

**Implication statistique et recherche en didactique.
Utilisation d'un outil non symétrique d'analyse de
données pour l'interprétation des résultats d'un
test d'évaluation**

*Statistical implication and educational research. An asymmetrical tool to analyse
the results of an evaluation test*

Maria-Gabriella Ottaviani et Silvia Zannoni



Édition électronique

URL : <http://journals.openedition.org/msh/2845>

DOI : 10.4000/msh.2845

ISSN : 1950-6821

Éditeur

Centre d'analyse et de mathématiques sociales de l'EHESS

Édition imprimée

Date de publication : 1 mars 2001

ISSN : 0987-6936

Référence électronique

Maria-Gabriella Ottaviani et Silvia Zannoni, « Implication statistique et recherche en didactique. Utilisation d'un outil non symétrique d'analyse de données pour l'interprétation des résultats d'un test d'évaluation », *Mathématiques et sciences humaines* [En ligne], 154 | Été 2001, mis en ligne le 10 février 2006, consulté le 23 juillet 2020. URL : <http://journals.openedition.org/msh/2845> ; DOI : <https://doi.org/10.4000/msh.2845>

IMPLICATION STATISTIQUE ET RECHERCHE EN DIDACTIQUE.
 UTILISATION D'UN OUTIL NON SYMÉTRIQUE D'ANALYSE DE DONNÉES
 POUR L'INTERPRÉTATION DES RÉSULTATS D'UN TEST D'ÉVALUATION

Maria Gabriella OTTAVIANI, Silvia ZANNONI¹

RÉSUMÉ – *L'analyse implicative révèle l'existence d'une hiérarchie de difficultés dans la solution d'exercices simples de mathématique proposés à des étudiants de la Faculté des Sciences Statistiques de l'Université de Rome "La Sapienza". La méthode a montré la structure des relations entre couples de variables, la façon qu'elles ont de se lier entre elles et a fait apprécier la capacité explicative des variables de structure. Une comparaison faite avec l'analyse des correspondances multiples montre que l'analyse implicative permet seule de pouvoir expliciter des réseaux de tendance comportementale et cognitive.*

MOTS-CLÉS – *Implication statistique, Graphe implicatif, Arbre hiérarchique, Structure des relations.*

SUMMARY – *Statistical implication and educational research. An asymmetrical tool to analyse the results of an evaluation test. The "implication statistical analysis" is an asymmetric method useful to study data structures, particularly those concerning education. It is used on simple mathematics exercises done by some students of the Faculty of Statistical Sciences in the University of Rome "La Sapienza". It shows the structure of binary variables, their connections to each other and the importance of structure variables to clarify students' difficulties. A comparison with multiple correspondence analysis shows that only the "implication statistical analysis" leads to trees and hierarchic structures useful in cognitive and behavioural studies.*

KEYWORDS – *Statistical implication, Non-symmetric graph, Hierarchic tree, Relation structure*

1. INTRODUCTION

L'analyse implicative ([6], [7], [8], [9]) se propose de donner un sens statistique et donc une mesure aux expressions du type: " si a alors b " dans le cas où l'implication est non stricte, c'est-à-dire dans la situation où l'ensemble A des individus qui vérifient a n'est pas inclus dans l'ensemble B des individus qui vérifient b . Intuitivement, la quasi-implication $a \sqsupseteq b$ est d'autant meilleure que l'inclusion de A dans B est

¹ Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università di Roma " La Sapienza ", Piazzale A. Moro, 5 – 00185 Roma, e-mail: mariagabriella.ottaviani@uniroma1.it, silviazannoni@hotmail.com. Travail effectué dans le cadre du projet de recherche MURST – 1998 prot.9813236997-003, responsable local M. G. Ottaviani.

proche de l'inclusion stricte (qui se réalise lorsque $A \sqsubseteq B = A$), eu égard aux tailles de A et de B ($n_a \sqsubseteq n_b$, étant la situation la plus favorable à l'inclusion). Nous ne donnerons ici que les grandes lignes de la théorie de l'analyse implicative dans le cas de variables binaires et nous renvoyons aux différents documents cités dans la bibliographie, dont une autre approche présentée ([2], [3]).

Pour modéliser le concept d'implication statistique dans le cas de deux variables binaires, R. Gras ([6], [7]) utilise un tableau de contingence 2×2 et examine sur la fréquence absolue des individus qui vérifient à la fois a et non b , c'est-à-dire sur $n_{a\bar{b}}$, étant donné que, dans le cas où $n_a \sqsubseteq n_b$ et $n_b \neq n$, $n_{a\bar{b}} = 0$ si et seulement si l'implication est stricte.

La quantité $n_{a\bar{b}}$ est la réalisation de la variable aléatoire $\text{Card}(X \sqsubseteq \bar{Y})$ qui suit pour une certaine modélisation ([11], et [7] p.30), la loi de Poisson de paramètre $n_a n_{\bar{b}} / n$. Cela permet de calculer la probabilité : $\Pr[\text{Card}(X \sqsubseteq \bar{Y}) \sqsubseteq n_{a\bar{b}}]$ et d'obtenir la nouvelle variable aléatoire :

$$Q(a, \bar{b}) = (n \text{Card}(X \sqsubseteq \bar{Y}) \sqsubseteq n_a n_{\bar{b}}) (n_a n_{\bar{b}})^{1/2}$$

qui suit approximativement la loi normale centrée réduite $N(0, 1)$, l'approximation étant d'autant meilleure que $n_a n_{\bar{b}} / n$ est suffisamment grand et qui pour $n_a n_{\bar{b}} \geq 3n$ est considérée comme bonne. Dans ce cas, la variable aléatoire $Q(a, \bar{b})$ permet de définir l'intensité d'implication :

$$\square(a, b) = 1 - [\Pr Q(a, \bar{b}) \sqsubseteq q(a, \bar{b})]$$

où $q(a, \bar{b})$ est la réalisation empirique de $Q(a, \bar{b})$. L'implication statistique $a \sqsubseteq b$ est alors admissible au niveau de confiance $1 - \square$, si et seulement si : $\square(a, b) \geq 1 - \square$ (en général, $\square = 0,05$ est la valeur de référence).

Du concept d'implication statistique dérivent deux représentations

- le *graphe implicatif*, visualisant la structure induite directement par l'intensité d'implication sur l'ensemble des variables prises deux à deux
- l'*arbre hiérarchique* (dit aussi *cohésitif*), obtenu à partir de l'intensité d'implication, arbre non symétrique où les classes de variables s'agrègent en fonction d'une distance ultramétrique sous-jacente, définie par la cohésion [8].

Dans le but d'analyser en détail la théorie de l'implication et d'en tester la pertinence et la sensibilité pour la recherche en didactique, nous avons disposé d'un questionnaire sur l'utilisation du repère cartésien, questionnaire soumis au début de l'année académique 1999-2000 aux étudiants fréquentant les 4 cours de " Statistique 1 " – enseignement de la première année – à la Faculté de Sciences Statistiques de l'Université de Rome "La Sapienza". En définitive, nous voulions utiliser l'implication statistique pour vérifier l'existence d'une hiérarchie de difficultés dans la solution des exercices simples proposés. Mais nous voulions aussi étudier comment les différentes réponses étaient liées entre elles ; si, par exemple, quelques-unes de celles-ci impliquaient d'autres réponses et avec quelle force. Puisque des tableaux binaires issus

d'un questionnaire peuvent être traités aussi par l'analyse des correspondances multiples, méthode symétrique, nous voulions confronter les résultats des deux analyses pour mettre en évidence leurs différences ou leurs complémentarités au point de vue méthodologique et conceptuel.

2. LE QUESTIONNAIRE

Le questionnaire qui fait l'objet de cette étude se base sur celui proposé par M. Polo dans sa thèse [12] et dans [13]. Celui-ci a été adapté en considérant le programme de mathématiques commun aux différentes filières de formation secondaire en Italie. Les questions posées aux étudiants des universités romaines demandaient (voir Annexe 1) :

- de donner les coordonnées des points d'intersection avec les axes et l'équation d'une droite et d'une parabole dessinées sur un repère cartésien, et
- de dessiner un repère cartésien étant données les coordonnées de 2 sommets d'un triangle représenté sur papier quadrillé.

Les questions choisies font référence à un thème de mathématiques que les professeurs de "Statistique 1" considèrent déjà acquis pendant le *cursus* du secondaire. La connaissance du repère cartésien et de l'équation d'une droite est en effet nécessaire aux étudiants de "Statistique 1" pour construire la représentation graphique des tableaux, le graphique de dispersion d'une série statistique bivariée, pour comprendre la notion de régression linéaire entre deux variables et pour représenter et interpréter la droite de régression.

Des informations supplémentaires, ou de structure ou de descripteurs, ont aussi été demandées aux étudiants concernant :

- 1 le type de "Maturità" (Baccalauréat = BAC) : scientifique, classique ou autre type
- 2 la note finale reçue au BAC ;
- 3 le lieu où le BAC a été obtenu : Rome, dans une autre commune du Lazio, ailleurs
- 4 leur situation universitaire : inscrit en première année, provenant d'une autre faculté, autres situations.

Il s'agit de descripteurs qui peuvent aider à mieux saisir le comportement des étudiants face au questionnaire proposé.

Le questionnaire a été soumis à 226 étudiants de la Faculté de Sciences Statistiques de l'Université de Rome "La Sapienza". Les étudiants interrogés suivaient, en octobre 1999, le cours de "Statistique 1" respectivement pour le "Corso di laurea"² en Sciences Statistiques et Actuarielles (71), en Sciences Statistiques Démographiques et Sociales et en Statistique (64), en Sciences Statistiques et Économiques (91).

² En Italie, le "corso di laurea" est une filière de formation universitaire qui décerne le diplôme universitaire de quatre ans dit "laurea".

3. OBJECTIFS DE LA RECHERCHE

À travers l'analyse des réponses au questionnaire, on voulait atteindre une double finalité.

Du point de vue statistique, en utilisant l'analyse implicative et l'analyse des correspondances multiples, on voulait comparer, sur une application, les résultats de ces deux méthodes toutes deux multidimensionnelles, globales et descriptives. Nous savons qu'elles permettent de prendre en compte le maximum d'informations sur l'ensemble des comportements de réponse de toute une population. Et ceci sans avoir besoin de soumettre une hypothèse à un test à des fins d'invalidation, mais plutôt de faire de la prospective quant à des liaisons encore inapparentes et de dégager une structure à laquelle on pourra attribuer un sens. Toutefois, tandis que l'analyse implicative rend compte d'une structure dynamique en raison de sa qualité non-symétrique, au sein de l'ensemble des comportements et doit dégager des chemins quasi-transitifs orientés par les implications d'un nœud du graphe vers ses successeurs, en revanche l'analyse des correspondances multiples dégage et hiérarchise des facteurs sous-jacents à la structure des comportements, afin d'observer les affinités de certaines variables relativement à ces facteurs et de dégager les variables les plus contributives à chacun d'entre eux.

En ce qui concerne l'utilisation des variables supplémentaires en analyse implicative comme en analyse de correspondances multiples, celles-ci devraient conduire à la découverte du rôle de ces descripteurs dans la structuration des comportements de réponse.

Sur le plan du contenu mathématique, l'analyse du questionnaire cherchait à vérifier des hypothèses d'acquisition posées *a priori* sur les connaissances des étudiants universitaires au début de la première année relativement au repère cartésien, à la droite et à la parabole.

Des situations différentes peuvent se présenter mais, en général, on peut penser qu'une réponse correcte à la deuxième question, c'est-à-dire la détermination de l'équation de la droite ou de la parabole, implique une réponse exacte à la première question, à savoir la détermination correcte des coordonnées des points d'intersection des deux courbes avec les axes. En effet, les équations de la droite et de la parabole ne peuvent être déterminées correctement que si les coordonnées des points d'intersection des deux courbes avec les axes ont été déterminées correctement, en prenant en compte le fait que pour la parabole, il faut déterminer les coordonnées d'un troisième point situé sur la courbe elle-même. La réciproque n'est pas toujours vraie : les coordonnées des points d'intersection d'une courbe avec les axes peuvent être déterminées mais l'absence de connaissance (ou l'oubli) de l'équation générique ou des erreurs de calcul peuvent conduire à une réponse fautive à la deuxième question. En outre on peut se demander si des obstacles à la lecture de propriétés de la parabole ont une relation avec des obstacles dans le cas de la droite.

La troisième question demande un raisonnement inverse par rapport à la première : dessiner un repère cartésien par rapport à un triangle dont les coordonnées de deux sommets sont données. On peut supposer, alors, l'existence d'un lien entre les réponses à la première question et celles à la dernière. En effet, si un étudiant est capable de déterminer les coordonnées des points d'intersection d'une courbe avec les axes et, donc, s'il a connaissance du fait que la coordonnée y du point d'intersection avec l'axe des abscisses est nulle et la coordonnée x du point d'intersection avec l'axe des ordonnées est également nulle, alors il devrait savoir procéder à l'envers, c'est-à-dire, qu'il devrait savoir dessiner le repère cartésien qui satisfait les conditions posées par la troisième question.

4. L'ANALYSE DU QUESTIONNAIRE

4.1 CODAGE DES RÉPONSES

Les réponses aux trois questions posées ont été transformées en variables que l'on peut traiter sur le plan statistique et prenant en compte différentes situations de réussite et d'échec.

Pour la question 1, les modalités de réponse considérées ont été : réussite (RE), indication d'une seule coordonnée correcte (Uc), échec - il y a au moins une erreur dans la réponse (EC), aucune réponse (NR).

Pour la question 2, le codage est le même, sauf pour la seconde modalité codée PC, dans le but d'indiquer que la réponse est partiellement correcte, autrement dit qu'il n'y a pas d'erreurs, mais la réponse n'est pas complète.

Le codage de la question 3 a aussi été établi en se référant aux modalités RE, PC, EC, NR, mais ici nous avons trouvé intéressant de mettre en évidence les différentes modalités de la réponse partielle à savoir :

- d'une part, la marque des axes est correcte et elle est la seule réponse (Ax)☐
- d'autre part, la marque des axes et de l'unité est indiquée correctement (AxUm)☐
- enfin, la marque des axes et de l'origine est indiquée correctement (AxOr).

On a jugé en outre judicieux de coder la réponse 3 en suivant l'approche de M. Polo dans sa thèse [12], c'est-à-dire en introduisant les trois modalités suivantes : il y a la marque des axes (AC3), la marque de l'origine (OR3), la marque des unités (UM), sans référence au fait que la réponse soit correcte ou non.

Le codage des informations supplémentaires ne pose pas de problèmes particuliers, sauf la variable *note au baccalauréat* qui a été découpée en trois classes.

On trouvera le codage complet du questionnaire en Annexe 2.

4.2 L'ANALYSE IMPLICATIVE

L'analyse implicative entre variables et entre classes de variables n'a concerné, dans cette étude, que des variables binaires. Ainsi chaque modalité est considérée comme une variable binaire qui prend la valeur 1 quand elle se manifeste chez un étudiant et 0 dans le cas contraire, de sorte que l'on obtient une matrice présence-absence de dimension $n \times m$, où n est le nombre des sujets (ici $n = 226$) et m est le nombre des variables binaires. Dorénavant le terme variable fera référence à une variable binaire.

Comme il a été déjà dit, à côté des questions de mathématiques, nous disposons également des informations de structure ou descripteurs de la population étudiante. Les variables binaires obtenues à partir de ces informations ont été utilisées tout d'abord comme variables supplémentaires. C'est-à-dire qu'elles n'ont pas participé au traitement qui conduit au graphe, ni à celui qui conduit à l'arbre, mais ont été utilisées pour déterminer les catégories d'étudiants qui offraient la contribution la plus forte à la constitution d'un chemin du graphe ou d'une classe de variables. Pour aider à la compréhension des résultats obtenus, nous avons ensuite analysé toutes les variables en les considérant comme principales. En conséquence, la première application a utilisé 18 variables principales et 15 variables supplémentaires. Quant à la seconde application elle a traité 33 variables toutes principales. Les variables utilisées pour l'analyse implicative sont indiquées dans l'Annexe 2.

Parmi les informations qui sont fournies en sortie du logiciel C.H.I.C. [4], le graphe implicatif et l'arbre hiérarchique sont particulièrement importants. Le premier donne une vision immédiate et une lecture aisée des relations d'implication entre un grand nombre de couples de variables ; le second met en évidence graphiquement les implications entre variables ou entre classes de variables.

Dans un premier temps, l'analyse des réponses a été conduite séparément pour le 4 "corsi di laurea" ; toutefois, comme des différences importantes entre les comportements de réponse des étudiants des 4 filières n'apparaissaient pas, l'analyse a été conduite sur la totalité des étudiants.

La figure 1 montre le graphe implicatif que l'on obtient, au seuil de 5 %, en utilisant les 18 variables principales et les 12 premières variables supplémentaires. Elle est une figure complexe qui représente, en même temps, les variables utilisées, les liaisons implicatives entre elles et une échelle qui donne la fréquence de la " présence " de chaque variable, c'est-à-dire que cette figure donne la vision globale des modalités de réponses au questionnaire, de leur enchaînement et de leur difficultés. Elle montre, dans notre propre lecture, deux sous-graphes (indiqués par les flèches 1 et 2) et deux liens implicatifs séparés entre deux couples de variables (indiqués par les flèches 3 et 4). Il s'agit des implications de la variable NRR2 vers NRP2 et de la variable PCp2 vers PCr2.

En ce qui concerne le comportement des étudiants, ces liens signifient que, en général, ceux qui n'ont pas donné l'équation de la droite ne donnent pas celle de la parabole, probablement en raison de la moindre familiarité des étudiants avec les

fonctions du second degré. En outre, l'étudiant qui a donné une réponse partiellement correcte pour la parabole a fourni également une réponse partiellement correcte pour la droite.

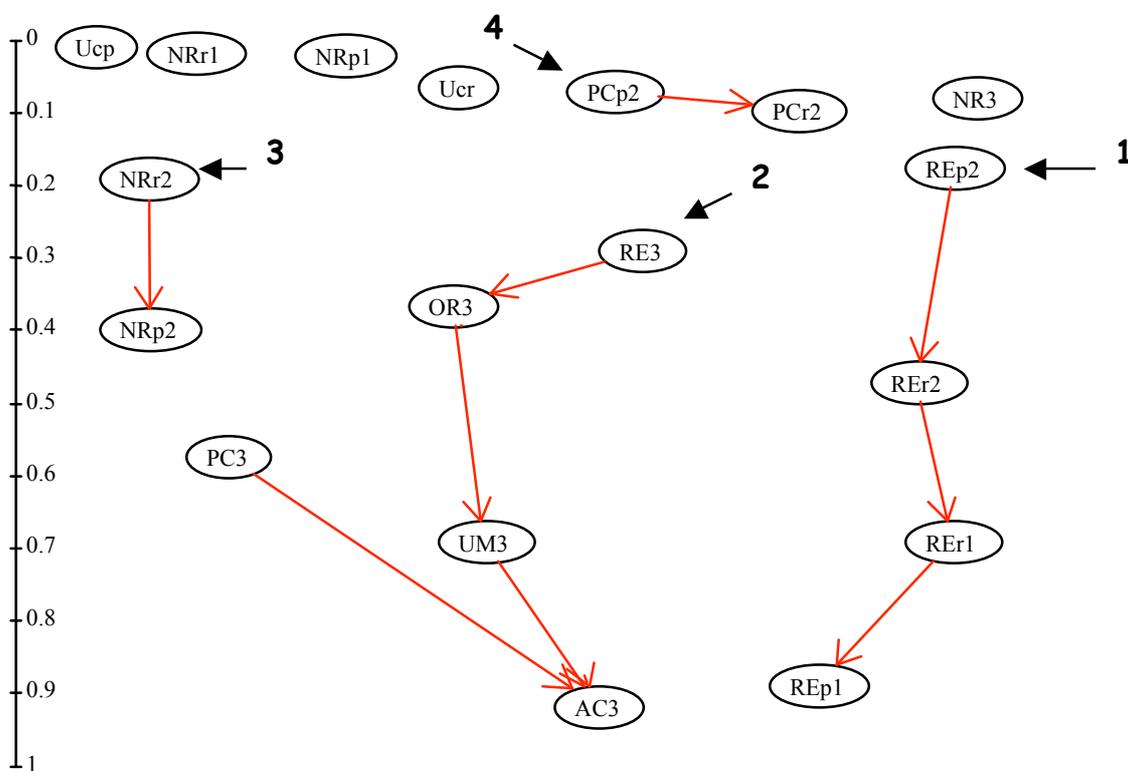


Figure 1. Graphe implicatif pour la totalité des étudiants au seuil de 5 % (18 variables principales)

À propos des deux sous-graphes 1 et 2 on peut observer que :

- le premier met en évidence une hiérarchie des réponses à la deuxième question et le lien de celles-ci avec les réponses à la première question. L'étudiant qui a donné correctement l'équation de la parabole répond correctement pour la droite et répond bien à la première question, c'est-à-dire détermine les coordonnées exactes des points d'intersection de la droite et de la parabole avec les axes. En effet, comme il était prévisible, sauf s'il y a erreur de calcul, dans le cas seulement où les coordonnées des points d'intersection avec les axes sont correctes, il est possible d'écrire correctement l'équation d'une courbe qui passe par ces points ; les tâches et les connaissances sont concomitantes au cours de l'enseignement comme nous le vérifions ☐
- le deuxième sous-graphe, entièrement relatif à la troisième question, montre que la réponse exacte suppose l'indication de l'origine, de l'unité et des axes et, en outre, il montre que ces trois éléments ont une difficulté décroissante. Nous y reviendrons plus loin.

L'existence d'une hiérarchie des difficultés entre les questions posées est plus visible si on subdivise le graphe par des lignes horizontales qui coupent l'échelle à gauche à 25%, 50% et 75% (figure 2).

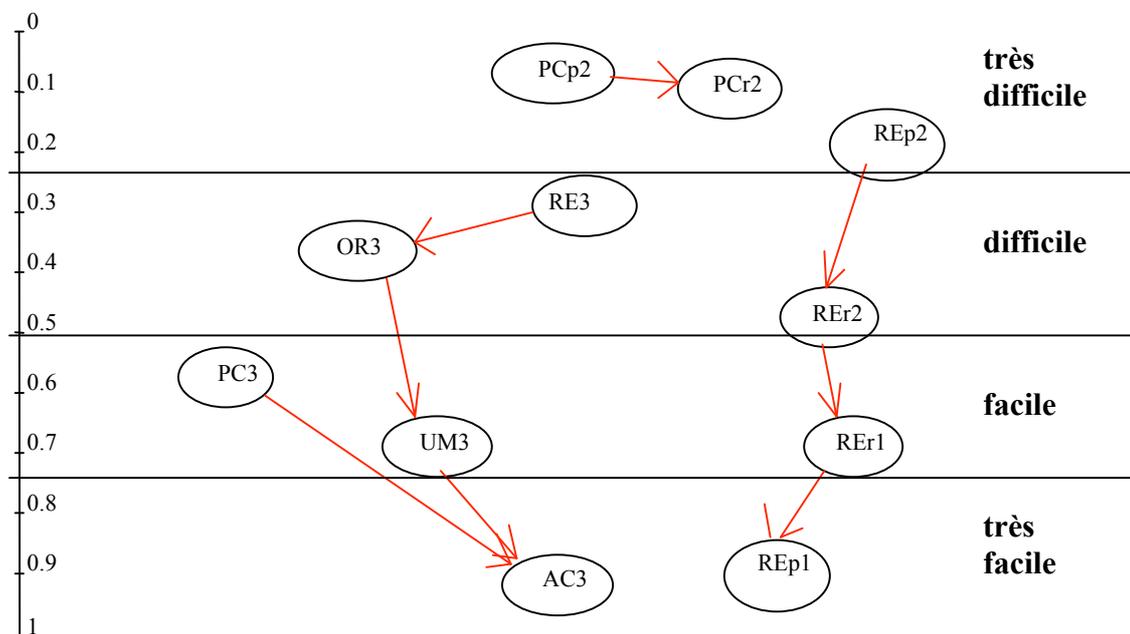


Figure 2. Graphe implicatif pour la totalité des étudiants au seuil de 5 % et échelle des difficultés (18 variables principales)

De cette façon, on produit une échelle des difficultés à partir de “ très difficile ” pour arriver graduellement jusqu’à la bande “ très facile ”. Alors, la position de la variable REp2 dans la première bande en haut témoigne de la difficulté que les étudiants ont rencontré dans l’écriture de l’équation de la parabole. La difficulté pour écrire l’équation de la droite, mise en évidence par la variable REr2, est moindre. Elle se trouve voisine de 50 %. Dans la bande “ facile ” se trouve la variable REr1 ; elle se réfère à la réponse exacte à la question sur les coordonnées des points d’intersection de la droite avec les axes. À cette question, 70 % environ des étudiants ont répondu correctement. Cet exercice, apparemment facile, a produit, au contraire, des difficultés plus grandes que celles que l’on attendait dans un ensemble de nouveaux diplômés. L’examen des questionnaires a montré que les erreurs sont liées à la non-considération du signe moins de l’abscisse du point d’intersection avec l’axe des x et révèle ainsi la persistance de l’obstacle épistémologique des nombres négatifs.

Le graphe implicatif est à même donc de valider ce que l’expérience nous permet d’exprimer *a priori*, à savoir qu’il est plus difficile pour un étudiant de trouver l’équation d’une parabole que celle d’une droite. En outre, il fournit un réseau orienté de comportements qui enchaîne et rend visibles les difficultés des étudiants.

L'arbre hiérarchique (figure 3) confirme et complète l'analyse du graphe implicatif.

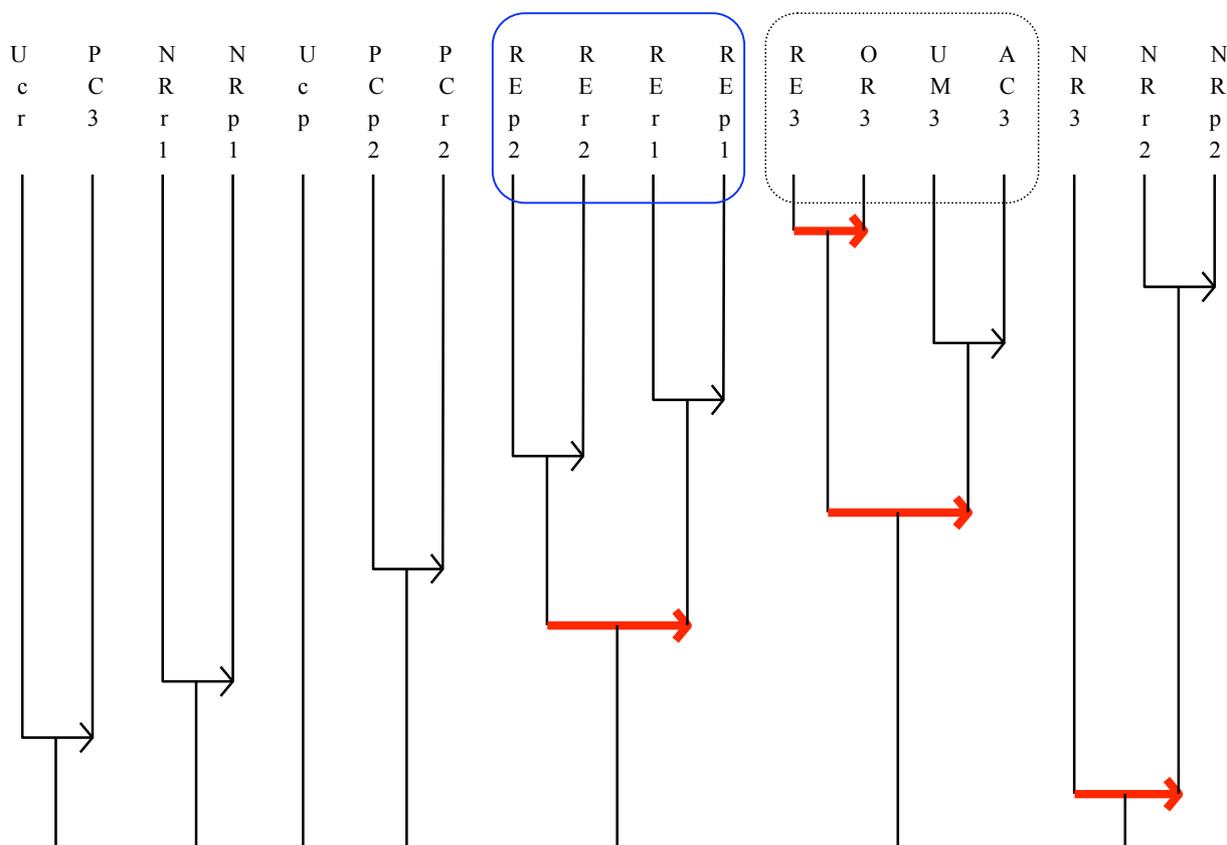


Figure 3. Arbre hiérarchique pour la totalité des étudiants (18 variables principales)

Deux classes des variables sont intéressantes□

- la première contient les variables qui représentent les réponses exactes aux questions sur la droite et sur la parabole□
- la deuxième contient les variables concernant la question sur le dessin du repère cartésien.

On voit que les deux classes ne se lient pas entre elles, ni *a fortiori* à des niveaux inférieurs de la cohésion, confirmant ainsi ce qui a été dit précédemment en ce qui concerne l'absence d'un lien implicatif entre les réponses aux deux premières questions et les réponses à la troisième. Ce constat montre bien que le concept de “ repère ” n’est pas, sur le plan cognitif, le fondement des réponses positives aux premières questions puisque le transfert n’est pas opérant dans la troisième. Le facteur de réussite tient donc plutôt à la nature algorithmique des tâches et à la familiarisation, même relative, des étudiants avec ces situations classiques dans l’enseignement secondaire.

L'analyse du groupe optimal et de la catégorie contributive à la constitution d'une classe montre que le BAC scientifique a une contribution importante à la solution des exercices sur la droite et sur la parabole. Ce n'est pas le cas pour la troisième question devant laquelle aucun BAC n'a une contribution importante, c'est-à-dire que tous les étudiants se trouvent dans les mêmes conditions en raison, sans doute, de l'absence d'apprentissage spécifique à ce type de problème.

Le graphe (figure 4) que l'on obtient en utilisant toutes les 33 variables en tant que principales, au seuil de 5 %, confirme ce résultat. En outre une nouvelle variable entre en jeu dans le sous-graphe concernant les questions sur la droite et sur la parabole une note supérieure à 90/100 (>90).

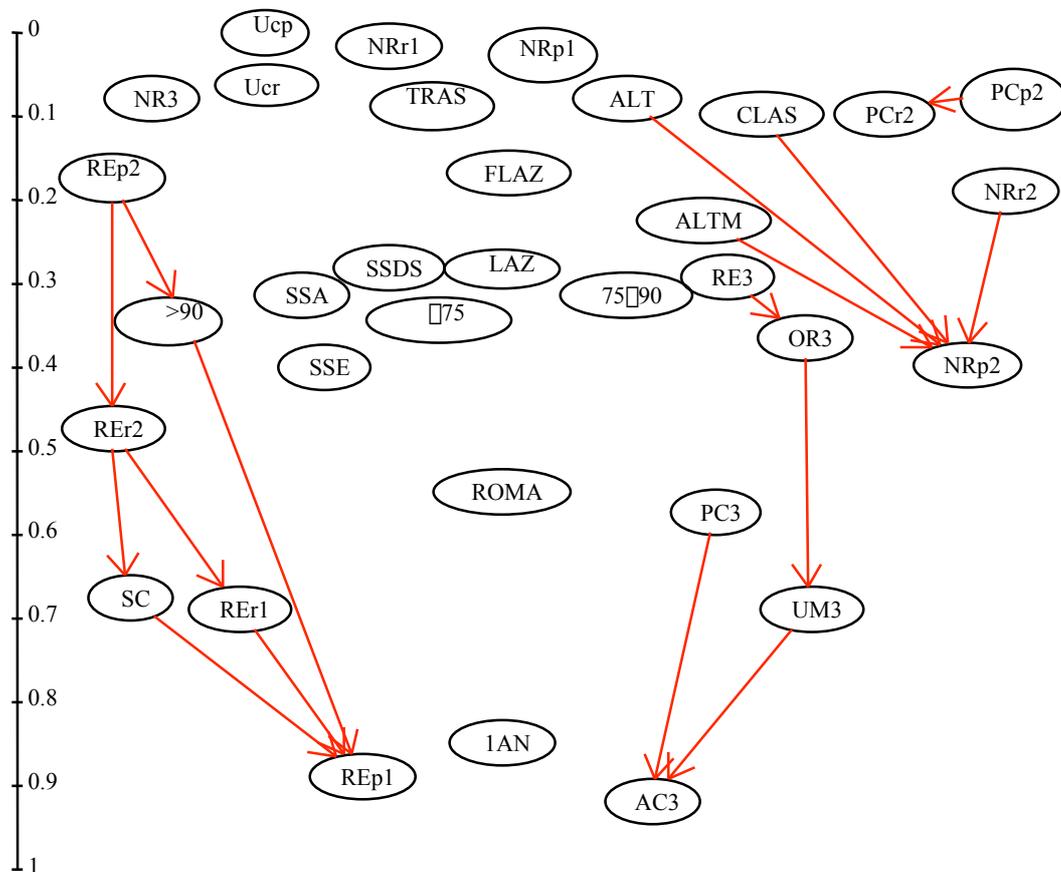


Figure 4. Graphe implicatif de toutes les variables en tant que principales, au seuil de 5 %

Les variables "BAC classique" (CLAS) et "Autre BAC" (ALTM) impliquent la non-réponse pour l'équation de la parabole. Ce qui signifie qu'en général, les étudiants qui ont un BAC non scientifique n'ont pas su écrire l'équation qui, bien qu'étant au programme de leur cursus, exige un niveau de connaissance et de savoir faire qu'ils n'ont pas acquis. En outre, dans le sous-graphe concernant la troisième question (dessin

du repère cartésien) les variables se référant au BAC n’entrent pas en jeu, confirmant l’absence d’un lien entre le type de diplôme et les réponses à la troisième question.

L’introduction en tant que variables principales des variables de structure apporte des changements dans l’arbre hiérarchique (figure 5).

Par rapport à l’arbre de la figure 3 on observe que la classe qui rassemblait les réponses exactes aux deux premières questions se partage en deux (voir figure 5).

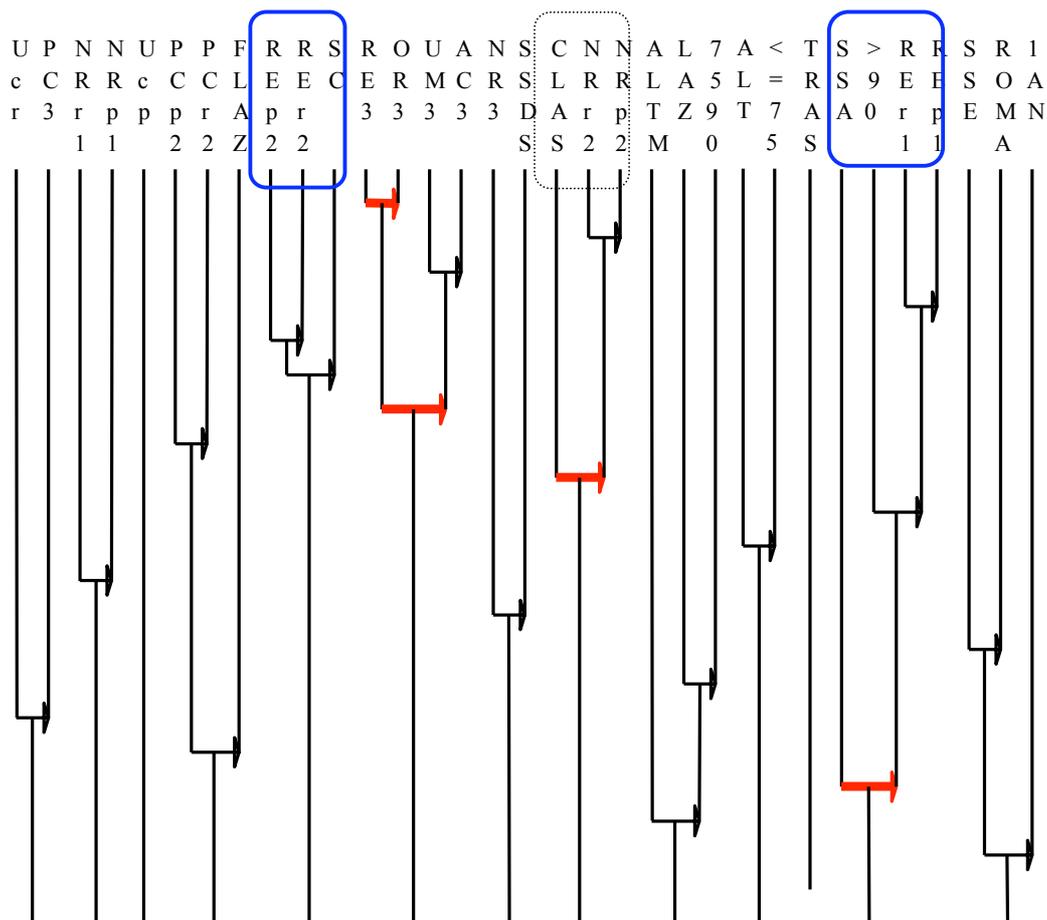


Figure 5. Arbre hiérarchique de toutes les variables en tant que principales

- la note au baccalauréat supérieure à 90/100 implique l’indication correcte des coordonnées des points d’intersection de la droite et de la parabole avec les axes ; on note aussi une intensité d’implication faible avec la filière SSA ;
- les réponses exactes pour l’équation de la droite et de la parabole impliquent le BAC scientifique.

La classe de variables concernant la troisième question reste inchangée, insensible, comme nous l'avons déjà dit, aux apprentissages antérieurs et beaucoup plus liée à la créativité et l'initiative.

L'analyse des réponses faite en agrégeant les quatre "corsi di laurea", montre que :

- en général, les étudiants des quatre filières ont rencontré les mêmes difficultés dans la solution des exercices proposés ;
- nous avons confirmé l'existence d'une hiérarchie de difficulté dans la solution des exercices proposés et l'implication des réponses concernant la parabole (questions plus difficiles) sur les réponses concernant la droite (questions plus faciles) ;
- l'"enchaînement" des réponses, visible dans le graphe implicatif, a permis de montrer des processus de solution des problèmes formant des réseaux orientés porteurs de sens ;
- nous avons observé que les étudiants face à une question "non standard" comme la dernière se comportent de la même façon, ceci donc sans lien avec la filière de formation secondaire suivie. Ce n'est pas le cas pour les questions les plus usuelles (les deux premières) pour lesquelles les réussites semblent liées à la filière scientifique et à une très bonne note à l'examen en fin de secondaire (>90).

4.3 L'ANALYSE DES CORRESPONDANCES MULTIPLES

L'analyse implicative utilisée s'est révélée un outil de synthèse des données montrant à la fois l'existence d'une relation entre couples de variables binaires et la façon qu'elles ont de se lier entre elles. L'approche a permis de faire apprécier la capacité explicative des variables supplémentaires.

Pour comparer les résultats obtenus par cette méthode asymétrique avec ceux produits par l'analyse des correspondances multiples (ACM), nous avons analysé un tableau de contingence de Burt, ce qui se ramène à l'analyse d'un nuage de p points-modalités dans R^p , où p est le nombre total des modalités, ou variables binaires utilisées. Dans ce cas, les variables actives sont au nombre de 22 et elles sont indiquées en Annexe 2. Il s'agit des modalités de réponse aux questions de mathématiques posées aux étudiants.

La procédure CORRESP de SAS fournit les valeurs propres et les coordonnées factorielles des points modalités et elle permet leur représentation sur les plans factoriels.

Les deux premiers axes représentent 22,75 % de l'inertie totale (taux d'inertie modifié : 76,2 %).

L'information retenue dans l'analyse est :

- axe 1 (facteur 1) 12,05 % (taux d'inertie modifié : 47,36)
- axe 2 (facteur 2) 10,70 % (taux d'inertie modifié : 28,84).

L'axe 1 oppose les réussites aux échecs ; l'axe 2 oppose les réussites partielles aux non-réponses (figure 6). La figure 6 comparée à l'arbre hiérarchique (figure 3), montre que quatre classes de modalités persistent :

1. non-réponse à la première question : droite et parabole (NRr1, NRp1) ;
2. non-réponse à la deuxième question : droite et parabole et à la troisième question (NRr2, NRp2, NR3) ;
3. réussite à la première et à la deuxième question : droite et parabole (REr1, REp1, REr2, REp2) ;
4. les modalités de la troisième question qui concernent la réussite totale ou partielle sont très proches de la classe 3.

Les modalités qui concernent toutes les autres réussites partielles forment une autre classe (Ucr, Ucp, PCr2, PCp2).

Les classes observées sur la figure 6 sont confirmées par une classification ascendante hiérarchique avec agrégation par la distance moyenne faite en utilisant les deux premières coordonnées factorielles.

Nous avons projeté les variables de structure sur le plan des facteurs 1 et 2, mais la qualité de représentation de chaque modalité sur ce plan est très faible. Les modalités les mieux représentées sont : BAC scientifique et une bonne note au BAC. Ces modalités sont placées, près des réussites. La modalité ALTM qui code les BAC ni scientifiques, ni classiques est placée entre échecs et non-réponses, s'opposant ainsi à BAC scientifique et BAC (figure 6).

L'analyse des correspondances multiples a permis de dégager les deux facteurs réussite/échec, réussite partielle/non-réponse, de reconstruire les classes constituées également par l'analyse implicative et de parvenir à une interprétation presque semblable du rôle des variables supplémentaires. Cependant, l'ACM ne présente pas la même capacité à structurer les comportements des étudiants. En particulier, elle ne permet pas de mettre en évidence la différence profonde qui sépare les performances respectives aux questions 1 et 2 d'une part et à la question 3 d'autre part. Elle est donc moins révélatrice que la méthode implicative sur le plan de la cognition, qualité exploratoire pourtant attendue par le didacticien et le cognitiviste.

CONCLUSION

L'analyse implicative est utile dans cette étude des connaissances des étudiants de première année de la Faculté de Sciences Statistiques de Rome "La Sapienza", cours de "Statistique 1".

L'apport le plus précieux de cette analyse par l'implication réside dans l'enchaînement entre modalités de réponses successives. Les modalités relatives à ces classes sont les mêmes que celles du plan des facteurs 1 et 2 obtenu par l'analyse des correspondances multiples. Mais dans ce plan, elles n'y présentent pas de liens orientés. Encore que la possibilité de croiser deux variables binaires, à la base de l'implication, n'impose pas de contraintes sur le choix des modalités utilisées pour coder chaque question. Les conditions qui sont à la base d'un tableau disjonctif complet ne sont pas nécessaires pour appliquer l'analyse implicative, ce qui lui donne une plus grande flexibilité. En ce qui concerne les variables supplémentaires, l'analyse implicative qui utilise toutes les variables en tant que principales, a donné des résultats semblables à ceux de l'ACM lorsque les variables actives servent à calculer les axes et que les variables supplémentaires sont utilisées pour les interpréter. Mais l'analyse implicative permet seule de pouvoir expliciter des réseaux de tendance comportementale et cognitive.

Remerciements □ Nous remercions les deux rapporteurs anonymes qui ont lu attentivement notre travail et nous ont permis de l'améliorer ainsi que Régis Gras pour l'aide qu'il nous a apportée dans la clarification de la méthode et J.-M. Bernard pour la pertinence de ses observations.

BIBLIOGRAPHIE

- [1] BAILLEUL, M., “ Mise en évidence de réseaux orientés de représentations dans deux études concernant des enseignants stagiaires en IUFM ”, *Actes des Journées sur la fouille dans les données par la méthode d'analyse statistique implicative*, IUFM de Caen 23-24 Juin 2000, R. Gras et M. Bailleul édit., Rennes, 2000.
- [2] BERNARD, J.-M., CHARRON, C., “ L'analyse implicative bayésienne, une méthode pour l'étude des dépendances orientées. I □ données binaires ”, *Mathématiques, Informatique et sciences humaines* 134, 1996, p. 5-38.
- [3] BERNARD, J.-M., POITRENAUD, S., “ L'analyse implicative bayésienne multivariée d'un questionnaire binaire □ quasi-implications et treillis de Galois simplifié ”, *Mathématiques, Informatique et sciences humaines* 147, 1999, p. 25-46.
- [4] COUTURIER, R., “ Traitement de l'analyse statistique implicative dans CHIC ”, *Actes des Journées la fouille dans les données par la méthode d'analyse statistique implicative*, R.Grass et M.Bailleul éds., 2001, p. 33-50.
- [5] DROESBEKE, J.-J., *Eléments de statistique*, Ellipses, 1992.
- [6] GRAS, R., *Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques*, Thèse d'état, Université de Rennes I, 1979.
- [7] GRAS, R. et coll., *L'implication statistique, nouvelle méthode exploratoire de données*, La Pensée Sauvage Éditions, 1996.

- [8] GRAS, R., KUNTZ, P., BRIAND, H., “ Les fondements de l'analyse statistique implicative et quelques prolongements pour la fouille de données ”, *Mathématique et Sciences humaines* 154-155, 2001, p. 9-29.
- [9] GRAS, R., LARHER, A., “ L'implication statistique, une nouvelle méthode d'analyse de données ”, *Mathématiques, Informatique et Sciences humaines* 120, 1992, p. 5-32.
- [10] LEBART, L., MORINEAU, A., PIRON, M., *Statistique exploratoire multidimensionnelle*, Paris, Dunod, 1995.
- [11] LERMAN, I.C., GRAS, R., ROSTAM, H., “Élaboration et évaluation d'un indice d'implication pour des données binaires”, I et II, *Mathématiques et Sciences humaines* 74, 1981, p. 5-35 et 75, 1981, p. 5-47.
- [12] POLO, M., *Le repère cartésien dans les systèmes français et italien* — étude didactique et application de méthodes d'analyse statistique multidimensionnelle. Thèse de l'Université de Rennes 1, 1996.
- [13] POLO, M., “ Traitement de résultats d'un questionnaire portant sur la représentation graphique présenté en France et en Italie ”, Actes du colloque *Méthode d'analyses quantitatives multidimensionnelles en didactiques des mathématiques*, IUFM de Caen, 27-28-29 janvier 1995, R. Gras éd., IRMAR, Rennes, 1995.
- [14] ROUANET, H., LE ROUX, B., *Analyse des données multidimensionnelles*, Paris, Dunod, 1993.

ANNEXE 1

Filière du BAC : CLASSIQUE ; SCIENTIFIQUE ; AUTRE FILIÈRE

Note au baccalauréat : /

Lieu où le baccalauréat a été obtenu :

- ROME
- AURE COMMUNE DU LAZIO
- AILLEURS

Situation universitaire :

- 1° ANNÉE
- PROVENANT D'UNE AUTRE FACULTÉ
- AUTRE SITUATION

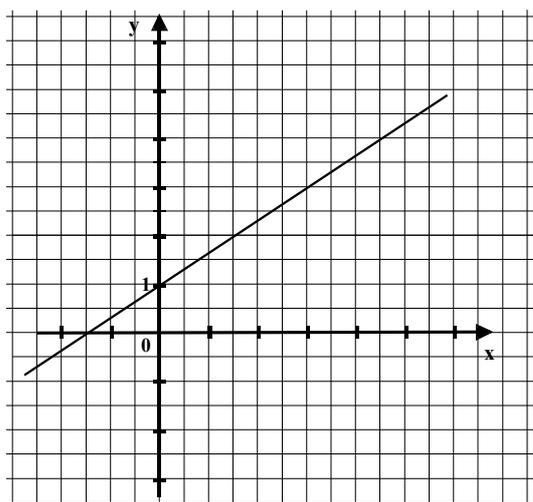
LE REPÈRE CARTÉSIEN

1° UTILISATION

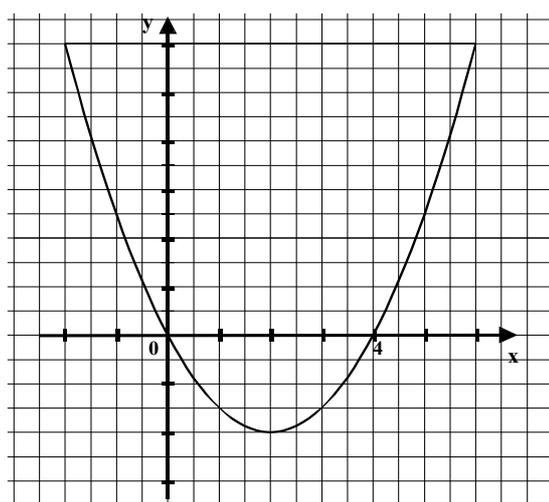
En utilisant les informations fournies par le repère cartésien,

a) Écrire les coordonnées des points d'intersection avec les axes de la droite R et de la parabole P.

droite R



parabole P



Coordonnées des points d'intersection des axes avec la droite R :

b) Écrire les équations des deux courbes.
 - équation de la droite R

Coordonnées des points d'intersection des axes avec la parabole P :

- équation de la parabole P :

2° UTILISATION

Dessiner un repère cartésien tel que les coordonnées des sommets A et B du triangle de la figure près sont $A = (2;0)$ et $B = (0;1)$

ANNEXE 2

Codage du questionnaire

QUESTION 1	Code	Variables utilisées	
		Analyse implicative	ACM
Réussite à la question 1 pour la droite	REr1	*	*
Il y a indication d'une seule coordonnée pour la droite et elle est correcte	Ucr	*	*
Echec à la question 1 pour la droite	Ecr1		*
La réponse n'est pas donnée pour la droite	NRr1	*	*
Réussite à la question 1 pour la parabole	REp1	*	*
Il y a indication d'une seule coordonnée pour la parabole et elle est correcte	Ucp	*	*
Echec à la question 1 pour la parabole	Ecp1		*
La réponse n'est pas donnée pour la parabole	NRp1	*	*
Question 2			
Réussite à la question 2 pour la droite	REr2	*	*
Réponse partiellement correcte pour la droite	PCr2	*	*
Echec à la question 2 pour la droite	Ecr2		*
Aucune réponse à la question 2 pour la droite	NRr2	*	*
Réussite à la question 2 pour la parabole	REp2	*	*
Réponse partiellement correcte pour la parabole	PCp2	*	*
Echec à la question 2 pour la parabole	Ecp2		*
Aucune réponse à la question 2 pour la parabole	NRp2	*	*
Question 3			
Réussite à la question 3	RE3	*	*
Réponse partiellement correcte à la question 3	PC3	*	
<i>Le marque des axes est correcte et elle est la seule réponse</i>	Ax		*
<i>Marque des axes et de l'unité indiquée correctement</i>	AxUm		*
<i>Marque des axes et de l'origine indiquée correctement</i>	AxOr		*
Echec à la réponse 3	EC3		*
Aucune réponse à la question 3	NR3	*	*
Il y a la marque des axes	AC3	*	
Il y a la marque de l'unité	UM3	*	
Il y a la marque de l'origine	OR3	*	

Informations supplémentaires	Code	Variables utilisées	
		Analyse implicative	ACM
BAC classique	CLAS	*	*
BAC scientifique	SC	*	*
Autre BAC	ALTM	*	*
BAC obtenu à Rome	ROM A	*	*
BAC obtenu dans une autre commune du Lazio	LAZ	*	*
BAC obtenu en dehors du Lazio	FLAZ	*	*
Aucune indication du lieu	AL		*
Inscription à la première année	1AN	*	*
Transfert d'une autre faculté	TRAS	*	*
Inscription à une année successive à la première	ALT	*	*
Note au baccalauréat ≤ 60	≤ 75	*	*
75---190	75-90	*	*
90---1100	>90	*	*
Aucune note	AN		*
Sciences Statistiques et Actuarielles	SSA	*	*
Sciences Statistiques, Démographiques et Sociales, Statistiques	SSDS	*	*
Sciences Statistiques et Economiques	SSE	*	*