnaire est généré à partir d'un plan factoriel complet en choisissant une structure d'alias. La structure d'alias détermine quels effets sont confondus. Par exemple, un plan factoriel fractionnaire 2^{5-2} avec les 5 facteurs peut être généré en utilisant un plan factoriel complet 2^3 impliquant trois facteurs (disons A, B et C), puis en choisissant de confondre les deux facteurs restants D et E avec les interactions générées par D = A*B et E = A*C. Ces deux expressions sont appelées les générateurs d'alias.

Ainsi, par exemple, lorsque le plan est exécuté et que l'expérimentateur estime les effets pour le facteur D, ce qui est réellement estimé est une combinaison de l'effet principal de D et de l'interaction à deux facteurs impliquant A et B.

Une caractéristique importante d'un plan fractionnaire est la relation de définition, qui donne à l'ensemble des colonnes d'interaction une valeur égale dans la matrice du plan à une colonne de signes plus, notée I.

Pour l'exemple ci-dessous, puisque D = AB et E = AC, alors ABD et ACE sont deux colonnes de signes plus, et par conséquent BCDE aussi. Dans ce cas, la relation de définition du plan factoriel fractionnaire est: I = ABD = ACE = BCDE

La relation de définition permet de déterminer le motif d'alias du plan.

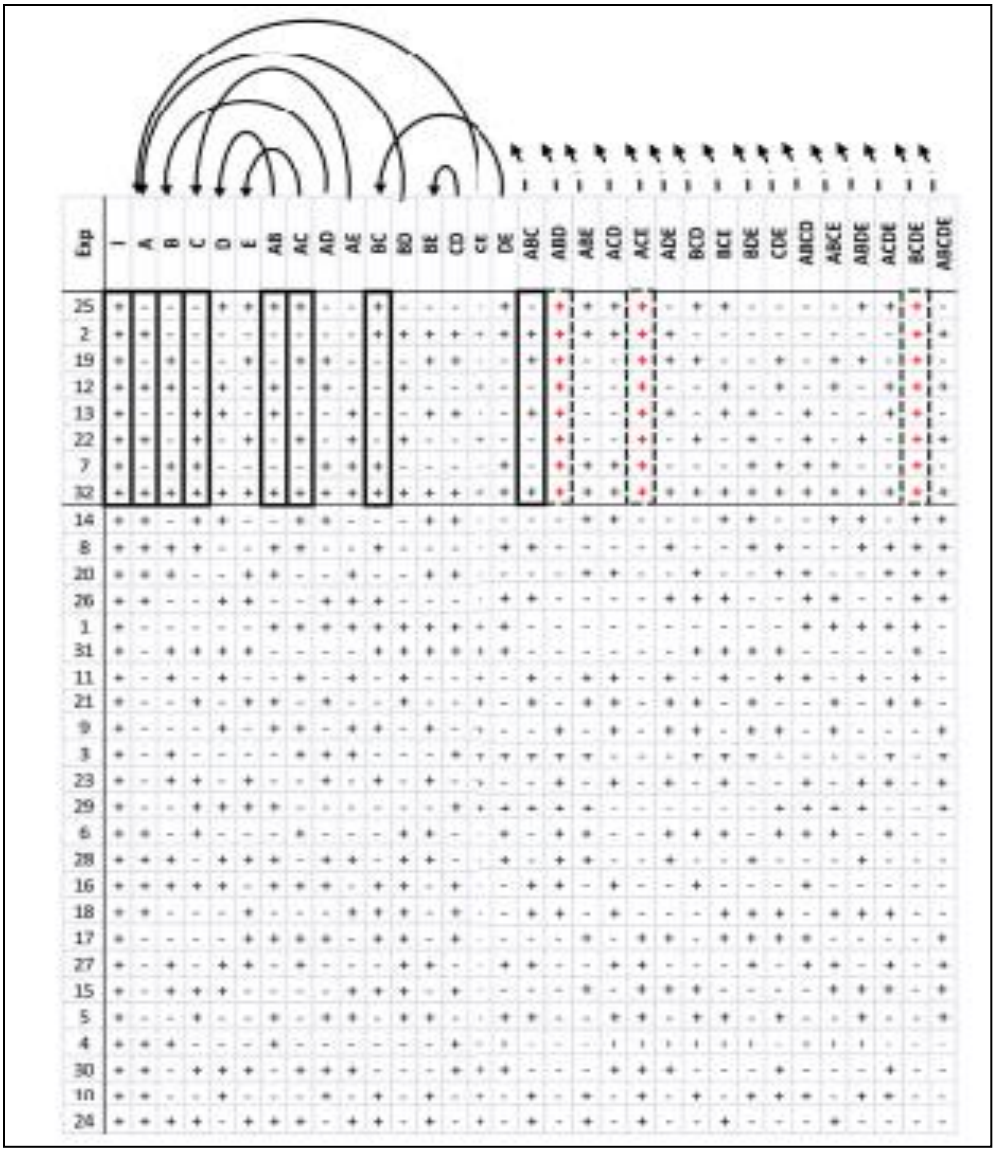


Figure III.2: Plan factoriel fractionnaire 2^{5-2} à partir un plan factoriel complet 2^3.

2.1. Résolution:

Une propriété importante d'un plan fractionnaire est sa résolution ou sa capacité à séparer les effets principaux et les interactions d'ordre faible les uns des autres.

Tableau III.1: Présente les résolutions des plans fractionnaires.

Résolution	Aptitude	Exemple
I	Intitule: une expérience comportant exactement une analyse ne teste qu'un niveau de facteur et ne permet donc même pas de distinguer les niveaux haut et bas de ce facteur.	2^{4-1} avec relation de définition I = A
II	Intitule: les effets principaux sont confondus avec d'autres effets principaux.	2^{2-1} avec relation de définition I = AB
III	Estimer les effets principaux, mais ceux-ci peuvent être confondus avec des interactions à deux facteurs.	2^{3-1} avec relation de définition I = ABC
IV	Estimer les effets principaux non confondus avec des interactions à deux facteurs.	2^{4-1} avec relation de définition I = ABCD
	Estimer l'effets des interactions à deux facteurs, mais ceux-ci peuvent être confondus avec d'autres interactions à deux facteurs.	
V	Estimer les effets principaux non confondus avec les interactions à trois facteurs (ou moins).	2^{5-1} Avec relation de définition I = ABCDE
	Estimer l'effet des interactions à deux facteurs non confondus avec les interactions à deux facteurs.	
	Estimer l'effets des interactions à trois facteurs, mais ils peuvent être confondus, avec d'autres interactions à deux facteurs.	
	Estimer les effets principaux non confondus avec des interactions à quatre facteurs (ou moins).	
	Estimer l'effet des interactions à deux facteurs	

VI	non confondus avec des interactions à trois facteurs (ou moins)	2^{6-1} avec relation de définition I = ABCDEF
	Estimer l'effet des interactions à trois facteurs, mais ils peuvent être confondus avec interactions à trois facteurs.	

Formellement, la résolution d'un plan correspond à la longueur minimale des mots dans la relation de définition excluant (I). Les plans fractionnaires les plus importants sont ceux des résolutions III, IV et V: les résolutions inférieurs à III ne sont pas utiles et les résolutions supérieurs à V sont inutiles, dans la mesure où l'expérimentation étendue ne présente aucun avantage pratique dans la plupart des cas. La structure 2^{5-2} de l'exemple ci-dessus correspond à la résolution III puisque sa relation de définition est I = ABD = ACE = BCDE.

2.2. Génération d'alias:

Les générateurs d'alias servent à calculer la structure des alias qui décrit les confusions des plans factoriels fractionnaires. Toute lettre multipliée par elle-même désigne l'identité I. I multiplié par toute lettre désigne cette lettre. Les lettres sont commutatives et associatives. Pour obtenir tous les alias pour chaque terme nous multiplions les termes de la relation de définition par les facteurs initiaux constituant la matrice extraite.

- **Commutativité** : AB = BA;
- **Associativité** : A(BC) = (AB)C = ABC;
- **AA = I quelque soit A**.

I = ABD = ACE = BCDE

A = ABD *A + BD, A + ACE * A = CE, A = BCDE * A = ABCDE

B = ABD *B = AD, B = ACE * B = ABCE, B = BCDE * B = CDE

C = ABD * C = ABCD, C = ACE * C = AE, C = BCDE* C = BDE

D = ABD * D = AB, D = ACE * D = ACDE, D = BCDE * D = BCE

E = ABD * E = ABDE, E = ACE * E = AC, E = BCDE * E = BCD

BE = ABD * BE = ADE, BE = ACE * BE = ABC, BE = BCDE * BE = CD

BC = ABD * BC = ACD, BC = ACE * BC = ABE, BC = BCDE * BC = DE

Tableau III.2: Matrices d'expériences avec contrastes pour un plan fractionnaire 2^{5-2}

Termes initiaux

	I	**A**	**B**	**C**	**AB**	**AC**	**BC**	**ABC**	
	ABD	**BD**	**AD**	**AB**	**D**	**E**	**DE**	**CD**	
	ACE	**CE**	**CDE**	**BDE**	**BCE**	**BCD**	**ABC**	**BC**	
Exp	**BCDE**	**ABCDE**	**ABCE**	**ABCD**	**ACDE**	**ABDE**	**ACD**	**ADE**	**Y_i**
1	+	-	-	-	+	+	+	-	Y_1
2	+	+	-	-	-	-	+	+	Y_2
3	+	-	+	-	-	+	-	+	Y_3
4	+	+	+	-	+	-	-	-	Y_4
5	+	-	-	+	+	-	-	+	Y_5
6	+	+	-	+	-	+	-	-	Y_6
7	+	-	+	+	-	-	+	-	Y_7
8	+	+	+	+	+	+	+	+	Y_8
Hi	h_0	h_1	h_2	h_3	h_4	h_5	h_6	h_7	

2.3. Contrastes h_i:

Nous reconnaissons bien le modèle d'un plan factoriel complet 2^3 dont nous pourrons utiliser la matrice des effets. Les nouveaux coefficients h_i calculés avec la matrice des effets par la méthode habituelle seront nommés contrastes. Un contraste h_i est une somme d'effets de facteurs principaux et d'interactions.

$$h_0 = \alpha_0 + \alpha_{ABD} + \alpha_{ACE} + \alpha_{BCDE}$$

$$h_1 = \alpha_A + \alpha_{BD} + \alpha_{CE} + \alpha_{ABCDE}$$

$$h_2 = \alpha_B + \alpha_{AD} + \alpha_{CDE} + \alpha_{ABCE}$$

$$h_3 = \alpha_C + \alpha_{AE} + \alpha_{BDE} + \alpha_{ABCD}$$

$$h_4 = \alpha_D + \alpha_{AB} + \alpha_{BCE} + \alpha_{ACDE}$$

$$h_5 = \alpha_E + \alpha_{AC} + \alpha_{BCD} + \alpha_{ABDE}$$

$$h_6 = \alpha_{BE} + \alpha_{CD} + \alpha_{ABC} + \alpha_{ADE}$$

$$h_7 = \alpha_{BC} + \alpha_{DE} + \alpha_{ABE} + \alpha_{ACD}$$

Le modèle mathématique qui lie les contrastes au différentes colonnes peut être décrit sous la même forme que celles des plans complets.

3. Estimation des effets et des interactions α_i

le calcul des contrastes h_i peut être effectué en utilisant la méthode habituelle du calcul des coefficients. Quant au calcul des effets des facteurs principaux et des interactions pour obtenir un modèle en fonction des différents termes, un ensemble d'hypothèses peuvent être appliquées pour déduire les différentes valeurs des α_i.

- **Hypothèse 1**: Les interactions d'ordre 3 et plus sont supposées négligeables.
- **Hypothèse 2**: S'il s'avère qu'un contraste est négligeable, tous les termes aliasés avec ce contraste sont aussi négligeables.
- **Hypothèse 3**: Si les effets de deux facteurs sont négligeables, l'effet de leur interaction est aussi considéré comme négligeable.
- **Hypothèse 4**: Si l'effet d'un des facteurs qui est composant d'une interaction est négligeable, l'interaction est généralement.

4. Exemple applicatif

Nous nous intéressons à étudier l'effet des 4 facteurs A, A, C et D sur une réponse y. La matrice du plan fractionnaire 2^{4-1} avec réponses est donnée par la figure (3.4).

Tableau III.3: Matrice d'expériences avec réponse pour le plan fractionnaire 2^{5-2}

Exp	I	A	B	C	D	Y
1	1	-1	-1	-1	-1	55
2	1	1	-1	-1	1	65
3	1	-1	1	-1	1	42
4	1	1	1	-1	-1	44
5	1	-1	-1	1	1	58
6	1	1	-1	1	-1	74
7	1	-1	1	1	-1	52
8	1	1	1	1	1	54

La matrice du plan fractionnaire 2^{4-1} avec interactions peut être obtenue à partir d'une matrice d'un plan complet 2^3. Le générateur d'alias pour ce plan est: I = ABCD. A partir de ce générateur nous pouvons définir tous les alias qui correspondant aux différents termes (A, B, C, AB, AC, BC et ABC) de la matrice en multipliant le générateur d'alias par les termes de matrice du plan 2^3).

$$A = A * I = A * ABCD = A * A * BCD = I * BCD = BCD$$

De la même façon les autres alias pou les autres termes peuvent être déterminer pour avoir:

$$B = ACD,\ C = ABD,\ AB = CD,\ AC = BD = BC = AD \text{ et } ABC = D$$

les contrastes sont calculés de la même manière que les effets. Les confusions des différents coefficients sont: $h_0 = \alpha_0 + \alpha_{ABCD}$, $h_1 = \alpha_A + \alpha_{BCD}$, $h_2 = \alpha_B + \alpha_{ACD}$, $h_3 = \alpha_C + \alpha_{ABD}$, $h_4 = \alpha_{AB} + \alpha_{CD}$, $h_5 = \alpha_{AC} + \alpha_{BD}$, $h_6 = \alpha_{BC} + \alpha_{AD}$, $h_7 = \alpha_D + \alpha_{ABC}$.

Tableau III.4: Matrice d'expériences avec contrastes pour le plan fractionnaire 2^{5-2}

Exp	I	A	B	C	AB	AC	BC	ABC	Y
	ABCD	BCD	ACD	ABD	CD	BD	AD	D	
1	1	-1	-1	-1	1	1	1	-1	55
2	1	1	-1	-1	-1	-1	1	1	65
3	1	-1	1	-1	-1	1	-1	1	42
4	1	1	1	-1	1	-1	-1	-1	44
5	1	-1	-1	1	1	-1	-1	1	58
6	1	1	-1	1	-1	1	-1	-1	74
7	1	-1	1	1	-1	-1	1	-1	52
8	1	1	1	1	1	1	1	1	54
Contraste	h_0=55.5	h_1=3.75	h_2=7.5	h_3=4	h_4= -2.75	h_5=0.75	h_6=1	h_7= -55.5	

Dans cette étude, les contrastes dont la valeur absolue est inférieure à 1.6 sont supposés négligeables. Un contraste présente une confusion de plusieurs effets. Pour estimer les effets des facteurs ou leurs interactions, nous appliquons les hypothèses sus-présentées:

$h_0 = \alpha_0 + \alpha_{ABCD}$, en appliquant **l'hypothèse 1**: $\alpha_{ABCD} = 0 \rightarrow \alpha_0 = h_0 = 55.5$.

$h_1 = \alpha_A + \alpha_{BCD}$, en appliquant **l'hypothèse 1**: $\alpha_{BCD} = 0 \rightarrow \alpha_A = h_1 = 3.75$.

$h_2 = \alpha_B + \alpha_{ACD}$, en appliquant **l'hypothèse 1**: $\alpha_{ACD} = 0 \rightarrow \alpha_B = h_2 = -7.5$.

$h_3 = \alpha_C + \alpha_{ABD}$, en appliquant **l'hypothèse 1**: $\alpha_{BD} = 0 \rightarrow \alpha_C = h_3 = 4$.

$h_7 = \alpha_D + \alpha_{ABC}$, en appliquant **l'hypothèse 1**: $\alpha_{ABC} = 0 \rightarrow \alpha_D = h_7 = -0.75$.

$|h_5| < 1.6$, en appliquant **l'hypothèse 2** $\rightarrow \alpha_{AC} = \alpha_{BD} = 0$.

$|h_6| < 1.6$, en appliquant **l'hypothèse 2** $\rightarrow \alpha_{BC} = \alpha_{AD} = 0$.

$|h_7| < 1.6$, en appliquant **l'hypothèse 2** $\rightarrow \alpha_D = 0$.

$\alpha_D = 0$, en appliquant **l'hypothèse 4** $\rightarrow \alpha_{CD} = 0$.

$h_4 = \alpha_{AB} + \alpha_{CD}$ et $\alpha_{CD} = 0 \rightarrow \alpha_{AB} = h_4 = -2.75$.

Le modèle mathématique peut donc être écrit comme suit: y = 55.5 + 3.75A - 7.5B + AC - 2.75 B.

Application 1: Screening de 5 Facteurs en 8 Essais

Optimisation d'un procédé de synthèse chimique dans l'industrie pharmaceutique. Budget limité à 8 essais pour identifier les facteurs critiques.

Données Complètes:

Facteurs et niveaux:

- A : Température (80°C vs 100°C);
- B : Pression (1 bar vs 2 bar);
- C : Temps de réaction (30 min vs 60 min);
- D : Concentration catalyseur (1% vs 2%);
- E : Type d'agitation (lente vs rapide).

Plan $2^{(5-2)}$ avec générateurs D = AB, E = AC:

Essai	A	B	C	D	E	Rendement (%)	Pureté (%)	Coût (€)
1	-1	-1	-1	+1	+1	65,2	95,8	120
2	+1	-1	-1	-1	-1	72,1	96,2	180
3	-1	+1	-1	-1	+1	58,4	94,5	160
4	+1	+1	-1	+1	-1	85,3	97,8	220
5	-1	-1	+1	+1	-1	62,7	95,1	140
6	+1	-1	+1	-1	+1	78,6	96,9	200
7	-1	+1	+1	-1	-1	55,9	93,8	180
8	+1	+1	+1	+1	+1	92,4	98,5	240

Informations supplémentaires:

- Précision des mesures : ±0,5%;
- Coût par essai : 500€;
- Production annuelle : 10 000 unités;
- Valeur économique : +5€ par % de rendement.

Questions:

1. Construire la matrice d'expériences avec l'ordre randomisé?
2. Calculer les effets principaux pour le rendement et la pureté?
3. Déterminer la structure d'alias complète?
4. Analyser la signification des effets (α=0,05) avec test de Student?

5. Proposer un modèle mathématique pour le rendement?
6. Effectuer une analyse économique et calculer le ROI?
7. Déterminer les conditions optimales avec l'optimiseur de réponse?
8. Élaborer un plan de validation expérimentale?

Solution Application 1

- L'objectif de l'application 1;

- Trouver quels paramètres sont importants pour améliorer le rendement, en faisant le moindre d'essais possible;

- Le problème concret.

- Vous avez un procédé chimique et 5 "boutons" à ajuster:

Bouton	Position Basse	Position Haute
Température	80°C	100°C
Pression	1 bar	2 bar
Temps	30 min	60 min
Concentration	1%	2%
Agitation	Lente	Rapide

Problème: Tester toutes les combinaisons = 32 essais → Trop long et cher!
Solution: Plan fractionnaire → 8 essais seulement!

L'Analyse des Résultats

Question 1: Quel paramètre est le PLUS important ?

Regardons la température:

- Quand température = 80°C : 65, 58, 62, 55 → moyenne = **60%**
- Quand température = 100°C : 72, 85, 78, 92 → moyenne = **81,75%**

Effet température = 81,75% - 60% = +21,75% → TRÈS IMPORTANT!.

Question 2: Et les autres paramètres?.

Même calcul pour chacun:

Paramètre	Effet sur le rendement	Importance
Température	+21,75%	TRÈS FORTE
Concentration	+10,25%	FORTE
Agitation	+4,75%	MOYENNE
Pression	+3,25%	FAIBLE
Temps	+1,75%	TRÈS FAIBLE

Le "Modèle Mathématique" - Version Cuisine

Rendement = 70,875 + 10,875×(Température) + 5,125×(Concentration) + 2,375×(Agitation) + 1,625×(Pression) + 0,875×(Temps)

Traduction:

- **70,875%** = rendement de base;
- **+10,875%** si température à 100°C (au lieu de 80°C);
- **+5,125%** si concentration à 2% (au lieu de 1%);
- etc.

L'Optimisation - La Meilleure Recette

Pour maximiser le rendement:

- Température = 100°C (+21,75%);
- Concentration = 2% (+10,25%);
- Agitation = Rapide (+4,75%);
- Pression = 2 bar (+3,25%);
- Temps = 60 min (+1,75%).

Rendement prédit = 91,75%

Intervalle de confiance:

On est sûr à 95% que le vrai rendement sera entre **90,43% et 93,07%**

L'Analyse Économique - Version Simple

Coût de l'étude:

- 8 essais × 500€ = **4 000€**

Gain annuel:

- Rendement amélioré : +21,67% (de 70,08% à 91,75%)
- Valeur : +5€ par % de rendement gagné
- Production : 10 000 unités/an

Calcul:

21,67% × 5€/% × 10 000 unités = 1 083 500€ de gain annuel

Retour sur investissement ROI:

ROI = (1 083 500€ - 4 000€) / 4 000€ = 269,875

Traduction: Pour chaque 1€ investi dans l'étude, vous gagnez **269€ par an**!

La Structure d'Alias - Le "Piège" du Plan Réduit

Le Problème:

Avec seulement 8 essais, on ne peut pas tout mesurer parfaitement. Certains effets sont "mélangés".

Exemple:

- L'effet "Température" est mélangé avec l'interaction "Pression×Concentration"
- Mais comme ces interactions sont généralement faibles, on suppose qu'elles sont nulles

La Vérification:

Dans votre cas, tous les effets principaux sont si forts qu'ils ressortent clairement malgré ce "mélange".

Les Conclusions Pratiques:

Ce Que Vous Avez Découvert :

1. **La température** est le paramètre LE PLUS important;
2. **La concentration** est très importante aussi;
3. **L'agitation** a un effet notable;
4. **Pression et temps** ont peu d'influence.

Les Actions à Prendre:

1. Surveiller particulièrement la température et la concentration;
2. Maintenir l'agitation rapide;
3. Économiser sur la pression et le temps (niveaux bas).

Le Gain Réel:

- Rendement passé de ~70% à ~92% → **+22%**
- Coût de l'étude : 4 000€
- Gain annuel : 1 083 500€
- ROI exceptionnel : 26 987%

L'Intelligence de Cette Méthode:

Ce Que Vous Avez Évité:

- 32 essais inutiles;
- Des semaines de travail perdues;
- Des coûts élevés sans certitude.

Ce Que Vous Avez Gagné:

- 8 essais seulement;
- Une compréhension claire des paramètres importants;
- Une optimisation scientifique
- Un gain économique énorme.

Application 2: Étude de Robustesse $2^{(4-1)}$

Validation d'une méthode analytique HPLC pour un nouveau principe actif. Vérification de la robustesse face aux variations normales.

Facteurs et niveaux :

- A: pH du tampon (6,0 vs 8,0);
- B: Concentration (10 mM vs 20 mM);
- C: Température (25°C vs 35°C);
- D: Temps d'analyse (5 min vs 10 min).

Plan $2^{(4-1)}$ avec générateur D = ABC :

Essai	A	B	C	D	Pureté (%)	Temps rétention (min)	Résolution	Area (%)
1	-1	-1	-1	+1	98,52	4,25	2,15	99,85
2	+1	-1	-1	-1	97,83	4,18	2,08	99,78
3	-1	+1	-1	-1	98,21	4,32	2,12	99,82
4	+1	+1	-1	+1	97,48	4,15	2,05	99,75
5	-1	-1	+1	+1	98,03	4,28	2,10	99,80
6	+1	-1	+1	-1	97,25	4,12	2,02	99,72
7	-1	+1	+1	-1	97,91	4,35	2,09	99,79
8	+1	+1	+1	+1	96,78	4,04	1,98	99,68

Points centraux (3 réplicats):

Essai	A	B	C	D	Pureté (%)
9	0	0	0	0	98,15
10	0	0	0	0	98,08
11	0	0	0	0	98,22

Spécifications qualité:

- Pureté $\geq$ 97,0%
- Temps rétention : 4,20 $\pm$ 0,15 min
- Résolution $\geq$ 2,0
- Area $\geq$ 99,70%

Questions:

1. Analyser les effets principaux et les interactions?
2. Effectuer le test de courbure avec les points centraux?
3. Vérifier la robustesse pour chaque paramètre qualité?
4. Déterminer les limites de contrôle qualité?
5. Établir les spécifications opérationnelles?
6. Rédiger le protocole de validation?

Solution Application 2

Qu'est-ce qu'une étude de robustesse ?

C'est comme tester si votre recette de cuisine fonctionne toujours bien même si:

- Vous changez un peu la température du four
- Vous utilisez un peu plus ou moins de sel
- Vous laissez cuire un peu plus ou moins longtemps

Objectif: Vérifier que la méthode analytique donne de bons résultats même avec de petites variations normales.

Ce qu'on Vérifie Exactement
Les 4 Facteurs Étudiés:

Facteur	Niveau Bas	Niveau Haut	Variation Normale
pH	6,0	8,0	±0,5
Concentration	10 mM	20 mM	±2,5 mM
Température	25°C	35°C	±2,5°C
Temps	5 min	10 min	±1,25 min

Les Critères de Qualité:

- **Pureté** ≥ 97,0%
- **Temps de rétention** : 4,20 ± 0,15 min
- **Résolution** ≥ 2,0
- **Surface du pic** ≥ 99,70%

<u>Comment On Fait le Test</u>

Le Plan Expérimental:

On fait seulement **8 essais** (au lieu de 16) en utilisant un plan fractionnaire.

<u>Analyse des Résultats</u>

1. Calcul des "Effets":

On regarde comment chaque facteur influence la pureté :

Facteur	**Influence sur la pureté**
pH	-0,83% (baisse faible)
Concentration	-0,31% (baisse très faible)
Température	-0,52% (baisse faible)
Temps	-0,08% (baisse négligeable)

2. Test de Courbure:

<u>Question:</u> Est-ce que la relation est linéaire ou courbée ?

<u>Réponse:</u> Il y a une **légère courbure** détectée, mais elle est très faible.

3. Vérification des Spécifications:

Paramètre	**Pire cas observé**	**Spécification**	**Conclusion**
Pureté	96,78%	≥97,0%	Très proche
Temps rétention	4,08-4,35 min	4,20±0,15 min	OK
Résolution	1,98	≥2,0	Limite
Surface	99,68%	≥99,70%	OK

La Méthode est-elle Robuste?

Oui, avec quelques précautions:

Ce qui est Bon:

- Les variations normales n'affectent pas significativement les résultats;
- La méthode est reproductible;
- La plupart des paramètres restent dans les spécifications.

Ce qu'il faut Surveiller:

- La pureté peut descendre très près de la limite (96,78%);

- La résolution est parfois à la limite inférieure.

Recommandations Pratiques

Conditions Opératoires Recommandées:

pH : 7,0 ± 0,3 (plus serré que ±0,5)

Concentration : 15 mM ± 2 mM

Température : 30°C ± 2°C

Temps : 7,5 min ± 1 min

Plan de Contrôle Qualité:

- Faire **1 contrôle** toutes les 20 analyses;
- Vérifier particulièrement la pureté et la résolution;
- Si pureté < 97,2% → revérifier l'étalonnage.

En Cas de Problème:

- Vérifier d'abord le pH et la température**;**
- Ce sont les facteurs les plus influents.

Pour Résumer en Une Phrase

"La méthode fonctionne bien dans les conditions normales, mais il faut faire attention au pH et à la température qui peuvent faire baisser un peu la pureté."

Les Avantages de Cette Étude

Ce Que Vous Avez Gagné:

1. Confiance dans votre méthode;
2. Connaissance des facteurs critiques;
3. Limites de contrôle bien définies;
4. Économie de temps et d'argent (8 essais au lieu de 16).

Application 3: Stratégie Séquentielle

Développement d'un nouveau matériau composite haute performance. 6 facteurs potentiels à étudier avec approche séquentielle.

Données Complètes:

Phase 1 : Screening $2^{(6-3)}$

Facteurs et niveaux :

- A : Température cuisson (180°C vs 220°C)
- B : Pression (50 bar vs 100 bar)
- C : Temps (30 min vs 60 min)
- D : % Fibres (40% vs 60%)
- E : Type résine (A vs B)
- F : Agent mouillant (1% vs 2%)

Plan $2^{(6-3)}$) avec générateurs D = AB, E = AC, F = BC

Essai	A	B	C	D	E	F	Résistance (MPa)	Module (GPa)	Coût (€/kg)
1	-1	-1	-1	+1	+1	+1	245	25,8	45,2
2	+1	-1	-1	-1	-1	+1	312	28,5	52,8
3	-1	+1	-1	-1	+1	-1	268	26,4	48,5
4	+1	+1	-1	+1	-1	-1	385	30,2	56,3
5	-1	-1	+1	+1	-1	-1	252	25,9	46,1
6	+1	-1	+1	-1	+1	-1	348	29,1	54,2
7	-1	+1	+1	-1	-1	+1	292	27,2	49,8
8	+1	+1	+1	+1	+1	+1	428	31,8	58,9

Phase 2: Plan complémentaire $2^{(6-3)}$

Nouveaux générateurs : D = -AB, E = -AC, F = -BC

Essai	A	B	C	D	E	F	Résistance (MPa)	Module (GPa)
9	-1	-1	-1	-1	-1	-1	238	25,5
10	+1	-1	-1	+1	+1	-1	308	28,2
11	-1	+1	-1	+1	-1	+1	261	26,1
12	+1	+1	-1	-1	+1	+1	378	29,8

13	-1	-1	+1	-1	+1	+1	248	25,7
14	+1	-1	+1	+1	-1	+1	342	28,9
15	-1	+1	+1	+1	+1	-1	285	26,9
16	+1	+1	+1	-1	-1	-1	415	31,2

Données économiques:

- Coût essai : 800€
- Production : 5 000 kg/an
- Prix vente : 80€/kg (résistance > 350 MPa)
- Prix vente : 60€/kg (résistance ≤ 350 MPa)
- Coût cible : ≤ 50€/kg

Questions:

1. Analyser le plan screening initial
2. Identifier les ambiguïtés de la structure d'alias
3. Analyser le plan combiné (16 essais)
4. Développer les modèles mathématiques finaux
5. Optimiser avec contraintes techniques et économiques
6. Calculer le retour sur investissement détaillé (**ROI**)
7. Élaborer le plan de déploiement industriel
8. Proposer un plan de contrôle statistique procédé

Solution Application 3

Imaginez que vous développez une **nouvelle recette de matériau composite** (comme du carbone ou de la fibre de verre). Vous avez **6 ingrédients** possibles mais vous ne savez pas lesquels sont vraiment importants.

Problème : Tester toutes les combinaisons coûterait **trop cher et trop de temps**.

Solution : Approche en **2 étapes** intelligente !

Étape 1: Le Screening (Dépistage)

Les 6 "Ingrédients" à Tester:

Ingrédient	Option 1	Option 2
A. Température	180°C	220°C
B. Pression	50 bar	100 bar
C. Temps	30 min	60 min
D. % Fibres	40%	60%
E. Type résine	A	B
F. Agent mouillant	1%	2%

Le Plan Malin:

Au lieu de faire **64 essais** (trop cher!), on fait seulement **8 essais** en mélangeant astucieusement les ingrédients.

Résultats des 8 premiers essais:

- Meilleur essai: **428 MPa** (très bon !)
- Pire essai: **245 MPa** (moyen)

Ce Qu'on Découvre:

- **A (Température)** et **B (Pression)** sont très importants;
- Les autres ont peu d'influence;
- **Problème :** On ne sait pas si A et B agissent ensemble (interaction).

Étape 2: Le Plan Complémentaire

Pourquoi faire plus d'essais?

On a un doute: est-ce que c'est vraiment A et B qui sont importants, ou est-ce une illusion due au plan réduit ?

Les 8 Essais Supplémentaires

On refait 8 essais avec une "recette" différente pour vérifier.

Résultat: On confirme que:

- **A (Température)** est crucial;
- **B (Pression)** est crucial;
- **L'interaction A×B** est importante (ils agissent ensemble);

- **C, D, E, F** sont peu importants.

Les Résultats Finaux

La "Recette" Optimale:

- **Température:** 220°C (niveau haut);
- **Pression:** 100 bar (niveau haut);
- **Temps:** 30 min (niveau bas - pour économiser);
- **% Fibres:** 40% (niveau bas - pour économiser);
- **Type résine:** A (niveau bas - pour économiser);
- **Agent mouillant:** 1% (niveau bas - pour économiser).

Performance Prédite:

- **Résistance: 405 MPa** (excellent !);
- **Coût: 48,2 €/kg** (dans le budget).

L'Analyse Économique

Combien Ça a Coûté ?.

- 16 essais × 800€ = **12 800€**

Combien Ça Rapporte ?

Avant l'optimisation:

- Résistance : environ 300 MPa
- Prix de vente : 60€/kg (car résistance moyenne)

Après optimisation:

- Résistance : 405 MPa → **produit haute performance**
- Prix de vente : 80€/kg
- Coût production : 48,2€/kg
- Marge: 31,8€/kg

Calcul du gain annuel:

5 000 kg/an × 31,8€/kg = 159 000€ de bénéfice par an

Le Retour sur Investissement (ROI):

- Gain annuel : 159 000€;
- Coût étude : 12 800€;
- ROI = (159 000 / 12 800) = 12,4.

Cela signifie: Pour chaque 1€ investi dans l'étude, vous gagnez 12,4€ par an !

Le Plan de Mise en Œuvre

Phase 1: Validation

- Faire 3 tests de confirmation;
- Vérifier que ça marche à plus grande échelle;
- Former les opérateurs.

Phase 2: Production

- Adapter les machines;
- Mettre en place les contrôles qualité;
- Produire les premiers lots.

Phase 3: Surveillance Continue

- Contrôler régulièrement la résistance;
- Vérifier que le procédé reste stable;
- Améliorer si nécessaire.

Le Système de Contrôle Qualité

Comment Surveiller la Production:

- **Fréquence :** 1 échantillon par lot;
- **Mesures :** Résistance, coût, apparence.

- **Seuils d'alerte:**

○ Résistance < 395 MPa → enquête;

○ Résistance < 385 MPa → arrêt production.

Les Avantages:

- Détection rapide des problèmes
- Qualité constante
- Satisfaction clients

Le Bilan Final

Ce Que Vous Avez Accompli:

1. Identifié les 2 facteurs vraiment importants sur 6;
2. Optimisé la recette pour performance maximale;
3. Réduit les coûts en fixant les autres facteurs au niveau économique;
4. Validé la solution avec une approche scientifique;
5. Calculé un retour sur investissement excellent.

Les Économies Réalisées:

- **Étude :** 48 essais économisés (64 - 16) = 38 400€ sauvés
- **Production :** Coût réduit grâce aux facteurs au niveau économique
- **Gain :** 159 000€/an de bénéfice supplémentaire

Exercices supplémentaires

Exercice 1: Choix de résolution

Scénario: Vous devez étudier 7 facteurs avec un budget de 16 essais maximum.

Options:

- Plan $2^{(7\text{-}3)}$: 16 essais, résolution IV
- Plan $2^{(7\text{-}4)}$: 8 essais, résolution III

Question: Quel plan choisiriez-vous et pourquoi ?

Réponse:

- **$2^{(7\text{-}3)}$ résolution IV** préférable car:

Effets principaux non confondus entre eux;

Interactions d'ordre 2 partiellement aliassées;

Meilleur compromis information/budget.

Exercice 2: Interprétation de structure d'alias

Structure donnée:

$$I = ABCDE$$
$$A = BCDE$$
$$B = ACDE$$
$$C = ABDE$$
$$D = ABCE$$
$$E = ABCD$$
$$AB = CDE$$
$$AC = BDE$$
$$AD = BCE$$
$$AE = BCD$$

Questions:

1. Quelle est la résolution de ce plan?.
2. Peut-on estimer l'interaction AB de manière fiable?.
3. Que risque-t-on si l'interaction BCDE est importante?.

Réponses:

1. Résolution V (plus petit mot a 5 lettres).
2. Non, AB est confondu avec CDE.
3. Confusion entre l'effet principal A et l'interaction BCDE.

Recommandations pratiques:

1. Toujours documenter la structure d'alias avant l'expérimentation.
2. Prévoir un budget pour une éventuelle phase de déconfusion.
3. Utiliser les connaissances métier pour interpréter les résultats ambigus.
4. Ne pas sur interpréter les effets marginaux dans les plans de faible résolution.

1. Limitations des plans factoriels

Les plans 2^k que nous avons vus jusqu'à présent sont excellents pour:

- Identifier les facteurs influents;
- Estimer les effets principaux et interaction;
- Faire du screening.

Mais ils sont limités pour:

- Capturer les courbures dans la réponse;
- Trouver des optimums précis;
- Modéliser des relations non linéaires**.**

<u>Solution:</u> les modèles du second ordre

Quand la courbure est détectée (via points centraux), nous devons utiliser un modèle plus complexe:

Modèle du second ordre:

$$Y = \beta_0 + \Sigma\beta_i X_i + \Sigma\beta_{ii} X_i^2 + \Sigma\Sigma\beta_{ij} X_i X_j + \varepsilon$$

Pour estimer ce modèle, nous avons besoin de plans spécifiques: les plans de surface de réponse**.**

2. Notion de surface de réponse et courbes isoréponses

<u>2.1. Surface de réponse:</u>

Définition: Représentation géométrique de la relation entre les facteurs et la réponse.

Pour 2 facteurs: Une surface en 3D.

```
Rendement (Z) ^
          |
          |      /\
          |     /  \
          |    /    \
          |   /      \
          | /__________________\--> Facteur X
 Facteur Y
```

Pour 3+ facteurs: Hyper-surface difficile à visualiser.

2.2. Courbes isoréponses (lignes de niveau):

Définition: Projection 2D de la surface de réponse montrant les combinaisons de facteurs donnant la même valeur de réponse.

Exemple pour $Y = f(X_1, X_2)$:

```
X2 ^
  |
  |  80    85    90
  |     \    |    /
  |       \  |  /
  |  75   \ | /   95
  |        \|/
  |_____________\__________> X2
  |          /|\
  |         / | \
  |        /  |  \
```

Interprétation:

- Les lignes relient les points de même rendement;
- La forme indique l'interaction entre les facteurs;
- Le centre des courbes fermées indique un optimum.

2.3. Types d'optimums:

Maximum local:

```
    /\
   /  \
  /    \
```

Minimum local:

```
\      /
 \    /
  \__/
```

Point selle:

```
  __    __
 /  \__/  \
/          \
```

3. Plans pour l'étude des modèles du second degré

3.1. Plan Box-Behnken:

Caractéristiques:

- Niveaux: -1, 0, +1 seulement;
- Nombre d'essais raisonnable;
- **Pas de points aux coins** → avantage pour les contraintes opérationnelles;
- Propriétés de rotabilité approximative.

Structure pour 3 facteurs:

12 essais + points centraux

Essai	X_1	X_2	X_3
1	-1	-1	0
2	-1	+1	0
3	+1	-1	0
4	+1	+1	0
5	-1	0	-1
6	-1	0	+1
7	+1	0	+1
8	+1	0	+1
9	0	-1	-1
10	0	-1	+1
11	0	+1	-1
12	0	+1	+1
Points centraux	0	0	0

Avantage:

- Economique en nombre d'essais;
- Evite les conditions extrêmes (coins);
- Bon pour l'exploration prudent.

Limitation:

- Ne couvre pas tout le domaine
- Variance de prédiction plus élevée aux coins

3.2. Plan Composite Centré (PCC):

Caractéristiques:

- Combinaison de trois parties:
  1. **Cube**: points des plans factoriels 2^k;
  2. **Points axiaux** (étoile): sur les axes;
  3. **Points centraux**: au centre.

Structure pour 2 facteurs:

```
      X₂
      ^
      |   (0,α)
      |
(-1,+1) +----+----+ (+1,+1)
      |   |   |
      |----+----+----> X₁
(-1,-1) +----+----+ (+1,-1)
      |
      |   (0,-α)
```

Nombre d'essais: $2^k + 2^k + n_0$

- **2^k :** points du cube;
- **2^k:** points axiaux;
- **n_0:** points centraux.

Paramètre α (alpha):

- **PCC orthogonal**: α calculé pour l'orthogonalité;
- **PCC rotable**: $\alpha = 2^{(k/4)}$ pour que la variance de prédiction ne dépende que de la distance au centre.

Exemple PCC 3 facteurs:

- Points cube: 8 (plan 2^3);
- Points axiaux: 6 (2×3);
- Points centraux: 6;

Total: 20 essais.

Avantages:

- Estimation précise des coefficients quadratiques;
- Propriétés d'optimalité;
- Couverture complète du domaine.

Limitations:

- Nombre d'essais plus important.
- Conditions extrêmes (coins) incluses.

4. Critères de qualité et d'optimalité d'un plan expérimental

4.1. Mesures de qualité d'un plan:

Matrice de variance-covariance:

$Var(\beta) = \sigma^2(X'X)^{-1}$;

La qualité dépend de la matrice $(X'X)^{-1}$.

Critères d'optimalité:

A-optimalité:

- Minimise la trace de $(X'X)^{-1}$;

- Minimise la variance moyenne des coefficients.

D-optimalité:

- Minimise le déterminant de $(X'X)^{-1}$;
- Minimise le volume de l'ellipsoïde de confiance;
- Très utilisé en pratique.

G-optimalité:

- Minimise la variance maximale de prédiction dans le domaine;
- Assure une précision homogène.

I-optimalité:

- Minimise la variance moyenne de prédiction sur le domaine;
- Bon pour l'optimisation.

Propriétés souhaitables:

Orthogonalité:

- Les colonnes de X sont orthogonales;
- Estimation independent des coefficients.

Rotabilit:

- La variance de prédiction ne dépend que de la distance au centre;
- Qualité de prédiction isotrope.

Uniformité:

- Répartition uniforme des points dans l'espace.

4.2. Calcul des plans optimaux:

Méthode des plans D-optimaux:

1. Définir le modèle (termes à inclure);
2. Définir le domaine expérimental;
3. Définir le nombre d'essais;
4. Algorithme d'échange pour sélectionner les points optimaux.

Avantages des plans optimaux:

- Adaptés aux contraintes pratiques;
- Nombre d'essais personnalisable;

- Peuvent inclure des points déjà réalisés.

Limitations:

- Propriétés théoriques moins garanties;
- Dépendent de l'algorithme utilisé.

4.3. Stratégie de modélisation et validation:

Procédure recommandée:

1. **Screening** → Plans factoriels/fractionnaires;
2. **Détection courbure** → Points centraux;
3. **Modélisation surface** → Plans de surface de réponse;
4. **Validation** → Essais de confirmation.

Validation du modèle:

Graphiques diagnostiques:

- Résidus vs valeurs ajustées;
- Normalité des résidus;
- Valeurs aberrantes.

Tests statistiques:

- R^2 et R^2 ajusté;
- Test de manque d'ajustement;
- Validation externe avec nouveaux points.

Interprétation pratique:

Analyse de la surface:

- Localisation des optimums;
- Plateaux de réponse;
- Sensibilité aux variations.

Courbes isoréponse :

- Visualisation des compromis;
- Zones d'opération robuste.

5. Synthèse du chapitre

Choix du plan de surface de réponse:

Situation	Plan recommandé	Avantages
Domaine cubique, pas de contraintes	PCC	Propriétés optimales, estimation précise
Éviter les conditions extrêmes	Box-Behnken	Sécurité, nombre d'essais réduit
Contraintes opérationnelles	D-optimal	Flexibilité, adaptation aux limites
Budget très limité	Hybride ou petit composite	Économique

Check-list de mise en œuvre:

1. Vérifier la présence de courbure (points centraux);
2. Choisir le plan adapté au domaine et contraintes;
3. Déterminer le nombre d'essais réalisable;
4. Valider les hypothèses du modèle;
5. Interpréter les résultats en contexte métier;
6. Confirmer par essais de validation.

Conclusion

Les plans de surface de réponse sont indispensables pour l'optimisation fine des procédés. Le choix entre Box-Behnken, PCC et plans optimaux dépend du contexte opérationnel, des contraintes et des objectifs de l'étude.

Des questions sur ces plans de surface de réponse ? Nous allons maintenant passer à la séance de TD pour mettre en pratique ces concepts.

Application 01: Optimisation d'un procédé chimique avec Box-Behnken

Énoncé:

Vous devez optimiser le rendement d'une réaction chimique. Après une étude préliminaire, 3 facteurs critiques ont été identifiés :

- A: Température (60°C vs 80°C);
- B: pH (6 vs 8);
- C: Temps de réaction (2h vs 4h).

Variable réponse: Rendement (%)

Contrainte: Éviter les conditions extrêmes (coins du domaine)

Étapes Minitab:

Étape 1: Création du plan Box-Behnken

1. Stat > DOE > Response Surface > Create Response Surface Design;
2. Choisir "Box-Behnken";
3. Number of factors: 3;
4. Cliquer sur Designs:
   - Default number of center points: 3;
   - Total runs: 15;
5. Factors:
   - A: Temperature (60, 80);
   - B: pH (6, 8);
   - C: Time (2, 4)

Étape 2: Matrice d'expériences et résultats

Essai	A	B	C	Rendement
1	-1	-1	0	72
2	-1	1	0	68
3	1	-1	0	85
4	1	1	0	82
5	-1	0	-1	65
6	-1	0	1	70

7	**1**	**0**	**-1**	**80**
8	**1**	**0**	**1**	**88**
9	**0**	**-1**	**-1**	**75**
10	**0**	**-1**	**1**	**78**
11	**0**	**1**	**-1**	**72**
12	**0**	**1**	**1**	**76**
13	**0**	**0**	**0**	**90**
14	**0**	**0**	**0**	**91**
15	**0**	**0**	**0**	**89**

Étape 3: Analyse du modèle

1. Stat > DOE > Response Surface > Analyze Response Surface Design;
2. Vérifier les termes significatifs.

Résultats:

text

Coefficients (données codées):

Terme	Coef	p-value
Constante	**90.00**	**0.000**
A	**6.25**	**0.002**
B	**-1.75**	**0.150**
C	**3.25**	**0.020**
A*A	**-2.50**	**0.080**
B*B	**-3.00**	**0.045**
C*C	**-2.00**	**0.120**
A*B	**1.50**	**0.250**
A*C	**2.00**	**0.150**
B*C	**-1.00**	**0.400**

$R^2 = 92{,}5\%$ $R^2(adj) = 85{,}0\%$

Étape 4: Optimisation

1. Stat > DOE > Response Surface > Response Optimiser;
2. Goal: Maximize, Target: 100;
3. Solution: A = +0,8, B = -0,6, C = +0,9;
4. Rendement prédit: 92,3% ± 2,1%.

Conditions optimales:

- Température: 76°C;
- pH: 6,8;
- Temps: 3,8h.

Application 02: Plan Composite Centré pour formulation alimentaire

Énoncé:

Optimisation d'une sauce avec 2 facteurs:

- A: Pourcentage de tomate (20% vs 30%);
- B: Temps de cuisson (15 min vs 25 min).

Variables réponses:

- Y_1: Score sensoriel (1-10);
- Y_2: Viscosité (cP);
- Y_3: Coût de production (€/L).

Étapes Minitab:

Étape 1: Création du PCC

1. Create Response Surface Design
2. Choisir "Central Composite"
3. Number of factors: 2
4. Dans Designs:

- Number of center points: 5
- Alpha: 1,414 (rotable par défaut)
- Total runs: 13

Étape 2: Données collectées

Essai	A	B	Score	Viscosité	Coût
1	-1	-1	6.5	350	1.8
2	+1	-1	7.8	420	2.2
3	-1	+1	7.2	480	2.0
4	+1	+1	8.5	520	2.4
5	-1.4	0	6.8	380	1.9
6	+1.4	0	8.2	450	2.3
7	0	-1.4	7.0	360	1.8
8	0	+1.4	7.9	500	2.2
9	0	0	8.8	440	2.1

10	**0**	**0**	**8.7**	**435**	**2.1**
11	**0**	**0**	**8.9**	**445**	**2.1**
12	**0**	**0**	**8.6**	**440**	**2.1**
13	**0**	**0**	**8.8**	**445**	**2.1**

Étape 3: Analyse pour chaque réponse

Modèle Score sensoriel: $\hat{Y} = 8,76 + 0,58A + 0,32B - 0,25A^2 - 0,15B^2 + 0,20AB$

Modèle Viscosité: $\hat{Y} = 441 + 35,2A + 52,8B - 12,5A^2 + 8,2B^2 + 10,5AB$

Modèle Coût: $\hat{Y} = 2,10 + 0,18A + 0,12B + 0,05A^2 + 0,03B^2 + 0,04AB$

Étape 4: Optimisation multi-réponses

1. Response Optimiser
2. Paramètres:
   - Score: Maximiser (cible 10, limite inf 6)
   - Viscosité: Cible 450 ± 30
   - Coût: Minimiser (limite sup 2,5)
3. Importance: Score (3), Viscosité (2), Coût (1)

Solution optimale:

- A = +0,65 (28,3% tomate)
- B = +0,45 (22,3 min)
- Désirabilité globale : 0,84

Prédictions:

- Score: 8,6 ± 0,3
- Viscosité: 448 ± 15 cP
- Coût: 2,18 ± 0,08 €/L

Application 03: Plan D-optimal avec contraintes opérationnelles

Énoncé:

Optimisation d'un procédé de séchage avec 3 facteurs et contraintes:

- A: Température (50°C vs 90°C);
- B: Débit d'air (100 m³/h vs 200 m³/h);
- C: Temps de séchage (10 min vs 30 min).

Contrainte: A × C ≤ 2400 (pour éviter la dégradation).

Variable réponse: Taux d'humidité résiduelle (%).

Budget: Maximum 15 essais.

Étapes Minitab:

Étape 1: Définition des contraintes

1. Create Response Surface Design
2. Choisir "Optimal design"
3. Designs:
   - Number of factors: 3
   - Number of runs: 15
   - Model: Quadratic
4. Factors: Définir les niveaux
5. Constraints: Ajouter A*C ≤ 2400

Étape 2: Plan généré et résultats

Le plan D-optimal évite automatiquement les combinaisons interdites.

Essais sélectionnés:

text

Essai	A	B	C	Humidité
1	-1	-1	-1	8.5
2	+1	-1	-1	6.2
3	-1	+1	-1	7.8
4	-1	-1	+0.5	5.5

5	**+0.5**	**+1**	**-1**	**6.8**
6	**-1**	**+0.5**	**+1**	**4.9**
7	**+1**	**+1**	**-0.5**	**5.2**
8	**+0.5**	**-1**	**+1**	**4.1**
9	**-0.5**	**+1**	**+1**	**4.3**
10	**+1**	**-0.5**	**+0.5**	**3.8**
11	**0**	**0**	**0**	**5.0**
12	**0**	**0**	**0**	**4.8**
13	**0**	**0**	**0**	**5.1**
14	**+0.8**	**-0.8**	**+0.2**	**4.5**
15	**-0.5**	**+0.5**	**-0.5**	**6.2**

Étape 3: Analyse et validation

Modèle final: $\hat{Y} = 4{,}97 - 0{,}85A - 0{,}62B - 1{,}25C + 0{,}15A^2 + 0{,}08B^2 + 0{,}25C^2 - 0{,}18AB - 0{,}12AC + 0{,}10BC$

Validation des contraintes:

- Toutes les combinaisons respectent $A \times C \leq 2400$
- Exemple : Point 10 → 81°C × 23 min = 1863 ≤ 2400 ✓

Étape 4: Optimisation

Conditions optimales:

- A = +0,9 (86°C);
- B = -0,7 (130 m³/h);
- C = +0,8 (26 min).

Humidité prédite: 3,5% ± 0,4%.

Vérification contrainte: 86 × 26 = 2236 ≤ 2400 ✓

Exercices supplémentaires

Exercice 1: Comparaison Box-Behnken vs PCC

Scénario: Vous devez optimiser un procédé avec 3 facteurs et un budget de 20 essais maximum.

Comparaison:

Critère	Box-Behnken	PCC
Nombre d'essais	15	20
Points extrêmes	Non	Oui
Estimation quadratique	Bonne	Excellente
Variance prédiction	Plus élevée aux coins	Uniforme (rotable)
Sécurité	Élevée	Modérée

Question: Quel plan choisir si le procédé est sensible aux conditions extrêmes?

Réponse: Box-Behnken, car il évite les points aux coins.

Exercice 2: Diagnostic de modèle quadratique

Sortie Minitab:

text;

Analyse de variance pour Rendement.

Source	Ddl	SC	MC	F	P
Régression	9	850.2	94.5	15.2	0.002
Linéaire	3	620.5	206.8	33.2	0.000
Carré	3	180.3	60.1	9.6	0.008
Interaction	3	49.4	16.5	2.6	0.140
Erreur résid	5	31.2	6.2		
Manque d'ajust	3	25.8	8.6	2.4	0.280
Erreur pure	2	5.4	2.7		
Total	14	881.4			

$R^2 = 96,5\%$ $R^2(adj) = 90,1\%$

Questions:

1. Le modèle est-il significatif?

2. La courbure est-elle présente?

3.Le manque d'ajustement est-il significatif ?

4.Le modèle est-il acceptable ?

Réponses:

1. **Oui** ($p = 0{,}002 < 0{,}05$);

2. **Oui** (carré $p = 0{,}008 < 0{,}05$);

3. **Non** ($p = 0{,}280 > 0{,}05$);

4. **Oui** (modèle significatif, pas de manque d'ajustement, R^2 élevé).

Recommandations pratiques:

1. Toujours commencer par une étude screening si nombreux facteurs
2. Utiliser des points centraux pour détecter la courbure
3. Choisir le plan de surface selon les contraintes et objectifs
4. Valider le modèle par des diagnostics statistiques
5. Confirmer l'optimum par des essais de validation

Introduction

Dans ce chapitre, nous abordons un type de plans d'expériences particulièrement important pour les formulations: les plans de mélange.

1. Spécificité des problèmes de mélange

Dans les chapitres précédents, les facteurs étaient indépendants. Dans les mélanges:

- Les facteurs représentent des proportions de composants;
- La somme des proportions est toujours égale à 1 (100%);
- On ne peut pas varier un composant indépendamment des autres.

1.1. Domaines d'application:

- **Industrie alimentaire**: Formulation de produits;
- **Pharmacie**: Développement de médicaments;
- **Cosmétique**: Création de crèmes, lotions;
- **Matériaux**: Alliages, composites, polymères;
- **Pétrochimie**: Carburants, lubrifiants.

1.2. Contrainte fondamentale:

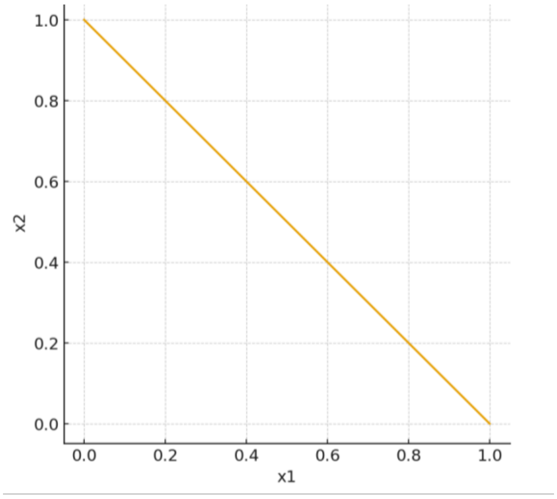
Pour q composants: $\mathbf{x_1 + x_2 + ... + x_q = 1}$

avec: $\mathbf{0 \leq x_i \leq 1}$ pour chaque composant.

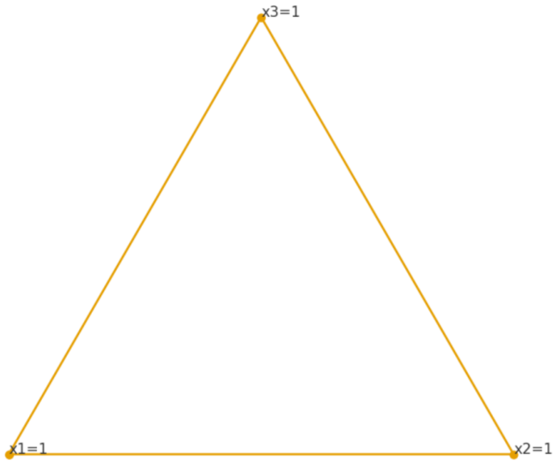
2. Représentation géométrique des mélanges

2.1. Espace expérimental:

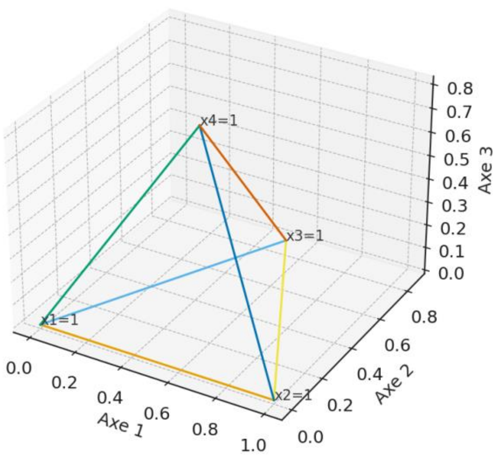
Pour 2 composants: Segment de droite



Pour 3 composants: Triangle équilatéral (simplexe)



Pour 4 composants: Tétraèdre



2.2. Points particuliers:

- **Sommet**: Mélange pur (un seul composant à 100%);

- **Arête**: Mélange binaire (2 composants seulement);
- **Face**: Mélange ternaire (3 composants);
- **Centre**: Mélange équilibré.

3. Domaine d'étude dans les plans de mélange

3.1. Contraintes supplémentaires:

En pratique, on ajoute souvent des contraintes:

- **Bornes inférieures**: $L_i \leq x_i$
- **Bornes supérieures**: $x_i \leq U_i$
- **Contraintes linéaires**: relations entre composants

3.2. Types de domaines:

Domaine complet (simplexe):

Aucune contrainte : $0 \leq x_i \leq 1$

3.2.1. Domaine contraint:

Exemple: $0{,}2 \leq x_1 \leq 0{,}6$

$0{,}1 \leq x_2 \leq 0{,}5$

$0{,}3 \leq x_3 \leq 0{,}7$

$x_1 + x_2 + x_3 = 1$

3.2.2. Domaine avec exclusion: Certaines zones interdites.

3.3. Représentation des domaines contraints

Le domaine expérimental devient un **polygone** à l'intérieur du simplex.

4. Modèles mathématiques des mélanges

4.1. Modèle linéaire (degré 1)

Formule: $E(Y) = \Sigma\, \beta_i\, x_i$.

Limitation: Ne capture pas les interactions.

4.2. Modèle quadratique (degré 2)

Formule: $E(Y) = \Sigma\, \beta_i\, x_i + \Sigma\, \Sigma\, \beta_{ij}\, x_i\, x_j$

Interprétation:

- β_i : effet du composant pur.
- β_{ij} : interaction/synergie entre composants i et j.

4.3. Modèle cubique complet (degré 3)

Formule: $E(Y) = \Sigma\, \beta_i\, x_i + \Sigma\, \Sigma\, \beta_{ij}\, x_i\, x_j + \Sigma\, \Sigma\, \Sigma\, \beta_{ijk}\, x_i\, x_j\, x_k$

Utilisation: Pour les relations complexes avec points d'inflexion

4.4. Modèles spéciaux:

Modèle de Scheffé: Modèles polynomiaux standards;

Modèle avec bornes: Adaptation aux contraintes.

5. Analyse d'un plan de mélange

5.1. Types de plans courants

5.1.1. Plan simplex-centré:

- Points aux sommets;
- Points au centre des arêtes;
- Point au centre du simplex.

5.1.2. Plan de lattice simplex:

- Répartition uniforme;
- Degré contrôlable.

5.1.3. Plans optimaux:

- Pour domaines contraints complexes
- Générés par algorithmes

5.2. Méthodologie d'analyse

1. Collecte des données selon le plan;
2. Ajustement du modèle (linéaire, quadratique, cubique);
3. Validation du modèle (R^2, R^2-adj, analyse résidus);
4. Interprétation des coefficients;
5. Optimisation de la formulation.

5.3. Représentation graphique

Courbes isoréponses: Lignes reliant les mélanges donnant la même réponse.
Surface de réponse 3D: Pour visualiser l'évolution de la réponse.

6. Aspects pratiques avancés

6.1. Variables de procédé:

Il est possible de combiner:

- Variables de mélange (proportions);
- Variables de procédé (température, temps, etc.);
- Approche: Plans d'expériences emboîtés.

6.2. Validation et robustesse:

Test de validation: Préparer 3-5 mélanges de confirmation.
Analyse de robustesse: Étudier la sensibilité aux variations de composition.

6.3. Coûts et contraintes économiques:

Intégration des coûts des matières premières dans l'optimisation.

7. Synthèse du chapitre

7.1. Avantages des plans de mélange:

- Prise en compte naturelle de la contrainte de somme;
- Optimisation systématique des formulations;
- Visualisation intuitive des résultats;

- Gestion des contraintes pratiques.

7.2. Limitations:

- Espace expérimental restreint;
- Interprétation des coefficients parfois délicate;
- Nombre d'essais peut être important pour les modèles complexes.

Check-list de mise en œuvre

1. Définir clairement les composants et leurs contraintes;
2. Choisir le plan adapté au nombre de composants et contraintes;
3. Sélectionner le degré du modèle (linéaire, quadratique, cubique);
4. Valider le modèle statistiquement;
5. Interpréter les résultats en termes de synergies/antagonismes;
6. Optimiser en considérant tous les critères importants;
7. Valider expérimentalement la solution optimale.

Conclusion

Les plans de mélange sont des outils indispensables pour l'optimisation des formulations. Leur maîtrise permet de développer des produits performants de manière rationnelle et efficace.

Application 01: Formulation d'un jus de fruits - Plan simplex

Énoncé:

Vous devez optimiser la formulation d'un jus de fruits avec 3 composants:

- **A:** Jus d'orange;
- **B:** Jus de pomme;
- **C:** Jus de raisin;

Contrainte: A + B + C = 100%;

Variable réponse: Score de préférence (1-10).

Étapes Minitab:

Étape 1: Création du plan simplex

1. Stat > DOE > Mixture > Create Mixture Design;
2. Choisir "Simplex Centroid";
3. Number of components: 3;
4. Components: Définir les noms;
5. Designs: Vérifier les 7 points proposés.

Étape 2: Matrice d'expériences et résultats

Essai	X_1	X_2	X_3	Score
1	1	0	0	6.5
2	0	1	0	5.8
3	0	0	1	4.2
4	0.5	0.5	0	8.2
5	0.5	0	0.5	7.1
6	0	0.5	0.5	6.0
7	0.333	0.333	0.333	7.8

Étape 3: Analyse du modèle

1. Stat > DOE > Mixture > Analyze Mixture Design;
2. Response: Score;

3. Terms: Sélectionner le modèle quadratique;
4. Graphs: Courbes isoréponses et surface de réponse.

Résultats:

Coefficients (modèle quadratique):

Terme	Coef
A	**6.500**
B	**5.800**
C	**4.200**
A*B	**3.400**
A*C	**1.200**
B*C	**0.800**

R^2 = 99,2% R^2(adj) = 97,5%

<u>Étape 4: Interprétation et optimisation</u>

Graphique des courbes isoréponses:

```
      A (100%)
     /\
    /  \
   / 8,0 \
  /--------\
 /   7,5    \
B ----------- C (100%)
```

Optimisation:

- Maximum trouvé sur l'arête A-B;
- Solution optimale : A=60%, B=40%, C=0%;
- Score prédit : 8,4 ± 0,3.

Application 02: Alliage métallique avec contraintes

Énoncé:

Développement d'un alliage avec contraintes sur les composants:

- Cuivre: 40-60%;
- Zinc: 20-40%;
- Étain: 10-30%.

Contraintes:

- $0{,}4 \leq Cu \leq 0{,}6$;
- $0{,}2 \leq Zn \leq 0{,}4$;
- $0{,}1 \leq Sn \leq 0{,}3$;
- $Cu + Zn + Sn = 1$.

Variable réponse: Résistance à la traction (MPa).

Étapes Minitab:

Étape 1: Création du plan avec contraintes

1. Create Mixture Design;
2. Choisir "Extreme Vertices" (pour domaines contraints);
3. Number of components: 3;
4. Components: Définir les bornes;
5. Designs: Minitab génère les points extrêmes.

Étape 2: Points du plan et résultats

Essai	Cu	Zn	Sn	Résistance
1	0.4	0.4	0.2	420
2	0.4	0.3	0.3	380
3	0.4	0.2	0.4	350
4	0.5	0.3	0.2	450
5	0.5	0.2	0.3	410
6	0.6	0.2	0.2	480
7	0.6	0.3	0.1	460
8	0.5	0.25	0.25	435

<u>Étape 3: Analyse avec modèle quadratique</u>

Résultats:

Coefficients:

Cu 450,0***

Zn 320,0**

Sn 280,0*

Cu*Zn 85,0**

Cu*Sn -45,0

Zn*Sn -25,0

R^2 = 94,8% R^2(adj) = 89,6%

<u>Étape 4: Optimisation</u>

Solution optimale:

- Cuivre : 55%;
- Zinc : 35%;
- Étain : 10%.

Résistance prédite: 465 ± 15 MPa

Domaine expérimental:

```
       Cu (60%)
      /\
     /  \
    / 460\
   /------\
  /  420   \
Zn -------- Sn (30%)
 (40%)
```

Application 03: Formulation de détergent - Multi-réponses

Énoncé:

Optimisation d'un détergent avec 4 composants:

- Tensioactif: 30-50%;
- Solvant: 20-40%;
- Additifs: 10-30%;
- Eau: 10-20%.

Variables réponses:

- Y1: Pouvoir nettoyant (1-10);
- Y2: Viscosité (cP);
- Y3: Coût (€/kg);
- Y4: Stabilité (jours).

Étapes Minitab:

Étape 1: Plan avec contraintes multiples

1. Create Mixture Design avec 4 composants;
2. Components avec bornes:
   - Tensioactif : 0,3-0,5;
   - Solvant : 0,2-0,4;
   - Additifs : 0,1-0,3;
   - Eau: 0,1-0,2;
3. Choisir "Optimal design" avec 20 essais.

Étape 2 : Analyse pour chaque réponse

Modèle Pouvoir nettoyant: $\hat{Y} = 7{,}2*T + 5{,}8*S + 4{,}5*A + 3{,}2*E + 2{,}5TS + 1{,}8TA - 0{,}5SE$

Modèle Viscosité: $\hat{Y} = 450*T + 320*S + 280*A + 150*E + 85TS - 45TE$

Modèle Coût: $\hat{Y} = 2{,}8*T + 1{,}5*S + 3{,}2*A + 0{,}5*E$

Modèle Stabilité: $\hat{Y} = 45*T + 38*S + 52*A + 28*E + 12TA + 8SA$

<u>Étape 3: Optimisation multi-critères</u>

1. Stat > DOE > Mixture > Response Optimiser;
2. Définir les objectifs:
   - Nettoyant : Maximiser (cible 10);
   - Viscosité : Cible 400 ± 50;
   - Coût : Minimiser (max 3,0);
   - Stabilité : Maximiser (min 40 jours).
3. Importance relative: Nettoyant(3), Viscosité(2), Coût(2), Stabilité(1).

Solution optimale:

- Tensioactif: 45%;
- Solvant: 25%;
- Additifs: 20%;
- Eau: 10%.

Prédictions:

- Nettoyant: 8,4 ± 0,4;
- Viscosité: 395 ± 20 cP;
- Coût: 2,45 ± 0,15 €/kg;
- Stabilité: 46 ± 3 jours;
- Désirablité globale: 0,82.

Exercices supplémentaires

Exercice 1: Diagnostic de modèle de mélange

Sortie Minitab:

Analyse de variance pour Score.

Source	ddl	SC	MC	F	p
Régression	5	15.896	3.179	25.43	0.012
Linéaire	2	5.947	2.973	23.78	0.014
Quadratique	3	9.949	3.316	26.53	0.011
Erreur résid.	1	0.125	0.125		
Total	6	16.021			

$R^2 = 99,2\%$ $R^2(adj) = 97,5\%$

Questions:

1. Le modèle est-il significatif ?
2. Les termes quadratiques sont-ils nécessaires ?
3. La qualité d'ajustement est-elle bonne ?

Réponses:

1. **Oui** ($p = 0,012 < 0,05$)
2. **Oui** ($p = 0,011 < 0,05$) → interactions importantes
3. **Excellente** ($R^2 = 99,2\%$)

Exercice 2: Interprétation de coefficients

Modèle: $\hat{Y} = 6,5A + 5,8B + 4,2C + 3,4AB + 1,2AC + 0,8BC$

Questions:

1. Quel composant pur donne le meilleur score?
2. Quelle est la synergie la plus forte?
3. Que se passe-t-il si on mélange A et C?

Réponses:

1. **A** (6,5) donne le meilleur score pur;
2. **A-B** (3,4) a la synergie la plus forte;
3. Le mélange A-C améliore le score (1,2) mais moins que A-B.

Feuille de route pour les mélanges

Recommandations pratiques:

1. Bien définir les contraintes avant de créer le plan
2. Commencer par un modèle linéaire et complexifier si nécessaire
3. Utiliser les graphiques isoréponses pour la compréhension
4. Valider toujours avec des points de confirmation
5. Considérer les aspects économiques dans l'optimisation

Pièges à éviter:

- Oublier la contrainte de somme = 1;
- Négliger les interactions entre composants;
- Surinterpréter les coefficients purs;
- Choisir un modèle trop complexe pour peu de données.

1. **Box, G.E.P., Hunter, J.S., & Hunter, W.G. (2005).** *Statistics for Experimenters: Design, Innovation, and Discovery* (2nd ed.). Wiley-Interscience, **ISBN:** 978-0-471-71813-0.

2. **Montgomery, D.C. (2017).** *Design and Analysis of Experiments* (9th ed.). John Wiley & Sons, **ISBN:** 978-1-119-32093-7.

3. **Goupy, J., & Creighton, L. (2006).** *Introduction aux plans d'expériences* (3ème éd.). Dunod, **ISBN:** 978-2-10-049639-2.

4. **Wu, C.F.J., & Hamada, M.S. (2009).** *Experiments: Planning, Analysis, and Optimization* (2nd ed.). Wiley, **ISBN:** 978-0-470-39882-6.

5. **Daniel, C. (1976).** *Applications of Statistics to Industrial Experimentation*. Wiley, **ISBN:** 978-0-471-19270-4.

6. **Box, G.E.P., & Draper, N.R. (2007).** *Response Surfaces, Mixtures, and Ridge Analyses* (2nd ed.). Wiley, **ISBN:** 978-0-470-05357-7.

7. **Myers, R.H., Montgomery, D.C., & Anderson-Cook, C.M. (2016).** *Response Surface Methodology: Process and Product Optimization Using Designed Experiments* (4th ed.). Wiley, **ISBN:** 978-1-118-91601-8.

8. **Cornell, J.A. (2002).** *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data* (3rd ed.). Wiley, **ISBN:** 978-0-471-39367-2.

9. **Scheffé, H. (1963).** *The Simplex-Centroid Design for Experiments with Mixtures*. Journal of the Royal Statistical Society. **DOI:** 10.1111/j.2517-6161.1963.tb00506.x

10. **Fisher, R.A. (1926).** *The Arrangement of Field Experiments*. Journal of the Ministry of Agriculture of Great Britain.

15. **Box, G.E.P., & Wilson, K.B. (1951).** *On the Experimental Attainment of Optimum Conditions*. Journal of the Royal Statistical Society. **DOI:** 10.1111/j.2517-6161.1951.tb00067.x

16. **Box, G.E.P., & Behnken, D.W. (1960).** *Some New Three Level Designs for the Study of Quantitative Variables*. Technometrics. **DOI:** 10.1080/00401706.1960.10489912

17. **Plackett, R.L., & Burman, J.P. (1946).** *The Design of Optimum Multifactorial Experiments*. Biometrika. **DOI:** 10.1093/biomet/33.4.305

18. **Droesbeke, J.J., & Tassi, P. (1997).** *Les plans d'expériences: de l'expérimentation à l'assurance qualité*. Editions Technip. **ISBN:** 978-2-7108-0738-9

19. **Pillet, M. (2013).** *Les plans d'expériences par la méthode Taguchi* (4ème éd.). Editions d'Organisation. **ISBN:** 978-2-212-54902-2

20. **Saporta, G. (2011).** *Probabilités, Analyse des Données et Statistique* (3ème éd.). Editions Technip. **ISBN:** 978-2-7108-0986-4

Sites Web et Ressources en Ligne

21. **NIST/SEMATECH (2023).** *e-Handbook of Statistical Methods*.

- **URL:** https://www.itl.nist.gov/div898/handbook
- **Section:** "Handbook of Experimental Design"

22. **Statistique & Société (2023).** *Ressources pédagogiques sur les plans d'expériences*.

- **URL:** https://www.statistique-et-societe.fr

23. **Minitab Blog (2023).** *Design of Experiments Articles and Tutorials*.

- **URL:** https://blog.minitab.com/en/category/design-of-experiments