

ANALYS : analyse de plans d'expériences

Version 2.2 pour Windows

A. Kobilinsky

Septembre 2000

Unité de Biométrie (BIA)
INRA
78352 JOUY-EN-JOSAS Cedex
FRANCE

Table des matières

1	Introduction	1
2	Fichiers d'entrée et sortie	1
2.1	Généralités	1
2.2	Fichier du plan d'expérience	4
2.3	Fichier des variables	4
2.4	Exemple	5
3	Facteurs, variables, modèle	5
3.1	Généralités	5
3.2	Information sur les facteurs	6
3.3	Contrastes: cas d'un facteur qualitatif à n niveaux	7
3.4	Contrastes: cas d'un facteur quantitatif	11
3.5	Variables	15
3.6	Modèle et parties de modèle	15
3.6.1	Ecriture du modèle	15
3.6.2	Parties de modèle	16
3.6.3	Suppression de termes	16
3.7	Moyennes	17
4	Reparamétrisation et définition des effets factoriels	18
4.1	Cas d'un unique facteur	18
4.2	Cas de deux facteurs	21
4.3	Obtention des moyennes	25
5	Etude d'un plan	26
6	Analyse de variance et moyennes	29
7	Etude des résidus	35
7.1	Exemples	36
7.2	Terminologie et théorie	41

7.3	Aide au diagnostic	43
7.4	Graphique des quantiles	44
7.5	Interprétation	44
8	Détection des effets actifs dans un plan sans répétitions	45
8.1	Graphique des quantiles demi-normaux (Daniel)	46
8.2	Analyse de Box et Meyer	50
8.2.1	Modèle et sur-modèle	52
8.2.2	Facteurs a priori actifs	53
8.2.3	Paramètres de la distribution a priori	54

Table des figures

1	Choix des fichiers à analyser	2
2	Facteurs, variables, modèles	6
3	Choix de la mesure pour un facteur quantitatif	14
4	Options pour l'étude des résidus	36
5	graphique des quantiles, exemple 7.1	37
6	Observé versus prédit, exemple 7.1	38
7	résidu versus prédit, exemple 7.1	38
8	graphique des quantiles, exemple 7.2, var Y1	40
9	Observé et résidu versus prédit, exemple 7.2, var Y1	40
10	graphique des quantiles après élimination unités 23, 13, exemple 7.2	45
11	graphique de Daniel sur l'ensemble des effets, exemple 8.1	47
12	graphique de Daniel après suppression de 4 points, exemple 8.1	47
13	graphique de Daniel sur ensemble des points, exemple 8.2	50
14	graphique de Daniel après suppression de 3 points, exemple 8.2	51
15	Etude de Box et Meyer d'un plan $2^3 \times 4$ en 4 blocs	51
16	Etude de Box et Meyer du plan pour la revivication des bactéries de [2]	52

Liste des tableaux

1	Données du fichier ex2p4r.txt	5
2	Plan pour l'exemple 3.2	10
3	Confusions dans le cas equipondéré sur A	12
4	Confusions dans le cas pondéré sur A	13
5	Contrastes retenus pour les deux facteurs	21
6	Tableaux 3×2 donnant les coefficients des contrastes	22
7	Définition d'un contraste $e(\text{lin } A.B)$	24
8	Analyse d'un $4 \times 4 \times 2$ en 4 blocs	27
9	Exemple 8. Efficacités avec modèle de degré 3	30
10	Exemple 8. Efficacités avec modèle de degré 4	31
11	Exemple 8. Analyse de variance avec modèle de degré 3.	33
12	Exemple 8. Contrastes utilisés dans la paramétrisation	34
13	Exemple 8. Moyennes avec modèle de degré 3.	35
14	Exemple avec deux facteurs croisés	36
15	Aide au diagnostic, exemple 7.1, var Y_1	37
16	Exemple 7.2. Lignes modifiées dans le tableau 8	39
17	Aide au diagnostic, exemple 7.2, var Y_1	39
18	Données simulées pour une fraction 2^{5-1}	46
19	Effets confondus par la fraction 2^{9-4} de l'exemple 8.2	49
20	Ensembles de facteurs actifs considérés dans l'exemple de la figure 15	53

1 Introduction

On a fréquemment recours pour analyser les plans d'expériences à des techniques qui sont difficiles ou même impossible à mettre en oeuvre avec des logiciels classiques de traitement du modèle linéaire. ANALYS permet de mettre en oeuvre commodément certaines de ces techniques, à partir de plans établis par PLANOR ou tout autre logiciel de construction.

Ainsi lors de l'analyse d'un plan de résolution 4, il précise les combinaisons linéaires d'interactions confondues entre elles et détermine la variance d'erreur en prenant en compte ces interactions, bien qu'elles ne soient pas toutes estimables.

Si le nombre de degrés de liberté résiduels est insuffisant, ANALYS permet d'utiliser des techniques spécifiques pour comparer les contrastes comparés entre eux et déterminer ceux qui correspondent à des effets "actifs".

Lorsque l'on utilise la technique d'agrégation de niveaux d'Addelman pour remplacer un facteur à 4 niveaux par un facteur à 3 niveaux, la paramétrisation utilisée doit généralement donner un poids double au niveau résultant de l'agrégation. L'impossibilité d'avoir recours à cette paramétrisation adaptée dans le cas d'un plan de résolution 4 peut faire perdre cette résolution et rendre le plan impossible à analyser. ANALYS donne la possibilité de reparamétrer simplement en prenant en compte un telle pondération.

Les efficacités factorielles qui précisent l'information apportée par le plan sur les différents effets factoriels sont aussi une sortie spécifique de ANALYS, très utile pour caractériser les plans non réguliers.

Les techniques utilisées par ANALYS sont décrites dans [2]. Cette notice donne quelques compléments et précise le mode d'utilisation.

2 Fichiers d'entrée et sortie

2.1 Généralités

Les fichiers d'entrée contiennent d'une part le plan d'expérience, d'autre part les variables observées.

Le plan est un tableau rectangulaire dont les lignes correspondent aux unités, les colonnes aux variables explicatives généralement appelées *facteurs*. Ces facteurs sont soit qualitatifs, soit quantitatifs. Les facteurs quantitatifs sont aussi appelés *covariables*.

Le tableau des variables observées peut figurer dans le même fichier que le plan ou dans un fichier séparé. Ses colonnes sont les variables observées, aussi appelées variables expliquées ou *réponses*. Ses lignes correspondent aux unités et doivent être impérativement dans le même ordre que celles du plan.

Les fichiers de sortie donnent les résultats des analyses. Le programme crée aussi un fichier spécial, suffixé par .PC, dans lequel sont stockés les paramètres introduits lors d'une analyse. Ce fichier peut être réutilisé ultérieurement comme fichier d'entrée pour éviter d'avoir à réintroduire les paramètres déjà introduits.

De façon standard, les graphiques sont affichés à l'écran et on peut alors demander leur impression. Il est aussi possible de les envoyer directement sur l'imprimante. Pour les enregistrer dans un fichier, il suffit de déclarer l'imprimante comme un fichier. Pour ce faire, après avoir dans le poste de travail sélectionné le dossier imprimante, cliquer à droite sur l'icône de l'imprimante, choisir propriétés, onglet détail ou port, et dans *imprimer vers*, choisir *FILE*. Le choix est visualisé par un petit dessin sous l'icône de l'imprimante. Avant chaque création de graphique est alors demandé le nom du fichier qui reçoit le graphique (il est impératif de donner un nom car l'annulation ne marche pas à cet endroit).

Si l'imprimante sélectionnée utilise du postscript, le fichier ainsi créé est un fichier postscript. Mais attention, il n'est généralement pas standard. Pour le réutiliser dans un texte, il faut généralement en supprimer les premières et dernières lignes non standard, éventuellement le convertir avec un programme ad hoc, de façon à obtenir un fichier postscript manipulable et convertissable dans un autre format.

ANALYS donne la possibilité d'étudier un plan d'expérience avant sa réalisation, pour détecter d'éventuels effets confondus et évaluer les efficacités d'estimation des différents effets factoriels. Dans ce cas, on précise qu'on étudie le plan seulement et les indications relatives aux variables, s'il y en a, sont ignorées.

Les renseignements sur les fichiers d'entrée et de sortie et certaines précisions générales sur la forme des sorties sont précisées dans une fenêtre de saisie qui apparaît au lancement du programme (fig. 1).

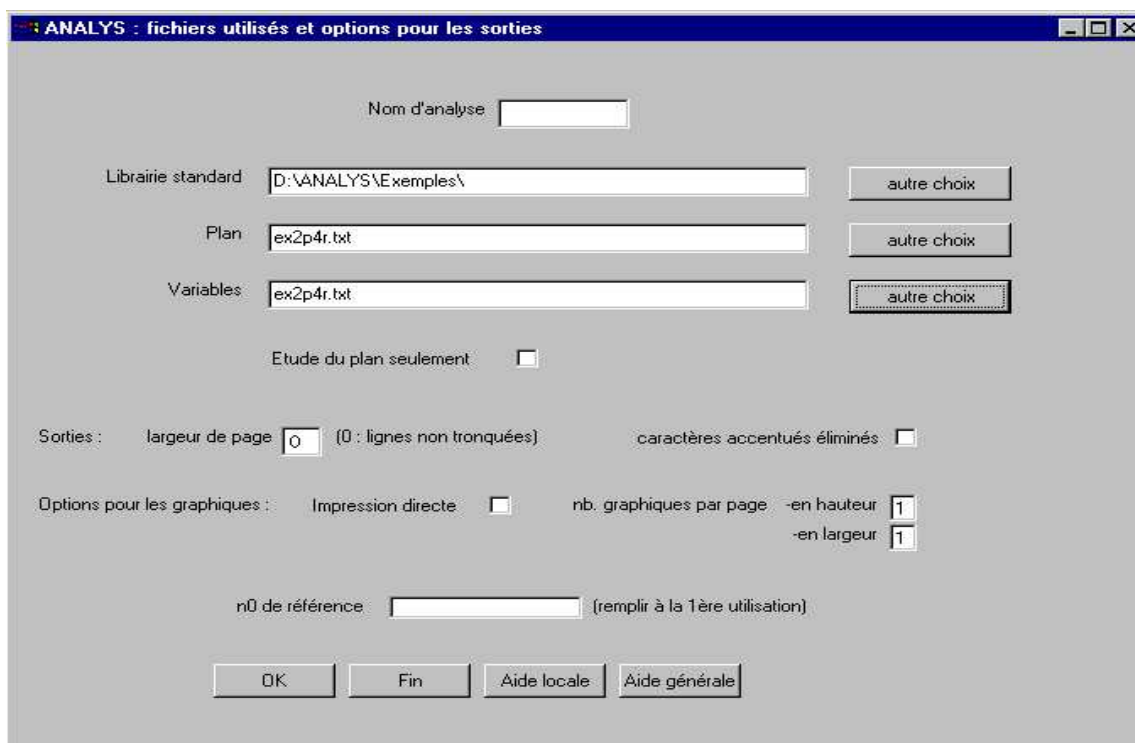


FIG. 1 – *Choix des fichiers à analyser*

Les champs figurant dans cette fenêtre sont les suivants:

Nom d'analyse : utilisé pour former le nom des fichiers créés, à savoir le fichier suffixé

par .PC contenant les paramètres de l'analyse (qui permet de refaire une analyse modifiée sans avoir à réintroduire toute l'information) et le fichier suffixé par .RES contenant de façon standard les résultats. Le nom de ce fichier de résultats peut être aisément modifié, par exemple pour refaire des analyses différentes sur un même jeu de données.

Si ce champs est laissée vide, le nom utilisé est identique à celui du fichier contenant le plan, plus précisément au préfixe de ce nom.

Librairie standard : chemin d'accès au répertoire (directory) qui contient en standard les fichiers lus ou créés par le programme. C'est dans cette librairie que sont stockés le fichier .PC et le fichier de résultats. Il faut cliquer sur le bouton "autre choix" pour changer ce chemin.

Plan : fichier donnant les valeurs des facteurs pour les différentes unités ainsi que les libellés de ces facteurs. Ces facteurs peuvent être qualitatifs ou quantitatifs. Les facteurs quantitatifs (les covariables) sont obligatoirement numériques. Les facteurs qualitatifs peuvent être numériques ou alpha-numériques.

Lorsque le fichier utilisé n'est pas dans la librairie standard, il est indispensable de préciser complètement son chemin d'accès.

Le changement de fichier peut être obtenu en cliquant sur le bouton "autre choix". Les différents types de fichier pouvant contenir le plan sont décrit au paragraphe 2.2.

Variables : fichier donnant les valeurs des variables observées pour les différentes unités ainsi que les libellés de ces variables. Par défaut, si la zone est vide, les variables sont lues sur le même fichier que le plan.

De même que pour le fichier du plan, il faut préciser complètement le chemin d'accès si ce fichier n'est pas dans la librairie standard.

Le changement de fichier peut être obtenu en cliquant sur le bouton "autre choix". Les différents types de fichier pouvant contenir le plan sont décrit au paragraphe 2.3.

Etude du plan seulement : on coche cette case pour étudier les propriétés du plan. Cela inactive le champs précisant le nom du fichier contenant les variables observées, inutilisé dans ce cas là. Le programme précise alors les confusions éventuelles entre effets factoriels et l'efficacité d'estimation pour ceux qui sont non confondus.

Sorties.

largeur de page. Donne la largeur maximum des lignes dans les fichiers de résultats.

Les lignes dépassant cette largeur sont coupées. Le symbole > est utilisé pour indiquer qu'un ligne est la continuation de la précédente. Les matrices sont découpées en blocs pour faciliter la lecture.

Largeur 0 : les lignes ne sont pas coupées et peuvent avoir une largeur quelconque. L'éditeur intégré, de même que la plupart des éditeurs classiques, permet sans difficulté de lire la partie de texte dépassant la largeur de l'écran, mais les fichiers de résultats ne peuvent être imprimés sans modification préalable dans le cas où le fichier de résultat contient des matrices trop larges.

Caractères accentués éliminés. Les caractères accentués peuvent parfois poser des problèmes à l'impression ou lors de la transmission par courrier électronique. C'est pourquoi on donne la possibilité de les éliminer en cochant la case.

Options pour les graphiques

Impression directe : si cette case est cochée, les graphiques sont envoyés directement à l'imprimante sans visualisation préalable sur l'écran.

nb. graphiques par page . Dans bien des cas, il est commode d'obtenir plusieurs graphiques par page. On indique ici les nombres de graphiques en largeur et hauteur sur la page.

Numéro de référence Chaque utilisateur se voit attribuer un numéro personnel de référence à 12 chiffres qu'il doit introduire lors de la première utilisation sur un microordinateur d'un des logiciels ANALYS ou PLANOR.

Les deux paragraphes qui suivent précisent le contenu et la forme des fichiers contenant le plan et les variables observées.

2.2 Fichier du plan d'expérience

Ce fichier donne les libellés des facteurs et covariables, puis leurs valeurs pour les différentes unités. Ce peut être un fichier .PS ou .PR obtenu par PLANOR, un fichier .PC créé lors d'une précédente analyse, un fichier excel (.XLS) ou un fichier texte (*).

Fichier texte: les libellés doivent être sur la ligne 1, les valeurs des facteurs et covariables pour chaque unité sur les lignes suivantes. Si ce fichier contient aussi les variables, celles-ci doivent figurer à droite du plan et un # doit séparer leurs libellés de ceux des facteurs et covariables. Valeurs et libellés sont séparés sur chaque ligne par un ou plusieurs blancs. Le tableau 1 donne un exemple.

Fichier excel: dans ce cas, le programme demande ultérieurement de préciser l'emplacement du plan et des libellés, qui peuvent donc être positionnés assez librement dans le fichier.

Les valeurs des covariables doivent être des nombres. Les niveaux des facteurs qualitatifs peuvent être numériques ou alphanumériques. Le programme permet de redéfinir ces valeurs ou niveaux. Dans le cas d'une covariable, ce sont les nouvelles valeurs qui sont prises en compte dans l'analyse. Dans le cas d'un facteur qualitatif, une telle redéfinition ne change rien à l'analyse mais permet seulement d'explicitier les niveaux si ceux qui apparaissent dans le fichier initial du plan ne sont pas suffisamment explicites.

2.3 Fichier des variables

Ce fichier donne, pour chaque unité, les valeurs des variables observées et les libellés de ces variables. Ce peut être un fichier excel (.XLS), un fichier texte ou un fichier .PC créé lors d'une précédente analyse. Les valeurs non numériques sont considérées comme des données manquantes. Les unités correspondantes sont ignorées si la variable ayant cette donnée manquante est analysée. On peut donc indiquer les valeurs manquantes par une lettre ou n'importe quel symbole non numérique. D'autre part, on peut ignorer des données jugées suspectes dans une analyse antérieure simplement en leur accolant un symbole non numérique (*S* comme Suspecte par exemple).

Fichier texte (*): les libellés doivent être sur la ligne 1, les données sur les lignes suivantes à raison d'une ligne par unité. Valeurs observées et libellés sont séparés sur chaque ligne par un ou plusieurs blancs.

Fichier excel (.XLS): l'emplacement des données et des libellés est précisé ensuite directement sur le fichier et peut donc être choisi librement.

Si le nom du fichier n'est pas précisé, ou s'il est identique au fichier du plan, les variables sont lues sur le même fichier que le plan. On sépare alors facteurs et covariables (à gauche) et variables (à droite) par un # sur la ligne 1 des libellés.

2.4 Exemple

Le tableau 1 donne le contenu du fichier *ex2p4r.txt* apparaissant dans la figure 1. Noter le # qui sépare les libellés des 4 facteurs pH, source de Carbone, dose d'Azote, et dose d'extrait de levure des deux variables NBG, nombre de bactéries par gramme, et logNBG qui est le logarithme à base 10 de NBG.

pH	sourceC	doseN	extrlev	#NBG	logNBG
-1	-1	-1	-1	8700000000	9.9395
-1	-1	-1	1	4600000000	9.6628
-1	-1	1	-1	8800000000	9.9445
-1	-1	1	1	4000000000	8.6021
-1	1	-1	-1	8400000000	9.9243
-1	1	-1	1	5800000000	9.7634
-1	1	1	-1	2600000000	9.415
-1	1	1	1	4900000000	8.6902
1	-1	-1	-1	4200000000	9.6232
1	-1	-1	1	4500000000	9.6532
1	-1	1	-1	9100000000	9.959
1	-1	1	1	8800000000	8.9445
1	1	-1	-1	12000000000	10.0792
1	1	-1	1	2400000000	9.3802
1	1	1	-1	6500000000	8.8129
1	1	1	1	3300000000	9.5185
-1	-1	-1	1	8600000000	9.9345
-1	-1	1	-1	4800000000	9.6812
-1	1	-1	1	4700000000	9.6721
-1	1	1	-1	1600000000	9.2041
1	-1	-1	1	2300000000	9.3617
1	-1	1	-1	6500000000	9.8129
1	1	-1	1	4800000000	9.6812
1	1	1	-1	1000000000	9.0000

TAB. 1 – Données du fichier *ex2p4r.txt*

3 Facteurs, variables, modèle

3.1 Généralités

Après validation de l'écran précisant les fichiers d'entrée apparaît l'écran de la figure 2 où figurent

- à gauche une liste des facteurs et éventuellement des variables, suivis de champs permettant d'apporter sur chacun d'eux des informations complémentaires;

- à droite des zones permettant de préciser le modèle plus éventuellement un sur-modèle et les moyennes que l'on veut calculer;
- en bas à droite, des cases à cocher pour obtenir,
 - lorsqu'on ne se limite pas à l'étude du plan, les graphiques de Daniel, l'analyse de Box-Meyer, l'étude des résidus. Un bouton d'option est associé à chacun de ces types d'étude.
 - lorsqu'on se limite à l'étude du plan, la sortie de la matrice X du modèle linéaire, les confusions d'effets, les efficacités.
- en bas à gauche une case à cocher pour ordonner les paramètres du modèle résultant de l'analyse de variance dans l'ordre décroissant de leurs valeurs absolues.

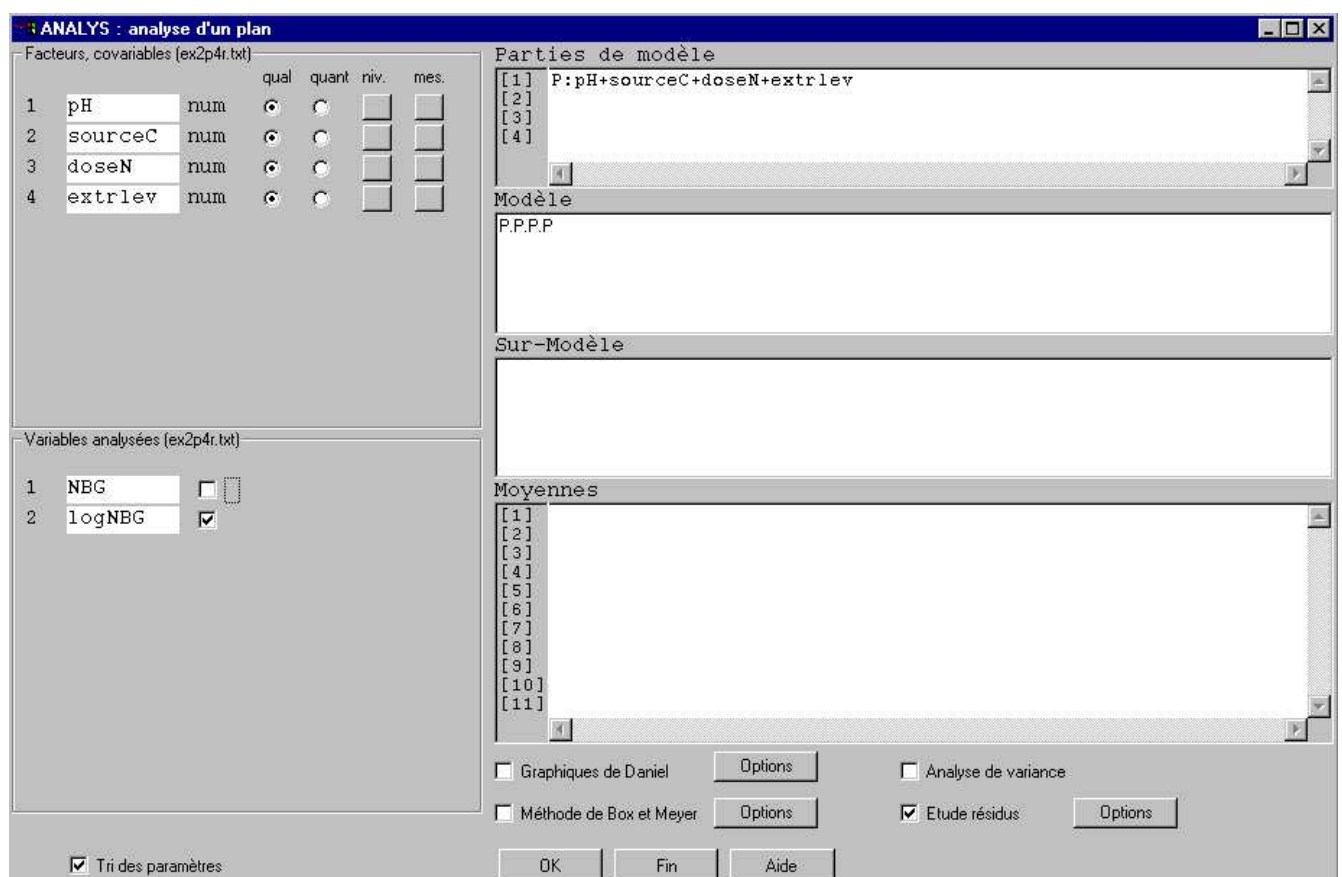


FIG. 2 – Facteurs, variables, modèles

3.2 Information sur les facteurs

On précise s'il sont qualitatifs ou quantitatifs (colonnes *qual* et *quant*). On peut modifier leurs niveaux (colonne *niv.*). On peut choisir les contrastes utilisés pour reparamétriser (colonne *mes.*). Ce choix, explicite dans le cas qualitatif, passe par l'introduction d'une mesure dans le cas quantitatif.

A droite du champs donnant le libellé du facteur, le programme indique si le facteur est numérique (*num*) ou alphanumérique (*alpha*). Un facteur est numérique si tous ses niveaux

dans le fichier contenant le plan sont numériques ainsi que les niveaux éventuellement introduits au moyen d'un bouton *niveaux* dans PLANOR ou ANALYS. Rappelons qu'un facteur quantitatif (une covariable) doit nécessairement être numérique.

Les niveaux.

On peut les examiner et éventuellement les modifier en appuyant sur le bouton associé de la colonne *niv*. Les niveaux qui figurent initialement sont ceux du fichier du plan. Il n'y a en règle générale pas à les modifier, mais si on les modifie il faut savoir que

- pour les facteurs quantitatifs, ce sont les niveaux modifiés qui sont pris en compte dans l'analyse,
- pour les facteurs qualitatifs, la modification des niveaux ne modifie que les sorties. L'analyse utilise les numéros des niveaux qui restent eux inchangés, même si par inadvertance on a mis deux fois le même niveau.

Contrastes.

La reparamétrisation utilisée dans l'analyse de variance dépend des contrastes choisis pour chaque facteur. Sur cette reparamétrisation, on trouvera une description détaillée dans [2] pour le cas où tous les facteurs sont croisés, où les cas qui s'y ramènent après d'éventuelles décompositions en pseudofacteurs. Ces cas excluent les hiérarchies non uniformes, pour lesquelles il faut utiliser la reparamétrisation plus générale décrite dans [3].

Dans les deux paragraphes qui suivent, on précise sur des exemples très simples comment s'opère, pour chaque facteur, le choix des contrastes utilisés. On illustre aussi l'impact que peut avoir ce choix sur l'analyse effectuée.

Le paragraphe 4 précise et justifie le choix des contrastes dans des cas plus généraux, puis montre comment on déduit de ces choix sur chaque facteur la paramétrisation utilisée pour l'ensemble du modèle.

Pour examiner et le cas échéant modifier les contrastes relatif à un facteur, on appuie sur le bouton associé de la colonne *mes.*. Deux cas se présentent alors selon que le facteur est qualitatif ou quantitatif.

3.3 Contrastes : cas d'un facteur qualitatif à n niveaux

Dans ce cas le programme propose les $m = n - 1$ contrastes retenus pour l'analyse sous forme de vecteurs colonnes notés c_1, c_2, \dots, c_m que l'utilisateur peut modifier librement.

Si ces contrastes ne sont pas orthonormés, ils sont modifiés pour être rendus orthonormés pour le produit scalaire usuel pondéré par des poids égaux ou inégaux. Le vecteur des poids est également donné en colonne. Les poids proposés initialement sont tous égaux à 1. L'utilisateur peut les modifier à sa guise. Quels que soient les poids introduits, le programme les divise par leur somme pour obtenir des poids de somme 1.

Exemple 3.1 . Facteur A à 3 niveaux 5, 6, 7. Lorsqu'on clique sur le bouton *mes.*, on

obtient le tableau suivant

Contrastes pour A

niv	Poids	c1	c2
5	1	-1	-1
6	1	1	-1
7	1	0	2

(1)

Le premier contraste c_1 oppose le niveau 6 au niveau 5, le second c_2 oppose le niveau 7 à la moyenne des niveaux 5 et 6. Ces deux contrastes sont orthogonaux pour le produit scalaire usuel

$$(-1, 1, 0)(-1, -1, 2)' = 0.$$

Ils sont aussi tous deux orthogonaux au vecteur $(1, 1, 1)'$ associé à la moyenne.

Pour préciser, considérons le cas simple d'un plan ne comportant que le facteur A et notons $\tau(5)$, $\tau(6)$, $\tau(7)$ les espérances de la réponse pour les niveaux 5, 6, 7 de ce facteur. Alors la moyenne et les deux contrastes de (1) sont définis, à une constante près (symbole \propto), par:

$$\begin{aligned} \mu &\propto \tau(5) + \tau(6) + \tau(7), \\ c_1 &\propto -\tau(5) + \tau(6), \\ c_2 &\propto -\tau(5) - \tau(6) + 2\tau(7). \end{aligned}$$

Dans les sorties du programme, on utilise les notations A^1 , A^2 pour désigner ces contrastes c_1 et c_2 introduits pour le facteur A .

Cas pondéré

Si le niveau 6 est deux fois plus représenté que les deux autres, il peut être judicieux de remplacer le tableau (1) proposé de façon standard par

Contrastes pour A

niv	Poids	c1	c2
5	1	-1	-1
6	2	0	1
7	1	1	-1

(2)

Dans ce cas la moyenne et les 2 contrastes associés sont

$$\begin{aligned} \mu &\propto \tau(5) + 2\tau(6) + \tau(7), \\ c_1 &\propto -\tau(5) + \tau(7), \\ c_2 &\propto -\tau(5) + 2\tau(6) - \tau(7). \end{aligned}$$

Noter que pour obtenir les coefficients ci-dessus on a multiplié le vecteur $(1, 1, 1)$ associé à la moyenne et les colonnes c_1 et c_2 par les poids correspondants. Les termes de droite des égalités sont donc de la forme $p(5)c(5)\tau(5) + p(6)c(6)\tau(6) + p(7)c(7)\tau(7)$ où $p(5)$, $p(6)$, $p(7)$ sont les poids et $c(5)$, $c(6)$, $c(7)$ les coefficients associés à la moyenne ou au contraste considéré. En particulier, les coefficients $p(5) = 1$, $p(6) = 2$, $p(7) = 1$ définissant la moyenne μ donnent un poids double au niveau 6 deux fois plus représenté.

L'orthogonalité de deux vecteurs $c_1 = (c_1(5), c_1(6), c_1(7))$ et $c_2 = (c_2(5), c_2(6), c_2(7))$ est dans ce cas pondéré définie par

$$0 = \langle c_1, c_2 \rangle = p(5)c_1(5)c_2(5) + p(6)c_1(6)c_2(6) + p(7)c_1(7)c_2(7).$$

Avec cette formule, on vérifie immédiatement que les colonnes c_1, c_2 dans (2) sont orthogonales et orthogonales au vecteur $(1,1,1)$ associé à μ .

Les contrastes définis par (1) et (2) ne sont pas normés. Le programme les norme pour la norme $\|c\| = \sqrt{\langle c, c \rangle}$ associée au produit scalaire, définie donc par

$$\|c\|^2 = \langle c, c \rangle = p(5)c(5)c(5) + p(6)c(6)c(6) + p(7)c(7)c(7).$$

Les poids $p(i)$ sont ici obtenus en divisant les poids introduits par leur somme et sont donc égaux à $(1/3, 1/3, 1/3)$ dans le cas standard (1) et $(1/4, 1/2, 1/4)$ dans le cas pondéré (2). Les contrastes normés figurent en tête des sorties que l'on reproduit ci-dessous pour les deux cas de figure considérés.

contrastes pour facteur A				contrastes pour facteur A			
Cas équi pondéré (1)				Cas pondéré (2)			
niv	c0	c1	c2	niv	c0	c1	c2
5	1	-1.22474	-0.70711	5	1	-1.41421	-1
6	1	1.22474	-0.70711	6	1	0.00000	1
7	1	0.00000	1.41421	7	1	1.41421	-1

Si les contrastes entrés par l'utilisateur ne sont pas orthogonaux, le programme le diagnostique et les remplace par des contrastes orthogonaux, obtenus par orthogonalisation successive du vecteur $(1,1,1)$ et des vecteurs c_1 et c_2 . Ainsi, si dans le tableau proposé en (1) on remplace le poids du niveau 6 par 2 sans modifier c_1 et c_2 (tableau de gauche ci-dessous), on obtient la sortie figurant à droite

Définition des contrastes pour A

niv	Poids	c1	c2
5	1	-1	-1
6	2	1	-1
7	1	0	2

ATTENTION: contrastes pour A non orthogonaux

contrastes orthonormés pour A

niv	c0	c1	c2
5	1	-1.50756	-0.85280
6	1	0.90453	-0.42640
7	1	-0.30151	1.70561

Rappelons que l'orthonormalité s'entend pour le produit scalaire pondéré par les poids $1/4, 1/2, 1/4$. Ainsi, le fait que c_1 soit normé et orthogonal à c_2 se traduit par les égalités

$$1 = (-1.50756)^2/4 + 0.90453^2/2 + (-0.30151)^2/4,$$

$$0 = (-1.50756) \times (-0.85280)/4 + 0.90453 \times (-0.42640)/2 + (-0.30151) \times 1.70561/4.$$

Pour certains dispositifs expérimentaux, le choix de contrastes utilisant une pondération du type (2) permet d'avoir des confusions d'effets plus facile à examiner comme le montre l'exemple suivant.

Exemple 3.2 . Fromage Arzúa Ulloa On étudie l'influence, sur la texture du fromage d'Arzúa-Ulloa, de 7 facteurs A, B, C, D, E, F, G caractéristiques du processus de fabrication

(moment de salage, addition de calcium, quantité de levain, etc . . .). Ces facteurs sont étudiés à 2 niveaux, sauf le premier d'entre eux, le moment de salage A , qui doit être impérativement étudié à 3 niveaux. Pour cette étude sont prévues 32 fabrications qui sont groupées par bloc de 4 correspondant aux fabrications réalisées sur une journée (facteur j). Un plan possible est celui du tableau 2. Son mode de construction est décrit de façon détaillée dans [2]. En fait il est déduit d'une fraction régulière $1/8$ de résolution 4 du plan factoriel complet 4×2^6 en agrégeant deux niveaux du facteur A à 4 niveaux pour le transformer en facteur à 3 niveaux. On a fait figurer dans le tableau 2 les pseudofacteurs A_1, A_2 utilisé pour définir A ainsi que les trois pseudofacteurs utilisés pour construire j .

n0	A_1	A_2	j_1	j_2	j_3	A	B	C	D	E	F	G	j
1	0	0	0	0	0	1	1	1	1	1	1	1	0
2	1	0	0	0	0	2	1	1	1	-1	1	1	0
3	0	1	1	1	1	2	1	1	1	1	-1	-1	7
4	1	1	1	1	1	3	1	1	1	-1	-1	-1	7
5	0	0	1	0	0	1	-1	1	1	-1	-1	-1	4
6	1	0	1	0	0	2	-1	1	1	1	-1	-1	4
7	0	1	0	1	1	2	-1	1	1	-1	1	1	3
8	1	1	0	1	1	3	-1	1	1	1	1	1	3
9	0	0	0	1	0	1	1	-1	1	-1	-1	1	2
10	1	0	0	1	0	2	1	-1	1	1	-1	1	2
11	0	1	1	0	1	2	1	-1	1	-1	1	-1	5
12	1	1	1	0	1	3	1	-1	1	1	1	-1	5
13	0	0	1	1	0	1	-1	-1	1	1	1	-1	6
14	1	0	1	1	0	2	-1	-1	1	-1	1	-1	6
15	0	1	0	0	1	2	-1	-1	1	1	-1	1	1
16	1	1	0	0	1	3	-1	-1	1	-1	-1	1	1
17	0	0	0	0	1	1	1	1	-1	-1	1	-1	1
18	1	0	0	0	1	2	1	1	-1	1	1	-1	1
19	0	1	1	1	0	2	1	1	-1	-1	-1	1	6
20	1	1	1	1	0	3	1	1	-1	1	-1	1	6
21	0	0	1	0	1	1	-1	1	-1	1	-1	1	5
22	1	0	1	0	1	2	-1	1	-1	-1	-1	1	5
23	0	1	0	1	0	2	-1	1	-1	1	1	-1	2
24	1	1	0	1	0	3	-1	1	-1	-1	1	-1	2
25	0	0	0	1	1	1	1	-1	-1	1	-1	-1	3
26	1	0	0	1	1	2	1	-1	-1	-1	-1	-1	3
27	0	1	1	0	0	2	1	-1	-1	1	1	1	4
28	1	1	1	0	0	3	1	-1	-1	-1	1	1	4
29	0	0	1	1	1	1	-1	-1	-1	-1	1	1	7
30	1	0	1	1	1	2	-1	-1	-1	1	1	1	7
31	0	1	0	0	0	2	-1	-1	-1	-1	-1	-1	0
32	1	1	0	0	0	3	-1	-1	-1	1	-1	-1	0

TAB. 2 – Plan pour l'exemple 3.2

A gauche figurent les pseudofacteurs A_1, A_2, j_1, j_2, j_3 ayant servi à la construction de A et de j

Pour montrer l'impact du choix des contrastes, on a utilisé 4 systèmes de contrastes EQP0, EQP1, P0, P1.

Les deux premiers systèmes (tableau 3) donnent des poids égaux aux niveaux de A . Ils diffèrent dans le choix de c_1 et c_2 . EQP0 garde les contrastes dits de Helmet proposés en standard, tandis que EQP1 modifie ces contrastes pour respecter la symétrie entre les niveaux 1 et 3, tous deux présents 8 fois dans le plan, alors que le niveau 2 résultant de l'agrégation apparaît 16 fois.

Les deux autres systèmes P0 et P1 (tableau 4) donnent un poids double au niveau 2. Dans le premier P0, on a gardé les contrastes c_1 et c_2 standards, tandis que dans le second P1 on a introduit des contrastes respectant la symétrie des niveaux 1 et 3. Avec les poids,

ces derniers sont orthogonaux et orthogonaux au vecteur $(1,1,1)$ associé à la moyenne. En revanche ceux de P_0 ne sont pas orthogonaux. Ils sont donc remplacés par des contrastes orthogonaux analogue à ceux figurant à droite de (3).

Les tableaux 3 et 4 précisent les effets confondus avec chacun de ces systèmes. Le modèle utilisé pour les obtenir est $P.P + j_1.j_2.j_3$ où P est la somme $A + B + C + D + E + F + G$ des 7 facteurs. Nous verrons au paragraphe 3.6 que cette formulation conduit à un modèle contenant les effets principaux, les interactions entre deux facteurs et les effets blocs. Le fait d'exprimer l'effet bloc sous forme du produit entre les pseudofacteurs j_1, j_2, j_3 permet de simplifier l'expression des confusions avec les effets blocs. Les paramètres associés à l'effet bloc sont alors les 7 interactions $j_1, j_2, j_1.j_2, j_3, j_1.j_3, j_2.j_3, j_1.j_2.j_3$. Les effets notés A et A^2 correspondent aux 2 contrastes sélectionnés pour A . Les effets comme $A.B, A^2.B$ sont les interactions de ces contrastes avec le facteur B . Le sens de ces effets dépend donc de façon étroite du système de contrastes utilisé.

Alors qu'avec les paramétrisations P_0 et P_1 utilisant les poids, les effets principaux ne se confondent avec aucune interaction, avec les systèmes EQP_0 et EQP_1 , cinq des effets principaux sont confondus. L'interprétation est donc quasiment impossible avec ces deux derniers systèmes.

Le choix de la pondération est crucial. Il n'en est pas de même du choix des contrastes c_1 et c_2 . Ainsi quand on passe de P_0 à P_1 on voit que les confusions concerne les mêmes paires d'interactions. Par exemple l'interaction $E.G$ est la seule confondue avec un contraste de l'interaction $A.C$ qu'on utilise P_0 ou P_1 . Il est en fait facile de montrer que l'espace de contrastes associé est le même dans les deux cas. On peut donc utiliser presque indifféremment P_0 ou P_1 . Le second système P_1 apparaît cependant légèrement préférable parce qu'il conduit à une expression plus simple des confusions. Avec P_1 , le contraste confondu est précisément $A^2.C$. Rappelons que A^2 traduit dans ce cas l'opposition entre le niveau dupliqué 2 et les deux autres 1, 3 et que $A^2.C$ est donc l'interaction entre ce contraste et C .

3.4 Contrastes : cas d'un facteur quantitatif

Sauf dans le cas où l'on choisit l'option "covariable non transformée", les puissances successives 1, 2, ... de la covariable sont transformées en polynômes orthonormés de degré 1, 2, ...

L'orthogonalité fait référence à une mesure qui doit être précisée.

En standard, la mesure utilisée est celle qui donne un poids 1 à chacune des valeurs distinctes prises par la covariable (mesure discrète sur les valeurs prises et uniforme). On peut choisir plutôt de donner à chacune des valeurs prises un poids proportionnel à son nombre d'occurrences dans les données.

Pour une covariable dont les niveaux ne sont pas rigoureusement contrôlés, il peut être préférable de spécifier le support de la mesure indépendamment des données (mesure discrète sur valeurs spécifiées). On peut dans ce cas introduire des poids non uniforme également.

Exemple 3.3 . Facteur quantitatif A à 4 niveaux Ce cas se présente fréquemment parce qu'il est facile de construire de bonnes fractions de plan lorsqu'on a un mélange de

<i>EQP0</i>				<i>EQP1</i>			
Contrastes pour <i>A</i>							
niv	Poids	c1	c2	niv	Poids	c1	c2
1	1	-1	-1	1	1	-1	1
2	1	1	-1	2	1	0	-2
3	1	0	2	3	1	1	1
A A^2 B $C + 0.333E.G$ $D + 0.333E.F$ $E + 0.333C.G + 0.333D.F - 0.333j_1.j_2.j_3$ $F + 0.333D.E$ $G + 0.333C.E$ $A.B$ $A^2.B$ $A.C - 0.817E.G$ $A^2.C + 0.471E.G$ $A.D - 0.817E.F$ $A^2.D + 0.471E.F$ $A.E - 0.817C.G - 0.817D.F + 0.817j_1.j_2.j_3$ $A^2.E + 0.471C.G + 0.471D.F - 0.471j_1.j_2.j_3$ $A.F - 0.817D.E$ $A^2.F + 0.471D.E$ $A.G - 0.817C.E$ $A^2.G + 0.471C.E$ $B.C + j_1.j_2$ $B.D + j_1.j_3$ $B.E$ $B.F - j_2$ $B.G - j_3$ $C.D + F.G + j_2.j_3$ $C.F + D.G - j_1$				A A^2 B $C + 0.33333E.G$ $D + 0.33333E.F$ $E + 0.33333C.G + 0.33333D.F - 0.33333j_1.j_2.j_3$ $F + 0.33333D.E$ $G + 0.33333C.E$ $A.B$ $A^2.B$ $A.C$ $A^2.C + 0.94281E.G$ $A.D$ $A^2.D + 0.94281E.F$ $A.E$ $A^2.E + 0.94281C.G + 0.94281D.F - 0.94281j_1.j_2.j_3$ $A.F$ $A^2.F + 0.94281D.E$ $A.G$ $A^2.G + 0.94281C.E$ $B.C + j_1.j_2$ $B.D + j_1.j_3$ $B.E$ $B.F - j_2$ $B.G - j_3$ $C.D + F.G + j_2.j_3$ $C.F + D.G - j_1$			

TAB. 3 – Confusions dans le cas equipondéré sur *A*

$P0$				$P1$					
Contrastes pour A									
niv	Poids	c1	c2	(*)	niv	Poids	c1	c2	
1	1	-1	-1		1	1	-1	1	
2	2	1	-1		2	2	0	-1	
3	1	0	2		3	1	1	1	
A				A					
A^2				A^2					
B				B					
C				C					
D				D					
E				E					
F				F					
G				G					
$A.B$				$A.B$					
$A^2.B$				$A^2.B$					
$A.C - 0.90453E.G$				$A.C$					
$A^2.C + 0.4264E.G$				$A^2.C + E.G$					
$A.D - 0.90453E.F$				$A.D$					
$A^2.D + 0.4264E.F$				$A^2.D + E.F$					
$A.E - 0.90453C.G - 0.90453D.F + 0.90453j_1.j_2.j_3$				$A.E$					
$A^2.E + 0.4264C.G + 0.4264D.F - 0.4264j_1.j_2.j_3$				$A^2.E + C.G + D.F - j_1.j_2.j_3$					
$A.F - 0.90453D.E$				$A.F$					
$A^2.F + 0.4264D.E$				$A^2.F + D.E$					
$A.G - 0.90453C.E$				$A.G$					
$A^2.G + 0.4264C.E$				$A^2.G + C.E$					
$B.C + j_1.j_2$				$B.C + j_1.j_2$					
$B.D + j_1.j_3$				$B.D + j_1.j_3$					
$B.E$				$B.E$					
$B.F - j_2$				$B.F - j_2$					
$B.G - j_3$				$B.G - j_3$					
$C.D + F.G + j_2.j_3$				$C.D + F.G + j_2.j_3$					
$C.F + D.G - j_1$				$C.F + D.G - j_1$					

TAB. 4 – Confusions dans le cas pondéré sur A

(*) Les contrastes introduits n'étant pas orthogonaux sont remplacés par des contrastes orthogonaux de façon analogue à ce qui est indiqué en (3).

facteurs à 2 niveaux et de facteurs à 4 niveaux qualitatifs ou quantitatifs.

Si on déclare le facteur A quantitatif et que l'on clique sur le bouton *mes.*, on obtient la fenêtre figurant à droite de la figure 3 qui permet de modifier les options standards.

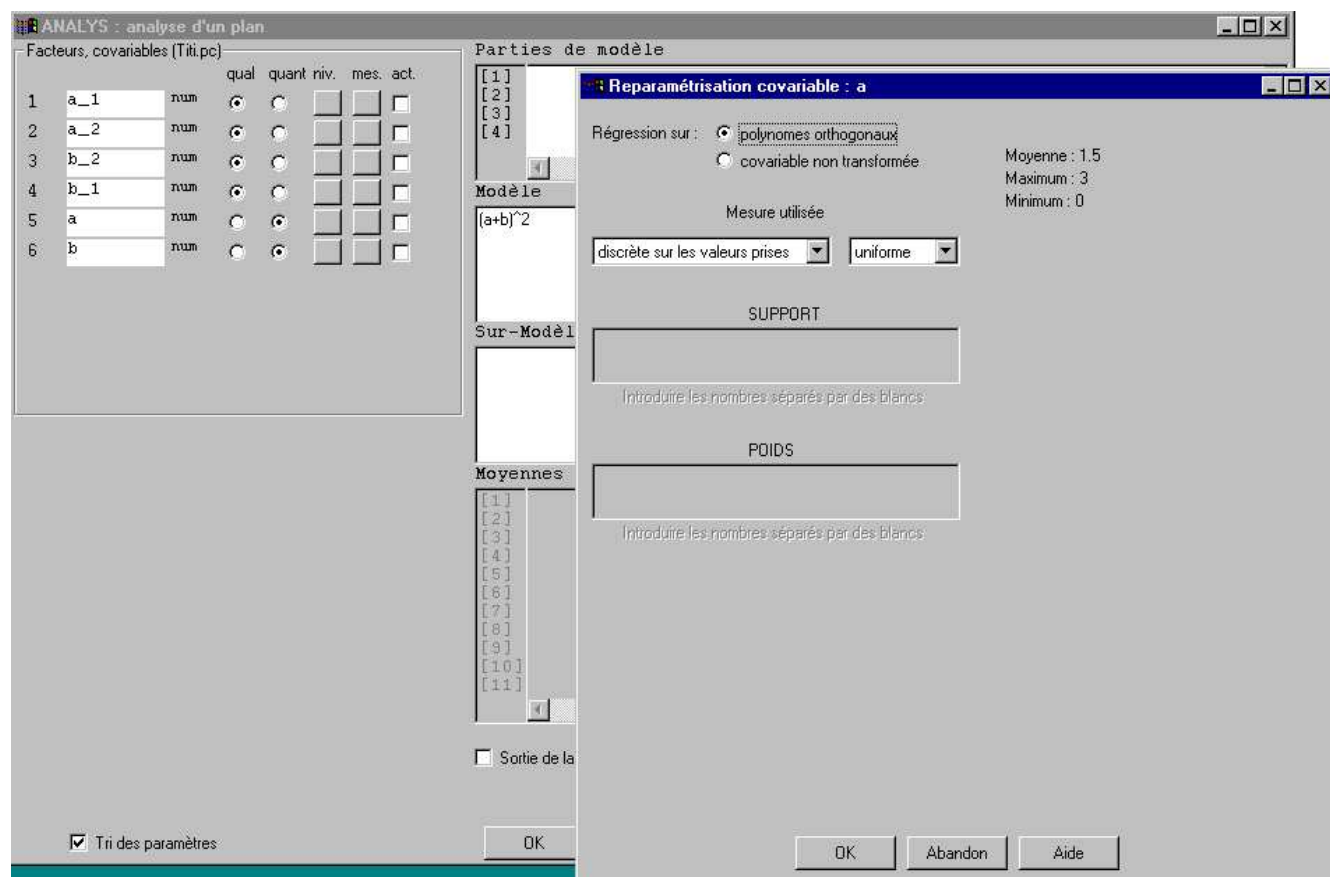


FIG. 3 – Choix de la mesure pour un facteur quantitatif

En tête des sorties sont précisés les polynômes orthogonaux retenus pour définir les effets linéaire, quadratique, cubique, etc ... Ainsi si les niveaux de A sont 5, 6, 7, 8, que le degré le plus fort de A dans le modèle est 2 et qu'on retient le choix standard d'une mesure equipondérée sur ces 4 niveaux, on obtient

coeff. pol. orth. pour A . Chaque colonne donne les coeff. à appliquer aux puissances successives de $A - 6.5$. Le libellé 0 , 1 , 2 ... à gauche précise cette puissance. Le libellé 0 , 1 , 2 , ... en haut de col. donne le degré du polynôme orthogonal.'

	0	1	2
0	1	0.00000	-1.25
1	0	0.89443	0.00
2	0	0.00000	1.00

Ceci signifie que les polynômes orthonormés de degré 0, 1, 2 sont:

$$P_0(A) = 1, \quad P_1(A) = 0.89443(A - 6.5), \quad P_2(A) = -1.25 + (A - 6.5)^2.$$

A ces polynômes sont associés la moyenne μ sur A et les contrastes notés $\text{lin } A$, $\text{quad } A$. Ces

contrastes peuvent être présentés sous la forme suivante analogue à (1):

Contrastes de degré 0, 1, 2

A	poids	0 μ	1 lin A	2 quad A
5	1	1	-1.3416	1
6	1	1	-0.4472	-1
7	1	1	0.4472	-1
8	1	1	1.3416	1

(4)

Les coefficients de la colonne lin A sont $P_1(5) = 0.89443(5 - 6.5)$, $P_1(6)$, $P_1(7)$, $P_1(8)$, ceux de la colonne quad A sont $P_2(5) = (5 - 6.5)^2 - 1.25$, $P_2(6)$, $P_2(7)$, $P_2(8)$, enfin les coefficients 1 de la colonne μ sont $P_0(5)$, $P_0(6)$, $P_0(7)$, $P_0(8)$. Il est facile de vérifier l'orthonormalité de ces trois colonnes.

3.5 Variables

On sélectionne les variables analysées avec le modèle retenu. Les lignes comportant une donnée non numérique pour ces variables sont éliminées. Si on veut analyser avec un même modèle plusieurs variables pour lesquelles il y a des valeurs manquantes pour des unités distinctes, il peut être judicieux d'analyser séparément les variables qui n'ont pas les mêmes données manquantes pour éviter de perdre d'information.

Lorsqu'une première analyse fait apparaître que certaines données sont suspectes, on peut éliminer les unités correspondantes d'une analyse ultérieure simplement en accolant à ces données un symbole non numérique (par exemple S comme Suspect)

3.6 Modèle et parties de modèle

La syntaxe est semblable à celle de PLANOR. Mais en outre dans ANALYS, les facteurs quantitatifs peuvent être introduits à différentes puissances, auxquelles sont associés les effets polynomiaux usuels. Le modèle est reparamétré en utilisant pour chaque facteur soit les contrastes standards, soit ceux introduits par l'utilisateur.

3.6.1 Ecriture du modèle

Un modèle ou une partie de modèle est une somme de termes, chaque terme étant un produit de facteurs ou covariables correspondant à un effet principal ou à une interaction. Le terme 1 est réservé à la constante.

Le symbole \wedge après une covariable permet de l'élever à différentes puissances. Sauf dans le cas où on laisse ces puissances non transformées, le terme associé à la puissance i est l'effet polynomial de degré i .

Exemple 3.4 . Si VAR (variété) et DENS (densité de semis) sont des facteurs qualitatifs, DOSE la covariable dose d'engrais azoté, le modèle :

$1 + \text{BL} + \text{VAR} + \text{DOSE} + \text{DOSE}^2 + \text{DENS} + \text{VAR}.\text{DOSE} + \text{VAR}.\text{DOSE}^2 + \text{DENS}.\text{DOSE} + \text{DENS}.\text{DOSE}^2 + \text{VAR}.\text{DENS}$

contient une constante notée symboliquement 1, un effet bloc BL, les effets principaux et interactions simples entre les facteurs VAR, DOSE, DENS où DOSE apparait avec un degré inférieur ou égal à 2. A DOSE, DOSE² sont associés les effets linéaires et quadratiques de DOSE. Les interactions VAR.DOSE et VAR.DOSE² sont les interactions de ces deux effets avec l'effet variété.

Dans ANALYS, le modèle est complété par tous les sous-termes. Ainsi la présence de DENS.DOSE² entraîne celle de 1, DENS, DOSE, DOSE², DENS.DOSE. Le modèle donné précédemment peut donc être raccourci en :

$\text{BL} + \text{VAR}.\text{DOSE}^2 + \text{DENS}.\text{DOSE}^2 + \text{VAR}.\text{DENS}$.

L'écriture peut aussi utiliser des parenthèses. En développant, on élimine dans chaque terme les facteurs qualitatifs redondants ($\text{VAR}.\text{VAR} \rightarrow \text{VAR}$), puis les termes redondants ($\text{VAR}.\text{DENS} + \text{DENS}.\text{VAR} \rightarrow \text{VAR}.\text{DENS}$). Le modèle ci-dessus peut donc s'écrire aussi

$\text{BL} + (1 + \text{VAR} + \text{DENS})(\text{VAR} + \text{DENS} + \text{DOSE} + \text{DOSE}^2)$

et si on complète le modèle par les sous termes

$\text{BL} + (\text{VAR} + \text{DENS})(\text{VAR} + \text{DENS} + \text{DOSE}^2)$

3.6.2 Parties de modèle

Pour raccourcir l'écriture de certains modèle, on peut donner un nom à une partie de ce modèle. Par exemple, si il y a 10 facteurs qualitatifs notés A, B, ..., J, il est commode d'écrire le modèle comprenant toutes les interactions de 2 de ces facteurs sous la forme :

$P.P$

où P est défini comme la somme formelle des 10 facteurs

$P: A+B+C+D+E+F+G+H+I+J$

Si l'un de ces facteurs, disons A, est quantitatif, le développement de P.P donne le terme A.A équivalent à A². Si l'on veut aussi inclure les interactions avec l'effet cubique de A, on peut écrire le modèle

$Q.Q + Q.A^3$

où la partie Q est la somme des facteurs B à J

$Q: B+C+D+E+F+G+H+I+J$

Le modèle étant complété par les sous-termes, la présence de B.A³ implique celle de B.A et B.A².

3.6.3 Suppression de termes

Il peut être commode dans certains cas de supprimer des termes du modèle. On fait alors suivre le modèle du signe ~ suivi de la partie à supprimer. Cette dernière est écrite avec les mêmes règles que le modèle. La complétion par les sous-termes n'est opérée qu'après la suppression.

Ainsi pour supprimer les interactions A.B et C.D du modèle P.P introduit au para-

graphe précédent 3.6.2 avec des facteurs tous qualitatifs, le modèle peut être écrit

$$(A+B+E+F+G+H+I+J)(C+D+E+F+G+H+I+J)$$

ou avec une partie de modèle Q: E+F+G+H+I+J

$$(A+B+Q)(C+D+Q)$$

ou en supprimant des termes

$$P.P \sim A.B+C.D.$$

La dernière expression est équivalente au précédente dans le cas de facteurs qualitatifs puisque, dans le développement de $P.P$, les produits $A.A$, $B.B$ sont remplacés par A et B . Mais si A est quantitatif, il n'y a plus équivalence puisque $P.P$ contient $A.A$, équivalent à A^2 , qui ne figure en revanche pas dans le développement de $(A+B+Q)(C+D+Q)$.

Si A , B , C sont trois facteurs quantitatifs, l'expression $(A+B+C)^3 \sim A^3$ donne par développement $A^3 + A^2B + A^2C + AB^2 + AC^2 + ABC + B^3 + B^2C + BC^2 + C^3 \sim A^3$, soit $A^2B + A^2C + AB^2 + AC^2 + ABC + B^3 + B^2C + BC^2 + C^3$. La complétion conduit alors à un modèle contenant $19 = 20 - 1$ termes, à savoir tous les monomes de degrés ≤ 3 à l'exception de A^3 :

$$1 + A + B + C + A^2 + B^2 + C^2 + AB + AC + BC + A^2B + A^2C + AB^2 + AC^2 + ABC + B^3 + B^2C + BC^2 + C^3.$$

Mais l'expression $(A+B+C)^3 \sim ABC + A^2B + AB^2$ conduit après développement à $A^3 + A^2B + A^2C + AB^2 + AC^2 + ABC + B^3 + B^2C + BC^2 + C^3 \sim ABC + A^2B + AB^2$, soit $A^3 + A^2C + AC^2 + B^3 + B^2C + BC^2 + C^3$. La complétion conduit alors à un modèle contenant $16 = 20 - 4$ termes, à savoir les monomes de degrés ≤ 3 qui diffèrent des trois monomes ABC , A^2B , AB^2 supprimés, mais aussi du monome AB de degré 2. En effet dans le modèle développé ne figure aucun terme d'interaction entre A et B et la complétion ne peut donc faire apparaître le terme $A.B$.

Pour éviter ce genre de surprise, une précaution simple consiste à rajouter 1 dans la somme élevée à la puissance 3, donc à utiliser le modèle

$$(1 + A + B + C)^3 \sim ABC + A^2B + AB^2.$$

Le développement de la partie à gauche de \sim donne alors bien tous les monomes de degré ≤ 3 et après suppression de ceux à droite de \sim , on trouve bien $17 = 20 - 3$ termes.

3.7 Moyennes

Lorsqu'un effet apparaît significatif dans l'analyse de variance, le procédé le plus simple pour préciser son action est d'examiner les moyennes associées. Celle ci s'obtiennent en faisant figurer le terme en question, débarrassé de ses éventuels exposants, sur une ligne de la fenêtre des moyennes.

Ainsi dans l'exemple 3.4, si les termes VAR.DENS et VAR.DOSE² apparaissent significatifs, on rajoutera les lignes suivantes dans la fenêtre des moyennes

```
VAR DENS
VAR DOSE
```

Le calcul des moyennes prend en compte le modèle. Tout se passe comme si

- on calculait la réponse calculée pour chaque combinaison de niveaux des facteurs introduits,
- puis on faisait les moyennes appropriées de ces réponses.

En fait, l'algorithme de calcul procède différemment. Il n'utilise que les termes du modèle contenant exclusivement les facteurs pour lesquels on demande la moyenne. Ainsi pour calculer les moyennes associée au terme VAR DENS, il n'utilise que les paramètres des termes 1, VAR, DENS, VAR.DENS, et pour calculer les moyennes associée à VAR.DOSE que les termes 1, VAR, DOSE, DOSE^2, VAR.DOSE, VAR.DOSE^2. Ces moyennes peuvent donc être calculées même si certains paramètres du modèle sont confondus.

4 Reparamétrisation et définition des effets factoriels

L'article [3] montre comment reparamétriser et définir des effets factoriels dans un cas très général. Une approche simplifiée est présentée dans [2] pour le cas où l'on peut se ramener à des facteurs croisés, après avoir éventuellement décomposé en pseudofacteurs certains facteurs. C'est uniquement ce dernier cas qui est pris en compte par ANALYS dans cette version.

La reparamétrisation est définie à partir des contrastes sélectionnés pour chaque facteur. Après avoir donné des compléments sur ces choix de contrastes (§ 4.1), on indique comment s'en déduit la paramétrisation globale (§ 4.2).

4.1 Cas d'un unique facteur

Considérons le cas d'un plan comportant un unique facteur A ayant N niveaux $1, \dots, N$ et notons $\tau(1), \dots, \tau(N)$ les espérances de la réponse y à ces N niveaux. L'espérance d'une observation y faite au niveau A (on note le niveau d'un facteur avec la même lettre que ce facteur) est donc $\tau(A)$, ce que l'on peut écrire:

$$E(y | A) = \tau(A)$$

On appelle aussi $\tau(A)$ la réponse théorique du traitement A . Dans le cas equipondéré où on accorde le même poids à tous les niveaux de A , l'effet moyen est défini comme la moyenne :

$$\mu = \frac{\tau(1) + \dots + \tau(N)}{N}. \quad (5)$$

Pour comparer les réponses théoriques, on introduit des contrastes, par exemple

$$\begin{aligned} \beta_1 &= \tau(2) - \tau(1), \\ \beta_2 &= 2\tau(3) - [\tau(1) + \tau(2)], \\ \beta_3 &= 3\tau(4) - [\tau(1) + \tau(2) + \tau(3)], \\ &\dots, \\ \beta_{N-1} &= (N-1)\tau(N) - [\tau(1) + \dots + \tau(N-1)] \end{aligned} \quad (6)$$

ou bien

$$\begin{aligned}\gamma_1 &= \tau(1) - \mu, \\ \gamma_2 &= \tau(2) - \mu, \\ &\dots, \\ \gamma_{N-1} &= \tau(N-1) - \mu\end{aligned}\tag{7}$$

Chacun de ces jeux de $N - 1$ contrastes permet d'obtenir de l'information sur toutes les différences entre traitements tout en évitant des redondances dans cet examen. En terme mathématique, ils définissent avec l'effet moyen μ un ensemble de N formes linéaires indépendantes du vecteur $\tau = (\tau(1), \dots, \tau(N))'$, ensemble à partir duquel on peut reconstituer les réponses théoriques.

Considérons à nouveau le cas $N = 3$ déjà considéré dans l'exemple 3.1. Les définitions (5), (6) et (7) conduisent aux égalités matricielles

$$\begin{bmatrix} 3\mu \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 0 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} \tau(1) \\ \tau(2) \\ \tau(3) \end{bmatrix} \quad \begin{bmatrix} \mu \\ \gamma_1 \\ \gamma_2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} \tau(1) \\ \tau(2) \\ \tau(3) \end{bmatrix}\tag{8}$$

qui implique en prémultipliant chaque égalité par l'inverse de la matrice qui y figure

$$\begin{bmatrix} \tau(1) \\ \tau(2) \\ \tau(3) \end{bmatrix} = \begin{bmatrix} 1 & -1/2 & -1/6 \\ 1 & 1/2 & -1/6 \\ 1 & 0 & 1/3 \end{bmatrix} \begin{bmatrix} \mu \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad \begin{bmatrix} \tau(1) \\ \tau(2) \\ \tau(3) \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \gamma_1 \\ \gamma_2 \end{bmatrix}\tag{9}$$

Contrastes orthogonaux.

Les contrastes β_1, β_2 ont une propriété que n'ont pas les paramètres γ_1 et γ_2 . Les vecteurs de coefficients associés à μ, β_1, β_2 , lignes de la matrice figurant à gauche de (8), sont orthogonaux pour le produit scalaire usuel. Cela implique l'orthogonalité des colonnes de la matrice apparaissant à gauche de (9) d'où découle l'orthogonalité du modèle associé si les niveaux de A sont équirépétés dans le plan. Pour illustrer considérons le cas où chaque niveau de A est expérimenté une seule fois. Le modèle associé est

$$\begin{aligned}\begin{bmatrix} E(y_1) \\ E(y_2) \\ E(y_3) \end{bmatrix} &= \begin{bmatrix} 1 & -1/2 & -1/6 \\ 1 & 1/2 & -1/6 \\ 1 & 0 & 1/3 \end{bmatrix} \begin{bmatrix} \mu \\ \beta_1 \\ \beta_2 \end{bmatrix} \\ E(y) &= X \theta\end{aligned}\tag{10}$$

L'estimateur des paramètres $\hat{\theta} = (X'X)^{-1}X'y$ a pour matrice de variance $\sigma^2(X'X)^{-1}$ qui est une matrice diagonale :

$$\text{cov}(\hat{\theta}) = \frac{\sigma^2}{3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 18 \end{bmatrix}$$

Normalisation des contrastes.

Si on remplace β_1, β_2 par $\alpha_1 = \beta_1/\sqrt{6}$ et $\alpha_2 = \beta_2/\sqrt{18}$, la matrice diagonale ci-dessus devient la matrice identité. Les paramètres μ, α_1, α_2 ainsi obtenus sont dits *normés*.

L'égalité matricielle qui les définit est

$$\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ -\sqrt{3/2} & \sqrt{3/2} & 0 \\ -1/\sqrt{2} & -1/\sqrt{2} & \sqrt{2} \end{bmatrix} \begin{bmatrix} \tau(1) \\ \tau(2) \\ \tau(3) \end{bmatrix}. \quad (11)$$

On peut aussi écrire cette définition sous la forme

$$\mu = A'_0\tau/3, \quad \alpha_1 = A'_1\tau/3, \quad \alpha_2 = A'_2\tau/3, \quad (12)$$

où A_0, A_1, A_2, τ sont les vecteurs colonnes suivants

$$\begin{array}{c} A \\ 1 \\ 2 \\ 3 \end{array} \quad \begin{array}{c} A_0 \\ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \end{array} \quad \begin{array}{c} A_1 \\ \begin{bmatrix} -\sqrt{3/2} \\ \sqrt{3/2} \\ 0 \end{bmatrix} \end{array} \quad \begin{array}{c} A_2 \\ \begin{bmatrix} -1/\sqrt{2} \\ -1/\sqrt{2} \\ \sqrt{2} \end{bmatrix} \end{array} \quad \begin{array}{c} \tau \\ \begin{bmatrix} \tau(1) \\ \tau(2) \\ \tau(3) \end{bmatrix} \end{array} \quad (13)$$

Les vecteurs de coefficients A_0, A_1, A_2 étant orthonormés, l'égalité matricielle donnant le vecteur τ des réponses théoriques à partir des paramètres s'obtient immédiatement en transposant la matrice figurant dans (11):

$$\begin{bmatrix} \tau(1) \\ \tau(2) \\ \tau(3) \end{bmatrix} = \begin{bmatrix} 1 & -\sqrt{3/2} & -1/\sqrt{2} \\ 1 & \sqrt{3/2} & -1/\sqrt{2} \\ 1 & 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix}. \quad (14)$$

De façon équivalente, on a en notant $A_i(A)$ la coordonnée d'indice A du vecteur A_i

$$\tau(A) = A_0(A)\mu + A_1(A)\alpha_1 + A_2(A)\alpha_2 \quad \text{pour } A = 1,2,3. \quad (15)$$

Le coefficient à appliquer à chaque paramètre est la coordonnée d'indice A du vecteur A_i associé à ce paramètre. Sous forme vectorielle, cela s'écrit

$$\begin{array}{c} \begin{bmatrix} \tau(1) - \mu \\ \tau(2) - \mu \\ \tau(3) - \mu \end{bmatrix} \\ \tau - \mu \mathbf{1} \end{array} = \alpha_1 \begin{array}{c} \begin{bmatrix} -\sqrt{3/2} \\ \sqrt{3/2} \\ 0 \end{bmatrix} \\ A_1 \end{array} + \alpha_2 \begin{array}{c} \begin{bmatrix} -1/\sqrt{2} \\ -1/\sqrt{2} \\ \sqrt{2} \end{bmatrix} \\ A_2 \end{array} \quad (16)$$

La variation de la réponse théorique τ sur les niveaux de A peut être mesurée par

$$\|\tau - \mu \mathbf{1}\|^2 = \frac{1}{3} [(\tau(1) - \mu)^2 + (\tau(2) - \mu)^2 + (\tau(3) - \mu)^2].$$

Compte tenu de l'orthonormalité des deux colonnes A_1, A_2 associées à α_1 et α_2 , elle se décompose de la façon suivante:

$$\|\tau - \mu \mathbf{1}\|^2 = \alpha_1^2 \|A_1\|^2 + \alpha_2^2 \|A_2\|^2 = \alpha_1^2 + \alpha_2^2.$$

Cette égalité montre que les paramètres normés α_1, α_2 sont directement comparables puisque α_1^2, α_2^2 mesurent des quantités similaires, savoir les parts de variation de la réponse théorique induits par chacun de ces deux paramètres.

La normation permet donc d'avoir des paramètres comparables entre eux. Cet avantage devient très important quand il y a plusieurs facteurs, donc de nombreux paramètres, et

nous verrons que les paramètres normés introduits ici sont la pierre de construction des paramètres normés dans le cas plus général.

Dans les plans où les niveaux de A sont équirépétés, α_1 et α_2 sont estimés avec un même écart-type. Considérons par exemple à nouveau le cas où chaque niveau de A est expérimenté une seule fois. Avec ces nouveaux paramètres normés, (10) devient

$$\begin{aligned} \begin{bmatrix} E(y_1) \\ E(y_2) \\ E(y_3) \end{bmatrix} &= \begin{bmatrix} 1 & -\sqrt{3/2} & -1/\sqrt{2} \\ 1 & \sqrt{3/2} & -1/\sqrt{2} \\ 1 & 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} \\ E(y) &= X \eta \end{aligned} \quad (17)$$

et l'estimateur du vecteur des paramètres normés $\hat{\eta} = (X'X)^{-1}X'y$ a pour matrice de variance

$$\text{cov}(\hat{\eta}) = \frac{\sigma^2}{3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

En principe, dès lors qu'il y a une certaine symétrie entre les niveaux de A , on est amené à choisir des plans où les nombres de répétitions sont identiques ou très voisins. Les paramètres normés ont donc alors des écart-types égaux ou très voisins, et cela facilite beaucoup la comparaison entre ces paramètres.

4.2 Cas de deux facteurs

Pour simplifier, supposons qu'on étudie deux facteurs A à 3 niveaux 1, 2, 3 et B à 2 niveaux 1, 2. Ce cas suffit pour bien comprendre le mécanisme définissant les paramètres dans le cas général. Le tableau 5 donne à gauche le jeu de vecteur A_0, A_1, A_2 précédemment utilisé dans (12) et (13) pour définir les paramètres μ, α_1, α_2 , à droite un jeu de vecteurs B_0, B_1 similaire pour le facteur B . On définit à partir de là les $6 = 3 \times 2$

A	A_0	A_1	A_2		B	B_0	B_1
1	$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} -\sqrt{3/2} \\ \sqrt{3/2} \\ 0 \end{bmatrix}$	$\begin{bmatrix} -1/\sqrt{2} \\ -1/\sqrt{2} \\ \sqrt{2} \end{bmatrix}$		1	$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} -1 \\ 1 \end{bmatrix}$
2					2		
3							

TAB. 5 – *Contrastes retenus pour les deux facteurs*

ensembles de coefficients à appliquer aux réponses théoriques $\tau(A,B)$, à un coefficient $1/6$ prêt, pour obtenir les paramètres utiles dans ce cas. Ces six ensembles sont donnés au tableau 6 sous forme de matrices 3×2 , avec en regard la matrice 3×2 des réponses correspondantes. La matrice associée à A_i, B_j n'est autre que le produit $A_i B_j'$ qu'on obtient en faisant la table de multiplication du vecteur colonne A_i par le vecteur ligne B_j' . L'effet correspondant est noté $e(A_i B_j)$. Le vecteur de coefficients correspondant est noté $A_i \otimes B_j$ et appelé le produit tensoriel des vecteurs A_i et B_j . La coordonnée d'indice (A,B) de ce vecteur est donc par définition le produit de la coordonnée d'indice A de A_i par la coordonnée d'indice B de B_j :

$$A_i \otimes B_j(A,B) = A_i(A)B_j(B) \quad (18)$$

Réponses théoriques

$\tau(A,B)$

$$\begin{bmatrix} \tau(1,1) & \tau(1,2) \\ \tau(2,1) & \tau(2,2) \\ \tau(3,1) & \tau(3,2) \end{bmatrix}$$

$$\begin{array}{ccc} \begin{array}{c} B'_0 \\ [1 \ 1] \\ A_0 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \\ A_0 B'_0 \end{array} & A_1 \begin{bmatrix} -\sqrt{3/2} \\ \sqrt{3/2} \\ 0 \end{bmatrix} \begin{bmatrix} -\sqrt{3/2} & -\sqrt{3/2} \\ \sqrt{3/2} & \sqrt{3/2} \\ 0 & 0 \end{bmatrix} & \begin{array}{c} B'_0 \\ [1 \ 1] \\ A_2 \begin{bmatrix} -1/\sqrt{2} \\ -1/\sqrt{2} \\ \sqrt{2} \end{bmatrix} \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{2} \\ -1/\sqrt{2} & -1/\sqrt{2} \\ \sqrt{2} & \sqrt{2} \end{bmatrix} \\ A_2 B'_0 \end{array} \\ \\ \begin{array}{c} B'_1 \\ [-1 \ 1] \\ A_0 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ -1 & 1 \\ -1 & 1 \end{bmatrix} \\ A_0 B'_1 \end{array} & A_1 \begin{bmatrix} -\sqrt{3/2} \\ \sqrt{3/2} \\ 0 \end{bmatrix} \begin{bmatrix} \sqrt{3/2} & -\sqrt{3/2} \\ -\sqrt{3/2} & \sqrt{3/2} \\ 0 & 0 \end{bmatrix} & \begin{array}{c} B'_1 \\ [-1 \ 1] \\ A_2 \begin{bmatrix} -1/\sqrt{2} \\ -1/\sqrt{2} \\ \sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \\ -\sqrt{2} & \sqrt{2} \end{bmatrix} \\ A_2 B'_1 \end{array} \end{array}$$

TAB. 6 – Tableaux 3×2 donnant les coefficients des contrastes

La formule définissant $e(A_i B_j)$ est donc

$$e(A_i B_j) = \frac{1}{6} \sum_{A,B} A_i(A) B_j(B) \tau(A,B) = \frac{1}{6} (A_i \otimes B_j)' \tau \quad (19)$$

Les vecteurs A_i sont ici orthonormés pour le produit scalaire pondéré par des poids égaux $p(i) = 1/3$, et de même les vecteurs B_j sont orthonormés pour le produit scalaire pondéré par des poids $q(j) = 1/2$:

$$\begin{aligned} \langle A_i, A_k \rangle &= \frac{1}{3} (A_i(1)A_k(1) + A_i(2)A_k(2) + A_i(3)A_k(3)) &= 0 & \text{si } i \neq k \\ & &= 1 & \text{si } i = k \\ \langle B_j, B_l \rangle &= \frac{1}{2} (B_j(1)B_l(1) + B_j(2)B_l(2)) &= 0 & \text{si } j \neq l \\ & &= 1 & \text{si } j = l \end{aligned}$$

On en déduit, en utilisant (18) que les produits tensoriels $A_i \otimes B_j$ sont orthonormés pour le produit scalaire pondéré par les poids produits $p(i)q(j)$ qui sont dans ce cas tous égaux à $1/6$:

$$\begin{aligned} \langle A_i \otimes B_j, A_k \otimes B_l \rangle &= \frac{1}{6} \sum_{A,B} A_i(A) B_j(B) A_k(A) B_l(B) &= 0 & \text{si } i \neq k \text{ ou } j \neq l \\ & &= 1 & \text{si } i = k \text{ et } j = l. \end{aligned}$$

Il est alors facile d'inverser la relation (19). On obtient la relation suivante:

$$\tau(A,B) = \sum_{i,j} A_i(A) B_j(B) e(A_i B_j) \quad (20)$$

qui montre que, pour obtenir la réponse théorique aux niveaux A et B des facteurs A et B , le coefficient à appliquer au paramètre $e(A_i B_j)$ est le produit des coordonnées d'indice A et B des vecteurs A_i, B_j associés.

Les égalités (20) s'écrivent aussi sous la forme vectorielle suivante

$$\tau = \sum_{i,j} A_i \otimes B_j e(A_i B_j) . \quad (21)$$

Le vecteur $A_0 \otimes B_0$ a toutes ses coordonnées égales à 1, i.e. $A_0 \otimes B_0 = \mathbf{1}$, et le paramètre associé $\mu = e(A_0 B_0)$ est la moyenne des 6 réponses théoriques. On peut donc écrire l'égalité qui précède sous la forme

$$\tau - \mu \mathbf{1} = \sum_{i,j | (i,j) \neq (0,0)} A_i \otimes B_j e(A_i B_j) . \quad (22)$$

De même que dans le cas d'un facteur, la variation de la réponse théorique τ peut être mesurée par $\|\tau - \mu \mathbf{1}\|^2$. Compte tenu de l'orthonormalité des vecteurs $A_i \otimes B_j$, elle se décompose sous la forme

$$\|\tau - \mu \mathbf{1}\|^2 = \sum_{i,j | (i,j) \neq (0,0)} e^2(A_i B_j) .$$

Cette égalité montre que les paramètres $e(A_i B_j)$ sont directement comparables puisque leurs carrés mesurent les parts de variation de la réponse théorique qu'ils induisent.

Usuellement on ne fait pas apparaître les vecteurs indicés par 0 dans la notation $e(A_i B_j)$. Ainsi on simplifie $e(A_1 B_0), e(A_2 B_0)$ en $e(A_1), e(A_2)$ et $e(A_0 B_1)$ en $e(B_1)$. Un autre allègement des notations consiste à supprimer l'indice 1 en écrivant $e(A), e(B), e(AB), e(A_2 B)$ à la place de $e(A_1), e(B_1), e(A_1 B_1), e(A_2 B_1)$. Enfin dans le programme, la notation $e()$ est omise et les indices sont précédés d'un signe \wedge : $A^\wedge 1$ remplace A_1 , que A soit quantitatif ou qualitatif.

Terminologie et interprétation des effets factoriels.

Les paramètres $e(A_i B_j)$ sont désignés sous le nom d'*effets factoriels*. On les classe en effets principaux de A si seul l'indice i est non nul, effet principal de B si seul l'indice j est non nul, enfin effets d'interaction $A.B$ si i et j sont tous deux non nuls.

L'effet principal $e(A)$ est par définition

$$e(A) = \sqrt{6} \left(\frac{\tau(2,1) + \tau(2,2)}{2} - \frac{\tau(1,1) + \tau(1,2)}{2} \right) .$$

Si on indique que l'on fait la moyenne sur un indice en remplaçant cet indice par un point, il s'écrit à une constante prêt:

$$e(A) \propto \tau(2, \cdot) - \tau(1, \cdot)$$

et de même

$$e(A_2) \propto \tau(3, \cdot) - \frac{\tau(1, \cdot) + \tau(2, \cdot)}{2} .$$

Ces effets principaux $e(A)$, $e(A_2)$ du facteur A s'obtiennent donc de la même façon que dans le cas où il n'y a qu'un seul facteur A , en remplaçant chaque réponse théorique $\tau(A)$ du cas un facteur par la moyenne correspondante $\tau(A, \cdot)$.

Les effets d'interaction sont les effets $e(A_i B_j)$ où $i \neq 0$, $j \neq 0$. Par définition

$$e(AB) = e(A_1 B_1) \propto [\tau(2,2) - \tau(2,1)] - [\tau(1,2) - \tau(1,1)]$$

et

$$e(A_2 B) = e(A_2 B_1) \propto \left[\tau(3,2) - \frac{\tau(2,2) + \tau(1,2)}{2} \right] - \left[\tau(3,1) - \frac{\tau(2,1) + \tau(1,1)}{2} \right].$$

Le paramètre $e(AB)$ s'interprète donc comme l'interaction avec B du contraste opposant les niveaux 2 et 1 de A , tandis que $e(A_2 B)$ est l'interaction avec B du contraste opposant le niveau 3 aux deux autres.

Ces contrastes peuvent servir directement à l'interprétation. Mais le plus souvent ils sont un intermédiaire technique commode pour les calculs. En pratique on teste globalement les hypothèses $\{e(AB) = 0, e(A_2 B) = 0\}$ d'absence d'interaction et $\{e(A) = 0, e(A_2) = 0\}$ d'absence d'effet principal. Si l'une de ces hypothèses est rejetée, on peut interpréter ensuite en examinant directement les estimations des 6 réponses théoriques $\tau(1,1), \dots, \tau(3,2)$ s'il y a interaction, des 3 moyennes $\tau(1, \cdot), \tau(2, \cdot), \tau(3, \cdot)$ s'il y a un effet principal. On peut aussi examiner des contrastes appartenant soit à l'interaction, soit à l'effet principal. Mais il n'y a alors pas de raison de se limiter aux seuls deux contrastes générateurs $e(AB), e(A_2 B)$ pour l'interaction ou $e(A), e(A_2)$ pour l'effet principal.

Cas de facteurs quantitatifs.

Supposons maintenant que A est quantitatif à 4 niveaux 5, 6, 7, 8 comme dans l'exemple 3.3. Si on définit les contrastes comme dans cet exemple, les effets $e(\text{lin } A)$, $e(\text{quad } A)$, $e(\text{lin } A.B)$, $e(\text{quad } A.B)$ se définissent comme précédemment à partir des contrastes donnés pour A en (4) et pour B à droite du tableau 5. Le tableau 7 récapitule les éléments à partir desquels est défini l'effet $e(\text{lin } A.B)$ d'interaction entre l'effet linéaire de A et B .

$$\begin{array}{ccc}
 \tau(A,B) & & B'_1 \\
 & & \begin{bmatrix} -1 & 1 \end{bmatrix} \\
 \begin{bmatrix} \tau(5,1) & \tau(5,2) \\ \tau(6,1) & \tau(6,2) \\ \tau(7,1) & \tau(7,2) \\ \tau(8,1) & \tau(8,2) \end{bmatrix} & \text{lin } A & \begin{bmatrix} -1.3416 \\ -0.4472 \\ 0.4472 \\ 1.3416 \end{bmatrix} \begin{bmatrix} 1.3416 & -1.3416 \\ 0.4472 & -0.4472 \\ -0.4472 & 0.4472 \\ -1.3416 & 1.3416 \end{bmatrix} \\
 \text{Réponses théoriques} & & \text{lin } A B'_1 \\
 & & \text{coefficients associés}
 \end{array}$$

TAB. 7 – Définition d'un contraste $e(\text{lin } A.B)$.

A la normation prêt, les coefficients du contraste $\text{lin } A$ sont $-3, -1, 1, 3$. En effet avec l'équipondération, le carré de la norme de $(-3, -1, 1, 3)$ est

$$\frac{1}{4}[(-3)^2 + (-1)^2 + 1^2 + 3^2] = 5$$

On norme ce vecteur en le divisant par $\sqrt{5}$, ce qui donne les coefficients figurant dans (4) et le tableau 7. Au coefficient de normalisation prêt, on a donc:

$$e(\text{lin } A.B) = [-3\tau(5,2) - \tau(6,2) + \tau(7,2) + 3\tau(8,2)] - [-3\tau(5,1) - \tau(6,1) + \tau(7,1) + 3\tau(8,1)] . \quad (23)$$

Le programme utilise la même notation que dans le cas qualitatif pour désigner les effets factoriels. Ainsi $A.B$ désigne l'effet $e(\text{lin } A.B)$, $A_2.B$ l'effet $e(\text{quad } A.B)$. Dans cette notation, l'indice fait référence au degré du polynome orthogonal, donc à la colonne correspondante dans (4).

Cas pondéré

Ce qui a été fait dans le cas de poids égaux se généralise immédiatement au cas où les poids sont inégaux

4.3 Obtention des moyennes

La moyenne générale est par définition le paramètre $e(A_0B_0)$, aussi noté $e(1)$. Elle est égale à la moyenne, pondérée si on a introduit des poids non égaux, des réponses théoriques $\tau(A,B)$:

$$e(1) = \sum_{A,B} p_A q_B \tau(A,B).$$

Les moyennes par niveau de A sont les quantités

$$\tau(A, \cdot) = \sum_B q_B \tau(A,B) .$$

Pour les calculer, on peut partir de la formule (20) et effectuer la moyenne sur B de chacun des termes:

$$\begin{aligned} \sum_B q_B A_i(A) B_j(B) e(A_i B_j) &= A_i(A) e(A_i B_j) \sum_B q_B B_j(B) \\ &= A_i(A) e(A_i B_j) \langle B_j, 1 \rangle \\ &= \begin{cases} A_i(A) e(A_i) & \text{si } j = 0 \\ 0 & \text{si } j \neq 0 \end{cases} \end{aligned}$$

La moyenne du terme où intervient le paramètre $e(A_i B_j)$ est donc nulle si $j > 0$, égale à $A_i(A) e(A_i)$ si $j = 0$. Par conséquent, la moyenne recherchée s'écrit

$$\tau(A, \cdot) = \sum_i A_i(A) e(A_i) \quad (24)$$

et ne fait intervenir dans son calcul que la moyenne générale $e(A_0) = e(1)$ et les paramètres appartenant à l'effet factoriel de A . On a un résultat semblable pour les moyennes par niveau de B :

$$\tau(\cdot, B) = \sum_j B_j(B) e(B_j) . \quad (25)$$

Ce type de formule se généralise immédiatement à des cas plus généraux. Par exemple, si il y a trois facteurs A , B , C , les moyennes par niveau de A sont de la forme

$$\tau(A, \dots) = \sum_i A_i(A) e(A_i)$$

et les moyennes par niveau du couple (A, B) sont

$$\tau(A, B, \dots) = \sum_{i,j} A_i(A) B_j(B) e(A_i B_j) . \quad (26)$$

Ces dernières font intervenir la moyenne générale $e(1)$, les paramètres $e(A_i), i \neq 0$ appartenant à l'effet factoriel de A , les paramètres $e(B_j), j \neq 0$ de l'effet factoriel de B , enfin les paramètres $e(A_i B_j), i \neq 0, j \neq 0$ de l'effet factoriel d'interaction entre A et B .

Les moyennes estimées s'obtiennent en remplaçant dans ces formules les paramètres $e(A_i B_j)$ par leurs estimations $\hat{e}(A_i B_j)$. On voit d'après ces formules qu'il n'est pas indispensable d'estimer tous les paramètres du modèle pour estimer les moyennes. Par exemple, les moyennes $\tau(A, \dots)$ par niveau de A peuvent être estimées dès lors que tous les paramètres de l'effet factoriel de A et la constante peuvent être estimés, même si les paramètres $e(A_i B_j), i \neq 0, j \neq 0$ de l'interaction entre A et B ne sont pas estimables. Cette possibilité d'estimer les moyennes par niveau d'un facteur même lorsque certaines interactions contenant ce facteur sont non estimables est constamment utilisée pour analyser des dispositifs de résolution 4, dans lesquels certaines interactions entre deux facteurs sont confondues, donc inestimables, alors que tous les effets principaux sont eux estimables.

5 Etude d'un plan

Pour étudier les propriétés d'un plan, confusions d'effets et efficacités factorielles d'estimation, on coche la case *Etude du plan seulement* sur l'écran de la figure 1. Ce paragraphe illustre les sorties obtenues dans quelques cas caractéristiques.

Exemple 5.1 . Le tableau 8 donne un plan obtenu par PLANOR. Il comporte 3 facteurs : A , B quantitatifs à quatre niveaux, C qualitatif à 2 niveaux. Les 32 traitements, expérimentés chacun une fois, sont répartis en 4 blocs. Le facteur bloc noté BL est déterminé par les pseudofacteurs $BL_1 = A_2 B_1$, $BL_2 = A_1 B_2 C$. Bien qu'ils soient inutiles pour analyser, on a fait figurer également ces pseudofacteurs dans le tableau. Par ailleurs, on a aussi fait figurer les valeurs de deux variables observées, Y_1 , Y_2 sur l'analyse desquelles on reviendra au paragraphe 6. Les effets pseudofactoriels confondus avec l'effet bloc sont donc $A_2 B_1$, $A_1 B_2 C$, $A_1 A_2 B_1 B_2 C$. En utilisant les relations entre ces effets pseudofactoriels et les effets polynomiaux [2], on peut facilement montrer que ces confusions ne génèrent qu'une faible perte d'efficacité si on retient le modèle de degré 3 qui s'écrit $P^{\wedge 3} + BL$ où P est la somme $1 + A + B + C$.

Le plan permet d'estimer aussi tous les termes du modèle de degré 4, mais avec une efficacité moindre pour certains effets polynomiaux.

Les tableaux 9 et 10 donnent les sorties d'ANALYS pour ces modèles de degré 3 et 4.

En fait, il n'est pas possible de faire figurer les termes en A^4 et B^4 ($A^{\wedge 4}, B^{\wedge 4}$) dans le modèle de degré 4 puisque ces facteurs n'ont que 3 niveaux. C'est pourquoi on a écrit le

n0	A ₁	A ₂	B ₁	B ₂	BL ₁	BL ₂	A	B	C	BL	#Y1	Y2
1	0	0	0	0	0	0	5	20	faible	0	26.36125	0.7035416667
2	1	0	0	0	0	1	7	20	faible	1	30.44386905	5.240327381
3	0	1	0	0	1	0	6	20	faible	2	22.84363095	5.610922619
4	1	1	0	0	1	1	8	20	faible	3	36.05125	12.77270833
5	0	0	1	0	1	0	5	24	faible	2	7.404345238	-1.321220238
6	1	0	1	0	1	1	7	24	faible	3	16.39339286	11.47544643
7	0	1	1	0	0	0	6	24	faible	0	14.26910714	1.815803571
8	1	1	1	0	0	1	8	24	faible	1	17.03315476	11.01247024
9	0	0	0	1	0	1	5	22	faible	1	9.703154762	1.132470238
10	1	0	0	1	0	0	7	22	faible	0	24.42910714	4.620803571
11	0	1	0	1	1	1	6	22	faible	3	17.53339286	8.775446429
12	1	1	0	1	1	0	8	22	faible	2	25.23434524	12.48877976
13	0	0	1	1	1	1	5	26	faible	3	0.36125	-0.5122916667
14	1	0	1	1	1	0	7	26	faible	2	8.223630952	8.670922619
15	0	1	1	1	0	1	6	26	faible	1	-2.136130952	3.790327381
16	1	1	1	1	0	0	8	26	faible	0	9.85125	13.72354167
17	0	0	0	0	0	1	5	20	fort	1	11.10458333	-4.173958333
18	1	0	0	0	0	0	7	20	fort	0	22.58696429	9.249613095
19	0	1	0	0	1	1	6	20	fort	3	23.75053571	3.849136905
20	1	1	0	0	1	0	8	20	fort	2	24.85791667	16.74770833
21	0	0	1	0	1	1	5	24	fort	3	13.55482143	-3.642291667
22	1	0	1	0	1	0	7	24	fort	2	11.38410714	13.83758929
23	0	1	1	0	0	1	6	24	fort	1	9.328392857	-2.528839286
24	1	1	1	0	0	0	8	24	fort	0	20.63267857	19.35104167
25	0	0	0	1	0	0	5	22	fort	0	12.27267857	-7.023958333
26	1	0	0	1	0	1	7	22	fort	1	19.09839286	10.25116071
27	0	1	0	1	1	0	6	22	fort	2	16.81410714	3.127589286
28	1	1	0	1	1	1	8	22	fort	3	28.91482143	16.62770833
29	0	0	1	1	1	0	5	26	fort	2	3.237916667	-7.162291667
30	1	0	1	1	1	1	7	26	fort	3	15.44053571	10.6541369
31	0	1	1	1	0	0	6	26	fort	0	13.59696429	1.559613095
32	1	1	1	1	0	1	8	26	fort	1	17.42458333	19.27604167

TAB. 8 – Analyse d'un $4 \times 4 \times 2$ en 4 blocs

modèle sous la forme $P^4 + BL \sim A^4 + B^4$ où P est la partie de modèle $1 + A + B + C$. Après développement, le terme P^4 donne les termes $A^a B^b$ pour lesquels $a + b \leq 4$ et les termes $A^a B^b C$ pour lesquels $a + b \leq 3$. Si on enlève A^4 et B^4 , il ne reste plus de terme ayant un exposant a ou b égal à 4.

Pour que ces modèles soient correctement interprétés, il ne faut naturellement pas oublier de déclarer A et B comme des facteurs quantitatifs avant d'analyser.

Définition des efficacités factorielles

L'efficacité compare la variance d'estimation donnée par le plan étudiée à celle que l'on obtiendrait avec un plan de référence, qui est le plan factoriel complet ou un plan en blocs complets. Pour rendre la comparaison possible, on multiplie la variance par le nombre d'unités pour se ramener à des variances *par unité*.

Si le modèle linéaire est $E(y) = X\beta$, la matrice de variance de $\hat{\beta}$ est $\sigma^2(X'X)^{-1}$ dans le cas où tous les paramètres sont estimables. La matrice de variance par unité est donc $\sigma^2 N(X'X)^{-1}$ s'il y a N unités.

La matrice correspondante pour le plan de référence est $\sigma^2 N_0(X'_0 X_0)^{-1}$. Si β est le vecteur des paramètres orthonormés introduits au paragraphe 4, on a $X'_0 X_0 = N_0 \mathbf{I}$ et cette matrice est donc $\sigma^2 \mathbf{I}$.

On est donc amené à comparer $\sigma^2 N(X'X)^{-1}$ à $\sigma^2 \mathbf{I}$ ou, en simplifiant par σ^2 , les matrices $N(X'X)^{-1}$ et \mathbf{I} . Lorsqu'on s'intéresse au sous-ensemble de β correspondant à un effet factoriel donné, on compare les sous-blocs diagonaux associés de ces deux matrices. Le premier V est donc le produit par N/σ^2 de la matrice de variance des paramètres considérés. Le second est bien sur la matrice identité de même taille que V .

Si il n'y a qu'un degré de liberté, V est un scalaire donnant à une constante prêt la variance du paramètre considéré. L'efficacité est alors définie comme le ratio $1/V$. Elle est d'autant plus grande que la variance d'estimation V est plus petite, donc la précision meilleure.

La généralisation au cas où il y a $m > 1$ degrés de liberté pour l'effet factoriel étudié s'opère en introduisant des *efficacités principales* qui sont les valeurs propres de $\mathbf{I}_m V^{-1}$, ou encore les inverses des valeurs propres de V . La plus faible de ces efficacités correspond au contraste le plus mal estimé pour cet effet factoriel.

L'efficacité moyenne d'estimation de cet effet factoriel peut être résumée par des critères tels que :

$$\begin{aligned} \text{tr} &= \frac{1}{\text{trace}(V)/m} = \frac{1}{(v_1 + \dots + v_m)/m}, \\ \text{dét} &= \frac{1}{\|V\|^{1/m}} = \frac{1}{(v_1 \dots v_m)^{1/m}}, \end{aligned}$$

en notant v_1, \dots, v_m les valeurs propres de V .

Lorsque le modèle ne comporte que des effets qualitatifs et qu'on utilise l'équipondération pour définir les contrastes relatifs à chaque facteur, on montre que les critères tr et $dét$ sont inférieurs à 1. Cela n'est pas nécessairement vrai s'il n'y a pas d'équipondération, ou qu'il y a des facteurs quantitatifs auxquels correspondent dans le modèle reparamétré des effets

polynomiaux. D'autre part, certaines des efficacités principales $1/v_1, \dots, 1/v_m$ peuvent être supérieures à 1 même si les critères *tr* et *dét* pour un effet factoriel sont inférieurs à 1

Outre les efficacités factorielles, qui donnent des indications très précises sur le plan, les sorties d'ANALYS comprennent aussi les valeurs propres de $X'X/N$, inverses des valeurs propres de $N(X'X)^{-1}$, dont on déduit des critères trace et déterminant globaux. Ces caractéristiques sont des équivalents globaux des caractéristiques introduites pour chaque effet factoriel.

Cas où le vecteur des paramètres n'est pas estimable.

Lorsque la matrice X du modèle linéaire n'est pas de plein rang, on ne peut pas inverser $X'X$ pour obtenir la matrice de variance de $\hat{\beta}$. Dans ce cas, le programme procède comme suit.

La matrice X est partitionnée en (X_0, X_1) où X_1 est formée des colonnes associées aux paramètres de l'effet factoriel étudié. On calcule alors

$$M = X_1'X_1 - X_1'X_0(X_0'X_0)^-X_0'X_1$$

où $(X_0'X_0)^-$ est une inverse généralisée de $X_0'X_0$. Les valeurs propres $\lambda_1, \dots, \lambda_m$ de M donnent les efficacités principales et on en déduit les critères trace et déterminant par les formules

$$\begin{aligned} \text{tr} &= \frac{m}{\frac{1}{\lambda_1} + \dots + \frac{1}{\lambda_m}}, \\ \text{dét} &= (\lambda_1 \dots \lambda_m)^{1/m} \end{aligned}$$

. Si l'une des valeurs propres λ_i est nulle, les paramètres de l'effet factoriel étudié ne sont pas tous estimables et les critères *tr* et *dét* sont nuls.

Le calcul des efficacités dans ce cas de figure est beaucoup plus long, puisqu'au lieu d'inverser la seule matrice $X'X$, on doit trouver des inverses généralisées des matrices $X_0'X_0$ associées à chaque effet factoriel.

6 Analyse de variance et moyennes

L'exemple qui suit reprend le plan et les données du tableau 8. En fait il s'agit de données simulées comme suit.

$$y_1 = 4(A - 6.5) + \gamma_C(B - 23) + \delta_{BL} + \varepsilon_1$$

où la pente γ_C de la droite reliant la réponse à B vaut $\gamma_0 = -4$ si C prend le niveau *faible* numéroté 0, vaut $\gamma_1 = -1.5$ si C prend le niveau *fort* numéroté 1 et le vecteur des effets blocs est $(\delta_0, \delta_1, \delta_2, \delta_3) = (18, 14, 15, 19)$. Le vecteur des erreurs ε_1 est un 32 échantillon d'une normale centrée d'écart-type 4, orthogonalisé à la matrice X du modèle de degré 3 utilisé pour l'analyse. De même

$$y_2 = 0.5(B - 23)(A - 6.5) + \gamma_C(A - 6.5) + \delta_{BL} + \varepsilon_2$$

Paramètres du premier écran de saisie

Nom d'analyse : EX442B
 Librairie standard : D:/planorw/exemples/
 Plan : D:/planorw/exemples/ex442b.dat
 Variables : D:/planorw/exemples/ex442b.dat
 Etude du plan seulement : oui

facteurs et covariables :

libellé	nat	type	niveaux
			...
A	num	cov	
B	num	cov	
C	alpha	fac	faible fort
BL	num	fac	

Parties de modèle : P:1+A+B+C
 Modèle : P³+BL
 Sur-Modèle :

----- ANALYSE -----
 Calcul de la matrice X associée au modèle

coeff. pol. orth. pour A . Chaque colonne donne les
 coeff. à appliquer aux puissances successives de A-6.5.
 Le libellé ⁰, ¹, ², ... à gauche précise cette
 puissance. Le libellé ⁰, ¹, ², ... en haut de col.
 donne le degré du polynome orthogonal. ... idem pour B-23 .

	⁰	¹	²	³		⁰	¹	²	³
⁰	1	0.00000	-1.25	0.00000	⁰	1	0.00000	-1.25	0.00000
¹	0	0.89443	0.00	-3.05596	¹	0	0.44721	0.00	-1.52798
²	0	0.00000	1.00	0.00000	²	0	0.00000	0.25	0.00000
³	0	0.00000	0.00	1.49071	³	0	0.00000	0.00	0.18634

contrastes pour facteur C

niv	c0	c1
faible	1	-1
fort	1	1

contrastes pour facteur BL

niv	c0	c1	c2	c3
0	1	-1.41421	-0.81650	-0.57735
1	1	1.41421	-0.81650	-0.57735
2	1	0.00000	1.63299	-0.57735
3	1	0.00000	0.00000	1.73205

Mesures d'efficacité par effet

EFFET -- ddl = 1. Critères tr : 1, dét : 1
 EFFET A -- ddl = 1. Critères tr : 1, dét : 1
 EFFET B -- ddl = 1. Critères tr : 1, dét : 1
 EFFET C -- ddl = 1. Critères tr : 1, dét : 1
 EFFET A² -- ddl = 1. Critères tr : 1, dét : 1
 EFFET A.B -- ddl = 1. Critères tr : 0.84, dét : 0.84
 EFFET A.C -- ddl = 1. Critères tr : 1, dét : 1
 EFFET B² -- ddl = 1. Critères tr : 1, dét : 1
 EFFET B.C -- ddl = 1. Critères tr : 1, dét : 1
 EFFET A³ -- ddl = 1. Critères tr : 1, dét : 1
 EFFET A².B -- ddl = 1. Critères tr : 1, dét : 1
 EFFET A².C -- ddl = 1. Critères tr : 1, dét : 1
 EFFET A.B² -- ddl = 1. Critères tr : 1, dét : 1
 EFFET A.B.C -- ddl = 1. Critères tr : 0.84, dét : 0.84
 EFFET B³ -- ddl = 1. Critères tr : 1, dét : 1
 EFFET B².C -- ddl = 1. Critères tr : 1, dét : 1
 EFFET BL -- ddl = 3. Critères tr : 0.8873239437, dét : 0.8902654597
 effic. principales : 0.840 1.000 0.840

Mesure globales d'efficacité du plan

trace	det	valmin (globaux)
0.961	0.982	0.600

Valeurs propres de la matrice d'information par unité (X'X/N)
 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000
 1.000 1.000 1.000 1.000 0.600 0.600 1.000 1.400 1.400

TAB. 9 - Exemple 8. Efficacités avec modèle de degré 3

Parties de modèle : P:1+A+B+C
 Modèle : P⁴+BL⁴A⁴+B⁴
 Sur-Modèle :

----- ANALYSE -----

Calcul de la matrice X associée au modèle

coeff. pol. orth. pour A ... coeffs à
 appliquer aux puissances de A-6.5. ... coeff. pol. orth. pour B ... coeff. à
 appliquer aux puissances de B-23. ...

	⁰	¹	²	³		⁰	¹	²	³
⁰	1	0.00000	-1.25	0.00000	⁰	1	0.00000	-1.25	0.00000
¹	0	0.89443	0.00	-3.05596	¹	0	0.44721	0.00	-1.52798
²	0	0.00000	1.00	0.00000	²	0	0.00000	0.25	0.00000
³	0	0.00000	0.00	1.49071	³	0	0.00000	0.00	0.18634

contrastes pour facteur C

niv	c0	c1
faible	1	-1
fort	1	1

contrastes pour facteur BL

niv	c0	c1	c2	c3
0	1	-1.41421	-0.81650	-0.57735
1	1	1.41421	-0.81650	-0.57735
2	1	0.00000	1.63299	-0.57735
3	1	0.00000	0.00000	1.73205

Mesures d'efficacité par effet

EFFET -- ddl = 1. Critères tr : 1, dét : 1
 EFFET A -- ddl = 1. Critères tr : 1, dét : 1
 EFFET B -- ddl = 1. Critères tr : 1, dét : 1
 EFFET C -- ddl = 1. Critères tr : 1, dét : 1
 EFFET A² -- ddl = 1. Critères tr : 1, dét : 1
 EFFET A.B -- ddl = 1. Critères tr : 0.5, dét : 0.5
 EFFET A.C -- ddl = 1. Critères tr : 1, dét : 1
 EFFET B² -- ddl = 1. Critères tr : 1, dét : 1
 EFFET B.C -- ddl = 1. Critères tr : 1, dét : 1
 EFFET A³ -- ddl = 1. Critères tr : 1, dét : 1
 EFFET A².B -- ddl = 1. Critères tr : 1, dét : 1
 EFFET A².C -- ddl = 1. Critères tr : 1, dét : 1
 EFFET A.B² -- ddl = 1. Critères tr : 1, dét : 1
 EFFET A.B.C -- ddl = 1. Critères tr : 0.84, dét : 0.84
 EFFET B³ -- ddl = 1. Critères tr : 1, dét : 1
 EFFET B².C -- ddl = 1. Critères tr : 1, dét : 1
 EFFET A³.B -- ddl = 1. Critères tr : 0.2, dét : 0.2
 EFFET A³.C -- ddl = 1. Critères tr : 1, dét : 1
 EFFET A².B² -- ddl = 1. Critères tr : 1, dét : 1
 EFFET A².B.C -- ddl = 1. Critères tr : 1, dét : 1
 EFFET A.B³ -- ddl = 1. Critères tr : 0.8, dét : 0.8
 EFFET A.B².C -- ddl = 1. Critères tr : 1, dét : 1
 EFFET B³.C -- ddl = 1. Critères tr : 1, dét : 1
 EFFET BL -- ddl = 3. Critères tr : 0.3554301834, dét : 0.512231666
 effic. principales : 0.840 1.000 0.160
 Mesure globales d'efficacité du plan
 trace det valmin (globaux)
 0.705 0.926 0.083

Valeurs propres de la matrice d'information par unité (X'X/N)
 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000
 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000
 1.000 0.600 1.000 1.400 0.083 1.917

TAB. 10 – Exemple 8. Efficacités avec modèle de degré 4

où $(\gamma_0, \gamma_1) = (4, 8)$, $(\delta_0, \delta_1, \delta_2, \delta_3) = (5, 5, 7, 8)$ et ε_2 est un 32 échantillon d'une normale centrée d'écart-type 3, orthonormalisé à la matrice X du modèle.

Le tableau 11 donne les éléments principaux de l'analyse de variance. On notera que A et B ont bien été déclarés comme des covariables (facteurs quantitatifs). Le modèle a été écrit $P^3 + BL$ où P est la somme $A + B + C$. En développant on retrouve tous les termes de degré 3 en A, B, C . Mais comme C est qualitatif, on remplace $A.C.C$ et $B.C.C, C.C.C$ respectivement par $A.C, B.C, C$. Après développement, on trouve donc

$$C + A^3 + B^3 + AB^2 + AC + A^2B + A^2C + BC + B^2C + ABC + BL .$$

C'est ce modèle, complété par les sous-termes, c'est à dire les monomes de degré inférieur à ceux figurant déjà, qui est utilisé pour effectuer l'analyse de variance dont les résultats figurent au tableau 11. A la fin de ce tableau figurent les estimations du vecteur $\hat{\beta}$ des paramètres estimés pour chaque variable observée. Ces paramètres sont ici triés par valeur absolue décroissante parce que la case *Tri des paramètres* figurant en bas à gauche de la figure 2 a été cochée.

Rappelons (§4) que ces paramètres sont formés à partir des bases de contrastes sélectionnées pour chacun des 3 facteurs. Ces bases sont définies dans le tableau 12 qui reprend certaines sorties volontairement omises dans le tableau 11. La notation \wedge^i utilisée dans ces tableaux fait référence soit à la puissance i ème de la covariable (à gauche des tableaux), soit au polynôme orthogonal de degré i (en haut des tableaux). Cette même notation apparaît aussi dans les *libellés des paramètres*. Si le facteur est quantitatif, elle fait référence au polynôme orthogonal de degré i ou à la puissance i ème suivant le choix effectué (cf figure 3. Pour un facteur qualitatif, elle fait référence au contraste c_i .

Le tableau 12 montre que les polynômes orthogonaux pour A sont définis par :

$$\begin{aligned} P_0(A) &= A^0 = 1, & P_1(A) &= 0.89443(A - 6.5), \\ P_2(A) &= (A - 6.5)^2 - 1.25, & P_3(A) &= 1.49071(A - 6.5)^3 - 3.05596(A - 6.5) \end{aligned}$$

et de même les polynômes orthogonaux pour B par :

$$\begin{aligned} P_0(B) &= B^0 = 1, & P_1(B) &= 0.44721(B - 23), \\ P_2(B) &= 0.25(B - 23)^2 - 1.25, & P_3(B) &= 0.18634(B - 23)^3 - 1.52798(B - 23) . \end{aligned}$$

Les modèles figurant dans le tableau 11 pour Y1 et Y2 ont été obtenus en travaillant pour chaque covariable avec les polynômes orthogonaux associé, ce qui est le choix retenu de façon standard. Le modèle estimé pour Y1 peut donc se réécrire (en négligeant les termes inférieurs en valeur absolue à 10^{-8})

$$\begin{aligned} E(Y1 | A, B, C) &= 16.5 - 6.149P_1(B) + 4.472P_1(A) + 2.795P_1(B).c_1(C) \\ &\quad + 1.443c_3(BL) - 1.414c_1(BL) - 0.4082c_2(BL). \end{aligned} \quad (27)$$

Dans cette expression, $c_1(C), c_3(BL), c_1(BL), c_2(BL)$ sont les coordonnées des colonnes correspondantes du tableau 12. Donc $c_1(C)$ est défini par $c_1(\text{faible}) = -1, c_1(\text{fort}) = 1$, $c_3(BL)$ par les égalités $c_3(0) = -0.57735, c_3(1) = -0.57735, c_3(2) = -0.57735, c_3(3) = 1.73205$, etc ...

Parce qu'on a pris la précaution d'orthogonaliser le vecteur des erreurs simulées à la matrice X du modèle, le modèle estimé (27) est identique au modèle initialement simulé.

```

...
A      num  cov
B      num  cov
C      alpha fac  faible fort
BL     num  fac

```

```

Variables analysées : Y1 Y2
Parties de modèle : P:A+B+C
Modèle : P^3+BL
Sur-Modèle :

```

----- ANALYSE -----

```

...
Variance(s) expliquée(s) 124.2 92.44 ddl = 18
Variance(s) d'erreur 10.01 4.901 ddl = 13
R^2 = 0.945 0.9631

```

```

Analyse avec l'erreur er ddl = 13
      Y1  Y2
CM erreur 10.011 4.901
ect-type 3.164 2.214

```

F de Snedecor			Probabilités (%)		Signification				
ddl	Y1	Y2	Y1	Y2	Y1	Y2			
1	870.2	255.0	0.0	0.0	***	***			
A	1	63.9	293.8	A	0.0	0.0	A	***	***
B	1	120.9	0.0	B	0.0	100.0	B	***	
C	1	0.0	0.0	C	100.0	100.0	C		
A^2	1	0.0	0.0	A^2	100.0	100.0	A^2		
A.B	1	0.0	8.6	A.B	100.0	1.2	A.B		*
A.C	1	0.0	32.6	A.C	100.0	0.0	A.C		***
B^2	1	0.0	0.0	B^2	100.0	100.0	B^2		
B.C	1	25.0	0.0	B.C	0.0	100.0	B.C	***	
A^3	1	0.0	0.0	A^3	100.0	100.0	A^3		
A^2.B	1	0.0	0.0	A^2.B	100.0	100.0	A^2.B		
A^2.C	1	0.0	0.0	A^2.C	100.0	100.0	A^2.C		
A.B^2	1	0.0	0.0	A.B^2	100.0	100.0	A.B^2		
A.B.C	1	0.0	0.0	A.B.C	100.0	100.0	A.B.C		
B^3	1	0.0	0.0	B^3	100.0	100.0	B^3		
B^2.C	1	0.0	0.0	B^2.C	100.0	100.0	B^2.C		
BL	3	4.5	3.1	BL	2.3	6.4	BL		*

variable Y1				variable Y2					
effet	+/-	+/-	+/-	effet	+/-	+/-	+/-		
	(95%)	(99%)	(99.9%)		(95%)	(99%)	(99.9%)		
	1.650E1	1.208	1.685	2.361	A	6.708E0	0.8455	1.179	1.652
B	-6.149E0	1.208	1.685	2.361		6.250E0	0.8455	1.179	1.652
A	4.472E0	1.208	1.685	2.361	A.C	2.236E0	0.8455	1.179	1.652
B.C	2.795E0	1.208	1.685	2.361	A.B	1.250E0	0.9225	1.286	1.802
BL^3	1.443E0	1.283	1.789	2.506	BL^3	1.010E0	0.8976	1.252	1.754
BL	-1.414E0	1.265	1.763	2.471	BL^2	8.165E-1	0.9101	1.269	1.778
BL^2	-4.082E-1	1.301	1.814	2.541	A^2.B	7.519E-10	0.8455	1.179	1.652
A.B	-1.384E-9	1.318	1.838	2.576	B^3	-6.904E-10	0.8455	1.179	1.652
A.B.C	-1.146E-9	1.318	1.838	2.576	B	5.842E-10	0.8455	1.179	1.652
B^3	2.516E-10	1.208	1.685	2.361	BL	-2.881E-10	0.8849	1.234	1.729
A^3	-2.516E-10	1.208	1.685	2.361	A.B.C	2.842E-10	0.9225	1.286	1.802
C	-1.875E-10	1.208	1.685	2.361	C	-1.875E-10	0.8455	1.179	1.652
A.B^2	1.398E-10	1.208	1.685	2.361	B^2	-1.875E-10	0.8455	1.179	1.652
A^2.B	-1.398E-10	1.208	1.685	2.361	A^2.C	1.875E-10	0.8455	1.179	1.652
A.C	1.118E-10	1.208	1.685	2.361	A^3	1.677E-10	0.8455	1.179	1.652
B^2	-2.064E-16	1.208	1.685	2.361	A^2	1.250E-10	0.8455	1.179	1.652
B^2.C	-1.568E-16	1.208	1.685	2.361	B^2.C	-1.250E-10	0.8455	1.179	1.652
A^2	1.060E-16	1.208	1.685	2.361	B.C	-1.090E-10	0.8455	1.179	1.652
A^2.C	-1.002E-17	1.208	1.685	2.361	A.B^2	-8.385E-11	0.8455	1.179	1.652

TAB. 11 – Exemple 8. Analyse de variance avec modèle de degré 3.

coeff. pol. orth. pour A . Chaque colonne donne les coeff. à appliquer aux puissances successives de A-6.5. Le libellé $\gamma_0, \gamma_1, \gamma_2, \dots$ à gauche précise cette puissance. Le libellé $\gamma_0, \gamma_1, \gamma_2, \dots$ en haut de col. donne le degré du polynome orthogonal.

					idem pour B (B-23)				
	γ_0	γ_1	γ_2	γ_3		γ_0	γ_1	γ_2	γ_3
γ_0	1	0.00000	-1.25	0.00000	γ_0	1	0.00000	-1.25	0.00000
γ_1	0	0.89443	0.00	-3.05596	γ_1	0	0.44721	0.00	-1.52798
γ_2	0	0.00000	1.00	0.00000	γ_2	0	0.00000	0.25	0.00000
γ_3	0	0.00000	0.00	1.49071	γ_3	0	0.00000	0.00	0.18634

contrastes pour facteur C

niv	c0	c1
faible	1	-1
fort	1	1

contrastes pour facteur BL

niv	c0	c1	c2	c3
0	1	-1.41421	-0.81650	-0.57735
1	1	1.41421	-0.81650	-0.57735
2	1	0.00000	1.63299	-0.57735
3	1	0.00000	0.00000	1.73205

TAB. 12 – Exemple 8. Contrastes utilisés dans la paramétrisation

Cela se vérifie facilement dans cet exemple en distinguant dans le modèle la contribution de A, celle du couple B, C, enfin celle de la constante et de l'effet bloc.

- Contribution de A

$$4.472P_1(A) = 4.472 \times 0.89443(A - 6.5) = 4(A - 6.5) . \quad (28)$$

- Contribution de B, C

$$\begin{aligned} -6.149P_1(B) + 2.795P_1(B).c_1(C) &= (-6.149 + 2.795c_1(C))P_1(B) \\ &= (-6.149 + 2.795c_1(C))0.44721(B - 23) . \end{aligned}$$

Comme

$$\begin{aligned} (-6.149 + 2.795c_1(C))0.44721 &= -4 = \gamma_0 \quad \text{si } C = \text{faible} \quad (c_1(C) = -1), \\ &= -1.5 = \gamma_1 \quad \text{si } C = \text{fort} \quad (c_1(C) = 1) . \end{aligned}$$

cette contribution se met sous la forme

$$\begin{aligned} -4(B - 23) & \quad \text{si } C = \text{faible} \\ -1.5(B - 23) & \quad \text{si } C = \text{fort} . \end{aligned} \quad (29)$$

- Contribution de la constante et du facteur bloc

$$\begin{aligned} 16.5 + 1.443c_3(\text{BL})\mu - 1.414c_1(\text{BL}) - 0.4082c_2(\text{BL}) \\ = 16.5 + 1.443c_3(0) - 1.414c_1(0) - 0.4082c_2(0) &= 18 \quad \text{si BL} = 0 \\ = 16.5 + 1.443c_3(1) - 1.414c_1(1) - 0.4082c_2(1) &= 14 \quad \text{si BL} = 1 \\ = 16.5 + 1.443c_3(2) - 1.414c_1(2) - 0.4082c_2(2) &= 15 \quad \text{si BL} = 2 \\ = 16.5 + 1.443c_3(3) - 1.414c_1(3) - 0.4082c_2(3) &= 19 \quad \text{si BL} = 3. \end{aligned} \quad (30)$$

Moyennes

L'analyse de variance (tableau 11) retrouve pour Y1 les effets nets A, B, B.C, ainsi que l'effet plus faible des blocs BL et pour Y2 les effets de A, les interactions A.C et A.B. Pour expliciter ces effets, le plus simple est de demander le calcul des moyennes ad hoc en faisant figurer les lignes suivantes dans la fenêtre *Moyennes* de la figure 2:

A
A B
A C
B C
BL

On obtient alors les résultats reportés au tableau 13. On peut retrouver ces résultats en

Calcul des moyennes					
Moyenne sur A		Moyenne sur B C		Moyenne sur A B	
Y1	Y2		Y1 Y2	Y1	Y2
5	10.5 -2.75	20	faible 28.5 6.25	5	20 18.75 -0.5
6	14.5 3.25	20	fort 21.0 6.25	5	22 13.25 -2.0
7	18.5 9.25	22	faible 20.5 6.25	5	24 7.75 -3.5
8	22.5 15.25	22	fort 18.0 6.25	5	26 2.25 -5.0
		24	faible 12.5 6.25	6	20 22.75 4.0
		24	fort 15.0 6.25	6	22 17.25 3.5
Moyenne sur A C		26	faible 4.5 6.25	6	24 11.75 3.0
Y1	Y2	26	fort 12.0 6.25	6	26 6.25 2.5
5	faible 10.5 0.25			7	20 26.75 8.5
5	fort 10.5 -5.75	Moyenne sur BL		7	22 21.25 9.0
6	faible 14.5 4.25	Y1	Y2	7	24 15.75 9.5
6	fort 14.5 2.25	0	18 5	7	26 10.25 10.0
7	faible 18.5 8.25	1	14 5	8	20 30.75 13.0
7	fort 18.5 10.25	2	15 7	8	22 25.25 14.5
8	faible 22.5 12.25	3	19 8	8	24 19.75 16.0
8	fort 22.5 18.25			8	26 14.25 17.5

TAB. 13 – Exemple 8. Moyennes avec modèle de degré 3.

utilisant des formules du type de (24), (25), (26). Ainsi les moyennes sur BL s’obtiennent en ne considérant dans le modèle (27) que la constante et les termes relatifs au facteur BL. Elles sont donc précisément données par (30). De même les moyennes sur A sont obtenues en rajoutant à la moyenne estimées 16.5 la contribution du facteur A donnée par (28) et celles sur (B,C) en rajoutant à 16.5 la contribution (29 des termes ne contenant que B ou C. Les moyennes du tableau 13 se retrouvent immédiatement en remplaçant A, B, C par les niveaux appropriés dans (28) et (29).

7 Etude des résidus

Les résidus sont les écarts entre les valeurs observées et les prédictions obtenues par le modèle estimé. Leur étude permet de détecter des points suspects que l’on peut alors éliminer d’une analyse ultérieure (une façon commode de le faire est d’accoler un S à la valeur suspecte dans le fichier analysé). L’examen du graphique valeurs estimées, résidus peut aussi permettre d’examiner si l’hypothèse que la variance d’erreur est constante (homoscédasticité) est acceptable ou pas.

On obtient l’étude de ces résidus en cochant la case *Etude résidus* en bas à droite de l’écran de la figure 2. On peut alors préciser les options retenues pour cette étude en appuyant sur le bouton *Options* à droite de cette case, ce qui fait apparaître l’écran de la figure 4. Les sorties peuvent comprendre, pour chacune des variables analysées

- un tableau fournissant pour chaque unité une aide au diagnostic. Cette aide comprend notamment le résidu réduit, une statistique du *t* de student qui s’en déduit, le *résidu studentisé*, et la probabilité associée. Ces aides complètent les graphiques.
- le graphique des quantiles des résidus studentisés qui fournit une aide graphique pour détecter des données suspectes,
- le graphique représentant les valeurs observées en fonction des valeurs calculées par le modèle
- le graphique représentant les résidus en fonction des valeurs calculées par le modèle.



FIG. 4 – Options pour l'étude des résidus

7.1 Exemples

On s'appuiera pour présenter les sorties sur les deux exemples suivants.

Exemple 7.1 . Le tableau 14 donne un exemple de plan comportant 2 facteurs qualitatifs A , B et deux variables analysées $Y1$ et $Y2$.

label	A	B	# $Y1$	$Y2$
$A1$	1	10	3.0	1.15
$A2$	1	10	7.2	1.07
$B1$	1	20	2.5	0.95
$B2$	1	20	3.0	0.86
$C1$	2	10	5.0	0.81
$C2$	2	10	4.0	0.97
$D1$	2	20	2.0	0.83
$D2$	2	20	3.0	1.06
$E1$	3	10	4.0	1.06
$E2$	3	10	3.0	1.24
$F1$	3	20	3.0	1.02
$F2$	3	20	3.5	0.77

TAB. 14 – Exemple avec deux facteurs croisés

Le tableau 15 et les figures 5, 6, 7 donnent les sorties correspondant à l'option standard.

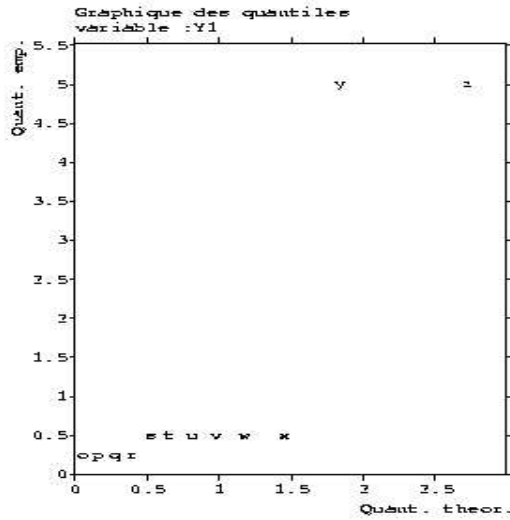
Exemple 7.2 . Nous reprenons les données du tableau 8 en modifiant la réponse $Y1$ pour les unités 13 (37.36125 au lieu de 0.36125) et 23 (39.328392857 au lieu de 9.328392857). Les lignes modifiées sont reportées dans le tableau 16.

Etude des résidus: fichier exABdat, variable Y1.

facteurs : A, B
 modèle : A.B
 écart-type d'erreur : 1.327 (ddl: 6)

	A	B	Y	YP	YR	NR	YRn	t	P	s	S
1	1	10	3.0	5.10	-2.10	0.707	-2.238	-5.020	0.004	!!	***
2	1	10	7.2	5.10	2.10	0.707	2.238	5.020	0.004	!!	***
3	1	20	2.5	2.75	-0.25	0.707	-0.266	-0.245	0.816		
4	1	20	3.0	2.75	0.25	0.707	0.266	0.245	0.816		
5	2	10	5.0	4.50	0.50	0.707	0.533	0.498	0.639		
6	2	10	4.0	4.50	-0.50	0.707	-0.533	-0.498	0.639		
7	2	20	2.0	2.50	-0.50	0.707	-0.533	-0.498	0.639		
8	2	20	3.0	2.50	0.50	0.707	0.533	0.498	0.639		
9	3	10	4.0	3.50	0.50	0.707	0.533	0.498	0.639		
10	3	10	3.0	3.50	-0.50	0.707	-0.533	-0.498	0.639		
11	3	20	3.0	3.25	-0.25	0.707	-0.266	-0.245	0.816		
12	3	20	3.5	3.25	0.25	0.707	0.266	0.245	0.816		

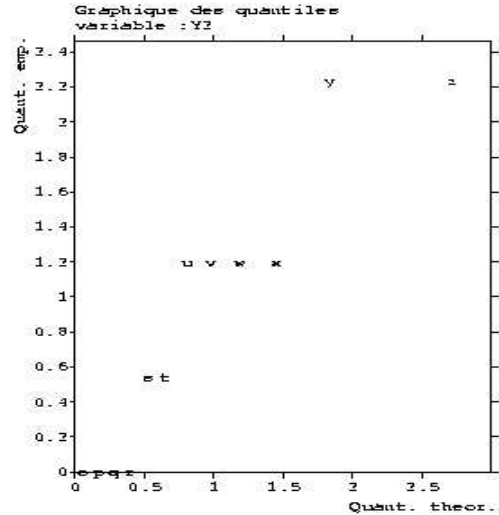
TAB. 15 – Aide au diagnostic, exemple 7.1, var Y1



Complément au graphique des quantiles, variable Y1

symboles et valeurs des quantiles pour les résidus les plus grands

symb	n0	Qemp	Prob	Qth
z	2	5.0200	0.9583	2.7222
y	1	5.0200	0.8750	1.8409
x	9	0.4983	0.7917	1.4440
w	8	0.4983	0.7083	1.1784
v	5	0.4983	0.6250	0.9735
u	7	0.4983	0.5417	0.8032
t	6	0.4983	0.4583	0.6546
s	10	0.4983	0.3750	0.5204
r	12	0.2446	0.2917	0.3962
q	3	0.2446	0.2083	0.2787
p	4	0.2446	0.1250	0.1655
o	11	0.2446	0.0417	0.0549



Complément au graphique des quantiles, variable Y2

symboles et valeurs des quantiles pour les résidus les plus grands

symb	n0	Qemp	Prob	Qth
z	8	2.2361	0.9583	2.7222
y	7	2.2361	0.8750	1.8409
x	3	1.1952	0.7917	1.4440
w	1	1.1952	0.7083	1.1784
v	4	1.1952	0.6250	0.9735
u	2	1.1952	0.5417	0.8032
t	10	0.5423	0.4583	0.6546
s	9	0.5423	0.3750	0.5204
r	6	0.0000	0.2917	0.3962
q	5	0.0000	0.2083	0.2787
p	12	0.0000	0.1250	0.1655
o	11	0.0000	0.0417	0.0549

FIG. 5 – graphique des quantiles, exemple 7.1

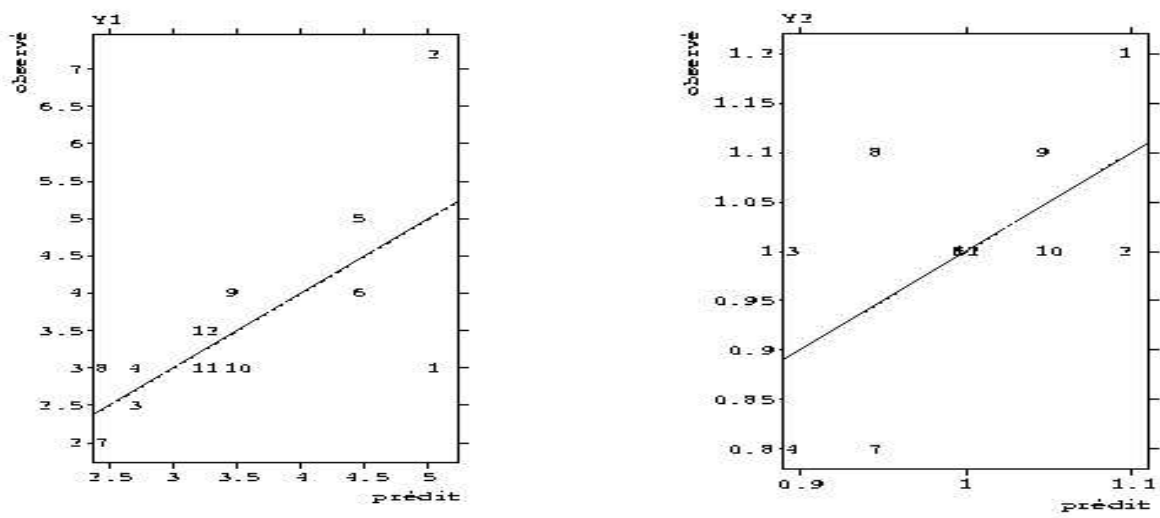


FIG. 6 – *Observé versus prédit, exemple 7.1*

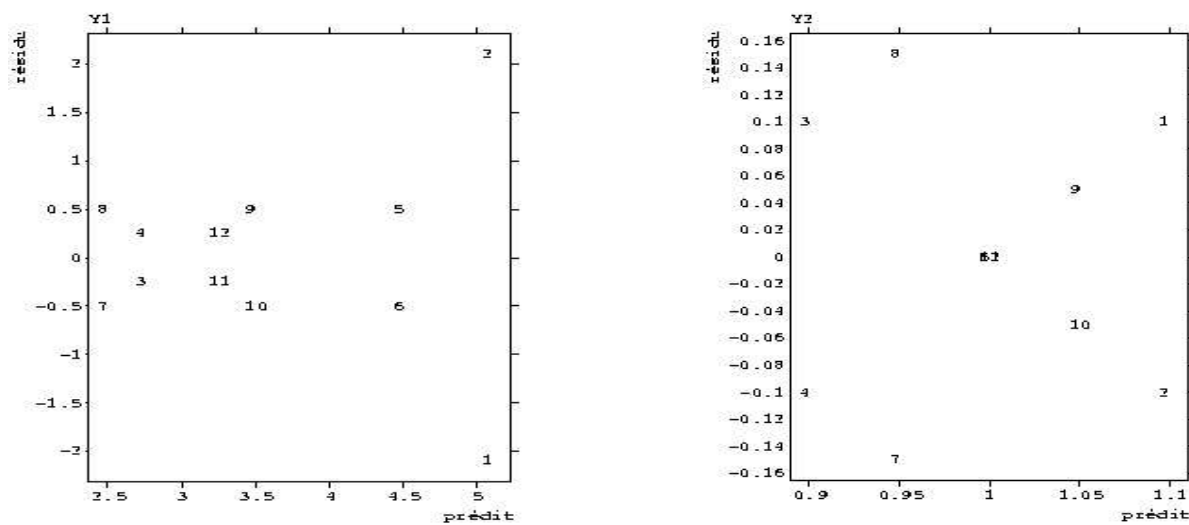


FIG. 7 – *résidu versus prédit, exemple 7.1*

n0	A ₁	A ₂	B ₁	B ₂	BL ₁	BL ₂	A	B	C	BL	#Y1	Y2
13	0	0	1	1	1	1	5	26	faible	3	37.36125	-0.5122916667
23	0	1	1	0	0	1	6	24	fort	1	39.328392857	-2.528839286

TAB. 16 – *Exemple 7.2. Lignes modifiées dans le tableau 8*

Le tableau 17 et les figures 8, 9 donnent les sorties correspondant à l'option standard.

Nom d'analyse : EX442B2

A, B sont des covariables, C un facteur, BL le facteur bloc.
La variable analysée est Y1, le modèle est défini par

Parties de modèle : $P : 1 + A + B + C$
Modèle : $P^3 + BL$

Etude des résidus pour la variable Y1

	A	B	C	BL	Y	YP	YR	NR	YRn	t	P	² s	S
1	5	20	0	0	26.36125	23.448	2.913	0.472	0.689	0.675	0.513		
2	7	20	0	1	30.44387	29.617	0.827	0.656	0.141	0.135	0.894		
3	6	20	0	2	22.84363	22.946	-0.102	0.656	-0.017	-0.017	0.987		
4	8	20	0	3	36.05125	41.165	-5.113	0.472	-1.210	-1.234	0.241		
5	5	24	0	2	7.40435	12.196	-4.791	0.656	-0.816	-0.805	0.436		
6	7	24	0	3	16.39339	14.906	1.487	0.736	0.226	0.217	0.832		
7	6	24	0	0	14.26911	13.656	0.613	0.736	0.093	0.089	0.930		
8	8	24	0	1	17.03315	18.767	-1.734	0.656	-0.295	-0.285	0.781		
9	5	22	0	1	9.70315	13.617	-3.914	0.656	-0.667	-0.652	0.527		
10	7	22	0	0	24.42911	18.156	6.273	0.736	0.952	0.948	0.362		
11	6	22	0	3	17.53339	18.906	-1.373	0.736	-0.208	-0.201	0.844		
12	8	22	0	2	25.23435	21.796	3.439	0.656	0.586	0.570	0.579		
13	5	26	0	3	37.36125	26.565	10.797	0.472	2.555	3.479	0.005	!!	*
14	7	26	0	2	8.22363	1.796	6.428	0.656	1.095	1.104	0.291		
15	6	26	0	1	-2.13613	12.017	-14.153	0.656	-2.411	-3.116	0.009	!!	
16	8	26	0	0	9.85125	11.448	-1.597	0.472	-0.378	-0.365	0.721		
17	5	20	1	1	11.10458	10.406	0.698	0.472	0.165	0.159	0.876		
18	7	20	1	0	22.58696	25.476	-2.890	0.656	-0.492	-0.477	0.642		
19	6	20	1	3	23.75054	26.161	-2.410	0.656	-0.411	-0.397	0.698		
20	8	20	1	2	24.85792	18.781	6.077	0.472	1.438	1.507	0.158		
21	5	24	1	3	13.55482	16.318	-2.763	0.656	-0.471	-0.456	0.656		
22	7	24	1	2	11.38411	18.228	-6.844	0.736	-1.039	-1.042	0.318		
23	6	24	1	1	39.32839	20.310	19.019	0.736	2.886	4.628	0.001	!!!	***
24	8	24	1	0	20.63268	25.619	-4.987	0.656	-0.850	-0.840	0.417		
25	5	22	1	0	12.27268	11.419	0.853	0.656	0.145	0.140	0.891		
26	7	22	1	1	19.09839	23.860	-4.761	0.736	-0.723	-0.709	0.492		
27	6	22	1	2	16.81411	17.228	-0.414	0.736	-0.063	-0.060	0.953		
28	8	22	1	3	28.91482	29.018	-0.103	0.656	-0.018	-0.017	0.987		
29	5	26	1	2	3.23792	7.031	-3.793	0.472	-0.898	-0.890	0.391		
30	7	26	1	3	15.44054	15.961	-0.520	0.656	-0.089	-0.085	0.933		
31	6	26	1	0	13.59696	14.776	-1.180	0.656	-0.201	-0.193	0.850		
32	8	26	1	1	17.42458	13.406	4.018	0.472	0.951	0.947	0.362		

TAB. 17 – Aide au diagnostic, exemple 7.2, var Y1

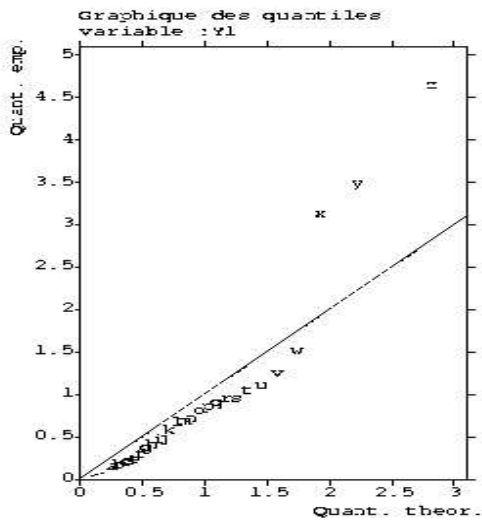
7.2 Terminologie et théorie

Pour expliciter, considérons l'exemple du plan de l'exemple 7.1 qui comporte deux facteurs A , B , analysé avec le modèle avec interaction $A.B$. Ce modèle peut s'écrire

$$E(y_{ij}) = \mu_{ij} \quad (31)$$

en notant y la réponse. Alors

- la valeur estimée (prédite, calculée) de y pour une unité qui a reçu le niveau i du facteur A et j du facteur B est l'estimateur des moindres carrés $\hat{\mu}_{ij}$, égal dans ce cas de figure à la moyenne y_{ij} . des observations de y sur l'ensemble des unités ayant reçu le traitement i, j : $\hat{\mu}_{ij} = y_{ij}$. Si on suppose la variance résiduelle constante,



Complément au graphique des quantiles, variable Y1

symboles et valeurs des quantiles pour les résidus les plus grands

symb	n0	Qemp	Prob	Qth
z	23	4.6275	0.9844	2.8143
y	13	3.4789	0.9531	2.2148
x	15	3.1161	0.9219	1.9260
w	20	1.5066	0.8906	1.7293
v	4	1.2341	0.8594	1.5777
u	14	1.1044	0.8281	1.4530
t	22	1.0421	0.7969	1.3462
s	10	0.9483	0.7656	1.2521
r	32	0.9471	0.7344	1.1677
q	29	0.8905	0.7031	1.0906
p	24	0.8399	0.6719	1.0194
o	5	0.8052	0.6406	0.9530
n	26	0.7087	0.6094	0.8906
m	1	0.6748	0.5781	0.8316

FIG. 8 – graphique des quantiles, exemple 7.2, var Y1

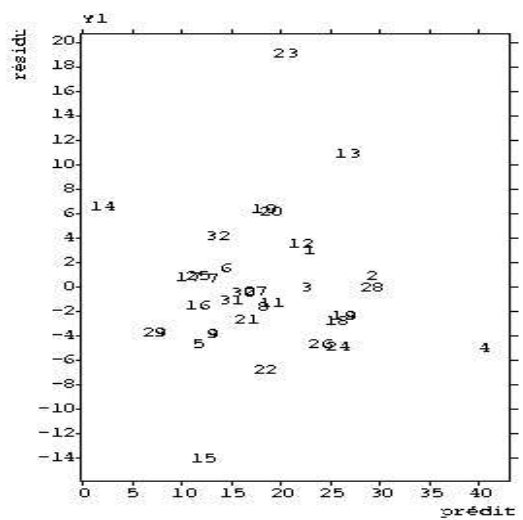
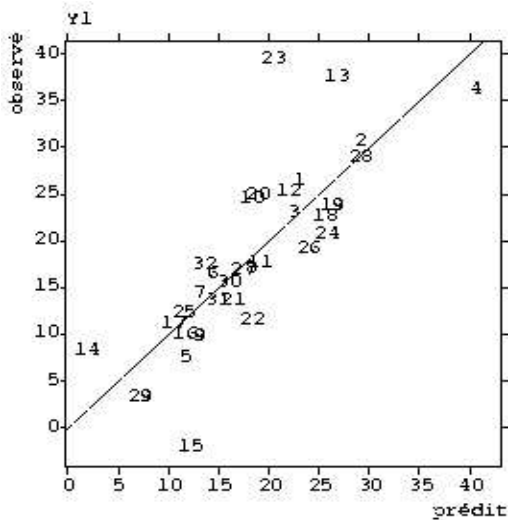


FIG. 9 – Observé et résidu versus prédit, exemple 7.2, var Y1

égale à σ^2 :

$$\text{var}(y_{ijk}) = \sigma^2 \quad (32)$$

la variance de $\hat{\mu}_{ij}$ est σ^2/n_{ij} où n_{ij} désigne le nombre de répétitions du traitement ij .

- le *résidu* correspondant à la $k^{\text{ème}}$ répétition du traitement ij est la différence $y_{ijk} - y_{ij\cdot}$ entre la valeur observée et la valeur estimée. Il est facile de vérifier que sa variance est ici $\sigma^2(1 - 1/n_{ij})$.
- la *norme associée au résidu* est la racine du coefficient de σ^2 dans cette variance, soit : $\sqrt{1 - 1/n_{ij}}$.
- le *résidu réduit* est donc

$$r_{ijk} = \frac{y_{ijk} - y_{ij\cdot}}{\hat{\sigma} \sqrt{1 - 1/n_{ij}}}$$

où $\hat{\sigma}^2$ est l'estimation usuelle de la variance résiduelle.

Soit q le nombre de degrés de liberté de cette variance résiduelle, dans cet exemple égal à $\sum_{ij}(n_{ij} - 1)$. Supposons en plus des hypothèses (31) et (32) que les observations sont indépendantes et normalement distribuées. Alors la quantité

$$t_{ijk} = \sqrt{\frac{q - 1}{q - r_{ijk}^2}} r_{ijk}$$

suit la distribution du t de student centrée avec $q - 1$ degrés de liberté. Cette statistique, appelée *résidu studentisé*, peut être utilisée pour tester une éventuelle aberration sur y_{ijk} . Ce résultat classique est redémontré à la fin de ce paragraphe.

- la *probabilité associée* à la statistique t_{ijk} est la probabilité p_{ijk} qu'un t de Student centré à $q - 1$ degrés de liberté dépasse t_{ijk} en valeur absolue.

S'il y a n données au total et que les statistiques t_{ijk} sont toutes centrées, on doit s'attendre à détecter au seuil 5% en gros 5% de données suspectes bien que toutes les données soient "normales", c'est à dire obéissent au modèle retenu pour l'analyse.

Aussi il convient d'utiliser des seuils d'autant plus sévères que n est plus grand. En utilisant un seuil égal à α/n , on est assuré que la probabilité de détection à tort d'au moins une donnée comme suspecte est inférieure à $\alpha = n\alpha/n$. Mais cette stratégie conduit généralement à être trop conservateur et ne pas détecter des données qui s'écartent notablement du modèle.

Test que l'observation i est aberrante

Pour tester qu'une observation i est aberrante, on peut comparer le modèle M retenu

$$\text{M: } E(y) = X\beta$$

au sur-modèle

$$\text{M1: } E(y) = X\beta + z\gamma$$

où

$$z = (0, \dots, 0, 1, 0, \dots, 0)'$$

est le vecteur indicateur de l'observation i dont les coordonnées sont nulles sauf la $i^{\text{ème}}$ égale à 1. Le paramètre supplémentaire γ permet à l'observation i de varier librement et donc de s'écarter du modèle M retenu.

Soit $P = X(X'X)^{-1}X'$ le projecteur orthogonal sur l'espace engendré par les colonnes de X et $Q = I - P$. Posant $\beta' = \beta + (X'X)^{-1}X'z\gamma$, on peut réécrire le modèle M1 sous la forme

$$M1': E(y) = X\beta' + Qz\gamma.$$

L'orthogonalité entre Qz et X dans cette forme simplifie les calculs.

Soit s^2 = la variance résiduelle du modèle M et q son nombre de degrés de liberté. Rappelons que

$$s^2 = y'Qy/q \quad \text{avec } q = \text{rang}(Q).$$

Le projecteur orthogonal sur l'image de Qz s'écrit

$$R = Qz(z'Qz)^{-1}z'Q.$$

. Comme $\text{rang}(R) = \text{rang}(Qz) = 1$, la variance résiduelle du modèle M1 est

$$s_1^2 = \frac{y'(I - P - R)y}{q - 1}$$

et la nullité de γ peut être testée par la statistique

$$F = \frac{y'Ry}{s_1^2}$$

qui suit sous l'hypothèse nulle un F de Fisher Snedecor à 1 et $q - 1$ degrés de liberté.

On a

$$(q - 1)s_1^2 = y'(I - P - R)y = y'Qy - y'Ry = qs^2 - \frac{(z'Qy)^2}{z'Qz}.$$

Notons y_i la $i^{\text{ème}}$ coordonnée de y et x_i la $i^{\text{ème}}$ ligne de X . Alors, si on note $\hat{\beta} = (X'X)^{-1}X'y$ l'estimateur de β dans le modèle M

$$z'Qy = y_i - x_i'\hat{\beta} = y_i - \hat{y}_i \quad \text{et}$$

$$z'Qz = 1 - x_i'(X'X)^{-1}x_i.$$

Soit v_i cette dernière quantité qui est à σ^2 près la variance du résidu $y_i - \hat{y}_i$:

$$v_i = (1 - x_i'(X'X)^{-1}x_i) = \frac{\text{var}(y_i - \hat{y}_i)}{\sigma^2}.$$

Le résidu réduit r_i s'obtient en divisant le résidu $y_i - \hat{y}_i$ par son écart-type estimé $s\sqrt{v_i}$:

$$r_i = \frac{y_i - \hat{y}_i}{s\sqrt{v_i}}.$$

Compte tenu de ce qui précède, on a

$$y'Ry = \frac{(z'Qy)^2}{z'Qz} = \frac{(y_i - \hat{y}_i)^2}{v_i} = s^2 r_i^2.$$

donc

$$(q - 1)s_1^2 = qs^2 - s^2 r_i^2 = (q - r_i^2)s^2$$

et finalement

$$F = \frac{q-1}{q-r_i^2} r_i^2$$

On peut aussi utiliser pour effectuer le test la quantité

$$t = \sqrt{\frac{q-1}{q-r_i^2}} r_i$$

qui suit sous les mêmes hypothèses un t de Student à $q-1$ degrés de liberté. Cette statistique s'obtient en divisant $y_i - \hat{y}_i$ par son écart-type estimé $s_1 \sqrt{v_i}$ puisque

$$\frac{y_i - \hat{y}_i}{s_1 \sqrt{v_i}} = \frac{s}{s_1} r_i = t.$$

On obtient bien sûr la même statistique à partir de l'estimateur des moindres carrés de γ dans le modèle M1 qui est proportionnel à $y_i - \hat{y}_i$:

$$\hat{\gamma} = \frac{z'Qy}{z'Qz} = \frac{y_i - \hat{y}_i}{v_i}.$$

7.3 Aide au diagnostic

Les colonnes du tableau d'aide au diagnostic (Tab. 17) sont les suivantes.

- Y : valeur observée
- YP : valeur prédite par le modèle
- YR : résidu = Y - YP
- NR : norme du résidu.
- YRn : résidu réduit $\frac{YR}{\hat{\sigma} NR}$
- t : résidu studentisé $\sqrt{\frac{q-1}{q-YRn^2}} YRn$
- P : probabilité de dépasser en valeur absolue le résidu studentisé
- s : signification du résidu studentisé : ! si $P < 0.05$, !! si $P < 0.01$, !!! si $P < 0.001$
- S : signification avec un seuil global : * si $P < 0.20/n$, ** si $P < 0.10/n$, *** si $P < 0.05/n$ où n est le nombre d'unités dans le plan.

7.4 Graphique des quantiles

Soit F la loi de répartition de la valeur absolue d'un t de student à $q-1$ degrés de liberté. Après avoir ordonné les unités du plan dans l'ordre croissant des valeurs absolues des résidus studentisés, on les représente avec

- en abscisse le quantile théorique: $F^{-1}(r/n + 1)$, où r est le numéro d'ordre

- en ordonnée le quantile empirique qui est la valeur absolue du résidu studentisé associé.

Le programme sort en complément du graphique un tableau donnant les coordonnées (Qemp, Qth) des points et les probabilités $r/n + 1$ (Prob) associées.

Si les résultats obéissent tous au modèle, les points obtenus sont sensiblement alignés le long de la première bissectrice. Les résultats en désaccord avec le modèle, donc suspects, donnent des points nettement au dessus de la bissectrice. Comme ils gonflent l'écart-type résiduel, les autres points ont tendance à se décaler vers le bas et se retrouvent donc au dessous de la bissectrice.

7.5 Interprétation

Dans l'exemple 7.1, les points 1 et 2 ressortent comme suspects pour la variable Y1, aussi bien dans le graphique des quantiles (figure 5) que dans le tableau 15 d'aide au diagnostic.

En fait ces deux points ont le même traitement défini par $A = 1$, $B = 10$. Leurs résidus sont de signes opposés et le diagnostic effectué signifie simplement que l'écart entre ces deux répétitions est anormalement important, compte tenu des autres écarts observés.

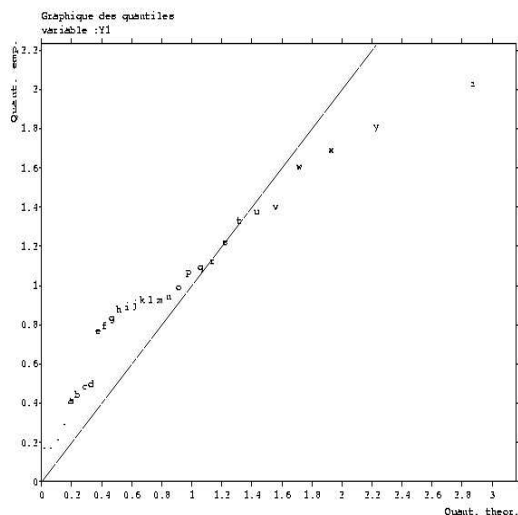
Pour essayer de diagnostiquer plus précisément lequel de ces deux points a un comportement anormal, on peut réanalyser avec le modèle additif $A + B$. C'est alors l'unité 2 qui ressort comme anormale.

Dans le graphique des quantiles de l'exemple 7.2 (fig. 8), 3 points s'écartent nettement de l'alignement : les points correspondant aux unités 23, 13, 15. Mais le tableau 17 donne un résidu studentisé de -3.116 avec $12 = 13 - 1$ degrés de liberté pour l'unité 15. La probabilité qu'un t de student à 12 ddl dépasse 3.116 est 0.9% . Il y a donc une probabilité non négligeable, sachant que l'on examine 32 tels résidus studentisés, d'en trouver au moins 1 pour lequel le test t est significatif à ce seuil. On décide donc de refaire un graphique des quantiles en éliminant seulement les unités 23 et 13, ce que l'on fait en accolant un symbole non numérique aux valeurs correspondantes du fichier et en reprenant l'analyse avec pour fichier plan le fichier ".PC" créé précédemment et pour fichier variables le fichier modifié avec les S. On obtient le graphique de la figure 10 qui suggère que le point 15 est "normal".

8 Détection des effets actifs dans un plan sans répétitions

La variance d'erreur est la variance des réponses quand on expérimente dans des conditions identiques, c'est à dire avec des niveaux constants des facteurs. C'est en comparant les contrastes estimés à l'écart-type déduit de cette variance que l'on détecte les contrastes significativement non nuls. Estimer la variance d'erreur est donc la première étape de l'analyse d'un plan d'expérience.

Si certains traitements sont répétés, on peut estimer cette variance d'erreur en comparant les répétitions de chaque traitement répété.



Complément au graphique des quantiles, variable Y1

symboles et valeurs des quantiles pour les résidus les plus grands

symb	n0	Qemp	Prob	Qth
		...		
<i>z</i>	21	2.0306	0.9833	2.8701
<i>y</i>	14	1.8154	0.9500	2.2281
<i>x</i>	13	1.6954	0.9167	1.9234
<i>w</i>	1	1.6104	0.8833	1.7174
<i>v</i>	11	1.4039	0.8500	1.5592
<i>u</i>	15	1.3827	0.8167	1.4296
<i>t</i>	30	1.3337	0.7833	1.3187
<i>s</i>	5	1.2242	0.7500	1.2213
<i>r</i>	26	1.1234	0.7167	1.1338

FIG. 10 – graphique des quantiles après élimination unités 23, 13, exemple 7.2

Dans les plans factoriels complets comportant beaucoup de facteurs, on fait souvent l'hypothèse que les interactions de 3 facteurs et plus sont négligeables et on utilise les contrastes correspondants pour estimer la variance d'erreur, évitant ainsi de coûteuses répétitions.

Dans les plans fractionnaires sans répétitions, il peut être difficile de faire a priori l'hypothèse que certains des contrastes étudiés sont nuls. Ainsi, dans la fraction 2^{5-1} à 5 facteurs A, B, C, D, E définie par $E = ABCD$, les 15 contrastes étudiés sont

$$\begin{array}{lllll}
 e(A) + e(BCDE), & e(B) + e(ACDE), & e(C) + e(ABDE), & e(D) + e(ABCE), & e(E) + e(ABCD) \\
 e(AB) + e(CDE), & e(AC) + e(BDE), & e(AD) + e(BCE), & e(AE) + e(BCD), & e(BC) + e(ADE) \\
 e(BD) + e(ACE), & e(BE) + e(ACD), & e(CD) + e(ABE), & e(CE) + e(ABD), & e(DE) + e(ABC)
 \end{array}$$

Chacun de ces 15 contrastes implique un effet principal ou une interaction de deux facteurs et, à moins de n'avoir déjà une bonne information sur le problème étudié, aucun ne peut a priori être considéré comme nul. Dans un tel cas de figure, deux procédures sont proposées pour l'analyse: une procédure graphique (Daniel) et une analyse Bayésienne (Box et Meyer). On trouvera dans [2] une description détaillée de ces procédures. Nous donnons ici quelques compléments sur l'utilisation du programme pour les mettre en oeuvre.

8.1 Graphique des quantiles demi-normaux (Daniel)

On l'obtient en cochant la case *Graphique de Daniel* en bas de l'écran de la figure 2. On y a recours quand l'analyse donne 0 degrés de liberté (ddl) pour estimer la variance résiduelle. Mais il est aussi utile si le modèle sélectionné ne laisse subsister qu'un nombre très réduit de degrés de liberté pour estimer la variance résiduelle. Dans ce cas, il faut compléter le modèle pour se retrouver avec exactement 0 ddl pour la variance résiduelle.

Si une première analyse fait ressortir des points s'écartant de l'alignement formé par les points associés aux effets les plus petits, on refait généralement l'analyse en éliminant ces points. Pour cela, on appuie sur le bouton *Options* à droite de la case *Graphique de*

Daniel et on précise le nombre k des effets à éliminer. Ce sont les k effets les plus grands en valeur absolue qui sont éliminés.

Exemple 8.1 . On reprend le plan 2^{5-1} de résolution 5 défini par $E = ABCD$. Le tableau 18 montre le fichier analysé avec le plan et la réponse Y . Celle ci a été obtenue par simulation, avec un écart-type d'erreur $\sigma = 1$ et des effets nuls à l'exception des effets principaux A , B , C et de l'interaction AB égaux respectivement à 2.25, -2.5 , -1.25 et 1 ([2], § 3.3.1).

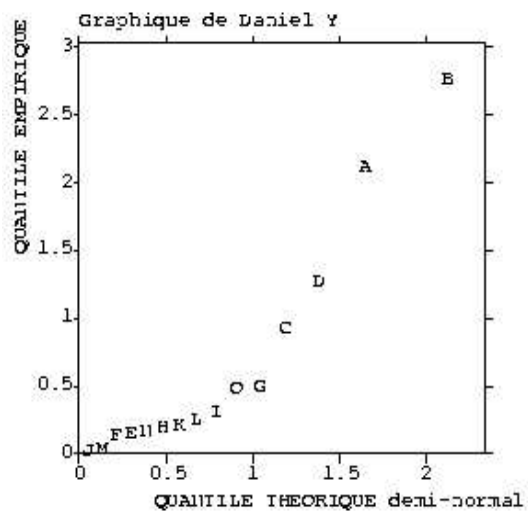
A	B	C	D	E	$\#Y$
1	1	1	1	1	12.847037729
1	1	1	-1	-1	12.447813377
1	1	-1	1	-1	15.714931181
1	1	-1	-1	1	16.091808862
1	-1	1	1	-1	17.256281020
1	-1	1	-1	1	17.332505682
1	-1	-1	1	1	20.030514679
1	-1	-1	-1	-1	17.088602385
-1	1	1	1	-1	6.964283530
-1	1	1	-1	1	7.590453774
-1	1	-1	1	1	8.194667060
-1	1	-1	-1	-1	10.151239848
-1	-1	1	1	1	13.248543868
-1	-1	1	-1	-1	14.167575764
-1	-1	-1	1	-1	16.767009750
-1	-1	-1	-1	1	18.106731492

TAB. 18 – *Données simulées pour une fraction 2^{5-1}*

La figure 11 donne à gauche le graphique de *daniel*. Le tableau à droite, obtenu en même temps que le graphique, précise pour chaque symbole du graphique l'effet associé, son ordonnée (quantile empirique Q_{emp}), son abscisse (quantile théorique Q_{th}) enfin la probabilité associée à ces quantiles (Prob). Cette dernière est $(i - 0.5)/(n - 1)$ où i est le rang de la valeur absolue de l'effet, $n - 1$ le nombre d'effets. Si on a pris soin de compléter convenablement le modèle pour que le nombre de ddl de l'écart-type résiduel soit nul, n est égal au nombre d'unités expérimentales.

Les effets B , A , C , AB , associés respectivement aux symboles B , A , D , C , s'écarte nettement de l'alignement formé par les autres effets. Si on les élimine en mettant un 4 dans la case *nombre de points supprimés* obtenue en appuyant sur le bouton *Options*, on obtient le graphique 12 où les points sont sensiblement alignés. La pente de la droite correspondante estime l'écart-type σ/\sqrt{n} des effets, où $n = 16$ est la taille de la fraction. Si on l'estime en régressant les 11 quantiles empiriques (ordonnée) sur les 11 quantiles théoriques (abscisses), dans un modèle ne comprenant pas de constante, on obtient l'estimation $\hat{\sigma}/4 = 0.265$, soit l'écart-type estimé $\hat{\sigma} = 1.06$ qui est très proche de l'écart-type 1 utilisé pour simuler les données.

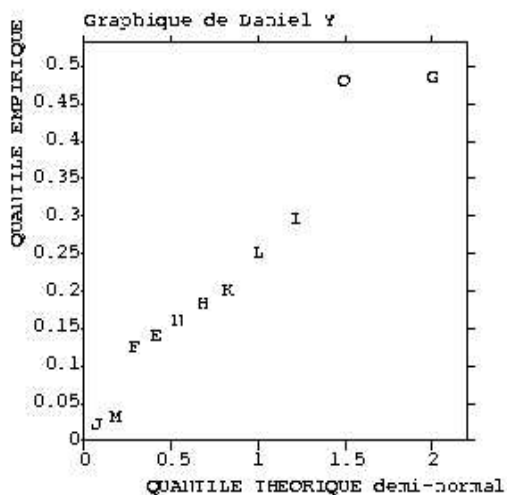
Exemple 8.2 . Pour chercher un milieu optimal de revivification de bactéries stressées par la chaleur, en vue de leur dénombrement, on a mis en place un plan 2^{9-4} de résolution 4. On



Complément au graphique de Daniel, var Y.

symp	effet	Qemp	Prob	Qth
B	+B	2.7497	0.9667	2.1285
A	+A	2.1012	0.9000	1.6452
D	+C	1.2682	0.8333	1.3832
C	+A.B	0.9239	0.7667	1.1919
G	+A.D	0.4831	0.7000	1.0364
O	+D.E	0.4780	0.6333	0.9026
I	+A.E	0.2940	0.5667	0.7833
L	+B.E	0.2496	0.5000	0.6742
K	+B.D	0.1980	0.4333	0.5726
H	+E	0.1803	0.3667	0.4766
N	+C.E	0.1575	0.3000	0.3849
E	+A.C	0.1379	0.2333	0.2963
F	+D	0.1221	0.1667	0.2100
M	+C.D	0.0307	0.1000	0.1254
J	+B.C	0.0197	0.0333	0.0417

FIG. 11 – graphique de Daniel sur l'ensemble des effets, exemple 8.1



Complément au graphique de Daniel, var Y.

Les 4 + grands contrastes en valeur absolue ne sont pas pris en compte dans le graphique.

symp	effet	Qemp	Prob	Qth
G	+A.D	0.48310	0.9545	2.0009
O	+D.E	0.47800	0.8636	1.4898
I	+A.E	0.29400	0.7727	1.2075
L	+B.E	0.24957	0.6818	0.9982
K	+B.D	0.19796	0.5909	0.8253
H	+E	0.18028	0.5000	0.6742
N	+C.E	0.15746	0.4091	0.5371
E	+A.C	0.13791	0.3182	0.4095
F	+D	0.12209	0.2273	0.2884
M	+C.D	0.03068	0.1364	0.1714
J	+B.C	0.01969	0.0455	0.0569

FIG. 12 – graphique de Daniel après suppression de 4 points, exemple 8.1

trouvera dans [2], § 3.3.4 la description détaillée de l'exemple et du plan. La variable Y analysée est le logarithme à base 10 du nombre d'UFC (Unités Formant Colonie) après croissance sur VRBL (Violet Red Bile Lactose agar). Pour simplifier, les facteurs ont été notés ici avec les 9 premières lettres de l'alphabet A, B, \dots, I .

Le modèle comprenant toutes les interactions de 2 facteurs s'écrit $P.P$ où P est la partie de modèle $A + B + C + D + E + F + G + H + I$. Pour obtenir dans les sorties que les effets principaux apparaissent avant les interactions, on l'écrit sous la forme $P + P.P$. L'analyse de ce modèle (partie gauche du tableau 19) confirme que le plan est de résolution 4. Ce plan permet d'estimer, outre la moyenne générale et les 9 effets principaux, 8 interactions de deux facteurs, puis 12 sommes de deux interactions et une somme de 4 interactions. Cela fait un total de 30 f.e.b. Il laisse donc 1 degré de liberté pour estimer la variance résiduelle ce qui est manifestement insuffisant pour tester les effets. Pour déterminer les

Analyse avec modèle $P + P.P$ P: A+B+C+D+E+F+G+H+I	Analyse avec modèle $P + P.P + P.P.P$ P: A+B+C+D+E+F+G+H+I
1 +	1 +
2 + A	2 + A + B.E.H + C.G.H + F.H.I
3 + B	3 + B + A.E.H + C.E.G + E.F.I
4 + C	4 + C + A.G.H + B.E.G + F.G.I
5 + D	5 + D
6 + E	6 + E + A.B.H + B.C.G + B.F.I
7 + F	7 + F + A.H.I + B.E.I + C.G.I
8 + G	8 + G + A.C.H + B.C.E + C.F.I
9 + H	9 + H + A.B.E + A.C.G + A.F.I
10 + I	10 + I + A.F.H + B.E.F + C.F.G
11 + A.B + E.H	11 + A.B + E.H - C.D.F - D.G.I
12 + A.C + G.H	12 + A.C + G.H - B.D.F - D.E.I
13 + A.D	13 + A.D - B.C.F - B.G.I - C.E.I - E.F.G
14 + A.E + B.H	14 + A.E + B.H - C.D.I - D.F.G
15 + A.F + H.I	15 + A.F + H.I - B.C.D - D.E.G
16 + A.G + C.H	16 + A.G + C.H - B.D.I - D.E.F
17 + A.H + B.E + C.G + F.I	17 + A.H + B.E + C.G + F.I
18 + A.I + F.H	18 + A.I + F.H - B.D.G - C.D.E
19 + B.C + E.G	19 + B.C + E.G - A.D.F - D.H.I
20 + B.D	20 + B.D - A.C.F - A.G.I - C.H.I - F.G.H
21 + B.F + E.I	21 + B.F + E.I - A.C.D - D.G.H
22 + B.G + C.E	22 + B.G + C.E - A.D.I - D.F.H
23 + B.I + E.F	23 + B.I + E.F - A.D.G - C.D.H
24 + C.D	24 + C.D - A.B.F - A.E.I - B.H.I - E.F.H
25 + C.F + G.I	25 + C.F + G.I - A.B.D - D.E.H
26 + C.I + F.G	26 + C.I + F.G - A.D.E - B.D.H
27 + D.E	27 + D.E - A.C.I - A.F.G - C.F.H - G.H.I
28 + D.F	28 + D.F - A.B.C - A.E.G - B.G.H - C.E.H
29 + D.G	29 + D.G - A.B.I - A.E.F - B.F.H - E.H.I
30 + D.H	30 + D.H - B.C.I - B.F.G - C.E.F - E.G.I
31 + D.I	31 + D.I - A.B.G - A.C.E - B.C.H - E.G.H
	32 + A.D.H + B.D.E + C.D.G + D.F.I

TAB. 19 – Effets confondus par la fraction 2^{9-4} de l'exemple 8.2

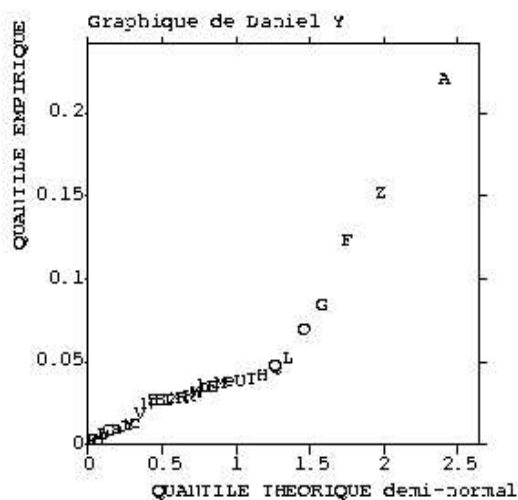
effets actifs, on demande donc le graphique de Daniel. Mais il convient alors de compléter le modèle pour ne pas laisser hors du graphique de Daniel l'unique contraste utilisé pour estimer l'écart-type résiduel.

On refait donc l'analyse de variance avec le modèle $P.P.P$ incluant toutes les interactions de trois facteurs, ou plutôt le modèle équivalent $P + P.P + P.P.P$ qui fait apparaître d'abord les f.e.b contenant un effet principal, ensuite celles qui contiennent une interaction au moins de deux facteurs, et en dernier la f.e.b qui ne contient que des interactions de 3 facteurs.

On trouve alors les 32 f.e.b figurant à droite du tableau 19. Une seule n'inclue aucun effet principal et aucune interaction de deux facteurs, à savoir la 32^{ème} égale à la somme $ADH + BDE + CDG + DFI$ de 4 interactions de 3 facteurs. Pour la faire apparaître dans le graphique de Daniel en évitant de compliquer la lecture avec la présence de toutes les autres interactions de 3 facteurs, il suffit d'ajouter une de ces interactions de trois facteurs, par exemple ADH au modèle $P.P$.

La figure 13 donne le graphique des quantiles et les données complémentaires correspondantes, obtenus avec ce modèle $P.P + ADH$. Les effets A (A), DE (Z), D (F) s'écartent vers le haut de façon assez nette. L'écart est moins net pour les effets AD (G) et $AH + BE + CG + FI$ (O). Si on refait le graphique en éliminant 3 points, on obtient le graphique 14 où les points sont sensiblement alignés. La régression sans constante du quantile empirique Q_{emp} sur le quantile théorique Q_{th} donne une pente 0.0344, et on peut donc estimer l'écart-type résiduel par $\hat{\sigma} = \sqrt{32} \times 0.0344 = 0.195$. Si on refait l'analyse de variance avec le modèle $A + D.E$ incluant les effets trouvés *actifs* dans le premier graphique de Daniel, on trouve un écart-type résiduel de 0.193 très voisin.

Dans les deux graphiques de Daniel, l'effet AD s'écarte moins nettement de l'alignement créé par les autres points. Cependant, il fait intervenir deux effets principaux A et D qui apparaissent actifs. On peut donc raisonnablement penser que l'interaction $A.D$ ne doit pas être négligée et continuer l'interprétation (étude des moyennes) avec le modèle $A.D + D.E$. L'écart-type résiduel estimé avec ce modèle est 0.1735. Il a 26 degrés de liberté. Le test de l'interaction $A.D$ apparaît significatif presque au seuil 1%, mais ce seuil n'a qu'une valeur indicative car la théorie classique conduisant aux seuils de signification n'est pas apte à prendre en compte une sélection du modèle utilisant les résultats de l'expérience. On notera que si l'on refait le graphique de Daniel en éliminant 4 points, donc l'effet $A.D$ en plus des effets A , $D.E$, D , la pente de la droite trouvée conduit à un écart-type estimé 0.175 très proche de l'écart-type résiduel du modèle $A.D + D.E$.

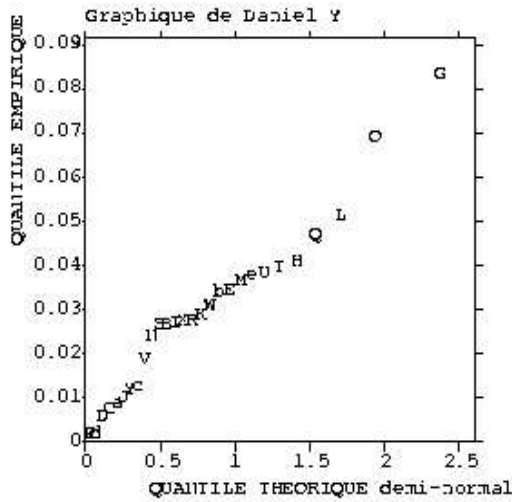


Complément au graphique de Daniel, var Y.

Modèle: $P.P+A.D.H$, où P est $A+B+C+D+E+F+G+H+I$.

symb	effet	Q_{emp}	Prob	Q_{th}
A	+A	0.21938	0.9839	2.4064
Z	+D.E	0.15063	0.9516	1.9744
F	+D	0.12188	0.9194	1.7473
G	+A.D	0.08313	0.8871	1.5856
O	+A.H + B.E + C.G + F.I	0.06875	0.8548	1.4571
L	+G	0.05125	0.8226	1.3490
Q	+A.I + F.H	0.04688	0.7903	1.2546
H	+E	0.04063	0.7581	1.1703
T	+B.F + E.I	0.03938	0.7258	1.0935
U	+B.G + C.E	0.03813	0.6935	1.0227
e	+A.D.H	0.03750	0.6613	0.9567
M	+A.G + C.H	0.03625	0.6290	0.8945
E	+A.C + G.H	0.03438	0.5968	0.8357
b	+D.G	0.03375	0.5645	0.7796

FIG. 13 – graphique de Daniel sur ensemble des points, exemple 8.2



Complément au graphique de Daniel, var Y.

Modèle: $P.P+A.D.H$, où P est $A + B + C + D + E + F + G + H + I$.

Les 3 + grands contrastes en valeur absolue ne sont pas pris en compte dans le graphique.

symb	effet	Qemp	Prob	Qth
G	+A.D	0.083125	0.9821	2.3690
O	+A.H + B.E + C.G + F.I	0.068750	0.9464	1.9307
L	+G	0.051250	0.9107	1.6996
Q	+A.I + F.H	0.046875	0.8750	1.5344
H	+E	0.040625	0.8393	1.4029
T	+B.F + E.I	0.039375	0.8036	1.2920
U	+B.G + C.E	0.038125	0.7679	1.1950
e	+A.D.H	0.037500	0.7321	1.1081
M	+A.G + C.H	0.036250	0.6964	1.0288
E	+A.C + G.H	0.034375	0.6607	0.9555
b	+D.G	0.033750	0.6250	0.8870

FIG. 14 – graphique de Daniel après suppression de 3 points, exemple 8.2

8.2 Analyse de Box et Meyer

On l’obtient en cochant la case “*Méthode de Box et Meyer*” au bas de l’écran de la figure 2. Cette méthode est décrite de façon synthétique dans [2] et de façon détaillée dans [1]. Elle vise à déterminer les facteurs “actifs” en associant un modèle à chaque sous-ensemble de facteurs actifs possible et en déterminant la probabilité a posteriori de ce modèle pour les observations effectuées.

Pour déterminer l’analyse effectuée, on doit préciser

- le modèle et le sur-modèle,
- les facteurs considérés a priori comme actifs,
- le couple de paramètres γ, π utilisé pour définir la distribution a priori.

La figure 15 donne un exemple d’analyse avec 4 facteurs traitement A, B, C, D et un facteur bloc, exemple qui sera repris dans ce qui suit.

L’écran représenté figure 16 montre comment définir l’analyse de Box et Meyer présentée dans [2], §3.3.4, tableaux 3.12 et 3.13. Le plan analysé est une fraction 2^{9-4} de résolution 4. Les facteurs étudiés sont relatifs aux conditions de “revivification” des bactéries stressées par un traitement thermique. L’objectif de cette revivification est de permettre la croissance de ces bactéries sur boîte de pétri et de pouvoir ainsi les dénombrer.

L’analyse bayésienne apporte un atout appréciable pour analyser les plans fractionnaires sans répétition. Néanmoins elle peut être impuissante à distinguer entre des effets totalement confondus par le plan d’expérience. C’est donc probablement avec les plans non réguliers qu’elle a le plus de chance de donner des informations utiles et ceci, à condition que les facteurs actifs soient peu nombreux.

Lorsqu’il y a beaucoup d’effets actifs pouvant interagir, aucune méthode d’analyse ne peut apporter de conclusions fiables si le plan est de résolution trop faible et ne comporte pas de répétitions.

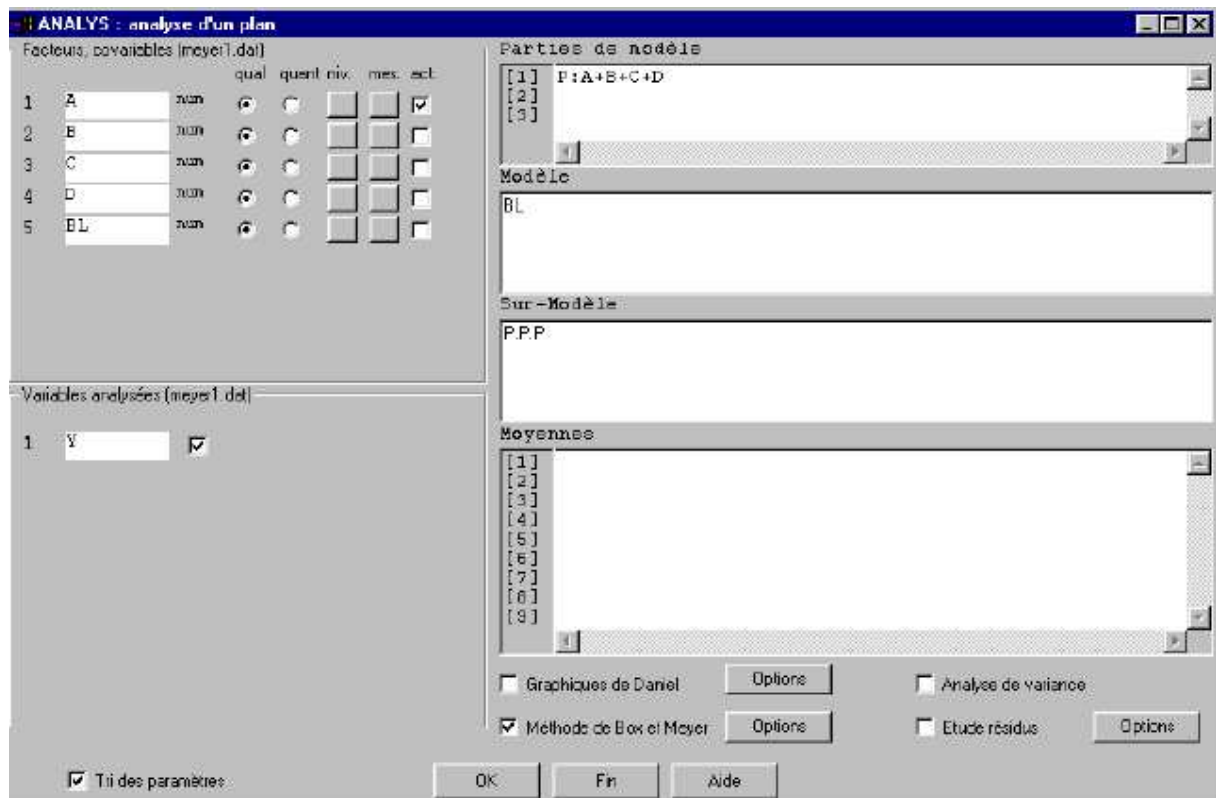


FIG. 15 – Etude de Box et Meyer d'un plan $2^3 \times 4$ en 4 blocs

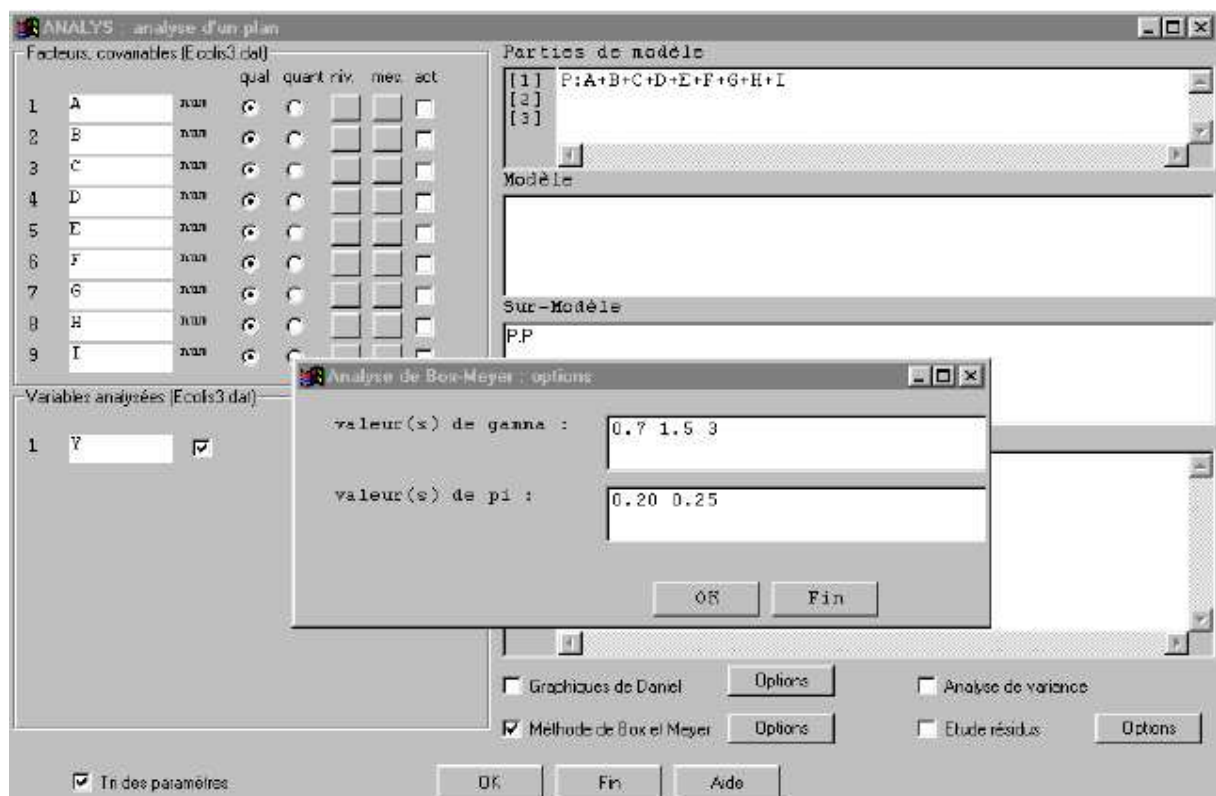


FIG. 16 – Etude de Box et Meyer du plan pour la revivication des bactéries de [2]

Les paragraphes suivants précisent comment est définie l'analyse, à partir des éléments reportés dans des écrans tels que ceux des figures 15 et 16.

8.2.1 Modèle et sur-modèle

Ces deux zones définissent les modèles dont on va déterminer les probabilités a posteriori.

La partie “*modèle*” est systématiquement prise en compte. Dans la figure 15, cette partie contient l'unique effet principal BL du facteur bloc qui est ainsi incorporé dans tous les modèles qui vont être examinés.

Les termes de la partie sur-modèle au contraire ne sont pas inclus systématiquement. Leur inclusion dépend de l'ensemble de facteurs actifs considéré. Sont inclus les termes dont tous les facteurs sont actifs. Tout terme d'interaction ou effet principal où apparaît un facteur non actif est donc supposé nul.

Dans la figure 15, il y a 4 facteurs traitements A, B, C, D . Le sur-modèle est $P.P.P$ où P , défini dans la zone “*partie de modèle*”, est la somme $A + B + C + D$. Ce sur-modèle comprend toutes les interactions entre les 4 facteurs à l'exception de l'interaction $A.B.C.D$ des 4 facteurs. Les $16 = 2^4$ ensembles possibles de facteurs actifs que l'on peut former avec ces 4 facteurs traitement figurent dans la colonne gauche du tableau 20. La

fact.actifs	modèle	ens. avec A
{A }	BL+A	←
{B }	BL+B	
{C }	BL+C	
{D }	BL+D	
{A,B }	BL+A.B	←
{A,C }	BL+A.C	←
{A,D }	BL+A.D	←
{B,C }	BL+B.C	
{B,D }	BL+B.D	
{C,D }	BL+C.D	
{A,B,C }	BL+A.B.C	←
{A,B,D }	BL+A.B.D	←
{A,C,D }	BL+A.C.D	←
{B,C,D }	BL+B.C.D	
{A,B,C,D }	BL+A.B.C+A.B.D+A.C.D+B.C.D	←

TAB. 20 – Ensembles de facteurs actifs considérés dans l'exemple de la figure 15

seconde colonne de ce tableau donne les modèles associés. Ces modèles ne sont écrit ici qu'avec les termes les plus complets, puisque comme indiqué au paragraphe 3.6.1, ils sont automatiquement complétés par les sous termes. Ainsi le modèle $BL + A.B.C$ est équivalent au modèle $BL + A + B + C + B.C + A.C + A.B + A.B.C$.

8.2.2 Facteurs a priori actifs

La connaissance préalable des phénomènes où des analyses préliminaires peuvent conduire à déclarer que certains facteurs sont a priori actifs dans l'analyse de Box et Meyer. Cela est indiqué en cochant la case de la colonne “act” associée à ces facteurs dans l'écran de la figure 2. Dans l'exemple de la figure 15, le facteur A est pris a priori comme actif. En conséquence les seuls ensembles de facteurs actifs à considérer sont ceux qui sont définis par \leftarrow dans la troisième colonne du tableau 20.

Dans les sorties, il est rappelé que le facteur A est a priori actif et donc inclu dans tous les modèles. Mais sa présence est ensuite sous-entendue et les groupes de facteurs actifs indiqués sont B, C, D, BC, BD, CD, BCD et le groupe vide. Le modèle associé au groupe vide est $(BL+A)$, celui associé à B est $(BL+AB)$, etc ...

8.2.3 Paramètres de la distribution a priori

La distribution a priori est déterminée en deux étapes.

1. Tout d'abord, on précise les probabilités a priori des différents modèles examinés. Box et Meyer ont choisi une distribution dépendant exclusivement du nombre de facteurs actifs associé au modèle. Si f est ce nombre, il propose de prendre une probabilité a priori proportionnelle à

$$\pi^f (1 - \pi)^{h-f},$$

où h est le nombre total de facteurs sur lesquels porte la recherche et donc 2^h le nombre total de modèles considéré. Comme il y a C_h^f sous-ensembles de f facteurs actifs possibles, la probabilité a priori d'avoir f facteurs est

$$P(f) \propto C_h^f \pi^f (1 - \pi)^{h-f}$$

et f suit a priori la loi binomiale de paramètre π .

Le paramètre π est donc l'espérance a priori de la proportion f/h . Ceci donne une indication pour son choix. On le prend égal à ce que l'on croit initialement être cette proportion. Naturellement, on peut faire plusieurs choix pour examiner dans quelle mesure la distribution a posteriori dépend du choix initial.

On notera que h , nombre total de facteurs inclus dans la recherche, ne prend pas en compte les facteurs systématiquement actifs cochés dans la colonne *act* (figure 2).

2. Chaque modèle se met sous la forme

$$y = X_0\beta_0 + X_1\beta_1 + \varepsilon \quad \text{où } \text{var}(\varepsilon) = \sigma^2.$$

Le vecteur β_0 comprend les paramètres associé à la zone modèle (dont la constante) qui figurent systématiquement. Le vecteur β_1 est formé des paramètres du sur-modèle retenus et qui ne figurent pas déjà dans la zone modèle.

Conditionnellement au modèle examiné, on fait a priori l'hypothèse que β_0 est distribué uniformément sur l'espace \mathbb{R}^{p_0} dans lequel il varie, tandis que β_1 a une loi normale centrée de matrice de variance $\gamma^2\sigma^2\mathbf{I}$. Le paramètre γ^2 est appelé paramètre d'inflation de la variance. L'information a priori sur β_1 pèse d'autant moins que γ est choisi plus grand.

En pratique, Box et Meyer suggère de comparer plusieurs valeurs de γ et de choisir la valeur qui donne au modèle réduit à la constante la probabilité a posteriori minimum. Il s'avère que cette valeur est aussi la valeur la plus probable a posteriori de γ si on suppose a priori γ localement uniformément distribué.

Références

- [1] Box G.E.P and Meyer R.D. (1993). Finding the active factors in fractionated screening experiments. *J. of Quality. Technology*, **25**, n° 2, 94-105.
- [2] Kobilinsky A. (1997). Les plans factoriels. Chap 3, p69–209. *In : Plans d'expériences : applications à l'entreprise*, Eds : Dreesbeke, Fine, Saporta. Technip, Paris. 509p.
- [3] Kobilinsky A. (2000). Reparametrisation of interest in non uniform factorial designs. To appear in *Linear Algebra and Applications*