BTS - Initiation aux plans d'expériences - Cours

1. Introduction

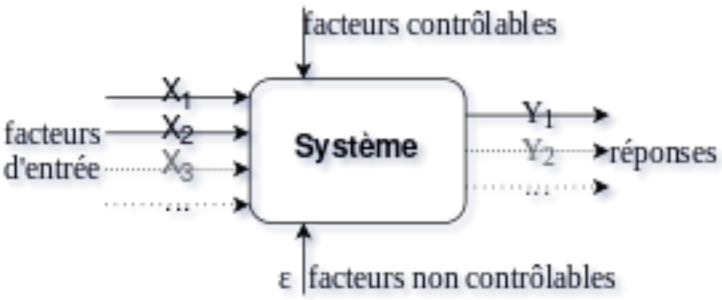
Définition 1 : *Les* ***plans d'expériences****, en anglais : DOE pour design of experiments, désignent un* ***modèle mathématique et une méthode*** *associée développés par Ronald A. Fisher dans les années 1920.*

Propriété 1 :

- **But :** obtenir un maximum de renseignements en un minimum d'essais expérimentaux.
- **Avantage :** les calculs réalisés sont facilement programmables.
- **Limite :** on n'obtient directement l'explication physico-chimique modélisé.

2. Le modèle mathématique

2.1. Relation entrées/sorties



Définition 2 : Notations :

- *les* ***facteurs d'entrée*** X_k *et les* ***réponses*** Y_l *sont des valeurs réelles (mesures) ;*
- *On cherche à modéliser* ***une des réponses*** *en fonction des entrées. On la note* Y.
- *On note* p *le nombre d'entrées :* $X_1, \dots, X_p$.
- *Le nombre de mesures effectuées (expérimentalement) pour chaque entrée* X_k *est noté* n_k*, pour* ***nombre de niveaux****.*
  Le nombre total d'expériences réalisables est alors $N = n_1 \times \cdots \times n_p$.
  Dans ce cours (et souvent), on utilise $n = 2$ *; on a alors* 2^p *expériences possibles : on parle de* ***plan complet*** 2^p*.*
- *On prend acte des facteurs contrôlables (fixés) dans l'effet global, et des* ***perturbations ε*** *(facteurs non contrôlables).*
- *Il faudra, selon les situations, parfois tenir compte de* ***interactions*** *entre les entrées.*

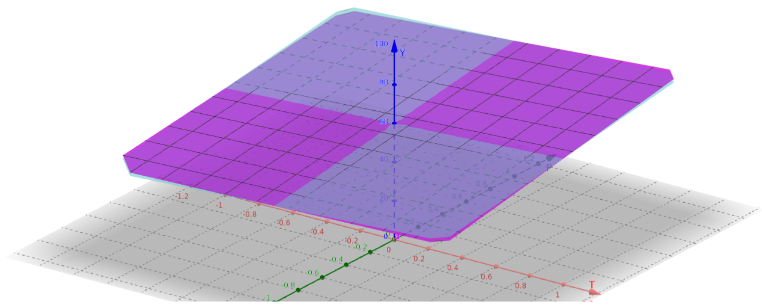
Modèle mathématique :

- *Pour un plan factoriel* 2^2*, sans interaction :* $Y = a_0 + a_1X_1 + a_2X_2 + \varepsilon$
- *Pour un plan factoriel* 2^2*, avec interaction :* $Y = a_0 + a_1X_1 + a_2X_2 + a_{1\,2}X_1X_2 + \varepsilon$
- *Pour un plan factoriel* 2^3*, sans interaction :* $Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + \varepsilon$
- *Pour un plan factoriel* 2^3*, avec interaction :*
  $Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + a_{1\,2}X_1X_2 + a_{1\,3}X_1X_3 + a_{2\,3}X_2X_3 + a_{1\,2\,3}X_1X_2X_3 + \varepsilon$

Notations du modèle mathématique :

- a_0 *est appelé* ***l'effet global*** *: il estime la* ***moyenne*** *des réponses, c'est la valeur centrale de Y.*
- a_k*, pour* $k \geqslant 1$ *est appelé* ***l'effet moyen de*** X_k
- $a_{k\,l}$ *est appelé* ***l'interaction*** *entre* X_k *et* X_l *; il estime l'influence de l'interaction entre les deux facteurs concernés.*
  Il en va de même pour 3 facteurs ou plus.
- ε *est appelé le* ***bruit*** *; c'est une variable aléatoire qui tient compte de tous les facteurs autres que ceux étudiés.*

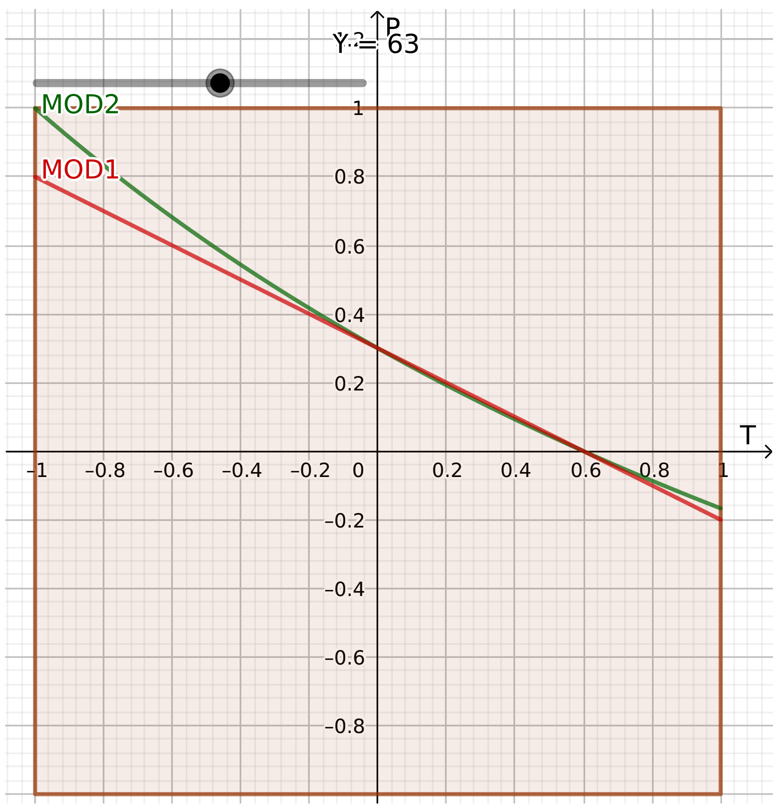
Exercice 1 :



Surface Y en fonction de T et P

Dans cet exercice, on analyse **deux modèles** décrivant le rendement Y (entre 0 et 100) d'une réaction chimique en fonction de deux facteurs d'entrée : la température $X_1 = T$ et la pression $X_2 = P$; ces deux entrées sont données, ici, en **unités arbitraires** (ua) : variant entre -1 et 1.

MOD1 : $Y = 60 + 5T + 10P + \varepsilon$ et **MOD2 :** $Y = 60 + 5T + 10P + 2TP + \varepsilon$



Courbes isoréponses T-P pour Y=63

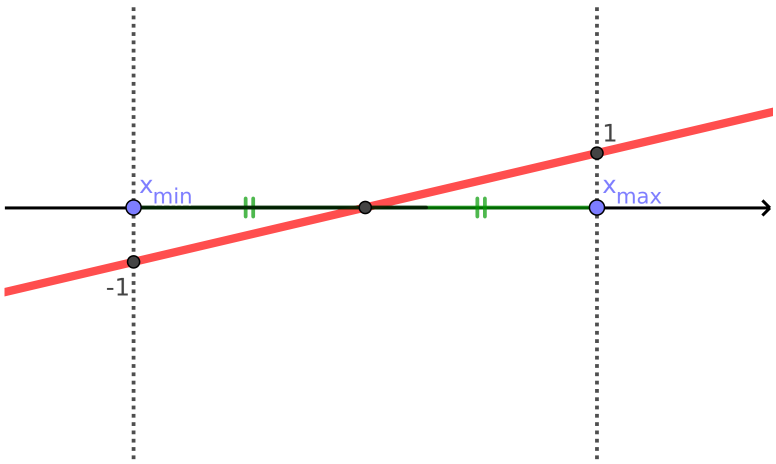
1. Identifier l'effet global et les effets moyens sur chaque modèle en précisant les valeurs.
2. Quel modèle fait apparaître une interaction entre les deux entrées ?
3. Donner les réponses, pour chaque modèle, lorsque $T = 0{,}5$ et $P = -0{,}2$.
4. Quelle entrée est la plus efficace pour améliorer le rendement ?
5. Exprimer P en fonction de T pour un rendement $Y = 70$; on néglige les perturbations.
6. **Généralisation :** exprimer P en fonction de T pour un rendement Y indéterminé.

Définition 3 : *Une telle courbe s'appelle* ***courbe isoréponse*** *(cf figure ci-dessus).*

Remarque 1 : Les modèles étudiés correspondent à des développements limités, au voisinage de 0, d'ordre 1 pour ceux sans interaction. Si on tient compte d'interactions, on rajoute des termes (d'ordre 2 pour X_kX_l et 3 pour $X_kX_lX_m$) à ce développement en 0. D'où l'idée de coder les valeurs des entrées par l'intervalle [-1;1] :

2.2. Codage affine des entrées (de Yates)

Définition 4 :



Codage de Yates

- *La valeur donnée à un facteur d'entrée pour réaliser un essai est appelée* ***niveau****.*
- *En pratique, le* ***domaine du facteur*** *(ensemble des valeurs du facteur d'entrée que l'on peut, ou veut, tester) est borné, de la forme* $[x_{\min};x_{\max}]$.
  Notant cela, Frack Yates (1902-1993) choisit de ***ramener chaque domaine dans l'intervalle [-1;1]*** *en appliquant une fonction affine (codage de Yates) : ramener ces valeurs dans un même intervalle permet de mieux les comparer (quelle entrée a le plus d'influence, etc).*
- *Le niveau -1 est appelé «****niveau bas****» et le 1 «****niveau haut****».*
  En pratique, on se limite souvent aux deux niveaux, le bas et le haut, pour tester chaque entrée.

Exercice 2 : Déterminer le codage de Yates, de la forme $X = ax + b$ pour les grandeurs suivantes :

1. P pour une pression p comprise entre 1 et 2 bar ;
2. T pour une température θ comprise entre 60°C et 80°C ;
3. L pour une longueur l comprise entre 1 micron et 10 microns ;
4. S pour un coefficient d'agrandissement/réduction s compris entre -20% et 50% ;
5. **Cas général :**
   X pour une valeur x comprise entre $x_{\min}$ et $x_{\max}$;

Propriété 2 : Codage de Yates $X \ : \quad [x_{\min};x_{\max}] \to [-1;1]$ $\boxed{x \mapsto X = \frac{2}{x_{\max}-x_{\min}}\left(x - \frac{x_{\min}+x_{\max}}{2}\right)}$

Remarque 2 : Réciproque (x en fonction de X) : on a $x = \frac{x_{\max}-x_{\min}}{2}X + \frac{x_{\min}+x_{\max}}{2}$.

3. Utilisation des plans d'expériences

3.1. Matrice des effets

Définition 5 : *Une fois les domaines des facteurs d'entrée choisis (niveaux bas et haut), on résume les résultats des expériences dans un tableau appelé* ***matrice des effets*** *ou bien* ***matrice d'expérience****. Ce tableau permet de déterminer les coefficients du modèle.*

Propriété 3 :

- L'effet global étant toujours présent, il est toujours au niveau 1.
- On a besoin d'au moins autant d'expérience que de coefficients à calculer.
- Mathématiquement, on privilégie le fait d'utiliser une **matrice de Hadamard** (le produit scalaire de deux colonnes est nul), qui minimise les écart-types sur les coefficients calculés (vis à vis des perturbations), comme dans l'exemple qui suit.
- ***Définition 6 :*** *La construction de cette matrice suit l'**algorithme de Yates** :*
  - *La colonne de l'effet global ne contient que des 1 ;*
  - *La colonne du premier effet contient -1, 1 en alternance*
  - *La colonne du deuxième effet contient -1, -1, +1, +1 en alternance, celle du troisième effet contient -1, -1, - 1, -1, +1, +1, +1, +1 en alternance, ...*
  - *Les colonnes contenant les interactions correspondent au produit de leurs effets.*

3.2. Plan factoriel 2^2 avec interaction

On étudie le rendement d'une réaction chimique en fonction de la température et de la pression.

N° essai	**Global**	**Pression** $P = X_1$	**Température** $T = X_2$	**Interaction T-P** X_1X_2	**Rendement** Y
1	+1	-1	-1	+1	60
2	+1	+1	-1	-1	78
3	+1	-1	+1	-1	63
4	+1	+1	+1	+1	89
Coefficients	a_0	a_1	a_2	$a_{1\,2}$	
niveau -1		2 bars	50°C		
niveau +1		4 bars	70°C		

Propriété 4 : En remplaçant les valeurs des entrées X_1 et X_2 pour chaque ligne (expérience) dans le modèle utilisé ici : $a_0 + a_1X_1 + a_2X_2 + a_{1\,2}X_1X_2 = Y$ (on néglige les perturbations ε) ; on voit que cette matrice équivaut au système $\begin{cases} a_0 - a_1 - a_2 + a_{1\,2} = 60 \\ a_0 + a_1 - a_2 - a_{1\,2} = 78 \\ a_0 - a_1 + a_2 - a_{1\,2} = 63 \\ a_0 + a_1 + a_2 + a_{1\,2} = 89 \end{cases}$

Méthode 1 : La résolution du système est simplifiée (par la construction choisie par Yates) :

- En additionnant toutes les lignes, on obtient : $4a_0 = 60 + 78 + 63 + 89 = 290$, donc $a_0 = 72{,}5$.
- Pour obtenir a_1, on multiplie les ligne 1 et 3 par -1, on a donc (attention à ne pas oublier de multiplier les valeurs de Y de ces lignes par -1 aussi) : $\begin{cases} -a_0 + a_1 + a_2 - a_{1\,2} = -60 \\ a_0 + a_1 - a_2 - a_{1\,2} = 78 \\ -a_0 + a_1 - a_2 + a_{1\,2} = -63 \\ a_0 + a_1 + a_2 + a_{1\,2} = 89 \end{cases}$. Ensuite, on additionne toutes les lignes ; on obtient : $4a_1 = -60 + 78 - 63 + 89 = 44$, d'où $a_1 = 11$.
- On procède de même pour a_2 (lignes 1 et 2 à multiplier par -1) et $a_{1\,2}$ (lignes 2 et 3).

Exercice 3 : Calculer les deux coefficients restants et écrire le modèle polynomial obtenu. Quel facteur d'entrée domine ? Le terme d'interaction est-il négligeable ?
Quels sont les résultats obtenus si l'on refait l'exercice en ne tenant pas compte du terme d'interaction dans les expériences et le modèle ?

Exercice 4 : On considère le plan d'expérience visant à étudier les conditions de fonctionnement d'un pistolet à peinture servant à appliquer une couche de vernis sur des objets. Deux facteurs sont étudiés : l'ouverture du pistolet (facteur A) variant de 1 cran à 3 crans et la pression (facteur B) variant de 1 bar à 2 bars. La réponse est la couleur obtenue allant de 0 (noir) à 60 (jaune).
Les expériences réalisées ont, dans l'ordre de l'algorithme de Yates, donné les résultats des rendements de 15, 20, 25 et 40. Le plan choisi est un plan factoriel complet 2^2 avec interaction.

1. Écrire la matrice des effets construite selon l'ordre de l'algorithme de Yates.
2. Donner les estimations ponctuelles des effets (ie calculer les coefficients) et écrire l'expression du modèle Y.
3. Vrai ou faux (justifier) ? «Pour une ouverture de 2,5 crans et une pression de 1,25 bar, on a une couleur attendue égale à Y = 23,1»
4. On ne souhaite pas descendre en dessous d'une couleur correspondant à Y=22. Donner une expression de la courbe isoréponse correspondante.

3.3. Plan factoriel 2^3 sans interaction

Remarque 3 : On utilise la même matrice que pour le plan factoriel 2^2 avec interaction précédent : cela permet de minimiser les écarts lors des calculs.

N° essai	Global	Pression $P = X_1$	Température $T = X_2$	Masse catalyseur $M = X_3$	Rendement Y
1	1	-1	-1	1	…
2	1	1	-1	-1	…
3	1	-1	1	-1	…
4	1	1	1	1	…
Coefficients	a_0	a_1	a_2	a_3	
niveau -1		2 bars	50°C	50 g	
niveau +1		4 bars	70°C	250 g	

3.4. Plan factoriel 2^3 avec interaction

Exercice 5 :

L'expérience consiste à déterminer en jars test les meilleures conditions de traitement d'eau polluée par des ions Cu^{2+} et Zn^{2+} après précipitation, floculation (agglomération de particules) et décantation. La faisabilité industrielle du procédé est estimée notamment à partir de la durée de décantation qui doit être la plus faible possible. Cette durée constitue la réponse.
Trois facteurs sont étudiés. La nature de l'hydroxyde (facteur A) est un facteur qualitatif. Sera utilisé de la chaux (-1) ou de la soude (+1). L'excès stœchiométrique en hydroxyde (facteur B) varie entre (×2) et (×4) et le pourcentage en volume de floculant (facteur C) varie de 2% à 10%.
Donner les estimations ponctuelles des effets (ie calculer les coefficients) et écrire l'expression du modèle Y.

Essai	**glob**	X_1	X_2	X_3	X_1X_2	X_1X_3	X_2X_3	$X_1X_2X_3$	Y
1	+1	-1	-1	-1	+1	+1	+1	-1	27
2	+1	+1	-1	-1	-1	-1	+1	+1	19,5
3	+1	-1	+1	-1	-1	+1	-1	+1	43,5
4	+1	+1	+1	-1	+1	-1	-1	-1	21,5
5	+1	-1	-1	+1	+1	-1	-1	+1	20,5
6	+1	+1	-1	+1	-1	+1	-1	-1	16,5
7	+1	-1	+1	+1	-1	-1	+1	-1	30
8	+1	+1	+1	+1	+1	+1	+1	+1	12,5
Niveau -1	-1	chaux	×2	2 %					
Niveau +1	1	soude	×4	10 %					
Coefficient	a_0	a_1	a_2	a_3	$a_{1\,2}$	$a_{1\,3}$	$a_{2\,3}$	$a_{1\,2\,3}$	