ANALYSE DE VARIANCE D'UN DISPOSITIF EN BLOCS NON ORTHOGONAL

PAR

Richard TOMASSONE

Station de Biométrie - C.N.R.F.

Pour appliquer de façon simple les méthodes statistiques aux résultats fournis par un plan d'expérience il faut que le dispositif soit orthogonal : il est alors possible de séparer les effets de chacun des facteurs, puis de tester leur effet séparément. Par exemple dans un dispositif en blocs complets il faut que chaque traitement figure une fois, ou un nombre égal de fois dans chaque bloc. Mais les conditions matérielles du déroulement de l'expérience sont telles qu'il peut exister des trous dans les tableaux de données : dans une expérience de comparaison de provenances une mortalité plus forte pour un groupe de provenances conduit à l'interprétation d'un dispositif non orthogonal. L'expérience ne peut alors plus s'interpréter avec les moyens classiques. Il faut faire appel à une méthode plus puissante : ce sont les résultats applicables à un dispositif en blocs que nous allons donner ; nous les appliquerons ensuite à l'interprétation d'une expérience citée dans la revue forestière.

Principes de la méthode

Les résultats de toute expérience à deux facteurs contrôlés, et un dispositif en blocs appartient à cette catégorie, peuvent se représenter sous forme d'un tableau à deux entrées :

TABLEAU 1.

Blocs / Traitements	1 . . . . . j . . . . . . b
1	⋮
.	
.	
.	
i	.... $\boxed{x_{ij}}$ ....
.	
.	
.	
t	⋮

Présentation schématique d'une analyse en blocs.

Généralement on repère une mesure par deux indices:

— l'indice i permet de repérer le numéro du traitement et peut varier entre 1 et t,

— l'indice j permet de repérer le numéro du bloc et peut varier entre 1 et b.

Le nombre noté symboliquement x_{ij} représente le résultat de la mesure du i^e traitement dans le j^e bloc.

Au sein d'une parcelle unitaire (1) on peut effectuer zéro, une ou plusieurs mesures, nous notons ce nombre n_{ij} et nous appelons le tableau formé par l'ensemble de ces nombres matrice d'incidence (2). L'interprétation est simple si tous ces nombres sont égaux (3), un peu plus délicate dans le cas où ils sont proportionnels (4), encore davantage s'ils sont différents et si aucun n'est nul (5). Enfin l'interprétation peut être matériellement impossible si les nombres b et t sont grands et les quantités n_{ij} absolument quelconques (6).

Nous n'étudions ici que le cas où il y a:

— soit une mesure par parcelle unitaire alors $n_{ij} = 1$

— soit aucune mesure par parcelle unitaire alors $n_{ij} = 0$

La méthode que nous allons exposer est connue depuis quelques années (7), mais on ne peut l'utiliser de façon routinière qu'à l'aide d'une calculatrice électronique à cause de la complexité *matérielle* des calculs. L'idée de départ est simple: pour définir l'effet des traitements on suppose que la différence entre deux traitements quelconques est indépendante du bloc; on démontre que cette hypothèse est équivalente à celle d'un effet additif: le résultat de la parcelle unitaire (i, j) est la somme d'une quantité imputable au traitement i, soit t_i, et d'une autre quantité due au bloc j, soit b_j. La méthode étudiée et mise

(1) Le nom de parcelle peut prêter à confusion, surtout dans le domaine forestier où il a un autre sens. Ici, nous appelons *parcelle unitaire*, la surface au sein de laquelle les conditions sont identiques: les variations observées ne sont dues qu'au « hasard ».

(2) Le calcul matriciel est un outil indispensable dans l'étude des plans d'expérience: la concision qu'il permet de donner aux formules est un aspect important de l'intérêt qu'il présente; en outre toutes les calculatrices modernes ont une bibliothèque de programmes de calcul matriciel. Pour une première étude cf. VIGNAL J. Calcul matriciel.

(3) Cf. VESSEREAU (A.). — Méthodes statistiques en biologie et en agronomie, p. 207-216 et 241-255.

(4) Cf. KEMPTHORNE (O.). — The design and analysis of experiments, p. 87-88.

(5) Cf. KENDALL (M.-G.). — The advanced Theory of Statistics, p. 220-224, tome II.

DUGUE (D.). — Traité de Statistique théorique et appliquée, p. 223-225.

(6) Cf. FEDERER (W.-T.). — Experimental design, p. 124-125.

(7) Cf. TOCHER (K.-D.). — The design and analysis of block experiments. Dans le cas où il y a une donnée manquante les livres classiques donnent l'estimation de cette donnée, cf. VESSEREAU, op. cit., p. 239-240. On peut aussi grâce à l'artifice de la covariance estimer ces données, cf. TOMASSONE: L'analyse de la covariance: théorie et application dans l'expérimentation forestière. La méthode exposée ici est plus générale.

au point par TOCHER revient à chercher l'estimation des paramètres $\hat{t}_{i.}$ ou ce qui est équivalent l'estimation des t composantes d'un vecteur **t**. Ensuite pour pouvoir comparer deux traitements quelconques on calcule la matrice des variances et des covariances de ces paramètres : Ω ; on démontre que :

$$\text{var}\ (\hat{t}_1 - \hat{t}_2) = (\omega_{11} - 2\omega_{12} + \omega_{22})\ S_E^2$$

où S_E^2 est la variance résiduelle. Evidemment les éléments de la matrice Ω sont les nombres ω_{ij}. L'avantage essentiel de cette méthode est qu'elle permet de comparer exactement les moyennes de deux traitements quelconques. La suite des calculs est indiquée au tableau 4.

Exemple

Nous utilisons les données du tableau 4 de l'article précédent de M. LACAZE J. Dans ce tableau t = 17, b = 4. Pour obtenir la matrice d'incidence N il suffit de remplacer la date d'installation par 1 et de remplir le reste du tableau de zéros. Tous les calculs ont été faits avec les données de l'article ci-dessus : pour montrer la suite des calculs nous allons extraire la comparaison des provenances SAINT-NECTAIRE (i = 7) et DALECARLIE (i = 14) :

TABLEAU 2.

Blocs / Traitements	1	2	3	4	Total T_i	Moyenne non corrigée
⋮ ⋮						
SAINT NECTAIRE i = 7	6,65	8,53	7,86		23,04	7,67
⋮ ⋮						
DALECARLIE i = 14		3,45			3,45	3,45
⋮ ⋮						
Total B_j	88,12	91,72	34,69	19,72	234,25	
Moyenne bloc	6,78	6,55	6,92	3,92		

Extrait des mesures de l'expérience de comparaison de provenances de pin sylvestre (*pinus Sylvestris* L.) (D. Amélioration 64.22).

L'analyse de variance a permis de calculer une estimation de la variance résiduelle $S_E^2 = 1{,}595$, avec 17 degrés de liberté, et on a conclu à un effet significatif des traitements au seuil 1 %. La matrice des variances et des covariances des paramètres (1), dont nous avons extrait les lignes et les

(1) Cette matrice est symétrique, on peut donc n'inscrire qu'un côté de la diagonale.

colonnes 7 et 14, est la suivante :

TABLEAU 3.

$i \rightarrow$ $\downarrow$	.....	7	......	14	......
:					
7		+ 0,38022	.....	0,03393	
:		:		:	
14				1,06910	
:		:			

Extrait de la matrice des variances et des covariances
(D. Amélioration 64-22).

Les valeurs ajustées des moyennes des provenances 7 et 14 sont :

SAINT-NECTAIRE : $\hat{t}_7 = 7{,}40$

DALECARLIE : $\hat{t}_{14} = 3{,}32$

$$\hat{t}_7 - \hat{t}_{14} = 4{,}08$$

Et la variance de la différence vaut :

$$\text{Var}\,(\hat{t}_7 - \hat{t}_{14}) = 1{,}595\,(0{,}38022 - 2 \times 0{,}03393 + 1{,}06910) = 2{,}203$$

donc :

$$\frac{\hat{t}_7 - \hat{t}_{14}}{\sqrt{\text{var}\,(t_7 - t_{14})}} = \frac{4{,}08}{1{,}48} = 2{,}76$$

ce qui est significatif au seuil de probabilité 5 % (pour 17 degrés de liberté).

Conclusion

La méthode que nous venons d'exposer a l'immense avantage de permettre la comparaison de chaque couple de moyenne. Elle est en outre générale à tous les dispositifs en blocs : blocs complets, incomplets équilibrés ou non. Une fois le programme de la calculatrice écrit on peut donc étudier théoriquement n'importe quel dispositif en blocs (1). A l'heure actuelle ce résultat est peut être trop général : il est inutile d'inverser une matrice d'ordre t si l'analyse peut se faire plus simplement (2). Mais, pour l'avenir, lorsque les calculatrices de la « troisième génération » entreront en service il serait peut-être bon d'avoir un modèle d'interprétation unique aussi bien pour les plans simples que pour les plans complexes (3).

(1) Pour les dispositifs en blocs incomplets, on n'étudie pas de cette façon l'analyse avec information entre blocs dans laquelle on suppose que les paramètres des blocs sont des variables aléatoires normales.

(2) La seule inversion sur la machine électronique de la Faculté de Nancy a demandé vingt-cinq minutes.

(3) Une calculatrice dernièrement mise en service peut inverser en moins d'une minute un système linéaire de cinquante équations à cinquante inconnues.

Lecture des données

↓

Calcul de la matrice des variances et des covariances des paramètres traitement

↓

Formation de la matrice d'incidence $\mathbf{N}$ $(t \times b)$

↓

Somme en ligne: vecteur $\mathbf{N}_i$ et matrice $\mathbf{N}_i^d$ $(t \times t)$

↓

Somme en colonne: vecteur $\mathbf{N}_j$ et matrice $\mathbf{N}_j^{-d}$ $(b \times b)$

↓

$\mathbf{W} = \mathbf{N}.\mathbf{N}_j^{-d}$ $(t \times b)$

↓

$$\Omega^{-1} = \mathbf{N}_i^d - \mathbf{WN'} + \frac{1}{n} \mathbf{N}_i.\mathbf{N}_i'$$

↓

Calculs préparatoires à l'analyse de variance

↓

Formation de la matrice des mesures $\mathbf{X}$ $(t \times b)$

↓

Somme des Carrés Σx^2

↓

Somme en ligne: vecteur traitement $\mathbf{T}$ $(t \times 1)$

↓

Somme en colonne: vecteur bloc $\mathbf{B}$ $(b \times 1)$

↓

Total général G, moyenne $\bar{\bar{x}}$ et terme de centrage G^2/n

↓

Analyse de variance

↓

$\mathbf{Q} = \mathbf{T\text{-}WB}$ $(t \times 1)$

↓

Résolution du système linéaire $\Omega^{-1}\ \mathbf{t}_o = \mathbf{Q}$

↓

Tableau d'analyse de variance

Source de variation	Somme des carrés	d. l.	Carré moyen	Test F
Blocs	$\mathbf{B'N}_j^{-d}\ \mathbf{B}\text{-}G^2/n$	$b - 1$		
Traitement	$\mathbf{Q'}\ \Omega\ \mathbf{Q}$	$t - 1$	S_T^2	S_T^2/S_E^2
Erreur (par différence)	(diff.)	(diff.)	S_E^2	
Total	$\Sigma x^2 - G^2/n$	$n - 1$	/	/

↓

Comparaison des moyennes ajustées (si test F significatif)

↓

Moyennes ajustées $\mathbf{t} = \mathbf{t}_o + \bar{x}\ \mathbf{1}_t$

↓

Variance de la différence de deux moyennes ajustées
$v_{ij} = (\omega_{ii} - 2\ \omega_{ij} + \omega_{jj})\ S_E^2$

↓

Test t de STUDENT

Les caractères notés en gras indiquent qu'on utilise un vecteur ou une matrice. Le ' correspond à une transposition.
Les chiffres indiqués entre parenthèses sont les dimensions de la matrice ou du vecteur.
$\mathbf{1}_t$ est un vecteur dont les t composantes sont égales à 1.
n est le nombre total de mesures.

Tableau 4: Suite des calculs pour l'analyse d'un dispositif en blocs non orthogonal.

Les résultats que nous avons énoncés permettent de donner la meilleure interprétation possible dans une situation désespérée: sans ce remède rien ne peut être déduit des mesures effectuées. Il ne faut tout de même pas se faire trop d'illusions sur la validité de l'interprétation: la meilleure interprétation statistique ne permettra jamais de retrouver toute l'information perdue par l'absence de certaines mesures.

BIBLIOGRAPHIE

DUGUE (D.). — Traité de Statistique Théorique et Appliquée, Masson, Paris, 1958.

FEDERER (W.-T.). — Experimental design: Theory and application, MacMillan, New York, 1963.

KEMPTHORNE (O.). — The design and analysis of experiments, John Wiley and Sons, New York, (3e Ed.), 1962.

KENDALL (M.-G.). — The advanced Theory of statistics, Londres, Griffin, (3e Ed.), 1951.

LACAZE (J.). — Sur la comparaison de diverses provenances de pin sylvestre (*pinus sylvestris* L.) représentées dans les arboretums forestiers, R.F.F., Août-Sept. 1964.

TOCHER (K.-D.). — The design and analysis of block experiments, Journ. Roy. Stat. Soc. B., XIV, p. 45-100, 1962.

TOMASSONE (R.). — L'analyse de covariance: Théorie et application dans l'expérimentation forestière. Annales des Sciences Forestières — T. XXI, Fasc. 2, 1964 (à paraître).

VESSEREAU (A.). — Méthodes statistiques en Biologie et en Agronomie, Baillère et Fils, Paris, 2e Ed., 1960.

VIGNAL (J.). — Calcul matriciel, Vuibert, Paris, 1962.