Troisièmes Rencontres Internationales – Terzo Convegno Internazionale - Third International Conference
A.S.I. Analyse Statistique Implicative – Analisi Statistica Implicativa – Implicative Statistic Analysis
Palermo (Italy) 6-8 Octobre 2005

Des variables supplémentaires et des élèves « fictifs », dans la fouille didactique de données avec CHIC

Pilar Orús, Pablo Gregori

Dpto. de Matemáticas, Universidad Jaume-I, Campus Riu Sec-ESTCE Castellón.
orus@mat.uji.es
gregori@mat.uji.es

Résumé. Les variables supplémentaires et les élèves « fictifs » peuvent nous aider dans la fouille des données avec CHIC. On montre ici une utilisation didactique de ces outils et leurs apports dans une analyse des réponses des étudiants de première année d'études universitaires en Espagne (Université Jaume I de Castelló), à un contrôle des connaissances initiales en mathématiques.

1 Introduction

Un contrôle sur les connaissances initiales en Mathématiques des élèves de premier cours de l'Université Jaume I de Castelló (UJI) en Espagne se déroule depuis l'année 2001. La première fois, il faisait partie d'un projet de Didactique en Mathématiques (Bosch et coll., 2001-2003) et d'une thèse de doctorat de C. Fonseca (2004). Les questions étaient choisies avec l'intention de prouver certaines hypothèses didactiques sur les discontinuités mathématiques et didactiques entre les enseignements secondaire et universitaire. Lors des cours suivants, le contrôle a fait partie de divers projets d'Amélioration Educative, avec l'objectif institutionnel des enseignants du Département des Mathématiques, de constater et d'analyser les connaissances en mathématiques de leurs étudiants par rapport à ce que l'on suppose acquis à la fin de l'enseignement secondaire, et ceci afin de leur fournir l'information leur permettant de prendre, chacun à titre individuel, les mesures didactiques opportunes.

1.2 Objectifs de l'étude

L'analyse statistique classique de l'échantillon nous permet de présenter des résultats qui résument la diversité de données et peuvent montrer les différences entre les groupes relativement à chaque variable.
Par ailleurs, l'analyse statistique classificatoire des similarités, l'analyse implicative et cohésitive, nous permettent en plus d'explorer les données, d'extraire et de découvrir également des relations entre les individus et les variables qui constituent l'étude.
Les concepts comme la contribution et la typicalité des individus et des variables dans la formation de classes dans la classification des similarités, ou bien dans la hiérarchie cohésitive aident à interpréter et comprendre la nature des rapports entre individus et variables.
L'objectif de notre contribution est de présenter, dans ce cadre, l'idée que l'introduction de certains individus "*fictifs*" et leur localisation dans des procédures comme la contribution et la typicalité des diverses analyses statistiques, nous offre des nouvelles opportunités d'interpréter didactiquement les résultats obtenus. Ces individus sont en réalité des caractères des questions ; avec leurs "réponses", ils constituent une matrice dite a priori du questionnaire. Cependant leur cardinal doit être suffisamment "petit" par rapport à l'échantillon pour que cette perturbation relativement petite de l'échantillon préserve (ne modifie pas sensiblement) les résultats globaux. Notre apport, à travers cet article, vise donc deux plans : celui de la didactique et celui de la méthodologie de l'analyse de données en tant que corollaire du premier.

1.3 Description de l'étude

Troisièmes Rencontres Internationales – Terzo Convegno Internazionale - Third International Conference
A.S.I. Analyse Statistique Implicative – Analisi Statistica Implicativa – Implicative Statistic Analysis
Palermo (Italy) 6-8 Octobre 2005

1.2.1 Les données

L'étude présentée ici s'appuie sur les données obtenues dans le contrôle de connaissances effectué avec les élèves de première année des diverses études universitaires de *l'Escola Superior de Tecnologia i Ciències Experimentals de la UJI*, au début du cours 2003-04.
Les groups d'étudiants impliqués dans l'étude sont les suivants, accompagnés des codes des groupes et des codes des élèves, codes que nous utiliserons pour les analyses ultérieures:

ÉTUDES (Filières)	**GROUPES**	**Codes ÉLÈVES**
Ingénierie Industrielle (IIND)	A, B et E	**Na***xx*, **Ne***xx*
Ingénierie Informatique (II)	A et B	**Fa***xx*, **Fb***xx*
Ingénierie Chimique (IQ)	A	**Qa***xx*
Ingénierie Tech. Agricole (ITA)	A	**Ta***xx*
Ingénierie Tech. en Design Industriel (ITDI)	A	**Da***xx*
Ingénierie Tech. Informatique en Gestion (ITIG)	A et B	**Ga***xx*, **Gb***xx*
Ingénierie Tech. Informatique en Systèmes (ITIS)	A	**Sa***xx*
Ingénierie Tech. Industrielle, esp. Mécanique (ITIMEC)	A	**Ma***xx*
Licence en Chimie (LQ)	A et B	**La***xx*, **Lb***xx*

Tab 1 - Groupes et types d'études des élèves qui ont répondu à la PROVA

Le contrôle (*PROVA* en Annexe) est constitué de 17 questions (correspondant à 21 items individuels) dont les réponses ont été codées par un nombre : 0 (réponse incorrecte ou pas de réponse) ou 1 (réponse correcte, quelle que soit la technique utilisée). D'un autre côté elles sont aussi codées qualitativement (selon la technique et/ou le type d'erreur commis), mais on laisse cette partie des données pour d'autres types d'analyses.
En conséquence on a un tableau de contingence de 21 variables et 690 individus, variables auxquelles on peut ajouter une nouvelle variable nommée TOTAL, qui attribue a chaque individu le nombre de réponses correctes (normalisé sur l'intervalle [0-10]) comme indicateur d'un niveau de connaissances en Mathématiques.

1.2.2 La MAP du questionnaire: les données fictives

Une classification, a priori, des questions du contrôle de la *PROVA*, selon le type de connaissance et de tâche, s'exprimé dans le tableau suivant :

TYPE de CONNAISSANCE	**TYPE de TACHE**	**Nº DE QUESTION**
ALGÈBRE ***Algb***	Problème : ***Probl***	P1, P6a, P6b, P8, P10, P12, P13a, P13b, P15, P16, P17b
	Graphique : ***Grafi***	P17a
	Exercice : ***Ejerc***	P2, P7
CALCUL ***Calcu***	Problème : ***Prob***l	P8, P14b, P17b
	Graphique : ***Grafi***	P3, P17a
	Exercice : ***Ejerc***	P3, P4, P5, P9, P11, P14a

Tab-2: Type de connaissances et de tâches des questions de la PROVA

Ces caractéristiques des questions seront utilisées pour définir les élèves fictifs de l'étude.

2 Analyse statistique classique de la variable « total »

Les résumés statistiques de la variable *TOTAL* (valeur min. 0, valeur max. 10), stratifiée par groupes et par type d'études, figurent dans le tableau suivant :

Group	Nº él.	Moyenne (sur 10)	R Min.	P25	Médiane(P50)	P75	Max.	Coeff. Var.
IIND E	23	**5,16**	1,43	3,81	4,76	6,67	8,57	34,97%
IIND A	50	**4,20**	1,43	2,86	4,05	5,24	9,05	40,21%
II A	38	**4,02**	0,95	2,86	3,81	4,76	9,52	46,57%
IIND B	27	**3,65**	1,91	2,86	3,33	4,29	6,19	34,47%
IQ A	49	**3,56**	0,95	1,91	3,33	4,76	8,10	50,23%
II B	39	**3,39**	0,00	2,38	3,33	4,76	6,19	43,29%
ITDI A	52	**3,28**	0,95	1,91	3,33	4,53	6,19	44,51%
LQ A	50	**3,15**	0,48	1,91	2,86	4,29	8,10	52,91%
ITIG A	94	**3,10**	0,95	2,38	2,86	3,81	6,19	40,43%
LQ B	34	**2,73**	0,95	1,43	2,62	3,81	6,67	52,28%
ITIS A	82	**2,57**	0,48	1,91	2,38	3,33	5,24	45,76%
ITIG B	42	**2,47**	0,48	1,43	2,15	3,33	5,24	49,26%
ITIMECA	55	**2,04**	0,00	0,95	1,91	2,38	5,24	54,69%
ITA A	55	**1,50**	0,00	0,95	1,43	1,91	3,81	67,66%
Total	**690**	**3,06**	**0,00**	**1,91**	**2,86**	**3,81**	**9,52**	**53,35%**

Tab -3: Statistiques de la variable TOTAL stratifiée par groupe et type d'études

L'ordre des groupes, en fonction de la réussite, permet mettre en évidence un niveau très faible de la réussite globale au questionnaire; seulement le groupe singulier E d'étudiants européens, des études d'Ingénierie Industrielle, ont une moyenne supérieure à 5. le reste des groupes est par-dessous de 5. Les groupes des ingénieries techniques (3 ans d'études universitaires) ont une moyenne inférieure à celles des groupes des études de licence ou d'ingénierie (5 ans), sauf le groupe A d'Ingénierie Tech. en Design Industriel (ITDI), placé devant des groupes de la licence en Chimie (LQ). Le groupe plus faible est celui d'Ingénierie Technique Agricole (ITA), avec une moyenne de 1,5 sur 10 et la note plus forte du groupe 3,81, n'arrive pas au 5.

Les groupes A ont toujours des moyennes plus fortes que les groupes B du même type d'études; il s'agit des groupes dits "de matin", préférés davantage par les élèves et choisis par les "bonnes élèves" qui ont priorité au moment de l'inscription.

Puisque les données ne semblent pas suivre la loi normale, une analyse de la variance ne peut pas être abordée, mais le test non paramétrique de Kruskal–Wallis peut être utilisé pour vérifier l'existence de différents niveaux de la variable TOTAL parmi les divers types d'études (comparaisons de médianes). Une représentation graphique utile pour l'analyse du niveau des étudiants est exploitée à travers les graphes « boxplots » au style de Tuckey, qui permettent de visualiser 5 niveaux significatifs (95% de confiance) des types d'études:

1: IIND 2: II, IQ, ITDI, LQ, ITIG 3: ITIS 4: ITIMEC 5: ITA

Troisièmes Rencontres Internationales – Terzo Convegno Internazionale - Third International Conference
A.S.I. Analyse Statistique Implicative – Analisi Statistica Implicativa – Implicative Statistic Analysis
Palermo (Italy) 6-8 Octobre 2005

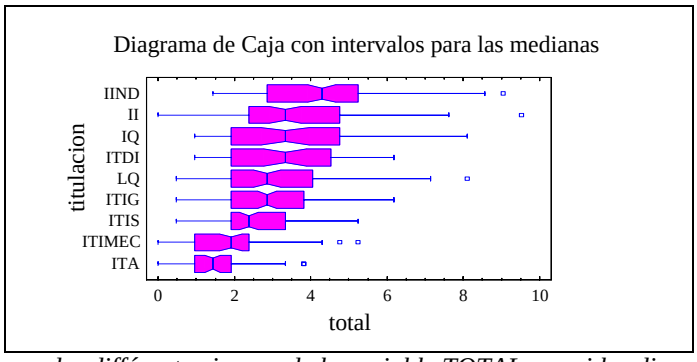


Figure- 1: Diagramme des différents niveaux de la variable TOTAL parmi les divers groupes et études.

3 Analyse classificatoire, implicative et cohésitive

3.1 Analyse classificatoire

Les variables ordinaires de l'échantillon, sont des variables binaires (réussite-échec) correspondant aux questions de la *PROVA*, (codées P1, P2, P3, P4, P5, P6a, P6b, P7, P8, P9, P10, P11, P12, P13a, P13b, P14a, P14b, P15, P16, P17a et P17b). On a retenu comme variables supplémentaires celles dont les données sont relatives à :

- Type d'études de l'étudiant à la UJI (Code: TITUL)
- Avez-vous suivi récemment un cours de préparation de Math.?:(MATPREV)
- Type d'études qui débouchent à l'entrée à l'Univ.): (TIPOACC)
- Résultat au test d'accès à l'Université: (NOTAACC)

Le traitement des données a été mené avec le logiciel CHIC et nous a permis réaliser d'abord une analyse des proximités des variables selon l'indice de I.C. LERMAN [Lerman 1981].

L'arbre de similarité montre des nœuds significatifs aux niveaux 1, 6, 9, 12, 15, 17, et 19, le niveau 12 étant le plus significatif, en agrégeant des clases aux niveaux 1, 6 et 9. Nous allons décrire le bloc de variables de ce niveau 12, à l'aide des caractéristiques des questions présentées au tableau TAB-2.

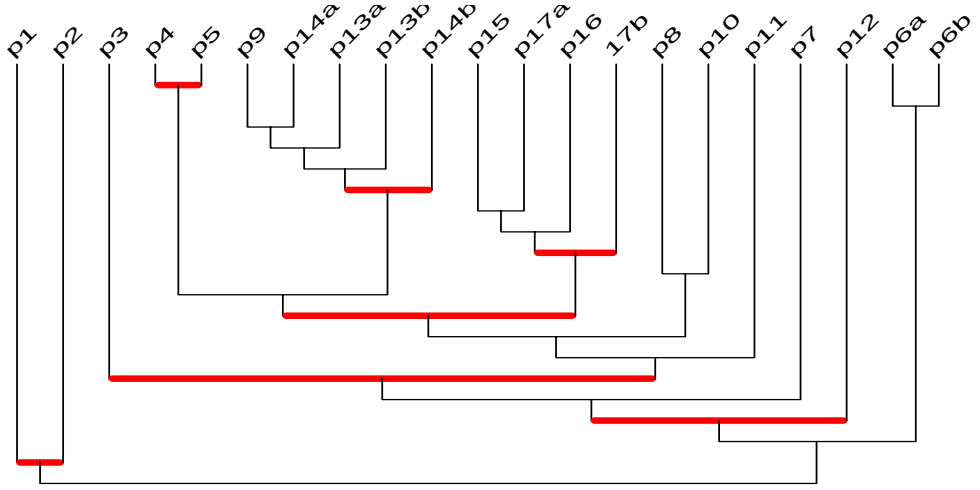
Les questions (P4, P5) de la clase du niveau 1, sont des *exercices* de *calcul*.

Les questions du niveau 9 (((P15, P17a), P16), P17b) sont des *problèmes* d'*algèbre*.

Le niveau 6 agrège des questions mixtes quant aux champs conceptuels et aux compétences en jeu : la classe des *exercices* de *calcul* (P9, P14a) s'unit selon une échelle de complexité de Guttman à celle de *problèmes* d'*algèbre* P13a et P13b, puis au *problème* de *calcul* P14b, pour former la classe significative ((((P9, P14a), P13a), P13b), P14b).

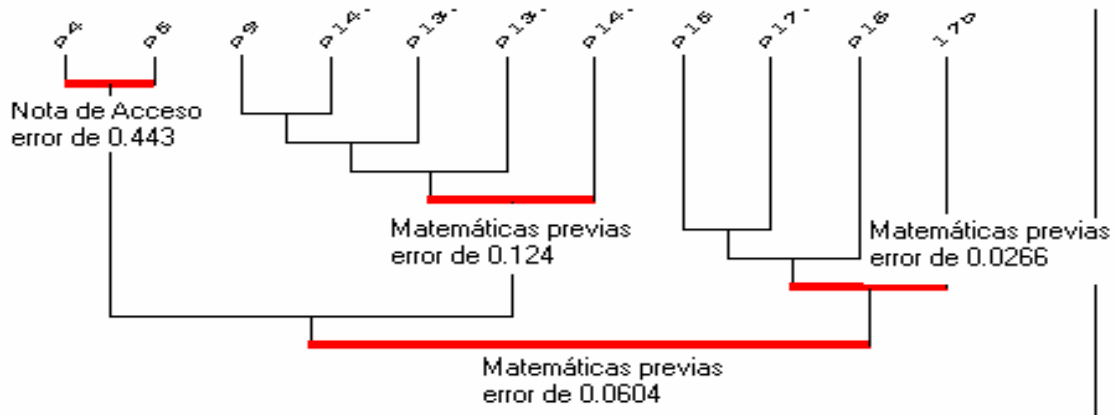
Examinons la contribution à la formation de la classe du niveau 12 : le groupe optimal est de 38 élèves, tous des élèves (sauf 1 élève de Ingénierie Tech. en Design Industriel, codé D*axx*) inscrits dans des études universitaires supérieures (5 années de durée: ingénierie ou licence), et dont la plupart (19) suivent des études d'Ingénierie Industrielle (les élèves codés N*pxx*). Le reste des étudiants appartient à l'Ingénierie Chimique (7 élèves, codés Q*pxx*), l'Ingénierie Informatique (6 élèves, codés F*pxx*), et la Licence en Chimie (5 élèves, codés L*pxx*)

Troisièmes Rencontres Internationales – Terzo Convegno Internazionale - Third International Conference
A.S.I. Analyse Statistique Implicative – Analisi Statistica Implicativa – Implicative Statistic Analysis
Palermo (Italy) 6-8 Octobre 2005



Arbre- 1: L'arbre de similarité des questions de la Prova-20003/04 élaboré par CHIC

Le groupe optimal pour la typicalité est composé seulement des élèves d'ingénierie industrielle ou informatique (Ne396, Fa63, Fa67, Qa296, Fa74, Ne406).
À l'égard des variables supplémentaires, c'est la variable MATPREV (*avoir suivi récemment un cours de préparation en Mathématiques*) qui contribue le plus à cette classe du niveau 12, au nœud le plus significatif (avec un risque de 0.060). Elle est à la fois sa variable supplémentaire typique, (avec un risque de 0.029).
L'analyse de la contribution des variables supplémentaires à la formation des classes (1, 6, 9) au nœud le plus significatif, montre que c'est aussi la variable MATPREV (*avoir suivi récemment un cours de préparation en Mathématiques*) qui contribue à chacune des classes (sauf pour le niveau 1: (P4, P5)) comme on l'observe dans l'arbre:

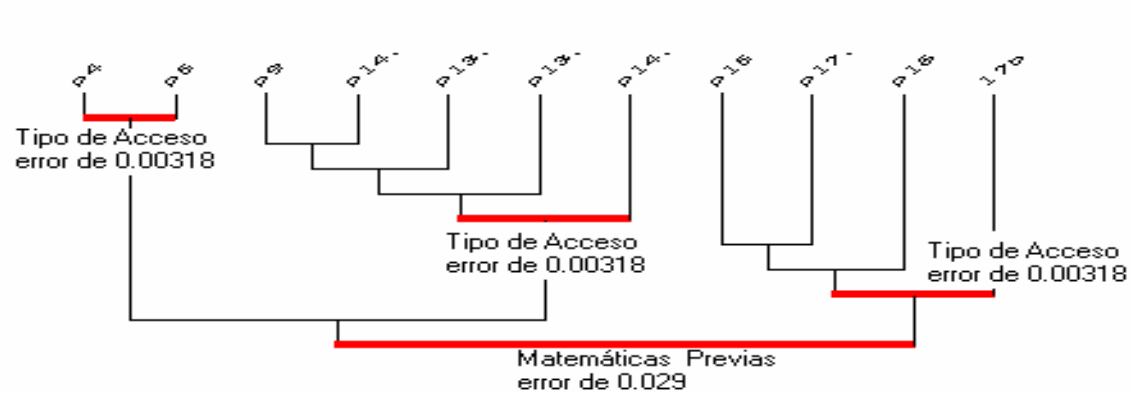


Arbre-2: Contribution des variables supplémentaires aux classes du Niveau 12 de l'arbre de similarité

L'analyse des variables supplémentaires aux nœuds significatifs (1, 6, 9) formant la classe du niveau 12, montre que la variable *Type d'accès à l'Université* (les études qui débouchent à l'entrée à l'Université)

Troisièmes Rencontres Internationales – Terzo Convegno Internazionale - Third International Conference
A.S.I. Analyse Statistique Implicative – Analisi Statistica Implicativa – Implicative Statistic Analysis
Palermo (Italy) 6-8 Octobre 2005

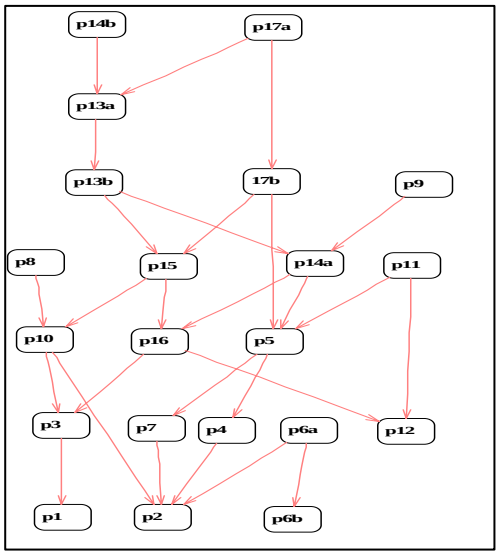
apparaît comme la variable la plus typique de chacune des classes formées au niveau 12, avec un risque d'erreur très petit, 0.003.



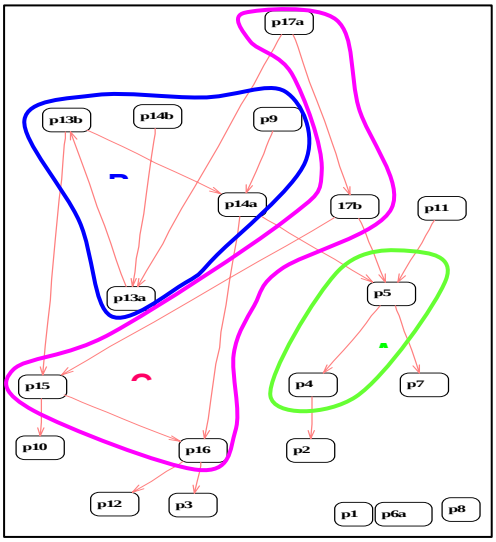
Arbre-3: Variables supplémentaires typiques des classes du Niveau 12 de l'arbre de similarité

3.2 Analyse implicative

Le résultat de l'analyse implicative de la *PROVA* donné par CHIC, en utilisant la loi de Poisson, désigne les règles entre les questions, en fonction des réponses des élèves.
Les graphes suivants représentent ces implications : à gauche, le graphe complet avec toutes les variables (les questions) et à droite le graphe restreint aux variables correspondant au nœud (12) de la classe la plus significative.
Dans le graphe implicatif de la classe 12 on peut distinguer les sous-graphes relativement homogènes **A**, **B,** et **C** qui, bien que liés, caractérisent des champs distincts Ils représentent des règles entre les questions qui constituent elles-mêmes les classes aux nœuds significatifs 1, 6 et 9 respectivement.



Graphe-1 : Graphe implicatif des questions



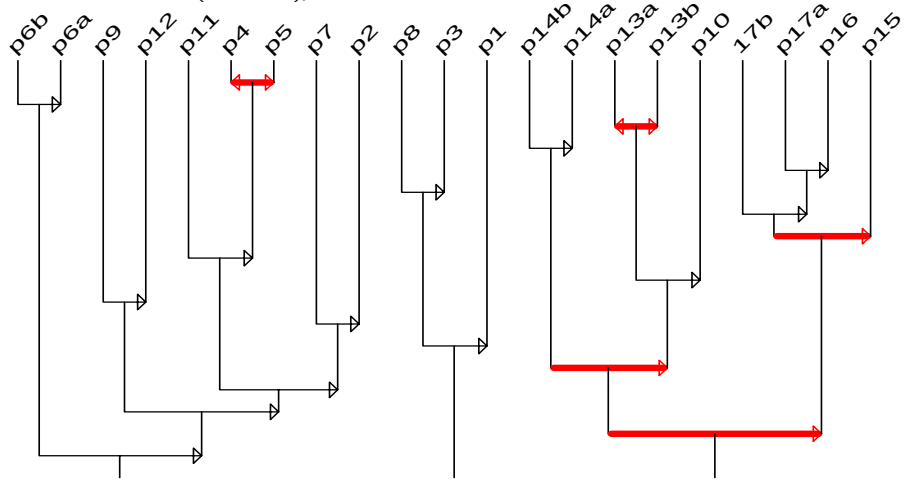
Graphe -2: Graphe implicatif du niveau 12

Troisièmes Rencontres Internationales – Terzo Convegno Internazionale - Third International Conference
A.S.I. Analyse Statistique Implicative – Analisi Statistica Implicativa – Implicative Statistic Analysis
Palermo (Italy) 6-8 Octobre 2005

L'implication entre des *exercices* de *calcul* P5 ⇒ P4 du sous-graphe **A**, montre que les élèves qui savent calculer une intégrale définie d'une fonction très simple ($\int 2axdx$), savent dériver une fonction un peu plus complexe ($(f(x) = 5 / (3x\text{-}2)^2$). L'implication de la question P11 sur (P4 → P5) s'explique aussi par ces deux caractéristiques, *exercice* et *calcul*, P11, un exercice de dérivation d'une fonction exponentielle ($f(x) = 8s^x$).
L'enchaînement de ces questions à P7 et P2 (*exercices* d'*algèbre*), peut être expliqué par le type de tâche : **A rassemble des exercices de calcul algébrique. C'est ce cadre qui est en jeu.**.
Les règles entre des *problèmes* d'*algèbre* (P17a → P17b → P15 → P16) du sous-graphe **C** le spécifie en tant qu'établissant la relation entre **activités de traduction selon deux registres, l'un graphique, l'autre algébrique**
Les différents niveaux d'agrégation des questions qui formaient la classe 6, se retrouvent et s'éclairent dans le sous-graphe **B**. La règle (P9 → P14a) pointe la relation entre deux *exercices* de *calcul* de limites. La chaîne (P14b → P13a → P13b) met en rapport trois *problèmes*. Ainsi, B spécifie des règles entre **capacités et connaissances de niveau supérieur, c'est-à-dire non algorithmiques**

3.3 Analyse de la cohésion

L'arbre cohésitif résultant de l'analyse des données de l'échantillon montre les nœuds significatifs aux niveaux 1, 3, 8, 14 et 17, le niveau 1 étant naturellement le plus significatif. Il est constitué d'une équivalence entre deux exercices de calcul (P4 et P5), comme le niveau 3.



Arbre -4 : Arbre cohésitif des questions de la Prova-20003/04 élaboré par CHIC

Les *problèmes* d'*algèbre* P17a, P17b, P15 et P16 se retrouvent liées aussi ensemble dans la même classe (niveau 8), mais l'enchaînement des questions que le graphe implicatif montrait (sous-graphe C), en partant de l'implication entre les paires des variables (P17a → P17b → P15 → P16) est nuancé dans l'analyse cohésitive: [P17b → (P17a → P16] ⇒ p15. Cette classe assemble, de façon non symétrique, deux tâches traductives : l'une en **langage graphique** et l'autre en **langage formel.**
Dans le niveau 14 (significatif) se retrouvent ensemble les questions du sous-graphe B (GRAPHE-2), enchaînées dans l'analyse implicative le long du chemin (P14b → P13a → P13b → P14a) et qui, dans l'analyse cohésitive, apparaissent plus fortement liées : P13a et P13b présentent une co-implication significative (une classe de quasi équivalence) au niveau 3 de l'arbre cohésitif et l'ensemble des questions du niveau 14, suivent le schéma d'implication suivant: [P14b →P14a] ⇒ [(P13a ⇔ P13b) → P10)]. La

Troisièmes Rencontres Internationales – Terzo Convegno Internazionale - Third International Conference
A.S.I. Analyse Statistique Implicative – Analisi Statistica Implicativa – Implicative Statistic Analysis
Palermo (Italy) 6-8 Octobre 2005

cohésion donc, regroupe plus fortement dans cette classe, les parties d'une même question : l'activité de *calcul* P14 (le *problème* (b) impliquant l'*exercice* (a)), implique significativement les *problèmes* d'*algèbre* P13 et P10.
Le dernier niveau significatif de l'arbre cohésitif, le niveau 17, établit une méta-règle entre ces deux catégories des questions, celles de niveau 14 et celles de niveau 8, en montrant en une classe généralisée toutes les questions que nous venons de caractériser. La cohésion continue à être significative, entre les *problèmes complexes* de *calcul* avec raisonnement P14-P13 et les *problèmes* où deux langages sont en jeu
Afin de poursuivre et de conforter l'interprétation de ces classes, nous avons adjoint certains critères associés aux questions (c'est-à-dire les variables) en les considérant comme des élèves fictifs, dont les réponses seraient les caractéristiques de chaque question (la matrice dite a priori). Cette stratégie de traitement consistant à considérer une nouvelle matrice a priori, dans l'analyse des données en didactique, a été initiée par G. Brousseau, dans les analyses multivariées (AFC et ACP) (Brousseau et Lacasta, 1995).

4 Application de l'introduction de données fictives

Nous présentons, dans cette partie, les nouveaux résultats de l'analyse obtenue en modifiant l'échantillon original de données par un ajout d'individus fictifs, qui, a priori, ne devraient guère perturber les résultats, mais qui seront une nouvelle source d'information pour l'interprétation de la formation des classes, etc.
La classification faite a priori des variables dans le TAB-3, nous a permis définir *les élèves fictifs* de la manière suivante:

- L'élève *Algb* fait correctement les questions classifiées par ce type de connaissance, et incorrectement le reste.
- L'élève *Calcu* fait correctement les questions classifiées par ce type de connaissance, et incorrectement le reste.
- Les élèves *Prob*, *Grafi* et *Ejerc* sont définis de même que les antérieures, mais dans ce cas, selon le type de tâche.

En résumé, les élèves fictifs que nous avons ajoutés à l'échantillon sont : Algèbre (*Algb*) , Calcul (*Calcu*), Problème (*Probl*), Graphique (*Grafi*), Exercice (*Ejerc*).

4.1 Analyse classificatoire (données + élèves fictifs)

Il faut signaler, tout d'abord, que l'arbre des similarités avec des élèves fictifs est le même que l'Arbre-1, celui des données initiales, ce qui nous permet d'en continuer l'analyse et, en particulier, d'examiner le rôle joué par ces élèves fictifs dans la constitution de classes et de règles.
La contribution des individus dans l'analyse des similarités est la suivante, une fois les élèves fictifs (é.f.) ajoutés à l'échantillon

- Dans la formation de la classe du niveau 12, le plus significatif de cette analyse, apparaissent les élèves fictifs *Algèbre*, *Calcul*, plus 47 élèves (parmi un échantillon de 690 élèves, plus 5 é.f.). Le détail de la contribution de ces élèves fictifs dans les autres niveaux significatifs à l'intérieur de la classe 12 illustrent ce fait :
- Le groupe optimal de la classe de niveau 1 (P4, P5) est constitué par 141 élèves plus les élèves fictifs *Calcul*, et *Exercice.*
- L'élève fictif *Calcul*, plus 67 élèves, forment le groupe optimal et contribuent à la formation de la classe de niveau 6 (P9, P14a, P13a, P13b, P14b).
- La classe de niveau 9 (P15, P17a, P16, P17b) est constituée par *Algèbre* et 19 élèves.

Les élèves fictifs apparaissent aussi en contribuant à la formation d'autres sous-classes (même si les niveaux ne sont pas eux-mêmes significatifs):

- Au niveau 11, se forme la classe qui réunit les classes des niveaux significatifs 1 et 6. *Calcul*, et *Exercice* appartiennent à son groupe optimal.

Troisièmes Rencontres Internationales – Terzo Convegno Internazionale - Third International Conference
A.S.I. Analyse Statistique Implicative – Analisi Statistica Implicativa – Implicative Statistic Analysis
Palermo (Italy) 6-8 Octobre 2005

• Les caractéristiques *Problème* et *Algèbre* se trouvent souvent ensemble dans la contribution des classes (P15, P16), (P8, P10), (P6a, P6b).
Le denier niveau significatif, celui de la classe (P1, P2), est caractérisé par *Algèbre.* Ainsi, nos interprétations précédentes ne sont pas démenties

4.2 Analyse implicative (données + élèves fictifs)

Dans le graphe implicatif Graphe-2',en sélectionnant les questions de la classe significative 12, nous retrouvons les sous-graphes **A**, **B,** et **C**, maintenant codés **A'**, **B',** et **C'** avec les traits suivantes:

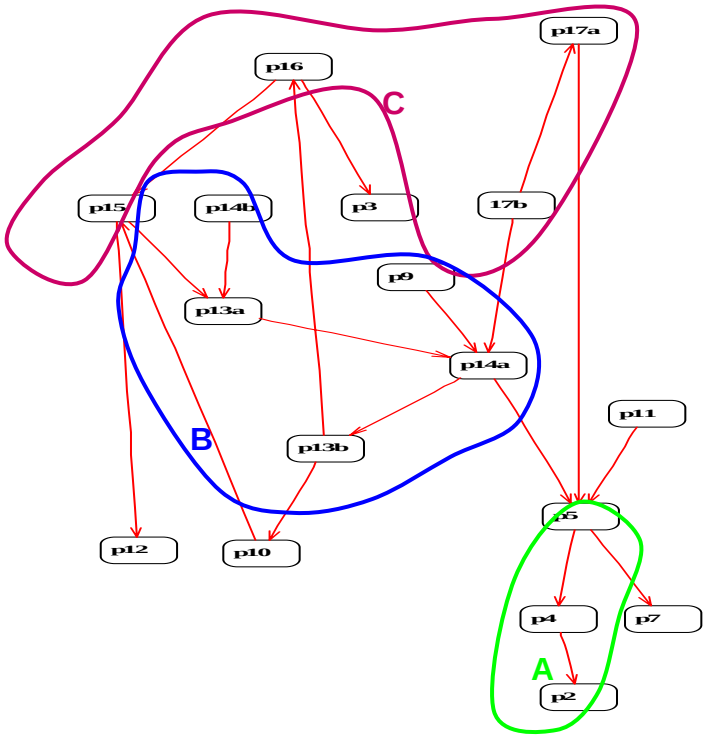
• L'implication des questions du sous-graphe **A'**, est la même que celle de **A**.

• Dans les sous-graphes **B'** et **B** l'unique implication qui change avec les élèves fictifs est l'implication P13b → P14a, qui apparaît sous sa forme réciproque P14a → P13b. Ce basculement est certainement un produit de la modification sensible des occurrences pour des variables ayant elles-mêmes des occurrences voisines.

• L'enchaînement des questions du sous-graphe **C** [P17a → P17b → P15 → P16], est brisé dans le sous-graphe **C'** en donnant deux implications P17a →P17b et P16→P15.

• L'enchaînement des questions du sous-graphe **C**, est légèrement brisé aussi dans le sous-graphe **C'** en donnant les implications P17a →P17b et P16→P15 (implication réciproque relative aux questions du tableau de données initial, P15→P16).

Le graphe implicatif avec les élèves fictifs, montre que se sont produits quelques changements dans les implications entre les variables observées dans les données initiales. Ce qui montre sa sensibilité aux changements d'occurrences et pourtant il ne semble pas non pertinent de continuer ce type d'analyse avec les élèves fictifs.

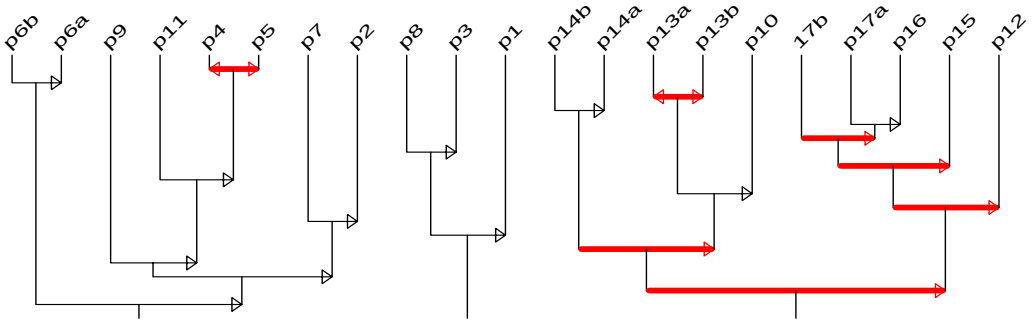


Graphe 2': Graphe implicatif du niveau 12, avec les élèves fictifs

Troisièmes Rencontres Internationales – Terzo Convegno Internazionale - Third International Conference
A.S.I. Analyse Statistique Implicative – Analisi Statistica Implicativa – Implicative Statistic Analysis
Palermo (Italy) 6-8 Octobre 2005

4.3 Analyse de la cohésion (données + élèves fictifs)

L'arbre cohésitif montre les nœuds significatifs aux niveaux 1, 3, 6, 8, 11, 14 et 17, le niveau1 étant le plus significatif (le niveau 11 ne l'était pas).



Arbre 4' : Arbre cohésitif des questions de la Prova-20003/04 , avec des élèves fictifs

La cohésion est très forte : jusqu'au niveau 10 elle est égale à 1 ; jusqu'au niveau 15 elle est supérieure à 0.99 et dans le dernier niveau significatif, le niveau 17, la cohésion est encore 0.858.
Les différences apparaissent uniquement, dans la catégorie des questions qui étaient au niveau 8; l'implication de ces questions dans l'Arbre-4 était [[P17b → (P17a → P16] ⇒ P15], tandis que dans l'Arbre 4', dans cette classe apparaissent davantage des implications significatives [P17b ⇒ (P17a → P16)] ⇒P15] ⇒ P12.
Les élèves fictifs *Problème* et *Algèbre*, contribuent avec 45 élèves de l'échantillon, à la formation de cette catégorie des questions et ils en sont aussi élèves typiques, avec 67 élèves.
L'analyse de la typicalité des classes significatives, montre que, à tous les niveaux de la classification cohésitive (**significatifs** ou pas), apparaissent les responsabilités des élèves fictifs; les caractéristiques *Problème* et *Algèbre* se présentent généralement ensemble (niveaux, 2, **3 6**, **8**, 10, **11** et 18) ; de même pour *Calcul*, et *Exercice* (niveaux **1**, 4, 7, 9, 13, 15), ce que nous avions déjà remarqué dan l'analyse des similarités.

5 Conclusions

La description de la façon dont nous avons fait fonctionner des *élèves fictifs* et des *variables supplémentaires*, dans les différentes analyses réalisées avec CHIC, nous a permis de pointer des informations nouvelles ou complémentaires que ces analyses ont apportées à l'analyse statistique classique (moyen, médiane, variance, etc) des données de l'échantillon; nous allons en résumer à mode des conclusions
Tandis que les résumés statistiques de la variable *TOTAL* ont montré des différences significatives entre les niveaux de la réussite globale des étudiants en fonction des types d'études, le calcul de la contribution des variables supplémentaires, montre que la variable *TITUL* (type d'études) ne contribue pas à la formation des classes significatives des questions dans les analyses de la similarité et de la cohésion des variables du questionnaire. Elle n'est pas non plus, variable typique de ces classes.
L'unique variable supplémentaire qui contribue à la formation de la classe la plus significative de l'analyse des similarités, est la variable MATPREV "*avoir suivi récemment un cours de préparation de Mathématiques*". Elle en est également typique (avec un risque d'erreur de 0,029) . Mais c'est la variable

Troisièmes Rencontres Internationales – Terzo Convegno Internazionale - Third International Conference
A.S.I. Analyse Statistique Implicative – Analisi Statistica Implicativa – Implicative Statistic Analysis
Palermo (Italy) 6-8 Octobre 2005

TIPOACC (*type d'accès à l'Université*) qui est la variable typique de chaque sous-classe de la classe 12 (avec un risque d'erreur très petit : 0,003).
Quant aux *élèves fictifs (e.f.),* nous pensons qu'ils nous ont permis de confirmer la caractérisation des classes significatives de variables construites par les analyses de similarité et de cohésion, en tant qu'élèves qui auraient contribué à la formation de ces classes ou comme des élèves typiques de ces classes. Mais il faut d'abord vérifier que le changement d'occurrences produit par l'introduction de ces élèves dans l'échantillon, ne modifie pas très sensiblement les résultats globaux des analyses abordées.
Dans l'analyse des similarités, l'arbre de classification des variables, avec les élèves fictifs, reste en effet invariable et les élèves fictifs sur le type des connaissances, *Algèbre* et *Calcul* caractérisent les classes significatives.
Tandis que dans l'analyse de la cohésion, apparaissent aussi les élèves fictifs sur le type de tâches, *Problèmes* ou *Exercices*, en nuançant la connaissance en jeu dans les paquets des questions des classes significatives, qui sont presque identiques (avec des élèves fictifs ou pas) à celles de l'analyse des similarités.
Les caractéristiques *Problème* et *Algèbre* apparaissent toujours ensemble et caractérisent de nombreuses classes de variables, dont les plus significatives sont la classe des questions (P17a, P17b, P16, P15) et la quasi-équivalence des questions (P13a, P13b). Les caractéristiques *Calcul* et *Exercices*, se présentent également ensemble, en caractérisant la quasi-équivalence des questions (P4 et P5) et l'implication des questions P14b et P14a.
Seule la classe ordonnée des questions (P13a, P13b, P14a, P14b) qui se trouvent dans la même classe significative, dans l'analyse de la similarité et de la cohésion) est caractérisée par un seul élève fictif, *Algèbre*.
Les analyses de similarité et de la cohésion faites à l'aide des élèves fictifs, semblent donc, nous apporter des informations complémentaires et justes en rapport à un paquet de questions (P4, P5, P13a, P13b, P14a, P14b, P17a, P17b, P16, P15) de la *PROVA*. En conséquence ils ont été outils pertinents dans la fouille des données de notre échantillon .
Mais, l'analyse implicative nous a posé certains problèmes du fait que se sont produits quelques changements dans les implications entre les variables observées dans les données initiales. Ce qui montre sa sensibilité aux changements d'occurrences.
Quelles sont donc les possibilités et les limites, dans l'utilisation de ces élèves fictifs? La question a été posée et nous espérons avoir apporté quelques éléments de réponse, nous attendons aussi les réponses des lecteurs.

Références

Brousseau G. et Lacasta E. (1995), L'analyse statistique des situations didactiques. Actes du Colloque Méthodes d'analyses statistiques multidimensionnelles en Didactique des Mathématiques, ARDM, pp 53-107.

Fonseca C., Gascón J. et Orús P. (2002), Las Organizaciones Matemáticas en el paso de Secundaria a la Universidad. Análisis de los resultados de una Prueba de Matemáticas a los alumnos de 1º de la UJI, Actas Jornadas de la CV. Universitat Jaume I, 2002. Societat d'Educació Matemática de la C.V.

Couturier, R. et Gras R. (1999), Introduction de variables supplémentaires dans une hiérarchie de classes et application à CHIC, Actes des 7èmes Rencontres de la Société Francophone de Classification, Nancy,1999, pp 87-92.

Gras R. (1995), Méthodes d'analyses statistiques multidimensionnelles en didactique des mathématiques, Actes du Colloque Méthodes d'analyses statistiques multidimensionnelles en Didactique des Mathématiques. ARDM, pp 53-107.

Gras, R et al (1996), L'implication statistique. Nouvelle méthode exploratoire de données, La Pensée Sauvage.

Lerman I.C. (1981), Classification et analyse ordinale des données, Dunod.

Troisièmes Rencontres Internationales – Terzo Convegno Internazionale - Third International Conference
A.S.I. Analyse Statistique Implicative – Analisi Statistica Implicativa – Implicative Statistic Analysis
Palermo (Italy) 6-8 Octobre 2005

Orús P et Groupe DIDENMAT (2004), Evaluación inicial de los conocimientos matemáticos de los alumnos de primero de la UJI, Actas del I Congreso de la Red Estatal de Docencia Universitaria, 2003. Publicacions de la Universitat Jaume I.

Orús P. et Pitarch I. (2000), Utilisation didactique des tableaux des données et du logiciel CHIC à l'école élémentaire. Actes des Journées sur La fouille dans les données par la méthode d'analyse statistique implicative, École Polytechnique de l'Université de Nantes 2000, pp.85-98.

Annexe 1

Questions de la *Prova* 2003/04, présentées ici sous forme de tableau.

1. Compras una camisa que marca 4000 ptas. y te hacen un descuento del 15%. Calcula lo que tendrás que pagar por la camisa. 90%	**2.** Busca soluciones del sistema de ecuaciones $\begin{cases} 2x + y = 1 \\ 3x + 2y = 3 \end{cases}$ 85,4%	**3.** Representa gráficamente la función $t(p) = 4p - p^2$ 43,6%
4. Calcula la derivada de la función $f(x) = \frac{5}{(3x-2)^2}$. 33,8%	**5.** Calcula la integral definida (donde x es la variable de integración y a es una constante): $\int_1^3 2axdx$ 36,7%	**6. a.** En la resolución de una ecuación llegas a la expresión **0x= 8**, ¿cómo interpretas este resultado? 33,5%
6. b. ¿Y si llegas a la expresión **0 x = 0**? 15,2%	**7.** Calcula el mínimo común múltiplo de 280 y 350. 39,1%	**8.** Una empresa tiene unos ingresos de $I(x) = 50x - x^2$ dólares, donde x representa las unidades producidas, y unos costes de $C(x) = 38x + 20$ dólares. ¿Cuántas unidades hay que producir para obtener beneficios? 20,3%
9.Las funciones $f(x) = 3x^4 + x$ y $g(x) = x^3 - 100x^2$ tienden a cero cuando x tiende a cero. Calcula el límite de la función cociente: $\frac{f(x)}{g(x)}$ cuando x tiende a cero. 15,8%	**10.** ¿Cómo compararías las siguientes ofertas de trabajo de repartir propaganda electoral? (a) Te pagan una cantidad fija de 50.000 ptas. más 10 ptas. por cada papeleta depositada en un buzón. (b) Te pagan 30.000 ptas. fijes más 15 ptas. por papeleta. 29,3%	**11.**Calcula la derivada de la siguiente función respecto de la variable x: $f(x) = 8s^x$ (donde s es un número real) **10,3%**
12. ¿Se pueden considerar como soluciones de la ecuación $\sqrt{3x-8} = 4 - \sqrt{x}$ los siguientes valores, $x = 4$ y $x = 36$?. Razona la respuesta. 47,7%	**13. a.** El volumen $C(t)$ de agua que mana de un grifo (en litros) viene dado por una función afín respecto del tiempo t (en segundos). Si en el primer segundo el agua recogida es de 3 litros, en el segundo es de 5 litros y en el tercero es de 7 litros, **a)** ¿Cuál es el volumen de agua recogida en un instante cualquiera t? 25,7%	**13. b. b)** ¿Cuál es el volumen de agua recogido en una hora? 19,1%
14. a. La cantidad de miles de unidades vendidas de un producto, $V(t)$, después de transcurridos t años de su lanzamiento comercial, viene dada por la función: $V(t) = 30 \cdot e^{\frac{-1.8}{t}}$ Calcula el límite de $V(t)$ cuando t tiende a infinito. 16,5%	**14. b.** Interpreta el resultado anterior en términos de ventas del producto en cuestión. **3,2 %**	**15.**Expresa en lenguaje algebraico el enunciado siguiente: "*El producto de tres números impares consecutivos es igual a* 1287" 26,5%

Troisièmes Rencontres Internationales – Terzo Convegno Internazionale - Third International Conference
A.S.I. Analyse Statistique Implicative – Analisi Statistica Implicativa – Implicative Statistic Analysis
Palermo (Italy) 6-8 Octobre 2005

16. La gráfica de $f(x) = (x-1)(x+1)(x+3)$, ¿en qué puntos corta al eje de las x? 34,2%	**17. a.** Dibuja las curvas $x^2 + y^2 = 4$, $y = 2 - x$ sobre los mismos ejes de coordenadas. **6,8 %**	**17. b.** Encuentra, de manera algebraica, los puntos donde se cortan. **9,0 %**

Tab -4 : Questions avec pourcentages de réussite, de la Prova 2003/04

Summary

Supplementary variables and «fictitious» students can help in the labour of data mining with CHIC. We show here a didactical use of this tools and their contribution in an analysis of results, with CHIC, obtained from a test of initial mathematical knowledge to students in their first course of university studies in Spain (Jaume I de Castelló University).