

ANALYSE EXPLORATOIRE A *PRIORI* ET ANALYSE CONFIRMATOIRE A *POSTERIORI* : PARADIGME POUR LA COMPARAISON DE DEUX STRUCTURES EN ANALYSE STATISTIQUE IMPLICATIVE.

Daniel PASQUIER¹, Régis GRAS²

A PRIORI EXPLORATORY ANALYSIS AND A POSTERIORI CONFIRMATORY ANALYSIS : PARADIGM FOR THE COMPARISON OF TWO STRUCTURES IN IMPLICATIVE STATISTICAL ANALYSIS.

RÉSUMÉ

L'analyse statistique implicative - A.S.I. - est une méthode d'analyse de données ayant pour objectif d'extraire des règles du type « si la variable a est observée dans la population alors généralement la variable b l'est aussi ». Nous proposons dans ce texte une étude originale permettant de mettre en évidence tant sur le plan théorique et méthodologique que sur le plan illustratif l'intérêt de la méthode A.S.I. pour comparer, sur une même population, des études successives d'un phénomène repéré par des variables psychologiques ce qui rend son utilisation possible dans le cadre d'études confirmatoires.

Mots clés : analyse statistique implicative, analyse confirmatoire, stabilité personologique.

ABSTRACT

The implicative statistical analysis - I.S.A. - is a data analysis method which aims at extracting rules such as "if variable a is observed in the population then variable b also generally is". In this paper, we propose an original work highlighting, from a theoretical, the methodological as well as a practical point of view, the advantages of the I.S.A. method in comparing successive studies of a phenomenon issued by psychological variables over a given population. We thus hope to point out how relevant this method would be when used in confirmatory studies.

Keywords : implicative statistical analysis, confirmatory analysis, personology stability.

1 Introduction

L'Analyse Statistique Implicative (ASI) est une méthode d'analyse de données ayant pour objectif premier, sur la base d'un ensemble de données croisant une population d'individus

¹ Université François Rabelais de Tours, EA 2114 : Psychologie des Ages de la Vie, 3 rue des Tanneurs - 37041 Tours cedex1, dpasquier@avenireentreprise.fr

² 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

(ou d'objets) et un ensemble de variables, d'extraire des règles du type « si la variable a est observée dans la population alors généralement la variable b l'est aussi ». Ces règles permettent d'inférer, avec une certaine confiance estimée par une mesure spécifique, des propriétés relationnelles non symétriques entre les variables en termes de quasi-implication, voire de causalité. Selon les situations d'observation, les variables peuvent être non seulement booléennes (par exemple : « vrai-faux » ou « présence-absence ») mais aussi, dans la théorie de l'ASI, numériques, d'intervalles, vectorielles ou floues.

Nous proposons dans ce texte une étude originale permettant de mettre en évidence, tant sur le plan théorique et méthodologique que sur le plan illustratif, l'intérêt de la méthode ASI pour comparer, sur une même population, des études successives d'un phénomène repéré par des variables psychologiques.

2 Principes généraux élaborés dans le domaine de la didactique

La situation générique est la suivante. A un ensemble E d'individus, de cardinal n, on fait correspondre des valeurs numériques (que l'on peut toujours normaliser pour les ramener dans l'intervalle [0 ; 1]) prises par des fonctions "attributs" d'un ensemble V, de cardinal m. Ces "attributs", tels que la variable a, élément de V, peuvent être des traits de caractère mais également des variables modales qui, par exemple, à l'individu x associent des valeurs telles que $a(x)$ soit un nombre de l'intervalle [0 ; 1].

On suppose connus deux états (ou plus), sous la forme de graphes implicatifs, de l'analyse des implications entre les éléments de V tels que les variables a_i et b_j . Ce graphe est, par nature, non symétrique : l'arc (a_i, b_j) du graphe représente la liaison implicative de a_i sur b_j . Par exemple, G_1 (resp. G_2, \dots) est le graphe implicatif à l'instant t_1 (resp. t_2, \dots) qui traduit l'ensemble des relations implicatives sur V, à un seuil implicatif retenu par l'analyste et mesuré par la pondération $\Psi(a_i, b_j)$. Mais ces relations peuvent être étendues à des conjonctions de variables de V.

On rencontre une telle situation, considérée comme prototypique, en éducation (didactique) où une même population d'élèves (ou deux échantillons extraits d'une même population-mère) est soumise à deux moments différents au même questionnaire ou mobilisant les mêmes outils cognitifs, l'un avant une intervention didactique, l'autre après celle-ci (ou une observation sur un échantillon et une seconde avec un autre de même taille). Il est intéressant d'étudier si, par exemple relativement à la même population, des relations de réussite sont modifiées de l'instant t_1 à l'instant t_2 . Si tel est le cas, on fera l'hypothèse de l'influence de l'intervention. On étudiera plus loin (§ 5) une situation en psychologie.

D'autres situations se présentent en didactique. Il s'agit de comparer les résultats lors de croisements de types différents :

- l'un dit « analyse *a priori* » où les lignes de la matrice de données sont composées de comportements généraux, de conceptions d'élèves relativement à des situations problèmes. Les colonnes sont constituées, comme dans la situation ci-dessus, d'items, de variables observables sur les sujets. Mais ces données sont construites par le chercheur qui présuppose des comportements attendus. Les intensités d'implication obtenues forment une matrice M dont il serait intéressant d'étudier l'évolution dans le temps ou les nuances selon des juges distincts ;

- l'autre dit « *a posteriori* » ou contingente où les lignes sont les sujets, les colonnes sont les items ou variables. A l'intersection ligne-colonne figure la valeur observée prise par un sujet selon une certaine variable. Une analyse statistique implicative conduit à une autre matrice d'intensités M dont il peut être intéressant d'étudier l'évolution dans le temps ou les nuances selon des groupes de sujets distincts (par exemple des classes).

De façon générale, nous disposons donc d'une suite de matrices pondérées, non symétriques, comparables à celles-ci (tab.1). Les graphes implicatifs ne sont qu'une traduction graphique partielle de ces matrices pour un seuil d'intensité d'implication donné.

TABLEAU 1 - Suite de matrices pondérées

Moment 1				Moment 2				Moment 3			
V	...	b ₁	...	V	...	b ₂	...	V	...	b ₃	...
...		
a ₁		$\psi(a_1, b_1)$...	a ₂		$\psi(a_2, b_2)$...	a ₃		$\psi(a_3, b_3)$...
...		

Les changements ou les différences peuvent s'avérer significatifs ou non. L'examen qualitatif de deux ou plusieurs graphes peut apporter des informations sur les changements éventuels des relations. Nous pourrions l'examiner dans le cas où l'on aurait, par exemple, plus de deux moments d'intervention ou plus de deux échantillons, dans le même cadre des variables de V . Mais pour préciser la significativité, une mesure de distance ou une recherche de la significativité de la différence entre les deux graphes serait opportune. Or les liaisons entre les variables étant le plus souvent différentes, la comparaison n'est pas possible sans revenir au tableau des mesures de qualité des relations implicatives sur V .

Dans Spagnolo, Gras et Régnier (2009), nous avons procédé de façon différente pour résoudre un problème où s'opposait une connaissance *a priori* et celle *a posteriori*. Mais elle ne faisait pas appel à la distance entre structures et nécessitait des calculs plus parcimonieux, mais tout aussi intéressants.

3 Analyses factorielles exploratoires et confirmatoires en psychologie

Dans le domaine de la psychométrie, on distingue deux types d'approche factorielle de l'analyse des tableaux de données. Historiquement première, rendue possible par l'analyse en composantes principales - A.C.P.- (Pearson, 1901 ; Hotelling, 1933), l'approche exploratoire recherche un résumé de l'information en faisant passer un système de droites dans le nuage de points, minimisant la somme des carrés de la distance de chaque point à chaque droite. On obtient ainsi une ou plusieurs nouvelles variables appelées composantes ou facteurs. Extraites par le calcul, ces variables sont dites latentes et on les interprète comme l'expression de processus mentaux susceptibles d'expliquer le niveau de performance à des familles d'épreuves (facteurs verbal, numérique, spatial...) ou à toutes les épreuves (facteur général).

Descriptive, l'A.C.P. aide à comprendre la structure d'une série de données, et constitue une source d'hypothèses à vérifier ultérieurement.

L'approche factorielle confirmatoire en psychologie prolonge le développement des modèles structuraux (Bacher, 1987, 1988), modèles linéaires qui englobent et généralisent les méthodes linéaires classiques : analyse factorielle, analyse en pistes causales et analyse de régression linéaire. On pose d'abord des hypothèses sur les relations factorielles entre variables et on regarde ensuite, après recueil des données, si elles sont infirmées ou confirmées par le modèle à l'aide de différents indices d'ajustement.

Ce papier propose la mise en œuvre de l'A.S.I. dans l'optique confirmatoire d'un modèle implicatif posé *a priori*, modèle renvoyant à une variable latente, dans notre cas à une variable latente socio-normative. Bien évidemment, y compris dans cette perspective confirmatoire, l'A.S.I. conserve son originalité, par rapport aux méthodes classiques, d'opérer sur des indices non symétriques. Dans le cas qui nous préoccupe, celui de l'autodescription personologique dans des situations contrastées, nous faisons l'hypothèse que les choix des descripteurs sont impliqués par la distance socio-normative à un profil prototypique élaboré par un groupe d'experts.

4 Critères de comparaison

A partir des mêmes variables A, B, C et D, sur deux groupes de sujets on obtient deux structures quasi-implicatives S_1 et S_2 . On postule que l'une des deux structures, S_1 , ait valeur de référence.

$$S_1 : B \Rightarrow A \Rightarrow C \Rightarrow D$$

En premier lieu, on peut s'intéresser à une comparaison globale entre les deux matrices qui croisent les mêmes variables et sous-tendent les structures S_1 et S_2 en calculant la distance entre ces deux matrices. Ces matrices sont carrées cette fois mais a priori non symétriques.

En second lieu, on pourra faire jouer d'autres critères de comparaison qui pourront s'appliquer soit à l'ensemble des matrices, soit à leurs parties correspondant à une séquence implicative issue de l'analyse, classe cohésitive ou chemin d'un graphe. On pourra introduire des critères de forme, de reclassement, de hauteur et de représentativité.

4.1 Calcul de la distance entre les deux matrices sous-jacentes à S_1 et S_2

Les matrices des indices d'intensité d'implication relatives à S_1 et à S_2 de valeurs normalisées selon la loi normale $N(0;1)$ sont a priori indépendantes si l'hypothèse nulle H_0 consiste à formuler qu'il n'y a pas interaction entre les deux familles d'observations. L'ensemble de cardinal m^2 des valeurs de leur différence $M_1 - M_2$ suit donc *a priori* une loi $N(0; \sqrt{2})^3$. Une partition de ces valeurs en k intervalles permet d'établir une comparaison entre les distributions observée et théorique, où l'intervalle réel $i =]u, v]$ contient m_i observations, alors que la distribution théorique sous H_0 en contiendrait $m^2 p_i$, et où p_i serait la

³ Bien entendu, il sera encore nécessaire de normaliser la différence des valeurs entre M_1 et M_2 afin de pouvoir utiliser la table de la loi normale centrée réduite $N(0; 1)$.

probabilité de l'intervalle i sous la loi $N(0; \sqrt{2})$. On est donc ramené à la situation générale d'ajustement de deux distributions à l'aide de celle du χ^2 à m^2-1 degrés de liberté.

Il suffit de former : $S = \sum_{i=1}^{i=k} \frac{(m_i - m^2 p_i)^2}{m^2 p_i}$ et de réfuter H_0 , au seuil α si $\Pr[\chi^2 > S] < \alpha$.

Si H_0 est réfutée, cela signifie que la différence entre l'analyse *a priori* et l'analyse *a posteriori*, est significativement grande, que les structures S_1 et S_2 sont différentes et, donc, qu'un phénomène discrimine les deux échantillons. Sinon, on considérera que les deux échantillons réagissent de façon comparable et, par suite, que les graphes ont une structure commune ou tout au moins voisine.

Remarque

On trouvera en annexe une autre modélisation quantitative de la distance ou significativité de la différence entre les deux matrices M_1 et M_2 . D'autres méthodes existent pour estimer les écarts entre matrices. Afin de ne pas alourdir le sujet, nous nous dispenserons de leur évocation (Saporta, 2006).

4.2 Effet de forme ou de séquence

La conservation de la forme séquentielle est la première condition pour décider d'un rapport d'homomorphie entre les structures. Ici, nul besoin d'un test statistique, une comparaison visuelle suffit. Par exemple $S_2 : A \Rightarrow D \Rightarrow B \Rightarrow C$ ne satisfait pas cette condition. On peut dans un second temps considérer les trois critères de comparaison suivants.

4.3 Effet de reclassement des intensités d'implication

S_1 présente trois indices d'intensité d'implication : $\Phi_{B,A}$ $\Phi_{A,C}$ $\Phi_{C,D}$.

Leurs valeurs les ordonnent de la manière suivante : $\Phi_{C,D} > \Phi_{B,A} > \Phi_{A,C}$.

Si dans S_2 le classement est identique, on valide le critère de manière absolue.

Si on observe des reclassements comme par exemple $\Phi_{B,A} > \Phi_{A,C} > \Phi_{C,D}$ alors on valide le critère de manière relative. On peut utiliser à cette fin le calcul d'une corrélation par rang (rho de Spearman par exemple). Si on ne montre pas de significativité statistique au coefficient de corrélation on pourra conclure à la non satisfaction du critère...

4.4 Effet de hauteur des intensités d'implication

Les trois indices d'intensité d'implication de S_1 se résument par leur moyenne et leur écart type. On peut calculer également moyenne et écart type de S_2 . La comparaison des deux moyennes peut se faire à l'aide du t de Student ou de tout autre technique de comparaison de moyenne adaptée aux données. La validation du critère passe par un test statistique susceptible de montrer une différence significative.

4.5 Effet de représentativité des répondants

Pour chaque séquence quasi-implicative, on dispose des contributions des sujets à la structuration des données en classes cohésitives ou en graphes implicatifs. On pourra pour S_1 résumer les indices de contributions par leur moyenne et leur écart type. Là encore, on peut calculer également moyenne et écart type de S_2 . La comparaison des deux moyennes peut se faire à l'aide du t de Student ou de tout autre technique de comparaison de moyenne adaptée aux données. La validation du critère passe par un test statistique susceptible de montrer une différence significative.

Décisions :

Si deux structures valident les cinq critères de manière stricte, alors on pourra conclure à leur isomorphie.

Si le critère de forme séquentielle n'est pas validé, on conclura à la divergence globale des deux structures, divergence dont l'ampleur peut être appréciée en prenant en compte les autres critères.

Enfin, si le critère de forme séquentielle est validée, mais pas tous les autres critères, on pourra conclure à l'homomorphie des deux structures associée à une formule xxx selon les critères secondaires validés : homomorphie à un critères ou homomorphie à deux critères...

5 Une application à la comparaison de deux structures quasi-implicatives

Le Q-ISAT est un questionnaire d'image de soi au travail. Il fournit différents indices psychométriques : distances socio-normatives dont l'intégrité professionnelle ajustée - IPA -, c'est-à-dire libérée du biais de désirabilité sociale, et profil de personnalité selon le modèle des cinq facteurs, à raison de deux scores par dimension, l'un pour son pôle positif et l'autre pour son pôle négatif.

Ainsi, la première dimension, la conscience se subdivise en un pôle positif D1P_conscience et un pôle négatif D1N_immaturité ; la seconde dimension, l'ouverture d'esprit en D2P_ouverture d'esprit et D2N_fermeture d'esprit ; la troisième dimension, la gentillesse en D3P_gentillesse et D3N_dureté ; la quatrième dimension, l'extraversion, en D4P_extraversion et D4N_introversion ; et enfin, la cinquième dimension, le névrosisme en D5P_stabilité émotionnelle et D5N_névrosisme.

Au cours d'un travail préalable mené à partir de données recueillies en consignes contextualisées opposées, soit " Vous êtes à votre travail et tout se passe bien pour vous ", soit " Vous êtes à votre travail et rien ne se passe bien pour vous ", on avait montré, par le biais d'une analyse implicative vectorielle que les profils des répondants n'étaient pas stables (Pasquier, Gras et Couturier, 2010).

Dans la présente étude, on étudiera de quelle manière les profils personnologiques varient en fonction de la contextualisation de la consigne. Cette variation sera déterminée par rapport à une variable repère, l'IPA. En d'autres termes, on se demande dans quelle mesure le passage d'une situation favorable à une situation de travail dégradée affecte l'image que des individus se font d'eux-mêmes au travail. Ceci nous amènera à comparer les deux structures implicatives obtenues dans chacune des conditions.

Les résultats sont établis sur la base d'un même échantillon qui a répondu au questionnaire successivement à deux reprises, en situation favorable puis en situation défavorable. Il comporte 91 répondants occasionnels. Les âges vont de 11 à 62 ans, avec une moyenne de 30 ans et un écart-type de 12,6 ans. On dénombre 68% de femmes et 32% d'hommes. Les niveaux de qualification se distribuent entre les six niveaux de la nomenclature interministérielle des niveaux de qualification, soit : 7% de niveaux 1, 46% de niveaux 2, 11% de niveaux 3, 13% de niveaux 4, 12% de niveaux 5 et 10% de niveaux 6.

Avant de passer à l'analyse statistique implicative, on décrira les caractéristiques des variables puis les corrélations entre IPA et les pôles des dimensions pour les deux structures.

6 Caractéristiques des variables et des structures

6.1 Caractéristiques des variables

Un premier tableau indique les caractéristiques des données brutes relatives aux deux structures, SF pour la situation favorable et SD pour la situation défavorable (tab. 2). Sur la seconde ligne, on constate que les scores moyens de distance socionormative (IPA) sont pratiquement stables d'une situation à l'autre.

TABLEAU 2 - *Caractéristiques des variables pour la situation favorable SF et la situation défavorable*

Variables	SF	SD	écart	t (p)	d	effet	Prep
IPA	0,17 (0,12)	0,14 (0,15)	0,03	1,49 (0,13 ; NS)	0,22 (-0,07 ; 0,51)	petit	0,85 (-0,19 ; 0,63)
D1P_conscience	4,92 (0,48)	4,25 (0,70)	-0,67	-7,53 (<0,01 ; TS)	-1,12 (-0,80 ; -1,42)	notable	1 (-0,68 ; -1,57)
D1N_immaturité	3,07 (0,50)	3,58 (0,83)	0,51	5,02 (<0,01 ; TS)	0,74 (1,04 ; 0,44)	intermédiaire	1 (0,32 ; 1,17)
D2P_ouverture	4,87 (0,33)	4,01 (0,68)	-0,86	-10,85 (<0,01 ; TS)	-1,60 (-1,27 ; -1,94)	notable	1 (-1,14 ; -2,09)
D2N_fermeture	3,27 (0,40)	3,61 (0,53)	0,34	4,88 (<0,01 ; TS)	0,72 (1,02 ; 0,42)	intermédiaire	1 (1,15 ; 0,30)
D3P_gentillesse	4,94 (0,45)	3,91 (0,82)	-1,03	-10,50 (<0,01 ; TS)	-1,55 (-1,22 ; -1,88)	notable	1 (-1,09 ; -2,07)
D3N_dureté	2,94 (0,43)	3,88 (0,76)	0,94	10,26 (<0,01 ; TS)	1,52 (1,85 ; 1,19)	notable	1 (1,06 ; 2)
D4P_extraversion	4,81 (0,45)	4,07 (0,67)	-0,74	-8,74 (<0,01 ; TS)	-1,29 (-0,97 ; -1,61)	notable	1 (0,85 ; 1,75)
D4N_introversion	3,26 (0,50)	4,20 (0,88)	0,94	8,86 (<0,01 ; TS)	1,31 (1,63 ; 0,99)	notable	1 (0,86 ; 1,62)
D5P_stabilité	4,93 (0,41)	3,96 (0,87)	-0,97	-9,62 (0 ; TS)	-1,42 (-1,09 ; -1,75)	notable	1 (0,97 ; 1,89)
D5N_névrosisme	3,12 (0,48)	4,59 (0,96)	1,47	13,06 (0 ; TS)	1,97 (2,29 ; 1,58)	notable	1 (1,60 ; 2,32)

SD. Dans chaque case, M et (ET). Pour le t de Student, (probabilité ; significativité). Pour le d de Cohen, (intervalles de confiance au seuil 0,95). Prep indique la probabilité de trouver un

effet de même signe dans une réplique, (intervalles de prédiction de l'effet dans la réplique à 95% de chances).

L'observation des écarts montre qu'avec le passage d'une situation à l'autre, pour chacune des dimensions, les scores des pôles positifs diminuent et inversement les scores des pôles négatifs augmentent. Cette évolution des écarts va jusqu'à inverser la préférence pour les pôles positifs pour les dimensions 4 et 5 : l'introversion l'emporte alors sur l'extraversion (4,20 vs 4,07) et le névrosisme prend le pas sur la stabilité émotionnelle (4,59 vs 3,96). Lorsque la situation est perçue comme défavorable, les gens ont tendance à se replier sur eux-mêmes et à s'angoisser.

6.2 Etude de la structure en situation favorable

Un second tableau indique les corrélations entre IPA et chacun des pôles des cinq dimensions pour la structure SF (tab. 3).

TABLEAU 3 - Situation favorable.

Pôles dimensions	R	p	limite 1(r)=	limite 2(r)=	d de Cohen	effet
D1P_conscience	0,47	<0,01	0,30	0,62	1,08	notable
D1N_immaturité	-0,61	<0,01	-0,72	-0,46	-1,54	notable
D2P_ouverture	-0,04	0,65	-0,25	-0,15	-0,04	négligeable
D2N_fermeture	-0,21	0,04	-0,40	-0,01	-0,44	petit
D3P_gentillesse	0,01	0,82	-0,19	0,22	0,03	négligeable
D3N_dureté	-0,46	<0,01	-0,61	-0,28	-1,05	notable
D4P_extraversion	0,15	0,14	-0,05	0,34	0,31	petit
D4N_introversion	<0,01	0,96	-0,19	0,21	0,01	négligeable
D5P_stabilité	0,24	0,01	0,03	0,42	0,50	intermédiaire
D5N_névrosisme	-0,12	0,21	-0,32	0,08	-0,24	petit

Corrélations entre IPA et chacune des variables candidates, probabilité, limites de l'intervalle de confiance au seuil 0,05 et d de Cohen suivi de la taille de l'effet pour SF.

Quand on prend comme critère la taille de l'effet, on observe que les deux pôles de la dimension conscience (conscience et immaturité) interviennent de manière notable dans l'explication statistique de l'IPA de même que le pôle négatif dureté. Un effet intermédiaire remplit également les autres critères de significativité : le pôle positif stabilité émotionnelle. Bien que son effet soit petit, le pôle négatif fermeture d'esprit remplit aussi les autres critères de significativité. Les deux autres petits effets ne remplissent pas les critères de probabilité ni de limite de confiance ; il s'agit du pôle positif extraversion et du pôle négatif névrosisme. Enfin, il reste trois pôles pour lesquels les effets sont négligeables et qui ne satisfont en rien les exigences fréquentistes : ouverture d'esprit, gentillesse et introversion.

Quand on considère les indices au moins égaux à 50 de la table des intensités implicatives de SF (tab. 4), on observe que les variables qui expliquent implicativement IPA sont D1P_conscience, D2P_ouverture d'esprit, D2N_fermeture d'esprit, D3P_gentillesse et D5P_stabilité.

On remarquera des différences entre le produit de l'ASI et le produit de l'étude des corrélations : on ne trouve que conscience, fermeture d'esprit et stabilité émotionnelle comme variables explicatives communes de l'IPA.

TABLEAU 4 - table des intensités d'implication pour SF.

	IPA	D1P	D1N	D2P	D2N	D3P	D3N	D4P	D4N	D5P	D5N
IPA	0	76	47	51	59	66	27	40	31	50	44
D1P	57	0	50	52	50	50	46	48	46	48	48
D1N	49	50	0	53	49	50	49	48	48	49	49
D2P	50	52	53	0	46	50	48	49	47	49	47
D2N	52	50	48	46	0	52	49	50	49	46	48
D3P	53	50	50	50	51	0	48	50	48	49	46
D3N	36	41	46	45	47	44	0	55	56	50	55
D4P	46	48	46	47	49	49	53	0	53	49	49
D4N	40	42	45	42	47	45	55	54	0	54	55
D5P	50	46	47	47	41	46	50	48	55	0	53
D5N	47	45	46	44	47	40	54	49	55	53	0

D1P : conscience ; D1N : immaturité ; D2P : ouverture d'esprit ; D2N : fermeture d'esprit ; D3P : gentillesse ; D3N : dureté ; D4P : extraversion ; D4N : introversion ; D5P : stabilité émotionnelle ; D5N: névrosisme.

Le graphe implicatif⁴ ordonne les variables selon une séquence quasi implicative (fig. 1).

On observe que c'est IPA, c'est-à-dire la distance socio-normative au prototype de l'intégrité professionnelle élaboré par les experts qui détermine les pôles des dimensions personologiques. La moyenne des contributions aux trois chemins est de 0,79 et l'écart type de 0,07.

⁴ On a choisi un seuil de .50 afin que l'ensemble de la structure apparaisse. Pour les implications de même valeur comme D5P on obtient une orientation sur la graphie du fait des décimales qui n'apparaissent pas sur le tableau issu de CHIC.

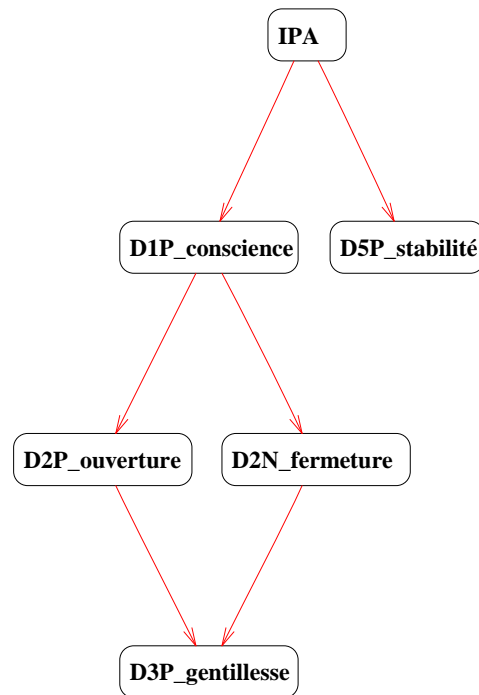


FIGURE 1 - Graphe implicatif en situation favorable.

6.3 Etude de la structure en situation défavorable

Le tableau 5 indique les corrélations entre IPA et chacun des pôles des cinq dimensions pour la structure SD.

TABLEAU 5 - Situation défavorable

Pôles dimensions	R	p	limite 1(r)=	limite 2(r)=	d de Cohen	effet
D1P_conscience	0,66	<0,01	0,53	0,76	1,79	notable
D1N_immaturité	-0,73	<0,01	-0,81	-0,61	-2,14	notable
D2P_ouverture	0,40	<0,01	0,21	0,56	0,88	notable
D2N_fermeture	-0,50	<0,01	-0,64	-0,32	-1,15	notable
D3P_gentillesse	0,31	<0,01	0,11	0,49	0,66	intermédiaire
D3N_dureté	-0,50	<0,01	-0,64	-0,33	-1,18	notable
D4P_extraversion	0,55	<0,01	0,39	0,68	1,33	notable
D4N_introversion	-0,37	<0,01	-0,54	-0,18	-0,81	notable
D5P_stabilité	0,44	<0,01	0,26	0,59	0,99	notable
D5N_névrosisme	-0,21	0,04	-0,40	<-0,01	-0,43	petit

Corrélations entre IPA et chacune des variables candidates, probabilité, limites de l'intervalle de confiance au seuil 0,05 et d de Cohen suivi de la taille de l'effet pour SD.

On remarque que toutes les variables psychologiques sont ici parties prenantes dans l'explication de l'IPA, quel que soit le critère de significativité considéré. Qu'en est-il d'un point de vue implicatif?

TABLEAU 6 - Table des intensités d'implication pour SD.

	IPA	D1P	D1N	D2P	D2N	D3P	D3N	D4P	D4N	D5P	D5N
IPA	0	86	11	77	21	72	20	79	25	79	34
D1P	70	0	31	68	36	66	32	64	32	65	36
D1N	21	25	0	24	71	26	75	32	69	23	71
D2P	63	66	32	0	33	70	29	63	33	68	30
D2N	36	37	64	33	0	37	65	39	59	34	62
D3P	62	67	30	74	34	0	24	61	32	73	26
D3N	31	30	72	25	68	24	0	37	67	27	70
D4P	67	65	34	66	37	61	38	0	31	63	36
D4N	35	29	66	29	62	31	67	30	0	34	64
D5P	68	67	27	73	30	73	26	64	34	0	23
D5N	42	37	65	30	62	30	67	38	61	27	0

D1P : conscience ; D1N : immaturité ; D2P : ouverture d'esprit ; D2N : fermeture d'esprit ; D3P : gentillesse ; D3N : dureté ; D4P : extraversion ; D4N : introversion ; D5P : stabilité émotionnelle ; D5N: névrosisme.

Les intensités d'implication supérieures à 0,50 s'appliquent aux cinq pôles positifs des dimensions personnologiques (tab. 6). La moyenne des contributions au chemin est de 0,77 et l'écart-type de 0,06.

Sur le graphe (Fig. 2) c'est une fois de plus l'IPA qui détermine les autres variables. En d'autres termes, dans l'approche corrélationnelle on a donné à la distance socionormative IPA un statut de variable expliquée par des dimensions personnologiques considérées comme variables explicatives, alors que l'analyse statistique implicative inverse les statuts, les dimensions du tableau personnologiques devenant implicativement expliquées par la distance socionormative. Les deux structures étant établies, on peut passer à leur comparaison.

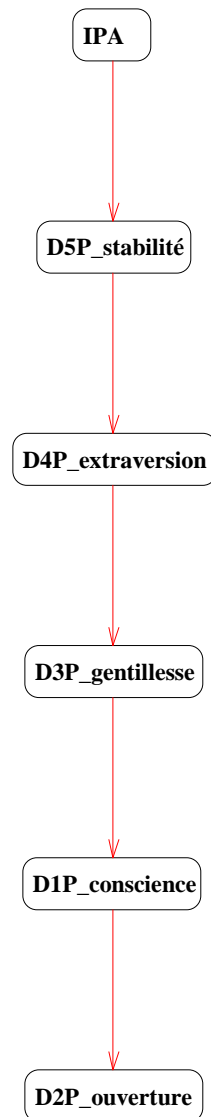


FIGURE 2 - *Graphe implicatif en situation défavorable.*

6.4 Comparaison des structures en situation favorable et en situation défavorable

Le calcul de la distance entre les deux matrices prend en compte l'ensemble des intensités d'implication des matrices alors que les comparaisons qui suivront ne prennent en compte que les intensités spécifiques aux structures quasi-implicatives étudiés, c'est-à-dire les liens entre IPA et les pôles des dimensions psychologiques.

6.4.1 Distance entre les deux matrices

Les différences normalisées entre les deux matrices ont été divisées en 10 intervalles. L'effectif théorique pour chaque intervalle est donc de 12,10 observations. Le tableau 7

indique les effectifs observés en regard des effectifs théoriques. Seuls les effectifs de l'intervalle 4 sont proches.

TABLEAU 7 - *Effectifs théoriques et observés*

effectifs	int. 1	int. 2	int. 3	int. 4	int. 5	int. 6	int. 7	int. 8	int. 9	int. 10
théoriques	12,10	12,10	12,10	12,10	12,10	12,10	12,10	12,10	12,10	12,10
observés	1	26	19	10	6	4	9	23	19	4

int. : intervalle.

La figure 3 montre la forme des écarts des effectifs observés par rapport aux effectifs théoriques, forme en M : il manque des observations aux intervalles des extrémités et du centre ; il y a trop d'observations pour les intervalles 2 et 3 , 8 et 9.

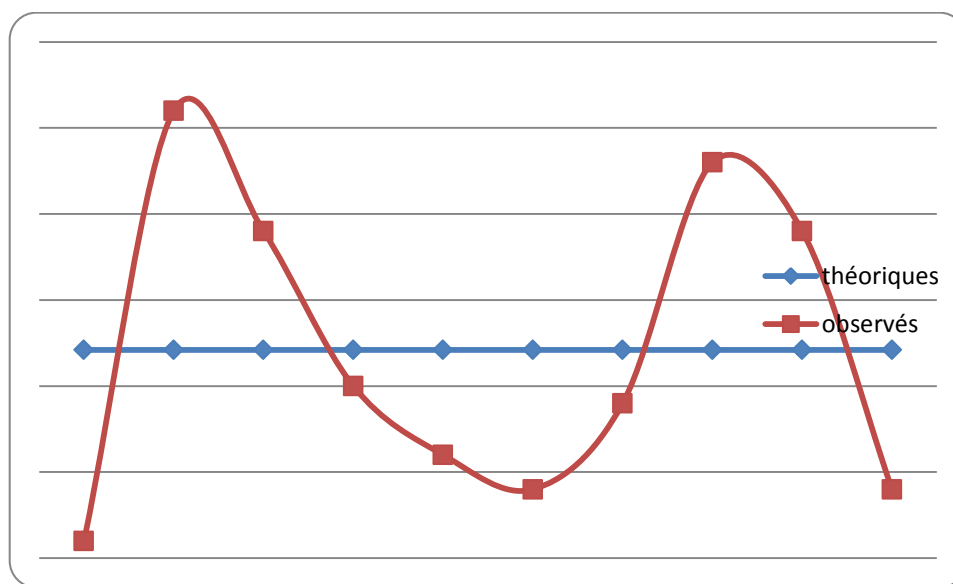


FIGURE 3 - *Courbe des écarts*

Le calcul du χ^2 donne la valeur 58.92 soit un rejet de l'hypothèse d'adéquation au risque $p < 0.005$. La loi sur la distribution de S_2 diffère donc fortement de la distribution sur S_1 , c'est-à-dire qu'on ne peut pas considérer que les deux matrices sont semblables.

6.4.2 Comparaison des chemins implicatifs

Quand on regarde les graphes au niveau le plus global, les deux structures semblent comparables en ce sens que c'est la distance socionormative qui implique les traits personnologiques et non l'inverse. En d'autres termes, la distance à la norme n'est pas réglée par le profil personologique alors que ce profil serait déterminé peu ou prou par cette distance. Cependant, au niveau des séquences implicatives des traits personnologiques on note une différence de structure entre les deux graphes : le premier présente un éclatement en chemins pouvant indiquer une certaine hétérogénéité entre profils, contrairement à ce que montre le 2ème graphe où un profil plus homogène semble sous-tendre les variables étudiées.

Enfin, quand on examine plus en détail les deux séquences implicatives, on observe des variations dans les pôles des dimensions. Les pôles conscience, ouverture, stabilité

émotionnelle et gentillesse sont mobilisés dans les deux cas, mais pas dans le même ordre. Le pôle extraversion n'apparaît qu'en situation défavorable et le pôle fermeture d'esprit qu'en situation favorable, parallèlement à l'ouverture d'esprit et en sens inverse si on regarde les deux corrélations. On remarquera que les autres pôles négatifs n'interviennent pas dans le schéma implicatif bien qu'ils entretiennent des corrélations élevées avec l'indice d'IPA.

6.4.3 Comparaison des classements des intensités d'implication

La corrélation entre les intensités d'implication relatives aux deux situations se situe à 0,52 à $p = 0,12$ entre les limites -0,16 et 0,86 de l'intervalle de confiance, soit un d de Cohen de 1,21 pour un effet notable bien que non significatif d'un point de vue fréquentiste.

De toutes les façons, 27% de variance expliquée invite à conclure à un effet de reclassement des indices d'intensités d'implication, effet lié au changement de consigne.

Dans le détail, l'observation des écarts entre les rangs (tab. 8) montre la stabilité relative des intensités des pôles D1P_conscience (1 ; 1), D2P_ouverture (4 ; 4), D3N_dureté (10 à 9) et D5N_névrosisme (7 à 6). D'autres intensités évoluent de manière modérée : D4N_introversion (9 à 7) et D3P_gentillesse (2 à 5). Enfin, trois indices se reclassent fortement; il s'agit de D1N_immaturité (6 à 10) qui perd 4 rangs, de D2N_fermeture d'esprit qui perd 5 rangs (3 à 8) et de D4P_extraversion qui gagne 5,5 rangs (8 à 2,5).

TABLEAU 8 - Reclassements des intensités

Pôle dimension	SF	SD	rang SF	rang SD	effet de reclassement
D1P_conscience	76	86	1	1	0
D1N_immaturité	47	11	6	10	-4
D2P_ouverture	51	77	4	4	0
D2N_fermeture	59	21	3	8	-5
D3P_gentillesse	66	72	2	5	-3
D3N_dureté	27	20	10	9	1
D4P_extraversion	40	79	8	2,5	5,5
D4N_introversion	31	25	9	7	2
D5P_stabilité	50	79	5	2,5	2,5
D5N_névrosisme	44	34	7	6	1

SF : situation favorable ; SD : situation défavorable.

6.4.4 Comparaison des hauteurs des intensités d'implication

On a calculé moyennes et écarts types des indices d'implication relatifs aux chemins des graphes en situation favorable et en situation défavorable (tab. 9).

TABLEAU 9 - Caractéristiques des intensités des indices des structures.

caractéristiques	SF	SD
M	49,10	50,40
ET	15,06	30,43

A l'écart de -1,30 correspond un t de Student de -0,12 qui renvoie un p de 0,90. Le d de Cohen se fixe à -0,05 (-0,93 ; 0,82). On ne montre pas de différence significative au niveau des moyennes. Par contre, la dispersion des intensités double quand on passe de la situation favorable à la situation défavorable. Le rapport des variances s'établit à 4,08 qu'on pourra

considérer comme significatif à la lecture de la table de F. Cela signifie que les distorsions des intensités d'implication sont plus fortes dans SD à la fois plus intenses, plus marquées. Le passage en situation défavorable induit une forme de désordre dans les logiques de choix personnologiques en ce sens que si ce choix reste globalement dérivé de la distance socio-normative, une plus grande diversité dans les implications se fait jour, diversité divergente entre des intensités d'implication qui se réduisent et celles qui s'amplifient. Il serait particulièrement intéressant d'appréhender ce processus de divergence en confrontant les écarts individuels à un réseau nomologique de variables susceptibles de lui donner sens.

6.4.5 Comparaison des représentativités des répondants

En SF, la moyenne des contributions aux trois chemins est de 0,79 et l'écart type de 0,07. En SD, la moyenne des contributions au chemin est de 0,77 et l'écart type de 0,06. A l'écart de -0,20 correspond un t de Student de -0,68 renvoie un p de 0,50. Le d de Cohen se fixe à -0,30 (-0,58 ; 1,18). On ne montre pas de différence significative au niveau des moyennes : la passation sous consigne défavorable n'affecte pas la représentativité des répondants.

6.5 Résumé des comparaisons

Prises dans leur globalité, on ne peut pas considérer les matrices comme semblables ; la passation en situation défavorable amène des distorsions au niveau des intensités d'implication. En ce qui concerne les liens entre IPA et les traits personnologiques, si pour les deux situations c'est bien la distance socio-normative qui implique les traits de personnalité, des variations apparaissent dans les traits impliqués et dans l'ordre donné par les graphes. D'une situation à l'autre, on observe des effets de reclassement des intensités, essentiellement à la baisse pour immaturité, fermeture d'esprit et à la hausse pour extraversion.

Les valeurs moyennes des intensités d'implication sont équivalentes alors que leur dispersion double d'importance quand on passe en situation défavorable ce qui renvoie à des comportements plus différenciés entre les individus dans cette dernière situation. On ne met pas en évidence d'écart de représentativité des répondants entre les deux situations.

On retiendra d'une part que les choix autodescriptifs dérivent de la distance socio-normative et d'autre part que les intensités des implications varient pour une large part en fonction des contextes évaluatifs. Le premier constat va dans le sens du primat du jeu des facteurs de la socialisation dans la genèse de l'image de soi *versus* le primat d'une « personnalité naturelle ». Le second constat renvoie une fois de plus à l'instabilité des descripteurs propre au modèle des cinq facteurs (Pasquier, 2007).

7 Conclusion

La nouvelle méthode statistique de comparaison de deux structures illustrées par des graphes d'implication, présentée dans le § 4.1 et appliquée dans le § 5.2, renforce les résultats fournis par des tests classiques. Elle a la vertu, à l'opposé de ces derniers, de prendre en compte une à une les différences positives ou négatives des intensités d'implication d'un moment à un autre. Appliquée ici à une situation en psychologie, elle présenterait un certain intérêt en économie où l'évolution de marchés serait à l'étude, en génétique où l'on s'intéresserait à l'évolution des espèces, etc.

Pour le psychologue, s'ouvre la possibilité de traiter des analyses confirmatoires sur un mode implicatif. Comme en analyse structurelle on établit le degré d'ajustement entre une structure à caractère théorique, ou au statut de modèle, posée *a priori* et une structure empirique établie à partir de données d'observation selon des indices globaux ou plus analytiques.

Ce mode de traitement implicatif ne s'oppose pas aux approches structurelles classiques ; il peut tout au contraire les compléter par des applications soit en parallèle dans une visée de validité concourante, soit en préalable afin d'ordonner un modèle théorique, soit après coup en utilisant les données issues de l'analyse structurelle pour des analyses implicatives...

8 Références

- [1] Bacher, F. (1987). Les Lisrel (I). *Travail Humain*, 50, 347-370. modèles structureaux en psychologie. Présentation d'un modèle:
- [2] Bacher, F. (1988). Les modèles structureaux en psychologie. Présentation d'un modèle: Lisrel (II). *Travail Humain*, 51, 273-288.
- [3] Gras, R., & Régnier, J.- C. (2009). Origine et développement de l'Analyse Statistique Implicative. In *Analyse Statistique Implicative, Une méthode d'analyse de données pour la recherche de causalités*, (Eds) R. Gras, J. C. Régnier et F. Guillet. Toulouse : Cépaduès, pp. 6-16.
- [4] Gras, R., & Kuntz, P. (2008). An overview of the Statistical Implicative. In *Statistical Implicative Analysis*, (Eds) R. Gras, E. Suzuki, F. Guillet et F. Spagnolo. Berlin-Heidelberg : Springer-Verlag.
- [5] Hotelling, H. (1933). Analysis of a complex of statistical variables into Principal Components. *Journal of Educationnal Psychology*, 24, 417-441.
- [6] Pasquier, D. (2007). *Évaluation de la personnalité à l'aide des questionnaires auto-descriptifs : 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 : Atelier national de reproduction des thèses.
- [7] Pasquier, D., Gras, R., & Couturier, R. (2010). L'analyse statistique implicative vectorielle ; une application au domaine de la description personologique en passation répétée. (Eds) J. C. Régnier, F. Spagnolo, B. Di Paola et R. Gras, *Actes du Colloque ASI 5*. Palerme : Université, pp.351-353.
- [8] Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophy Magazine*, 2, 559-572
- [9] Saporta G. (2006), *Probabilités, Analyse de Données et statistique*, Paris : Ed. Technip,
- [10] Spagnolo, F., Gras, R., & Régnier, J. C. (2009). Mesurer l'écart entre une analyse a priori et la contingence en didactique, In *Analyse Statistique Implicative, Une méthode d'analyse de données pour la recherche de causalités*, (Eds) R. Gras, J. C. Régnier et F. Guillet. Toulouse : Cépaduès, pp. 165-174.

Annexe : espace vectoriel normé de matrices d'implication

On cherche à définir une topologie sur les matrices carrées M de dimensions $m \times m$ dont les colonnes sont les vecteurs $(\vec{V}_1, \vec{V}_2, \dots, \vec{V}_m)$. Pour tout $i=1, 2, \dots, m$, \vec{V}_i a pour composantes les m valeurs $V_j(i) = \psi(a_j \Rightarrow a_i)$. Ce sont donc les valeurs des intensités d'implication-inclusion de chacune des variables sur la variable relative au vecteur colonne. L'ensemble de telles matrices peut être muni d'une structure d'espace vectoriel bien que les opérations somme et produit par un scalaire n'aient pas une résonance sémantique dans le domaine de l'implication. Les valeurs numériques, en effet, ont une nature comparable aux probabilités (elles le sont si ce sont des intensités d'implication ordinaires). Cependant, cet espace vectoriel peut être normé de la façon suivante à partir des normes de ses vecteurs-colonnes soit

$$\|\vec{V}_i\| = \left[\sum_{j=1}^{j=m} V_j^2(i) \right]^{\frac{1}{2}}, \text{ norme euclidienne.}$$

et la norme que nous choisissons pour M est alors : $\|M\| = \sup_i \|\vec{V}_i\|$

La relation entre les deux graphes implicatifs G_1 et G_2 peut alors être étudiée à partir de l'examen de la distance $\|M_1 - M_2\|$. Cette distance permet d'étudier également les positions respectives d'autres matrices par rapport à M_1 . En particulier, si l'on recueille une suite temporelle de données et qu'il lui correspond une suite de matrices respectives M_0, M_1, \dots, M_k , il est possible de mettre en évidence la proximité de l'une des séquences par rapport à l'instant initial 0.