4, 4. Covariance .

Objekttyp: **Chapter**

Zeitschrift: **L'Enseignement Mathématique**

Band (Jahr): **6 (1960)**

Heft 1: **L'ENSEIGNEMENT MATHÉMATIQUE**

PDF erstellt am: **12.07.2024**

On montre aisément que $\mathbf{SC}\{\hat{\theta}_1, \hat{\theta}_2\}$ n'est autre que $\mathbf{SC}EM$ du modèle additif. Chacune de ces deux sommes de carrés est en effet due à un sous-espace de $\mathsf{V}^\star$ ayant deux dimensions et orthogonal tant à $\hat{\mu}$, $\hat{\Delta}\lambda$, $\hat{\Delta}_1\gamma$, $\hat{\Delta}_2\gamma$ qu'aux différences « internes » $x_{2i} - x_{2i-1}$, et un tel espace est unique.

$\hat{\theta}_1$ et $\hat{\theta}_2$ (et leurs combinaisons linéaires, notamment $\hat{\theta}_3$) portent le nom de « contrastes de non-additivité » (le terme « contrastes d'interaction » n'est guère heureux). On remarquera que, si les M_i déterminent entièrement les paramètres « orthogonalement estimables » μ, $\Delta\lambda$, $\Delta_1\gamma$, $\Delta_2\gamma$, θ_1, θ_2, ceux-ci à leur tour déterminent entièrement les M_i. C'est en le posant au moyen des paramètres $\mu, \ldots, \theta_2$ que le problème de la classification (2×3) (avec un nombre quelconque d'observations par cellule) se traite le plus aisément. Toutefois, ces paramètres ne restent orthogonalement estimables que si toutes les cellules renferment un même nombre d'observations.

Remarque. — Ces considérations s'étendent immédiatement aux autres problèmes de classification.

4, 4. *Covariance* [15]).

4, 41. Supposons que l'on dispose d'une observation de chacune de n variables aléatoires indépendantes, normales, de même variance σ^2, réparties en s groupes, le $i^{ème}$ groupe étant formé de n_i éléments, avec

$$\begin{cases} \mathbf{E}\,\mathbf{x}_{i,j} = T_i + \beta v_{i,j} \\ i = 1, \ldots, s\ ; \quad j = 1, \ldots, n_i\ ; \quad n_1 + \ldots + n_s = n\ ; \end{cases}$$

les $v_{i,j}$ étant des nombres certains [16]). On a ici

$$\mathfrak{b}_H = \| T_1, \ldots, T_s, \beta \|^T ,$$

et, en supposant que dans chaque groupe il y a au moins deux valeurs $v_{i,j}$ distinctes,

$$p = r = s + 1 .$$

On se convainc aisément que la matrice $\mathfrak{G}_H$ n'est pas diagonale. Toutefois, il est aisé d'orthogonaliser le problème; il suffit de poser

$$\begin{aligned} v_{i,-} &= (1/n_i) \sum_j v_{i,j} \\ u_{i,j} &= v_{i,j} - v_{i,} \ , \end{aligned}$$

ce qui entraîne

$$\begin{cases} \mathbf{E}\, x_{i,j} = A_i + \beta\, u_{i,j} \\ i = 1, \ldots, s\ ; \quad j = 1, \ldots, n_i\ ; \quad \sum_j u_{i,j} = 0 \\ A_i \equiv T_i - \beta\, v_{i,-}\ . \end{cases}$$

En fait, cela revient à référer B à une base $\mathfrak{K}$ telle que

$$\mathfrak{b}_K = \begin{Vmatrix} 1 & 0 & \ldots & 0 & -v_{1,-} \\ 0 & 1 & & & \\ \vdots & & \ddots & & \\ 0 & & & 1 & -v_{s,-} \\ 0 & \ldots & & 0 & 1 \end{Vmatrix} \mathfrak{b}_H$$

On a alors

$$\mathfrak{A}_K = \begin{Vmatrix} 1 & & & & u_{1,1} \\ \vdots & & & & \vdots \\ 1 & & & & u_{1,n_1} \\ & 1 & & & u_{2,1} \\ & \vdots & & & \vdots \\ & 1 & \ldots & & u_{2,n_2} \\ & & \ldots & & \vdots \\ & & & 1 & u_{s,1} \\ & & & \vdots & \vdots \\ & & & 1 & u_{s,n_s} \end{Vmatrix}$$

(les termes non écrits étant nuls).

4, 42. Il est alors utile de poser

$$\begin{aligned} S_{i;t,q} &= \sum_j (u_{i,j})^t\, (x_{i,j})^q\ , \\ S_{;t,q} &= \sum_i S_{i;t,q}\ , \\ S'_{i;t,q} &= \sum_j (v_{i,j})^t\, (x_{i,j})^q\ , \\ S'_{;t,q} &= \sum_i S'_{i;t,q}\ , \end{aligned}$$

ce qui conduit à

$$\begin{aligned} \mathfrak{A}_K^T\, \mathfrak{x} &= \| S_{1;0,1}, \ldots, S_{s;0,1}, S_{;1,1} \|\ , \\ \mathfrak{G}_K &= \operatorname{diag}\,(n_1, \ldots, n_s, S_{;2,0})\ , \end{aligned}$$

d'où

$$\begin{cases} \hat{\mathbf{A}}_i = \mathbf{S}_{i;0,1}/n_i \\ \hat{\beta} = \mathbf{S}_{;1,1}/S_{;2,0} \, , \end{cases}$$

$\mathbf{SC}T = \mathbf{S}_{;0,2}$ (avec n degrés de liberté),

$\mathbf{SC}\{\hat{\mathbf{A}}_i\} = \mathbf{red}\,[A_i] = (\mathbf{S}_{i;0,1})^2/n_i$ (avec 1 degré de liberté),

$\mathbf{SC}\{\hat{\beta}\} = \mathbf{red}\,[\beta] = (\mathbf{S}_{;1,1})^2/S_{;2,0}$ (avec 1 degré de liberté),

$\mathbf{SC}N = \sum_i SC\{\hat{\mathbf{A}}_i\} + SC\{\hat{\beta}\}$ (avec $s + 1$ degrés de liberté),

$$\mathbf{SC}E = \mathbf{SC}T - \mathbf{SC}N$$

$$= \sum_{i=1}^{s} \frac{n_i\,\mathbf{S}_{i;0,2} - (\mathbf{S}_{i;0,1})^2}{n_i} - \frac{\sum_i \left[n_i\,\mathbf{S}'_{i;1,1} - S'_{i;1,0}\,\mathbf{S}_{i;0,1}\right]}{\sum_i \left[n_i\,S'_{i;2,0} - \left(S'_{i;1,0}\right)^2\right]}$$

(avec $n - s - 1$ degrés de liberté).

σ^2 est estimé par $\mathbf{SC}E/(n - s - 1)$.

4, 43. En fait, l'intérêt se porte généralement sur l'estimation des différences $T_i - T_k$, les paramètres A_i n'ayant pas d'intérêt propre. On a, en posant $v_{i,-} - v_{k,-} = d_{i,k}$,

$$(\mathbf{T}_i - \mathbf{T}_k)^{\wedge} = \left[(\hat{\mathbf{A}}_i + \hat{\beta}\,v_{i,-}) - (\hat{\mathbf{A}}_k + \hat{\beta}\,v_{k,-})\right]^{\wedge}$$

$$= \hat{\mathbf{A}}_i - \hat{\mathbf{A}}_k + \hat{\beta}\,d_{i,k}$$

$$= \mathfrak{l}^{\star}\,\mathfrak{x} \equiv \left\| l_{1,1}\,, \ldots, \, l_{s,n_s} \right\| \mathfrak{x}$$

moyennant

$$l_{t,q} = d_{i,k}\,u_{t,q}/S_{;2,0} \qquad \text{si } t \neq i,\ t \neq k\,,$$

$$l_{i,q} = d_{i,k}\,u_{i,q}/S_{;2,0} + 1/n_i\,,$$

$$l_{k,q} = d_{i,k}\,u_{k,q}/S_{;2,0} - 1/n_k\,,$$

d'où

$$\mathfrak{l}^{\star}\,\mathfrak{l} = (d_{i,k})^2/S_{;2,0} + 1/n_i + 1/n_k\,.$$

Il résulte de là que l'expression

$$\Delta_{i,k} = \frac{\left[(\hat{\mathbf{T}}_i - \hat{\mathbf{T}}_k) - (T_i - T_k)\right]}{\sqrt{[\mathbf{SC}E/(n - s - 1)}\,.\,\sqrt{[(d_{i,k})^2/S_{;2,0} + 1/n_i + 1/n_k]}}$$

est une aléatoire $\mathbf{t}_{n-s-1}$.

On voit que l'erreur-type de $\Delta_{i,k}$, produit de σ par

$$\varepsilon_{i,k} = \left[\frac{(d_{i,k})^2}{S_{;2,0}} + \frac{1}{n_i} + \frac{1}{n_k}\right]^{1/2},$$

varie, en général, avec la paire (i, k), même si les n_i sont égaux entre eux ($n_1 = \ldots = n_s = \nu$). Toutefois, dans ce dernier cas, on utilise d'ordinaire une expression approchée de $\varepsilon_{i,k}$, obtenue en considérant les $v_{i,j}$ comme n observations indépendantes d'une même aléatoire, de moyenne v et écart-type θ [17]). En désignant par $\mathbf{M}$ la valeur moyenne prise par rapport à la distribution de cette aléatoire (fictive), on a

$$\mathbf{M}\,(v_{i,-} - v_{k,-})^2 = \mathbf{M}\,[(v_{i,-} - v) - (v_{k,-} - v)]^2$$

$$= 2\,\mathbf{M}\,(v_{i,-} - v)^2 = 2\,\theta^2/\nu\ ;$$

$$\mathbf{M}\,S_{;2,0} = \mathbf{M} \sum_i \sum_j (v_{i,j} - v_{i,-})^2 = \sum_i (\nu - 1)\,\theta = k\,(\nu - 1)\,\theta^2\ ;$$

$$\left\{\frac{\mathbf{M}\,(v_{i,-} - v_{k,-})^2}{\mathbf{M}\,S_{;2,0}} + \frac{1}{n_i}\,\frac{1}{n_k}\right\}^{1/2} = \left\{\frac{2}{\nu}\left[1 + \frac{1}{k\,(\nu - 1)}\right]\right\}^{1/2};$$

c'est cette dernière expression que l'on utilise comme valeur approchée de $\varepsilon_{i,k}$.

4, 44. Il est utile aussi de calculer la somme de carrés due au sous-espace $\mathbf{U}^\star$ engendré par les estimateurs des contrastes $T_i - T_k$, afin de pouvoir construire une épreuve globale de la nullité de ceux-ci. On pourrait évidemment construire une base orthogonale de $\mathbf{U}^\star$, mais les calculs nécessaires sont d'une grande complication. Il est beaucoup plus simple d'appliquer le théorème du § 2, 31, en procédant comme suit: on considère le modèle

$$\mathbf{E}\,\mathbf{x}_{i,j} = T + \beta\,v_{i,j}$$

et on calcule la somme de carrés de l'erreur qui lui est associée, c'est-à-dire (voir § 4, 211)

$$(\mathbf{SC}E)_{T_i = T} = \mathbf{S}_{;0,2} - \frac{(\mathbf{S}_{;0,1})^2}{n} - \frac{(\mathbf{S}_{;1,1})^2}{S_{;2,0}}\ ;$$

or ce modèle s'obtient à partir du modèle initial en supposant que, pour tout vecteur $\mathfrak{l}^\star$ de $\mathbf{U}^\star$, on a $\mathfrak{E}\,\mathfrak{l}^\star = 0$; par conséquent

$$\begin{aligned} \mathbf{SC}\,\mathbf{U}^\star &= (\mathbf{SC}E)_{T_i=T} - \mathbf{SC}E \\ &= \sum_{i=1}^{n} \frac{(\mathbf{S}_{i;0,1})^2}{n_i} - \frac{(\mathbf{S}_{;0,1})^2}{n} \,. \end{aligned} \tag{19}$$

Si on pose

$$\mathbf{x}_{i,-} = \frac{1}{n}\sum_j \mathbf{x}_{i,j}\,, \quad \mathbf{x}_{-,-} = \frac{1}{n}\sum_i \sum_j \mathbf{x}_{i,j} = \sum_i \left(\frac{n_i}{n}\right) \mathbf{x}_{i,-}\,,$$

la formule (19) peut s'écrire

$$\begin{aligned} \mathbf{SC}\,\mathbf{U}^\star &= \sum_{i=1}^{s} \frac{\left(\sum_j \mathbf{x}_{i,j}\right)^2}{n_i} - \frac{1}{n}\left(\sum_{i=1}^{n} \sum_j \mathbf{x}_{i,j}\right)^2 \\ &= n \sum_{i=1}^{s} \frac{n_i}{n} (\mathbf{x}_{i,-})^2 - n \left(\sum_{i=1}^{s} \frac{n_i}{n} \mathbf{x}_{i,-}\right)^2 \\ &= \sum_{i=1}^{s} n_i (\mathbf{x}_{i,-} - \mathbf{x}_{-,-})^2 \,. \end{aligned} \tag{20}$$

L'expression (20), moins aisée à mettre en nombres que l'expression (19), est plus parlante qu'elle.

L'hypothèse $T_1 = T_2 = \ldots = T_s$ s'éprouve en comparant

$$\frac{\mathbf{SC}\,\mathbf{U}^\star/(s-1)}{\mathbf{SC}E/(n-s-1)}$$

à la distribution de **F** à $(s - 1)$ et $(n - s - 1)$ degrés de liberté. La table d'analyse de variance se présente ainsi:

$$\begin{array}{llr} SCT = S_{;0,2} & & n \text{ d.l.} \\ SCN \begin{cases} \text{red}\,[T] = (S_{;0,1})^2/n \\ \text{red}\,[\beta] = (S_{;1,1})^2/S_{;2,0} \\ \text{red}\,[T_1 - T_2, \ldots, T_1 - T_s \mid T, \beta] = \mathrm{SC}\,\mathbf{U}^\star \end{cases} & (s+1) & \begin{cases} 1 \text{ d.l.} \\ 1 \text{ d.l.} \\ (s-1) \text{ d.l.} \end{cases} \\ SCE = SCT - SCN \quad (n-s-1) \text{ d.l.} & & \end{array}$$

Si plusieurs paires (i, j) correspondent à une même valeur de $v_{i,j}$, on peut introduire, en outre, la décomposition habituelle de SCE en SCint et $SCEM$.