

Plans factoriels complets, plans fractionnaires

Cas des facteurs ayant deux modalités

Frédéric Bertrand¹

¹IRMA, Université de Strasbourg
Strasbourg, France

ENSAI 3^e Année
2017-2018

Sommaire

Sommaire

- 1 Plans factoriels complets
 - Données, notations
 - Estimation
 - Analyse des résultats
- 2 Principe des plans fractionnaires
 - Introduction
 - Construction en trois étapes
 - Propriétés du plan proposé

- 3 Étude des confusions des plans fractionnaires

- Une confusion en engendre d'autres
- Lister les confusions induites par une confusion donnée
- Choix des confusions ; cas de plusieurs confusions

- 4 Choix et dépouillement d'un plan fractionnaire

- Nombre de facteurs étudiés et nombre d'essais
- Dépouillement des résultats
- Essais complémentaires

Sources

Ce cours s'appuie sur les références :

- D. Collombier, *Plans d'expérience factoriels : construction et propriétés des fractions de plans*, aux éditions Springer, 1996.
- F. Husson et J. Pagès, *Statistiques générales pour utilisateurs*, aux PUR, 2005.
- le livre de Douglas C. Montgomery, **Design and Analysis of Experiments**, 7th Edition, aux éditions Wiley, 2009.

Sommaire

1 Plans factoriels complets

2 Principe des plans fractionnaires

3 Étude des confusions des plans fractionnaires

4 Choix et dépouillement d'un plan fractionnaire

Notations, modèles et matrice des effets

Introduction

Introduction

Ces plans présentent les caractéristiques suivantes :

- ce sont les plus simples sur le plan technique ; ils se prêtent donc bien à l'introduction d'éléments méthodologiques ;
- ils sont souvent utilisés ;
- il est nécessaire de les connaître pour aborder les plans fractionnaires.

Introduction

Considérons le plan complet à 3 facteurs ayant chacun 2 modalités.

Ce plan comporte 2^3 essais et est souvent noté 2^3 .

Plus généralement, le plan complet à p facteurs à 2 modalités est noté 2^p .

Introduction

Voici la seconde :

F1	F2	F3	Y
1	1	1	Y_{111}
1	1	2	Y_{112}
1	2	1	Y_{121}
1	2	2	Y_{122}
2	1	1	Y_{211}
2	1	2	Y_{212}
2	2	1	Y_{221}
2	2	2	Y_{222}

TABLE – 1. Notations pour le plan 2^3 .

Choix d'un modèle et matrice des effets

Deux modèles pour analyser ces données

Le modèle complet

Il contient tous les effets : 3 effets principaux, 3 interactions d'ordre 2 et l'interaction d'ordre 3.

Puisqu'il n'y a pas de répétitions, ce modèle est saturé (autant de paramètres que de données, ici $8 = 7 +$ la constante). Soit :

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}$$

Utiliser un tel modèle revient à considérer que la variance résiduelle est négligeable devant les effets des facteurs. De façon pragmatique, ce peut-être une première étape dans l'analyse des résultats.

Le modèle réduit d'ordre 2

Il ne contient que les effets principaux et les interactions d'ordre 2.

Il revient à négliger a priori toute interaction d'ordre supérieur à 2, ce qui s'avère raisonnable en pratique. L'interaction d'ordre 3 du modèle précédent est vue maintenant comme le résidu.

Soit :

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \epsilon_{ijk}$$

Dans ce modèle à 3 facteurs, la variance résiduelle est estimée avec un seul degré de liberté.

Le modèle réduit d'ordre 2 (suite)

Ceci est très peu : pour s'en convaincre, on peut consulter quelques valeurs de la loi de Student (utilisée pour le test de signification d'un coefficient) : les deux valeurs suivantes $t_{1;0,975} = 12,71$ et $t_{3;0,975} = 3,18$ montrent une grande différence entre 1 et 3 degrés de liberté.

Le modèle réduit d'ordre 2 (suite)

En raccourci, on peut dire que pour être significatif, un coefficient doit être 4 fois plus grand avec 1 ddl qu'avec 3 ddl.

Aussi, les tests sont-ils beaucoup moins puissants (i.e. aptes à mettre en évidence des effets significatifs) avec 1 ddl qu'avec 3 : ces deux situations sont donc très différentes en pratique.

Dès que le nombre de facteurs p croît au-delà de trois, les résidus du modèle réduit d'ordre 2 présentent rapidement beaucoup plus de degrés de liberté (5 ddl pour $p = 4$; 16 ddl pour $p = 5$).

Matrice des effets

On appelle matrice des effets la matrice X de l'écriture du modèle $Y = X\beta + \epsilon$. Dans le cas de facteurs à deux modalités, cette matrice présente des propriétés remarquables.

Effets principaux

Pour les effets principaux, par exemple F_1 , la contrainte $\sum_j \alpha_j = 0$ conduit à $\alpha_2 = -\alpha_1$; chaque effet principal est représenté par une seule colonne qui ne contient donc que les valeurs 1 et -1.

Interactions

Pour les interactions, par exemple entre F_1 et F_2 , les contraintes $\sum_i (\alpha\beta)_{ij} = 0$ et $\sum_j (\alpha\beta)_{ij} = 0$ conduisent à exprimer tous les paramètres en fonction d'un seul (d'où un seul degré de liberté associé à une interaction et donc une seule colonne dans X), en l'occurrence $(\alpha\beta)_{11}$ (cf. Tableau 12.2).

En outre ces contraintes illustrent bien la notion d'interaction : si, pour $F_1 = 1$, c'est la modalité 1 de F_2 qui est meilleure ($(\alpha\beta)_{11} > 0$) alors c'est l'inverse pour $F_1 = 2$ ($(\alpha\beta)_{21} = -(\alpha\beta)_{11} < 0$).

Matrice des effets

Il en résulte que dans la matrice des effets, la colonne correspondant à une **interaction** s'obtient en **multipliant** entre elles les **colonnes des effets principaux correspondants**, propriété très utilisée en pratique.

Plus généralement, la colonne d'une interaction d'ordre quelconque est obtenue en multipliant entre elles toutes les colonnes des effets principaux correspondant.

Tout ceci n'est valable que pour des facteurs à deux modalités : pour des facteurs à plus de modalités, plusieurs colonnes (autant que de ddl) sont nécessaires pour représenter un effet.

Matrice des effets

$$F_1 \quad \begin{array}{c} 1 \\ 2 \\ \sum \end{array} \quad \begin{array}{c} 1 \\ (\alpha\beta)_{11} \\ -(\alpha\beta)_{11} \\ 0 \end{array} \quad \begin{array}{c} 2 \\ -(\alpha\beta)_{11} \\ (\alpha\beta)_{11} \\ 0 \end{array} \quad \begin{array}{c} \sum \\ 0 \\ 0 \\ 0 \end{array}$$

TABLE – 2. Relations entre les coefficients d'interaction dans le cas de facteurs à deux modalités

Matrice des effets

Dans le cas du plan 2^3 , on obtient ainsi la matrice du tableau 3 reproduit sur le transparent suivant.

Les trois premières colonnes (qui correspondent aux effets principaux) sont celles de la matrice des essais à condition de coder 2 (dans le tableau 1) par -1 .

Matrice des effets

Les colonnes représentant les interactions sont obtenues par multiplication de celles des effets principaux correspondants.

On ajoute à la fin la colonne constante.

Matrice des effets

		Effets							
		1	2	3	12	13	23	123	1
1	+1	+1	+1	+1	+1	+1	+1	+1	+1
2	+1	+1	-1	-1	+1	-1	-1	+1	+1
3	+1	-1	+1	-1	+1	-1	-1	+1	+1
4	+1	-1	-1	-1	-1	+1	+1	+1	+1
5	-1	+1	+1	-1	-1	+1	-1	-1	+1
6	-1	+1	-1	-1	+1	-1	+1	+1	+1
7	-1	-1	+1	+1	-1	-1	-1	+1	+1
8	-1	-1	-1	+1	+1	+1	-1	-1	+1

TABLE – 3. Matrice X associée au modèle saturé du plan 2^3

Matrice des effets

Enfin, le **carré de la norme euclidienne d'une colonne** est égal au **nombre d'essais** n .

Toutes ces propriétés peuvent être résumées dans l'écriture matricielle : $(X'X) = (XX') = nI_d$, en notant I_d la matrice identité de bonne dimension.



Sommaire

- 1 Plans factoriels complets
 - Données, notations
 - **Estimation**
 - Analyse des résultats
- 2 Principe des plans fractionnaires
- 3 Étude des confusions des plans fractionnaires
- 4 Choix et dépouillement d'un plan fractionnaire



Matrice des effets

Une matrice qui vérifie cette propriété est dite matrice de Hadamard.

Il n'existe des matrices de Hadamard que pour n multiple de 4.

Pour $n = 2^p$, la matrice du modèle saturé du plan 2^p est une matrice de Hadamard. Pour les autres valeurs de n les matrices de Hadamard se trouvent dans des tables.



Estimation

Lorsque X est une matrice de Hadamard, l'estimation des paramètres est particulièrement simple. En effet :

$$(X'X)^{-1} = (1/n)I_d.$$

L'estimation des coefficients β devient alors : $\hat{\beta} = (1/n)X'Y$.
Au coefficient $1/n$ près, un coefficient est estimé en multipliant la colonne de X correspondante par le vecteur Y .



Exemple

$$Y = \begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \end{pmatrix} = \begin{pmatrix} 26 \\ 18 \\ 8 \\ 8 \\ 6 \\ 6 \\ 4 \\ 4 \end{pmatrix} \quad \hat{\beta} = \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\beta}_1 \\ \hat{\gamma}_1 \\ (\alpha\beta)_{11} \\ (\alpha\gamma)_{11} \\ (\beta\gamma)_{11} \\ (\alpha\beta\gamma)_{111} \\ \hat{\mu} \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \\ 1 \\ 3 \\ 1 \\ 1 \\ 1 \\ 10 \end{pmatrix}$$

TABLE – 4. Exemple numérique d'estimation des paramètres dans le cas d'un plan 2³



Exemple

Chaque terme du vecteur $\hat{\beta}$ est obtenu en effectuant le produit scalaire entre Y et la colonne correspondante de X puis en divisant par $n = 8$. Ce produit scalaire s'interprète simplement. Ainsi l'estimation de α_1 , obtenue par le produit scalaire entre Y et la première colonne du tableau 4, s'écrit :

$$\begin{aligned} \hat{\alpha}_1 &= \frac{1}{8}[Y_{111} + Y_{112} + Y_{121} + Y_{122} - Y_{211} - Y_{212} - Y_{221} - Y_{222}] \\ &= \frac{1}{2}[Y_{1..} - Y_{2..}] \\ &= \frac{Y_{1..} - Y_{2..}}{2} \end{aligned} \quad (1)$$



Exemple

On reconnaît, au coefficient 1/2 près, la différence de moyenne de Y entre les modalités 1 et 2 de F_1 . L'interprétation du coefficient d'un effet principal est donc très simple.

Exemple

Pour un coefficient d'interaction, par exemple $\alpha\beta_{11}$, on obtient, par le produit scalaire entre Y et la 4^e colonne du tableau 4 :

$$\begin{aligned} (\alpha\beta)_{11} &= \frac{1}{8}[Y_{111} + Y_{112} - Y_{121} - Y_{122} - Y_{211} - Y_{212} + Y_{221} + Y_{222}] \\ &= \frac{1}{4}[Y_{1.} - Y_{2.} - Y_{21.} + Y_{22.}] \\ &= \frac{1}{2}[(Y_{1.} - Y_{1..}) - (Y_{.1} - Y_{..})] \end{aligned}$$

dans laquelle on reconnaît la notion d'interaction usuelle.



Exemple

Comme les estimations précédentes, celle du coefficient de l'interaction d'ordre 3 noté $(\alpha\beta\gamma)_{111}$, s'obtient par le produit scalaire entre Y et la colonne du tableau 4 correspondant à 123. Soit :

$$\begin{aligned} (\alpha\beta\gamma)_{111} &= 1/8[Y_{111} - Y_{112} - Y_{121} + Y_{122} - Y_{211} + Y_{212} + Y_{221} - Y_{222}] \\ &= 1/8[(Y_{111} - Y_{112} - Y_{121} + Y_{122}) - (Y_{211} - Y_{212} - Y_{221} + Y_{222})] \end{aligned}$$

La seconde écriture aide à comprendre ce que représente une interaction d'ordre 3. On reconnaît la différence entre l'interaction entre $F2$ et $F3$ lorsque $F1 = 1$ et cette même interaction lorsque $F1 = 2$.



Variance des estimateurs

Elles se trouvent sur la diagonale de la matrice $\sigma^2(X'X)^{-1}$ qui, pour une matrice de Hadamard, vaut $(\sigma^2/n)Id$. Toutes ces variances sont donc égales à σ^2/n .

Théorème

Cette variance est la plus petite possible : aucun autre plan en n essais ne fournit de meilleurs estimateurs.



Interprétation

Il n'en reste pas moins que, en pratique, lorsque l'on met en œuvre des plans qui permettent de tester la signification des interactions d'ordre 3 celles-ci sont rarement significatives. Pour toutes ces raisons, la plupart des utilisateurs négligent les interactions d'ordre 3. A fortiori, il en est bien sûr de même pour les interactions d'ordre supérieur à 3.



Variance des estimateurs

On peut avoir une intuition de ce résultat en remarquant que cette variance est celle de l'estimateur d'une moyenne à partir d'un échantillon de taille n , situation dans laquelle toutes les données ne servent à estimer qu'une seule moyenne.

Ce qui est remarquable ici est que, à partir du même ensemble de données, on estime tous les effets avec cette même précision.

Cette idée sera étendue avec les plans fractionnaires.



Propriété remarquable

La relation $\hat{\beta} = (1/n)X'Y$ montre clairement que l'estimation d'un coefficient particulier ne dépend pas de la présence ou non des autres effets dans le modèle.

Ainsi, les **estimations** $\{5, 4, 1\}$ des effets principaux sont **les mêmes**, que l'on utilise le **modèle saturé** ou le **modèle réduit** aux seuls effets principaux.

Propriété remarquable

Cette propriété est attendue par l'utilisateur. Les plans qui la possède sont appelés plans orthogonaux et elle est toujours vraie dans le cas d'un plan complet (quels que soient les nombres de modalités des facteurs), ce qui est l'un des attraits des plans complets.

Cette propriété repose sur le résultat suivant : l'estimation des coefficients de F_1 ne dépend pas de la présence ou non de F_2 dans le modèle lorsque F_1 et F_2 ne sont pas liés.

Propriété remarquable

En pratique, cette liaison peut s'apprécier en calculant le tableau de contingence croisant les modalités de F_1 (en lignes) avec celles de F_2 (en colonnes) dans lequel, à l'intersection de la ligne i et de la colonne j , on trouve le nombre d'essais réalisés selon ces modalités (cf. Tableau 5 à gauche).

Si le tableau des effectifs est constant, ce qui est le cas dans notre exemple, mais aussi plus généralement pour tout plan complet et ce quel que soit le couple de facteurs, les deux facteurs ne sont pas liés : on dit qu'ils sont orthogonaux.

Propriété remarquable

	Plan 2^3		Plan [P]					
	F_2		F_2					
	1	2			1	2		
F_1	1	2	2	2	F_1	1	3	1
	2	2	2	2	2	2	2	2

TABLE – 5. Tableau d'effectifs croisant les facteurs F_1 et F_2 . À gauche pour le plan 2^3 , chaque combinaison de modalités est représentée par le même nombre d'essais. Ce n'est pas le cas, à droite, pour le plan [P].

Propriété remarquable

A contrario, il y aurait liaison si, par exemple, la modalité 1 de F_1 était plus souvent présente en conjonction avec la modalité 1 de F_2 qu'avec la modalité 2 de F_2 ; c'est le cas avec le plan [P] du tableau 5 dans lequel, par rapport au plan 2^3 du tableau 1, au lieu de réaliser l'essai 3 ($F_1 = 1, F_2 = 2, F_3 = 1$), on avait réalisé un 3^e essai semblable aux deux premiers ($F_1 = 1, F_2 = 1, F_3 = 1$).

Intuitivement, on « sent » bien que, dans ce cas, une valeur élevée de Y pour la modalité $F_1 = 1$ (par rapport à la valeur de Y pour $F_1 = 2$) est peut-être imputable à de plus fortes valeurs de Y pour la modalité $F_2 = 1$.

Propriété remarquable

En comparant les estimations fournies par un programme exécuté pour chacun des deux modèles, on peut constater que les estimations de α_1 ne sont pas les mêmes selon que l'on introduit ou non F_2 dans le modèle (plus profondément, on peut comparer les estimateurs $\hat{\beta} = (X'X)^{-1}X'Y$ dans les deux cas).

Propriété remarquable

L'estimateur $\hat{\beta}$ gère bien ce problème si F_2 est dans le modèle (il est sans biais) et, bien sûr, ne le gère pas si F_2 n'est pas dans le modèle. Ce problème est appelé « confusion » entre facteurs expérimentaux.

La **confusion** est ici **partielle** ; en pratique, de telles confusions sont gênantes et on essaie d'éviter de se mettre dans ces situations par une planification rigoureuse.

Propriété remarquable

Un autre type de confusion, dite totale, est utilisé dans des stratégies expérimentales : c'est l'objet du paragraphe suivant. Remarquons enfin que l'intérêt de disposer de 2 facteurs non confondus entre eux est annulé si un troisième facteur est confondu partiellement avec chacun d'eux.

Sommaire

- 1 Plans factoriels complets
 - Données, notations
 - Estimation
 - Analyse des résultats
- 2 Principe des plans fractionnaires
- 3 Étude des confusions des plans fractionnaires
- 4 Choix et dépouillement d'un plan fractionnaire

Analyse des résultats

Remarquons tout d'abord que, dans le cas d'un facteur à deux modalités, il revient au même de tester le facteur dans son ensemble par un test F de Fisher ou l'unique coefficient de ce facteur par un test t de Student.

Il est facile de constater que, dans ce cas, $F_{obs} = (t_{obs})^2$. En outre, la loi $F_{1; \nu}$ (ayant un seul degré de liberté au numérateur) et la loi t_{ν} sont liées par la relation suivante : si $\mathcal{L}(X) = t_{\nu}$, alors $\mathcal{L}(X^2) = F_{1; \nu}$.

Analyse des résultats

Dès lors que les facteurs sont un tant soit peu nombreux, on peut hésiter quant à la stratégie de dépouillement des résultats. Ce problème a donné lieu à de nombreux travaux, conduisant à des propositions de stratégie plus ou moins sophistiquées. Nous donnons ci-après quelques points de repères simples.

Démarche inférentielle classique

C'est celle que nous avons adoptée jusqu'ici. On postule un modèle a priori (c'est-à-dire avant de réaliser l'expérience ou, tout au moins, indépendamment des résultats Y), modèle comportant un terme d'erreur.

On réalise les tests et on prend en compte, dans la décision finale, uniquement des effets significatifs. Cette démarche est la plus pure sur le plan statistique.

Dans l'exemple, elle consiste par exemple à postuler le modèle comportant les effets principaux et les interactions d'ordre 2 et conduit à la conclusion qu'aucun effet n'est significatif (cf. Tableau 6, ligne 3 - modèle 1 : toutes les probabilités critiques sont supérieures à 5%).

Tableau de résultats de l'exemple

Effet	α_1	β_1	γ_1	$\alpha\beta_{11}$	$\alpha\gamma_{11}$
Coeffs \forall le modèle	5	4	1	3	1
probabilité critique	Mod. 1 0,126	0,156	0,500	0,205	0,500
	Mod. 2 0,007	0,016		0,040	
Effet	$\beta\gamma_{11}$	$\alpha\beta\gamma_{111}$	$\widehat{\sigma^2}$	<i>ddl</i>	$t_{v,0,975}$
Coeffs \forall le modèle	1	1			
probabilité critique	Mod. 1 0,500	8	8	1	12,71
	Mod. 2	8	4	4	2,776

TABLE – 6. Analyse des données du tableau 4 par deux modèles. Estimation des coefficients (ligne 1) et probabilités critiques (lignes 2 et 3) ; les cases vides sont « sans objet »

Démarche non inférentielle

On utilise le modèle saturé, ce qui permet de disposer de l'estimation du coefficient de chaque effet.

On sélectionne les effets les plus importants uniquement d'après la valeur de leur coefficient ; éventuellement on peut disposer d'un seuil a priori exprimé en valeur de Y : par exemple, Y étant un rendement exprimé en %, on peut considérer qu'un effet qui se traduit par un coefficient inférieur à 2% ne nous intéresse pas (même s'il est significatif) car il est négligeable d'un point de vue économique.

Démarche non inférentielle

Dans l'exemple, cela revient à considérer uniquement les effets principaux des facteurs 1 et 2 ainsi que leur interaction (cf. Tableau 6, ligne 2 : coefficients). Ce cas est bien sûr facile (il a été construit pour !) puisqu'il montre clairement deux groupes de coefficients.

Démarche inférentielle empirique

On considère d'abord un modèle comportant beaucoup d'effets, le cas extrême étant celui du modèle saturé.

On sélectionne les effets les plus importants comme précédemment et on construit un nouveau modèle avec ces seuls effets (on peut aussi retirer un à un les effets les plus faibles). On réalise alors les tests avec la nouvelle variance résiduelle.

Démarche inférentielle empirique

Dans l'exemple, après avoir analysé le modèle 1 ou le modèle saturé, on est conduit, selon cette démarche, à tester le modèle 2 (restreint aux effets dont les coefficients sont supérieurs à 1 dans le modèle 1) :

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}.$$

Les 3 effets sont maintenant significatifs (cf. Tableau 6). Que s'est-il passé ?

Démarche inférentielle empirique

Par rapport au modèle 1, on incorpore dans la variance résiduelle du modèle 2 trois effets supplémentaires : la SC_R s'accroît donc mécaniquement. Mais le nombre de degrés de liberté augmente aussi.

Dans ce cas très particulier où les effets que l'on ajoute sont chacun équivalents à celui qui s'y trouvait déjà (cf. les coefficients indiqués dans la première ligne), l'augmentation du nombre de degrés de liberté compense exactement l'augmentation de la SC_R .

Démarche inférentielle empirique

Finalement, dans ce cas très particulier, l'estimation de σ^2 est constante et, par voie de conséquence, la variance des estimateurs aussi (elle vaut, pour n'importe quel coefficient de chacun des deux modèles : $\widehat{\sigma^2}/n = 1$).

Toujours dans ce cas particulier, la statistique de test d'un coefficient est égale au coefficient lui-même puisque les variances des estimateurs sont égales à 1.

Dans ces données, un seul élément change donc d'un modèle à l'autre dans les tests des coefficients : le nombre de degrés de liberté qui passe de 1 à 4, ce qui a une influence énorme sur le seuil de signification $t_{\nu,0,975}$ qui passe de 12,71 à 2,776.

Démarche inférentielle empirique

Cet exemple illustre très bien la faible puissance des tests avec un seul degré de liberté et le gain appréciable apporté par 3 degrés de liberté supplémentaires.

En pratique, face à ces données, on aurait conclu à l'influence des trois effets F_1 , F_2 et F_1F_2 .

En quoi cette pratique est-elle empirique et critiquable du point de vue de la théorie des tests ?

Sommaire

Démarche inférentielle empirique

Le principe même de changer de modèle jusqu'à observer des effets significatifs engendre un risque 0 : réel sûrement supérieur à celui annoncé même si on ne peut le calculer.

Plus précisément, en choisissant de ne faire entrer dans la résiduelle que des effets petits, on a tendance à sous-estimer 0-2 ; mais, d'un autre côté , on améliore sensiblement la puissance des tests du fait de l'augmentation des degrés de liberté : la stratégie est donc raisonnable. Elle reste néanmoins empirique car on ne peut afficher avec certitude les risques associés aux tests.

- 1 Plans factoriels complets
- 2 Principe des plans fractionnaires
- 3 Étude des confusions des plans fractionnaires
- 4 Choix et dépouillement d'un plan fractionnaire

Sommaire

Démarche inférentielle empirique

Le principe même de changer de modèle jusqu'à observer des effets significatifs engendre un risque 0 : réel sûrement supérieur à celui annoncé même si on ne peut le calculer.

Plus précisément, en choisissant de ne faire entrer dans la résiduelle que des effets petits, on a tendance à sous-estimer 0-2 ; mais, d'un autre côté , on améliore sensiblement la puissance des tests du fait de l'augmentation des degrés de liberté : la stratégie est donc raisonnable. Elle reste néanmoins empirique car on ne peut afficher avec certitude les risques associés aux tests.

- 1 Plans factoriels complets
- 2 Principe des plans fractionnaires
- 3 Étude des confusions des plans fractionnaires
- 4 Choix et dépouillement d'un plan fractionnaire

Sommaire

Principe des plans fractionnaires

Introduction

- 1 Plans factoriels complets
- 2 Principe des plans fractionnaires
 - Introduction
 - Construction en trois étapes
 - Propriétés du plan proposé
- 3 Étude des confusions des plans fractionnaires
- 4 Choix et dépouillement d'un plan fractionnaire

Introduction

Soit le problème suivant : proposer un plan d'expérience permettant d'étudier 3 facteurs à 2 modalités à l'aide de 4 essais.

Il est clair que si l'on n'était pas limité par le nombre d'essais, le plan complet 2^3 s'imposerait.

Ce cas très simple illustre la difficulté la plus fréquemment rencontrée en expérimentation : le plan complet est intéressant de par ses propriétés statistiques mais conduit à un nombre d'essais prohibitif dès lors que le nombre de facteurs est un tant soit peu grand (2^p dans le cas de p facteurs à 2 modalités).

Sommaire

- 1 Plans factoriels complets
- 2 **Principe des plans fractionnaires**
 - Introduction
 - **Construction en trois étapes**
 - Propriétés du plan proposé
- 3 Étude des confusions des plans fractionnaires
- 4 Choix et dépouillement d'un plan fractionnaire

Introduction

Indiquons d'emblée que l'on ne peut pas espérer faire en 4 essais aussi bien qu'avec 8. Mais le problème ne se pose pas en ces termes puisque nous ne pouvons pas, pour des raisons économiques, faire plus de 4 essais.

Le problème est donc : peut-on en 4 essais, mettre en évidence des informations utiles voire suffisantes ?

Construction en trois étapes

Étape 1

- 1 On considère le nombre de facteurs maximum tel que le plan complet limité à ces facteurs est compatible avec le nombre d'essais. Ces facteurs sont dits « facteurs de base » et désignés par F_1 et F_2 .

Exemple

Dans l'exemple considéré précédemment, nous avons $n = 4$, on peut donc construire un plan complet avec 2 facteurs.

Exemple

n° essai	Matrice des essais		Matrice des effets			
	F1	F2	F1	F2	F1F2	I
1	+1	+1	+1	+1	+1	+1
2	+1	-1	+1	-1	-1	+1
3	-1	+1	-1	+1	-1	+1
4	-1	-1	-1	-1	+1	+1

TABLE – 7. Plan 2^2 : matrice des essais et matrice des effets pour le modèle saturé. Les modalités des facteurs ne sont plus notées 1 et 2 mais +1 et -1.

Étape 2

- On décide de faire varier le facteur F3 comme l'interaction F1F2.

Le facteur F3 est dit confondu (totalement) avec l'interaction F1F2.

On obtient ainsi, la matrice d'un plan en 4 essais à 3 facteurs, plan que l'on note 2^{3-1} exprimant qinsi qu'il met en jeu 3 facteurs à deux modalités, l'un des facteurs étant confondu avec un effet issu des deux autres (cf. Tableau 8).

Sommaire

- Plans factoriels complets
- Principe des plans fractionnaires**
 - Introduction
 - Construction en trois étapes
 - Propriétés du plan proposé**
- Étude des confusions des plans fractionnaires
- Choix et dépouillement d'un plan fractionnaire

Exemple

n° essai	F1	F2	F3
1	+1	+1	+1
2	+1	-1	-1
3	-1	+1	-1
4	-1	-1	+1

TABLE – 8. Matrice des essais du plan 2^{3-1}

Propriétés du plan proposé

Orthogonalité

La qualité majeure de ce plan est que tous ses facteurs sont orthogonaux deux à deux.

Cette orthogonalité peut être vérifiée en calculant les tableaux croisant les modalités de deux facteurs (comme cela a été fait dans le cours sur les plans factoriels complets).

L'orthogonalité entre $F1$ et $F2$ est acquise par construction puisque le point de départ est le plan complet avec $F1$ et $F2$; pour les autres, on vérifie que les tableaux croisés contiennent un essai par case.

Modèle et précision de l'estimation

Pour déduire ce plan, le seul modèle possible est :

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k.$$

Ce modèle est saturé : 4 paramètres sont estimés à partir de 4 essais.

On ne peut pas faire de test de signification des effets. La variance de l'estimation de chaque coefficient est optimale (cf. cours sur les plans factoriels complets) et vaut : $1/4 \times \sigma^2$.

Confusions

Il est facile de voir que la matrice X des effets de ce modèle est exactement celle du modèle à 2 facteurs avec interaction (cf. Tableau 9 à droite) : c'est la même colonne qui représente dans un cas $F3$ et dans l'autre $F1F2$.

Il résulte donc de la confusion entre $F3$ et $F1F2$ que les deux paramètres ne sont pas estimables séparément.

Autrement dit, lorsque l'on estime γ_k à partir du modèle ci-dessus, on estime en fait la somme : $\gamma_k + (\alpha\beta)_{ij}$.

Confusions

Il en résulte que le plan 2^{3-1} est intéressant en pratique uniquement si l'on a de bonnes raisons **a priori** de négliger l'interaction $F1F2$.

Pourquoi la dénomination de fractionnaire ? Parce que ce plan est une fraction du plan complet 2^3 (cf. Tableau 9).

Sommaire

- 1 Plans factoriels complets
- 2 Principe des plans fractionnaires
- 3 Étude des confusions des plans fractionnaires
- 4 Choix et dépouillement d'un plan fractionnaire

Sommaire

- 1 Plans factoriels complets
- 2 Principe des plans fractionnaires
- 3 Étude des confusions des plans fractionnaires
 - Une confusion en engendre d'autres
 - Lister les confusions induites par une confusion donnée
 - Choix des confusions ; cas de plusieurs confusions
- 4 Choix et dépouillement d'un plan fractionnaire

Exemple

n°	1	2	3	n°	1	2	3
1	+1	+1	+1	1	+1	+1	+1
2	+1	+1	-1	4	+1	-1	-1
3	+1	-1	+1	6	-1	+1	-1
4	+1	-1	-1	7	-1	-1	+1
5	-1	+1	+1	8	-1	-1	-1
6	-1	+1	-1	5	-1	+1	+1
7	-1	-1	+1	3	+1	-1	+1
8	-1	-1	-1	2	+1	+1	-1

TABLE – 9. Le plan 2^3 et ses deux fractions 2^{3-1} formées des lignes (1,4,6,7) ou des lignes (8,5,3,2). Les deux fractions sont dites opposées car se déduisent l'une de l'autre par multiplication par -1 .

Modèle complet

Ecrivons le modèle complet (i.e. incluant tous les effets possibles) d'un plan à 3 facteurs.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}$$

Ce modèle n'est évidemment pas estimable à partir du plan 2^{3-1} (8 paramètres pour 4 données).

Confusions

On peut tout de même calculer la matrice X associée à ce modèle et au plan 2^{3-1} (cf. Tableau 10).

Elle permet de constater que

- $F3$ est confondu avec $F1F2$ (les colonnes correspondantes de X sont les mêmes) ;
- $F2$ est confondu avec $F1F3$;
- $F1$ est confondu avec $F2F3$;
- l'interaction d'ordre 3 $F1F2F3$ est confondue avec la constante.

Exemple

n°	1	2	3	12	13	23	123	I
1	1	1	1	1	1	1	1	1
2	1	-1	-1	-1	-1	1	1	1
3	-1	1	-1	-1	1	-1	1	1
4	-1	-1	1	1	-1	-1	1	1

TABLE – 10. Matrice X des effets du modèle complet pour le plan 2^{3-1} .

Sommaire

- 1 Plans factoriels complets
- 2 Principe des plans fractionnaires
- 3 Étude des confusions des plans fractionnaires
 - Une confusion en engendre d'autres
 - Lister les confusions induites par une confusion donnée
 - Choix des confusions ; cas de plusieurs confusions
- 4 Choix et dépouillement d'un plan fractionnaire

Confusions

Ainsi, lorsque l'on décide de confondre $F3$ avec $F1F2$, cela engendre implicitement d'autres confusions.

Avec ce plan de petite taille, il est facile d'observer ces nouvelles confusions sur la matrice X .

Mais, dès que le nombre d'essais est élevé (16 et au-delà), l'examen visuel de X devient très difficile et il est nécessaire de disposer d'une méthode générale.

Lister les confusions induites par une confusion donnée

Définition

On définit la multiplication entre deux colonnes comme leur multiplication terme à terme.

Loi de composition

Ainsi, dans le tableau 10, la multiplication entre la colonne 1 et la colonne 2, multiplication notée 12 (lire « un deux » et non pas « douze » !), conduit à la colonne d'interaction entre 1 et 2, notée elle aussi 12 (ce qui tombe bien !).

Ainsi, pour écrire que l'on confond F3 avec F1F2 on écrit : $3 = 12$.

Loi de composition

Cette opération de multiplication est

- commutative ($12 = 21$) ;
- la colonne I joue le rôle d'élément neutre ($I2 = 2$) ;
- comme les colonnes ne comportent que les valeurs $+1$ et -1 le produit d'une colonne par elle-même donne I (par exemple $22 = 2^2 = I$).

Énumération des confusions

Munie de cette opération, la méthode pour énumérer les confusions engendrées par une confusion donnée est la suivante :

- Expliciter l'effet confondu avec la constante : dans l'exemple, de $3 = 12$ on déduit $33 = 123$ soit $I = 123$. Cette confusion est appelée générateur d'alias, le terme « alias » étant synonyme de « confusion ».
- On multiplie les deux termes du générateur d'alias par chacun des effets du modèle saturé dérivant des seuls facteurs de base.

Exemple

Dans le cas 2^{3-1} , ces effets sont 1, 2 et 12. Soit :

$$1/ = 1123 = 23 \Rightarrow 1 = 23$$

$$2/ = 2123 = 13 \Rightarrow 2 = 13$$

$$12/ = 12123 = 3 \Rightarrow 3 = 12.$$

On résume ces confusions par le tableau 11.

Exemple

effets du plan 2^2	1	2	12	/
effets impliquant F3	23	13	3	123

TABLE – 11. Confusions dans le plan 2^{3-1} . Les effets confondus entre eux sont dans une même colonne.



Exemple

Avec **R**, on calcule aisément $X'X$ avec **X** correspondant au modèle complet (cf. Tableau 12).

À l'intersection de la ligne i et de la colonne k on trouve n si les effets i et k sont confondus et 0 sinon.

Dans cette matrice, les termes de la ligne et de la colonne correspondant à la constante sont les sommes des colonnes de **X** ; les autres termes sont, au coefficient n près, les coefficients de corrélation entre les colonnes de **X**.

Exemple

	1	2	3	12	13	23	123	/
1	4	0	0	0	0	4	0	0
2	0	4	0	0	4	0	0	0
3	0	0	4	4	0	0	0	0
12	0	0	4	4	0	0	0	0
13	0	4	0	0	4	0	0	0
23	4	0	0	0	0	4	0	0
123	0	0	0	0	0	0	4	4
/	0	0	0	0	0	0	4	4

TABLE – 12. Matrice $X'X$ correspondant à la matrice **X** du tableau 10.



Sommaire

- 1 Plans factoriels complets
- 2 Principe des plans fractionnaires
- 3 Étude des confusions des plans fractionnaires
 - Une confusion en engendre d'autres
 - Lister les confusions induites par une confusion donnée
 - Choix des confusions ; cas de plusieurs confusions
- 4 Choix et dépouillement d'un plan fractionnaire

Énumération des confusions

Cette propriété est très utile pour découvrir les propriétés d'un plan que l'on a pas soi-même construit et dont on ne connaît pas les confusions de base (ou les générateurs d'alias).

Choix des confusions ; cas de plusieurs confusions

Principe

On étudie p facteurs (à 2 modalités) ; on ne peut réaliser que n essais ; $n = 2^q$.

Procédure

On choisit q facteurs de base.

Il reste donc à confondre les $k = p - q$ autres facteurs avec des interactions entre facteurs de base. Comment choisir ces confusions ?

Le principe est de confondre les nouveaux facteurs avec des interactions a priori négligeables, c'est-à-dire d'ordre élevé.

Exemple 1 : Plan 2^{4-1}

On étudie $p = 4$ facteurs (F_1, F_2, F_3, F_4) avec $n = 8$ essais.

Les trois premiers sont les facteurs de base. Les effets du modèle saturé impliquant ces facteurs de base sont sur la première ligne du tableau 13.

On choisit de confondre F_4 avec l'interaction d'ordre 3. Il en résulte le générateur d'alias $I = 1234$ dont on déduit les autres confusions.

Exemple 1 : Plan 2^{4-1}

effets plan 2^3	1	2	3	12	13	23	123	I
effets avec F4	234	134	124	34	24	14	4	1234

TABLE – 13. Confusions dans le plan 2^{4-1} . Les effets confondus entre eux sont dans une même colonne.

Exemple 1 : Plan 2^{4-1}

En revanche, les interactions d'ordre 2 sont confondues entre elles deux à deux ; des ambiguïtés pourront survenir lors du dépouillement.

Avant d'appliquer ce plan, il reste à affecter les facteurs réels aux numéros (par exemple F1 est le facteur matière première, F2 est le facteur process, etc.).

Si l'on a des idées a priori sur les interactions les plus importantes, on réalisera cette affectation de façon à ce que ces interactions ne soient pas confondues entre elles (i.e. ne figurent pas dans la même colonne).

Exemple 1 : Plan 2^{4-1}

Dans ce plan, les effets principaux sont confondus avec des interactions d'ordre 3 ; celles-ci étant généralement négligeables, les effets principaux seront donc bien estimés.

« bien » est ici à comprendre au sens de « non confondu avec un autre effet » c'est à dire sans biais. Un plan dont les effets principaux sont confondus avec des interactions d'ordre supérieur ou égal 3 est dit plan de résolution IV. Ce chiffre est celui de l'ordre le moins élevé des effets confondus avec la constante.

Exemple 2 : Plan 2^{5-2}

On dispose toujours de $n = 8$ essais mais cette fois pour étudier 5 facteurs.

Nous repartons du plan précédent 2^{4-1} ; il faut choisir à quelle colonne affecter F5.

Cette colonne ne doit pas, bien sûr, contenir d'effet principal ce qui exclut les colonnes 1, 2, 3 et 123. A priori, celles qui restent sont équivalentes ; si l'on choisit la colonne incluant 12 et 34, ceci induit deux générateurs d'alias : $I = 125$ et $I = 345$.

Exemple 2 : Plan 2^{5-2}

Le tableau 14 explicite l'ensemble des confusions de ce plan.

On obtient :

- la 3^e ligne en multipliant le générateur d'alias $I = 125$ par les effets du plan 2^3 ;
- la 4^e ligne en multipliant $I = 345$ par les effets du plan 2^3 .

Exemple 2 : Plan 2^{5-2}

effets du plan 2^3	1	2	3	12
conf gén par $I = 1234$	234	134	124	34
conf gén par $I = 125$	25	15	1235	5
conf gén par $I = 345$	1345	2345	45	12345

effets du plan 2^3	13	23	123	/
conf gén par $I = 1234$	24	14	4	1234
conf gén par $I = 125$	235	135	35	125
conf gén par $I = 345$	145	245	1245	345

TABLE – 14. Confusions dans le plan 2^{5-2} . Les effets confondus entre eux sont dans une même colonne.

Exemple 2 : Plan 2^{5-2}

Le plan obtenu comporte bien sûr plus de confusions que le précédent.

Mais surtout plusieurs de ces confusions concernent un effet principal et une interaction d'ordre 2.

Un tel plan, dont seuls les effets principaux ne sont pas confondus entre eux, est dit de résolution III.

Exemple 2 : Plan 2^{5-2}

L'affectation de facteurs réels aux numéros de 1 à 5 reste à faire. Si une interaction particulière est attendue, on s'arrangera pour qu'elle ne soit pas confondue avec un effet principal.

Remarque

Lorsque l'on doit ajouter plusieurs facteurs, il est généralement plus avantageux de les confondre avec les interactions d'ordre le plus élevé moins 1.

Sommaire

- 1 Plans factoriels complets
- 2 Principe des plans fractionnaires
- 3 Étude des confusions des plans fractionnaires
- 4 **Choix et dépouillement d'un plan fractionnaire**

Sommaire

- 1 Plans factoriels complets
- 2 Principe des plans fractionnaires
- 3 Étude des confusions des plans fractionnaires
- 4 **Choix et dépouillement d'un plan fractionnaire**
 - Nombre de facteurs étudiés et nombre d'essais
 - Dépouillement des résultats
 - Essais complémentaires

Choix et dépouillement d'un plan

Nombre de facteurs étudiés et nombre d'essais

Introduction

L'exemple précédent montre que l'on peut ajouter un nouveau facteur tant qu'il reste une colonne disponible c'est-à-dire non occupée par un effet principal.

Ainsi, avec $n = 8$ essais, on peut étudier simultanément 7 facteurs. C'est bien le maximum possible puisque dans ce cas, le modèle ne comportant que les effets principaux est saturé (8 paramètres, incluant la constante, estimés avec 8 données).

Résolution et effectif

Si l'on augmente le nombre d'essais, le nombre de facteurs que l'on peut étudier simultanément augmente d'autant. Mais dans ce cas, on néglige a priori toutes les interactions.

On peut souhaiter se limiter à des plans de résolution IV. Dans ce cas, le nombre de facteurs que l'on peut étudier simultanément est $n/2$.

Résolution et effectif

Enfin, on peut être encore plus exigeant quant aux confusions et souhaiter que ni les effets principaux ni les interactions d'ordre 2 ne soient confondus entre eux. Un tel plan est dit de résolution V.

En effet, on constate que le générateur $I = ABCDE$ (qui met en jeu une interaction d'ordre 5) ne peut conduire qu'à des confusions de type $A = BCDE$ ou $AB = CDE$. Ce type de plan est évidemment très confortable dans l'interprétation de ses résultats à tel point que certains considèrent que, dans beaucoup d'applications, réaliser plus d'essais que le nombre exigé par un plan de résolution V n'est pas nécessaire.

Résolution et effectif

Le nombre maximum de facteurs à deux modalités que l'on peut étudier avec un plan fractionnaire est donné tableau 15.

Exemple

avec 16 essais, on peut étudier 5 facteurs en résolution 5, 8 facteurs en résolution 4 et 15 facteurs en résolution 3.

Frédéric Bertrand

Résolution et effectif

r	nombre d'essais $n = 2^r$	Nombre maximum de facteurs		
		Résol. V	Résol. IV	Résol. III
3	8	3	4	7
4	16	5	8	15
5	32	6	16	31
6	64	8	32	
7	128	11		
8	256	17		
9	512	23		

TABLE – 15. Nombre maximum de facteurs selon le nombre d'essais et la résolution.

Frédéric Bertrand

Remarque

Le choix de la résolution d'un plan dépend généralement de l'état de connaissance du problème que l'on examine.

Dans une étude préliminaire, on souhaite généralement tester beaucoup de facteurs ce que permet un plan de résolution III.

Dans une étude finale, les facteurs influents ont été sélectionnés et l'on s'intéresse à leurs interactions : un plan de résolution III ne peut convenir.

Frédéric Bertrand

Sommaire

Dépouillement des résultats

- 1 Plans factoriels complets
- 2 Principe des plans fractionnaires
- 3 Étude des confusions des plans fractionnaires
- 4 Choix et dépouillement d'un plan fractionnaire
 - Nombre de facteurs étudiés et nombre d'essais
 - Dépouillement des résultats
 - Essais complémentaires

Estimation et confusions

Compte tenu des confusions entre effets, le dépouillement par analyse de variance d'un plan fractionnaire pose des problèmes spécifiques.

La difficulté fondamentale est que l'on ne peut estimer les effets isolément : chaque estimation de coefficient correspond en fait à une somme d'effets.

Ainsi, dans l'analyse de la variance du plan 2^{3-1} l'estimation qu'un programme indiquerait comme étant celle de γ_k est en fait celle de la somme $\gamma_k + \alpha\beta_{ij}$.

Sélection des facteurs influents

Nous ne reprenons pas ici la discussion sur la sélection des effets, à partir des seules estimations ou de tests de signification.

Nous nous plaçons dans la situation où, à l'issue du traitement statistique, certains effets sont considérés négligeables (parce qu'ils sont petits et/ou non significatifs) et d'autre non. Le problème qui demeure est alors dû au fait que ces effets représentent en fait une somme d'effets.

Sélection des facteurs influents

Nous indiquons cinq règles de dépouillement qui relèvent à la fois de la nécessité (sans elles on ne peut rien dire) et de l'importance des effets observés en pratique. Nous illustrons ces règles à partir d'une succession de cas types résumés dans le tableau 16.

Un cas type est réduit à l'essentiel : il indique les effets (ou, plus rigoureusement les sommes d'effets que l'on doit prendre en compte car ils sont non négligeables (grands et/ou significatifs ; on dit aussi influents) ce qui est noté NN.

Ainsi la 3^e ligne, cas n°1, correspond à une situation dans laquelle l'analyse statistique n'a conduit à aucun effet significatif.

Règle 1

On considère négligeables :

- 1 **Règle 1. Les composantes d'une somme d'effets négligeables.**

Par exemple, si l'estimation de la somme $\gamma_k + \alpha_i\beta_{ij}$ est très faible, on considère que ces deux effets sont très faibles.

Ceci n'est absolument pas une implication absolue puisque deux effets peuvent se compenser.

Adopter cette règle revient simplement à considérer qu'une telle compensation est improbable.

Exemple

Dans le cas 1, cette règle suffit pour interpréter. Toutes les sommes d'effets sont négligeables : on retiendra qu'aucun des facteurs étudiés n'agit. Cette situation est un cas d'utilisation efficace d'un plan fractionnaire : avec peu d'essais, on se rend compte que les facteurs ont sans doute été mal choisis, ou qu'une variation résiduelle importante masque tout ; il eût été dommage de faire un plan complet pour en arriver là.

Règle 2

On considère négligeables :

- 1 **Les interactions d'ordre strictement supérieur à 2.**
Cette règle, que nous avons souvent indiquée, est la plus solide en ce sens que, en pratique, dans les plans qui permettent de les estimer, on observe rarement de telles interactions.

Exemple

Ainsi, dans le cas 2 on ne retient que les effets 1 et 2.

Règle 3

On considère négligeables :

- 1 **Les interactions entre deux effets négligeables.**
On peut admettre que si des facteurs n'agissent pas séparément, ils n'agissent pas en interaction. Ce qui dérive du fait qu'en général les effets d'interaction sont plus faibles que ceux des effets principaux.
Ceci n'est évidemment pas une vérité absolue mais découle d'observations en maints domaines.

Exemple

Le cas 3 pose le problème suivant : l'ensemble 12 + 34 est non négligeable.

Est-ce dû à ces deux interactions ? Ou seulement à l'une d'entre elles et si oui laquelle ?

Les règles précédentes conduisent à retenir les effets principaux des facteurs 1 et 2 (règle 2) et à ne pas retenir ceux des facteurs 3 et 4 (règle 1).

Ainsi, on considère que le caractère non négligeable de 12 + 34 est dû uniquement à 12.



Règle 4

On considère négligeables :

- 1 Les interactions dont l'un des deux effets est négligeable.



Exemple

Le cas 4 pose un problème analogue au cas précédent.

Les facteurs 1, 2 et 3 sont influents (règle 2) ; le facteur 4 ne l'est pas (règle 1).

L'ensemble 12 + 34 est encore non négligeable. On peut admettre que si 4 n'est pas influent, alors il n'agit pas non plus avec un autre facteur qui lui est influent.

Cette règle est bien sûr plus risquée que la précédente ; d'ailleurs, en pratique, lorsque le plan permet de les estimer, on observe de temps en temps de telles interactions significatives.



Règle 5

On considère négligeables :

- 1 Toutes les interactions.



Exemple

Nous illustrons cette règle à partir du plan 2^{5-2} construit en 10. Sélectionnons les colonnes 1, 2 et 12. On peut décider que ces 3 effets sont influents. Mais, pour la colonne 12, n'est-ce pas plutôt l'effet de 5 qui est influent ? Auquel cas on peut revenir sur la colonne 1 et se demander si ce n'est pas plutôt l'interaction 25 qui joue plutôt que l'effet de 1. La situation est inextricable sauf si l'on néglige a priori toutes les interactions.

Exemple

2^{4-1} cas	1	2	3	12	13	23	123	1234	1234	1234	effets considérés
1	—	—	—	—	—	—	—	—	—	—	aucun
2	NN	NN	—	—	—	—	—	—	—	—	1+2
3	NN	NN	—	NN	—	—	—	—	—	—	1+2+3
4	NN	NN	NN	NN	—	—	—	—	—	—	1+2+4
2^{5-2}	1	2	3	12	13	23	123	1	1234	1234	effets considérés
234	134	124	34	24	24	14	4	4	1234	1234	effets considérés
25	15	1235	5	235	135	135	35	125	1245	1245	effets considérés
cas	1345	2345	45	12345	145	245	1245	345	1245	1245	effets considérés
5	NN	NN	NN	NN	—	—	NN	—	—	—	1,2,3,4,5

TABLE – 16. Cinq situations de dépouillement. En haut, le plan 2^{4-1} de résolution IV ; en bas, un plan 2^{5-2} . La première partie de chaque tableau récapitule les confusions d'effets. La seconde (lignes numérotées de 1 à 4 puis 5) énumère des situations à l'issue de l'estimation ; NN = non négligeable ; — = négligeable.

Remarque

Il faut bien admettre que ces règles sont empiriques et ne constituent nullement des théorèmes.

La pratique montre qu'elles sont généralement bonnes, ce qui ne les empêche donc pas de pouvoir être inadéquates dans tel ou tel cas particulier.

On adoucit un peu le risque d'une mauvaise décision en tentant de trouver une interprétation aux effets conservés, et, plus difficilement, à l'absence des effets considérés comme négligeables.

Sommaire

- 1 Plans factoriels complets
- 2 Principe des plans fractionnaires
- 3 Étude des confusions des plans fractionnaires
- 4 Choix et dépouillement d'un plan fractionnaire
 - Nombre de facteurs étudiés et nombre d'essais
 - Dépouillement des résultats
 - Essais complémentaires

Essais complémentaires

Introduction

Une autre façon de sortir d'une interprétation inextricable est de réaliser des essais complémentaires.

Il existe beaucoup de stratégies pour choisir des essais complémentaires en fonction des ambiguïtés que l'on veut lever. Nous nous limitons à donner une règle à la fois simple et très générale.

Doublement par le plan opposé

Règle.

À partir d'un plan de résolution III, on obtient un plan de résolution IV en le complétant par le plan opposé (pour obtenir le plan opposé on remplace les modalités +1 par -1 et réciproquement).

Il est facile de vérifier cette propriété en considérant ce que deviennent les colonnes des effets principaux et celles des interactions d'ordre 2 quand on passe du plan $P1$ de résolution III au plan $P2$ juxtaposant ce plan et son opposé noté $-P1$.

Doublement par le plan opposé

Il suffit de remarquer que, entre $P1$ et $-P1$: les colonnes des effets d'interaction sont identiques alors que les colonnes des effets principaux sont opposées. Il en résulte que :

- deux interactions confondues dans $P1$ seront encore confondues dans $P2$;
- un effet principal et une interaction confondus dans $P1$ seront orthogonaux dans $P2$;
- deux effets orthogonaux dans $P1$ seront orthogonaux dans $P2$.

Doublement par le plan opposé

Cette règle sera efficace pour séparer les effets principaux et les interactions d'ordre 2, mais n'aidera aucunement à séparer des interactions confondues dans $P1$.

Le plan $-P1$ est aussi un plan fractionnaire ; en le juxtaposant à $P1$ on obtient un plan $P2$ fractionnaire. Cette règle ouvre la voie à une démarche séquentielle que nous illustrons par un petit exemple fictif.

Démarche séquentielle

Soit le plan 2^{6-3} et ses résultats (cf. Tableau 17).

n	$F1$	$F2$	$F3$	$F4$	$F5$	$F6$	Y
1	1	1	1	1	1	1	42
2	1	1	-1	-1	1	-1	46
3	1	-1	1	-1	-1	1	44
4	1	-1	-1	1	-1	-1	48
5	-1	1	1	-1	-1	-1	42
6	-1	1	-1	1	-1	1	38
7	-1	-1	1	1	1	-1	120
8	-1	-1	-1	-1	1	1	108

TABLE – 17. Plan 2^{6-3} construit à partir des confusions : $4 = 123$; $5 = 12$; $6 = 13$.



Démarche séquentielle

Ces données, concaténées aux précédentes, constituent un plan 2^{6-2} .

Les confusions de ce plan, limitées aux effets principaux et aux interactions d'ordre 2 (les autres sont négligées) sont dans le tableau 20 : on retrouve les effets principaux isolés et les ensembles d'interactions confondues dans le plan 2^{6-3} .



Démarche séquentielle

Les confusions dans ce plan sont rassemblées Tableau 18, avec, en dernière ligne, les estimations des coefficients du modèle saturé. Les quatre premières lignes sont identiques à celles du tableau 15 bas.

Les colonnes 1, 2 et 12 sont non négligeables.

Est-ce dû à l'ensemble des effets (1, 2 et 12) ?

Ou à l'ensemble des effets (1, 5 et 15) ?

La situation est inextricable, sauf à négliger toutes les interactions. On décide de réaliser le plan opposé. À l'issue de l'expérience, on obtient huit nouvelles valeurs de Y (cf. Tableau 19).

Démarche séquentielle

Comme on néglige les interactions d'ordre supérieur à 2, le modèle de dédoublement contient un terme d'erreur avec deux degrés de liberté.

Le carré moyen résiduel associé permet de réaliser le test de signification de chaque coefficient : ainsi, en bas du Tableau 20, outre les estimations des coefficients, figurent aussi les probabilités critiques correspondantes.



Exemple

effets de 2^3 générateur	1	2	3	12	13	23	123	1
$I=1234$	234	134	124	34	24	14	4	1234
$I=125$	25	15	1235	5	235	135	35	125
$I=345$	1345	2345	45	12345	145	245	1245	345
$I=136$	36	1236	16	236	6	126	26	136
$I=246$	1246	46	2346	146	12346	346	1346	246
$I=2356$	12356	356	256	1356	1256	56	156	2356
$I=1456$	456	12456	13456	2456	3456	123456	23456	1456
Estim. effets	-16	-19	1	18	-3	-1	-1	61

TABLE – 18. Confusion des effets dans le plan 2^{6-3} du tableau 17 et estimation des coefficients.

Exemple

n	F1	F2	F3	F4	F5	F6	Y
9	-1	-1	-1	-1	-1	-1	86
10	-1	-1	1	1	-1	1	106
11	-1	1	-1	1	1	-1	60
12	-1	1	1	-1	1	1	56
13	1	-1	-1	1	1	1	50
14	1	-1	1	-1	1	-1	54
15	1	1	-1	-1	-1	1	28
16	1	1	1	1	-1	-1	32

TABLE – 19. Plan 2^{6-3} opposé du plan du tableau 17.

Exemple

effet	1	2	3	4	5	6	12	34
coeff	-17	-17	0,003	0,184	2	7	-1	11
p. crit.	0,003	0,003	0,184	0,184	0,020	0,423	0,423	0,008
effet	13	14	15	16	25	26		
	24	23	46	45	36	35		
coeff	-2	-2	-2	-1	1	-1		
p. crit.	0,184	0,184	0,184	0,423	0,423	0,423		

TABLE – 20. Résultats de l'analyse des données du plan 2^{6-2} concaténant les plans des tableaux 17 et 18. Rappels des confusions, estimation des coefficients et probabilités critiques.

Exemple

Le tableau 20 met en évidence les effets principaux 1, 2 et 5 et la colonne 12 + 34 ; dans cette dernière, on retiendra 12 qui est associée à deux effets principaux influents.

Finalement, on considère comme influents : 1, 2, 5 et 12.

Remarque

Bien entendu, l'exemple ci-dessus a été choisi pour bien fonctionner ; à l'issue d'une telle démarche, il est possible qu'il reste encore des confusions difficiles à démêler mais elle ne peuvent concerner que des interactions.

On retiendra de cet exemple, l'idée d'une démarche séquentielle.

Si un premier plan ($P1$), comportant un nombre très restreint d'essais s'avère insuffisant, on réalise des essais supplémentaires ($-P1$) qui, analysés avec les premiers, constituent encore un plan ($P2$) optimal. Par rapport à une mise en œuvre directe du plan ($P2$) on n'a rien perdu. On gagne en se ménageant la possibilité d'arrêter l'étude avant.