

# L'analyse statistique implicative vectorielle ; une application au domaine de la description personologique en passation répétée.

**Daniel Pasquier\*, Régis Gras\*\*, Raphaël Couturier\*\*\***

\*15, rue du Grand Carré F-45800 Saint-Jean de Braye  
Unité de Recherche « Psychopathologie Clinique » (URPC), Tunis

dpasquier@avenireentreprise.fr

<http://www.avenireentreprise.fr>

\*\*Equipe Connaissances et Décision, Laboratoire d'Informatique de Nantes Atlantique

Ecole Polytechnique de l'Université de Nantes, UMR 6241  
La Chantrerie BP 60601 44306 Nantes cedex

regisgra@club-internet.fr

\*\*\* Institut Universitaire de Technologie de Belfort,  
BP 527, rue E. Gros, 90016 Belfort cedex

raphael.couturier@iut-bm.univ.fcomte.fr

**Résumé.** Nous nous plaçons ici dans le cadre de la méthode d'analyse de données, l'analyse statistique implicative (A.S.I.). A l'instar de ce que nous avons fait pour passer des variables binaires aux variables numériques ou aux variables-intervalles, nous étendons le champ des traitements aux variables à valeurs vectorielles. Nous établissons un indice permettant de mesurer la qualité d'une règle entre variables vectorielles. Nous traitons des exemples portant, l'un sur l'examen des critères de convergence des économies de l'Union Européenne, l'autre, plus développé, appliqué au domaine de la description personologique en passation répétée.

**Abstract.** We stand here as part of the method of data analysis, statistical analysis implicative (ASI). Like what we did to pass variables to binary numerical variables or variables-intervals, we extend the scope of treatment for vector-valued variables. We establish an index to measure the quality of a rule between vector variables. We treat examples bearing, one on the examination of economic convergence criteria of the European Union, the other, more developed, applied to the field description in personological award repeatedly.

## 1. Introduction et problématique

Certaines situations réelles, y compris celles pour lesquelles l'analyse implicative a créé son modèle statistique (la didactique des mathématiques), conduisent au traitement d'autres types de variables que les variables binaires. C'est ainsi que des extensions successives, pour répondre à des applications comme dans (Gras et al. 1996) nous ont conduits à intégrer dans la même théorie, des variables modales, numériques, floues, des variables sur intervalles, des variables-intervalles, des variables-rangs et à élaborer des outils de représentation graphique. Dans cet article, nous étudions, toujours dans le cadre de l'analyse implicative, le cas de variables à valeurs vectorielles<sup>1</sup>.

On veut comparer l'évolution temporelle d'un système de variables à  $d$  dimensions. Par exemple, sans pratiquer de test d'hypothèse mais sans l'exclure, on envisage :

- une suite de  $d$  notes obtenues dans des disciplines, indépendantes a priori, examinée à des périodes régulières (trimestre) sur un ensemble  $E$  d'élèves soumis à un traitement expérimental. On cherche à savoir si, globalement, les performances à l'instant  $t$  impliquent celles obtenues à l'instant  $t+1$ ,  $t+2$ , etc. et donc si le traitement expérimental est influent ;

- des mesures (analyses médicales) de  $d$  paramètres, indépendants a priori, obtenues à des périodes régulières afin d'étudier l'effet d'un traitement sur une population  $E$  de patients. On cherche s'il existe des règles d'association entre les moments d'observation et donc s'il existe un effet du traitement ;

- on fait l'observation à  $d$  instants (année, séquence, phase,...) de valeurs de paramètres dont on veut extraire d'éventuelles règles d'évolution d'un moment à un autre ; par exemple, l'évolution comparée de la fonction en entreprise occupée par des individus à  $d$  instants par rapport aux salaires respectivement octroyés à ces instants.

La représentation de ces suites est vectorielle. A une observation sur un sujet correspond le vecteur des observations de chacune de ses  $d$  composantes. Ces observations sont quantitatives, de nature binaire, modale ou intervalle. Une règle d'association entre deux vecteurs aura d'autant plus de sens que les composantes vectorielles présenteront une sémantique commune (par exemple : « progrès », « régression », etc.). Elle se substituera d'autant plus avantageusement à l'analyse implicative classique entre les composantes, modalités de chaque vecteur, que celles-ci présenteront des caractéristiques indépendantes dans leur essence même.

Mais les deux approches peuvent être complémentaires et informatives : l'analyse vectorielle est globale, l'analyse classique est ponctuelle. La première vise, en quelque sorte, à partir de profils individuels (la suite des composantes vectorielles) à dégager des règles d'association entre profils synthétiques. La seconde permet de dégager des règles d'association entre attributs binaires ou non, élémentaires ou composites (obtenu par la conjonction d'attributs élémentaires). Mais un attribut composite n'est pas un profil, il n'en est que la contraction. Ainsi, l'analyse vectorielle ne se ramène pas à l'analyse classique.

Dans l'exemple 2, les mesures répétées de  $n$  dimensions psychologiques posées a priori comme indépendantes ouvrent des évolutions temporelles et la représentation de ces suites de mesures est vectorielle. On cherche à savoir si un profil de traits psychologiques, relevant du modèle des cinq facteurs, obtenu dans un premier contexte implique un profil semblable dans un second contexte. In fine, on teste la stabilité du profil psychologique au travers de contextes différents. Cette stabilité conditionne la possibilité de poser des pronostics de réussite scolaire et/ou professionnelle à partir d'un profil de dimensions psychologiques.

Dans un premier temps, on rappellera les principes de l'analyse statistique implicative vectorielle. On décrira ensuite les questionnaires utilisés et les variables qu'ils permettent d'obtenir. Puis on passera aux contextes des passations répétées et à la description de l'échantillon des répondants avant de présenter les résultats obtenus.

---

<sup>1</sup> Ce texte a emprunté, tout en les modifiant, certains paragraphes à Gras R. et Couturier, [2009], « Analyse statistique implicative entre variables vectorielles »

## 2. L'analyse implicative vectorielle

### 2-1 Détermination d'un indice « vectoriel » dans le cas binaire

Les situations précédentes relèvent de variables numériques (voir le § 2-4) qu'il est possible de garder comme telles mais qu'il est aussi possible de binariser en ramenant chaque valeur à sa position par rapport à une norme : au-dessus, au-dessous. Mais d'autres situations peuvent se présenter sous une forme immédiatement binaire : par ex., une épreuve composée de  $d$  modalités où les observations seraient réussite-échec, présence-absence, etc. 2.

On cherche donc à comparer des vecteurs de type  $\vec{a} = (a_1, a_2, \dots, a_d)$ , vecteur représentatif d'une variable vectorielle  $\vec{a}$ , et  $\vec{b} = (b_1, b_2, \dots, b_d)$ , vecteur représentatif d'une variable vectorielle  $\vec{b}$ . On veut en extraire par exemple la règle  $\vec{a} \Rightarrow \vec{b}$ . Pour ce faire, on associe à chaque observation selon  $n$  sujets,

- d'une part, un vecteur à  $d$  dimensions, associé à la variable  $\vec{a}$ , de la somme sur les  $n$  sujets de chacune des  $d$  composantes, c'est-à-dire le vecteur-ligne :  $\vec{A} = \left( \sum_{i=1}^{i=n} a_1(i), \sum_{i=1}^{i=n} a_2(i), \dots, \sum_{i=1}^{i=n} a_d(i) \right)$ .  $\vec{A}$  admet

ainsi pour  $j$ ème composante scalaire le nombre  $\sum_{i=1}^{i=n} a_j(i)$ . Ce nombre représente donc le nombre de fois (cas binaire) où apparaît la variable composante  $a_j$  (ou la somme de ses pondérations dans le cas numérique) dans l'ensemble de la population. Soit aussi  $\vec{A}(i)$  le vecteur-ligne des  $d$  composantes du sujet  $i$  de la forme  $(a_1(i), a_2(i), \dots, a_j(i), \dots, a_d(i))$ .  $a_j(i)$  est la  $j$ ème composante du vecteur  $\vec{A}(i)$ , c'est-à-dire la valeur prise par le sujet  $i$  selon la variable composante  $a_j$ . On fait de même avec la variable  $\vec{b}$  afin d'obtenir le vecteur-ligne  $\vec{B}_i$  ;

- d'autre part, deux ensembles aléatoires de vecteurs-lignes, choisis au hasard et indépendamment l'un de l'autre, dont les sommes vectorielles-lignes  $\vec{X}$  et  $\vec{Y}$  respectives coïncident exactement et respectivement, composante par composante, avec celles de  $\vec{A}$  et de  $\vec{B}$ . Par analogie avec le cas des variables  $a$  et  $b$  uniques (une seule composante), ces deux ensembles sont à comparer aux ensembles  $X$  et  $Y$  de sujets, de même cardinaux que ceux de  $A$  et  $B$  supports des variables respectives  $a$  et  $b$ . Mais, bien entendu, ce ne sont pas généralement, les mêmes sujets qui les ont constitués.

A ces vecteurs  $\vec{A}$  et  $\vec{B}$ , on associe le vecteur  $\vec{W}$  des contre-exemples observés aux implications successives selon les vecteurs-sujets  $\vec{A}(i)$  et  $\vec{B}(i)$ , donc d'indice identique. Par exemple, un contre-exemple en  $i$  à  $a_j \Rightarrow b_j$  apparaît lorsque  $a_j(i) = 1$  alors que  $b_j(i) = 0$ , c'est-à-dire lorsque  $a_j(i) \wedge \bar{b}_j(i) = 1$ . Par suite, le vecteur des contre-exemples observés aura pour composantes scalaires les  $j$  scalaires de

la forme  $\sum_{i=1}^{i=n} a_j(i) \wedge \bar{b}_j(i)$  que nous notons, comme dans le cas de la variable unique  $N_{a_j \wedge \bar{b}_j}$ . Aux

vecteurs aléatoires  $\vec{X}$  et  $\vec{Y}$ , on associe de la même façon, le vecteur  $\vec{W}$  des contre-exemples aléatoires

de composantes  $\sum_{i=1}^{i=n} x_j(i) \wedge \bar{y}_j(i)$  notés également  $N_{a_j \wedge \bar{b}_j}$ .

<sup>2</sup> S'il y a nécessité d'étendre la méthodologie employée dans le cas d'une variable binaire, la généralisation au cas de variable numérique se fera comme dans le cas de la composante unique en utilisant la valeur corrigée des indices définis dans (Lagrange, 1998) ou (Régner et Gras, 2005).

Or,  $N_{a_j \wedge \bar{b}_j}$  suit, en tant que variable aléatoire relative à la règle présumée  $a_j \Rightarrow b_j$ , et comme nous l'avons prouvé par ailleurs (Gras R. et Régnier J.C., 2009), la loi de Poisson de paramètre  $\frac{n_{a_j}}{n} \cdot n_{\bar{b}_j}$  ou bien la loi binomiale de paramètres  $n$  et  $\frac{n_{a_j}}{n} \cdot \frac{n_{\bar{b}_j}}{n}$ . Alors, du fait de l'indépendance des variables composantes, la probabilité pour que les nombres de tous les contre-exemples aléatoires ne soient pas inférieurs aux nombres de contre-exemples respectivement observés est le produit des probabilités pour que cette propriété soit vérifiée sur chaque composante, à savoir :

$$\prod_{j=1}^{j=d} [1 - \Pr[N_{a_j \wedge \bar{b}_j} \leq n_{a_j \wedge \bar{b}_j}]]$$

On définit alors l'intensité d'implication de la règle  $\bar{a} \Rightarrow \bar{b}$  par <sup>3</sup>

$$\Phi(\bar{a}, \bar{b}) = \left[ \prod_{j=1}^{j=d} [1 - \Pr[N_{a_j \wedge \bar{b}_j} \leq n_{a_j \wedge \bar{b}_j}]] \right]^{1/d}$$

Par suite, compte tenu de la définition de l'indice classique de l'intensité d'implication d'une variable vers une autre :  $\Phi(\bar{a}, \bar{b}) = \left[ \prod_{j=1}^{j=d} \Phi(a_j, b_j) \right]^{1/d}$

Le logiciel C.H.I.C. (Couturier, 2009) permet le calcul des intensités d'implication dans le cas vectoriel. Nous en aurons deux applications dans les § 2-4 et 3.

## 2-2 Remarques

\* On peut comparer cette extension à la situation originelle où l'on a, pour une variable unique, un vecteur à une dimension. Dans le cas vectoriel présenté ici, une quasi-inclusion du vecteur représentatif de  $a$  dans un parallélépipède de dimension  $d$  définie par sa diagonale  $\bar{b}$  se substitue à la quasi-inclusion de l'ensemble  $A$  dans l'ensemble  $B$ , supports respectifs de  $a$  et  $b$ .

\* Toutes les formules où figure l'intensité d'implication  $\Phi(\bar{a}, \bar{b})$  peuvent être réécrites en remplaçant  $\Phi$  par  $\Psi$  qui symbolise l'implication-inclusion sur des bases entropiques [Gras et als, 01].

\* De même, ces formules peuvent être étendues au cas des variables numériques selon la transformation établie par (Lagrange, 1998), qui modifie l'écart-type de la loi de Poisson.

\* On pourrait s'intéresser à l'événement où  $k$  composantes seulement parmi les  $d$  satisferaient l'inégalité entre la valeur aléatoire des contre-exemples et la valeur observée. Ce qui correspond à une exigence affaiblie quant à la règle  $\bar{a} \Rightarrow \bar{b}$ . Pour cela, il suffit d'envisager toutes les parties à  $k$  éléments parmi  $\{(a_1, b_1), (a_2, b_2), \dots, (a_d, b_d)\}$ , de faire la moyenne arithmétique des sommes des produits correspondants dans le crochet donnant  $\Phi(\bar{a}, \bar{b})$  et d'en prendre la puissance correspondante :

Par exemple, si  $d=3$  et  $k=2$ , on aurait :

$$\Phi(\bar{a}, \bar{b}) = \left\{ \frac{1}{3} [\Phi(a_1, b_1) \cdot \Phi(a_2, b_2) + \Phi(a_2, b_2) \cdot \Phi(a_3, b_3) + \Phi(a_3, b_3) \cdot \Phi(a_1, b_1)] \right\}^{1/2}$$

\* Le respect de la sémantique quant au jugement associé à la mise en évidence d'une implication peut être délicat. Il est loisible, pour certaines des variables composantes de la prémisse ou de la

<sup>3</sup> Une formule acceptable et alternative de celle-ci est  $1 - \prod_{j=1}^{j=d} [\Pr[N_{a_j \wedge \bar{b}_j} \leq n_{a_j \wedge \bar{b}_j}]]^{1/d}$

conclusion, de changer l'ordre défini a priori sur les valeurs (binaires ou non) que prennent ces composantes. Ce changement viserait à optimiser l'intensité d'implication  $\Phi(\bar{a}, \bar{b})$ . Par exemple, on pourrait permuter les valeurs 0 et 1 dans le cas binaire ou, de façon générale, passer à la complémentation à 1 de chaque valeur.

### 2-3 Application numérique

Une population est constituée de 85 Sujets, dont certains possèdent les mêmes caractéristiques que leur prototype. Ainsi on observe, dans le tableau 1, 20 sujets répondant 0 en a1 et de même selon b1, 1 en a2 et b2.

**Tableau 1. Réponses des sujets.**

	$\bar{a}$		$\bar{b}$	
	a <sub>1</sub>	a <sub>2</sub>	b <sub>1</sub>	b <sub>2</sub>
20 Sujets 1	0	1	0	1
30 Sujets 2	0	0	1	0
30 Sujets 3	1	0	1	1
5 Sujets 4	1	1	0	0
Totaux	<b>35</b>	<b>25</b>	<b>60</b>	<b>50</b>

On observe :  $n_{a_1 \wedge \bar{b}_1} = n_{a_2 \wedge \bar{b}_2} = 5$ . En utilisant le modèle de Poisson, on obtient en se limitant au calcul par l'intensité d'implication classique (et non pas « entropique ») :

$$\Pr[N_{a_1 \wedge \bar{b}_1} \leq 5] = 0.05 = \Pr[N_{a_2 \wedge \bar{b}_2} \leq 5] \text{ d'où } \Phi(\bar{a}, \bar{b}) = (0.95 \times 0.95)^{1/2} = 0.95$$

### 2-4 Indices vectoriels dans le cas de vecteurs à composantes numériques

Nous avons établi (Lagrange, 1998) et (Gras et Régner, 2009) un indice respectant l'approche statistique du cas binaire aux cas où :

- \* les variables sont modales (degré de possession d'un attribut ou d'adéquation à celui-ci),
- \* les variables sont numériques (nombre d'occurrences, valeurs de contingence)

L'indice adopté permet d'accorder une mesure à l'implication entre de telles variables. De plus, cet indice restreint au cas binaire coïncide avec l'indice mesurant traditionnellement la règle où prémisses et conclusions sont binaires.

Si a et b sont deux variables numériques, dont les valeurs observées respectivement  $\{a_i\}_{i \in E}$  et  $\{b_i\}_{i \in E}$  dans une population E de taille n et dont les moyennes et variances respectives sont  $m_a, m_b, s_a, s_b$ , alors l'indice de base est :

$$q(a, \bar{b}) = \frac{\sum_{i \in E} a_i \bar{b}_i - \frac{m_a m_b}{n}}{\sqrt{\frac{(n^2 s_a^2 + m_a^2)(n^2 s_b^2 + m_b^2)}{n^3}}}$$

et l'intensité de la règle  $\bar{a} \Rightarrow \bar{b}$  est dans une approximation gaussienne :

$$\Phi(a, b) = 1 - \Pr[Q(a, \bar{b}) \leq q(a, \bar{b})] = \frac{1}{\sqrt{2\pi}} \int_{q(a, \bar{b})}^{\infty} e^{-\frac{t^2}{2}} dt$$

expression dans laquelle  $Q(a, \bar{b})$  est la valeur centrée réduite de  $N_{a \wedge \bar{b}}$ , pour l'une des lois retenues dans le modèle : binomiale ou de Poisson ou gaussienne.

Considérant maintenant des vecteurs de type  $\vec{a} = (a_1, a_2, \dots, a_d)$ , vecteur représentatif d'une variable vectorielle  $\vec{a}$ , et  $\vec{b} = (b_1, b_2, \dots, b_d)$ , vecteur représentatif d'une variable vectorielle  $\vec{b}$  où  $a_1, a_2, \dots, a_d$  et  $b_1, b_2, \dots, b_d$  sont  $2d$  variables numériques. L'intensité d'implication de la règle  $\vec{a} \Rightarrow \vec{b}$  est encore :

$$\Phi(\vec{a}, \vec{b}) = \left[ \prod_{j=1}^{j=d} \varphi(a_j, b_j) \right]^{1/d}$$

### 2-5 Exemple 1

Nous disposons de « critères de convergence » entre États de l'Union européenne au 25-3-98, critères qui permettaient de comparer les économies des pays et d'en définir leur capacité à respecter la politique monétaire commune conditionnant l'ouverture à l'euro.

**Tableau 2. Critères de convergence.**

	96Infl	96défic	96dette	96Taux	97Infl	97défic	97dette	97Taux	98Infl	98défic	98dette	98Taux
Allemagne	0,22	0,41	0,48	0,62	0,35	0,23	0,5	0,7	0,38	0,1	0,52	0,75
Autriche	0,29	0,35	0,55	0,63	0,33	0,26	0,54	0,71	0,33	0,15	0,55	0,75
Belgique	0,27	0,43	1	0,65	0,29	0,33	1	0,73	0,29	0,31	1	0,76
Danemark	0,25	0,68	0,56	0,72	0,42	0,83	0,53	0,79	0,47	1	0,5	0,75
Espagne	0,4	0,29	0,55	0,87	0,46	0,25	0,56	0,8	0,49	0,18	0,57	0,84
Finlande	0,19	0,42	0,45	0,71	0,26	0,54	0,46	0,75	0,44	0,82	0,45	0,73
France	0,22	0,34	0,44	0,63	0,2	0,18	0,48	0,7	0,22	0	0,49	0,73
Grèce	1	0	0,88	1	1	0	0,89	1	1	0,18	0,91	1
Irlande	0,13	0,71	0,57	0,73	0,26	0,86	0,54	0,79	0,73	1	0,5	0,83
Italie	0,51	0,08	0,98	0,94	0,44	0,23	1	0,86	0,47	0,1	1	0,89
Luxembourg	0,19	1	0,05	0,63	0,26	1	0,06	0,7	0,36	1	0,06	0,75
Pays-Bas	0,15	0,52	0,61	0,62	0,4	0,46	0,59	0,7	0,51	0,33	0,59	0,73
Portugal	0,31	0,43	0,51	0,86	0,38	0,26	0,51	0,8	0,49	0,18	0,51	0,83
Royaume Uni	0,31	0,27	0,43	0,79	0,42	0,37	0,44	0,89	0,51	0,59	0,44	0,93
Suède	0,14	0,4	0,6	0,8	0,4	0,56	0,63	0,95	0,33	0,87	0,63	0,87

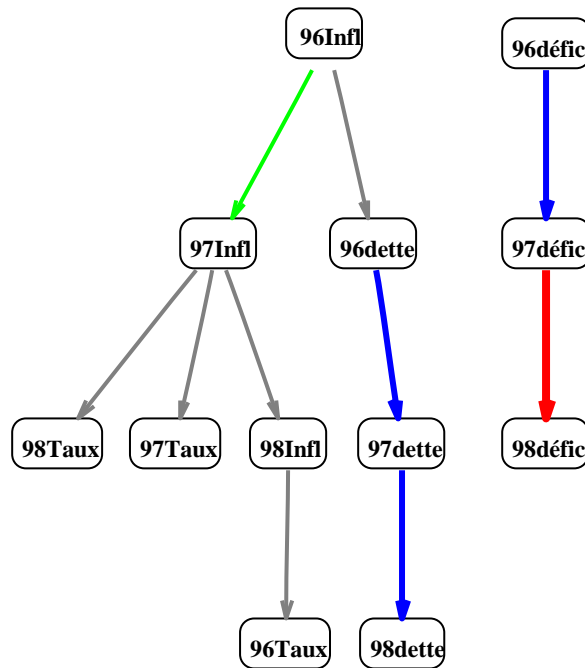
Ces critères s'expriment selon plusieurs composantes : l'inflation, le déficit public (% du PIB), la dette publique (% du PIB) et le taux d'intérêt à long terme. Le tableau<sup>4</sup> ci-dessus donne les différentes valeurs observées sur 3 années : 1996, 1997 et 1998 sur les 15 pays de l'Union.

Nous avons encore affaire à des variables numériques : l'une pour 1996, une autre pour 1997, une troisième pour 1998. Les valeurs indiquées TAB. 4 ont été toutes ramenées à des nombres de

<sup>4</sup> Tableau extrait du QUID 99, avec estimation des taux d'intérêt de la Grèce pour les trois années en question.

l'intervalle [0,1] par division de chaque colonne par le maximum de la colonne. Seul le déficit a nécessité de translater les valeurs de telle façon que nous n'ayons que des nombres positifs, alors même que dans la plupart des pays, le déficit était toujours négatif.

Un premier traitement par C.H.I.C. des 12 variables numériques (4 pour chacune des 3 années) conduit au graphe implicatif suivant construit aux seuils .85 à .70 :



**Figure 1.** Graphe implicatif des critères de convergence.

On observe une indépendance des 3 critères : inflation, dette et taux d'intérêt par rapport au critère déficit. Sur le plan économique, et cela paraît surprenant, le déficit n'aurait pas d'influence sur la dette et l'inflation. On constate, en revanche, la liaison attendue entre l'inflation et les taux d'intérêt, l'ensemble se présentant relativement peu lié à la dette.

On observe également qu'alors que déficit, inflation et dette croissent de 96 à 98, comme on pouvait l'attendre, les taux d'intérêt diminuent pendant cette période. Nous y reviendrons.

Arrêtant ici l'exploitation de ce graphe sur lequel bien d'autres considérations peuvent être énoncées, nous constatons que les natures des critères, bien qu'en partie liées, ne rendent pas bien compte du comportement global des économies de 96 à 98. D'où l'intérêt de considérer 3 variables vectorielles numériques : A code l'année 1996, présente les composantes A1, A2, A3 et A4 respectivement correspondant aux 4 critères, puis B code 1997, avec les composantes B1 à B4, et C code 1998 avec les composantes C1 à C4.

Le logiciel C.H.I.C. dégage les intensités d'implication suivantes :

$$\Phi (A \Rightarrow B) = 0,71 ; \Phi (A \Rightarrow C) = 0,70 ; \Phi (B \Rightarrow C) = 0,74$$

les intensités des relations réciproques étant plus faibles. Les trois variables s'enchaînent donc transitivement, au seuil 0,70, confirmant la croissance globale des valeurs des critères avec le temps, croissance freinée nécessairement par la décroissance des taux d'intérêt dont cette approche ne rend

plus compte. Cette restriction illustre le propos tenu dans le dernier paragraphe du § 2. La relation d'ordre dérivée de la sémantique des critères-composantes devrait être la même. Or pour les trois critères : inflation, déficit et dette, la sémantique de ces variables oriente l'évolution du critère vers la croissance, alors que pour le taux, dans les données, l'évolution est décroissante. Si nous complétons, alors, à 1 chaque valeur du taux énoncé, nous obtenons des intensités d'implication supérieures aux précédentes, ce qui confirme le bien-fondé de la remarque au sujet des sémantiques d'ordre :

$$\Phi (A \Rightarrow B) = 0,78 ; \Phi (A \Rightarrow C) = 0,77 ; \Phi (B \Rightarrow C) = 0,79.$$

### 3. Une étude de la stabilité de profils personnologiques en passation répétée

#### 3.1 Le Q-ISAT et les variables disponibles

Le questionnaire auto-descriptif utilisé est le Questionnaire d'Image de Soi au Travail - Q-I.S.A.T. De format Q-sort, il comprend 100 descripteurs de la personnalité. Le choix de la méthodologie Q (Stephenson, 1935a, 1935b, 1953) se justifie par le fait que se décrire est une activité subjective et que cette méthodologie permet d'accéder aux facteurs opérants de cette subjectivité.

Les descripteurs sont cohérents avec le modèle des cinq facteurs qui postule qu'on peut décrire une personnalité de manière exhaustive à l'aide de cinq facteurs bipolaires et orthogonaux : (1) conscience, (2) ouverture d'esprit, (3) gentillesse, (4) extraversion, (5) névrosisme. Dans le Q-I.S.A.T., chaque pôle des dimensions est décrit par 10 adjectifs. La passation est informatisée sous Excel et on obtient un profil pour les pôles positifs et un profil pour les pôles négatifs. D'autres indices sont disponibles qui ne seront pas décrits ici.

#### 3.2 Les passations répétées et l'échantillon des répondants

Les répondants passent le questionnaire successivement à deux reprises. Au cours de la première passation, on leur demande de se décrire à leur travail dans une situation favorable où tout se passe bien. Au cours de la seconde passation, on leur demande toujours de se décrire à leur travail, mais dans une situation défavorable où rien ne se passe bien.

L'échantillon comporte 91 répondants occasionnels. Cet échantillon n'a aucune ambition à représenter une quelconque population parente. Les âges vont de 11 à 62 ans, avec une moyenne de 30 ans et un écart-type de 12,6 ans. On dénombre 56 femmes et 35 hommes. Enfin, les niveaux de qualification se distribuent entre les six niveaux de la nomenclature interministérielle des niveaux de qualification, soit : 6 niveaux 1, 42 niveaux 2, 10 niveaux 3, 12 niveaux 4, 11 niveaux 5 et 9 niveaux 6.

#### 3.3 Résultats des analyses

On a procédé à l'analyse des scores des pôles positifs des cinq dimensions. Les valeurs brutes pour chaque variable varient de 1 à 7. Elles ont été ramenées à des nombres de l'intervalle [0,1] par division de chaque colonne par le maximum de la colonne.

Une première façon de considérer la stabilité des scores d'une passation à l'autre consiste à observer les corrélations (tab. 1).

**Tableau 3.** Corrélations des scores des dimensions obtenus aux deux passations.

dimensions	corrélations	% variance
F1-D1	0,14	1,96



F2-D2	0,11	1,21
F3-D3	0,28	7,84
F4-D4	0,27	7,29
F5-D5	0,16	2,56

*F: situation favorable. D: situation défavorable; 1 à 5: numéros des dimensions.*

Si les cinq corrélations sont positives, leurs valeurs absolues restent très modérées et les parts de variance expliquée vont de 1,21% à 7,84%.

Dans la mesure où les liens sont faibles, on peut aussi s'intéresser à l'ampleur des différences entre les scores (tab. 2). Les premières lignes du tableau indiquent les caractéristiques des distributions puis les valeurs d'écart des moyennes des scores d'une passation à l'autre.

La seconde partie du tableau donne des indicateurs d'évaluation de ces écarts. En premier lieu, on observe des *d* de Cohen variant de 0,88 à 1,29 traduisant, selon la convention admise, des tailles d'effets notables. Les écarts de surface des courbes vont de 12% à 21%. Enfin, les *t* de Student<sup>5</sup> vont de 5,72 à 8,79 et renvoient tous des valeurs de *p* très significatives.

**Tableau 4.** Ecarts des scores des deux dimensions obtenus aux deux passations.

<b>favorable</b>	<b>F1</b>	<b>F2</b>	<b>F3</b>	<b>F4</b>	<b>F5</b>
<b>m</b>	0,82	0,89	0,85	0,83	0,86
<b>σ</b>	0,08	0,06	0,08	0,08	0,07
<b>défavorable</b>	<b>D1</b>	<b>D2</b>	<b>D3</b>	<b>D4</b>	<b>D5</b>
<b>m</b>	0,73	0,76	0,71	0,70	0,71
<b>σ</b>	0,12	0,13	0,15	0,12	0,16
<b>écart</b>	0,09	0,13	0,14	0,13	0,16
<b>d de Cohen</b>	0,88	1,29	1,17	1,28	1,22
<b>taille de l'effet</b>	notable	notable	notable	notable	notable
<b>comparaison</b>	-12%	-17%	-20%	-19%	-21%
<b>t de Student</b>	5,72	8,74	7,98	8,66	8,79
<b>p</b>	<0,01	<0,01	<0,01	<0,01	<0,01

On peut conclure que les scores obtenus pour chaque dimension aux deux passations ne présentent pas une grande stabilité: les répondants, lorsqu'ils passent à leur auto-description en situation défavorable, choisissent moins de descripteurs des pôles positifs des cinq dimensions.

Bien qu'implication et corrélation linéaire vont plutôt "dans le même sens", on montre la non-coïncidence des deux concepts et donc leur différence informative (Gras et al, 1996). Dans une logique implicite, dans le cas d'une structure personologique stable, le score sur un pôle à la première passation devrait impliquer le score sur ce pôle à la seconde passation avec une intensité d'implication

<sup>5</sup> Statistique donnée à titre indicatif, par "habitude", tout en reconnaissant qu'elle n'est pas adaptée au cas traité.

consistante, de l'ordre de 80. Concrètement, pour les cinq pôles, on obtient des graphes implicatifs dont les implications vont dans le sens  $D \Rightarrow F$ , de la situation défavorable vers la situation favorable, avec des indices allant de 53 à 60, résultat qui confirme le peu de stabilité du choix des répondants d'une situation auto-descriptive à l'autre.

Le raisonnement tenu jusqu'à ce point de l'analyse considère, d'un point de vue analytique, l'évolution des scores obtenus à chacun des pôles des dimensions. On peut passer à un point de vue plus synthétique par le biais de l'analyse implicative vectorielle qui permet d'évaluer l'évolution non plus de chaque dimension, mais du profil des dimensions d'une passation à l'autre, afin d'appréhender le comportement global des répondants.

On considérera deux variables vectorielles : F code les dimensions en situation de passation favorable et D les mêmes dimensions en situation de passation défavorable. Le logiciel CHIC-vectoriel dégage les intensités d'implication suivantes :  $F \Rightarrow D$  0,53 et  $D \Rightarrow F$  0,56. Sous l'option entropique ces intensités s'annulent.

On retrouve le sens de l'implication de D vers F, avec une validation de la sémantique: "la passation en situation défavorable abaisse la fréquence des choix des descripteurs propres aux pôles positifs des dimensions personnologiques". En d'autres termes les profils personnologiques ne sont pas stables et varient en fonction du contexte de l'auto-description demandée.

#### 4. Discussion

On distinguera deux axes de discussion : (a) le résultat des analyses dans le cadre de la problématique de l'évaluation personologique à l'aide de questionnaires auto-descriptifs et (b) l'apport informationnel du recours à l'analyse vectorielle dans le cadre des modèles psychométriques.

(a) L'analyse des données ne permet de justifier un modèle psychométrique postulant la stabilité des auto-descriptions personnologiques : les scores obtenus pour chaque pôle positif, ainsi que leur profil global, évoluent à la baisse quand on passe d'une situation favorable à une situation défavorable. Pour une part les choix des items dépendent du contexte situationnel. Dans une recherche précédente, pratiquée sur un autre échantillon et avec un autre schéma expérimental l'un d'entre nous (Pasquier, 2007) était parvenu à la même conclusion d'une instabilité du profil personologique en fonction des situations.

Dans ces conditions, les pronostics de réussite scolaire et/ou professionnelle réalisés à l'aide des questionnaires auto-descriptifs ne sont pas légitimes : un prédicteur flou, ambigu, élastique ne peut produire que des prédictions illusives. Les corrélations éventuellement obtenues relèvent d'artefacts de désirabilité sociale et de distance socio-normative (Pasquier, 2007; Pasquier et Mazilescu, 2009). Ces points ne seront pas développés dans le présent cadre.

(b) On a vu l'intérêt de l'analyse implicative vectorielle pour comparer globalement les relations implicatives entre deux séries de profils personnologiques issues de deux situations différentes intervenant diachroniquement. En psychométrie, la notion de profil intervient certes dans le domaine de la description de la personnalité, mais aussi dans le domaine des contenus cognitifs, des facteurs conatifs, des variables affectivo-émotionnelles et d'une manière générale à chaque fois que la validation du modèle psychométrique passe par une problématique multi-dimensionnelle.

Bien évidemment, le modèle multi-dimensionnel ne suffit pas en lui-même à déclencher une analyse vectorielle. Encore faut-il avoir des dimensions a priori indépendantes, ce qui exclut les rotations obliques dans le paramétrage des analyses factorielles, et des passations répétées dans le cadre par exemple des études longitudinales, tout aussi bien du développement (psychologie génétique) que de l'involution (gérontologie), ou de l'évolution de carrière (G.R.H.)...

## 5 Conclusion

Nous avons étendu la variété des variables prises en compte par la méthode d'analyse statistique implicative en conceptualisant la relation implicative entre deux variables vectorielles. De telles variables peuvent présenter des modalités binaires ou numériques représentées par les composantes vectorielles. Cette souplesse permet de dégager et quantifier une règle d'association dissymétrique entre deux vecteurs en exprimant par une mesure la dynamique inhérente aussi bien à l'ensemble des variables vectorielles qu'à celle de leurs modalités. Des exemples simples, traités à la main et des exemples réels traités par CHIC illustrent la construction théorique de l'extension de l'A.S.I. aux vecteurs.

## Références

Couturier, R., & Ag Almouloud, S. (2009), Historique et fonctionnalités de CHIC, *Analyse Statistique Implicative, Une méthode d'analyse de données pour la recherche de causalités, sous la direction de Régis Gras, réd, invités Régis Gras, Jean-Claude Régnier, Fabrice Guillet, Cépaduès Ed. Toulouse, ISBN : 978.2.85428.8971*, p. 279-294.

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

Gras, R., & Régnier, J. C. (2009), Origine et développement de l'Analyse Statistique Implicative, *Analyse Statistique Implicative, Une méthode d'analyse de données pour la recherche de causalités, sous la direction de Régis Gras, réd, invités Régis Gras, Jean-Claude Régnier, Fabrice Guillet, Cépaduès Ed. Toulouse, ISBN : 978.2.85428.8971*, p. 6-16.

Gras, R., & Couturier, R. (2009), Analyse statistique implicative entre variables vectorielles, *Analyse Statistique Implicative, Une méthode d'analyse de données pour la recherche de causalités, sous la direction de Régis Gras, réd, invités Régis Gras, Jean-Claude Régnier, Fabrice Guillet, Cépaduès Ed. Toulouse, ISBN : 978.2.85428.8971*, p. 195-205.

Lagrange, J.B. (1998). Analyse implicative d'un ensemble de variables numériques; application au traitement d'un questionnaire aux réponses modales ordonnées, *Revue de Statistique Appliquée XLVI-1*, Paris, I.H.P., 71-93.

Pasquier, D. (2007). Évaluation de la personnalité à l'aide des questionnaires autodescriptifs: approche critique du postulat de la stabilité de l'image de soi. Pour une autre utilisation et une autre lecture du modèle des cinq facteurs. Lille : A.N.R.T.

Pasquier, D., & Mazilescu, C.A. (2009). La validité prédictive d'un questionnaire de personnalité basé sur le modèle des cinq facteurs : les médiations sous-jacentes. *Cognition, Brain, Behavior. An Interdisciplinary Journal*. XIII, 2, 165-177.

Gras, R., Suzuki, E., Guillet, F., & Spagnolo, F. (Eds), *Statistical Implicative Analysis*, Springer-Verlag, ISBN 978-3-540-78982-6.

Ratsimba-Rajohn, H., (2009). Guide d'utilisation des principales fonctionnalités du logiciel CHIC, *Analyse Statistique Implicative, Une méthode d'analyse de données pour la recherche de causalités, sous la direction de Régis Gras, réd, invités Régis Gras, Jean-Claude Régnier, Fabrice Guillet, Cépaduès Ed. Toulouse, ISBN : 978.2.85428.8971*, p295-316.

Stephenson, W. (1935a). Technique of factor analysis. *Nature*, 136, 297.

Stephenson, W. (1935b). Correlating persons instead of tests. *Character and Personality*, 4, 17-24.

Stephenson, W. (1953). *The study of behaviour : Q-technique and its methodology*. Chicago : University of Chicago Press.