Introduction à la planification d'expérience

Plans d'expériences usuels

Frédéric Bertrand[1]

[1] IRMA, Université de Strasbourg
Strasbourg, France

ENSAI 3e Année
2015-2016

Référence

Ce cours s'appuie essentiellement sur

1. le livre de Hardeo Sahai et Mohammed I. Ageel, **The Analysis of Variance : Fixed, Random and Mixed Models**, aux éditions Birkhäuser, 2000.
2. le livre de Hardeo Sahai et Miguel Ojeda, **Analysis of Variance for Random Models**, **Volume I: Balanced data**, aux éditions Birkhäuser, 2004.
3. le livre de Hardeo Sahai et Miguel Ojeda, **Analysis of Variance for Random Models**, **Volume II: Unbalanced data**, aux éditions Birkhäuser, 2005.

Référence

Ce cours s'appuie également sur

4. le livre de Douglas C. Montgomery, **Design and Analysis of Experiments**, 7þEdition, aux éditions Wiley, 2009.
5. le livre de C. P. Doncaster et A. J. H. Davey, **Analysis of Variance and Covariance : How to Choose and Construct Models for the Life Sciences**, aux éditions Cambridge University Press, 2007.

Sommaire

Sommaire

Connaître les différents modèles usuels

Avant de réaliser une expérience, ou lorsque vous participez à l'élaboration du protocole expérimental, il faut toujours se poser la question de savoir avec quels outils statistiques, et donc quels modèles, vous pourrez analyser les résultats une fois les essais réalisés.

Il existe plusieurs situation classiques « universelles », dans le sens où elles apparaissant aussi bien en biologie, en agronomie, dans le domaine médical que dans l'industrie par exemple.
Il est donc fortement probable que vous soyez confronté à l'une d'entre elles.

Définition : Équilibre

Équilibre : un plan équilibré a le même nombre d'observation pour chaque combinaison des modalités.
L'équilibre est une propriété particulièrement souhaitable des modèles contenant des facteurs croisés, lorsque la perte de l'équilibre amène généralement, mais pas nécessairement, à une perte d'orthogonalité. S'en suivent des difficultés pour partitionner les sources de variation de la réponse qui ne peuvent se traiter que grâce aux générales du modèle linéaire et aux non mixtes non-restreints (Searle 1971).

Les plans équilibrés ou orthogonaux peuvent, quant à eux, être analysés à l'aide des techniques spécifiques et des modèles restreints.

Définition : Model-1 and Model-2 designs (Newman et al. 1997)

Model 1 :

Les plans d'ANOVA sans répétitions complètes, comme les plans aléatoires complets en blocs S à deux facteurs A et B, confondent la variation résiduelle non mesurée (pas de répétitions) avec celle associée au terme d'interaction de degré le plus élevé entre le facteur bloc et les facteurs traitements S'*A*B. L'analyse de ce plan en complets complets à l'aide d'un Modèle-1, amène à incorporer dans le modèle tous les termes d'interaction entre les blocs et les facteurs hormis celui de degré le plus elévé, c'est-à-dire S'*A et S'*B. Ces termes serviront de terme d'erreur pour les statistiques de test des effets principaux des facteurs A et B.

Définition : Model-1 and Model-2 designs (Newman et al. 1997)

Model 2 :

Par contraste, dans un Modèle-2, nous considérons que $MS[S*A]$, $MS[S*B]$ and $MS[S*B*A]$ estiment la même quantité et par conséquent c'est le terme regroupant ces trois quantités qui servira de terme d'erreur pour le test des deux effets principaux.

Remarque

L'utilisation d'un Modèle-1 est adéquate s'il y a des raisons biologiques pour supposer qu'il existe des interactions entre les blocs et les traitements. L'existence de ces interactions pourra être testée lors de l'étude du modèle. Par exemple, le plan en blocs aléatoires complets S'|B|A permet de tester l'interaction S*A avec le rapport F = MS[S'*A]/MS[S'*B*A], et l'interaction S*B avec le rapport F = MS[S'*B]/MS[S'*B*A].

Remarque

Il faut néanmoins noter que de tels tests ont généralement une faible puissance. En utilisant un Modèle-1, nous réalisons le test des effets fixes en utilisant les interactions blocs-traitements. Cette stratégie a un coût, puisque la puissance de ces tests est généralement réduite par rapport à un Modèle-2 et son terme d'erreur dont le nombre de degrés de liberté est bien plus important à cause du regroupement qui a été réalisé.

Remarque

- Les plans aléatoires en blocs complets peuvent être analysés avec un Model-1 ou un Model-2.
- Les plans split plot (en parcelles divisées) sont généralement analysés avec un modèle-2.
- Les plans pour mesures répétées sont généralement analysés avec un Modèle-1.

Sommaire

Plans en bloc complets

- Plan aléatoire en blocs complets pour un facteur : Y = S'|A.
- Plan aléatoire en blocs complets pour deux facteurs : Y = S'|B|A.
- Plan aléatoire en blocs complets pour trois facteurs : Y = S'|C|B|A.

Il existe des variantes incomplètes de ces plans, les plans en blocs incomplets, les carrés latins ou gréco-latins.

Plans splits plots

Dans les plans splits plot, les sujets sont emboîtés dans les facteurs par rapport auxquels les mesures ne sont pas répétées.

- Modèle split-plot à deux facteurs (i) Y = B|P'(S'|A)
- Modèle split-plot à trois facteurs (i) Y = C|P'(S'|B|A)
- Modèle split-plot à trois facteurs (ii) Y = C|B|P'(S'|A)
- Mdèle split-split-plot (i) Y = C|Q'(B|P'(S'|A))
- Mdèle split-split-plot (ii) Y = C|P'(B|S'(A))
- Modèle split-plot à deux facteurs (ii) Y = B|S'(A)
- Modèle split-plot à trois facteurs (iii) Y = C|B|S'(A)
- Modèle split-plot avec emboîtement Y = C|S'(B(A))
- Modèle split-plot à trois facteurs (ii) Y = C|S'(B|A)

Plans pour mesures répétées

- Mesures répétées à un facteur Y = S’|A
- Mesures répétées par rapport aux deux facteurs Y = S’|B|A
- Mesures répétées par rapport à l’un des deux facteurs Y = B|S’(A)
- Trois facteurs avec mesures répétées et emboîtement Y = C(B)|S’(A)
- Trois facteurs avec mesures répétées par rapportà deux facteurs croisés Y = C|B|S’(A)
- Modèle emboîté avec mesures répétées et facteur croisé Y = C|S’(B(A))
- Modèle à trois facteurs avec mesures répétées suivant l’un des facteurs Y = C|S’(B|A)

Sommaire

Un nouveau modèle

Il est souvent d'usage d'utiliser des plans aléatoires en blocs complets afin de réduire l'erreur résiduelle d'une expérience en supprimant la variabilité due à une variable de nuisance connue et contrôlée.

Il existe d'autres types de plans s'appuyant sur le principe de l'utilisation de blocs. Par exemple, supposons qu'un expérimentateur s'intéresse à l'effet de cinq formulations différentes d'un combustible de fusée utilisé dans un système d'éjection sur la vitesse de combustion observée.

Un nouveau modèle (suite)

Chaque formulation est obtenue à partir d'un lot de matières premières qui ne permet que de tester cinq formules. De plus les formulations sont préparées par différents opérateurs qui possèdent des compétences et une expérience variées.

Par conséquent, il semble qu'il y ait deux variables de nuisances dont nous allons chercher à compenser les effets en les « moyennant » grâce au plan. Le plan approprié pour ce problème consiste à tester chaque formulation une et une seule fois pour chacun des lots et pour chacun des cinq opérateurs.

Un nouveau modèle (suite)

Le plan résultant de ces choix s'appelle un **carré latin.** Il a été reproduit dans le tableau ci-dessous :

Lots de Mat. Pre.	Opérateurs 1	2	3	4	5
1	$A = 24$	$B = 20$	$C = 19$	$D = 24$	$E = 24$
2	$B = 17$	$C = 24$	$D = 30$	$E = 27$	$A = 36$
3	$C = 18$	$D = 38$	$E = 26$	$A = 27$	$B = 21$
4	$D = 26$	$E = 31$	$A = 26$	$B = 23$	$C = 22$
5	$E = 22$	$A = 30$	$B = 20$	$C = 29$	$D = 31$

Un nouveau modèle (suite)

Remarquons que le plan d'expérience est un arrangement carré et que les cinq formulations (ou traitements) sont représentés par les lettres romaines A, B, C, D et E d'où le nom de carré latin. Nous constatons qu'aussi bien les lots de matière première que les opérateurs sont orthogonaux aux traitements.

Un nouveau modèle (suite)

Le carré latin est utile pour éliminer deux sources de nuisance, c'est-à-dire permettre une répartition orthogonale en bloc suivant deux facteurs. De ce fait, les lignes et les colonnes du carré latin sont en fait des restrictions sur la randomisation.

En général, un carré latin pour p facteurs, ou un carré latin $p \times p$, est un carré à p lignes et à p colonnes. Chacunes des p^2 cellules contient l'une des p lettres qui correspondent aux traitements, chacune des lettres n'appaissant qu'une fois dans chaque ligne et chaque colonne.

Un nouveau modèle (suite)

Voici des exemples de carré latin :

4×4

A	B	D	C
B	C	A	D
C	D	B	A
D	A	C	B

5×5

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

6×6

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

Modèle statistique

Le modèle statistique pour un carré latin est :

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + \mathcal{E}_{ijk},$$

où $i = 1, \ldots, I, j = 1, \ldots, J$ et k est une fonction de i et j donnée par le carré.
Notons $n = I \times J$ le nombre total de mesures ayant été effectuées.

Définition

Un carré gréco-latin est un tableau carré de n lignes et n colonnes remplies avec n^2 paires distinctes, et où chaque ligne et chaque colonne ne contient qu'un seul exemplaire. Il s'agit de la superposition de deux carrés latins orthogonaux. Si les deux carrés latins n'étaient pas orthogonaux, alors une paire pourrait apparaître plus d'une fois. On dit aussi carré bilatin.

Exemple de carré Gréco-Latin

Aα	Bδ	Cβ	Dε	Eγ
Bβ	Cε	Dγ	Eα	Aδ
Cγ	Dα	Eδ	Aβ	Bε
Dδ	Eβ	Aε	Bγ	Cα
Eε	Aγ	Bα	Cδ	Dβ