1 Reparamétrisation dans les plans factoriels non uniformes

Considérons un plan à deux facteurs croisés A et B et soit μ_{ij} l'espérance de la réponse lorsqu'on expérimente avec le niveau i de A, le niveau j de B. Traditionnellement, on réécrit cette espérance sous la forme:

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

où les paramètres μ moyenne générale, α_i effet principal du niveau i de A, β_j effet principal du niveau j de B, γ_{ij} effet d'interaction entre i et j sont définis à partir de la moyenne générale $\mu_{\bullet\bullet}$ et des moyennes marginales $\mu_{i\bullet}$, $\mu_{\bullet j}$:

$$\mu = \mu_{\bullet\bullet}, \quad \alpha_i = \mu_{i\bullet} - \mu_{\bullet\bullet}, \quad \beta_j = \mu_{\bullet j} - \mu_{\bullet\bullet}, \quad \gamma_{ij} = \mu_{ij} - \mu_{i\bullet} - \mu_{\bullet j} + \mu_{\bullet\bullet}\ .$$

Cette reparamétrisation et l'analyse qui s'en suit dépend de façon essentielle de la pondération utilisée pour les moyennes. On peut imaginer d'affecter un poids p_{ij} à chaque observation et de définir les moyennes par niveau i sous la forme $\mu_{i\bullet} = \sum_j p_{ij}\mu_{ij}$. Mais sous sa forme la plus générale, cette définition a l'inconvénient de pouvoir donner des effets principaux non nuls pour A même si l'espérance de la réponse ne dépend en fait que du niveau j du facteur B. Il est donc essentiel d'utiliser une pondération de la forme $p_{ij} = p_i q_j$ qui évite l'écueil précédent.

Lorsqu'il y a des facteurs croisés mais aussi des hiérarchies entre facteurs et que le nombre de niveaux d'un facteur hiérachisé n'est pas le même pour chaque niveau du facteur hiérarchisant (plan non uniforme), la situation se complique et il devient difficile de voir comment définir des moyennes et reparamétrer dans le cas général. Cela explique que les logiciels habituels d'ANOVA ne donnent pas dans cette situation des résultats cohérents, comme cela se voit dans l'exemple du tableau 1 où un même jeu de donnée conduit pour trois logiciels d'usage courant à trois carrés moyens complètement différents pour l'effet du facteur B.

Dans le cas de deux facteurs croisés, les moyennes sont effectuées sur le tableau croisant les niveaux des deux facteurs, i.e. le produit cartésien entre les ensembles de niveaux. La donnée de cet ensemble d'unités est complété par les poids $p_{ij} = p_i q_j$. Cet ensemble d'unité avec la pondération définit ce que j'ai appelé *plan de référence*, qui sert de base pour le calcul des moyennes. Dans le cas plus général, c'est la notion de limite projective qui remplace le produit cartésien et m'a permis de proposer un plan de référence et donc une reparamétrisation orthogonale simple. J'invite le lecteur intéressé à lire l'introduction de ma publication sur le sujet ou mon précédant rapport pour le concours DR1.

Plan et Données			
A	C	B	y
1	1	1	54
1	1	2	14
1	2	1	21
1	2	1	17
1	2	2	36
1	2	2	28
1	3	1	24
1	3	1	25
1	3	2	18
1	3	2	15
2	1	1	17
2	1	1	12
2	1	2	21
2	1	2	25
2	2	1	15
2	2	1	14
2	2	2	18

factorial effect	*ddl*	Carrés Moyens		
		SAS type III	Splus	MINITAB
A	1	314.29	0	314.29
B	1	42.75	0	30.03
$A.B$	1	291.84	0	291.84
$A.C$	3	84.53	84.53	84.53
$A.B.C$	3	317.67	317.67	317.67

TAB. 1 – *Exemple avec C hiérarchisé par A et B croisé avec A et C*

```
SAS

data d;
infile 'nonunif1.don';
input A C B V;
run;
proc glm data=d;
class A C B;
model V=A C(A)  A*B B C*B(A)/ ss3 e3;
lsmeans A C(A) A*B B C*B(A);
run;
```

```
Splus

d<-read.table("nonunif1.don",header=T)
d$a<-factor(d$a)
d$b<-factor(d$b)
d$c<-factor(d$c)
result<-aov(v~a/c*b,d)
drop1.aov(result,scope=result$call)
```

TAB. 2 – *Programmes utilisés pour calculer les carrés moyens dans pour le tableau 1*