Méthode de calcul de quelques chiffres fournis par le logiciel EVA.

Logiciel qui exploite les résultats des tests de l'évaluation nationale faite en début d'année scolaire pour tous les élèves de Seconde.

A. PICHEREAU, Lycée M. de Valois, Angoulême

Pour les mathématiques ces tests se présentent sous la forme de 6 exercices décomposés en 44 questions ou items. La réponse de l'élève à chaque question est codée sous la forme : 0,1,2,3,9.
0 = question non traitée ;
1 = réponse juste ;
2 = réponse juste mais différente de celle attendue ;
3 = réponse inexacte mais reflétant un intérêt pédagogique particulier ;
9 = réponse inexacte.
A chaque élève est donc associée une 44-liste d'éléments dans {0,1,2,3,9}. Un ensemble (ou une sélection) d'élèves et un ensemble (ou une sélection) d'items étant choisis, le logiciel offre la possibilité de plusieurs traitements.
On ne s'intéresse ici qu'au traitement HIERARCHIE permettant d'obtenir des partitions d'un ensemble d'élèves. Il ne s'agit pas pour l'instant de décrire la méthode permettant d'obtenir une hiérarchie puis d'en déduire des partitions, **il s'agit, une partition étant obtenue, de décrire les méthodes de calcul qui permettent d'obtenir les chiffres que le logiciel lui associe.**
Ci-contre, 3 exemples de 3 partitions correspondant à une même hiérarchie.

Pour chaque partition apparaît <u>deux informations</u> :
1) un pourcentage r, qui augmente avec le nombre d'éléments de la partition.
2) un tableau (α_{ij}) , α_{ij} étant la "contribution " au groupe i de l'item j.

On constate facilement que pour tout i $\sum_{j=1}^{4} |\alpha_{ij}| = 100$

(sommation par ligne).
Le mot "contribution" est utilisé dans la documentation fournie avec le logiciel ; par contre cette documentation n'indique pas de quelle manière précise sont déterminés r et $\alpha_{i\,j}$. Cependant page 85 on peut y lire : *"la valeur du rapport variance intergroupe/variance intragroupe doit être la plus grande possible".*

<u>Méthode de calcul de r et $\alpha_{i\,j}$</u>

Tout d'abord les codes 2 sont transformés en 1 et les codes 3 et 9 sont transformés en 0 : le résultat de chaque élève est donc sur une 4-liste à 2 éléments {0,1}
(0 = échec ; 1 = réussite).

<u>Calcul de r.</u>

Chaque résultat d'un élève est considéré comme un élément de l'espace vectoriel $\mathbb{R}^4$, muni de la distance euclidienne :

$$d^2(x,y)=\|x-y\|^2= \sum_{i=1}^{4}(a_i - b_i)^2$$

si $x = (a_1, a_2, a_3, a_4)$ et $y = (b_1, b_2, b_3, b_4)$

Hiérarchie à indice symétrique

6 élèves de 2 13 — ***EXEMPLE 1***
4 items de Exercice 7A : Calculs algébriques

50.0%	Mat23A	Mat24A	Mat25A	Mat26A
Groupe n°1 (2)	76.2	-19.0	-4.8	0.0
Groupe n°2 (4)	-76.2	19.0	4.8	0.0

Groupe n°1	23A	24A	25A	26A
S. FILIEUL	1	9	1	1
A. FIDELE	1	2	9	1

Groupe n°2	23A	24A	25A	26A
S. DOUIT	9	1	3	1
S. DESCU.	9	1	2	1
S.DESSIM	9	1	2	2
F. DONAT	9	1	2	1

Hiérarchie à indice symétrique

6 élèves de 2 13 — ***EXEMPLE 2***
4 items de Exercice 7A : Calculs algébriques

78.6%	Mat23A	Mat24A	Mat25A	Mat26A
Groupe n°1 (1)	35.6	-55.6	8.9	0.0
Groupe n°2 (1)	48.5	3.0	-48.5	0.0
Groupe n°3 (4)	-76.2	19.0	4.8	0.0

Groupe n°1	23A	24A	25A	26A
S. FILIEUL	1	9	1	1

Groupe n°2	23A	24A	25A	26A
A. FIDELE	1	2	9	1

Groupe n°3	23A	24A	25A	26A
S. DOUIT	9	1	3	1
S. DESCU.	9	1	2	1
S.DESSIM	9	1	2	2
F. DONAT	9	1	2	1

Hiérarchie à indice symétrique

6 élèves de 2 13 — ***EXEMPLE 3***
4 items de Exercice 7A : Calculs algébriques

100.0%	Mat25A	Mat23A	Mat24A	Mat26A
Groupe n°1 (1)	8.9	35.6	-55.6	0.0
Groupe n°2 (1)	-48.5	48.5	3.0	0.0
Groupe n°3 (1)	-76.2	-19.0	4.8	0.0
Groupe n°4 (3)	44.4	-44.4	11.1	0.0

Groupe n°1	25A	23A	24A	26A
S. FILIEUL	1	1	9	1

Groupe n°2	25A	23A	24A	26A
A. FIDELE	9	1	2	1

Groupe n°3	25A	23A	24A	26A
S. DOUIT	3	9	1	1

Groupe n°4	25A	23A	24A	26A
S. DESCU.	2	9	1	1
S.DESSIM	2	9	1	2
F. DONAT	2	9	1	1

Remarquons que cette distance n'est absolument pas utilisée par le logiciel pour obtenir les hiérarchies et les partitions correspondantes.
Soit E l'ensemble des élèves et $G_1, G_2, \ldots G_k$ une partition de E en k groupes.

On appelle variation générale la quantité $V_g = \sum_{x\in E} \|x - \overline{x}\|^2$.

On appelle variation du groupe G_i la quantité
$V_i = \sum_{x\in G_i} \|x - \overline{x_i}\|^2$, avec $\overline{x}$ = moyenne de E , $\overline{x_i}$ = moyenne du groupe G_i.

(La moyenne est la somme, au sens habituel dans $\mathbb{R}^4$ des résultats des élèves divisée par le nombre d'élèves ; variance = variation moyenne).
Il y a une relation simple entre V_g et V_i :

$$Vg = \sum_{i=1}^{k} V_i + \sum_{i=1}^{k} n_i \|\overline{x_i} - \overline{x}\|^2 ,$$

n_i étant l'effectif du groupe G_i.
Ce résultat classique en analyse de variance est énoncé sous la forme suivante :

variation générale=variation intragroupe+variation intergroupe.

Il se démontre en remarquant que
$\|x_{ij} - \overline{x}\| = \|x_{ij} - \overline{x_i} + \overline{x_i} - \overline{x}\|$,
x_{ij} étant le $j^{ème}$ élève du groupe i. **Le nombre r donné par le logiciel est :** $\boxed{r = \frac{\text{variation intergroupe}}{\text{variation générale}}}$.

On a évidemment r compris entre 0 et 1 et r prend la valeur 1 si et seulement si la variation intragroupe est nulle, c'est à dire si pour tout i on a $V_i = 0$: chaque groupe est alors constitué d'individus tous identiques.
C'est d'ailleurs ce qui se passe dans l'exemple 3 où les variations des 4 groupes sont nulles (Rappel : 9 et 3 sont remplacés par 0, et 2 par 1).
Vérifions la formule ci-dessus dans le cas de l'exemple 2.
On a $\overline{x}$ = (1/3,5/6,2/3,1).
La variation générale est
4/9+25/36+1/9+4/9+1/36+4/9+1/9+1/36+4/9+ 3(1/9+1/36+ 1/9) = 7/2
La variation de G_1 et la variation de G_2 sont nulles .
La moyenne de G_3 est $\overline{x}_{G_3}$ = (0,1,3/4,1).
La variation de G_3 est $(3/4)^2+(1/4)^2 \times 3 = 3/4$.
La variation intergroupe est donc 7/2- (0+0+3/4) = 11/4.
d'où r = (11/4)/(7/2) = 11/14 = 0,7857....

Calcul des α_{ij}
a) On calcule m_j = moyenne de réussite à l'item j (moyenne prise sur tous les élèves de E)
b) On calcule γ_{ij} =(moyenne de réussite du groupe i à l'item j)- m_j .

c) $\alpha_{ij} = \text{sign}(\gamma_{ij}) \times \frac{\gamma_{ij}^2}{\sum_{j=1}^{4} \gamma_{ij}^2} \times 100$.

avec sign(ε) = 1 si ε>0, -1 si ε<0, 0 si ε = 0.
Remarquons que $\overline{x} = (m_1, m_2, m_3, m_4)$
L'aspect c) explique évidemment le fait que, pour tout i, $\sum_j |\alpha_{ij}| = 100$
l'aspect b) entraîne que $\alpha_{ij} > 0$ a pour signification : la moyenne de réussite du groupe i à l'item j est supérieure à la moyenne générale.

Vérifions ces formules sur l'exemple 3.
On a $m_1 = 2/3$, $m_2 = 1/3$, $m_3 = 5/6$, $m_4 = 1$ (l'item n°1 étant le 25A, etc)

Le tableau des γ_{ij} est donc

1/3	2/3	-5/6	0
-2/3	2/3	1/6	0
-2/3	-1/3	1/6	0
1/3	-1/3	1/6	0

On en déduit le tableau des $\frac{\alpha_{ij}}{100}$

4/45	16/45	-25/45	0
-16/33	16/33	1/33	0
-16/21	-4/21	1/21	0
4/9	-4/9	1/9	0

d'où le tableau des α_{ij} à 1/100 près

8,88	35,55	-55,55	0
-48,48	48,48	3,03	0
-76,19	19,04	4,76	0
44,44	-44,44	11,11	0

Terminons cette analyse par la remarque suivante : dès que la partition est constituée de 2 groupes, on a pour tout j : $\alpha_{1j} = -\alpha_{2j}$

Voici la preuve de ce résultat : notons n_i l'effectif du groupe i et $\Sigma_{i,j}$ le nombre d'élèves du groupe i ayant réussi l'item j.
On peut alors écrire :

$$\gamma_{1j} = \frac{\Sigma_{1,j}}{n_1} - \frac{\Sigma_{1,j} + \Sigma_{2,j}}{n_1 + n_2} \quad \text{et} \quad \gamma_{2j} = \frac{\Sigma_{2,j}}{n_2} - \frac{\Sigma_{1,j} + \Sigma_{2,j}}{n_1 + n_2}$$

d'où $\gamma_{1j} = -\frac{n_2}{n_1}\gamma_{2j}$.
Les séquences $(\gamma_{1j})_{j=1,4}$ et $(\gamma_{2j})_{j=1,4}$ sont donc proportionnelles, de même lorsqu'on passe au carré.

Donc les séquences $(|\alpha_{1j}|)_{j=1,4}$ et $(|\alpha_{2j}|)_{j=1,4}$ sont proportionnelles ; or chacune a pour somme 100,
donc $|\alpha_{1j}| = |\alpha_{2j}|$ pour tout j, d'où
si $\gamma_{1j} = \gamma_{2j} = 0$ on a $\alpha_{1j} = -\alpha_{2j} = 0$
si $\gamma_{1j} \neq 0$, sign (γ_{1j}) = - sign (γ_{2j}), donc $\alpha_{1j} = -\alpha_{2j}$.