

STATISTIQUE ET ANALYSE DES DONNÉES

J. R. MATHIEU

Un exemple de comparaison des puissances dans le modèle linéaire

Statistique et analyse des données, tome 2, n° 1 (1977), p. 23-36

http://www.numdam.org/item?id=SAD_1977__2_1_23_0

© Association pour la statistique et ses utilisations, 1977, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » 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/>

UN EXEMPLE DE COMPARAISON DES PUISSANCES
DANS LE MODELE LINEAIRE

par

J.R. MATHIEU*

* : Laboratoire de statistique
118, route de Narbonne - 31077 TOULOUSE

I - INTRODUCTION :

Il se présente, dans les applications, des plans à un ou plusieurs facteurs croisés dans lesquels certains facteurs sont définis par un paramètre quantitatif (dose, pourcentage, ...) si bien que les différents niveaux de chacun de ces facteurs sont définis par différentes valeurs d'un paramètre réel.

Dans une telle situation expérimentale, lorsque l'on ne sait rien de la façon dont les facteurs "influencent" les observations, on exploite les données numériques à l'aide du modèle d'analyse de la variance.

Supposons toutefois que l'on possède l'information supplémentaire que, tel facteur parmi ceux qui sont définis par un paramètre quantitatif, a, pour chacun de ses niveaux expérimentés, un effet différentiel s'exprimant comme une fonction polynomiale (de degré connu et de coefficients inconnus) de la valeur prise au niveau considéré par ce paramètre quantitatif, qui joue ainsi tout à fait le rôle de variable explicative ; on est alors amené à affiner le modèle initialement choisi, à savoir celui d'analyse de la variance sur plans croisés, en y intégrant cette information supplémentaire, le modèle ainsi obtenu étant vis-à-vis de l'effet principal de certains facteurs, un modèle de régression. On se pose alors le problème de savoir quel est le gain de puissance ainsi obtenu.

On considère donc pour analyser un jeu de données deux modèles : le premier qui est celui d'un plan d'analyse de la variance à facteurs croisés et qui ignore toute information supplémentaire sur l'effet des facteurs, le second qui est plus fin (c'est-à-dire à espace paramétrique plus restreint) parce qu'il intègre une information supplémentaire précisant la "forme" des effets différentiels de certains facteurs. Dans chacun de ces deux modèles on peut éprouver une hypothèse linéaire ayant la même signification concrète, par exemple que tel facteur a un effet nul ; cette hypo-

thèse se formule différemment, suivant celui des deux modèles que l'on considère mais elle est, dans les deux cas, passible du test de Fisher que l'on prend chaque fois au niveau α .

Supposons que l'hypothèse est fautive : une même alternative simple donnée s'exprime dans chacun des deux modèles par des valeurs spécifiant les paramètres inconnus de ce modèle : on peut alors calculer en ce "point" la puissance de chacun des deux tests de Fisher et déterminer le rapport des puissances : ce rapport représente, en ce point considéré de l'alternative, le gain de puissance obtenu en utilisant le second modèle de préférence au premier, c'est-à-dire le gain de puissance apporté par l'information supplémentaire qui a permis, partant d'un modèle "usuel" d'analyse de la variance de l'"affiner" jusqu'à un modèle plus spécifique.

On s'intéresse donc au rapport des puissances lorsque le paramètre inconnu décrit l'hypothèse alternative : ce problème a déjà été évoqué dans de nombreux travaux y compris à propos du modèle linéaire comme par exemple dans [1].

II - NOTATIONS :

On considère un modèle linéaire dans lequel le paramètre appartient à R^q : l'observation x écrite sous forme de matrice colonne est donc considérée comme une réalisation d'une v.a. X que l'on suppose

$$N_n[A\theta, \sigma^2 I_n]$$

où A est une matrice connue $n \times q$, de rang q ($q < n$)

I_n désigne la matrice identité d'ordre n .

θ et σ^2 sont inconnus : $\theta \in R^q$, $\sigma^2 \in R^{**}$

On note V le sous-espace de R^n , image de R^q par l'application linéaire h de matrice A .

Soit l'hypothèse :

$\theta \in \theta_1$ où θ_1 est un sous-espace de R^q de rang $q - r_1$

W_1 désigne l'image de θ_1 par h

R^n et R^q sont munis de produit scalaire :

$$x, y \in R^n \quad \langle x, y \rangle = {}^t x y = \sum_{i=1}^n x_i y_i$$

$$(1) \quad \theta, \phi \in R^q, \quad \langle \theta, \phi \rangle = \langle h(\theta), h(\phi) \rangle = {}^t \theta {}^t A A \phi$$

h est alors une isométrie de R^q sur V .

La statistique du test de Fisher de θ_1 contre $R^q - \theta_1$ est :

$$T_1(x) = (n-q) \frac{\|x_{W_1^\perp}\|^2}{\|x - x_V\|^2}$$

où x_V et $x_{W_1^\perp}$ désignent les projections de x respectivement sur V et sur le sous-espace inclus dans V et orthogonal à W_1 .

Le test rejette θ_1 lorsque

$$T_1(x) \geq r_1 f_{1,\alpha}$$

On notera $F(n_1, n_2, \gamma, \cdot)$ la fonction de répartition de la loi de Fisher à n_1 et n_2 d.d.l. et de paramètre de décentrage γ :

$$F(r_1, n-q, 0, f_{1,\alpha}) = 1 - \alpha.$$

Le test ci-dessus a pour puissance :

$$p(\theta, \sigma^2) = 1 - F(r_1, n-q, \lambda_1, f_{1, \alpha})$$

$$\text{avec } \lambda_1 = \lambda_1(\theta, \sigma^2) = \frac{\|\theta_{\theta_1^\perp}\|^2}{2\sigma^2}$$

où $\theta_{\theta_1^\perp}$ est la projection de θ sur l'orthogonal de θ_1 .

III- RESTRICTION DE L'ESPACE PARAMETRIQUE :

Supposons maintenant qu'une information supplémentaire permette de considérer que l'espace paramétrique se restreint de $R^q \times R^{**}$ à $\theta_2 \times R^{**}$ où θ_2 est un sous-espace de R^q de rang $q - r_2$: on obtient donc un nouveau modèle linéaire dans lequel X est considéré comme $N_n[A\theta, \sigma^2 \cdot I_n]$

$$\text{avec } \theta \in \theta_2, \sigma^2 \in R^{**}$$

On pose $\theta = \theta_1 \cap \theta_2$ et on note $q - r_2 - r$ le rang de θ ($r > 0$).

Dans ce modèle, on éprouve l'hypothèse θ contre l'hypothèse $\theta_2 - \theta$ avec le test de Fisher : en notant W et W_2 les images respectives de θ et θ_2 par h , le test de niveau α de θ contre $\theta_2 - \theta$ rejette θ lorsque

$$T(x) \geq rf_\alpha$$

$$\text{avec } T(x) = [n - (q - r_2)] \frac{\|x_{W^\perp}\|^2}{\|x_{W_2}\|^2}$$

$$\text{et } F(r, n - (q - r_2), 0, f_\alpha) = \alpha.$$

où W^\perp désigne le sous-espace inclus dans W_2 et orthogonal à W .

La puissance du test de Fisher est alors :

$$p(\theta, \sigma^2) = 1 - F[r, n - (q - r_2), \lambda, f_\alpha]$$

$$\text{avec } \lambda = \lambda(\theta, \sigma^2) = \frac{\|\theta_{\theta^\perp}\|^2}{2\sigma^2}$$

où θ^\perp désigne le sous-espace inclus dans θ_2 et orthogonal

à θ .

Le problème envisagé dans l'introduction est donc celui de la comparaison de $p(\theta, \sigma^2)$ et $p_1(\theta, \sigma^2)$ lorsque (θ, σ^2) parcourt $\theta_2 \times \mathbb{R}^{**}$:

$$(2) \quad \frac{p(\theta, \sigma^2)}{p_1(\theta, \sigma^2)} = \frac{1-F[r, n-(q-r_2), \lambda, f_\alpha]}{1-F[r_1, n-q, \lambda_1, \lambda_1, \alpha]}$$

Définition : Si θ_1 et θ_2 sont tels que les sous-espaces inclus dans $\theta_1 + \theta_2$ et orthogonaux respectivement à θ_1 et θ_2 sont orthogonaux entre eux, on dira que θ_1 et θ_2 sont faiblement séparables.

Cette définition et la propriété qui suit ont déjà été énoncées dans [2].

Propriété : θ_1 et θ_2 sont faiblement séparables si et seulement si, pour tout θ appartenant à θ_2 ,

$$\theta_{\theta^\perp} = \theta_{\theta_1^\perp}$$

En effet, notons θ' le sous-espace inclus dans θ_1 et orthogonal à θ , on a :

$$\theta_1 + \theta_2 = \theta \oplus \theta^\perp \oplus \theta'$$

Pour tout θ on a

$$\theta = \theta_{\theta_1} + \theta_{\theta_1^\perp} = \theta_\theta + \theta_{\theta'} + \theta_{\theta_1^\perp}$$

De plus pour tout θ appartenant à θ_2 , on a :

$$\theta = \theta_\theta + \theta_{\theta^\perp}$$

Donc pour tout θ appartenant à θ_2

$$\theta_{\theta^\perp} = \theta_{\theta'} + \theta_{\theta_1^\perp}$$

Par suite $\forall \theta \in \theta_2, \theta_{\theta_1}^\perp = \theta_{\theta_1}^\perp \iff \forall \theta \in \theta_2, \theta_{\theta_1} = 0 \iff$

$$\theta_2 \perp \theta' \iff \theta_1^\perp \perp \theta'$$

Donc, à condition que θ_1 et θ_2 soient faiblement séparables et seulement à cette condition, $\lambda(\theta, \sigma^2)$ et $\lambda_1(\theta, \sigma^2)$ sont égaux sur $\theta_2 \times R^{**}$ et par suite pour étudier le rapport des puissances, lorsque l'alternative parcourt $\theta_2 \times R^{**}$, λ_1 et λ sont égaux.

Un cas particulier où θ_1 et θ_2 sont faiblement séparables est celui où $\theta_1 \subset \theta_2$; en effet

$$\theta = \theta_1 \text{ et } \theta' = \{0\} .$$

Dans cette situation l'espace paramétrique subit une restriction qui n'affecte pas l'hypothèse nulle θ_1 mais seulement l'alternative qui se réduit de

$$(R^k - \theta_1) \times R^{**} \quad \text{à} \quad (\theta_2 - \theta_1) \times R^{**}$$

IV - EXEMPLES D'APPLICATIONS :

Applications aux plans croisés :

Dans ce qui suit les notations qui ne sont pas définies explicitement sont celles utilisées dans [3].

Soit un plan à k facteurs croisés :

$K = \{1, 2, \dots, k\}$ est l'ensemble des facteurs

$E = \prod_{i=1}^k \{1, 2, \dots, v_i\}$ est l'ensemble des cellules

$$\text{Card}(E) = v = \prod_{i=1}^k v_i$$

$m(e)$ est l'effet dans la cellule e .

Le paramètre inconnu m appartient donc à R^v .

Le produit scalaire défini par (1) a pour expression :

$$(3) \quad \langle m, m' \rangle = \sum_{e \in E} n(e) m(e) m'(e)$$

où $n(e)$ est le nombre d'observations de la cellule e .

Supposons le plan équilibré : $\forall e \in E, n(e) = n_0$ ($n_0 > 1$)

$$\text{Alors } \langle m, m' \rangle = n_0 \sum_{e \in E} m(e) m'(e)$$

Au sens du produit scalaire (3), les sous-espaces $\Omega_I, I \subset K$, sont donc orthogonaux deux à deux et m_I est donc la projection orthogonale de m sur Ω_I

\mathfrak{F} étant une famille donnée de parties de K ,

$$\Theta_1 = \{m / \forall I \in \mathfrak{F}, m_I = 0\}$$

L'appartenance de m à Θ_1 , signifie donc la nullité des effets différentiels ou d'interactions de certains facteurs.

J étant une partie donnée de K , différente de la partie vide soit Θ_2 l'hypothèse définie par :

$m \in \Theta_2 \iff \forall j \in J$ il existe une application connue b_j de E_j dans R et un réel ϕ_j tel que

$$(4) \quad m_{\{j\}}(e) = \phi_j (b_j(e_j) - \bar{b}_j) \quad e_j = 1, 2, \dots, v_j$$

$$\text{avec } \bar{b}_j = \sum_{e_j=1}^{v_j} \frac{b_j(e_j)}{v_j}$$

Dire que m appartient à Θ_2 , c'est donc affirmer que pour chaque facteur j de J , il existe un "codage" connu des niveaux de ce facteur, défini par la fonction b_j ; tel que les effets différentiels des niveaux de ce facteur soient donnés par (4).

Montrons que Θ_1 et Θ_2 sont faiblement séparables :

On notera \bar{J} la famille des parties de K constituée des ensembles $\{i\}, i \in J$.

$$m \in \Theta \iff m(e) = \sum_{\{i\} \in \bar{J} - \mathfrak{F}} \phi_i (b_i(e_i) - \bar{b}_i) + \sum_{I \in \mathfrak{F} \cup \bar{J}} m_I(e).$$

$$m \in \Theta^\perp \iff m(e) = \sum_{\{i\} \in \bar{J} \cap \mathfrak{F}} \phi_i (b_i(e_i) - b_i) + \sum_{I \in \mathfrak{F} - \bar{J}} m_I(e)$$

donc $\theta^1 \perp \theta_1 \implies \theta^1 \perp \theta'$

Par suite, le rapport des puissances est déterminé par l'expression (2) avec $\lambda = \lambda_1$ et pour :

$$n = n_0 v$$

$$q = v$$

$$r_1 = \sum_{I \in \mathcal{F}} \dim(\Omega_I)$$

$$r_2 = \sum_{j \in J} (v_j - 2)$$

$$r = \text{Card}(\bar{J} \cap \mathcal{F}) + \sum_{I \in \mathcal{F} - \bar{J}} \dim(\Omega_I)$$

Dans le cas particulier du plan à 1 facteur, θ_1 et θ_2 sont faiblement séparables quelles que soient les valeurs de $n(e)$:

en effet, la décomposition

$$m = m_\phi + m_{\{1\}}$$

où $m_{\{1\}}(e)$ est l'effet différentiel du niveau e

est, quelles que soient les quantités $n(e)$, orthogonale pour le produit scalaire défini par (3).

Applications aux carrés eulériens :

Soit E^* un carré eulérien d'ordre p et de degré k :

$$m(e) = m_\phi + \sum_{i=1}^k m_{\{i\}}(e_i)$$

Le paramètre inconnu m appartient à R^q avec $q = 1+k(p-1)$

Le produit scalaire défini par (1) a pour expression :

$$\langle m, m' \rangle = \sum_{e \in E^*} m(e) m'(e).$$

Par suite, au sens de ce produit scalaire, m_ϕ , $m_{\{i\}}$, $m_{\{i'\}}$, $i \neq i'$, sont orthogonaux deux à deux.

I étant une partie de K , différente de la partie vide soit

$$\theta_1 = \{m / \forall i \in I, m_{\{i\}} = 0\}$$

J étant une partie de K et les applications b_i , $i \in J$, étant données, on pose :

$$\Theta_2 = \{m/V_j \in I, m_{\{j\}}(e) = \phi_j(b_j(e_j) - \bar{b}_j)\}$$

On vérifie de la même façon que précédemment que Θ_1 et Θ_2 sont faiblement séparables : le rapport des puissances est donc de la forme (2) :

$$\begin{aligned} \lambda &= \lambda_1 \\ n - q &= (p-1)(p+1-k) \\ r_1 &= (p-1) \text{Card}(I) \\ r_2 &= (p-2) \text{Card}(J) \\ r &= \text{Card}(I \cap J) + (p-1) \text{Card}(I-J) \end{aligned}$$

Bibliographie :

- [1] D.J. BARTHOLOMEW : "Ordered tests in analysis of variance"
Biometrika (1961) , 48, 3 and 4, pp. 325-332.
- [2] J.R. MATHIEU : "Sur certaines propriétés des hypothèses séparables"
Publications du Laboratoire de Statistique :
n° 02-76 Toulouse (1976)
- [3] J.R. BARRA : "Notions fondamentales de statistique mathématique"
Dunod (1971)

APPLICATION NUMERIQUE

A titre d'illustration numérique du gain de puissance par restriction de l'espace paramétrique, on a considéré un plan croisé à 2 facteurs ayant respectivement v_1 et v_2 niveaux et présentant deux observations par cellule :

On peut donc appliquer les résultats du paragraphe IV : on a envisagé successivement chacun des deux cas suivants :

$$\mathcal{F} = \{\{1\}\}$$

$$\mathcal{F} = \{\{2\}\}$$

\mathcal{F} étant fixé, l'hypothèse $\theta_1 \times \mathbb{R}^{*+}$ s'en déduit : lorsque $\mathcal{F} = \{\{1\}\}$ (respectivement $\{\{2\}\}$), l'hypothèse exprime que le facteur 1 (respectivement 2) est sans effet.

De plus on envisage les cas suivants :

$$J = \{1\}$$

$$J = \{2\}$$

$$J = \{1,2\}$$

Lorsque $J = \{1\}$ (respectivement $\{2\}$), le facteur 1 (respectivement le facteur 2) est tel que $m_{\{1\}}$ (respectivement $m_{\{2\}}$) satisfait l'expression (4) où le paramètre ϕ_1 (respectivement ϕ_2) est un paramètre réel inconnu ; lorsque $J = \{1,2\}$, chacun des deux facteurs satisfait cette condition. La donnée de J détermine θ_2

Pour chacune des valeurs données à v_1 et v_2 et pour chacune des trois valeurs possibles de J on a effectué, pour quelques valeurs du paramètre λ de décentrage, le rapport des puissances du test de $(\theta_2 \cap \theta_1) \times \mathbb{R}^{*+}$ contre $(\theta_2 - \theta_1 \cap \theta_2) \times \mathbb{R}^{*+}$ et du test de $\theta_1 \times \mathbb{R}^{*+}$ contre $(\mathbb{R}^v - \theta_1) \times \mathbb{R}^{*+}$, ces deux tests étant de niveau $\alpha = 0,05$.

Lorsque $v_2 = 2$, l'application $m_{\{2\}}$ est une application de $\{1,2\}$ dans \mathbb{R} et satisfait donc toujours la condition (4) : par suite lorsque $J = \{2\}$, θ_2 coïncide avec \mathbb{R}^v et le rapport des puissances vaut constamment 1 ; de plus

θ_2 est identique pour $J = \{1\}$ et $J = \{1,2\}$; c'est pourquoi lorsque $v_2=2$, seul est envisagé le cas $J = \{1\}$.

De plus lorsque $v_1 = v_2$, les cas de figure où $\mathfrak{F} = \{\{2\}\}$ se ramènent tous à l'un des cas de figure où $\mathfrak{F} = \{\{1\}\}$

	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	10
3		{1}	1,300	1,352	1,429	1,476	1,502	1,517	1,527	1,528	1,527	1,521	1,499	1,469	1,402	1,335	1,113	
3		{2}	1,025	1,029	1,034	1,037	1,040	1,041	1,042	1,043	1,043	1,043	1,043	1,043	1,041			
3		{1}	1,282	1,326	1,387	1,421	1,438	1,442	1,444	1,439	1,432	1,422	1,392	1,358	1,291	1,231	1,057	
3		{2}	1,013	1,016	1,020	1,023	1,025	1,026	1,028	1,029	1,029	1,030	1,030	1,030	1,028			
3		{1,2}	1,296	1,344	1,407	1,441	1,459	1,466	1,466	1,462	1,454	1,444	1,413	1,377	1,306	1,242	1,059	
4		{1}	1,466	1,559	1,691	1,776	1,829	1,859	1,875	1,879	1,875	1,865	1,823	1,767	1,644	1,526	1,162	
4		{2}	1,027	1,031	1,037	1,040	1,042	1,043	1,044	1,044	1,044	1,044	1,043	1,040				
4		{1}	1,442	1,523	1,633	1,697	1,732	1,748	1,751	1,746	1,734	1,717	1,665	1,606	1,486	1,381	1,092	
4		{2}	1,008	1,010	1,013	1,015	1,017	1,018	1,019	1,020	1,021	1,021	1,022	1,022	1,021			
4		{1}	1,449	1,532	1,644	1,709	1,745	1,762	1,765	1,760	1,748	1,731	1,678	1,617	1,495	1,387	1,093	
4		{2}	1,015	1,018	1,023	1,026	1,028	1,029	1,031	1,031	1,032	1,032	1,032	1,032				
4		{2}	1,272	1,314	1,366	1,392	1,403	1,405	1,402	1,395	1,385	1,374	1,342	1,307	1,242	1,186	1,038	
4		{1,2}	1,287	1,331	1,386	1,414	1,427	1,430	1,427	1,419	1,409	1,397	1,363	1,327	1,257	1,197	1,040	
4		{1}	1,428	1,503	1,601	1,654	1,681	1,691	1,689	1,679	1,664	1,645	1,590	1,530	1,415	1,317	1,066	
4		{2}	1,010	1,012	1,015	1,018	1,019	1,021	1,022	1,023	1,023	1,024	1,024	1,024	1,023			
4		{1,2}	1,437	1,514	1,614	1,669	1,697	1,706	1,704	1,694	1,679	1,660	1,603	1,542	1,424	1,323	1,067	
5		{1}	1,571	1,688	1,871	1,993	2,068	2,106	2,138	2,143	2,142	2,128	2,074	1,997	1,829	1,671	1,199	
5		{2}	1,026	1,029	1,034	1,037	1,039	1,040	1,040	1,040	1,040	1,040	1,038	1,035				

5	3	{ {1} }	{1}	1,546	1,650	1,806	1,902	1,956	1,978	1,992	1,986	1,975	1,953	1,886	1,806	1,645	1,503	1,121
5	3	{ {1} }	{2}	1,006	1,007	1,009	1,011	1,012	1,013	1,014	1,015	1,016	1,016	1,017	1,017	1,017	1,017	
5	3	{ {1} }	{1,2}	1,553	1,663	1,817	1,911	1,965	1,993	2,002	1,998	1,984	1,964	1,894	1,814	1,651	1,507	1,122
5	3	{ {2} }	{1}	1,015	1,018	1,022	1,025	1,027	1,028	1,029	1,030	1,030	1,030	1,030	1,029			
5	3	{ {2} }	{2}	1,266	1,305	1,351	1,374	1,382	1,382	1,377	1,369	1,358	1,346	1,313	1,279	1,215	1,162	1,030
5	3	{ {2} }	{1,2}	1,280	1,321	1,371	1,395	1,405	1,405	1,400	1,391	1,380	1,368	1,333	1,296	1,229	1,172	1,031
5	4	{ {1} }	{1}	1,533	1,634	1,773	1,853	1,897	1,916	1,918	1,908	1,890	1,867	1,795	1,715	1,560	1,427	1,091
5	4	{ {1} }	{2}	1,007	1,008	1,011	1,013	1,014	1,016	1,017	1,017	1,018	1,018	1,019	1,019	1,018		
5	4	{ {1} }	{1,2}	1,539	1,642	1,782	1,863	1,908	1,927	1,929	1,919	1,901	1,877	1,804	1,723	1,565	1,431	1,091
5	4	{ {2} }	{1}	1,010	1,012	1,015	1,017	1,019	1,020	1,021	1,022	1,022	1,023	1,023	1,023			
5	4	{ {2} }	{2}	1,419	1,490	1,581	1,629	1,650	1,656	1,651	1,639	1,622	1,602	1,546	1,486	1,375	1,282	1,053
5	4	{ {2} }	{1,2}	1,428	1,501	1,593	1,642	1,665	1,671	1,666	1,654	1,636	1,616	1,558	1,497	1,383	1,287	1,054
5	5	{ {1} }	{1}	1,523	1,620	1,750	1,823	1,860	1,874	1,872	1,860	1,840	1,815	1,741	1,661	1,511	1,384	1,075
5	5	{ {1} }	{2}	1,007	1,008	1,011	1,013	1,014	1,015	1,016	1,017	1,017	1,018	1,018	1,018	1,017		
5	5	{ {1} }	{1,2}	1,528	1,626	1,757	1,831	1,869	1,884	1,882	1,869	1,849	1,824	1,749	1,668	1,516	1,388	1,076