

4. Exemples.

Objekttyp: **Chapter**

Zeitschrift: **L'Enseignement Mathématique**

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

Heft 1: **L'ENSEIGNEMENT MATHÉMATIQUE**

PDF erstellt am: **25.09.2024**

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

On calcule et vérifie \mathfrak{S} comme ci-dessus, puis:

$$\begin{aligned} x a_1 &= n & y a_1 + u a_2 &= p & z a_1 + v a_2 + w a_3 &= q \\ A &= a_1 + a_2 + a_3 & T &= n + p + q & t a_1 + r a_2 + s a_3 &= T \\ c_1 &= x & c_2 &= y + u & c_3 &= z + v + w \\ w b_3 &= a_3 & v b_3 + u b_2 &= a_2 & z b_3 + y b_2 + x b_1 &= a_1 \\ b_1 c_1 + b_2 c_2 + b_3 c_3 &= A \\ SCN &= a_1^2 + a_2^2 + a_3^2. \end{aligned}$$

4. EXEMPLES.

4, 1. Les épreuves de Student.

4, 11. Soit $\mathbf{x}_1, \dots, \mathbf{x}_n$ un échantillon simple et fortuit d'une population normale de moyenne μ et écart-type σ . La théorie des modèles linéaires s'applique ici, avec

$$r = p = 1, \quad \mathfrak{X} = \|\mathbf{1}, \dots, \mathbf{1}\| \quad b_H = \|\mu\|,$$

$$\mathfrak{X}^T \mathfrak{X} = \|n\|, \quad \mathfrak{X}^T \mathfrak{x} = \sum_1^n \mathbf{x}_i,$$

et le système normal se réduit à

$$n \hat{\mu} = \sum_1^n \mathbf{x}_i, \quad \hat{\mu} = \frac{1}{n} \sum_1^n \mathbf{x}_i = \mathbf{m}.$$

On a alors

$$SCN = (\sum \mathbf{x}_i)^2/n, \quad SCT = \sum_1^n \mathbf{x}_i^2,$$

$$SCE = \sum_1^n \mathbf{x}_i^2 - \left(\sum_1^n \mathbf{x}_i \right)^2/n = \sum_1^n (\mathbf{x}_i - \mathbf{m})^2 = \frac{n \sum_1^n \mathbf{x}_i^2 - \left(\sum_1^n \mathbf{x}_i \right)^2}{n}.$$

Si $\mu = a$, l'expression

$$\frac{(\mathbf{m} - a) \sqrt{(n-1)}}{\sqrt{(SCE/n)}} = \sqrt{(n-1)} \frac{\sum_1^n \mathbf{x}_i - na}{\sqrt{(n SCE)}}$$

est une aléatoire t_{n-1} ; SCE/σ^2 est une aléatoire χ_{n-1}^2 .

4, 12. Soit $(\mathbf{x}_1, \dots, \mathbf{x}_q)$ un échantillon simple et fortuit d'une population normale de moyenne μ_1 et écart-type σ , $(\mathbf{x}_{q+1}, \dots, \mathbf{x}_n)$ un échantillon simple et fortuit d'une population normale de moyenne μ_2 et écart-type σ , les deux échantillons étant mutuellement indépendants. La théorie des modèles linéaires s'applique encore:

$$\mathbf{E} \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_q \\ \mathbf{x}_{q+1} \\ \vdots \\ \mathbf{x}_n \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},$$

$$r = p = 2, \quad \mathbf{b}_H^T = \|\mu_1, \mu_2\|, \quad \mathfrak{A}^T \mathfrak{A} = \begin{pmatrix} q & 0 \\ 0 & n - q \end{pmatrix}.$$

Si on pose

$$\sum_1^q \mathbf{x}_i^s = \mathbf{S}_{1,s}, \quad \sum_{q+1}^n \mathbf{x}_i^s = \mathbf{S}_{2,s},$$

on a

$$\hat{\mu}_1 = \mathbf{S}_{1,1}/q, \quad \hat{\mu}_2 = \mathbf{S}_{2,1}/(n - q),$$

$$\text{SCE} = \frac{q \mathbf{S}_{1,2} - (\mathbf{S}_{1,1})^2}{q} + \frac{(n - q) \mathbf{S}_{2,2} - (\mathbf{S}_{2,1})^2}{n - q},$$

de sorte que, sous l'hypothèse $\mu_1 - \mu_2 = \Delta$, l'expression

$$\frac{(\hat{\mu}_1 - \hat{\mu}_2 - \Delta) \sqrt{(n - 2)}}{\sqrt{\left[\left(\frac{1}{q} + \frac{1}{n - q}\right) \text{SCE}\right]}}$$

est une aléatoire t_{n-2} ; SCE/σ^2 est une aléatoire χ_{n-2}^2 .

4, 2. Problèmes de régression.

4, 211. Supposons que, u_1, \dots, u_s étant des constantes certaines deux à deux distinctes, on ait $n = \sum_1^s k_i$ variables aléatoires $\mathbf{x}_{i,j}$ ($i = 1, \dots, s; j = 1, \dots, k_i$), normales, de variance commune σ^2 , indépendantes, avec

$$\mathbf{E} \mathbf{x}_{i,j} = \alpha + \beta u_i. \tag{17}$$

Ce cas rentre dans le cadre général des modèles linéaires moyennant

$$b_H = \|\alpha \quad \beta\|^T, \quad \mathfrak{A} = \left\| \begin{array}{cccc} 1 & \dots & 1 & 1 & \dots & 1 & \dots & 1 & \dots & 1 \\ u_1 & \dots & u_1 & u_2 & \dots & u_2 & \dots & u_s & \dots & u_s \end{array} \right\|^T$$

(\mathfrak{A} composée de s groupes ayant respectivement k_1, k_2, \dots, k_s lignes identiques entre elles); ici, $r = p = 2$. Si l'on procède comme au § 3, 232, en posant

$$\sum_{i=1}^s \sum_{j=1}^{k_i} u_i^q x_{i,j}^t = S_{q,t}; \quad S_{0,0} \equiv n; \quad S_{q,0} = \sum_1^i k_i u_i^q;$$

$$L_{u,u} = \frac{n S_{2,0} - S_{1,0}^2}{n}; \quad L_{u,x} = \frac{n S_{1,1} - S_{1,0} S_{0,1}}{n};$$

on a, successivement,

$$\begin{aligned} \mathfrak{G} &= \left\| \begin{array}{cc} n & S_{1,0} \\ S_{1,0} & S_{2,0} \end{array} \right\| \\ \mathfrak{E} &= \left\| \begin{array}{cc} \sqrt{n} & S_{1,0}/\sqrt{n} \\ 0 & \sqrt{L_{u,u}} \end{array} \right\| & \mathfrak{A}^T \mathfrak{x} &= \left\| \begin{array}{c} S_{0,1} \\ S_{1,1} \end{array} \right\| & \hat{\mathfrak{b}}_K &= \left\| \begin{array}{c} S_{0,1}/\sqrt{n} \\ L_{u,x} \end{array} \right\| \\ \hat{\mathfrak{b}}_H^T &= \left\| \begin{array}{cc} \frac{S_{0,1}}{n} - \frac{S_{1,0}}{n} \frac{L_{u,x}}{L_{u,u}} & \frac{L_{u,x}}{L_{u,u}} \end{array} \right\| \end{aligned}$$

d'où la table d'analyse de variance:

$$\begin{array}{ll} SCT = S_{0,2} & n \text{ d.l.} \\ \left\{ \begin{array}{l} \text{red } [\alpha] = (S_{0,1})^2/n \\ \text{red } [\beta | \alpha] = (L_{u,x})^2 \end{array} \right. & \left\{ \begin{array}{l} 1 \text{ d.l.} \\ 1 \text{ d.l.} \end{array} \right. \\ SCN = \text{red } [\alpha] + \text{red } [\beta | \alpha] & 2 \text{ d.l.} \\ SCE = SCT - SCN & (n - 2) \text{ d.l.} \end{array}$$

4, 212. On peut traiter ce même problème d'une manière un peu différente, en posant

$$\begin{aligned} u'_i &= u_i - S_{1,0}/n, \\ \alpha + \beta u_i &\equiv \alpha' + \beta' u'_i \quad (\beta' = \beta, \quad \alpha' = \alpha + \beta S_{1,0}/n); \end{aligned}$$

ceci revient à changer de base dans \mathbf{B} , et nous écrivons

$$b_{H'} = \|\alpha' \quad \beta'\|.$$

On a alors, en marquant de l'apostrophe les expressions propres à la forme actuelle du modèle considéré,

$$\mathfrak{X}' = \left\| \begin{array}{cccc} 1 & \dots & 1 & 1 & \dots & 1 & \dots & 1 & \dots & 1 \\ u'_1 & \dots & u'_1 & u'_2 & \dots & u'_2 & \dots & u'_s & \dots & u'_s \end{array} \right\|^T,$$

$$S'_{1,0} = 0, \quad S'_{2,0} = L_{u,u}, \quad S'_{1,1} = \mathbf{L}_{u,x},$$

$$\mathfrak{G}' = \text{diag}(n, L_{u,u}); \quad \mathfrak{X}'^T \mathfrak{X}' = \left\| \begin{array}{cc} S_{0,1} & \mathbf{L}_{u,x} \end{array} \right\|^T$$

$$\hat{\alpha}' = S_{0,1}/n \quad \hat{\beta}' = \mathbf{L}_{u,x}/L_{u,u} (= \hat{\beta}).$$

La table d'analyse de la variance ne change évidemment pas.

La méthode du § 3, 232 constitue, en quelque sorte, une orthogonalisation a posteriori: on part de \mathfrak{b}_H , les vecteurs $\varepsilon_i^T \hat{\mathfrak{b}}_H (\in \mathbf{V}^*)$ ne sont pas orthogonaux, mais les calculs introduisent d'eux-même une base \mathfrak{K} telle que les vecteurs $\varepsilon_i^T \hat{\mathfrak{b}}_K$ soient orthogonaux. Ici, nous venons de procéder à une orthogonalisation a priori: nous avons d'emblée introduit une base \mathfrak{D}' telle que la matrice \mathfrak{G}' relative à cette base soit diagonale, ce qui garantit l'orthogonalité des vecteurs $\varepsilon_i^T \hat{\mathfrak{b}}_{H'}$. Cette seconde méthode est souvent préférable à la première. C'est sur elle que reposent, notamment, les procédés de « codage linéaire » utilisés, dans les manuels d'analyse statistique, pour l'étude des plans factoriels à facteurs quantitatifs (plans factoriels « de régression »).

4, 213. Il arrive que l'on désire contrôler, par les observations elles-mêmes, la validité de la relation (17). Le modèle basé sur (17) est alors considéré comme un cas particulier du modèle défini par

$$\mathbf{E} x_{i,j} = M_i;$$

dans ce modèle plus général, l'espace des erreurs, \mathbf{V}_* , est engendré par les fonctionnelles de \mathfrak{X} qui sont de la forme $(x_{i,j} - x_{i,k})$; il admet donc la base suivante:

$$x_{i,1} - x_{i,2}, \dots, x_{i,1} - x_{i,k_i}, \quad i = 1, 2, \dots, s,$$

laquelle s'orthogonalise en

$$\left\{ \begin{array}{l} \delta_{i,t}^* \mathfrak{X} \equiv x_{i,1} + \dots + x_{i,t-1} - (t-1)x_{i,t} \\ i = 1, \dots, s; \quad t = 2, \dots, k_i. \end{array} \right.$$

On a alors

$$\begin{aligned}
 SCint \equiv SCV_* &= \sum_{i=1}^s \sum_{t=2}^{k_i} SC \{ \delta_{i,t}^* \} \\
 &= \sum_{i=1}^s \sum_{t=2}^{k_i} (\delta_{i,t}^*)^2 / (t-1)t \\
 &= \sum_{i=1}^s \frac{k_i \sum_{j=1}^{k_i} x_{i,j}^2 - \left(\sum_{j=1}^{k_i} x_{i,j} \right)^2}{k_i}
 \end{aligned} \tag{18}$$

avec $\sum_i (k_i - 1) = n - s$ degrés de liberté. Le contrôle envisagé n'est donc possible que si l'un au moins des entiers k_i est > 1 , et il ne présente, en pratique, quelque intérêt que si $n - s$ est, au moins, de l'ordre de s . On le fait alors en comparant $SCEM$ à $SCint$ au moyen des tables de **F**. Si l'on procède ainsi, il sied d'utiliser $SCint$, et non SCE , comme dénominateur des divers F calculés.

Remarque. — Le calcul qui a conduit à l'expression (18) de $SCint$ est valide dans des conditions extrêmement générales.

4, 214. On peut évidemment éprouver des hypothèses très diverses relativement à α et β (ou, ce qui revient au même, à α' et β')¹⁴). Ainsi, l'on pourrait éprouver l'hypothèse $\beta = a$, a étant un nombre donné; il suffit d'appliquer la formule du § 2, 23, en remplaçant, au besoin, SCE et $(n - r)$ par $SCint$ et $(n - s)$. Le seul point un peu délicat est le calcul de $l^* l$; or, on a

$$\hat{\beta} = \| l_{1,1}, \dots, l_{1,k_1}, \dots, l_{s,k_s} \| \quad \varepsilon \equiv l^* \varepsilon$$

moyennant

$$l_{i,j} = (n u_i - S_{1,0}) / n L_{u,u} ;$$

on a donc

$$l^* l = \sum_i \sum_j l_{i,j}^2 = 1 / L_{u,u} ;$$

par conséquent, sous l'hypothèse $\beta = a$, l'expression

$$\frac{(\hat{\beta} - a) \sqrt{(n - s)}}{\sqrt{l^* l \cdot SCint}} = (\hat{\beta} - a) \sqrt{\frac{(n - s) L_{u,u}}{SCint}}$$

est une valeur observée d'une aléatoire t_{n-s} .

On éprouverait de même, par exemple, l'hypothèse que, pour des valeurs données u_0 et x_0 , on a $\alpha + \beta u_0 = x_0$ (on considérerait l'expression $\hat{\alpha} + \hat{\beta} u_0$, qui, sous cette hypothèse, a comme moyenne x_0).

4, 22. Supposons que, u et v étant deux variables certaines, on ait

$$E x_{u,v} = \beta_0 + \beta_1 u + \beta_2 v, \tag{19}$$

et que les observations aient été faites aux « points » $(0, 0)$, $(2, 0)$, $(2, 1)$, $(1, 2)$, $(0, 2)$ et $(1, 1)$; les observations sont, ici encore, censées être des valeurs observées d'aléatoires normales, indépendantes, de même variance σ^2 . La théorie générale s'applique alors, avec

$$b_H = \|\beta_0, \beta_1, \beta_2\|^T, \quad \mathcal{X}_H = \left\| \begin{array}{cccccc} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 2 & 2 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 & 2 & 1 \end{array} \right\|^T, \quad n = 6; \quad p = r = 3.$$

Il est commode de traiter ce problème par orthogonalisation a priori; on rapporte donc \mathbf{B} à une base \mathfrak{R} telle que les colonnes de \mathcal{X}_K soient deux à deux orthogonales; si l'on pose $b_K = \|\gamma_0, \gamma_1, \gamma_2\|^T$, cela revient à chercher deux polynômes du premier degré, $\varphi(u)$ et $\psi(u, v)$, tels que

$$\begin{aligned} \gamma_0 + \gamma_1 \varphi(u) + \gamma_2 \psi(u, v) &\equiv \beta_0 + \beta_1 u + \beta_2 v, \\ \sum_1^6 \varphi(u_i) &= 0, \quad \sum_1^6 \psi(u_i, v_i) = 0, \quad \sum_1^6 \varphi(u_i) \psi(u_i, v_i) = 0. \end{aligned}$$

On peut prendre

$$\varphi(u) = u - 1, \quad \psi(u, v) = -5 + u + 4v,$$

ce qui correspond à

$$b_H = \left\| \begin{array}{ccc} 1 & -1 & -5 \\ 0 & 1 & 1 \\ 0 & 0 & 4 \end{array} \right\| b_K.$$

On a alors

$$\begin{aligned} E \left\| \begin{array}{c} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \\ \mathbf{x}_5 \\ \mathbf{x}_6 \end{array} \right\| &= \left\| \begin{array}{ccc} 1 & -1 & -5 \\ 1 & 1 & -3 \\ 1 & 1 & 1 \\ 1 & 0 & 4 \\ 1 & -1 & 3 \\ 1 & 0 & 0 \end{array} \right\| \left\| \begin{array}{c} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{array} \right\| \\ \mathcal{X}_K^T \mathfrak{K} &\equiv \left\| \begin{array}{c} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \mathbf{A}_3 \end{array} \right\| = \left\| \begin{array}{c} \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3 + \mathbf{x}_4 + \mathbf{x}_5 + \mathbf{x}_6 \\ -\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3 - \mathbf{x}_5 \\ -5\mathbf{x}_1 - 3\mathbf{x}_2 + \mathbf{x}_3 + 4\mathbf{x}_4 + \mathbf{x}_5 \end{array} \right\| \end{aligned}$$

d'où

$$\mathfrak{S}_K = \left\| \begin{array}{ccc} 6 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 60 \end{array} \right\|, \quad \begin{array}{l} \hat{\gamma}_0 = \mathbf{A}_1/6 \\ \hat{\gamma}_1 = \mathbf{A}_2/4 \\ \hat{\gamma}_2 = \mathbf{A}_3/60 \end{array}$$

$$\mathbf{SC} \{ \hat{\gamma}_0 \} = \mathbf{A}_1^2/6, \quad \mathbf{SC} \{ \hat{\gamma}_1 \} = \mathbf{A}_2^2/4, \quad \mathbf{SC} \{ \hat{\gamma}_2 \} = \mathbf{A}_3^2/60.$$

$$\mathbf{SC} \{ \hat{\gamma}_0 \} = \mathbf{red}[\beta_0], \quad \mathbf{SC} \{ \hat{\gamma}_1 \} = \mathbf{red}[\beta_1 | \beta_0], \quad \mathbf{SC} \{ \hat{\gamma}_2 \} = \mathbf{red}[\beta_2 | \beta_0, \beta_1]$$

$$\mathbf{SC} N = \mathbf{SC} \{ \hat{\gamma}_0 \} + \mathbf{SC} \{ \hat{\gamma}_1 \} + \mathbf{SC} \{ \hat{\gamma}_2 \},$$

$$\mathbf{SCE} = \mathbf{SCT} - \mathbf{SC} N.$$

4, 3. Problèmes de classification.

4, 31. Supposons que l'on dispose des valeurs observées de douze aléatoires normales, indépendantes, de même variance σ^2 , classées suivant deux critères: « lignes », de « valeurs » L_1 et L_2 , et « colonnes », de « valeurs » C_1, C_2, C_3 , suivant le schéma

| | C_1 | C_2 | C_3 |
|-------|------------|---------------|------------------|
| L_1 | x_1, x_2 | x_3, x_4 | x_5, x_6 |
| L_2 | x_7, x_8 | x_9, x_{10} | x_{11}, x_{12} |

On suppose a priori qu'il y a additivité, c'est-à-dire qu'il existe cinq nombres $\lambda_1, \lambda_2, \gamma_1, \gamma_2, \gamma_3$ tels que la valeur moyenne d'une observation de la ligne L_i et de la colonne C_k soit $\lambda_i + \gamma_k$ ($i = 1, 2; k = 1, 2, 3$). On a donc, par hypothèse,

$$\mathbf{E} \left\| \begin{array}{c} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \\ \mathbf{x}_5 \\ \mathbf{x}_6 \\ \mathbf{x}_7 \\ \mathbf{x}_8 \\ \mathbf{x}_9 \\ \mathbf{x}_{10} \\ \mathbf{x}_{11} \\ \mathbf{x}_{12} \end{array} \right\| = \left\| \begin{array}{ccc} 1 & 1 & \\ 1 & 1 & \\ 1 & & 1 \\ 1 & & 1 \\ 1 & & & 1 \\ 1 & & & 1 \\ & 1 & 1 & \\ & 1 & 1 & \\ & 1 & & 1 \\ & 1 & & 1 \\ & 1 & & & 1 \\ & 1 & & & 1 \end{array} \right\| \left\| \begin{array}{c} \lambda_1 \\ \lambda_2 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{array} \right\|$$

Si on appelle $S_{i,-}$ la somme des observations de la ligne L_i , et $S_{-,j}$ celle des observations de la colonne C_k , les équations normales s'écrivent:

$$6 \hat{\lambda}_1 + 2 \hat{\gamma}_1 + 2 \hat{\gamma}_2 + 2 \hat{\gamma}_3 = S_{1,-} \quad (a)$$

$$6 \hat{\lambda}_2 + 2 \hat{\gamma}_1 + 2 \hat{\gamma}_2 + 2 \hat{\gamma}_3 = S_{2,-} \quad (b)$$

$$2\hat{\lambda}_1 + 2\hat{\lambda}_2 + 4\hat{\gamma}_1 = \mathbf{S}_{-,1} \quad (c)$$

$$2\hat{\lambda}_1 + 2\hat{\lambda}_2 + 4\hat{\gamma}_2 = \mathbf{S}_{-,2} \quad (d)$$

$$2\hat{\lambda}_1 + 2\hat{\lambda}_2 + 4\hat{\gamma}_3 = \mathbf{S}_{-,3} \quad (e)$$

Ici, manifestement, $p = 5$, $r = 4$ [en effet, $(a) + (b) \equiv (c) + (d) + (e)$]. On est donc amené à mettre en évidence quatre combinaisons estimables fondamentales; on peut prendre

$$\begin{aligned} \mu &= 6(\lambda_1 + \lambda_2) + 4(\gamma_1 + \gamma_2 + \gamma_3) \\ \Delta\lambda &= \lambda_1 - \lambda_2 \\ \Delta_1\gamma &= \gamma_1 - \gamma_2, \quad \Delta'\gamma = \gamma_1 - \gamma_3. \end{aligned}$$

On constate immédiatement que les estimateurs privilégiés de ces quatre combinaisons sont deux à deux orthogonaux, à l'exception près de la dernière paire; l'orthogonalité complète est atteinte en remplaçant $\Delta'\gamma$ par

$$\Delta_2\gamma = (\lambda_1 + \lambda_2 - 2\lambda_3)/2 = (\lambda_1 + \lambda_2)/2 - \lambda_3.$$

Alors:

$$\begin{aligned} \hat{\mu} &= \mathbf{S}_{1,-} + \mathbf{S}_{2,-} (= \mathbf{S}_{-,1} + \mathbf{S}_{-,2} + \mathbf{S}_{-,3}), \\ \Delta\hat{\lambda} &= (\mathbf{S}_{1,-} - \mathbf{S}_{2,-})/6 \\ \Delta_1\hat{\gamma} &= (\mathbf{S}_{-,1} - \mathbf{S}_{-,2})/4 \\ \Delta_2\hat{\gamma} &= (\mathbf{S}_{-,1} + \mathbf{S}_{-,2} - 2\mathbf{S}_{-,3})/8. \end{aligned}$$

Ici, on peut calculer

$$\mathbf{SCint} = (1/2)[(\mathbf{x}_1 - \mathbf{x}_2)^2 + \dots + (\mathbf{x}_{11} - \mathbf{x}_{12})^2],$$

avec six degrés de liberté, puis, avec deux degrés de liberté,

$$\mathbf{SCEM} = \mathbf{SCE} - \mathbf{SCint}.$$

On notera que, ici, on a

$$\mathbf{SC}\{\hat{\mu}\} = \mathbf{red}[\mu] = \mathbf{red}[\mu | \Delta\lambda] = \dots = \mathbf{red}[\mu | \Delta\lambda, \Delta_1\gamma, \Delta_2\gamma],$$

et des relations analogues pour les autres paramètres; ceci en vertu de l'orthogonalité de leurs estimateurs privilégiés.

\mathbf{SCEM} peut servir à éprouver l'hypothèse d'additivité, mais on ne le voit clairement qu'en étudiant le modèle non additif.

4, 32. Ne supposons donc plus a priori qu'il y ait additivité; admettons que

$$\mathbf{E} \mathbf{x}_{2i-1} = \mathbf{E} \mathbf{x}_{2i} = M_i ,$$

de sorte que $p = 6$, $\mathbf{b}_H = \parallel M_1 M_2 M_3 M_4 M_5 M_6 \parallel^T$. On voit aisément que

$$\mathbb{G} = \text{diag} (2, \dots, 2)$$

d'où

$$\hat{\mathbf{M}}_i = (\mathbf{x}_{2i-1} + \mathbf{x}_{2i})/2 ,$$

$$SC \{ \hat{\mathbf{M}}_i \} = (\mathbf{x}_{2i-1} + \mathbf{x}_{2i})^2/2 ,$$

$$SCN = \sum_1^6 SC \{ \hat{\mathbf{M}}_i \} ,$$

$$SCE = \sum_1^{12} \mathbf{x}_i^2 - SCN = (1/2) \sum_1^6 (\mathbf{x}_{2i} - \mathbf{x}_{2i-1})^2 .$$

Donc, pour le modèle général actuel, SCE vaut l'expression SCint du modèle additif.

L'hypothèse d'additivité (c'est-à-dire, répétons-le, l'hypothèse qu'il existe cinq nombres $\lambda_1, \lambda_2, \gamma_1, \gamma_2, \gamma_3$ tels que

$$M_1 = \lambda_1 + \gamma_1 , \quad M_2 = \lambda_1 + \gamma_2 , \dots , \quad M_6 = \lambda_2 + \gamma_3)$$

est satisfaite si et seulement si

$$\theta_1 \equiv M_1 - M_2 - M_4 + M_5 = 0 , \quad \theta_2 \equiv M_1 - M_3 - M_4 + M_6 = 0 .$$

On doit donc former, pour éprouver cette hypothèse,

$$\hat{\theta}_1 = \hat{\mathbf{M}}_1 - \hat{\mathbf{M}}_2 - \hat{\mathbf{M}}_4 + \hat{\mathbf{M}}_5 , \quad \hat{\theta}_2 = \hat{\mathbf{M}}_1 - \hat{\mathbf{M}}_3 - \hat{\mathbf{M}}_4 + \hat{\mathbf{M}}_6 ,$$

puis $SC \{ \hat{\theta}_1, \hat{\theta}_2 \}$, et éprouver si $SC \{ \hat{\theta}_1, \hat{\theta}_2 \}$ est, ou non, significativement plus grande que SCE (SCint du modèle additif). $\hat{\theta}_2$ n'est pas orthogonal à $\hat{\theta}_1$, mais bien

$$\hat{\theta}_3 = 2 \hat{\theta}_2 - \hat{\theta}_1 = \mathbf{M}_1 + \mathbf{M}_2 - 2\mathbf{M}_3 + \mathbf{M}_4 + \mathbf{M}_5 - 2\mathbf{M}_6 ;$$

or,

$$SC \{ \hat{\theta}_1 \} = \hat{\theta}_1^2/8 , \quad SC \{ \hat{\theta}_3 \} = \hat{\theta}_3^2/24 ,$$

$$SC \{ \hat{\theta}_1, \hat{\theta}_2 \} = SC \{ \hat{\theta}_1, \hat{\theta}_3 \} = SC \{ \hat{\theta}_1 \} + SC \{ \hat{\theta}_3 \} ,$$

ce qui permet d'éprouver l'additivité.

On montre aisément que $SC \{ \hat{\theta}_1, \hat{\theta}_2 \}$ n'est autre que $SCEM$ du modèle additif. Chacune de ces deux sommes de carrés est en effet due à un sous-espace de V^* 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 » $\mu, \Delta\lambda, \Delta_1\gamma, \Delta_2\gamma, \theta_1, \theta_2$, ceux-ci à leur tour déterminent entièrement les M_i . C'est en le posant au moyen des paramètres μ, \dots, θ_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 $i^{\text{ème}}$ groupe étant formé de n_i éléments, avec

$$\begin{cases} \mathbf{E} x_{i,j} = T_i + \beta \varphi_{i,j} \\ i = 1, \dots, s; \quad j = 1, \dots, n_i; \quad n_1 + \dots + n_s = n; \end{cases}$$

les $\varphi_{i,j}$ étant des nombres certains ¹⁶). On a ici

$$b_H = \| T_1, \dots, T_s, \beta \|^T,$$

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

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

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

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

ce qui entraîne

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

En fait, cela revient à référer \mathbf{B} à une base \mathcal{R} telle que

$$b_K = \left\| \begin{array}{cccc|c} 1 & 0 & \dots & 0 & -\varphi_{1,-} \\ 0 & 1 & & & \\ \vdots & & \ddots & & \\ 0 & & & 1 & -\varphi_{s,-} \\ 0 & \dots & & 0 & 1 \end{array} \right\| b_H$$

On a alors

$$\mathcal{R}_K = \left\| \begin{array}{cccc|c} 1 & & & & u_{1,1} \\ \vdots & & & & \vdots \\ 1 & & & & u_{1,n_1} \\ & 1 & & & u_{2,1} \\ & \vdots & & & \vdots \\ & 1 & \dots & & u_{2,n_2} \\ & & \dots & & \vdots \\ & & & 1 & u_{s,1} \\ & & & \vdots & \vdots \\ & & & 1 & u_{s,n_s} \end{array} \right\|$$

(les termes non écrits étant nuls).

4, 42. Il est alors utile de poser

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

ce qui conduit à

$$\begin{aligned} \mathcal{R}_K^T \mathcal{R} &= \| S_{1;0,1}, \dots, S_{s;0,1}, S_{;1,1} \|, \\ \mathbb{G}_K &= \text{diag} (n_1, \dots, n_s, S_{;2,0}), \end{aligned}$$

d'où

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

$SCT = S_{;0,2}$ (avec n degrés de liberté),

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

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

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

$SCE = SCT - SCN$

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

(avec $n - s - 1$ degrés de liberté).

σ^2 est estimé par $SCE/(n - s - 1)$.

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 $v_{i,-} - v_{k,-} = d_{i,k}$,

$$\begin{aligned} (T_i - T_k)^\wedge &= [(\hat{A}_i + \hat{\beta} v_{i,-}) - (\hat{A}_k + \hat{\beta} v_{k,-})]^\wedge \\ &= \hat{A}_i - \hat{A}_k + \hat{\beta} d_{i,k} \\ &= l^\star \mathfrak{z} \equiv \| l_{1,1}, \dots, l_{s,n_s} \| \mathfrak{z} \end{aligned}$$

moyennant

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

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

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

d'où

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

Il résulte de là que l'expression

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

est une aléatoire 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 ($n_1 = \dots = n_s = \nu$). Toutefois, dans ce dernier cas, on utilise d'ordinaire une expression approchée de $\varepsilon_{i,k}$, obtenue en considérant les $\nu_{i,j}$ comme n observations indépendantes d'une même aléatoire, de moyenne ν 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

$$\begin{aligned} \mathbf{M} (\nu_{i,-} - \nu_{k,-})^2 &= \mathbf{M} [(\nu_{i,-} - \nu) - (\nu_{k,-} - \nu)]^2 \\ &= 2 \mathbf{M} (\nu_{i,-} - \nu)^2 = 2 \theta^2 / \nu; \end{aligned}$$

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

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

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}^* 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}^* , 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

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

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

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

or ce modèle s'obtient à partir du modèle initial en supposant

que, pour tout vecteur l^* de U^* , on a $\mathcal{C}l^* = 0$; par conséquent

$$\begin{aligned} SCU^* &= (SCE)_{T_i=T} - SCE \\ &= \sum_{i=1}^n \frac{(S_{i;0,1})^2}{n_i} - \frac{(S_{;0,1})^2}{n}. \end{aligned} \quad (19)$$

Si on pose

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

la formule (19) peut s'écrire

$$\begin{aligned} SCU^* &= \sum_{i=1}^s \frac{(\sum_j x_{i,j})^2}{n_i} - \frac{1}{n} \left(\sum_{i=1}^n \sum_j x_{i,j}\right)^2 \\ &= n \sum_{i=1}^s \frac{n_i}{n} (x_{i,-})^2 - n \left(\sum_{i=1}^s \frac{n_i}{n} x_{i,-}\right)^2 \\ &= \sum_{i=1}^s n_i (x_{i,-} - x_{-,-})^2. \end{aligned} \quad (20)$$

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

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

$$\frac{SCU^*/(s-1)}{SCE/(n-s-1)}$$

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

$$\begin{array}{l} SCT = S_{;0,2} \qquad \qquad \qquad n \text{ d.l.} \\ \left. \begin{array}{l} \text{red } [T] = (S_{;0,1})^2/n \\ \text{red } [\beta] = (S_{;1,1})^2/S_{;2,0} \\ \text{red } [T_1 - T_2, \dots, T_1 - T_s | T, \beta] = SCU^* \end{array} \right\} (s+1) \begin{cases} 1 \text{ d.l.} \\ 1 \text{ d.l.} \\ (s-1) \text{ d.l.} \end{cases} \\ SCE = SCT - SCN \quad (n-s-1) \text{ d.l.} \end{array}$$

Si plusieurs paires (i, j) correspondent à une même valeur de $\rho_{i,j}$, on peut introduire, en outre, la décomposition habituelle de SCE en SC_{int} et SC_{EM} .