



Construction et randomisation de plans factoriels réguliers avec le package R PLANOR

Hervé Monod, Annie Bouvier, André Kobilinsky

INRA, Unité MIA-Jouy, France

Rencontres R 2012 Bordeaux

2 juillet 2012





Contexte statistique

- **Plans d'expériences** \Rightarrow étape préalable au recueil de données
- **Analyse prévue** : anova, modèle linéaire et extensions
- **Données du problème** :
 - **s facteurs contrôlés** à m_i modalités, $i = 1, \dots, s$
 - modèle d'analyse et effets à estimer spécifiés
 - taille N du plan fixée
- **Approche** :
 - plan = sélection de N combinaisons (x_1, \dots, x_s)
 - **construction algébrique** (facteurs contrôlés)
 - **randomisation** (autres facteurs)
 - \rightarrow plans factoriels orthogonaux dits réguliers



Bibliographie

Plans factoriels réguliers

- Yates (1933, 1937), Fisher (1942) : plans factoriels avec confusion d'effets
- Patterson (1976), Bailey (1977), Franklin (1977) : **design key**
- Kobilinsky (1991), Kobilinsky et Monod (1991), Pistone and Rogantin (2008) : mixed fractional factorial designs

Applications aujourd'hui

- *Biometrics, 2011* : Split-Plot Designs for Robotic Serial Dilution Assays
- *Genetics, 2010* : Statistical design and analysis of RNA sequencing data
- *Biometrika, 2011* : Nested orthogonal array-based Latin hypercube designs
- *2011 : Annals of Statistics, JASA, JSPI, Technometrics, etc*



Plan

1 Exemples de plans factoriels

- Plan en blocs complets
- Carré latin
- Plan factoriel fractionnaire

2 Aspects théoriques

3 Implémentation

4 Discussion



Plan

1 Exemples de plans factoriels

- Plan en blocs complets
- Carré latin
- Plan factoriel fractionnaire

2 Aspects théoriques

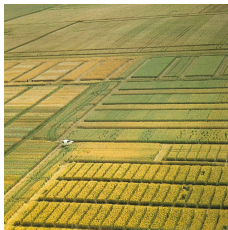
3 Implémentation

4 Discussion



Plan en blocs complets randomisés

Objectif : comparer 6 variétés en utilisant 4 blocs de 6 parcelles
 Les variétés sont notées A, B, C, D, E, F.



E	C	A	F	D	B
B	E	D	C	F	A
D	E	C	B	F	A
B	D	F	C	E	A



Sous R

Script :

```
> library(planor)
> set.seed(123)
> pbc.key <- planor.designkey(
  factors = list(bloc=1:4,
                 variety=LETTERS[1:6]),
  model = ~bloc+variety,
  nunits = 24)
[...]
```

```
> pbc.plan <- planor.design(pbc.key,
  randomize=~bloc/UNITS)
[...]
```



Sous R

```
> pbc.plan@design
      UNITS bloc variety
1         1     1       E
2         2     1       C
3         3     1       A
.
.
.
24        24     4       A
```



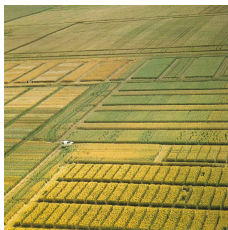

Plan en carré latin

Objectif :

comparer 3 variétés et 2 doses d'engrais en 6 lignes \times 6 colonnes

Les variétés sont notées A, B, C.

Les doses d'engrais sont notés 1 et 2.



B2	A1	C1	C2	B1	A2
C2	B1	A1	A2	C1	B2
A1	C2	B2	B1	A2	C1
C1	B2	A2	A1	C2	B1
A2	C1	B1	B2	A1	C2
B1	A2	C2	C1	B2	A1



Sous R

Script :

```
> set.seed(123)
> latin.key <- planor.designkey(
  factors = list(ligne=1:6,
                 colonne=1:6),
                 variete=LETTERS[1:3]),
  dose=1:2),
  model = ~ligne+colonne+variete*dose,
  nunits = 36)
[...]
> latin.plan <- planor.design(latin.key,
  randomize=~ligne+colonne)
```





Plan factoriel fractionnaire

Objectif : analyse de sensibilité d'un modèle d'épidémiologie animale

(Courcoul, Monod *et al.*, 2011, J. Theor. Biol.)

Plan d'expérience numérique :

facteurs = paramètres du modèle épidémio

- 12 facteurs à 4 niveaux
- 7 facteurs à 2 niveaux
- Modèle : effets principaux + interactions entre 2 facteurs

$\implies 4^{12} \times 2^7 = 2^{31}$ combinaisons possibles

Contrainte : $\leq 2^{12} = 4096$ simulations



Sous R

Script :

```
> set.seed(123)
> frac.key <- planor.designkey(
  factors = LETTERS[1:19],
  nlevels = c(rep(4,12),rep(2,7)),
  model = ~(A+[...]+S)^2,
  nunits = 4096)
[...]
> frac.plan <- planor.design(frac.key)
```



Solution :

```
> frac.plan@design
      A B C D E F G H I J K L M N O P Q R S
1     1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2     1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1
.     .
.     .
.     .
4096 4 4 4 4 1 4 2 2 2 2 4 3 2 1 1 2 1 1 1
>
```

Propriété : Plan complet pour tous les sous-ensembles de 4 facteurs
 ⇒ **Tableau orthogonal** de force 4 (\equiv résolution 5)



Plan

1 Exemples de plans factoriels

- Plan en blocs complets
- Carré latin
- Plan factoriel fractionnaire

2 Aspects théoriques

3 Implémentation

4 Discussion



Principes

- facteur $F_i \equiv$ groupe abélien T_i
- unités \equiv groupe abélien U
- plan : morphisme de U dans $T = T_1 \times \dots \times T_s$
- matrice clé = matrice Φ du morphisme
- matrice clé $\Phi^* =$ duale de Φ
 \Rightarrow les confusions d'effets sont données par $\ker(\Phi)^*$
- Problème : trouver Φ^* adaptée au modèle



Matrice clé sur un cas simple

4 facteurs à 2 modalités \Rightarrow sélectionner 8 traitements parmi $2^4 = 16$

<i>Unit</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D = ABC</i>
000	0	0	0	0
001	0	0	1	1
010	0	1	0	1
011	0	1	1	0
100	1	0	0	1
101	1	0	1	0
110	1	1	0	0
111	1	1	1	1

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

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



Conséquences sur la matrice X du modèle linéaire

$$X = \begin{pmatrix} 1 & A & B & C & D & AB & AC & AD & BC & BD & CD \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Effets factoriels *confondus* (*aliased*) 2 par 2



Conséquences : aliasing

$$1 = ABCD$$

$$A = BCD$$

$$B = ACD$$

$$C = ABD$$

$$D = ABC$$

$$AB = CD$$

$$AC = BD$$

$$AD = BC$$

interactions d'ordre ≥ 3 nulles \Rightarrow effets principaux estimables



Conséquences pour la modélisation

Avec le modèle :

$$\begin{aligned} Y = & 1 + A + B + C + D \\ & + A.B + A.C + A.D + B.C + B.D + C.D \\ & + \varepsilon \end{aligned}$$

Tous les effets principaux sont estimables



Plan

1 Exemples de plans factoriels

- Plan en blocs complets
- Carré latin
- Plan factoriel fractionnaire

2 Aspects théoriques

3 Implémentation

4 Discussion



Spécificités R

Entrées :

- 1 facteurs, niveaux, relations de hiérarchie
- 2 modèle(s) et liste(s) de termes à estimer
- 3 taille du plan

Sorties :

- 1 matrices clés, effets confondus, plan (dataframe)

Programmation :

- codage en R et C++
- utilisation des packages biganalytics et bigmemory
- classes S4 emboîtées (matrice clé, listes de matr. clés, plan, etc.)
- méthodes : analyse de la matrice clé, impression





Grandes étapes de l'algorithme

Etapes :

- 1 décomposition des facteurs selon les nombres premiers p impliqués
- 2 détermination des effets ne devant pas appartenir à $\ker(\Phi_p^*)$ à partir du modèle
- 3 recherche backtrack colonne par colonne de Φ_p^* , telle que $\text{Ker}(\Phi^*) \cap I = \emptyset$
- 4 analyse de la (des) solutions
- 5 construction du plan et randomisation



Plan

- 1 Exemples de plans factoriels
 - Plan en blocs complets
 - Carré latin
 - Plan factoriel fractionnaire
- 2 Aspects théoriques
- 3 Implémentation
- 4 Discussion



Plans factoriels réguliers : existant

Des logiciels tels que

- SAS : proc factex, + Kuhfeld and Tobias (2005) *Technometrics*
- R library FrF2 (2010)

Atouts de PLANOR

- nombres de modalités des facteurs libres
- modèles asymétriques autorisés
- hiérarchies entre facteurs prises en compte
- randomisation pour des structures très générales
- plans multistrates (e.g. split-plot)