

PLANS D'EXPERIENCES EN BLOCS

Application aux expériences agronomiques

INTRODUCTION

Exemple : essai variétal

Répartition des variétés dans l'essai

Bloc 1	Bloc 2	Bloc 3	Bloc 4
12 24 17 21 22	6 14 24 27 30	11 10 12 1 21	7 19 2 5 3
7 3 20 4 9	4 1 17 11 13	2 9 13 20 22	4 28 29 6 14
11 26 25 18 13	15 3 22 12 25	19 17 14 23 25	1 16 22 11 26
16 23 6 30 15	2 18 20 16 7	4 16 5 24 15	9 12 18 23 30
8 28 27 10 5	10 8 5 29 9	26 28 18 3 7	20 17 24 21 13
1 2 14 19 29	19 26 28 21 23	29 27 8 6 30	27 10 8 25 15

Essai variétal réalisé par le GEVES en 1993 :

- 30 variétés de tournesol à comparer,
- 6 lignes de 20 parcelles : 120 parcelles,
- chaque variété est répétée 4 fois,
- la variable principale mesurée sur les parcelles est le rendement en grains (q/ha).

Définition de certaines notions à l'aide de deux exemples. Le premier exemple est un essai variétal. On cultive des variétés de tournesol sur des parcelles puis on mesure le rendement sur chaque parcelle. L'objectif de l'essai est de comparer les variétés et de sélectionner les meilleures. Les numéros indiqués sur le schéma du dispositif représentent les variétés cultivées sur les parcelles.

- Une **variable réponse** est une variable que l'on cherche à étudier. Dans l'essai, la variable réponse principale est le rendement.
- Un **facteur** est une variable qui peut influencer les variables réponses.
- Les **niveaux** (ou **modalités**) d'un facteur sont les valeurs prises par ce facteur dans l'expérience.
- Un **facteur traitement** est un facteur dont on cherche à étudier l'influence sur les variables réponses.
- Dans l'essai, il y a un facteur traitement, le facteur variété, qui a 30 niveaux.
- Les **unités expérimentales** sont les éléments sur lesquels on effectue les observations.
- Dans l'essai, les unités expérimentales sont les parcelles.
- Un **bloc** est un groupe d'unités expérimentales considérées homogènes.
- Un **facteur bloc** est un facteur dont les niveaux sont des blocs.
- Dans l'essai, il y a un facteur bloc qui a 4 niveaux.

Exemple : comparaison de populations de tournesol

On croise 3 populations de tournesol d'origines géographiques différentes avec 2 testeurs, pour connaître la valeur en croisement des populations. La teneur en huile est mesurée sur les hybrides issus des croisements.

Il y a 2 facteurs traitements :

- origine, à 3 niveaux,
- testeur, à 2 niveaux.

Les **traitements** sont les combinaisons des niveaux des facteurs traitements.

Traitement	Niveau de testeur	Niveau de origine
1	1	1
2	1	2
3	1	3
4	2	1
5	2	2
6	2	3

Remarque

Cet exemple vient du module FPSTAT 2/3 sur le modèle linéaire.

Expérience

Dans une expérience :

- on étudie l'influence d'un ou plusieurs facteurs traitements sur une ou plusieurs variables réponses, expérimentales, puis on mesure les variables réponses sur chaque unité expérimentale,
- on applique les traitements sur les unités expérimentales, alors que ce n'est pas le cas en *échantillonnage*.

Planification

Planifier une expérience consiste à définir le protocole de l'expérience.

Choisir le **plan d'expériences** consiste à choisir :

- les traitements,
- les unités expérimentales,
- la répartition des traitements entre les unités expérimentales.

- Pourquoi prendre en compte des blocs dans une expérience ?
- Quels facteurs blocs utiliser pour avoir les résultats les plus précis ?
- Une fois les facteurs blocs choisis, comment répartir les traitements entre les blocs ?
- Comment répartir les traitements entre les blocs lorsqu'il y a plusieurs facteurs traitements ?
- Comment analyser les données une fois l'expérience réalisée ?

Contenu du module

Dans ce module :

- on considère des expériences n'ayant que des facteurs qualitatifs,
- on présente comment répartir les traitements sur les unités expérimentales, à traitements et unités expérimentales fixes,
- on insiste sur la prise en compte de l'hétérogénéité entre unités par l'utilisation de facteurs blocs.

Plan du module

- Dispositifs en blocs à un facteur traitement expérimentales par des facteurs blocs
- Analyse d'un dispositif en blocs avec le modèle à effets blocs fixes
- Construction de dispositifs en blocs
- Analyse d'un dispositif en blocs avec le modèle à effets blocs aléatoires
- Plans factoriels en blocs

Ce module n'aborde pas les aspects liés au choix des traitements et des unités expérimentales (par exemple le choix de la surface des parcelles dans une expérience agronomique), sauf l'influence du nombre de répétitions sur la précision de l'analyse.

La partie I porte sur la prise en compte de facteurs blocs lorsqu'il n'y a qu'un facteur traitement. La partie II porte sur la prise en compte de facteurs blocs lorsqu'il y a plusieurs facteurs traitements. L'analyse d'un dispositif en blocs avec le modèle à effets blocs fixes (B) est présentée avant la planification (C), car pour planifier une expérience, il faut savoir comment l'analyser.

I. DISPOSITIFS EN BLOCS À UN FACTEUR TRAITEMENT

A. DESCRIPTION DE L'HÉTÉROGÉNÉITÉ DES UNITÉS EXPÉRIMENTALES PAR DES FACTEURS BLOCS

Plan

- Pourquoi prendre en compte des blocs ?
- Plans en blocs
- Plans en lignes-colonnes
- Plans en répliques
- Carrés semi-latins
- Conclusion

La partie A présente les systèmes de blocs des principaux dispositifs à un facteur traitement. Les systèmes de blocs sont présentés en allant du plus simple au plus compliqué : les plans en blocs ont un facteur bloc (non trivial), les plans en lignes-colonnes et les plans en répliques ont deux facteurs blocs, les carrés semi-latins en ont 3.

Voir Bailey (1991) pour une présentation plus détaillée des systèmes en blocs.

1. Pourquoi prendre en compte des blocs ?

Un plan complètement aléatoire à r répétitions est

- chaque traitement est répété r fois,

- les traitements sont répartis sur les unités de

manière complètement aléatoire.

Si on avait planifié l'essai du GEVES sur tournesol à l'aide d'un plan complètement aléatoire, le modèle

d'analyse serait :

$$\hat{y}_i(t) = \mu_i + \varepsilon_i$$

avec :

$\hat{y}_i(t)$ rendement sur la parcelle i ,

t variété présente sur la parcelle i ,

μ_i rendement moyen de la variété i ,

ε_i erreurs résiduelles supposées indépendantes avec

$$\varepsilon_i \sim \mathcal{N}(0, \sigma_{PCA}^2).$$

Les erreurs résiduelles sont dues à l'hétérogénéité des parcelles et aux erreurs techniques.

Le modèle peut être écrit sous la forme :

$$y_i(t) = \mu + \tau_i + \varepsilon_i$$

avec :

μ = moyenne des μ_i ,

$\tau_i = \mu_i - \mu$, **effet de la variété i** ,

Les effets des variétés vérifient la contrainte :

$$\sum_{i=1}^t \tau_i = 0$$

Notation abrégée du modèle :

$$y_i = \text{„variété“}$$

La notation abrégée du modèle évoque celle utilisée pour spécifier les modèles lorsqu'on utilise un logiciel statistique comme SAS ou S-PLUS. Elle permet de noter les modèles de manière simplifiée. Elle sera présentée régulièrement pour les différents modèles considérés. La moyenne générale et l'erreur résiduelle y sont définies implicitement.

1. Pourquoi prendre en compte des blocs ?

Un plan complètement aléatoire à r répétitions est

- chaque traitement est répété r fois,

- les traitements sont répartis sur les unités de

manière complètement aléatoire.

Si on avait planifié l'essai du GEVES sur tournesol à l'aide d'un plan complètement aléatoire, le modèle

d'analyse serait :

$$\hat{y}_i(t) = \mu_i + \varepsilon_i$$

avec :

$\hat{y}_i(t)$ rendement sur la parcelle i ,

t variété présente sur la parcelle i ,

μ_i rendement moyen de la variété i ,

ε_i erreurs résiduelles supposées indépendantes avec

$$\varepsilon_i \sim \mathcal{N}(0, \sigma_{PCA}^2).$$

Les erreurs résiduelles sont dues à l'hétérogénéité des parcelles et aux erreurs techniques.

Notation : $\hat{y}_{\bullet}(t)$ = moyenne des rendements observés de la variété t .

- **Estimer** un paramètre inconnu consiste à donner une valeur à ce paramètre à l'aide des observations. Les estimateurs sont en général des variables aléatoires.
- Pour comparer les variétés, on cherche à estimer les paramètres $\tau_t - \tau_{t'}$. L'estimateur de $\tau_t - \tau_{t'}$ est noté $\widehat{\tau_t - \tau_{t'}}$.

- Avec un plan complètement aléatoire, on aurait :

$$\widehat{\tau_t - \tau_{t'}} = \hat{y}_{\bullet}(t) - \hat{y}_{\bullet}(t')$$

- L'**espérance** de $\widehat{\tau_t - \tau_{t'}}$ est égale à :

$$E(\widehat{\tau_t - \tau_{t'}}) = \tau_t - \tau_{t'}$$

Le **biais** de $\widehat{\tau_t - \tau_{t'}}$ est défini par

$$E(\widehat{\tau_t - \tau_{t'}}) - (\tau_t - \tau_{t'}). \text{ L'estimateur n'est pas biaisé.}$$

- La **variance** de $\widehat{\tau_t - \tau_{t'}}$ mesure la précision de

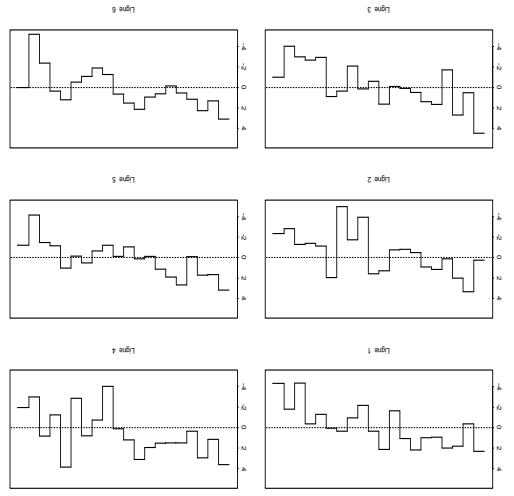
$$\widehat{\tau_t - \tau_{t'}}.$$

$$\text{Var}(\widehat{\tau_t - \tau_{t'}}) = \frac{p}{2} \sigma_{PCA}^2$$

Les résidus de l'analyse en plan complètement aléatoire intègrent l'ensemble des effets terrains. L'essai présente un gradient d'hétérogénéité de gauche à droite. Avec ce gradient d'hétérogénéité, le plan complètement aléatoire n'est pas très satisfaisant. Il peut conduire à un plan où les répétitions d'une variété sont situées d'un même côté du dispositif. Si par exemple les répétitions d'une variété se trouvent sur les parcelles de gauche du dispositif, cette variété sera favorisée par les effets terrains.

Résidus de l'analyse en plan complètement aléatoire

$$\text{Résidus : } \hat{\varepsilon}_i = \widehat{y_i(t)} - (\mu + \tau_i)$$



Gradient d'hétérogénéité non pris en compte dans le modèle

L'essai du GEVES a été planifié selon un **plan en blocs complets** : chaque variété est dans chaque bloc une fois.

Répartition des variétés en quatre blocs complets

	Bloc 1	Bloc 2	Bloc 3	Bloc 4
12	24	17	21	22
7	3	20	4	9
11	26	25	18	13
16	23	6	30	15
8	28	27	10	5
1	2	14	19	29
16	23	6	30	15
11	26	25	18	13
7	3	20	4	9
19	17	14	23	25
4	16	5	24	15
26	28	18	3	7
19	26	28	21	23
29	27	8	6	30
27	10	8	25	13

Modèle :

$$y_{ij(t)} = \mu_{ti} + \varepsilon_{ij}$$

avec :

$y_{ij(t)}$ rendement sur la parcelle j du bloc i ,

t variété présente sur cette parcelle,

μ_{ti} rendement moyen de la variété t dans le bloc i ,

ε_{ij} erreurs résiduelles supposées indépendantes avec

$$\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_{\text{BGC}}^2).$$

Le modèle peut être écrit sous la forme :

$$y_{ij(t)} = \mu + \tau_i + \beta_j + \gamma_{ti} + \varepsilon_{ij}$$

avec :

$\mu = \mu_{\bullet\bullet}$ moyenne générale,

$\tau_i = \mu_{i\bullet} - \mu_{\bullet\bullet}$ effet de la variété t_i ,

$\beta_j = \mu_{\bullet j} - \mu_{\bullet\bullet}$ effet du bloc j ,

$\gamma_{ti} = \mu_{ti} - \mu_{i\bullet} - \mu_{\bullet i} + \mu_{\bullet\bullet}$ interaction entre la variété t

et le bloc j ,

Les paramètres vérifient les contraintes :

$$\sum_i \tau_i = 0$$

$$\sum_i \beta_j = 0$$

$$\sum_i \gamma_{ti} = 0$$

Ce plan évite que les répétitions d'une variété soient situées d'un même côté du dispositif.

Hypothèse : on suppose que les interactions entre les variétés et les blocs sont nulles.

Le modèle devient :

$$y_{ij(t)} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

Notation abrégée :

$$"y = \text{variété} + \text{bloc}"$$

L'effet bloc n'est plus intégré dans la résiduelle.

Si les effets blocs ne sont pas négligeables, la prise en

compte des effets blocs dans le modèle a pour conséquence de diminuer la variance résiduelle.

$$\widehat{\tau_i - \tau_j} = \hat{y}_{i\bullet\bullet(t)} - \hat{y}_{j\bullet\bullet(t)}$$

$$E(\widehat{\tau_i - \tau_j}) = \tau_i - \tau_j$$

$$\text{Var}(\widehat{\tau_i - \tau_j}) = \frac{p}{2} \sigma_{\text{PBC}}^2$$

Notation : PBC = plan en blocs complets. La résiduelle du modèle intègre les différences entre parcelles non expliquées par les effets blocs. La variance résiduelle est donc d'autant plus faible que les parcelles sont semblables à l'intérieur des blocs. La variance des comparaisons entre variétés du plan en blocs complets et celle du plan complètement aléatoire font intervenir des variances résiduelles différentes.

L'effet bloc n'est plus intégré dans la résiduelle.

Si les effets blocs ne sont pas négligeables, la prise en compte des effets blocs dans le modèle a pour conséquence de diminuer la variance résiduelle.

$$\widehat{\tau_i - \tau_j} = \hat{y}_{i\bullet\bullet(t)} - \hat{y}_{j\bullet\bullet(t)}$$

$$E(\widehat{\tau_i - \tau_j}) = \tau_i - \tau_j$$

$$\text{Var}(\widehat{\tau_i - \tau_j}) = \frac{p}{2} \sigma_{\text{PBC}}^2$$

Tableau d'analyse de variance PCA

Terme	ddl	Somme	Carre	F	Proba
Terme	29	537,2	18,5	3,5	0,00
Carres	Moyen				
Residuelle	90	472,3			

Tableau d'analyse de variance PBC

Terme	ddl	Somme	Carre	F	Proba
Terme	29	537,2	18,5	5,8	0,00
Carres	Moyen				
Residuelle	87	276,0			
Bloc	3	196,3	65,4	20,6	0,00

Notations : ddl = nombre de degrés de liberté, F = statistique du test F, Proba = probabilité associée au test F, PCA = plan complètement aléatoire, PBC = plan en blocs complets. La prise en compte du facteur bloc a permis de réduire la variance résiduelle et d'augmenter la puissance du test des effets variétés, c'est à dire la capacité du test à mettre en évidence des différences entre les effets des variétés (la valeur du F est plus forte pour l'analyse PBC que pour l'analyse PCA).

Tableau d'analyse de variance PCA

Terme	ddl	Somme	Carre	F	Proba
Terme	29	537,2	18,5	3,5	0,00
Carres	Moyen				
Residuelle	90	472,3			

Tableau d'analyse de variance PBC

Terme	ddl	Somme	Carre	F	Proba
Terme	29	537,2	18,5	5,8	0,00
Carres	Moyen				
Residuelle	87	276,0			
Bloc	3	196,3	65,4	20,6	0,00

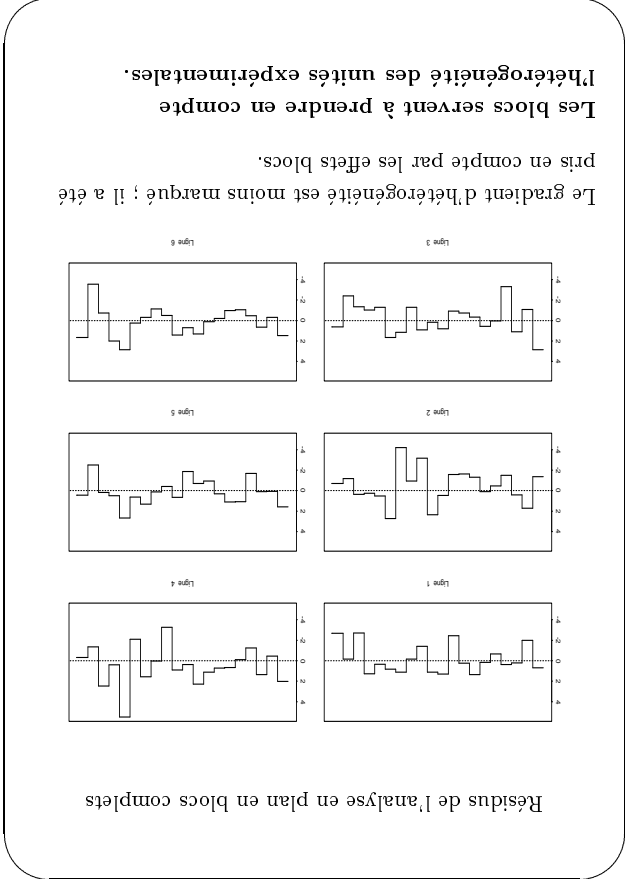
Précisions des analyses

Analyse	etd	σ^2
PCA	1,6	5,2
PBC	1,3	3,2

etd = écart-type de l'estimateur de la différence entre les effets de deux variétés :

$$etd = \sqrt{\text{Var}(t_i - t_j)}$$

La prise en compte d'un facteur bloc a permis de gagner en précision sur les comparaisons entre variétés.



Les résidus de l'analyse en plan en blocs complets n'intègrent pas les effets blocs. Le gradient est moins marqué que pour l'analyse en plan complètement aléatoire : il a été pris en compte par les effets blocs.

2. Plans en blocs

Dispositif

Un **plan en blocs** est un dispositif qui prend en compte l'hétérogénéité des unités expérimentales à l'aide d'un système de blocs.

v traitements à comparer, N unités expérimentales

La **taille** d'un bloc est le nombre d'unités qui le composent.

Choix des blocs :

- les unités d'un même bloc doivent être aussi homogènes que possible et doivent le rester pendant toute la durée de l'expérience
- si possible, on définit des blocs de même taille
 - b blocs de k unités
 - $v = k \rightarrow$ **plan en blocs complets**
 - $v > k \rightarrow$ **plan en blocs incomplets**
- les blocs peuvent être :
 - clairement définis : expériences en mesures répétées sur animaux
 - plus subjectifs : blocs utilisés pour contrôler des effets terrains

Exemple : essai variétal du GEVES sur tournesol ; plan en blocs complets avec $b = 4, v = k = 30$

Plan de l'essai

Bloc 1	Bloc 2	Bloc 3	Bloc 4
12 24 17 21 22 6 14 24 27 30 11 10 12 1 21 7 19 2 5 3	1 2 14 19 29 19 26 28 21 23 29 27 8 6 30 30	8 28 27 10 5 10 8 5 29 9 26 28 18 3 7 20 17 24 21 13	16 23 6 30 15 2 18 20 16 7 4 16 5 24 15 9 12 18 23 30
11 26 25 18 13 15 3 22 12 25 19 17 14 23 25 1 16 22 11 26	7 3 20 4 9 4 1 17 11 13 2 9 13 20 22 4 28 29 6 14		

Exemple : expérience réalisée pour comparer 21 modèles de buse pour leur aptitude à répartir de façon homogène la solution pulvérisée ; plan en blocs incomplets avec $v = 21, b = 6, k = 5$

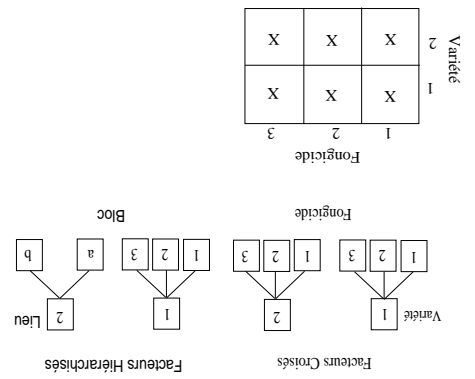
Plan de l'expérience

Bloc 1	Bloc 2	Bloc 3	Bloc 4	Bloc 5	Bloc 6
5 12 13 15 4	14 19 12 15 6	2 18 12 15 7	16 3 15 8 12	17 15 20 9 1	12 11 10 21 15

Expérience réalisée par l'ITCF en 1999

- Il s'agit de comparer des modèles et des marques de buses de pulvérisation, en conditions contrôlées dans une installation pilote. Pour chaque test, cinq buses d'un modèle donné sont fixées à une rampe sous laquelle sont réparties 10 gobelets. On met le circuit sous pression pendant 1m, puis on mesure le volume d'eau dans chaque gobelet.
 - L'installation était disponible pendant trois demi-journées. Ceci permettait de réaliser une trentaine d'essais, ce qui rendait impossible la répétition de toutes les buses. De plus, pour contrôler de possibles dérives au cours du temps, les séries de cinq essais consécutifs ont été considérées comme des blocs.
- Dans l'essai variétal du GEVES, les blocs servent à contrôler des effets terrains.
- Précisions sur l'expérience de comparaison des buses de pulvérisation.

Rappel sur les facteurs croisés et emboîtés (ou hiérarchisés)



Chaque variété est expérimentée avec même niveau du facteur bloc

Chaque fongicide reçoit le même niveau du facteur lieu (et les répétitions des niveaux des 2 facteurs entre unités ne sont pas identiques).

Variété	2	X	X	X
	1	X	X	X
		1	2	3
		Fongicide		

Présentation de notions qui servent à décrire les systèmes de blocs des dispositifs.

Remarque
Ce transparent est inspiré d'un transparent du module FPSTAT 2/3 sur le modèle linéaire.

Chaque variété est expérimentée avec même niveau du facteur bloc

Chaque fongicide reçoit le même niveau du facteur lieu (et les répétitions des niveaux des 2 facteurs entre unités ne sont pas identiques).

Structure en blocs

La **structure en blocs** d'un dispositif est l'ensemble des facteurs blocs.

Deux facteurs blocs particuliers toujours présents :

- **facteur constant U** : prend le niveau 1 pour chaque unité expérimentale ; les autres facteurs blocs sont tous emboîtés dans le facteur constant,
- **facteur unité** : prend un niveau différent pour chaque unité expérimentale et est emboîté dans tous les autres facteurs blocs.

Les facteurs constant et unité sont d'habitude implicites. Nous les explicitons ici car ils interviennent dans les représentations graphiques des structures en blocs présentées dans les prochains transparents.

Remarque
La notation U vient du fait que ce facteur est parfois appelé univers.

Représentation graphique de la structure en blocs

- Chaque facteur bloc est représenté par un point.
 - Si un facteur B est emboîté dans un facteur A :
 - A est placé au dessus de B ,
 - A et B sont reliés par un trait.
 - Un point pour le facteur constant U en haut du diagramme.
 - Un point pour le facteur unité en bas du diagramme.
- ⇒ Deux facteurs croisés ne sont pas reliés et sont en général placés au même niveau.

Cette représentation graphique s'appelle un **diagramme de Hasse**.

Structure en blocs d'un plan en blocs

Modèle d'analyse

Le modèle d'analyse dépend de la structure en blocs.

Regles pour écrire le modèle

- Additivité entre effets blocs et effets traitements
 - Un terme pour les effets traitements
 - Un terme d'effets blocs pour chaque facteur bloc (donc pour chaque point du diagramme de Hasse)
 - Deux points particuliers :
 - facteur constant $U \Rightarrow$ moyenne générale
 - facteur unité \Rightarrow erreur résiduelle
- Dans ce module, on suppose que les effets traitements sont fixes.

Modèle pour un plan en blocs

“ $y =$ traitement + bloc”

$$y_{ij(t)} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

avec :

$y_{ij(t)}$ réponse sur l'unité j du bloc i ,
 t traitement appliqué à cette unité,

μ moyenne générale,

τ_i effet du traitement t_i ,

β_j effet du bloc j ,

ε_{ij} erreurs résiduelles supposées indépendantes avec $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$.

Modèle d'analyse

Le modèle d'analyse dépend de la structure en blocs.

Regles pour écrire le modèle

- Additivité entre effets blocs et effets traitements
 - Un terme pour les effets traitements
 - Un terme d'effets blocs pour chaque facteur bloc (donc pour chaque point du diagramme de Hasse)
 - Deux points particuliers :
 - facteur constant $U \Rightarrow$ moyenne générale
 - facteur unité \Rightarrow erreur résiduelle
- Dans ce module, on suppose que les effets traitements sont fixes.

Modèle pour un plan en blocs

“ $y =$ traitement + bloc”

$$y_{ij(t)} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

avec :

$y_{ij(t)}$ réponse sur l'unité j du bloc i ,
 t traitement appliqué à cette unité,

μ moyenne générale,

τ_i effet du traitement t_i ,

β_j effet du bloc j ,

ε_{ij} erreurs résiduelles supposées indépendantes avec $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$.

On est parfois amené à utiliser un modèle différent du modèle adapté à la structure en blocs, par exemple :
 – modèles spatiaux,
 – modèles avec effets de voisinage,
 – modèles où on ajoute ou on supprime des effets blocs. Nous verrons au D que l'analyse en modèle mixte permet de résoudre en partie le problème de la suppression d'effets blocs du modèle.

3. Plans en lignes-colonnes

Dispositif

Un **plan en lignes-colonnes** est un dispositif qui prend en compte l'hétérogénéité des unités expérimentales à l'aide de deux systèmes de blocs croisés.

v traitements à comparer, N unités expérimentales

$$(N = dp)$$

Les unités forment un tableau à p lignes et q colonnes

On prend en compte des effets blocs associés aux lignes et aux colonnes.

Exemple : essai variétal pour comparer 13 variétés de

lin ; plan en lignes-colonnes avec $v=13, p=4, q=13$

Plan de l'essai

	7	4	9	3	13	8	10	12	1	11	5	6	2
Lignes	5	9	1	12	6	2	3	7	13	8	11	10	4
	8	13	6	5	3	9	7	11	10	4	2	12	1
	6	5	11	1	2	12	9	13	8	3	10	4	7
	Colonne	7	4	9	3	13	8	10	12	1	11	5	6

Essai réalisé par le GEVES

Effets lignes : effets terrains

Effets colonnes : effets dus aux pratiques culturales

Exemple : expérience réalisée pour comparer

l'ingestibilité de 4 ensilages de maïs ; plan en

lignes-colonnes avec $v=4, p=4, q=4$

Plan de l'expérience

	4	3	2	1
Lots	1	4	3	2
	2	1	4	3
	3	2	1	4
Périodes	4	3	2	1

Expérience réalisée par l'INRA.

Il y a 4 ensilages et 12 vaches réparties en 4 lots de 3 vaches.

Chaque lot est nourri successivement avec chaque ensilage.

La variable réponse est la quantité de matière sèche ingérée par période et par lot.

A l'échelle des lots et des périodes, le dispositif est un plan en lignes-colonnes.

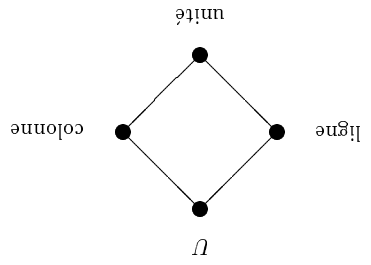
A l'intersection d'une ligne et d'une colonne il n'y a qu'une unité expérimentale.
 Dans l'essai du GEVES sur lin, les lignes sont des blocs complets, mais ce n'est pas forcément le cas dans un plan en lignes-colonnes.

Les effets lignes sont dus à des différences entre lots ; les effets colonnes sont dus à des effets périodes. Ce plan n'a pas été randomisé et peut poser des problèmes s'il y a des arrière effets.

Remarque

A l'échelle des vaches et des périodes, le dispositif a une structure en blocs plus compliquée que celle d'un plan en lignes-colonnes.

Structure en blocs d'un plan en lignes-colonnes



Modèle pour un plan en lignes-colonnes

" y = traitement + ligne + colonne"

$$y_{ij(t)} = \mu + \tau_i + \lambda_j + \omega_j + \varepsilon_{ij}$$

avec :

$y_{ij(t)}$ réponse sur l'unité située en ligne i et colonne j ,
 t traitement appliqué à cette unité,

μ moyenne générale,

τ_i effet du traitement t ,

λ_j effet de la ligne i ,

ω_j effet de la colonne j ,

ε_{ij} erreurs résiduelles supposées indépendantes avec

$$\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2).$$

Les effets lignes et colonnes sont retirés de la résiduelle.

Si le nombre de traitements est élevé ou si les unités expérimentales sont hétérogènes, un dispositif en blocs incomplets, par exemple un plan en réplics, peut permettre de gagner en précision par rapport à un plan en blocs complets.

4. Plans en réplics

Dispositif

Un **plan en réplics** est un dispositif qui permet de diviser des blocs complets en sous-blocs plus homogènes. v traitements à comparer, N unités expérimentales. Il y a r blocs complets, appelés **réplics**. Si le nombre de traitements est important, les unités peuvent être hétérogènes à l'intérieur des réplics \Rightarrow chaque réplic est divisée en s **sous-blocs** incomplets de plus petite taille. Si possible, on définit des sous-blocs de même taille k .

Dans l'essai sur maïs de l'INRA, il y a 64 génotypes et deux réplics qui sont divisées en sous-blocs de taille 8. Un plan en réplics offre une certaine liberté pour la taille des sous-blocs. On peut choisir celle-ci en tenant compte de la disposition des unités et des connaissances que l'on a de leurs hétérogénéités. On peut également la choisir à l'aide d'analyses rétrospectives effectuées les années précédentes. Pour les essais variétaux, certains auteurs recommandent une taille de sous-blocs égale approximativement à la racine carrée du nombre de variétés (Paterson et Hunter, 1983), d'autres recommandent une taille d'environ 10 parcelles (Oger et Roisin, 1992).

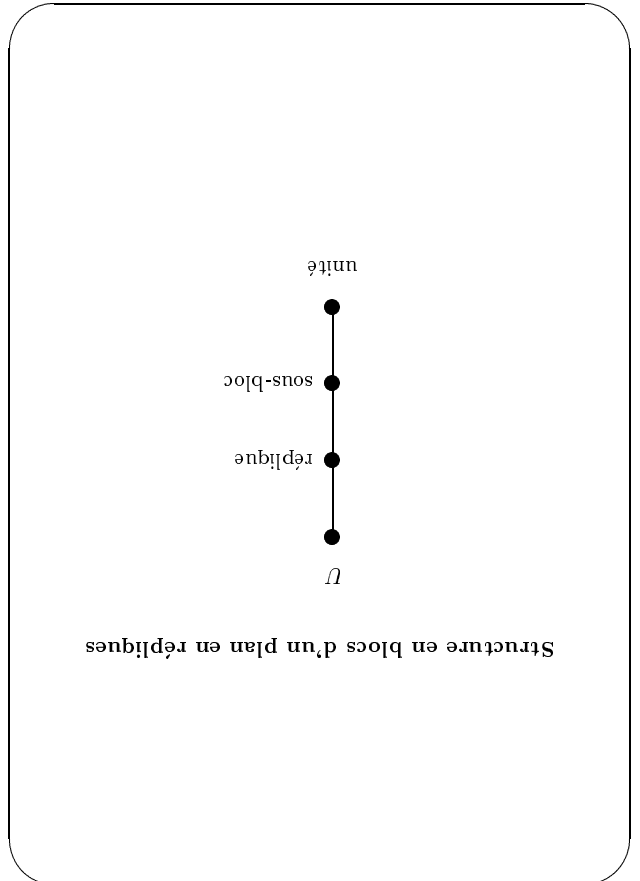
Exemple : essai en champ réalisé par l'INRA pour comparer 64 génotypes de maïs ; plan en réplics avec $v=64$, $r=2$, $k=8$

Réplique 1				Réplique 2											
33	43	50	42	46	34	52	10	6	47	16	32	48	3	15	34
48	41	7	54	21	45	58	26	35	53	51	39	38	43	41	24
62	23	36	4	64	24	16	2	64	37	54	27	22	50	25	9
20	17	63	12	3	49	39	9	13	45	23	55	52	12	19	60
11	57	19	38	5	59	37	32	36	46	59	31	7	44	28	49
29	60	35	15	18	61	22	28	10	4	14	26	61	5	8	20
27	31	1	47	14	13	51	30	30	58	2	33	11	56	17	18
8	6	40	56	25	55	53	44	29	62	40	57	63	21	1	42

Lorsque v n'est pas un multiple de k , on peut ajouter quelques traitements à l'expérience ou utiliser des sous-blocs de tailles k et $k+1$.

Exemple : plan en réplics avec des sous-blocs de tailles 3 et 4 ; $v=11$, $r=4$, $k=3$ ou 4

Sous-bloc 1				Sous-bloc 2				Sous-bloc 3															
Rép. 1	5	6	3	8	9	10	1	2	7	11	4	Rép. 2	5	2	8	4	11	3	1	9	10	7	6
Rép. 3	2	11	6	1	4	5	9	7	10	3	8	Rép. 4	10	1	8	4	6	2	7	9	3	5	11



Modèle pour un plan en répétitions

“ y = traitement + répétition + sous-bloc”

$$y_{ijl(t)} = \mu + \tau_t + \rho_i + \beta_{ij} + \varepsilon_{ijl}$$

avec :

- $y_{ijl(t)}$ réponse sur l'unité l du sous-bloc j de la répétition i ,
- t traitement appliqué à cette unité,
- μ moyenne générale,
- τ_t effet du traitement t ,
- ρ_i effet de la répétition i ,
- β_{ij} effet du sous-bloc j de la répétition i ,
- ε_{ijl} erreurs résiduelles supposées indépendantes avec $\varepsilon_{ijl} \sim \mathcal{N}(0, \sigma^2)$.

A la différence du modèle du plan en blocs complets, les effets des sous-blocs sont intégrés dans le modèle plutôt que dans la résiduelle.

Les carrés semi-latins sont parfois appelés carrés quasi-latins, plans latinisés. Dans un plan en lignes-colonnes, il n'y a qu'une unité à l'intersection d'une ligne et d'une colonne, alors que dans un carré semi-latin, il y a un sous-bloc à l'intersection d'une ligne et d'une colonne. Une alternative à l'utilisation d'un carré semi-latin pour contrôler un double gradient d'hétérogénéité est d'utiliser un plan en blocs complets ou un plan en répétitions avec des unités de forme allongée :

– les blocs permettent de contrôler le gradient dans une direction, – la forme allongée des unités permet de contrôler le gradient dans l'autre direction.

Un carré semi-latin est un dispositif qui permet de croiser deux systèmes de blocs complets sans faire autant de répétitions qu'un plan en lignes-colonnes. v traitements à comparer, N unités expérimentales Chaque traitement est répété r fois. Les unités forment un tableau à r lignes et r colonnes. A l'intersection d'une ligne et d'une colonne il y a un sous-bloc de taille $k = v/r$. Les lignes et les colonnes sont des blocs complets. Leurs intersections forment des sous-blocs incomplets. On prend en compte des effets blocs associés aux lignes, aux colonnes, aux sous-blocs.

5. Carrés semi-latins

Dispositif

L1	20	28	27	4	6	32	10	8	14	1	8	22	24	19	13	17	16	21	29	7	23	11	18	3	25	9	30	15	2		
L2	6	23	15	7	24	19	30	13	3	10	27	4	17	12	16	28	9	25	8	18	20	31	22	14	32	5	21	26	11	1	29
L3	9	29	17	8	12	16	25	50	20	7	26	11	32	18	21	27	14	6	5	13	4	3	2	31	23	13	22	28	19	24	10
L4	11	31	13	22	21	18	3	14	5	2	28	15	25	29	25	9	19	12	32	1	10	26	24	30	16	6	8	27	4	20	17

Exemple : essai variétal réalisé par l'ITCF pour comparer 32 variétés d'orge d'hiver ; carré semi-latin avec $v=32$, $r=4$, $k=8$

Lorsque v n'est pas un multiple de r , on peut ajouter quelques traitements au dispositif ou utiliser des sous-blocs de tailles k et $k+1$.

Exemple : carré semi-latin avec des sous-blocs de tailles 3 et 4 ; $v=11$, $r=3$, $k=3$ ou 4

Ligne 1	9	6	4	1	7	10	3	5	11	8	2
Ligne 2	8	3	2	1	6	11	5	9	7	10	4
Ligne 3	10	5	11	7	2	9	8	4	1	3	6

Le carré semi-latin de l'ITCF a 32 variétés, 4 lignes et 4 colonnes. A l'intersection d'une ligne et d'une colonne, il y a un sous-bloc de taille 8. Les colonnes sont composées de huit colonnes élémentaires.

Le plan présenté n'est pas un carré semi-latin car les colonnes ne sont pas complètes.
 Dans un carré semi-latin, il n'est pas possible de choisir la taille des sous-blocs, qui est égale à $k = v/r$. Si k est trop élevé pour obtenir des sous-blocs homogènes, on peut alors préférer un plan en répliques ou un plan latinisé, qui laissent plus de choix pour k .

Si les colonnes ou les lignes ne peuvent pas être complètes mais contiennent chaque traitement au plus une fois \Rightarrow **plan latinisé**.

Exemple : plan latinisé à 3 lignes et 4 colonnes ; $v=12, r=3, k=3$

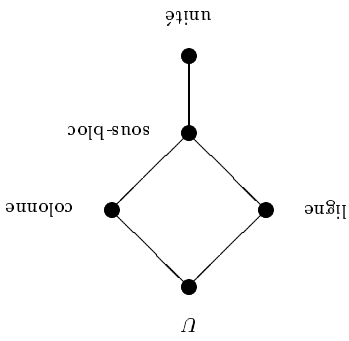
	Colonne 1	Colonne 2	Colonne 3	Colonne 4
Ligne 1	5	11	8	10
Ligne 2	12	4	3	9
Ligne 3	9	6	7	4

Structure en blocs et modèle d'un carré semi-latin

Facteurs blocs : U , ligne, colonne, sous-bloc, unité

Exercice : écrire le diagramme de Hasse et le modèle d'un carré semi-latin.

Diagramme de Hasse :
 Corrigé de l'exercice



Le facteur sous-bloc s'interprète comme l'interaction entre le facteur ligne et le facteur colonne.
 Modèle :

" y = traitement + ligne + colonne + sous-bloc"

$$y_{ijl(t)} = \mu + \tau_i + \lambda_j + \omega_l + \beta_{ij} + \varepsilon_{ijl}$$

avec :
 $y_{ijl(t)}$: réponse sur l'unité l du sous-bloc situé en ligne i et colonne j ,
 t traitement appliqué à cette unité,
 μ moyenne générale,
 τ_i effet du traitement i ,
 λ_j effet de la ligne j ,
 ω_l effet de la colonne l ,
 β_{ij} effet du sous-bloc situé en ligne i et colonne j ,
 ε_{ijl} erreurs résiduelles supposées indépendantes avec $\varepsilon_{ijl} \sim \mathcal{N}(0, \sigma^2)$.

6. Conclusion

- Les blocs servent à prendre en compte l'hétérogénéité des unités expérimentales.
- Choix de la structure en blocs :
 - la structure en blocs doit décrire le mieux possible les hétérogénéités connues entre unités,
 - les unités d'un même bloc doivent être aussi homogènes que possible et doivent le rester pendant toute la durée de l'expérience,
 - prendre en compte les contraintes sur les façons de faire varier les traitements entre les unités, éviter d'utiliser des facteurs blocs qui ont peu de chances de correspondre à de véritables hétérogénéités entre unités.
- Le modèle d'analyse dépend de la structure en blocs.
 - Avantages et inconvénients de prendre en compte des blocs :
 - avantages : meilleure répartition des traitements et réduction possible de σ^2 ,
 - inconvénients : augmentation du nombre de paramètres du modèle d'analyse.
 - Le choix de la structure en blocs est une étape de la planification.

B. ANALYSE D'UN DISPOSITIF EN BLOCS AVEC LE MODÈLE À EFFETS BLOCS FIXES

Plan

1. Exemple
2. Estimation de contrastes
3. Coefficients d'efficacité d'un dispositif
4. Matrice d'information
5. Tests et intervalles de confiance
6. Logiciels
7. Conclusion

Cette partie présente :

- des rappels sur le modèle linéaire ; voir le module FPSTAT 2/3 pour plus d'informations sur le modèle linéaire,
 - comment le plan d'expériences influence l'analyse.
- Plan de cette partie : exemple d'analyse avec le modèle à effets blocs fixes puis présentation de la théorie de l'analyse.

L'inconvénient de prendre en compte des blocs est d'augmenter le nombre de paramètres du modèle d'analyse. Nous verrons au B que lorsque les effets blocs sont supposés fixes, ceci a comme conséquences :

- de diminuer le nombre de degrés de liberté résiduels,
- d'avoir à ajuster les estimations des effets traitements pour les effets de blocs incomplets.

Nous verrons au D qu'il y a moins de risques à introduire des effets blocs dans le modèle lorsque les effets blocs sont considérés comme aléatoires.

D'un point de vue pratique, cet inconvénient signifie surtout qu'il faut éviter d'utiliser des facteurs blocs qui ont peu de chance de contrôler de véritables hétérogénéités entre unités.

1. Exemple

Plan de l'essai

	Colonne 1	Colonne 2	Colonne 3	Colonne 4
L1	20262827	4	3	32105114
L2	20262827	4	8	22241913
L3	20262827	4	6	22241913
L4	20262827	4	2	22241913
L5	20262827	4	8	22241913
L6	20262827	4	3	22241913
L7	20262827	4	2	22241913
L8	20262827	4	6	22241913
L9	20262827	4	2	22241913
L10	20262827	4	8	22241913
L11	20262827	4	3	22241913
L12	20262827	4	2	22241913
L13	20262827	4	6	22241913
L14	20262827	4	2	22241913
L15	20262827	4	8	22241913
L16	20262827	4	3	22241913
L17	20262827	4	2	22241913
L18	20262827	4	6	22241913
L19	20262827	4	2	22241913
L20	20262827	4	8	22241913
L21	20262827	4	3	22241913
L22	20262827	4	2	22241913
L23	20262827	4	6	22241913
L24	20262827	4	2	22241913
L25	20262827	4	8	22241913
L26	20262827	4	3	22241913
L27	20262827	4	2	22241913
L28	20262827	4	6	22241913
L29	20262827	4	2	22241913
L30	20262827	4	8	22241913
L31	20262827	4	3	22241913
L32	20262827	4	2	22241913
L33	20262827	4	6	22241913
L34	20262827	4	2	22241913
L35	20262827	4	8	22241913
L36	20262827	4	3	22241913
L37	20262827	4	2	22241913
L38	20262827	4	6	22241913
L39	20262827	4	2	22241913
L40	20262827	4	8	22241913
L41	20262827	4	3	22241913
L42	20262827	4	2	22241913
L43	20262827	4	6	22241913
L44	20262827	4	2	22241913
L45	20262827	4	8	22241913
L46	20262827	4	3	22241913
L47	20262827	4	2	22241913
L48	20262827	4	6	22241913
L49	20262827	4	2	22241913
L50	20262827	4	8	22241913
L51	20262827	4	3	22241913
L52	20262827	4	2	22241913
L53	20262827	4	6	22241913
L54	20262827	4	2	22241913
L55	20262827	4	8	22241913
L56	20262827	4	3	22241913
L57	20262827	4	2	22241913
L58	20262827	4	6	22241913
L59	20262827	4	2	22241913
L60	20262827	4	8	22241913
L61	20262827	4	3	22241913
L62	20262827	4	2	22241913
L63	20262827	4	6	22241913
L64	20262827	4	2	22241913
L65	20262827	4	8	22241913
L66	20262827	4	3	22241913
L67	20262827	4	2	22241913
L68	20262827	4	6	22241913
L69	20262827	4	2	22241913
L70	20262827	4	8	22241913
L71	20262827	4	3	22241913
L72	20262827	4	2	22241913
L73	20262827	4	6	22241913
L74	20262827	4	2	22241913
L75	20262827	4	8	22241913
L76	20262827	4	3	22241913
L77	20262827	4	2	22241913
L78	20262827	4	6	22241913
L79	20262827	4	2	22241913
L80	20262827	4	8	22241913
L81	20262827	4	3	22241913
L82	20262827	4	2	22241913
L83	20262827	4	6	22241913
L84	20262827	4	2	22241913
L85	20262827	4	8	22241913
L86	20262827	4	3	22241913
L87	20262827	4	2	22241913
L88	20262827	4	6	22241913
L89	20262827	4	2	22241913
L90	20262827	4	8	22241913
L91	20262827	4	3	22241913
L92	20262827	4	2	22241913
L93	20262827	4	6	22241913
L94	20262827	4	2	22241913
L95	20262827	4	8	22241913
L96	20262827	4	3	22241913
L97	20262827	4	2	22241913
L98	20262827	4	6	22241913
L99	20262827	4	2	22241913
L100	20262827	4	8	22241913

Carté semi-latin de l'ITCF avec 32 variétés d'orge d'hiver, 4 lignes et 4 colonnes.

La variable réponse principale est le rendement (q/ha).

Il y a une donnée manquante, indiquée entre parenthèses.

Essai réalisé en 1999.

L'analyse comprend 4 étapes :

- modélisation,
- tests (analyse de variance),
- estimation,
- validation (analyse des résidus).

Modèle d'analyse

“y = variété + ligne + colonne + sous-bloc”

Si la parcelle l du sous-bloc situé en ligne i et colonne j reçoit la variété t :

$$y_{ijl(t)} = \mu + \tau_i + \lambda_j + \omega_l + \beta_{ij} + \varepsilon_{ijl}$$

Les ε_{ijl} sont supposés indépendants avec $\varepsilon_{ijl} \sim \mathcal{N}(0, \sigma^2)$.

On suppose que les effets variétés sont fixes.

Dans cette partie, on suppose que les effets blocs sont fixes.

On est dans le cadre du **modèle linéaire** :

- l'espérance de chaque observation est une combinaison linéaire des paramètres,
- les observations ne sont pas corrélées et ont la même variance.

Analyse de variance

Analyse en plan en blocs complets (pour comparaison)

Terme	ddl	Somme	Carres	Moyen	F	Proba
Variété	31	3186,2	102,8	4,1	0,00	
Ligne	3	2947,5	982,5	39,5	0,00	
Résiduelle	92	2285,6	24,8			

Analyse en carré semi-latin

Terme	ddl	Somme	Carres	Moyen	F	Proba
Variété	31	2850,1	91,9	8,4	0,00	
Ligne	3	2949,4	983,1	89,5	0,00	
Colonne	3	556,6	185,5	16,9	0,00	
Sous-bloc	9	853,8	94,9	8,6	0,00	
Résiduelle	80	879,0	11,0			

L'analyse en plan en blocs complets a été effectuée en prenant le facteur ligne comme facteur bloc. L'analyse en plan en blocs complets d'un carré semi-latin n'est en principe pas adaptée ; elle est présentée ici pour comparaison. Ces tableaux donnent les sommes de carrés de type III. Ces sommes de carrés sont calculées à partir d'estimations des effets ajustées pour tous les autres effets du modèle. En général, il est préférable d'utiliser les sommes de carrés de type III. Dans l'analyse en carré semi-latin, les effets lignes, colonnes et sous-blocs sont tous significatifs. Les effets lignes sont les plus importants. La prise en compte des effets colonnes et sous-blocs dans l'analyse en carré semi-latin a permis :
 – de réduire la variance résiduelle,
 – d'augmenter la puissance du test des effets variétés, c'est à dire la capacité du test à mettre en évidence des différences entre les capacités des variétés (la valeur du F est plus forte pour l'analyse en carré semi-latin que pour l'analyse en plan en blocs complets).

Estimation des moyennes des variétés

Estimations

Var	PBC	CSL	Ajust	Var	PBC	CSL	Ajust
7	92,4	89,2	-3,2	4	81,3	80,7	-0,6
17	91,6	90,4	-1,2	25	80,9	82,6	1,7
8	90,7	92,7	2,0	22	80,7	81,2	0,5
21	89,1	90,0	0,9	1	80,5	83,2	2,7
28	88,0	89,0	0,9	9	79,8	81,5	1,7
27	86,8	86,2	-0,6	13	79,6	76,8	-2,8
19	86,8	84,7	-2,1	16	79,6	78,3	-1,2
32	85,7	88,0	2,3	18	79,4	80,1	0,8
10	85,2	83,8	-1,4	14	79,0	81,1	2,1
11	84,8	85,8	0,9	15	78,0	76,5	-1,5
20	84,4	86,0	1,5	2	77,3	75,8	-1,5
23	84,3	81,4	-2,9	26	76,6	78,8	2,3
6	84,0	82,8	-1,2	24	76,4	74,3	-2,1
12	82,9	82,2	-0,7	5	74,0	76,5	2,5
29	82,6	84,5	1,9	31	73,4	73,8	0,5
30	81,4	79,7	-1,7	3	72,6	71,2	-1,4

La prise en compte des colonnes et des sous-blocs a modifié le classement des variétés ...

Précision des estimations

etd	etd	etd	etd	etd	etd	etd	etd
Var	PBC	CSL	Ajust	Var	PBC	CSL	Ajust
7	92,4	89,2	-3,2	4	81,3	80,7	-0,6
17	91,6	90,4	-1,2	25	80,9	82,6	1,7
8	90,7	92,7	2,0	22	80,7	81,2	0,5
21	89,1	90,0	0,9	1	80,5	83,2	2,7
28	88,0	89,0	0,9	9	79,8	81,5	1,7
27	86,8	86,2	-0,6	13	79,6	76,8	-2,8
19	86,8	84,7	-2,1	16	79,6	78,3	-1,2
32	85,7	88,0	2,3	18	79,4	80,1	0,8
10	85,2	83,8	-1,4	14	79,0	81,1	2,1
11	84,8	85,8	0,9	15	78,0	76,5	-1,5
20	84,4	86,0	1,5	2	77,3	75,8	-1,5
23	84,3	81,4	-2,9	26	76,6	78,8	2,3
6	84,0	82,8	-1,2	24	76,4	74,3	-2,1
12	82,9	82,2	-0,7	5	74,0	76,5	2,5
29	82,6	84,5	1,9	31	73,4	73,8	0,5
30	81,4	79,7	-1,7	3	72,6	71,2	-1,4

... et a permis de gagner en précision sur les comparaisons entre variétés.

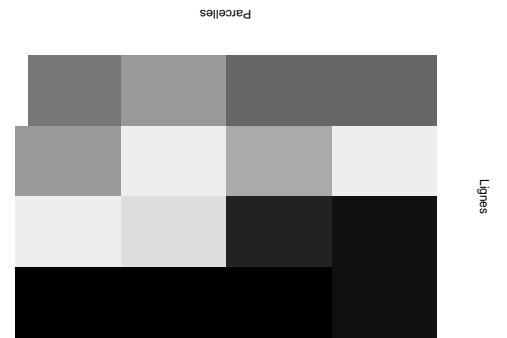
Les etd sont en principe tous égaux dans une analyse en plan en blocs complets. Ce n'est pas le cas ici à cause de la donnée manquante. Les etd ne sont en général pas tous égaux pour un dispositif en blocs incomplets (voir plus loin).

Remarque

– les écarts-types des différences (etd).
 – la variance résiduelle.

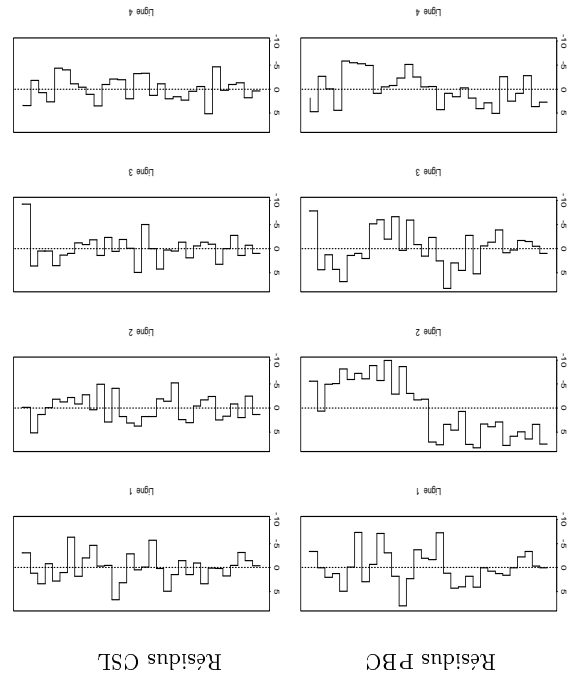
Notations : Var = variété, PBC = plan en blocs complets, CSL = carré semi-latin, Ajust = Ajustement = différence entre l'estimation CSL et l'estimation PBC, etd = écart-type des différences.
 Les variétés sont classées par ordre décroissant des estimations de l'analyse en plan en blocs complets. La variance en carré semi-latin a changé le classement des variétés. La variance en carré semi-latin a permis de réduire :
 – la plus rabaisée.
 La prise en compte des colonnes et des sous-blocs dans l'analyse en carré semi-latin a permis de réduire :
 – la plus rabaisée.
 Les variétés sont classées par ordre décroissant des estimations de l'analyse en plan en blocs complets. La variance en carré semi-latin a changé le classement des variétés. La variance en carré semi-latin a permis de réduire :
 – la plus rabaisée.
 La prise en compte des colonnes et des sous-blocs dans l'analyse en carré semi-latin a permis de réduire :
 – la plus rabaisée.
 Les variétés sont classées par ordre décroissant des estimations de l'analyse en plan en blocs complets. La variance en carré semi-latin a changé le classement des variétés. La variance en carré semi-latin a permis de réduire :
 – la plus rabaisée.
 La prise en compte des colonnes et des sous-blocs dans l'analyse en carré semi-latin a permis de réduire :
 – la plus rabaisée.

Estimation des effets terrains



L'effet terrain correspond à la somme de l'effet ligne, de l'effet colonne et de l'effet sous-bloc. Plus le grisé est foncé, plus l'estimation de l'effet terrain est élevée. On retrouve que les effets lignes sont importants. La variété 1, qui est la variété qui a été la plus remontée par l'analyse en carré semi-latin, était située sur des parcelles où les effets terrains étaient plutôt défavorables (C1 L3, C2 L1, C3 L4, C4 L2). Au contraire, la variété 7, qui est la variété qui a été la plus rabaisée par l'analyse en carré semi-latin, était située sur des parcelles où les effets terrains étaient plutôt favorables (C1 L2, C2 L3, C3 L1, C4 L4).

Analyse des résidus



Résidus CSL plus faibles et distribués de manière plus indépendante, en particulier en ligne 2.

Les résidus de l'analyse en plan en blocs complets intègrent les effets colonnes et sous-blocs, alors que ceux de l'analyse en carré semi-latin n'intègrent pas ces effets.

Corrigé de l'exercice

$$E(\widehat{\tau_1 - \tau_2}) = \tau_1 - \tau_2$$

$$\text{Var}(\widehat{\tau_1 - \tau_2}) = \frac{p}{2} \sigma_{\text{PBC}}^2$$

(PBC = plan en blocs complets ; sur l'exemple $r = 2$)

2. Estimation de contrastes

Exemple : plan en blocs complets

Plan :

1	2	3	1	2	3
Bloc 1			Bloc 2		

Les différences entre les effets traitements sont estimées à l'aide des moyennes brutes.

$$\widehat{\tau_1 - \tau_2} = (\bar{y}_{11(1)} + \bar{y}_{21(1)})/2 - (\bar{y}_{12(2)} + \bar{y}_{22(2)})/2$$

$$= \tau_1 - \tau_2 + (\varepsilon_{11} + \varepsilon_{21} - \varepsilon_{12} - \varepsilon_{22})/2$$

Les effets blocs s'annulent.

Exercice : calculer l'espérance et la variance de $\widehat{\tau_1 - \tau_2}$.

$$E(\widehat{\tau_1 - \tau_2}) =$$

$$\text{Var}(\widehat{\tau_1 - \tau_2}) =$$

Cet exemple a le même nombre de traitements et d'unités expérimentales que l'exemple précédent. Sur ce transparent, on montre que l'estimateur le plus intuitif, celui basé sur la différence des moyennes brutes, ne convient pas pour un plan en blocs incomplets, car il est biaisé par des effets blocs.

Exemple : plan en blocs incomplets

Plan :

1	2	3	1	2	3
Bloc 1		Bloc 2		Bloc 3	

$$\widehat{\delta} = \tau_1 - \tau_2$$

Estimateur basé sur des moyennes brutes :

$$\delta^{\text{brut}} = (\bar{y}_{11(1)} + \bar{y}_{22(1)})/2 - (\bar{y}_{12(2)} + \bar{y}_{31(2)})/2$$

$$= \tau_1 - \tau_2 + (\beta_2 - \beta_3)/2 + (\varepsilon_{11} + \varepsilon_{22} - \varepsilon_{12} - \varepsilon_{31})/2$$

$$E(\delta^{\text{brut}}) = \tau_1 - \tau_2 + (\beta_2 - \beta_3)/2$$

Les effets blocs ne s'annulent pas : estimateur biaisé. Il faut modifier les estimations des effets traitements pour qu'elles ne soient pas biaisées par les effets blocs ; on dit qu'il faut **ajuster** les estimations des effets traitements pour les effets blocs.

Plan :

1	2	3	1	2	3
Bloc 1		Bloc 2		Bloc 3	

Plusieurs possibilités pour ajuster les estimations des effets traitements pour les effets blocs :

- comparaison directe
 $\hat{\delta}^{direct} = \hat{y}_{11(1)} - \hat{y}_{12(2)} = \tau_1 - \tau_2 + \varepsilon_{11} - \varepsilon_{12}$
 Exerc. : calculer l'espérance et la variance de $\hat{\delta}^{direct}$.
 $E(\hat{\delta}^{direct}) = E(\tau_1 - \tau_2) = \tau_1 - \tau_2$
 $Var(\hat{\delta}^{direct}) = Var(\tau_1 - \tau_2) = 2\sigma_{PB1}^2$
- comparaison indirecte
 $Var(\hat{\delta}^{indirect}) =$
 $E(\hat{\delta}^{indirect}) =$
 Exerc. : calculer l'espérance et la variance de $\hat{\delta}^{indirect}$.
 $\hat{\delta}^{indirect} = (\hat{y}_{22(1)} - \hat{y}_{21(3)}) + (\hat{y}_{32(3)} - \hat{y}_{31(2)})$
 $= \tau_1 - \tau_2 + \varepsilon_{22} - \varepsilon_{21} + \varepsilon_{32} - \varepsilon_{31}$
 Exerc. : calculer l'espérance et la variance de $\hat{\delta}^{indirect}$.
 $E(\hat{\delta}^{indirect}) = \tau_1 - \tau_2$
 $Var(\hat{\delta}^{indirect}) = \frac{2}{3} \times \frac{4}{3} \sigma_{PB1}^2 = \frac{8}{9} \sigma_{PB1}^2$
- estimateur des moindres carrés

Contrastes

Un **contraste** sur les effets traitements est une combinaison linéaire des effets traitements $c_1\tau_1 + \dots + c_p\tau_p$ telle que $c_1 + \dots + c_p = 0$.

Exemple : différence entre les effets de deux traitements : $\tau_1 - \tau_2$.

Un effet traitement centré τ_i est un contraste sur les moyennes des traitements $(\mu + \tau_i)$.

Dans ce module, on considère des expériences dont l'objectif est d'estimer des contrastes sur les moyennes des traitements.

avec :

- ' = transposition
- $c = (c_1 \dots c_p)'$
- $\tau = (\tau_1 \dots \tau_p)'$

Norme du vecteur c :

$$\|c\| = \sqrt{c_1^2 + \dots + c_p^2}$$

Exemples :

- pour $\tau_1 - \tau_2$, $\|c\|_2 = 2$,
- pour un effet traitement centré, $\|c\|_2 = \frac{p}{\sqrt{p-1}}$.

Les expériences considérées dans ce module ont pour objectif principal de comparer des traitements entre eux. Ces comparaisons peuvent s'exprimer comme des contrastes sur les effets traitements. La norme d'un contraste intervient dans la variance de son estimateur, présentée plus loin.

Plus généralement, un contraste sur un vecteur x est une combinaison linéaire $c'x$ telle que $\sum_i c_i = 0$.

Corrigé de l'exercice

Notation : $PB1 =$ plan en blocs incomplets.

Plusieurs estimateurs des différences entre effets traitements non biaisés sont possibles. Seul celui des moindres carrés doit être utilisé car il est le plus précis. C'est celui qui est donné comme solution par la théorie du modèle linéaire. Une comparaison directe est une différence entre deux observations d'un même bloc. Une comparaison indirecte combine plusieurs comparaisons directes. Dans cet exemple, l'estimateur des moindres carrés de $\tau_1 - \tau_2$ est une combinaison linéaire de $\hat{\delta}^{direct}$ et $\hat{\delta}^{indirect}$, dont les poids sont proportionnels aux inverses des variances $\hat{\delta}^{direct}$ et $\hat{\delta}^{indirect}$.

Les comparaisons directes sont plus précises que les comparaisons indirectes. Si deux traitements apparaissent sous vent dans un même bloc, la comparaison de ces traitements sera précise car elle fera intervenir plusieurs comparaisons directes. Les nombres de fois où les traitements apparaissent dans un même bloc sont donc des caractéristiques importantes du plan.

Comparaison plan en blocs complets - plan en blocs incomplets.

Comparaison du plan en blocs complets - plan en blocs incomplets.

Avantage du plan en blocs incomplets : l'utilisation de blocs incomplets peut permettre de réduire la variance résiduelle.

Corrigé de l'exercice

$$E(\hat{\delta}^{direct}) = \tau_1 - \tau_2, \quad Var(\hat{\delta}^{direct}) = 2\sigma_{PB1}^2$$

$$E(\hat{\delta}^{indirect}) = \tau_1 - \tau_2, \quad Var(\hat{\delta}^{indirect}) = 4\sigma_{PB1}^2$$

Estimabilité d'un contraste

Un contraste c^T est **estimable** s'il existe une combinaison linéaire des observations dont l'espérance est égale à c^T

\Rightarrow il existe une combinaison linéaire des observations qui est un estimateur sans biais de c^T .

Exemple :

Bloc 1	1	2	3	4
Bloc 2	1	2	3	4

• $\tau_1 - \tau_2$ est-il estimable ?

$E(y_{11(1)} - y_{12(2)}) = \tau_1 - \tau_2 \Rightarrow \tau_1 - \tau_2$ est estimable.

• $\tau_1 - \tau_3$ est-il estimable ?

Non : aucune combinaison linéaire des observations dont l'espérance est égale à $\tau_1 - \tau_3$. Par exemple :

$E(y_{11(1)} - y_{21(3)}) = \tau_1 - \tau_3 + \beta_1 - \beta_2$

Exercice : on considère le plan :

Bloc 1	1	2	3
Bloc 2	1	2	3

$\tau_1 - \tau_3$ est-il estimable ?

Connexité d'un dispositif

Un dispositif en blocs est dit **connexe** si tous les contrastes sur les effets traitements sont estimables.

Un plan en blocs n'est pas connexe si et seulement si :

- on peut diviser les traitements en deux groupes G_1 et G_2 ,
- aucun bloc ne contient à la fois un traitement de G_1 et un traitement de G_2 .

Exemple de plan qui n'est pas connexe :

Bloc 1	1	3
Bloc 2	2	4
Bloc 3	3	5
Bloc 4	4	6
Bloc 5	5	1
Bloc 6	6	2

Exercice : préciser les groupes G_1 et G_2 .

Dans la suite on considère, sauf mention du contraire, des plans connexes.

Corrigé de l'exercice

G_1 comprend les traitements impairs ; G_2 comprend les traitements pairs. Il suffit de rajouter un bloc avec un traitement de G_1 et un traitement de G_2 pour rendre le plan connexe.

Écriture du modèle sous forme matricielle

Plan :

1	2	3	1	2	3
Bloc 1			Bloc 2		

Modèle pour chaque observation :

$$y_{ij(t)} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

Modèle sous forme matricielle :

$$\begin{pmatrix} y_{11(1)} \\ y_{12(2)} \\ y_{13(3)} \\ y_{21(1)} \\ y_{22(2)} \\ y_{23(3)} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \mu + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \end{pmatrix}$$

Exercice : compléter les éléments de la matrice associée aux effets blocs.

avec :

- y vecteur des observations, y vecteur de uns (ici de taille 6),
- μ moyenne générale,
- X matrice d'incidence des traitements,
- τ vecteur des effets traitements,
- Z matrice d'incidence des blocs,
- β vecteur des effets blocs,
- ε vecteur des erreurs résiduelles.

y peut être considéré comme un vecteur de \mathbb{R}^N .

Matrice de variance-covariance de $\varepsilon = \text{Var}(\varepsilon)$:

- éléments diagonaux : $\text{Var}(\varepsilon_{ij}) = \sigma^2$,
- éléments non diagonaux : $\text{Cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = 0$.

$\varepsilon \sim \mathcal{N}(0, \sigma^2 I_N)$

avec I_N matrice identité de taille N

On pourra insister sur le fait que :

- la matrice Z dépend de la structure en blocs,
- la matrice X dépend de la répartition des traitements entre les unités.

Corrigé de l'exercice

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

L'écriture matricielle est utile pour effectuer les calculs théoriques et les calculs numériques du modèle linéaire.

Notation : $\text{Im}(X'Q_Z X) =$ sous-espace engendré par les colonnes de $X'Q_Z X$, $y_{i\bullet} =$ moyenne des $y_{ij(t)}$ du bloc i .
 Si les stagiaires ne sont pas habitués aux notations matricielles, on peut se limiter à utiliser ce transparent pour montrer que :

- il existe des expressions matricielles très générales pour calculer les estimateurs de contrastes et leur variance,
- pour un plan en blocs, $\hat{c}'\tau$ s'exprime à partir des observations centrées par blocs $(Q_Z y)$,
- la variance des contrastes dépend de la matrice $X'Q_Z X$, qui joue donc un rôle central pour évaluer la qualité d'un dispositif.

La notation utilisée pour le modèle est valable pour les différents modèles présentés au A ; le terme $Z\beta$ regroupe les différents effets blocs du dispositif.

Quelques précisions :

Dans le cadre du modèle linéaire gaussien, on fait l'hypothèse que les erreurs résiduelles suivent une distribution normale. La méthode du maximum de vraisemblance et celle des moindres carrés sont alors équivalentes.

On utilise une inverse généralisée de $X'Q_Z X$ au lieu d'une inverse, car $X'Q_Z X$ n'est pas inversible. Une inverse généralisée de $X'Q_Z X$ est une matrice qui vérifie :

$$X'Q_Z X (X'Q_Z X)^- X'Q_Z X = X'Q_Z X.$$

Q_Z s'interprète comme la matrice du projecteur sur le sous-espace orthogonal aux colonnes de Z .

Formule matricielle de l'estimateur d'un contraste

Modèle :

$$y = \Pi \mu + X \tau + Z \beta + \varepsilon$$

Méthode d'estimation : méthode du maximum de vraisemblance ou méthode des moindres carrés.

Un contraste $c'\tau$ est estimable si et seulement si $c \in \text{Im}(X'Q_Z X)$. Alors :

$$\hat{c}'\tau = c'(X'Q_Z X)^- X'Q_Z y$$

$$E(\hat{c}'\tau) = c'\tau$$

$$\text{Var}(\hat{c}'\tau) = c'(X'Q_Z X)^- c\sigma^2$$

avec :

$(X'Q_Z X)^-$ inverse généralisée de $X'Q_Z X$,

$Q_Z = I_N - Z(Z'Z)^- Z'$,

pour un plan en blocs, Q_Z retire les moyennes des blocs : composantes de y : $y_{ij(t)}$ - composantes de $Q_Z y$: $y_{ij(t)} - y_{i\bullet}$.

Retour à l'exemple

Plan :

1	2	3
Bloc 1	Bloc 2	Bloc 3

Modèle :

$$y = \Pi \mu + X \tau + Z \beta + \varepsilon$$

$$X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, Z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

$$Q_Z = I_6 - \frac{1}{2} \begin{pmatrix} J_2 & 0 & 0 \\ 0 & J_2 & 0 \\ 0 & 0 & J_2 \end{pmatrix}$$

avec

I_6 : matrice identité de taille 6

J_2 : matrice carrée remplie de uns de taille 2

Corrigé de l'exercice

$$\text{Var}(\widehat{\tau_1 - \tau_2}) = \frac{3}{2} d_{12}^2 (I_3 - \frac{1}{3} J_3) d_{12}^2 \sigma^2 = \frac{3}{4} \sigma^2$$

$$X'Q_Z X = \frac{3}{1} I_3 - \frac{3}{2} J_3, (X'Q_Z X)^{-1} = \frac{3}{2} I_3 - \frac{3}{1} J_3$$

$$\widehat{\tau_1 - \tau_2} = \frac{3}{2} d_{12}^2 (I_3 - \frac{1}{3} J_3) \times$$

$$\begin{pmatrix} (y_{11(1)} - y_{1\bullet}) + (y_{22(1)} - y_{2\bullet}) \\ (y_{12(2)} - y_{1\bullet}) + (y_{31(2)} - y_{3\bullet}) \\ (y_{21(3)} - y_{2\bullet}) + (y_{32(3)} - y_{3\bullet}) \end{pmatrix}$$

$$= (2)(y_{11(1)} - y_{12(2)}) + y_{22(1)} - y_{21(3)} + (y_{32(3)} - y_{31(2)})/3$$

$$\text{où } d_{12} = (1, -1, 0)'$$

Exercice : calculer la variance de $\widehat{\tau_1 - \tau_2}$.

3. Coefficients d'efficacité d'un dispositif

Coefficient d'efficacité pour un contraste

Variance de l'estimateur d'un contraste :
 pour un PBC :

$$\text{Var}(\widehat{c\tau}) = \frac{\|c\|_2^2}{r} \sigma_{PBC}^2$$

pour un DBI :

$$\text{Var}(\widehat{c\tau}) = \frac{\|c\|_2^2}{r} \sigma_{DBI}^2$$

$e(c)$ = coefficient d'efficacité du dispositif pour le contraste $c\tau$

$$e(c) = 1 \text{ pour un PBC,}$$

$0 \leq e(c) \leq 1$ pour un DBI \Rightarrow facteur d'inflation de la variance.

La notation A vient du mot *average*. Le coefficient d'efficacité moyen tient compte de l'ensemble des comparaisons entre traitements. Lorsqu'on choisit un dispositif en blocs incomplets, c'est en espérant que le gain dû à la réduction de la variance résiduelle compensera le coût d'ajustement pour les effets blocs.

Coefficient d'efficacité moyen
deux :
$$A = \frac{1}{2} \sum_{t < t'} \text{Var}(t_t - t_{t'})$$

Pour un PBC :

$$A = \frac{r}{2} \sigma_{\text{PBC}}^2$$

Pour un DBI :

$$A = \frac{r}{2} \sigma_{\text{DBI}}^2$$

$e =$ **coefficient d'efficacité moyen** du dispositif

$e = 1$ pour un PBC.

$0 \leq e \leq 1$ pour un DBI \Rightarrow facteur d'inflation de la variance

Par convention, on pose $e = 0$ si le plan n'est pas connexe.

Comparaison PBC – DBI

Pour un DBI :

- $0 \leq e \leq 1 \Rightarrow$ "coût" en précision dû à l'ajustement pour les effets blocs ;
- si le DBI réduit la variabilité des unités dans les blocs, $\sigma_{\text{PBC}}^2 \geq \sigma_{\text{DBI}}^2 \Rightarrow$ "gain" en précision dû à la réduction de la variance résiduelle.

La matrice d'information dépend :
 – de la structure en blocs (QZ),
 – de la répartition des traitements (X).

La matrice d'information et le coefficient d'efficacité ne dépendent pas de paramètres inconnus et peuvent être calculés au moment de planifier. La variance résiduelle dépend en partie du choix de la structure en blocs. Elle est inconnue au moment de la conception du dispositif, mais un bon choix de structure en blocs doit permettre de la minimiser. La formule $e = \frac{r}{a} \frac{\sum_{i=1}^r \theta_i^2}{1}$ est utile pour calculer le coefficient d'efficacité d'un dispositif et c'est cette formule qui a été utilisée pour calculer les valeurs de e données dans la suite. Les valeurs propres θ_i peuvent être calculées à l'aide des principaux logiciels statistiques.

Remarques

Les valeurs propres θ_i et vecteurs propres z_i de M sont définis par :

$$M z_i = \theta_i z_i$$

M a une valeur propre nulle associée au vecteur propre $\mathbb{1}$.

4. Matrice d'information

Matrice d'information sur τ pour le modèle à effets blocs fixes :

$$M = X'QZX$$

La précision des estimations dépend du plan

d'expériences au travers de :

- la variance résiduelle σ^2 ,
- la matrice d'information M .

La matrice d'information joue un rôle central pour évaluer la qualité d'un dispositif.

Elle détermine en particulier le coefficient d'efficacité du dispositif par la relation :

$$e = \frac{r}{a} \frac{\sum_{i=1}^r \theta_i^2}{1}$$

avec θ_i valeurs propres de M qui ne sont pas associées à $\mathbb{1}$.

Notations : v = nombre de traitements, k = taille des blocs, r = nombre de répétitions des traitements.

Matrice d'information pour un plan en blocs

Modèle :

$$y = \mathbb{1}\mu + X\tau + Z\beta + \varepsilon$$

Matrice d'information :

$$M = X'X - X'Z'Z'X/k$$

Un plan **équilibré** est un plan dans lequel les traitements sont expérimentés un même nombre de fois.

Pour un plan équilibré :

$$M = rI_v - X'Z'Z'X/k = rI_v - C/k$$

Matrice des concomitances :

$$C = X'Z'Z'X$$

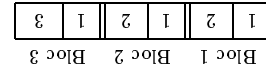
• C_{tt} = nombre de répétitions du traitement t

• $C_{tt'}$ = **concomitance** entre les traitements t et t'

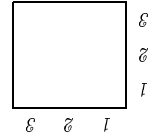
Un plan **binaire** est un plan dans lequel chaque bloc contient chaque traitement au plus une fois.

Pour un plan binaire, la concomitance entre les traitements t et t' correspond au nombre de blocs qui contiennent les traitements t et t' .

Exercice : on considère le plan :

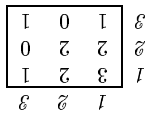


Ecrire la matrice des concomitances :



Pour un plan en blocs, M dépend :

- des nombres de répétitions des traitements,
- des concomitances entre traitements,
- de la taille des blocs.



Corrigé de l'exercice

Les quantités dont dépend la matrice M sont importantes car elles influencent la variance des estimations des contrastes. On retrouve que la précision de l'analyse dépend des concomitances entre traitements ; nous avons vu que ces dernières déterminent les nombres de comparaisons directes entre traitements.

Notations : p = nombre de lignes, q = nombres de colonnes, $X'LL'X$ est la matrice des concomitances en lignes, $X'WW'X$ est la matrice des concomitances en colonnes.

Remarque
L'expression de M vient du fait que le projecteur sur le sous-espace orthogonal aux colonnes de L et W s'écrit :

$$Q_{L,W} = I_N - LL'/q - WW'/p + \Pi \Pi' / (dp).$$

On en déduit l'expression à l'aide la formule $M = X'Q_{L,W}X$.

Matrice d'information pour un plan en lignes-colonnes

Modèle :

$$y = \Pi \tau + X \tau + L \lambda + W \omega + \varepsilon$$

avec :

L matrice d'incidence des lignes,

λ vecteur des effets lignes,

W matrice d'incidence des colonnes,

ω vecteur des effets colonnes.

Matrice d'information :

$$M = X'X - X'LL'X/b - X'WW'X/d + X' \Pi \Pi' X / (bd)$$

M dépend :

- des nombres de répétitions des traitements,
- des concomitances entre traitements en lignes,
- des concomitances entre traitements en colonnes,
- des nombres de lignes et de colonnes.

Matrice d'information pour un plan en répétitives

Modèle :

$$y = \Pi \tau + X \tau + R p + Z \beta + \varepsilon$$

avec :

R matrice d'incidence des répétitives,

p vecteur des effets répétitives,

Z matrice d'incidence des sous-blocs,

β vecteur des effets sous-blocs.

Matrice d'information :

$$M = X'Q_Z X$$

Même expression que la matrice d'information d'un plan en blocs incomplets.

Dans un plan en répétitives, le facteur sous-bloc est emboîté dans le facteur répétitive. Avec le modèle à effets blocs fixes, l'effet d'une répétitive s'exprime donc comme la moyenne des effets des sous-blocs qui la composent. Par conséquent, ajuster les estimations des effets traitements pour les effets répétitives et sous-blocs revient à ajuster les estimations des effets traitements pour les effets sous-blocs seulement. La matrice d'information du plan en répétitives a donc la même expression que celle d'un plan en blocs incomplets et dépend des nombres de répétitions des traitements, des concomitances entre traitements dans les sous-blocs et de la taille des sous-blocs.

Dans un carré semi-latin, le facteur sous-bloc est emboîté dans le facteur ligne et le facteur colonne. Ajuster les estimations des effets traitements pour les effets lignes, colonnes et sous-blocs revient à ajuster les estimations des effets traitements pour les effets sous-blocs seulement (pour le modèle à effets blocs fixes).

Matrice d'information pour un carré semi-latin

Modèle :

$$y = \Pi \mu + X \tau + L \lambda + W \omega + Z \beta + \varepsilon$$

avec :

L matrice d'incidence des lignes,

λ vecteur des effets lignes,

W matrice d'incidence des colonnes,

ω vecteur des effets colonnes,

Z matrice d'incidence des sous-blocs,

β vecteur des effets sous-blocs,

Matrice d'information :

$$M = X' \hat{Q} Z X$$

Même expression que la matrice d'information d'un plan

en blocs incomplets.

5. Tests et intervalles de confiance

Test global sur les traitements

Modèle :

$$y = \Pi \mu + X \tau + Z \beta + \varepsilon, \varepsilon \sim \mathcal{N}(0, \sigma^2 I_N)$$

On rejette H_0 “ $\tau = 0$ ” si :

$$F = \frac{CM_\tau}{CM_\varepsilon} \geq F(v - 1, ddf_\varepsilon, \alpha)$$

CM_τ : carré moyen associé aux traitements (de type III)

CM_ε : carré moyen résiduel

ddf_ε : nombre de degrés de liberté résiduels

α : risque de première espèce (0.05)

$F(v - 1, ddf_\varepsilon, \alpha)$: α -quantile maximum de la loi de Fisher à $v - 1$ et ddf_ε degrés de liberté

On se limite ici à présenter les tests et intervalles de confiance portant sur les effets traitements. On considère implicitement la contrainte $\sum_i \tau_i = 0$; H_0 “ $\tau = 0$ ” est donc l'hypothèse d'égalité de tous les effets traitements. La notation utilisée pour le modèle est valable pour les différents modèles présentés au A ; le terme $Z\beta$ regroupe les différents effets blocs du dispositif. En général, il est préférable d'utiliser les sommes de carrés de type III, car ces sommes de carrés sont calculées à partir d'estimations des effets ajustées pour tous les autres effets du modèle.

Remarque

Ce transparent est inspiré d'un transparent du module FPSTAT 1.

Test	Réalité	
	H_0 vraie	H_0 fausse
H_0 vraie	$1 - \alpha$	β
H_0 fausse	α	$1 - \beta$
	<i>risque 1^{ère} espèce</i>	<i>risque 2^{ème} espèce</i>

Le **risque de première espèce** est la probabilité de détecter à tort des différences entre les effets traitements lorsque ces différences sont nulles.

On choisit α : le risque de première espèce est contrôlé.

La **puissance** est la probabilité de détecter des différences entre les effets traitements lorsque ces différences ne sont pas nulles.

La **puissance dépend du plan d'expériences au travers** :

- de la variance des estimations des effets traitements,
- du nombre de degrés de liberté résiduels,
- du nombre de degrés de liberté associés aux traitements.

Tests de comparaisons par paires

- H_0 " $\tau_i - \tau_{i'} = 0$ " est rejetée si :

$$|\widehat{\tau_i - \tau_{i'}}| \geq t(dll_{\alpha}, \alpha) \sqrt{\widehat{\text{Var}(\tau_i - \tau_{i'})}}$$

$t(dll_{\alpha}, \alpha)$: α -quantile maximum de la loi de Student à dll_{α} degrés de liberté

Plus petite différence significative :

$$pdds(\alpha) = t(dll_{\alpha}, \alpha) \sqrt{\widehat{\text{Var}(\tau_i - \tau_{i'})}}$$

- **Intervalle de confiance** de $\tau_i - \tau_{i'}$ au niveau $1 - \alpha'$:

= intervalle aléatoire qui contient $\tau_i - \tau_{i'}$ avec la probabilité $1 - \alpha'$

$$\widehat{\tau_i - \tau_{i'}} - pdds(\alpha') \leq \tau_i - \tau_{i'} \leq \widehat{\tau_i - \tau_{i'}} + pdds(\alpha')$$

- On peut utiliser des méthodes de *comparisons multiples* pour contrôler le risque de première espèce sur l'ensemble des comparaisons.

La longueur des intervalles de confiance dépend du plan d'expériences au travers :

- de la variance des comparaisons entre traitements,
- du nombre de degrés de liberté résiduels.

6. Logiciels S-PLUS

```

> csl.dat <- read.table("csl.txt",header=T)
> csl.dat$VARIABLE <- factor(csl.dat$VARIABLE)
> csl.dat$LIGNE <- factor(csl.dat$LIGNE)
> csl.dat$COLONNE <- factor(csl.dat$COLONNE)
> aov.csl <- aov(RDT ~ VARIABLE + LIGNE*COLONNE,
data=csl.dat,na.action=na.omit)
> round(drop1.aov(csl,scope=aov.csl$call),2)
Single term deletions

Model:
RDT ~ VARIABLE + LIGNE * COLONNE
Df Sum of Sq  RSS F Value Pr(>F)
<none>
VARIABLE 31 2650.11 3729.14 8.37 0
LIGNE 3 2949.37 3828.40 89.47 0
COLONNE 3 556.59 1435.61 16.88 0
LIGNE:COLONNE 9 853.83 1732.86 8.63 0
> dummy.coef(aov.csl)
$(Intercept) :
82.14796
(Intercept)

```

Instructions et sorties S-PLUS de l'analyse avec le modèle à effets blocs fixes du carré semi-latin de l'ITCF à 32 variétés, 4 lignes et 4 colonnes.

Les premières instructions permettent de lire les données et de déclarer les facteurs. La commande aov permet d'effectuer l'analyse. La commande drop1.aov(csl, scope=aov.csl\$call) donne les sommes de carrés de type III. La commande summary(aov.csl) (qui n'a pas été utilisée ici) donne les sommes de carrés de type I. La commande dummy.coef(aov.csl) donne les estimations des effets centres.

```

$VARIABLE:
71103 52126 NH 71104 1997 TH  Altees  Angora  BIDON  CAROLA
1.067737 -6.375102 -10.965956 -1.399312 -5.650556 0.6628958
CASTRO CRUNCH  Clarine  EPONA  ESTEREL  EVEREST  Labaa
7.064899 10.57349 -0.697011 1.649208 3.625564 0.02598209 -5.303633
Ladoga  Lambic  MAEVA MAJESTIC  MARILOR  MATHIAS  MUSCAT
-1.091695 -5.670102 -3.837722 8.247278 -2.037482 2.551416 3.831656
MIXEL Naturelle PLAISANT Pacific Platine Reine SARAH
7.880564 -0.9574243 -0.7702854 -7.848584 0.4679889 -3.314095 4.053188
SPHINX SUMATRA Splendid Sunrise Vanessa
6.816215 2.326035 -2.473641 -8.327424 5.865905
$LIGNE:
1 1 2 3 4
7.380477 -1.007023 -6.048273 -0.3251803
$COLONNE:
1 1 2 3 4
1.455477 2.463602 -2.848586 -1.070493
$LIGNE:COLONNE":
11 21 31 41 12 22
-2.63419 5.847149 -3.176886 -0.03607215 -1.99459 4.045581
-0.7399604 -1.31103 3.208561 -3.686102 0.02222745 0.4553142
14 24 34 44
1.42022 -6.206627 3.894619 0.8917879
> sed.calc(aov.csl,"VARIABLE")
means
Variances moyenne des comparaisons 2 à 2 : 6.40067221871711
sed.min sed.moyenne sed.max
2.343908 2.528978 2.777894
> motif()
> hist(aov.csl$residuals)
> box()

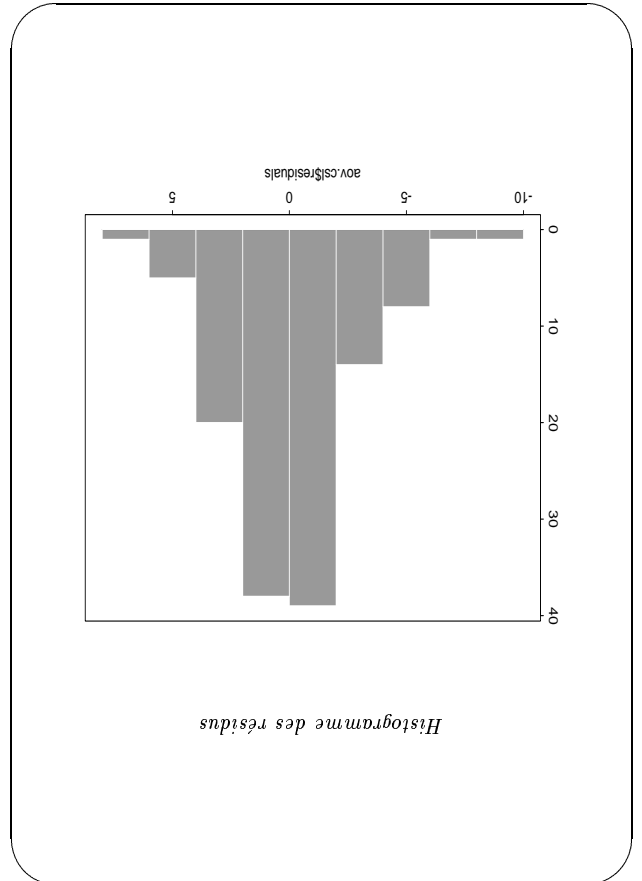
```

La fonction sed.calc est une fonction que nous avons écrite et qui permet d'estimer les écarts-types des différences (sed en anglais) pour un dispositif en blocs incomplets. Elle est donnée plus loin.

Les dernières instructions permettent de tracer un histogramme des résidus.

Remarque

D'autres représentations graphiques peuvent être effectuées pour examiner les résidus (par exemple graphes des résidus en fonction des valeurs prédites, boîtes à pattes ; voir le module FPSTAT 2/3 sur le modèle linéaire).



```
sed.calc <- fonction(object,facteur,type="means",output="resume")
{
# estimations des variances d'estimation pour les niveaux d'un facteur
# ENTREEES
# object: objet de classe aov ou varcomp
# facteur: nom du facteur auquel on s'intéresse, dont l'effet principal
# doit faire partie de la partie fixe du modèle de "object"
# type: "means" si l'on s'intéresse aux comparaisons deux à deux
# entre les niveaux du facteur; "effects" si l'on s'intéresse
# aux effets
# out: "resume" pour obtenir les valeurs min-max de la sed; "details"
# pour obtenir toutes les sed
# SORTIE
# suivant l'option choisie dans output
# auteur: H. Monod
p <- length(object) > 0
if(sum(class(object)=="aov") > 0)
{
vov <- solve(object$diag,p)
vov <- vov %*% t(vov) * sum(object$residuals_2/object$df
select <- object$ssigm[[facteur]]
}
else
{
vov <- object$cov.fixed
select <- object$ssigm.fixed[[facteur]]
}
n <- length(select)+1
vov <- vov[select,select]
contrastes <- object$contrasts[[facteur]]
vov <- contrastes %*% vov %*% t(contrastes)
VNC <- sum(diag(vcov))*2/(n-1)
takevc <- (1:n[2])[outer(1:n,1:n,">")]
V11 <- matrix(diag(vcov),n,n)
V11 <- t(V11)
V11+V11-2*vcov)[takevc]
```

Fonction S-PLUS pour estimer les écarts-types des dif-
férences.

```

if (type=="means") { out <- vcomp220.5;cat("means\n")}
else if (type=="effects") { out <- diag(vcov).0.5;cat("effects\n")}
}
if (output=="resume")
{
out <- c(sed.min=min(out),sed.moyenne=mean(out),sed.max=max(out))
}
else if (type=="means")
{
outvec <- out
out <- matrix(NA,n,n)
out[takevec] <- outvec
}
cat(paste("variance moyenne des comparaisons 2 à 2: ",VMC,"n\n"))
}
out
}

```

```

data csl;
infile 'csl.txt';
input position variete $ ligne colonne rdt x y;
run;

proc glm data=csl;
class variete ligne colonne;
model rdt = variete ligne colonne / sst;
lsmeans variete ligne colonne;
output out=resu r=resid;
run;

proc chart data=resu;
vbar resid;
run;

The SAS System
General Linear Models Procedure
Class Level Information
Class Levels Values
VARIETE 32 7103-52 7104-19 A1ess Angora BIDON CAROLA CASTRO
GRUNCH Clarine EPONA ESTEREL EVEREST Labea Ladoga
Lambic MAEVA MAJESTIC MARILOR MATHIAS MUSCAT NIKEL
NATURELLI PLAISANT Pacific Platine Reine SARAH SPHINX
SUMATRA Splendid Sunrise Vanessa
LIGNE 4 1 2 3 4
COLONNE 4 1 2 3 4

```

Instructions et sorties SAS de l'analyse avec le modèle à effets blocs fixes du carré semi-latin de l'ITCF à 32 variétés, 4 lignes et 4 colonnes.

Les premières commandes permettent de lire les données. La procédure glm permet d'effectuer l'analyse. L'option sst permet d'obtenir les sommes de carrés de type III. Les options pdiff et tdiff de lsmeans (qui n'ont pas été utilisées dans l'analyse présentée) permettent de tester si les différences entre effets sont égales à 0. Les dernières commandes permettent de faire un histogramme des résidus.

Remarque

D'autres représentations graphiques peuvent être effectuées pour examiner les résidus (par exemple graphes des résidus en fonction des valeurs prédites, boîtes à pattes ; voir le module FPSTAT 2/3 sur le modèle linéaire).

Number of observations in data set = 128

NOTE: Due to missing values, only 127 observations can be used in this analysis.

General Linear Models Procedure

Dependent Variable: RDT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	46	7547.79286	164.08245	14.93	0.0001
Error	80	879.02490	10.98781		
Corrected Total	126	8426.81776			
R-Square = 0.895687 C.V. = 3.31479 Root MSE = 82.0963					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
VARIETE	31	2850.1198	91.93910	8.37	0.0001
LIGNE	3	2949.37041	983.12347	89.47	0.0001
COLONNE	3	556.58606	185.52869	16.88	0.0001
LIGNE*COLONNE	9	853.83481	94.87053	8.63	0.0001

General Linear Models Procedure

Least Squares Means

VARIETE RDT

LSMEAN

7103-52	83.2156973
7104-19	75.7728584
Atess	71.1920110
Angora	80.7486486
BIDON	76.4974050
CAROLA	82.8108565
CASTRO	89.2128593
CRUNCH	92.7214477
Clarine	81.4509496
EPONA	83.7971691
ESTRELL	85.7735247
EVEREST	82.1739428
Labea	76.8443279
Ladoga	81.0562655
Lambic	76.4778584
MAEVA	78.3102392
MAJESTIC	90.3952392
MARILOR	80.1104787
MATHIAS	84.6993770
MUSCAT	85.9796163
NIKEL	90.0285247
Naturel1	81.1905364
PLAISANT	81.3776753
Pacific	74.2993770
Patine	82.6159496
Reine	78.8338659
SARAH	86.2011486
SPHINX	88.9641759
SUMATRA	84.4739956
Splendid	79.6743193
Sunrise	73.8205364
Vanessa	88.0138659

```

General Linear Model Procedure
Least Squares Means

LIGNE      RDT      LSMEAN
1          89.5284375
2          81.1409375
3          76.0996875
4          81.8227804

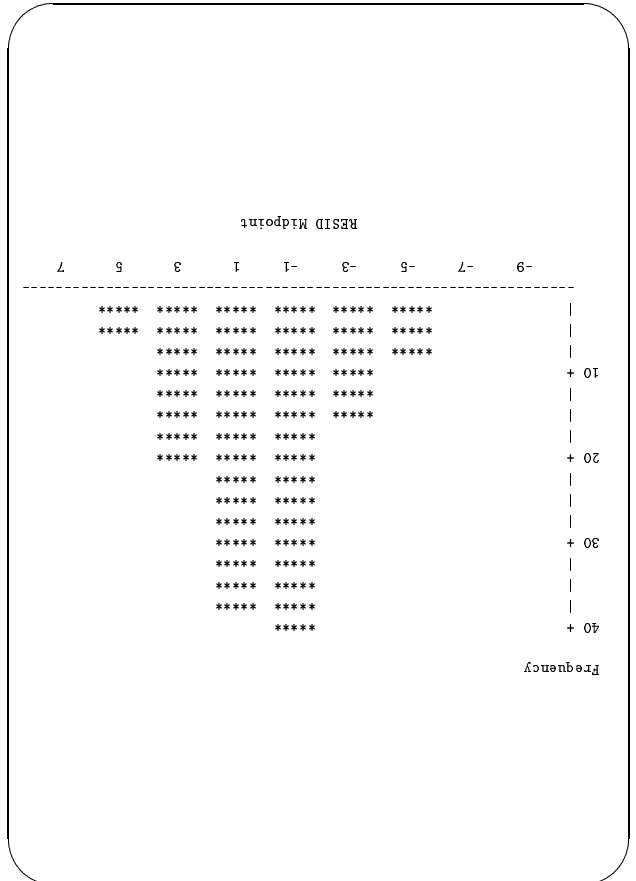
COLONNE    RDT      LSMEAN
1          83.6034375
2          84.6115625
3          79.2993750
4          81.0774679
    
```

```

General Linear Model Procedure
Least Squares Means

LIGNE      COLONNE    RDT      LSMEAN
1          1          88.3497238
2          2          89.9974490
3          3          89.8884125
4          4          89.8781647
1          1          88.4435631
2          2          87.6501199
3          3          74.6062494
4          4          73.8638176
1          1          74.3782781
2          2          77.8232889
3          3          73.2732922
4          4          78.9238138
1          1          83.2421851
2          2          82.9753522
3          3          79.4295089
4          4          81.6440756
    
```

La moyenne d'une combinaison ligne - colonne est la somme des estimations de la moyenne générale, de l'effet ligne, de l'effet colonne et de l'interaction ligne*colonne.



7. Conclusion

- L'analyse avec le modèle à effets blocs fixes ajuste les estimations des effets traitements pour les effets blocs.
- L'introduction d'effets de blocs incomplets dans le modèle :
 - a un coût : ajustement pour les effets blocs, réduction du nombre de degrés de liberté, résiduels,
 - mais peut permettre de réduire la variance résiduelle.
- Le plan d'expériences influence la variance des estimations et la puissance des tests.
- Les concomitances entre traitements sont des caractéristiques importantes du plan d'expériences.
- La matrice d'information joue un rôle central pour évaluer la qualité d'un dispositif.

C. CONSTRUCTION DE DISPOSITIFS EN BLOCS

Plan

1. Principal critère de comparaison entre dispositifs
2. Répartition des traitements entre les blocs
3. Méthodes de construction
4. Nombre de répétitions
5. Randomisation
6. Logiciels
7. Conclusion

1. Principal critère de comparaison entre dispositifs

Le dispositif est choisi pour optimiser l'analyse.

Le principal critère de comparaison entre dispositifs est la variance moyenne des comparaisons deux à deux :

$$A = \frac{a(a-1)}{2} \sum_{i > j} \text{Var}(\bar{t}_i - \bar{t}_j) = \frac{r}{2} e^2$$

On cherche un dispositif qui minimise A :

- choix du nombre de répétitions r
- ⇒ on agit sur la variance résiduelle r et e
- choix de la structure en blocs
- ⇒ on agit sur la variance résiduelle σ^2 et e
- choix de la répartition des traitements entre les blocs
- ⇒ on agit sur e

inconnus et peut être calculé au moment de planifier.

Le coefficient d'efficacité e ne dépend pas de paramètres Le poids plus important aux comparaisons traitements - témoins, critères existent, par exemple des critères qui donnent un est de comparer tous les traitements entre eux. D'autres Le critère A est approprié lorsque l'objectif de l'expérience de leur variance.

particulier grâce à la randomisation. La qualité des estimations des effets traitements peut alors être mesurée à l'aide On suppose que les estimations ne sont pas biaisées, en répétitions des traitements.

Notations : v = nombre de traitements, r = nombre de

2. Répartition des traitements entre les blocs

Une fois la structure en blocs et les traitements choisis, on cherche à répartir les traitements entre les blocs. La répartition des traitements entre les blocs est souvent appelée plan. On cherche des plans qui, pour la structure en blocs choisie :
• maximisent e : plans optimaux,
• ou ont une valeur élevée de e : plans efficaces.
Les plans ① et ② sont deux plans en répétitions. Quel est le meilleur plan ?

Plan ① table with 3 rows (Rép. 1, 2, 3) and 12 columns

Plan ② table with 3 rows (Rép. 1, 2, 3) and 12 columns

Cette partie 2 donne des idées intuitives sur le choix de la répartition des traitements entre les blocs. Les méthodes de construction de plans seront présentées dans la partie suivante 3.
Le critère utilisé pour choisir la répartition des traitements entre les blocs est le coefficient d'efficacité.

Plan ①

Plan ① table with 3 rows (Rép. 1, 2, 3) and 12 columns

Matrice des concordances C

Matrice des concordances C table with 12 rows (1 to 12) and 12 columns

e = 0 : certaines différences entre effets traitements ne sont pas estimables.
Les concordances minimum et maximum diffèrent de 3.

Le plan ① est un mauvais plan, pour lequel e = 0. Le plan n'est pas connexe si bien que certains contrastes ne sont pas estimables (avec le modèle à effets blocs fixes), par exemple $\tau_1 - \tau_4$.

Ce plan est optimal. Le logiciel qui l'a construit indiquait que le coefficient d'efficacité du plan était égal à une borne supérieure du coefficient d'efficacité. L'écart entre concommitances minimum et maximum est faible.

Remarque

On a calculé le coefficient d'efficacité du plan à l'aide de la formule :

$$e = \frac{v - 1}{v} \frac{\sum_{i=1}^r \theta_i^2}{v - 1}$$

où les θ_i sont les valeurs propres de M qui ne sont pas associées à Π .

Rép. 1	1	4	8	5	3	7	10	2	6	11	12	9
Rép. 2	2	1	5	12	7	4	3	6	9	8	11	10
Rép. 3	2	9	4	3	1	11	8	6	7	12	10	5

Plan \otimes

Matrice des concommitances C

1	3	1	1	1	1	0	0	1	0	0	1	0
2	1	3	0	1	1	1	0	1	0	1	0	1
3	1	0	3	0	1	1	1	0	0	1	1	0
4	1	1	0	3	0	1	0	0	1	0	0	1
5	1	1	1	0	3	0	1	0	0	1	0	1
6	0	1	1	0	0	3	1	1	1	0	0	0
7	0	0	1	1	1	1	3	1	0	0	0	1
8	1	0	0	1	0	1	1	3	0	1	1	0
9	0	1	1	1	0	1	0	0	3	0	1	1
10	0	1	0	0	1	1	0	1	0	3	1	1
11	1	0	1	0	0	0	0	1	1	1	3	1
12	0	0	0	1	1	0	1	0	1	1	1	3

$e = 0,68$, plan optimal.

Les concommitances minimum et maximum ne diffèrent que de 1.

Lorsqu'on attache la même importance à tous les traitements, les plans optimaux en général :

- sont équilibrés,
- sont binaires,
- équilibrent le mieux possible les concommitances entre traitements.

3. Méthodes de construction

Plans optimaux ou efficaces

Construction de plans en blocs

Plans en blocs complets

Un plan en blocs complets est un plan en blocs dans lequel chaque bloc contient chaque traitement une fois. Plans optimaux.

Les plans en blocs complets sont des plans orthogonaux.

Rappel : un plan binaire est un plan tel que chaque bloc contient chaque traitement au plus une fois. D'une façon générale, on compare deux traitements quand ils apparaissent souvent dans un même bloc. Lorsqu'on porte le même intérêt à tous les traitements, il faut donc équilibrer le mieux possible les concommitances pour toutes les paires de traitements. Bien qu'il y ait des exceptions à cette règle, on peut retenir que, en général, le plan est d'autant plus efficace que les concommitances sont proches les unes des autres. La partie 3 présente comment construire des plans (répartitions des traitements entre les blocs) optimaux ou efficaces. Elle présente :

- des plans qui sont connus pour être optimaux ou efficaces, – puis des méthodes de construction algorithmiques générales.

Exemple : plan en blocs complets avec 2 traitements et 2 blocs de taille 2.

Plan :

B1	B2
1 2	1 2

Modèle :

$$y_{ij}(t) = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

Contraintes : $\tau_2 = -\tau_1$, $\beta_2 = -\beta_1$

Si on prend en compte les contraintes sur les paramètres quand on écrit le modèle sous forme matricielle, on obtient :

$$\tilde{y} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \mu + \begin{pmatrix} -1 \\ 1 \\ -1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \tau_1 + \begin{pmatrix} -1 \\ 1 \\ -1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \beta_1 + \varepsilon$$

Le sous-espace associé aux effets traitements est orthogonal au sous-espace associé aux effets blocs :

• les effets traitements sont dits **orthogonaux** aux effets blocs,

• le facteur traitement est dit **orthogonal** au facteur bloc,

• le plan est dit **orthogonal**.

Conséquences de l'orthogonalité :

- même carré moyen des effets traitements que les effets blocs soient présents ou non dans le modèle (sous SAS : type I = type II = type III),
- les contrastes sur les effets traitements sont estimés à partir de moyennes brutes, sans avoir besoin d'ajuster pour les effets blocs,
- même précision de cette estimation que les effets blocs soient présents ou non dans le modèle,
- les estimations de contrastes sur les effets traitements et sur les effets blocs sont non corrélées (indépendantes sous l'hypothèse de normalité des erreurs).

L'interprétation des résultats d'un plan en blocs complets est simple et très intuitive.

Remarque

Le point 3 signifie que la variance des l'estimateur d'un contraste ne dépend pas du fait que les effets blocs soient présents ou non dans le modèle. Cependant, l'estimation de cette variance peut en dépendre.

Notations : v = nombre de traitements, b = nombre de blocs, k = taille des blocs, r = nombre de répétitions des traitements.

Dans un PBIE, les comparaisons entre traitements ont toutes la même variance, qui est minimale. Les PBIE sont donc des plans très intéressants, mais ils ne peuvent être construits que pour des valeurs très particulières de v, b, k (il faut en particulier que $r(k-1)/(v-1)$ soit un entier et que $v \leq b$) et nécessitent souvent un nombre de répétitions trop élevé pour pouvoir être utilisés. Ils sont cependant intéressants à connaître pour des raisons théoriques. Par exemple la formule du coefficient d'efficacité pour un PBIE fournit une borne supérieure du coefficient d'efficacité d'un plan en blocs quelconque.

Remarque

On peut retrouver la valeur de λ de la manière suivante. Dans un bloc, il y a $k(k-1)/2$ paires de traitements. La somme de toutes les concomitances entre traitements est donc égale à $bk(k-1)/2$.

Il y a $v(v-1)/2$ paires de traitements. La somme de toutes les concomitances entre traitements est donc égale à $\lambda v(v-1)/2$.

On a donc $bk(k-1)/2 = \lambda v(v-1)/2$, et comme $bk = rv$, on obtient la valeur de λ .

Plans en blocs incomplets équilibrés (PBIE)

Un plan en blocs incomplets équilibrés est un plan en blocs binaire, équilibré et dans lequel les concomitances ont toutes la même valeur λ .

$$\lambda = \frac{r(k-1)}{v-1}$$

$$v \leq b$$

$$e = \frac{v(k-1)}{v-1}k$$

Plans optimaux.

Le premier plan a été construit en partant d'un plan en blocs complets et en supprimant chaque traitement une fois dans des blocs différents.

Le deuxième plan a été construit en générant toutes les combinaisons de deux traitements possibles.

Corrigé de l'exercice

Pour le premier PBIE, $\lambda = 2$. Pour le deuxième PBIE, $\lambda = 1$.

Exemples :

PBIE avec $v = 4, b = 4, k = 3, r = 3; e = 0.89$

1	2	3
1	2	4
1	3	4
2	3	4

PBIE avec $v = 4, b = 6, k = 2, r = 3; e = 0.67$

1	2
1	3
1	4
2	3
2	4
3	4

Exercice : donner la valeur de λ pour ces deux PBIE.

Plans cycliques

Un **plan cyclique** est un plan en blocs équilibré construit par développement cyclique d'un bloc initial. Le coefficient d'efficacité dépend uniquement du bloc initial.

Exemples : plans cycliques avec $v = 6, b = 6, k = 3, r = 3$

0	2	3
1	3	4
2	4	5
3	5	0
4	0	1
5	1	2

Concomitances : 1, 2
 $e = 0,78$

0	1	2

Blocs
Concomitances : 0, 1, 2
 $e = 0,74$

Exercice : compléter le plan cyclique ci-dessus.

Construction de plans en lignes-colonnes

Carés latins

Un **caré latin** est un plan en lignes-colonnes dans lequel chaque ligne et chaque colonne contiennent chaque traitement une fois.

Plans optimaux.

Plans orthogonaux.

Exemple : expérience réalisée par l'INRA pour comparer l'ingestibilité de 4 ensilages de maïs ; caré latin avec $v=4, p=4, q=4$

4	3	2	1
1	4	3	2
2	1	4	3
3	2	1	4

Concomitances en lignes et en colonnes : v (blocs complets)

$e = 1$

Comme les FBE, ces plans ne sont pas disponibles sous-vent. Le coefficient d'efficacité est celui du plan en blocs incomplets associé aux colonnes (sur l'exemple, $e = \frac{13-1}{13-14}$).

Carres de Youden

Un **carre de Youden** est un plan en lignes-colonnes dans lequel les lignes sont des blocs complets et les colonnes forment un plan en blocs incomplets équilibré. Certains carres de Youden peuvent être construits en supprimant des lignes d'un carre latin.

Plans optimaux.

Exemple : essai varietal du GEVES sur lin : carre de Youden avec $v = 13, p = 4, q = 13, r = 4$

7	4	9	3	13	8	10	12	1	11	5	6	2
5	9	1	12	6	2	3	7	13	8	11	10	4
8	13	6	5	3	9	7	11	10	4	2	12	1
6	5	11	1	2	12	9	13	8	3	10	4	7

Colonnes

Lignes

Concomitances en lignes : 4
 Concomitances en colonnes : 1
 $e = 0.81$

Construction de plans en répliques

Latrices carrés

r répliques avec k sous-blocs de k unités.
 $v = k^2$

Plans optimaux.

Si k est un nombre premier ou une puissance d'un nombre premier, on peut construire jusqu'à $k + 1$ répliques.

$r = k + 1 \Rightarrow$ concomitances égales à 1 : latrice équilibré.

$$e = \frac{(k+1)(r-1)}{rk+2r-k-1}$$

Notations : $v =$ nombre de traitements, $k =$ taille des sous-blocs, $r =$ nombre de répliques.
 Le principal inconvénient des latrices est de n'être disponibles que pour des valeurs de v, r, k, s très particulières. En particulier, v doit être le carré d'un nombre, donc est limité, pour les valeurs inférieures à 50, à 4, 9, 16, 25, 36 et 49. Dans un latrice, les concomitances sont égales à 0 ou 1. Deux sous-blocs de deux répliques différentes ont toujours exactement un traitement en commun. Lorsque $r = k + 1$, le plan est un FBE avec $\lambda = 1$.

Remarque

Il n'existe pas de carrés latins mutuellement orthogonaux de taille 6 et on ne sait pas encore s'il existe plus de deux carrés latins mutuellement orthogonaux de taille 10.

Pour construire un lattice carré, on utilise des carrés latins mutuellement orthogonaux de taille k .
 Des carrés latins **mutuellement orthogonaux** (clmo) sont des carrés latins tels que si on superpose deux quelconques de ces carrés latins, chaque symbole du premier carré latin se trouve dans la même case que chaque symbole du deuxième carré latin exactement une fois.

Lorsque k est un nombre premier ou une puissance d'un nombre premier, on peut construire jusqu'à $k - 1$ carrés latins mutuellement orthogonaux.

Exemple : trois carrés latins mutuellement orthogonaux de taille 4

1	2	3	4
1	4	3	2
2	1	4	3
3	4	2	1

1	2	3	4
1	3	4	2
2	4	1	3
3	1	2	4

1	2	3	4
1	4	3	2
2	1	4	3
3	2	1	4

Exemple de construction d'un lattice carré ($v=9$) :

1	2	3
4	5	6
7	8	9

1	4	7
2	5	8
3	6	9

a	b	c
b	c	a
c	a	b

1	6	8
2	4	9
3	5	7

a	b	c
c	a	b
b	c	a

1	6	8
2	4	9
3	5	7

Exercice : compléter la quatrième réplique.
 Concomitances : 1
 $e = 0.75$

Notation : clmo = carrés latins mutuellement orthogonaux.
 Les traitements sont placés dans la première réplique dans un ordre standard. Dans la seconde réplique, les sous-blocs sont constitués à l'aide des colonnes de la première réplique. On génère ensuite $r - 2$ carrés latins mutuellement orthogonaux de taille k . Pour construire une nouvelle réplique :
 – on superpose un des carrés latins mutuellement orthogonaux sur la première réplique,
 – les traitements de la première réplique qui sont associés à un même symbole du carré latin sont placés dans un même sous-bloc de la nouvelle réplique.
 La répartition des unités et des sous-blocs sur le terrain peut bien être différente de celle présentée sur le trans-parent.

Corrigé de l'exercice

1	5	9
2	6	7
3	4	8

Alpha-plans

Un **alpha-plan** est un plan en répliques construit par permutations cycliques d'une réplique initiale.

Les alpha-plans ont été développés pour permettre de construire des plans en répliques pour des valeurs de v, r, k, s beaucoup plus générales que celles qui permettent de construire des latrices.

Les plans obtenus ont souvent des coefficients d'efficacité élevés lorsque l'expérience a peu de répliques et beaucoup de traitements.

Exemple de construction d'un alpha-plan ($v=12$) :

1	5	9
2	6	10
3	7	11
4	8	12

Tableau T

0	0	0
0	1	3
0	2	1

Tableau α

Réplique 1	0	0	0	0	0	0	0	0	0
	1	5	9	4	8	12	4	5	11
	2	6	10	3	7	11	3	8	10
	3	7	11	2	7	9	2	7	9
	4	8	12	1	6	12	1	6	12

Exercice : compléter la troisième réplique.

Concomitances : 0, 1
 $e = 0,67$

1	7	10
2	8	11
3	5	12
4	6	9
0	2	1

Réplique 3

Corrigé de l'exercice

Nous avons vu que dans certains cas, il est nécessaire d'utiliser deux tailles de sous-blocs : k et $k + 1$. Pour construire de tels dispositifs, on construit un alpha-plan pour $(k + 1)$ s traitements. Les numéros de traitements les plus élevés sont disposés dans la dernière colonne du tableau T et n'apparaissent donc jamais dans un même sous-bloc. En retirant ces numéros du dispositif, on est sûr que tous les sous-blocs sont de tailles k et $k + 1$.

Les traitements qui se trouvent dans une même colonne du tableau T n'apparaissent jamais dans un même sous-bloc. Cet alpha-plan n'est pas optimal : le plan α donné au 2 est optimal et a un coefficient d'efficacité de 0,68. Cet alpha-plan est cependant très efficace.

Les méthodes algorithmiques de ce tableau. Le tableau α est en général construit à l'aide

Le choix du tableau α est une étape déterminante de cette méthode : en effet, les concomitances entre traitements et donc le coefficient d'efficacité du plan, dépendent entièrement

du tableau α situé en ligne i et en colonne j .

Les traitements sont rangés dans un tableau T à s lignes et k colonnes dans un ordre systématique. On définit un second tableau, appelé tableau α , à r lignes et k colonnes, dont les éléments sont compris entre 0 et $s - 1$. Pour construire la façon suivante : chaque colonne j du tableau T est décalée vers le haut (ou le bas) de α_{ij} positions, où α_{ij} est l'élément du tableau α situé en ligne i et en colonne j .

La méthode de construction des alpha-plans peut être adaptée pour construire des carrés semi-latins ou des plans latinisés. On parle alors d'alpha-plans latinisés. Dans l'exemple, aucun traitement ne doit se trouver plus d'une fois dans une même ligne. Pour obtenir un alpha-plan latinisé, il faut que, pour chaque colonne du tableau T, les décalages effectués dans les différentes répétitions soient tous différents. Aucune valeur du tableau α ne doit donc être répétée dans une même colonne. Sur l'exemple, on vérifie que les lignes forment des blocs complets : le plan obtenu est un carré semi-latin.

Exemple de construction d'un alpha-plan latinisé ($v=12$) :

Tableau α

0	0	0
1	2	1
3	1	2
2	3	3

Aucune valeur du tableau α ne doit être répétée dans une même colonne.

Concomitances : 0, 1, 2 $e = 0,65$

Colonne 1	Colonne 2	Colonne 3	Colonne 4
0 0 0	1 2 1	3 1 2	2 3 3
4 8 12	1 6 9	3 5 10	2 7 11
3 7 11	4 5 12	2 8 9	1 6 10
2 6 10	3 8 11	1 7 12	4 5 9
1 5 9	2 7 10	4 6 11	3 8 12

Alpha-plans latinisés

Construction de carrés semi-latins

1. Choix d'un plan initial
2. Echange
 - permutation aléatoire des traitements de deux unités
3. Calcul de e
 - si e a augmenté, l'échange est accepté
 - sinon, l'échange est refusé
4. Itération des étapes 2 et 3
5. Arrêt
 - par comparaison à une borne supérieure pour e
 - quand on ne peut plus améliorer le plan

Exemple d'algorithme d'échanges

Méthodes de construction algorithmiques

Les méthodes de construction algorithmiques reposent sur des algorithmes d'optimisation, où le critère à optimiser est le coefficient d'efficacité ou un autre critère de variance. Le transparent donne un exemple simple d'algorithme. Des algorithmes plus sophistiqués existent, par exemple des algorithmes basés sur le recuit simulé. Les échanges de l'algorithme sont en général effectués avec des contraintes. Par exemple pour un plan en répétition, on permute les traitements de deux unités qui se trouvent dans une même répétition et qui ne se trouvent pas dans un même sous-bloc. Une borne supérieure pour le coefficient d'efficacité d'un dispositif en blocs incomplets peut être obtenue à l'aide de la formule du coefficient d'efficacité d'un F.B.I.E. En pratique, les logiciels utilisent des bornes plus fines.

Remarque
OPTEX de SAS et CycDesign permettent de construire des plans par des méthodes algorithmiques. L'algorithme de CycDesign, utilisé pour construire un plan en répétitions pour $v = 12, k = 3, s = 3, r = 4, t = 3$, a généré le plan α donné au 2. On sait que ce plan est optimal car son coefficient d'efficacité est égal à une borne supérieure pour e .

Exemple : construction d'un plan en blocs ayant 4 traitements et 6 blocs de taille 2

Les échanges consistent à permuter les traitements de deux unités de deux blocs différents.

Borne supérieure d'efficacité : 0,67

- Plan initial :

B1	B2	B3	B4	B5	B6
1 2	1 3	1 2	4 3	2 4	3 4

 $e = 0,55$
- Itération 1
 - Plan initial :

1 2	1 3	1 2	4 3	2 4	3 4
-----	-----	-----	-----	-----	-----

 $e = 0,55$
 - Echange :

1 2	1 2	1 2	4 3	3 4	3 4
-----	-----	-----	-----	-----	-----

 $e = 0$
 - L'échange est refusé.
 - Plan final :

1 2	1 3	1 2	4 3	2 4	3 4
-----	-----	-----	-----	-----	-----

 $e = 0,55$
- Itération 2
 - Plan initial :

1 2	1 3	1 2	4 3	2 4	3 4
-----	-----	-----	-----	-----	-----

 $e = 0,55$
 - Echange :

1 2	1 3	1 4	2 3	2 4	3 4
-----	-----	-----	-----	-----	-----

 $e = 0,67$
 - L'échange est accepté.
 - Plan final :

1 2	1 3	1 4	2 3	2 4	3 4
-----	-----	-----	-----	-----	-----

 $e = 0,67$

Borne supérieure d'efficacité atteinte \Rightarrow plan optimal.

- Itération 2
 - Plan initial :

1 2	1 3	1 2	4 3	2 4	3 4
-----	-----	-----	-----	-----	-----

 $e = 0,55$
 - Echange :

1 2	1 3	1 4	2 3	2 4	3 4
-----	-----	-----	-----	-----	-----

 $e = 0,67$
 - L'échange est accepté.
 - Plan final :

1 2	1 3	1 4	2 3	2 4	3 4
-----	-----	-----	-----	-----	-----

 $e = 0,67$
- Le plan obtenu atteint la borne supérieure d'efficacité \Rightarrow arrêt de l'algorithme.
- Plan final :

1 2	1 3	1 4	2 3	2 4	3 4
-----	-----	-----	-----	-----	-----

 $e = 0,67$
- Plan final :

1 2	1 3	1 4	2 3	2 4	3 4
-----	-----	-----	-----	-----	-----

 $e = 0,67$

Borne supérieure d'efficacité atteinte \Rightarrow plan optimal.

Le plan obtenu est un BIF.

4. Nombre de répétitions

On étudie en général l'influence de r sur :

- la *pdds* ou la longueur des intervalles de confiance,
- la puissance du test F .

pdds et longueur des intervalles de confiance

Pour un dispositif en blocs incomplets :

$$pdds(0,05) \approx t(dll_\epsilon, 0,05) \sqrt{\frac{r}{2} \hat{\sigma}^2}$$

Exemple : *pdds* pour un alpha-plan avec $v = 20$, $s = 4$, $k = 5$ et $2 \leq r \leq 4$

r	e	$t(dll_\epsilon, 0,05)$	$pdds(0,05)$
2	0,75	2,16	$2,50\sqrt{\hat{\sigma}^2}$
3	0,80	2,05	$1,87\sqrt{\hat{\sigma}^2}$
4	0,82	2,02	$1,58\sqrt{\hat{\sigma}^2}$

Notations : *pdds* = plus petite différence significative, $t(dll_\epsilon, 0,05)$ = 0,05-quantile maximum de la loi de Student avec dll_ϵ degrés de liberté, e = coefficient d'efficacité. L'étude de l'influence du nombre de répétitions sur la *pdds* est utile si l'on a de l'information sur $\hat{\sigma}^2$ au moment de planifier. On est alors capable d'évaluer par la *pdds* l'ordre de grandeur des différences entre les effets traités que l'on pourra mettre en évidence, en fonction du nombre de répétitions. Le calcul effectué ici suppose que le dispositif est analysé avec un modèle à effets blocs fixes. Le coefficient d'efficacité augmente avec le nombre de répétitions : on se rapproche de plus en plus d'un plan où les concomitances sont équilibrées. La *pdds* diminue grâce à l'augmentation de r , de e et du nombre de degrés de liberté résiduels. En général, on gagne plus à augmenter r quand r est petit que quand r est grand. Sur l'exemple, supposons que l'on s'attende à une variance résiduelle égale à 1 ($\hat{\sigma}^2 = 1$). Alors en passant de 2 à 4 répétitions, on passe d'une plus petite différence significative de 2,50 à 1,58 (ce calcul est effectué ici sans tenir compte du problème des comparaisons multiples).

Puissance du test F

Pour étudier la puissance au moment de planifier

(puissance a priori), on doit fixer une alternative à H_0 (" $r = 0$ ", par exemple H_1)

$$"r_1 = -\Delta/2, r_2 = \dots = r_{v-1} = 0, r_v = \Delta/2"$$

Pour un plan en blocs complets, on peut calculer la

puissance :

$$1 - \beta = P[F \geq \mathcal{F}(v-1, dll_\epsilon, \alpha)]$$

sous H_1 (Dean et Voss, 1999).

Exemple : puissance a priori d'un plan en blocs complets avec $v = 5$, $\Delta = 2\sigma$, $\alpha = 0,05$ et $3 \leq r \leq 5$

r	dll_ϵ	$1 - \beta$
3	0,24	0,46
4	0,46	0,60
5	16	0,60

Remarque
Les quantiles des lois de Student ou de Fisher sont fournis par les principaux logiciels statistiques tels que SP-PLUS et SAS.

On se limite souvent à étudier la puissance en examinant le nombre de degrés de liberté résiduels du test F . Pour étudier la puissance plus précisément, on peut calculer une puissance a priori. Le calcul de puissance effectué dans le transparent a tendance à sous-estimer la puissance, car il suppose que seulement deux traitements n'ont pas un effet nul. Il demande de préciser l'ordre de grandeur de la différence Δ que l'on veut pouvoir mettre en évidence et de la variance résiduelle $\hat{\sigma}^2$; ici on a supposé que $\Delta = 2\sigma$. Une fois l'expérience réalisée, il est possible de calculer une puissance a posteriori à l'aide des estimations des effets traités et de la variance résiduelle.

Le nombre de degrés de liberté résiduels du test F , par la précision sur le carré moyen résiduel, qui est lié au nombre de degrés de liberté résiduels.

Le nombre de répétitions agit sur le test F à deux niveaux : ci sont non nulles.

Rappel : la puissance du test F est la probabilité de détecter des différences entre les effets traités quand celles-ci sont non nulles.

Notations : dll_ϵ = nombre de degrés de liberté résiduels, $\mathcal{F}(v-1, dll_\epsilon, \alpha)$ = α -quantile maximum de la loi de Fisher, α = risque de première espèce, $1 - \beta$ = puissance.

5. Randomisation

La **randomisation** est une procédure qui rend aléatoire la répartition des traitements sur les unités expérimentales.

Pourquoi randomiser ?

Exemple : plan complètement aléatoire à 3 traitements et 2 répétitions.

Plan :

1	1	2	3
1	2	2	3

Plan non randomisé

Modèle :

$$y_{(i)} = \mu + \tau_i + \delta_i + \eta_i$$

avec :

δ_i effet fixe de l'unité i , $\sum_i \delta_i = 0$,

η_i erreurs techniques supposées indépendantes avec $\eta_i \sim \mathcal{N}(0, \sigma_\eta^2)$.

Plan randomisé

On permute aléatoirement les unités.

La randomisation rend aléatoires les effets unités exercés sur les réponses d'un traitement.

Modèle :

$$y_{(i)} = \mu + \tau_i + \Delta_i + \eta_i$$

Les Δ_i sont des variables aléatoires :

- qui prennent les valeurs $\delta_1, \dots, \delta_6$ avec la même probabilité,
- centrées, de variance σ_Δ^2 , avec $\text{Cov}(\Delta_i, \Delta_{i'}) = -\frac{\sigma_\Delta^2}{6}$ ($i \neq i'$),
- supposées indépendantes des erreurs techniques η_i .

$$\widehat{\tau_1 - \tau_2} = \hat{y}_{\bullet(1)} - \hat{y}_{\bullet(2)} = \tau_1 - \tau_2 + (\Delta_1 + \Delta_2 - \Delta_3 - \Delta_4)/2 + (\eta_1 + \eta_2 - \eta_3 - \eta_4)/2$$

$$E(\widehat{\tau_1 - \tau_2}) = \tau_1 - \tau_2$$

$$\text{Var}(\widehat{\tau_1 - \tau_2}) = \sigma_2^2 + \sigma_4^2$$

$\widehat{\tau_1 - \tau_2}$ n'est pas biaisé.
 On peut estimer $\text{Var}(\widehat{\tau_1 - \tau_2})$ sans biais.
 En pratique, on utilise le modèle :

$$y^{(i)} = \mu + \tau_i + \varepsilon_i$$

où les ε_i sont supposés indépendants avec $\varepsilon_i \sim \mathcal{N}(0, \sigma_2^2)$.

$E(\widehat{\tau_1 - \tau_2})$ est calculée en moyennant à la fois sur les erreurs techniques et sur les randomisations.

- Les facteurs blocs ne peuvent pas contrôler toutes les hétérogénéités.
- Les hétérogénéités non contrôlées peuvent biaiser les estimations.

En randomisant :

- on traite les traitements de façon équitable par rapport aux hétérogénéités non contrôlées,
- on évite les biais sur :
 - les estimations des effets traitements,
 - l'estimation de la variance résiduelle,
 - l'estimation de la variance des estimateurs des effets traitements,
- on conforte les hypothèses du modèle sur les espérances, variances et covariances des observations.

Les randomisations présentées dans cette partie conviennent les hypothèses du modèle d'analyse, pour le modèle adapté à la structure en blocs.
 Dans certaines situations, des plans systématiques peuvent être préférables (voir par exemple Edmondson, 1993).

A chaque structure en blocs correspond une procédure de randomisation appropriée.

Règles pour randomiser un plan

- Le choix des traitements, de la structure en blocs et de la répartition des traitements entre les blocs définit un *plan initial*.

- On permute aléatoirement les *unités* du plan initial, en respectant la structure en blocs. Pour chaque facteur bloc B dans le diagramme de Hasse (sauf U):

- si B est directement emboîté dans un seul facteur A : permutations aléatoires et indépendantes des niveaux de B à l'intérieur de chaque niveau de A ,
- si B est directement emboîté dans plusieurs facteurs, pas de permutations spécifiques à effectuer pour B .

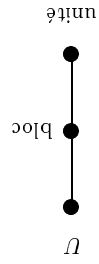
- On permute aléatoirement les *niveaux du facteur traitement* : cette étape est facultative et à effectuer avec précaution si les traitements sont structurés.

Ces règles de randomisation ne modifient pas le coefficient d'efficacité du plan.

La randomisation des traitements est facultative. Par exemple dans les essais conduits en réseau, la randomisation des traitements évite qu'un déséquilibre dans la répartition des traitements entre les blocs soit répété d'un essai à l'autre.

Randomisation d'un plan en blocs

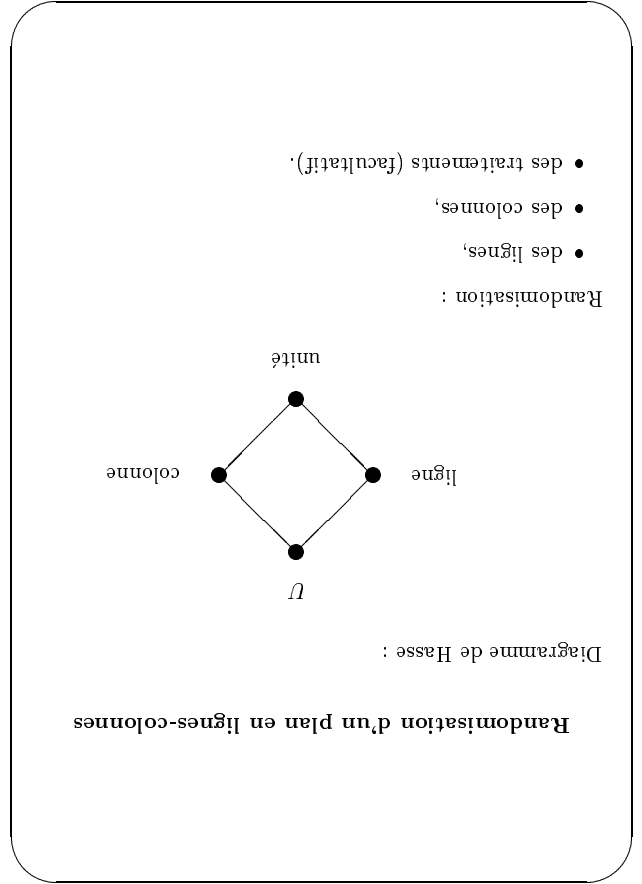
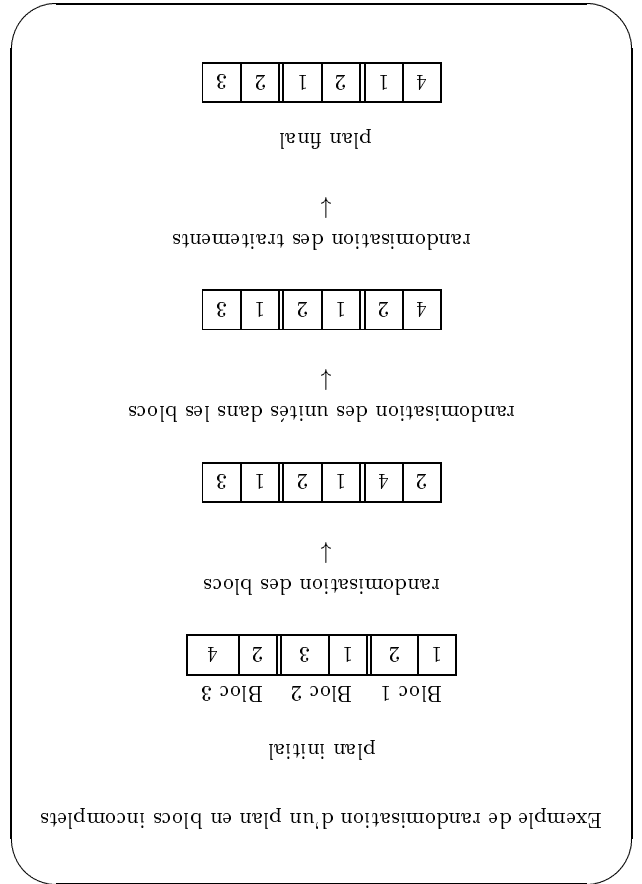
Diagramme de Hasse :



Randomisation :

- des blocs,
- des unités dans les blocs,
- des traitements (facultatif).

Pour un plan en blocs complets, il est inutile de randomiser les blocs et les traitements.



Le facteur unité est le croisement des facteurs ligne et colonne. Il n'y a qu'une unité à l'intersection d'une ligne et d'une colonne, donc il n'y pas de randomisation à effectuer au niveau des unités.

Exemple de randomisation d'un plan en lignes-colonnes

plan initial

1	2	3
1	2	3
1	2	3

↑
randomisation des lignes

1	2	3
3	1	2
2	3	1

↑
randomisation des colonnes

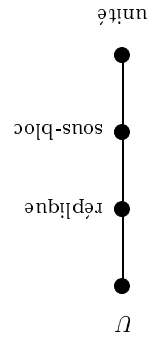
plan final

3	1	2
3	1	2
2	3	1

Dans cet exemple, la randomisation des traitements, qui est facultative, n'a pas été effectuée.

Randomisation d'un plan en répétitions

Diagramme de Hasse :

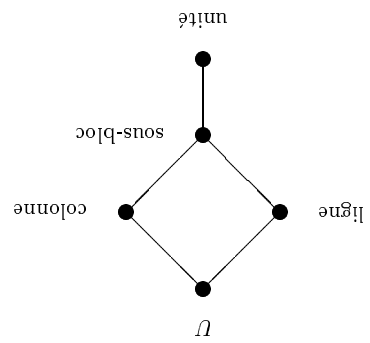


Randomisation :

- des répétitions,
- des sous-blocs dans les répétitions,
- des unités dans les sous-blocs,
- des traitements (facultatif).

Randomisation d'un carré semi-latin

Diagramme de Hasse :



Exercice : écrire les étapes de la randomisation d'un carré semi-latin.

6. Logiciels

CycDesign

- Construction :
 - de plans en blocs : plans cycliques ou construction algorithmique,
 - de plans en lignes-colonnes : plans cycliques ou construction algorithmique,
 - de plans en répliques : alpha-plans ou construction algorithmique,
 - de carrés semi-latins et de plans latinisés : alpha-plans latinisés ou construction algorithmique.
- Permet de randomiser les plans.

Corrigé de l'exercice
Randomisation :

- des lignes,
- des colonnes,
- des unités dans les sous-blocs,
- des traitements (facultatif).

Il n'y a pas de randomisation des sous-blocs, car le facteur sous-bloc est à l'intersection de ligne et de colonne.

CycDesign est un logiciel interactif sous Windows, conçu par des statisticiens australiens.

```

Block design
Resolvable
Design parameters
Number of treatments = 30
Number of units/block = 5
Number of replicates = 3
Alpha design
Random number seed for design generation = 990791192
Average efficiency factors (upper bounds)
Block 0.784346 (0.78553)
Concurrence 0
Block 180
Randomization 1
Random number seed for randomization = 990791395
Treatment randomization:
Group 1:
30 10 11 12 8 13 15 16 17 4 2 28 14 5 29 7
21 22 23 24 25 26 1 27 4 2 28 14 5 29 7
Replicate randomization:
1 3 2
Block randomization:
3 5 2 6 4 1
1 4 5 6 3 2
2 4 3 5 6 1

```

Sortie de CycDesign pour la construction d'un alpha-plan à 30 variétés, 3 répétitions et 6 sous-blocs de taille 5 par répétition. Le logiciel donne les paramètres du plan, le coefficient d'efficacité du plan retenu avec une borne supérieure pour ce coefficient, les concomitances (concurrency en anglais), des informations sur la randomisation et le plan.

```

Blocks of the design printed in rows
rep 1 -----
plot 1 2 3 4 5
block +-----+
1 | 5 1 10 18 8
2 | 21 28 2 12 11
3 | 7 30 25 17 29
4 | 15 3 6 23 13
5 | 19 24 14 20 9
6 | 4 16 26 22 27
rep 2 -----
plot 1 2 3 4 5
block +-----+
1 | 20 12 6 22 30
2 | 26 8 2 24 17
3 | 9 18 28 29 3
4 | 1 11 19 15 4
5 | 14 10 7 16 23
6 | 5 21 27 25 13
rep 3 -----
plot 1 2 3 4 5
block +-----+
1 | 3 7 24 12 1
2 | 21 15 9 30 16
3 | 13 20 28 4 8
4 | 11 23 22 17 5
5 | 14 26 25 6 18
6 | 2 29 10 27 19

```

ALPHA+

- Construction :
- d'alpha-plans,
- d'alpha-plans latinisés.

- Permet de randomiser les plans.

OPTEX (SAS)

Construction algorithmique :

- de plans en blocs,
- de plans adaptés à une matrice de variance-covariance des observations.

```

data t1;
do treatmt = 1 to 7;
output;
end;
proc optex data=t1 coding=orth;
generate criterion=a;
class treatmt;
model treatmt;
blocks structure=(7)3;
run;
output out=des1gn1;
proc print data=des1gn1;
run;

```

The SAS System

Design Number	D-efficiency	A-efficiency	G-efficiency	Standard Error
1	97.9821	96.0784	90.7485	0.6547
2	97.9821	96.0784	90.7485	0.6547
3	97.9821	96.0784	90.7485	0.6547
4	97.9821	96.0784	90.7485	0.6547
5	97.9821	96.0784	90.7485	0.6547
6	97.9821	96.0784	90.7485	0.6547
7	97.9821	96.0784	90.7485	0.6547
8	97.9821	96.0784	90.7485	0.6547
9	97.9821	96.0784	90.7485	0.6547
10	97.9821	96.0784	90.7485	0.6547

Instructions SAS pour construire un plan en blocs incomplets à 7 traitements et 7 blocs de taille 3. Le premier tableau de la sortie peut être ignoré.

ALPHA+ est un logiciel sous Windows conçu par des statisticiens écossais et australiens. ALPHA+ et CycDesign ont été conçus à l'origine par les inventeurs des alpha-plans, en particulier H. D. Patterson et E. R. Williams.

ALPHA+ permet de construire des plans en répliques et des plans latinisés basés sur la méthode des alpha-plans. L'utilisation et les sorties d'ALPHA+ sont proches de celles de CycDesign.

OPTEX est une procédure du module QC (Quality Control) de SAS.

Il existe d'autres logiciels ou commandes de logiciels qui permettent de construire des plans d'expériences, mais ils sont plutôt destinés à la construction de plans factoriels.

OPTEX effectuée 10 recherches et donne le meilleur plan obtenu après ces recherches. La colonne "Treatment A-efficiency" donne le coefficient d'efficacité. Le plan obtenu est un plan en blocs incomplets équilibré.
 G. Douaire (ENSA de Rennes) a développé un logiciel spécialisé dans la construction de PBIF.

The SAS System

Design Number	Treatment	Block Design
1	1	3
2	2	1
3	3	5
4	4	7
5	5	2
6	6	2
7	7	3
8	8	3
9	9	6
10	10	6
11	11	4
12	12	4
13	13	5
14	14	5
15	15	5
16	16	6
17	17	6
18	18	6
19	19	7
20	20	7
21	21	7

7. Conclusion

- Le plan est choisi pour optimiser l'analyse.
- Dans un plan orthogonal, les estimations des effets traitements n'ont pas à être ajustées pour les effets blocs.
- Répéter les traitements permet d'améliorer la précision des estimateurs des effets traitements et d'avoir des degrés de liberté résiduels.
- La répartition des traitements entre les blocs est générée en maximisant le coefficient d'efficacité du dispositif.
- Des plans optimaux ou efficaces peuvent être construits par des méthodes théoriques ou algorithmiques.
- Si on attache la même importance à tous les traitements, les plans efficaces sont en général équilibrés, binaires et équilibreront le mieux possible les concomitances entre traitements.

Abréviations : Caract. = caractéristiques, Optim. = optimalité, Pl. = plan, Bl. = bloc, compl. = complet, tjs = toujours, Concom. = concomitance, prfs = parfois, l. = ligne, c. = colonne, Car. = carré, rép. = répétitive, s. = semi, latin. = latinisé.

- La randomisation permet d'éviter les biais sur les estimations des effets traitements et sur l'estimation de la variance des estimateurs des effets traitements.
- A chaque structure en blocs correspond une procédure de randomisation appropriée.
- Principaux plans

Type de plan	Plan	Caract.	Optim.
Pl. en blocs	PBC	Bl. compl.	tjs
	PBE	Concom.	tjs
		égales	
		Pl. cycliques	prfs
Pl. en l.-c.	Car. latins	L. et c. compl.	tjs
	Car. de Youden	L. compl.	tjs
		PBE en c.	
Pl. en rép.	Latines car.	$v = k^2$	tjs
		α -pl.	prfs
		Car. s.-latins	α -pl. latin.
			prfs

D. ANALYSE D'UN DISPOSITIF EN BLOCS AVEC LE MODELE À EFFETS BLOCS ALÉATOIRES

Plan

1. Exemple
2. Effets blocs aléatoires
3. Informations intra-bloc et inter-bloc
4. Estimation
5. Conséquences de l'utilisation d'un modèle mixte sur la planification
6. Logiciels
7. Conclusion

Cette partie présente l'analyse avec le modèle à effets blocs aléatoires de manière simplifiée. Elle n'aborde pas en particulier :
 – la prédiction d'effets aléatoires,
 – les tests adaptés aux modèles mixtes.
 Plan de cette partie : présentation d'un exemple puis de la théorie de l'analyse.

Précisions sur l'exemple.

- Il s'agit de comparer des modèles et des marques de buses de pulvérisation, en conditions contrôlées dans une installation pilote. Pour chaque test, cinq buses d'un modèle donné sont fixées à une rampe sous laquelle sont réparties 10 gobelets. On met le circuit sous pression pendant 1m, puis on mesure le volume d'eau dans chaque gobelet.
- L'installation était disponible pendant trois demi-journées. Ceci permettait de réaliser une trentaine d'essais, ce qui rendait impossible la répétition de toutes les buses. De plus, pour contrôler de possibles dérives au cours du temps, les séries de cinq essais consécutifs ont été considérées comme des blocs.
- Le dispositif utilisé est un plan en blocs incomplets avec 6 blocs de taille 5. Les buses ne sont expérimentées qu'une fois, sauf la buse 15 qui se trouve dans tous les blocs et la buse 12 qui se trouve dans cinq blocs. Répéter les buses 12 et 15 a l'avantage de rendre le plan connexe. Cependant, ce plan n'est pas optimal, car il répartit les répétitions de façon très déséquilibrée entre les différentes buses.

Comme pour le modèle à effets blocs fixes, l'analyse comprend 4 étapes :

- modélisation,
- tests,
- estimation,
- validation.

Pour simplifier, nous ne présentons pas de transparents sur la validation pour l'exemple.

1. Exemple

- Expérience réalisée par l'ITCF en 1999.
- Comparaison de 21 modèles de buse pour leur aptitude à répartir de façon homogène la solution pulvérisée.
- La variable mesurée est le coefficient de variation des masses d'eau recueillies sur 10 gobelets.
- Un bloc correspond à une série de cinq essais consécutifs.

Plan de l'expérience

Bloc 1	5	12	13	15	4
Bloc 2	14	19	12	15	6
Bloc 3	2	18	12	15	7
Bloc 4	16	3	15	8	12
Bloc 5	17	15	20	9	1
Bloc 6	12	11	10	21	15

Données

buse	1	7.435000	cv
12	1	6.623333	
13	1	3.945000	
15	1	9.711667	
4	1	10.130000	
14	2	6.456667	
19	2	5.558333	
12	2	7.473333	
15	2	10.121667	
6	2	8.450000	
2	3	6.371667	
18	3	5.021667	
12	3	6.953333	
15	3	9.721667	
7	3	9.335000	
16	4	4.266667	
3	4	8.508333	
15	4	10.118333	
8	4	5.433333	
12	4	6.776667	
17	5	7.226667	
15	5	9.880000	
20	5	8.065000	
9	5	7.535000	
1	5	10.723333	
12	6	6.640000	
11	6	6.230000	
10	6	11.915000	
21	6	9.646667	
15	6	10.075000	

Notation : cv = coefficient de variation.

La modélisation des effets blocs par des effets aléatoires sera justifiée plus loin.

Modèle d'analyse

$$y_i = \text{buse} + \text{bloc}$$

$$y_{ij(t)} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

avec :

$y_{ij(t)}$ réponse sur l'unité j du bloc i ,
 t buse appliquée à cette unité,

μ moyenne générale,

τ_i effet fixe de la buse t_i ,

β_j effet aléatoire du bloc i ,

ε_{ij} erreur résiduelle.

Les β_j et ε_{ij} sont supposés indépendants avec :

- $\beta_j \sim \mathcal{N}(0, \sigma_\beta^2)$

- $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$

Le modèle comprend des effets fixes et des effets aléatoires \Rightarrow **modèle mixte**.

Les paramètres σ_β^2 et σ_ε^2 s'appellent les **composantes de la variance**.

Analyse de variance

Terme	ddl	Somme Carré	F	Proba
Buse	20	108,20	5,41	91,07
Bloc	5	0,44	0,09	1,48
Résiduelle	4	0,24	0,06	
Carrés Moyens				

Estimation des composantes de la variance

Bloc	Résiduelle
0,02	0,05

Estimation des moyennes des buses

Estimations

Bus	PCA	Fixe	Mixte	Bus	PCA	Fixe	Mixte
13	3.9	4.2	4.0	5	7.4	7.7	7.5
16	4.3	4.2	4.3	6	8.4	8.1	8.3
18	5.0	5.1	5.1	20	8.1	8.1	8.1
19	5.6	5.2	5.4	3	8.5	8.5	8.5
8	5.4	5.4	5.4	7	9.3	9.4	9.4
14	6.5	6.1	6.3	21	9.6	9.7	9.7
11	6.2	6.3	6.3	15	9.9	9.9	9.9
2	6.4	6.4	6.4	4	10.1	10.4	10.2
12	6.9	6.9	6.9	1	10.7	10.8	10.7
17	7.2	7.3	7.2	10	11.9	12.0	11.9
9	7.5	7.6	7.6				

Différence Mixte - Fixe maximum en valeur absolue : 0.2

L'analyse en modèle mixte a peu modifiée le classement des buses par rapport à l'analyse avec le modèle à effets blocs fixes ...

Precision des estimations

etd	etd	max	min	max
PCA	0.17	0.37	0.39	
Fixe	0.15	0.41	0.46	
Mixte	0.14	0.34	0.37	

... mais a quand même permis de gagner en précision.

2. Effets blocs aléatoires

Pourquoi modéliser les effets blocs par des effets aléatoires ?

- Randomisation des blocs ⇒ effets blocs aléatoires équilibrés.
- On fait l'hypothèse que ces effets blocs suivent une loi normale.

Exemple : modèle mixte pour un plan en blocs

$$y = \text{traitement} + \text{bloc}$$

- fixe : traitement
- aléatoire : bloc

$$y_{ij(t)} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

avec :
 $y_{ij(t)}$: réponse sur l'unité j du bloc i ,
 t traitement appliqué à cette unité,
 μ moyenne générale,
 τ_i effet fixe du traitement i ,
 β_j effet aléatoire du bloc j , $\beta_j \sim \mathcal{N}(0, \sigma_{\beta}^2)$,
 ε_{ij} erreur résiduelle, $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_{\varepsilon}^2)$.

Les β_j et ε_{ij} sont supposés indépendants.

Les effets blocs peuvent être modélisés par des effets aléatoires car la randomisation rend aléatoires les effets blocs qui sont exercés sur les réponses d'un traitement. Pour une même expérience randomisée, les deux modélisations des effets blocs, en fixes ou en aléatoires, sont possibles.

Notations : b = nombre de blocs, N = nombre d'unités expérimentales.

Corrigé de l'exercice

$$\text{Cov}(y_{ij(i)}, y_{i'j'(i)}) = \begin{cases} \sigma_\beta^2 + \sigma_\varepsilon^2 & \text{si } i = i', j = j' \\ \sigma_\beta^2 & \text{si } i = i', j \neq j' \\ 0 & \text{sinon} \end{cases}$$

Dans le modèle à effets blocs fixes, la variance des observations est égale à σ_ε^2 et les observations sont supposées indépendantes. Dans le modèle mixte par contre, on suppose que la variance des observations est égale à $\sigma_\beta^2 + \sigma_\varepsilon^2$ et que la covariance entre deux observations est égale à σ_β^2 pour deux observations d'un même bloc et à 0 pour deux observations de deux blocs différents.

Écrire éventuellement la matrice de variance-covariance de y au tableau.

Remarque

On peut retrouver les expressions de l'espérance et de la variance de y à l'aide des formules :

$$E(Az) = A E(z), \text{Var}(Az) = A \text{Var}(z) A'$$

où A est une matrice fixe et z est un vecteur aléatoire.

Exemple : modèle mixte pour un plan en lignes-colonnes

“ y = traitement + ligne + colonne”

- fixe : traitement
- aléatoire : ligne, colonne

$$y_{ij(i)} = \mu + \tau_i + \lambda_j + \omega_j + \varepsilon_{ij}$$

avec :

$y_{ij(i)}$ réponse sur l'unité située en ligne i et colonne j ,
 t traitement appliqué à cette unité,
 μ moyenne générale,
 τ_i effet fixe du traitement t_i ,
 λ_j effet aléatoire de la ligne j , $\lambda_j \sim \mathcal{N}(0, \sigma_\lambda^2)$,
 ω_j effet aléatoire de la colonne j , $\omega_j \sim \mathcal{N}(0, \sigma_\omega^2)$,
 ε_{ij} erreur résiduelle, $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$.
 Les variables aléatoires λ_i, ω_j et ε_{ij} sont supposées indépendantes.

Exercice : compléter les valeurs de $\text{Cov}(y_{ij(i)}, y_{i'j'(i)})$.

$$\text{Cov}(y_{ij(i)}, y_{i'j'(i)}) = \text{Cov}(\beta_i + \varepsilon_{ij}, \beta_{i'} + \varepsilon_{i'j'}) = \begin{cases} \text{si } i = i', j = j' \\ \text{si } i = i', j \neq j' \\ \text{sinon} \end{cases}$$

Modèle sous forme matricielle :

$$y = \Pi \mu + X \tau + Z \beta + \varepsilon$$

$$\beta \sim \mathcal{N}(0, \sigma_\beta^2 I_b)$$

$$\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2 I_N)$$

β et ε sont supposés indépendants.

$$E(y) = \Pi \mu + X \tau$$

$$\text{Var}(y) = V = \sigma_\beta^2 Z Z' + \sigma_\varepsilon^2 I_N$$

3. Informations intra-bloc et inter-bloc

Exemple : plan en blocs incomplets

Plan :

1	2	3	1	2	3
Bloc 1		Bloc 2		Bloc 3	

Modèle :

$$y = \mu + X\tau + Z\beta + \varepsilon, \text{ Var}(y) = \sigma_\tau^2 Z'Z' + \sigma_\beta^2 I_6$$

On a :

$$V = (2\sigma_\tau^2 + \sigma_\beta^2) O_1 O_1' + (2\sigma_\tau^2 + \sigma_\beta^2) O_2 O_2' + \sigma_\beta^2 O_3 O_3'$$

- $O_1 = \mathbb{I} / \sqrt{6}$

- les colonnes de O_2 sont des vecteurs ortho-normalisés de $\text{Im}(Z)$ orthogonaux à \mathbb{I} :

$$O_2' = \begin{pmatrix} 1 & 1 & 1 & 1 & -1 & -1 \\ \frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} & 0 & 0 \end{pmatrix} / 2$$

- les colonnes de O_3 sont des vecteurs ortho-normalisés orthogonaux à $\text{Im}(Z)$:

$$O_3' = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} / \sqrt{2}$$

Le terme intra-bloc vient du fait que l'information intra-bloc repose sur des contrastes sur les observations d'un même bloc.

Les variables aléatoires $\beta_{11(1)} / \sqrt{2}, \beta_{12(2)} / \sqrt{2}, \beta_{21(3)} / \sqrt{2}$ et $\beta_{22(3)} / \sqrt{2}$ ont les propriétés suivantes :

– leur espérance est une combinaison linéaire des paramètres, – elles sont non corrélées et de même variance (σ_ε^2) :

$$\text{Var}(O_3 y) = O_3' V O_3 = \sigma_\varepsilon^2 I_3$$

Elles vérifient donc les propriétés d'un modèle linéaire et on peut leur appliquer des techniques de modèles linéaires.

L'estimateur intra-bloc de $\tau_1 - \tau_2$ est obtenu à l'aide de la méthode des moindres carrés.

- Information intra-bloc : information apportée par $O_3 y$:

$$\frac{\hat{y}_{11(1)} - \hat{y}_{12(2)}}{\varepsilon_{11} - \varepsilon_{12}} = \frac{\hat{y}_{21(3)} - \hat{y}_{22(3)}}{\varepsilon_{21} - \varepsilon_{22}} = \frac{\hat{y}_{31(2)} - \hat{y}_{32(2)}}{\varepsilon_{31} - \varepsilon_{32}}$$

L'information intra-bloc est basée sur les 3 différences intra-bloc → élimination des effets blocs.

Ces équations vérifient les propriétés d'un modèle linéaire.

Variance : σ_ε^2

Nombre de degrés de liberté total : $N - b = 3$

- Estimateur intra-bloc de $\tau_1 - \tau_2$:

= estimateur de $\tau_1 - \tau_2$ obtenu avec l'information intra-bloc :

$$\delta_{\text{intra}} = (2(\hat{y}_{11(1)} - \hat{y}_{12(2)}) + \hat{y}_{22(1)} + \hat{y}_{32(3)} - \hat{y}_{31(2)}) / 3 = \delta_{\text{fixe}}$$

$$E(\delta_{\text{intra}}) = \tau_1 - \tau_2, \text{ Var}(\delta_{\text{intra}}) = \frac{4}{3} \sigma_\varepsilon^2$$

Même information sur $\tau_1 - \tau_2$ qu'avec le modèle à effets blocs fixes.

• Information inter-bloc :

= information apportée par $O^2 y$:

$$y_{1\bullet} - y_{2\bullet} = \frac{\tau_2 - \tau_3}{2} + \beta_1 - \beta_2 + \varepsilon_{1\bullet} - \varepsilon_{2\bullet}$$

$$y_{1\bullet} + y_{2\bullet} - 2y_{3\bullet} = \frac{2\tau_1 - \tau_2 - \tau_3}{2} + \beta_1 + \beta_2 - 2\beta_3 + \frac{\sqrt{3}}{2} \frac{\varepsilon_{1\bullet} + \varepsilon_{2\bullet} - 2\varepsilon_{3\bullet}}{\sqrt{3}}$$

L'information inter-bloc est basée sur 2 contrastes sur les moyennes par blocs.

Ces équations vérifient les propriétés d'un modèle linéaire.

Variance : $k\sigma_\beta^2 + \sigma_\varepsilon^2 = 2\sigma_\beta^2 + \sigma_\varepsilon^2$

Nombre de degrés de liberté total : $b - 1 = 2$

• Estimateur inter-bloc de $\tau_1 - \tau_2$:

= estimateur de $\tau_1 - \tau_2$ obtenu avec l'information inter-bloc :

$$\delta_{inter} = 2(y_{2\bullet} - y_{3\bullet})$$

$$= \tau_1 - \tau_2 + 2(\beta_2 - \beta_3) + 2(\varepsilon_{2\bullet} - \varepsilon_{3\bullet})$$

Exercice : compléter l'espérance et la variance de δ_{inter} .

$$E(\delta_{inter}) =$$

$$\text{Var}(\delta_{inter}) =$$

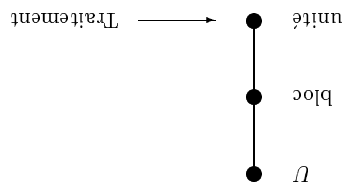
Représentation à l'aide du diagramme de Hasse

ddl Variance

Dans le diagramme de Hasse, l'information intra-bloc est associée au facteur unité et l'information inter-bloc est associée au facteur bloc. Il existe également une information associée au facteur U , qui correspond à la somme des observations. Cette information est souvent ignorée, car elle ne permet d'estimer que la moyenne générale et n'apporte pas d'information sur les différences entre effets traitements. Les flèches du diagramme indiquent que l'information intra-bloc et l'information inter-bloc contiennent de l'information sur les effets traitements.

La variance de l'information inter-bloc est plus élevée que celle de l'information intra-bloc. L'information inter-bloc est donc moins précise que l'information intra-bloc, d'autant moins que σ_β^2 est élevée.

Exemple : plan en blocs complets



Plan en blocs complets : toute l'information sur les effets traitements est dans l'information intra-bloc et est utilisée par l'analyse avec le modèle à effets blocs fixes. Plan en blocs incomplets : une partie de l'information sur les effets traitements se trouve dans l'information inter-bloc et n'est pas utilisée par l'analyse avec le modèle à effets blocs fixes.

Généralisation

- Modèle :

$$y = \mathbb{1}\mu + X\tau + \sum_{p \neq U} Z_p \beta_p$$

$$\text{Var}(y) = V = \sum_{p \neq U} \sigma_p^2 Z_p Z_p'$$

avec :

P facteur bloc,
 Z_p matrice d'incidence des blocs de P ,
 β_p vecteur des effets aléatoires des blocs de P de variance σ_p^2 .

- La matrice V peut s'écrire sous la forme :

$$V = \sum_{p \neq U} \zeta_p O_p O_p'$$

où les colonnes des O_p sont des vecteurs orthormés de l'espace des observations \mathbb{R}^N .

- La strate P est le sous-espace de \mathbb{R}^N engendré par les colonnes de O_p .

Une strate pour chaque facteur bloc (donc pour chaque point du diagramme de Hasse).

Dans un plan en blocs complets, les constatés sur les moyennes par blocs n'apportent pas d'information sur les effets traitements : l'information inter-bloc n'apporte pas d'information sur les effets traitements.

La structure en blocs du dispositif doit vérifier certaines propriétés pour que l'on puisse effectuer une décomposition en strates. Nous ne détaillons pas ces propriétés. Les structures en blocs présentes dans ce module pour lesquelles chaque facteur bloc a des blocs de même taille vérifient ces propriétés.

Remarque

La formule :

$$V = \sum_{p \neq U} \zeta_p O_p O_p'$$

peut être interprétée comme une diagonalisation de V .

L'information de la strate P vérifie les propriétés d'un modèle linéaire :

- $E(O_p y) = O_p \Pi \mu + O_p X \tau$: l'espérance des composantes de $O_p y$ est une combinaison linéaire des paramètres,
- $\text{Var}(O_p y) = O_p V O_p = \xi P I_{\text{dof}} P$, avec dof nombre de degrés de liberté de la strate P : les composantes de $O_p y$ sont non corrélées et de même variance (ξP).

On peut donc appliquer à cette information des techniques de modèles linéaires (estimation avec la méthode des moindres carrés, test F, \dots).

Lorsque des blocs sont complets, les contrastes sur les moyennes par blocs n'apportent pas d'information sur les effets traitements.

Remarques

Géométriquement, $O_p y$ peut être interprété comme la projection de y sur la strate P . Il existe des méthodes relatives simples pour calculer les matrices de projection sur les strates. Nous ne les présentons pas dans ce module.

- L'information de la strate P est l'information apportée par $O_p y$.

$O_p y$ vérifie les propriétés d'un modèle linéaire.

ξP est la variance dans la strate P .

- L'information de la strate P utilise des contrastes sur les moyennes par blocs de P .

- Si G est emboîté dans P , l'information de la strate G utilise des contrastes intra-blocs de P .

- L'information d'une strate est (généralement)

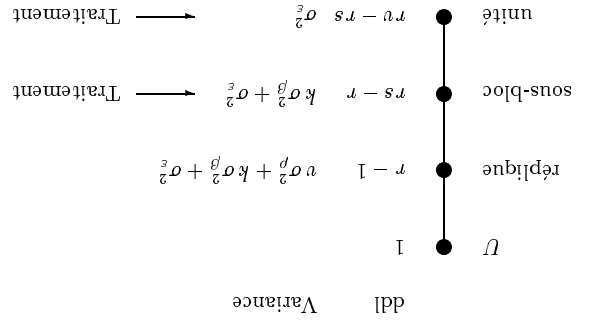
d'autant plus précise que la strate est basse dans le diagramme de Hasse.

- Si un facteur bloc est complet équilibré pour les traitements, la strate correspondante n'apporte pas d'information sur les effets traitements.

- Dans un dispositif en blocs incomplets, l'information sur les effets traitements est répartie entre plusieurs strates.

- La strate intra-bloc apporte la même information sur les effets traitements que l'analyse avec le modèle à effets blocs fixes.

Strates d'un plan en répliques



Compléments pour les formateurs :

- Variance dans la strate P :

$$\xi_P = \sum_{G \succ P} k_G \sigma_G^2$$

avec :

k_G taille des blocs de G ,

$G \succ P$ signifie que G est emboîté dans P ou est égal à P .

- Nombre de degrés de liberté de la strate P :

$$\text{dof}_P = n_P - \sum_{G \succ P} \text{dof}_G$$

avec :

n_P nombre de niveaux de P ,

$G \succ P$ signifie que G est emboîté dans G .

- Les informations des strates ne sont pas corrélées.

Comment les informations des strates sont-elles utilisées par les différentes méthodes d'analyse ?

- L'analyse avec le modèle à effets blocs fixes n'utilise que l'information intra-bloc (strate unité).
- Dans une **analyse de variance en strates** (ou **multi-strates**), on analyse *séparément* les informations des strates à l'aide de techniques de modèles linéaires.
- Nous allons voir dans la partie suivante que l'analyse avec le modèle à effets blocs aléatoires *combine* les informations des strates.

4. Estimation

Exemple

Bloc 1	Bloc 2	Bloc 3
1	2	3
2	1	2
3	3	1

- Plan :
- Notation : $\delta = \tau_1 - \tau_2$
- Estimateur de $\tau_1 - \tau_2$ basé sur des moyennes brutes : $\delta_{\text{brut}} = (\hat{y}_{11(1)} + \hat{y}_{22(1)})/2 - (\hat{y}_{12(2)} + \hat{y}_{31(2)})/2$
- Estimateur intra-bloc de $\tau_1 - \tau_2$:

$$\begin{aligned} \delta_{\text{intra}} &= \delta_{\text{hxe}} \\ &= 2(\hat{y}_{11(1)} - \hat{y}_{12(2)}) + \\ &= (\hat{y}_{22(1)} - \hat{y}_{21(3)} + \hat{y}_{32(3)} - \hat{y}_{31(2)})/3 \\ E(\delta_{\text{intra}}) &= \tau_1 - \tau_2 \\ \text{Var}(\delta_{\text{intra}}) &= \frac{3}{4} \sigma_\varepsilon^2 \end{aligned}$$

L'estimateur $\hat{\delta}_{mixte}$ est une combinaison linéaire de $\hat{\delta}_{intra}$ et $\hat{\delta}_{inter}$, dont les poids sont proportionnels aux inverses des variances de $\hat{\delta}_{intra}$ et $\hat{\delta}_{inter}$. Il utilise à la fois l'information intra-bloc et l'information inter-bloc, si bien qu'il est plus précis que $\hat{\delta}_{intra}$ et $\hat{\delta}_{inter}$. On dit que l'analyse en modèle mixte récupère l'information inter-bloc. Dans cet exemple, $\hat{\delta}_{mixte}$ s'exprime comme une combinaison linéaire de $\hat{\delta}_{intra}$ et $\hat{\delta}_{inter}$, mais en général son expression est plus compliquée.

Le calcul de $\hat{\delta}_{mixte}$ nécessite des estimations des composantes de la variance. Il est donc important d'avoir des estimations précises de ces paramètres pour bien comparer les traitements. Lorsque les effets blocs sont importants ($\sigma_{\beta}^2 \rightarrow +\infty$), les estimations des effets traitements sont ajustées pour les effets blocs ; lorsque les effets blocs sont faibles ($\sigma_{\beta}^2 \rightarrow 0$), les estimations des effets traitements sont proches de celles qui auraient été obtenues avec une analyse sans effets blocs. Un plan en blocs incomplets, analysé avec le modèle à effets blocs aléatoires, assure donc une protection contre les effets blocs à moindre coût. De plus, l'analyse en modèle mixte évite dans une certaine mesure d'avoir à sélectionner le modèle (éventuelle suppression des effets blocs). Par ailleurs, on constate qu'on récupère d'autant plus d'information inter-bloc que les effets blocs sont faibles.

Bloc 1	Bloc 2	Bloc 3
1	2	3
2	1	2
3	3	1

• Plan :

• Estimateur inter-bloc de $\tau_1 - \tau_2$:

$$\hat{\delta}_{inter} = 2(y_{2\bullet} - y_{3\bullet})$$

$$E(\hat{\delta}_{inter}) = \tau_1 - \tau_2$$

$$\text{Var}(\hat{\delta}_{inter}) = 4(2\sigma_{\beta}^2 + \sigma_{\epsilon}^2)$$

• Estimateur de $\tau_1 - \tau_2$ obtenu avec le modèle mixte :

$$\hat{\delta}_{mixte} = \frac{w_{intra}\hat{\delta}_{intra} + w_{inter}\hat{\delta}_{inter}}{w_{intra} + w_{inter}}$$

avec

$$w_{intra} = 1/(4\sigma_{\beta}^2/3)$$

$$w_{inter} = 1/(8\sigma_{\beta}^2 + 4\sigma_{\epsilon}^2)$$

$$E(\hat{\delta}_{mixte}) = \tau_1 - \tau_2$$

$$\text{Var}(\hat{\delta}_{mixte}) = \frac{6\sigma_{\beta}^2 + 3\sigma_{\epsilon}^2}{4\sigma_{\beta}^2} \times \frac{6\sigma_{\beta}^2 + 4\sigma_{\epsilon}^2}{3}$$

$$\text{Si } \sigma_{\beta}^2 \rightarrow +\infty, \hat{\delta}_{mixte} \rightarrow \hat{\delta}_{intra}$$

$$\text{Si } \sigma_{\beta}^2 \rightarrow 0, \hat{\delta}_{mixte} \rightarrow \hat{\delta}_{brut}$$

Généralisation

Modèle :

$$y = \mathbb{1}\mu + X\tau + \sum_{p \neq u} Z_p \beta_p$$

$$\text{Var}(y) = V = \sum_{p \neq u} \sigma_p^2 Z_p Z_p'$$

avec :

P Facteur bloc,

Z_p matrice d'incidence des blocs de P ,

β_p vecteur des effets aléatoires des blocs de P de

variance σ_p^2 .

1. Estimation des composantes de la variance

Méthode d'estimation : méthode du maximum de vraisemblance restreint (REML) :

• la vraisemblance restreinte est la vraisemblance

de $Q_X y$,

$$Q_X = I_N - X(X'X)^{-1}X'$$

• on estime les σ_p^2 par maximisation de la

vraisemblance restreinte (en général à l'aide de méthodes numériques)

$$\rightarrow V = \sum_{p \neq u} \sigma_p^2 Z_p Z_p'$$

Après l'exemple, l'estimation des paramètres du modèle à effets blocs aléatoires est présentée de manière générale. L'estimation comprend deux étapes. La première est l'estimation des composantes de la variance. La méthode REML est la méthode la plus utilisée, mais d'autres méthodes existent. Le terme "restreint" vient du fait qu'on utilise une partie restreinte des données, celle qui est orthogonale aux effets fixes (Q_X s'interprète comme le projecteur sur le sous-espace orthogonal aux colonnes de X). Les équations à résoudre pour maximiser la vraisemblance restreinte sont en général trop compliquées pour qu'on puisse obtenir des solutions analytiques : on doit utiliser des méthodes numériques.

La deuxième étape de l'estimation est l'estimation des effets fixes, en fixant les composantes de la variance à leur valeur estimée dans la première étape.
 La formule de l'estimateur de c^T pour le modèle mixte ressemble à celle vue pour le modèle à effets blocs fixes, mais Q_Z est remplacé par V^{-1} .
 Si V était connue, on aurait

$$\text{Var}(\widehat{c^T}) = c^T(X^T V^{-1} X)^{-1} c.$$

Comme \widehat{V} est une estimation de V , la formule ci-dessus est approchée ; elle sous-estime la variance de $\widehat{c^T}$, de façon sensible surtout pour les petits dispositifs.

Méthode d'estimation : méthode du maximum de vraisemblance ou méthode des moindres carrés généralisés.

2. Estimation des effets fixes

Si un contraste c^T est estimable :

$$\widehat{c^T} = c^T(X^T V^{-1} X)^{-1} X^T V^{-1} y$$

$$E(\widehat{c^T}) = c^T$$

$$\text{Var}(\widehat{c^T}) \simeq c^T(X^T V^{-1} X)^{-1} c$$

Matrice d'information sur τ pour le modèle mixte :

$$M^V = X^T V^{-1} X$$

Lorsque les blocs sont complets, les différences entre moyennes par bloc n'apportent pas d'information sur les effets traités ; il n'y a donc pas d'information inter-bloc à récupérer. Dans l'analyse d'un plan en répétitions, les effets des répétitions sont souvent modélisés par des effets fixes car les répétitions sont complètes et sont en général peu nombreuses. De même dans l'analyse d'un carré semi-latin, les effets des lignes et des colonnes sont souvent modélisés par des effets fixes.

- On peut modéliser les effets d'un facteur bloc soit par des effets fixes, soit par des effets aléatoires.
- Blocs complets : pas d'information inter-bloc sur les effets traitements \Rightarrow pas de différence entre les deux modélisations des effets blocs pour l'estimation des effets traitements.
- Blocs incomplets :
 - modélisation des effets blocs par des effets aléatoires \Rightarrow récupération de l'information inter-bloc \Rightarrow les estimations des effets traitements sont généralement plus précises,
 - récupération d'information inter-bloc d'autant plus élevée que la variance des effets blocs est faible,
 - dans quelques cas, la modélisation des effets blocs par des effets fixes est préférable : nombre de blocs trop faible pour avoir une estimation fiable de la variance des effets blocs.

5. Conséquences de l'utilisation d'un modèle mixte sur la planification

a. Sur le choix entre un plan en blocs complets et un plan en répétitions

- Si l'hétérogénéité des unités est importante : un plan en répétitions permet souvent de gagner en précision par rapport à un plan en blocs complets.
 - Si l'hétérogénéité des unités est modérée : les estimations des effets sous-blocs aléatoires : les estimations des effets traitements sont proches de celles qui auraient été obtenues avec une analyse en plan en blocs complets.
- Avec le modèle à effets blocs aléatoires, il y a moins de risques à utiliser des blocs incomplets.

Les plans optimaux pour le modèle à effets blocs aléatoires peuvent dépendre des composantes de la variance. Comme ces paramètres sont en général inconnus au moment de planifier, il est difficile de construire des plans optimaux. On planifie alors pour que :

- l'information sur les traitements se trouve dans les plus basses, car ces strates sont les plus précises ; par exemple dans un plan en répétitions, comme les répétitions sont complètes, l'information sur les traitements se trouve dans des strates situées au dessous de la strate répétitive, car l'information intra-bloc soit la plus précise possible, car c'est la strate intra-bloc qui contient le plus d'information sur les effets traitements ; par exemple un plan en répétitions est construit en maximisant le coefficient d'efficacité, donc en maximisant la précision de l'information intra-bloc.

b. Sur la recherche de plans optimaux

- Matrice d'information $= X'V^{-1}X$
 - ⇒ Le plan optimal peut dépendre des composantes de la variance
 - On obtient des plans efficaces en :
 - plaçant l'information sur les traitements dans les strates les plus basses,
 - optimisant l'information intra-bloc.
- ⇒ Objectif commun de la planification en effets blocs fixes ou aléatoires : optimiser le coefficient d'efficacité intra-bloc.

Instructions et sorties S-PLUS de l'analyse avec le modèle à effets blocs aléatoires de l'expérience de comparaison des buses de pulvérisation de l'ITCF.

La commande `is.random` permet de déclarer les facteurs à effets aléatoires. La fonction `varcomp` permet d'effectuer l'analyse en modèle mixte. La fonction `lme` (qui n'a pas été utilisée dans l'analyse présentée) permet également d'effectuer des analyses avec un modèle mixte.

6. Logiciels

S-PLUS

```
> testcv.dat <- read.table("testcv.txt", header=T)
> testcv.dat$buse <- factor(testcv.dat$buse)
> testcv.dat$bloc <- factor(testcv.dat$bloc)
> is.random(testcv.dat$bloc) <- T
> aov.pbl <- varcomp(cv ~ buse + bloc,
data = testcv.dat, method = "reml")
> round(aov.pbl$variances, 2)
Bloc Residuals
0.02
> select.pbl <- aov.pbl$assign.fixed$buse
> contr.pbl <- aov.pbl$contrasts$buse
```

La commande `contr.pbl` `coef(aov.pbl)[select.pbl]` donne les estimations des effets centrés des buses. Le vecteur `aov.pbl$beta` contient les prédictions des effets blocs.

```
> round(contr.pbl %*% coef(aov.pbl)[select.pbl], 2)
[1,
1 3.16
2 -1.17
3 0.92
4 2.66
5 -0.04
6 0.71
7 1.79
8 -2.16
9 -0.03
10 4.36
11 -1.32
12 -0.69
13 -3.53
14 -1.28
15 2.36
16 -3.32
17 -0.33
18 -2.52
19 -2.18
20 0.50
21 2.09
> sed.calc(aov.pbl, "buse")
means
Variance moyenne des comparaisons 2 à 2: 0.120323003479821
sed.min sed.moyenne sed.max
0.1432984 0.3447957 0.3689909
> round(aov.pbl$beta, 2)
[1] -0.10 0.16 -0.03 0.01 -0.02 -0.02
> motif()
> hist(aov.pbl$residuals)
> box()
```

```

data testcv;
infile 'testcv.txt';
input buse bloc cv;
run;

proc mixed data=testcv;
class buse bloc;
model cv = buse;
random bloc / solution;
lsmeans buse;
run;

The SAS System

The MIXED Procedure

Class Level Information

Class      Levels Values
-----
BUSE      21  1 2 3 4 5 6 7 8 9 10 11 12 13
BLOC      6  1 2 3 4 5 6

RML Estimation Iteration History

Iteration  Evaluations  Objective  Criterion
-----
0          1 -10.88322801
1          3 -11.24710622  0.00025005
2          1 -11.24855706  0.00000061
3          1 -11.24856052  0.00000000

Convergence criteria met.
    
```

Instructions et sorties SAS de l'analyse avec le modèle à effets blocs aléatoires de l'expérience de comparaison des buses de pulvérisation de l'ITCF.

L'analyse en modèle mixte s'effectue avec la procédure mixed. L'option solution de random imprime les prédictions des effets blocs. Les options diff ou pdiff de lsmeans (qui n'ont pas été utilisées ici) permettent d'estimer les différences entre effets buses et les écarts-types des différences.

```

Covariance Parameter Estimates (RML)

Cov Parm      Estimate
-----
BLOC          0.01999878
Residual      0.05467180

Model Fitting Information for CV

Description      Value
-----
Observations      30.0000
Res Log Likelihood -2.6462
Akaike's Information Criterion -4.6462
Schwarz's Bayesian Criterion -4.8434
-2 Res Log Likelihood 5.2923

Solution for Random Effects

Effect BLOC      Estimate      SE Pred      DF      Pr > |t|
-----
BLOC 1          -0.10420457    0.11428339    1.61    -0.91
BLOC 2          -0.16196880    0.11428339    1.61    1.42
BLOC 3          -0.03238001    0.11428339    1.61    -0.28
BLOC 4          0.01409471    0.11428339    1.61    0.12
BLOC 5          -0.01554886    0.12464057    1.17    -0.12
BLOC 6          -0.02393006    0.11428339    1.61    -0.21
Tests of Fixed Effects

Source      MDF      DDF      Type III F      Pr > F
-----
BUSE      20      4      103.18      0.0002
    
```

Effect	BUSE	LSMEAN	Std Error	DF	t	Pr > t
BUSE 1	10.73888186	0.26496617	8.67	40.53	0.0001	40.53
BUSE 2	6.40404701	0.26025467	7.85	24.61	0.0001	24.61
BUSE 3	8.49423829	0.26025467	7.85	32.64	0.0001	32.64
BUSE 4	10.23420457	0.26025467	7.85	39.32	0.0001	39.32
BUSE 5	7.53920457	0.26025467	7.85	28.97	0.0001	28.97
BUSE 6	8.28803120	0.26025467	7.85	31.85	0.0001	31.85
BUSE 7	9.36738001	0.26025467	7.85	35.99	0.0001	35.99
BUSE 8	5.41923829	0.26025467	7.85	20.82	0.0001	20.82
BUSE 9	7.55054886	0.26496617	8.67	28.50	0.0001	28.50
BUSE 10	11.93893006	0.26025467	7.85	45.87	0.0001	45.87
BUSE 11	6.25393006	0.26025467	7.85	24.03	0.0001	24.03
BUSE 12	6.89022343	0.12147253	8.81	56.72	0.0001	56.72
BUSE 13	4.04920457	0.26025467	7.85	15.56	0.0001	15.56
BUSE 14	6.29469820	0.26025467	7.85	24.19	0.0001	24.19
BUSE 15	9.93805667	0.11155759	8.56	89.08	0.0001	89.08
BUSE 16	4.25257229	0.26025467	7.85	16.34	0.0001	16.34
BUSE 17	7.24221586	0.26496617	8.67	27.33	0.0001	27.33
BUSE 18	5.05404701	0.26025467	7.85	19.42	0.0001	19.42
BUSE 19	5.39636420	0.26025467	7.85	20.73	0.0001	20.73
BUSE 20	8.08054886	0.26496617	8.67	30.50	0.0001	30.50
BUSE 21	9.67059706	0.26025467	7.85	37.16	0.0001	37.16

7. Conclusion

- Dans un dispositif en blocs, l'information sur les effets traitements peut être répartie entre plusieurs strates.
- L'analyse avec le modèle à effets blocs fixes n'utilise que l'information intra-bloc.
- L'analyse avec le modèle à effets blocs aléatoires d'analyse conseillée pour les dispositifs en blocs incomplets, sauf pour de petits dispositifs.

II. PLANS FACTORIELS EN BLOCS

Plan

1. Introduction
2. Plans factoriels en blocs complets
3. Confusion totale d'effets principaux avec des effets blocs
4. Confusion totale d'interactions avec des effets blocs
5. Confusion partielle entre effets traitements et effets blocs
6. Logiciels
7. Conclusion

Cette partie ne présente qu'une petite partie des plans factoriels. Le but est essentiellement de montrer ce qui peut changer dans la prise en compte des blocs pour la répartition des traitements, lorsque les traitements sont en fait les combinaisons des niveaux de plusieurs facteurs. Voir le module FPSTAT 4 et Droesbeke, Fine et Saporita (1997) pour plus d'informations sur les plans factoriels.

1. Introduction

Un **plan factoriel** est un dispositif dans lequel on étudie *simultanément* plusieurs facteurs traitements.

Pourquoi étudier plusieurs facteurs dans la même expérience ?

☞ *Economie de temps et de moyens* : répondre à plusieurs questions simultanément.

☞ *Généralité des résultats* :

- chaque facteur est étudié dans une diversité contrôlée de situations,
- des interactions entre facteurs peuvent être détectées.

Aspect séquentiel : Les manips avec beaucoup de facteurs pour identifier les plus pertinents ; manips suivantes avec les facteurs les plus pertinents étudiés plus précisément (par exemple sur plus de sites-années en agronomie).

Avec un facteur quantitatif, il faut au moins $p+1$ niveaux distincts pour ajuster un polynôme de degré p . Par exemple pour estimer les paramètres d'une droite (polynôme de degré 1), il faut au moins deux points de mesure distincts.

Comment choisir les facteurs ?

- ☞ Pré-sélection : lister les facteurs pouvant avoir une influence sur le phénomène étudié.
- ☞ Si possible les ordonner par ordre d'intérêt décroissant.
- ☞ Conserver un nombre de facteurs compatible avec les ressources expérimentales.

Comment choisir les niveaux des facteurs ?

- ☞ *Facteur qualitatif* : sélectionner/échantillonner des modalités représentatives de la diversité du facteur.
- ☞ *Facteur quantitatif* : au moins $p+1$ niveaux distincts nécessaires pour ajuster un polynôme de degré p .

⇒ *Compromis à trouver entre nombre de facteurs/nombre de niveaux par facteur/nombre d'unités expérimentales*

Cet exemple est repris du module 2/3 de FPSTAT. On dispose de 3 origines géographiques. On croise des plantes de chaque population avec 2 variétés appelées testeur 1 et testeur 2, pour connaître la valeur en croisement de chaque population. Le caractère étudié, la teneur en huile, est mesuré sur les hybrides issus de chaque croisement. Les deux facteurs sont croisés (on ne considérera pas le cas de facteurs traitements emboîtés dans ce module).

Exemple à deux facteurs croisés (sans facteur bloc)

Comparaison de populations de tournesol
– 2 testeurs (facteur A)
– 3 origines géographiques (facteur B)
– 12 parcelles, plan complètement aléatoire

Teneur en huile

Origine		par testeur		
Afrique	Hongrie	Maroc	Moyennes	
(1)	(2)	(3)	Moyennes	par testeur
43.54	44.25	47.28	45.39	47.17
45.30	42.55	49.40	45.73	47.47
44.42	43.40	48.34	45.42	48.48
47.21	44.34	47.75	47.73	46.28
Moyennes	45.95	44.41	48.48	46.28

Traitements

On appelle **traitements** les combinaisons de niveaux des facteurs à étudier.

Dans l'exemple, $2 \times 3 = 6$ traitements :

Traitement	Niveau de A	Niveau de B	Effet
T11	1	1	T11
T12	1	2	T12
T13	1	3	T13
T21	2	1	T21
T22	2	2	T22
T23	2	3	T23

Le plan utilisé est un plan factoriel :

- **complet** : tous les traitements sont expérimentés,

- **équilibré** : les traitements sont expérimentés un

même nombre de fois.

Les traitements ont une **structure factorielle** :

	Moyennes	Origine 1	Origine 2	Origine 3	Moyennes
Testeur 1	T11	T12	T13	T1•	
Testeur 2	T21	T22	T23	T2•	
Moyennes	T•1	T•2	T•3	T••	origines

On veut donc structurer l'interprétation des résultats :

- ☞ quelles différences moyennes entre niveaux de A ?

- ☞ quelles différences moyennes entre niveaux de B ?

- ☞ y a-t-il une interaction entre A et B ?

Effets factoriels

Chaque effet traité peut être écrit sous la forme :

$$Tab = \mu + \alpha_a + \beta_b + \gamma_{ab}$$

avec :

→ $\mu = \tau_{\bullet\bullet}$: moyenne

→ $\alpha_a = \tau_{a\bullet} - \tau_{\bullet\bullet}$: effet du niveau a du facteur A

→ $\beta_b = \tau_{\bullet b} - \tau_{\bullet\bullet}$: effet du niveau b du facteur B

→ $\gamma_{ab} = \tau_{ab} - \tau_{a\bullet} - \tau_{\bullet b} + \tau_{\bullet\bullet}$: interaction entre les

niveaux a et b de A et B

Ces paramètres vérifient les contraintes :

$$\checkmark \sum_a \alpha_a = 0$$

$$\checkmark \sum_b \beta_b = 0$$

$$\checkmark \sum_a \gamma_{ab} = \sum_b \gamma_{ab} = 0$$

⇒ Trois effets factoriels :

- effet principal de A (effets α_a) : 1 degré de liberté,
- effet principal de B (effets β_b) : 2 degrés de liberté,
- interaction A,B (effets γ_{ab}) : 2 degrés de liberté.

L'effet principal d'un facteur correspond aux comparaisons entre les niveaux de ce facteur, en moyenne sur l'ensemble des niveaux des autres facteurs présents dans le dispositif. L'interaction entre deux facteurs A et B correspond aux écarts au modèle qui suppose l'additivité des effets de A et de ceux de B .

Changements dans les notations :

- b n'est plus le nombre de blocs, qui sera noté np dans la suite,
- β_b n'est plus un effet bloc, qui sera noté π_i dans la suite.

Remarque

Les paramètres estimés par les logiciels ne vérifient pas généralement les contraintes du transparent et sont donc difficilement interprétables. Sous S-PLUS, il faut utiliser la fonction `dummy.coef` ou `model.tables` pour récupérer des paramètres vérifiant ces contraintes. Sous SAS, on peut utiliser `estimate` ou plus simplement `lsmeans` qui donne les estimations des $\tau_{a\bullet}$, $\tau_{\bullet b}$, τ_{ab} .

Les modèles sont présentés pour le moment en l'absence de facteurs blocs.
 Le modèle factoriel complet est équivalent au modèle avec un paramètre par traitement :

$$y_{i(ab)} = \tau_{ab} + \varepsilon_i$$

que nous avons utilisé jusqu'à présent. Seule la paramétrisation est différente.
 Le modèle sous forme factorielle est spécifié différemment selon les logiciels statistiques :

$$\text{SAS model } y = A B A*B$$

$$\text{S-PLUS } y \sim A+B+A:B$$

Modélisations

Modèle factoriel complet :

$$"y = A + B + A.B"$$

Sur l'unité i recevant les niveaux a et b de A et B :

$$y_{i(ab)} = \mu + \alpha_a + \beta_b + \gamma_{ab} + \varepsilon_i$$

Modèles factoriels incomplets :

Certaines interactions sont supposées nulles

⇒ ce sont des **sous-modèles** du modèle complet

Exemple : *modèle factoriel additif*

(Hypothèse : absence d'interactions)

$$"y = A + B"$$

$$y_{i(ab)} = \mu + \alpha_a + \beta_b + \varepsilon_i$$

Analyse de variance

Terme	ddl	SC	CM	F	Proba
Testeur	1	9.49	9.49	6.21	0.05
Origine	2	33.75	16.87	11.04	0.01
Testeur.origine	2	3.95	1.97	1.29	0.34
Résiduelle	6	9.17	1.53		

Estimations

Effets testeurs :

T1	T2
-0.89	0.89

Effets origines :

Afrique	Hongrie	Maroc
-0.33	-1.86	2.20

Moyennes traitements :

Afrique	Hongrie	Maroc
44.42	43.40	48.34
T2	T1	
47.47	45.415	48.61

$$etd=1.24$$

$$etd=0.87$$

$$etd=0.71$$

La différence entre les variétés 1 et 2 est estimée par :

• dans le modèle $y =$ Variété :

$$(6 + 10 + 11 + 12 + 19 + 15 + 18)/7 - (13 + 15 + 31)/3$$

• dans le modèle $y =$ Variété + Sol :

$$w_1[(6 + 10 + 11)/3 - (13 + 15)/2] + w_2[(12 + 19 + 15 + 18)/4 - 31]$$

où w_1 et w_2 sont des pondérations de somme 1 ;

• dans le modèle complet :

$$0.5[(6 + 10 + 11)/3 - (13 + 15)/2] + 0.5[(12 + 19 + 15 + 18)/4 - 31]$$

$y_{i(ab)}$	Variété 1	Variété 2	Variété 3
Sol 1	6	10	11
Sol 2	12	19	15
Sol 3	13	15	31
Sol 4	14	22	18
Sol 5	18	9	12

Pour un plan factoriel complet mais non équilibré, les formules des estimateurs sont plus complexes car les estimations des effets factoriels doivent être ajustées pour les autres effets. De plus elles dépendent des termes retenus dans le modèle. FPSTAT, module 2/3, propose l'exemple suivant, sur l'étude du nombre de jours avant germination pour des variétés de carottes.

L'estimation des effets est très simple pour un plan factoriel complet et équilibré : les valeurs données dans le transparent s'obtiennent à partir des moyennes par niveau de A ou de B .

Notations : SC = somme des carrés, CM = carré moyen, etd = écart-type des différences.

Dans le modèle complet, on teste d'abord l'interaction, qui est ici non significative. Les deux effets principaux sont significatifs.

Les coefficients d'efficacité factoriels représentent l'application de la décomposition factorielle aux coefficients d'efficacité. Ils permettent de juger plus finement de la qualité d'un dispositif quand le modèle est factoriel.

Critères de comparaison entre dispositifs

Pour comparer des plans factoriels :

☞ le coefficient d'efficacité global e ne suffit pas,

☞ on utilise les **coefficients d'efficacité factoriels**.

Dans un dispositif en blocs (information intra-bloc) :

$$\widehat{\text{Var}}(c^T) = \|e\|_2^2 \sigma^2 / (r e(c))$$

avec $e(c)$ coefficient d'efficacité pour le contraste c^T

vérifiant $0 \leq e(c) \leq 1$.

Définitions

• Un plan factoriel est **factoriellement équilibré** si

$e(c)$ a la même valeur pour tous les contrastes c^T d'un même effet factoriel T .

• Cette valeur, qui dépend uniquement du plan, est

appelée **coefficient d'efficacité de l'effet**

factoriel T et notée e_T .

Par convention, si au moins un degré de liberté de l'effet

factoriel T est non estimable dans la strate unité, on

pose $e_T = 0$.

2. Plans factoriels en blocs complets

Définition

Plan factoriel complet :

• m facteurs traitements A, B, \dots à n_A, n_B, \dots , niveaux

$\Rightarrow v = n_A \times n_B \times \dots$ traitements

• chaque traitement est répété r fois

$\Rightarrow rv$ unités expérimentales

en blocs complets :

• un facteur bloc noté P à n_P niveaux

• chaque bloc contient chaque traitement un même nombre de fois (une fois en général et dans la suite)

Orthogonalité

Exemple :

- plan factoriel en blocs complets
- 2 facteurs traitements A et B à 2 niveaux
- niveaux notés 1, 2
- traitements notés ab

Plan :

11 12 21 22	11 12 21 22
Bloc 1	Bloc 2

Modèle :

$$y_{ij(ab)} = \mu + \alpha_a + \beta_b + \gamma_{ab} + \pi_i + \varepsilon_{ij}$$

$$y = A + B + A.B + P$$

Modèle sous forme matricielle (les contraintes sur les paramètres sont prises en compte) :

$$y = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \mu + \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \\ -1 \\ -1 \\ 1 \\ 1 \end{pmatrix} \alpha_1 + \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \\ 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} \beta_1 + \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} \gamma_1 + \begin{pmatrix} -1 \\ -1 \\ -1 \\ -1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \pi_1 + \varepsilon$$

Plan :

11 12 21 22	11 12 21 22
Bloc 1	Bloc 2

Modèle sous forme matricielle (les contraintes sur les paramètres sont prises en compte) :

$$y = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \mu + \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \\ -1 \\ -1 \\ 1 \\ 1 \end{pmatrix} \alpha_1 + \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \\ 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} \beta_1 + \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} \gamma_1 + \begin{pmatrix} -1 \\ -1 \\ -1 \\ -1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \pi_1 + \varepsilon$$

Les effets factoriels des traitements et les effets blocs sont deux à deux orthogonaux \Rightarrow le dispositif est dit **orthogonal**.

Remarque
 Il existe aussi des dispositifs orthogonaux complets mais non équilibrés, évoqués dans FPSTAT, module 2/3. Deux facteurs croisés A et B sont orthogonaux si $r_{ab} = r_{a+}r_{+b}/r_{++}$, où :

- r_{ab} est le nombre de répétitions du traitement de niveaux a et b de A et B,

- $r_{a+} = \sum_b r_{ab}$,

- $r_{+b} = \sum_a r_{ab}$,

- $r_{++} = \sum_{ab} r_{ab}$.

Conséquences de l'orthogonalité

Analyse de variance

Pour chaque effet factoriel, le carré moyen ne dépend pas des autres effets factoriels retenus dans le modèle (sous SAS : type I = type II = type III).

Exemple tournesol :

Terme	ddl	SC	CM	F	Proba
Bloc	1	3,60	3,60	3,23	0,13
Testeur	1	9,49	9,49	8,52	0,03
Origine	2	33,75	16,87	15,15	0,01
Testeur.origine	2	3,95	1,97	1,77	0,26
Résiduelle	5	5,57	1,11		

Nombre de degrés de liberté résiduels pour le modèle complet :

- plan complètement aléatoire : $(r - 1)v$
- plan en blocs complets : $(r - 1)(v - 1)$

Estimation

- Les effets factoriels sont estimés à partir de moyennes brutes, sans avoir besoin d'ajuster pour d'autres termes du modèle.
- L'estimation d'un contraste associé à un terme du modèle ne dépend pas des autres termes retenus.
- La précision de cette estimation ne dépend pas des autres termes retenus.
- Les estimations de contrastes associés à des termes distincts sont non corrélées (indépendantes sous l'hypothèse de normalité des erreurs).

Exemple tournesol :

$$\alpha_1 = y_{(1\bullet)} - y_{(\bullet\bullet)} = 45,39 - 46,28 = -0,89$$

$$\gamma_{12} = y_{(12)} - y_{(1\bullet)} - y_{(\bullet 2)} + y_{(\bullet\bullet)}$$

$$= 43,40 - 45,39 - 44,41 + 46,28 = -0,12$$

L'interprétation des résultats d'un plan factoriel en blocs complets est simple et très intuitive. Les propriétés mentionnées ne sont plus vérifiées si l'on rajoute dans le modèle des covariables ou d'autres termes non compris dans les effets des facteurs.

Remarque

Pour que les formules soient justes, d_1 et d_2 doivent être des vecteurs contrastes, et d_3 doit être un vecteur qui vérifie les mêmes contraintes que γ .

Exemple tournesol :

$$\begin{aligned} \text{Var}(\widehat{d_1 \alpha}) &= \|d_1\|_2^2 \sigma^2 / r_A \\ \text{Var}(\widehat{d_2 \beta}) &= \|d_2\|_2^2 \sigma^2 / r_B \\ \text{Var}(\widehat{d_3 \gamma}) &= \|d_3\|_2^2 \sigma^2 / r \end{aligned}$$

avec :

$r_A = n_B r = 6$ nombre de répétitions de chaque niveau de A
 $r_B = n_A r = 4$ nombre de répétitions de chaque niveau de B

De plus :

$$\begin{aligned} \text{Cov}(\widehat{\alpha_a}, \widehat{\beta_b}) &= 0 \\ \text{Cov}(\widehat{\alpha_a}, \widehat{\gamma_{a'v'}}) &= 0 \\ \text{Cov}(\widehat{\beta_b}, \widehat{\gamma_{a'v'}}) &= 0 \end{aligned}$$

Coefficients d'efficacité factoriels :

$$e_A = e_B = e_{A,B} = 1$$

Commentaires sur les inconvénients :

- si le nombre de traitements est élevé, les blocs complets sont nécessairement de grande taille ;
- dans la partie qui suit, on aborde justement les situations où les niveaux de certains facteurs ne varient qu'entre blocs ;
- un exemple est la comparaison d'inévitables techniques, que l'on pourrait aborder en faisant varier des facteurs densité de semis, doses de produits divers, dates d'apport, etc. ; cette approche est inutilement coûteuse si seule une petite partie des combinaisons est réellement pertinente.

Avantages et inconvénients des plans factoriels en blocs complets

Avantages

- Simplicité d'interprétation
- Indépendance des estimateurs (sous l'hypothèse de normalité) associée à des effets factoriels distincts
- Tous les effets factoriels des traitements sont estimés dans la strate intra-bloc avec un coefficient d'efficacité égal à 1.

Inconvénients

- Faible contrôle de l'hétérogénéité entre unités si le nombre de traitements est élevé
- Parfois, difficultés à faire varier les niveaux de certains facteurs entre unités d'un même bloc
- Parfois, certains traitements sont inintéressants, ou impossibles à réaliser.

3. Confusion totale d'effets principaux avec des effets blocs

Confusion

Exemple : plan en blocs avec 2 traitements et 4 blocs de taille 2.

1	1	2	2
B1	B2	B3	B4

Modèle sous forme matricielle (les contraintes sur les paramètres sont prises en compte) :

$$y = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \tau + \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \\ -1 \\ -1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \varepsilon$$

Le sous-espace associé aux effets traitements est inclus dans le sous-espace associé aux effets blocs :

- les effets traitements sont dits **totalément confondus** avec les effets blocs,

- le facteur traitement est dit **totalément confondu** avec le facteur bloc.

Pourquoi confondre totalement un effet principal F avec des effets blocs ?

- Difficulté/impossibilité techniques à faire varier les modalités de F à l'intérieur des blocs
- Moindre intérêt pour les comparaisons des modalités de F
- Intérêt principal sur les effets principaux des autres facteurs et sur les interactions

Trois exemples dans la suite :

1. cas le plus simple : un seul facteur bloc, avec un facteur traitement confondu
2. dispositif à parcelles divisées (split-plot)
3. dispositif à bandes croisées (criss-cross)

$$\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}$$

Le sous-espace associé aux effets traitements est inclus dans le sous-espace associé aux effets blocs car :

Plan avec un facteur bloc emboîté dans un facteur traitement

Essai variété

- Facteur bloc P : $n_P = 6$ blocs de $k = 9$ unités
- Facteur irrigation A à $n_A = 2$ modalités : Irrigué/Non irrigué
- Facteur variété B à $n_B = k = 9$ modalités présentes une fois par bloc
- Contrainte : irriguer toutes les parcelles d'un même bloc simultanément
- Chaque traitement est répété $r = n_P/n_A = 3$ fois.

Plan initial, avant randomisation :

I	1	2	3	4	5	6	7	8	9
I	1	2	3	4	5	6	7	8	9
I	1	2	3	4	5	6	7	8	9
NI	1	2	3	4	5	6	7	8	9
NI	1	2	3	4	5	6	7	8	9
NI	1	2	3	4	5	6	7	8	9

Le facteur irrigation est totalement confondu avec le facteur bloc.

Randomisation

1. Blocs entre eux

I	1	2	3	4	5	6	7	8	9
NI	1	2	3	4	5	6	7	8	9
NI	1	2	3	4	5	6	7	8	9
NI	1	2	3	4	5	6	7	8	9
I	1	2	3	4	5	6	7	8	9

2. Parcelles dans les blocs

I	3	6	1	4	7	8	9	5	2
NI	6	2	5	4	1	8	9	3	7
NI	1	6	7	8	9	4	3	2	5
I	5	4	8	7	6	1	9	2	3
NI	2	1	8	4	7	5	3	6	9
I	5	9	1	8	7	6	3	4	2

Les blocs sont complètement randomisés, sans contrainte sur la répartition des niveaux de A entre les blocs. Si l'on impose que les deux niveaux de A soient présents une fois chacun dans les paires de blocs successifs, on considère implicitement un facteur réplique dont les modalités sont constituées de ces paires de blocs. Le dispositif est alors un split-plot, que nous verrons plus loin.

En restreignant de plus la randomisation des variétés dans les blocs d'une certaine façon, on obtient un criss-cross. Comme ces variétés s'illustrent, la structure en blocs que nous présentons également plus loin.

Comme ces variétés s'illustrent, la structure en blocs qu'il faut utiliser dans l'analyse dépend des contraintes imposées à la randomisation.

Modèle

$$y = A + B + A.B + P$$

avec :

$$y_{ij}(ab) = \mu + \alpha_a + \beta_b + \gamma_{ab} + \pi_i + \varepsilon_{ij}$$

$y_{ij}(ab)$ réponse sur la parcelle j du bloc i
 a et b niveaux de A et B sur cette parcelle
 π_i effet aléatoire du bloc i , de variance σ_π^2

Propriétés

- Plan factoriel complet pour 2 facteurs A et B , équilibré ($r = 3$)
- Le facteur B est *orthogonal* au facteur bloc.
- Le facteur A est *totalement confondu* avec le facteur bloc.
- L'interaction $A.B$ est *orthogonale* aux effets blocs.

Diagramme de Hasse

Les effets factoriels sont estimés dans des strates différentes, donc avec des niveaux de précision différents.

Les blocs sont complets pour le facteur B :

- les différences entre observations d'un même bloc apportent de l'information sur l'effet principal de $B \Rightarrow$ la strate unité contient de l'information sur l'effet principal de B ,
- les différences entre moyennes par bloc n'apportent pas d'information sur l'effet principal de $B \Rightarrow$ la strate bloc n'apporte pas d'information sur l'effet principal de B .

Le facteur A est totalement confondu avec le facteur bloc :

- les différences entre observations d'un même bloc n'apportent pas d'information sur l'effet principal de $A \Rightarrow$ la strate unité ne contient pas d'information sur l'effet principal de A ,
- les différences entre moyennes par blocs apportent de l'information sur l'effet principal de A .

Il est plus difficile de justifier pourquoi l'interaction $A.B$ est estimée dans la strate unité.

Analyse de variance en strates

Dans un tableau d'analyse de variance en strates (ou multi-strates) on effectue des tests F dans chaque strate.

Strate	Terme fact.	ddl	Carré	Moyen	F
Bloc		(5)	$n_A - 1 = 1$	CM_A	$\frac{CM_A}{CM_{Bloc}}$
A	Résid. Bloc	$n_A(r - 1) = 4$	CM_{Bloc}		
Parcelle		(48)	$n_B - 1 = 8$	CM_B	$\frac{CM_{B}}{CM_{Bloc}}$
B			$(n_A - 1)(n_B - 1) = 8$	$CM_{A,B}$	$\frac{CM_{A,B}}{CM_{Bloc}}$
A,B	Résid. Intra	$n_A(r - 1)(n_B - 1) = 32$	CM_{Intra}		

avec

$$E(CMR_{Bloc}) = 9\sigma_2^2 + \sigma_\varepsilon^2$$

$$E(CMR_{Intra}) = \sigma_\varepsilon^2$$

Le test F pour A est effectué avec un carré moyen résiduel différent de celui des tests F pour B et A,B .

Précision des estimations

Effet principal de A

$$\text{Var}(\widehat{d_1^i \alpha}) = \|d_1\|_2^2 (k\sigma_2^2 + \sigma_\varepsilon^2) / r_A$$

$$= \|d_1\|_2^2 (\sigma_2^2 + \sigma_\varepsilon^2 / k) / r$$

$$= e_A = 0$$

Effet principal de B

$$\text{Var}(\widehat{d_2^j \beta}) = \|d_2\|_2^2 \sigma_2^2 / r_B$$

$$= e_B = 1$$

Interaction A,B

$$\text{Var}(\widehat{d_3^k \gamma}) = \|d_3\|_2^2 \sigma_2^2 / r$$

$$= e_{A,B} = 1$$

avec, pour l'exemple : $k = 9, r = 3, r_A = 27, r_B = 6$

Notation : CMR = carré moyen résiduel.
 On peut faire des tests F dans chaque strate car l'information de chaque strate peut être analysée à l'aide de techniques de modèles linéaires.
 Il faut remarquer que l'effet principal de A est doublement pénalisé :

- il est estimé dans une strate de plus grande variabilité,
- le nombre de degrés de liberté de la résiduelle est très faible.

Chaque effet factoriel des traitements est estimable dans une strate \Rightarrow l'analyse en modèle mixte et l'analyse de variance en strates sont équivalentes pour l'estimation des effets factoriels des traitements.

On retrouve des expressions similaires à celles du plan factoriel en blocs complets. La seule différence est que la variance pour l'effet principal de A intègre la variance entre blocs.
 Plus généralement, le plan présente vérifie toutes les propriétés d'orthogonalité mentionnées pour le plan factoriel en blocs complets, mais avec la nécessité de prendre en compte des variances résiduelles différentes suivant les termes d'effets factoriels des traitements. Ce plan, comme le split-plot et le criss-cross présentés plus loin, est pour ces raisons considéré comme un dispositif orthogonal.

Principales conséquences

Par rapport à un plan en blocs complets pour les deux facteurs :

Pour la conduite de l'essai :

- plus grande simplicité, moindre coût

• moindre risque d'erreurs dans la réalisation de l'essai

Pour l'effet principal de A :

- effet totalement confondu avec les effets blocs

• estimation dans la strate inter-bloc

• variabilité résiduelle plus élevée

• moins de degrés de liberté résiduels

D'où

- tests moins puissants

• estimations moins précises

Pour l'effet principal de B et l'interaction A, B :

- effets orthogonaux aux effets blocs

• estimation dans la strate intra-bloc

- coefficient d'efficacité factoriel égal à 1

Dispositif à parcelles divisées (split-plot)

Essai d'inoculation sur haricot (Dagnelle, 1981)

- $r = 4$ répétitions (facteur R) de $s = 6$ sous-blocs (facteur P) de $k = 3$ parcelles

- Facteur variété A à $n_A = s = 6$ modalités

- Facteur fertilisation B à $n_B = k = 3$ modalités

présentes une fois par sous-bloc : témoin, inoculation de Rhizobium, apport d'azote minéral

- Le facteur variété A est totalement confondu avec le facteur sous-bloc :

- on privilégie l'étude de l'effet principal de B et de l'interaction A, B ,
- mise en place de l'essai plus facile.

- La variable mesurée est le rendement (kg de gousses fraîches par parcelle).

Nous utilisons des termes cohérents avec le reste du modèle. Dans la présentation traditionnelle du split-plot, ce que nous appelons réplique est appelé bloc ; ce que nous appelons sous-bloc est appelé grande parcelle ; et ce que nous appelons parcelle est appelé sous-parcelle.

Comparaison avec le cas précédent :

- cette fois-ci les niveaux da A sont groupés par le système de répétitions et non complètement randomisés entre sous-blocs ;
- pour le facteur confondu et à l'échelle des sous-blocs, on fait donc un "plan en blocs complets" plutôt qu'un "plan complètement aléatoire", et on contrôle donc mieux l'hétérogénéité entre sous-blocs ;
- l'effet principal du facteur confondu est donc estimé avec une meilleure précision, mais on perd quelques degrés de liberté pour estimer la variance résiduelle de la strate dans laquelle il est estimé.

Plan de l'essai

Rép.1	A	6	4	3	5	2	1	2	3	1	2	3	1
	B	2	3	1	2	3	1	2	3	1	2	3	1
	B	2	3	1	2	3	1	2	3	1	2	3	1
	A	5	6	3	2	1	1	2	3	1	2	3	2
	A	5	6	3	2	1	1	2	3	1	2	3	2
	B	3	2	1	3	2	1	3	2	1	3	2	1
	B	3	2	1	3	2	1	3	2	1	3	2	1
	A	3	4	5	1	2	1	2	3	1	2	3	2
	A	3	4	5	1	2	1	2	3	1	2	3	2
	B	3	2	1	3	2	1	3	2	1	3	2	1
	B	3	2	1	3	2	1	3	2	1	3	2	1
	A	4	5	6	3	2	1	4	5	6	3	2	1
	A	4	5	6	3	2	1	4	5	6	3	2	1
	B	3	1	2	3	1	2	3	1	2	3	1	2
	B	3	1	2	3	1	2	3	1	2	3	1	2

La randomisation est celle d'un plan en répétitions.

Modèle

$$y_{ijl(ab)} = \mu + \alpha_a + \beta_b + \gamma_{ab} + \rho_i + \pi_{ij} + \varepsilon_{ijl}$$

avec :

$y_{ijl(ab)}$ réponse sur la parcelle l du sous-bloc j de la répétition i

a et b niveaux de A et B sur cette parcelle

ρ_i effet aléatoire de la répétition i de variance σ_p^2

π_{ij} effet aléatoire du sous-bloc j de la répétition i , de variance σ_{π}^2

Diagramme de Hasse

Exercice : préciser quels effets factoriels des traitements sont estimables dans les strates.

Analyse de variance en strates

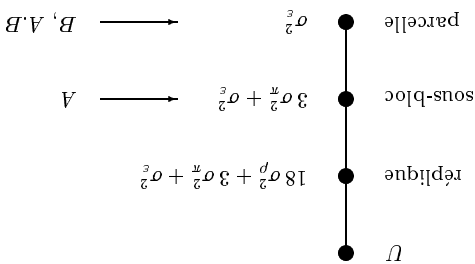
Strate	ddl	CM	F	P
Terme fact.				
Réplique	$r - 1 = 3$	5.8		
Sous-bloc	$n_A - 1 = 5$	29.6	23.2	0.00
A	$(r - 1)(n_A - 1) = 15$	1.3		
Parcelle	$n_B - 1 = 2$	5.3	15.0	0.00
B				
A, B	$(n_A - 1)(n_B - 1) = 10$	0.8	2.4	0.03
Res. Intra	$(r - 1)n_A(n_B - 1) = 36$	0.4		

avec $E(\text{CMR}_{\text{Sous-bloc}}) = 3\sigma_\pi^2 + \sigma_\epsilon^2$
 $E(\text{CMR}_{\text{Intra}}) = \sigma_\epsilon^2$

Coefficients d'efficacité factoriels

$e_A = 0, e_B = e_{A,B} = 1$

Corrigé de l'exercice



Notation : F = probabilité associée au test F .
 On pourra comparer avec le tableau d'analyse de variance du plan avec un facteur bloc emboîté dans un facteur traitement, pour mieux comprendre ce qui distingue ces dispositifs : la strate "bloc" a été divisée en deux dans le split-plot : la strate "réplique" et la strate "sous-bloc".
 Les effets répliques peuvent être supposés fixes ou aléatoires, sans que cela ne modifie le tableau d'analyse de variance, ni les estimations d'effets traitements, ni leurs précisions. En pratique, il est préférable de les considérer comme des effets fixes.

Remarque

Analyse des données brutes, sans la transformation logarithmique proposée par Dagnelie (1981).

Le criss-cross n'est pas un plan avec un facteur bloc emboîté dans un facteur traitement car les lignes sont groupées en répliques. Ce n'est pas un split-plot parce que, en plus, on ne randomise pas les parcelles dans les lignes indépendamment. Les trois types de plans correspondent donc à des systèmes de contraintes sur la randomisation différents. Le choix entre ces dispositifs, lorsqu'il est possible, se fait donc en tenant compte d'abord des contraintes de réalisation, ensuite de l'importance relative accordée aux effets principaux et aux interactions des facteurs.

Dispositif à bandes croisées (criss-cross)
Essai de comparaison entre variétés sous deux doses d'engrais

- $r = 3$ répliques (facteur R) de 2 lignes (facteur L) et 9 colonnes (facteur C)
- Facteur engrais A à $n_A = 2$ modalités (2 doses D1 et D2), totalement confondu avec le facteur ligne
- Facteur variété B à $n_B = 9$ modalités, totalement confondu avec le facteur colonne

Plan initial :

D1	1	2	3	4	5	6	7	8	9
D2	1	2	3	4	5	6	7	8	9
D1	1	2	3	4	5	6	7	8	9
D2	1	2	3	4	5	6	7	8	9
D1	1	2	3	4	5	6	7	8	9
D2	1	2	3	4	5	6	7	8	9

La structure en blocs est constituée du facteur réplique, qui emboîte les deux facteurs ligne et colonne (diagramme de Hasse deux transparents plus loin). La randomisation décrite ici est celle qui est associée à cette structure en blocs, selon les règles décrites dans la partie I.C. Par rapport au split-plot, c'est l'étape 3 qui est différente.

Randomisation
1. Répliques entre elles (sans effet)
2. Lignes dans les répliques :

D1	1	2	3	4	5	6	7	8	9
D2	1	2	3	4	5	6	7	8	9
D1	1	2	3	4	5	6	7	8	9
D2	1	2	3	4	5	6	7	8	9
D1	1	2	3	4	5	6	7	8	9
D2	1	2	3	4	5	6	7	8	9

3. Colonnes dans les répliques :

D1	8	2	6	3	4	5	7	9	1
D2	8	2	6	3	4	5	7	9	1
D1	3	5	4	6	8	1	9	2	7
D2	3	5	4	6	8	1	9	2	7
D1	3	9	1	5	2	7	6	4	8
D2	3	9	1	5	2	7	6	4	8

Modèle

$$y = A + B + A.B + R + R.L + R.C^m$$

$$y_{ijl(ab)} = \mu + \alpha_a + \beta_b + \gamma_{ab} + p_i + \lambda_{ij} + \omega_l + \varepsilon_{ijl}$$

avec

$y_{ijl(ab)}$ réponse dans la réplique l , en ligne j et colonne i

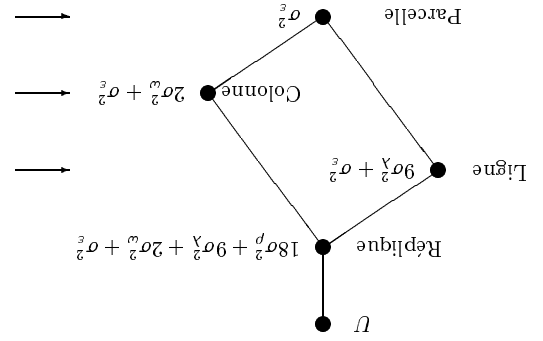
a et b niveaux de A et B sur cette parcelle

p_i effets répliques

λ_{ij} effets aléatoires des lignes

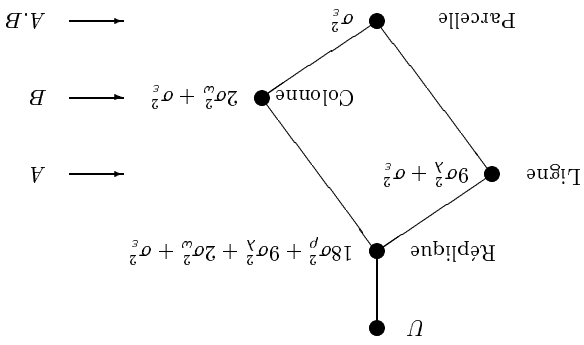
ω_l effets aléatoires des colonnes

Diagramme de Hasse



Exercice : préciser quels effets factoriels des traitements sont estimables dans les strates.

Corrigé de l'exercice



Les effets principaux de A et B sont estimés avec une précision plus faible que l'interaction $A.B$. Comme les split-plots, les criss-cross sont généralement utilisés à causes de contraintes pratiques.
Par rapport au split-plot, une partie de la strate Parcelle forme maintenant la strate Colonne.

Il n'y a que 2 degrés de liberté résiduels dans la strate ligne. C'est surtout le facteur confondu avec les lignes qui est pénalisé par ce criss-cross.

Analyse de variance en strates

Strate	ddl	
Terme factoriel		
Réplique	$r - 1 = 3$	
Résiduelle Réplique		
Ligne dans Réplique	$n_A - 1 = 1$	
Résiduelle Ligne	$(r - 1)(n_A - 1) = 2$	
Colonne dans Réplique	$n_B - 1 = 8$	
Résiduelle Colonne	$(r - 1)(n_B - 1) = 16$	
Parcelle		
A,B	$(n_A - 1)(n_B - 1) = 8$	
Résiduelle Intra	$(r - 1)(n_A - 1)(n_B - 1) = 16$	

Coefficients d'efficacité factoriels

$e_A = 0, e_B = 0, e_{A,B} = 1$

4. Confusion totale d'interactions avec des effets blocs

Confondre totalement des interactions avec des effets blocs :

→ ces interactions ne sont pas estimables intra-bloc ;
 → les autres effets factoriels des traitements sont orthogonaux aux effets blocs et sont estimables intra-bloc avec un coefficient d'efficacité factoriel égal à 1.

Pourquoi confondre des interactions ?

- On souhaite utiliser des blocs incomplets.
- Les interactions à confondre sont négligeables.
- On s'intéresse surtout aux effets principaux ou aux interactions d'ordres inférieurs.
- C'est une méthode de base pour construire des plans en blocs incomplets plus généraux.

On impose $1 \leq q \leq m - 1$ car sinon les blocs sont de taille supérieure ou égale au nombre de traitements et on peut donc faire des blocs complets.

Comment confondre totalement une interaction avec des effets blocs ?
 En confondant l'effet principal d'un **pseudofacteur** avec les effets blocs.
 Le pseudofacteur est défini à partir des facteurs qui composent l'interaction.
 Conditions nécessaires :

- les m facteurs traitements ont le même nombre de niveaux n ,
- les blocs sont de taille n^q , avec $1 \leq q \leq m - 1$,
- et n est un nombre premier (ou une puissance d'un nombre premier).

Présentation de ces méthodes par des exemples.

Exemple avec 3 facteurs à 2 niveaux

- 3 facteurs A, B, C à 2 niveaux \Rightarrow 8 traitements
 niveaux notés 0 et 1
 traitements notés abc (avec a niveau de A , etc.)
- 1 répétition, 2 blocs (facteur P) de taille 4

Tableau des traitements :

Traitement	Niveau de	Valeur de
abc	A	B
	C	$a + b + c \text{ mod } 2$
000	0	0
001	0	1
010	0	1
011	0	1
100	1	0
101	1	0
110	1	0
111	1	1

La valeur de $a + b + c \text{ mod } 2$ définit un nouveau facteur à 2 niveaux appelé **pseudofacteur $\overline{A.B.C}$** .

Modèle factoriel complet :

$$y = A + B + C + A.B + A.C + B.C + A.B.C + P$$

Correspondance entre l'interaction $A.B.C$ et l'effet principal du pseudofacteur $\overline{A.B.C}$

γ_{abc} : interaction entre les niveaux a, b, c de A, B, C

Contraintes :

$$\sum_a \gamma_{abc} = \sum_b \gamma_{abc} = \sum_c \gamma_{abc} = 0$$

$$\Rightarrow \gamma_{000} = \gamma_{011} = \gamma_{101} = \gamma_{110} = -\gamma_{001} = -\gamma_{010} = -\gamma_{100} = -\gamma_{111}$$

$\Rightarrow 1$ ddl pour l'interaction $A.B.C$

\Rightarrow Les γ_{abc} qui ont une même valeur de $a + b + c \pmod 2$ sont égaux.

\Rightarrow L'interaction $A.B.C$ peut être modélisée à l'aide de l'effet principal du pseudofacteur $\overline{A.B.C}$.

Répartition des traitements entre les blocs

Bloc 0	Bloc 1
$a + b + c = 0 \pmod 2$	$a + b + c = 1 \pmod 2$
000 011 101 110	001 010 100 111

Le pseudofacteur $\overline{A.B.C}$ est totalement confondu avec le facteur bloc :

\Rightarrow l'interaction $A.B.C$ est totalement confondu avec les effets blocs ;

\Rightarrow on écrit $P = \overline{A.B.C}$: relation de définition ou **d'alias** du plan ;

\Rightarrow l'interaction $A.B.C$ n'est pas estimable intra-bloc.

On montre que les autres effets factoriels :

$$A, B, C, A.B, A.C, B.C,$$

sont orthogonaux aux effets blocs et sont estimables

intra-bloc.

$$e_A = e_B = e_C = e_{A.B} = e_{A.C} = e_{B.C} = 1, e_{A.B.C} = 0$$

Le principe se généralise pour des répliques de k blocs, ou plus (mais une puissance de 2). Nous ne le détaillons pas ici.

met dans le même bloc tous les traitements qui ont la même valeur de $b + d \pmod 2$.

Si l'on veut confondre d'autres interactions avec les effets blocs, on utilise le même principe avec des équations différentes. Par exemple pour confondre l'interaction $B.C$, on

d'en confondre des différentes. Ce sera l'objet de la section suivante.

Avec plusieurs répliques, on a le choix entre confondre la même interaction dans toutes les répliques, ou au contraire

pour estimer la résiduelle sauf si l'on fait l'hypothèse, par exemple, que les interactions entre deux facteurs sont nulles. Avec une seule réplique, il n'y a pas de degré de liberté

Exemple avec 2 facteurs à 3 niveaux

- 2 facteurs A et B à 3 niveaux \Rightarrow 9 traitements
niveaux notés 0, 1, 2
traitements notés ab (avec a niveau de A , etc.)
- 1 répétition, 3 blocs (facteur F) de taille 3

Tableau des traitements :

Traitement	Niveau de	Facteur	Facteur	Facteur
Facteur	Niveau de	Facteur	Facteur	Facteur
Facteur	Niveau de	Facteur	Facteur	Facteur
00	0	0	0	0
01	0	1	1	2
02	0	2	2	1
10	1	0	1	1
11	1	1	1	2
12	1	2	2	0
20	2	0	0	2
21	2	1	1	0
22	2	2	2	1

Les valeurs de $a + b \pmod 3$ et $a + 2b \pmod 3$ définissent deux **pseudofacteurs** à 3 niveaux $\overline{A.B}$ et $\overline{A.B^2}$.

Modèle factoriel complet :

$$y = A + B + A.B + P$$

Correspondance entre effets factoriels et effets des pseudofacteurs

- Effet principal de A : 2 ddl
 \Rightarrow comparaisons entre moyennes par niveau de A
- Effet principal de B : 2 ddl
 \Rightarrow comparaisons entre moyennes par niveau de B
- Interaction $A.B$: 4 ddl qui peuvent se décomposer en :
 \Rightarrow comparaisons entre moyennes par niveau du pseudofacteur $\overline{A.B}$ (2 ddl)
 \Rightarrow comparaisons entre moyennes par niveau du pseudofacteur $\overline{A.B^2}$ (2 ddl)
- La décomposition pseudofactorielle est plus fine que celle en effets factoriels.

Répartition des traitements entre les blocs : deux possibilités

1. Confondre totalement le pseudofacteur $\overline{A.B}$ avec le facteur bloc ($P = \overline{A.B}$)

Bloc 0	Bloc 1	Bloc 2
$a + b = 0$	$a + b = 1$	$a + b = 2$
mod 3	mod 3	mod 3
00 12 21	01 10 22	

Exercice : compléter les traitements expérimentés dans le troisième bloc.

2. Confondre totalement le pseudofacteur $\overline{A.B^2}$ avec les blocs ($P = \overline{A.B^2}$)

Bloc 0	Bloc 1	Bloc 2
$a + 2b = 0$	$a + 2b = 1$	$a + 2b = 2$
mod 3	mod 3	mod 3
00 11 22	02 10 21	01 12 20

Dans chaque cas :

- 2 ddl de l'interaction sont totalement confondus avec les effets blocs ;
- les effets principaux et les 2 autres ddl de l'interaction sont orthogonaux aux effets blocs et sont estimables intra-bloc ;
- $e_A = e_B = 1, e_{A.B} = 0$.

Bloc 2	02 11 20
--------	----------

5. Confusion partielle entre effets traitements et effets blocs

Principe :

- Plan factoriel à plusieurs répétitions.
- Dans chaque répétition, on confond totalement des effets factoriels (ou pseudofactoriels) traitements avec les effets sous-blocs.
- Les effets traitements confondus sont différents d'une répétition à l'autre.

Conséquence : sur l'ensemble du plan, les effets traitements confondus ne sont pas totalement confondus avec les effets sous-blocs :

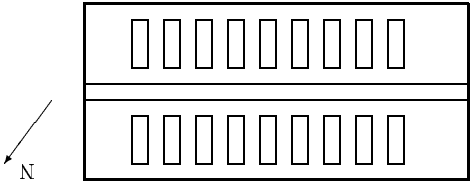
⇒ ils sont (en général) estimables dans la strate unité, ⇒ ils ne sont pas orthogonaux aux effets sous-blocs et sont estimables avec un coefficient d'efficacité factoriel > 1 .

Exemple : essai sous serre

Facteurs traitements :

- ☞ A : 3 variétés de laitue (notées 0, 1, 2)
- ☞ B : 3 solutions nutritives (notées 0, 1, 2)
- ☞ les traitements sont notés 00, 01, etc.

Unités expérimentales :



Structure en blocs choisie :

- ☞ 2 lignes (facteur R) de 3 sous-blocs (facteur P) de 3 unités
- ☞ réplique/sous-bloc/unité

Modèle factoriel complet :

$$y_{ij} = A + B + A.B + R + R.P$$

Propositions de plans

Plan ① : confusions partielles des effets principaux

Réplique 1 : $P = A$

Réplique 2 : $P = B$

Plan ② : confusion totale d'une partie de l'interaction

Réplique 1 : $P = \overline{A.B^2}$

Réplique 2 : $P = \overline{A.B^2}$

Plan ③ : confusion partielle de l'interaction

Réplique 1 : $P = \overline{A.B^2}$

Réplique 2 : $P = \overline{A.B}$

Il y a bien sûr d'autres possibilités. On pourra en particulier demander aux stagiaires sur quel type de plans vu précédemment on retombe, si l'on confond l'effet principal de A avec les effets sous-blocs dans les deux répliques (réponse : split-plot).

Plans obtenus, avant randomisation :

00	01	02	10	11	12	20	21	22
00	10	20	01	11	21	02	12	22

①

00	11	22	02	10	21	01	12	20
00	11	22	02	10	21	01	12	20

②

00	12	21	01	10	22	02	11	20
00	11	22	02	10	21	01	12	20

③

Comparaison des plans

Propriété :

Soit r_G le nombre de répétitions où le facteur (ou pseudofacteur) G est confondu avec le facteur sous-bloc ;

alors :

$$e_G = 1 - \frac{r}{r_G}$$

Coefficients d'efficacité factoriels :

Effacités	ddl	Plan
globale et factorielles	①	②
	③	
e	8	2/3
e_A	2	0.5
e_B	2	0.5
$e_{A,B}$	4	1
	0	0.5

La propriété présentée n'est vérifiée qu'à condition d'utiliser des facteurs ou pseudofacteurs traitements orthogonaux.

Commentaires

Le coefficient d'efficacité e d'un dispositif est nul par définition des qu'au moins un contraste traitement est non estimable (exemple : plan ②).
De même, le coefficient d'efficacité d'un effet factoriel est nul par définition si au moins un contraste de cet effet factoriel est non estimable (exemple : $e_{A,B}$, plan ③).
Dans le cas où tous les contrastes sont estimables, le coefficient e est égal à $(\sum_P \text{ddl}_P) / (\sum_P \text{ddl}_P e_{P_1}^2)$.

6. Logiciels

S-PLUS

```
> split.dat <- read.table("split.txt", header=T)
> split.dat$variete <- factor(split.dat$variete)
> split.dat$fert <- factor(split.dat$fert)
> split.dat$sbloc <- factor(split.dat$sbloc)
> aov.split <- aov(rdt ~ variete*fert +
  Error: sbloc %in% repl
  Df Sum of Sq Mean Sq F Value Pr(>F)
  Residuals 3 17.52966 5.84322
  Error: repl
  Df Sum of Sq Mean Sq F Value Pr(>F)
  Residuals 15 19.1427 1.27618
  variete 5 148.1967 29.63933 23.22508 1.437739e-06
  Error: Within
  Df Sum of Sq Mean Sq F Value Pr(>F)
  fert 2 10.63452 5.317260 15.00073 0.00001826
  variete:fert 10 8.37541 0.837541 2.36282 0.02879090
  Residuals 36 12.76080 0.354467
  > model.tables(aov.split, type="means", se=T)
  Tables of means
  Grand mean
  5.588
```

```
variete
  1 2 3 4 5 6
7.252 7.463 5.355 5.47 4.322 3.49
  Fert
  1 2 3
5.02 5.767 5.889
variete:fert
  Dim 1 : variete
  Dim 2 : fert
  1 2 3
1 6.280 7.315 8.160
2 6.843 7.525 8.020
3 4.625 6.290 5.150
4 4.818 5.535 6.058
5 4.190 4.348 4.428
6 3.363 3.588 3.520
Standard errors for differences of means
variete 0.4612
repl:c. 12.0000
fert 0.1719
repl:c. 24.0000
variete:fert
  When comparing means with same levels of:
  variete otherwise
  0.421 0.5752
  repl:c. 4
```

Instructions et sorties S-PLUS de l'analyse de variance en strates de l'essai d'inoculation sur haricot. Analyse des données par Dagnelie (1981).
L'analyse en strates est effectuée avec la fonction aov, en précisant la structure en blocs à l'aide du terme Error(repl/sbloc). La fonction model.tables donne les estimations des moyennes et des écarts-types des différences. Cette fonction ne peut être utilisée que pour un plan orthogonal.

```

data split;
infile 'split.txt';
input variete fert repl sbloc rdt;
run;

/* proc glm data=split;
class variete fert repl;
model rdt = variete|fert repl variete*repl;
random variete*repl / test;
(test h=variete e=variete*repl);
run; */

proc mixed data=split;
class variete fert repl;
model rdt = variete|fert repl;
random variete*repl;
lsmeans variete|fert;
run;

```

SAS

Instructions et sorties SAS de l'analyse de variance en strates de l'essai d'inoculation sur haricot. Analyse des données brutes, sans la transformation logarithmique proposée par Dagnelie (1981).
 L'analyse peut être effectuée avec la procédure glm ou la procédure mixed. Les transparents ne présentent que les sorties obtenues avec la procédure mixed. Pour obtenir des tests corrects, les effets sous-blocs doivent être introduits dans le modèle sous la forme variete*repl.
 Avec la procédure glm, le test des effets des variétés peut être obtenu soit avec la commande random variete*repl / test, soit avec la commande test h=variete e=variete*repl.
 Les options diff ou pdiff de lsmeans de la procédure mixed (non utilisées dans l'analyse présentée) permettent d'estimer les différences entre effets et les écarts-types des différences.

```

The MIXED Procedure
Class Level Information
Class      Levels  Values
VARIETE   6      1 2 3 4 5 6
FERT      3      1 2 3
REPL      4      1 2 3 4
RMML Estimation Iteration History
Iteration  Evaluations  Objective  Criterion
0          1          59.31345844
1          1          49.55908602  0.00000000
Convergence criteria met.

```

Seules les sorties de la procédure mixed sont présentées.

Covariance Parameter Estimates (REML)			
Cov Parm	Estimate		
VAR1*REPL	0.30723701		
Residual	0.35446667		
Model Fitting Information for RDT			
Description	Value		
Observations	72.0000		
Res Log Likelihood	-71.6454		
Akaike's Information Criterion	-73.6454		
Schwarz's Bayesian Criterion	-75.5772		
-2 Res Log Likelihood	143.2908		
Tests of Fixed Effects			
Source	NDF	DDF	Type III F Pr > F
VAR1	5	15	23.23 0.0001
FERT	2	36	15.00 0.0001
VAR1*FERT	10	36	2.36 0.0288
REPL	3	15	4.58 0.0181

Least Squares Means						
Effect	VAR1	FERT	FERT LSMEAN	Std	DF	t Pr > t
VAR1	1	7.25	0.33	15	22.24	0.0001
VAR1	2	7.46	0.33	15	22.88	0.0001
VAR1	3	5.35	0.33	15	16.42	0.0001
VAR1	4	5.47	0.33	15	16.77	0.0001
VAR1	5	4.32	0.33	15	13.25	0.0001
VAR1	6	3.49	0.33	15	10.70	0.0001
FERT	1	5.02	0.17	36	30.23	0.0001
FERT	2	5.77	0.17	36	34.73	0.0001
FERT	3	5.89	0.17	36	35.47	0.0001
VAR1*FERT	1	6.28	0.41	36	15.44	0.0001
VAR1*FERT	2	7.31	0.41	36	17.99	0.0001
VAR1*FERT	3	8.16	0.41	36	20.06	0.0001
VAR1*FERT	4	6.84	0.41	36	16.82	0.0001
VAR1*FERT	5	7.52	0.41	36	18.50	0.0001
VAR1*FERT	6	8.02	0.41	36	19.72	0.0001
VAR1*FERT	3	4.62	0.41	36	11.37	0.0001
VAR1*FERT	3	6.29	0.41	36	15.46	0.0001
VAR1*FERT	3	5.15	0.41	36	12.66	0.0001
VAR1*FERT	4	4.82	0.41	36	11.84	0.0001
VAR1*FERT	4	6.06	0.41	36	14.89	0.0001
VAR1*FERT	4	4.19	0.41	36	10.30	0.0001
VAR1*FERT	5	4.35	0.41	36	10.69	0.0001
VAR1*FERT	5	4.43	0.41	36	10.89	0.0001
VAR1*FERT	6	3.36	0.41	36	8.27	0.0001
VAR1*FERT	6	3.59	0.41	36	8.82	0.0001
VAR1*FERT	6	3.52	0.41	36	8.65	0.0001

Utilisation de CycDesign pour construire un plan pour l'essai sous serre à deux facteurs, deux répliques et des sous-blocs de taille trois.

CycDesign

CycDesign permet de construire des plans factoriels en blocs randomisés.

```

Block design
Resolvable
Factorial
Design parameters
  Number of factors      = 2
  Factor 1 levels       = 3
  Factor 2 levels       = 3
  Number of units/block = 3
  Number of replicates  = 2
Random number seed for design generation = 984650259
Factorial design criteria
Criterion (Upper bounds)
  Factor 1      1.000000
  Factor 2      1.000000
  Randomization 1
Random number seed for randomization = 984650308
    
```

Le plan obtenu est le plan © de la partie 5.

Il existe d'autres logiciels qui permettent de construire des plans factoriels, par exemple PLANOR (Kobylinsky, 1994), FACTEX (SAS, module QC).

```

Factorial design randomization:
  Factor 1:      2
  Factor 2:      3
  Replicate randomization:
    1 2
  Block randomization:
    2 1 3
    2 1 3
Blocks of the design printed in rows
rep 1 -----
plot 1 2 3
block +-----+
  1 | (3,3) (2,2) (1,1)
  2 | (2,3) (1,2) (3,1)
  3 | (1,3) (2,1) (3,2)
rep 2 -----
plot 1 2 3
block +-----+
  1 | (1,2) (3,3) (2,1)
  2 | (1,3) (2,2) (3,1)
  3 | (3,2) (2,3) (1,1)
    
```

7. Conclusion

- Les plans factoriels sont des dispositifs dans lesquels on étudie simultanément plusieurs facteurs
- Dans un split-plot ou un criss-cross, on confond totalement des effets principaux avec des effets blocs, en général à cause de contraintes pratiques ; les effets confondus sont estimés avec moins de précision.
- On peut construire des plans factoriels en blocs "sur mesure" en confondant des effets factoriels des traitements avec des effets blocs.
- Certains plans d'expériences n'ont pas été abordés dans ce module, par exemple les plans factoriels fractionnaires et pour surfaces de réponse.

COMPLÉMENTS, RÉFÉRENCES BIBLIOGRAPHIQUES

Complément : effets de voisinage

Il y a des **effets de voisinage** lorsque les réponses des unités dépendent des traitements appliqués aux unités voisines.

- Expériences agronomiques : interférences entre traitements voisins ; par exemple effets de compétition liés à la hauteur des variétés.
- Mesures répétées : effet du traitement précédent.

Effets de voisinage \Rightarrow biais sur les comparaisons entre traitements.

Méthodes physiques pour limiter les effets de voisinage

- Utiliser des parcelles avec des rangs de bordure
- Augmenter la taille des unités expérimentales

Complément à présenter aux stagiaires en fonction de leur intérêt pour le sujet et du temps disponible.

Les effets de voisinage sont gênants car ils sont une source de biais. On peut limiter les effets de voisinage par des méthodes statistiques. Les rangs de bordure sont des rangs qui semés en bordure de parcelle mais qui ne sont pas récoltés.

Méthodes statistiques

Analyse

Modèle :

réponse = traitement + *voisinage* + bloc + erreur

- Effets de voisinage pour chaque traitement

$$y_{ij(t)} = \mu + \tau_i + \phi_t + \phi_{t''} + \beta_i + \varepsilon_{ij}$$

ϕ_t : effet de voisinage du traitement t

t' et t'' : traitements voisins de t

- Effets de voisinage proportionnels aux différences de hauteur entre parcelles voisines

$$y_{ij(t)} = \mu + \tau_i + \delta (h_{ij-1} - h_{ij+1})/2 + h_{ij+1}/2 - h_{ij} + \beta_i + \varepsilon_{ij}$$

h_{ij} : hauteur sur la parcelle ij

h_{ij-1} : hauteur sur la parcelle de gauche

h_{ij+1} : hauteur sur la parcelle de droite

δ : paramètre à estimer (coefficient de régression de la covariable "différence de hauteur")

Plan d'expériences

Les plans équilibrés pour les voisinages équilibrés Les voisinages entre traitements.

Plan en blocs complets partiellement équilibré pour les voisinages avec des parcelles de bordure pour $v = 12$ et $r = 3$:

Bloc 1	7	9	2	8	10	11	6	4	5	12	1	3	7	9
Bloc 2	8	4	3	2	12	10	6	9	5	1	11	7	8	4
Bloc 3	11	3	5	7	1	9	10	2	6	8	12	4	11	3

Les plans équilibrés pour les voisinages ont des propriétés d'optimalité :

- pour le *biais* des estimations des effets traitements lorsque le modèle d'analyse ne comprend pas d'effets de voisinage,
- pour la *variance* des estimations des effets traitements et des effets de voisinage lorsque le modèle d'analyse comprend des effets de voisinage.

Les effets de voisinage peuvent être pris en compte dans l'analyse statistique. Les deux modèles présentés supposent que chaque unité subit un effet de voisinage venant du traitement situé à droite et un effet de voisinage venant du traitement situé à gauche. Le premier modèle fait moins d'hypothèses sur les effets de voisinage que le deuxième mais utilise plus de paramètres. D'autres modélisations des effets de voisinage existent.

Les effets de voisinage peuvent également être pris en compte à l'aide du plan d'expériences. Dans l'exemple, aucun traitement n'a plusieurs fois un même traitement comme voisin. L'analyse porte sur les données des parcelles qui ne sont pas des parcelles de bordure. Voir Azais, Bailey et Monod (1993) pour la construction de plans équilibrés pour les voisinages ; les auteurs ont écrit un programme qui permet de générer les plans.

Complément : modèles spatiaux

Les **modèles spatiaux** cherchent à modéliser les effets terrains plus finement qu'avec des effets blocs.

Sur la parcelle en ligne i et colonne j :

$$y_{ij(t)} = \tau + \zeta_{ij}(\text{effet terrain}) + \varepsilon_{ij}(\text{erreur exper.})$$

Les effets terrains ζ_{ij} sont modélisés par des variables

aléatoires :

- centres,

- corrélés, dont la covariance $\text{Cov}(\zeta_{ij}, \zeta_{i'j'})$ diminue

avec la distance entre les parcelles (i, j) et (i', j') .

Exemple : modèle autorégressif ARI en ligne

$$E(\zeta_{ij}) = 0$$

$$\text{Cov}(\zeta_{ij}, \zeta_{i'j'}) = \sigma_{\zeta}^2 \phi^{|j-j'|} \text{ si } i = i'$$

$$= 0 \text{ si } i \neq i'$$

Sous forme matricielle :

$$y = X\tau + \zeta + \varepsilon$$

avec

$$\text{Var}(\zeta) = \sigma_{\zeta}^2 \begin{pmatrix} S & 0 & \dots & 0 \\ 0 & S & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & S \end{pmatrix}$$

et

$$S = \begin{pmatrix} 1 & \phi & \dots & \phi^{k-1} \\ \phi & 1 & \dots & \phi^{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{k-1} & \phi^{k-2} & \dots & 1 \end{pmatrix}$$

La matrice de variance de y dépend de plusieurs paramètres \Rightarrow analyse en modèle mixte.

Conséquences sur la planification

Analyse par un modèle spatial \Rightarrow en général :

- les plans optimaux sont des plans équilibrés pour les voisinages,
- mais les dispositifs en blocs randomisés restent bien adaptés.

Références bibliographiques

- Dispositifs en blocs à un facteur traitement*
- Azaïs, J.-M., Bailey, R. A. et Monod, H. (1993). A catalogue of efficient neighbour-designs with border plots. *Biometrics*, 49, 1252-1261.
- Bailey, R. A. (1991). Stratification and randomization. Cours ECAS-91 : *Design of experiments*.
- Bergonzini, J.-C. et Duby, C. (1995). *Analyse et planification des expériences. Les dispositifs en blocs*. Masson, Paris.
- Edmondson, R. N. (1993). Systematic row-and-column designs balanced for low order polynomial interactions between rows and columns. *Journal of the Royal Statistical Society*, B, 55, 707-723.
- John, J. A. et Williams, E. R. (1995). *Cyclic and computer generated designs*. Chapman et Hall, Londres.
- Monod, H., David, O., Gouet, J.-P., Philipppeau, G., Piraux, F. et Schott, J.-J. (2001). *Alpha-plans, carrés semi-latins et autres dispositifs en répliques. Comment les utiliser ?* ITCF, Boigneville.
- Oger, R. et Roisin, C. (1992). The influence of block size on the efficiency of incomplete block designs in Belgian cereal variety trials. *Buletin Ocenj Odmian*, 24-25, 159-165.

- Paterston, H. D. et Hunter, E. R. (1983). The efficiency of incomplete block designs in National List and Recommended List cereal variety trials. *Journal of Agricultural Science, Cambridge*, 101, 427-433.
- Philipppeau, G. (1977). *Théorie des plans d'expérience. Application à l'agronomie*. ITCF, Paris.
- Whitaker, D., Williams, E. R. et John, J. A. (1997). *CycDesign: a package for the computer generation of experimental designs*. CSIRO, Canberra.
- <http://www.fhp.csiro.au/figr/software/cycdesign/>
- Williams, E. R. et Matheson, A. C. (1994). *Experimental design and analysis for use in tree improvement*. CSIRO, Canberra.
- Williams, E. R. et Talbot, M. (1993). *ALPHA+, Experimental designs for variety trials*. Design user manual. CSIRO, Canberra et BIOS, Edinburgh.
- Plans factoriels*

Box, G. E. P., Hunter, W. G. et Hunter J. S. (1978). *Statistics for experimenters. An introduction to design, data analysis, and model building*. Wiley, New York.

Durier, C., Hennequet-Antier, C. et Schaeffer, B. (2002). *Plans d'expériences. Dans Daudin, J.-J. et Duby, C. (Eds) Techniques mathématiques pour l'industrie agro-alimentaire. A paraître.*

Kobylinskiy, A. (1994). *PLANOR : programme de génération automatique de plans d'expériences réguliers*. Document interne, laboratoire de biométrie, INRA, Versailles.

Livres généraux sur les plans d'expériences

Cochran, W. G. et Cox, G. M. (1957). *Planning of experiments*. Wiley, New York.

Cox, D. R. (1958). *Planning of experiments*. Wiley, New York.

Dagnèlle, P. (1981). *Principes d'expérimentation*. Presses Agronomiques de Gembloux, Gembloux.

Dean, A. et Voss, D. (1999). *Design and analysis of experiments*. Springer-Verlag, New York.

Droesbeke, J.-J., Fine, J. et Saporta, G. (1997). *Plans d'expériences. Applications à l'entreprise*. Editions Technip, Paris.

John, P. W. M. (1971). *Statistical design and analysis of experiments*. Mac Millan, New York.

Kempthorne, O. (1952). *The design and analysis of experiments*. Krieger, Malabar, Floride.

Mead, R. (1988). *The design of experiments*. Statistical principles for practical application. Cambridge University Press, Cambridge.

Silvey, S. D. (1980). *Optimal design: an introduction to the theory of parameter estimation*. Chapman and Hall, London.

Modèle linéaire

Colin, F., Concorde, D., Dessaint, F., Henneguet, C., Kaan, J., Laurens, F., Lyazrhi, F., Mangin, B., Monpied, P., Robert, N. et Wavresky, P. (1993). *Le modèle linéaire*. Module FPSTAT 2/3. INRA, Paris.

<http://www.inra.fr/bia/T/schnu/FPstat/node14.html>

Coursol, J. (1980). *Technique statistique des modèles linéaires. I - Aspects théoriques*. CIMPA, Nice.

Gouet, J.-P. et Philippéau, G. (1992). *Comment interpréter les résultats d'une analyse de variance ? ITCF*, Boigneville.

Prum, B. (1996). *Modèle linéaire. Comparaison de groupes et régression*. Les éditions INSERM, Paris.

Autres

Dagnèlle, P. (1980). *Théorie et méthodes statistiques*. Presses Agronomiques de Gembloux, Gembloux.

Kempthorn, R. A. et Fox, P. N. (1997). *Statistical methods for plant variety evaluation*. Chapman et Hall, Londres.

Pearce, S. G. (1983). *The agricultural field experiment. A statistical examination of theory and practice*. Wiley, New York.

Searle, S. R., Casella, G. et McCullouch, C. E. (1992). *Variance components*. Wiley, New York.

Tomassone, R., Derwin, C. et Masson, J.-P. (1993). *Biométrie, modélisation de phénomènes biologiques*. Masson, Paris.