Plans d’expériences incomplets

Résumé

L’objet de ce chapitre est d’étudier certains plans factoriels particuliers, assez courants dans la pratique statistique, notamment industrielle. Ces plans d’expériences particuliers étant nombreux, nous avons du faire des choix qui, comme toujours, comportent un certain arbitraire. Nous détaillerons successivement la méthode des blocs, les plans en carrés latins et gréco-latins et les plans à plusieurs facteurs à deux niveaux (de type Plackett et Burman). Dans un dernier paragraphe, nous donnerons quelques indications sur d’autres dispositifs expérimentaux également courants.

1 Introduction

La principale référence bibliographique de ce chapitre est l’ouvrage de John (1998). Un autre ouvrage de référence, très complet sur les plans d’expériences et rédigé en français, est celui de Droesbeke *et al.* (1997). Enfin, on trouvera divers compléments intéressants dans les ouvrages de Azaïs & Bardet (2005), de Bergonzini & Duby (1995) de Goupy & Creighton (2006) et de Saporta (2006).

La méthode des blocs consiste à introduire un facteur exogène (autre que le(s) facteur(s) d’intérêt), appelé le facteur bloc. Chaque bloc pris en compte (autrement dit chaque niveau du facteur bloc) est une “unité expérimentale” supposée homogène relativement à l’expérimentation considérée (des exemples sont donnés dans le premier paragraphe). Ce type de dispositif a pour avantage de permettre de contrôler une partie de la variabilité résiduelle du modèle étudié (donc de son erreur).

Les plans en carrés latins et en carrés gréco-latins sont des plans factoriels fractionnaires (plans factoriels incomplets dans lesquels on n’observe qu’une fraction de l’ensemble des cellules du plan). Ils sont définis par le croisement de trois facteurs (cas des carrés latins) ou de quatre facteurs (cas des carrés gréco-latins), ces facteurs ayant nécessairement tous le même nombre de niveaux. Cette particularité permet d’obtenir des propriétés intéressantes avec un nombre restreint d’observations.

Dans les plans à plusieurs facteurs à deux niveaux, on peut étudier un grand nombre de facteurs (jusqu’à dix et plus) avec peu d’observations (8, 12, 16...), grâce à des dispositifs expérimentaux particuliers, à savoir des plans (très) incomplets, équilibrés et orthogonaux : il s’agit des plans de Plackett et Burman qui permettent de faire des tests (essentiellement sur les effets principaux, mais pas uniquement) et d’estimer les paramètres des modèles retenus.

La mise en œuvre de ces plans peut se faire, de façon standard, au moyen de la procédure GLM du logiciel SAS.

2 La méthode des blocs

2.1 Principes

2.1.1 *Objectif*

L’idée essentielle à l’origine de la méthode des blocs est de *réduire la variabilité résiduelle* d’un modèle pour un plan factoriel. En effet, même dans le cas très simple d’une analyse de variance à un seul facteur avec un plan complet et équilibré, si le nombre de répétitions au sein de chaque niveau du facteur est assez élevé (disons supérieur à 5), il peut se faire que les valeurs de la variable réponse soit relativement dispersées, même pour un niveau fixé du facteur, ce qui va entraîner une estimation de la variance de l’erreur du modèle ($\hat{\sigma}^2$) assez importante. Cela aura alors les conséquences suivantes :

- la statistique du test de l’effet du facteur (statistique de Fisher) aura un grand dénominateur et sera donc relativement petite : il sera ainsi assez difficile de mettre en évidence cet effet ;
- les erreurs-types des estimateurs des paramètres du modèle (proportionnelles à $\hat{\sigma}$) seront grandes, de sorte que les estimations correspondantes seront peu précises.

L’introduction d’un facteur bloc, facteur exogène, a pour but de diminuer la variabilité résiduelle du modèle considéré (donc $\hat{\sigma}^2$) et d’atténuer ainsi les problèmes évoqués ci-dessus.

2.1.2 Définition

On appelle *bloc* une condition expérimentale homogène relativement au phénomène étudié (le phénomène mesuré par la variable réponse Y). L'idée est donc de constituer un certain nombre, noté B ($B \geq 2$), de blocs distincts ; dans chacun d'eux, on observe la variable réponse pour différents niveaux du (des) facteur(s). En fait, l'ensemble des blocs considérés constitue un facteur supplémentaire.

Cette définition étant un peu vague, nous la précisons en donnant quelques exemples.

2.1.3 Exemples

- En agronomie, si l'on souhaite comparer les rendements (variable réponse) de différentes variétés d'une culture donnée (le facteur est la variété) dans une certaine zone géographique, un bloc sera une parcelle de terrain, homogène relativement aux facteurs non contrôlés susceptibles d'influencer le rendement : altitude, ensoleillement, humidité... Ainsi, un bloc-parcelle devra être sensiblement situé à la même altitude, avoir partout la même exposition, la même humidité...
- En médecine, si l'on souhaite étudier l'effet de différents médicaments (le facteur est le type de médicament) sur la guérison d'une maladie donnée (la variable réponse est le taux de guérison parmi les patients atteints de cette maladie), un bloc sera un ensemble de malades homogènes selon, par exemple, les facteurs âge et sexe (autrement dit un ensemble de malades de même sexe et appartenant à la même tranche d'âge).
- Dans une expérience de laboratoire se traduisant par une mesure très précise (la variable réponse) faite par un opérateur utilisant une certaine machine, un bloc sera constitué par un couple opérateur-machine.

2.1.4 Notion de randomisation

En statistique, la notion de *randomisation* (ce terme anglais est passé dans la langue française et signifie *répartition au hasard*) est très importante dans les plans d'expériences. Elle n'est pas propre à la méthode des blocs, même si elle est fondamentale dans ce cas. Son principe est de répartir au hasard les traitements qui seront administrés aux différents individus d'un même bloc. Nous préciserons la notion de randomisation dans les différents plans d'expériences étudiés par la suite.

2.2 Plans en blocs complets équilibrés

Nous nous intéressons ici au cas d'un seul facteur F, comportant J niveaux ($J \geq 2$) indicés par j. Toutefois, ce qui est décrit dans ce cas s'applique de la même manière dans le cas de plusieurs facteurs croisés.

2.2.1 Principe

On constitue B blocs au sein desquels chacun des niveaux du facteur est observé une fois et une seule. Chaque niveau est ainsi observé B fois, de sorte que, pour le seul facteur F, le plan est un plan complet, équilibré, avec B répétitions. De son côté, le plan croisant le facteur F et le facteur bloc est un plan complet, équilibré, sans répétition.

EXEMPLE 1 *Considérons un facteur F à 3 niveaux ($J = 3$) pour lequel on a constitué 5 blocs ($B = 5$). Nous donnons, dans le tableau ci-dessous, les 15 observations réalisées.*

	niveaux du facteur F →	*1*	*2*	*3*
	1	*6*	*5*	*8*
	2	*10*	*9*	*12*
Blocs	*3*	*5*	*4*	*8*
	4	*9*	*7*	*10*
	5	*10*	*7*	*11*

Remarque. — De même qu'on peut envisager de considérer plusieurs facteurs croisés et d'observer une fois et une seule chaque cellule dans chaque bloc, il est aussi possible d'observer plusieurs fois chaque niveau de F dans chaque bloc (le nombre d'observations de chaque niveau dans chaque bloc restant toujours le même pour conserver un plan équilibré). Toutefois, cela augmente d'autant le nombre total d'observations, donc le coût de l'expérimentation.

2.2.2 Modèle

Indiçons par b les B blocs considérés et notons Y_{bj} la v.a.r. réponse dans le bloc b, au niveau j de F. Nous utilisons le paramétrage centré et nous écrivons le modèle sous la forme

$$Y_{bj} = \mu + \alpha_b^1 + \alpha_j^2 + U_{bj}$$

avec $U_{bj} \sim \mathcal{N}(0, \sigma^2)$, les U_{bj} étant indépendantes. Comme d'habitude, les paramètres vérifient les contraintes : $\sum_{b=1}^{B} \alpha_b^1 = \sum_{j=1}^{J} \alpha_j^2 = 0$.

Il s'agit en fait d'un modèle à 2 facteurs croisés (le facteur bloc et le facteur F), sans interaction. En général, on ne considère pas les effets d'interactions entre le facteur et les blocs, d'une part car ils n'ont pas de signification concrète, d'autre part car il ne serait pas possible d'estimer la variance du modèle (et donc de faire des tests) dans un modèle avec interactions (qui serait un modèle saturé, c'est-à-dire comportant autant de paramètres que d'observations).

Remarque. — Dans le modèle ci-dessus, on notera qu'il n'y a pas d'indice i, indice des individus dans la notation utilisée dans tout ce cours. Cela tient au fait qu'il n'y a, dans ce type de plan, qu'une seule observation par bloc et par niveau du facteur.

2.2.3 Randomisation

Dans le cas d'un plan en blocs à un seul facteur, complet et équilibré, la randomisation consiste, dans chaque bloc, à tirer au hasard les niveaux du facteur auquel seront observés les différents individus.

2.3 Plans en blocs incomplets équilibrés

Toujours dans le cas d'un seul facteur et d'un dispositif en blocs, si le nombre de niveaux J du facteur est assez élevé, ainsi que le nombre B de blocs, un plan complet équilibré peut être trop coûteux pour l'expérimentateur. D'où l'idée, dans ce cas, de considérer un plan incomplet, mais toujours équilibré. Par construction, ces plans vont nécessairement vérifier diverses contraintes.

2.3.1 Principe

- Le nombre J de niveaux du facteur F doit, dans ce cas, être au moins égal à 3 : $J \geq 3$ (nous en précisons plus loin la raison, mais, de toutes façons, ce type de plan est conçu pour un facteur avec de nombreux niveaux).
- Le nombre de blocs considéré doit, de son côté, être au moins égal à J : $B \geq J$.
- Dans chaque bloc, on n'observe pas tous les niveaux de F (sinon, on aurait affaire à un plan complet), mais seulement un nombre restreint K, toujours le même pour équilibrer l'expérience ; comme il est nécessaire d'observer au moins deux niveaux distincts de F dans chaque bloc (sinon, il y aurait confusion des effets de F et des effets des blocs), il vient : $2 \leq K < J$ (on voit ainsi pourquoi il est nécessaire d'avoir $J \geq 3$). Le nombre total d'observations est donc $n = BK$.
- Chaque niveau j du facteur F est observé dans le même nombre R de blocs (R est donc le nombre d'observations réalisées pour chaque niveau du facteur, autrement dit le nombre de répétitions : le plan est ainsi équilibré). Il vient $2 \leq R < B$ et le nombre total d'observations vérifie $n = JR = BK$ (on obtient ainsi $B \geq 3$; la condition plus restrictive $B \geq J$ provient de considérations sur le rang de la matrice d'incidence ; voir, par exemple, John, 1998).

Pour obtenir un maximum d'homogénéité du plan considéré, on lui impose, en plus, la condition suivante :

- Chaque couple (j, j') de niveaux de F ($j \neq j'$) est observé dans le même nombre L de blocs. Cette condition entraîne l'égalité suivante : $BK(K-1) = LJ(J-1)$. En effet :

  . le nombre total de couples de niveaux de F est $\dfrac{J(J-1)}{2}$;

  . le nombre de "places" nécessaires dans l'ensemble des blocs est donc : $L\dfrac{J(J-1)}{2}$;

  . le nombre de couples expérimentables dans un bloc est : $C_K^2 = \dfrac{K(K-1)}{2}$;

  . le nombre total de couples expérimentables est donc : $B\dfrac{K(K-1)}{2}$;

  . on en déduit : $BK(K-1) = LJ(J-1)$.

2.3.2 Conséquences

Comme on a vu $n = JR = BK$, on déduit de l'égalité ci-dessus : $L = R\dfrac{K-1}{J-1}$. Par conséquent, pour un nombre J de niveaux de F donné, les entiers B, K et R doivent être choisis de telle sorte que $B \geq J$, que $JR = BK$

et que $R\frac{K-1}{J-1}$ soit entier.

Ainsi, les dispositifs en blocs incomplets équilibrés, caractérisés par les 5 entiers J, B, K, R et L, répondent à des règles précises de construction et sont, pour cette raison, en nombre limité.

EXEMPLE 2 *L'exemple ci-dessous correspond au cas $J = 5$, $B = 10$, $K = 3$, $R = 6$, $L = 3$ et $n = 30$.*

	niveaux de F →	*1*	*2*	*3*	*4*	*5*
	1	*52*	*51*	*60*	–	–
	2	*56*	*61*	–	*61*	–
	3	*49*	*54*	–	–	*65*
	4	*46*	–	*56*	*55*	–
blocs	*5*	*48*	–	*53*	–	*52*
	6	*44*	–	–	*52*	*57*
	7	–	*45*	*51*	*51*	–
	8	–	*46*	*52*	–	*52*
	9	–	*47*	–	*50*	*50*
	10	–	–	*29*	*29*	*30*

Pour des valeurs "raisonnables" du nombre J de niveaux du facteur F considéré, on peut lister les caractéristiques de tous les plans en blocs incomplets et équilibrés qu'il est possible de construire. Ainsi, pour les cinq premières valeurs possibles de J, nous donnons, dans le tableau ci-après, les caractéristiques de tous ces plans.

J	B	K	R	L	n
3	3	2	2	1	6
4	4	3	3	2	12
4	6	2	3	1	12
5	5	4	4	3	20
5	10	2	4	1	20
5	10	3	6	3	30
6	6	5	5	4	30
6	10	3	5	2	30
6	15	2	5	1	30
6	15	4	10	6	60
6	20	3	10	4	60
7	7	3	3	1	21
7	7	4	4	2	28
7	7	6	6	5	42
7	21	2	6	1	42

Pour les structures de plans avec $J \geq 8$, on se reportera aux tables de Fisher & Yates (1963).

2.3.3 Modèle

Indiçons encore par j chaque niveau du facteur et par b chaque bloc. La v.a.r. réponse Y_{bj} n'est observée que si le couple (b, j) est tel que le niveau j est expérimenté dans le bloc b. On écrit alors, toujours avec le paramétrage centré,

$$Y_{bj} = \mu + \alpha_b^1 + \alpha_j^2 + U_{bj},$$

avec $U_{bj} \sim \mathcal{N}(0, \sigma^2)$, les U_{bj} étant indépendantes. Autrement dit, le modèle est le même que dans le cas complet.

2.3.4 Randomisation

Encore une fois, la randomisation est assez systématique dans ce genre de situation expérimen-tale. Elle se fait, ici, à trois niveaux :

- les libellés des traitements (autrement dit les colonnes du tableau de l'exemple 8) sont tirés au hasard ;
- les configurations des lignes sont affectées aléatoirement aux blocs ;
- dans chaque bloc, on tire au hasard le traitement subi par chaque individu.

Remarque. — Pour simplifier l'exposé, on n'a considéré, dans ce paragraphe, qu'un seul facteur F. Dans la pratique des plans d'expériences, on rencontre fréquemment des dispositifs en blocs incomplets et équilibrés associés à plusieurs facteurs. Cela complique le dispositif sans en changer le principe.

Remarque. — Nous verrons, au chapitre 6, les modèles à effets aléatoires et les modèles mixtes. On peut, d'une part utiliser la méthode des blocs avec un modèle mixte, d'autre part considérer, dans certains cas, que le facteur bloc est lui-même un facteur à effets aléatoires.

3 Les plans en carrés latins et gréco-latins

Il s'agit de plans factoriels à trois facteurs croisés (cas des carrés latins) ou à quatre facteurs croisés (cas des carrés gréco-latins), incomplets, équilibrés et correspondant au cas particulier où tous les facteurs ont le même nombre de niveaux.

3.1 Les plans en carrés latins

3.1.1 Contexte

On considère un modèle d'ANOVA à trois facteurs croisés, dans lequel chacun des facteurs possède le même nombre de niveaux, noté J. On suppose qu'il n'y a pas d'effets d'interactions dans le modèle ou, du moins, que ces effets ne sont pas d'un intérêt majeur pour l'expérimentateur (en effet, il ne sera pas possible de les prendre en compte dans le modèle). On ne va donc considérer que des modèles additifs et ne s'intéresser qu'aux effets principaux.

3.1.2 Principe

Dans chaque cellule correspondant au croisement de deux niveaux des deux premiers facteurs, on réalise une observation et une seule pour un niveau déterminé du troisième facteur. On a ainsi J^2 cellules observées parmi les J^3 possibles, et une seule observation dans chaque cellule observée : c'est un plan incomplet fractionnaire (de fraction $1/J$, voir la remarque 27).

Dans les plans en carrés latins, le choix du niveau du troisième facteur couplé avec chacune des cellules obtenues par croisement des deux premiers facteurs se fait au moyen d'une table appelée **carré latin**. Le principe d'une telle table est illustré dans l'exemple ci-dessous ; il s'agit d'avoir une fois et une seule chacun des niveaux de F_3 dans chaque ligne et dans chaque colonne de la table croisant F_1 et F_2, de telle sorte que le plan obtenu soit équilibré.

EXEMPLE 3 *Considérons le cas $J = 4$. La table ci-dessous est un carré latin.*

niveaux de F_2 →		*1*	*2*	*3*	*4*
	1	*a*	*b*	*c*	*d*
niveaux de F_1	*2*	*b*	*c*	*d*	*a*
	3	*c*	*d*	*a*	*b*
	4	*d*	*a*	*b*	*c*

Les lettres contenues dans la table représentent les niveaux du facteur F_3 ; on notera que les trois facteurs jouent des rôles symétriques.

Par permutation des lignes (sauf la première, qui représente un codage arbitraire des niveaux de F_3) et permutation des colonnes, on génère, à partir du carré ci-dessus, un sous-ensemble (de cardinal $3! \times 4! = 144$) de tous les carrés latins d'ordre quatre. Il existe quatre "modèles" de tables non équivalentes par permutations des lignes et des colonnes, donc 576 carrés latins d'ordre quatre au total (pour plus de détails, on pourra se reporter aux tables de Fisher & Yates, déjà mentionnées).

Remarque. — La dénomination de *carré latin* provient du fait que, pour symboliser un tel plan d'expériences, on utilise un carré dans lequel figure des lettres latines.

Remarque. — En fait, chacun des trois facteurs considérés ayant J niveaux, le plan complet comprend J^3 cellules dont on n'observe seulement que la fraction $\frac{1}{J}$, soit J^2 observations. Ce type de plans, croisant plusieurs facteurs ayant le même nombre de niveaux (souvent 2 ou 3), dans lesquels on n'observe qu'une fraction donnée des cellules observables, s'appelle un *plan fractionnaire*.

Remarque. — Jusqu'à présent, nous n'avons fait que décrire les carrés latins. De façon plus précise, un carré latin est un plan d'expériences qui vérifie les caractéristiques suivantes :

1. le plan croise trois facteurs, à J niveaux chacun ($J \geq 2$), ces facteurs jouant des rôles symétriques ;

2. le plan factoriel est incomplet, fractionnaire, de fraction $\frac{1}{J}$;
3. chaque niveau de chaque facteur est observé exactement J fois (de sorte qu'il s'agit d'un plan équilibré) ;
4. chaque niveau de chaque facteur est observé simultanément une fois et une seule avec chaque niveau de chacun des deux autres facteurs : c'est cette dernière propriété qui nécessite certaines précautions dans la construction d'un plan en carré latin et qui explique l'usage de tables comme celle donnée dans l'exemple ci-dessus.

Ainsi, la table ci-dessous ne constitue pas un carré latin, dans la mesure où elle permet de vérifier les propriétés 1, 2 et 3, mais pas la propriété 4 :

niveaux de $F_2 \rightarrow$		1	2	3	4
	1	a	a	a	a
niveaux de F_1	2	b	b	b	b
	3	c	c	c	c
	4	d	d	d	d

Remarque. — Enfin, signalons qu'un Sudoku est un carré latin d'ordre 9, dans lequel on a remplacé les lettres latines par les chiffres arabes de 1 à 9 et dans lequel on a rajouté une contrainte (n'ayant rien à voir avec les plans d'expériences) sur les 9 sous-carrés d'ordre 3 disjoints qui le composent. Le but du jeu est de retrouver l'intégralité d'un tel carré à partir d'une partie restreinte qui en est donnée.

Remarque. — En guise d'exercice sur les carrés latins, on trouvera en Annexe C un jeu mathéma-tique paru dans le quotidien "Le Monde".

3.1.3 Modèle

Comme indiqué plus haut, seul le modèle additif est pris en compte dans ce type de plan. En utilisant le paramétrage centré, ce modèle s'écrit sous la forme

$$Y_{jk\ell} = \mu + \alpha_j^1 + \alpha_k^2 + \alpha_\ell^3 + U_{jk\ell},$$

avec $\sum_{j=1}^J \alpha_j^1 = \sum_{k=1}^J \alpha_k^2 = \sum_{\ell=1}^J \alpha_\ell^3 = 0$, $U_{jk\ell} \sim \mathcal{N}(0, \sigma^2)$, les $U_{jk\ell}$ étant des v.a.r. indépendantes. Comme dans les modèles considérés dans le paragraphe 4.1, il n'y a pas d'indice i ici car on ne fait qu'une seule observation par cellule observée.

3.1.4 Tests des effets

Compte tenu de la structure particulière du plan dans ce cas, la décomposition de la somme des carrés totale est elle-même particulière. Cette décomposition est explicitée dans le tableau d'analyse de la variance ci-dessous (elle est analogue à celle obtenue, au chapitre précédent, pour les plans complets et équilibrés, dans le cas d'un modèle additif).

sources de variation	sommes des carrés	d.d.l.	carrés moyens
F_1	$SSF_1 = J\sum_{j=1}^J (\overline{y}_{j\bullet\bullet} - \overline{y}_{\bullet\bullet\bullet})^2$	$J-1$	$MSF_1 = \frac{SSF_1}{J-1}$
F_2	$SSF_2 = J\sum_{k=1}^J (\overline{y}_{\bullet k\bullet} - \overline{y}_{\bullet\bullet\bullet})^2$	$J-1$	$MSF_2 = \frac{SSF_2}{J-1}$
F_3	$SSF_3 = J\sum_{\ell=1}^J (\overline{y}_{\bullet\bullet\ell} - \overline{y}_{\bullet\bullet\bullet})^2$	$J-1$	$MSF_3 = \frac{SSF_3}{J-1}$
Erreur	SSE	$(J-1)(J-2)$	$MSE = \frac{SSE}{(J-1)(J-2)} = \hat{\sigma}^2$
Total	$SST = \sum_A (y_{jk\ell} - \overline{y}_{\bullet\bullet\bullet})^2$	$J^2 - 1$	—

Dans le tableau ci-dessus, la somme des carrés relative aux erreurs s'écrit :

$$SSE = \sum_A [y_{jk\ell} - (\overline{y}_{j\bullet\bullet} + \overline{y}_{\bullet k\bullet} + \overline{y}_{\bullet\bullet\ell} - 2\,\overline{y}_{\bullet\bullet\bullet})]^2$$

(A désigne le sous-ensemble de J^3 correspondant aux indices (j, k, ℓ) observés).

Les estimations des paramètres (lorsque les effets correspondants sont significatifs) sont données ci-dessous :

$$\hat{\mu} = \overline{y}_{\bullet\bullet\bullet} = \frac{1}{J^2}\sum_A y_{jk\ell} \; ; \;\; \hat{\alpha}_j^1 = \overline{y}_{j\bullet\bullet} - \overline{y}_{\bullet\bullet\bullet}, \text{ avec } \overline{y}_{j\bullet\bullet} = \frac{1}{J}\sum_{k=1}^J y_{jk\ell}$$

(pour un j et un k fixés, ℓ est aussi fixé) ;

$$\hat{\alpha}_j^2 = \overline{y}_{\bullet k\bullet} - \overline{y}_{\bullet\bullet\bullet} \; ; \;\; \hat{\alpha}_j^3 = \overline{y}_{\bullet\bullet\ell} - \overline{y}_{\bullet\bullet\bullet}.$$

3.1.5 Randomisation

Encore une fois, elle est quasiment systématique dans ce type de modèles. Elle se pratique à plusieurs niveaux :
- parmi les trois facteurs dont on dispose, on peut tirer au sort celui qui sera en lignes, celui qui sera en colonnes et celui qui sera le troisième ;
- on peut aussi tirer au sort les étiquettes des niveaux de chaque facteur ;
- partant d'un carré latin donné, on peut tirer au sort une permutation sur les lignes et une autre sur les colonnes ;
- enfin, on peut aussi tirer au sort l'affectation de chaque individu à une cellule.

3.2 Les plans en carrés gréco-latins

On les appelle également plans en *carrés latins orthogonaux*.

3.2.1 Contexte

Il s'agit maintenant d'une ANOVA à quatre facteurs croisés, chaque facteur possédant toujours le même nombre de niveaux noté J. Là encore, on ne fera intervenir aucune interaction dans les modèles envisagés qui seront donc tous additifs.

3.2.2 Principe

Sur les J^4 cellules envisageables, seules J^2 sont observées et elles le sont une seule fois (on considère donc la fraction $\frac{1}{J^2}$ du plan complet : c'est encore un plan fractionnaire). En fait, si l'on désigne par F_1, F_2, F_3 et F_4 les quatre facteurs considérés, (F_1, F_2, F_3) et (F_1, F_2, F_4) constituent chacun un carré latin, les différents niveaux de F_3 et de F_4 étant couplés une fois et une seule, de telle sorte que ces deux carrés latins soient "orthogonaux".

EXEMPLE 4 *Les carrés gréco-latins sont plus délicats à construire que les carrés latins. Nous donnons ci-dessous un exemple de carré gréco-latin d'ordre quatre ($J = 4$).*

niveaux de F_2 →		*1*	*2*	*3*	*4*
	1	$a\,\alpha$	$b\,\beta$	$c\,\gamma$	$d\,\delta$
niveaux de F_1	*2*	$c\,\beta$	$d\,\alpha$	$a\,\delta$	$b\,\gamma$
	3	$d\,\gamma$	$c\,\delta$	$b\,\alpha$	$a\,\beta$
	4	$b\,\delta$	$a\,\gamma$	$d\,\beta$	$c\,\alpha$

Les lettres latines contenues dans la table représentent les niveaux du facteur F_3, tandis que les lettres grecques représentent ceux du facteur F_4.

On notera encore que les quatre facteurs jouent des rôles symétriques et que, par permutation des lignes et des colonnes, on génère une partie (un tiers) de l'ensemble de tous les carrés gréco-latins d'ordre quatre (sur lesquels on trouvera encore d'autres précisions dans les tables de Fisher & Yates).

3.2.3 Modèle

Le modèle (additif) pour un plan en carré gréco-latin s'écrit, avec le paramétrage centré, sous la forme suivante :

$$Y_{jk\ell m} = \mu + \alpha_j^1 + \alpha_k^2 + \alpha_\ell^3 + \alpha_m^4 + U_{jk\ell m},$$

avec $\sum_{j=1}^J \alpha_j^1 = \sum_{k=1}^J \alpha_k^2 = \sum_{\ell=1}^J \alpha_\ell^3 = \sum_{m=1}^J \alpha_m^4 = 0$, $U_{jk\ell m} \sim \mathcal{N}(0, \sigma^2)$, les $U_{jk\ell m}$ étant des v.a.r. indépendantes.

Test de significativité du modèle ci-dessus, tests relatifs à chacun des effets principaux, estimation des paramètres dans le modèle retenu et randomisation généralisent de façon naturelle ce qui a été explicité dans le cas des carrés latins.

Remarque. — De façon claire, la dénomination de *carré gréco-latin* provient du fait que, pour symboliser un tel plan d'expériences, on utilise un carré dans lequel figure des lettres latines et des lettres grecques.

Remarque. — Ici encore, on peut préciser la définition des carrés gréco-latins : un carré gréco-latin est un plan d'expériences qui vérifie les caractéristiques suivantes :

1. le plan croise quatre facteurs, à J niveaux chacun ($J \geq 3$), ces facteurs jouant des rôles symétriques (on vérifie, de façon immédiate, qu'il n'existe pas de carré gréco-latin d'ordre 2) ;
2. le plan factoriel est incomplet, fractionnaire, de fraction $\frac{1}{J^2}$;

3. chaque niveau de chaque facteur est observé exactement J fois (il s'agit encore d'un plan équilibré) ;
4. chaque niveau de chaque facteur est observé simultanément une fois et une seule avec chaque niveau de l'un quelconque des trois autres facteurs.

Remarque. — En partant d'un carré gréco-latin, si on supprime l'un quelconque des quatre facteurs, le plan d'expériences ainsi obtenu est un carré latin. Si on en supprime deux (ou si on supprime un facteur dans un carré latin), le plan obtenu est un plan à deux facteurs croisés à J niveaux chacun, complet, équilibré, sans répétition.

Remarque. — Dans un carré latin ou dans un carré gréco-latin, un (voire plusieurs) des facteurs peut être un bloc.

Remarque. — Nous introduirons au chapitre 6 les modèles mixtes faisant intervenir à la fois des effets fixes (comme dans tous les modèles envisagés jusqu'ici) et des effets aléatoires. Dans les plans en carrés latins ou en carrés gréco-latins, on peut très bien considérer que certains des trois ou quatre facteurs intervant sont à effets aléatoires, ce qui conduit à généraliser le modèle ci-dessus à un modèle mixte.

Remarque. — Donnons quelques précisions sur l'existence des carrés latins et gréco-latins. On peut construire des carrés latins d'ordre J (J niveaux pour chaque facteur), pour toute valeur de J à partir de 2. Les tables de Fisher & Yates en donnent toutes les configurations pour J variant de 4 à 12. Concernant les carrés gréco-latins, on ne peut en construire qu'à partir de $J = 3$. De plus, il faut noter la particularité suivante : il n'existe pas de carré gréco-latin d'ordre six [1]. Les tables de Fisher & Yates en donnent toutes les configurations pour J vallant 3, 4 et 5, puis pour J vallant 7, 8 et 9.

1. Cette particularité remonte au mathématicien suisse Leonhard EULER (1707–1783) qui chercha vainement à disposer 36 officiers de 6 grades différents (facteur F_3), appartenant à 6 régiments différents (facteur F_4), dans une grille à 6 lignes (facteur F_1) et 6 colonnes (facteur F_2), de telle sorte que chaque grade et chaque régiment apparaisse une fois et une seule dans chaque ligne et dans chaque colonne. En 1782, Euler conjectura le résultat suivant : il n'existe pas de carré gréco-latin d'ordre $J = 4k + 2, k \in \mathbb{N}$. Le résultat est trivial pour $k = 0$, soit $J = 2$. Pour $k = 1$, soit $J = 6$, il n'a été démontré qu'en 1900, par Gaston Tarry, mathématicien français né à Villefranche de Rouergue. Par contre, cette conjecture est fausse pour $k \geq 2$ et ce n'est qu'en 1960 que Bose *et al.* ont montré la possibilité de construire des carrés gréco-latins pour tout $J > 2$, sauf pour $J = 6$.

4 Les plans à plusieurs facteurs à deux niveaux

4.1 Introduction

Dans la pratique des plans d'expériences, il est fréquent d'avoir un nombre important de facteurs à prendre en compte (par exemple plus de cinq) tout en étant limité dans le nombre d'observations réalisables (problèmes de coût, de temps...).

En fait, il existe des plans d'expériences spécifiques pour répondre à ce type de situations. Pour ce faire, on se limite en général à des facteurs à deux niveaux (*bas* et *haut*) ou, éventuellement, à trois niveaux (*bas*, *intermédiaire* et *haut*). Dans ce paragraphe, nous ne traiterons que des facteurs à deux niveaux et nous noterons p le nombre total de facteurs pris en compte.

4.2 Cas $p = 2$

Il s'agit d'un cas très simple d'analyse de variance à deux facteurs croisés, chacun à deux niveaux. Nous renvoyons donc au chapitre 3 pour plus de détails. Notons seulement les deux particularités suivantes :

- les écritures des tests et des estimations se simplifient, dans la mesure où l'on a $J = K = 2$; en particulier, les effets de chaque facteur étant supposés centrés, il y a un seul paramètre à estimer par facteur ;
- dans le cas d'un plan complet, équilibré, sans répétition (une seule observation par cellule, soit quatre au total), on ne peut pas faire de test si l'on estime les interactions (il y a, dans ce cas, quatre paramètres à estimer et le modèle est saturé) ; on doit donc supprimer les interactions si l'on veut pouvoir faire des tests.

4.3 Cas $p = 3$

On a donc ici $J = K = L = 2$; pour le cas général, nous renvoyons encore au chapitre 3. Dans ce paragraphe, nous allons nous intéresser, tout d'abord, au cas d'un *plan complet, équilibré, sans répétition* (huit observations au total, soit une par cellule), puis au cas d'un plan incomplet à 4 observations.

4.3.1 Plan d'expériences complet sans répétition

Les huit observations se présentent sous la forme suivante :

$\ell \to$	1		2	
$j \quad k \to$	1	2	1	2
1	y_{111}	y_{121}	y_{112}	y_{122}
2	y_{211}	y_{221}	y_{212}	y_{222}

4.3.2 Modèle

Si l'on prend en compte l'effet général (un paramètre), les effets principaux de chacun des trois facteurs (trois paramètres en tout) et les effets d'interactions d'ordre deux (3 paramètres), avec le paramètre de variance, on arrive à huit paramètres, soit le nombre d'observations. Le modèle prenant en compte ces trois types d'effets est donc un modèle saturé et il n'est pas possible, dans ce cas, d'estimer ou de tester les interactions d'ordre trois. Ce modèle s'écrit, selon le paramétrage centré :

$$Y_{jk\ell} = \mu + \alpha^1_j + \alpha^2_k + \alpha^3_\ell + \gamma^{12}_{jk} + \gamma^{13}_{j\ell} + \gamma^{23}_{k\ell} + U_{jk\ell}.$$

Il est clair qu'un modèle additif, sans aucune interaction, est plus approprié à ce dispositif expéri-mental.

4.3.3 Tableau d'analyse de la variance pour le modèle additif

Compte tenu que chacun des trois facteurs pris en compte n'a que deux niveaux, les moyennes partielles, servant à estimer et à tester les différents effets, ont des expressions particulières. Par exemple, on a, par définition :

$$\begin{aligned}
\overline{y}_{jk\bullet} &= \frac{1}{2}(y_{jk1} + y_{jk2})\ ; \quad \overline{y}_{j\bullet\bullet} = \frac{1}{4}\sum_{k=1}^{2}\sum_{\ell=1}^{2} y_{jk\ell}\ ; \\
\overline{y}_{\bullet\bullet\bullet} &= \frac{1}{8}\sum_{j=1}^{2}\sum_{k=1}^{2}\sum_{\ell=1}^{2} y_{jk\ell} = \frac{1}{2}(\overline{y}_{1\bullet\bullet} + \overline{y}_{2\bullet\bullet}).
\end{aligned}$$

On en déduit des égalités du type :

$$\overline{y}_{1\bullet\bullet} - \overline{y}_{\bullet\bullet\bullet} = \frac{1}{2}(\overline{y}_{1\bullet\bullet} - \overline{y}_{2\bullet\bullet}) = -(\overline{y}_{2\bullet\bullet} - \overline{y}_{\bullet\bullet\bullet}).$$

Pour le tableau d'analyse de la variance, on part de l'égalité suivante :

$$\begin{aligned}
y_{jk\ell} - \overline{y}_{\bullet\bullet\bullet} &= y_{jk\ell} - (\overline{y}_{j\bullet\bullet} + \overline{y}_{\bullet k\bullet} + \overline{y}_{\bullet\bullet\ell} - 2\,\overline{y}_{\bullet\bullet\bullet}) + (\overline{y}_{j\bullet\bullet} - \overline{y}_{\bullet\bullet\bullet}) \\
&+ (\overline{y}_{\bullet k\bullet} - \overline{y}_{\bullet\bullet\bullet}) + (\overline{y}_{\bullet\bullet\ell} - \overline{y}_{\bullet\bullet\bullet}).
\end{aligned}$$

On en déduit :

$$\begin{aligned}
\sum_{j=1}^{2}\sum_{k=1}^{2}\sum_{\ell=1}^{2}(y_{jk\ell} - \overline{y}_{\bullet\bullet\bullet})^2 &= \sum_{j=1}^{2}\sum_{k=1}^{2}\sum_{\ell=1}^{2}[y_{jk\ell} - (\overline{y}_{j\bullet\bullet} + \overline{y}_{\bullet k\bullet} + \overline{y}_{\bullet\bullet\ell} - 2\,\overline{y}_{\bullet\bullet\bullet})]^2 \\
&+ 4\sum_{j=1}^{2}(\overline{y}_{j\bullet\bullet} - \overline{y}_{\bullet\bullet\bullet})^2 + 4\sum_{k=1}^{2}(\overline{y}_{\bullet k\bullet} - \overline{y}_{\bullet\bullet\bullet})^2 \\
&+ 4\sum_{\ell=1}^{2}(\overline{y}_{\bullet\bullet\ell} - \overline{y}_{\bullet\bullet\bullet})^2
\end{aligned}$$

(le plan étant équilibré, on vérifie sans difficulté la nullité des doubles produits). Cette décomposi-tion peut se réécrire sous la forme

$$SST = SSE + SSF_1 + SSF_2 + SSF_3,$$

où SST et SSE désignent respectivement la somme des carrés totale et la somme des carrés relative aux erreurs dans la modèle additif, tandis que SSF_1, SSF_2 et SSF_3 désignent respectivement les sommes des carrés relatives à chacun des trois facteurs. On notera que les égalités du type

$$SSF_1 = 4\sum_{j=1}^{2}(\overline{y}_{j\bullet\bullet} - \overline{y}_{\bullet\bullet\bullet})^2 = 2\,(\overline{y}_{1\bullet\bullet} - \overline{y}_{2\bullet\bullet})^2$$

découlent des égalités signalées plus haut, ce qui permet de simplifier le calcul de ces sommes de carrés.

On obtient encore le tableau d'analyse de la variance sous la forme suivante :

sources de variation	sommes des carrés	d.d.l.	carrés moyens	valeurs des statistiques de Fisher
F_1	SSF_1	1	SSF_1	$\dfrac{SSF_1}{MSE}$
F_2	SSF_2	1	SSF_2	$\dfrac{SSF_2}{MSE}$
F_3	SSF_3	1	SSF_3	$\dfrac{SSF_3}{MSE}$
Erreur	SSE	4	$MSE = \dfrac{SSE}{4} = \hat{\sigma}^2$	—
Total	SST	7	—	—

Remarque. — Dans les généralisations à plus de trois facteurs, on garde le caractère équilibré des plans d'expériences étudiés et donc ce type de tableau de décomposition.

4.3.4 Plan d'expériences incomplet, fractionnaire, équilibré

Si l'on souhaite juste estimer les effets μ, α_j^1, α_k^2 et α_ℓ^3, sans faire de test ni calculer d'erreur-type, on peut, à la limite, se contenter de quatre observations (deux par niveau de chaque facteur) selon le dispositif incomplet, fractionnaire de fraction $\frac{1}{2}$, suivant :

observations	F_1	F_2	F_3
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1

(les chiffres à l'intérieur de la table désignent les niveaux des facteurs à prendre en compte).

Dans l'écriture vectorielle du modèle additif (selon le paramétrage centré), sous la forme $Y = \mathbf{X}_c\beta_c + U$, le vecteur β_c a pour transposé $\beta_c' = (\mu \; \alpha_1^1 \; \alpha_1^2 \; \alpha_1^3)$ et la matrice $\mathbf{X}_c$ s'écrit :

$$\mathbf{X}_c = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

On notera, tout d'abord, que les trois dernières colonnes de la matrice $\mathbf{X}_c$ correspondent au dispositif présenté au-dessus, en remplaçant la valeur 2 par la valeur -1 (ces trois colonnes sont ainsi centrées). Ensuite, les colonnes de cette matrice sont deux à deux orthogonales, ceci n'étant possible que parce que le plan est équilibré (il s'agit d'une condition nécessaire, mais pas suffisante). Enfin, on notera que ce dispositif est associé à ce qu'on appelle la table L_4 que nous présentons ci-dessous (1 est remplacé par $+$ et -1 par $-$) et dont la version L_8 sera détaillée au point suivant.

La table L_4 :

	F_1	F_2	F_3
1	+	+	+
2	+	−	−
3	−	+	−
4	−	−	+

4.4 Cas $4 \leq p \leq 6$

Dans la pratique, ce cas est, bien sûr, plus intéressant que les précédents dans la mesure où l'on peut raisonablement envisager, tant que l'on a $p \leq 3$, un plan complet, équilibré avec au moins deux répétitions. Au delà, c'est d'autant plus difficile que p augmente.

4.4.1 Dispositif expérimental

L'idée est ici de généraliser ce qui a été fait plus haut pour trois facteurs et seulement quatre observations. Ainsi, avec huit observations, il est possible d'étudier des modèles additifs comportant jusqu'à six facteurs explicatifs, chacun ayant toujours seulement deux niveaux. En effet, il y a un paramètre à estimer par facteur, un paramètre pour l'effet général et un paramètre pour la variance. Toutefois, le dispositif expérimental est assez spécifique (même s'il n'est pas unique, voir plus bas) et les observations se présentent, par exemple, sous la forme décrite dans le tableau ci-dessous.

observations	F_1	F_2	F_3	F_4	F_5	F_6
1	1	1	1	1	1	1
2	1	1	2	1	2	2
3	1	2	1	2	2	1
4	1	2	2	2	1	2
5	2	1	1	2	1	2
6	2	1	2	2	2	1
7	2	2	1	1	2	2
8	2	2	2	1	1	1

4.4.2 Modèle

En considérant toujours le paramétrage centré, le modèle avec six facteurs s'écrit sous la forme :

$$Y_{hijk\ell m} = \mu + \alpha_h^1 + \alpha_i^2 + \alpha_j^3 + \alpha_k^4 + \alpha_\ell^5 + \alpha_m^6 + U_{hijk\ell m}.$$

Sous forme matricielle, il s'écrit encore $Y = \mathbf{X}_c\beta_c + U$, avec $\beta_c' = (\mu\ \alpha_1^1\ \alpha_1^2\ \alpha_1^3\ \alpha_1^4\ \alpha_1^5\ \alpha_1^6)$ et :

$$\mathbf{X}_c = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & 1 & -1 \\ 1 & -1 & 1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 \end{pmatrix}.$$

On notera les particularités suivantes de la matrice $\mathbf{X}_c$:

- à l'exception de la première colonne, qui ne contient que des 1, les autres colonnes correspondent à celles du dispositif présenté plus haut, en remplaçant chaque valeur 2 par la valeur -1 ;
- toujours à l'exception de la première colonne, les autres colonnes de $\mathbf{X}_c$ sont centrées (parce que le plan est équilibré) ;
- les colonnes de $\mathbf{X}_c$ sont deux à deux orthogonales (pour obtenir ce résultat, il est nécessaire, mais pas suffisant, que le plan soit équilibré).

4.4.3 Estimations et tests

Dans le modèle avec six facteurs, l'estimation de l'effet général μ est $\hat{\mu} = \overline{y}$ (moyenne des huit observations).

L'estimation de α_1^s ($s = 1, \dots, 6$) est :

$$\hat{\alpha}_1^s = \overline{y}_1^s - \overline{y} = \frac{\overline{y}_1^s - \overline{y}_2^s}{2},$$

où $\overline{y}_1^s$ (respectivement $\overline{y}_2^s$) est la moyenne des quatre observations réalisées au niveau 1 (respectivement au niveau 2) de F_s et où $\overline{y}$ vérifie $\overline{y} = \dfrac{\overline{y}_1^s + \overline{y}_2^s}{2}$.

Enfin, l'estimation de σ^2 vérifie

$$\hat{\sigma}^2 = \|\hat{U}\|^2 = \|Y - \mathbf{X}_c\hat{\beta}_c\|^2, \text{ avec } \hat{\beta}_c{}' = (\hat{\mu}\ \hat{\alpha}_1^1\ \hat{\alpha}_1^2\ \hat{\alpha}_1^3\ \hat{\alpha}_1^4\ \hat{\alpha}_1^5\ \hat{\alpha}_1^6)$$

(il n'y a pas de dénominateur car le degré de liberté de $\hat{U}$ est 1).

Pour le test de significativité du facteur F_s ($s = 1, \dots, 6$), équivalent au test de l'hypothèse nulle $\{H_0 : \alpha_1^s = 0\}$, la valeur de la statistique du test est :

$$f = \frac{2\,(\overline{y}_1^s - \overline{y}_2^s)^2}{\hat{\sigma}^2}.$$

4.4.4 Base orthogonale de $\mathbb{R}^8$

Considérons maintenant l'espace vectoriel réel $\mathbb{R}^8$ muni de la base canonique $\mathcal{C}$ et de la métrique identité (associée à la matrice identité d'ordre 8 relativement à $\mathcal{C}$). Considérons d'autre part la matrice $\mathbf{P}$ ci-dessous, carrée d'ordre

8 :

$$\mathbf{P} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \end{pmatrix}.$$

On peut vérifier que les colonnes de $\mathbf{P}$ sont deux à deux orthogonales. Il s'ensuit que $\mathbf{P}$ est régulière, est donc une matrice de passage (de la base canonique à une nouvelle base, notée $\mathcal{B}$), que les coordonnées de cette nouvelle base sur $\mathcal{C}$ sont fournies par les colonnes de $\mathbf{P}$ et que $\mathcal{B}$ est une base orthogonale de $\mathbb{R}^8$. On notera encore que $\mathbf{P}$ est obtenue en rajoutant une colonne orthogonale à la matrice $\mathbf{X}_c$ définie plus haut et que ses colonnes sont centrées, à l'exception de la première. On obtient ainsi la propriété suivante.

PROPOSITION 5. — *Tout plan qui expérimente six facteurs avec huit observations, de telle sorte que la matrice d'incidence associée soit constituée du premier vecteur-colonne de* $\mathbf{P}$ *et de six autres de ses vecteurs-colonnes (n'importe lesquels, d'où la non unicité du plan) permet d'étudier les effets de ces six facteurs dans les mêmes conditions qu'avec le modèle défini plus haut.*

4.4.5 Notion de confusion des effets et triangle des interactions

Supposons que, dans le modèle présenté plus haut, on souhaite prendre en considération un effet d'interaction, par exemple entre les deux facteurs F_1 et F_2. Cet effet sera mesuré par un paramètre noté γ^{12} et l'on réécrira le modèle sous la forme matricielle suivante :

$$Y = \mathbf{X}^*\beta^* + U^*, \text{ avec } \mathbf{X}^* = (\mathbf{X}_c | X^{12}) \text{ et } \beta^* = \begin{pmatrix} \beta_c \\ \overline{\gamma^{12}} \end{pmatrix}.$$

De façon standard, la colonne X^{12} est obtenue par produit des colonnes X^1 et X^2 et vaut :

$$X^{12} = X^1 \times X^2 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \\ -1 \\ -1 \\ 1 \\ 1 \end{pmatrix} = X^4,$$

colonne associée au facteur F_4. Ainsi, on ne pourra estimer dans ce modèle ni α_1^4 (effet de F_4) ni γ^{12} (effet d'interaction entre F_1 et F_2), mais seulement la somme $\alpha_1^4 + \gamma^{12}$. On dit qu'il y a confusion de ces deux effets. Avec le plan à seulement 8 observations considéré ici, il faut donc choisir, entre ces deux effets, celui qui sera introduit dans le modèle.

Pour savoir à quel facteur correspond toute interaction d'ordre deux dans ce modèle, on dresse un tableau, appelé *triangle des interactions*, qui fournit le résultat. Dans le tableau ci-dessous, on a considéré 7 facteurs (voir la remarque 39). De plus, pour une meilleure lisibilité, on a remplacé les facteurs F_1 à F_7 par les lettres de a à g. Pour la même raison, on a fait figurer le tableau entier (un carré) plutôt que sa partie supérieure ou inférieure (un triangle).

	a	b	c	d	e	f	g
a	–	d	f	b	g	c	e
b	d	–	e	a	c	g	f
c	f	e	–	g	b	a	d
d	b	a	g	–	f	e	c
e	g	c	b	f	–	d	a
f	c	g	a	e	d	–	b
g	e	f	d	c	a	b	–

La lecture du tableau est très simple : il y a confusion de l'effet d'interaction entre a et b et de l'effet du facteur d ; de même pour l'interaction entre c et e et le facteur b...

Remarque. — **: Où l'algèbre rejoint la statistique...** Considérons l'ensemble

$$H = \{1, a, b, c, d, e, f, g\}$$

muni d'une opération interne, notée $*$, dont le tableau ci-dessus est la table, et rajoutons lui les propriétés suivantes :
• 1 est l'élément neutre de $*$: $\forall x \in H, 1 * x = x * 1 = x$;
• le produit de tout élément avec lui-même est égal à 1 : $\forall x \in H, x * x = 1$.

Alors, H est un groupe commutatif pour l'opération $*$ (immédiat).

Nous ne développerons pas davantage les liens entre plans d'expériences et algèbre, mais il est clair que l'étude des propriétés mathématiques des plans d'expériences nécessite un large usage de l'algèbre (voir, par exemple, Collombier, 1996).

4.4.6 La table L_8

On appelle ainsi la table à 8 lignes et 7 colonnes obtenue à partir de la matrice $\mathbf{P}$ définie plus haut en supprimant la première colonne (qui ne correspond pas à un facteur) et en remplaçant $+1$ par $+$ (niveau *haut* du facteur) et -1 par $-$ (niveau *bas* du facteur). Elle indique comment expérimenter de 4 à 6 facteurs à deux niveaux (en supprimant trois, deux ou une colonne) avec seulement 8 observations. Nous donnons ci-dessous un modèle de table L_8.

	a	b	c	d	e	f	g
1	$+$	$+$	$+$	$+$	$+$	$+$	$+$
2	$+$	$+$	$-$	$+$	$-$	$-$	$-$
3	$+$	$-$	$+$	$-$	$-$	$+$	$-$
4	$+$	$-$	$-$	$-$	$+$	$-$	$+$
5	$-$	$+$	$+$	$-$	$+$	$-$	$-$
6	$-$	$+$	$-$	$-$	$-$	$+$	$+$
7	$-$	$-$	$+$	$+$	$-$	$-$	$+$
8	$-$	$-$	$-$	$+$	$+$	$+$	$-$

Il existe différentes tables L_8, équivalentes du point de vue de l'expérimentation. L'une d'entre elles est obtenue, à partir d'une ligne fixe appelée *générateur*, par un procédé systématique de permutations circulaires (ce procédé se généralisant aux tables plus importantes décrites dans le point suivant). En fait, le générateur comportant quatre signes $-$ pour trois signes $+$, et les permutations circulaires d'un ensemble de sept éléments ne permettant de générer que sept lignes, on doit rajouter une ligne (la première) ne comportant que des $+$. Nous donnons ci-dessous cette table.

	a	b	c	d	e	f	g
1	$+$	$+$	$+$	$+$	$+$	$+$	$+$
2	$-$	$-$	$-$	$+$	$-$	$+$	$+$
3	$+$	$-$	$-$	$-$	$+$	$-$	$+$
4	$+$	$+$	$-$	$-$	$-$	$+$	$-$
5	$-$	$+$	$+$	$-$	$-$	$-$	$+$
6	$+$	$-$	$+$	$+$	$-$	$-$	$-$
7	$-$	$+$	$-$	$+$	$+$	$-$	$-$
8	$-$	$-$	$+$	$-$	$+$	$+$	$-$

Remarque. — Si l'on se contente d'estimations ponctuelles, sans faire de test ni construire d'intervalle de confiance, la table L_8 indique le plan à considérer pour étudier jusqu'à 7 facteurs avec un modèle sans interaction (ou 6 facteurs avec une interaction...).

4.5 Cas $p > 6$

Comme on a construit une table L_4 et une table L_8 (et même plusieurs, mais équivalentes), il est possible de construire une table L_{12}, une table $L_{16}\ldots$ L'indice de ces tables est le nombre total d'observations réalisées ($n = 4$; $n = 8$; $n = 12$; $n = 16\ldots$) et est un multiple de 4 : $n = 4 \times m,\ m \in \mathbb{N}^*$. La valeur 4 correspond au carré du nombre de niveaux considérés (2) et permet de conserver des plans équilibrés lorsqu'on augmente le nombre d'observations, ainsi que le caractère orthogonal des colonnes de $\mathbf{X}_c$.

La table L_{12} permet de construire un plan avec lequel on peut étudier jusqu'à dix facteurs sans interaction (onze si l'on ne fait que des estimations). Le triangle des interactions se généralise et permet d'introduire certaines interactions, à condition de supprimer les facteurs correspondant à la confusion des effets. De même, la table L_{16} permet d'étudier jusqu'à 14 facteurs...

Tous ces plans sont appelés des *plans de Plackett et Burman*, car ils ont été introduits pour la première fois par ces deux auteurs (Plackett & Burman, 1946). Ils présentent tous la particularité d'être des *plans incomplets, équilibrés et orthogonaux* :
– ils sont incomplets, puisque p facteurs à 2 niveaux nécessitent un minimum de 2^p observations pour obtenir un plan complet et qu'on en est loin dans les exemples présentés ici ;
– ils sont équilibrés, car chaque niveau (*bas* ou *haut*) de chaque facteur est

observé $\frac{n}{2}$ fois, si n est le nombre total d'observations du plan considéré ;
- ils sont orthogonaux, car la matrice d'incidence $\mathbf{X}_c$ associée à chacun de ces plans est orthogonale (ses colonnes sont deux à deux orthogonales).

Remarque. — Les plans de Plackett et Burman étudiés dans ce paragraphe sont, pour certains d'entre eux, des plans fractionnaires. En effet, avec p facteurs comportant tous deux niveaux ($p \geq 3$), un plan complet sans répétition doit comporter 2^p observations. Un plan fractionnaire n'expérimentera qu'une fraction déterminée, $\frac{1}{2^k}$, de ces observations, de sorte que l'on fera seulement 2^{p-k} observations ($k = 1, \ldots, p-2$). Il s'agit des plans obtenus avec les tables L_4, L_8, L_{16}... Les plans de Plackett et Burman offrent donc plus de possibilités en proposant également des tables telles que L_{12}.

5 Compléments

Dès l'introduction de ce chapitre, nous avons indiqué que les dispositifs expérimentaux permettant de répondre à des situations concrètes particulières sont extrêmement nombreux et que nous nous sommes volontairement contentés ici de présenter les plus courants. Nous signalons ci-dessous quelques-uns des autres dispositifs également courants.

Tout d'abord, de façon analogue à ce qui a été développé pour des facteurs à deux niveaux, on peut considérer des plans incomplets, équilibrés et orthogonaux pour plusieurs facteurs à trois niveaux. Il existe encore des tables permettant de construire les dispositifs expérimentaux correspondant. Ainsi, la table L_9 permet d'étudier jusqu'à 3 facteurs à trois niveaux (et même 4, si l'on n'estime pas σ^2 et qu'on ne fait pas de tests) avec seulement 9 observations. La table L_{18} permet d'étudier jusqu'à 8 facteurs à trois niveaux avec 18 observations. La table L_{27} permet d'étudier jusqu'à 12 facteurs à trois niveaux (éventuellement 13), avec 27 observations... De la même manière, il existe des dispositifs permettant d'étudier simultanément plusieurs facteurs à deux niveaux et plusieurs facteurs à trois niveaux.

Un autre dispositif courant dans la pratique est le plan dit en *cross-over*. Dans sa version la plus simple, ce dispositif consiste à étudier un facteur à deux niveaux (par exemple, traitement et placébo) en deux périodes. Dans la première période, une partie de l'échantillon considéré reçoit le traitement, l'autre partie recevant le placebo. Dans la deuxième période, les rôles sont inversés (autrement dit croisés, d'où le terme de *cross-over*). Comme dans le cas des blocs, on réduit ainsi la variabilité résiduelle, ce qui permet d'améliorer l'efficacité du dispositif.

Pour mémoire, signalons encore d'autre dispositifs : les plans *split plot*, ou en parcelles divisées, les plans Taguchi, les surfaces de réponses... Pour plus de détails sur ces dispositifs, on se reportera, par exemple, à Azaïs & Bardet (2005), à Droesbeke *et al.* (1997), à John (1998) ou à Saporta (2006).