REVUE DE STATISTIQUE APPLIQUÉE

JEAN LESAVRE

Régression quadratique dans les dispositifs expérimentaux utilisés en agronomie

Revue de statistique appliquée, tome 16, nº 3 (1968), p. 75-97

<http://www.numdam.org/item?id=RSA_1968__16_3_75_0>

© Société française de statistique, 1968, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (http://www.sfds.asso.fr/publicat/rsa.htm) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

RÉGRESSION QUADRATIQUE DANS LES DISPOSITIFS EXPÉRIMENTAUX UTILISÉS EN AGRONOMIE

par Jean LESAVRE

Ingénieur de Recherches à l'UVEX
Compagnie de Saint-Gobain
Chalon-sur-Saône

Il y a une dizaine d'années nous travaillions en collaboration avec les agronomes du Groupe Saint-Gobain. Cette collaboration, saisonnière et épisodique, qui a laissé chacun sur son terrain : l'agronome sur ses champs d'expérience, et le statisticien sur les chiffres que le premier voulait bien lui soumettre, nous avait déjà permis quelques fructueuses réflexions. Et comme les mathématiques couvrent le vaste domaine des sciences, il s'est trouvé que des méthodes mises au point pour l'agronomie se sont révélées applicables au domaine verrier.

De ce dialogue entre spécialistes sont nés les plans d'expérience barycentrés -[1]- Une critique de la méthode des blocs, et la constatation de l'existence fréquente d'un effet linéaire dû très certainement au terrain et non aux traitements, nous ont conduit, dans cette étude publiée en 1965, à la mise au point de plans aussi peu aléatoires que possible, puisqu'ils sont barycentrés - ni plus ni moins que les carrés latins.

Or, chose curieuse, si les carrés latins sont admis par tous, on préfère assez généralement, dès que le dispositif perd sa structure carrée, se confier au hasard plutôt qu'à la systématique d'une règle, or le barycentrage n'est qu'un reflet atténué de la "latinité" des carrés latins, puisque dans chaque colonne et dans chaque ligne on trouve une fois et une seule chaque traitement. Ce type de combinatoire impose comme conséquence le barycentrage, mais la réciproque n'est pas vraie

Sur cette querelle des "tireurs au sort" et des "systématiques", se greffe une plainte des expérimentateurs : "vos analyses de variance ne sont pas assez fines, il nous faudrait un ajustement plus étroit aux chiffres souvent si péniblement collectés" ; bref, n'y a t-il pas moyen de diminuer cet écart-type résiduel qui est l'étalon de mesure de ces différences, qui nous paraissent, à nous, bien suffisamment significatives, et que vous trouvez, vous, statisticiens, si peu souvent significatives, à votre niveau de probabilité de 5 %" ?

Le statisticien, ou plutôt le mathématicien, va répondre : "Vous n'êtes pas satisfaits de l'ajustement linéaire, eh bien ! passons à l'échelle au-dessus, c'est-à-dire tentons sur vos chiffres un ajustement quadratique". Ce n'est pas qu'en passant du 1er au second degré on ait trouvé là une formule magique, c'est simplement que dans l'ordre des complications c'est la plus simple qui se présente à nous : après le plan horizontal, si j'ose dire, c'est-à-dire hypothèse du terrain homogène et randomisation totale, nous trouvons le plan de pente quelconque et la régression linéaire, c'est-à-dire hypothèse d'un gradient de fertilité constant, puis aujourd'hui la quadrique et donc la régression quadratique, c'est-à-dire une hypothèse assez difficile à préciser pour le profane, mais qui assimile la surface des rendements du champ à une surface du second degré.

Ces deux dernières hypothèses, de linéarité et de quadraticité, nous inclineront à faire choix de plans barycentrés, dont nous savons qu'ils existent, et même de plans barycentrés qui seraient aussi quadratiquement orthogonaux. Ces derniers vocables désignent une réalité assez claire pour nous, mais il est infiniment peu probable que de tels plans existent. Le barycentrage a l'avantage de nous mettre à l'abri des variations linéaires rencontrées dans le champ : les américains - [2]- diraient que le barycentrage permet d'obtenir un plan "robuste" contre les effets linéaires. Nous doutons que des plans puissent être totalement robustes contre des effets quadratiques.

Cette méthode de la régression quadratique, sur laquelle nous fondons les plus grands espoirs, arrive à son heure pour remplacer l'analyse de variance classique, qui demande un plan très strict, mais peu de calcul dans le dépouillement des résultats. Cette nouvelle application d'une méthode par ailleurs bien classique se contentera d'un plan assez quelconque, mais exigera des calculs absolument inabordables au pauvre calculateur manuel, même armé d'une machine de bureau. Une calculatrice électronique est indispensable : songez qu'il faudra inverser une matrice, symétrique sans doute, mais d'ordre 10 au minimum. Bien sûr l'astuce mathématique rejoindra le sens agronomique pour essayer de rendre cette matrice aussi facile à inverser que possible, c'est-à-dire à la diagonaliser au maximum. Voyons cela d'un peu plus près.

PRINCIPE DE LA REGRESSION QUADRATIQUE

Nous supposons que le champ est rapporté à deux axes Ox et Oy. Pour simplifier, nous supposerons que les parcelles limitées par des droites parallèles aux axes ont un rendement uniforme. Nous nous rendons bien compte de la fragilité de cette hypothèse, mais il faut bien s'en contenter. Sinon on serait amené à considérer le rendement d'une parcelle comme l'intégrale d'une certaine surface, ce qui n'est pas impossible avec le calcul électronique, mais laissons cette complication pour l'instant.

Chaque centre de parcelle sera caractérisé par 3 paramètres :

1/ le traitement de cette parcelle,

2/ son abscisse,

3/ son ordonnée,

et un résultat, par exemple le poids récolté dans la parcelle.

Bien entendu, toutes les parcelles sont d'égale superficie. Mais c'est une facilité qu'on se donne, cela n'est pas indispensable. Si les parcelles étaient légèrement inégales, il suffirait, par exemple, de diviser le poids récolté par la surface de la parcelle, donc de ramener au m^2, ou à l'are. Il ne faudrait pas, bien sûr, que cette inégalité soit trop importante, car on introduirait de nouvelles causes d'hétérogénéité.

Bien entendu, il sera admis que l'ensemble des parcelles forment un champ rectangulaire, limité par des parallèles aux axes Ox et Oy. Mais là encore c'est une facilité qu'on se donne, le champ pourra avoir une forme quelconque, sous la réserve que ses limites restent des segments de droite parallèles aux Ox et Oy.

Il sera encore admis que les traitements sont en nombres égaux, mais ce n'est pas nécessaire, les répétitions d'un même traitement pourront être différentes d'un traitement à l'autre.

Bref, avant même de préciser ce qu'est la régression quadratique, nous voyons que toutes les contraintes usuelles qui pèsent si lourdement sur les essais agronomiques sont déjà considérablement assouplies, pratiquement anéanties. Ce sont donc des avantages indirects offerts par cette méthode de régression, ou si l'on analyse mieux la chose, ces avantages sont en fin de compte procurés par les possibilités extraordinaires du calcul électronique. Il faut tout de suite remarquer que la simple régression linéaire offrirait déjà à peu près les mêmes avantages, mais qu'en fait, elle n'est pas encore entrée dans la pratique courante. Espérons que nous n'allons tout de même pas trop loin et que la méthode de la régression linéaire, et mieux encore celle de la régression quadratique, feront partie très prochainement de l'arsenal des méthodes courantes de dépouillement de résultats agronomiques.

Choix de la formule quadratique à ajuster

En effet un choix est nécessaire, il correspond à diverses hypothèses qu'on peut faire sur le comportement de nos traitements. Dans un premier temps, nous allons faire l'hypothèse que les traitements, et le témoin s'il y a lieu, réagissent de la même façon au terrain : les traitements auront bien sûr des différences entre deux, mais le terrain les marquera tous de la même façon, et il n'y aura qu'une seule formule quadratique pour tous les traitements. Nous ajusterons donc une formule du type :

$$z = z_T + ax + by + cx^2 + dy^2 + exy + \text{résidu}$$

Ce résidu sera gaussien, de moyenne nulle et on cherchera à en minimiser la variance

En pratique cette formule d'ajustement se révèle assez peu fructueuse, et l'on se voit contraint de prendre certaines précautions :

1/ x et y seront des variables centrées, équidistantes de l'unité - c'est-à-dire qu'on préfère travailler sur des unités réduites, ou normalisées, plutôt que sur des longueurs ou largeurs en mètres, ou décamètres par exemple. Ainsi si (x) est à 3 niveaux, on a : -1, 0 et +1
si (y) est à 4 niveaux, on a : -3/2, -1/2, +1/2 et +3/2
Si le plan n'est pas classique, ou si certaines données d'un plan classique sont manquantes, on ne cherchera pas à centrer parfaitement les variables (x) et (y), on se contentera d'un centrage approximatif, mais on conservera soigneusement l'équidistance de 1,0.

2/ x^2 et y^2, représentent les carrés de variables réduites et généralement centrées. Là encore on préfère travailler sur des variables centrées ; visiblement x^2 et y^2, des carrés toujours positifs ne sont pas centrés. On utilisera un décalage pour opérer ce centrage : cette méthode est classique, c'est celle des polynômes orthogonaux de Tchebycheff-Fisher. La formule qu'on va chercher à ajuster sera du type :

$$z = z_T + ax + by + c\,(x^2 - \alpha^2) + d\,(y^2 - \beta^2) + e\,xy + \varepsilon_R$$

x et y sont des variables centrées et normalisées.

α^2 dépend du nombre de colonnes du plan et

β^2 dépend du nombre de lignes du plan. Ils sont parfaitement connus et fixés dès que le plan est fixé. La formule dépend donc de 5 paramètres Comme on doit déterminer en même temps n traitements, il faut donc que les essais soient assez nombreux pour permettre la détermination de (n + 5) quantités. Ce sera généralement le cas. Nous voyons tout de suite une restriction : nous excluons implicitement tous les plans qui ne comporteraient que 2 lignes ou 2 colonnes, car alors on sera dans l'impossibilité de calculer un effet "carré" ; dans ce cas nos paramètres (c) ou (d) ne pourraient être que nuls, et d'aucune utilité dans l'ajustement. Bref nous commençons avec un plan comportant au minimum 9 essais : 3 lignes et 3 colonnes ; s'il y a, par exemple, 3 traitements répétés 3 fois, la solution est possible puisque : 3 + 5 = 8, mais la résiduelle n'aura qu'un seul degré de liberté, ce qui est peu.

Dans un deuxième temps nous voudrions ne faire aucune hypothèse sur les réactions traitements - terrain. Chaque traitement, éventuellement le témoin, serait représenté par une certaine formule quadratique particulière. On aurait donc n formules :

$$z = z_T + a_T x + b_T y + c_T (x^2 - \alpha^2) + d_T (y^2 - \beta^2) + e_T xy + \varepsilon_R$$

Nous voyons donc que cela fait 6 paramètres par traitements : il faudrait donc au moins 6 répétitions par traitements. Nul doute que les agronomes ne trouvent le procédé beaucoup trop coûteux. On peut cependant faire quelque chose dans cette voie. Du fait que des répétitions inégales sont maintenant possibles, sans complication supplémentaire, on peut répéter le témoin mettons une dizaine de fois, et ajuster une formule quadratique pour le témoin. Les traitements répétés un nombre trop faible de fois, pourront être groupés en blocs d'importance suffisante pour ajuster une formule quadratique. On peut espérer alors comparer ces diverses formules quadratiques, et fixer un peu les réactions traitements-terrain.

Dans le but de ne pas trop alourdir notre texte nous ne donnerons qu'un exemple de régression quadratique. Il sera traité in extenso, et nous espérons ainsi être dispensé de faire la théorie générale, tout en donnant des indications pratiques bien suffisantes pour qu'on puisse en faire l'application à des cas nouveaux. En fait, cet exemple ne sera que faiblement convaincant, mais nous ne pouvons qu'insister sur la souplesse de cette régression quadratique, qui peut s'affranchir en fait de toutes les contraintes classiques d'une planification d'essais agronomiques, répartition en blocs, régularité du plan, etc... On pourra en effet, avec cette méthode, envisager une randomisation totale, si chère à tant d'agronomes et de statisticiens, au détriment toutefois de la rigueur dans l'analyse de variance.

Bref, la nouvelle contrainte sur le plan d'expériences sera légère et assez subtile : il faudrait, et ce n'est même pas nécessaire, rendre aussi diagonale que possible une certaine matrice qu'on a la nécessité d'inverser. Par contre même si finalement nous allons, après régression quadratique, aboutir à une analyse de variance encore assez voisine de celle que nous avons l'habitude de faire, il n'est plus possible de faire des calculs "à la plume". Si l'agronome utilisait déjà le calcul électronique à cause du volume de ses calculs, il lui est maintenant indispensable à cause de la difficulté intrinsèque du dépouillement du plan le plus élémentaire.

EXEMPLE DE REGRESSION QUADRATIQUE

C'est l'exemple déjà précédemment analysé dans l'article de la Revue de Statistique Appliquée - 1965 - Vol. XIII - N° 1 - à la page 108

Il s'agit d'une culture d'orge - la parcelle est de 2,16 ares.

- les résultats (tableau B) sont en Kg - une légère modification des chiffres originaux nous a permis de les rendre tous multiples de 0,28 (pure simplification de calcul)

- le plan (tableau A) est un 7 × 4 barycentré, faisant intervenir un témoin T et six traitements, qui sont des engrais.

- le tableau N nous fournit une numérotation des essais pour qu'on puisse s'y retrouver.

(N)

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16
17	18	19	20
21	22	23	24
25	26	27	28

(A)

T	+N	NPK	+P
NP	+K	+P	NK
NK	T	+K	NP
NPK	NP	+N	+K
+N	NK	T	NPK
+P	NPK	NK	+N
+K	+P	NP	T

(B)

37,24	55,72	40,04	44,80
39,48	45,08	49,28	47,04
44,24	38,64	49,84	48,72
49,00	50,68	56,28	55,16
52,36	52,36	42,84	52,92
51,80	47,88	51,52	55,44
47,32	52,92	52,92	44,80

les abscisses des quatre colonnes sont : - 1,5, - 0,5, + 0,5, + 1,5
et $(x^2 - 1,25)$ donne : + 1,0, - 1,0, - 1,0, + 1,0
De même les ordonnées sont : - 3, - 2, - 1,0 + 1, + 2, + 3
et $(y^2 - 4,00)$ donne : + 5, 0, - 3, - 4, - 3, 0, + 5

On a maintenant la possibilité d'établir la matrice X et le vecteur colonne Y (ici ce serait plutôt Z ou B).

Cette matrice X est classique dans les problèmes de régression. Elle comporte 28 lignes, c'est-à-dire autant que de résultats - 7 colonnes pour les traitements + 5 colonnes pour la régression quadratique, soit 12 colonnes au total. Nous pensons qu'il est tout de même utile de la donner in extenso. (Voir page 80).

Elle se compose de deux parties : la 1ère partie, composée de 7 colonnes, est relative aux traitements. Elle comporte des 1 et des 0 - pour plus de simplicité les 0 ne sont pas figurés - la 2ème partie est formée des abscisses et ordonnées de leurs carrés recentrés par décalage, et de leur produit (xy) terme d'interaction.

L'étape suivante du calcul consiste à établir la matrice des moments : c'est le produit de la matrice X par sa transposée X', soit (X'X). Ici cette matrice est d'ordre 12, elle est symétrique (voir page 81). Cette matrice n'est pas exactement la matrice des moments centrés classique. Mais elle se comporte et se présente exactement de la même manière. En effet les traitements sont représentés par des 1 (présence) ou des 0 (absence), cette variable du genre booléen n'est pas centrée, c'est évident. Mais du fait qu'une parcelle ne peut contenir à la fois plusieurs traitements, le produit de deux vecteurs "traitements" sera toujours nul. Ou si l'on veut le 1 est exclusif et impose des 0 dans tout le reste de la ligne, pour les autres traitements.

Matrice X — Vecteur

	+N	+P	+K	NK	NP	NPK	T	x	y	x^2 - 1,25	y^2 - 4	xy	Z
1							1	-1,5	-3	+1	5	+ 4,5	37,24
2	1							-0,5	-3	-1	5	+1,5	55,72
3						1		+0,5	-3	-1	5	-1,5	40,04
4		1						+1,5	-3	+1	5	-4,5	44,80
5					1			-1,5	-2	+1	0	+3,0	39,48
6			1					-0,5	-2	-1	0	+1,0	45,08
7		1						+0,5	-2	-1	0	-1,0	49,28
8				1				+1,5	-2	+1	0	-3,0	47,04
9				1				-1,5	-1	+1	-3	+1,5	44,24
10							1	-0,5	-1	-1	-3	+0,5	38,64
11			1					+0,5	-1	-1	-3	-0,5	49,84
12					1			+0,5	-1	+1	-3	-1,5	48,72
13						1		-1,5	0	+1	-4	0	49,00
14					1			-0,5	0	-1	-4	0	50,68
15	1							+0,5	0	-1	-4	0	56,28
16			1					+1,5	0	+1	-4	0	55,16
17	1							-1,5	1	+1	-3	-1,5	52,36
18				1				-0,5	1	-1	-3	-0,5	52,36
19							1	+0,5	1	-1	-3	+0,5	42,84
20						1		+1,5	1	+1	-3	+1,5	52,92
21		1						-1,5	2	+1	0	-3,0	51,80
22						1		-0,5	2	-1	0	-1,0	47,88
23				1				+0,5	2	-1	0	+1,0	51,52
24	1							+1,5	2	+1	0	+3,0	55,44
25			1					-1,5	3	+1	5	-4,5	47,32
26		1						-0,5	3	-1	5	-1,5	52,92
27					1			+0,5	3	-1	5	+1,5	52,92
28							1	+1,5	3	+1	5	+4,5	44,80

Somme des carrés de tous les termes du vecteur colonne Z = 66.491,1968
terme correctif = 65.700,1408
Somme générale des carrés = 791,0560

pour : 27 Degrés de Liberté
Total = 1.356,32
Moyenne = 48,44

Matrice $(X'X)$

	+N	+P	+K	NK	NP	NPK	T	x_L	y_L	x_Q	y_Q	$x_L y_L$
+N	4							0	0	0	-2	3
+P		4						0	0	0	10	-10
+K			4					0	0	0	-2	-4
NK				4				0	0	0	-6	-1
NP					4			0	0	0	-2	3
NPK						4		0	0	0	-2	-1
T							4	0	0	0	4	10
x	0	0	0	0	0	0	0	35	0	0	0	0
y	0	0	0	0	0	0	0	0	112	0	0	0
x^2	0	0	0	0	0	0	0	0	0	28	0	0
y^2	-2	10	-2	-6	-2	-2	4	0	0	0	336	0
xy	3	-10	-4	-1	3	-1	10	0	0	0	0	140

Donc cette matrice $X'X$ se décompose en une 1ère matrice, diagonale, ne comportant que des 4 dans la diagonale (4 est en effet le nombre constant des répétitions de chaque traitement) - en une 2ème matrice le plus souvent diagonale ; si le plan est rectangulaire et complet, cette matrice est diagonale ; ici elle comporte les termes 35, 112, 28, 336 et 140 et des zéros partout ailleurs - et enfin en deux matrices rectangles (t, 5) (5, t) symétriques, dont les termes seront généralement différents de 0. Ici la matrice s'abaisse à l'ordre 9. En effet du fait que nous avons des blocs de 7 traitements : les colonnes, l'effet linéaire (x) et l'effet carré (x^2 - 1,25) sont bien orthogonaux aux traitements. Un certain effet cubique le serait aussi : nous le laissons se fondre dans la résiduelle. Du fait du barycentrage l'effet linéaire (y) est aussi orthogonal aux traitements. Voilà donc le faible avantage apporté par le barycentrage.

Au point de vue du calcul, on ne peut même pas dire que la méthode des blocs avec traitements barycentrés apporte grand'chose : de toutes façons la matrice $(X'X)$ est trop difficile à inverser à la plume, et il faut passer au calcul électronique. Par contre au point de vue conception du plan et réalisation concrète de l'essai, il ne nous paraît pas indifférent de savoir qu'ici un effet linéaire quelconque selon les (x) et les (y) et qu'un effet carré selon les (x) n'ont aucune influence sur la détermination des traitements (et réciproquement). Cette orthogonalité des traitements et des propriétés du terrain est tout de même précieuse, puisque les uns n'ont pas la possibilité de biaiser les autres ; mais on voit bien qu'elle n'est pas nécessaire. C'est un pur souci d'élégance et de clarté qui nous fait préférer ces plans équilibrés et barycentrés à des plans randomisés.

Supposons maintenant que nous fassions encore un plan $7 \times 4 = 28$ parcelles, que nous utilisions une randomisation totale, et que nous ayons la malchance de devoir négliger le résultat (certainement) aberrant ou manquant d'une parcelle, que devient la matrice $(X'X)$?

La 1ère sous-matrice considérée, celle des 7 traitements conservera sa structure diagonale, mais l'un des 4 de la diagonale sera remplacé par un 3, pour la répétition manquante d'un des traitements. La 2ème sous-matrice, celle de l'ajustement quadratique, ne sera plus diagonale du fait de la parcelle manquante. Les deux matrices rectangles (7, 5) (5, 7) ne comporteront pratiquement plus de terme nul. On aura donc une matrice que nous ne pouvons pas nous empêcher de trouver quelque peu cafouilleuse, mais dont effectivement l'inversion ne va pas présenter de difficultés particulières, sauf malchance extraordinaire due à une randomisation par trop singulière.

Enfin dernière question sur cette matrice $(X'X)$, peut-on espérer rendre les matrices (t, 5) (5, t) identiquement nulles ? Autrement dit, peut-on espérer trouver des plans, des dispositions sur le terrain, tels les traitements soient orthogonaux aux effets linéaires, aux effets quadratiques et au produit des deux effets linéaires ?

Les plans en carrés latins, du fait des blocs rencontrés tant dans les lignes que dans les colonnes, répondent bien partiellement à la question : les traitements y sont bien orthogonaux aux effets linéaires et aux effets quadratiques proprement dits, mais leur point faible est bien connu : ils ne rendent pas compte d'une interaction possible entre les facteurs ligne et colonne. Mais compte tenu de cette remarque on peut tenter un dépouillement beaucoup plus fin des carrés latins.

DÉPOUILLEMENT AFFINÉ DES CARRÉS LATINS

Là encore nous ne chercherons pas à faire la théorie dans le cas général, mais nous allons donner in extenso un exemple de carrés latins, ce qui permettra à la fois de saisir la méthode et d'appliquer cette nouvelle méthode à des cas particuliers.

Le carré latin considéré n'est pas un carré latin agronomique. Il s'agit d'essais de fluorescence X sur une plaque de verre, mais l'analogie est évidente entre le champ et la plaque de verre, entre les coups comptés et les récoltes faites sur les parcelles. C'est un carré latin $4 \times 4 = 16$, du type bidiagonal, où les traitements (T, A, B, C) sont en fait des températures d'essai. T désignant un véritable témoin, comme dans les essais agronomiques. La numérotation, le dispositif et les résultats sont donnés dans les 3 tableaux suivants : (P) (Q) et (R).

(P)

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

(Q)

T	A	B	C
A	T	C	B
B	C	T	A
C	B	A	T

(R)

542	712	657	675
640	538	713	647
561	667	529	566
615	560	627	516

<u>Estimation des traitements</u> :

$$T = 1/4\,(542 + 538 + 529 + 516) = 531,25$$

$$A = 1/4\,(640 + 712 + 627 + 566) = 636,25$$

$$B = 1/4\ (561 + 560 + 657 + 647) = 606,25$$

$$C = 1/4\ (615 + 667 + 713 + 675) = 667,50$$

Les abscisses : x sont comptées positivement de gauche à droite et sont fixées aux valeurs conventionnelles : - 1,5, - 0,5, + 0,5, + 1,5.

Les ordonnées : y sont comptées positivement du haut vers le bas et sont fixées aux mêmes valeurs conventionnelles, équidistantes de l'unité.

Pour les traitements on les supposera toujours rangés dans le même ordre : T, A, B, C, ce qui permettra de calculer un effet linéaire, un effet quadratique, un effet cubique ; l'ensemble des 3 effets représentera la variabilité "traitements", globalement nous désignons les traitements par la lettre : z. Nous avons distingué les effets linéaires, quadratiques, et cubiques par la méthode classique des polynômes orthogonaux de Tchebycheff-Fischer, 3.

Dans cette optique la matrice X (voir page 84), habituellement considérée pour l'établissement de la régression comporte 16 lignes : les 16 éléments du carré latin - et 10 colonnes : une colonne de 1,0 (pour le calcul de la moyenne), trois colonnes pour les effets abscisses et ordonnées du carré latin soient :

	x_L	x_Q	x_C	L : Linéaire
et	y_L	y_Q	y_C	Q : Quadratique
				C : Cubique

De même les effets traitements sont au nombre de trois :

$$z_L \qquad z_Q \qquad z_C$$

Le vecteur colonne Y rappelle les données numériques déjà fournies sous la forme d'un tableau carré (4 × 4) (tableau R). Nous avons rangé ces colonnes dans un ordre insolite qui s'expliquera de lui-même un peu plus loin. La 11ème colonne contient le produit (x linéaire) par (y linéaire) ; elle est ignorée dans le dépouillement classique. La 12ème colonne est justement le vecteur Y, les résultats numériques à analyser.

Puis on établit la matrice des moments centrés, ou des produits croisés, en multipliant la matrice X par sa transposée X'. De même on calcule le vecteur colonne X'Y. Cette matrice est symétrique et d'ordre 10. Il faut calculer la matrice inverse : ici du fait du plan en carré latin et des orthogonalités qu'on rencontre, la matrice est diagonale, donc directement inversible en prenant l'inverse de chaque terme, ce qui est tout de même simple. On obtient donc en multipliant la matrice inverse par le vecteur X'Y tous les renseignements souhaitables :

La moyenne : $z_0 = 1/16 \times 9765,0 = 610,312\ 5$

puis : $x_L = 1/20 \times 93,5 = 4,675$

$x_Q = 1/16 \times (-241,0) = -15,062\ 5$

$x_C = 1/7,2 \times (-30,3) = -4,208\ 333...$

$y_L = 1/20 \times (-509,5) = -25,475$

$y_Q = 1/16 \times (43,0) = 2,687\ 5$

Matrice X (Corrélation)

	z_0	x_L	x_Q	x_C	y_L	y_Q	y_C	z_Q	z_L	z_C	$x_L y_L$	Vecteur Y
1 T	1	-1,5	1	-0,3	-1,5	1	-0,3	1	-1,5	-0,3	+2,25	542
2 A	1	-0,5	-1	+0,9	-1,5	1	-0,3	-1	-0,5	+0,9	0,75	712
3 B	1	+0,5	-1	-0,9	-1,5	1	-0,3	-1	+0,5	-0,9	-0,75	657
4 C	1	+1,5	1	+0,3	-1,5	1	-0,3	1	+1,5	+0,3	-2,25	675
5 A	1	-1,5	1	-0,3	-0,5	-1	+0,9	-1	-0,5	+0,9	0,75	640
6 T	1	-0,5	-1	+0,9	-0,5	-1	+0,9	1	-1,5	-0,3	0,25	538
7 C	1	+0,5	-1	-0,9	-0,5	-1	+0,9	1	+1,5	+0,3	-0,25	713
8 B	1	+1,5	1	+0,3	-0,5	-1	+0,9	-1	+0,5	-0,9	-0,75	647
9 B	1	-1,5	1	-0,3	+0,5	-1	-0,9	-1	+0,5	-0,9	-0,75	561
10 C	1	-0,5	-1	+0,9	+0,5	-1	-0,9	1	+1,5	+0,3	-0,25	667
11 T	1	+0,5	-1	-0,9	+0,5	-1	-0,9	1	-1,5	-0,3	+0,25	529
12 A	1	+1,5	1	+0,3	+0,5	-1	-0,9	-1	-0,5	+0,9	+0,75	566
13 C	1	-1,5	1	-0,3	+1,5	1	+0,3	1	+1,5	+0,3	-2,25	615
14 B	1	-0,5	-1	+0,9	+1,5	1	+0,3	-1	+0,5	-0,9	-0,75	560
15 A	1	+0,5	-1	-0,9	+1,5	1	+0,3	-1	-0,5	+0,9	+0,75	627
16 T	1	+1,5	1	+0,3	+1,5	1	+0,3	1	-1,5	-0,3	2,25	516
	Moyenne	Effets colonnes			Effets Lignes			Effets traitements			Inter-action	
	1	2	3	4	5	6	7	8	9	10	11	12

Matrice des moments centrés X'X

	z_0	x_L	x_Q	x_C	y_L	y_Q	y_C	z_Q	z_L	z_C	xy	VECTEUR X'Y
z_0	16										0	9765,
x_L		20									0	+93,5
x_Q			16								0	-241,0
x_C				7,2							0	-30,3
y_L					20						0	-509,5
y_Q						16					0	+43,0
y_C							7,2				0	+113,1
z_Q								16			0	-175,0
z_L									20	0	-18	+757,5
z_C									0	7,2	2,4	+271,5
xy	0	0	0	0	0	0	0	0	-18	2,4	25	-510,25

$$
\begin{aligned}
y_C &= 1/7{,}2 \times 113{,}1 = 15{,}708\ 333\ldots \\
z_L &= 1/20 \times 757{,}5 = 37{,}875 \\
z_Q &= 1/16 \times (-175{,}0) = -10{,}937\ 5 \\
\text{et :}\quad z_C &= 1/7{,}2 \times 271{,}5 = 37{,}708\ 333\ldots
\end{aligned}
$$

De même on peut faire l'analyse de variance correspondant à cette régression : ici elle est particulièrement simple et la matrice des carrés, ou des variances, se réduit aux termes de la diagonale principale. La somme des carrés, relative au vecteur Y se répartit entre un terme dominant, dû à la moyenne, souvent appelé terme correctif, entre $3 \times 3 = 9$ termes (abscisses, ordonnées, traitements) (linéaire, quadratique, et cubique) et un résidu qu'on obtient par différence. Ici on peut même le chiffrer directement et facilement en considérant les deux carrés latins, orthogonaux au carré latin considéré, et formant la famille complète des carrés latins orthogonaux (ici 4×4 bidiagonal).

Les termes non figurés sont nuls.

Somme des carrés	(Y)	:	6.025.021,000 0	D.d.L. = 16
terme correctif	(z_0)	:	5.959.701,562 5	= 1
4.194,6875	x_L	:	437,112 5	= 1
	x_Q	:	3.630,062 5	= 1
	x_C	:	127,512 5	= 1
14.871,6875	y_L	:	12.979,512 5	= 1
	y_Q	:	115,562 5	= 1
	y_C	:	1.776,612 5	= 1
40.842,1875	z_L	:	28.690,312 5	= 1
	z_Q	:	1.914,062 5	= 1
	z_C	:	10.237,812 5	= 1
1er carré latin orthogonal		:	2.641,187 5	= 3
2ème carré latin orthogonal		:	2.769,687 5	= 3

On pourra vérifier que le 1er terme est égal à la somme des 12 autres. L'ennui de cette méthode est l'escamotage de nos traitements T.A.B.C, dont nous n'avons pas d'estimation directe. On les rétablira par application de la formule de régression trouvée, et on pourra contrôler l'exactitude de cette formule en appliquant la méthode classique de dépouillement des carrés latins. Ainsi par exemple, nous pouvons calculer la valeur de T, en remarquant que le traitement Témoin est toujours le premier dans la séquence T, A, B, C. Ses valeurs repères sont donc :

- 1,5 pour l'effet linéaire (z_L)

+ 1,0 pour l'effet quadratique (z_Q)

et - 0,3 pour l'effet cubique (z_C)

Il ne reste plus qu'à multiplier ces valeurs par les coefficients de régression correspondant, qu'on a calculés ci-dessus.

$$\text{Donc } T = \underset{\text{moyenne}}{610{,}3125} - 1{,}5\,\underset{z_L}{(+37{,}875)} + 1{,}0\,\underset{z_Q}{(-10{,}9375)} - 0{,}3\,\underset{z_C}{(+37{,}7083.)} =$$

soit T = 531,25 qui est égal à la détermination classique : 1/4 [542,0 + 538,0 + 529,0 + 516,0].

Pour éviter toute ambiguïté, nous donnons les valeurs repères concernant les 4 traitements.

	T	A	B	C
Effet Linéaire	-1,5	-0,5	+0,5	+1,5
Effet Quadratique	+1,0	-1,0	-1,0	+1,0
Effet Cubique	-0,3	+0,9	-0,9	+0,3

Ces valeurs étaient déjà données dans la matrice X de la page 84. On y reconnait les coefficients permettant l'ajustement des polynômes orthogonaux de Tchebycheff, à cela près qu'on a divisé les coefficients entiers habituels par le coefficient généralement appelé λ, qui a précisément pour fonction de nous faire travailler sur des nombres entiers, ce qui est commode en calcul à la plume, très utile pour la précision des calculs, mais complique l'estimation des carrés emportés par la régression.

Mais l'avantage de ce calcul artificiel, et un peu plus compliqué, des traitements tient dans ce fait qu'on peut maintenant, sans aucune difficulté, introduire dans la formule d'ajustement un terme croisé en [xy], bref un terme d'interaction, qui ne saurait être totalement orthogonal à tous les autres termes déjà considérés. La matrice (page 84), aux 16 lignes, comporte une nouvelle colonne, la 11ème, formée par le produit des variables fictives $(x_L \times y_L)$. La matrice X'X des moments centrés est d'ordre 11 (page 84) mais heureusement on s'aperçoit qu'elle s'abaisse à l'ordre 3. En effet, à cause du plan en carré latin, la matrice X'X est une matrice diagonale, en ce qui regarde des dix premières colonnes : moyenne, effets lignes, effets colonnes et effets traitements seront toujours orthogonaux pour un carré latin. Il nous reste à examiner comment se situe le vecteur nouveau $(x_L\, y_L)$: il est orthogonal à la moyenne, et aux effets lignes et colonnes du plan, mais il n'a aucune raison d'être orthogonal aux effets traitements. Ici, il se trouve fortuitement que l'effet $(x_L\, y_L)$ et l'effet quadratique des traitements (z_Q) sont aussi orthogonaux, c'est pourquoi nous l'avons placé en tête des effets traitements. Seuls restent donc plus ou moins en interdépendance 3 effets $[z_L]$ $[z_C]$ et $[x_L\, y_L]$. Nous avons donc une matrice irréductible d'ordre 3 seulement (au lieu de 11).

Pour cette matrice partielle et son vecteur correspondant X'Y on applique la méthode générale d'inversion, puis de multiplication pour obtenir les coefficients de la régression.

Matrice directe irréductible

$$\begin{bmatrix} 20 & 0 & -18 \\ 0 & 7,2 & 2,4 \\ -18 & 2,4 & 25 \end{bmatrix}$$

Vecteur X'Y

$$\begin{bmatrix} +757,5 \\ +271,5 \\ -510,25 \end{bmatrix}$$

Matrice inverse

$$\begin{bmatrix} 0,15125 & -0,0375 & 0,1125 \\ -0,0375 & -0,1527.. & -0,0416... \\ 0,1125 & -0,0416.. & 0,125 \end{bmatrix}$$

Coefficients de régression

z_L = 46,9875

z_C = 34,3333...

(xy) = 10,125

Nous pouvons donc, grâce à cette nouvelle régression qui comporte le terme non orthogonal en $(x_L \, y_L)$ rétablir les valeurs de nos traitements T, A, B et C, en utilisant les valeurs repères déjà signalées. En ce qui concerne le terme d'interaction, et en se rapportant au tableau matriciel de la page 84, on voit que T est caractérisé par :

$$1/4 \; (2,25 + 0,25 + 0,25 + 2,25) = + 1,25$$

$$\text{A par : } 1/4 \; (0,75 + 0,75 + 0,75 + 0,75) = + 0,75$$

$$\text{B par : } 1/4 \; (-0,75 - 0,75 - 0,75 - 0,75) = - 0,75$$

$$\text{C par : } 1/4 \; (-2,25 - 0,25 - 0,25 - 2,25) = - 1,25$$

Ainsi pour T l'application de la nouvelle formule de régression donne :

$$T = 610,3125 - 1,5 \, (46,9875) + 1,0 \, (-10,9375) - 0,3 \, (34,33\ldots) + 1,25 (10,125)$$
$$= 531,25$$

soit :

$$T = 518,59375 + 12,65625 = 531,25$$

$$A = 628,65625 + 7,59375 = 636,25$$

$$B = 613,84375 - 7,59375 = 606,25$$

$$C = 680,15625 - 12,65625 = 667,50$$

On retrouve les estimations connues des traitements, et c'est heureux. Mais nous avons cependant fait un pas en avant, car maintenant nous pouvons dissocier dans les traitements ce qui revient aux traitements proprement dits et ce qui revient à l'interaction lignes-colonnes du terrain. Bien sûr le découpage est quelque peu arbitraire et pourra être discuté par les utilisateurs puisqu'il n'y a pas orthogonalité. Mais l'estimation affinée des traitements :

$$T = 518,59375$$

$$A = 628,65625$$

$$B = 613,84375$$

$$C = 680,15625$$

est la meilleure possible dans cette hypothèse du jeu d'une interaction lignes-colonnes.

L'analyse de variance afférente à cette régression particulière va perdre beaucoup de sa simplicité classique. Pourtant, dans une optique matricielle, l'établissement des divers carrés reste assez facile : il suffit de penser que l'on a à faire le carré d'une forme linéaire, on aura donc la somme des carrés de tous les éléments composants plus deux fois la somme de tous les produits deux à deux. Matriciellement cette élévation au carré est obtenue par une multiplication mettant en jeu un vecteur ligne, une matrice de coefficients, et le même vecteur disposé en colonne.

Le vecteur ligne (ou colonne) est le vecteur des coefficients de régression rangés dans leur ordre logique - allant ici de la moyenne générale à l'interaction $(x_L \, y_L)$.

La matrice des coefficients est symétrique, et elle nous est déjà bien connue, c'est la matrice des moments centrés $X'X$. Le résultat comporte une seule ligne (ou colonne) mais cette présentation, d'un scalaire en quelque sorte, est sans intérêt. Respectons plutôt la matrice $X'X$ des coefficients, multiplions chaque élément de chaque ligne par l'élément correspondant du vecteur des coefficients de la régression et recommençons la même opération par colonne en suivant bien entendu l'ordre logique. Ainsi chaque terme de la diagonale principale est bien multiplié par le carré d'un coefficient de régression, et les termes non diagonaux vont par paires en raison de la symétrie de la matrice des moments centrés devenue ici matrice des coefficients de la forme quadratique. On a donc, par cette méthode, un tableau imagé des carrés emportés par chaque terme de régression. Si le terme diagonal de la matrice $X'X$ est le seul non nul de sa ligne (ou de sa colonne) on retombe bien sur la méthode ordinaire de décomposition de la somme de carrés générale en somme de carrés de formes linéaires orthogonales. Cela se complique là où les formes linéaires ne sont plus orthogonales. Nous fractionnons donc le travail en deux parties :

a - pour les comparaisons orthogonales, nous retrouvons les valeurs déjà mentionnées ci-dessus :

Somme des carrés		: 6.025.021,000 0	16
T.C.	z_0	: 5.959.701,562 5	1
Effets colonnes	x_{LQC}	: 4.194, 6875	3
Effets lignes	y_{LQC}	: 14.871, 6875	3
Effet traitement	z_Q	: 1.914, 0625	1

d'où un reste à expliquer : 44.339,000 0 pour D.d.L. = 8

b - pour les comparaisons non orthogonales, on a le schéma suivant :

$$\begin{array}{c} z_L \\ z_C \\ (xy) \end{array} \begin{bmatrix} 44.156,503.125 & 0 & -8.563,471\ 875 \\ 0 & 8.487,200\ 000 & 834,300\ 000 \\ -\ 8.563,471.875 & 834,300\ 000 & 2.562,890\ 625 \end{bmatrix} \text{D.d.L} = 3$$

La somme de ces neuf termes est égale à : 39.748,250 0

D'où un résidu de : 4.590,75 pour D.d.L. = 5

Il y a un ennui qui saute aux yeux : ce sont les carrés négatifs et il y en aura automatiquement chaque fois qu'on réussira à augmenter le contraste, ici c'est essentiellement l'effet linéaire traitements qui est augmenté, et comme la somme des carrés est constante, il faut bien que s'introduisent quelque part des termes négatifs.

Bref, à la réflexion, personne ne devrait trouver choquante l'existance de ces carrés négatifs, car ce ne sont pas de vrais carrés mais des produits croisés, qui s'introduisent automatiquement dans l'élévation au carré d'une forme linéaire, où les coefficients sont soit positifs, soit négatifs. Pourquoi seraient-ils tous de même signe ?

Dans le dépouillement classique la variance résiduelle est :

5.410,875/6 = 901,8125 avec D.d.L. = 6

Dans le dépouillement quadratique la variance résiduelle est :

4.590,75/5 = 918,15 avec D.d.L. = 5

C'est généralement ainsi que se passent les choses : la variance résiduelle n'est pas abaissée notablement, et de plus, on perd un degré de liberté. Néanmoins, le test F de Fisher Snedecor est sensiblement plus net avec le dépouillement quadratique.

Ici l'effet total traitements était, pour le carré latin classique :

40.842,1875 = 28.690,3125 + 1.914,0625 + 10.237,8125

D'où F = 45,29 pour F_0 (table) = 4,76

Pour le dépouillement quadratique

54.557,765.625 = 44.156,503125 + 1914,0625 + 8.487,2000

D'où F = 59,42 pour F_0 (table) = 5,41

Le résultat n'est pas différent par l'une ou l'autre des deux méthoses, mais il n'est pas difficile d'imaginer que la méthode de l'ajustement quadratique sera ordinairement la plus sensible.

AUTRE PRESENTATION DU MEME DEPOUILLEMENT DES CARRES LATINS

Nous pouvons donner ici une autre forme au dépouillement des résultats par la méthode de l'ajustement quadratique : nous reprenons la méthode indiquée plus haut où les traitements sont repérés dans la matrice X par des 1 pour la présence et des 0 pour l'absence. Voir page 90.

La matrice des moments X'X qui en résulte, voir page 90, est d'ordre 11, exactement la même chose que dans le premier cas. Mais la matrice irréductible qu'il va falloir inverser s'élève à l'ordre 5 (au lieu de 3), ce qui est un peu plus compliqué.

Nous donnons les éléments du calcul : Matrice - Vecteur - Matrice inverse - Coefficient de régression - Etude détaillée des carrés emportés par la régression.

Matrice directe irréductible					Vecteur X'Y	Matrice inverse					Coefficient de régression
4	0	0	0	5	+2125,0	0,445.312	+0,117.187	-0,117.187	-0,195.312	-0,156.250	T = 518, 593.750
0	4	0	0	3	+2545,0	+0,117.187	+0,320.312	-0,070.312	-0,117.187	-0,093.750	A = 628, 656.250
0	0	4	0	-3	+2425,0	-0,117.187	-0,070.312	+0,320.312	+0,117.187	0,093.750	B = 613, 843.750
0	0	0	4	-5	+2670,0	-0,195.312	-0,117.187	+0,117.187	+0,445.312	0,156.250	C = 680, 156.250
5	3	-3	-5	25	-510,25	-0,156.250	-0,093.750	+0,093.750	+0,156.250	0,125.000	$x_L y_L$ = 10, 125.000

Tableau des carrés emportés par la régression				
1.075.757,910.0				+26.253,808.590
	1.580.834,722.8			+19.095,433.593
		1.507.216,597.6		-18.645,503.904
			1.850.450,097.6	-34.432,910.155
26.253,808.590	19.095,433.593	-18.645,503.904	-34.432,910.155	2.562,890.625

Somme des carrés de la diagonale : 6.014.259, 328.125 + 2.562,890.625 = 6.016.822,218.750

Somme des carrés non diagonaux : 2 × -7.729,171.875 = -15.458,343.750

Somme des carrés du tableau = 6.001,363,875.000

Terme correctif = 5.959.701,562.500

Somme des carrés : traitements + interaction = 41.662,312.500

Matrice X (Corrélation)

Vecteur

	T	A	B	C	x_L	x_Q	x_C	y_L	y_Q	y_C	$x_L x_L$	Y
1 T	1				-1,5	1	-0,3	-1,5	1	-0,3	2,25	542
2 A		1			-0,5	-1	+0,9	-1,5	1	-0,3	0,75	712
3 B			1		+0,5	-1	-0,9	-1,5	1	-0,3	-0,75	657
4 C				1	+1,5	1	+0,3	-1,5	1	-0,3	-2,25	675
5 A		1			-1,5	1	-0,3	-0,5	-1	+0,9	0,75	640
6 T	1				-0,5	-1	+0,9	-0,5	-1	+0,9	0,25	538
7 C				1	+0,5	-1	-0,9	-0,5	-1	+0,9	-0,25	713
8 B			1		+1,5	1	+0,3	-0,5	-1	+0,9	-0,75	647
9 B			1		-1,5	1	-0,3	+0,5	-1	-0,9	-0,75	561
10 C				1	-0,5	-1	+0,9	+0,5	-1	-0,9	-0,25	667
11 T	1				+0,5	-1	-0,9	+0,5	-1	-0,9	+0,25	529
12 A		1			+1,5	1	+0,3	+0,5	-1	-0,9	+0,75	566
13 C				1	-1,5	1	-0,3	+1,5	1	+0,3	-2,25	615
14 B			1		-0,5	-1	+0,9	+1,5	1	+0,3	-0,75	560
15 A		1			+0,5	-1	-0,9	+1,5	1	+0,3	+0,75	627
16 T	1				+1,5	1	+0,3	+1,5	1	+0,3	2,25	516
	Effets traitements				Effets colonnes			Effets Lignes			Inter-action	

Les termes non figurés sont nuls

Matrice X'X

Vecteur

	T	A	B	C	x_L	x_Q	x_C	y_L	y_Q	y_C	$x_L y_L$	X'Y
T	4	0	0	0	0	0	0	0	0	0	5,00	2125,00
A	0	4	0	0	0	0	0	0	0	0	3,00	2545,00
B	0	0	4	0	0	0	0	0	0	0	-3,00	2425,00
C	0	0	0	4	0	0	0	0	0	0	-5,00	2670,00
x_L					20						0	+93,5
x_Q						16					0	-241,0
x_C							7,2				0	-30,3
y_L								20			0	-509,5
y_Q									16		0	+43,0
y_C										7,2	0	+113,1
$x_L y_L$	5,00	3,00	-3,00	-5,00	0	0	0	0	0	0	25,0	-510,25

Les termes non figurés sont nuls.

On remarquera que la somme des carrés des traitements + interaction est :

41.662,312.500 et qu'elle est bien égale à :

39.748,250 : somme des 9 termes de la matrice (3 ×3) de la page 86

1.914,062.500 : somme des carrés relative à l'effet quadratique des traitements qui se trouve être orthogonal à l'interaction.

De même on remarquera que la somme des carrés à affecter aux traitements corrigés est :

6.014.259,328.125 - T.C. = 54.557,765.625, chiffre connu.

Cette troisième méthode est donc rigoureusement équivalente à la précédente : elle a l'avantage de fournir directement les valeurs ajustées des traitements. Mais il faut bien remarquer que les "traitements" ne sont pas des "comparaisons orthogonales" au sens statistique précis de ce terme, comme le sont les colonnes 2 à 11 de la matrice de la page 84 - En effet ce ne sont même pas des comparaisons : la somme des termes n'est pas nulle. Ils sont pourtant orthogonaux entre eux.

EXEMPLE DE REGRESSION QUADRATIQUE (Suite)

Revenons donc à l'exemple introduit plus haut, à la page 79. Malgré certaines précautions prises (barycentrage) qui permettent une estimation non biaisée, ou indépendante, des effets linéaires en x et y - et de l'effet quadratique en x -, nous sommes dans le cas général où les traitements et les effets dus au terrain sont en interdépendance. La matrice $(X'X)$ ici d'ordre 12 s'abaisse du fait de l'orthogonalité à l'ordre 9. C'est une matrice symétrique du type biflèche, où l'on rencontre des termes dominants dans la diagonale principale et dans deux colonnes seulement ; on peut rejeter ces deux colonnes à l'extrême droite, d'où la désignation de matrice flèche et même biflèche.

L'ennui est que l'inverse de cette matrice est une matrice pleine où aucun coefficient n'a de raison d'être nul. Nous avons donc à manipuler une matrice (9 × 9 = 81), symétrique bien sûr, et à la multiplier par le vecteur colonne $X'Z$, pour obtenir les coefficients de la régression quadratique (oui, quadratique en ce qui concerne le terrain - mais en fait linéaire en ce qui concerne le programme de calcul, car les calculateurs numériques ne connaissent que celle-ci, et jouent sur les variables d'entrée pour avoir des régressions autres que linéaires).

Nous ne pensons pas utile de donner la matrice inverse $(X'X)^{-1}$ hélas si foisonnante. Nous nous contentons de donner le vecteur colonne $X'Z$, les coefficients correspondants de la régression.

Les 28 termes non-ortho. :	-55,310.824		
Somme des carrés résiduelle :	137,686.507	pour 16	D.d.L.
Somme des carrés totale :	66.491,196.800	pour 28	D.d.L.

Du fait de la non-orthogonalité de la majeure partie de nos traitements et effets terrain, l'analyse de variance a perdu sa belle simplicité : nous ne voyons pas d'autre moyen que de la donner in extenso

	XZ	Coefficient de régression	Carré emporté par la régression	
+N	219, 80	54, 558.543	11.906, 538.566	
+P	198, 80	51, 244.161	10.503, 856.187	
+K	197, 40	49, 536.914	9.815, 623.592	
NK	195, 16	48, 441.874	9.386, 460.820	66.155, 975.518
NP	191, 80	47, 558.543	9.047, 260.124	
NPK	189, 84	47, 399.041	8.986, 676.352	
T	163, 52	40, 340.922	6.509, 559.877	
x_L	40, 88	1, 168.000	47, 747.840	
y_L	131, 04	1, 170.000	153, 316.800	209, 845.440
x_Q	-15, 68	-0, 560 000	8, 780.800	
y_Q	-111, 44	-0, 287.167	27, 708.118	43, 000.159
$x_L y_L$	-50, 40	0, 330.498	15, 292.041	

sous forme d'un tableau carré (12 × 12) ou mieux (9 × 9), comme nous l'avons fait pour le carré latin (voir ci-dessus). Ce tableau encore assez simple aura la structure de la matrice flèche ($X'X$). On y rencontre 7 × 2 × 2 = 28 termes non diagonaux qui ne sont pas nuls, donc relatifs à la non-orthogonalité. Ces termes ont une somme faible : -55,310.824 ; elle est négative, ce qui indique que nous avons augmenté le contraste pour la séparation des traitements.

Matrice ($X'X$) des carrés non orthogonaux

	+N	+P	+K	NK	NP	NPK	T	y_Q	$x_L y_L$
+N	11.906, 538.566							+31, 334.779	+54, 094.453
+P		10.503, 856.187						-147, 156.097	-169, 360.880
+K			9.815, 623.592					+28, 450.691	-65, 487.386
NK				9.386, 460.820				+83, 465.320	-16, 009.938
NP					9.047, 260.124			+27, 314.447	+47, 153.997
NPK						8.986, 676.352		+27, 222.839	-15, 665.283
T							6.509, 559.877	-46, 338.256	+133, 325.902
y_Q	+31, etc..	-147,	+28,	+83,	+27,	+27,	-46,	27, 708.118	0
$x_L y_L$	+54, etc..	-169,	-65,	-16,	+47,	-15,	+133,	0	15, 292.041

Les termes non figurés sont nuls.

Ce tableau concerne les sommes de carrés relatives à 9 degrés de liberté, globalement.

Du fait du manque d'orthogonalité il n'est pas possible de préciser plus, du moins en toute rigueur.

- La somme des carrés relative aux traitements ajustés est ,

66.155,975.518 - 65.700,140.800 = 455,834.718

au lieu de = 413,089.600

- La somme des carrés résiduelle est donc de : 137,686.507 pour 16 degrés de liberté, ce qui fournit une variance résiduelle de 8,605.406, un écart-type de : 2,93, et un coefficient de variation de : 6,05 %.

- Le F de Fisher-Snedecor est :

Variance relative aux traitements : 455,834.718/6 = 75,972.453 d'où F = 75,972.453/8,605.406 = 8,83 à comparer avec F (6,16) au niveau 5 %, soit : 2,74.

- Nous rappelons que nous avions :

F = 3,94 pour 2,66, pour le dépouillement classique,

et F = 7,22 pour 2,70, pour la régression linéaire.

L'avantage de la régression quadratique est très faible dans cet exemple, mais on augmente tout de même le contraste des traitements et l'on diminue légèrement la variance résiduelle de base.

En effet, avec cette variance, on voit qu'on serait amené à ne tenir compte que des effets linéaires seulement (47,75 et 153,32).

- Toutefois la régression quadratique nous permet de remonter sensiblement l'estimation de + P, très légèrement celle de + K, et de baisser très légèrement les estimations des 5 autres traitements. Elle est utile pour préciser les choses, mais elle n'est pas indispensable.

EXEMPLE D'UN PLAN QUADRATIQUEMENT ORTHOGONAL

Nous savons que le barycentrage, très naturel des carrés latins, un peu plus artificiel des rectangles barycentrés, permet d'utiliser des plans qui sont "robustes" contre les effets linéaires. Nous ne pensons pas que les plans robustes à la fois contre les effets linéaires et les effets quadratiques soient bien nombreux. Pourtant une recherche nullement systématique, et très lacunaire, nous a permis d'en trouver quelques uns, où effets linéaires et quadratiques du terrain sont bien orthogonaux entre eux (aucune difficulté) et orthogonaux aux effets linéaire et quadratique des traitements (rangés dans un certain ordre) mais malheureusement il n'y a plus d'orthogonalité pour les effets d'ordre supérieur (cubique, bicarré, etc). Or nous souhaitons des plans tels que la totalité des effets traitements soit bien orthogonale aux effets du terrain. Cette condition semble nous ramener inexorablement aux carrés latins, mais par définition le carré latin sera toujours biaisé par un effet combiné du type $x_{LIN} \times y_{LIN}$, c'est même là sa faiblesse congénitale : l'interaction est supposée nulle.

Si l'on considère un carré latin impair, on constate que les parcelles de la colonne centrale, et celles de la ligne centrale ne pèsent d'aucun poids pour la détermination d'un effet croisé du type : x_L y_L (x ou y sont nuls). On pourrait croire que ce fait constitue un avantage et qu'on va pouvoir tracer un carré latin facilement. On est donc ramené à équilibrer un carré latin d'ordre pair : admettons que ce soit possible, on voit alors qu'on est obligé de mettre toujours le (2 n + 1)ème traitement dans les branches de la croix centrale et qu'au centre même on ne sait plus trop quoi mettre : ce dispositif ne saurait convenir : où sont les blocs orthogonaux, attendus par lignes et par colonnes ?

Quand on analyse les origines de nos difficultés pour les carrés latins d'ordre pair, on voit qu'elles tiennent aux termes rencontrés dans les deux diagonales. Le plus petit carré possible est celui d'ordre 4 : 4 × 4 = 16 parcelles. On tourne la difficulté en mettant dans les deux diagonales le même traitement, de préférence le témoin, qui sera ainsi répété 8 fois, tandis que les deux traitements restant ne sont répétés que 4 fois. Nous n'avons pas pu trouver mieux comme exemple de plan "quadratiquement orthogonal" - du moins pour le moment.

Nous donnons donc le plan : (4 × 4) = 16 = (8 + 4 + 4) et nous traitons un exemple numérique par les deux méthodes : dépouillement classique du carré latin, et dépouillement par ajustement d'une fonction quadratique à deux variables. Bien entendu, nous n'avons encore pas réalisé concrètement de tels plans : l'exemple ne peut être que "synthétique". Mais nous l'avons voulu aussi "réel" que possible : nous avons repris les valeurs numériques d'un essai fait au laboratoire de la Croix de Berny le 15 juin 1960, et nous les avons arrangées - simple changement de place dans le plan de l'essai.

T_1	A	B	T_2
B	T_1	T_2	A
A	T_2	T_1	B
T_2	B	A	T_1

50,65	54,65	54,45	53,50	213,25
52,50	49,90	50,75	64,20	217,35
52,05	49,60	46,25	50,70	198,60
46,25	47,90	49,35	46,05	189,55
201, 45,	202, 05	200, 80	214, 45	818,75

T_1 = 192,85 T_2 = 200,10 A = 220,25 B = 205,55

A/ <u>Dépouillement en carré latin classique</u> en distinguant les deux témoins T_1 et T_2.

		Degrés de Liberté
Somme des carrés générale	: 293,999 843 750	15
Somme des carrés colonnes	: 31,964 218 750	3
Somme des carrés Lignes	: 124,964 218 750	3
Somme des carrés traitements	: 101,026 718 750	3
Somme des carrés résiduelle	: 36,044 687 500	6

D'ou une variance résiduelle de 6,007 447 916...

D'où une variance résiduelle de 6,007.916...

Le test F de Fisher-Snedecor indique clairement qu'il y a, au niveau habituel de 5 %, un effet lignes et un effet traitements.

B/ <u>Dépouillement par ajustement quadratique</u> en confondant les deux témoins T_1 et T_2.

Du fait de l'orthogonalité la matrice se réduit à la diagonale principale (2ème colonne de notre tableau) : le vecteur produit est donné à la 3ème colonne ; puis viennent, à la 4ème colonne, le coefficient de régression, et à la 5ème le carré emporté par ce terme de régression. A la 6ème et dernière colonne des sommes partielles font réapparaître des quantités en général déjà connues par le dépouillement classique du carré latin.

1 Désignation de l'effet	2 Matrice	3 Vecteur	4 Coefficient de Régression	5 Carré emporté par la régression	6
T	8	392,95	49,118.750	19.301,212.812	Somme <u>brute</u> des
A	4	220,25	55,062.500	12.127,515.625	carrés traitements
B	4	205,55	51,387.500	10.562,700.625	41.991,429.062
x_{LIN}	20	18,875	0,943.750	17,813.281	Somme des carrés
$x_{QUA.}$	16	13,050	0,815.625	10,643.906	effets colonnes
$x_{CUB.}$	7,2	5,025	0,697.916	3,507.031	31,964.218
$y_{LIN.}$	20	-44,925	-2,246.250	100,912.781	Somme des carrés
$y_{QUA.}$	16	-13,150	-0,821.875	10,807.656	effets lignes
y_{CUB}	7,2	9,765	1,356.250	13,243.781	124,964.218
$x_L y_L$	25	-16,4625	-0,658.500	10,840.556	

Somme des carrés toutes mesures : 42.190,972.500
Somme des carrés résiduelle : 31,774.443 par différence
Par ailleurs terme correctif : 41.896,972.656
d'où somme des carrés traitements : 94,456.406

D'où le tableau récapitulatif :

		Degrés de Liberté
Somme des carrés générale	: 293,999.843.750	15
Somme des carrés colonnes	: 31,964.218.750	3
Somme des carrés lignes	: 124,964.218.750	3
Somme des carrés traitements	: 94,456.406.250	2
Somme des carrés interaction	: 10,840.556.250	1
Somme des carrés résiduelle	: 31,774.443.750	6

d'où une variance résiduelle de 5,295.740.625.

Les conclusions avec cet autre type de dépouillement restent ici inchangées. Mais il est clair qu'on a un peu diminué la variance résiduelle et qu'on a beaucoup augmenté la variance des traitements, quoiqu'en perdant un degré de liberté ce qui tend à réduire le contraste (F monte en effet de 4,76 à 5,14). Quant à l'interaction $x_L y_L$ elle reste d'un niveau très acceptable, mais serait-elle significative que nous nous en soucierons pas puisque le plan est orthogonal et qu'elle ne peut biaiser les traitements.

Il ne semble pas qu'un plan (6 × 6) = 36 parcelles puisse être quadratiquement orthogonal. Par contre nous donnons ci-dessous un plan (8 × 8) = 64 parcelles qui est lui aussi, comme le (4 × 4), quadratiquement orthogonal. Bien entendu les deux diagonales sont occupées par le même traitement et il semble que le dispositif indiqué ne soit pas unique, mais on doit conserver la contrainte d'un même traitement dans les deux diagonales.

Plan (8 × 8) quadratiquement orthogonal

T_1	F	E	D	C	B	A	T_2
A	T_1	D	E	B	C	T_2	F
B	C	T_1	A	F	T_2	D	E
C	B	F	T_1	T_2	A	E	D
D	E	A	T_2	T_1	F	B	C
E	D	T_2	F	A	T_1	C	B
F	T_2	C	B	E	D	T_1	A
T_2	A	B	C	D	E	F	T_1

Pour l'instant il ne nous apparaît pas qu'il y ait pléthore de plans quadratiquement orthogonaux, mais peut être les conditions d'existence ne sont-elles pas aussi draconiennes que nous le pensons, et quelque lecteur amateur de curiosités mathématiques, pourra-t-il nous donner un démenti en nous apportant nombre de ces plans qui seraient précieux dans le cas où l'on a décidé de juger en faisant l'hypothèse d'un ajustement quadratique ?

CONCLUSION

Si l'on consent à faire la double convention :

- chaque parcelle est représentée par un rendement moyen

- la "surface" de ces rendements moyens est une quadrique d'équation :

$$z = z_T + ax + by + c\,(x^2 - \alpha^2) + d\,(y^2 - \beta^2) + e\,xy + \varepsilon_R$$

nous avons montré que la méthode de régression linéaire multiple était la méthode convenable pour faire cet ajustement. On obtient ainsi des valeurs plus précises pour les traitements, et des coefficients de régression qui donnent l'allure de la quadrique d'ajustement. Ces coefficients traduisent donc l'hétérogénéité du terrain.

A la limite, nous avons montré qu'il existe des plans qui seraient quadratiquement orthogonaux, où traitements et effets terrain sont estimables sans biais, du moins si on se contente des effets linéaires, quadratiques et produits des linéaires. Dans ce cas l'analyse de variance classique reste possible, et chaque degré de liberté, c'est-à-dire chaque effet élémentaire, emporte un certain carré dans la régression.

Pratiquement cette méthode d'ajustement quadratique - par régression linéaire multiple - nous permettra le dépouillement des plans les plus quelconques. Nous avons vû comment on pouvait affiner ainsi le dépouillement des carrés latins et sortir de l'impasse où l'on se trouve si une interaction lignes-colonnes a le malheur d'être prédominante. Plus couramment nous pensons que cette régression quadratique s'appliquera à des plans totalement randomisés, avec l'inconvénient de renoncer à une analyse de variance sérieuse : tout élément étant biaisé par tous les autres. Néanmoins cette méthode conduit à l'obtention d'une variance résiduelle de la régression, qu'il sera parfaitement légitime d'employer comme la variance résiduelle d'une analyse de variance classique.

Enfin cette méthode d'ajustement s'accomode de répétitions inégales, de données manquantes, d'un plan non rectangulaire, etc...

Bref, on ne sera plus aussi exigeant sur les plans (encore qu'il faille éviter ceux qui seraient systématiquement idiots) pour les essais déployés sur une surface, si l'on admet de traiter les résultats par régression quadratique. Ces essais sur des "surfaces" peuvent être des essais agronomiques, sur des champs, mais ce serait aussi bien l'étude de la corrosion des briques d'un four, la surface étant alors la paroi du four.

Mais il faut renoncer aux "comparaisons" rigoureusement orthogonales et à l'analyse de variance simple qui en découle. Nous n'avons fait que déplacer le problème, et l'on ne sort guère du dilemme :

- plans très élaborés, dépouillement simple et rigide,

- plans quelconques, dépouillement exigeant une effarante complexité dans les calculs - méthode aujourd'hui à notre portée grâce aux calculatrices électroniques.

REFERENCES BIBLIOGRAPHIQUES

[1] Jean LESAVRE - "Régression linéaire et plans barycentrés dans les dispositifs expérimentaux utilisés en agronomie". Revue de statistique appliquée 1965, Vol XIII N° 1 pages 95 à 118.

[2] Daniel CUTHBERT and Frank WILCOXON - "Factorial 2^{p-q} Plans Robust against Linear and Quadratic Trends". Technometrics May 1966, Volume 8 Number 2 pages 259 à 278.

[3] E.S. PEARSON and H.O. HARTLEY - Biometrika Tables for Statisticians, Volume I Cambridge University Press (pour trouver les coefficients permettant le calcul des effets linéaires, quadratiques, etc...)

[4] N.R. DRAPER and H. SMITH - Applied Regression Analysis. John Wiley and Sons, Inc. Third Printing, July 1967, notamment :

5,3 The Use of "Dummy" Variables in Multiple Regression
5,5 Orthogonal Polynomials
7,5 The addition of a cross-product term to the Model
9 Multiple Regression Applied to Analysis of Variance Problems