

REVUE DE STATISTIQUE APPLIQUÉE

JEAN LESAVRE

Une méthode rapide d'ajustement linéaire. La régression orthogonalisée. Son application à l'industrie du verre

Revue de statistique appliquée, tome 13, n° 2 (1965), p. 75-82

http://www.numdam.org/item?id=RSA_1965__13_2_75_0

© Société française de statistique, 1965, 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/>

UNE MÉTHODE RAPIDE D'AJUSTEMENT LINÉAIRE LA RÉGRESSION ORTHOGONALISÉE SON APPLICATION A L'INDUSTRIE DU VERRE

Jean LESAVRE

Ingénieur à la Direction des Recherches de la Compagnie de SAINT-GOBAIN(*)

I - POSITION DU PROBLEME

Actuellement lorsque nous voulons étudier les variations d'une propriété en fonction d'une composition, nous avons à notre disposition une méthode basée sur l'exploitation d'un plan factoriel. On peut, soit réaliser un plan factoriel, soit le compléter de répétitions du point moyen et d'une "étoile" sur les axes passant par le point moyen. Ces derniers plans sont connus sous le nom de "Central composite rotatable designs". Même si l'on se contente de plans factoriels du type 2^n , qui sont les plus économiques, on arrive très vite à un nombre d'essais considérable, c'est-à-dire dans notre cas à la nécessité de fondre et d'affiner un grand nombre de verres. Pratiquement pour un verre à n composants, il n'y a que $n-1$ composants indépendants qu'on peut prendre comme facteurs, on a un minimum de 2^{n-1} essais à réaliser ; pour 7 composants, par exemple on aurait $2^6 = 64$ verres à fabriquer. C'est beaucoup trop si l'on ne se trouve pas à un point singulier, mais dans le cas courant où l'approximation linéaire est probable on devrait se contenter d'un nombre d'essais bien inférieur. A la limite si l'on veut déterminer n coefficients pour n composants d'un type de verre, on peut penser que n verres suffiraient. En effet chacun sait que n équations à n inconnues est un problème généralement soluble. Si l'on a plus de n équations pour n inconnues on se trouve dans le cas général de la régression multiple.

II - REGRESSION LINEAIRE

Cette technique de la régression comporte l'établissement d'une matrice dite matrice des corrélations. Elle est symétrique. Son inversion est le point délicat et pénible d'une chaîne d'opérations par ailleurs assez simples.

Pour ajuster une équation de la forme :

$$y = a + b_1 x_1 + b_2 x_2 + \dots + b_k x_k$$

à un nuage de points $(y_j, x_{1j}, x_{2j}, \dots, x_{kj})$

où $j = 1, 2, \dots, n$ et où $n-1$ est supérieur à k , on utilise la méthode des moindres carrés. Pour cela on établit la matrice de corrélations, qui est d'ordre $k + 1$ (matrice A) et le vecteur colonne (B) :

* Chef du Centre de Calcul à l'Usine Verrière Expérimentale de CHALON sur SAONE

$$A = \begin{vmatrix} n & \sum x_{1j} & \sum x_{2j} & \dots & \sum x_{kj} \\ \sum x_{1j} & \sum x_{1j}^2 & \sum x_{1j}x_{2j} & \dots & \sum x_{1j}x_{kj} \\ \sum x_{2j} & \sum x_{1j}x_{2j} & \sum x_{2j}^2 & \dots & \sum x_{2j}x_{kj} \\ \dots & \dots & \dots & \dots & \dots \\ \sum x_{kj} & \sum x_{1j}x_{kj} & \sum x_{2j}x_{kj} & \dots & \sum x_{kj}^2 \end{vmatrix} \quad \text{et } B = \begin{vmatrix} \sum y_j \\ \sum x_{1j}y_j \\ \sum x_{2j}y_j \\ \dots \\ \sum x_{kj}y_j \end{vmatrix}$$

Matrice des corrélations vecteur colonne

où tous les \sum sont par rapport aux j .

On doit bien entendu calculer également $\sum y_j^2$.

Cette matrice symétrique recèle quelques pièges : en particulier la quantité n dans la diagonale n'est pas toujours de l'ordre de grandeur qui serait souhaitable. On peut facilement ramener cette matrice d'ordre $(k+1)$ à une matrice d'ordre k , en utilisant les variables centrées. Il n'est du reste pas nécessaire de faire effectivement la translation pour placer l'origine au barycentre du nuage de points. La formule classique avec le terme correctif du type : $n \bar{x}^2$ est parfaitement applicable. Quant aux termes de la 1ère ligne ou de la 1ère colonne, ils deviennent nuls - à l'exception du terme diagonal n .

On obtient une matrice \underline{A} et un vecteur colonne \underline{B} , qui ne contiennent plus que des termes du second degré, carrés dans la diagonale, produits croisés aux autres places.

On a :

$$\sum X_{1j} = \sum X_{2j} = \dots = \sum X_{kj} = \sum Y_j = 0$$

$$\underline{A} = \begin{vmatrix} \sum X_{1j}^2 & \sum X_{2j}X_{1j} & \dots & \sum X_{kj}X_{1j} \\ \sum X_{1j}X_{2j} & \sum X_{2j}^2 & \dots & \sum X_{kj}X_{2j} \\ \dots & \dots & \dots & \dots \\ \sum X_{1j}X_{kj} & \sum X_{2j}X_{kj} & \dots & \sum X_{kj}^2 \end{vmatrix} \quad \text{et } \underline{B} = \begin{vmatrix} \sum X_{1j}Y_j \\ \sum X_{2j}Y_j \\ \dots \\ \sum X_{kj}Y_j \end{vmatrix}$$

La multiplication de \underline{A}^{-1} , matrice inverse, par le vecteur colonne \underline{B} donne les coefficients de la régression linéaire b, b, \dots, b . Quant à a , c'est la moyenne des y_j .

III - REGRESSION ORTHOGONALISEE

On voit tout de suite qu'un système de valeurs numériques tel que toutes les sommes de produits croisés soient nulles nous faciliterait à l'extrême la tâche de l'inversion de la matrice, qui serait alors une matrice diagonale. Or de tels systèmes existent : ils sont fournis, par exemple, par la technique des ajustements polynomiaux par la méthode des polynômes orthogonaux. Pour le but précis que nous poursuivons il est à peine nécessaire de rappeler ce que sont les polynômes orthogonaux.

Par n points on peut en principe faire passer une courbe de degré $n-1$ et l'ajustement d'une courbe de degré k , inférieur ou égal à $n-1$, peut se faire par plusieurs méthodes. L'une des plus intéressantes est justement celles des polynômes orthogonaux de Tchebycheff, que les statisticiens connaissent bien à la suite des travaux de Ronald Fisher.

Concrètement, la méthode suppose que les n observations sont équidistantes sur l'axe des abscisses, et qu'elles sont de plus centrées sur l'origine. On doit commencer par les polynômes de degrés les plus bas. Ainsi nous rencontrons :

$$\begin{aligned}\Phi_0(x) &= 1 \\ \Phi_1(x) &= \lambda_1 x \\ \Phi_2(x) &= \lambda_2 \left[x^2 - \frac{1}{12} (n^2 - 1) \right] \\ \Phi_3(x) &= \lambda_3 \left[x^3 - \frac{1}{20} (3n^2 - 7)x \right] \\ \Phi_4(x) &= \lambda_4 \text{ (etc..)}\end{aligned}$$

n est le nombre total de points auxquels on veut ajuster une moyenne, une formule linéaire, puis quadratique, puis cubique, etc.. L'expression des polynômes orthogonaux se complique singulièrement dès que k s'élève. Il existe une relation de récurrence entre Φ_{k+1} et Φ_k ce qui simplifie les choses - mais nous n'avons nul besoin de connaître ces polynômes. Pour calculer le coefficient à affecter au polynôme $\Phi_k(x)$ on se sert de tables donnant λ_k , et une collection de n valeurs numériques formant en quelque sorte un vecteur de l'espace à n dimensions. On fait le produit scalaire de ce vecteur par le vecteur expérimental $(y_1, y_2, y_3, \dots, y_n)$, et on divise le résultat par le carré du vecteur - valeur qu'on trouve aussi dans la table - pour obtenir le coefficient du polynôme Φ_k .

Cette méthode des polynômes orthogonaux permet également le calcul des sommes de carrés dues aux effets linéaire, quadratique, cubique, etc. et c'est là son principal intérêt en statistique appliquée. Pour cela on élève au carré le produit scalaire (vecteur de la table \times vecteur expérimental) et l'on divise par le carré du vecteur de la table. De cette façon on sait la fraction de la variance des y expliquée par l'effet linéaire, l'effet quadratique, etc..

La détermination du coefficient d'un polynôme de degré k , ou de la somme de carrés affectée à l'effet de ce polynôme, est unique et invariable. En raison de l'orthogonalité des polynômes et des vecteurs qu'on en tire, elle n'est pas influencée par les autres déterminations, pour des degrés inférieurs ou supérieurs. Autrement dit on dispose avec ces polynômes orthogonaux d'une méthode permettant, dans un ajustement polynomial, le calcul sélectif et indépendant du coefficient, ou de la somme de carrés, à attribuer au polynôme de degré K .

Pour nous, ce qui compte, ce sont ces tables donnant les n vecteurs orthogonaux dans l'espace à n dimensions. L'un de ces vecteurs est la 1ère bissectrice $(1, 1, 1, \dots, 1)$; il est sans intérêt pour nous. Mais les $(n-1)$ autres vecteurs forment un système tout trouvé de valeurs numériques ordonnées telles que leur produit scalaire, $\sum x_{1j} \cdot x_{2j} = 0$ par exemple, sera toujours nul. Nous avons donc, avec ces tables, un répertoire de valeurs numériques tel que toutes les sommes de leurs produits croisés soient nulles. Une transformation linéaire conservant cette propriété, il nous sera possible d'ajuster nos facteurs aux niveaux désirés. Et nous obtenons un plan d'expériences, assez simple à réaliser, et dont on pourra, presque sans calcul, tirer une formule linéaire. En effet, la matrice des corrélations sera une matrice diagonalisée, donc facile à inverser ; cette diagonalisation étant obtenue par l'emploi de vecteurs orthogonaux, nous donnons à ce plan très particulier le nom de *régression orthogonalisée*.

Un exemple, sans doute, fera mieux comprendre les particularités de la méthode.

IV - EXEMPLE PRATIQUE DE REGRESSION ORTHOGONALISEE

Nous le prenons bien entendu dans l'industrie du verre. Et nous nous excuserons de ne pas le donner totalement en clair. Les compositions sont plus ou moins secrètes, plus ou moins brevetables également.

Le plan concerne 10 verres, 9 verres + 1 verre moyen, composés de silice et de 6 autres composants, que nous appellerons A, B, C, D, E, F, tout simplement. La silice, comme c'est assez naturel, est choisie comme facteur de bouclage à 100 % ; la quantité de silice est donc liée à A, B, C, D, E, F, par la relation : $S = 100 - (A+B+C+D+E+F)$. On établit le plan sans s'occuper de la silice qui est obtenue par différence. Nous voulons un nuage de 9 verres. Nous prenons donc le tableau des vecteurs orthogonaux pour $n = 9$, et nous obtenons le tableau I. Ce tableau est celui du répertoire à un signe près : en effet on peut changer tous les signes d'un vecteur, on obtient le vecteur opposé, qui a les mêmes propriétés d'orthogonalité ; c'est ce que nous avons fait pour Φ_5 . L'affectation des polynômes orthogonaux aux composants peut se faire dans un ordre quelconque.

TABLEAU N° I

n=9	Φ_0	Φ_1	Φ_2	Φ_3	Φ_4	$-\Phi_5$	Φ_6	Φ_7	Φ_8		
Bouclage par différence	1	-4	28	-14	14	+4	4	-1	1		
	1	-3	7	7	-21	-11	-17	+6	-8		
	1	-2	-8	13	-11	+4	22	-14	+28		
	1	-1	-17	9	9	+9	1	+14	-56		
	1	0	-20	0	18	0	-20	0	70		
	1	1	-17	-9	9	-9	1	-14	-56		
	1	2	-8	-13	-11	-4	22	+14	+28		
	1	3	7	-7	-21	+11	-17	-6	-8		
	1	4	28	14	14	-4	4	+1	+1		
Limites de	SiO ₂	//	A	B	D	C	E	Disp	F	Disp	
	de	?	//	6,2	2,00	4,6	4,0	1,9		8,6	
	à	?	//	11,8	4,83	7,4	6,0	4,1		11,4	

Les limites sont approximativement connues. On aura souvent avantage à les fixer de telle façon que l'ajustement linéaire puisse se faire sans traîner trop de décimales. On remarquera que la somme des produits croisés deux à deux de 2 colonnes est toujours nulle : c'est cela la propriété d'orthogonalité. Elle est conservée par une transformation linéaire, ce qui nous permet donc d'ajuster nos compositions aux limites désirées. Par exemple :

$$\Phi_3 \text{ et } D \quad \quad \quad -14 \text{ doit représenter } 4,6$$

$$\text{et} \quad \quad \quad +14 \text{ doit représenter } 7,4$$

Ces deux points suffisent à établir la transformation linéaire. Il y a 6 transformations linéaires à effectuer, c'est assez fastidieux, mais sans difficultés. Par différence on trouvera la teneur en SiO₂ (silice) de chaque verre. Chaque verre voit maintenant sa composition représentée dans ce système par une ligne. On obtient le tableau II, complété par un verre moyen.

TABLEAU II

SiO ₂	A	B	C	D	E	F	Références	Y
65,32	6,20	4,83	5,75	4,60	3,40	9,90	8312	70
66,31	6,90	3,59	4,00	6,70	1,90	10,60	8308	60
65,89	7,60	2,71	4,50	7,30	3,40	8,60	8314	62
61,82	8,30	2,18	5,50	6,90	3,90	11,40	8313	50
64,05	9,00	2,00	5,95	6,00	3,00	10,00	8311	55
66,82	9,70	2,18	5,50	5,10	2,10	8,60	8306	60
63,69	10,40	2,71	4,50	4,70	2,60	11,40	8305	40
62,51	11,10	3,59	4,00	5,30	4,10	9,40	8309	50
57,52	11,80	4,83	5,75	7,40	2,60	10,10	8307	60
63,77	9,00	3,18	5,05	6,00	3,00	10,00	8310	55

Ainsi nous avons pris la précaution de *fabriquer* des variables indépendantes (aux erreurs de composition près) et le dépouillement des résultats sera très simplifié, puisque de cette façon la matrice des corrélations est devenue diagonale, donc inversible. La variable y a les valeurs indiquées à la dernière colonne du tableau II. C'est cette quantité que nous voulons relier par une formule linéaire aux teneurs en divers composants.

Tout d'abord si l'on prend les 10 termes de y , on obtient une moyenne de 56,20 et une somme de carrés des écarts de 609,60, qu'il va nous falloir répartir dans une analyse de variance. Si l'on prend 9 termes (exclusion du point moyen, N° 8310) on obtient une moyenne de 56,33 et une somme de carrés des écarts qui est : 608,00. La différence, soit 1,60, représente la variance entre le point moyen "théorique" c'est-à-dire en fait entre le point moyen des 9 verres et le point moyen expérimental (verre N° 8310). Si cette variance était trop forte, on serait amené à suspecter la valeur de l'ajustement par une formule linéaire.

Revenons à notre "nuage" de 9 points. On n'a pas utilisé Φ_6 et Φ_8 , mais dans le dépouillement ces 2 vecteurs vont nous permettre de calculer une variance de dispersion - avec 2 degrés de liberté.

La matrice de corrélations se présente ainsi réduite à sa diagonale :

29,40 (A)	et	1980,00 (Dispersion Φ_6)
9,36 (B)		12870,00 (Dispersion Φ_8)
5,00 (C)		
9,90 (D)		voir tableau I
4,68 (E)		
8,58 (F)		

Ces quantités sont purement et simplement les sommes de carrés des écarts relatives aux divers composants, et aux composants fictifs Φ_6 et Φ_8 .

Il s'agit maintenant de calculer le vecteur de la corrélation entre y et les diverses colonnes, indépendantes, du tableau II, et du tableau I pour Φ_6 et Φ_8 . On obtient le vecteur, à la 1ère colonne du tableau III.

TABLEAU III

	Vecteur corrélation	Coefficient de la régression	Variance
A	-72,80	-2,48	180,27
B	+36,25	+3,87	140,46
C	+18,40	+3,68	67,64
D	+12,60	+1,27	16,04
E	-7,20	-1,54	11,08
F	-39,80	-4,64	184,62
Φ_6	-96,00		4,65
Φ_8	-204,00		3,23
$\Sigma =$			607,99

En divisant le vecteur corrélation par les carrés correspondants de la diagonale de la matrice des corrélations (29,40 puis 9,36 etc..) on obtient les coefficients de la régression linéaire désirée, sans s'occuper de Φ_6 et Φ_8 , qui sont fictifs.

En élevant au carré les termes du vecteur corrélation et en divisant ce résultat par les éléments successifs de la diagonale de la matrice, on obtient les sommes de carrés, qui sont d'ailleurs ici des variances, tous les degrés de liberté étant égaux à l'unité (1 pour l'effet linéaire, 1 pour l'effet quadratique, 1 pour l'effet cubique, etc..). Si les calculs sont justes, on a donc la vérification très simple : leur somme est égale à 608,00.

Nous pouvons conclure à une variance résiduelle qui serait $1/3$ ($1,60 + 4,65 + 3,23$) = 3,16 avec 3 degrés de liberté. On voit alors que les composants A, B et F sont significatifs au niveau 1 %, C est significatif au niveau 5 %, D est significatif au niveau 10 %. On pourrait en principe négliger (E).

Mais en fait nous avons fait ici un travail d'ajustement mathématique bien plus que statistique. Nous sommes à la recherche d'une formule linéaire représentant telle propriété en fonction de la composition, il nous importe peu que certains coefficients soient très faibles, ou non significatifs statistiquement parlant.

Dans le cas pris comme exemple nous aboutissons à la formule :

$$y - 56,33 = -2,48 a + 3,87 b + 3,68 c + 1,27 d - 1,54 e - 4,64 f$$

ou a, b, c, d, e, f sont des variables centrées :

$$\begin{array}{ll}
a = A - 9,00 & e = E - 3,00 \\
b = B - 3,18 & f = F - 10,00 \\
c = C - 5,05 & s = S - 63,77 \\
d = D - 6,00 &
\end{array}$$

V - REMARQUE SUR LES FORMULES LINEAIRES OBTENUES PAR REGRESSION DANS LE CAS DES PROBLEMES DE COMPOSITION BOUCLANT A 100 %

Il est assez choquant de voir une formule où la silice n'intervient pas, malgré un pourcentage généralement dominant. Si le mathématicien est heureux de se limiter aux variables réellement indépendantes, le praticien éprouve quelque gêne devant cette formule, qui lui paraît si peu objective, si peu intrinsèque, puisqu'elle semble négliger un composant, et même le plus important.

S'il est vain d'espérer revenir à n variables indépendantes, on peut toutefois rétablir la symétrie, c'est-à-dire les n composants. On peut en effet faire une "translation" de tous les coefficients sans affecter la formule.

La formule est du type $y - y_0 = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_{n-1} X_{n-1}$, où $X_1, X_2, \dots, X_{n-1}, X_n$ sont des variables centrées liées par la relation :

$$X_1 + X_2 + X_3 + X_4 \dots + X_{n-1} + X_n = 0$$

Ajoutons la translation β (sur les α , sans oublier $\alpha_n = 0$)

On obtient :

$$y - y_0 = (\alpha_1 + \beta)X_1 + (\alpha_2 + \beta)X_2 + \dots + (\alpha_{n-1} + \beta)X_{n-1} + \beta X_n$$

qui est équivalent à :

$$y - y_0 = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_{n-1} X_{n-1} + \beta(X_1 + X_2 + X_3 + \dots + X_{n-1} + X_n)$$

le second membre, malgré les apparences, ne dépend pas de β , puisque le coefficient de β est identiquement nul. On a donc une infinité de formules équivalentes. Nous en choisissons une, en nous imposant une condition : que le vecteur $(\alpha_1 + \beta) (\alpha_2 + \beta) \dots (\alpha_{n-1} + \beta) (\beta)$ soit de longueur minimum, par exemple. On retrouve alors une formule complète et unique, en faisant :

$$\beta = -1/n \sum_1^{n-1} \alpha_i ,$$

qui est non seulement la condition de longueur minimum mais aussi celle pour obtenir un vecteur de somme nulle.

Dans notre exemple :

$$\sum_1^{n-1} \alpha_i = 0,16 \quad , \quad \beta = -0,023 \quad ,$$

d'où :

$$y - 56,33 = -0,023 s - 2,503 a + 3,847 b + 3,657 c + 1,247 d - 1,563 e - 4,663 f$$

$$\text{avec } s + a + b + c + d + e + f = 0$$

Ici le coefficient de s est effectivement très minime, mais il n'en est pas toujours ainsi. On vérifiera que la somme des coefficients est bien nulle.

VI - CONCLUSION

Nous pensons tenir là, une méthode très souple, très puissante, très économique.

On peut en effet la restreindre à n'être plus qu'une méthode d'ajustement rigoureux, sans conserver aucun degré de liberté pour l'erreur : alors pour n coefficients réels, $n-1$ facteurs indépendants, il suffit de n essais, c'est-à-dire n verres.

On peut au contraire tout comme pour les "central composite rotatable designs" amplifier le plan en l'agréant de répétitions du point moyen et de points sur les axes passant par le barycentre du nuage.

Quant au dépouillement de ce plan d'expériences il devient vraiment enfantin, puisque nous avons su élaborer par un choix judicieux des niveaux des facteurs une matrice diagonale, directement inversible.

REFERENCE BIBLIOGRAPHIQUE

M. G. KENDALL - THE ADVANCED THEORY of STATISTICS - Vol. II - Edition 1961 - pages 356 et suivantes, sous le titre : ORTHOGONAL REGRESSION ANALYSIS