

Questions-Réponses : Correspondances multiples

Résumé

Cette fiche utilise l'enquête menée par M. Le Berre, J. Reynaud-Feurly et H. Zucchetto (IREM de Lyon) pour décrire l'usage des modules utilisant des variables qualitatives. Cette deuxième partie illustre les manipulations associées à l'analyse des correspondances multiples d'un tableau faiblement structuré. On souligne les précautions à prendre dans la manipulation des cartes factorielles de modalités.

Plan

1 — Implantation des données.....	2
1.1 — Vérifier les codes.....	3
1.2 — Lire les données.....	4
1.3 — Tableaux de Burt et tests χ^2	8
2 — L'analyse des correspondances multiples	9
2.1 — Axe factoriel	11
2.2 — Cartes factorielles	16
3 — Tableaux croisés	20
3.1 — Tests univariés.....	21
3.2 — Test multivarié.....	22

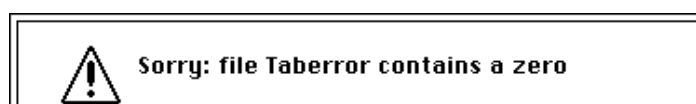
D. Chessel et J. Thioulouse

	A	B	C
1	1		2
2	40		1
3	2		2
4	1		1
5	2		2
6	0		1
7	0		2
8	2		2
9	2		2
10	2		1
11	2		1
12	2		2
13	40		1
14	1		2
15	2		2
16	1		2
17	1		1
18	1		1

Si on essaie de lire Taberror avec l'option ReadCateg du module CategVar :



on obtient :



Lancer l'option :



On obtient :

```
File: Taberror - Column: 1
Zero is not a valid number for category
Some categories are missing
old = 0 -> new = 1 || old = 1 -> new = 2 ||
old = 2 -> new = 3 || old = 40 -> new = 4 ||
-----
File: Taberror - Column: 2
No modification
etc...
```

Le module renumérote dans l'ordre naturel les numéros de modalités à partir de 1 en éditant la liste des changements de code et assure les fonctions normales de l'option ReadCateg.

1.2 — Lire les données

On obtient le même résultat qu'en utilisant :



Le listing permet de reporter les codes des questions et des modalités de réponse :

1 - Combien d'actions de formation suivies ?

[1] Cat. : 1 Num: 13 Freq.: 0.149 aucun
[2] Cat. : 2 Num: 36 Freq.: 0.414 2 5
[3] Cat. : 3 Num: 24 Freq.: 0.276 3 5< 10
[4] Cat. : 4 Num: 14 Freq.: 0.161 4 >10

2 - Lectures tendance scientifique

[5] Cat. : 1 Num: 43 Freq.: 0.494 faible
[6] Cat. : 2 Num: 44 Freq.: 0.506 fort

3 - Lectures tendance pédagogique

[7] Cat. : 1 Num: 45 Freq.: 0.517 faible
[8] Cat. : 2 Num: 42 Freq.: 0.483 fort

4 - Lectures tendance informatique

[9] Cat. : 1 Num: 51 Freq.: 0.586 faible
[10] Cat. : 2 Num: 36 Freq.: 0.414 fort

5 - Faites-vous des mathématiques par plaisir

[11] Cat. : 1 Num: 26 Freq.: 0.299 non
[12] Cat. : 2 Num: 36 Freq.: 0.414 oui : jeux
[13] Cat. : 3 Num: 10 Freq.: 0.115 oui : études
[14] Cat. : 4 Num: 15 Freq.: 0.172 oui: jeux + études

6 - Souhaits lectures personnelles

[15] Cat. : 1 Num: 39 Freq.: 0.448 non
[16] Cat. : 2 Num: 48 Freq.: 0.552 oui

7 - Souhaits stages

[17] Cat. : 1 Num: 42 Freq.: 0.483 non
[18] Cat. : 2 Num: 45 Freq.: 0.517 oui

8 - Souhaits conférences

[19] Cat. : 1 Num: 57 Freq.: 0.655 non
[20] Cat. : 2 Num: 30 Freq.: 0.345 oui

9 - Souhaits groupe IREM

[21] Cat. : 1 Num: 76 Freq.: 0.874 non
[22] Cat. : 2 Num: 11 Freq.: 0.126 oui

10 - Souhaits reprise études

[23] Cat. : 1 Num: 59 Freq.: 0.678 non
[24] Cat. : 2 Num: 28 Freq.: 0.322 oui

11 - Pratique activités hors-programme

[25] Cat. : 1 Num: 58 Freq.: 0.667 non
[26] Cat. : 2 Num: 29 Freq.: 0.333 oui

12 - Pratique activités type PAE

[27] Cat. : 1 Num: 81 Freq.: 0.931 non
[28] Cat. : 2 Num: 6 Freq.: 0.069 oui

13 - Pratique activités inter-disciplinaires

[29] Cat. : 1 Num: 65 Freq.: 0.747 non
[30] Cat. : 2 Num: 22 Freq.: 0.253 oui

14 - Aimez-vous enseigner dans une classe non scientifique ?

[31] Cat. : 1 Num: 22 Freq.: 0.253 non
[32] Cat. : 2 Num: 53 Freq.: 0.609 oui
[33] Cat. : 3 Num: 12 Freq.: 0.138 Non-réponse

15 - Une culture mathématique pour un enseignant c'est :

[33] Cat. : 1 Num: 30 Freq.: 0.345 D'abord un état d'esprit, ensuite avoir des connaissances
[34] Cat. : 2 Num: 29 Freq.: 0.333 D'abord un état d'esprit, ensuite connaître l'histoire
[35] Cat. : 3 Num: 20 Freq.: 0.230 Pour l'essentiel, avoir des connaissances
[36] Cat. : 4 Num: 8 Freq.: 0.092 Connaître l'évolution et les maths d'aujourd'hui

16 - Un minimum de connaissances informatiques est un élément de culture mathématique

[37] Cat. : 1 Num: 15 Freq.: 0.172 pas d'accord
[38] Cat. : 2 Num: 32 Freq.: 0.368 plutôt pas d'accord
[39] Cat. : 3 Num: 33 Freq.: 0.379 tout à fait d'accord
Missing data Num: 7 Freq.: 0.0805

17 - La culture mathématique est la connaissance d'un langage spécifique

[40] Cat. : 1 Num: 10 Freq.: 0.115 Non, pas d'accord
[41] Cat. : 2 Num: 25 Freq.: 0.287 plutôt pas d'accord
[42] Cat. : 3 Num: 44 Freq.: 0.506 D'accord
Missing data Num: 8 Freq.: 0.092

18 - Les idées mathématiques dépendent peu de l'époque

[43] Cat. : 1 Num: 14 Freq.: 0.161 Non, pas d'accord
[44] Cat. : 2 Num: 26 Freq.: 0.299 plutôt pas d'accord
[45] Cat. : 3 Num: 24 Freq.: 0.276 plutôt d'accord
[46] Cat. : 4 Num: 7 Freq.: 0.0805 tout à fait d'accord
Missing data Num: 16 Freq.: 0.184

19 - Je suis plus en accord avec

[47] Cat. : 1 Num: 65 Freq.: 0.747 pas de niveau d'arrêt pour l'enseignement des maths
[48] Cat. : 2 Num: 18 Freq.: 0.207 arrêt subordonné à l'acquisition d'un niveau minimal
Missing data Num: 4 Freq.: 0.046

20 - L'objectif dans une série scientifique est :

[49] Cat. : 1 Num: 15 Freq.: 0.172 donner un regard critique
[50] Cat. : 2 Num: 17 Freq.: 0.195 introduire des concepts spécifiques
[51] Cat. : 3 Num: 41 Freq.: 0.471 préparer aux cursus ultérieurs
Missing data Num: 14 Freq.: 0.161

21 - L'objectif dans une série littéraire est :

[52] Cat. : 1 Num: 69 Freq.: 0.793 donner un regard critique
[53] Cat. : 2 Num: 11 Freq.: 0.126 introduire des concepts spécifiques
[54] Cat. : 3 Num: 7 Freq.: 0.0805 autre

22 - L'objectif dans une série économique et sociale est :

[55] Cat. : 1 Num: 35 Freq.: 0.402 donner un regard critique
[56] Cat. : 2 Num: 23 Freq.: 0.264 introduire des concepts spécifiques
[57] Cat. : 3 Num: 18 Freq.: 0.207 préparer aux cursus ultérieurs
Missing data Num: 11 Freq.: 0.126

23 - La théorie des nombres est un secteur de recherche dynamique

[58] Cat. : 1 Num: 58 Freq.: 0.667 peu convaincant
[59] Cat. : 2 Num: 27 Freq.: 0.31 convaincant
Missing data Num: 2 Freq.: 0.023

24 - L'arithmétique est une source de problèmes motivants

[60] Cat. : 1 Num: 19 Freq.: 0.218 peu convaincant
[61] Cat. : 2 Num: 40 Freq.: 0.460 assez convaincant
[62] Cat. : 3 Num: 26 Freq.: 0.299 très convaincant
Missing data Num: 2 Freq.: 0.023

25 - Les élèves n'ont pas des connaissances suffisantes sur les nombres

[63] Cat. : 1 Num: 14 Freq.: 0.161 peu convaincant
[64] Cat. : 2 Num: 35 Freq.: 0.402 assez convaincant
[65] Cat. : 3 Num: 37 Freq.: 0.425 très convaincant
Missing data Num: 1 Freq.: 0.0115

26 - L'arithmétique est un champ riche et diversifié

[66] Cat. : 1 Num: 10 Freq.: 0.115 peu convaincant
[67] Cat. : 2 Num: 45 Freq.: 0.517 assez convaincant
[68] Cat. : 3 Num: 30 Freq.: 0.345 très convaincant
Missing data Num: 2 Freq.: 0.023

27 - Favoriser les projets professionnels est un objectif majeur

[69] Cat. : 1 Num: 38 Freq.: 0.437 non
[70] Cat. : 2 Num: 49 Freq.: 0.563 oui

28 - Permettre des actions de remédiation est un objectif majeur

[71] Cat. : 1 Num: 55 Freq.: 0.632 non
[72] Cat. : 2 Num: 32 Freq.: 0.368 oui

29 - Aménager des passerelles entre sections est un objectif majeur

[73] Cat. : 1 Num: 49 Freq.: 0.563 non
[74] Cat. : 2 Num: 38 Freq.: 0.437 oui

30 - Permettre d'aller plus loin en maths est un objectif majeur

[75] Cat. : 1 Num: 17 Freq.: 0.195 non
[76] Cat. : 2 Num: 70 Freq.: 0.805 oui

31 - Permettre de traiter des thèmes hors-programme est un objectif majeur

[77] Cat. : 1 Num: 41 Freq.: 0.471 non
[78] Cat. : 2 Num: 46 Freq.: 0.529 oui

32 - Accepteriez vous un enseignement facultatif

[79] Cat. : 1 Num: 21 Freq.: 0.241 non
[80] Cat. : 2 Num: 63 Freq.: 0.724 oui
Missing data Num: 3 Freq.: 0.0345

33 - Accepteriez vous de prendre en charge un enseignement sur un sujet mathématique que vous connaissez mal

[81] Cat. : 1 Num: 10 Freq.: 0.115 oui
[82] Cat. : 2 Num: 23 Freq.: 0.264 oui, si intérêt
[83] Cat. : 3 Num: 34 Freq.: 0.391 oui, si formation
[84] Cat. : 4 Num: 8 Freq.: 0.092 non
[85] Cat. : 5 Num: 10 Freq.: 0.115 oui, si intérêt et formation
Missing data Num: 2 Freq.: 0.023

34 - L'aide de documents-élèves directement utilisables est

[86] Cat. : 1 Num: 27 Freq.: 0.31 indispensable
[87] Cat. : 2 Num: 41 Freq.: 0.471 utile
[88] Cat. : 3 Num: 16 Freq.: 0.184 pas déterminant
Missing data Num: 3 Freq.: 0.0345

35 - L'aide de comptes-rendus d'activité avec les élèves est

[89] Cat. : 1 Num: 8 Freq.: 0.092 indispensable
[90] Cat. : 2 Num: 53 Freq.: 0.609 utile
[91] Cat. : 3 Num: 24 Freq.: 0.276 pas déterminant
Missing data Num: 2 Freq.: 0.023

36 - L'aide du travail en équipe disciplinaire est

[92] Cat. : 1 Num: 30 Freq.: 0.345 indispensable
[93] Cat. : 2 Num: 43 Freq.: 0.494 utile
[94] Cat. : 3 Num: 11 Freq.: 0.126 pas déterminant
Missing data Num: 3 Freq.: 0.0345

37 - L'aide du travail en équipe interdisciplinaire est

[95] Cat. : 1 Num: 45 Freq.: 0.517 utile
[96] Cat. : 2 Num: 34 Freq.: 0.391 pas déterminant
Missing data Num: 8 Freq.: 0.092

38 - L'aide de stage d'établissement est

[97] Cat. : 1 Num: 56 Freq.: 0.644 utile
[98] Cat. : 2 Num: 24 Freq.: 0.276 pas déterminant
Missing data Num: 7 Freq.: 0.0805

39 - L'aide de formation sur l'histoire des maths est

[99] Cat. : 1 Num: 11 Freq.: 0.126 indispensable
[100] Cat. : 2 Num: 39 Freq.: 0.448 utile
[101] Cat. : 3 Num: 29 Freq.: 0.333 pas déterminant
Missing data Num: 8 Freq.: 0.092

40 - L'aide d'un stage d'approfondissement thématique est

[102] Cat. : 1 Num: 8 Freq.: 0.092 indispensable
[103] Cat. : 2 Num: 56 Freq.: 0.644 utile

[104] Cat. : 3 Num: 17 Freq.: 0.195 pas déterminant
Missing data Num: 6 Freq.: 0.069

41 - L'aide d'un stage de mise à jour des connaissances est

[105] Cat. : 1 Num: 22 Freq.: 0.253 indispensable
[106] Cat. : 2 Num: 45 Freq.: 0.517 utile
[107] Cat. : 3 Num: 15 Freq.: 0.172 pas déterminant
Missing data Num: 5 Freq.: 0.0575

42 - L'aide d'une formation à l'interdisciplinarité est

[108] Cat. : 1 Num: 9 Freq.: 0.103 indispensable
[109] Cat. : 2 Num: 39 Freq.: 0.448 utile
[110] Cat. : 3 Num: 32 Freq.: 0.368 pas déterminant
Missing data Num: 7 Freq.: 0.0805

Auxiliary binary output file TabModa: Indicator vector of modalities

It contains variable number for each modality

It has 110 rows (modalities) and one column

Auxiliary ASCII output file Tab.123: labels (two characters) for 110 modalities

It contains one label for each modality

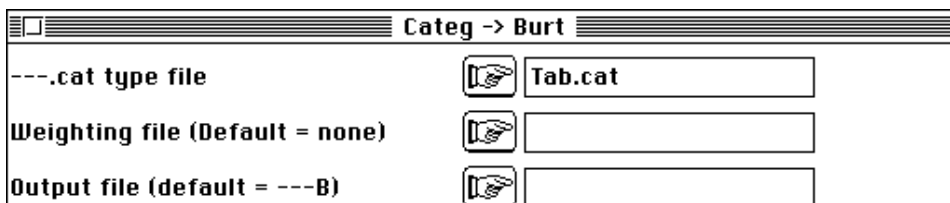
It has 110 rows (modalities) and labels 1a,1b, ..., 2a, 2b, ...

Variable number 1,2, ..., A, ..., Z,+, Modality number a,b, ..., z,+

Le fichier Tab.123 est un fichier d'étiquettes des modalités (110) qui contient 1a, 1b, 1c, 1d, 2a, 2b, 3a, 3b, 4a, 4b, 5a, 5b, ..., +a, +b, +c. Chaque étiquette de deux caractères désigne d'abord la variable (1, 2, ..., Y, Z, + au delà de 36) puis la modalité (a, b, ..., + au delà de 26). Rien n'empêche, suivant les cas, de modifier à l'éditeur de texte ce fichier pour le rendre plus efficace. Le module ne signale pas qu'il a créé un fichier TabModa qui contient comme seule variable qualitative le numéro de la variable à laquelle appartient chacune des modalités (110 lignes, 1 colonne, valeurs comprises entre 1 et 42). Tab et TabModa étant des fichiers de variables qualitatives vérifiés, ils sont accompagnés des fichiers Tab.cat et TabModa.cat par lesquels tous les modules accèdent à ce type de fichier. Cela évite des erreurs dans le passage des noms de fichiers. TabModa.cat permettra, en particulier de trier les modalités par variables dans les dessins multifenêtrés. TabModa.123 étiquette les classes de modalités, donc les variables. Rien n'empêche de remplacer le caractère utilisé (1, 2, ..., Y, Z, + au delà de 36) par un code explicite (Formation, Lec_Scient, Lec_Pédago, Lec_Info, Plaisir, >Lec_Perso, >Stages, >Conf, >IREM, >Etudes, >Hors_Prog, PAE, Inter_Disc, Non_Scient, *Culture, *Info, *Langage, *Epoque, Arrêt, Obj_Scient, Obj_Litt, Obj_Eco, +Nombre, +Arith, +Eleves, +Motivant, +Diversifié, •Prof, •Remédiation, •Passerelles, •maths, •Hors_Prog, =facultatif, =sujet, /Doc, /CR, /Equipe, /Stage, /Histoire, /Thématique, /Connais., /Interdisc).

1.3 — Tableaux de Burt et tests Khi2

Une bonne précaution est de s'assurer que les variables qualitatives présentent des corrélations significatives. Utiliser :



Le fichier de pondération pourrait indiquer que chaque ligne du tableau correspond à plusieurs individus ayant fait les mêmes réponses. On obtient :



Error: too many characters in listing window

Le listing est trop long pour la fenêtre d'édition :



Do you want to save listing to file
I-CategVar/12/23/50 ?

Il est directement conservé dans le fichier de sortie :

File TabB contains Burt matrix from file Tab
It has 110 rows (categories) and 110 columns (categories)

File TabB.blo contains the number of modalities for each variable)
It has 42 rows (categories) and 1 column

Khi2 tests on Burt's crossed table TabB

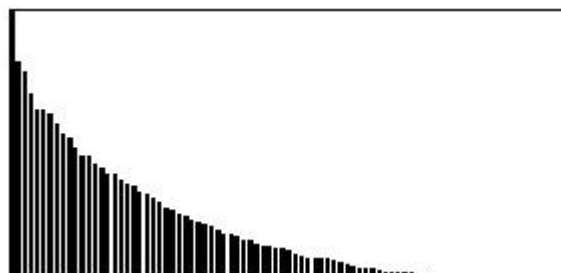
Var 2 versus var 1:	Khi 2 = 1.4624e+00	ddl = 3	Proba = 0.695104
Var 3 versus var 1:	Khi 2 = 6.9341e+00	ddl = 3	Proba = 0.072707
Var 3 versus var 2:	Khi 2 = 8.4118e+00	ddl = 1	Proba = 0.003857
etc...			
Var 42 versus var 37:	Khi 2 = 2.8453e+01	ddl = 2	Proba = 0.000002
Var 42 versus var 38:	Khi 2 = 5.0072e+00	ddl = 2	Proba = 0.079727
Var 42 versus var 39:	Khi 2 = 5.6797e+00	ddl = 4	Proba = 0.223126
etc...			

On trouve le résultat des $42 \times 41/2$ (861) tests croisés. Ceux qui sont très significatifs sont rares et la redondance entre réponses aux diverses questions est faible en moyenne au niveau d'un couple de variables. On retrouve ici l'absence de types d'opinion tranchés et exprimés sur plusieurs variables.

2 — L'analyse des correspondances multiples

L'analyse des correspondances multiples (ACM) est l'équivalent sur variables qualitatives de l'ACP sur variables quantitatives. On l'exécute simplement par le module MCA :

Multiple Correspondence Analysis	
.cat type file	 Tab.cat
Option: Row weighting file	 <input type="text"/>



cm/MCA: Multiple correspondence analysis

Uniform row weights

File Tab.cmpl contains the row weights
It has 87 rows and 1 column

Vérifier que les valeurs présentes dans ce fichier sont toutes égales à 0.0115 (1/87).

File Tab.cmpc contains the column weights (1/V)*DM
It has 110 rows and 1 column

Marginal distributions by variable:

Variable number 1 has 4 categories

[1] Category: 1 Weight: 0.149
[2] Category: 2 Weight: 0.414
[3] Category: 3 Weight: 0.276
[4] Category: 4 Weight: 0.161

...

[110] Category: 3 Weight: 0.4

On retrouve le listing de ReadCateg car les pondérations utilisées sont uniformes. Noter que les quatre premières valeurs du fichier Tab.cmpc sont 0.003558, 0.009852, 0.006568 et 0.003831, soit les fréquences des modalités divisées par le nombre total de variables ($0.003558 * 42 = 0.149$). Le triplet de l'analyse comporte un tableau individus-modalités, alors qu'on part d'un tableau individus-variables (codage par numéro de modalité).

File Tab.cmta contains the table processed by MCA
It has 87 rows and 110 columns (categories)

On n'utilise pas en général ce tableau. Noter cependant ce qu'il contient. Les quatre premières valeurs de la ligne 1 sont -1, 1.4167, -1 et -1. L'individu 1 porte pour la variable 1 la modalité 2. La variable 1 a 4 modalités. En codage disjonctif complet, le fait que l'individu 1 porte la modalité 2 est transcrit 0, 1, 0, 0. Cette valeur 0 ou 1 est divisée par le poids de la modalité (0 reste 0) et on retranche 1. Les modalités non portées donnent donc -1 et la modalité portée donne $1/0.414 - 1 = 1.4167$. Toujours sur la ligne 1 et pour les colonnes 37 à 39, on trouve $6.6613e-16$, $6.6613e-16$ et $6.6613e-16$, soit 0 trois fois. L'individu 1 est codé valeur manquante sur la variable 16 qui a 3 modalités. (colonne 37 à 39). Une donnée manquante donne 0 pour chaque modalité et ne participe pas à la définition de l'inertie.

Total inertia: 1.55063

Num.	Eigenval.	R. Iner.	R. Sum		Num.	Eigenval.	R. Iner.	R. Sum	
01	+9.5827E-02	+0.0618	+0.0618		02	+7.7635E-02	+0.0501	+0.1119	
03	+7.3872E-02	+0.0476	+0.1595		04	+6.5991E-02	+0.0426	+0.2021	
...									
39	+1.1397E-02	+0.0074	+0.9142		40	+1.1071E-02	+0.0071	+0.9213	

L'inertie totale théorique vaut $m/v - 1$ (m modalités, v variables) soit $110/42 - 1$, soit 1.6190. On a trouvé 1.55 et on peut dire que les données manquantes ont enlevé 4 % de l'inertie attendue, ce qui est tolérable. Il ne faudrait pas croire que les données manquantes sont justifiées au delà de 10% (en gros). Seules les 40 premières valeurs propres sont éditées.

File Tab.cmvp contains the eigenvalues and relative inertia for each axis
 --- It has 110 rows and 2 columns

File Tab.cmco contains the column scores
 --- It has 110 rows and 4 columns
 File :Tab.cmco ----- Minimum/Maximum -----
 Col.: 1 Mini = -1.40255 Maxi = 1.23103
 ...
 Col.: 4 Mini = -1.12181 Maxi = 0.665233

File Tab.cml i contains the row scores
 --- It has 87 rows and 4 columns
 File :Tab.cml i ----- Minimum/Maximum -----
 Col.: 1 Mini = -0.678159 Maxi = 0.644918
 ...
 Col.: 4 Mini = -0.731137 Maxi = 0.455431

On obtient un fichier de coordonnées des lignes (individus) et des colonnes (modalités). Les coordonnées des colonnes sont centrées par variables pour la pondération des modalités. La fin du listing est particulière à ce type d'analyse.

```
Variable : 1
> Categ= 1 Weight= 0.149 -0.265 -0.928 0.508 0.147
> Categ= 2 Weight= 0.414 -0.564 -0.128 -0.220 -0.038
> Categ= 3 Weight= 0.276 0.445 0.599 -0.031 0.292
> Categ= 4 Weight= 0.161 0.934 0.165 0.147 -0.540
-----> r= 0.337 0.239 0.062 0.074
```

Pour s'en servir il faut connaître le principe de l'analyse.

2.1 — Axe factoriel

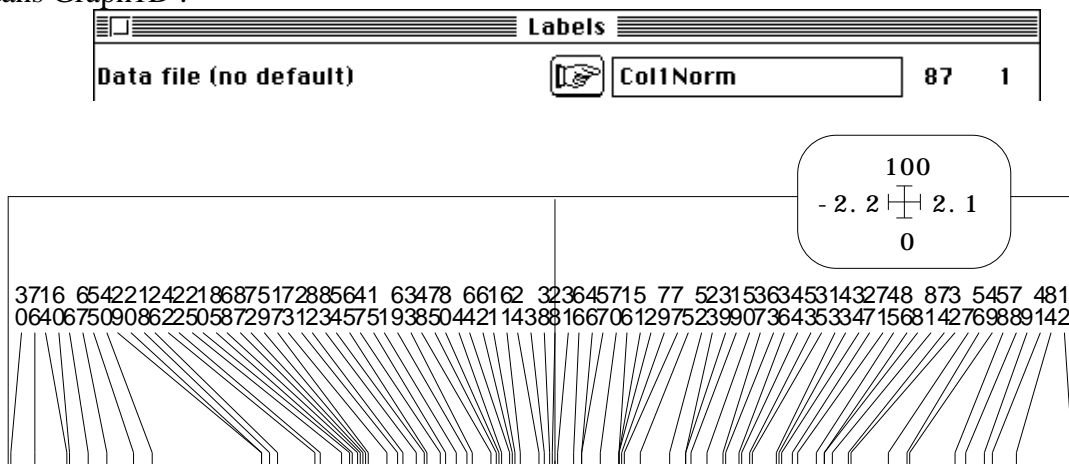
Considérons la première coordonnées factorielle des individus, soit la colonne 1 de Tab.cml i. Mettons la dans un fichier Col1 (87 - 1) :

Row-Col Selection	
Input file	Tab.cml i 87 4
Selection of rows (default = all)	<input type="text"/>
Selection of columns (default = all)	1
Output file	Col1

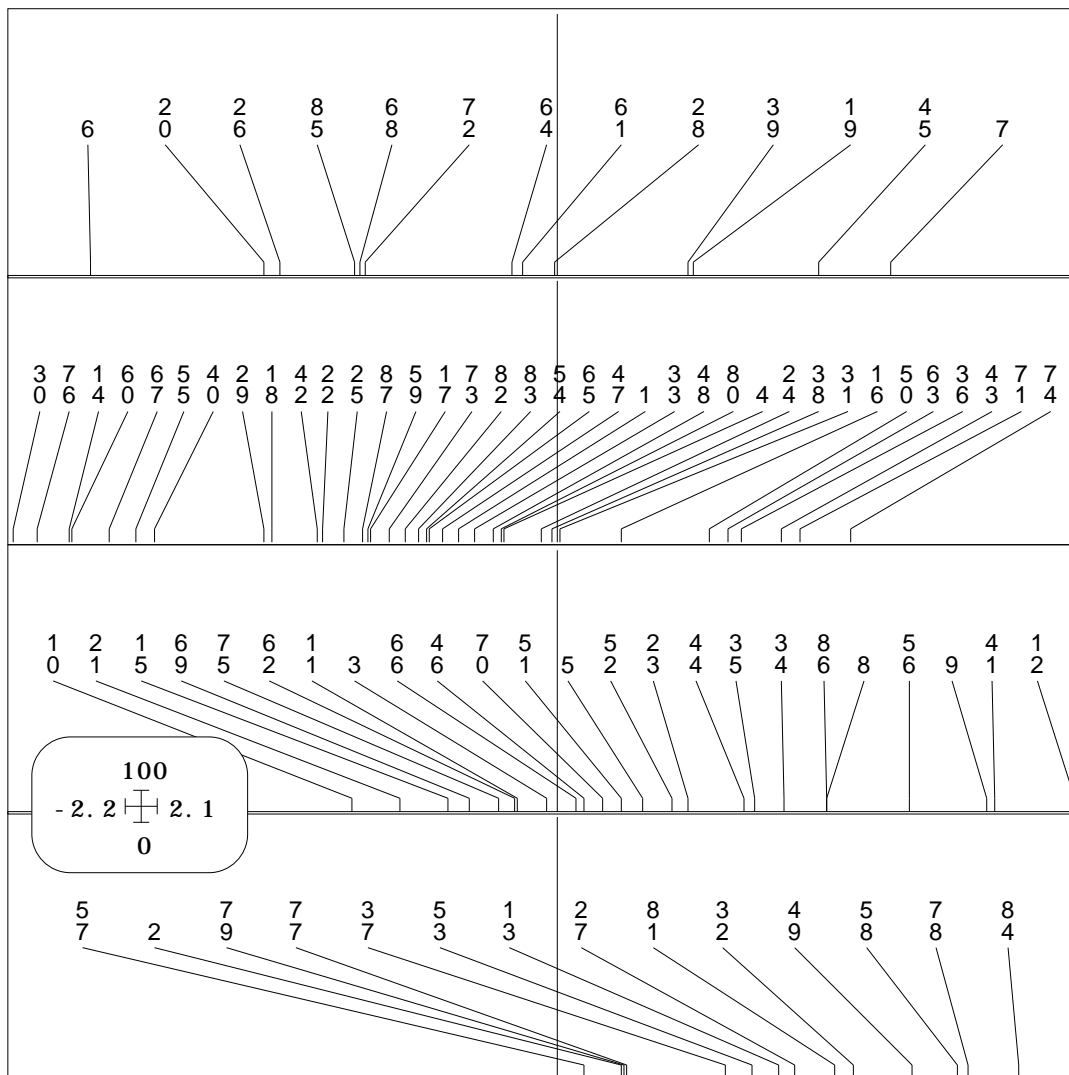
Calculer sa variance (MatAlg, option Diagonal Inner Product) et observer qu'elle vaut 0.09583, soit exactement la première valeur propre. Elle est de plus centrée. La normaliser dans Col1Norm :

Centring	
Input file	Col1 87 1
Option: file for row weighting	<input type="text"/>
Option for H matrix (no default)	3
Output file	Col1Norm


Ce code numérique positionne les 87 points sur un axe. Représenter cette ordination dans Graph1D :




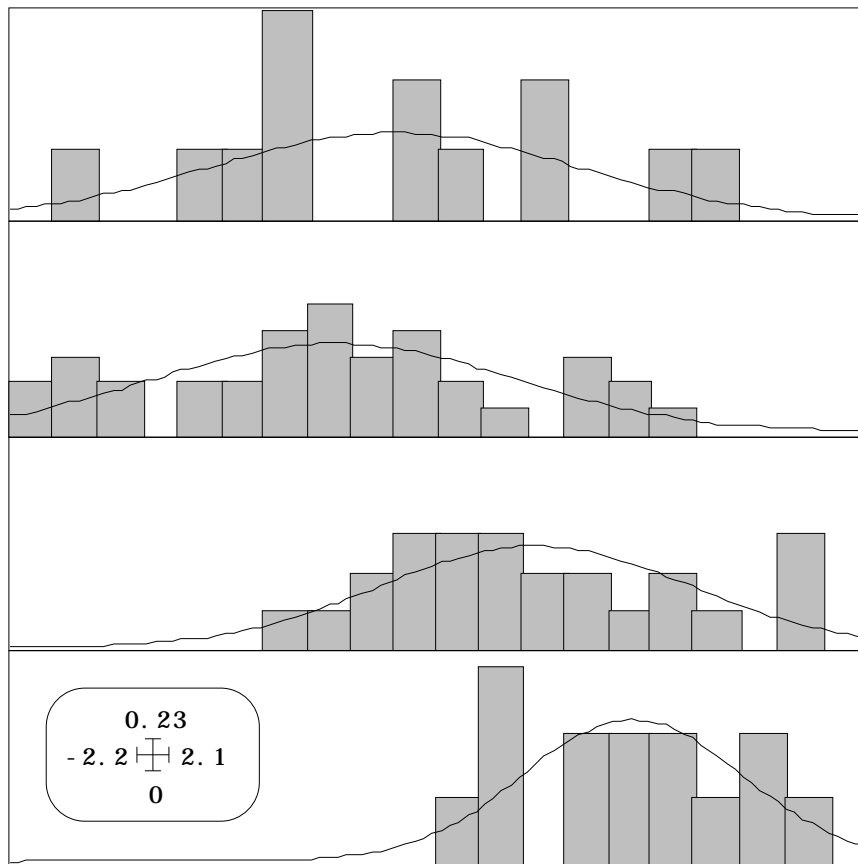
A gauche l'individu 30 et à droite l'individu 12. Séparer les 87 individus par paquet suivant la modalité portés par la première variable de Tab :



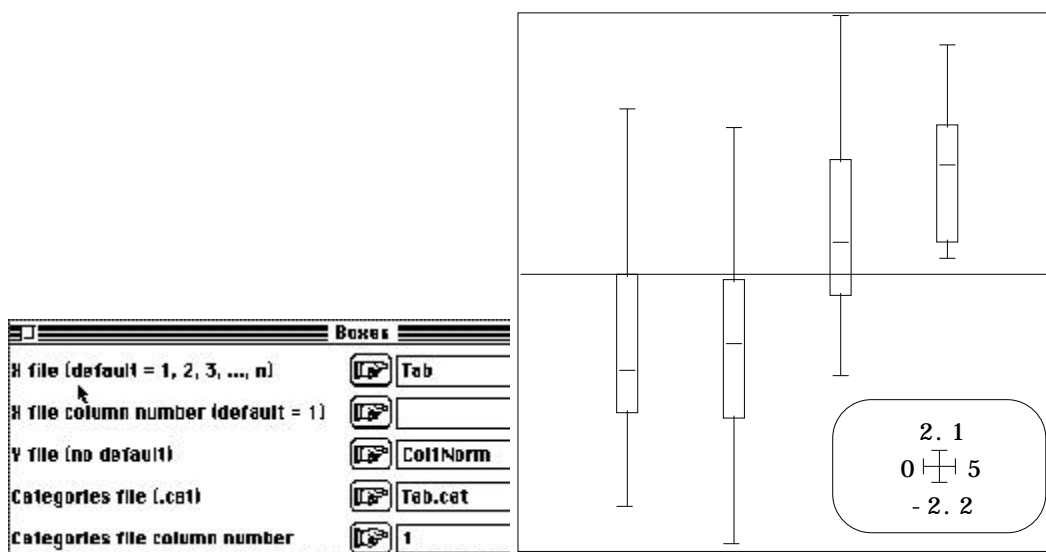
On a utilisé :

Row selection file (.cat): 

Selection col. number: 



Les individus 4 (plus de 10 actions de formation suivies) sont manifestement à droite. Utiliser l'option Histograms du même module (ci-dessus). On peut faire cela 42 fois mais c'est fastidieux. On peut approcher la même réalité par CurveClass (option Boxes) :



On peut se contenter pour résumer cette opération de la position moyenne sur l'axe des porteurs de chacune des quatre modalités et du pourcentage de variance expliquée par la partition en 4 paquets. Calculer les moyennes

CateRowSum-Mean	
Input file	Col1Norm 87 1
.cat file	Tab.cat
Column number for selection	1
Output file	A1
Option: sum (1) or mean (2)	2

Input file: A1 Row: 4 Col: 1	Input file: Tab.cmco Row: 110 Col: 4
1 -0.2653	1 -0.2653 -0.9278 0.5081 0.1467
2 -0.5640	2 -0.5640 -0.1284 -0.2199 -0.0377
3 0.4449	3 0.4449 0.5992 -0.0309 0.2918
4 0.9340	4 0.9340 0.1645 0.1465 -0.5395
	5 -0.3999 0.2423 0.3601 0.1599

```

Variable : 1
> Categ= 1 Weight= 0.149 -0.265 -0.928 0.508 0.147
> Categ= 2 Weight= 0.414 -0.564 -0.128 -0.220 -0.038
> Categ= 3 Weight= 0.276 0.445 0.599 -0.031 0.292
> Categ= 4 Weight= 0.161 0.934 0.165 0.147 -0.540
-----> r= 0.337 0.239 0.062 0.074

```

Les coordonnées des colonnes de l'analyse sont les moyennes conditionnelles par paquet d'individus portant la même modalité (moyenne sur les coordonnées des lignes normalisées) et sont éditées sur le listing de l'analyse. La ligne r donne les pourcentages de variance expliquée. 33.7 % de la variance des positions des individus est expliquée par la partition des réponses à la première question. D'où l'intérêt de la lecture conjointe des positions des modalités de toutes les variables (module GraphIDClass) :

Gauss curves	
Data file (no default)	Col1Norm 87 1
Number of classes (default = 10)	
Categories file (.cat)	Tab.cat
Variable label file (optional)	
Draw class labels (yes = 1, no = 2)	1

Ce qui est essentiel, c'est le caractère d'optimalité de la figure obtenue (figure 1). Dans chaque fenêtre, implicitement les individus sont placés sur le même axe dans la même position et chaque variable qualitative définit une partition sur cette unique variable quantitative. Les mêmes individus sont répartis par paquet, chaque groupe donnant une moyenne et une courbe de Gauss résumé. Chaque variable définit un rapport de corrélation et c'est la moyenne des rapports de corrélation qui est optimisée.

On lit simplement la figure. A droite on trouve ceux qui ont suivi de nombreuses formations (variable 1, modalité 4), à gauche ceux qui pensent que l'aide du travail en équipe disciplinaire n'est pas déterminant (variable 36, modalité 3). A droite on rencontre les tendances à forte lecture scientifique (2), à forte lecture pédagogique (3), à faire des mathématiques par plaisir (jeux+études, 5), les souhaits de stages (7), de conférences (8), de formation sur l'histoire des maths (39), d'approfondissement thématique (40), de mises à jour des connaissances (41) et de formation à l'interdisciplinarité (42). A gauche on retrouve les tendances inverses, le refus d'enseigner des sujets mal connus (33, modalité 4), le refus de l'objectif "donner un

regard critique” en section littéraire (21, modalité 2-3), l’idée que les mathématiques dépendent peu de l’époque (18, modalité 4).

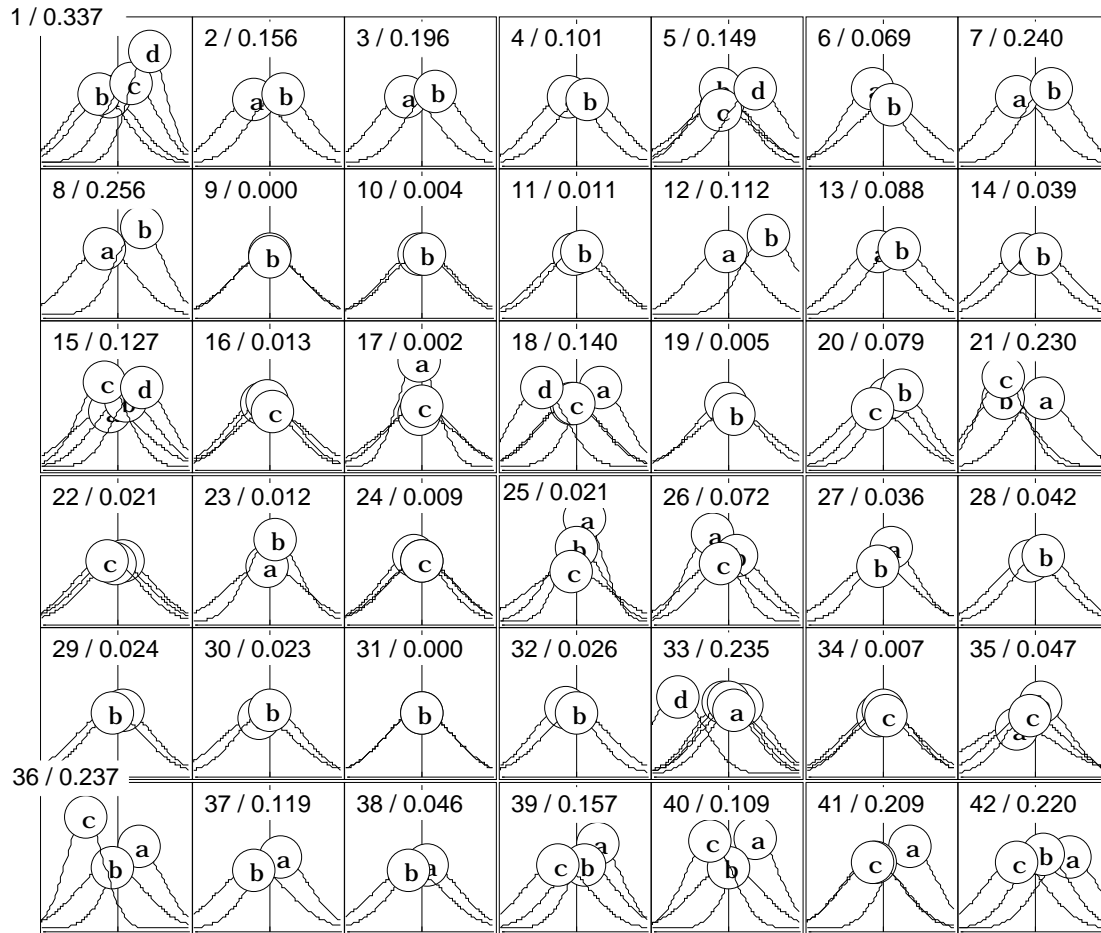


Figure 1 : En abscisse, coordonnée normalisée du facteur 1 de l’ACM de Tab. Représentation des moyennes par classes définies par les modalités des variables. Numéros des variables et rapports de corrélation. Les rapports dépassant 15% sont encadrés.

Formation	1-->	r=	0.337	0.239	0.062	0.074
Lec_Scient	2-->	r=	0.156	0.057	0.127	0.025
Lec_Pédago	3-->	r=	0.196	0.040	0.049	0.038
Lec_Info	4-->	r=	0.101	0.044	0.036	0.001
Plaisir	5-->	r=	0.149	0.257	0.142	0.144
>Lec_Perso	6-->	r=	0.069	0.085	0.001	0.007
>Stages	7-->	r=	0.240	0.090	0.029	0.002
>Conf	8-->	r=	0.256	0.010	0.000	0.005
>IREM	9-->	r=	0.000	0.127	0.016	0.106
>Etudes	10-->	r=	0.004	0.003	0.005	0.042
>Hors_Prog	11-->	r=	0.011	0.015	0.012	0.001
PAE	12-->	r=	0.112	0.001	0.072	0.007
Inter_Disc	13-->	r=	0.088	0.015	0.050	0.030
Non_Scient	14-->	r=	0.039	0.006	0.027	0.013
*Culture	15-->	r=	0.127	0.194	0.040	0.070
*Info	16-->	r=	0.013	0.287	0.136	0.067
*Langage	17-->	r=	0.002	0.191	0.004	0.016
*Epoque	18-->	r=	0.140	0.099	0.053	0.045
Arrêt	19-->	r=	0.005	0.015	0.038	0.153
Obj_Scient	20-->	r=	0.079	0.105	0.002	0.188
Obj_Litt	21-->	r=	0.230	0.225	0.061	0.015
Obj_Eco	22-->	r=	0.021	0.023	0.023	0.146
+Nombre	23-->	r=	0.012	0.216	0.019	0.079

+Arith	24-->	r=	0.009	0.186	0.184	0.023
+Elèves	25-->	r=	0.021	0.026	0.055	0.311
+Diversifié	26-->	r=	0.072	0.017	0.155	0.064
•Prof	27-->	r=	0.036	0.002	0.019	0.033
•Remédiation	28-->	r=	0.042	0.023	0.079	0.067
•Passerelles	29-->	r=	0.024	0.009	0.002	0.020
•maths	30-->	r=	0.023	0.029	0.098	0.004
•Hors_Prog	31-->	r=	0.000	0.005	0.253	0.061
=facultatif	32-->	r=	0.026	0.030	0.071	0.001
=sujet	33-->	r=	0.235	0.104	0.029	0.059
/Doc	34-->	r=	0.007	0.040	0.104	0.118
/CR	35-->	r=	0.047	0.212	0.089	0.072
/Equipe_disc.	36-->	r=	0.237	0.055	0.228	0.190
/Equipe_inter.	37-->	r=	0.119	0.052	0.008	0.219
/Stage	38-->	r=	0.046	0.025	0.098	0.017
/Histoire	39-->	r=	0.157	0.087	0.011	0.035
/Thématique	40-->	r=	0.109	0.004	0.208	0.054
/Connais.	41-->	r=	0.209	0.011	0.341	0.016
/Interdisc	42-->	r=	0.220	0.001	0.067	0.133






File Tab.cmrc contains the correlation ratios between the categorical variables and the factor scores
It has 42 rows and 4 columns

La liste des rapports de corrélation permet donc de sélectionner les variables participant à la définition des axes. Ces valeurs sont faibles en moyenne. Ces moyennes inter-variables sont simplement les valeurs propres (0.095, 0.078, 0.074, 0.060). Aucune relation franche entre les variables n'est signalée mais l'axe 1 est typiquement une ordination quantitative faite à partir de multiples correspondances dispersées entre plusieurs variables (tendances).

On retrouve l'axe d'investissement personnel qui intègre cependant des positions pédagogiques, l'absence de souhaits de renouvellement des connaissances (tant culturelle que thématique) accompagnant une vision disciplinaire assez stricte. Un grand nombre de variables sont indépendantes de cet ordination principale.

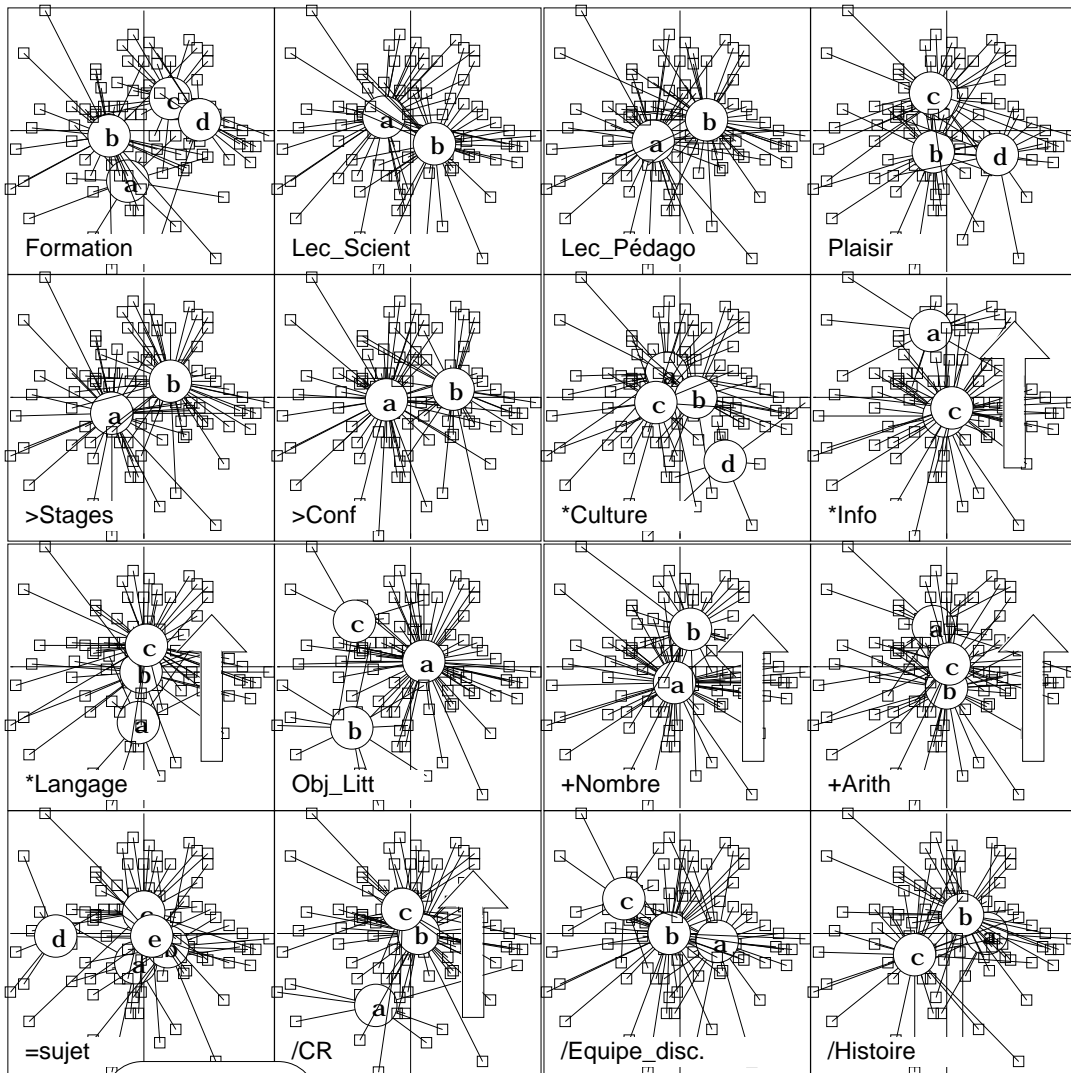
2.2 — Cartes factorielles

Le second facteur détermine un nouveau code numérique qui optimise la moyenne des rapports de corrélation, sous contrainte d'être non corrélé au précédent. La normalisation ne joue aucun rôle dans l'expression de ce principe et on peut directement partir des coordonnées ordinaires. On peut représenter les deux premiers axes (ScatterClass) :

Stars	
XY coordinates file	 Tab.cml 87 4
X-axis column number (default = 1)	 <input type="text"/>
Y-axis column number (default = 2)	 <input type="text"/>
Categories file (.cat)	 Tab.cat
Variable label file (or #)	 TabModa.123

Utiliser une sélection de variables :

Row & col. selection
 Col. selection: 15a17;21;23;24;33;35;36;39

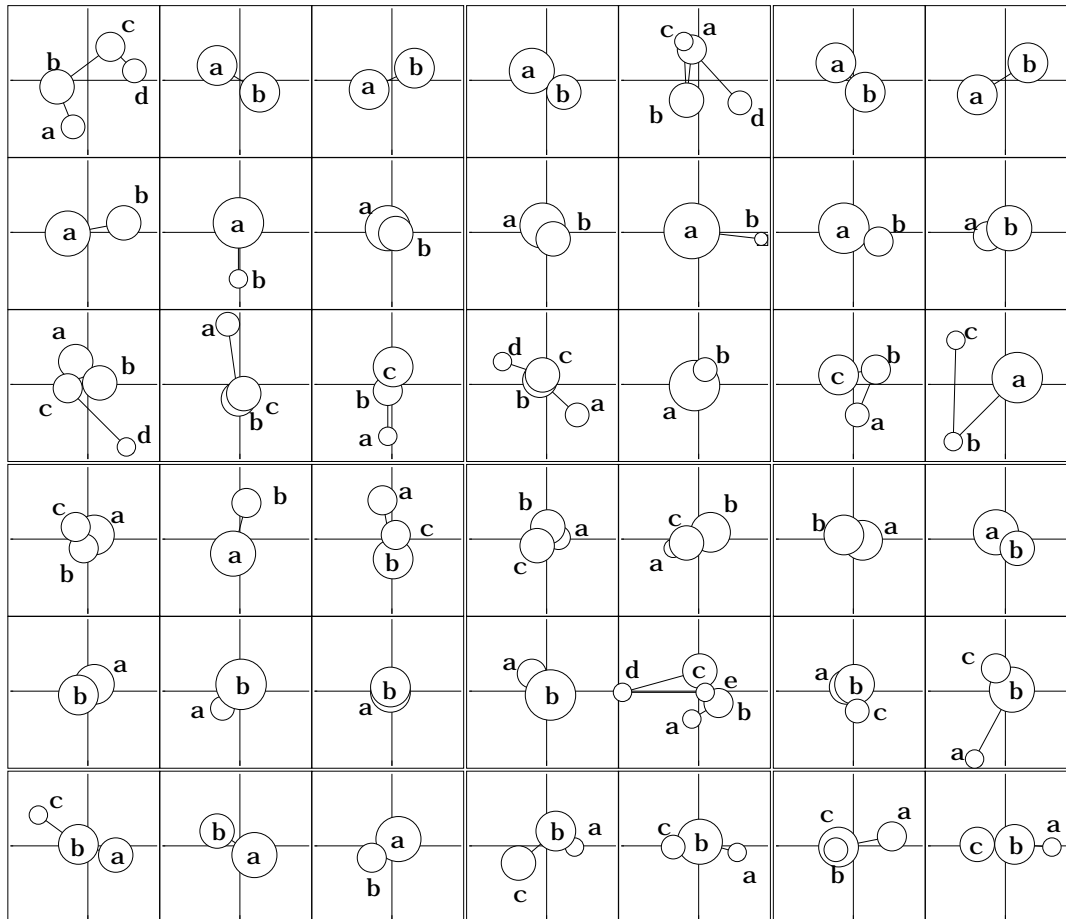


0. 62
 - 0. 68 | 0. 65
 - 0. 72

On notera que la disposition des individus se fait dans le plan pour que la corrélation soit nulle, ce qui est une contrainte de fonctionnement, mais que la disposition des moyennes se fait dans des directions orthogonales ce qui n'est pas toujours le cas. Une fois que le principe de positionnement des modalités par moyennes est compris, le plus simple est d'interpréter les cartes des modalités seulement, ce qui est plus rapide. Par exemple, dans Curves on peut représenter les modalités par variables :

Lines			
X file (default = 1, 2, 3, ..., n)		Tab.cmco	110 4
X file column number (default = 1)			
Y file (no default)		Tab.cmco	110 4

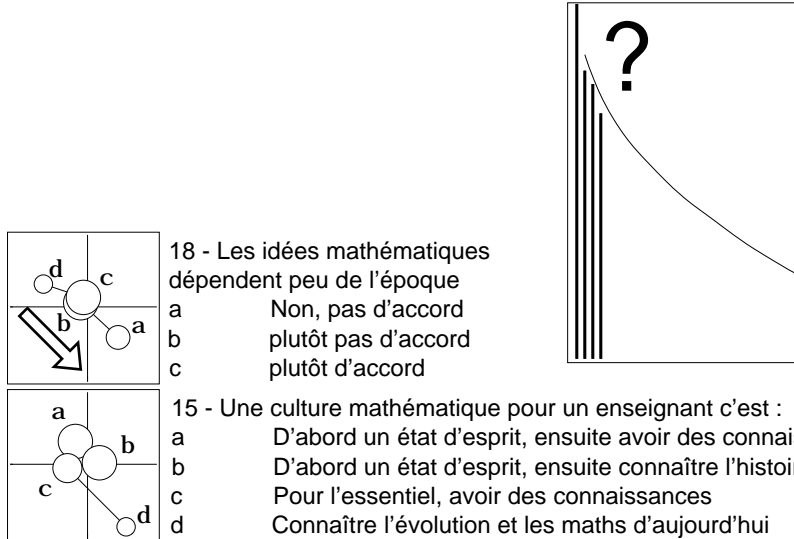
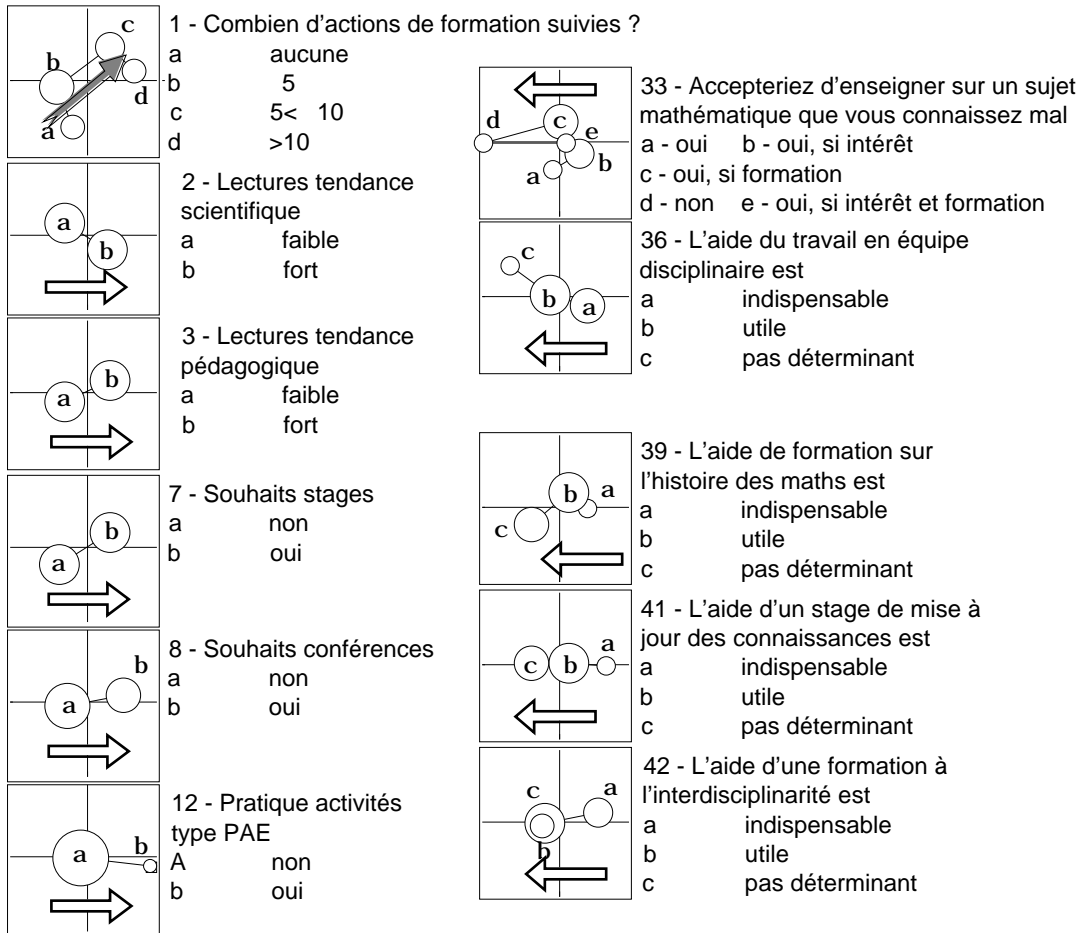
Row & col. selection	
Col. selection:	<input type="text" value="2"/>
Row selection method:	<input checked="" type="radio"/> File <input type="radio"/> Keyboard
Row selection file (.cat):	<input type="text" value="TabModa.cat"/>
Selection col. number:	<input type="text" value="1"/>



$$\begin{matrix} 1. & 4 \\ -1. & 5 \\ -1. & 4 \end{matrix}$$

Pour obtenir la figure ci-dessus, on a superposé les poids des modalités par Scatters, ce qui souligne le centrage par variables en utilisant la pondération marginale des modalités :

Values	
HV coordinates file	<input type="text" value="Tab.cmco"/> 110 4
H-axis column number (default = 1)	<input type="text"/>
Y-axis column number (default = 2)	<input type="text"/>
G values file	<input type="text" value="Tab.cmpc"/> 110 1



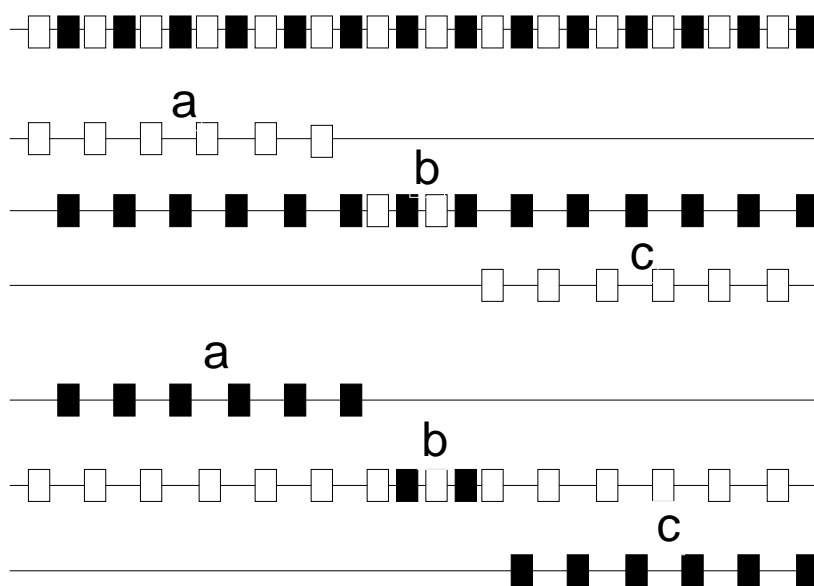
L'axe 1, ainsi représenté, exprime globalement que ce qui exprime le plus de différences entre les enseignants interrogés touche à la nécessité et à la possibilité de renouveler leur propre position. Ceux qui ont suivi le plus de formations sont ceux qui en espèrent le plus et en pensent le plus grand bien. Le jugement porté sur la nécessité de la formation est cohérent avec le souhait formulé de formation personnelle et le niveau effectif d'une telle formation (depuis "cela ne sert à rien, je n'en veux pas, je n'en fais pas" à l'inverse "c'est indispensable, j'en veux et j'en fais beaucoup"). La cohérence dans le questionnaire se fait avec l'idée d'une évolution tant pédagogique (je pourrais enseigner un contenu que je ne connais pas encore ou je m'y refuserais) ou

inter-disciplinaire (il est indispensable ou “pas déterminant” de se former à l’interdisciplinarité).

Mais cette représentation pose clairement la question de la liaison de cette ordination qu’on peut appeler indice d’investissement avec le bloc des quatre questions sur la culture mathématique, petit bloc formant la deuxième partie du questionnaire.

On voit bien se refaire l’association entre “les idées mathématiques évoluent” — la question avec une double négation ne pouvait appartenir qu’à une questionnaire posé par des mathématiciens à d’autres mathématiciens ! — et “la culture c’est connaître les mathématiques d’aujourd’hui”. Mais il apparaît que l’ordination se fait en bas à droite orthogonalement à l’ordre des modalités de la question 1 qui sert de référence. On peut se s’interroger sur la pertinence du dépouillement du facteur 2. Le point d’interrogation souligne sur le graphe des valeurs propres cette question. D’autant plus que le facteur 3 reprend un effet Guttman partiel sur le 1.

On notera que l’axe 1 est l’essentiel de l’analyse. Bien que nous ayons limité la redondance dans le premier bloc de variables et que la structure soit faible dans le troisième, les deux indication d’investissement ont été associées automatiquement. L’analyse préfère un grand nombre de correspondances cohérentes et dispersées à l’association forte entre deux variables qui disparaît dans la vision simultanée des 42 variables. Elle porte bien son nom d’analyses des correspondances multiples. On en voit aussi les limites. Il est possible de se faire piéger par des superposition de structures sans liaison :



En dépit de la non corrélation des coordonnées, il faut se souvenir que deux facteurs peuvent dire la même chose alors qu’un seul peut en dire plusieurs.





3 — Tableaux croisés

Nous examinerons donc de manière séparée la relation entre les parties 1 et 2 du questionnaire à partir de la question :

Les deux tendances minoritaires apparaissant dans la partie 2 peuvent elles l’une et l’autre être repérées sur l’axe d’investissement de la partie 1 ?

Extraire les 14 premières colonnes de Tab dans un fichier P1 et les 4 suivantes dans un fichier P2. Lire ces deux fichiers avec ReadCateg de CategVar. on retrouve les deux distributions marginales consignées avec les codes page 5 et 6.

3.1 — Tests univariés

CrossingCateg	
First ---.cat type file (Table X)	 P1.cat
Second ---.cat type file (table Y)	 P2.cat
Weighting file (Default = none)	 <input type="text"/>
Output file (HtDY)	 P12

File P12BB contains crossed Burt's matrix $X'DY$ from coding matrix X and coding matrix Y. It has 32 rows (categories) and 14 columns (categories)

On peut éditer le fichier P12BB avec Excel™ :

3	4	5	1	0	6	6	2	4	3	2	2	5	1
12	11	7	6	7	14	13	6	12	14	6	12	10	4
9	9	5	1	5	9	8	1	5	18	4	7	7	2
6	5	3	0	3	3	6	1	4	9	2	5	2	0
21	12	9	1	10	16	14	4	13	22	2	15	12	6
9	17	11	7	5	16	19	6	12	22	12	11	12	1
14	15	12	4	5	19	17	6	13	20	4	13	12	6
16	14	8	4	10	13	16	4	12	24	10	13	12	1
18	18	11	4	13	20	12	6	12	30	9	17	11	5
12	11	9	4	2	12	21	4	13	14	5	9	13	2
7	11	6	2	5	12	7	1	6	16	1	8	7	2
12	12	9	3	5	10	17	7	8	18	6	12	10	3
5	1	3	1	4	1	5	0	6	4	3	3	2	1
6	5	2	2	1	9	4	2	5	6	4	3	5	1
16	14	8	1	8	12	15	5	13	20	4	12	10	1
14	15	12	7	7	20	18	5	12	24	10	14	14	6
12	13	13	4	6	19	11	6	8	22	5	12	15	4
18	16	7	4	9	13	22	4	17	22	9	14	9	3
22	16	15	4	9	23	20	8	16	27	6	22	15	5
8	13	5	4	6	9	13	2	9	17	8	4	9	2
24	27	19	6	14	27	29	7	22	40	11	21	22	7
6	2	1	2	1	5	4	3	3	4	3	5	2	0
17	20	16	6	11	25	20	5	19	27	9	18	15	5
13	9	4	2	4	7	13	5	6	17	5	8	9	2
20	17	16	5	9	23	20	6	15	31	7	15	20	6
10	12	4	3	6	9	13	4	10	13	7	11	4	1
28	25	20	8	15	32	28	10	22	41	13	24	24	7
2	4	0	0	0	0	5	0	3	3	1	2	0	0
24	20	16	5	11	27	26	6	18	34	11	19	18	7
6	9	4	3	4	5	7	4	7	10	3	7	6	0
8	9	3	2	3	5	11	2	8	9	3	4	8	3
17	19	11	6	8	22	19	6	13	29	8	20	13	4

Access for X: file P1.cat
 Access for Y: file P2.cat
 File P12BB.bli contains row indicator (number of modalities for each variable)
 It has 14 rows (categories) and 1 column
 File P12BB.bco contains column indicator (number of modalities for each variable). It has 4 rows (categories) and 1 column

 Khi2 tests on Burt's crossed table P12BB

Var	1 versus var	1: Khi 2 =	6.8887e+00	ddl =	9	Proba =	0.650
...							
Var	2 versus var	4: Khi 2 =	1.1318e+01	ddl =	3	Proba =	0.010
Var	3 versus var	1: Khi 2 =	8.6540e-01	ddl =	3	Proba =	0.835
...							
Var	4 versus var	2: Khi 2 =	1.1450e+01	ddl =	2	Proba =	0.003
...							
Var	8 versus var	4: Khi 2 =	7.6827e+00	ddl =	3	Proba =	0.052
...							
Var	11 versus var	4: Khi 2 =	6.9070e+00	ddl =	3	Proba =	0.073
...							
Var	12 versus var	2: Khi 2 =	7.5960e+00	ddl =	2	Proba =	0.022

Chaque table de contingence donne un test Khi2. La plupart ne sont pas significatifs. Un seul l'est fortement, celui qui relie tendance à la lecture informatique et opinion sur la question culture mathématique et informatique.

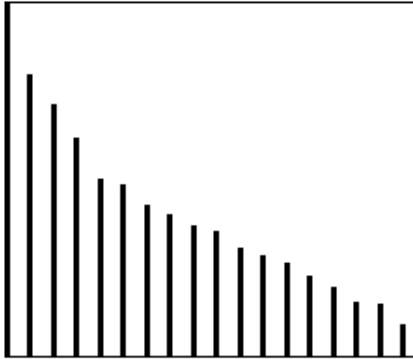
3.2 — Test multivarié

Effectuer d'abord les deux analyses des correspondances multiples :

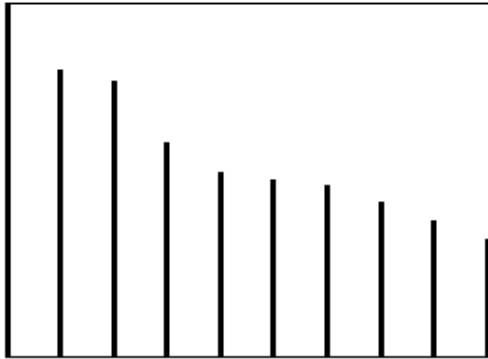
Multiple Correspondence Analysis

.cat type file

Option: Row weighting file



Number of axes ?



Number of axes ?

Multiple Correspondence Analysis

.cat type file

Coupler ensuite les deux analyses dans le module CoInertia :

Matching two statistical triplets			
First data input file		P1.cmta	87 32
Second data input file		P2.cmta	87 14
Output file name		A	

First input statistical triplet: table P2.cmta
 Number of rows: 87, columns: 14
 Second input statistical triplet: table P1.cmta
 Number of rows: 87, columns: 32

File A.<>ta contains the crossed table YtDnX
 It has 32 rows and 14 columns

File A.<>pl contains the weight of each row
 It has 32 rows and 1 column

File A.<>pc contains the weight of each column
 It has 14 rows and 1 column

File A.<>ma contains the name of the input statistical triplet

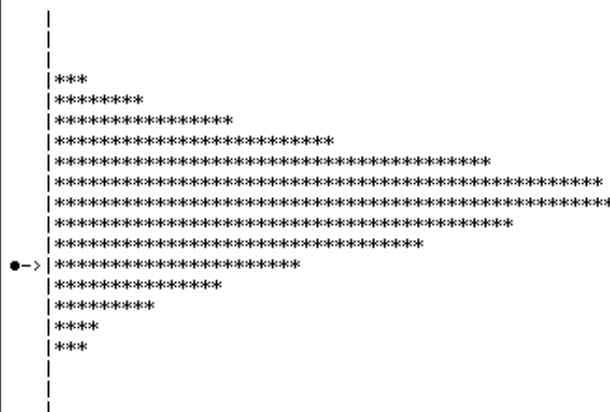
1 ---> P2.cmta
 2 ---> P1.cmta

total inertia: 0.037865

L'ordre des deux tableaux dans le couple a été changé, pour affecter en premier celui qui a plus de colonnes.

Coinertia test - Fixed D	
---.<>ma input file	A.<>ma
Select a number of permutations	1000

```
number of random matching: 1000 Observed: 0.037865
Histogramm: minimum = 0.020632, maximum = 0.048424
number of simulation X<Obs: 858 (frequency: 0.858000)
number of simulation X>=Obs: 142 (frequency: 0.142000)
```



Le test de permutation sur la trace largement discuté dans un article récent ¹ ne permet pas d'associer significativement les deux tableaux au niveau multivarié. La réponse à la question posée est donc négative. Dans le cas inverse, l'analyse de co-inertie aurait permis un bilan d'association des deux tableaux.

Reste donc à faire un bilan de l'enquête à partir des éléments statistiques. On pourra souligner que l'acquisition des données a garanti la représentativité du groupe de personnes étudiées et mis en évidence des opinions consensuelles là où on n'en attendait pas toujours. Dans l'analyse de l'hétérogénéité de l'échantillon, on n'a pas trouvé de partitions mais une très grande diversité d'opinion personnelle. Par exemple, les questions de l'objectif de l'enseignement des mathématiques dans les séries donne le petit tableau :

	Sections		
Littéraires	Socio-éco	Scientifiques	
69	35	15	Donner un regard critique
11	23	17	Introduire des concepts nouveaux
7	18	41	Préparer au cursus ultérieurs

Pratiquement, ces questions n'apparaissent jamais dans la discussion, car si les personnes interrogées utilisent peut-être de préférence l'une ou l'autre des modalités de réponse, l'important est dans le consensus structurel (consensus sur les différences organisées) d'un objectif variant de manière continue en fonction des sections.

Autour du consensus, on a trouvé un grand nombre de combinaisons d'appréciations personnelles qui laisse deviner un individualisme n'autorisant guère la définition de types d'opinion. L'informatique laisse ou non les personnes indifférentes, l'arithmétique, l'histoire des mathématiques ou les objectifs pédagogiques proposés appelle des réponses variées. On voit parfois une minorité se dessiner sur plusieurs critères mais jamais d'archétypes, tels qu'on en espère pour montrer l'efficacité des méthodes multivariées.

On espérait enfin définir des types d'attitude permettant de définir des groupes de personnes plus spécialement concernées par des fonctions nouvelles. On a surtout trouvé une axe d'investissement personnel qui oppose des personnes ouvertes à un grand nombre de questions à des personnes ayant manifestement des centres d'intérêt extérieurs au sujet de l'enquête. Notons encore que la question 1 — qui fait état d'un fait dans le nombre d'actions de formation suivies — joue un rôle prédominant (c'est le seul rapport de corrélation supérieur à 30% dans l'analyse globale) alors que, mélangée à de nombreuses autres, rien ne l'y prédispose. Cela donne du poids à la définition de cet axe principal. Quelque soit l'objectif d'une action novatrice, il y a bien des chances qu'on est affaire aux mêmes acteurs potentiels.

Références

¹ Kazi-Aoual, F., Hitier, S., Sabatier, R. & Lebreton, J.D. (1994) Refined approximations to permutation tests for multivariate inference. *Computational Statistics and Data Analysis* : (In press).