JOURNAL DE LA SOCIÉTÉ FRANÇAISE DE STATISTIQUE

JEAN-MARC AZAÏS
HERVÉ MONOD

Discussion et commentaires. La planification des expériences : choix des traitements et dispositif expérimental

Journal de la société française de statistique, tome 141, n° 1-2 (2000), p. 31-33

<http://www.numdam.org/item?id=JSFS_2000__141_1-2_31_0>

DISCUSSION ET COMMENTAIRES

La planification des expériences : choix des traitements et dispositif expérimental

Jean-Marc AZAÏS [1]

Hervé MONOD [2]

Nous félicitons Pierre Dagnelie pour son article qui souligne avec clarté et avec des exemples précis l'importance toujours actuelle de plusieurs principes de base de la planification expérimentale. Nous apprécions particulièrement la Section 7, qui illustre de manière très pédagogique le problème de la précision nécessaire pour détecter un maximum. Nous concentrerons nos compléments et commentaires sur les exemples présentés dans la section 9, qui aborde les relations délicates entre randomisation, structure en blocs, et contraintes expérimentales.

Concernant la randomisation proposée en 9.1, il s'agit comme le fait remarquer l'auteur, de la randomisation d'un plan en parcelles divisées (le fameux « split-plot »). La randomisation proposée assure effectivement un estimateur sans erreur systématique, mais débouche sur un modèle à deux strates d'erreur :

- une erreur attachée aux séries de quatre essais consécutifs (les « grandes parcelles » du split-plot, **2112**,**2122**, **2121**, **2111** par exemple)
- une erreur attachée aux unités à l'intérieur des séries (les « petites parcelles »).

En conséquence, comme c'est classique dans tout plan en parcelles divisées, la randomisation impliquera une précision moindre pour tous les contrastes liés aux deux premiers facteurs, car ils ne sont estimables que dans le modèle (ou strate) inter-séries. De plus, il faut souligner que, dans l'exemple tel qu'il est présenté, on tombe dans le piège dénoncé en partie 6 : même en supposant l'absence d'interaction entre les deux premier facteurs, l'analyse de la variance du modèle inter-série sera

ddl :	3	(4 Données)
facteur 1 :	1	
facteur 2 :	1	
résiduelle :	1	

1. Laboratoire de Statistique et UMR-CNRS C5583, Université Paul Sabatier, Probabilités, 118, route de Narbonne, 38062 Toulouse Cedex; azaïs@cict.fr.

2. INRA-Unité de Biométrie, Domaine de Vilvert, 78352 Jouy-en-Josas Cedex; Herve.Monod@jouy.inra.fr

Il ne reste donc qu'un seul degré de liberté pour estimer la variabilité résiduelle dans la strate inter-séries. Comme vu en section 6, la précision sur les deux premiers facteurs en sera très dégradée.

Pour l'exemple en 9.2, les blocs sont cette fois clairement identifiés. Dans ce type de situation, et en l'absence de contraintes techniques sur les changements de modalités du quatrième facteur (notons le D), la pratique généralement recommandée consiste à confondre les effets blocs avec des interactions d'ordre élevé des facteurs traitements. La proposition de l'auteur, cependant, est différente.

De façon analogue au cas précédent, c'est maintenant l'effet principal du facteur D qui est confondu avec les effets blocs, et il n'y a cette fois-ci aucun degré de liberté dans la strate inter-blocs pour estimer la variance résiduelle. Il est donc impossible de tester ou de calculer des intervalles de confiance sur l'effet principal de D. Il est également impossible, par exemple, de tester ou de calculer des intervalles de confiance sur la différence entre deux traitements qui ont des modalités de D différentes.

Pour éviter de sacrifier ainsi l'étude de l'un des facteurs, il sera très souvent préférable de confondre des interactions d'ordre élevé avec les effets blocs. Pierre Dagnelie a raison de souligner qu'une telle solution est moins aisée à trouver dans la situation décrite, mais il y a tout de même plusieurs possibilités.

Nous notons A, B, C, D les facteurs traitements et P le facteur bloc. Les méthodes sont plus simples pour des facteurs ayant tous un nombre de niveaux qui est une puissance de 2. Pour l'instant nous allons donc ajouter un quatrième niveau factice au facteur C, que l'on abandonnera par la suite.

Une première méthode consiste à rechercher une fraction régulière pour facteurs à deux niveaux, après avoir assimilé les quatre modalités des facteurs C, D et P aux combinaisons de niveaux de paires de pseudofacteurs (Monod et Bailey, 1992) à deux niveaux $(C1, C2)$, $(D1, D2)$ et $(P1, P2)$. On peut choisir, par exemple, comme relations de définition :

$$P1 = A * B * C1 \quad P2 = B * C2 * D1 \tag{1}$$

Comme $P1 * P2 = A * C1 * C2 * D1$, on voit que seules des interactions triples et plus sont confondues avec les blocs. Tous les effets principaux et interactions doubles sont estimables intra-bloc.

Dans cette méthode les niveaux des facteurs sont tous codés par des -1 (niveau bas) et des $+1$ (niveau haut). Les unités ayant les mêmes valeurs de $P1$ et $P2$ données par la formule (1) sont placées dans le même bloc. Après suppression du niveau fictif, recodages des niveaux et avant randomisation, le plan est :

bloc 1 : 1111, 1112, 1123, 1124, 1233, 1234, 2131, 2132, 2213, 2214, 2221, 2222;
bloc 2 : 1113, 1114, 1121, 1122, 1231, 1232, 2133, 2134, 2211, 2212, 2223, 2224;
bloc 3 : 1131, 1132, 1213, 1214, 1221, 1222, 2111, 2112, 2123, 2124, 2233, 2234;
bloc 4 : 1133, 1134, 1211, 1212, 1223, 1224, 2113, 2114, 2121, 2122, 2231, 2234.

Une solution qui permet d'estimer intra-bloc tous les effets mais avec une légère perte d'efficacité fait appel à la juxtaposition de fractions régulières et à la notion de macroblocs définies dans Monod (1991), Kobilinsky et Monod (1995) :

- deux des blocs reçoivent les traitements vérifiant $A{*}B{*}C1{*}C2{*}D1{*}D2 = 1$, en confondant $C1 * D1$ (et donc $A * B * C2 * D2$) avec l'effet bloc;
- les deux autres blocs reçoivent les traitements vérifiant $A * B * C1 * C2 * D1 * D2 = -1$, en confondant $C2 * D2$ (et donc $A * B * C1 * D1$) avec l'effet bloc.

Dans le plan obtenu (après élimination des traitements avec le niveau factice de C), les effets principaux et les interactions entre deux et trois facteurs sont estimables intra-blocs, si l'on suppose que l'interaction entre les quatre facteurs est nulle. De plus, seuls l'effet principal de D (efficacité 0.92) et les interactions CD (eff : 0.73) et ABD (eff : 0.90) perdent (légèrement) en précision par rapport à un plan orthogonal. Avant randomisation le plan est :

bloc 1 : 1123, 1112, 1134, 1221, 1214, 1232, 2121, 2114, 2132, 2223, 2212, 2234;
bloc 2 : 1122, 1113, 1131, 1224, 1211, 1233, 2124, 2111, 2133, 2222, 2213, 2231;
bloc 3 : 1124, 1111, 1133, 1223, 1212, 1234, 2123, 2112, 2134, 2224, 2211, 2233;
bloc 4 : 1121, 1114, 1132, 1222, 1213, 1231, 2122, 2113, 2131, 2221, 2214, 2232.

Une troisième solution est le recours à un algorithme de recherche de plans optimaux. Un essai rapide avec la procédure PROC OPTEX de SAS montre que l'on peut ainsi obtenir des plans qui perdent les propriétés d'orthogonalité des plans construits algébriquement, mais qui ont de très bons coefficients d'efficacité pour chacun des effets factoriels du modèle excluant uniquement l'interaction $A * B * C * D$. Une solution avant randomisation est par exemple :

bloc 1 : 1234, 1122, 2214, 1133, 2222, 1113, 1223, 1212, 1114, 2134, 2112, 2123;
bloc 2 : 2124, 1224, 1211, 1232, 2111, 1112, 1121, 2221, 2132, 1134, 2212, 2234;
bloc 3 : 2122, 2113, 1222, 1213, 2211, 1111, 1132, 2232, 2223, 1231, 2131, 2233;
bloc 4 : 1131, 1233, 1214, 1124, 2213, 2114, 1221, 2224, 1123, 2121, 2231, 2133.

RÉFÉRENCES BIBLIOGRAPHIQUES

KOBILINSKY, A. et MONOD, H. [1995]. Juxtaposition of regular factorial designs and the complex linear model. *Scand. J. Statist.*, 22, 223–254.

MONOD H. [1990]. Construction de plans d'expérience factoriels (Annexe de l'article de A. Kobilinsky, « Plans factoriels en blocs »). *Journées MODULAD – Applications industrielles de l'analyse des données (Lannion-Tregastel, 14–15 juin 1990)*, N. Valette et P. Villoing eds, pp. 215–221. INRIA, Rocquencourt.

MONOD H. et BAILEY R. A. [1992]. Pseudofactors : normal use to improve design and facilitate analysis. *Applied Statist.*, 41, 317–336.