

REVUE DE STATISTIQUE APPLIQUÉE

M. HAKIM

Analyse multivariante des plans d'expérience

Revue de statistique appliquée, tome 21, n° 2 (1973), p. 71-92

http://www.numdam.org/item?id=RSA_1973__21_2_71_0

© Société française de statistique, 1973, 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/>

ANALYSE MULTIVARIATE DES PLANS D'EXPÉRIENCE

M. HAKIM

Dans cet article on donne la généralisation de la méthode d'analyse de la variance des plans d'expérience, en utilisant la méthode des moindres carrés généralisée, au cas où les observations sont des vecteurs à p composantes, en supposant toujours les hypothèses classiques de normalité, d'indépendance des observations et d'une matrice commune des variances et covariances.

Dans une première partie on considère le modèle linéaire général et dans la deuxième partie on applique les résultats trouvés aux plans d'expérience usuels.

1. GENERALITES SUR LES MODELES LINEAIRES :

1.1. — Estimation des paramètres :

Considérons le modèle linéaire général :

$$E(Y) = BA \quad (I.1.1)$$

où :

$Y(p, N)$ est la matrice des observations des variables dépendantes

$$y_{\alpha}(p, 1) \quad \alpha = 1, 2, \dots, N.$$

$A(q, N)$ est la matrice des valeurs connues des variables indépendantes. Ces valeurs étant dans le cas d'un plan d'expérience : 1 pour la présence de la modalité du facteur considéré et 0 pour son absence.

$B(p, q)$ est la matrice des coefficients inconnus des variables indépendantes ou matrice des paramètres du plan dans le cas des plans d'expérience.

Exemple :

Pour clarifier ces notations donnons ici un exemple de modèle linéaire général, un autre exemple de plan d'expérience est donné à la fin du § II.1.

On étudie l'effet d'un engrais sur le poids de grains et le poids de paille dans une production de blé ([1] p.p. 227)

On désigne par y^1 et y^2 respectivement le poids de grains et le poids de paille obtenus avec la quantité a_α d'engrais $\alpha = 1, 2 \dots, 8$.

Dans le modèle

$$E y_\alpha = a_\alpha b + \mu \quad \alpha = 1, 2 \dots, 8$$

avec

$$y_\alpha = \begin{pmatrix} y_\alpha^1 \\ y_\alpha^2 \end{pmatrix} \quad b = \begin{pmatrix} b^1 \\ b^2 \end{pmatrix} \quad \mu = \begin{pmatrix} \mu^1 \\ \mu^2 \end{pmatrix}$$

Qu'on résume par $EY = BA$

On a :

$$B = (b, \mu) = \begin{pmatrix} b^1 & \mu^1 \\ b^2 & \mu^2 \end{pmatrix}$$

$$A = \begin{pmatrix} a_\alpha \\ 1 \end{pmatrix} = \begin{pmatrix} 24 & 11 & 5 & 12 & 7 & 14 & 11 & 18 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$Y = (y_\alpha) = \begin{pmatrix} 40 & 17 & 9 & 15 & 6 & 12 & 5 & 9 \\ 53 & 19 & 10 & 29 & 13 & 27 & 19 & 30 \end{pmatrix}$$

On suppose que les variables y_α , $\alpha = 1, 2 \dots$, sont indépendantes et suivent toutes une loi normale de même matrice de covariance Σ inconnue. On suppose aussi que le rang r de la matrice du plan A est tel que

$$n = N-r \geq p$$

Soit $E(p,p)$ la matrice erreur autour du modèle :

$$E = (Y-BA)(Y-BA)' = \begin{pmatrix} e_j^i \end{pmatrix} \quad (1.1.2)$$

avec

$$e_j^i = \sum_{\alpha=1}^N \left(y_\alpha^i - \sum_{k=1}^q a_\alpha^k b_k^i \right) \left(y_\alpha^j - \sum_{k=1}^q a_\alpha^k b_k^j \right)$$

1) Estimation de B :

Pour estimer B on utilise la méthode des moindres carrés généralisée qui consiste à minimiser, par rapport aux éléments de B, le déterminant de la matrice erreur, |E|.

On obtient [5] :

$$\frac{\delta |E|}{\delta B} = - 2 |E| E^{-1} (YA' - BAA') = 0_{(p, q)}$$

Soit

$$\hat{B} A A' = YA' \quad (1.1.3)$$

On montre ([1], [5]) qu'on obtient les mêmes équations (1.1.3) par la méthode du maximum de vraisemblance.

Pour résoudre le système (1.1.3), [15], si la matrice A est régulière c.a.d $r = q < N$, alors AA' est inversible et l'estimation \hat{B} de B est donnée par :

$$\hat{B} = YA' (AA')^{-1} \quad (1.1.4)$$

Dans les plans usuels, la matrice A est singulière ; $r < q$; On impose alors, pour avoir l'unicité de la solution du système (1.1.3), des conditions aux paramètres du plan qu'on peut formuler par :

$$\begin{matrix} B & K \\ (p, q) & (q, 1) \end{matrix} = \begin{matrix} 0 \\ (p, 1) \end{matrix}$$

Où K est la matrice des conditions sur B supposées vraies, de rang $1 = q - r$ et telle que la matrice

$$\begin{matrix} G \\ (q, N + 1) \end{matrix} = (A \ K) \text{ soit de rang } q.$$

Alors le système donnant \hat{B} s'écrit

$$\hat{B} AA' = YA' \text{ et } \hat{B} KK' = 0$$

Soit

$$\hat{B} (AA' + KK') = \hat{B} GG' = YA'$$

Or G est régulière, GG' est inversible, d'où :

$$\hat{B} = YA' (GG')^{-1}. \quad (1.1.5)$$

On montre [5] que les \hat{b}_j^i ainsi trouvées sont des estimateurs efficaces des b_j^i

2) Estimation de Σ :

On applique la méthode du maximum de vraisemblance ; on trouve [1] :

$$\hat{\Sigma} = \frac{1}{N} \hat{E} \quad (1.1.6)$$

Où \hat{E} est la matrice erreur autour du modèle :

$$\begin{aligned} \hat{E} &= (Y - \hat{B}A) (Y - \hat{B}A)' = YY' - YA'\hat{B}' \\ &= YY' - \hat{B} AA' \hat{B}' \end{aligned} \quad (1.1.7)$$

La matrice \hat{E} a une distribution de Wishart $W(n, \Sigma)$ ($n = N - r > p$) de densité [1] :

$$\frac{|\hat{E}|^{1/2} (n-p-1)^{p/2} \exp - 1/2 \text{tr} \Sigma^{-1} \hat{E}}{2^{np/2} \Pi^{\frac{p(p-1)}{4}} |\Sigma|^{n/2} \prod_{i=1}^p \Gamma \left[\frac{(n+1-i)}{2} \right]} \quad (1.1.8)$$

1.2. Test d'hypothèses :

Les hypothèses qu'on désire tester en analyse de la variance sont des hypothèses linéaires sur la matrice des paramètres B (signification des effets et comparaisons de moyennes). On peut les formuler par :

$$H_0 : \underset{(p, q)}{BC} = \underset{(q, s)}{0} \quad (1.2.1)$$

Où $C(q, s)$ est une matrice, supposée de rang $s < q$, formée de 0,1 et - 1 contre l'alternative :

$$H_1 : BC = \eta \text{ quelconque.} \quad (1.2.2)$$

Basé sur le rapport de maximum de vraisemblance, on réestime B dans le modèle (1.1.1) avec les conditions (1.2.1) en utilisant les multiplicateurs de Lagrange pour les $p \times s$ éléments de BC ; on minimise alors, par rapport aux éléments de B, la quantité :

$$Q = |(Y-BA)(Y-BA)'| + \sum_{i=1}^p \sum_{j=1}^s \lambda_j^i \sum_{k=1}^q b_k^i c_j^k \quad (1.2.3)$$

On obtient [5] le minimum pour B_0 :

$$B_0 = \hat{B} - \hat{B}C [C'(AA')^{-1}C]^{-1} C'(AA')^{-1} \quad (1.2.4)$$

Où \hat{B} est donnée par (1.1.4) (en supposant A régulière)

La matrice erreur E_o dans (1.1.1) et (1.2.1) s'écrit alors :

$$\begin{aligned} E_o &= (Y - B_o A) (Y - B_o A)' = YY' - B_o AY' \\ &= YY' - \hat{B} AY' + \hat{B}C [C' (AA')^{-1} C]^{-1} C' \hat{B}' \end{aligned}$$

Soit

$$E_o = E + H \quad (1.2.5)$$

avec

$$\begin{aligned} H &= \hat{B} C [C' (AA')^{-1} C]^{-1} C' \hat{B}' \\ &= Y A' (AA')^{-1} C [C' (AA')^{-1} C]^{-1} C' (AA')^{-1} AY' \end{aligned} \quad (1.2.6)$$

H est appelée matrice de l'hypothèse H_o . D'après le théorème de Cochran généralisé [5], H suit, si H_o est vraie, la même distribution que $\sum_{i=1}^s x_i x_i'$ où $x_i \sim N(0, \Sigma)$ qu'on note dans la suite distribution de Wishart généralisée à d.d./ et qui suit une distribution $W(s, \Sigma)$ si $s \geq p$.

Pour tester H_o contre H_1 , ayant E et H, plusieurs critères sont possibles dont on cite 3 d'entre eux :

Rejeter H_o au risque α si

$$\left\{ \begin{array}{l} C_1 : U = \frac{|E|}{|E + H|} \leq U_\alpha(p, s, n) \\ C_2 : V = \lambda_{\max} H(E + H)^{-1} \geq V_\alpha(t, \nu_1, \nu_2) \\ C_3 : W = \text{trace } H E^{-1} \geq W_\alpha \end{array} \right. \quad (1.2.7)$$

Avec

$$\text{Prob. } [U \leq U_\alpha/H_o] = \text{Prob. } [V \geq V_\alpha/H_o] = \text{Prob. } [W \geq W/H_o] = \alpha$$

et où λ_{\max} désigne le maximum des valeurs propres.

C_1 est dû à Wilks [16] et est basé sur le rapport des vraisemblances maximales ($\Lambda = U^{N/2}$), C_2 est due à S.N. Roy [12] et C_3 à Hotelling H. et Lawley D.N. ([7], [9]).

Les distributions explicites de U, V et W ne sont pas données pour toutes les valeurs de p, s et n même quand H_o est vraie : Anderson T.W [1] a donné la distribution de U sous forme d'un produit de variables β_i , et a montré que cette

distribution reste inchangée si on change p, s et n en $p' = s, s' = p$ et $n' = n + s - p$. On peut donc supposer $s \geq p$ sans nuire à la généralité. Il a donné les valeurs de $U_\alpha(p, s, n)$ à partir des tables de Fisher-Snedecor pour $p = 1$ et 2 et dans une table si $p = s = 3$ et $p = s = 4, n \leq 10$. Pour n grand on peut utiliser la distribution asymptotique : $m \log U$ suit si H_0 est vraie un χ^2 à ps d.d.l. où $m = 1/2(p + 1 - s - 2n)$.

D.I. Heck [6] a donné les $V_\alpha(t, \nu_1, \nu_2)$ où $t = \min(p, s) \nu_1 = 1/2(|s-p|-1) \nu_2 = 1/2(n-p-1)$, dans des abaques pour $\alpha = 0,01; 0,05$; t de 2 à 5; ν_1 de $-1/2$ à 10 et ν_2 de 5 à 100.

Pour W , on a que NW suit asymptotiquement un χ^2 à ps d.d.l.

Pour la puissance de ces 3 tests, il est connu [1] qu'elle dépend au plus de $t = \min(s, p)$ paramètres de décentralisation qu'on peut prendre égaux aux valeurs propres non nulles de $\eta^1 \Sigma^{-1} \eta$ soit d_1, d_2, \dots, d_t . Cette puissance est fonction monotone croissante de d_1, d_2, \dots, d_t dans les 3 tests ([3], [11], [14]).

Ces tests sont sans biais, mais aucun d'eux n'est uniformément plus puissant que les 2 autres : pour cela on rejette H_0 si un de ces tests est significatif.

Remarque :

Si on veut tester des hypothèses sur h composantes des paramètres ($h < p$) seulement, on peut utiliser une matrice $M(h, p)$ formée des lignes correspondantes de la matrice unité $I(p, p)$ et modifier l'hypothèse (1.2.1) sous la forme $MBC = 0(h, s)$.

Exemple :

On traite les données de l'exemple 1 § 1 :

Pour $\zeta' = (\zeta_1, \dots, \zeta_N)$ on désigne par :

$$\zeta = \frac{1}{N} \sum_{\alpha=1}^N \zeta_\alpha, \quad s_\zeta^2 = \frac{1}{N} \sum_{\alpha} (\zeta_\alpha - \bar{\zeta})^2$$

et pour

$$\eta' = (\eta_1, \dots, \eta_N)$$

par :

$$s_{\zeta \eta} = \frac{1}{N} \sum_{\alpha} (\zeta_\alpha - \bar{\zeta}) (\eta_\alpha - \bar{\eta}) .$$

On a :

$$(AA')^{-1} = \frac{1}{Ns_a^2} \begin{pmatrix} 1 & -\bar{a} \\ -\bar{a} & \frac{\sum a_\alpha^2}{N} \end{pmatrix} = \frac{1}{255,5} \begin{pmatrix} 1 & -12,75 \\ -12,75 & 194,5 \end{pmatrix}$$

$$YA' = \begin{pmatrix} \sum_{\alpha} y_{\alpha}^1 a_{\alpha} & N\bar{y}^1 \\ \sum_{\alpha} y_{\alpha}^2 a_{\alpha} & N\bar{y}^2 \end{pmatrix} = \begin{pmatrix} 1799 & 200 \\ 3097 & 113 \end{pmatrix}$$

D'où l'estimation \hat{B} de B :

$$\hat{B} = YA' (AA')^{-1} \begin{pmatrix} \hat{b}^1 & \hat{\mu}^1 \\ \hat{b}^2 & \hat{\mu}^2 \end{pmatrix} = \begin{pmatrix} \frac{s_{ay}^1}{s_a^2} & \bar{y}^1 - \hat{b}^1 \bar{a} \\ \frac{s_{ay}^2}{s_a^2} & \bar{y}^2 - \hat{b}^2 \bar{a} \end{pmatrix}$$

Soit

$$\hat{B} = \begin{pmatrix} 1,4 & -3,75 \\ 2,14 & -2,3 \end{pmatrix}$$

On trouve pour la matrice erreur \hat{E} :

$$\hat{E} = YY' - YA'\hat{B}' = N \begin{pmatrix} s_{y^1}^2 - \frac{s_{y^1 a}^2}{s_a^2} & s_{y^1 y^2} - \frac{s_{y^1 a} s_{y^2 a}}{s_a^2} \\ s_{y^1 y^2} - \frac{s_{y^1 a} s_{y^2 a}}{s_a^2} & s_{y^2}^2 - \frac{s_{y^2 a}^2}{s_a^2} \end{pmatrix}$$

Soit :

$$\hat{E} = 8 \begin{pmatrix} 47,81 & 17,36 \\ 17,36 & 10,81 \end{pmatrix}$$

Pour tester l'hypothèse

$$H_0 : b = \begin{pmatrix} b^1 \\ b^2 \end{pmatrix} = 0$$

qui s'écrit $H_0 : B C = 0$ avec $C = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

On a :

$$C' (AA')^{-1} C = \frac{1}{N s_{a_2}}$$

D'où la matrice H de cette hypothèse :

$$H = \hat{B} C [C'(AA')^{-1} C] C' \hat{B}' = N s_{a_2} \hat{b} \hat{b}'$$

Soit :

$$H = 8 \begin{pmatrix} 62,60 & 95,68 \\ 95,68 & 146,26 \end{pmatrix}$$

et

$$H E^{-1} = \begin{pmatrix} -4,57 & 16,19 \\ -6,98 & 24,74 \end{pmatrix}$$

Or ici H est singulière et de rang $s = 1$, donc les valeurs propres de HE^{-1} sont :

$$\lambda_2 = 0 \quad \text{et} \quad \lambda_1 = \text{tr. } H E^{-1} = 20,17$$

Et les trois critères de (1.2.7) :

$$U = \prod_i \frac{1}{(1 + \lambda_i)} = \frac{1}{1 + \lambda_1}$$

$$V = \frac{\max \lambda_i}{1 + \max \lambda_i} = \frac{\lambda_1}{1 + \lambda_1}$$

$$W = \sum_i \lambda_i = \lambda_1$$

sont équivalents. On a que :

$$\frac{n+1-p}{p} \times \frac{1-U_{p,1,n}}{U_{p,1,n}}$$

suit un F (p, n + 1 - p)

Soit ici p = 2, n = 8 - 2 = 6 et $\frac{1-U}{U} = \lambda_1$:

$\frac{5}{2}\lambda_1$ suit un F(2, 5) ici $\frac{5}{2}\lambda_1 = 50,4$ donc hautement significatif, et on rejette l'hypothèse H_0 (avec un risque inférieur à 1 %).

1.3. Equation d'analyse de la variance généralisée :

Dans ce paragraphe on va donner une décomposition de la matrice YY' , qui généralise l'équation d'analyse de la variance dans le cas univariate, et qui permet d'avoir la matrice H pour les hypothèses classiques en analyse de variance.

Supposons que dans le modèle (1.1.1) on fasse une partition de la matrice $B(p, q)$ en l sous matrices $B^i(p, q_i)$ $i = 1, 2, \dots, l$ (avec $\sum_{i=1}^l q_i = q$) qui correspondent à l facteurs :

$$B = (B^1 \ B^2 \ \dots \ B^l) \quad (1.3.1)$$

et que l'on décompose A de la même façon en l sous matrices $A_i(q_i, N)$:

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_l \end{bmatrix} \quad (1.3.1)'$$

alors on obtient :

$$BA = B^1 A_1 + B^2 A_2 + \dots + B^l A_l$$

et

$$A'B' = (A_1)'(B^1)' + (A_2)'(B^2)' + \dots + (A_l)'(B^l)'$$

La matrice E peut s'écrire alors :

$$E = YY' - Y(A_1)'(\hat{B}^1)' - Y(A_2)'(\hat{B}^2)' \dots - Y(A_i)'(\hat{B}^i)'$$

On obtient donc la relation :

$$YY' = E + Y(A_1)'(\hat{B}^1)' + Y(A_2)'(\hat{B}^2)' + \dots + Y(A_i)'(\hat{B}^i)' \quad (1.3.2)$$

Si on désigne par b_{i_s} , $s = 1, \dots, q_i$, les q_i colonnes de la matrice B^i représentant les q_i niveaux du $i^{\text{ème}}$ facteur, on a :

$$A_i(q_i, N) = (a_{i_s \alpha})$$

Avec $a_{i_s \alpha} = 1$ si y_α est obtenu avec une combinaison faisant intervenir le niveau i_s .

$$a_{i_s \alpha} = 0 \text{ dans le cas contraire.}$$

Ce qui fait qu'une colonne T_{i_s} de $Y(A_i)'(pxq_i)$ représente la somme des résultats y_α dans la classe correspondante au $i_s^{\text{ème}}$ niveau du $i^{\text{ème}}$ facteur pour $s = 1, \dots, q_i$. Soit donc :

$$Y(A_i)'(\hat{B}^i)' = \sum_{s=1}^{q_i} T_{i_s} \hat{b}_{i_s} \quad (1.3.3)$$

2. ANALYSE MULTIVARIATE DES PLANS USUELS

2.1. Plan factoriel à 2 facteurs : "ligne" à 1 niveaux et "colonne" à k niveaux.

Soit y_{ij} le vecteur à p composantes, observé à la $i^{\text{ème}}$ ligne et $j^{\text{ème}}$ colonne. Considérons le modèle linéaire sans interaction :

$$E(y_{ij}) = \mu + \alpha_i + \beta_j \quad (2.1.1)$$

avec

$$\sum_{i=1}^1 \alpha_i = \sum_{j=1}^k \beta_j = 0(p, 1) \quad (2.1.2)$$

1) Estimation des paramètres μ , α_i , β_j .

On a conformément aux notations du § 1.1. :

$$Y(p, kl) = (y_{ij}) \text{ où } j \text{ croît plus vite que } i$$

$$B(p, k + l + 1) = \mu_j \alpha_1 \dots \alpha_1 : \beta_1 \dots \beta_k$$

et A et K sont données par :

$$A = \begin{array}{c} 1 \\ \vdots \\ l \\ (k+l+1, kl) \\ \vdots \\ k \end{array} \left[\begin{array}{ccc|cc|cc} & k & & & k & & & & k \\ 1 & 1 & 1 & 1 & & 1 & & 1 & 1 \\ \hline 1 & 1 & 0 & 0 & 0 & 0 & & 0 & 0 \\ 0 & & 0 & 1 & 1 & & & & 0 \\ \vdots & & & & & & & & \\ 0 & & 0 & 0 & 0 & 0 & 1 & 1 & \\ \hline 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ \vdots & & & & & & & & \\ 0 & & 1 & 0 & 1 & 0 & 0 & 1 & \end{array} \right]$$

$$K = \begin{array}{c} 1 \\ \vdots \\ l \\ (k+l+1, 2) \\ \vdots \\ k \end{array} \left[\begin{array}{cc} 0 & 0 \\ \hline 1 & 0 \\ \vdots & \\ 1 & 0 \\ \hline 0 & 0 \\ \vdots & \\ 0 & l \end{array} \right]$$

d'où :

$$A A' = \begin{array}{c} (k+1+1, k+1+1) \end{array} \left[\begin{array}{ccc|cc} k+1 & k & & 1 & 1 \\ \hline k & k & & 1 & 1 \\ & & k & 0 & \\ k & 0 & & k & 1 \\ \hline 1 & 1 & 1 & 0 & l \\ & & & & 1 \\ 1 & 1 & 1 & 1 & 0 \end{array} \right]$$

En désignant par L_i , C_j et G les sommes respectivement dans la $i^{\text{ème}}$ ligne, $j^{\text{ème}}$ colonne et le total général

$$L_i = \sum_{j=1}^k y_{ij} \quad C_j = \sum_{i=1}^1 y_{ij} \quad \text{et} \quad G = \sum_{ij} y_{ij}$$

On a :

$$Y A' = (G : L_1 \quad L_2 \quad \dots \quad L_l : C_1 \quad C_2 \quad \dots \quad C_k)$$

Compte tenu des conditions (2.1.2) :

$$BAA' = (k1 \mu : k \mu + \alpha_1 : k \mu + \alpha_2 \dots k \mu + \alpha_l : l \mu + \beta_1 \dots l \mu + \beta_k)$$

(y, k + l + 1)

Et les équations normales (1.1.3) s'écrivent alors en désignant par m, a_i, b_j les estimateurs de μ, α_i, β_j respectivement :

$$\begin{aligned} G - k1 m &= 0 \\ L_i - k m - k a_i &= 0 \quad i = 1, 2, \dots, l \\ C_j - l m - l b_j &= 0 \quad j = 1, 2, \dots, k \end{aligned}$$

Soit

$$\begin{aligned} m &= \frac{1}{k1} G = y_{..} \\ a_i &= \frac{1}{k} L_i - m = y_{i.} - y_{..} \quad i = 1, 2, \dots, l \\ b_j &= \frac{1}{l} C_j - m = y_{.j} - y_{..} \quad j = 1, 2, \dots, k \end{aligned} \tag{2.1.3}$$

2) Test de signification des effets :

Les hypothèses HL_0 : pas d'effet ligne et HK_0 : pas d'effet colonne se formalisent par :

$$HL_0 : \alpha_1 = \alpha_2 \dots = \alpha_l = 0(p, 1)$$

$$HK_0 : \beta_1 = \beta_2 \dots = \beta_k = 0(q, 1)$$

L'équation d'Analyse de la variance généralisée, d'après (1.3.2) et (1.3.3) est :

$$YY' = E + \sum_{i=1}^l L_i a_i' + \sum_{j=1}^k C_j b_j' + Gm' \tag{2.1.4}$$

d.d.l. $kl = (k-1)(l-1) + (l-1) + (k-1) + 1$

Soit

$$\begin{aligned}
 YY' &= \sum_{ij} (y_{ij} - y_{i.} - y_{.j} + y_{..}) (y_{ij} - y_{i.} - y_{.j} + y_{..}) \\
 &+ k \sum_{i=1}^l (y_{i.} - y_{..}) (y_{i.} - y_{..})' + l \sum_{j=1}^k (y_{.j} - y_{..}) (y_{.j} - y_{..})' + G m'
 \end{aligned}$$

D'après le théorème de Cochran généralisé [5] on a :

$$E = \sum_{ij} (y_{ij} - y_{i.} - y_{.j} + y_{..}) (y_{ij} - y_{i.} - y_{.j} + y_{..})'$$

suit dans le modèle (2.1.1) une distribution de Wishart $W(n, \Sigma)$ avec $n = (k - 1)(l - 1)$ d.d.l. ($n > p$)

$$H_1 = k \sum_{i=1}^l (y_{i.} - y_{..}) (y_{i.} - y_{..})'$$

est la matrice de l'hypothèse HL_0 , et suit, si cette hypothèse est vraie, une distribution de Wishart généralisée à $(l - 1)$ d.d.l.

$$H_k = l \sum_{j=1}^k (y_{.j} - y_{..}) (y_{.j} - y_{..})'$$

est la matrice de l'hypothèse HK_0 et suit, si cette hypothèse est vraie, une distribution de Wishart généralisé à $(k - 1)$ d.d.l.

Et E, H_1, H_k sont indépendantes en probabilité

Pour tester HL_0 et HK_0 on procède à un critère C_1, C_2 ou C_3 de (1.2.7) en prenant pour matrice H ; H_1 pour HL_0 avec $s = l - 1$ et H_k pour HK_0 avec $s = k - 1$.

Exemple :

Les observations y_{ij}^1 et y_{ij}^2 sont respectivement la production d'orge la 1^{ère} et la 2^{ème} année dans la localité i et avec la variété j . $i = 1, 2, \dots, 6$; $j = 1, 2, \dots, 5$. (cf. [1] page 217).

On a le tableau des observations :

i \ j	1	2	3	4	5
1	81	105	120	110	98
	81	882	80	87	84
2	147	142	151	192	146
	100	116	112	148	108
3	82	77	78	131	90
	103	105	117	140	130
4	120	121	124	141	125
	90	62	96	126	76
5	99	89	69	89	104
	66	50	97	62	80
6	87	77	79	102	96
	68	67	67	92	94

Pour l'estimation des paramètres on a :

$$m = y_{..} = \begin{pmatrix} 109,6 \\ 93,16 \end{pmatrix}$$

Les a_i données par $a_i = y_{i.} - y_{..}$ sont :

$$\begin{array}{cccccc} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ \begin{pmatrix} -6,26 \\ -10,36 \end{pmatrix} & \begin{pmatrix} 46,53 \\ 23,63 \end{pmatrix} & \begin{pmatrix} -17,46 \\ 25,83 \end{pmatrix} & \begin{pmatrix} -17,13 \\ -1,36 \end{pmatrix} & \begin{pmatrix} -19,06 \\ -22,16 \end{pmatrix} & \begin{pmatrix} -20,86 \\ -15,56 \end{pmatrix} \end{array}$$

Les b_j données par $b_j = y_{.j} - y_{..}$ sont :

$$\begin{array}{ccccc} b_1 & b_2 & b_3 & b_4 & b_5 \\ \begin{pmatrix} -6,4 \\ -7 \end{pmatrix} & \begin{pmatrix} -7,23 \\ -12,83 \end{pmatrix} & \begin{pmatrix} -5,56 \\ 1,66 \end{pmatrix} & \begin{pmatrix} 18,43 \\ 16 \end{pmatrix} & \begin{pmatrix} 0,76 \\ 2,16 \end{pmatrix} \end{array}$$

Pour le test on trouve :

$$E = \begin{pmatrix} 3.279 & 802 \\ 802 & 4.017 \end{pmatrix} \sim W(20, \Sigma)$$

$$H_1 = \begin{pmatrix} 18.011 & 7.188 \\ 7.188 & 10.345 \end{pmatrix} \sim W(5, \Sigma)$$

$$H_k = \begin{pmatrix} 2.788 & 2.550 \\ 2.550 & 2.863 \end{pmatrix} \sim W(4, \Sigma)$$

D'où :

1) Pour tester l'hypothèse HL_0 : par d'effet "ligne" (ou pas d'influence de la localité), on a :

Les valeurs propres de $H_1 E^{-1}$ sont :

$$\lambda_1 = 5,66 \quad \text{et} \quad \lambda_2 = 1,90$$

D'où avec le critère U on a :

$$U = \frac{E}{E + H_1} = \prod_i \frac{1}{1 + \lambda_i} = \frac{1}{1 + \lambda_1} = \frac{1}{1 + \lambda_2} = 0,052$$

qui est à comparer avec $U_\alpha(2,5, 20)$

Or

$$\frac{1 - \sqrt{U_{2,5,20}}}{\sqrt{U_{2,5,20}}} \times \frac{19}{5} \text{ suit un } F(10,38) \quad (\text{cf. [1]})$$

Ce qui donne

$$U_\alpha(2,5,20) = \begin{cases} 0,42 & \text{à } \alpha = 5 \% \\ 0,33 & \text{à } \alpha = 1 \% \end{cases}$$

Donc on rejette l'hypothèse HL_0 même à 1 %.

avec le critère V on a :

$$V = \frac{\lambda_1}{1 + \lambda_1} = 0,85$$

alors que

$$V_{\alpha} \left(2, 1, \frac{17}{2} \right) = \begin{cases} 0,55 & \text{à } \alpha = 5 \% \\ 0,58 & \text{à } \alpha = 1 \% \end{cases} \quad (\text{cf. [6]})$$

Aussi on rejette H_{L_0} même à 1 %.

2) Pour tester l'hypothèse H_{K_0} : pas d'effet "colonne" (ou pas de différence significative entre les variétés) on a :

Les valeurs propres de $H_k E^{-1}$ sont :

$$\lambda_1 = 1,22 \quad \text{et} \quad \lambda_2 = 0,10$$

D'où avec le critère U, on a :

$$U = 0,41$$

et

$$U_{\alpha} (2, 4, 20) = \begin{cases} 0,46 & \text{à } \alpha = 5 \% \\ 0,33 & \text{à } \alpha = 1 \% \end{cases}$$

Donc l'hypothèse H_{K_0} est à rejeter à 5 % (différence significative entre les variétés) et pas de différence significative à 1 %.

avec le critère V on a :

$$V = 0,55$$

et

$$V \left(2, 1/2, \frac{17}{2} \right) = \begin{cases} 0,53 & \text{à } \alpha = 5 \% \\ 0,57 & \text{à } \alpha = 1 \% \end{cases}$$

Donc aussi même conclusion qu'avec U.

2.2. Plan factoriel à plusieurs facteurs orthogonaux :

Une analyse semblable peut être effectuée dans le cas d'un plan factoriel à plusieurs facteurs orthogonaux.

Considérons par exemple un plan à 3 facteurs, le premier à n_1 niveaux, le 2^{ème} à n_2 niveaux et le 3^{ème} à n_3 niveaux, avec une seule observation par case c.à.d. par combinaison d'un niveau de chaque facteur. Soit y_{ijk} ($p, 1$) le vecteur observé avec le $i^{\text{ème}}$ niveau du 1^{er} facteur, le $j^{\text{ème}}$ du second facteur et de $k^{\text{ème}}$ du 3^{ème} facteur et considérons le modèle linéaire suivant :

$$E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k + \theta_{ij} + \eta_{ik} + \xi_{jk} \quad (2.2.1)$$

avec

$$\sum_{i=1}^{n_1} \alpha_i = \sum_{j=1}^{n_2} \beta_j = \sum_{k=1}^{n_3} \gamma_k = 0(p, 1)$$

$$\sum_{i=1}^{n_1} \theta_{ij} = \sum_{i=1}^{n_1} \eta_{ik} = 0(p, 1) \quad \forall j \text{ et } k$$

(2.2 2)

$$\sum_{j=1}^{n_2} \theta_{ij} = \sum_{j=1}^{n_2} \zeta_{jk} = 0(p, 1) \quad \forall i, k$$

$$\sum_{k=1}^{n_3} \eta_{ik} = \sum_{k=1}^{n_3} \zeta_{jk} = 0(p, 1) \quad \forall i, j$$

Ici on a :

$Y(p, n_1 \times n_2 \times n_3) = (y_{ijk})$ où k croit le plus vite, ensuite j puis i .

$$B = (\mu : \alpha_i : \beta_j : \gamma_k : \theta_{ij} : \eta_{ik} : \zeta_{jk})$$

$$(p, 1 + n_1 + n_2 + n_3 + n_1 n_2 + n_1 n_3 + n_2 n_3)$$

et A est formée de 0 et de 1, facile mais longue à écrire.

En résolvant les équations (1.1.3) dans ce cas, on obtient les estimations $m, a_i, b_j, c_k, t_{ij}, U_{ik}, V_{jk}$ respectivement de $\mu, \alpha_i, \beta_j, \gamma_k, \theta_{ij}, \eta_{ik}, \zeta_{jk}$ suivantes :

$$m = \frac{1}{n_1 n_2 n_3} \sum_{i, j, k} y_{ijk} = y_{...}$$

$$a_i = \frac{1}{n_2 n_3} \sum_{j, k} y_{ijk} - m = y_{i..} - y_{...}$$

$$b_j = \frac{1}{n_1 n_3} \sum_{i, k} y_{ijk} - m = y_{.j.} - y_{...}$$

$$c_k = \frac{1}{n_1 n_2} \sum_{i, j} y_{ijk} - m = y_{..k} - y_{...}$$

$$t_{ij} = \frac{1}{n_3} \sum_k y_{ijk} - m - a_i - b_j = y_{ij.} - y_{i..} - y_{.j.} + y_{...}$$

$$u_{ik} = \frac{1}{n_2} \sum_j y_{ijk} - m - a_i - c_k = y_{i.k} - y_{i..} - y_{..k} + y_{...}$$

$$v_{jk} = \frac{1}{n_1} \sum_i y_{ijk} - m - b_j - c_k = y_{.jk} - y_{.j.} - y_{..k} + y_{...}$$

2) Test de signification des effets

L'équation d'analyse de la variance généralisée s'écrit dans ce cas :

$$YY' = E + H_0 + H_\eta + H_\zeta + H_\alpha + H_\beta + H_\gamma + G m' \quad (2.2.4)$$

où

$$E = \sum_{ijk} (y_{ijk} - y_{ij.} - y_{i.k} - y_{.jk} + y_{ij.} + y_{.j.} + y_{..k} - y_{...}) ()'$$

est la matrice erreur autour du modèle (2.2.1) et a une distribution de Wishart

$$W(n, \Sigma) \text{ avec } n = (n_1 - 1)(n_2 - 1)(n_3 - 1) \text{ d.d.l. } (n \geq p).$$

$$H_0 = n_3 \sum_{ij} (y_{ij.} - y_{i..} - y_{.j.} + y_{...}) (\leftarrow)'$$

est la matrice de l'hypothèse de non interaction du 1^{er} et 2^{ème} facteur :

$$\theta_{ij} = 0 \quad \forall i, j, \text{ ayant une distribution de Wishart généralisée à}$$

$$s_1 = (n_1 - 1)(n_2 - 1) \text{ d.d.l.}$$

$$H_\eta = n_2 \sum_{ik} (y_{i.k} - y_{i..} - y_{..k} + y_{...}) (\leftarrow)'$$

est la matrice de l'hypothèse $\eta_{ik} = 0 \quad \forall i, k$ distribuée comme un Wishart généralisé avec $s_2 = (n_1 - 1)(n_3 - 1) \text{ d.d.l.}$

$$H_\zeta = n_1 \sum_{jk} (y_{.jk} - y_{.j.} - y_{..k} + y_{...}) (\leftarrow)'$$

est la matrice de l'hypothèse $\zeta_{jk} = 0$ avec $s_3 = (n_2 - 1)(n_3 - 1) \text{ d.d.l.}$

$$H_\alpha = n_2 n_3 \sum_i (y_{i..} - y_{...}) (y_{i..} - y_{...})'$$

est la matrice de l'hypothèse $\alpha_i = 0$ avec $s_4 = (n_1 - 1)$ d.d.l.

$$H_\beta = n_1 n_3 \sum_j (y_{.j} - y_{...}) (y_{.j} - y_{...})'$$

est la matrice de l'hypothèse $\beta_j = 0$ avec $s_5 = (n_2 - 1)$ d.d.l.

$$H_\gamma = n_1 n_2 \sum_k (y_{..k} - y_{...}) (y_{..k} - y_{...})'$$

est la matrice de l'hypothèse $\gamma_k = 0$ avec $s_6 = (n_3 - 1)$ d.d.l.

Pour tester chacune de ces hypothèse on applique l'un des critères C_1 , C_2 ou C_3 de (1.2.7).

2.3. Plan en carré latin :

Dans un plan en carré latin (3 facteurs "ligne", "colonne" et "traitement" à c niveaux chacun, chaque combinaison des 3 facteurs se trouve une et une seule fois) le modèle s'écrit :

$$E(y_{ij}) = \mu + \alpha_i + \beta_j + \gamma_k \quad k \in (ij) \quad (2.3.1)$$

avec

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = 0 \quad (2.3.2)$$

1) Estimation des paramètres :

Appliquons directement la méthode des moindres carrés généralisée, on a :

$$E = \sum_{ij} (y_{ij} - \mu - \alpha_i - \beta_j - \gamma_k) (y_{ij} - \mu - \alpha_i - \beta_j - \gamma_k)'$$

d'où

$$\frac{\delta |E|}{\delta \mu} = -2 |E| E^{-1} \sum_{ij} (y_{ij} - \mu - \alpha_i - \beta_j - \gamma_k) = 0$$

$$\frac{\delta |E|}{\delta \alpha_i} = -2 |E| E^{-1} \sum_j (y_{ij} - c^\mu - c \alpha_i - \sum_j \beta_j - \sum_k \gamma_k) = 0$$

$$\frac{\delta |E|}{\delta \beta_j} = -2 |E| E^{-1} \sum_i (y_{ij} - c^\mu - \sum_i \alpha_i - c \beta_j - \sum_k \gamma_k) = 0$$

$$\frac{\delta |E|}{\delta \gamma_k} = -2 |E| E^{-1} \sum_{(k)} (y_{ij} - c^\mu - \sum_i \alpha_i - \sum_j \beta_j - c \gamma_k) = 0$$

Ce qui donne en désignant par L_i , C_j , T_k et G le total des observations respectivement dans le $i^{\text{ème}}$ niveau "ligne", $j^{\text{ème}}$ niveau "colonne", $k^{\text{ème}}$ niveau traitement et le total général, et par m , a_i , b_j , t_k les estimateurs de μ , α_i ; β_j , et γ_k respectivement, et compte tenu des conditions (2.3.2) :

$$G - c^2 m = 0$$

$$L_i - c m - c a_i = 0$$

$$C_j - c m - c b_j = 0$$

$$T_k - c m - c t_k = 0$$

Soit :

$$m = \frac{1}{c^2} G = y_{..}$$

$$a_i = \frac{1}{c} L_i - m = y_{i.} - y_{..}$$

(2.3.3)

$$b_j = \frac{1}{c} C_j - m = y_{.j} - y_{..}$$

$$t_k = \frac{1}{c} T_k - m = y_{(k)} - y_{..}$$

où

$$y_{(k)} = \frac{1}{c} T_k$$

2) Test de signification des effets :

L'équation d'Analyse de la variance généralisée s'écrit ici :

$$\begin{aligned} \mathbf{YY}' &= \mathbf{E} + \sum_{i=1}^c L_i a_i' + \sum_{j=1}^c C_j b_j' + \sum_{k=1}^c T_k t_k' + \mathbf{G} m' \\ &= \mathbf{E} + \mathbf{H}_1 + \mathbf{H}_c + \mathbf{H}_t + \mathbf{G} m' \end{aligned}$$

avec

$$\mathbf{E} = \sum_{ij} (y_{ij} - y_{i.} - y_{.j} - y_{(k)} + 2 y_{..}) (\leftarrow)'$$

est la matrice erreur autour du modèle (2.3.1) qui suit une distribution de Wishart $W(n, \Sigma)$ avec $n = (c - 1)(c - 2)$ supposé $\geq p$.

$$\mathbf{H}_1 = c \sum_{i=1}^c (y_{i.} - y_{..}) (y_{i.} - y_{..})'$$

est la matrice de l'hypothèse de non signification de l'effet "ligne". formulée par $\alpha_i = 0 \forall_i$, et qui suit si cette hypothèse est vraie, une distribution de Wishart généralisée avec $s = (c - 1)$ d.d.l.

$$H_c = c \sum_{j=1}^c (y_{.j} - y_{..}) (y_{.j} - y_{..})'$$

est la matrice de l'hypothèse de non signification de l'effet colonne, formulée par $\beta_j = 0 \forall_j$ et qui suit si cette hypothèse est vraie une distribution de Wishart généralisée avec $s = (c - 1)$ d.d.l.

$$H_t = c \sum_{k=1}^c (y_{(k)} - y_{..}) (y_{(k)} - y_{..})'$$

est la matrice de l'hypothèse de non signification de l'effet traitement, formulée par $\gamma_k = 0 \forall_k$, et qui suit si cette hypothèse est vraie, une distribution de Wishart généralisée avec $s = (c - 1)$ d.d.l.

On peut donc tester chacune de ces hypothèses en appliquant un des critères C_1 , C_2 ou C_3 du § (1.2.7).

REFERENCES

- [1] ANDERSON T.W – Introduction to multivariate statistical analysis J. Wiley 1958.
- [2] COX D.R. – Planning of experiments J. Wiley 1960.
- [3] DASGUPTA S. ANDERSON T.W. et MUDHOLKAR G.S. – Monotonicity of the power function of some tests of multivariate linear hypothesis Ann. Math. Stat. (35) 1964 pp 200 – 5.
- [4] DUGUE D. – Traité de statistique théorique et appliquée Masson 1959.
- [5] HAKIM M. – Analyse multivariate des plans d'expérience avec données manquantes. Thèse de 3^{ème} cycle, Fac. Sciences. Paris 1965.
- [6] HECK D.L. – Charts of some upper percentage points of the distribution of the largest characteristic roots. Ann. Math. Stat (31) 1960 pp 625 – 42.
- [7] HOTELLING H. – A generalised T – test and measure of multivariate, dispersion. Proc. Secd. Berkley Symp. Prob. Stat. pp 23 – 41.
- [8] KEMPTHORNE O. – The Design and Analysis of Experiments Wiley 1952
- [9] LAWLEY D.N. – A generalisation of Fisher's z – test Biometrika (31) 1940
- [10] PILLAI K.C.S. – On the distribution of the largest characteristic root of a matrix in multivariable analysis. Biometrika (52) 1965.

- [11] POSTEN H.O., BARGMANN R.E. – Power of the likelihood-ratio test of the general linear hypothesis in multivariate analysis *Biometrika* (51) 1964.
- [12] ROY S.N. – Some aspects of multivariate analysis Wiley 1957.
- [13] SMITH H. HUGHES J.B. and GNANADESIKAW R. – Multivariate analysis of variance MANOVA *Biometrics* 1962.
- [14] SRIVASTAVA J.N. – On the monotonicity property of the three main tests for multivariate analysis. *J. Roy. Stat Soci. Serie B* (26) 1964.
- [15] ULMO J. – I – Regression linéaire multiple 1959 – 1960 .
II – Eléments de la théorie des plans d'expérience 1960 – 1961.
Cours de 3^{ème} cycle Publ. de l'I.S.U.P.
- [16] WILKS – Certain généralisations in the analysis of variance *Biometrika* (24) 1932.