BIOMETRIE-PRAXIMETRIE — V, 1, 20-25, 1964

ANALYSE DE LA COVARIANCE D'EXPERIENCES 3^3 EN 27 UNITES SUR ORDINATEUR IBM 704

par

Monique CHARBONNEL et Raymond VAN DEN DRIESSCHE

Centre National de la Recherche Scientifique, Paris
et
Office de la Recherche Scientifique et Technique Outre-Mer, Bondy (Seine)

L'analyse de la variance d'une expérience factorielle 3^3 en 27 unités (qui peuvent être aussi bien des bocaux, des boîtes de Pétri, que des parcelles de terrain), dont chacune a reçu un des 27 traitements issus de la combinaison de trois facteurs à trois niveaux, peut se faire selon un processus systématisé, à l'aide d'une machine à calculer de bureau, même si elle englobe la décomposition des effets principaux et des interactions à deux facteurs en degrés simples de liberté (Van den Driessche, 1961).

L'analyse de la covariance d'un essai 3^3 débouche, par contre, sur de longs calculs qui risquent de trop accaparer l'expérimentateur, au détriment du temps qu'il consacrera à l'interprétation des résultats. Or les avantages de la covariance étant connus, il a semblé nécessaire de programmer cette analyse. L'ordinateur utilisé est un IBM 704 modèle 32 K à 6 armoires de bandes, non connecté à d'autres ordinateurs.

Une seule variable indépendante est prise en considération. Le nombre de degrés de liberté résiduels est déjà très faible (7 ou 5 selon qu'il y ait absence ou présence de confusion) dans une covariance totale de 3^3; aussi la covariance multiple n'a pas été envisagée.

L'entrée des données se fait par cartes perforées: 5 cartes maîtresses, 4 cartes détail et 1 carte interproblèmes. Une série de problèmes en provenance d'une même expérience se voit précéder de 5 autres cartes maîtresses. Un feuillet de données à perforer est utilisé par variable dépendante, ou par couple variable dépendante-variable indépendante. Les données sont inscrites de préférence en nombres de 3 chiffres (4 lorsque le chiffre des milliers ne dépasse pas 2) et sans virgule. Des multiplicateurs de cadrage restituent, néanmoins, les moyennes et erreurs-type dans l'unité désirée. Des données peuvent être négatives; leur signe tombe, dans ce cas, dans la case des milliers et le nombre négatif en question ne peut comporter plus de 3 chiffres. L'analyse est identifiée par un maximum de 15 caractères alphanumériques. La transformation éventuelle des données s'opère à la retranscription et, de plus, les tableaux de résultats sont à retransformer manuellement.

Lors de la perforation des données, on définit le terme résiduel (E) qui servira d'erreur expérimentale (Yates, 1937, page 18) dans l'analyse de la variance. Ce sera la somme des composantes de l'interaction à trois facteurs

$$E = ABC\,W + ABC\,X + ABC\,Y + ABC\,Z$$

avec 8 degrés de liberté, lorsque le dispositif 3^3 n'est pas confondu; ou bien la somme de trois composantes non confondues

$$E = ABC\,W + ABC\,X \qquad + ABC\,Z$$

avec 6 degrés de liberté, lorsque l'une de ces quatre composantes (ici $ABC\,Y$) est confondue avec les différences entre blocs incomplets de 9 unités expérimentales. A la notation de Yates (1937), on peut préférer celle de Finney (1947) et écrire l'erreur résiduelle d'une expérience 3^3 non confondue comme suit:

$$E = AB^2C^2 + AB^2C + ABC^2 + ABC$$

Dans certaines expériences, un choix malheureux des niveaux peut avoir pour conséquences une absence de réponse quadratique aux facteurs étudiés (A_q, B_q, ou C_q non significatifs à un seuil de probabilité préalablement choisi) et une réponse exclusivement linéaire (A_l, B_l ou C_l *significatifs*). Dans de tels cas il est loisible de joindre les composantes quadratiques des interactions à deux facteurs (A_qB_l, A_lB_q, A_qB_q, etc.), chacune avec un degré de liberté, aux termes résiduels E donnés plus haut. L'erreur est alors estimée avec 15 degrés de liberté:

$$E = ABC\,W + ABC\,X + ABC\,Z + A_qB_l + A_lB_q + A_qB_q + A_qC_l + \ldots + B_qC_q$$

Il est toutefois exclu d'opérer cette addition de composantes lorsque certains facteurs entraînent une réponse en forme de branche de parabole.

Le programme se présente sous forme de 232 cartes binaires renfermant les instructions de calcul + 128 cartes « dictionnaire » groupant les abréviations et la signification des différents effets factoriels, par exemple:

$$ABC\,Z_2 = 200 + 110 + 020 + 101 + 011 + 221 + 002 + 212 + 122$$

Deux analyses de variance en découlent, l'une sur la variable indépendante et l'autre sur la variable dépendante. Elles comprennent, chacune, les tableaux de moyennes affectées d'erreurs-type, les sommes de carrés et F correspondants, obtenus par décomposition des effets principaux et des interactions à deux facteurs. Ces décompositions sont, dans l'ordre : 1° la linéaire + la quadratique (Yates, 1937); 2° la générale + l'additionnelle (Yates, Lipton, Sinha et Das Gupta, 1959). On sait que le choix entre les deux méthodes de décomposition s'opère avant l'interprétation des résultats, de façon à assurer leur indépendance (*). Ce choix peut être influencé par la facilité de l'interprétation et il semble que certaines modalités de l'expérience, telles que la présence ou l'absence d'un niveau zéro, la nature quantitative ou qualitative des facteurs, peuvent également guider ce choix. D'autre part, le test séparé des composantes $AB\,J$ et $AB\,I$ de l'interaction AB, ainsi que celui des composantes homologues $AC\,J$, $AC\,I$, $BC\,J$ et $BC\,I$ permet, dans certains cas, de mettre le doigt sur de fausses interactions. Ces composantes, encore notées AB et AB^2 (Finney, 1947), correspondent respectivement au résultat de la subdivision d'un treillis cubique $3 \times 3 \times 3$ par les deux carrés latins orthogonaux

$$\begin{array}{ccc} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{array} \quad \text{et} \quad \begin{array}{ccc} 0 & 1 & 2 \\ 2 & 0 & 1 \\ 1 & 2 & 0 \end{array}$$

(*) Lorsque les indices 0, 1 et 2 des réponses Y désignent les niveaux, la composante linéaire d'un effet principal $= Y_2 - Y_0$; la quadratique $= (Y_0 + Y_2)/2 - Y_1$; la générale $= (Y_1 + Y_2)/2 - Y_0$; et l'additionnelle $= Y_2 - Y_1$.

appliqués, dans le plan *ab*, aux trois niveaux du troisième facteur localisés sur l'axe *c*. Les deux autres interactions du premier ordre se décomposent d'une façon analogue en se plaçant dans les plans *ac* et *bc*.

Dans l'analyse de la covariance, le calcul de

$$T_{yy} - [(T_{xy} + E_{xy})^2/(T_{xx} + E_{xx})] + (E_{xy}{}^2/E_{xx})$$

donne les carrés moyens réduits et les F réduits, après division par le carré moyen réduit de l'erreur qui est égal à

$$s'^2 = [E_{yy} - (E_{xy}{}^2/E_{xx})]/(\text{degrés de liberté résiduels} - 1)$$

Les notations utilisées sont celles de Cochran (1957).

En vue de permettre une interprétation correcte du tableau 3×3 de moyennes ajustées issu d'une interaction significative, le programme donne les erreurs-type des effets principaux et des interactions (Finney, 1946). Ainsi, l'erreur-type des 9 moyennes ajustées d'une interaction AB est:

$$\left[\frac{s'^2}{3}\left(1 + \frac{AB\,J_{xx} + AB\,I_{xx}}{4\,E_{xx}}\right)\right]^{1/2}$$

Le coefficient de régression E_{xy}/E_{xx}, le coefficient résiduel de corrélation totale $E_{xy}/(E_{xx} \cdot E_{xy})^{1/2}$, et le coefficient de variation réduit

$$CV' = \left[100\,s'\left(1 + \frac{A_{xx} + B_{xx} + C_{xx} + AB\,J_{xx} + AB\,I_{xx} + AC\,J_{xx} + AC\,I_{xx} + BC\,J_{xx} + BC\,I_{xx}}{18\,E_{xx}}\right)^{1/2}\right] / \bar{y}$$

sont également donnés (dans la dernière expression $\bar{y}$ est la moyenne générale). Le coefficient de variation est toutefois omis des résultats lorsque des composantes d'interaction à deux facteurs sont fusionnées avec l'interaction à trois facteurs.

Le dépouillement de plusieurs problèmes se fait simultanément, bien entendu. Le programme est mis en mémoire; le « dictionnaire » et les cartes de données passent sur une bande magnétique. Le temps de calcul est d'environ 15 secondes pour la variance et 30 secondes pour la covariance. L'impression des résultats de la variance couvre deux feuilles, celle de la covariance six feuilles de papier paravent pour tabulatrice.

En présence de plusieurs répétitions du dispositif 3^3, l'analyse séparée, répétition par répétition, constitue, par la juxtaposition des réponses, un atout majeur dans le dépistage de fausses interactions, et le test d'homogénéité des variances résiduelles précède ainsi, logiquement, l'analyse d'ensemble.

Les expérimentateurs, ayant accès à des ordinateurs IBM 704, peuvent se procurer une copie du programme en écrivant au Laboratoire de Calcul Numérique du C.N.R.S. - 23, rue du Maroc, Paris 19e.

SUMMARY

The variance analysis of a 3^3 experiment is easily performed on a desk calculator, even when all main effects and first order interactions are divided into single degrees of freedom components.

On the other hand, covariance analysis of a 3^3 leads to a rather entangled computing routine and, as far as we know, is seldom brought to completion. As the advantages of covariance hardly need be stressed, a program has been worked out for the IBM 704 computer.

Input of the data is by means of punched cards. Scaling factors make it possible to avoid decimal places. However, when necessary, transformation of the data has to be made beforehand on the data sheets. Moreover, transformed observations have their two-way tables of means expressed in the same units; therefore, a manual retransformation of these means is required after completion of the analysis. Input includes the definition of the error term. For example, all the components of the three-factor interaction in a design free of confounding, or three of them when the fourth is confounded with blocks. Occasionally twelve components of interaction are considered when no curvature is expected, possibly as a result of a poor choice of the levels.

A complete analysis of variance is provided by the program on the dependent variable as well as on the independent one. The analysis includes for each variable, the two-way tables of means with their standard errors, the sums of squares and F *values obtained after partitioning all the factorial effects into* linear + quadratic *and* general + additional *components. Albeit the experimenter has to choose between the two partitioning systems in order to secure independence of the results, he should find it useful to substantiate his choice on interpretation arguments. The presence or absence of a zero level, the quantitative versus qualitative nature of the factors, the interaction components themselves may be of some guidance. Also testing separately the* I- *and* J-*components of first order interactions seems rewarding in locating spurious interactions.*

The covariance program proceeds with the calculation of the reduced mean squares and corresponding F *values. To facilitate a correct interpretation of the tables of adjusted means in the presence of interaction, standard errors are estimated for each main effect and interaction. The regression coefficient, the residual coefficient of total correlation, and the effective reduced coefficient of variation are given.*

As many problems as desired may be run at one time and a shorter output is provided for the analysis of variance when no covariate is available. Computation time is approximately 30 seconds for a covariance and 15 seconds for a variance

analysis. When more than one replication of the 3^s is to be examined, a separate analysis of each replication informs us of the consistency of the responses and leads the way to a test of the homogeneity of the residual variances before any attempt is made towards an over-all analysis.

BIBLIOGRAPHIE

Cochran W.G., « Analysis of covariance: its nature and uses », *Biometrics*, 13, 261-281, 1957.

Finney D.J., « Standard errors of yields adjusted for regression on an independent measurement », *Biometrics*, 2, 53-55, 1946.

Finney D.J., « The construction of confounding arrangements », *Emp. J. Exp. Agric.*, 15, 107-112, 1947.

Van den Driessche R., « Analyse systématisée des expériences factorielles », *Biom.-Prax.*, 2, 245-259, 1961.

Yates F., *The design and analysis of factorial experiments*, T.C. n° 35 of the Comm. Bur. of Soils, Comm. Agric. Bur., Farnham Royal, Bucks., U.K., 1937.

Yates F., Lipton S., Sinha P. and Das Gupta K.P., « An exploratory analysis of a large set of $3 \times 3 \times 3$ fertilizer trials in India », *Emp. J. Exp. Agric.*, 27, 263-275, 1959.