# Workshop


# Plans d'expériences

15 - 17 Septembre 2009

Hostellerie La Magnaneraie
VILLENEUVE LÈS AVIGNON

Michelle SERGENT

Laboratoire de Méthodologie de la Recherche Expérimentale
Université Paul Cézanne Aix Marseille III

# Session 1 : Le screening de facteurs

# Elaborer une stratégie expérimentale

choisir, pour chaque problématique, le meilleur plan d'expériences

## Les objectifs

- **Criblage de facteurs**
- Etude quantitative de facteurs
- Optimisation

# Criblage de facteurs

➡ Stratégie permettant d'identifier rapidement les quelques facteurs réellement influents (h) parmi un grand nombre de facteurs (k) potentiellement influents

$$h << k$$

L'hypothèse du *"principe de parcimonie"* (ou *"rasoir d'Ockham"*) doit être respectée :

En présence d'un très grand nombre de facteurs, il est raisonnable de penser que quelques-uns seulement seront influents.

*Le nombre de facteurs actifs est très faible*

# Etat de l'art :

## Méthodes de criblage pour des facteurs à 2 niveaux

1) Matrices d'Hadamard

2) Group screening

3) Bifurcation séquentielle

4) Matrices supersaturées

## Matrices d'HADAMARD $2^k//N$ :

Une **matrice d'Hadamard** est une matrice carrée dont les éléments sont +1 ou -1 et dont les lignes sont toutes orthogonales entre elles et pour laquelle la matrice d'information $X'X$ est telle que :

$$X'X^* = N\,\mathbf{I}_N$$

avec, $\mathbf{I}_N$ : matrice identité d'ordre N.

*k facteurs à 2 niveaux*

$N \geq k + 1$

N ≡ multiple de 4

$N = 2^r$ *(matrice géométrique)*

$N \neq 2^r$ *(matrice non géométrique)*

# Les méthodes :

- **Matrices d'HADAMARD $2^k//N$ :**

Modèle postulé : $\eta = X\beta$

Modèle vrai : $\eta = X\beta + X_1\beta_1$

$B = (X'X)^{-1} X'Y$ $\qquad$ $E[B] = ?$

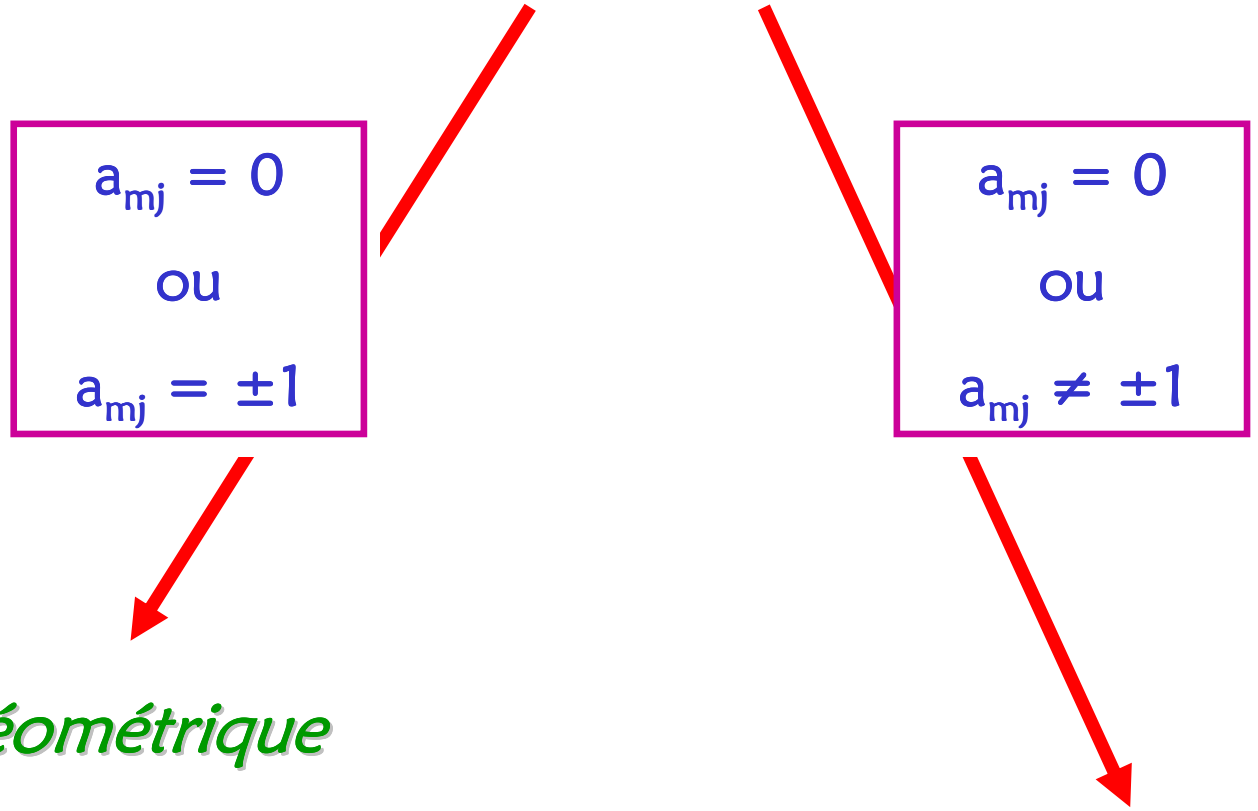
$E[B] = \beta + A\,\beta_1$

$$E[b_0] = \beta_0$$

$$E[b_i] = \beta_i \pm \alpha_{m,j}\,\beta_{m,j} \quad \ldots \quad m, j \neq i$$

# Matrices d'Hadamard

## *(de résolution III)*




# Matrices d'Hadamard géométriques

Modèle postulé : $\eta = X\beta$

Modèle vrai : $\eta = X\beta + X_1\beta_1$

$\Rightarrow$ $B = (X'X)^{-1} X'Y$ $\qquad$ $E[B] = ?$

$E[B] = \beta + A\,\beta_1$

$2^3//4$

$$E[b_0] = \beta_0$$
$$E[b_1] = \beta_1 + \beta_{23}$$
$$E[b_2] = \beta_2 + \beta_{13}$$
$$E[b_3] = \beta_3 + \beta_{12}$$

$2^7//8$

$$E[b_0] = \beta_0$$
$$E[b_1] = \beta_1 - \beta_{34} - \beta_{26} - \beta_{57}$$
$$\cdots\cdots\cdots\cdots$$
$$E[b_7] = \beta_7 - \beta_{23} - \beta_{15} - \beta_{46}$$

# Matrices d'Hadamard non géométriques

**$2^{11}//12$**

$$E[b_0] = \beta_0$$

$$E[b_1] = \beta_1 \pm \mathbf{0.33}\ \beta_{m,j} \quad \ldots \quad m, j \neq 1$$

$$E[b_{11}] = \beta_{11} \pm \mathbf{0.33}\ \beta_{m,j} \quad \ldots \quad m, j \neq 11$$

**$2^{82}//84$**

$$E[b_0] = \beta_0$$

$$E[b_i] = \beta_i \pm \mathbf{a}_{m,j}\ \beta_{m,j} \quad \ldots \quad m, j \neq i$$

$$a_{m,j} = \mathbf{0.04; 0.10; 0.20; 0.28}$$

# Les méthodes :

- **Matrices d'HADAMARD de résolution IV :**

## *La résolution :*

*Une matrice d'expériences est de résolution L si aucun effet m-indicé n'est confondu avec un ou plusieurs effets comprenant (L-m) indices.*

$R_L$ — *Chiffre romain*

*Matrices d'expériences d'Hadamard* ⇒ $R_{III}$

# Les méthodes :

## Matrices d'HADAMARD de résolution IV :

"Fold over" (= repliement)

$$\begin{pmatrix} \mathbf{H} \\ -\mathbf{H} \end{pmatrix}$$

Avec la matrice **H** :

$$E[\,l_0\,] = \beta_0$$
$$E[\,l_i\,] = \beta_i \pm a_{m,j}\,\beta_{m,j} \quad \ldots \quad m, j \neq i$$

Avec la matrice **- H** :

$$E[\,l'_0\,] = \beta_0$$
$$E[\,l'_i\,] = \beta_i \pm a_{m,j}\,\beta_{m,j} \quad \ldots \quad m, j \neq i$$

$$E[(l_0+l'_0)/2\,] = \beta_0$$
$$E[(l_i+l'_i)/2] = \beta_i$$
$$E[(l_i-l'_i)/2] = \pm a_{m,j}\,\beta_{m,j} \qquad m, j \neq i$$

⇒ Matrices $R_{IV}$

# La probabilité qu'un facteur soit actif est faible

k facteurs à 2 niveaux ( $N \leq 100$)

- Matrices d'expériences d'Hadamard $\mathbf{R_{III}}$

  $N \geq k + 1$

  $N \equiv 0$ (modulo 4)

- Matrices d'expériences d'Hadamard $\mathbf{R_{IV}}$

  $N \geq 2\,(k + 1)$

  $N \equiv 0$ (modulo 4)

# Les méthodes :

*La probabilité qu'un facteur soit actif est TRES faible*

Identifier rapidement les quelques facteurs réellement influents

- Criblage par groupes
- Groupes Multiples
- Matrices Supersaturées
- Bifurcation séquentielle

## Criblage par groupes :

- Les k facteurs sont partitionnés en plusieurs groupes de taille non égale,
- A la 1ère étape, chaque groupe est traité comme un "facteur" (facteur groupé), que l'on teste, et les groupes "non actifs" sont éliminés.
- Les facteurs des "groupes actifs" seront traités ensuite :
  - individuellement (processus à 2 étapes)
  - en les divisant à nouveau en groupes de taille plus petite (processus à plusieurs étapes)

# Hypothèses :

- Il n'y a pas d'effet d'interaction entre les facteurs

- Chacun des facteurs a 2 niveaux et on connaît la direction des éventuels effets :

$$\beta_j \geq 0$$

- Un groupe est actif, si un ou plusieurs facteurs de ce groupe est actif.

- Si un groupe est sans influence
  $\Rightarrow$ tous les facteurs de ce groupe sont éliminés.

***Remarque** : pour mettre en oeuvre les techniques de criblage par groupes, quelles qu'elles soient, il est nécessaire de connaître le sens de variation de la réponse en fonction du sens de variation du facteur et il est "préférable" d'avoir une idée a priori de l'impact de chaque facteur sur la réponse (et/ou de sa probabilité) .*

# Les méthodes :

## Criblage par groupes multiples :

- Chaque facteur est affecté à plusieurs groupes à la 1ère étape,
- Un facteur est potentiellement influent si chaque groupe auquel le facteur appartient est actif.
- Les facteurs pour lesquels tous les groupes associés sont "actifs" sont étudiés dans une 2ème étape.

# Les méthodes :

## Bifurcation séquentielle :

≅ Approche "dichotomique"

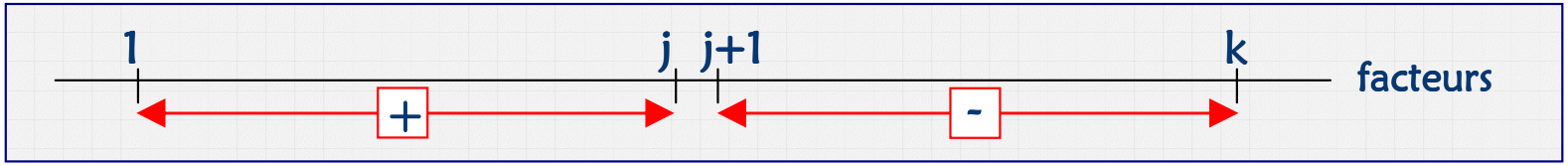
- y[k] : réponse quand tous les facteurs [1, …, k] sont tous au niveau (+)
- y[0] : réponse quand tous les facteurs [1, …, k] sont tous au niveau (-)

- Si y[0] = y[k] ➡ pas de facteur influent !
- Sinon, certains facteurs sont influents.

*Lesquels ?*

- Sinon, certains facteurs sont influents. *Lesquels ?*

- y[j] : réponse expérimentale quand :
  - les facteurs [1, …, j] sont tous à leur niveau (+)
  - les autres facteurs : [j+1, …, k] étant à leur niveau (-).




- y [k/2] : k/2 facteurs (+) et k/2 facteurs (-)

comparaison de y[0] et y [k/2]

y[k] et y [k/2]

et on bifurque …..

# Bifurcation séquentielle :

Méthode très efficace !

Cette méthode permet de cribler un très grand nombre de facteurs avec moins de simulations que de facteurs.

## Hypothèses :

➡ Seulement quelques facteurs sont réellement influents

"parsimony principle" ou " effect sparsity "

➡ On connaît la direction des éventuels effets, c.a.d le signe des coefficients $\beta_i$ :

$$\beta_j \geq 0$$

# Les méthodes :

## Matrices supersaturées :

Le nombre d'expériences N est inférieur au nombre d'information demandées NI,

$$NI = 1 + \sum (s_j - 1)$$

*$s_j$ : nombre de niveaux du facteur j*

Les matrices d'expériences supersaturées les plus utilisées, *actuellement*, sont celles dont tous les facteurs ont 2 niveaux :

$$N < NI = 1 + 2k$$

# Les méthodes :

- **Matrices supersaturées :**

## Méthodes de construction :

| | |
|---|---|
| Lin (1991, 1995) | algorithme |
| Lin (1993) | matrices d'Hadamard non géométriques |
| Wu (1993) | matrices d'Hadamard non géométriques |
| Deng, Lin et Wang (1994) | matrices d'Hadamard |
| Liu et Zhang | |
| Nguyen (1996) | BBID |
| Yamada et Lin (1997) | bases orthogonales |
| Cheng (1997) | algorithme |
| Cela (1998) | algorithme génétique |

# Les méthodes :

## Matrices supersaturées :

### Méthodes de traitement :

- Régression Stepwise
- Ridge régression
- Régression PLS
- $R^2$ après calcul de toutes les régressions
- Algorithme génétique
- Approche bayésienne